Distribution: Restricted

Delivery report

Development and validation of a global GPP/NPP model using MERIS and Sentinel-3 data (TerrA-P)

ATBD v1

Iain Colin Prentice, Rebecca Thomas

Study accomplished under the authority of ESA

May 2017

DISTRIBUTION LIST

Philippe Goryl, ESA-ESRIN

Iain Colin Prentice, ICL
Rebecca Thomas, ICL

Else Swinnen, VITO
Roel Van Hoolst, VITO
Bruno Smets, VITO

Ivan janssens, UA
Sara Vicca, UA
Manuela Balzarolo, UA

SUMMARY

The TerrA-P project will implement and validate a new global monitoring system for primary production by land ecosystems, comprising 10-daily composites of gross primary production (GPP) and annual composites of above-ground biomass production (ABP) by C_3 and C_4 plants. Spectral reflectance data will eventually be provided by Sentinel-3. These data will be used to drive a recently developed universal, first-principles light use efficiency model (the 'P' model) for GPP.

Initial model calibration was carried out using GPP derived from a selection of 17 eddy-covariance carbon dioxide (CO_2) flux measurement sites. These sites were selected for their relatively homogeneous surrounding vegetation and long measurement records. The calibration used a merged input data set based on spectral reflectances from SeaWiFS and MERIS for the fraction of incident photosynthetically active radiation absorbed by green plant tissues (fAPAR). The required meteorological data for calibration were derived from direct measurements at the flux sites.

This document is version 1 of the Algorithm Theoretical Basis Document (ATBD) for the proposed TerrA-P products. It describes the design criteria adopted for the new monitoring system based on a user survey and also on a set of scientific requirements, which go significantly beyond the current state of the art. The document briefly reviews the history of light use efficiency (LUE) models, and summarizes the strengths and weaknesses of various existing LUE models including those used operationally. The principle and derivation of the P model are described. It is shown how this model – which has the mathematical form of a LUE model – can nonetheless be derived from the standard model of photosynthesis. The P model has been shown to achieve comparable accuracy (in comparison to GPP derived from eddy-covariance carbon dioxide flux measurements) to other models, while requiring fewer parameters to be estimated. Optimal agreement of modelled and observed GPP in the calibration data set was obtained after minimal adjustment of the value used for the intrinsic quantum efficiency of C_3 photosynthesis.

The monitoring system will be further developed to make use of remotely sensed environmental inputs as far as possible to replace the interpolated meteorological inputs that are currently used in operational LUE models. The system will include data quality flags, and a specification of per-pixel uncertainties based on the established principles of uncertainty measurement and propagation. A preliminary description of the proposed method to assign per-pixel uncertainties is provided. The system will be validated using an extensive data set of GPP data derived from eddy-covariance flux measurements in different biomes and climatic regions, and the most comprehensive available global set of quality-controlled data on annual ABP. These developments will form the basis for version 2 of the ATBD.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

3-PG	Physiological Principles Predicting Growth
ABP	Above-ground biomass production
ATBD	Algorithm Theoretical Basis Document
AVHRR	Advanced Very High Resolution Radiometer
BESS	Breathing Earth System Simulator
BIOME-BGC	Biome–BioGeochemical Cycles model
BPLUT	Biome Parameter Look-Up Table
CASA	Carnegie-Ames-Stanford Approach model
C-Fix	Carbon Fixation model
CGLOPS-1	Copernicus Global Land Operations Lot1
CRO	Cropland
CUE	Carbon use efficiency
DBF	Deciduous broadleaf forest
DMP	Dry Matter Productivity algorithm
EBF	Evergreen broadleaf forest
EC-LUE	Eddy Covariance-Light Use Efficiency model
ECMWF	European Centre for Medium-range Weather Forecasting
ENF	Evergreen needleleaf forest
ET	Evapotranspiration
FACE	Free Air Carbon dioxide Enrichment
fAPAR, FPAR	Fractional absorbed photosynthetically active radiation
FvCB	Farquhar, von Caemmerer and Berry model
GLO-PEM	Global Production Efficiency Model
GMAO	Global Modeling and Assimilation Office
GPP	Gross primary production
GVI	Global vegetation index
IPAR	Incident photosynthetically active radiation
LAI	Leaf area index
LPJ	Lund-Potsdam-Jena model
LPX	Land-surface Processes and eXchanges model
LUE	Light use efficiency
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
MTCI	MERIS Total Chlorophyll Index
NASA	National Aeronautics and Space Administration
NDVI	Normalized difference vegetation index
NOAA	National Oceanic and Atmospheric Administration
NPP	Net primary production
OSH	Open shrubland
P	Production model
PPFD	Photosynthetic photon flux density
RMSE	Root mean squared error of prediction
SDBM	Simple Diagnostic Biosphere Model
SeaWiFS	Sea-Viewing Wide Field-of-View Sensor
TerrA-P	Development and validation of a global GPP/NPP model using MERIS and Sentinel-3 data
VOC	Volatile organic compound
vpd	vapour pressure deficit

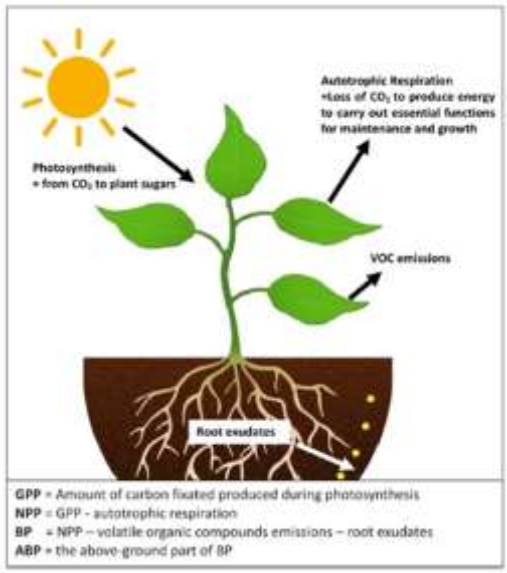
LIST OF SYMBOLS

C_3	photosynthetic pathway whereby carbon is first incorporated into three-carbon sugars
C_4	photosynthetic pathway whereby carbon is first incorporated into four-carbon sugars
c_a	ambient partial pressure of carbon dioxide
c_i	leaf-internal partial pressure of carbon dioxide
C:N	ratio of carbon to nitrogen content
CO_2	carbon dioxide
c^*	unit cost of maintaining electron transport capacity (P model)
D	vapour pressure deficit
e_a	absolute vapour pressure of water
e_s	saturation vapour pressure of water
f_{BG}	fraction of carbon allocated below ground
g_1	stomatal sensitivity parameter (Medlyn et al., 2011 model)
J_{max}	maximum rate of electron transport (FvCB model)
m	intermediate term in the P model
K	effective Michaelis-Menten coefficient of Rubisco
K_C	Michaelis-Menten coefficient of Rubisco for carboxylation
K_O	Michaelis-Menten coefficient of Rubisco for oxygenation
O	partial pressure of oxygen
pH	measure of acidity: minus log (hydrogen ion concentration)
Q_{10}	ratio of reaction rate at temperature T + 10 (˚C or K) to reaction rate at temperature T
R	universal gas constant
R^2	coefficient of determination
T	absolute temperature (K)
T_C	Celsius temperature (˚C)
u^2	standard uncertainty
V_{cmax}	maximum velocity of carboxylation (FvCB model)
β	ratio of cost factors for carboxylation and transpiration (P model)
Γ*	CO_2 compensation point in the absence of mitochondrial respiration (FvCB model)
$\delta^{13}C$	normalized ratio of the stable carbon isotopes ^{13}C and ^{12}C relative to a standard reference
$\Delta H_{\Gamma*}$	activation energy of Γ*
ΔH_{K_C}	activation energy of K_C
ΔH_{K_O}	activation energy of K_O
η*	viscosity of water relative to its value at 25˚C
Θ	curvature parameter in the relation between electron transport and absorbed PPFD (FvCB model)
ξ	stomatal sensitivity parameter (P model)
$\phi_0(C_3)$	intrinsic quantum efficiency of C_3 photosynthesis
$\phi_0(C_4)$	intrinsic quantum efficiency of C_4 photosynthesis
χ	ratio of leaf-internal to ambient carbon dioxide
$\partial f(x_i)/\partial x_i$	sensitivity of GPP to variable x_i

CHAPTER 1 BACKGROUND OF THE DOCUMENT

1.1. SCOPE AND OBJECTIVES

The TerrA-P project aims to implement and validate a new global monitoring system for primary production by land ecosystems. The project focuses on gross primary production (GPP) and above-ground biomass production (ABP), defined as follows:



Gross primary production is the rate of total carbon fixation (photosynthesis) by the ecosystem. This is the most fundamental measure of primary production, as all other ecosystem functions depend on it. Also, thanks to the availability of eddy-covariance flux measurements, GPP data are available – at time scales from half-hourly up to multi-annual – for some hundreds of locations worldwide (albeit with a bias towards temperate regions), and for croplands as well as for natural and managed ecosystems.

Figure 1 (left): schematic illustrating the different aspects of primary production.

Above-ground biomass production is the rate of production of plant matter, excluding roots. This is a practically important measure, because this is the production rate of forage for grazing animals; it is closely related to the production rate of timber for harvest; and it can be converted (through data on the harvest index – the ratio of harvestable yield to ABP – for different crops) to estimates of crop yield. There are data on ABP, occasionally at a monthly time scale but more commonly at the annual time scale, for many ecosystems, especially crops and managed forests but also for natural ecosystems.

We chose not to focus either on (total) biomass production or on net primary production (NPP), for the following reasons:

Biomass production is the total rate of production of plant matter, including roots. For most crops the root production is of less interest than the above-ground production. Even for root crops there are data on the harvest index, so ABP can be used to predict harvestable yield by a simple conversion. There are some data on BP but in most cases the root production has not been measured directly, but rather inferred from above-ground measurements. Inevitably this increases the uncertainty of BP data.

Net primary production is defined as the difference between GPP and plant respiration. Formerly, NPP was assumed to equivalent to BP. Most data that claim to be NPP are, in fact, BP. But it is now

understood that a fraction of NPP – under some circumstances this can be as much as 20% – is "lost" from the plant by other pathways than respiration, in the form of volatile organic compounds (such as isoprene and monoterpenes) and/or root exudates, which do not contribute to biomass production. Moreover, published data sets of "NPP" are generally of poor quality.

The aims of the project will be achieved by using spectral reflectance data from Sentinel-3 as input to a recently developed universal, first-principles light use efficiency (LUE) model for GPP, called the 'P' model (for 'production'). Initial calibration has been carried out against GPP data derived from a selection of site-based eddy-covariance carbon dioxide (CO_2) flux measurements. This calibration used a merged data set from SeaWiFS and MERIS (Global Vegetation Index, GVI) to provide the fraction of incident photosynthetically active radiation that is absorbed by green tissues (fAPAR), a key input to the P model. The model also requires meteorological data, which for calibration purposes have been derived from direct measurements at the flux sites. The model thus calibrated will provide the initial basis for a global system in which meteorological data, at least initially, are obtained by interpolation from a coarse-resolution grid.

The project will create a monitoring system for 10-daily GPP and annual ABP by C_3 and C_4 plants. The system will eventually make use of remotely sensed environmental inputs as far as possible. It will include data quality flags and – an innovative feature – a specification of per-pixel uncertainties. It will be validated using an extensive data set of GPP derived from eddy-covariance flux measurements in different biomes and climatic regions, and the most comprehensive available global set of quality-controlled data on annual ABP.

This is version 1 of the Algorithm Theoretical Basis Document (ATBD) for TerrA-P. Version 2 will be produced following developments during the course of the project.

1.2. CONTENT OF THE DOCUMENT

The document is organized in the following way.

- **Chapter 1** describes the background of the document.
- **Chapter 2** describes the criteria adopted for new primary production data products, taking into account both the requirements articulated by users, and scientific considerations.
- **Chapter 3** is a selective review of existing approaches to monitoring primary production from space, including those currently used operationally.
- **Chapter 4** describes the P model and the proposed manner of its implementation, including proposed approaches to calculate uncertainties; introduces the initial GPP calibration data set; and presents the results of the calibration.

CHAPTER 2 CRITERIA FOR NEW PRIMARY PRODUCTION DATA PRODUCTS

2.1. INTERPRETATION OF USER REQUIREMENTS

TerrA-P has conducted a survey targeting various potential users of the proposed new products, including current users of the Copernicus Global Land Operations Lot1 (CGLOPS-1) of the Copernicus Global Land Service. The survey yielded the following key information for product design.

The *principal products currently in use* are MODIS GPP and net primary production (NPP), CGLOPS-1 products, and models. The new products should thus aim to reproduce (at least) the functionality of these existing, widely-used products.

Different users work at different *geographic scales*, from subnational to global. The new products should accordingly be global, gridded products allowing flexibility of application.

About three-quarters of users surveyed agreed with the proposed strategy to *focus on GPP* and *ABP*. Some caveats were mentioned, including the fact that GPP cannot be directly derived from flux measurements (ecosystem respiration has to be factored out through a 'partitioning' method, of which there are several that give somewhat different results); and the fact that total biomass production (including production below ground) may sometimes be of greater interest than ABP. We propose to deal with uncertainty in partitioning by using the full range of alternative partitioning methods as a measure of uncertainty in observed GPP. For biomass production, however, the extreme paucity and low reliability of data on below-ground production argues for maintaining a focus on ABP. We note that ABP is quantitatively related to below-ground production by, for example, root crops just as it is quantitatively related to above-ground production by grain crops.

Users were approximately equally divided in their preferences for *units of dry matter versus carbon*. For maximum comparability with existing products, and with the main data sources for each quantity, we propose supplying GPP in carbon units and ABP in dry matter units. Climate modellers preferred carbon units, but are likely to be more interested in 10-daily GPP than annual ABP.

Most users saw the need to consider *C_3 versus C_4 photosynthesis* but, not surprisingly, there was no specific proposal as to how the prevalence of the two pathways could be specified on a per-pixel basis. We propose to circumvent this problem by providing both as alternatives for every pixel.

Most users asked for a *data quality layer*, and *information on uncertainty*. A per-pixel uncertainty layer was not explicitly requested. However, a numbers of users in different ways indicated a need for quantitative, per-pixel uncertainty information. In our view a systematic approach to per-pixel uncertainty should be a significant part of product development, and would satisfy this need.

The most popular *sampling frequency* was 10-daily. A number of users voted for daily, but daily data on spectral reflectances are not meaningful because many dates, in most locations, will be

affected by clouds. This problem can be largely circumvented by providing 10-daily composites. Most users asked for *data availability* in near-real time, that is, 3 to 5 days after acquisition.

The preferred *spatial resolution* is 300 m. Some users work with coarser resolutions which, however, can be readily obtained by post-processing of a 300 m product.

The average request in terms of *relative accuracy* was close to 20%, which likely provides a realistic target accuracy for both 10-daily GPP and annual ABP.

Table 1 summarizes the user requirements, and the specifications adopted that are consistent with these requirements as far as is technically feasible.

Table 1: Summary of user requirements and specifications adopted for TerrA-P products.

User requirement	**Specification adopted**
Geographic scale subnational to global	Global gridded product
Focus on GPP and ABP	Focus on GBP and ABP
Carbon or dry-matter units	Carbon for GPP, dry matter for ABP
Distinction of C_3 and C_4 photosynthesis	Provide results for both C_3 and C_4 plants
Data quality layer, information on uncertainty	Provide per-pixel uncertainty estimates
Daily to 10-daily sampling frequency	10-daily sampling frequency
Data availability in near-real time	Data available 3-5 days after acquisition
Spatial resolution 300 m or coarser	300 m grid with facility for post-processing
Relative accuracy ca 20%	Target relative accuracy 20%

2.2. SCIENTIFIC REQUIREMENTS

A priori we determined that new products should as far as possible possible satisfy a number of additional criteria, summarized here. These requirements go significantly beyond the current state of the art in LUE-based modelling.

Explicit relationship to the standard model of photosynthesis. The Farquhar, von Caemmerer and Berry (1980) (FvCB) model is the standard model of C_3 photosynthesis. Modifications exist to describe C_4 photosynthesis. There are thousands of published field measurements of the parameters defined in the FcVB model. All current ecophysiological theory, and the great majority of biophysical land-surface schemes for climate modelling, make use of the FvCB model. Therefore, a newly developed remotely sensed GPP product should be explicitly defined in terms of the FvCB model.

There is no such general model for plant respiration and other carbon "losses" from GPP. Thus models for biomass production should be based on GPP, with modifications to account for these losses as fractions of GPP.

Representation of physiological effects of CO_2. Models based on remotely sensed data, including thiose in operational use, generally do not consider the effect of changing atmospheric CO_2 concentration on the LUE of photosynthesis. Thus, they only consider a CO_2 effect in so far as it is manifested by changes in foliage cover that can be seen from space. As a direct consequence, products such as MODIS GPP and NPP severely underestimate the generally increasing trend in

primary production due to rising CO_2 (De Kauwe et al., 2016a). Newly developed products should explicitly include the effect of CO_2 concentration on LUE.

No discontinuities at biome boundaries. Although the convention of defining different model parameter values per biome is widely entrenched in remote sensing applications, it inevitably leads to discontinuities at boundaries defined by an external classification. This is a highly undesirable property, because biomes intergrade. Imposed biome boundaries are arbitrary and differently located according to different land cover products. New products should attempt to avoid such discontinuities.

A demonstrated level of accuracy assessed by comparison to relevant measurements. Eddy-covariance measurements of CO_2 flux can be processed ('partitioned') to yield estimates of daily, 10-daily, monthly or annual GPP. Flux measurement sites vary in public availability status, and in the length of records. Some sites are more suitable than others for model calibration and validation, because in areas of complex terrain or land use patterns there can be a severe problem in attempting to match remotely sensed spectral reflectance data with the (time-varying) footprint of the flux tower. Thus, model calibration and validation should be based on an informed selection of flux measurement sites.

CHAPTER 3 REVIEW OF SELECTED EXISTING APPROACHES

3.1. BACKGROUND AND HISTORY

The basis for nearly all algorithms designed to calculate primary production based on remotely sensed data is the general LUE model first proposed by Monteith (1972, 1977). Monteith based this model on field measurements of crop growth in both tropical and temperate climates, showing that growth is proportional to the time-integral of the light absorbed by the crop.

The general LUE model can be applied either to GPP or to NPP. More formally, in a remote sensing context, the general LUE model states that primary production is proportional to the product of incident photosynthetic photon flux density (PPFD) and fractional green vegetation cover, also called fractional absorbed photosynthetically active radiation (fAPAR or FPAR). fAPAR depends on Leaf Area Index (LAI) but is closer to actual reflectance measurements than LAI, and more directly related to primary production. In the remote sensing literature, incident PPFD (μmol m^{-2} s^{-1}) is more often described as 'incident photosynthetically active radiation' (IPAR) (W m^{-2}). The former term is more accurate because photosynthesis depends on the *number* of photons absorbed, rather than their energy (which varies with their wavelength). However, the two units can be interconverted, if it is assumed that the solar spectrum is constant.

NPP is the remainder of GPP after autotrophic (plant) respiration has converted approximately half of GPP back to CO_2. Traditionally, NPP has been regarded as synonymous with biomass production, i.e. the production of plant tissues. However, it is now recognized that a fraction (which can be as much as 20%) of NPP is lost from plants in the form of exudation from roots (a carbon subsidy to microbes in the rhizosphere, which enables plants to increase their uptake of soil nutrients) and emissions of volatile organic compounds (VOCs) such as isoprene and monoterpenes from leaves (which confer protection against both oxidants, including ozone, and high leaf temperatures). We therefore make a distinction between NPP and biomass production. The latter is of greater interest than NPP *sensu stricto* to users in forestry and agriculture.

Pioneering examples of remotely sensed primary production models are the Simple Diagnostic Biosphere Model (SDBM) of Knorr and Heimann (1995), the Carnegie-Ames-Stanford Approach (CASA) model of Potter et al. (1993), and the Global Production Efficiency Model (GLO-PEM) of Prince and Goward (1995). These models used spectral reflectance data from the Advanced Very High Resolution Radiometer (AVHRR) to infer fAPAR. A constant maximum LUE was specified, then reduced by scalars representing aspects of temperature and moisture conditions that can reduce LUE. The SDBM was combined with an atmospheric tracer transport model and deployed in an inverse mode, using observations of the seasonal cycle of atmospheric CO_2 concentration at different latitudes to estimate a single global maximum value for the LUE of NPP, and a single global Q_{10} value to quantify the dependence of soil organic matter decomposition on temperature. In GLO-PEM, theoretical maximum LUE values for GPP were determined based on the FvCB model. One value was assigned for C_3 plants and another for C_4 plants. These values were modified following the FvCB model's estimation of photorespiratory carbon loss as a function of temperature. GLO-PEM also made use of a number of other remote-sensing approaches to estimate meteorological variables, including IPAR. Unusually, this modelling approach – further

developed by Goetz et al. (1999) – aimed to derive all required meteorological and biophysical variables from remotely sensed platforms.

The Physiological Principles Predicting Growth (3-PG) model of Landsberg and Waring (1997) is also an LUE model and can be driven by remotely sensed and meteorological data. This model, focused on forest production, makes a number of empirically well-supported simplifications to predict GPP, NPP and forest growth.

All of the models discussed so far included an estimate of soil moisture availability as one of the factors reducing LUE.

In the following subsections, we review the principal literature on LUE models and note the key features of various widely used models, including those used operationally, and some other models that have introduced potentially useful innovations.

3.2. THE MODIS GPP AND NPP PRODUCTS

By far the most widely used remotely sensed primary production data products for scientific applications today are the MODIS GPP and NPP (MOD17) products (Running et al., 2004; Zhao et al., 2005). The most recent (2015) user's guide to the MOD17 products can be found at:
http://www.ntsg.umt.edu/sites/ntsg.umt.edu/files/modis/MOD17UsersGuide2015_v3.pdf
MODIS GPP specifies a maximum LUE which varies per biome. This value is multiplied by two scalars formulated as ramp functions: a linearly declining function of daily vapour pressure deficit (vpd), between biome-specific limit values where the function equals one or zero, and a linearly increasing function of daily minimum temperature, between biome-specific limit values where the function equals zero or one. Soil moisture effects are not considered: thus implicitly, soil moisture effects are considered to be accounted for in the remotely sensed fAPAR, and/or the effect of vpd on photosynthesis. The calculations make use of a land-cover classification and a Biome Parameter Look-Up Table (BPLUT) which recognizes 10 biomes. The required meteorological data, including IPAR (incoming solar shortwave radiation multiplied by 0.45), are supplied by the Global Modeling and Assimilation Office (GMAO) of the US National Aeronautics and Space Administration (NASA) at a 1˚ x 1.25˚ grid resolution. These data are smoothly interpolated to the finer (1 km) spatial resolution of the remotely sensed fAPAR data. The spatial and temporal resolution (8 days) of the product are set by the fAPAR data, which are obtained from the MODIS FPAR/LAI product.

Biome-specific model parameter values for MODIS GPP were based on literature-derived estimates used in the process-based BIOME-Biogeochemical Cycles model (BIOME-BGC: Running and Hunt, 1993). Thus, CO_2 flux data were not used in the initial calibration of the remote-sensing based model. However, the MODIS GPP product has been very extensively and independently evaluated by comparison with GPP derived from flux measurements. A number of these evaluation studies are cited in the user guide. Verma et al. (2014) included MODIS GPP – and also a version called MOD17-Tower, which was pre-calibrated against flux measurements – in a systematic global comparison with flux data-derived GPP. MODIS GPP was included in the set of seven LUE models compared globally with flux data-derived GPP by Yuan et al. (2014). Tang et al. (2015) undertook a comprehensive evaluation of MODIS GPP against flux data-derived GPP for forest ecosystems. Tagesson et al. (2017) showed that MODIS GPP greatly underestimates flux data-derived GPP in the Sahel, apparently due to unrealistically low maximum LUE assigned to semi-arid ecosystems.

The approach taken in the MODIS NPP product to derive NPP from GPP relies on separately modelling autotrophic respiration, which is separated into maintenance and growth components. NPP is then the difference between modelled GPP and modelled total autotrophic respiration.

Maintenance respiration in the model depends on leaf area index (LAI), obtained from the MODIS FPAR/LAI product, and biome-specific values of specific leaf area, the ratio of fine root mass to leaf mass, base maintenance respiration rates for leaves and fine roots, and a Q_{10} value that specifies the steepness of an assumed exponential relationship to temperature. In the current version of MODIS NPP, the Q_{10} of leaf maintenance respiration is a decreasing function of growth temperature. This is said to reflect respiratory acclimation, although Medlyn (2011) pointed out that the use of this function in fact further steepens the modelled negative response of NPP to temperature. Other components of maintenance respiration assume a fixed Q_{10}. Additional biome-specific information is used to infer annual growth respiration, which is then combined with annually integrated estimates of GPP and autotrophic respiration to yield annual NPP. Evaluation of NPP is considerably more difficult than evaluation of GPP, as the available data are far more limited. A number of compilations of annual NPP measurements have been made, however, and these have been used as a benchmark for MODIS NPP.

The MODIS GPP and NPP products were the first global and widely disseminated products of their kind. They represented a major technical advance, and now underpin a large number of high-profile scientific publications. However, criticisms have been made of some recent studies in which key assumptions underlying these products were overlooked. The study by Zhao and Running (2010) for example was criticized by Samanta et al. (2011) and Medlyn (2011), as the decline in NPP during the 2000 to 2009 period reported by Zhao and Running (2010) was not present in the remotely sensed data. It was, instead, a consequence of the model's high sensitivity of maintenance respiration to temperature. Prentice (2013) noted that this sensitivity must be too high, because the interannual variability of MODIS NPP is so large as to fully account for the observed year-to-year variability of the atmospheric CO_2 growth rate – allowing no room for the effect of temperature variability on the soil decomposition rate, which is generally understood to be the key factor modulating the CO_2 growth rate (e.g. Wenzel et al., 2014).

Another emerging problem is the weak increase over time shown in the MODIS GPP and NPP products, which contrasts with the strong increase shown by most process-based models (Anav et al., 2015) and with evidence for increasing LUE as a principal driver of the high-latitude seasonal cycle of atmospheric CO_2 (Graven et al., 2013; Thomas et al., 2017). In process-based models, GPP and NPP increase primarily as a consequence of the rising atmospheric CO_2 concentration. Smith et al. (2016) asserted that process-based models overestimate the stimulatory effect of rising CO_2 concentrations on NPP. They reached this conclusion by comparing process-based model outputs with the much weaker trend shown by MODIS NPP. However, MODIS GPP and NPP do not allow CO_2 concentration to influence LUE; therefore, the only possible CO_2 influence in these products is via increasing fAPAR. There is a worldwide 'greening' trend which has been attributed in part to the effect of CO_2 in increasing NPP and water use efficiency (Donohue et al., 2013; Ukkola et al., 2015; Zhu et al., 2016). But Free Air Carbon Enrichment (FACE) experiments have shown that the principal effect of enhanced CO_2 on primary production in forests is through increased LUE whereas increased LAI or fAPAR are much less important. Thus, the discrepancy noted by Smith et al. (2016) does not mean that process-based models overestimate the effect of CO_2 on NPP. Instead it means that MODIS NPP underestimates this effect (De Kauwe et al., 2016a), as a consequence of its design.

3.3. C-FIX AND THE DRY MATTER PRODUCTIVITY PRODUCT

Veroustraete et al. (1994, 2002) introduced the C-Fix model, a pioneer effort and the forerunner of the present-day Dry Matter Productivity (DMP) algorithm (Swinnen et al., 2015) currently provided by CGLOPS1. The original version of C-Fix was a LUE model for GPP in which a maximum LUE (a

single global value) was reduced through multiplication by scalars representing the effects of suboptimal temperature and water availability. Values for NPP were obtained from GPP using an empirical linear function of temperature to account for the fraction of GPP lost to autotrophic respiration.

C-Fix also included an attempt to account for an effect of rising atmospheric CO_2 concentration on photosynthesis via the FvCB model. However, the effect was implemented by way of the CO_2-dependence of the Rubisco-limited rate of photosynthesis, which is the rate measured at saturating light intensity (De Kauwe et al., 2016b). This rate depends steeply on CO_2. But it is not the rate actually realized in the field because the light- and Rubisco-limited rates are approximately balanced for typical daytime field conditions (Maire et al., 2012). The light-limited rate depends on CO_2 as well, but less steeply than the Rubisco-limited rate. The original CO_2 response function in C-Fix therefore presumably overestimates the effect of CO_2 on GPP. However this overestimation may be tempered by the fact that no distinction is made between leaf-internal and ambient CO_2. The steepness of the response is thereby less than it would have been if leaf-internal CO_2 had been used.

Another limitation of C-Fix is its assumption of a universal, strongly peaked response of GPP to temperature with an optimum around 22˚C – thus not accounting for thermal acclimation, and necessarily underestimating GPP in hot climates. But C-Fix was designed for use in temperate forests and has in fact only been applied in temperate regions. Veroustraete et al. (2002) successfully compared C-Fix GPP predictions with flux data from forests in different regions of Europe.

C-Fix was not deployed operationally, but it provided the initial basis for the present operational DMP product (Swinnen et al., 2015). In DMP, IPAR is determined from solar shortwave radiation by applying a factor 0.48, and APAR is obtained by multiplying IPAR by a remotely sensed estimate of fAPAR as in other models. Daily meteorological data are supplied by the European Centre for Medium range Weather Forecasting (ECMWF) on a 0.25˚ grid, and bilinearly interpolated to the remote-sensing pixels. A single global maximum value is assigned to the maximum LUE. This is modified by a temperature function and (for NPP) a further temperature function. These two functions are unchanged from C-Fix: production is therefore likely to be underestimated in hot climates, due to the 'temperate' location of the peak of the temperature response function for LUE. There is no effect of water availability (apart from that manifested in changes in fAPAR) and the CO_2 response of C-Fix is not implemented. C_4 photosynthesis is not distinguished.

3.4. SOME RECENT DEVELOPMENTS AND TRENDS

Numerous LUE models have been developed in recent years with the expressed intention of achieving improved consistency with CO_2 flux-based measurements of GPP. These measurements, when suitably analysed, can provide more information about the controls of LUE than is utilized in an 'end-of-pipe' comparison of modelled and measured values of GPP.

The EC-LUE model (Yuan et al., 2007) requires only four quantities as input: fAPAR (which is estimated from the Normalized Difference Vegetation Index, NDVI), IPAR, air temperature and the Bowen ratio, which was inferred from other remotely sensed measurements (Yuan et al., 2007). In a later version the Bowen ratio was replaced by the ratio of actual evapotranspiration to net radiation (Yuan et al., 2010), which proved to be more robustly estimated than the Bowen ratio. EC-LUE has the merit of simplicity, as well as outperforming MODIS GPP in comparisons with flux measurements.

Further innovations in recent LUE models include consideration of seasonal variations in maximum LUE (Garbulsky et al., 2010; Lin et al., 2017), accounting for thermal acclimation (McCallum et al., 2013), the inclusion of an effect of diffuse light fraction on LUE (Donohue et al., 2014), and the use of a MODIS canopy conductance product to include effects of vpd and soil moisture availability (Yebra et al., 2015).

These various recent model developments differ in the extent to which they achieve good empirical results by increasing complexity and adding to the number of unknown parameters, *versus* the alternative of trying to make modelling more robust through the application of theory that can often simplify models and reduce the number of unknown parameters. McCallum et al. (2009), for example, argued for the inclusion of all of those processes that have been shown to improve model performance. An opposite view (Prentice et al., 2015) is that this is a flawed approach that tends always to increase uncertainty, rather than to reduce it, as the number of parameters increases and the transparency of the model decreases.

3.5. THE SCARF MODEL

Ogutu et al. (2013) introduced the Southampton Carbon Flux (SCARF) model, a new LUE model with a number of specific advantages for potential operational use. First, the remote sensing data used to drive the model are the MERIS Total Chlorophyll Index (MTCI). MTCI was considered by Ogutu et al. (2013) to offer an improvement over more standard 'greenness' measures as it explicitly relates to the abundance of green, photosynthesizing tissues. They argued that other measures of fAPAR used in LUE models include light absorption by non-green tissues that do not contribute to GPP. Second, the model substantially avoids spatial discontinuities and the use of a look-up table for biome-specific parameters by (a) adopting universal intrinsic quantum efficiency values defined in terms of the FvCB model for C_3 and C_4 plants respectively, and (b) applying universal temperature and CO_2 response functions for C_3 plants, and vpd response functions for C_3 and C_4 plants. The CO_2 and temperature response functions for C_3 plants were derived from the FcVB model, but an empirical formulation was used for the vpd response functions. A look-up table (together with a number of simplifying assumptions) was used to estimate the fraction of C_3 versus C_4 photosynthesis on a per-pixel basis. The model was evaluated successfully against GPP data derived from flux measurements across Europe and the USA.

Ogutu and Dash (2013) showed that the fidelity of flux measurements to the FvCB model at two study sites was close enough that reasonable estimates of 'green' fAPAR could be obtained by inversion of the model, i.e. estimating the fAPAR required to produce the observed patterns of GPP. This finding strongly supports the notion that LUE models could avoid the need for multiple unknown parameters (including the need for look-up tables for vegetation types, apart from the issue of C_3 versus C_4 photosynthesis) through application of the FvCB model.

3.6. THE BESS MODEL

The Breathing Earth System Simulator (BESS) by Ryu et al. (2011) represents an advanced modelling system for GPP and evapotranspiration (ET) and included many novel features. The model was designed to be the 'first system that harmonizes and utilizes MODIS Atmosphere and Land products on the same projection and spatial resolution over the global land' (Ryu et al., 2011, p. 1), thereby utilizing remotely sensed solar radiation and other meteorological data (from MODIS) and avoiding the need to interpolate such data from a coarse spatial grid. The model was described as calibration-free, that is, no parameters were to be estimated from flux data; all were to be

specified independently. A global selection of 33 flux measurement sites was used to independently evaluate the model's predictions of both GPP and ET.

BESS is more complex than any of the other models discussed in this review. It includes an explicit radiative transfer model for solar radiation in the canopy; a 'two-leaf' model that distinguishes the properties of sun and shade leaves, which has been claimed to provide better accuracy, especially in modelling the differential penetration of diffuse versus direct light into the canopy and the consequences for photosynthesis; consideration of foliar clumping effects on photosynthetic light absorption (making use of a satellite-derived foliar clumping index product); and an extended FvCB model, including light-, Rubisco- and triose phosphate utilization-limited rates of photosynthesis. However, this complexity comes at a considerable cost, both computational, and in terms of data availability.

The model was set up on the Microsoft Azure cloud computing system, as it was considered to be infeasible on the supercomputing resources available at Berkeley. Many compromises were unavoidably made. A look-up table was used to provide values of many parameters, including the carboxylation capacity (V_{cmax}) over much of the Earth's surface. For some biomes V_{cmax} was estimated from foliage N, which in turn was estimated from vegetation albedo – this approach relying on the (questionable) relationship between foliar N and $V_{cmax.}$ The ratio of leaf-internal to ambient CO_2 was set at constant values for C_3 and C_4 plants respectively, thus disregarding the well-established effect of vpd on this ratio. Some external data, not available from MODIS, were obtained from a coarsely gridded re-analysis product. Thus, although BESS includes many advanced features and the ideal of obtaining all required information from remote sensing remains worth pursuing, this ideal was not in fact realized. This approach does not appear to provide a useful way forward for the development of operational systems at the present time.

CHAPTER 4 THE P MODEL: DESCRIPTION OF THE PROPOSED METHOD

4.1. THE P MODEL

The P model is fully derived and described by Wang et al. (2016a), and aspects of the underlying theory have been applied by Keenan et al. (2016) and Wang et al. (2016b). Unlike any of the LUE models discussed above, the P model possesses all of the following desirable attributes for a 'next-generation' primary production monitoring system:

- An explicit derivation from the FvCB model, and a clear relationship to a well-established functional form for stomatal behaviour – both elements required for a prediction of GPP.
- A representation of physiological CO_2 effects on photosynthesis that is consistent with both the FvCB model and results from FACE experiments.
- No distinctions among plant functional types and biomes (except for the well-established differences between C_3 and C_4 plants), eliminating the need for spatial discontinuities induced by the use of a land-cover classification and look-up table.
- Demonstrated success in representing flux-derived GPP across different biomes at monthly time scales.

The model is extremely parameter-sparse, while achieving a fidelity to data comparable with other models. This combination of simplicity with accuracy has been achieved through the development of new theory that accounts for the observed environmental dependencies of the ratio (henceforth termed χ) of the leaf-internal (c_i) to ambient (c_a) partial pressures of CO_2 in C_3 plants; and the acclimation of photosynthetic parameters in space and time. Both aspects of the theory rely on eco-evolutionary optimality concepts to derive testable hypotheses, which in turn yield good agreement with observations from field measurements and field experiments.

4.1.1. PREDICTING χ WITH THE LEAST-COST HYPOTHESIS

Prentice et al. (2014) tested a quantitative theory based on a hypothesis first proposed by Wright et al. (2004), that plants should minimize the sum of the unit costs (per unit of carbon assimilation) of maintaining the capacities for carboxylation (proportional to V_{cmax}) and water transport (proportional to the maximum rate of transpiration). Transpiration is a requirement for photosynthesis because stomata have to open in order to allow CO_2 to diffuse towards the chloroplasts. In so doing, they draw water from the soil to replenish that lost by evaporation through the stomata. Prentice et al. (2014) showed that this 'least-cost' criterion leads to an optimal value of χ as a function of environmental variables (temperature and vpd) that is independent of PAR, and almost independent of c_a. This value is given by:

$$\chi = \Gamma^*/c_a + (1 - \Gamma^*/c_a)\,\xi/(\xi + \sqrt{D}), \quad (1a)$$

$$\xi = \sqrt{\{\beta(K + \Gamma^*)/1.6\eta^*\}} \quad (1b)$$

where Γ^* and K are respectively the photorespiratory compensation point and the effective Michaelis-Menten coefficient of Rubisco (both known functions of temperature and atmospheric

pressure), β is an empirical constant (estimated from $\delta^{13}C$ data: Wang et al., 2016a), η* is the viscosity of water relative to its value at 25˚C (a known function of temperature), and D is the vpd. As $\Gamma^* << c_a$ under field conditions, and $K >> \Gamma^*$, equation (1) is well approximated by:

$$\chi = \xi/(\xi + \sqrt{D}) \quad (2a)$$

$$\xi = \sqrt{\{\beta K/1.6\eta^*\}} \quad (2b)$$

Equation (2a) is identical with the 'universal stomatal model' proposed by Medlyn et al. (2011) and tested against a globally distributed set of gas-exchange measurements by Lin et al. (2012). Lin et al. (2012) also showed that ξ (there called g_1) increases approximately exponentially with temperature. This was predicted by Medlyn et al. (2011). But it also follows from equation (2b), because of the temperature dependencies of both K (increasing) and η* (decreasing).

Wang et al. (2016a, b) noted that the optimal value of χ should also depend on elevation, acting through the effects of changing atmospheric pressure on the partial pressures of both oxygen, which competes with CO_2 for the Rubisco catalytic sites, and water vapour. Wang et al. (2016a) used a large data set of leaf stable carbon isotope ($\delta^{13}C$) measurements to show that all three environmental dependencies are correctly predicted by the model. The predicted partial derivatives of ln $\chi/(1-\chi)$ are 0.055 K^{-1} for temperature, –0.5 for ln vpd, and –0.08 km^{-1} for elevation. These partial derivatives were independently estimated from the $\delta^{13}C$ data by multiple linear regression, yielding 95% confidence intervals that enclose the predicted values: (0.052, 0.046) for temperature, (–0.61, –0.48) for ln vpd, and (–0.08, –0.13) for elevation.

4.1.2. PREDICTING GPP WITH THE CO-ORDINATION HYPOTHESIS

Wang et al. (2016a) also made use of the co-ordination hypothesis, which proposes that acclimation (on time scales of weeks to months) should ensure the similarity of Rubisco- and light-limited photosynthetic rates. This long-standing idea is well supported by independent studies (Haxeltine and Prentice, 1996; Dewar, 1996; Maire et al., 2012) and has a number of implications that are useful for modelling. These include a simple method to predict the spatial and temporal acclimation of V_{cmax} as a function of IPAR and temperature, meaning that V_{cmax} does not have to be specified independently. A variant of this principle is already included in the widely used Lund-Potsdam-Jena (LPJ) dynamic global vegetation model (Sitch et al., 2003) and models derived from LPJ, including the LPX global carbon cycle model (Stocker et al., 2012), although its implications have not been much explored by users of these models.

Wang et al. (2016a) further showed that a cost-benefit analysis of the maximum electron transport capacity J_{max} – which can be measured in the field by artificially increasing c_a to a high level – leads to a predictable optimal ratio of J_{max} to V_{cmax} that declines steeply with growth temperature, in accordance with experimental findings. The mathematical optimization was performed using the Smith formula relating the electron transport rate to absorbed PAR at the leaf level. Inclusion of this acclimation of J_{max} has been found to exert a modest but significantly beneficial effect on the prediction of V_{cmax}. Similar results have been obtained recently using the alternative empirical light response curve (non-rectangular hyperbola with curvature parameter Θ) that is more commonly used in conjunction with the FvCB model. This alternative has not yet been implemented in the P model. Regardless of which light-response curve is used, the practical consequence is the J_{max}, like V_{cmax}, does not need to be independently specified.

Together, the elements described above define a model to predict GPP. This is achieved simply by equating the light- and Rubisco-limited rates of photosynthesis in the FvCB model (implicitly over

an acclimation period of days to weeks, compatible with the time scale of remotely sensed fAPAR products), and re-arranging to eliminate χ, V_{cmax} and J_{max} (Wang et al., 2016a):

$$GPP = \phi_0(C_3) \times fAPAR \times IPAR \times m \sqrt{\{1 - (c^*/m)^{2/3}\}} \quad (3a)$$

where $\phi_0(C_3)$ is the dimensionless intrinsic quantum efficiency of C_3 photosynthesis (taken to be 0.085 by Wang et al., 2016a), c^* is a parameter representing the unit cost of maintaining the capacity for electron transport, and

$$m = \{c_a - \Gamma^*\} / \{c_a + 2\Gamma^* + 3\Gamma^* \sqrt{[1.6\, \eta^*\, D\, \beta^{-1}\, (K + \Gamma^*)^{-1}]}\} \quad (3b)$$

Equation (3) has the mathematical form of a LUE model: that is, for a given set of environmental conditions (ambient atmospheric CO_2 concentration, temperature, atmospheric pressure and vpd) modelled GPP is proportional to the absorbed PPFD. But unlike other LUE models, equation (3) is explicitly defined in terms of the FvCB model of photosynthesis. Although GPP at time scales of minutes to hours (as seen, for example, during the diurnal cycle of CO_2 flux) has a saturating response to IPAR, GPP at longer (e.g. weekly) time scales has a linear response to IPAR conferred by the acclimation of V_{cmax}. This principle was previously articulated by Haxeltine and Prentice (1996) and Dewar (1996), and can provide a theoretical basis for LUE models (Medlyn, 1998). Equation (3) gives mathematical expression to the principle, and has proved to be at least as effective in terms of simulating flux-derived monthly GPP as other LUE models – as shown in Wang et al. (2016), and in Table 1.

Table 2: Goodness of fit (R^2) and root-mean-squared error of prediction (RMSE) statistics for P model (Wang et al., 2016) predictions of monthly GPP, compared with results from several LUE models tested against flux measurements by Yuan et al. (2014) (data provided by H. Wang).

	R^2		RMSE	
	This study	Yuan et al.	This study	Yuan et al.
All ecosystems	0.551	0.553 ± 0.096	2.094	2.428 ± 0.275
Shrubland	0.772	0.255 ± 0.175	2.165	1.866 ± 0.915
Deciduous broadleaf forest	0.588	0.703 ± 0.094	2.766	2.919 ± 0.450
Evergreen broadleaf forest	0.341	0.119 ± 0.063	2.046	2.961 ± 0.801
Evergreen needleleaf forest	0.535	0.501 ± 0.108	1.856	2.384 ± 0.437
Grassland	0.572	0.631 ± 0.076	2.025	2.109 ± 0.280
Mixed forest	0.700	0.637 ± 0.068	1.824	2.339 ± 0.325

4.1.3. EFFECTS OF CO_2 IN THE P MODEL

It follows from the co-ordination hypothesis that the benefit of rising CO_2 for the LUE of GPP by C_3 plants will be limited to its effect on the *light-limited* rate of photosynthesis. This effect is predicted by the P model, with no additional parameter requirements, including its well-known interaction with temperature and vpd. Effects of CO_2 on different photosynthesis metrics, as measured in 12 FACE experiments, were the subject of a meta-analysis by Ainsworth and Long (2005). They showed (for a 200 ppm increase in c_a) that LUE changed by an average of + 12 ± 9 %, instantaneous

water-use efficiency by + 54 ± 17 %, and stomatal conductance by – 20 ± 3 %. Corresponding predictions with the P model were + 17%, + 55% and – 15% (data provided by H. Wang).

4.1.4. SOIL MOISTURE EFFECTS

In common with many LUE models, including those used operationally, the P model does not take account of soil moisture effects – except in so far as they are manifested through changes in fAPAR. Analysis of flux-based GPP measurements at most of the sites considered by Wang et al. (2016a), including a number of sites with pronounced dry seasons, has shown that there is no regular fall-off of LUE with drought. However, this is not universal. Some ecosystems (for example, some tropical savannas) do show reduced LUE in the dry season. Moreover, extreme droughts to which ecosystems are not well adapted are expected to suppress LUE by a combination of reduced χ and reduced V_{cmax}, as has been widely observed in drying-down experiments (Zhou et al., 2013). Therefore, in common with other products the initial P model implementation is likely to overestimate dry-season GPP in some ecosystems, and to underestimate the effects of extreme droughts on GPP (B. Stocker, unpublished results).

4.1.5. C_4 PHOTOSYNTHESIS

The simplest way to implement C_4 photosynthesis makes just two modifications to equation (3). First, a generic ϕ_0 value suitable for C_4 plants must be chosen (taken to be 0.055 in current work). Second, c_a is made arbitrarily large. These two changes lead to a simplified equation for C_4 photosynthesis:

$$\mathrm{GPP} = \phi_0(C_4) \times \mathrm{fAPAR} \times \mathrm{IPAR} \qquad (4)$$

C_4 photosynthesis can benefit from rising CO_2 under conditions of limited water availability, because water-use efficiency increases even if photosynthesis does not. However, this benefit is expected to be fully realized in increasing fAPAR.

4.1.6. MODELLING ABOVE-GROUND BIOMASS PRODUCTION

The translation from GPP to ABP (in carbon units, easily modified to dry matter units) can be summarized by the formula ABP = $(1 - f_{BG})$ x CUE x GPP, where f_{BG} is the fraction of NPP allocated below ground (including root exudation, as well as allocation to the maintenance and turnover of roots) and CUE is the carbon use efficiency, i.e. the ratio of NPP to GPP. (This formula disregards the fraction of NPP allocated to VOC emission, which is normally much smaller than f_{BG}.)

The additional terms required to calculate ABP from GPP are much less well understood from a theoretical and quantitative point of view than the terms in the equations for GPP itself. However, there is evidence for two competing effects of temperature on the ratio of ABP to GPP. On the one hand, acclimation of V_{cmax} results in a weakly increasing V_{cmax} with growth temperature (Atkin et al., 2015); and leaf dark respiration varies approximately in proportion to V_{cmax}, according to the FvCB model. This effect, by itself, would cause CUE to decrease with increasing temperature. On the other hand, warm conditions increase the availability of soil nutrients by enhancing the rates of microbial metabolism, thus diminishing the need for plants to allocate carbon below ground – and thereby reducing f_{BG} (Gill and Finzi, 2017).

A statistical analysis of the NPP data provided by Michaletz et al. (2014) yielded the equations NPP ≈ 0.542 GPP + 0.0127 (T_C – 25) and ABP ≈ 0.535 GPP + 0.0074 (T_C – 25), where T_C is the growth temperature in ˚C (analysis by H. Wang). These preliminary results suggest that the effect of increased nutrient availability with temperature dominates over the effect of increased leaf respiration rate. There is substantial scatter around these relationships that likely reflects variations in soil fertility and management regimes (Vicca et al., 2012; Campioli et al., 2015). However, this general approach – applied to an improved data set under development by the University of Antwerp group – will allow us to derive an empirically well-founded, albeit approximate, prediction scheme for ABP. It is expected that soil fertility, indexed by pH and/or C:N ratio, will be a factor in the scheme.

The general approach outlined above is preferred to trying to explicitly model total autotrophic respiration, which has proved to be a major problem in the MODIS NPP product. Our approach implicitly assumes that because of the ubiquitous acclimation of autotrophic respiration to temperature, the instantaneous response of respiration rates to temperature (as expressed in the Q_{10} factor employed by many models, including MODIS NPP) is largely irrelevant to predicting the ratio of either ABP or NPP to GPP – just as the instantaneous responses of photosynthetic rates to light and temperature do not determine the responses of actual photosynthetic rates under field conditions.

4.2. DATA NEEDS TO IMPLEMENT THE P MODEL

Here we list the data requirements for the initial global-scale implementation of the P model.

- *fAPAR* per pixel and dekad or month: these data have been obtained initially from SeaWiFS and MERIS GVI data, and will eventually be obtained from Sentinel-3.
- *Meteorological data*: daily total 'global' (solar shortwave) radiation, daily minimum and maximum temperature, and daily absolute vapour pressure. These data will be obtained initially at 0.25˚ resolution from the ECMWF data stream, and interpolated to the pixel scale, in the same way as is currently implemented for the DMP product. IPAR will be estimated from global radiation as in the DMP product, and converted to PPFD for input to the P model.
- *Ambient partial pressure of CO_2*: a single global, time-varying, value will be obtained from the international CO_2 monitoring network and used as input to the P model. (However, see the discussion in chapter 4.3 regarding the uncertainty of c_a under field conditions.)
- *C_3 versus C_4 photosynthesis*: the need for data will be circumvented by providing both values for every pixel.

Subsequent developments under consideration for later implementation include:

- Implementation of a remotely sensed soil moisture effect on GPP, through a modification of the P model. Although in relatively moist soils the actual soil moisture content has little effect on GPP, very low soil moisture can negatively affect GPP in addition to the effect of high vpd (as noted in section 4.1.4 above). This soil moisture effect is not incorporated in existing operational products which, therefore, share a common bias towards overpredicting GPP under drought conditions. This bias could be mitigated by using a remotely sensed soil moisture product as an additional driver of the P model.
- Substitution of remotely sensed land surface temperature for air temperature as a driver of the P model. This is desirable for two reasons. First, it would avoid dependence on temperature values interpolated from a coarser grid. Second, leaf temperature rather than air temperature is the actual driver of photosynthesis. Most photosynthesis is carried out by the leaves in the

upper canopy. Remotely sensed land surface temperature over continuous vegetation provides information on the temperature of the upper-canopy leaves which is, therefore, directly relevant to GPP. Over sparse canopies, remotely sensed land surface temperature combines signals from the vegetation and soil surface and therefore may need to be corrected in some way based on the observed fAPAR.

- Use of either a remotely sensed digital elevation model, or remotely sensed surface pressure, to account for atmospheric pressure effects on photosynthesis via the P model. This would improve the accuracy of GPP values for high elevations.
- For the ABP model: use of global soils data, to be translated into scalars reflecting the effect of soil fertility on the ratio of ABP to GPP. Soil fertility can be indexed e.g. by pH (low-pH soils tend to be less fertile) or C:N ratio (more fertile soils tend to have narrower C:N ratios). Machine learning methods have been used to combine national soil survey mapping with soil profile measurements at 250 m resolution (https://www.soilgrids.org) and provide the best available source of global data on soil types and properties.

4.3. AN APPROACH TO ESTIMATING PER-PIXEL UNCERTAINTY IN GPP

Two independent methods will be applied to generate per-pixel uncertainties, taking into account that on the one hand, the P model algorithm is derived from first principles and consists of a single equation, which can be differentiated with respect to all of the uncertain quantities that it contains; and on the other hand, the algorithm's credibility is assured by its ability to predict independently measured GPP, which will be quantified.

Thus two methods will be used to quantify uncertainties in GPP. The first method will be a classical Type B uncertainty evaluation, derived analytically and producing a per-pixel uncertainty value explicitly considering the known sources of uncertainty in different quantities entering the model and combining them using established principles. The second can be considered as a Type A evaluation in so far as data from different observation periods at a flux measurement site, and data from flux different sites, can be considered as stochastic realizations of the same underlying processes. Uncertainty estimates obtained by the second (Type A) method may include consequences of processes that are not explicitly included in the GPP algorithm, potentially leading to wider uncertainty estimates than those obtained by the first (Type B) method. A criterion for this approach to work well is that the estimated uncertainty should be based on a sufficiently large and diverse ensemble of flux sites. Implicitly, errors in flux measurements and their partitioning to GPP would be included in the Type A uncertainty assessment.

4.3.1. UNCERTAINTY EVALUATION BASED ON THE P MODEL ALGORITHM

Equations (3) and (4) contain various input variables and parameters whose uncertainty needs to be quantified. In addition, a number of the photosynthetic parameters are temperature-dependent. Uncertainties in the temperature dependencies can be isolated from uncertainty in the temperature data by applying the following standard formulae:

$$\Gamma^* = \Gamma^*[25] \exp\{(\Delta H_{\Gamma^*}/R)(1/298.15 - 1/T)\} \quad (5)$$

$$\eta^* = \exp\{580\,[1/(T - 138)] - [1/(160]\} \quad (6)$$

$$K = K_C\,(1 + O/K_O) \quad (7)$$

$$K_C = K_C[25] \exp\{(\Delta H_{K\_C}/R)(1/298.15 - 1/T)\} \quad (8)$$

$$K_O = K_O[25] \exp\{(\Delta H_{K\_O}/R)(1/298.15 - 1/T)\} \quad (9)$$

where R is the universal gas constant (8.314 46 J mol^{-1} K^{-1}), T is the canopy temperature (K), K_C is the Michaelis-Menten coefficient for carboxylation (Pa), K_O is the Michaelis-Menten coefficient for oxygenation (Pa), O is the partial pressure of oxygen (209 460 µmol mol^{-1} x atmospheric pressure in Pa); $\Gamma^*[25]$, $K_C[25]$ and $K_O[25]$ are the values of Γ^*, K_C and K_O, respectively, at 298.15 K; and ΔH_{Γ^*}, ΔH_{K_C} and ΔH_{K_O} are the corresponding activation energies (J mol^{-1}). Moreover, if D is estimated from absolute water vapour pressure (e_a) and saturation vapour pressure (e_s), then:

$$D = e_s(T_C) - e_a \quad (10)$$

where $e_s = e_s(0) \exp\{17.27\, T_C/(T_C + 237.3)\}$ (Pa) and $T_C = T - 273.15$ K.

Those quantities that are either defined precisely, or known with an uncertainty that is effectively negligible in this context, have been assigned numerical values above and will not be considered further. In the following section we describe the sources of data to estimate standard uncertainties for each of the remaining quantities for use in the initial GPP product. The formulation above allows each of the sources of uncertainty to be considered independent and, therefore, uncertainties from each source to be combined using the standard formula:

$$u^2(y) = \sum_i (\partial f/\partial x_i)^2\, u^2(x_i) \quad (11)$$

where $u(y)$ is the standard uncertainty of GPP, $\partial f/\partial x_i$ is the sensitivity of GPP to variable x_i (obtained by differentiating equation (1) with respect to each uncertain variable and evaluating the partial derivative at the current central value of x_i), and $u(x_i)$ is the standard uncertainty of x_i.

4.3.2. DATA UNCERTAINTIES

fAPAR: standard uncertainties provided with the EO products (initially, MERIS GVI) will form part of the data input.

Uncertainties of IPAR, T and e_a are provided with the meteorological data by ECMWF. T will be corrected for the expected offset between canopy and air temperature following the general theoretical form presented by Michaletz et al. (2016) and its intrinsic uncertainty will be combined with an uncertainty estimate based on the statistics of the fitted global relationship between canopy and air temperature, based on point measurements, presented in Michaletz et al. (2016).

Spatial and temporal variations in c_a can be obtained via the integrated CO_2 measurement data set maintained by the US National Oceanic and Atmospheric Administration (NOAA):
http://www.esrl.noaa.gov/gmd/ccgg/globalview/co2/co2_intro.html
The largest uncertainty in c_a however derives *not* from the systematic deviations from a single global c_a that are observed by the monitoring network, but rather from local, non-measured variations due to ground-level sources and sinks in soils and vegetation and local industrial and/or transport sources. This uncertainty will be quantified approximately based on a literature survey.

Parameter uncertainties

β has been estimated based on leaf $\delta^{13}C$ data, from the intercept of the regression of $\ln \chi/(1 - \chi)$ against environmental predictors (Wang et al., 2016a). The uncertainty of this estimate will be assessed from the standard error of the intercept, and inflated to account for uncertainty in the

conversion from stable isotope measurements to χ. *c** has been estimated from published values of electron transport capacity (J_{max}) and carboxylation capacity (V_{cmax}) under a variety of experimental growth conditions (Kattge and Knorr, 2007; Wang et al., 2016a). The uncertainty of *c** will be estimated based on a regression of experimentally determined J_{max}/V_{cmax} values against growth temperature.

The remaining parameters of equations (3) and (4) are standard elements of the FcVB photosynthesis model. They are accurately measured and show little variation among different plant species and measurement techniques. Nonetheless, they are subject to some uncertainty, which will be taken into account as described below.

$\phi_0(C_3)$ and $\phi_0(C_4)$: published surveys of measurements on various species provide the basis to quantify the uncertainty in these parameters (Skillman, 2008; Zhu et al., 2010), which is approximately normally distributed across species within each photosynthetic pathway.

Γ*[25], K_C[25], K_O[25] and the corresponding activation energies: although most recent modelling studies have used the *in vivo* values determined by Bernacchi et al. (2001), two other modern experimental data sets give slightly different reference values and activation energies (De Kauwe et al., 2016b). There are also published studies showing of variation in Rubisco kinetic properties across species from different environments (e.g. Hermida-Carerra et al., 2016). An estimate of the uncertainty of all six quantities will be derived based on this recent literature.

4.3.3. COMBINING UNCERTAINTIES

Derivatives of equation (1) with respect to each uncertain quantity will be obtained analytically. For constant quantities such as the two ϕ_0 values the derivative will be pre-calculated. For quantities that vary in time and/or space, the derivative will be evaluated as part of the standard workflow. The outputs of the workflow per time-step and pixel will include the central estimate of GPP from equation (1) and its composite standard uncertainty from equation (2) with elements calculated in the manner described above.

4.3.4. UNCERTAINTY EVALUATION BASED ON A COMPARISON OF MODELLED AND MEASURED GPP

Sections 4.3.1 to 4.3.3 above outline the proposed method for calculation of a Type B, per-pixel uncertainty for modelled GPP. The alternative Type A approach involves defining a model for total uncertainty based on a statistical comparison of observed and modelled GPP across space and time. The comparative analysis of modelled and measured GPP by Wang et al. (2016a) indicated no bias, and visually suggested that a suitable model for the empirical distribution of uncertainties might consider relative uncertainty (fraction of estimated GPP) as a constant. This conclusion will be re-visited based on the new evaluation in TerrA-P, and used to define a Type A uncertainty calculation. The two proposed methods to quantify uncertainty will be complementary and act as a cross-check on one another.

4.4. A PRELIMINARY CALIBRATION DATA SET FOR GPP

The University of Antwerp group has selected from the most recent synthesis data set (http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/) 17 flux sites in different biomes that have data in the public domain and are characterized by multi-year records and relatively homogeneous

vegetation footprints. These sites, tabulated below (Table 3), provide the basis for preliminary calibration.

Table 3: The selected initial calibration set of eddy-covariance flux measurement sites. VEG = IGBP vegetation type: EBF = evergreen broadleaf forest, ENF = evergreen needleleaf forest, OSH = open shrubland, CRO = cropland, DBF = deciduous broadleaf forest.

CODE	NAME	LAT (˚)	LONG (˚)	ELEV (m)	VEG
AU-Tum	Tumbarumba	–35.6566	148.1517	645	EBF
CA-NS3	UCI-1964 burn site	55.9117	–98.3822	260	ENF
CA-NS6	UCI-1989 burn site	55.9167	–98.9644	244	OSH
CA-Obs	Saskatchewan – Western Boreal, Mature Black Spruce	53.9872	–105.1178	629	ENF
DE-Geb	Gebesee	51.1001	10.9143	162	CRO
DE-Hai	Hainich	51.0792	10.4530	430	DBF
DE-Kli	Klingenberg	50.8929	13.5225	478	CRO
FI-Hyy	Hyyttiälä	61.8475	24.2950	181	ENF
FR-Fon	Fontainebleau-Barbeau	48.4764	2.7801	103	DBF
FR-LBr	Le Bray (after 28 June 1998)	44.7171	–0.7693	61	ENF
FR-Pue	Puechabon	43.7414	3.5958	270	EBF
IT-Cpz	Castelporziano	41.7052	12.3761	68	EBF
NL-Loo	Loobos	52.1666	5.7436	25	ENF
US-Ha1	Harvard Forest EMS Tower (HFR1)	42.5378	–72.1715	340	DBF
US-MMS	Morgan Monroe State Forest	39.3232	–86.4131	275	DBF
US-UMB	University of Michigan Biological Station	45.5598	–84.7138	234	DBF
US-WCr	Willow Creek	45.8059	–90.0799	520	DBF

Approximately 20 additional flux sites have been targeted as having multi-year records and relatively homogeneous footprints, but requiring liaison with the flux-site principal investigators, for subsequent validation. Validation will later include on the order of 100 additional public-domain sites of shorter duration and/or more complex footprints. The inclusion of this larger set (accepting there may be issues arising due to the heterogeneous footprints of some sites) is important in order to obtain a wider spread of ecosystem types.

4.5. CALIBRATION RESULTS

Simulations were set up for each of the 17 calibration sites, using local meteorological measurements of daily total incoming shortwave radiation (converted to PPFD using a factor of 0.49) and monthly mean temperature and vapour pressure (converted to vpd using the standard method, using mean daily minimum and maximum temperatures). Annual values of CO_2 were prescribed. Low-temperature inhibition of photosynthesis was represented in the simplest possible way, by setting GPP to zero during periods with subfreezing temperatures.

The parameters $\phi_0(C_3)$ and $\phi_0(C_4)$ can in principle be calibrated within a certain range. This is because (a) there is some natural variation (as well as methodological and measurement uncertainties) in their values (Skillman, 2008), and (b) the absolute magnitudes of fAPAR vary among remotely sensed products, even if the spatial patterns are similar. $\phi_0(C_3)$ was estimated from the comparison of P model estimates with the GPP data by varying its value in the model

between 0.05 and 0.1125. The optimized value (yielding the smallest sum of RMSE across sites) was 0.084, only marginally different from the standard value of 0.085 used by Wang et al. (2016a). Figure 2 shows the effect of varying $\phi_0(C_3)$ on the sum of RMSE across sites. $\phi_0(C_4)$ was not calibrated but should be estimated using data from C_4 plant-dominated vegetation, which is not represented in the calibration set.

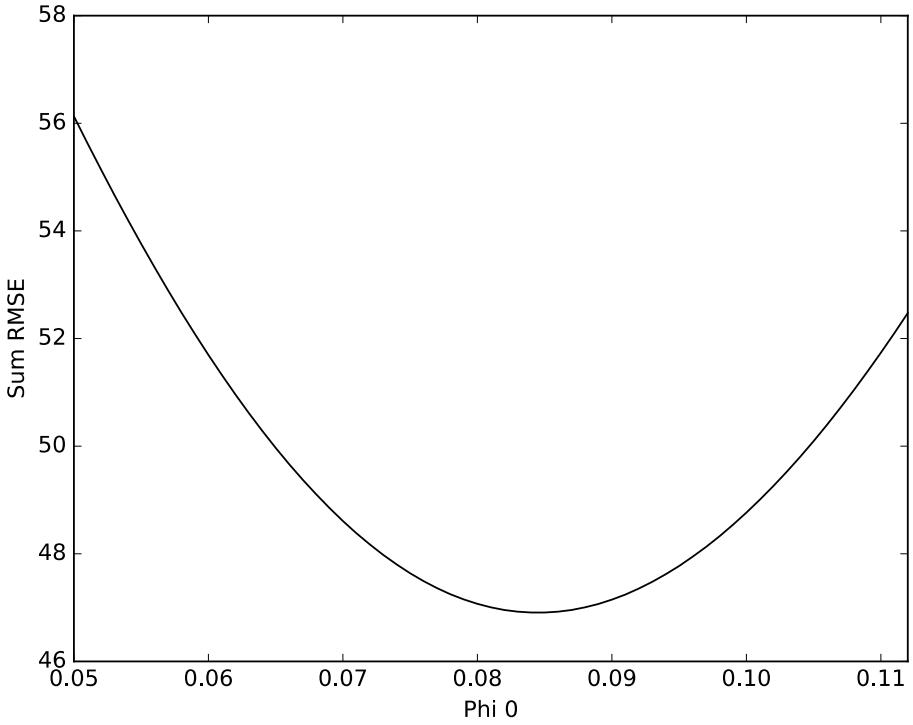
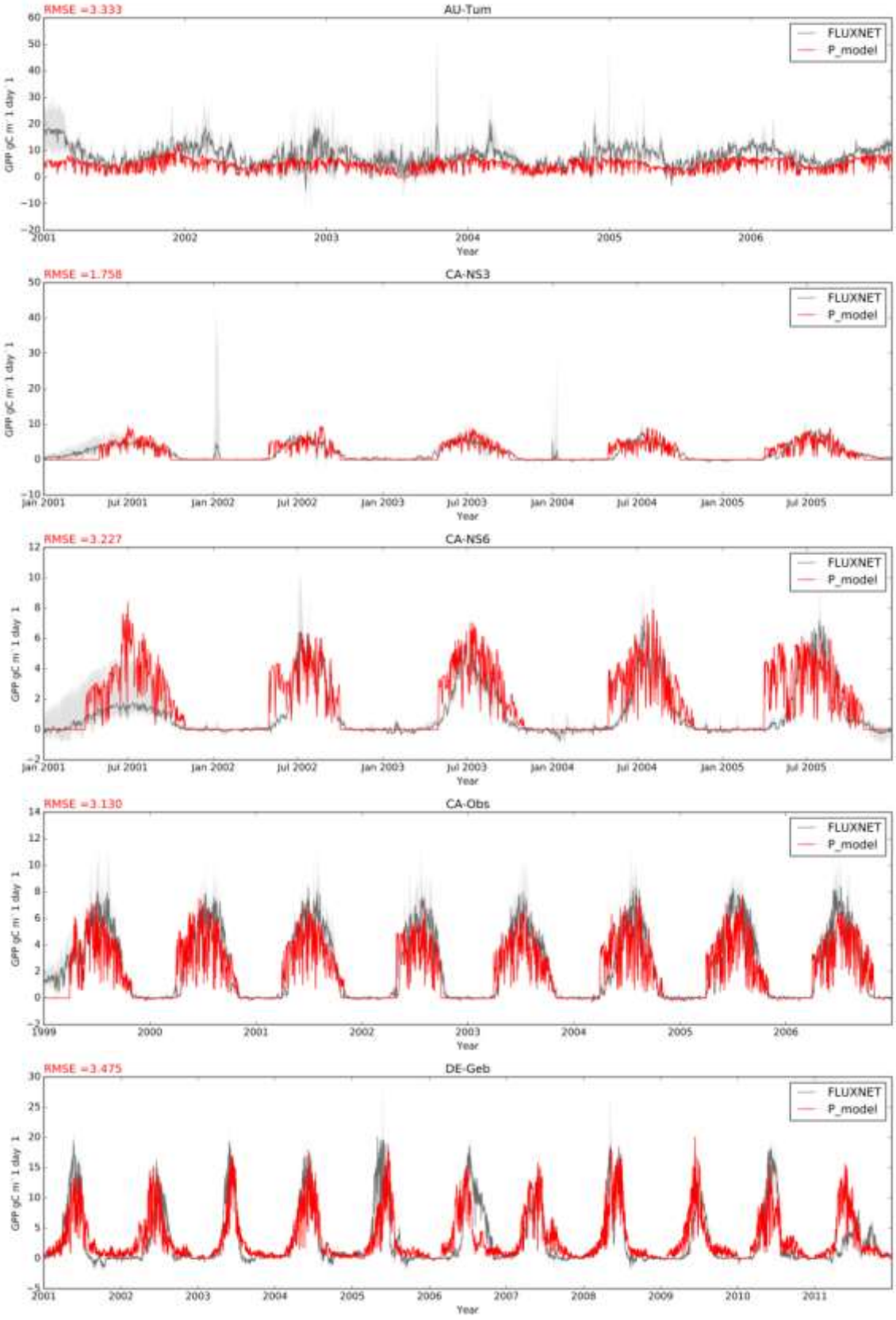
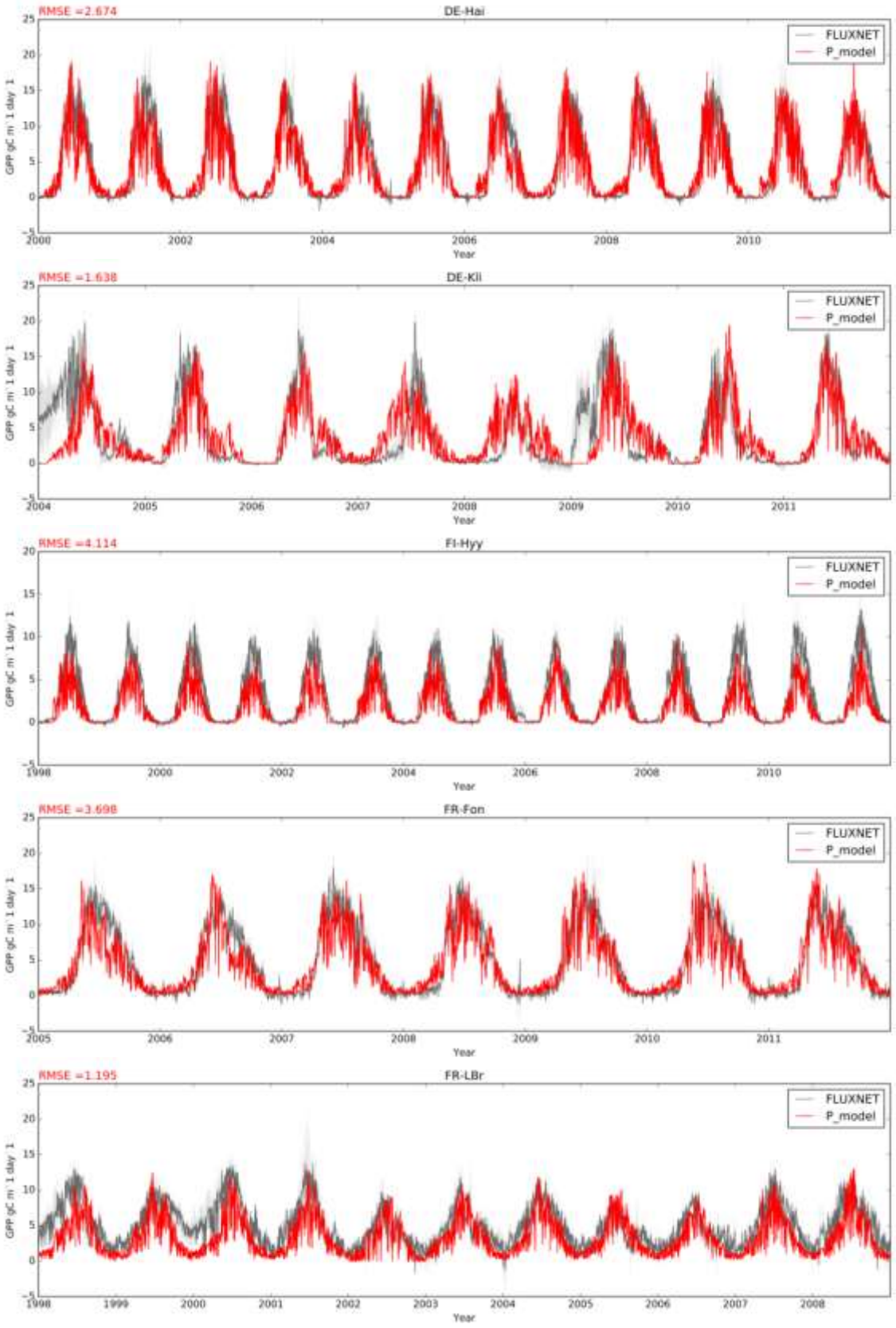


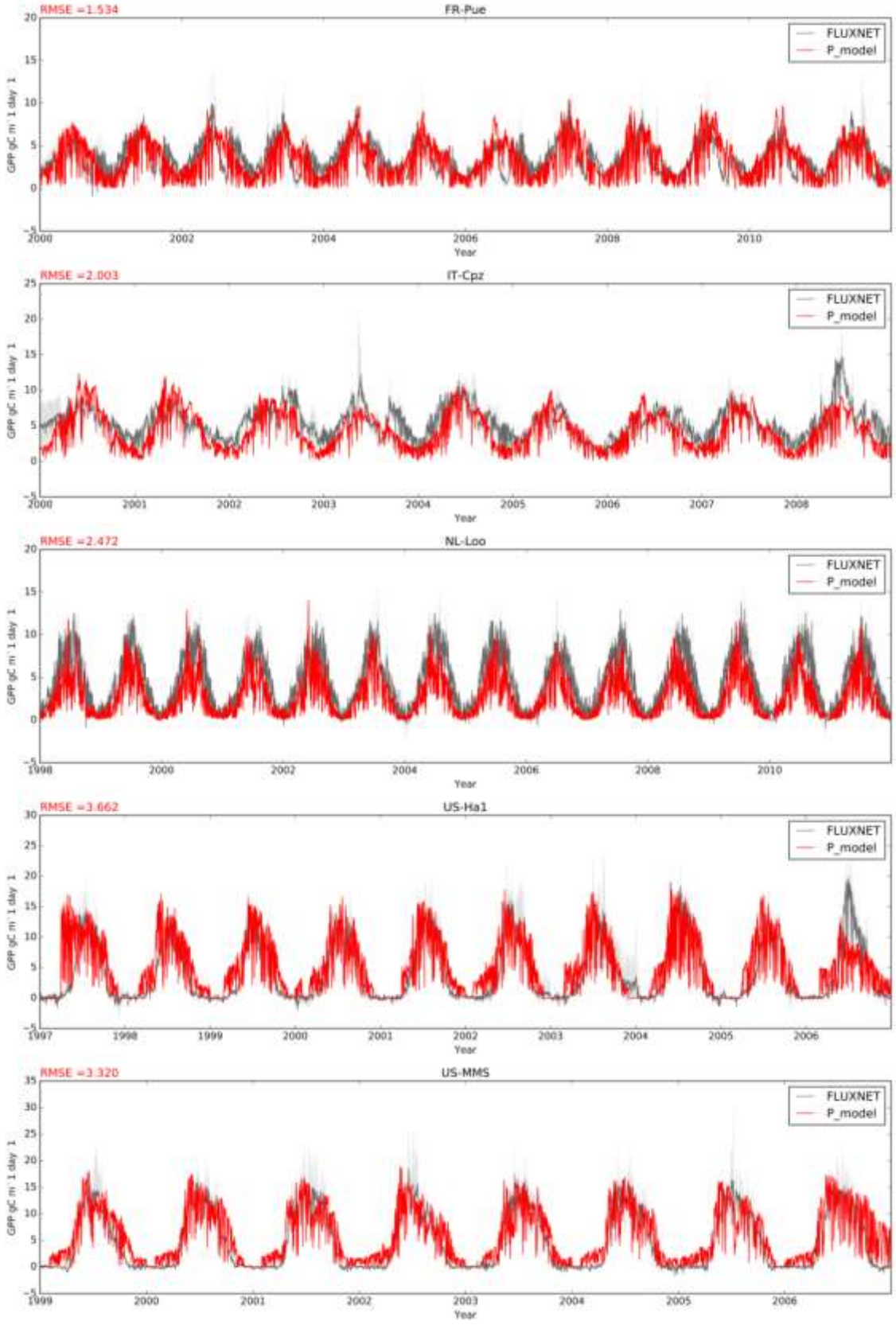
Figure 2: Effect of varying $\varphi_0(C_3)$ on the summed daily RMSE between flux-derived and modelled GPP at the 17 calibration sites.

Figure 3 shows the results of data-model comparison in the form of time series of GPP from the flux measurements, and from the P model with optimized $\phi_0(C_3)$. Visual agreement and RMSE values are generally satisfactory. There are some mismatches, which do not appear to be related to vegetation type; for example, there is no indication in this comparison that GPP is systematically either under-or overestimated in any one vegetation type. Mismatches include underestimation of peak-season GPP by the model at a few sites; and in some sites and years, the simulation of positive GPP around the start and/or end of the growing season at times when the flux-derived GPP is close to zero.

The cause of the first problem is currently unknown. The second problem could be due in part to a mismatch between the tower footprint and the remote sensing pixel. It may also indicate that the simple threshold approach we have adopted to stand for low-temperature inhibition is too simple. Both issues will be investigated during later stages of the project.


RMSE =3.333
AU-Tum
FLUXNET
P_model
GPP gC m⁻² 1 day⁻¹
Year
RMSE =1.758
CA-NS3
RMSE =3.227
CA-NS6
RMSE =3.130
CA-Obs
RMSE =3.475
DE-Geb


RMSE =2.674
DE-Hai
FLUXNET
P_model
GPP gC m⁻¹ 1 day⁻¹
Year
RMSE =1.638
DE-Kli
FLUXNET
P_model
GPP gC m⁻¹ 1 day⁻¹
Year
RMSE =4.114
FI-Hyy
FLUXNET
P_model
GPP gC m⁻¹ 1 day⁻¹
Year
RMSE =3.698
FR-Fon
FLUXNET
P_model
GPP gC m⁻¹ 1 day⁻¹
Year
RMSE =1.195
FR-LBr
FLUXNET
P_model
GPP gC m⁻¹ 1 day⁻¹
Year


RMSE =1.534
FR-Pue
FLUXNET
P_model
GPP gC m⁻¹ 1 day⁻¹
Year
RMSE =2.003
IT-Cpz
FLUXNET
P_model
GPP gC m⁻¹ 1 day⁻¹
Year
RMSE =2.472
NL-Loo
FLUXNET
P_model
GPP gC m⁻¹ 1 day⁻¹
Year
RMSE =3.662
US-Ha1
FLUXNET
P_model
GPP gC m⁻¹ 1 day⁻¹
Year
RMSE =3.320
US-MMS
FLUXNET
P_model
GPP gC m⁻¹ 1 day⁻¹
Year

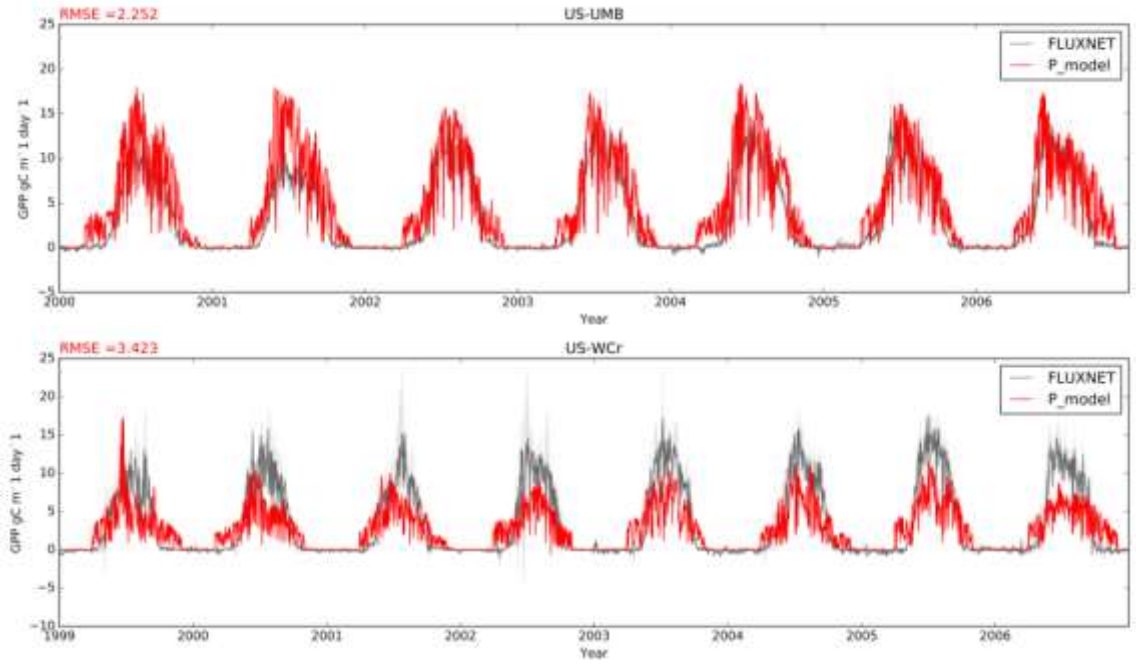


Figure 3: Comparison of flux-derived GPP (grey) and P model-simulated GPP (red) at the calibration sites. The dark grey traces represent the mean GPP from the alternative FLUXNET partitioning methods; the light grey bands represent the range of GPP. The red traces represent modelled GPP.

CHAPTER 5 VALIDATION APPROACH

This chapter describes the validation approach of the products. A two-fold validation will be done:

(1) The point location products will be validated with in-situ measurements of GPP, and where possible against aboveground biomass production (NPP) to assess the accuracy and the limitations of the products.

(2) The global spatial GPP/NPP products are compared against other operational EO products (MODIS, C-GLOPS1) to assess how much and where the new products differ from the existing ones.

5.2. VALIDATION METHOD AGAINST IN-SITU DATA

This validation aims to assess the capability of model simulations to describe carbon dynamics for a variety of ecosystems and to identify potential ways to improve these simulations. More specifically we will test how well the simulations: (i) describe annual carbon dynamics for different ecosystems (forest, grassland and cropland) and climates (ii) encompass the daily and seasonal trends of GPP; (iii) describe interannual variability; (iii) describe the main environmental functions controlling GPP and NPP.

5.2.1. DESCRIPTION OF THE IN SITU DATA

Model outputs will be evaluated against sites selected from two in-situ databases:

- **FLUXNET 2015 database**

For GPP, we will use the FLUXNET 2015 database. The FLUXNET database contains data of ecosystem CO_2 fluxes obtained with the eddy-covariance technique. This well-established technique provides GPP data from the post-processing of the direct measurement of net ecosystem CO_2 exchange. FLUXNET integrates data from regional networks, international projects and field-sites of research institutes and provides a highly standardized data treatment and data analysis (including uncertainty estimations) for measuring sites distributed globally. These data have been extensively used for the development and evaluation of ecological models at regional or global scale (Balzarolo et al 2014).

The FLUXNET GPP data are well suited to be used as validation product for several reasons: (i) they are available at both very high time resolution (half-hourly) and aggregated at longer time step (daily to annual), (ii) are available for multiple years (up to >10 years for the most intensively studied sites), (iii) typically measure an area of the ecosystem (footprint) comparable to the resolution of remote sensed products, (iv) data are provided with uncertainty estimations, and (v) sites are available globally and for all types of terrestrial ecosystems (e.g. forests, grasslands, croplands, wetlands, tundra). FLUXNET has been established in 1998 (Baldocchi et al 2001) and since then techniques have improved and different database version have been produced. In this project, we will use the latest version (FLUXNET 2015) and publically available data, for which data use is free and open provided that proper acknowledgment is given to site PIs and funding agencies

(Tier 1, see http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/). About 200-250 GPP site are available now.

- **Biomass production dataset**

For biomass production, we will use a dataset recently released by the UA partner of this consortium (UA, Vicca et al 2012 and Campioli et al 2015). This dataset is best suited for validation purposes because (i) it provides quality-controlled data for both aboveground and belowground NPP, (ii) it provides standardized uncertainty estimates, (iii) provides ecosystem level data (e.g. dominant and codominant species, overstory and understory) compatible with the spatial footprint of the FLUXNET 2015 sites, thus comparable to remotely sensed data and (iv) provides NPP data paralleled by GPP for the same year of measurements. This dataset is published and has been made publically available (Vicca et al 2012 and Campioli et al 2015).

Site selection

Table 4 lists the 17 FLUXNET sites selected as ideal sites for the calibration of GPP model (calibration sites). These sites are ideal because they have:

- a long time series (>5 years, daily time series) with good data quality. GPP data quality was checked by following the standardized methodology defined in the FLUXNET (Reichstein et al., 2005; Papale et al., 2006)
- a large homogenous footprint area at least 1 km x 1 km spatial resolution to ensures reliable comparisons between in-situ and remotely sensed data

Additional sites for calibration and validation of products will be selected from FLUXNET 2015 database, Tier 1 sites. For validation NPP products, sites where both GPP and NPP estimates are available will be selected from biomass production dataset (Campioli et al 2015 and Vicca et al 2012).

Table 4: Coordinates and characteristics of the FLUXNET 2015 sites selected for the validation of the model outputs. ENF—Evergreen Needleleaf Forest; EBF—Evergreen Broadleaf Forest; DBF—Deciduous Broadleaf Forest; and CRO—Cropland.

SITE_ID	SITE_NAME	2015_DATA_START	2015_DATA_END	LOCATION_LAT	LOCATION_LONG	LOCATION_ELEV	IGBP
AU-Tum	Tumbarumba	Tier1: 2001	Tier1: 2014	-35.6566	148.1517		EBF
CA-NS3	UCI-1964 burn site	Tier1: 2001	Tier1: 2005	55.9117	-98.3822	260	ENF
CA-NS6	UCI-1989 burn site	Tier1: 2001	Tier1: 2005	55.9167	-98.9644	244	OSH
CA-Obs	Saskatchewan - Western Boreal, Mature Black Spruce	Tier2: 1997	Tier2: 2010	53.9872	-105.118	628.94	ENF
DE-Geb	Gebesee	Tier1: 2001	Tier1: 2014	51.1001	10.9143	161.5	CRO
DE-Hai	Hainich	Tier1: 2000	Tier1: 2012	51.0792	10.453	430	DBF
DE-Kli	Klingenberg	Tier1: 2004	Tier1: 2014	50.8929	13.5225	478	CRO
FI-Hyy	Hyytiala	Tier1: 1996	Tier1: 2014	61.8475	24.295	181	ENF
FR-Fon	Fontainebleau-Barbeau	Tier1: 2005	Tier1: 2014	48.4764	2.7801	103	DBF
FR-LBr	Le Bray (after 6/28/1998)	Tier1: 1996	Tier1: 2008	44.7171	-0.7693	61	EN

							F
FR-Pue	Puechabon	Tier1: 2000	Tier1: 2014	43.7414	3.5958	270	EBF
IT-Cpz	Castelporziano	Tier1: 1997	Tier1: 2009	41.7052	12.3761	68	EBF
NL-Loo	Loobos	Tier1: 1996	Tier1: 2013	52.1666	5.7436	25	ENF
US-Ha1	Harvard Forest EMS Tower (HFR1)	Tier1: 1991	Tier1: 2012	42.5378	-72.1715	340	DBF
US-MMS	Morgan Monroe State Forest	Tier1: 1999	Tier1: 2014	39.3232	-86.4131	275	DBF
US-UMB	Univ. of Mich. Biological Station	Tier1: 2000	Tier1: 2014	45.5598	-84.7138	234	DBF
US-WCr	Willow Creek	Tier1: 1999	Tier1: 2014	45.8059	-90.0799	520	DBF

5.2.2. VALIDATION METHOD

Validation will be performed at different temporal resolution but always at high spatial resolution i.e. at site level (we will not perform landscape or regional validations as this will require up-scaling of in-situ data with large propagation of uncertainty).

For GPP, the following metrics will be tested:

1. Between-sites differences in average annual GPP (modeled GPP against in-situ GPP)
2. Interannual variation in modeled GPP against in-situ observation of GPP (only for the years for which the remote sensed product matches observations)
3. Seasonal variability in GPP: daily (considering mean or cumulated values over 24 hours), 10-days and monthly time scales

For NPP, test 1 and 2 will be done, while tests 3 will not be possible because NPP is estimated annually.

For the comparison between modeled and observed data we will use various statistical approaches (depending on the exact objective). For example, the following tests can be applied to evaluate model performance at annual scale in relation to the uncertainty of in-situ data: (i) paired-t-test and Wilcoxon signed-rank tests between pairs of modeled versus observed variables, (ii) ordinary least squares and major axis regression (major-axis regressions are best suited to evaluate the performance of methodologies with uncertainty of similar magnitude), (iii) regressions weighted by the inverse of the data uncertainty (so that sites with larger uncertainty have a lower weight than sites with smaller uncertainty). To verify that errors are not systematically related to environmental and ecosystem properties, the differences between modeled and ground data will be related to a large set of site variables (e.g. MAP, MAT, LAI, biomes types). To ensure that the impact of each given site is independent of the flux magnitude, these analyses will be performed using both absolute [e.g. for GPP: $(GPP_M - GPP_O)$] and relative difference between the modelled and observed data [e.g. for GPP: $(GPP_M - GPP_O)/((GPP_M + GPP_O)/2)$]. In addition to these tests, the performance in predicting seasonal GPP (e.g. daily, monthly GPP) can be verified using the correlation coefficient (R^2), efficiency (E), root mean square error (RMSE) and bias as previously done in Balzarolo et al 2014. This last approach was also used in geoland2 GMES Land Monitoring Core Service for the validation of the Land Surface Models against in-situ carbon fluxes. Furthermore, the model capability in predicting seasonal changes in GPP could be tested by applying a spike detection method as proposed in Vicca et al 2016.

5.3. BENCHMARK METHOD TO OTHER DATA SETS

5.3.1. REFERENCE DATA SETS

At global scale a comparison will be done with two operational EO GPP/NPP products:

- **C-GLOPS1 Dry Matter Production (DMP) product**

This DMP is the agromic equivalent of NPP, calculated at 10-daily time steps and expressed in kg DM/ha/day. The DMP is a product generated by a Monteith Light Use Efficiency (LUE) model first implemented by Veroustraete (1994) but modified and improved in the MARSOP and Copernicus Global Land Service project to run on SPOT-VGT/PROBA-V imagery and ECMWF meteorological data. Within C-GLOPS, the product has been extensively validated against in-situ, MODIS and other modelled datasets of GPP and NPP. As the DMP is a VITO in-house product, we are well familiar with its algorithm and quality. The ATBD and validation report of the DMP are available through the C-GLOPS website http://land.copernicus.eu. The current online version 1 is based on 1 km SPOT-VGT and PROBA-V fAPAR data derived from the MARSOP project. In the meantime, a second version of the DMP is in development and has been validated. Besides a number of algorithmic changes, this version is based on an improved dataset of SPOT-VGT and PROBA-V fAPAR. This product will be made publically available in the end of May 2017, but currently only in "demonstration" mode. Hence, both products will be to compare against the P-model output.

- **MODIS GPP (MOD17A2) and NPP (MOD17A3)**

These products are described in detail by Running et al. (1999), Heinsch et al. (2003), Zhao et al. (2005) is also a variant of the satellite based Monteith approach. The GPP, available at 8-daily timesteps, is reduced at yearly base with the maintenance and growth respiration linking daily biomass and annual growth of plant tissues to the satellite-derived estimates of leaf area index (LAI) to generate yearly NPP maps. The MYD17A3H Version 6 product, recently released by MODIS, provides information about annual (yearly) Net Primary Production at 500 meter pixel resolution. Annual NPP is derived from the sum of the 45, 8-day Net Photosynthesis (PSN) products (MYD17A2H) from the given year. The PSN value is the difference of the GPP and the Maintenance Respiration (MR) (GPP-MR). The MODIS products can be accessed and downloaded through the Google Earth Engine platform.

5.3.2. METHODS

The methods for the benchmarking of the different EO-derived vegetation production data sets are based on guidelines, protocols and metrics defined by the Land Product Validation (LPV) group of the Committee on Earth Observation Satellite (CEOS) for the validation of satellite-derived land products. The following aspects will be evaluated.

(1) *Product completeness*: the missing values or pixels flagged as invalid over land were quantified, overall and over different biomes. An aggregated version of the ESA CCI land cover map will used for this purpose.
(2) Spatial consistency analysis:
    - *Spatial distribution of the GPP/NPP values*: global maps of metrics expressing the similarity and difference between different global GPP/NPP time series will be computed. The metrics include the Root Mean Squared Error (RMSE).

- *Magnitude of the retrievals*: global yearly averages will be calculated for GPP and NPP.

(3) Global statistical analysis:

- *Histograms of bias*: histograms of residuals between products.
- *Distribution per biome type*: statistical distributions of GPP/NPP values and residuals, computed over biomes for the different data sets. An aggregated version of the ESA CCI land cover map will used for this purpose.
- *Global statistics*: Scatterplots between the different datasets will be produced at a global scale and per biome. Metrics (e.g. coefficient of determination, agreement coefficient, orthogonal regression) among different data sets are computed per biome.

(4) Temporal consistency analysis

- *Temporal variation*: Statistical metrics among different data sets are computed per scene to evaluate the time evolution of the metrics.

REFERENCES

Ainsworth, E.A. and S.P. Long (2005) What have we learned from 15 years of free-air CO_2 enrichment? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO_2. *New Phytologist* **165**: 351–372.

Anav, A., P. Friedlingstein, C. Beer, P. Ciais, A. Harper, C. Jones, G. Murray-Tortarolo, D. Papale, N.C. Parazoo and P. Peylin (2015) Spatiotemporal patterns of terrestrial gross primary production: a review. *Reviews of Geophysics* **53**: 785–818.

Atkin, O.K., K.J. Bloomfield, P.B. Reich, M.G. Tjoelker, G.P. Asner, D. Bonal, G. Bönisch, M. Bradford, L.A. Cernusak, E.G. Cosio, D. Creek, K.Y. Crous, T. Domingues, J.S. Dukes, J.J.G. Egerton, J.R. Evans, G.D. Farquhar, N.M. Fyllas, P.P.G. Gauthirt, E. Gloor, T.E. Gimeno, K.L. Griffin, R. Guerrieri, M.A. Heskel, C. Huntingford, F.Y. Ishida, J. Kattge, H. Lambers, M.J. Liddell, J. Lloyd, C.H. Lusk, R.E. Martin, A.P. Maksimov, T.C. Maximov, Y. Malhi, B.E. Medlyn, P. Meir, L.M. Mercado, N. Mirotchnik, D. Ng, Ü. Niinemets, O.S. O'Sullivan, O.L. Philips, L. Poorter, P. Poot, I.C. Prentice, N. Salinas, L.M. Rowland, M.G. Ryan, S. Sitch, M. Slot, N.G. Smith, M.H. Turnbull, M.C. VanderWel, F. Valladares, E.J. Veneklaas, L.K. Weerasinghe, C. Wirth, I.J. Wright, K. Wythers, J. Xiang, S. Xiang and J. Zaragoza-Castells (2015) Global variability in leaf respiration in relation to climate, plant functional types and leaf traits. *New Phytologist* **206**: 614–636.

Bernacchi, C.J., E.L. Singsaas, C. Pimentel, A.R. Portis Jr and S.P. Long (2001) Improved temperature response functions for models of Rubisco-limited photosynthesis. *Plant, Cell and Environment* **24**: 253–259.

Campiolo, M. S. Vicca, S. Luyssaert, J. Bilcke, E. Ceschia, F.S. Chapin III, P. Ciais, M. Fernández-Martínez, Y. Malhi, M. Obersteiner, D. Olefeldt, D. Papale, S.L. Piao, J. Peñuelas, P.F. Sullivan, X. Wang, T. Zenonen and I.A. Janssens (2015) Biomass production efficiency controlled by management in temperate and boreal ecosystems. *Nature Geoscience* **8**: 843–846.

De Kauwe, M.G., T.F. Keenan, B.E. Medlyn, I.C. Prentice and C. Terrer (2016a) Satellite based estimates underestimate the effect of CO_2 fertilization on NPP. *Nature Climate Change* **6**: 892–893.

De Kauwe, M.G., Y.-S. Lin, I.J. Wright, B.E. Medlyn, K.Y. Crous, D.S. Ellsworth, V. Maire, I.C. Prentice, O.K. Atkin, A. Rogers, Ü. Niinemets, S. Serbin, P. Meir, J. Uddling, H.F. Togashi, L. Tarvainen, L.K. Weerasinghe, B.J. Evans, F.Y. Ishida and T. F. Domingues (2016b) A test of a "one-point method" for estimating maximum carboxylation capacity. *New Phytologist* **210**: 1130–1144.

Dewar, R.C. (1996) The correlation between plant growth and intercepted radiation: an interpretation in terms of optimal plant nitrogen content. *Annals of Botany* **78**: 125–136.

Donohue, R.J., I.H. Hume, M.L. Roderick, T.R. McVicar, J. Beringer, L.B. Hutley, J.C. Gallant, J.M. Austin, E. can Gorsel, J.R. Cleverly, W.S. Meyer and S.K. Arndt (2014) Evaluation of the remote-sensing-based DIFFUSE model for estimating photosynthesis of vegetation. *Remote Sensing of Environment* **155**: 349–365.

Donohue, R.J., M.L. Roderick, T.R. McVicar and G.D. Farquhar (2013) Impact of CO_2 fertilization on maximum foliage cover across the globe's warm, arid environments. *Geophysical Research Letters* **40**: 1–5.

Farquhar, G.D., S. von Caemmerer and J. Berry (1980) A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. *Planta* **149**: 78–90.

Garbulsky, M.F., J. Peñuelas, D. Papale, J. Ardö, M.L. Goulden, G. Kiely, A.D. Richardson, E. Rotenberg, E.M. Veenendaal and I. Filella (2010) Patterns and controls of the variability of radiation use efficiency and primary productivity across terrestrial ecosystems. *Global Ecology and Biogeography* **19**: 253–267.

Gill, A.L. and A.C. Finzi (2016) Belowground carbon flux links biogeochemical cycles and resource-use efficiency at the global scale. *Ecology Letters* **19**: 1419–1428.

Goetz, S.J., S.D. Prince, S.N. Goward, M.M. Thawley and J. Small (1999) Satellite remote sensing of primary production: an improved production efficiency modeling approach. *Ecological Modelling* **122**: 239–255.

Graven, H.D., R.F. Keeling, S.C. Piper, P.K. Patra, B.B. Stephens, S.C. Wofsy, L.R. Welp, C. Sweeney, P.P. Tans, J.J. Kelley, B.C. Daube, E.A. Kort, G.W. Santoni and J.D. Bent (2013) Enhanced seasonal exchange of CO_2 by northern ecosystems since 1960. *Science* **341**: 1085–1089.

Haxeltine, A. and I.C. Prentice (1996) A general model for the light use efficiency of primary production. *Functional Ecology* **10**: 551–561.

Hermida-Carrera, C., M.V. Kapralov and J. Galmés (2016) Rubisco catalytic properties and temperature response in crops. *Plant Physiology*: doi:http://dx.doi.org/10.1104/pp.16.01846.

Kattge, J. and W. Knorr (2007) Temperature acclimation in a biochemical model of photosynthesis: a reanalysis of data from 36 species. *Plant Cell and Environment* **30:** 1176–1190.

Keenan, T.F., I.C. Prentice, J.G. Canadell, C. Williams, H. Wang, M.R. Raupach and G.J. Collatz (2016) Recent pause in the growth rate of atmospheric CO_2 due to enhanced terrestrial uptake. *Nature Communications* **7**: 13428.

Knorr, W. and M. Heimann (1995) Impact of drought stress and other factors on seasonal land biosphere CO_2 exchange studied through an atmospheric tracer transport model. *Tellus B* **47**: 471–489.

Landsberg, J.J. and R.H. Waring (1997) A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. *Forest Ecology and Management* **95**: 209–228.

Lin, X., B. Chen, J. Chen, H. Zhang, S. Sun, G. Xu, L. Guo, M. Ge, J. Qu, L. Li and Y. Kong (2017) Seasonal fluctuations of photosynthetic parameters for light use efficiency models and the impacts on gross primary production estimation. *Agricultural and Forest Meteorology* **236**: 22–35.

Lin, Y.-S., B.E. Medlyn, R.A. Duursma, I.C. Prentice, H. Wang, D. Eamus, C.V.M. Barton, J. Bennie, D. Bonal, A. Bosc, M.S.J. Broadmeadow, L.A. Cernusak, P. De Angelis, A.C. Lola da Costa, J.E. Drake, D.S. Ellsworth, M. Freeman, O. Ghannoum, T.E. Gimeno, Q. Han, K. Hikosaka, L.B. Hutley, J.W. Kelly, K. Kikuzawa, P. Kolari, K. Koyama, J.-M. Limousin, M.L. Linderson, M. Löw, C. Macinnis-Ng, N.K. Martin-StPaul, P. Meir, T.N. Mikkelsen, P. Mitchell, J.B. Nippert, T.W. Ocheltree, Y. Onoda, M. Op de Beeck, V. Resco de Dios, A. Rey, A. Rogers, L. Rowland, N. Salinas, S.A. Setterfield, W. Sun, L. Tarvainen, S. Tausz-Posch, D.T. Tissue, J. Uddling, G. Wallin, J.M. Warren, L. Wingate and J. Zaragoza-Castells (2015) Optimal stomatal behaviour around the world: synthesis of a global stomatal conductance database. *Nature Climate Change* **5**: 459–464.

Maire, V., P. Martre, J. Kattge, F. Gastal, G. Esser, S. Fontaine and J.-F. Soussana (2012) The coordination of leaf photosynthesis links C and N fluxes in C_3 plant species. *PLOS One* **7**: e38345.

McCallum, I., O. Franklin, E. Moltchanova, L. Merbold, C. Schmullius, A. Shvidenko, D. Schepaschenko and S. Fritz (2013) Improved light and temperature responses for light-use-efficiency-based GPP models. *Biogeosciences* **10**: 6577–6590.

McCallum, I., W. Wagner, C. Schmullius, A. Shvidenko, M. Obersteiner, S. Fritz and S. Nilsson (2009) Satellite-based terrestrial production efficiency modeling. *Carbon Balance and Management* **4**: 8.

Medlyn, B.E. (1998) Physiological basis of the light use efficiency model. *Tree Physiology* **18**: 167–176.

Medlyn, B.E. (2011) Comment on "Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009". *Science* **333**: 1093.

Medlyn, B.E., R.A. Duursma, D. Eamus, D.S. Ellsworth, I.C. Prentice, C.V.M. Barton, K.Y. Crous, P. De Angelis, M. Freeman and L. Wingate (2011) Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Global Change Biology* **17**: 2134–2144. [Corrigendum: *ibid.* **18**: 3476.]

Michaletz, S.T., D. Cheng, A.J. Kerkhoff and B.J. Enquist (2014) Convergence of terrestrial plant production across global climate gradients. *Nature* **512**: 39–43.

Michaletz, S.T., M.D. Weiser, N.G. McDowell, J. Zhou, M. Kaspari, B.R. Helliker and B.J. Enquist (2016) The energetic and carbon economic origins of leaf thermoregulation. *Nature Plants* doi: 10.1038/NPLANTS.2016.129.

Monteith, J. (1972) Solar radiation and productivity in tropical ecosystems. *Journal of Applied Ecology* **19**: 747–766.

Monteith, J.L. (1977). Climate and the efficiency of crop production in Britain. *Philosophical Transactions of the Royal Society of London* **281**: 277–294.

Ogutu, B.O. and J. Dash (2013) An algorithm to derive the fraction of photosynthetically active radiation absorbed by photosynthetic elements of the canopy ($FAPAR_{ps}$) from eddy covariance flux tower data. *New Phytologist* **197**: 511–523.

Ogutu, B.O., J. Dash and T.P. Dawson (2013) Developing a diagnostic model for estimating terrestrial vegetation gross primary productivity using the photosynthetic quantum yield and Earth Observation data. *Global Change Biology* **19**: 2878–2892.

Potter, C.S., J.T. Randerson, C.B. Field, P.A. Matson, P.M. Vitousek, H.A. Mooney and S.A. Klooster (1993) Terrestrial ecosystem production: a process model based on global satellite and surface data. *Global Biogeochemical Cycles* **7**: 811–841.

Prentice, I.C. (2013) Ecosystem science for a changing world. *Grantham Institute for Climate Change, Discussion Paper* **4**, 16 pp.

Prentice, I.C., N. Dong, S.M. Gleason, V. Maire and I.J. Wright (2014) Balancing the costs of carbon gain and water loss: testing a new quantitative framework for plant functional ecology. *Ecology Letters* **17**: 82–91.

Prentice, I.C., X. Liang, B. Medlyn and Y. Wang (2015) Reliable, robust and realistic: the three R's of next-generation land-surface modelling. *Atmospheric Chemistry and Physics* **15**: 5987–6005.

Prince, S.D. and S.N. Goward (1995) Global primary production: a remote sensing approach. *Journal of Biogeography* **22**: 815–835.

Running, S.W. and E.R. Hunt (1993) Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models. In: J.R. Ehleringer and C.B. Field (eds) *Scaling Physiological Processes: Leaf to Globe*, Academic Press, New York, 141–158.

Running, S., F. Nemani, M. Heinsch, M. Zhao, M. Reeves and H. Hashimoto (2004) A continuous satellite-derived measure of global terrestrial primary production. *BioScience* **54**: 547–560.

Ryu, Y., D.D. Baldocchi, H. Kobayashi, C. van Ingen, J. Li, T.A. Black, J. Beringer, E. van Gorsel, A. Knohl, B.E. Law and O. Roupsard (2011) Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales. *Global Biogeochemical Cycles* **25**: GB4017.

Samanta, A., M.H. Costa, E.L. Nunes, S.A. Vieira, L. Xu and R.B. Myneni (2011) Comment on "Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009". *Science* **333**: 1093.

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

Skillman, J.B. (2008) Quantum yield variation across the three pathways of photosynthesis: not yet out of the dark. *Journal of Experimental Botany* **59**: 1647–1661.

Smith, W.K., S.C. Reed, C.C. Cleveland, A.P. Ballantyne, W.R.L. Anderegg, W.R. Wieder, Y.Y. Liu and S.W. Running (2016) Large divergence of satellite and Earth system model estimates of global terrestrial CO_2 fertilization. *Nature Climate Change* **6**: 306–310.

Stocker, B.D., F. Joos, R. Roth, R. Spahni, L. Bouwman, S. Zaehle, Xu-Ri and I.C. Prentice (2013) Multiple greenhouse gas feedbacks from the land biosphere under future climate change scenarios. *Nature Climate Change* **3**: 666–672.

Swinnen, E., R. van Hoolst and H. Eerens (2015) Algorithm Theoretical Basis Document: Dry Matter Productivity (DMP). Copernicus, 37 pp.

Tagesson, T., J. Ardö, B. Cappelaere, L. Kergoat, A. Abdi, S. Horion and R. Fensholt (2017) Modelling spatial and temporal dynamics of gross primary production in the Sahel from earth-observation-based photosynthetic capacity and quantum efficiency. *Biogeosciences* **14**: 1333–1348.

Tang, X., H. Li, N. Huang, X. Li, X. Xu, Z. Ding and J. Xie (2015) A comprehensive evaluation of MODIS-derived GPP for forest ecosystems using the site-based FLUXNET database. *Environmental Earth Sciences* **74**: 5907z–5918.

Thomas, R.B., I.C. Prentice, H. Graven, P. Ciais, J.B. Fisher, M. Huang, D.N. Huntzinger, A. Ito, A. Jacobson, A. Jain, J. Mao, A. Michalak, S. Peng, B. Poulter, D.M. Ricciuto, X. Shi, C. Schwalm, H. Tian and N. Zeng (2016) Increased light-use efficiency in northern terrestrial ecosystems indicated by CO_2 and greening observations. *Geophysical Research Letters* **43**: 11339–11349.

Ukkola, A.M., I.C. Prentice, T.F. Keenan, A.I.J.M. van Dijk, N.R. Viney, R.B. Myneni and J. Bi (2015) Reduced streamflow in water-stressed climates consistent with CO_2 effects on vegetation. *Nature Climate Change* **6**: 75-78.

Verma, M., M.A. Friedl, A.D. Richardson, G. Kiely, A. Cescatti, B.E. Law, G. Wohlfahrt, B. Gielen, O. Roupsard, E.J. Moors, P. Toscano, F.P. Vaccari, D. Gianelle, G. Bohrer, A. Varlagin, N. Buchmann, E. van Gorsel, L. Montagnani and P. Propastin (2014) Remote sensing of annual terrestrial gross primary productivity from MODIS: an assessment using the FLUXNET La Thuile data set. *Biogeosciences* **11**: 2185–2200.

Veroustraete, F. (1994) On the use of a simple deciduous forest model for the interpretation of climate change effects at the level of carbon dynamics. *Ecological Modelling* **75**/**76**: 221–237.

Veroustraete, F., H. Sabbe and H. Eerens (2002) Estimation of carbon mass fluxes over Europe using the C-Fix model and Euroflux data. *Remote Sensing of Environment* **83**: 377–400.

Vicca, S., S. Luyssaert, J. Peñuelas, M. Campioli, F.S. Chapin III, P. Ciais, A. Heinemeyer, P. Högberg, W.L. Kutsch, B.E. Law, Y. Malhi, D. Papale, S.L. Piao, M. Reichstein, E.D. Schulze and I.A. Janssens (2012) Fertile forests produce biomass more efficiently. *Ecology Letters* **15**: 520–526.

Wang, H., I.C Prentice, W.M. Cornwell, T.F. Keenan, T.W. Davis, I.J. Wright, B.J. Evans and C. Peng (2016a) A universal model for carbon dioxide uptake by plants. *bioRxiv* http://dx.doi.org/10.1101/040246 (preprint; revised version in revision for *Nature Plants*).

Wang, H., I.C. Prentice, T.W. Davis, T.F. Keenan, I.J. Wright and C. Peng (2016b) Photosynthetic responses to altitude: an explanation based on optimality principles. *New Phytologist* **213**: 976–982.

Wenzel, S., P.M. Cox, V. Eyring and P. Friedlingstein (2014) Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 Earth system models. *Journal of Geophysical Research* **119**: 794–807.

Wright, I.J., P.B. Reich and M. Westoby (2003) Least-cost input mixtures of water and nitrogen for photosynthesis. *The American Naturalist* **161**: 98–111.

Yebra, M., A.I.J.M. van Dijk, R. Leuning and J.P. Guerschman (2015) Global vegetation gross primary production estimation using satellite-derived light-use efficiency and canopy conductance, *Remote Sensing of Environment* **163**: 206–216.

Yuan, W., W. Cai, J. Xia, J. Chen, S. Liu, W. Dong, L. Merbold, B. Law, A. Arain, J. Beringer, C. Bernhofer, A. Black, P.D. Blanken, A. Cescatti, Y. Chen, D. Gianelle, I.A. Janssens, M. Jung, T. Kato, G. Kiely, D. Liu, B. Marcolla, L. Montagnani, A. Raschi, O. Roupsard, A. Varlagin and G. Wohlfahrt (2007) Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. *Agricultural and Forest Meteorology* **143**: 189–207.

Yuan, Y., S. Liu, G. Yu, J.-M. Bonnefond, J. Chen K. Davis, A.R. Desai, A.H. Goldstein, D. Gianelle, F. Rossi, A.E. Suyker and S.B. Verma (2010) Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data. *Remote Sensing of Environment* **114**: 1416–1431.

Yuan, W.,W. Cai, J. Xia, J. Chen, S. Liu, W. Dong, L. Merbold, B. Law, A. Arain, J. Beringer, C. Bernhofer, A. Black, P.D. Blanken, A. Cescatti, Y. Chen, L. François, D. Gianelle, I.A. Janssens, M. Jung, T. Kato, G. Kiely, D. Liu, B. Marcolla, L. Montagnani, A. Raschi, O. Roupsard, A. Varlagin and G. Wohlfahrt (2014) Global comparison of light use efficiency models for simulating terrestrial vegetation gross primary production based on the LaThuile database. *Agricultural and Forest Meteorology* **192**: 108–120.

Zhao, M., F.A. Heinsch, R.R. Nemani and S.W. Running (2005) Improvements of the MODIS terrestrial gross and net primary production data set. *Remote Sensing of Environment* **95**: 164–176.

Zhao, M. and S.W. Running (2010) Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science* **329**: 940–943.

Zhou, S., R. Duursma, B.E. Medlyn, J.W.G. Kelley and I.C. Prentice (2013) How should we model plant responses to drought? An analysis of stomatal and non-stomatal responses to water stress. *Agricultural and Forest Meteorology* **182–183**, 204–214.

Zhu, X.-G., S.P. Long and D.R. Ort (2010) Improving photosynthetic efficiency for greater yield. *Annual Review of Plant Biology* **61**: 235-261.

Zhu, Z. S. Piao, R.B. Myneni, M. Huang, Z. Zeng, J.G. Canadell, P. Ciais, S. Sitch, P. Friedlingstein, A. Arneth, C. Cao, L. Cheng, E. Kato, C. Koven, Y. Li, X. Lian, Y. Liu, R. Liu, J. Mao, Y. Pan, S. Peng, J. Peñuelas, B. Poulter, T.A.M. Pugh, B.D. Stocker, N. Viovy, X. Wang, Y. Wang, Z. Xiao, H. Yang, S. Zaehle and N. Zeng (2016) Greening of the Earth and its drivers. *Nature Climate Change* **6**: 791–795.

ANNEX A: PROTOTYPE CODE FOR APPLICATION WITH MERIS GVI AND METEO DATA

The model runs by using the following scripts:

1. Start_model
2. Main_model
3. Run_site
4. Run_pmod_site
5. Calc_gpp
6. Calc_m

The first four scripts control the flow of data inputs and outputs, while the final two perform the calculations. For each site, daily vpd and temperature, monthly fAPAR and annual CO_2 are used to calculate GPP. The Python codes to calculate GPP and *m* in equation (3), and the constants used for the calculations, are shown below.

```
---------------------------------- GPP calc ----------------------------------

######## Import modules#########

import numpy as np
from constants import c_star
from m_calc import M_CALC

###############
# Main program
###############

class GPP_CALC:
    """
    Name: GPP_CALC
    Features: Calculates GPP for given data
    """
    def __init__(self):
            """
            Name:     GPP_CALC.__init__
            Input:    None.
            Features: Initialise class variables

            """
            self.gpp = None
            self.old_frac = None
            self.gpp_old_eq = None

        #       ### Class Function Definitions ####

    def run_grid(self, ppfd, this_fapar, temp, aCO2, alpha, elv, beta, phi_0, absG, tMax = False, tMin
= False, vap = False, vpd = False):
            """
```

```
        Name: GPP_CALC.run_grid
        Input: - float, ambient CO2 (aCO2) ppm
               - float, faparn/a
               - float, temperature (temp) C
               - float, elevation (elv) m
               - float, monhtly maximum daily temp (tMax) C
               - float, monthly minimum daily tamperature (tMin) C
               - float, vapour pressure (vap) Pa
               - float, vapour pressure defict (vpd) Pa This is data input if running at site scale, but
calcuated at global
        Outputs: - float, GPP
                 - float, C13 discrimination
        Depends: - SPLASH --> EVAP-->alpha
                 - SPLASH --> PPFD

        """

        if tMax is not False:
          this_tMax = tMax
        else:
           this_tMax = False

        if tMin is not False:
          this_tMin = tMin
        else:
           this_tMin = False

        if vap is not False:
          this_vap = vap
        else:
           this_vap = False

        if vpd is not False:
          this_vpd = vpd
        else:
           this_vpd = False



        my_M=M_CALC()
        my_M.run_grid(temp,aCO2,elv,beta,this_tMax,this_tMin,this_vap,this_vpd)
        self.M = my_M.m
        self.GAMMA_ST = my_M.GSTAR
        self.K1_ = my_M.K_1
        self.ETA_ST = my_M.ESTAR
        self.vpd_Pa_ = my_M.vpd_Pa
        self.m_frac = np.sqrt(1.0 - ((c_star / self.M) ** ( 2.0 / 3.0 )))


        # GPP is zero for monthly mean temperatures below 0C
```

```
        self.lue = phi_0 *\
                self.M * \
                self.m_frac *\
                (alpha ** (1.0/4.0))

        if type(temp) ==np.float64:
           if temp < 0.0:
              self.lue = 0.0
        else:
            neg_t = np.where(temp < 0.0)
            self.lue[neg_t] = 0.0

        self.gpp = self.lue * \
                absG * \
                ppfd * \
                this_fapar
```

---------- M calc ----------

```
import numpy as np

import os
import sys

CURRENT_DIR = os.path.abspath(os.curdir)
PARENT_DIR = os.path.abspath(CURRENT_DIR + "/../")

evap_path = os.path.join(PARENT_DIR, 'SPLASH/working/py_version') # the SPLASH code location
sys.path.append(evap_path)


from evap import EVAP

from constants import gamma_25, t_25, Ha, R, eta_const, kco, kPo

class M_CALC:
   """
   Name: M_CALC
```

```
    Features: Calculates M for GPP
    """

    def __init__(self):
        """
        Name:     M_CALC.__init__
        Input:    - float, latitude, degrees (lat)
                  - float, latitude, degrees (lon)

        """
        self.m = None
        self.vpd_Pa = None

      #### Class Function dEfinitions ####

    def run_grid(self, temp, aCO2, elv, beta, tMax, tMin, vap, vpd):
        """
        Name: M_CALC.run
        Input: - float, ambient CO2 (aCO2)
        Outputs: - float, M (for GPP claucation)
        Depends: - GAMMA_STAR
                 - ETA_STAR
                 - VPD_CALC
                 - K1_CALC
        """
        evap = EVAP(25.0, elv)
        self.patm = evap.elv2pres(elv)


          pp_CO2 = aCO2 * 1E-6 * self.patm #* 101.325 * 1000 # CO2 in Pa--> partial pressure for
average atmospheroc pressure


        if type(temp) == np.float64:
           kelv2cel = 273.15
        else:
           kelv2cel = np.tile(273.15, (pp_CO2.shape[0], pp_CO2.shape[1]))


        # Calculate VPD is not given as data input
        if vpd is False:
            this_vpd = self.calc_vpd(tMax, tMin, vap)
        else:
            this_vpd = vpd


        k_temp = temp + kelv2cel  #temp in kelvin
        k_temp25 = 25.0 + kelv2cel


        m_calc = (pp_CO2 - self.gamma_star(k_temp)) / \
```

```
                (pp_CO2 + (2 * self.gamma_star(k_temp)) + \
                (3 * self.gamma_star(k_temp)) * \
                np.sqrt(1.6 * self.eta_star(temp, self.patm, 25.0, kPo, elv) * \
                this_vpd * \
                (beta ** -1.0) * \
                ((self.k1(k_temp, elv, self.patm) + self.gamma_star(k_temp)) ** -1.0)))


        self.vpd_Pa = this_vpd
        self.m = m_calc
        return m_calc

    def gamma_star(self,k_temp):
        """
        Name: GAMMA_STAR.run
        Input: -Float, Air temp (Tk) Given in degC, need kelvin
        Output: - Float, Gamma*
        """
        ttg=((k_temp - t_25) * Ha) / (R * k_temp * t_25)
        gs=gamma_25 * np.exp(ttg)
        self.GSTAR = gs

        return gs

    def eta_star(self,temp, patm, temp25, kPo, elv):
        """
        Name: ETA_STAR.run
        Input: -Float, Air temp (Tk)
        Output: - Float, ETA*
        """
        # Calculated as eta/eta25
        # viscosity correction factor = viscosity( temp, press )/viscosity( 25 degC, 1013.25 Pa)

        #es = np.exp(eta_const * (k_temp - t_25))

        evap = EVAP(25.0, elv)

        self.ns = evap.viscosity_h2o(temp, patm)
        self.ns25 = evap.viscosity_h2o(temp25, kPo)

        es = self.ns/self.ns25
        self.ESTAR = es
#         print es
        return es

    def calc_vpd(self,tMax, tMin, vap):
        """
        Name: VPD_CALC.run
        Input: -Float, monlty average daily maximum temp (Tmax) deg C
            -Float, monlty average daily minimum temp (Tmin) deg C
            -Float, monlty average vapour pressure temp (vap) hPa (1 Pa = 0.01 hPa)
```

```
        Output: - Float, D (jn equation)
    """

    tMaxMin=(8.635 * (tMax + tMin))/(0.5 * (tMax + tMin) + 237.3)

    vpd_out=(0.611 * np.exp(tMaxMin) - (0.10 * vap)) * 1000 # kPa to Pa (units for m are all Pa)

    return vpd_out

def k1(self, k_temp, elv, patm):
    """
    Name:     K0_CALC.__init__
    Input:    - float, latitude, degrees (lat)
              - float, latitude, degrees (lon)
              -Float, monlty average temperature (Tk)
              -Float, elevation at grid square (elv)
              Output: - Float, K1
    Depends: - Kc
             - P0
             - K0
    """
    k1_out = self.kc(k_temp) * (1 + ( self.p0(elv, self.patm) / self.k0(k_temp)))
    self.K_1 = k1_out
    return k1_out

def kc(self,k_temp):

    """
    Name: k1_calc.kc
    Input: -Float, monlty average temperature (Tk)
         - Const = 79.430 kJ mol-1 = energy of activation for carboxylation
    Output: - Float, Kc
    """

    dhatKc = 79430 #J/mol
    kc25 = 39.97 # Pa at 25 deg C and 98.716KPa

    tempFrac=((k_temp - t_25) * dhatKc) / (R * k_temp * t_25)

    Kc_out = kc25 * np.exp(tempFrac)

    return Kc_out

def p0(self,elv, patm):
    """
    Name: k1_calc.po
    Input: -Float, elevation at grid point
    Output: - Float, P0. O2 partial pressure
    """

    #P0_out=21000  * np.exp( - 0.114 * (elv * 1E-3))
```

```
        P0_out = kco * (1e-6) * patm

        return P0_out

    def k0(self,k_temp):
        """
        Name: K0_CALC.run
        Input: -Float, monlty average temperature (Tk)
            - Const = 36.380 kJ mol-1 = energy of activation for oxygenation
        Output: - Float, K1
        """

        dhatK0 = 36380 # J/mol
        k025 = 27480 # Pa, at 25 deg C and 98.716KPa

        tempFracK0=((k_temp - t_25) * dhatK0) / (R * k_temp * t_25)

        K0_out=k025 * np.exp(tempFracK0)

        return K0_out

    def print_vals(self):

        print " m: %0.10f" % (self.m)



---------------------------------- constants ----------------------------------


gamma_25 = 4.220 # Pa  Gamma* at 25C
t_25 = 298.15 # K 25C in Kelvin
Ha = 37830.0 #J/mol Activiation energy for Gamma*
R = 8.3145 # J/mol/K Universal gas constant
eta_const = -0.0227 # No unit  Constant for viscosity of water relative to its value at 25C (See Wang 2015)
kco =  2.09476e5 #ppm. US standard pressure. (From Beni's code - Ref Bernacchi et al 2001)
kPo = 101325.0  # Standard atmopsheric pressure (Pa), Allen 1973

Ho = 36.38 # Bernacchi 2001 energy of activation for oxygenation
Hc = 79.43 # Bernacchi 2001 energy of activation for carboxylation

c_star = 0.41 # unit of carbon cost for maintenance of electron transport capacity (obs Jmax:Vc max)
a_hat = 4.4 # Standard value for c13 discrimination - diffusion component (Wang 2015 Eq s42)
b_hat = 27.0 # Standard value for c13 discrimination - biochemical component (Wang 2015 Eq s42)

c_molmass =  12.0107  # g C / mol C

C3_phi0 = 0.084 * c_molmass # # mol/mol # gC/mol Intrinisic quantum yield of photosynthesis for C3 plants
```

C4_phi0 = 0.055 * c_molmass # # mol/mol # gC/mol Intrinisic quantum yield of photosynthesis for C4 plants

---

Additional functions from the SPLASH model (doi:10.5194/gmd-10-689-2017, https://bitbucket.org/labprentice/splash) are used in the calculation of *m* (with constants listed below).

→To convert CO2 form ppm to Pa, the following function is used:
Evap.elv2pres

```
def elv2pres(self, z):
    """
    Name:     EVAP.elv2pres
    Input:    float, elevation above sea level (z), m
    Output:   float, atmospheric pressure, Pa
    Features: Calculates atm. pressure for a given elevation
    Depends:  Global constants
              - kPo
              - kTo
              - kL
              - kMa
              - kG
              - kR
    Ref:      Allen et al. (1998)
    """
    self.logger.debug("estimating atmospheric pressure at %f m", numpy.nanmean(z))
    p = kPo*(1.0 - kL*z/kTo)**(kG*kMa/(kR*kL))
return p
```

→In eta_star, the following function is used:
Evap.viscosity_h20

```
def viscosity_h2o(self, tc, p):
    """
    Name:     LUE.viscosity_h2o
    Input:    - float, ambient temperature (tc), degrees C
              - float, ambient pressure (p), Pa
    Return:   float, viscosity of water (mu), Pa s
    Features: Calculates viscosity of water at a given temperature and
              pressure.
    Depends:  density_h2o
    Ref:      Huber, M. L., R. A. Perkins, A. Laesecke, D. G. Friend, J. V.
              Sengers, M. J. Assael, ..., K. Miyagawa (2009) New
              international formulation for the viscosity of H2O, J. Phys.
              Chem. Ref. Data, Vol. 38(2), pp. 101-125.
    """
    # Define reference temperature, density, and pressure values:
    tk_ast = 647.096      # Kelvin
    rho_ast = 322.0       # kg/m^3
```

```
    mu_ast = (1e-6)      # Pa s

    # Get the density of water, kg/m^3
    rho = self.density_h2o(tc, p)

    # Calculate dimensionless parameters:
    tbar = (tc + 273.15)/tk_ast
    tbarx = tbar**(0.5)
    tbar2 = tbar**2
    tbar3 = tbar**3
    rbar = rho/rho_ast

    # Calculate mu0 (Eq. 11 & Table 2, Huber et al., 2009):
    mu0 = 1.67752
    mu0 += 2.20462/tbar
    mu0 += 0.6366564/tbar2
    mu0 += -0.241605/tbar3
    mu0 = 1e2*tbarx/mu0

    # Create Table 3, Huber et al. (2009):
    hj0 = (0.520094, 0.0850895, -1.08374, -0.289555, 0., 0.)
    hj1 = (0.222531, 0.999115, 1.88797, 1.26613, 0., 0.120573)
    hj2 = (-0.281378, -0.906851, -0.772479, -0.489837, -0.257040, 0.)
    hj3 = (0.161913,  0.257399, 0., 0., 0., 0.)
    hj4 = (-0.0325372, 0., 0., 0.0698452, 0., 0.)
    hj5 = (0., 0., 0., 0., 0.00872102, 0.)
    hj6 = (0., 0., 0., -0.00435673, 0., -0.000593264)
    h = hj0 + hj1 + hj2 + hj3 + hj4 + hj5 + hj6
    h_array = numpy.reshape(numpy.array(h), (7, 6))

    # Calculate mu1 (Eq. 12 & Table 3, Huber et al., 2009):
    mu1 = 0.
    ctbar = (1./tbar) - 1.
    for i in range(6):
       coef1 = numpy.power(ctbar, i)
       coef2 = 0.
       for j in range(7):
          coef2 += h_array[j][i]*numpy.power((rbar - 1.), j)
       mu1 += coef1*coef2
    mu1 = numpy.exp(rbar*mu1)

    # Calculate mu_bar (Eq. 2, Huber et al., 2009)
    #   assumes mu2 = 1
    mu_bar = mu0*mu1

    # Calculate mu (Eq. 1, Huber et al., 2009)
    mu = mu_bar*mu_ast    # Pa s

    return mu
```

---

where desnsity_h20:

```
def density_h2o(self, tc, p):
    """
    Name:     EVAP.density_h2o
    Input:    - float, air temperature (tc), degrees C
              - float, atmospheric pressure (p), Pa
    Output:   float, density of water, kg/m^3
    Features: Calculates density of water at a given temperature and
              pressure
    Ref:      Chen et al. (1977)
    """
    # self.logger.debug(
    #     ("calculating density of water at temperature %f Celcius and "
    #      "pressure %f Pa") % (tc, p))

    # Calculate density at 1 atm (kg/m^3):
    po = 0.99983952
    po += (6.788260e-5)*tc
    po += -(9.08659e-6)*tc*tc
    po += (1.022130e-7)*tc*tc*tc
    po += -(1.35439e-9)*tc*tc*tc*tc
    po += (1.471150e-11)*tc*tc*tc*tc*tc
    po += -(1.11663e-13)*tc*tc*tc*tc*tc*tc
    po += (5.044070e-16)*tc*tc*tc*tc*tc*tc*tc
    po += -(1.00659e-18)*tc*tc*tc*tc*tc*tc*tc*tc
    # self.logger.debug("water density at 1 atm calculated as %f kg/m^3", po)

    # Calculate bulk modulus at 1 atm (bar):
    ko = 19652.17
    ko += 148.1830*tc
    ko += -2.29995*tc*tc
    ko += 0.01281*tc*tc*tc
    ko += -(4.91564e-5)*tc*tc*tc*tc
    ko += (1.035530e-7)*tc*tc*tc*tc*tc
    # self.logger.debug("bulk modulus at 1 atm calculated as %f bar", ko)

    # Calculate temperature dependent coefficients:
    ca = 3.26138
    ca += (5.223e-4)*tc
    ca += (1.324e-4)*tc*tc
    ca += -(7.655e-7)*tc*tc*tc
    ca += (8.584e-10)*tc*tc*tc*tc
    # self.logger.debug("temperature coef, Ca, calculated as %f", ca)

    cb = (7.2061e-5)
    cb += -(5.8948e-6)*tc
    cb += (8.69900e-8)*tc*tc
    cb += -(1.0100e-9)*tc*tc*tc
    cb += (4.3220e-12)*tc*tc*tc*tc
    # self.logger.debug("temperature coef, Cb, calculated as %f bar^-1", cb)

    # Convert atmospheric pressure to bar (1 bar = 100000 Pa)
```

```
    pbar = (1.0e-5)*p
    # self.logger.debug("atmospheric pressure calculated as %f bar", pbar)

    pw = (ko + ca*pbar + cb*pbar**2.0)
    pw /= (ko + ca*pbar + cb*pbar**2.0 - pbar)
    pw *= (1e3)*po
    return pw
```

→ constants

```
kG = 9.80665     # gravitational acceleration, m/s^2 (Allen, 1973)
kL = 0.0065      # temperature lapse rate, K/m (Allen, 1973)
kMa = 0.028963   # molecular weight of dry air, kg/mol (Tsilingiris, 2008)
kPo = 101325     # standard atmosphere, Pa (Allen, 1973)
kR = 8.31447     # universal gas constant, J/mol/K (Moldover et al., 1988)
kTo = 298.15     # base temperature, K (Berberan-Santos et al., 1997)
```
