

# Sensitivity Analysis for the Estimation of Gross Primary Production Using Remote Sensing Data

Nicolas Weaver

FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS AND  
COMPUTER SCIENCE  
Master of Science  
in  
Applied Mathematics

TU Delft Supervisor and Chair:

Deltares Supervisor:

Deltares Supervisor and Thesis Committee Member:

Thesis Committee Member:

Prof.dr.ir. Martin Verlaan

Anna Spinosa

Dr.ir. Ghada El Serafy

Dr. Neil Budko



Delft University of Technology

September, 2024

## Abstract

Gross Primary Production (GPP) is the amount of carbon dioxide ( $\text{CO}_2$ ) that is fixed by an ecosystem through photosynthesis. It is a key variable to understand the carbon cycle and biodiversity conservation. In-situ measurements of GPP are possible at a local scale. To estimate GPP on a global scale, models based on remote sensing data are used. One of them is the Penman-Montheith-Leuning (PML) model, which is constructed on physical processes.

In this thesis, a sensitivity analysis of the PML model is performed for the case study of Torgnon subalpine grassland. It is defined around three aspects: the application of the PML model to Torgnon subalpine grassland, the impact of the input data on the output data, and the reliability of the calibration procedure. We found out that the PML model can be applied to the study site, but the value of the parameters presented in the literature are not optimal ( $R^2 = 0.57$  using the literature calibration and  $R^2 = 0.94$  in the optimised case).

Moreover, the model is extremely sensitive to the Leaf Area Index (LAI), and, to a lesser extent, to temperature and Photosynthetically Active Radiations (PAR). However, it is weakly dependent on the  $\text{CO}_2$  concentration of the atmosphere. The analysis of the calibration procedure showed that a simplified PML model, that does not consider the  $\text{CO}_2$  concentration and Vapour Pressure Deficit (VPD), performs similarly to the initial model. An ensemble estimation of GPP is obtained by using an ensemble of optimal parameters. It suggests that errors in the calibration principally impact the spring and summer estimations.

## Acknowledgement

I would like to thank all of the people that helped me going through this thesis.

First of all, I would to thank Martin Verlaan for being my university supervisor and chair of my thesis committee. His regular advice and reading of drafts helped me to go smoothly in my thesis. I am also grateful to the other thesis committee members Neil Budko and Ghada El Serafy for taking their time going through my thesis.

I would also like to think the people from Deltares for accepting me in their structure. In particular, Anna Spinoso for welcoming me and introducing me to primary production and remote sensing topics, Ghada El Serafy, Mario Alberto Fuentes Monjaraz and Valeria Mobilia for guiding me when Anna had to take a step back. Also, I want to thank the interns at Deltares with whom I enjoyed sharing many lunches, coffee breaks, and YD drinks.

I am thankful to all of my friends that I saw again and all the new ones that I met during this Thesis. The people from that awesome ultimate Frisbee club Force Elektro, the anacleto from the Masters, and those français aux Pays-Bas with whom I shared many moments. I can't name you all, but you all count for me.

Lastly, I wish to thank my family and my friends that are in France. Even if I have not had the time to see you a lot during this thesis because of the distance, I know that you will always be there for me and that we can always count on each other.

# Contents

<b>1</b>	<b>Introduction</b>	<b>11</b>
<b>2</b>	<b>The modelling of GPP</b>	<b>13</b>
2.1	The Carbon Cycle . . . . .	13
2.2	The PML model . . . . .	13
2.3	Other models . . . . .	15
2.3.1	Light Use Efficiency (LUE) models . . . . .	15
2.3.2	Statistical models . . . . .	16
2.3.3	Artificial Intelligence . . . . .	17
<b>3</b>	<b>Mathematical methods</b>	<b>18</b>
3.1	Methods for data processing . . . . .	18
3.1.1	Whittaker Smoother . . . . .	18
3.2	Methods for the sensitivity of input variables . . . . .	19
3.2.1	Pearson’s correlation . . . . .	19
3.2.2	Rank correlation . . . . .	19
3.2.3	One variable at the time procedure . . . . .	20
3.3	Methods for the calibration procedure . . . . .	20
3.3.1	Metrics . . . . .	20
3.3.2	Grid search . . . . .	21
3.3.3	Cross-validation . . . . .	21
3.4	Ensemble forecasting with parameter uncertainties . . . . .	22
<b>4</b>	<b>Data availability and processing</b>	<b>23</b>
4.1	Remote Sensing Data . . . . .	23
4.1.1	The MODIS satellite characteristics . . . . .	23
4.1.2	The Global Land Data Assimilation System (GLDAS) dataset . . . . .	23
4.1.3	Available Products in Google Earth Engine . . . . .	24
4.2	In-situ measurements . . . . .	24
4.3	Data processing . . . . .	25
4.3.1	Climatic variables . . . . .	25
4.3.2	CO <sub>2</sub> atmospheric concentration . . . . .	26
4.3.3	Photosynthetically Active Radiations (PAR) . . . . .	27
4.3.4	Leaf Area Index . . . . .	27
<b>5</b>	<b>Results and analysis of the PML model</b>	<b>29</b>
5.1	Estimation using parameters from literature . . . . .	29
5.2	Correlation analysis . . . . .	29
5.3	One at the time procedure and asymptotic analysis of parameters . . . . .	31
5.3.1	Temperature and VPD constraint . . . . .	31
5.3.2	GPP response to PAR and CO <sub>2</sub> atmospheric concentration . . . . .	32
5.3.3	GPP response to LAI . . . . .	33
5.4	Calibration of parameters and uncertainties . . . . .	34
5.4.1	One-dimensional equations . . . . .	35
5.4.2	Two-dimensional equations . . . . .	38

5.4.3	Three-dimensional equations with $\eta \neq \infty$ . . . . .	40
5.4.4	Three-dimensional equations with $\eta \rightarrow \infty$ . . . . .	42
5.4.5	Reset of the VPD constraint . . . . .	46
5.5	Ensemble estimation and test set evaluation . . . . .	47
5.5.1	Evaluation on the test set . . . . .	48
5.5.2	Distribution of the optimised parameters . . . . .	48
5.5.3	Ensemble estimation . . . . .	50
<b>6</b>	<b>Discussion</b> . . . . .	<b>53</b>
6.1	Uncertainties in the methodology . . . . .	53
6.1.1	Uncertainties in situ measurements . . . . .	53
6.1.2	Uncertainties in input variables . . . . .	53
6.1.3	Differences between self implemented PML model in GEE product . . . . .	54
6.2	Diversity of data . . . . .	54
6.2.1	Temporal diversity of data . . . . .	54
6.2.2	Spatial diversity of data . . . . .	55
<b>7</b>	<b>Conclusion</b> . . . . .	<b>57</b>
<b>A</b>	<b>MOD17 functions</b> . . . . .	<b>58</b>
<b>B</b>	<b>Remote Sensing Data</b> . . . . .	<b>59</b>
<b>C</b>	<b>Spearman's correlation matrices</b> . . . . .	<b>61</b>
<b>D</b>	<b>Computation of the limits and asymptotic equations</b> . . . . .	<b>62</b>
<b>E</b>	<b>Other RMSE figures</b> . . . . .	<b>67</b>

## Acronyms

**APAR** Absorbed Photosynthetically Active Radiations

**ATP** adenosine triphosphate

**CLr** Red-edge Index

**CO<sub>2</sub>** carbon dioxide

**EBV** Essential Biodiversity Variables

**EC** Eddy Covariance

**EPPEs** Ensemble Prediction and Parameter Estimation System

**ET** evapotranspiration

**EV** Environmental Variables

**EVI** Enhanced Vegetation Index

**EVI2** Two-bands Enhanced Vegetation Index

**GEE** Google Earth Engine

**GLDAS** Global Land Data Assimilation System

**GPP** Gross Primary Production

**GSFC** Goddard Space Flight Center

**LAI** Leaf Area Index

**LUE** Light Use Efficiency

**MODIS** Moderate Resolution Imaging Spectroradiometer

**NADPH** nicotinamide adenine dinucleotide phosphate

**NASA** National Aeronautics and Space Administration

**NCEP** National Centers for Environmental Prediction

**NDVI** Normalised Difference Vegetation Index

**NEE** Net Ecosystem Exchange

**NIR<sub>v</sub>** Near-Infrared Reflectance of terrestrial vegetation

**NOAA** National Oceanic and Atmospheric Administration

**NPP** Net Primary Production

**OVAT** One-Variable-at-a-Time

**PAR** Photosynthetically Active Radiations

**PML** Penman-Montheith-Leuning

**RMSE** Root Mean Square Error

**RNN** Recurrent Neural Networks

**RS** Remote Sensing

**SPPT** Stochastically Perturbed Parameterization Tendencies

**SW** Short-Wave

**VI** Vegetation Index

**VPD** Vapour Pressure Deficit

## List of Figures

1	Smoothed temperature and VPD time series. The blue and red curves respectively correspond to the in-situ and remote sensing measurements. . . . .	26
2	CO <sub>2</sub> concentration from the NOAA dataset linearly interpolated to daily values. . .	26
3	Time series of PAR expressed as a photon flux density in $\mu mol m^{-2} day^{-1}$ from the GLDAS-2.1 dataset . . . . .	27
4	Time series of LAI. The blue curve is the MCD15 unprocessed product. The orange curve is the smoothed LAI. The red dots are the totally filtered outliers (weight set to 0). . . . .	28
5	Time series of the PML model using parameter values founded in literature. The black curve is the reference data. The blue and green curves are the output of the PML model using the methodology describe in section 4.3. They respectively use the in-situ measurements, and the remote sensing estimation of temperature and VPD. The red curve is the product available in GEE. The purple curve is the MOD17 product. . . . .	30
6	Correlation matrix of the variables and GPP. . . . .	31
7	Correlation matrix over the seasons. The color bar is the same as in figure 6. . . . .	32
8	Temperature and VPD constraints. The values of the parameters for the visualisation of both functions are from [26] (temperature) and [23]. . . . .	33
9	Evolution of GPP with respects to specific variables and their associated parameter. The red curve represents the asymptote when the variable or the parameter goes to infinity. The green curve parameter value is the one from literature. The black vertical lines delimit the interval that contains the dataset values. . . . .	34
10	Evolution of GPP with respect to LAI and for several values of $k_q$ . The red curve represents the asymptote when the variable or the parameter goes to zero. The green curve parameter value is the one from literature. The black and gray vertical lines delimit the interval that contains the raw and smoothed dataset values respectively. . . . .	35
11	RMSE in the cases when there is only one limiting factor. The red dot corresponds to the minimum of the function. . . . .	37
12	Visualisation of RMSE proprieties in the case where light availability is not a limiting factor. Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to $V_{m,25}$ in function of $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\eta$ in function of $V_{m,25}$ . . . . .	39
13	Visualisation of RMSE proprieties in case the case where the catalytic capacity of RuBisCO is not a limiting factor. Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\beta$ in function of $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\eta$ in function of $\beta$ . . . . .	41
14	Visualisation of RMSE proprieties in the case where the catalytic capacity of RuBisCO, the CO <sub>2</sub> and the light availabilities are limiting factors. The extinction coefficient of PAR is fixed at $k_q = 0.72$ . Each curve represents a different value of $\beta$ . Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to $V_{m,25}$ in function of $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\eta$ in function of $V_{m,25}$ . . . . .	42

15	Visualisation of RMSE proprieties in the case where the catalytic capacity of Ru-BisCO, the CO <sub>2</sub> and the light availabilities are limiting factors. The extinction coefficient of PAR is fixed at $k_q = 0.72$ . Each curve represents a different value of $V_{m,25}$ . Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\beta$ in function of $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\eta$ in function of $\beta$ . . . . .	43
16	Visualisation of RMSE proprieties in the case where the CO <sub>2</sub> availability is not a limiting factor. Each curve represents a different value of $k_q$ . Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to $V_{m,25}$ in function of $\beta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\beta$ in function of $V_{m,25}$ . . . . .	44
17	Visualisation of RMSE proprieties in the case where the catalytic capacity of Ru-BisCO, the CO <sub>2</sub> and the light availabilities are limiting factors. The extinction coefficient of PAR is fixed at $k_q = 0.72$ . Each curve represents a different value of $V_{m,25}$ . Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\beta$ in function of $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\eta$ in function of $\beta$ . . . . .	45
18	Visualisation of RMSE proprieties in the case where the catalytic capacity of Ru-BisCO, the CO <sub>2</sub> and the light availabilities are limiting factors. The extinction coefficient of PAR is fixed at $k_q = 0.72$ . Each curve represents a different value of $V_{m,25}$ . Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\beta$ in function of $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\eta$ in function of $\beta$ . . . . .	46
19	Visualisation of the mean value of the tested VPD constraints time-series, depending on the minimal threshold $VPD_{min}$ , and the angle of the slope. . . . .	47
20	Visualisation of the VPD constraint with the general shape of the function (20a), and it's value over time (20b). . . . .	48
21	RMSE on the test set for the 55195 lowest RMSE scores on the training set. The pale triangle points are the results when the optimal VPD constraint is applied. They separated in two obvious clusters (blue and pink). The round dark points are the evaluation but without any constraint. The navy blue and the red correspond respectively to the first and second cluster. . . . .	49
22	Time series of the estimated GPP. The black curve is the in-situ reference. The blue and red curves are the ensemble means of the optimised PML model respectively without and with the VPD constraint. . . . .	49
23	Violin plots of the parameter values in the ensemble. The blue and the pink violins are the parameter values in the first and second cluster respectively. The red and green horizontal lines are respectively the literature value and the optimal value in the train set. . . . .	50
24	Ensemble time-series of the estimated GPP. The black curve is the in-situ reference. The pink curves each represent one prediction of the ensemble. The blue curve is the ensemble mean of the optimised PML model without the VPD constraint. . . . .	51

25	Ensemble’s standard deviation time series. The blue curve is the value of the standard deviation over time. The red curve is the average standard deviation value. . . . .	51
26	Time series of the PML model when the training data comes from different season. The black curve is the in situ data. The blue, orange, green, and red curves respectively represent training in Autumn, Winter, Spring and Summer. . . . .	55
27	Spatial patterns of CO <sub>2</sub> at four different dates. Screenshots of a NASA video ( <a href="https://svs.gsfc.nasa.gov/12072">https://svs.gsfc.nasa.gov/12072</a> ) . . . . .	56
28	Ramp functions of minimal daily air temperature ( <i>TMIN</i> ) and Vapour Pressure Deficit ( <i>VPD</i> ) (picture from [88]) . . . . .	58
29	Spectral Bands of Sentinel-2 with their resolution [66]. . . . .	59
30	Modis bands (picture from [89]) . . . . .	60
31	Spearman’s correlation matrix of the variables and GPP. The color bar is the same as in figure 6. . . . .	61
32	Spearman’s correlation matrix over the seasons. The color bar is the same as in figure 6. . . . .	62
33	Visualisation of RMSE proprieties in case the case where the catalytic capacity of RuBisCO is not a limiting factor. Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\beta$ in function of $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\eta$ in function of $\beta$ . . . . .	67
34	Visualisation of RMSE proprieties in the case where the catalytic capacity of RuBisCO, the CO <sub>2</sub> and the light availabilities are limiting factors. The extinction coefficient of PAR is fixed at $k_q = 0.72$ . Each curve represents a different value of $\eta$ . Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to $V_{m,25}$ in function of $\beta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to $\beta$ in function of $V_{m,25}$ . . . . .	68

## List of Tables

1	List of the parameters and variables in the PML model with their units. <i>ppm</i> is parts-per-million. . . . .	15
2	Comparative table between MODIS and Sentinel-2 characteristics. . . . .	23
3	Datasets available in the Earth Engine Data Catalog [71]. Dates are in the DD-MM-YYYY format. . . . .	24
4	Value of the parameters in Torgnon grassland according to literature. . . . .	29
5	RMSE and Pearson’s $R^2$ of the GEE products and the self-implemented PML model using parameter values from literature. . . . .	30
6	Critical parameter values . . . . .	37
7	Grid proprieties for the evaluation of the RMSE for $\beta = \infty$ having $\eta_c$ and $V_{m,25_c}$ respectively being the critical values for $\eta$ and $V_{m,25}$ . . . . .	38
8	Grid proprieties for the evaluation of the RMSE for $V_{m,25} = \infty$ , $k_q = 0.72$ having $\beta_c$ and $\eta_c$ respectively being the critical values for $\beta$ and $\eta$ . . . . .	39
9	Grid proprieties for the evaluation of the RMSE for $k_q = 0.72$ having a small number of evaluations for $\beta$ . . . . .	40
10	Grid proprieties for the evaluation of the RMSE for $k_q = 0.72$ having a small number of evaluations for $V_{m,25}$ . . . . .	40
11	Grid proprieties for the evaluation of the RMSE for $\eta = \infty$ having $\beta_c$ , and $V_{m,25_c}$ respectively being the critical values for $\beta$ , and $V_{m,25}$ . The results of this configuration is illustrated in figure 16. . . . .	44
12	Grid proprieties for the evaluation of the RMSE for $\eta = \infty$ having $\beta_c$ , and $V_{m,25_c}$ respectively being the critical values for $\beta$ , and $V_{m,25}$ . The results of this configuration is illustrated in figure 17. . . . .	45
13	Grid proprieties for the evaluation of the RMSE for $\eta = \infty$ having $\beta_c$ , and $V_{m,25_c}$ respectively being the critical values for $\beta$ , and $V_{m,25}$ . The results of this configuration is illustrated in figure 18. . . . .	46
14	Grid proprieties for the evaluation of the RMSE using seasonal training . . . . .	54
15	Optimal parameters with RMSEs on the test and train set using seasonal training . . . . .	55
16	Biome-Property-Look-Up-Table (BPLUT) for MODIS GPP/NPP algorithm with NCEP-DOE reanalysis II and the Collection5 FPAR/LAI as inputs. The full names for the University of Maryland land cover classification system (UMD_VEG_LC) in MOD12Q1 dataset (fieldname: Land_Cover_Type_2) are, Evergreen Needleleaf Forest (ENF), Evergreen Broadleaf Forest (EBF), Deciduous Needleleaf Forest (DNF), Deciduous Broadleaf Forest (DBF), Mixed forests (MF), Closed Shrublands (CShrub), Open Shrublands (OShrub), Woody Savannas (WSavanna), Savannas (Savanna), Grassland (Grass), and Croplands (Crop) [88]. . . . .	58

# 1 Introduction

Ecosystems are sources of many benefits for species, including humans. They provide essential resources like freshwater, food, and clean air. They also act as natural buffers against hazards such as floods and landslides [1]. To measure the health and changes of ecosystems, Essential Biodiversity Variables (EBV) have been created. They cover many aspects of environmental sciences, such as genetic composition, species populations and traits, community composition, and ecosystems structure and functions [2].

Gross Primary Production (GPP) is defined as the total mass of carbon that an ecosystem produces in a given surface and length of time through photosynthesis [3]. It is usually expressed in grams of carbon per square meter per day ( $\text{gC}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ ). It is a key indicator for measuring the dynamics of carbon dioxide ( $\text{CO}_2$ ) variations. Therefore, the assessment of GPP leads to a better understanding of the evolution of the environment [4]. Due to its importance, it is considered as an ecosystem and species physiology EBV [5].

However, GPP can only be directly measured at the leaf level scale [6]. At the canopy scale, the estimation of GPP is done by measuring the Net Ecosystem Exchange (NEE) using the Eddy Covariance (EC) technique. NEE is the net  $\text{CO}_2$  flux from an ecosystem to the atmosphere. The concentrations of  $\text{CO}_2$  and water are measured near EC towers with a high frequency. From those values, many algorithms exist to derive GPP near the tower of measurement ([7],[8],[9],[10]). But, the footprint zone that influences the  $\text{CO}_2$  measurements has a typical length scale of only 100m to 2000m around the tower [11].

To estimate GPP on a larger scale — for example continental or global scale — Remote Sensing (RS) data are used. Satellites capture images of the Earth at different wavelength bands, from which we can derive indicators relating on the state of vegetation [12]. Historically, the used RS data in models were captured by the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument. It contains 36 spectral bands having a resolution from 250 m to 1000 m [13]. Therefore, models that uses MODIS data like MOD17, estimates GPP with a 1 km spatial resolution [14].

Different strategies for modeling GPP using RS data exist. For instance, the Light Use Efficiency (LUE) models compute GPP by estimating the Absorbed Photosynthetically Active Radiations (APAR) and the actual LUE. Processed-based models use biophysical non-linear equations to estimate GPP. Statistical models use equations containing a combination of parameters to calibrate. The complexity of the equations can vary from the linear regression to machine learning structures. The input RS data of such models can either be Vegetation Index (VI), that is transformation of image bands representing vegetation properties; or physical vegetation variables - like the Leaf Area Index (LAI) - that are estimated using satellite images. Moreover, Environmental Variables (EV) are conjointly used with RS information to derive GPP [15].

However, the parameter values, assumptions, and the choice of relevant variables of any model are subject to change and error [16]. A sensitivity analysis is "The study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input" [17]. With this, modelers can determine which inputs are the most significant for the output, and what the consequences are if the input gets changed [18]. Therefore, sensitivity analyses have to be performed to fully understand the solution of a model [19]. For instance, in the case of LUE models, Pei *et al.* relate that the input variables for the representation of environmental stress factors differ from one model to another [15].

The aim of this study is to perform a sensitivity analysis on the estimation of GPP. Among the many GPP models that utilises RS data, processed-based models have the most complex mathe-

mathematical process chain. We can consequently identify which intermediate variables and mechanisms are the most subject to changes in input data. Therefore, this study will focus on the Penman-Monteith-Leuning (PML) model, a commonly used processed based model.

There are two types of input in the PML model: input variables and input parameters. Input variables, like temperature, are provided directly by datasets. Therefore, a first objective is to quantify how the input data affect the model's output data. Input parameters have to be determined through a calibration procedure. They represent biophysical proprieties of the plants. Hence, we will also study the reliability of this calibration procedure. This will lead to an ensemble prediction of GPP by evaluating the model with a set of adjusted parameters.

To conduct this study, we will implement the PML model in the Torgnon subalpine grassland in northern Italy. The values of parameters will initially correspond with the ones founded in literature.

Thus, the research questions to tackle are:

1. Can we apply the PML model to Torgnon subalpine grassland?
2. How does input data affect the model's output?
3. Is the calibration procedure reliable? Can similar low objective function scores predict similar outputs?

Indeed, the PML model is applied and calibrated globally, where the parameters are computed per vegetation type (grassland, savannas, wetlands...). However, the study site is a special mountainous grassland. Moreover, in literature, the influence of variables is not explicitly studied, and the calibration procedure is only performed, not analysed.

This thesis is structured as follow:

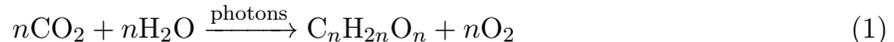
- Chapter 2 presents how GPP is modelled in literature,
- Chapter 3 explains the specific mathematical tools used in this thesis,
- Chapter 4 introduces the used data and it's processing,
- Chapter 5 displays the results obtained,
- Chapter 6 discusses those results, drawing limits and perspectives,
- Chapter 7 concludes this study.

## 2 The modelling of GPP

### 2.1 The Carbon Cycle

Plants grow using the photosynthesis process. It is a chemical reaction that uses the energy from photons to happen. Therefore, the light energy is converted into chemical energy. Only the radiations in the 400nm-700nm spectral band can be used by plants for photosynthesis. The energy of those radiations is called Photosynthetically Active Radiation (PAR).

A simple representation of the chemical equation is:



Hence, GPP corresponds to the amount of carbon that is on the left-hand side of this chemical equation.

The full process of photosynthesis is complex, but can be divided into two phases: the light dependent reactions and the light independent reactions. In the light dependent phase, the reaction uses the energy of photons to create some intermediate chemical components: adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide phosphate (NADPH). The light independent phase consists of a cycle that transforms carbon dioxide into carbohydrates. The capture of  $\text{CO}_2$  is possible thank to the RuBisCO enzyme. It is a catalyst of the photosynthesis reaction. Its regeneration is possible thank to the ATP and NADPH produced in the light independent phase [20].

### 2.2 The PML model

Several processed-based models exist to estimate GPP. They describe the exchanges processes between the atmosphere and the vegetation. Among those models, we will work with the Penman-Montheith-Leuning (PML) model that is described below. It estimates simultaneously evapotranspiration (ET) and GPP. There are two versions of the PML model: PML-V1 developed by Zhang *et al.* [21] and PML-V2 created by Gan *et al.* [22], and improved by Zhang *et al.* [23]. For our study, only the latest version is considered.

The gross assimilation rate of carbon  $A_g$  is calculated using Thornley's model that considers it as a rectangular hyperbola function of PAR and the atmospheric  $\text{CO}_2$  concentration  $C_a$  [24]:

$$A_g = \frac{A_m \beta \eta C_a I}{A_m \beta I + \beta \eta C_a I + \eta A_m C_a} \quad (2)$$

where  $A_m$  is the maximum photosynthetic rate at saturation of PAR and  $\text{CO}_2$ ,  $\beta$  is the initial slope of the light response curve to assimilation rate (also know as quantum efficiency),  $\eta$  is the carboxylation efficiency, and  $I$  is the PAR expressed as a Photosynthetic Photon Flux Density ( $\mu\text{mol m}^{-2}\text{s}^{-1}$ ). By dividing by  $\beta I \eta C_a$  we obtain:

$$A_g = \frac{A_m}{1 + A_m \left( \frac{1}{\beta I} + \frac{1}{\eta C_a} \right)} \quad (3)$$

With the formulation in equation 3, we can see more clearly that this model is built around a maximum production restricted by limiting factors. Here, the limiting factors are the quantity of available  $\text{CO}_2$  to capture and the amount of available light.

We approximate  $A_m$  by the maximum catalytic capacity of RuBisCO per unit leaf area  $V_m$  [25]. This value mainly depends on temperature  $T$  and is expressed as:

$$A_m = V_m = V_{m,25} \frac{\exp(a(T - 25))}{1 + \exp(b(T - 41))} \quad (4)$$

where  $V_{m,25}$  is the value of  $V_m$  at 25°C, and  $a = 0.031$  and  $b = 0.115$  are temperature coefficients [26]. Those parameters will be considered fixed, and will not be studied.

The intensity of light vanishes through the vegetation depth, as leaves create shadow. This vanishing process is modelled using Beer-Lambert law:

$$I = I_0 \exp(-k_q l) \quad (5)$$

having  $I_0$  the flux density at the top of the canopy,  $k_q$  is the extinction coefficient, and  $l$  is the unit leaf area from top to the bottom of the canopy. It is the ratio of the one side green leaf area over the ground surface area, on an infinitesimal surface.

These equations correspond to the values at the local leaf scale. Therefore, to get the gross assimilation rate at the canopy scale  $A_{c,g}$ , we integrate  $A_g$  over the leaf area index (LAI). The LAI is the ratio of the one green leaf area over the full canopy surface. It quantifies the amount of vegetation. Thus, we obtain:

$$A_{c,g} = \int_0^{LAI} A_g dl \quad (6)$$

$$= \int_0^{LAI} \frac{V_m \beta \eta C_a I_0 \exp(-k_q l)}{(V_m \beta + \beta \eta C_a) I_0 \exp(-k_q l) + \eta V_m C_a} dl \quad (7)$$

$$A_{c,g} = \frac{V_m \eta C_a}{k_q (V_m + \eta C_a)} \left[ k_q LAI + \ln \left( \frac{V_m \beta I_0 + \eta C_a \beta I_0 + V_m \eta C_a}{V_m \beta I_0 + \eta C_a \beta I_0 + V_m \eta C_a \exp(k_q LAI)} \right) \right] \quad (8)$$

Moreover, Zhang *et al.* implemented a constraint depending on the Vapour Pressure Deficit (VPD) [23]. VPD is defined as the difference between water vapour pressure at saturation and the observed water vapour pressure. Therefore, the closer it is to zero, the more humid the atmosphere is. This constraint takes the form of a ramp function:

$$f(VPD) = \begin{cases} 1 & \text{if } VPD < VPD_{min} \\ \frac{VPD_{max} - VPD}{VPD_{max} - VPD_{min}} & \text{if } VPD_{min} < VPD < VPD_{max} \\ 0 & \text{if } VPD_{max} < VPD \end{cases} \quad (9)$$

In other words, if  $VPD < VPD_{min}$ , then the assimilation is not restricted because air is humid enough. On the contrary, if  $VPD > VPD_{max}$ , the assimilation is no longer possible because the air is too dry. Therefore,  $VPD_{min}$  and  $VPD_{max}$  are respectively the minimum and maximum thresholds for the assimilation constraint.

Thus, the expression of GPP is:

$$GPP = f(VPD) A_{c,g} \quad (10)$$

The complete process chain also includes equations related to the estimation of ET. They are not described here but are available in the appendix of [23]. The list of the different parameters and variables are available in table 1. We distinguish 4 categories: parameters to calibrate, fixed parameters, RS variables and environmental variables from meteorological data.

Parameter or Variable	Definition	Unit	Estimation Method
$\beta$	Quantum efficiency	$\mu\text{mol CO}_2 (\mu\text{mol PAR})^{-1}$	Calibration
$\eta$	Carboxylation efficiency	$\mu\text{mol m}^{-2}\text{s}^{-1} (\text{ppm})^{-1}$	Calibration
$k_q$	Extinction coefficient of PAR	-	Calibration
$V_{m,25}$	Maximum catalytic capacity of RuBisCO per unit leaf area at 25 °C	$\mu\text{mol m}^{-2}\text{s}^{-1}$	Calibration
$VPD_{min}$	Minimum VPD value for assimilation constraint	Pa	Calibration
$VPD_{max}$	Maximum VPD value for assimilation	Pa	Calibration
$a, b$	Temperature coefficients	-	Fixed
$LAI$	Leaf Area Index	$\text{m}^{-2}/\text{m}^{-2}$	Remote sensing
$I_0$	Photosynthetically Active Radiations (PAR)	$\mu\text{mol m}^{-2}\text{s}^{-1}$	Meteorological data
$C_a$	Concentration of CO <sub>2</sub> in the atmosphere	<i>ppm</i>	Meteorological data
$VPD$	Vapour Pressure Deficit	Pa	Meteorological data
$T$	Temperature of the atmosphere	°C	Meteorological data

Table 1: List of the parameters and variables in the PML model with their units. *ppm* is parts-per-million.

To sum up, the PML model relies on three concepts: There is a maximum photosynthesis rate per vegetation unit, this rate is limited by the meteorological context (PAR, CO<sub>2</sub>, temperature, VPD), and the amount of available light vanishes with the amount of vegetation. This last concept introduces non-linearity in the model.

## 2.3 Other models

Processed based models are not the only way to estimate GPP. Other categories exist such as Light Use Efficiency (LUE) models, simple statistical model and artificial intelligence.

### 2.3.1 Light Use Efficiency (LUE) models

LUE models compute GPP as the product of the Absorbed PAR (APAR) and the actual LUE. Not all of the PAR is actually absorbed, so we introduce the Fraction of PAR (fPAR) as the ratio of APAR and PAR. In LUE models, this last variable is modeled by VI's. The actual LUE is

expressed as the maximum LUE multiplied by constraint functions depending on Environmental Variables (EV) such as air temperature or humidity ([15]). Thus, LUE models are characterised by the equations:

$$GPP = APAR \times \varepsilon_g \quad (11)$$

$$APAR = PAR \times fPAR \quad (12)$$

$$\varepsilon_g = \varepsilon_{max} \times f(EV) \quad (13)$$

Having  $\varepsilon_g$  being the actual LUE,  $\varepsilon_{max}$  the theoretical maximum LUE,  $EV$  the list of environmental variables used in a specific model. The chosen function and the list of EVs are specific to every models [15]. The maximum LUE and the function depend on parameters that are constant per vegetation type.

**MOD17** Among the LUE models, the MOD17 model is one of the most widely used to estimate GPP [14]. For this one, fPAR is approximated by the popular Normalised Difference Vegetation Index (NDVI) [27], and the EVs are minimal air daily temperature  $T$ , and the Vapour Pressure Deficit  $VPD$ . The EV functions are ramp functions. We therefore have to define two thresholds per function:  $EV_{min}$  and  $EV_{max}$ . Consequently, the MOD17 equation is:

$$GPP = PAR \times NDVI \times \varepsilon_{max} \times f(T) \times f(VPD) \quad (14)$$

$$f(T) = \begin{cases} 0 & \text{if } T < T_{min} \\ \frac{T - T_{min}}{T_{max} - T_{min}} & \text{if } T_{min} < T < T_{max} \\ 1 & \text{if } T_{max} < T \end{cases} \quad (15)$$

$$f(VPD) = \begin{cases} 1 & \text{if } VPD < VPD_{min} \\ \frac{VPD_{max} - VPD}{VPD_{max} - VPD_{min}} & \text{if } VPD_{min} < VPD < VPD_{max} \\ 0 & \text{if } VPD_{max} < VPD \end{cases} \quad (16)$$

The variables  $PAR$ ,  $T$ , and  $VPD$  are available in several databases. The variable  $NDVI$  is computed using RS data. The values of  $\varepsilon_{max}$ ,  $T_{min}$ ,  $T_{max}$ ,  $VPD_{min}$  and  $VPD_{max}$  are constant per vegetation type and available in a Look-up Table. The visualisation of the ramp functions and the Look-up Table are present in Appendix A.

### 2.3.2 Statistical models

Statistical models find a strong relationship between VIs, environmental variables (EV) and GPP. A first structure of such models are linear regression. In this case, the model's equation is simply:

$$GPP = a \times VI \times EV + b \quad (17)$$

where  $a$  and  $b$  are parameters to tune. In most cases, the VI that is the most correlated to GPP depends on the study area. For instance, NDVI tends to be related to GPP in low biomass areas [28], [29], whereas, in dense biomass areas, the Enhanced Vegetation Index (EVI) [30] shows strong connection [31], [32]. Moreover, other VIs have been used to predict GPP. For example, Cai *et al.* used a Two-bands Enhanced Vegetation Index (EVI2) [33] in the northern European forests [34]. Bagley *et al.* created the Near-Near-Infrared Reflectance of terrestrial vegetation (NIRv) that

models well GPP [35]. Spinosa *et al.* compared the correlation of several VIs and EVs combination with GPP in a southern Spanish wetland. They concluded that the Red-edge Index (CLr) [36] leads to the best estimation [37].

### **2.3.3 Artificial Intelligence**

More recently, advanced artificial intelligence models have emerged. One of them is the random forest [38]. This method has showed good results in predicting different ecosystem structure EBVs such as soil moisture [39], Net Primary Production (NPP) [40] and GPP [41] [42] [43]. Likewise, due to the temporal dependency of GPP, Recurrent Neural Networks (RNN) may constitute a new generation of models [44], [45], [46].

### 3 Mathematical methods

#### 3.1 Methods for data processing

##### 3.1.1 Whittaker Smoother

The input RS data is noisy, irregular, and with differences in the quality of measurements. The Whittaker smoother ([47]) is a method to filter and smooth time-series, that will be applied on the LAI dataset.

Let  $y$  be the noisy data of size  $N$ , and  $z$  the optimal smoothness of  $y$ . Because not all the measurements are of the same quality, we introduce the weight vector  $w$ . For an order 1 smoother, the residuals  $S$  and smoothed metric  $R$  are respectively defined as:

$$S = \sum_{i=1}^N w_i (y_i - z_i)^2 \quad (18)$$

$$R = \sum_{i=1}^N (z_i - z_{i-1})^2 \quad (19)$$

Using matrix notation we rewrite  $S$  and  $R$  as:

$$S = (y - z)^T W (y - z) \quad (20)$$

$$R = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & 0 & \\ & 0 & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix} z = Dz \quad (21)$$

Where  $W = \text{diag}(w)$ . We therefore introduce the cost function  $Q$  as:

$$Q = S + \lambda R \quad (22)$$

having  $\lambda$  being a smoothness scalar.

The optimal smoothness is founded when  $Q$  is minimal with respect to  $z$ , which yields:

$$\frac{\partial Q}{\partial z} = 0 \quad (23)$$

$$\iff -2W(y - z) + 2\lambda D^T D z = 0 \quad (24)$$

$$\iff (W + \lambda D^T D)z = Wy \quad (25)$$

Equation 25 is of the form  $Ax = b$  and can be solved using matrix decomposition.

The order of the smoother corresponds to the number of previous measurements the metric  $R$  takes into account. It is also the order of the difference method that is used. When the employed order is greater than one, the structure of equation 25 is unchanged, but the matrix  $D$  gets modified. For example, an order 2 smoother yields:

$$R = \sum_{i=1}^N (z_i - 2z_{i-1} + z_{i-2})^2 \quad (26)$$

$$D = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & 0 & \\ & 0 & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{N,N-2} \quad (27)$$

## 3.2 Methods for the sensitivity of input variables

### 3.2.1 Pearson's correlation

Pearson's correlation coefficient, denoted as  $r$ , is a statistical measure that quantifies the linear relationship between two continuous variables [48]. Let  $(X, Y)$  be a pair of variables containing  $N$  samples. The Pearson correlation variable is defined as [49]:

$$r_{xy} = \frac{cov(X, Y)}{\sigma_x \sigma_y} \quad (28)$$

$$= \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (29)$$

Where  $cov(X, Y)$ , is the covariance between  $X$  and  $Y$ ,  $\sigma_x$  and  $\sigma_y$  are respectively the standard deviations of  $X$  and  $Y$ , and  $\bar{x}$  and  $\bar{y}$  are the mean values of  $X$  and  $Y$ .

The value of  $r$  ranges from -1 to 1 where:

- $r = 1$  : Indicates a perfect positive linear relationship,
- $r = -1$  : Indicates a perfect negative linear relationship,
- $r = 0$  : Indicates no linear relationship.

The closer the absolute value is to 1, the more linear the relationship is.

However, Pearson's correlation is limited to linear relationships. Variables can have correlation score close to 0, but still be dependent [50].

### 3.2.2 Rank correlation

Unlike Pearson's correlation that uses the nominal values of variables, a rank correlation uses the ordinal values. Let  $X$  be a time-series. Time-series are originally ordered by time, which, in the general case, is not by value. Let  $R(X)$  be the  $X$  vector sorted by increasing values. Therefore, the rank of  $X_i$  is the index of  $R(X_i)$ . Rank correlation test have the advantage to go beyond the linear relationships, but it does not measure the differences in values within the sample [51].

For instance, the Spearman rank correlation coefficient between  $(X, Y)$  is Pearson's correlation coefficient applied to  $(R(X), R(Y))$  [52]:

$$\rho = r_{R(x)R(y)} = \frac{cov(R(X), R(Y))}{\sigma_{R(x)} \sigma_{R(y)}} \quad (30)$$

Therefore, a Spearman’s score of 1 (respectively -1) is for a monotonous increasing (decreasing) relationship between  $X$  and  $Y$ .

### 3.2.3 One variable at the time procedure

The One-Variable-at-a-Time (OVAT) procedure is used to understand the impact of individual input variables on the output of a model. This method involves shifting one input variable at a time, while keeping all of the other variables constant, by taking their temporal average value. This approach isolates the effect of a single variable on the model’s output [53].

The main limitation of this method is that it does not capture the interactions between variables and should be used with caution [54]. For example, the LAI finds it’s peak in summer, when the daylight (PAR) is also at high values. By only varying the LAI, we do not consider the evolution of PAR that goes with it.

However, this method gives a clear picture of the mathematical contribution of the variable in the model. For instance, it shows how the model can evolve with variable values going beyond the ones in the dataset. Also, in the PML model, each variable is associated to a parameter. By changing the value of this parameter, we can illustrate it’s impact on the estimate across a range of variable values.

Moreover, we can derive asymptotic equations, corresponding to extreme values of the variables and parameters. Since the PML model is based on a maximum assimilation ( $V_m$ ) that can be limited by the availability of certain factors (PAR, CO<sub>2</sub> concentration) and the climate conditions (temperature, VPD), those asymptotic equations represent the model without constraints. Therefore, those asymptotic models reduce the dimensionality of the equation. Consequently, the calibration procedure of the simplified models is performed on a fewer numbers of parameters.

## 3.3 Methods for the calibration procedure

The goal of the calibration procedure is to find the combination of parameters that minimises an objective function. Those objective functions are also called metrics.

### 3.3.1 Metrics

A metric is a function that transforms elements of a metric-space (here the vector space  $\mathbb{R}^n$ ) into  $\mathbb{R}$  [55]. In particular, it is a way of measuring the distance between a reference and a predicted time-series. For this study, two metrics are considered: the Root Mean Square Error (RMSE) and Pearson’s coefficient of determination ( $R^2$ ).

**Root Mean Square Error (RMSE)** The Root Mean Square Error (RMSE) computes the Euclidean (order 2) norm of the difference vector. In the case of GPP, it is computed as:

$$\text{RMSE}(GPP, \widehat{GPP}) = \sqrt{\frac{\sum_{i=1}^N (GPP_i - \widehat{GPP}_i)^2}{N}} \quad (31)$$

Having,

- $GPP_i$  being the reference value of GPP on the  $i^{th}$  day,
- $\widehat{GPP}_i$  being the predicted value of GPP on the  $i^{th}$  day,

- $N$  the sample size.

Because of the square component in the formula, the outliers, that is to say model points that deviate from the reference data, are strongly penalised [56]. The RMSE is expressed in the same units as the variable (here  $\text{gC m}^{-2} \text{ day}^{-1}$ ).

**Pearson’s coefficient of determination** Pearson’s coefficient of determination, also known as the  $R^2$  coefficient, is a metric to determine the linear correlation between variables. It is defined as:

$$R^2(GPP, \widehat{GPP}) = 1 - \frac{\sum_{i=1}^N (GPP_i - \widehat{GPP}_i)^2}{\sum_{i=1}^N (GPP_i - \overline{GPP})^2} \quad (32)$$

Where  $\overline{GPP}$  is the mean value of the reference GPP. Contrary to the RMSE, it is a dimensionless number. Its maximum value is 1 corresponding to a perfect estimate. When the value is negative, then the model performs worse than the average model.

### 3.3.2 Grid search

The grid search is an exhaustive optimisation method that consist in evaluating the objective function on all the points of a grid. For every parameter to optimise, we define a set of finite points that will constitute one dimension of the grid. So, each one of those points is a candidate for being the optimal value of the parameter. We therefore get a finite number of evaluation of the model, for which the objective function can be computed. Thus, the minimum of this finite set can easily be selected.

The main issue of this method is the ”curse of dimensionality”, that is the explosion of computational time and memory usage when the number of dimensions increases [57]. However, the computation can easily be parallelised [58]. In our case, the dimension space to optimise is equal to the number of parameters, that is to say 6 ( $\beta$ ,  $\eta$ ,  $V_{m,25}$ ,  $k_q$ ,  $VPD_{min}$ , and  $VPD_{max}$ ). The parallelisation is performed using the outer product. In the execution of the model, instead of performing  $n$  times a scalar product for  $n$  parameter values, we create a parameter vector of size  $n$ , and use the outer product between the parameter vector and the variable vector. This method allows multiple evaluation of the model at the same time. The process is much faster, but is limited by the available memory usage.

Nevertheless, the grid search method gives a clear picture of the objective function to optimise. Since our aim is to study the reliability of the calibration procedure (and not only it’s execution), this method can lead to a better understanding of the behaviour of the objective function near the global minimum. Indeed, with this exhaustive evaluation of the objective function, we can select all of the combinations of parameters that results in low score. This creates an ensemble of parameters that can be analysed.

To reduce, the useless computational cost and to increase the comprehension of the obtained result, we first decrease the dimensionality of the equation using the methodology describe in section 5.4. Then we increase it by discarding known in advance degenerative values.

### 3.3.3 Cross-validation

Cross-validation is a validation of model technique that allows the generalisation of a model to an independent dataset. It relies on the splitting of the available data into a ”training” and a ”test” set.

The training set is used for the tuning of the parameters. The test set is utilised for the evaluation of the model, once the tuning is done [59].

More specifically, in the case of a grid search, the training phase will consist of drawing and selecting the minimum of the objective function taking into account data from the train set. The testing phase will be the evaluation of the model and the objective function for the selected value of parameters, using the test dataset.

One cross-validation method is the  $k$ -fold cross-validation. It consists in splitting the training data into  $k$  subsets. In turns,  $k - 1$  of those subsets are used for the training and one is left for the testing. This operation is repeated  $k$  times, once every subset has been once the testing subset. This procedure reduces the overfitting of a specific train set [60].

### 3.4 Ensemble forecasting with parameter uncertainties

Ensemble forecasting is a technique used on deterministic models with high uncertainties. For example, this is the case in weather and hydrological forecasting, where the studied systems are complex. Indeed, there are uncertainties in the construction of models that could lead to errors in the output [61]. With the example of forecasting, errors in initial conditions induces mistakes in the long-term predictions. To cope with this issue, an ensemble prediction would apply an ensemble set of perturbations on the initial conditions. The model will then be applied for every perturbed cases [62]. This creates a prediction ensemble.

Another source of uncertainty in modelling concerns the parameters. This is the case we will focus on in this study. Likewise, perturbation can be applied to the parameters, and an ensemble prediction can be made. Many perturbations schemes exist. For instance, the Stochastically Perturbed Parameterization Tendencies (SPPT) uses a multiplicative noise on the parameter’s tendency (i.e. the variation of the parameter over time) [63]. For complex models depending on hundreds of parameters, a bayesian optimisation approach may be taken. The Ensemble Prediction and Parameter Estimation System (EPPES) is one particular method that also predicts uncertainty in the parameter values. We can therefore apply a perturbed parameter scheme to obtain an ensemble estimation [64].

The previous methods that we talked about are based on Monte-Carlo estimations. The approach we use in this study is different. Indeed, the studied model ”only” has six parameters to optimise, who each matches to a different process. Therefore, we will use an exhaustive procedure for looking at the almost optimal parameter values, using the grid search optimisation. The term ”almost” signifies the tolerance of small errors in the retained values. Nevertheless, this method has an exponential computational complexity with respect to the number of dimensions in the grid (number of parameters).

To obtain an ensemble spread, we will select not one combination of parameter values as it is done in classical optimisation, but we will select an ensemble of almost optimal parameter values by evaluating the RMSE in the calibration phase. For this, we will tolerate all combination of parameters that result in a objective function value that is lower than 1% more than the minimal value of the objective function. In a mathematical formulation, we select all parameters  $\tilde{p}$  such that:

$$RMSE(\tilde{p}) \leq 1.01 \times \min_{p \in \text{Grid}} RMSE(p) \quad (33)$$

## 4 Data availability and processing

### 4.1 Remote Sensing Data

The available remote sensing data used in this study will be provided by two datasets: MODIS and GLDAS-2.1.

#### 4.1.1 The MODIS satellite characteristics

MODIS is a sensor embedded into two National Aeronautics and Space Administration (NASA) satellites: Terra launch in 1999, and Aqua launched in 2002. Therefore, MODIS data has been available and used for research over the last 20 years.

There are four types of resolution for remote sensing images: spatial, temporal, spectral and radiometric.

The spatial resolution corresponds to the size of the pixels. This depends on the spectral band that is used. For MODIS data, bands B1 and B2 have a 250m resolution, B3 to B7 have a 500m resolution and bands B8 to B36, a 1000m resolution.

Unlike spatial resolution, temporal resolution is not a constant value across the globe. Surely, each Terra and Aqua satellite have a 5 day revisit. But, due to overlap in adjacent orbits, the actual temporal resolution spans from 1 to 2 days.

The spectral resolution concerns which wavelength of the electromagnetic spectrum is captured. MODIS has 36 bands spanning from violet (band 8 at 405nm) to thermal infrared (band 36 at 14000nm) with no overlapping. The visualisation of the spectral bands is available in the appendix B.

Radiometric resolution is the capacity of the instrument to measure a diverse range of the energy of an electromagnetic wave, for a given spectral band. It is expressed in bits. For example, an 8-bit radiometric resolution of a black and white image, will lead to  $2^8 = 256$  discrete values of gray [65]. MODIS satellites have a radiometric resolution of 12 bits per spectral band, which leads to 4096 discrete values of intensity [66].

Information about MODIS characteristics are summarised in Table 2. For comparison, the same characteristics are also given for the Sentinel-2 satellite, another remote sensing data source for estimating GPP in statistical models.

<b>Characteristics</b>	<b>MODIS</b>	<b>Sentinel-2</b>
Date of first capture	18-02-2000	28-03-2017
Spatial resolution (m)	250, 500, 1000	10, 20, 60
Spectral resolution	36 bands	13 bands
Temporal resolution (days)	1	2-5
Radiometric resolution	12 bits	12 bits

Table 2: Comparative table between MODIS and Sentinel-2 characteristics.

#### 4.1.2 The Global Land Data Assimilation System (GLDAS) dataset

The Global Land Data Assimilation System (GLDAS) is a meteorological dataset that combines in ground based and satellite observations [67]. It is a collaborative project between the NASA,

the Goddard Space Flight Center (GSFC), the National Oceanic and Atmospheric Administration (NOAA), and the National Centers for Environmental Prediction (NCEP). The spatial resolution is off 0.25°, corresponding to 27830m. Therefore, this dataset is not locally precise, but still provides a global availability. So, a generalisation of the results to a larger scale is possible. The 2.1 version of this dataset is used in this study.

### 4.1.3 Available Products in Google Earth Engine

The "raw" remote sensing data is captured by the instrument owners, i.e. NASA for MODIS. Before sharing their images, the owners have to preprocess the data. Several processing levels are available. Level-0 is the unprocessed data. Level-1 focuses on decompressing, radiometric correction, and geometric corrections. Level-2 includes an atmospheric correction. Level-3 adds some map projections and temporal homogenisation. Level-4 are products that are indirectly captured by remote sensing data, and that have received extra processing [68]. For our study, we will work with the Level-4 MCD15 and MOD17 datasets. Nevertheless, despite being a level-4 product, MCD15 still needs supplemental processing before being used.

Some useful products of certain algorithms that uses remote sensing data are available with the Google Earth Engine (GEE) catalog. This is the case of the GPP and ET of PML model at a global scale [23], the GPP and NPP of MOD17 [69], the environmental variables from GLDAS-2.1 (temperature, air pressure, Short-Wave (SW) radiations, and specific humidity (Qair)) [67], and the surface classification dataset MCD12 [70]. Details of the available products in the GEE catalog are available in table 3. The data availability in time of the in-situ measurements is in the data time-window of the diverse remote sensing products. Therefore, only the in-situ measurements define the data's temporal range.

Datasets	Beginning date	End Date	Spatial resolution	Temporal resolution
PML GPP& ET	26-02-2000	26-12-2020	500m	8 days
MOD17 GPP&NPP	18-02-2000	10-02-2023	500m	8-days
MCD15A3H.061 LAI	04-07-2002	Today	500m	4 days
MCD12Q1.061 surface classification	01-01-2001	01-01-2022	500m	Yearly
GLDAS-2.1 Temperature, Pressure, SW radiations, Qair	01-01-2000	Today	500m	3 Hourly

Table 3: Datasets available in the Earth Engine Data Catalog [71]. Dates are in the DD-MM-YYYY format.

## 4.2 In-situ measurements

As said in the introduction, GPP can be measured locally. This study focuses on the Torgnon subalpine grassland. Torgnon is mountainous site, having an average elevation of 2168 m, and a

cold mean temperature of 3.27°C [72]. Therefore, the GPP and meteorological conditions are greatly influenced by seasons, having a snowy context in Winter. The data availability is from 01-01-2016 to 01-01-2020, which constitutes four full years of study. Torgnon is a member of the FLUXNET group. It is a network of flux measurements of carbon dioxide, and water vapour between the soil and the atmosphere [11]. Moreover, Torgnon is one of the study sites that served for the training of the PML model available in GEE [23].

The measure of GPP is performed indirectly by first measuring the Net Ecosystem Exchange (NEE) using the Eddy Covariance (EC) method. From those measurements we can derive GPP using various mechanisms. For this study, we use Papale’s methodology [8], having the specific processing for the Torgnon grassland developed by Spinosa *et al.* [37]. The procedure mainly consist of data correction (outlier detection, velocity friction filtering and gap filling), physical equation to derive GPP (carbon fluxes partitioning), and a daily aggregation. The obtained time-series of GPP is presented in figure 5 (black curve).

Even though, those measurements are subject to uncertainties, they constitute the reference data for our study.

### 4.3 Data processing

In order to compare more precisely the PML output of this study with the one available on GEE, we use the same input datasets and processing as described in Zhang *et al.* article [23].

#### 4.3.1 Climatic variables

The considered climatic variables are temperature, VPD, CO<sub>2</sub> atmospheric concentration, and PAR. The first two can either be obtained with in-situ measurements or with remote sensing data. In the first case, the methodology developed by Spinosa *et al.* [37] is used. The time series have a 30 minutes time step and are aggregated to a daily time step by averaging their values over 24 hours.

In the second case, we use the environmental variables available in the GLDAS-2.1 dataset with a 3-hourly time step [67]. VPD is not directly available and needs to be computed. For this we used the same methodology as Zhang *et al.* [23]. The vapour pressure of water is computed using Tetens equation [73]

$$e_s = 0.6108 \exp \frac{17.27T}{T + 237.3} \times 10^3 \quad (34)$$

having  $T$  being the temperature in °C and  $e_s$  the vapour pressure of water in Pa. The maximum and minimal daily vapour pressure of water ( $e_{s,max}$  and  $e_{s,min}$  respectively) are computed using the maximum and minimum daily temperature.

The saturation water pressure  $e_a$  is calculated with Bolton’s formula [74]:

$$e_a = \frac{qP}{0.622 + 0.378q} \quad (35)$$

where  $q$  is the relative humidity (kg/kg) and  $P$  the pressure in Pa. Both variables are available in the GLDAS-2.1 dataset. Finally, daily VPD is obtained by subtracting the saturation water pressure to the average of the maximum and minimum daily vapour pressure of water:

$$VPD = \frac{e_{s,max} + e_{s,min}}{2} - e_a \quad (36)$$

Many literature models evaluate GPP on an 8-day period, resulting in smoother results [23], [34]. To imitate this smoothness, an 8-days moving average is applied to the daily value of the time series.

The visualisation of temperature and VPD in both in-situ and remote sensing measurements is given in figure 1. We observe in both cases an underestimation of the GLDAS data.

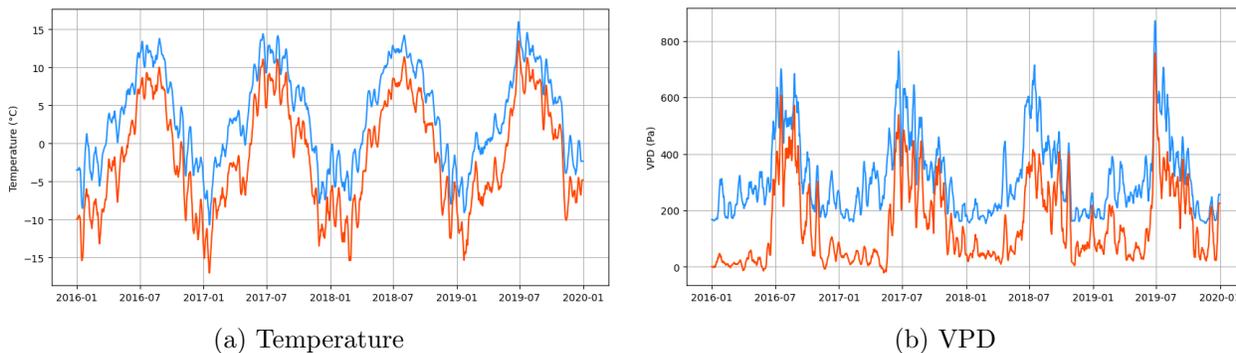


Figure 1: Smoothed temperature and VPD time series. The blue and red curves respectively correspond to the in-situ and remote sensing measurements.

### 4.3.2 CO<sub>2</sub> atmospheric concentration

CO<sub>2</sub> atmospheric concentration is available in the NOAA globally averaged marine surface monthly mean data [75]. The values are given monthly and need to be interpolated to daily. This dataset is global so not specific to the study site. However, it makes the upscaling of the model to a regional scale easier.

A visualisation of the time series is available in figure 2. Over the four years of study, the CO<sub>2</sub> atmospheric concentration raises of about 10 ppm corresponding to a relative augmentation of 2.5%. We can also notice the sinusoidal variation of CO<sub>2</sub> over the seasons. Because the CO<sub>2</sub> variations are significant at a larger scale then the four years of this study, it is likely that the uncertainties in the atmospheric concentration of CO<sub>2</sub> will have a minor impact in the estimation of GPP.

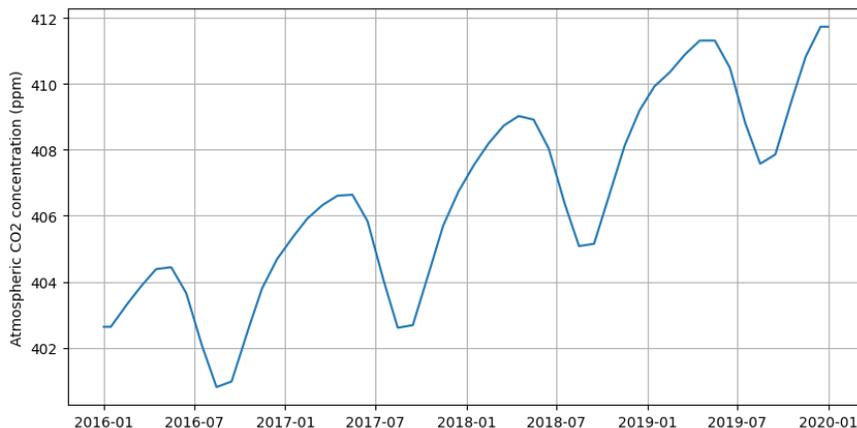


Figure 2: CO<sub>2</sub> concentration from the NOAA dataset linearly interpolated to daily values.

### 4.3.3 Photosynthetically Active Radiations (PAR)

Photosynthetically Active Radiations (PAR) are the sunlight radiations that are in the 400-700nm band. They correspond to 45% of the shortwave downward radiation. Similarly to temperature and VPD, the shortwave downward radiations are taken in the GLDAS-2.1 dataset, with a 3-hour temporal resolution [67]. Shortwave radiation is an instantaneous variable (expressed as a power per surface area), so the aggregation must be done by averaging the value over the day. The units are then converted into a photosynthetic photon flux density ( $\mu\text{mol.m}^{-2}.\text{s}^{-1}$ ).

Similarly to the smoothing of temperature and VPD, the PAR signal is recomputed using an 8-days moving average. This has a strong effect in filtering high frequencies information. The visualisation of the PAR time series is presented in figure 3. We see that the radiations follow the natural season cycle. It is bounded by a sinusoidal function representing the inclination of Earth with respect to the sun. However, the presence optical masks, like clouds, lowers the received amount of light.

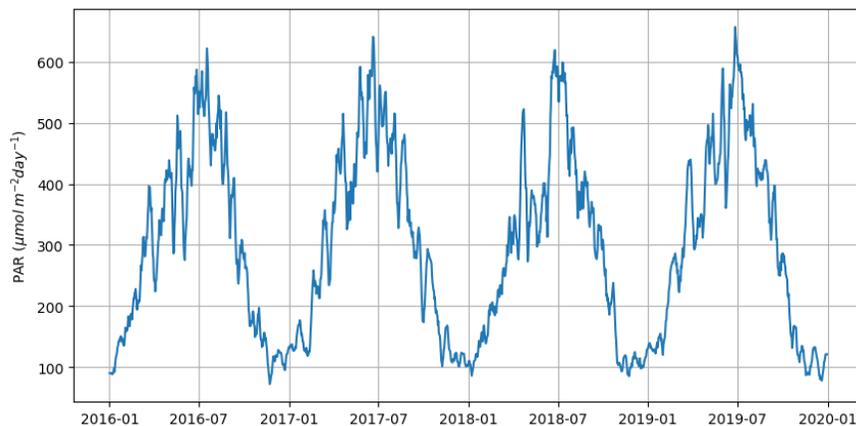


Figure 3: Time series of PAR expressed as a photon flux density in  $\mu\text{mol m}^{-2}\text{day}^{-1}$  from the GLDAS-2.1 dataset

### 4.3.4 Leaf Area Index

LAI is estimated using the MCD15A3H dataset, with a 4-day resolution. This dataset is a level-4 product of the MODIS instrument. However, the available LAI is contaminated by noise coming from clouds, shadows, aerosols and cirrus. To remove it, we use a weighted Whittaker smoother [47]. A description of the Whittaker smoother is given in section 3.1.1. In practice, the module `whittaker_eilers` from python is used. It is based on the sparse LU-decomposition solver from `scipy`.

The goal of the weighted smoother is twofold. First the weights model the quality of the measurements and filter outliers. A weight is applied to each measurement, regardless of the neighbors. Second, the smoothing keeps the low frequencies in the signal and removes noise. By using an order 3 smoother, the values of the 3 previous measurements are used to rectify the current measurement.

The initial weights are set as:

- 1 if the image is not contaminated,
- 0.5 if the image is contaminated by aerosols or cirrus,

- 0.05 if the image is contaminated by snow, cloud or shadow

The updating of the weights at the  $k$ -th iteration proceeds as follow:

$$r_{i,k} = LAI_{i,k} - LAI_i \quad (37)$$

$$MAD_k = \text{median}(|r_{i,k}|) \quad (38)$$

$$w_{i,k+1} = \begin{cases} \left(1 - \left(\frac{r_{i,k}}{6MAD_k}\right)^2\right)^2 w_{i,k} & \text{if } 0 < r_{i,k} < 6MAD_k \\ w_{i,k} & \text{if } -6MAD_k < r_{i,k} < 0 \\ 0 & \text{if } 6MAD_k < |r_{i,k}| \end{cases} \quad (39)$$

where  $w_{i,k}$  is the weight value of the point  $i$  at iteration  $k$ ,  $LAI_i$  is the value of the LAI from the MCD15 product of the point  $i$ , and  $LAI_{i,k}$  is the value of the smoothed LAI of the point  $i$  at iteration  $k$ .

This algorithm is applied for 2 iterations, with the smoothing scalar  $\lambda$  set at 100. The time series of the unprocessed and smoothed LAI products is shown in figure 4. We observe a much smoother data and many filtered measurements. The increasing and decreasing phases happening in Spring and Summer are regular. Also, the amplitude of the smoothed signal less important.

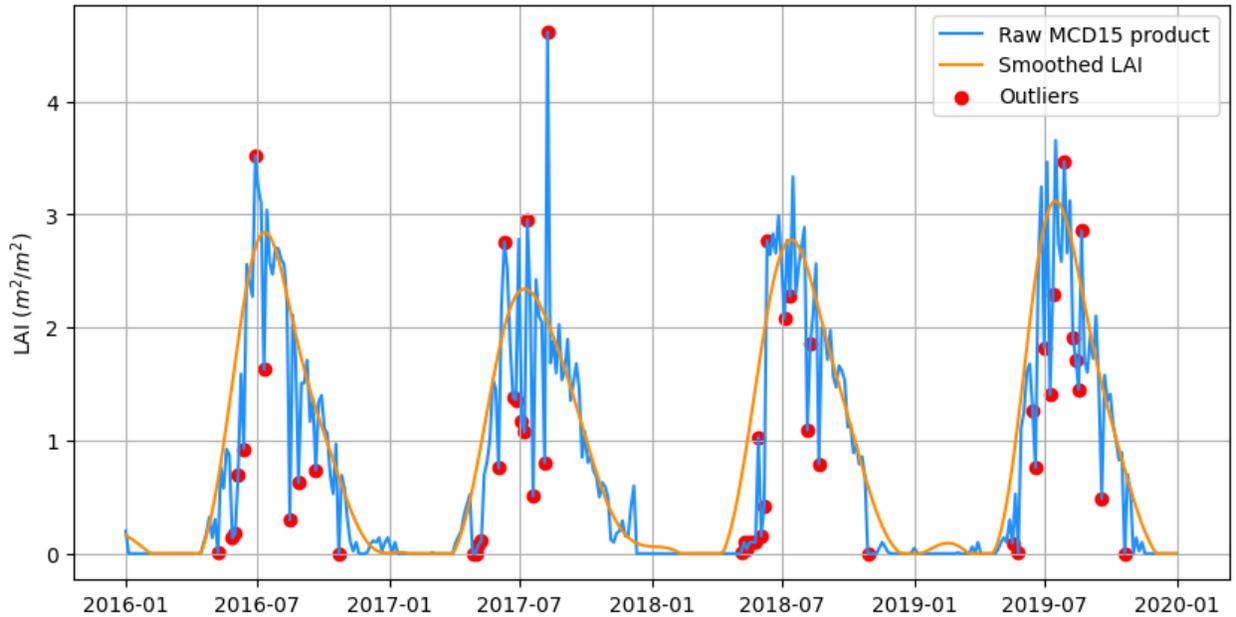


Figure 4: Time series of LAI. The blue curve is the MCD15 unprocessed product. The orange curve is the smoothed LAI. The red dots are the totally filtered outliers (weight set to 0).

## 5 Results and analysis of the PML model

### 5.1 Estimation using parameters from literature

GPP depends on both the meteorological and environmental context, and on how plants adapt to this context. In the PML model, the climate situation is represented by the variables. The ability of primary production of the vegetation, given the climate conditions, is modeled by the parameters. Hence, the values of the parameters in the PML model are specific to the type of plants.

Zhang *et al.* use the land classification of the MCD12 product to cluster vegetation types. According to this same classification, the Torngon site is considered being a grassland. Therefore, the corresponding parameter values are given in table 4 [23].

Parameter	$\beta$	$\eta$	$k_q$	$V_{m,25}$	$VPD_{max}$	$VPD_{min}$
Unit	$\mu\text{mol C}$ $(\mu\text{mol PAR})^{-1}$	$\mu\text{mol C m}^{-2}\text{s}^{-1}$ $(\text{ppm})^{-1}$	-	$\mu\text{mol m}^{-2}\text{s}^{-1}$	Pa	Pa
Values	0.03	0.03	0.72	46.4	3630	660

Table 4: Value of the parameters in Torngon grassland according to literature.

Therefore, we can compute the daily GPP time-series using those values. Their representation is shown in figure 5. We can notice first of all, that the model identifies correctly the seasonal aspect of GPP, having low amounts in Autumn and Winter, growth in Spring, a peak in July, and a decrease for the rest of Summer. However, there is a global overestimation, especially around the maximum values. In Autumn and Winter, our LAI is estimated as 0. This explains the flat zero value of the PML output in Autumn and Winter. Moreover, there is a minor difference between the GEE PML product and our self-implemented PML. This can be explained by the difference in the processing of LAI. Indeed, the value of LAI is sensitive to the tuning of the Whittaker smoother, resulting in different input data. Further discussion on this difference is discussed in section 6.1.3. Furthermore, there is a minor difference in the PML output if we rather use in-situ or remote sensing measurements of Temperature and VPD (respectively orange and green curve in the figure). However, the ramp function from literature is barely constraining GPP. Indeed, the minimum threshold for which the constraint is applied ( $VPD_{min}$ ) is 660 Pa. In our dataset, this value is only exceeded once. Therefore, the constraint is active only for one day out of the four years studied.

Besides, the MOD17 model tends to underestimate the reference data. Its estimation pattern is different from the three PML outputs.

In addition, table 5 gives the RMSE and Pearson’s coefficient of determination  $R^2$  for the different outputs. The definitions of both metrics are given in section 3.3.1. The products from GEE perform significantly better than our own PML, highlighting the importance of the optimal processing of LAI. Also, as observed in the figure, the source of input temperature and VPD exerts minimal impact on the PML output.

### 5.2 Correlation analysis

To have a first understanding of the link between the variables themselves, and between the variables and the GPP, we perform a correlation analysis. The correlation matrix is shown in figure 6. The

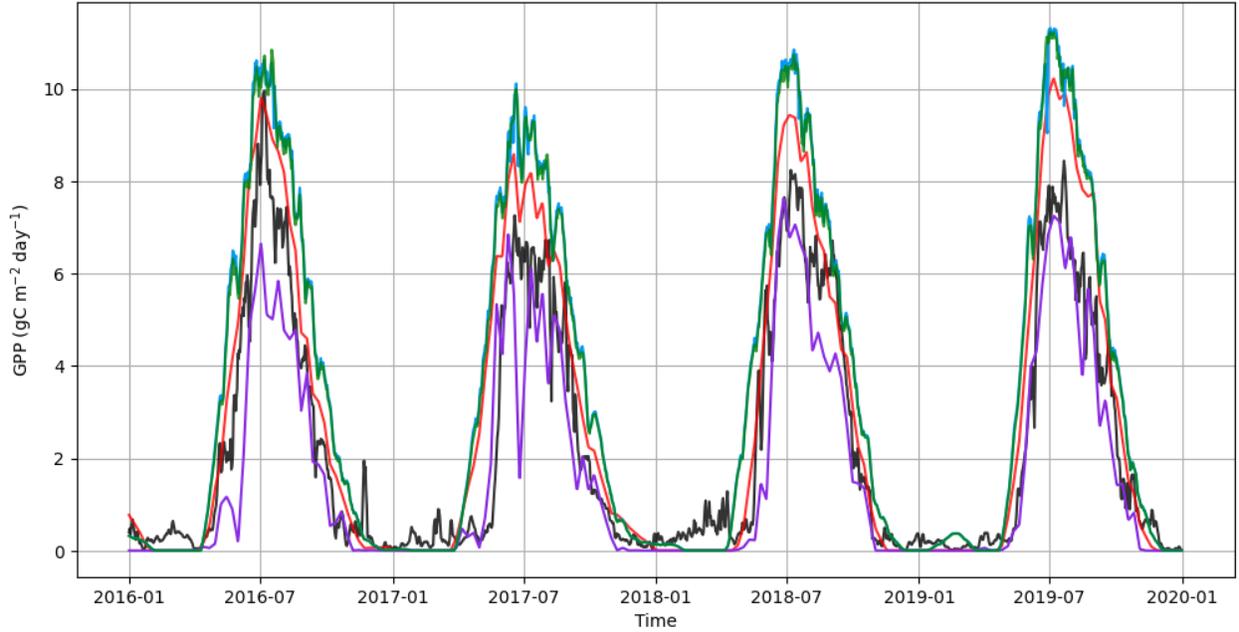


Figure 5: Time series of the PML model using parameter values founded in literature. The black curve is the reference data. The blue and green curves are the output of the PML model using the methodology describe in section 4.3. They respectively use the in-situ measurements, and the remote sensing estimation of temperature and VPD. The red curve is the product available in GEE. The purple curve is the MOD17 product.

	<b>In-situ EVs</b>	<b>RS EVs</b>	<b>GEE product</b>	<b>MOD17</b>
<b>RMSE (gC m<sup>-2</sup> day<sup>-1</sup>)</b>	1.80	1.73	1.15	1.04
<b>R<sup>2</sup></b>	0.53	0.57	0.81	0.84

Table 5: RMSE and Pearson’s R<sup>2</sup> of the GEE products and the self-implemented PML model using parameter values from literature.

correlation score is written in each cell of the matrix. However, the color of each cell corresponds to the absolute value of the correlation score.

This matrix shows that the atmospheric concentration of CO<sub>2</sub> is poorly (in absolute value) and negatively correlated to the other variables and to GPP. This goes against the physical model of photosynthesis that is influenced by that the amount of CO<sub>2</sub>. This can be explained by the stability of the quantity of carbon dioxide within four years in the used dataset. Moreover, the global aspect of the dataset may not be applicable to the specific Torgnon case. Other influential variables have more fluctuating values, which explain the variations of GPP. This is why temperature, VPD, and PAR are highly correlated to GPP. Moreover, LAI is extremely correlated to GPP with a correlation score over 0.9. More specifically, the smoothed time series has a higher correlation than the raw one, which highlights the need of processing LAI.

To get a better view of the seasonal aspect, the correlation matrix is computed per season (see figure 7). The correlation scores remain good in Spring, Summer and Autumn, but plummet in Winter. Therefore, in Winter, the PML model and the used variables cannot predict the low GPP

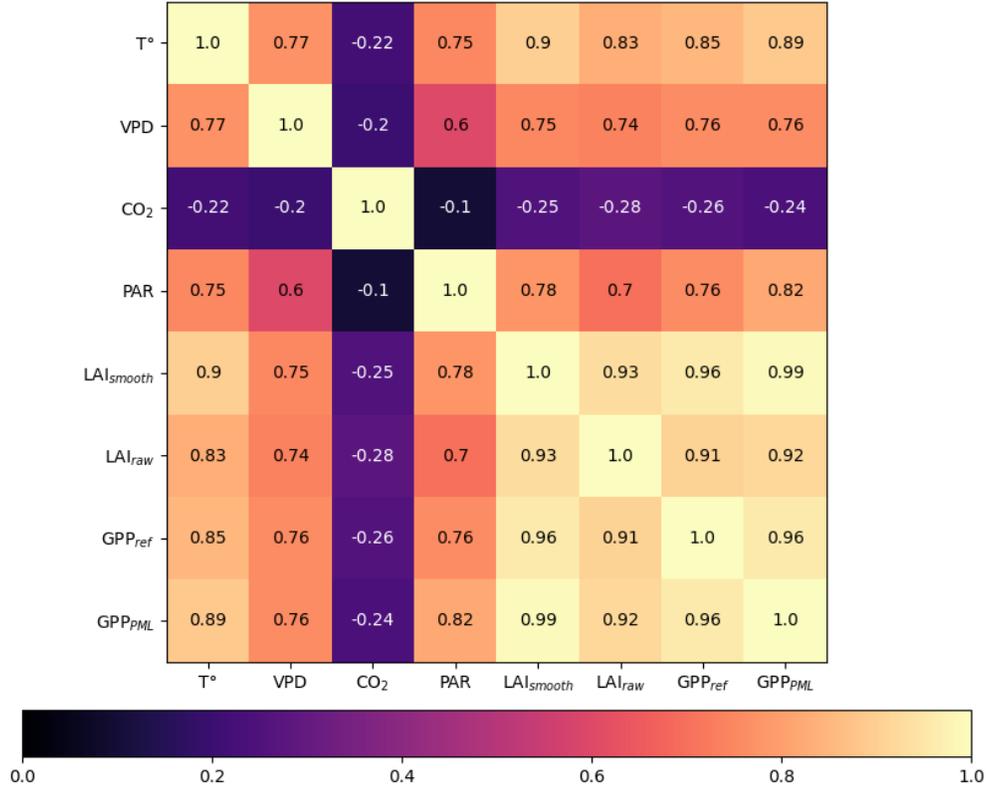


Figure 6: Correlation matrix of the variables and GPP.

values. This is probably due to the extreme Winter conditions in the Alps, with the presence of snow, low temperatures and pressure, and a small amount of vegetation ( $LAI \approx 0$ ).

Similar conclusions can be derived from the Spearman’s rank correlation matrix in appendix C.

### 5.3 One at the time procedure and asymptotic analysis of parameters

To understand the link between input and output data, we perform an OVAT procedure with an asymptotic analysis. The detailed steps for the computation of the limits and asymptotic equations that will follow are presented in appendix D. For each variable, we select a range of possible values. We consider the other variables constant by taking their mean value over the study period. We then evaluate the model on the range values. This last step is repeated by changing the value of relevant parameters. In this analysis, the VPD constraint is ignored for simplicity.

#### 5.3.1 Temperature and VPD constraint

In the PML model, both temperature and VPD act has constraints to the growth of GPP. The temperature constraint is given in equation 4. This constraint only acts on the value of the maximum catalytic capacity of RuBisCO  $V_m$ . The VPD constraint is a ramp function given in equation 9, that acts on the global assimilation rate. The visualisation of both function is represented in figure 8.

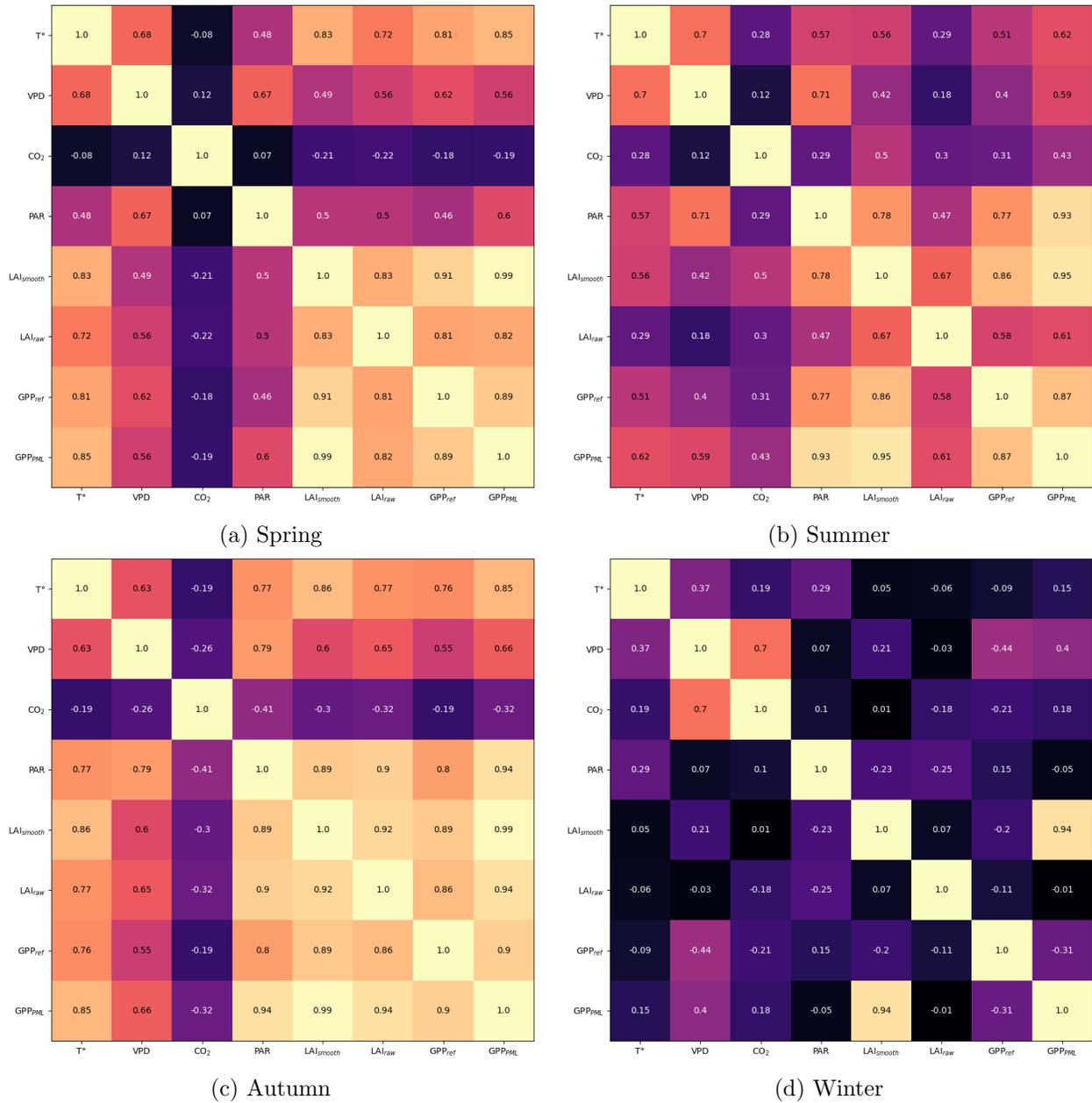


Figure 7: Correlation matrix over the seasons. The color bar is the same as in figure 6.

### 5.3.2 GPP response to PAR and CO<sub>2</sub> atmospheric concentration

The assimilation of carbon at the canopy scale without the VPD constraint is expressed in equation 8. The parameter  $\beta$  corresponds to the initial slope of carbon assimilation response to PAR. It is therefore linked to the quantity of PAR  $I_0$ . The primary production due to light exposure appears in the logarithmic terms of equation 8. When either  $I_0$  or  $\beta$  goes to infinity, the GPP response is bounded by the value:

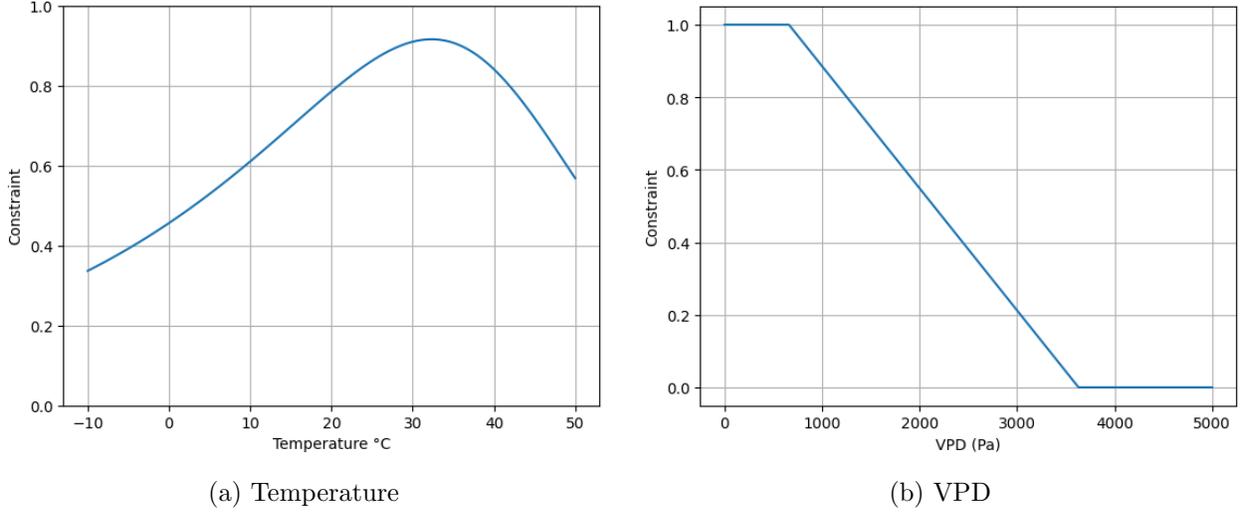


Figure 8: Temperature and VPD constraints. The values of the parameters for the visualisation of both functions are from [26] (temperature) and [23].

$$\lim_{I_0, \beta \rightarrow +\infty} A_{c,g} = \frac{V_m \eta C_a}{V_m + \eta C_a} LAI \quad (40)$$

Likewise, the parameter  $\eta$  corresponds to the initial slope of carbon assimilation response to the atmospheric  $\text{CO}_2$  concentration. This production is also bounded when the variable or the parameter goes to infinity:

$$\lim_{C_a, \eta \rightarrow +\infty} A_{c,g} = \frac{V_m}{k_q} \left[ k_q LAI + \ln \left( \frac{\beta I_0 + V_m}{\beta I_0 + V_m \exp(k_q LAI)} \right) \right] \quad (41)$$

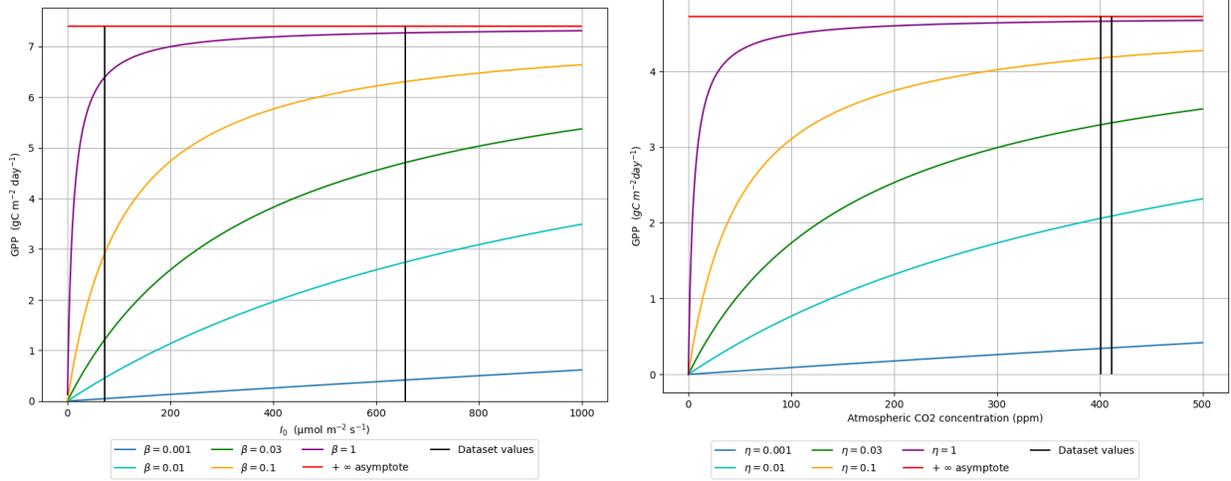
The visualisation of the GPP value when those two variables vary are shown in figure 9. We notice that for high values of  $\beta$  and  $\eta$ , the GPP reaches quickly the limit value. Therefore, for the calibration phase, the differences between a large and an extraordinary large value of  $\beta$  and  $\eta$  will have little consequences in the output.

Furthermore, the values of  $I_0$  in our dataset are where the GPP is in the transient state. Therefore, the parameter  $\beta$  relates on how much GPP will vary with variations of  $I_0$ .

Moreover, due do the small variations of  $C_a$  in the dataset, the GPP response to it is very stable. Therefore, uncertainties in  $C_a$  have a negligible impact on GPP. Hence, the global atmospheric  $\text{CO}_2$  dataset that is used, which has a poor temporal resolution (monthly) is sufficient for this study.

### 5.3.3 GPP response to LAI

When building the PML model, equation 5 highlights that the  $k_q LAI$  expresses the extinction factor of PAR. The LAI states how much vegetation there is to stop the light, and the parameter  $k_q$  shows how effective this extinction happens. When either LAI or  $k_q$  goes to zero, we can linearise the PML model as follow:



(a) Evolution of GPP with respect to  $I_0$  for several values of  $\beta$ . (b) Evolution of GPP with respect to  $C_a$  for several values of  $\eta$ .

Figure 9: Evolution of GPP with respects to specific variables and their associated parameter. The red curve represents the asymptote when the variable or the parameter goes to infinity. The green curve parameter value is the one from literature. The black vertical lines delimit the interval that contains the dataset values.

$$A_{c,g} \underset{LAI, k_q \rightarrow 0^+}{\sim} \frac{V_m \eta C_a \beta I_0}{V_m \eta C_a + V_m \beta I_0 + \eta C_a \beta I_0} LAI \quad (42)$$

The fact that  $k_q$  does not appear in the linearisation shows that for low values of LAI, this coefficient does not have an impact on primary production. This means that GPP is proportional to the amount of vegetation, in the case where light does not extinct.

Also for high values of LAI we reach the following limit:

$$\lim_{LAI \rightarrow +\infty} A_{c,g} = \frac{V_m \eta C_a}{k_q (V_m + \eta C_a)} \ln \left( \frac{(V_m + \eta C_a) \beta I_0 + V_m \eta C_a}{V_m \eta C_a} \right) \quad (43)$$

Figure 10 shows that the linearisation is efficient over the hole dataset for low values of  $k_q$ . Also, for the values of LAI in Autumn and Winter, the value of  $k_q$  is not significant. This means that we need to Spring and Summer data in the train set for an accurate calibration of this parameter.

Moreover, when  $k_q$  is high but not very high, we reach the limit in equation 43 within the dataset values, without that limit being 0. Therefore, the exploration of the value of  $k_q$  going beyond 2 will probably neutralise the estimate.

## 5.4 Calibration of parameters and uncertainties

The PML model relies on six parameters:  $\beta$ ,  $\eta$ ,  $V_{m,25}$ ,  $k_q$ ,  $VPD_{min}$ , and  $VPD_{max}$ . They model the photosynthesis proprieties of vegetation, in a given meteorological context. The goal of studying the reliability of this calibration procedure is to determine what are the physical proprieties we can extract from this procedure. Also, it provides information on the influence of the linked variables for estimating GPP.

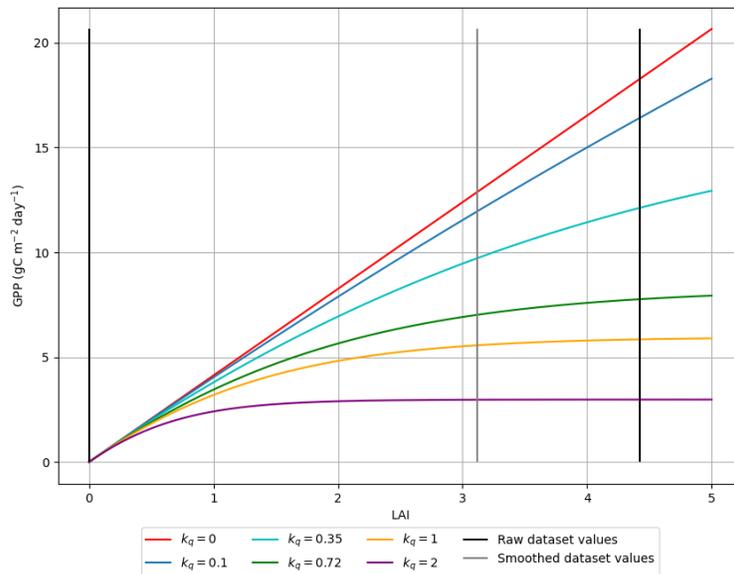


Figure 10: Evolution of GPP with respect to LAI and for several values of  $k_q$ . The red curve represents the asymptote when the variable or the parameter goes to zero. The green curve parameter value is the one from literature. The black and gray vertical lines delimit the interval that contains the raw and smoothed dataset values respectively.

For that, we will need to have a global picture of the objective function we want to minimise. Therefore, classical optimisation algorithm such as the gradient descent cannot be applied, as they only give one local minimum. Nevertheless, as six parameters are being optimised, the objective function is six-dimensional, which is quite challenging to represent. Consequently, we will start studying simplified cases that reduces the number of dimensions. Once the understanding of the simplified model is done, we will increase the number of dimensions. To do so, we start by ignoring the VPD constraint. Then, by using the asymptotic equations, we decrease the number of dimensions. Next, we perform the grid search optimisation on the simplified equations. This method is memory and computational expensive but gives a global picture of the objective function. Once this picture is analysed, the dimensionality can be increased. Eventually, the VPD constraint is reset and an optimal ramp function is founded.

In this section, we will denote  $p^*$  the optimal value of any parameter  $p$ .

#### 5.4.1 One-dimensional equations

The PML model is built on the availability of PAR and  $\text{CO}_2$ , and the maximum catalytic capacity of RuBisCO  $V_m$ . When we set two of those three quantities to an infinite value, it means that the remaining variable is the only limiting factor of  $\text{CO}_2$  absorption. We therefore get the following equations:

$$\lim_{\beta, V_{m,25} \rightarrow +\infty} A_{c,g} = \eta C_a LAI \quad (44)$$

$$\lim_{\beta, \eta \rightarrow +\infty} A_{c,g} = V_{m,25} f(T) LAI \quad (45)$$

$$\lim_{\eta, V_{m,25} \rightarrow +\infty} A_{c,g} = \frac{\beta I_0}{k_q} (1 - e^{-k_q LAI}) \xrightarrow[k_q \rightarrow 0]{} \beta I_0 LAI \quad (46)$$

Where  $f(T)$  is the temperature constraint defined in equation 4. In equations 44 and 45, when  $\beta$  goes to infinity, the vanishing of the light  $k_q$  also disappears from the equation. Physically speaking, when the ecosystem has an infinite quantum efficiency, then the amount of available light is no longer a relevant factor. Therefore, the quantity of light that is extinct also becomes irrelevant. In the case of equation 46, the light availability is the only limiting factor, so both  $\beta$  and  $k_q$  are present. However, if we put our self in the extreme case where light does not extinct ( $k_q = 0$ ), then the simplified equation becomes one -dimensional.

Figure 11, shows the RMSE of those equation. All of the four equations are strictly convex and the minimum is easy to find.

**Critical values** Using the one-dimensional equations, we can derive "critical values" for the parameters. Let  $\beta_c, \eta_c, V_{m,25_c}$  be the critical values for  $\beta, \eta$  and  $V_{m,25}$ . Let  $b$  represent the rest of the parameters in the PML equation. By definition:

$$\forall \beta < \beta_c, RMSE(PML(\beta, b)) > RMSE(PML(\beta_c, b)) \quad (47)$$

$$\forall \eta < \eta_c, RMSE(PML(\eta, b)) > RMSE(PML(\eta_c, b)) \quad (48)$$

$$\forall V_{m,25} < V_{m,25_c}, RMSE(PML(V_{m,25}, b)) > RMSE(PML(V_{m,25_c}, b)) \quad (49)$$

Having  $b = (\eta, V_{m,25}, k_q, VPD_{min}, VPD_{max})$  in equation 47,  $b = (\beta, V_{m,25}, k_q, VPD_{min}, VPD_{max})$  in equation 48, and  $b = (\beta, \eta, k_q, VPD_{min}, VPD_{max})$  in equation 49.

Those critical values are equal to the parameter values that minimises the MSE in the one-dimensional case. It is the parameter that solves the equation:

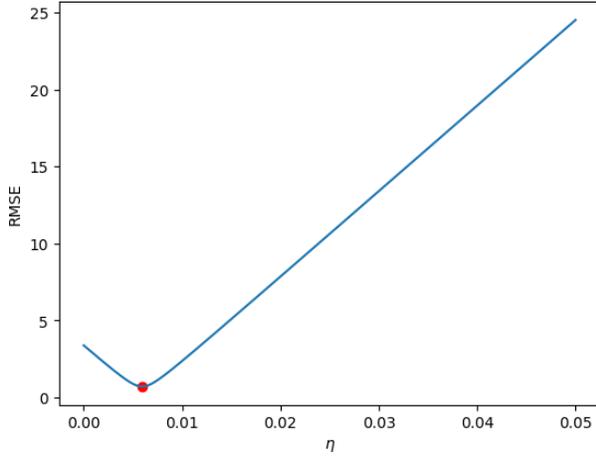
$$\sum_i \frac{\partial PML_i(a_c)}{\partial a_c} (GPP_i - PML_i(a_c)) = 0 \quad (50)$$

where  $i$  is a point in the training dataset,  $GPP_i$  is the reference value, and  $a_c$  is the critical value of a parameter  $a$ . The proof of the existence and the expression of the critical values cannot be given analytically because of the non-linearity of the model. But it can be verified numerically as it will be the case in the following sections.

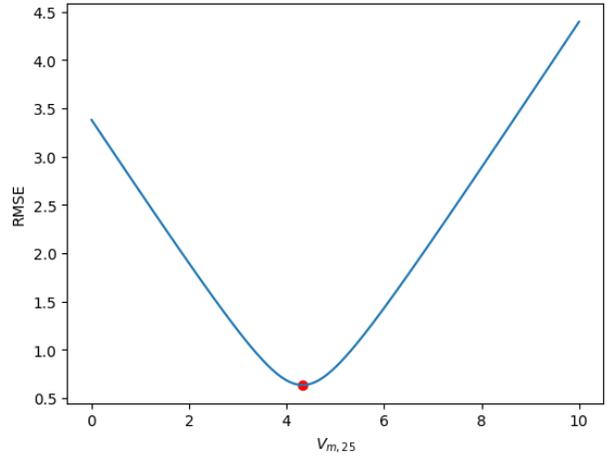
For the parameter  $k_q$ , a similar numerical proof can be given to derive an upper bound value, but this last one depends on the value of  $\beta$ . Indeed, it is impossible to have a one-dimensional version of the PML model that only depends on  $k_q$ . Therefore, no general formula can be derived.

Similarly, critical values for the VPD parameters will also depend on the other parameters. However, a minimum threshold that is bigger than the maximum value in the time series, will always result in a VPD constraint of 1 (so never active). Likewise, a maximum threshold that is lower than the minimum value in the time series will result in a VPD constraint of 0, so always active.

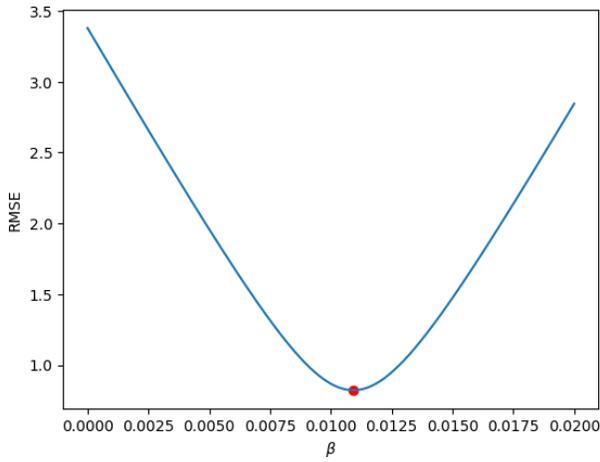
The founded critical values on the training set are summed up in Table 6.



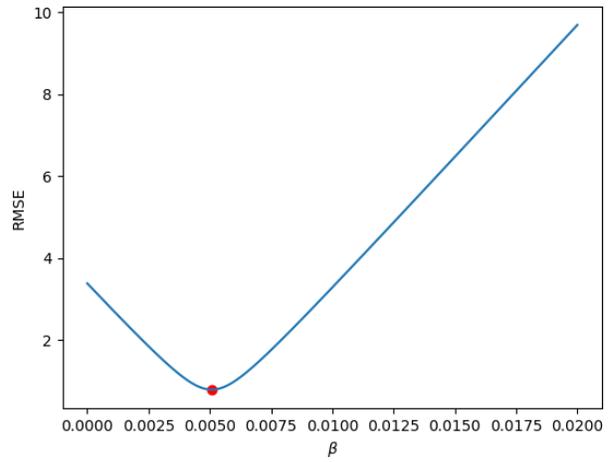
(a) RMSE of infinite  $\beta$  and  $V_{m,25}$  model



(b) RMSE of infinite  $\beta$  and  $\eta$  model



(c) RMSE of infinite  $\beta$ ,  $V_{m,25}$ , and  $k_q = 0.72$  model



(d) RMSE of infinite  $\beta$ ,  $V_{m,25}$ , and minimal  $k_q$  model

Figure 11: RMSE in the cases when there is only one limiting factor. The red dot corresponds to the minimum of the function.

Parameter	Critical value	Type of bound	Method to derive it
$\beta$	$0.00561 \mu\text{mol C m}^{-2}\text{s}^{-1} (\mu\text{mol PAR})^{-1}$	Lower	$\frac{\sum_i GPP_i I_{0_i} LAI_i}{\sum_i I_{0_i}^2 LAI_i^2}$
$\eta$	$0.00581 \mu\text{mol C m}^{-2}\text{s}^{-1} (ppm)^{-1}$	Lower	$\frac{\sum_i GPP_i C_{a_i} LAI_i}{\sum_i C_{a_i}^2 LAI_i^2}$
$V_{m,25}$	$4.40 \mu\text{mol C m}^{-2}\text{s}^{-1}$	Lower	$\frac{\sum_i GPP_i f(T) LAI_i}{\sum_i f(T)^2 LAI_i^2}$
$k_q$	Depends on $\beta$	Upper	
$VPD_{min}$	734 Pa	Upper	$\max(VPD)$
$VPD_{max}$	0 Pa	Lower	$\min(VPD)$

Table 6: Critical parameter values

### 5.4.2 Two-dimensional equations

For the two-dimensional case, we will only look at the cases where  $\beta \rightarrow \infty$  (equation 51) and  $V_{m,25} \rightarrow \infty$  (equation 52).

$$\lim_{\beta \rightarrow +\infty} A_{c,g} = \frac{V_m \eta C_a}{V_m + \eta C_a} LAI \quad (51)$$

$$\lim_{V_{m,25} \rightarrow +\infty} A_{c,g} = \eta C_a \left[ LAI + \frac{1}{k_q} \ln \left( \frac{\beta I_0 + \eta C_a}{\beta I_0 + \eta C_a e^{k_q LAI}} \right) \right] \quad (52)$$

The case of  $\eta \rightarrow \infty$  will be discussed in section 5.4.4. For now, we assume that the value of the coefficient of extinction is the literature value (0.72). With this the non-linearity with respect to LAI is kept.

The RMSE is evaluated for the following values of parameters (Table 7):

Parameter	Min value	Max value	Number of points	Type of scale
$\eta$	$0.9 \times \eta_c$	5	1000	Log scale
$V_{m,25}$	$0.9 \times V_{m,25c}$	10	498	Log scale

Table 7: Grid proprieties for the evaluation of the RMSE for  $\beta = \infty$  having  $\eta_c$  and  $V_{m,25c}$  respectively being the critical values for  $\eta$  and  $V_{m,25}$ .

In the case where  $\beta$  goes to infinity, the PML model only depends on  $V_m$  and  $\eta$ . Figure 12 displays aspects of the objective function. In order to work with one-dimensional curves instead of surfaces, the minimum of the RMSE with respect to one of the two parameters is represented per picture. On image (a), the scale is purposely zoomed in at the extreme. We notice that the optimised score of the RMSE is very close to the optimal value when  $\eta$  goes to infinity. Indeed, there is a only a difference of 0.005% between both values. So the implementation of the CO<sub>2</sub> information is very limited. On the contrary, the graph (b) shows a clear minimum with respect to  $V_{m,25}$ , having a strong difference between the optimal RMSE and the RMSE when  $V_{m,25}$  diverges. We can also notice some waves and irregularities in the functions, making their representation non convex. However, those last ones are due to the discretisation of the grid. There is indeed, a small difference between the optimal parameter that is present in the grid, and the optimal parameter in  $\mathbb{R}$ .

The figures (c) and (d) represent the argument of the minimum with respect to  $\eta$  and  $V_{m,25}$ . We can note that the optimal parameters are indeed bounded by the critical values. The functions  $\eta \rightarrow \arg \min_{V_{m,25}} \text{RMSE}(PML(\eta, V_{m,25}))$  and  $V_{m,25} \rightarrow \arg \min_{\eta} \text{RMSE}(PML(\eta, V_{m,25}))$  are decreasing, bounded by the limit values, being the critical values. Furthermore, the horizontal shape of the curves or low parameter values is justified by the limited range of the parameters. Therefore, those function are injective.

With a finer discretisation, the results could have been more precise (fewer waves, larger range...). However, the computing of those results is limited by time and especially memory. Thus, this coarse grid is sufficient for this analysis.

For  $V_m$  going to infinity, the GPP now depends on  $\beta$ ,  $\eta$  and  $k_q$ . For simplification, we use  $k_q = 0.72$  as in the literature. The used grid is described in table 8.

Figure 13 represents the same facets of the objective function as in figure 12. Similar comments can be given on the critical values, the monotony of the argument of the minimum, the convexity,

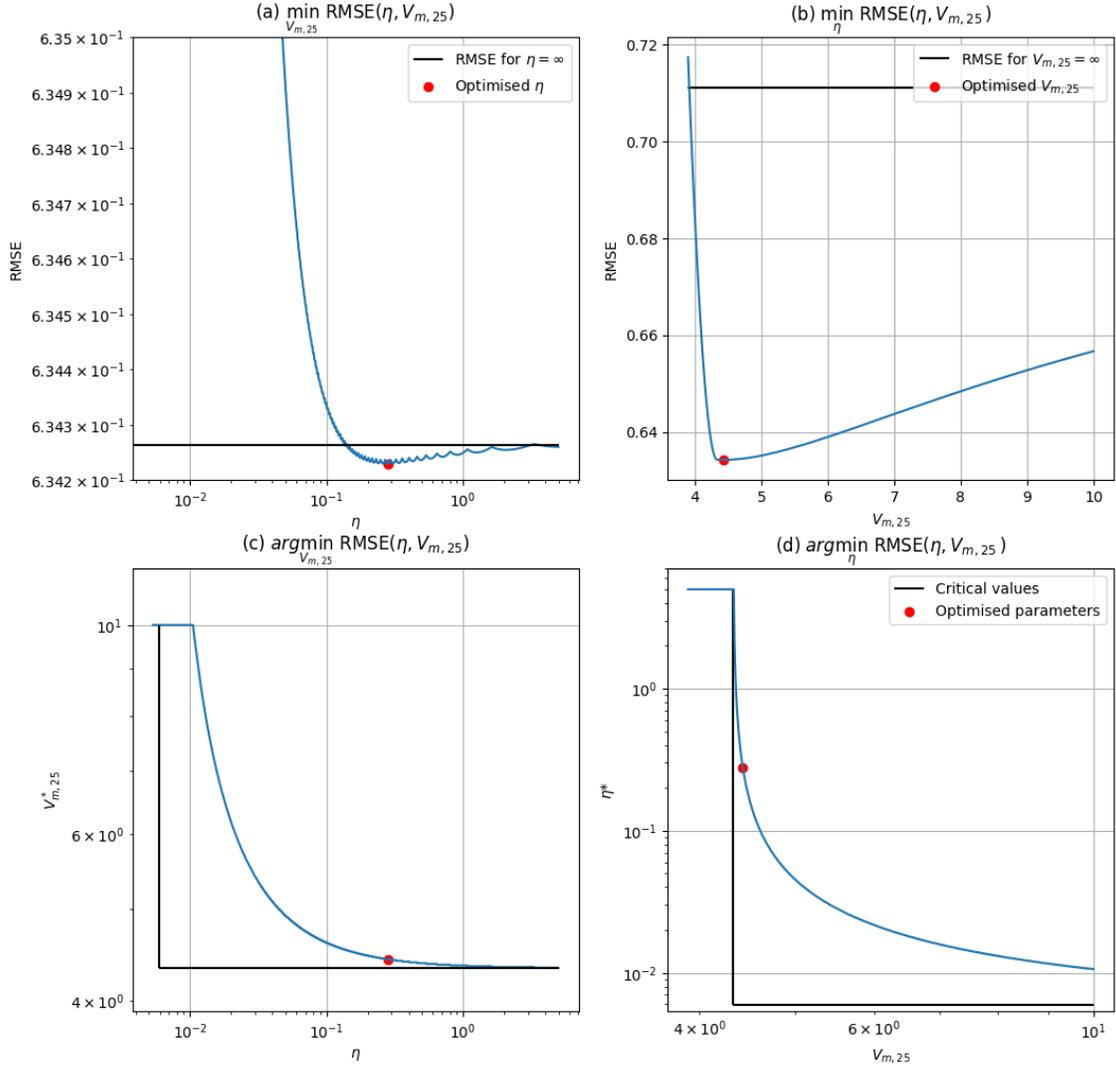


Figure 12: Visualisation of RMSE proprieties in the case where light availability is not a limiting factor. Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $V_{m,25}$  in function of  $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\eta$  in function of  $V_{m,25}$ .

Parameter	Min value	Max value	Number of points	Type of scale
$\beta$	$0.9 \times \beta_c$	2	501	Log scale
$\eta$	$0.9 \times \eta_c$	5	1000	Log scale

Table 8: Grid proprieties for the evaluation of the RMSE for  $V_{m,25} = \infty$ ,  $k_q = 0.72$  having  $\beta_c$  and  $\eta_c$  respectively being the critical values for  $\beta$  and  $\eta$ .

and the "wave" aspect in the RMSE. Furthermore, there is a clear optimal value for  $\beta$ . Nevertheless, the optimal value of the RMSE with respect to  $\eta$  is significantly lower than the optimal value when it goes to infinity. This can be explained by the fact that, in the case of assimilation without light

constraints (equation 51),  $V_m \eta C_a / (V_m + \eta C_a)$  is the GPP per unit leaf area. In the case of  $V_m \rightarrow \infty$ ,  $\eta$  becomes the only controllable parameter of that phenomenon.

Moreover, the value of  $\beta^*$  when  $\eta$  diverges, is different from the computed critical value. This is because,  $\beta$ 's critical value is a decreasing function with respect to the extinction coefficient  $k_q$ . Indeed, in case of light extinction, we have

$$\beta_c(k_q) = \frac{\sum_i GPP_i I_{0_i} (1 - e^{-k_q LA I_i}) / k_q}{\sum_i I_{0_i}^2 (1 - e^{-k_q LA I_i})^2 / k_q^2} \quad (53)$$

where  $\beta_c(k_q)$  is the critical value of  $\beta$  depending on  $k_q$ . Figure 33 in the appendix E shows a similar graph when  $k_q = 0$ .

### 5.4.3 Three-dimensional equations with $\eta \neq \infty$ .

Now, we consider that all the efficiencies play an active roll in gross primary production. We still set the vanishing term at 0.72. Because of the memory usage and the difficult representation of three simultaneously varying parameter, we did this optimisation and representation of RMSE on two different grids (Tables 9 and 10).

Parameter	Min value	Max value	Number of points	Type of scale
$\beta$	$0.9 \times \beta_c$	$\sqrt{10}$	6	Log scale
$\eta$	$0.9 \times \eta_c$	100	241	Log scale
$V_{m,25}$	$0.9 \times V_{m,25_c}$	10	252	Log scale

Table 9: Grid proprieties for the evaluation of the RMSE for  $k_q = 0.72$  having a small number of evaluations for  $\beta$ .

To display information about the RMSE, we use similar curves as in section 5.4.2, where each layer represent one value of the third parameter. For instance, in figure 14, each colour is for one value of  $\beta$ . Notice that the value  $\beta = \infty$  that has been studied in section 5.4.2 is also included in the graph. The same analysis applies for every values of  $\beta$ . More particularly, for every values of  $\beta$ , the optimal RMSE is very close to the RMSE when  $\eta \rightarrow \infty$ . On the other hand, the value of  $V_{m,25}^*$  clearly depends on the value of  $\beta$ .

Parameter	Min value	Max value	Number of points	Type of scale
$\beta$	$0.9 \times \beta_c$	$\sqrt{10}$	252	Log scale
$\eta$	$0.9 \times \eta_c$	100	241	Log scale
$V_{m,25}$	$0.9 \times V_{m,25_c}$	10	6	Log scale

Table 10: Grid proprieties for the evaluation of the RMSE for  $k_q = 0.72$  having a small number of evaluations for  $V_{m,25}$ .

Likewise, in figure 15 every colour represents one value of  $V_{m,25}$ . In this situation, the optimal value of  $\beta$  is quite stable per layer of  $V_{m,25}$ , with the exception of the blue curve, where  $V_{m,25}$  is below it's critical value.

Moreover,  $\eta^*$  is very dependent on the value of  $V_{m,25}$ . However, the global minimum in picture (a) is for when  $\eta \rightarrow \infty$ .

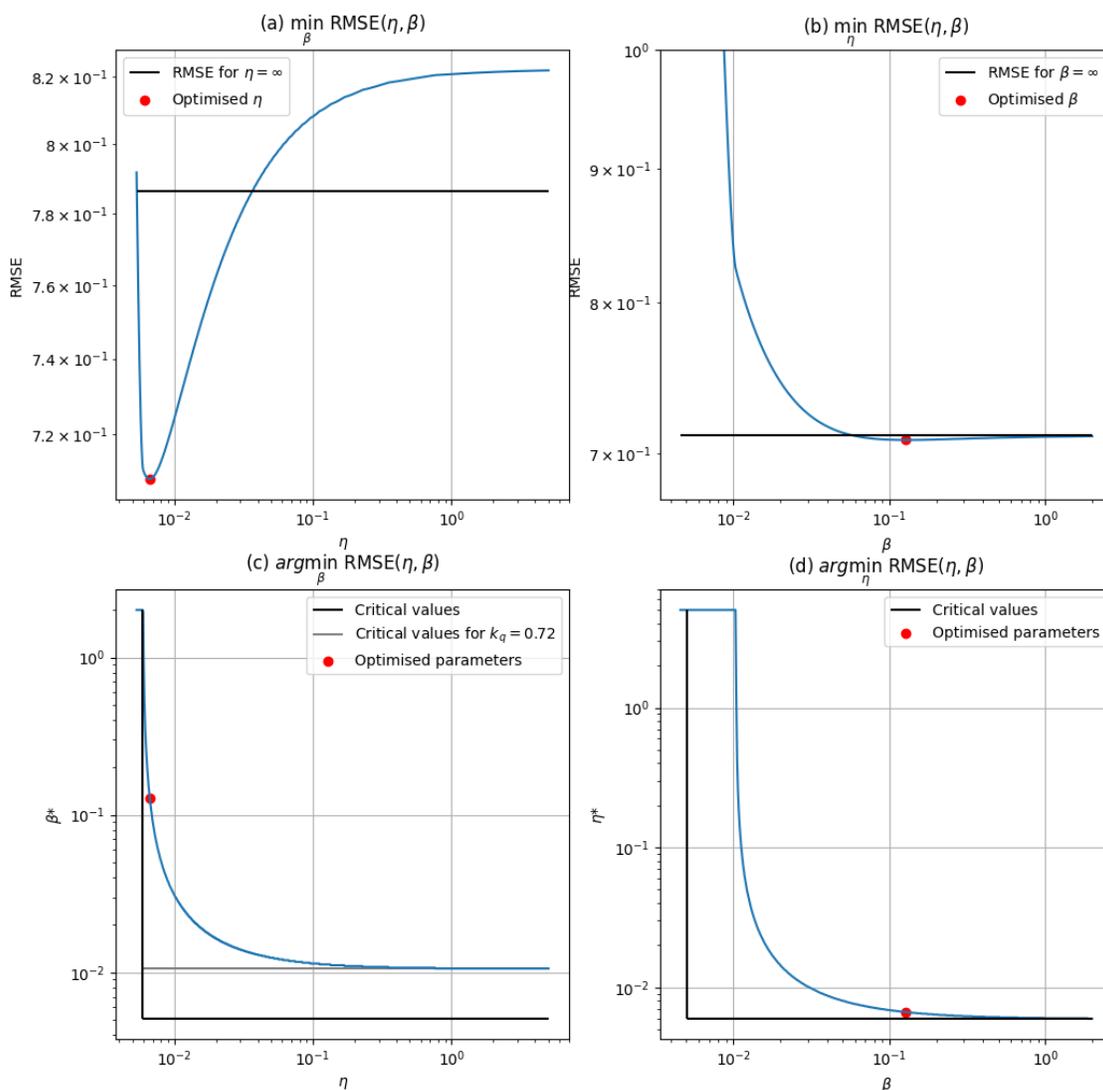


Figure 13: Visualisation of RMSE proprieties in case the case where the catalytic capacity of RuBisCO is not a limiting factor. Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\beta$  in function of  $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\eta$  in function of  $\beta$ .

Thus, we will consider henceforth that  $\eta^* = \infty$ . This leads to a simplified model that has a marginal impact on the accuracy, and is easier to calibrate. We should nevertheless keep in mind that there is large range of acceptable values of  $\eta$  for having a well calibrated model.

A similar figure representing a layers of  $\eta$  is given in figure 34 of Appendix E.

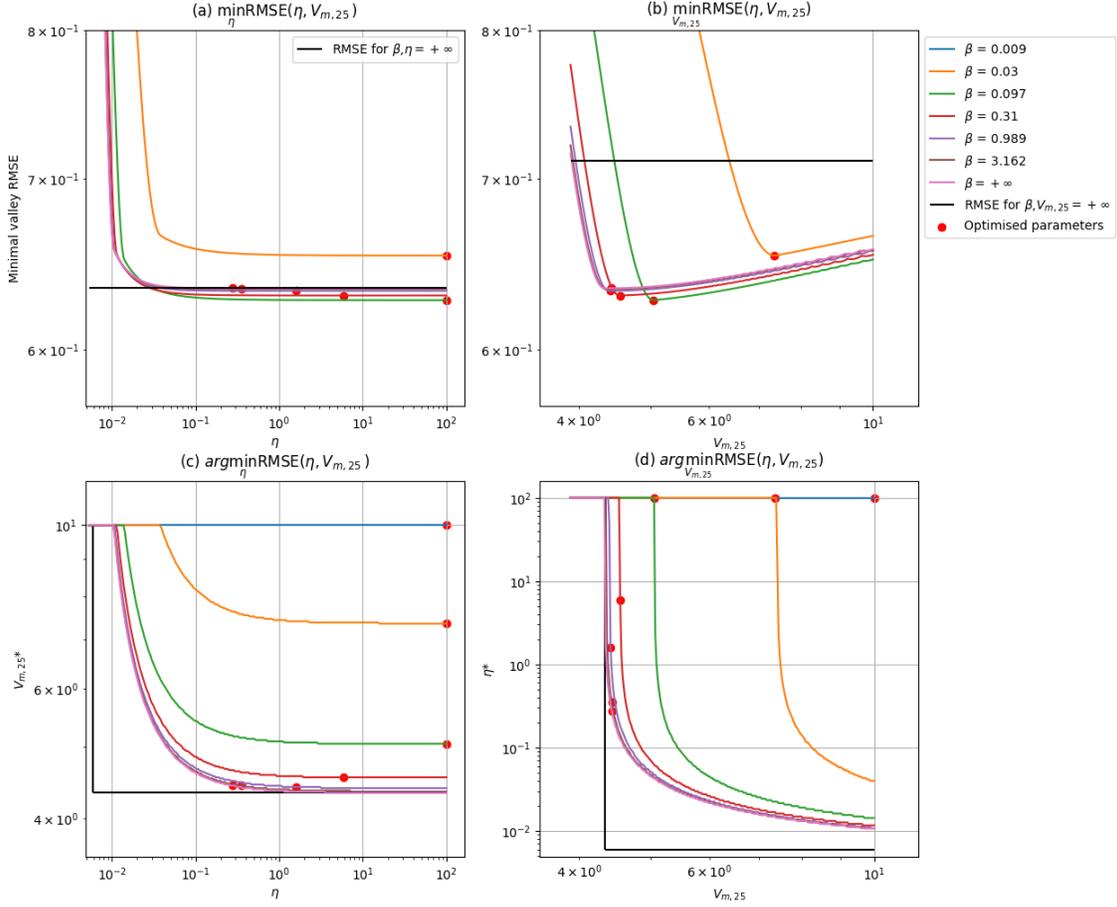


Figure 14: Visualisation of RMSE proprieties in the case where the catalytic capacity of RuBisCO, the  $\text{CO}_2$  and the light availabilities are limiting factors. The extinction coefficient of PAR is fixed at  $k_q = 0.72$ . Each curve represents a different value of  $\beta$ . Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $V_{m,25}$  in function of  $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\eta$  in function of  $V_{m,25}$ .

#### 5.4.4 Three-dimensional equations with $\eta \rightarrow \infty$

The simplified model is presented in equation 54. This time we will also consider variations in the light extinction.

$$\lim_{\eta \rightarrow +\infty} A_{c,g} = V_m \left[ LAI + \frac{1}{k_q} \ln \left( \frac{\beta I_0 + V_m}{\beta I_0 + V_m e^{k_q LAI}} \right) \right] \quad (54)$$

Figures 16, 17, and 18 represent the RMSE characteristics but with a different point of view every time. The proprieties of the grids are respectively given in Tables 11, 12, and 13.

First of all, figure 16 displays layers of  $k_q$ . For each curve, the optimal value of  $\beta$  is always significantly different: it suggests that  $\beta^*$  is strongly dependent on  $k_q$ . Furthermore, there is a small, but not negligible difference between the optimal value of the RMSE and the limit value

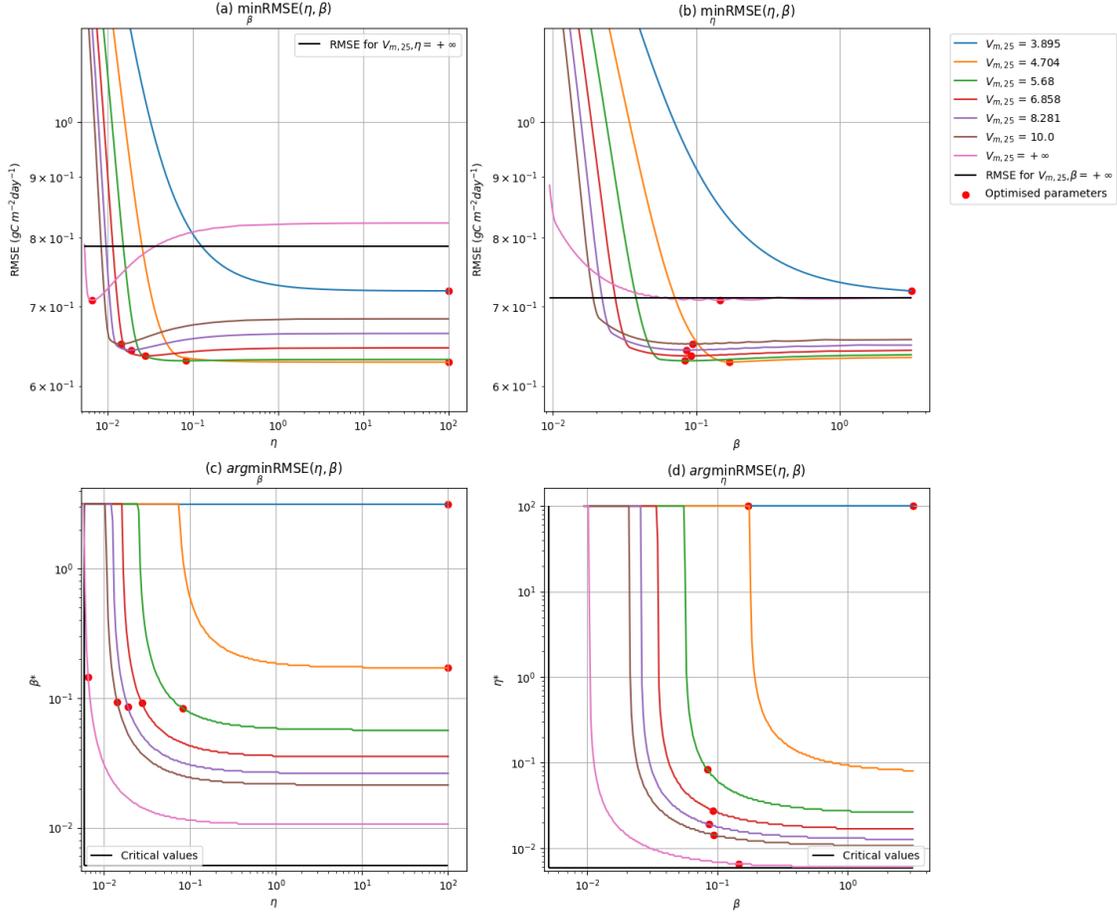


Figure 15: Visualisation of RMSE proprieties in the case where the catalytic capacity of RuBisCO, the  $\text{CO}_2$  and the light availabilities are limiting factors. The extinction coefficient of PAR is fixed at  $k_q = 0.72$ . Each curve represents a different value of  $V_{m,25}$ . Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\beta$  in function of  $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\eta$  in function of  $\beta$ .

when  $\beta$  goes to infinity. On the other hand,  $V_{m,25}^*$  does not majorly vary on the set of curves (figures b and c). More largely, the observations done when  $k_q = 0.72$  and  $k_q = 0$  can be generalised to all the other values. For instance, we observe the same monotony and injectivity propriety between  $V_{m,25}^*(\beta)$  and  $\beta$ , and between  $\beta^*(V_{m,25})$  and  $V_{m,25}$ . However, the values of  $V_{m,25}^*$  and  $\beta^*$  seem to increase with  $k_q$ . Indeed, if the amount of light in the model is diminishing, then the global estimation of GPP will lower. To balance that phenomenon, the efficiencies must rise.

In figure 17, we have the confirmation that  $\beta^*$  as a strong dependency relation to  $k_q$ . Reversely  $k_q^*$  also depends on  $\beta$  (figure a and c), but seems stable with the variations of  $V_{m,25}$  (figure b and d). We also have the validation, that  $k_q^*(\beta)$  is a strictly increasing function of  $\beta$ , and  $\beta^*(k_q)$  a strictly increasing function of  $k_q$ .

Finally, figure 18 is in agreement with the previous statements. For each curve corresponding to a different value of  $\beta$ , the optimal value of  $k_q$  also varies. On the other hand,  $V_{m,25}^*$  remains in a

Parameter	Min value	Max value	Number of points	Type of scale
$\beta$	$0.9 \times \beta_c$	$\sqrt{10}$	261	Log scale
$V_{m,25}$	$0.9 \times V_{m,25,c}$	70	262	Log scale
$k_q$	0.1	2	7	Linear scale

Table 11: Grid proprieties for the evaluation of the RMSE for  $\eta = \infty$  having  $\beta_c$ , and  $V_{m,25,c}$  respectively being the critical values for  $\beta$ , and  $V_{m,25}$ . The results of this configuration is illustrated in figure 16.

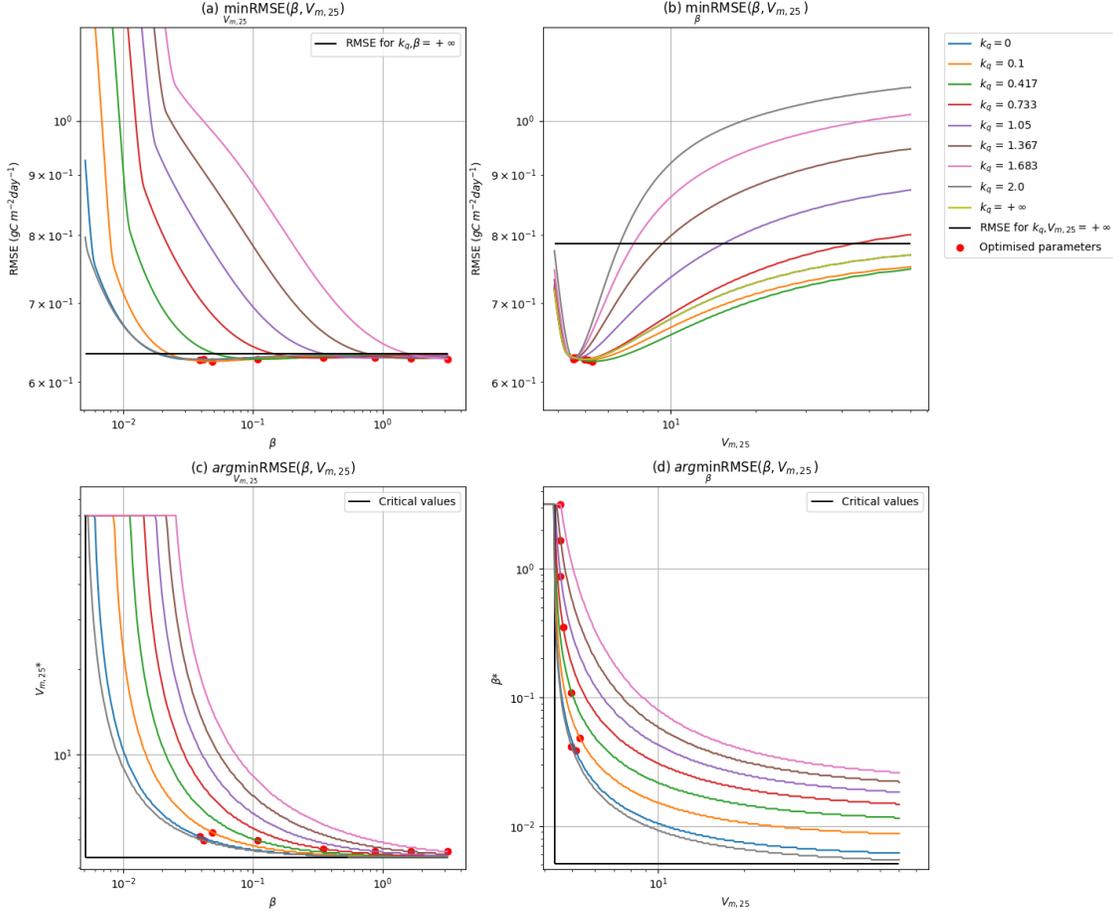


Figure 16: Visualisation of RMSE proprieties in the case where the  $\text{CO}_2$  availability is not a limiting factor. Each curve represents a different value of  $k_q$ . Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $V_{m,25}$  in function of  $\beta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\beta$  in function of  $V_{m,25}$ .

small interval. At last, the same monotony propriety for the argmin functions is observed.

Parameter	Min value	Max value	Number of points	Type of scale
$\beta$	$0.9 \times \beta_c$	$\sqrt{10}$	261	Log scale
$V_{m,25}$	$0.9 \times V_{m,25,c}$	70	7	Log scale
$k_q$	0.1	2	263	Linear scale

Table 12: Grid proprieties for the evaluation of the RMSE for  $\eta = \infty$  having  $\beta_c$ , and  $V_{m,25,c}$  respectively being the critical values for  $\beta$ , and  $V_{m,25}$ . The results of this configuration is illustrated in figure 17.

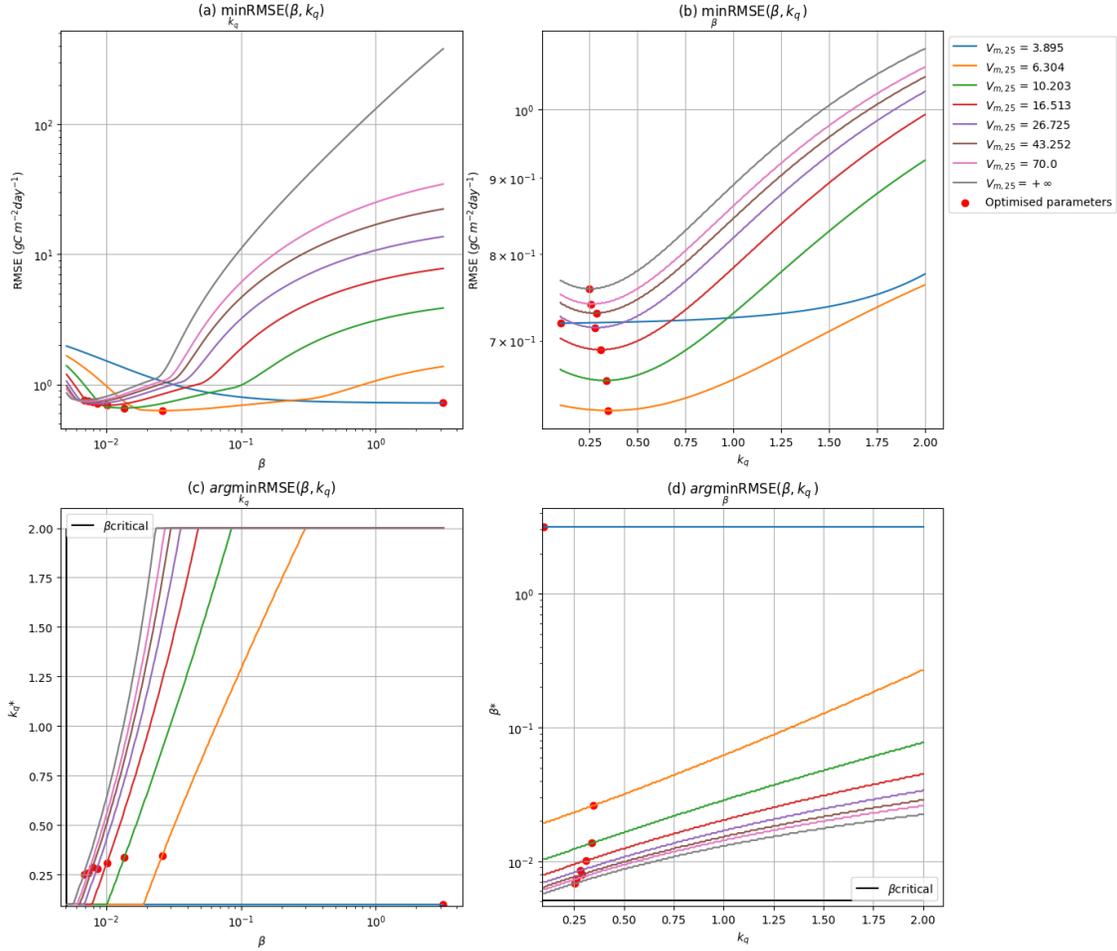


Figure 17: Visualisation of RMSE proprieties in the case where the catalytic capacity of RuBisCO, the  $\text{CO}_2$  and the light availabilities are limiting factors. The extinction coefficient of PAR is fixed at  $k_q = 0.72$ . Each curve represents a different value of  $V_{m,25}$ . Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\beta$  in function of  $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\eta$  in function of  $\beta$ .

Parameter	Min value	Max value	Number of points	Type of scale
$\beta$	$0.9 \times \beta_c$	$\sqrt{10}$	261	Log scale
$V_{m,25}$	$0.9 \times V_{m,25,c}$	70	7	Log scale
$k_q$	0.1	2	263	Linear scale

Table 13: Grid proprieties for the evaluation of the RMSE for  $\eta = \infty$  having  $\beta_c$ , and  $V_{m,25,c}$  respectively being the critical values for  $\beta$ , and  $V_{m,25}$ . The results of this configuration is illustrated in figure 18.

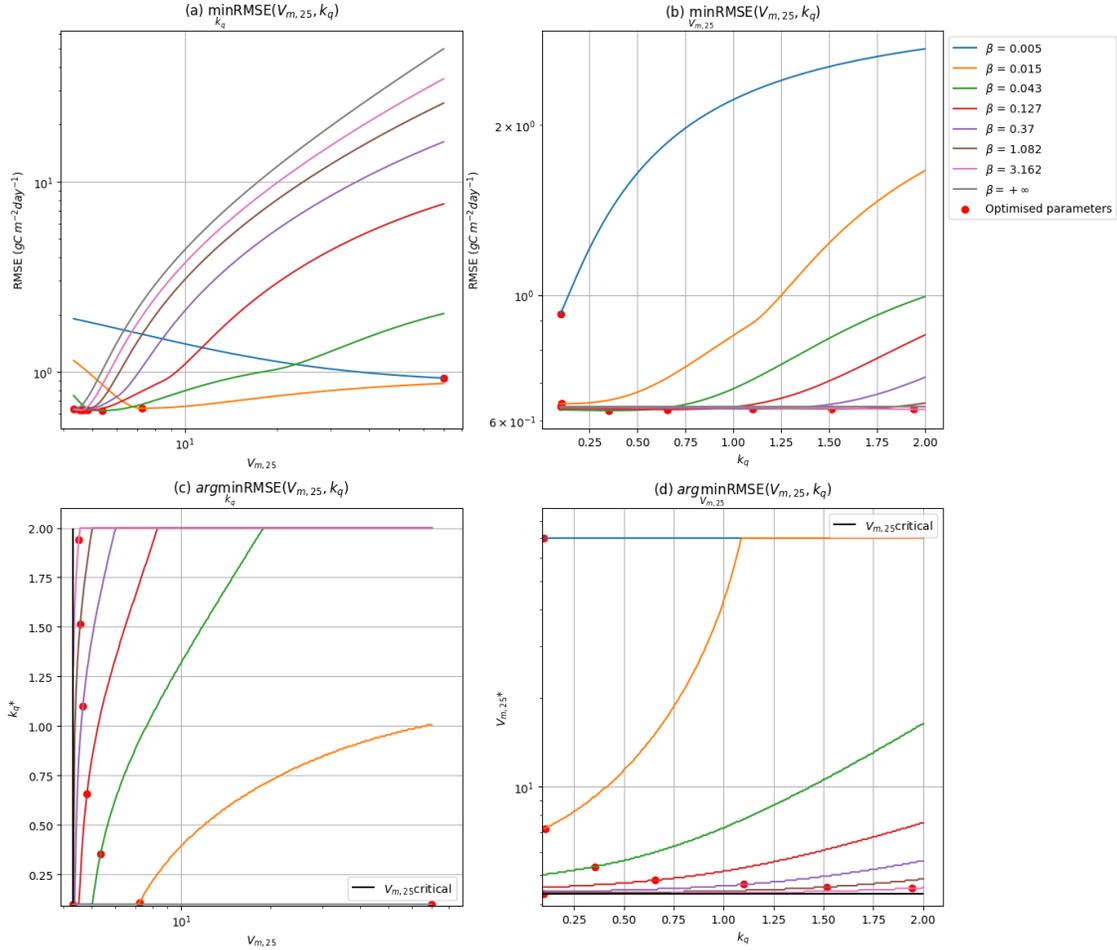


Figure 18: Visualisation of RMSE proprieties in the case where the catalytic capacity of RuBisCO, the CO<sub>2</sub> and the light availabilities are limiting factors. The extinction coefficient of PAR is fixed at  $k_q = 0.72$ . Each curve represents a different value of  $V_{m,25}$ . Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\beta$  in function of  $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\eta$  in function of  $\beta$ .

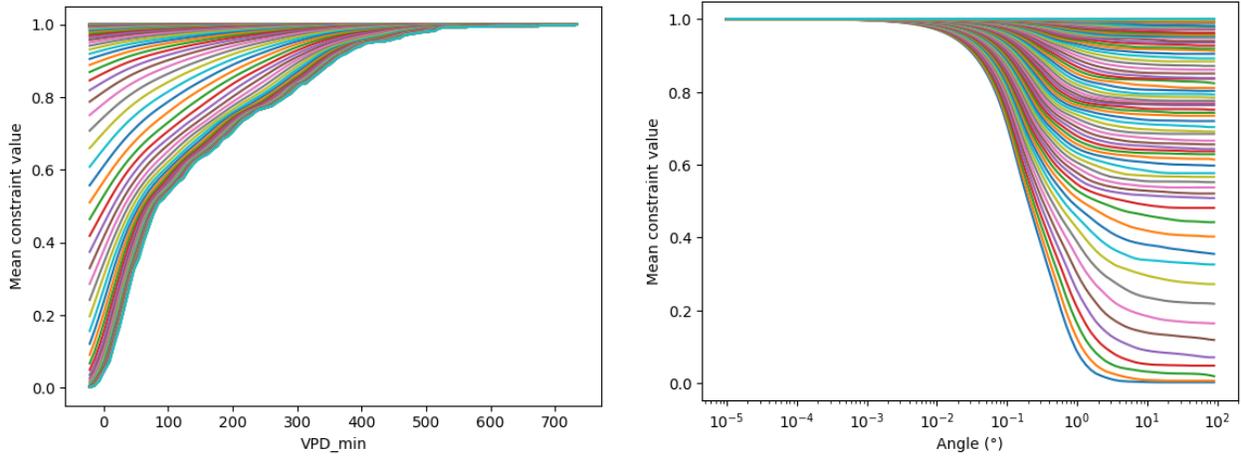
#### 5.4.5 Reset of the VPD constraint

The VPD constraint defined in equation 9 depends on two parameters: a maximum and a minimum threshold. However, we can also define a ramp function using one threshold ( $VPD_{min}$ ) and the

angle on the slope. The angle is defined as:

$$\alpha = \arctan \left( \frac{1}{VPD_{max} - VPD_{min}} \right) \quad (55)$$

Therefore, we can create many VPD constraints by defining the minimum threshold and angle of the slope. For this we chose to span  $VPD_{min}$  from the minimal value in the set to the maximum value, with a number of 100 points. The angle was span in a logspace starting from  $10^{-5}$  to  $90^\circ$  with 100 points as well. This builds us 10000 ramp function that are tested. The mean constraint value that is applied to the time-series is represented in figure 19.



(a) Mean value of the VPD constraint time-series depending on  $VPD_{min}$ . Every curve represent one angle value

(b) Mean value of the VPD constraint time-series depending on the angle. Every curve represent one  $VPD_{min}$  value

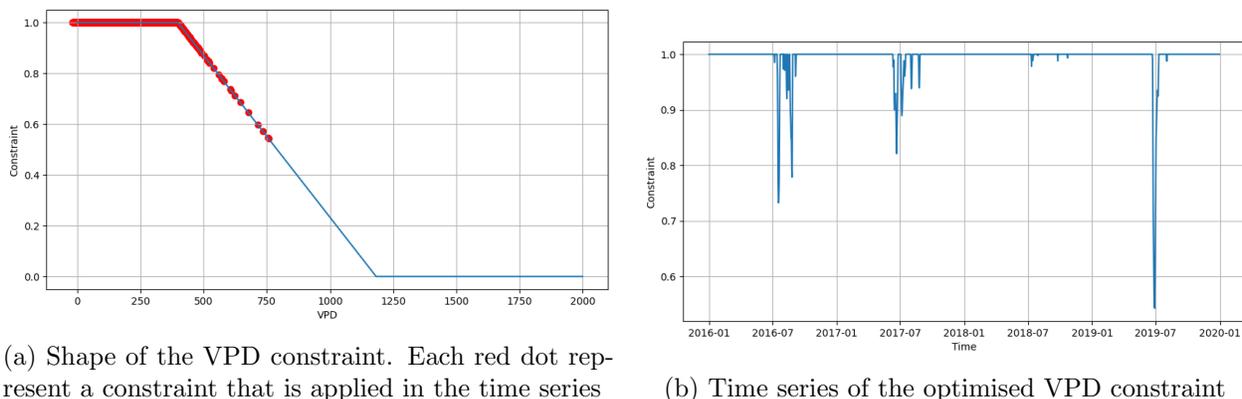
Figure 19: Visualisation of the mean value of the tested VPD constraints time-series, depending on the minimal threshold  $VPD_{min}$ , and the angle of the slope.

Using the previous optimisation, we select all the combination of parameters that result in a RMSE lower than 1% more then the minimal founded value. This is the case for 55195 combinations. The VPD constraints are applied to the output of each of those previous combinations, and the RMSE is recomputed on the training data. The constraint that results in the best estimation is kept. This constraint is shown in figure 20.

Notice that the constraint is never fully activated, and is partially activated only 7% of the time. This is justified by the low VPD context in Torgnon and the literature values.

### 5.5 Ensemble estimation and test set evaluation

The previous optimisation analysis described the behaviour of the objective function with respect to the parameters. We selected all the combinations of parameters that were close enough to the optimal combination. In this section, we will analyse this selection and investigate the effects of the perturbations on the test set and the global time series.



(a) Shape of the VPD constraint. Each red dot represent a constraint that is applied in the time series

(b) Time series of the optimised VPD constraint

Figure 20: Visualisation of the VPD constraint with the general shape of the function (20a), and it's value over time (20b).

### 5.5.1 Evaluation on the test set

The 55195 best points that have been founded on the training set are now evaluated on the test set. Figure 21 displays the RMSE scores on the test set. We first notice that, in the case where the computed optimal VPD constraint is applied, two clear clusters appear. We will see further on that they correspond to the application two different VPD parameters. When the VPD constraint is removed (or when the parameters are the same as in the literature), we no longer observe a distinction between the two initial clusters. Moreover, the obtained results are significantly better. This suggests that the tuning of the VPD constraint creates an overfitting situation. Also, it is interesting to notice that there is a slight correlation between low RMSE scores on the train set and low RMSE scores in the test set.

Figure 22 presents the time series of the GPP prediction of the optimised PML model. The difference between the prediction with and without the VPD constraints only happens on specific local points. With the VPD constraint, we observe big drops that significantly underestimate GPP. Those points are not present in the training phase of the PML parameters. Thus, we have a confirmation that the VPD constraint creates outliers that importantly increase the RMSE.

### 5.5.2 Distribution of the optimised parameters

Now that we have identified two clusters in the predictions, we will analyse the distribution of the parameter values. Figure 23 displays the violin plots of the parameter values in the ensemble. A violin plot is a more advanced form of box plots. In addition to the statistical distribution of values and quantiles, a violin plot also provides information about density shapes [76].

For the three parameters  $\beta$ ,  $V_{m,25}$  and  $k_q$ , there are no significant differences between the first and second cluster. In both of them, the obtained values of  $\beta$  are low, with the optimum value being close to the one in the literature. A few extreme values are also observed. Contrariwise, the literature value of  $V_{m,25}$  ( $46 \mu\text{mol m}^{-2}\text{s}^{-1}$ ) is much larger than the computed candidate values. This can be explained by the fact that we chose an infinite value for the carboxylation efficiency ( $\eta$ ), which narrows the range of acceptable values for  $V_{m,25}$ . The distribution of the values of  $k_q$  is homogeneous, which makes it difficult to predict.

On the contrary, the clusters identified in the RMSE on the test set are clearly visible for the

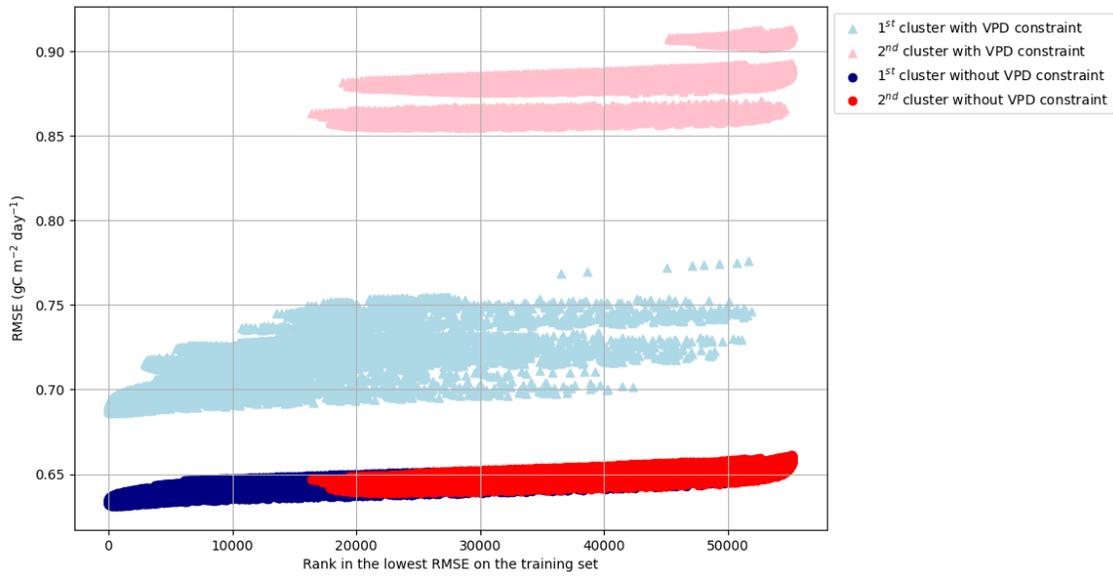


Figure 21: RMSE on the test set for the 55195 lowest RMSE scores on the training set. The pale triangle points are the results when the optimal VPD constraint is applied. They separated in two obvious clusters (blue and pink). The round dark points are the evaluation but without any constraint. The navy blue and the red correspond respectively to the first and second cluster.

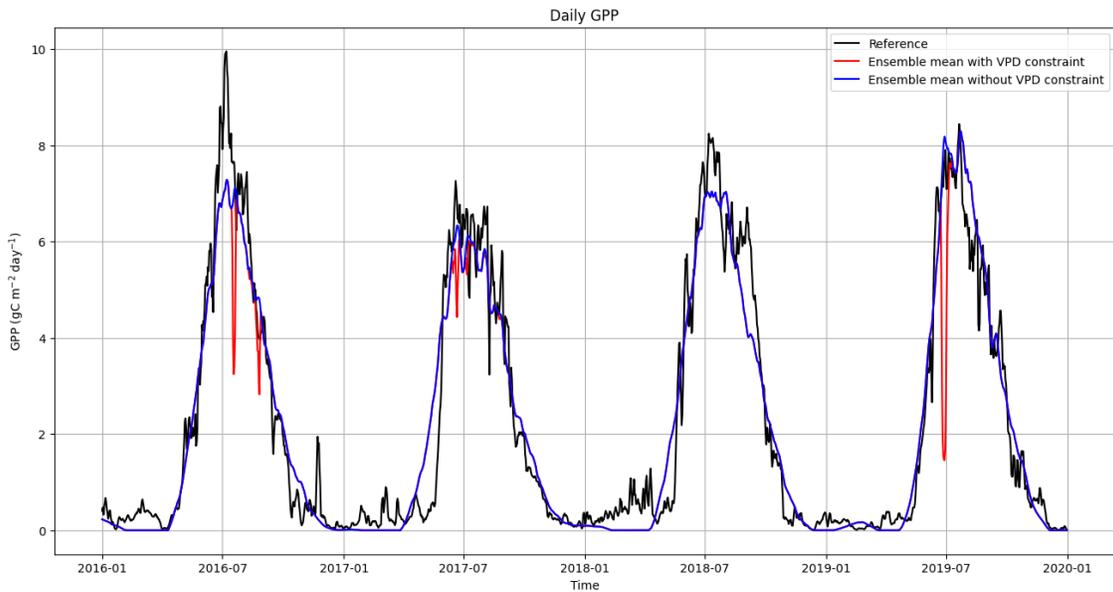


Figure 22: Time series of the estimated GPP. The black curve is the in-situ reference. The blue and red curves are the ensemble means of the optimised PML model respectively without and with the VPD constraint.

VPD parameters. This confirms that they are the cause of the separation of the two groups. Thus, in the Torgnon region, the measured VPD is too low to have an effect on GPP. The modelling of the

VPD effect can not only add complexity to the model, but also can worsen the estimates because of overfitting.

It is also important to mention that the parameter  $\eta$  is not represented in this figure because we chose a fixed value being  $+\infty$ . However, with the results obtained in section 5.4.3, we would expect a homogeneous distribution that would be divergent. Furthermore, the distribution of  $V_{m,25}$  would have a larger range.

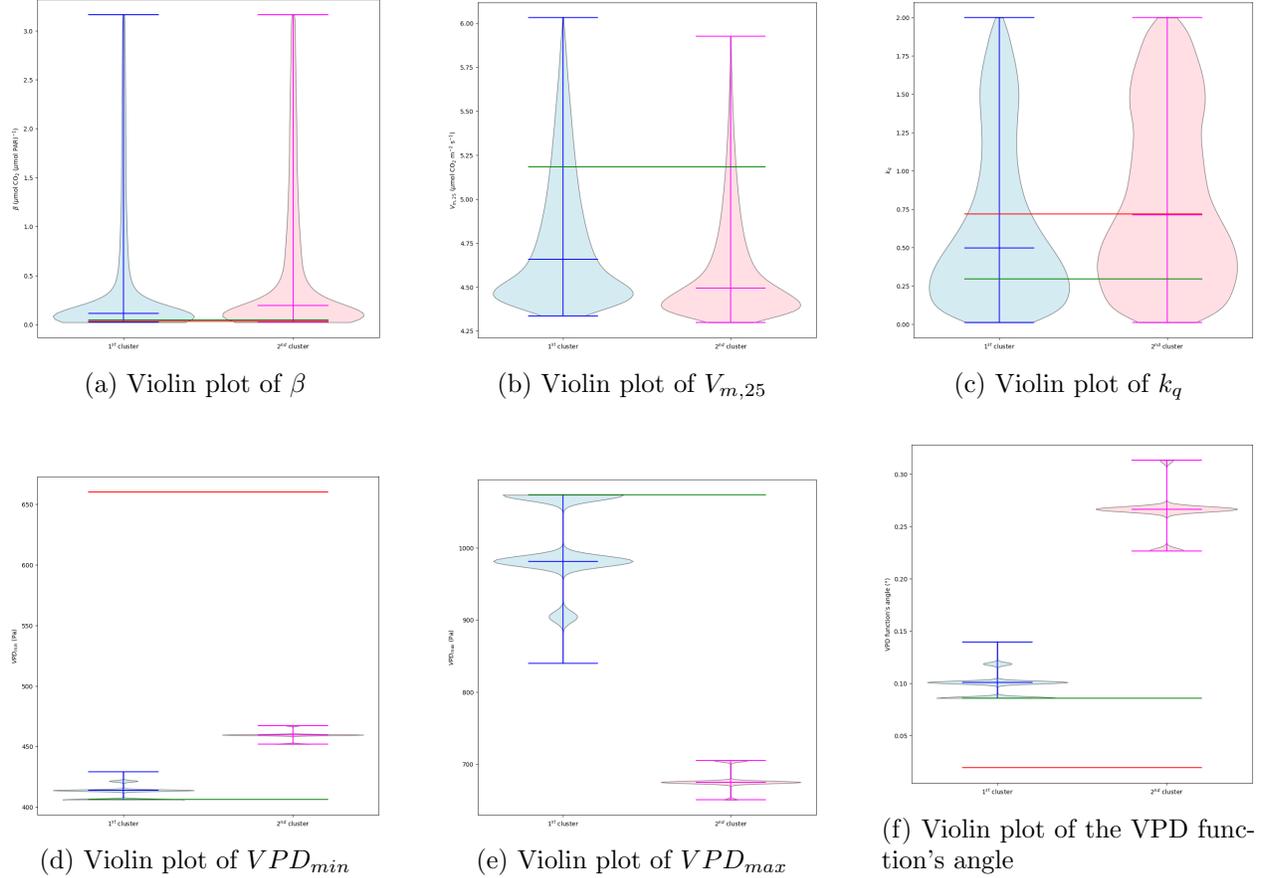


Figure 23: Violin plots of the parameter values in the ensemble. The blue and the pink violins are the parameter values in the first and second cluster respectively. The red and green horizontal lines are respectively the literature value and the optimal value in the train set.

### 5.5.3 Ensemble estimation

An ensemble estimation helps to understand to what extend errors in the parameter estimation lead to errors in the output prediction. Figure 24 shows the time series prediction of the ensemble.

The first observation is that the reference data is not located within the ensemble. This means that the uncertainties in the predictions goes beyond the uncertainties in the parameters. This is especially the case in the Winter season, where the spread is very narrow. This is explained by the extremely low LAI values that puts the estimation to zero.

Secondly, the ensemble spread is narrow. If we look at the standard deviation of the ensemble

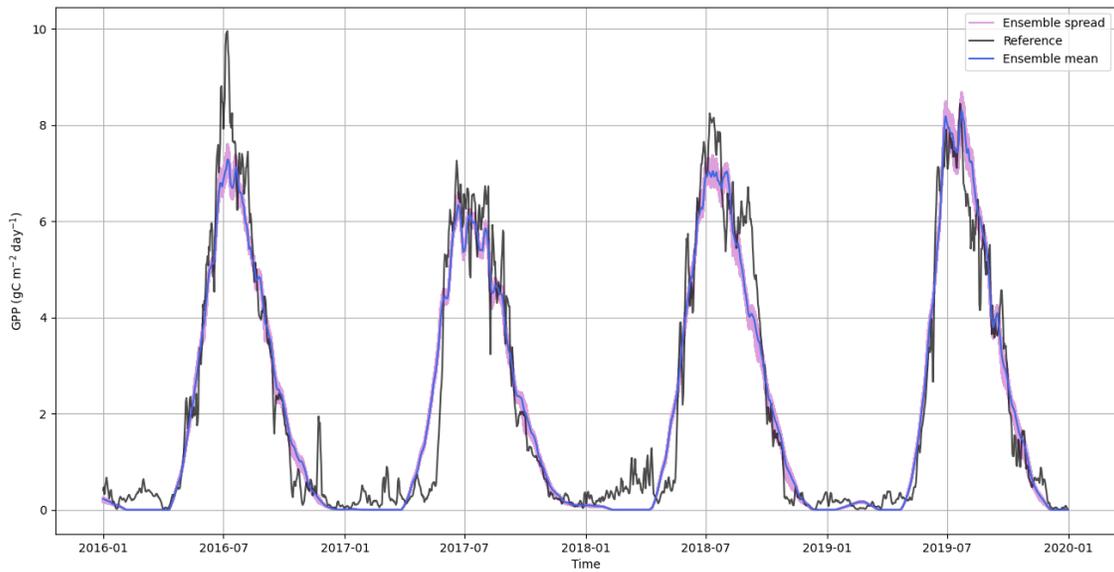


Figure 24: Ensemble time-series of the estimated GPP. The black curve is the in-situ reference. The pink curves each represent one prediction of the ensemble. The blue curve is the ensemble mean of the optimised PML model without the VPD constraint.

over time (figure 25), we notice a larger spread in the Spring and Summer season, when the primary production is high. Reversely, the spread is low in Winter, when the productivity is low. Thus, the outputs of the model are similar despite a large range in the parameter values. This is because the value of one parameter influences another. Therefore, it is hard to assess the ground value of the parameters.

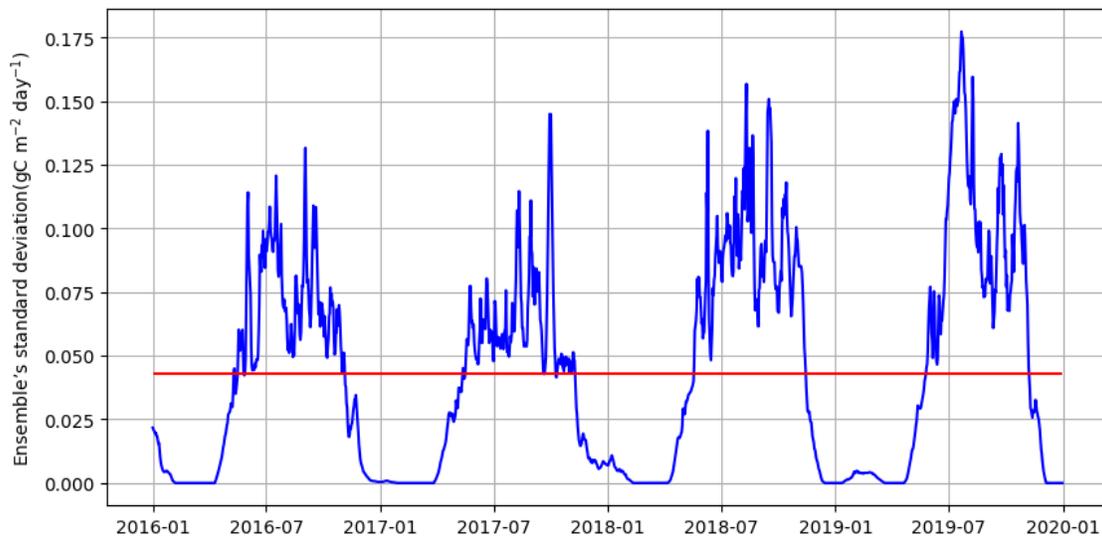


Figure 25: Ensemble's standard deviation time series. The blue curve is the value of the standard deviation over time. The red curve is the average standard deviation value.

Finally, the ensemble mean has a slight higher RMSE and lower  $R^2$  ( $0.643 \text{ gC m}^{-2} \text{ day}^{-1}$  and  $0.940$ ) than the optimal combination of parameters on the train set ( $0.634 \text{ gC m}^{-2} \text{ day}^{-1}$  and  $0.942$ ). Moreover, still on the test set, the best and worst predictions of the ensemble have respectively a RMSE of  $0.631$  and  $0.661 \text{ gC m}^{-2} \text{ day}^{-1}$ . So a 1% difference between the minimum and maximum RMSE value in the train set leads to a 4.5 % in the test set.

Thus, the calibration procedure is reliable for the estimation of GPP, having the optimal values in the train set remaining optimal on the test set. However, it is harder to derive the biophysical properties of the ecosystem.

## 6 Discussion

### 6.1 Uncertainties in the methodology

#### 6.1.1 Uncertainties in situ measurements

The in-situ measurements of GPP were considered as the reference ground truth data. However, those measurements are also subject to noise and uncertainties.

The first source of errors concerns the acquisition of the Net Ecosystem Exchange of CO<sub>2</sub> (NEE). Those can come from errors in flux data (turbulence, measurements system [77], [78]), footprint variations in irregular ecosystems [79], and instrumental limits (measurement frequency...).

Moreover, GPP is estimated by following an algorithm that processes the NEE data. For example, Papale’s method, that was applied to derive the in-situ measurements, uses an algorithm to filter outlier spikes. Also, it uses turbulence correction and gap filling to derive GPP. The first and last procedures are destruction and reconstruction of data that result in a better data quality but that is still imperfect. The  $u^*$  correction is a specific method for solving underestimations issues during calm nights. The selection of the thresholds for this method is the main source of inaccuracy. Overall, Papale estimates that his method has a 10 % uncertainty [8].

Furthermore, Baldocchi remarks that several studies have quantified the measurement error of NEE in different landscapes [80]. In flat landscapes, the error is low (about 30 gC.m<sup>-2</sup>.year<sup>-1</sup> out of the 200 gC.m<sup>-2</sup>.year<sup>-1</sup> in the Harvard forest [81]). However, he points out that larger uncertainties (between 20 and 60 %) happen in more uneven topography ([82], [83], and [84]). This can be explained by winder conditions.

In our case, the Torgnon measurement tower is located in the mountains at more than 2000m of altitude. So the measurement errors linked to complex topography are likely to be present.

#### 6.1.2 Uncertainties in input variables

As we saw in chapter 4, there are several ways of retrieving the input data. Those may have an influence in the obtained results. For instance, we observed in figure 1 underestimations in temperature and VPD. However, those have a minor impact in the output. For VPD, this is due to the low values in the Torgnon regions. Therefore the atmosphere is not dry enough to have an impact on primary production.

The used dataset for concentration of CO<sub>2</sub> in the atmosphere has a poor spatial and temporal resolution. However, looking at the order of magnitude of the variations over the study period, the impact of the fluctuation is limited.

On the contrary, variations of PAR have a large outcome in the prediction of GPP. The original ”raw” time series contains many high frequencies terms. Those are filtered using a moving 8 days average. Therefore, the smoothing of PAR can have an impact in the *a posteriori* sensitivity analysis, and in the calibration procedure.

On an even larger note, the LAI has the biggest impact on GPP and is the most uncertain data. The MCD15 dataset is subject to uncertainties in the methodology of estimation of LAI, and with presence of noise in the measurements (clouds, snow...). To cope with this issue, a Whittaker smoother is applied to the signal. This reconstruction is not perfect and it’s parametrisation relies on arbitrarily choices. Moreover, the change of the smoother’s parameters have large consequences on the processed time-series. The optimal smoothing of the LAI would be an interesting subject to investigate in further details.

### 6.1.3 Differences between self implemented PML model in GEE product

As we can see in figure 5, the PML model that was implemented in this study has a different output than the available PML product in GEE. This is the case despite using the same datasets (GLDAS 2.1, MCD15) and parameter values (the values for grassland ecosystems).

First of all, the online product has a temporal resolution of eight days, which is less precise than the our daily resolution. Also, the data processing follows a slightly different methodology. Unlike in this study, the computed product uses a spike filter on the meteorological data [85]. Furthermore, the precise settings of the Whittaker smoother is a sophisticated and computationally expensive process [86].

On a less important scale, all of the GEE data was captured on a set of pixels corresponding to the footprint (area for which the GPP measurements are captured). So, the GPP from GEE, the LAI, and the GLDAS data were spatially aggregated as the average value of the pixels. However, the PML model is not linear. So, the GPP computed from the spatial mean of the input variables is different from the spatial mean of the GPP computed for all pixels.

Thus, those differences in the methodology result in differences in the computed product.

## 6.2 Diversity of data

The results derived in this study fit into a specific framework. Indeed, the data is limited in space and in time, with unequal quality.

### 6.2.1 Temporal diversity of data

First of all, the used data is limited in time. GPP is a seasonal variable, so we need at least a full year of data yo have suitable results. In our case, we have four. This is sufficient to capture the yearly characteristics but not the long term changes. For instance, after four years, the average temperature and CO<sub>2</sub> concentration have a marginal variations. They are however noteworthy when longer periods (decades for example) are considered.

Within a year, the data is not equally worthy. We saw with the correlation analysis that all of the input data of the PML model is not correlated with the reference GPP in Winter. Also, the non zero values in the in-situ measurements is not compatible with the zero LAI values. This is either due to a modelling problem (assumptions that no longer stand in Winter), an issue with methodology in the measurements, or an underestimated LAI.

The seasonal analysis can be pushed in further details by calibrating the model using only one season as the training data. For this we only consider the simplified model when  $\eta \rightarrow \infty$  and we use the literature’s VPD constraint. The optimisation is done using a grid search whose proprieties are described in Table 14.

Parameter	Min value	Max value	Number of points	Type of scale
$\beta$	0.0051	$\sqrt{10}$	102	Log scale
$V_{m,25}$	3.90	70	103	Log scale
$k_q$	0.1	2	104	Linear scale

Table 14: Grid proprieties for the evaluation of the RMSE using seasonal training

The time series of the calibrated model for each season are illustrated in figure 26. We notice that when the training is done using only Spring, or only Summer, or only Autumn data, the

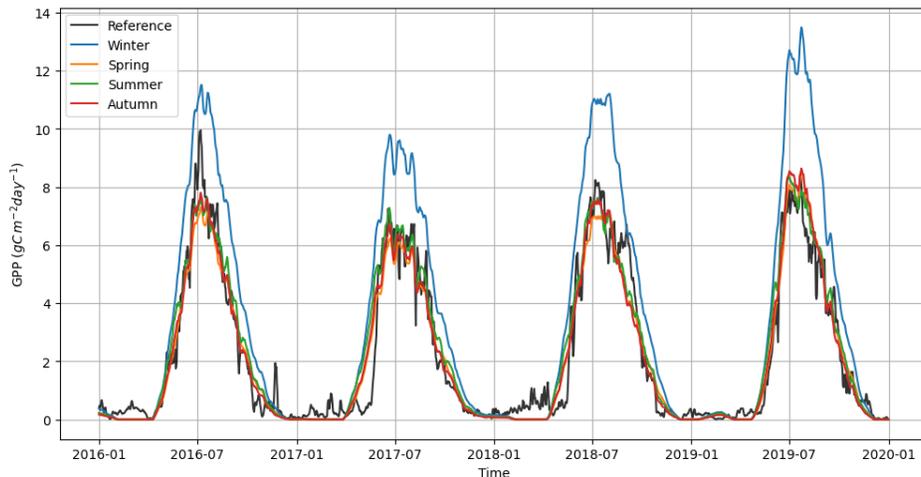


Figure 26: Time series of the PML model when the training data comes from different season. The black curve is the in situ data. The blue, orange, green, and red curves respectively represent training in Autumn, Winter, Spring and Summer.

estimates are accurate. Nevertheless, as expected, the model using Winter data overestimates the in-situ measurements. This is due to the low values of input variables in Winter, coupled with non zero values in the reference data.

Finally, the calibrated parameters and RMSEs for each scenario are presented in Table 15. We have the confirmation that, when using the Winter data, we have an overfitting of results. Moreover, the extinction coefficient is relevant only in Summer. In the three other seasons, it reaches the minimum value of the search grid. This underlines the fact that light extinction does no longer effect GPP when the LAI is close to zero. In addition, the values for  $\beta$  and  $V_{m,25}$  are different per season. It underlines the difficulty to estimates those biophysical factors. Also, the PML model simplifies the characteristics of the plants by assuming them constant through time. However, those proprieties can also have seasonal fluctuation. It is for example the case of the leaf maximum photosynthetic capacity at 25°C  $V_{m,25}$  [87].

Season	$\beta$	$V_{m,25}$	$k_q$	RMSE test set	RMSE train set
Winter	3.16	6.67	0.1	2.21	0.29
Spring	1.47	4.24	0.1	0.61	0.74
Summer	0.028	8.37	0.75	0.65	0.81
Autumn	0.022	5.96	0.1	0.70	0.36

Table 15: Optimal parameters with RMSEs on the test and train set using seasonal training

### 6.2.2 Spatial diversity of data

The PML model and the used datasets are designed to be global. But this study is focused on the application to the particular case of Torgnon subalpine grassland. It has a different behaviour from other grasslands because of the mountainous context. So, the specific region very likely overfits the results. We can for example observe it with the significant differences between the estimates

using the literature’s calibration, and the estimates specifically optimised for Torgnon. For instance, if we use the constraint derived in the literature, the VPD does not restraint GPP. However, the optimised constraint on the training data gives VPD a role to play. We then saw on the test set (figure 21) that it was an overfit. This is probably caused by the particularly low VPD in the region.

It would therefore be interesting to perform the same study on other highlands. The results can be compared, and the drawn conclusions of this study may or may not be extrapolated.

In addition, the calibration can be performed simultaneously on different study sites. In that way, the global phenomena may have larger influence on GPP. This is particularly the case of the CO<sub>2</sub> concentration of the atmosphere. The dataset that was used describes the global temporal patterns, but not the spatial patterns. The low (negative) correlations between the measured GPP and the CO<sub>2</sub> data may not be observed using distanced study sites. Figure 27 shows the map of concentration of CO<sub>2</sub> in the atmosphere at different time periods. The distribution is not homogeneous. Nevertheless, the scale of the spatial dispersion is in the same range as the scale in the NOAA dataset we used. So, the impact of the CO<sub>2</sub> concentration in the atmosphere on GPP can still remain marginal.

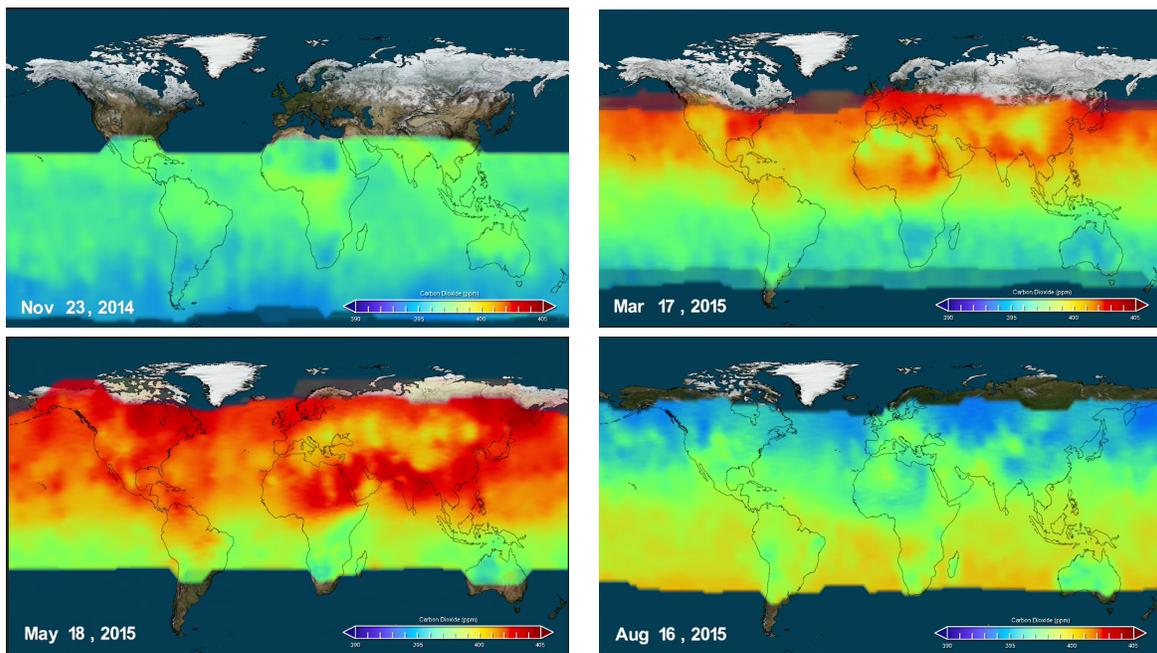


Figure 27: Spatial patterns of CO<sub>2</sub> at four different dates. Screenshots of a NASA video (<https://svs.gsfc.nasa.gov/12072>)

## 7 Conclusion

This study focused on the application and the sensitivity analysis of the input variables and parameters of the PML model to the Torgnon subalpine grassland.

First of all, the PML model can be applied to the Torgnon subalpine grassland, but the values of parameters founded in literature are not optimal. Moreover, using those literature values, the VPD is no longer a limiting factor for the carbon assimilation.

Furthermore, the sensitivity analysis showed that the inputs are not equally important. LAI is by far the most influential variable. Temperature and PAR are also meaningful but in a lesser way. In the case of the low VPD context, the associated constraint is not as consequential as it should be.

The time-series of CO<sub>2</sub> concentration in the atmosphere has a minor impact on GPP. This is explained by either a low quality of the dataset, a study period being too short, or a phenomenon that does not limit in practice the assimilation.

Finally, the calibration procedure is reliable in the sense that the founded optimal values in the train set are also optimal in the test set. Similar low objective function scores predict similar outputs. However, little conclusions can be said about the founded parameter values. The values resulting in a good estimate stand on a large range. Moreover, the parameters are strongly dependent with each other. So, uncertainty in the estimate of one parameter will lead to uncertainty in the estimate of another parameter.

## A MOD17 functions

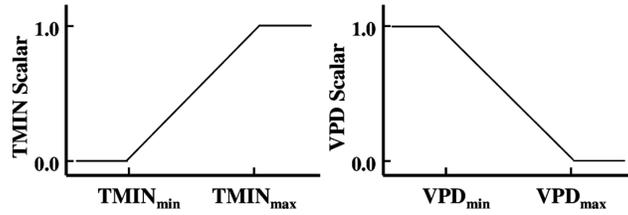
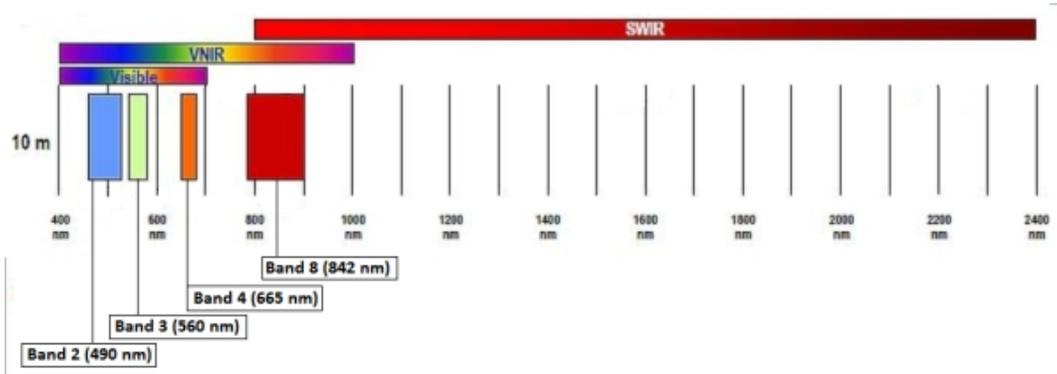


Figure 28: Ramp functions of minimal daily air temperature ( $TMIN$ ) and Vapour Pressure Deficit ( $VPD$ ) (picture from [88])

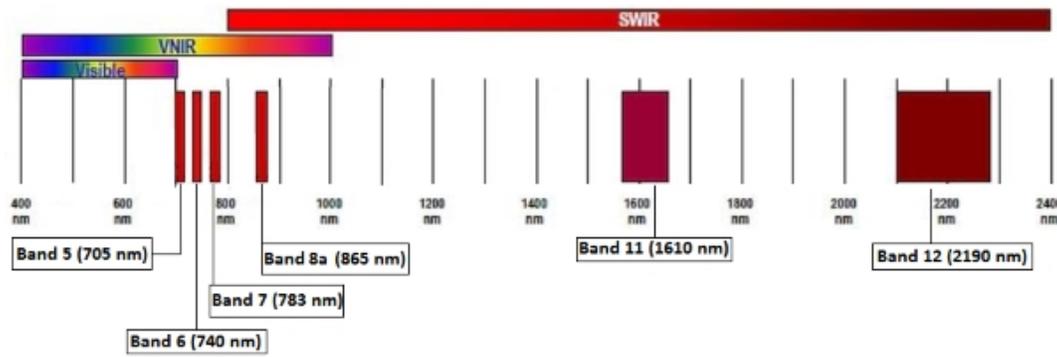
UMD_VEG_LC	ENF	EBF	DNF	DBF	MF	CShrub	OShrub	WSavanna	Savanna	Grass	Crop
LUE <sub>max</sub> (KgC/m <sup>2</sup> /d/MJ)	0.000962	0.001268	0.001086	0.001165	0.001051	0.001281	0.000841	0.001239	0.001206	0.000860	0.001044
T <sub>min_min</sub> (C)	-8.00	-8.00	-8.00	-6.00	-7.00	-8.00	-8.00	-8.00	-8.00	-8.00	-8.00
T <sub>min_max</sub> (C)	8.31	9.09	10.44	9.94	9.50	8.61	8.80	11.39	11.39	12.02	12.02
VPD <sub>min</sub> (Pa)	650.0	800.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0
VPD <sub>max</sub> (Pa)	4600.0	3100.0	2300.0	1650.0	2400.0	4700.0	4800.0	3200.0	3100.0	5300.0	4300.0

Table 16: Biome-Property-Look-Up-Table (BPLUT) for MODIS GPP/NPP algorithm with NCEP-DOE reanalysis II and the Collection5 FPAR/LAI as inputs. The full names for the University of Maryland land cover classification system (UMD\_VEG\_LC) in MOD12Q1 dataset (fieldname: Land\_Cover\_Type\_2) are, Evergreen Needleleaf Forest (ENF), Evergreen Broadleaf Forest (EBF), Deciduous Needleleaf Forest (DNF), Deciduous Broadleaf Forest (DBF), Mixed forests (MF), Closed Shrublands (CShrub), Open Shrublands (OShrub), Woody Savannas (WSavanna), Savannas (Savanna), Grassland (Grass), and Croplands (Crop) [88].

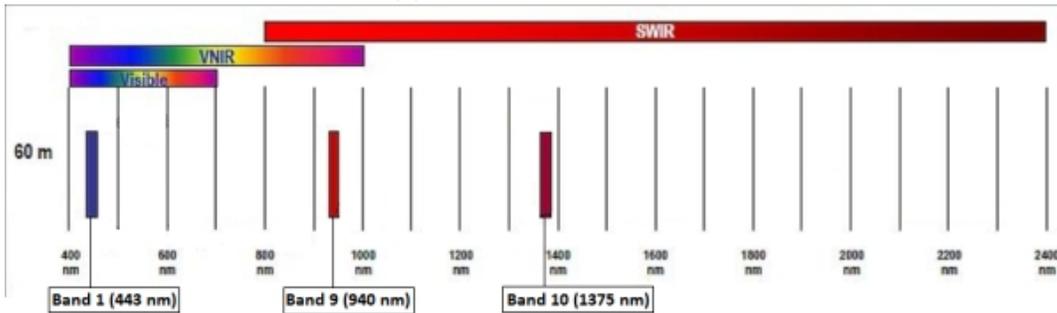
## B Remote Sensing Data



(a) 10m resolution bands



(b) 20m resolution bands



(c) 60m resolution bands

Figure 29: Spectral Bands of Sentinel-2 with their resolution [66].

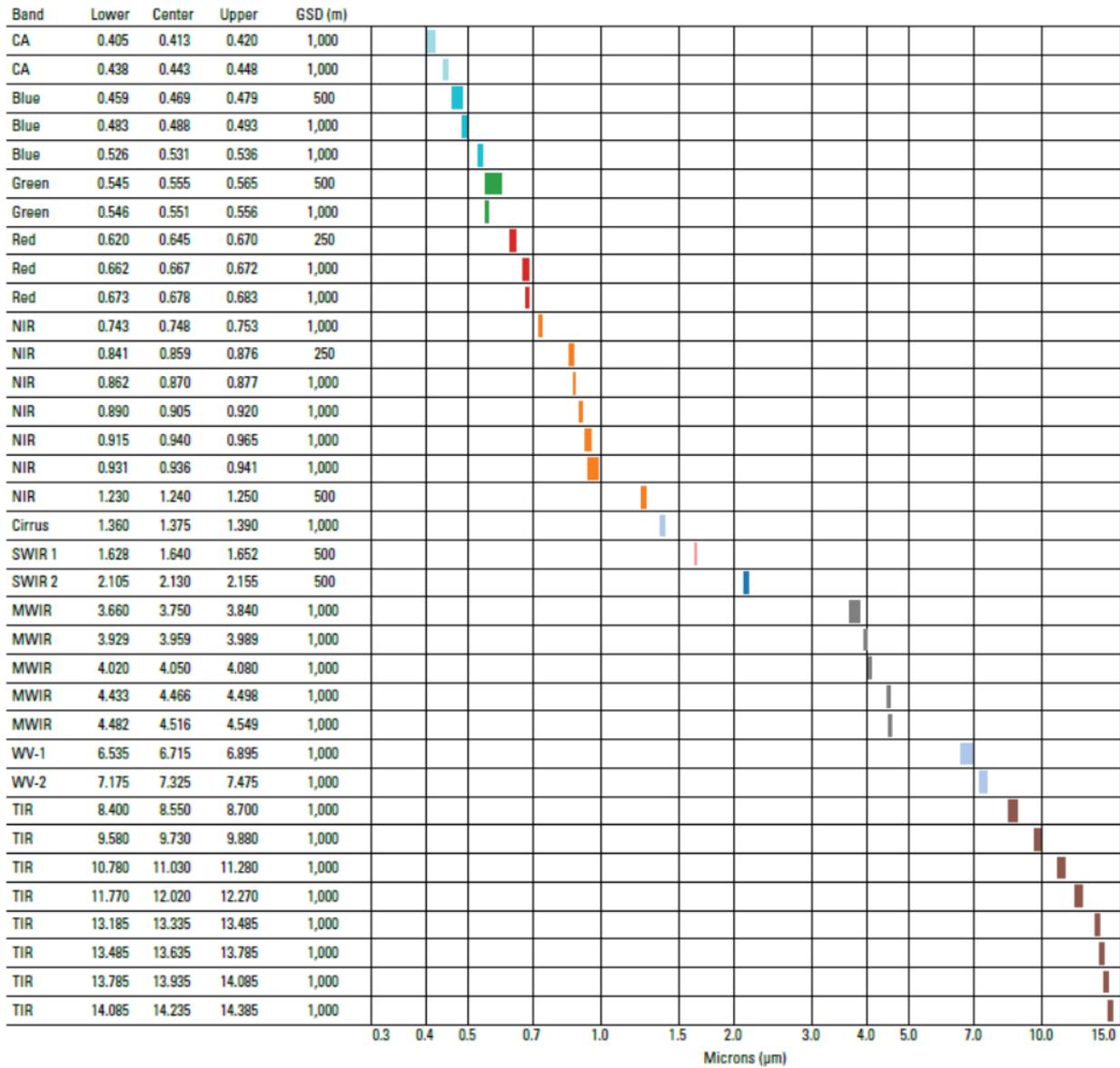


Figure 30: Modis bands (picture from [89])

## C Spearman's correlation matrices

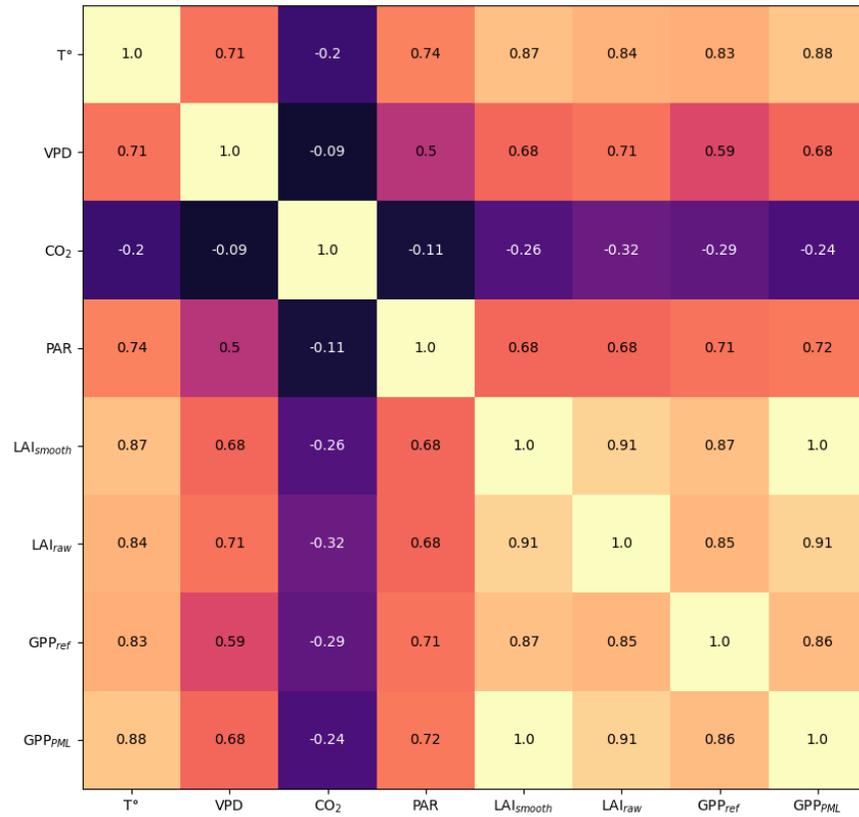


Figure 31: Spearman's correlation matrix of the variables and GPP. The color bar is the same as in figure 6.

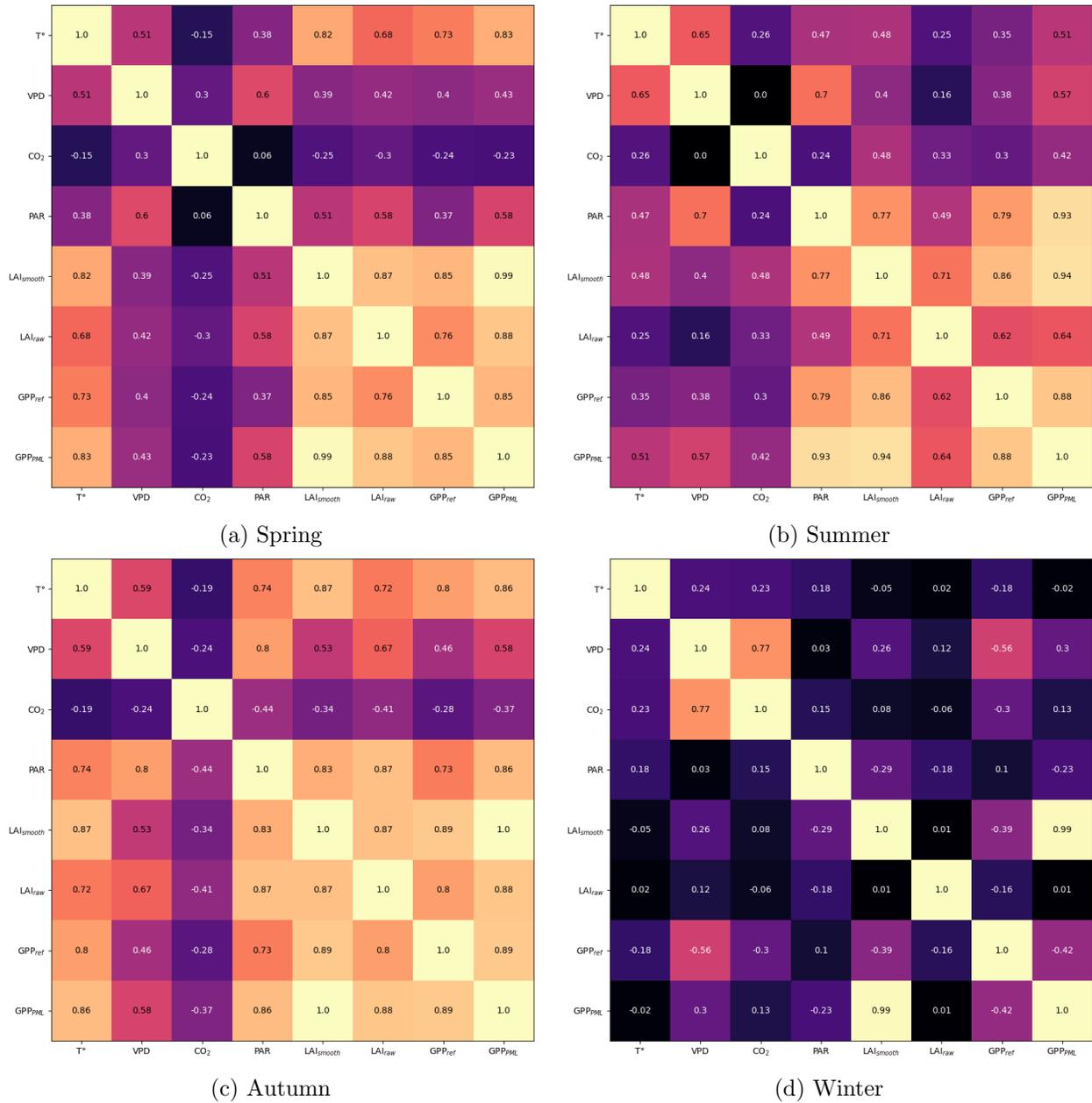


Figure 32: Spearman's correlation matrix over the seasons. The color bar is the same as in figure 6.

## D Computation of the limits and asymptotic equations

## Limits of the gross assimilation rate $A_{cg}$

$$A_{cg} = \frac{V_m n c_a}{b_q (V_m + n c_a)} \left[ b_q LAI + \ln \left( \frac{(V_m + n c_a) \beta I_0 + V_m n c_a}{(V_m + n c_a) \beta I_0 + V_m n c_a e^{b_q LAI}} \right) \right]$$

• For  $\beta, I_0 \rightarrow +\infty$

$$\ln \left( \frac{(V_m + n c_a) \beta I_0 + V_m n c_a}{(V_m + n c_a) \beta I_0 + V_m n c_a e^{b_q LAI}} \right) = \ln \left( 1 + \frac{V_m n c_a}{(V_m + n c_a) \beta I_0} \right) - \ln \left( 1 + \frac{V_m n c_a e^{b_q LAI}}{(V_m + n c_a) \beta I_0} \right)$$

$\xrightarrow{\beta, I_0 \rightarrow +\infty} 0$

So,  $\lim_{\beta, I_0 \rightarrow +\infty} A_{cg} = \frac{V_m n c_a}{V_m + n c_a} LAI$

• For  $\beta, I_0 \rightarrow 0$ .

$$\ln \left( \frac{(V_m + n c_a) \beta I_0 + V_m n c_a}{(V_m + n c_a) \beta I_0 + V_m n c_a e^{b_q LAI}} \right) \xrightarrow{\beta, I_0 \rightarrow 0} \ln \left( \frac{V_m n c_a}{V_m n c_a e^{b_q LAI}} \right) = -b_q LAI$$

So,  $\lim_{\beta, I_0 \rightarrow 0} A_{cg} = 0$

• For  $n, c_a \rightarrow +\infty$

$$\ln \left( \frac{(V_m + n c_a) \beta I_0 + V_m n c_a}{(V_m + n c_a) \beta I_0 + V_m n c_a e^{b_q LAI}} \right) = \ln \left( \frac{V_m \beta I_0 + \frac{V_m \beta I_0}{n c_a}}{V_m e^{b_q LAI} + \beta I_0 + \frac{V_m \beta I_0}{n c_a}} \right) \xrightarrow{n, c_a \rightarrow +\infty} \ln \left( \frac{V_m + \beta I_0}{V_m e^{b_q LAI} + \beta I_0} \right)$$

and,  $\frac{V_m n c_a}{V_m + n c_a} \xrightarrow{n, c_a \rightarrow +\infty} V_m$ .

So,  $\lim_{n, c_a \rightarrow +\infty} A_{cg} = \frac{V_m}{\beta} \left[ b_q LAI + \ln \left( \frac{\beta I_0 + V_m}{\beta I_0 + V_m e^{b_q LAI}} \right) \right]$

①

• For  $\eta, C_a \rightarrow 0$

$$\lim_{\eta, C_a \rightarrow 0} \ln \left( \frac{(V_m + \eta(C_a)) \beta I_0 + V_m \eta C_a}{(V_m + \eta(C_a)) \beta I_0 + V_m \eta C_a e^{b_q L A I}} \right) = \ln \left( \frac{V_m \beta I_0}{V_m \beta I_0} \right) = 0$$

and  $\frac{V_m \eta C_a}{b_q (V_m + \eta(C_a))} \underset{\eta, C_a \rightarrow 0}{\sim} \frac{\eta C_a}{b_q}$

So  $A_{c, g} \underset{\eta, C_a \rightarrow 0}{\sim} \frac{\eta C_a}{b_q} [b_q L A I]$

$$A_{c, g} \underset{\eta, C_a \rightarrow 0}{\sim} \eta C_a L A I$$

• For  $V_m \rightarrow +\infty$

$$\ln \left( \frac{(V_m + \eta(C_a)) \beta I_0 + V_m \eta C_a}{(V_m + \eta(C_a)) \beta I_0 + V_m \eta C_a e^{b_q L A I}} \right) = \ln \left( \frac{\eta(C_a + \beta I_0) + \frac{\eta C_a \beta I_0}{V_m}}{\eta C_a e^{b_q L A I} + \beta I_0 + \frac{\eta C_a \beta I_0}{V_m}} \right)$$

$$\xrightarrow{V_m \rightarrow +\infty} \ln \left( \frac{\eta(C_a + \beta I_0)}{\eta C_a e^{b_q L A I} + \beta I_0} \right)$$

$$\text{So } \lim_{V_m \rightarrow +\infty} A_{c, g} = \frac{\eta C_a}{b_q} \left[ b_q L A I + \ln \left( \frac{\beta I_0 + \eta C_a}{\beta I_0 + \eta C_a e^{b_q L A I}} \right) \right]$$

• For  $V_m \rightarrow 0$

With the same reasoning as for  $\eta$  and  $C_a$ ,  $A_{c, g} \underset{V_m \rightarrow 0}{\sim} V_m L A I$

• For  $k_q, LAI \rightarrow 0$

The Taylor approximation (order 1) of the exponential function and the  $x \mapsto \ln(1+x)$  function are respectively:

$$\exp(x) = 1 + x + o(x)$$

$$\ln(1+x) = x + o(x)$$

Therefore,

$$\ln \left( \frac{(V_{m+n}(a) \beta I_0 + V_m \eta c_a)}{(V_{m+n}(a) \beta I_0 + V_m \eta c_a e^{k_q LAI})} \right)$$

$$= \ln \left( \frac{(V_{m+n}(a) \beta I_0 + V_m \eta c_a)}{(V_{m+n}(a) \beta I_0 + V_m \eta c_a (1 + k_q LAI + o(k_q LAI)))} \right)$$

$$= \ln \left( \frac{1}{1 + \frac{V_m \eta c_a k_q LAI + o(k_q LAI)}{(V_{m+n}(a) \beta I_0 + V_m \eta c_a)}} \right)$$

$$= - \frac{V_m \eta c_a k_q LAI}{(V_{m+n}(a) \beta I_0 + V_m \eta c_a)} + o(k_q LAI)$$

$$\text{So, } A_{c,q} = \frac{V_m \eta c_a}{k_q (V_{m+n}(a))} \left[ k_q LAI - \frac{V_m \eta c_a k_q LAI}{(V_{m+n}(a) \beta I_0 + V_m \eta c_a)} + o(k_q LAI) \right]$$

$$= \frac{V_m \eta c_a}{(V_{m+n}(a))} \left[ \frac{(V_{m+n}(a) \beta I_0 + V_m \eta c_a - V_m \eta c_a)}{(V_{m+n}(a) \beta I_0 + V_m \eta c_a)} \right] LAI + o(LAI)$$

$$\text{Thus, } A_{c,q} \underset{k_q LAI \rightarrow 0}{\sim} \frac{V_m \eta c_a \beta I_0}{V_m \eta c_a + V_m \beta I_0 + \eta c_a \beta I_0} LAI$$

(3)

• For  $k_q \text{ LAI} \rightarrow +\infty$

$$\ln \left( \frac{(V_m + n(a)) \beta I_0 + V_m n(a)}{(V_m + n(a)) \beta I_0 + V_m n(a) e^{k_q \text{ LAI}}} \right)$$

$$= \ln \left( (V_m + n(a)) \beta I_0 + V_m n(a) \right) - \ln \left( 1 + \frac{(V_m + n(a)) \beta I_0}{V_m n(a) e^{k_q \text{ LAI}}} \right) - \ln \left( V_m n(a) e^{k_q \text{ LAI}} \right)$$

$$= \ln \left( (V_m + n(a)) \beta I_0 + V_m n(a) \right) - \ln \left( V_m n(a) e^{k_q \text{ LAI}} \right) - \frac{(V_m + n(a)) \beta I_0}{V_m n(a) e^{k_q \text{ LAI}}} + o \left( \frac{1}{e^{k_q \text{ LAI}}} \right)$$

$$\text{So, } A_{c,g} = \frac{V_m n(a)}{k_q (V_m + n(a))} \left[ \ln \left( 1 + \frac{(V_m + n(a)) \beta I_0}{V_m n(a)} \right) - \frac{(V_m + n(a)) \beta I_0}{V_m n(a) e^{k_q \text{ LAI}}} + o \left( e^{-k_q \text{ LAI}} \right) \right]$$

Thus, we have,  $\lim_{\text{LAI} \rightarrow +\infty} A_{c,g} = \frac{V_m n(a)}{k_q (V_m + n(a))} \ln \left( 1 + \frac{(V_m + n(a)) \beta I_0}{V_m n(a)} \right)$

And,  $A_{c,g} \underset{k_q \rightarrow +\infty}{\sim} \frac{V_m n(a)}{k_q (V_m + n(a))} \ln \left( 1 + \frac{(V_m + n(a)) \beta I_0}{V_m n(a)} \right) \xrightarrow{k_q \rightarrow +\infty} 0$

• For  $k_q$  and  $\beta \rightarrow +\infty$

Let  $A_L = \frac{V_m n(a)}{V_m + n(a)}$ . So,  $A_{c,g} = \frac{A_L}{k_q} \left[ k_q \text{ LAI} + \ln \left( \frac{\beta I_0 + A_L}{\beta I_0 + A_L e^{k_q \text{ LAI}}} \right) \right]$ .

We assume that  $k_q \approx \frac{\beta I_0}{A_L} \rightarrow +\infty$  (H1)

$$\text{So } \ln \left( \frac{\beta I_0 + A_L}{\beta I_0 + A_L e^{k_q \text{ LAI}}} \right) = \ln(\beta I_0 + A_L) - \ln(\beta I_0 + A_L e^{k_q \text{ LAI}})$$

$$= \ln(\beta I_0) + \ln \left( 1 + \frac{A_L}{\beta I_0} \right) - \ln \left( A_L e^{k_q \text{ LAI}} \right) - \ln \left( 1 + \frac{\beta I_0}{A_L e^{k_q \text{ LAI}}} \right)$$

$$= \ln(\beta I_0) + \frac{A_L}{\beta I_0} + o \left( \frac{A_L}{\beta I_0} \right) - \ln(A_L) - k_q \text{ LAI} - \frac{\beta I_0}{A_L e^{k_q \text{ LAI}}} + o \left( \frac{\beta I_0}{A_L e^{k_q \text{ LAI}}} \right)$$

Hence,  $A_{c,g} = \frac{A_L \ln(\beta I_0)}{k_q} + \frac{A_L^2}{k_q \beta I_0} - \frac{A_L \ln(A_L)}{k_q} - \frac{\beta I_0}{k_q e^{k_q \text{ LAI}}} + \frac{1}{k_q} o \left( \frac{A_L}{\beta I_0} \right) + \frac{1}{k_q} o \left( \frac{\beta I_0}{A_L e^{k_q \text{ LAI}}} \right)$  (1)

Using (H1),  $\frac{\beta I_0}{e^{k_q \text{ LAI}}} \rightarrow 0$  and  $\frac{A_L \ln(\beta I_0)}{k_q} \rightarrow 0$

Thus, all of the terms in equation (1) converge to 0.

So,  $\lim_{k_q \rightarrow +\infty} A_{c,g} = 0$

(9)

## E Other RMSE figures

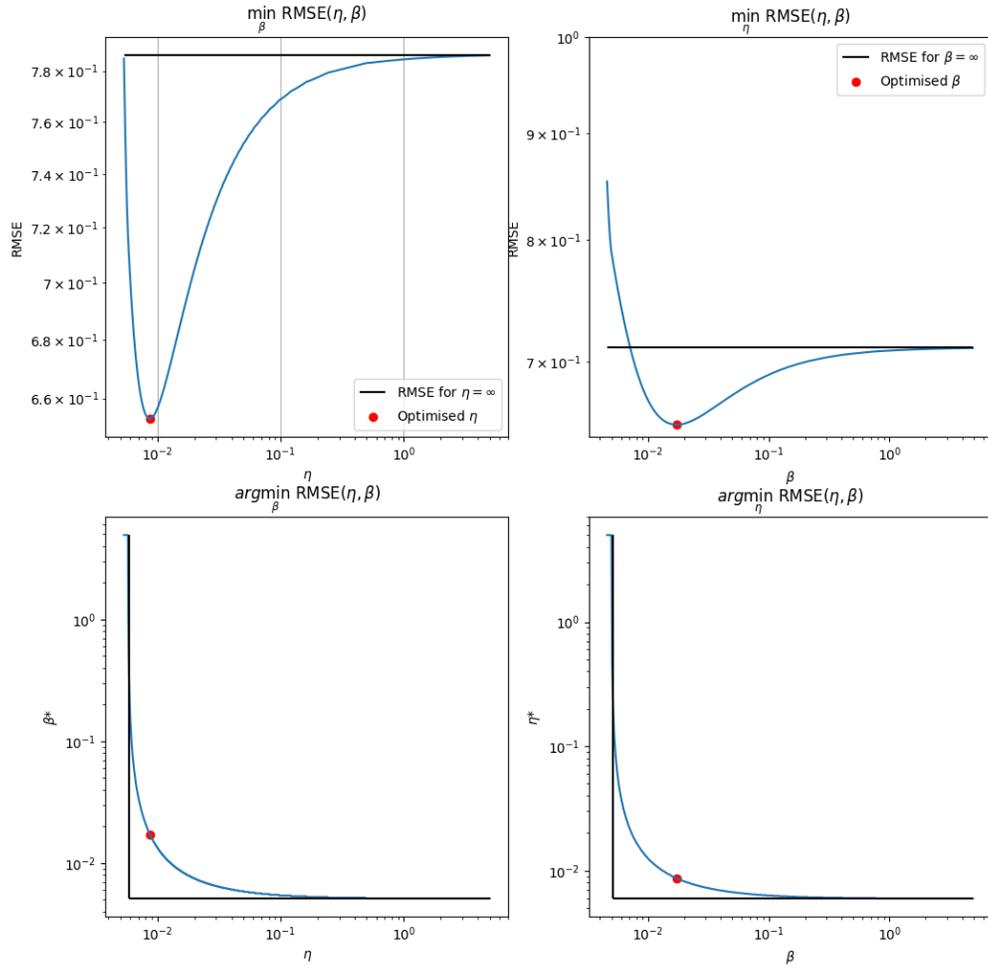


Figure 33: Visualisation of RMSE properties in case the catalytic capacity of RuBisCO is not a limiting factor. Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\beta$  in function of  $\eta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\eta$  in function of  $\beta$ .

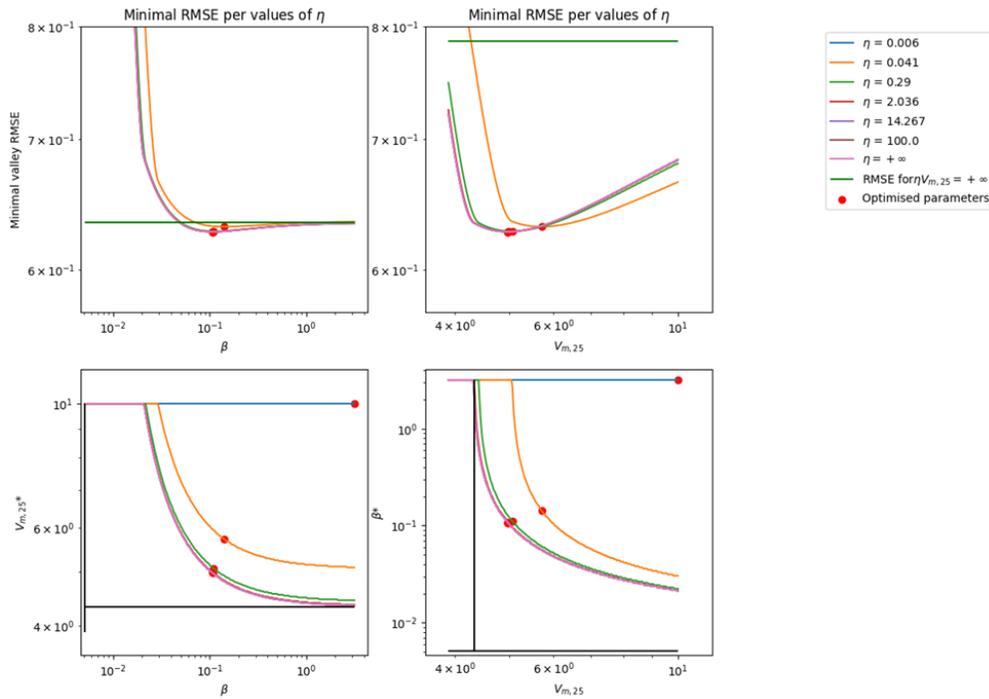


Figure 34: Visualisation of RMSE proprieties in the case where the catalytic capacity of RuBisCO, the  $\text{CO}_2$  and the light availabilities are limiting factors. The extinction coefficient of PAR is fixed at  $k_q = 0.72$ . Each curve represents a different value of  $\eta$ . Figure (a) and (c) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $V_{m,25}$  in function of  $\beta$ . Figure (b) and (d) display respectively the minimum and the argument of the minimum of the RMSE with respect to  $\beta$  in function of  $V_{m,25}$ .

## References

- [1] Millennium Ecosystem Assessment. *Ecosystems and Human Well-being: Synthesis*. Washington, DC: Island Press, 2005.
- [2] H. M. Pereira et al. “Essential Biodiversity Variables”. In: *Science* 339.6117 (2013), pp. 277–278. DOI: [10.1126/science.1229931](https://doi.org/10.1126/science.1229931). eprint: <https://www.science.org/doi/pdf/10.1126/science.1229931>. URL: <https://www.science.org/doi/abs/10.1126/science.1229931>.
- [3] Christian Beer et al. “Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate”. In: *Science* 329.5993 (2010), pp. 834–838. DOI: [10.1126/science.1184984](https://doi.org/10.1126/science.1184984). eprint: <https://www.science.org/doi/pdf/10.1126/science.1184984>. URL: <https://www.science.org/doi/abs/10.1126/science.1184984>.
- [4] Wei Li et al. “Recent Changes in Global Photosynthesis and Terrestrial Ecosystem Respiration Constrained From Multiple Observations”. In: *Geophysical Research Letters* 45.2 (2018), pp. 1058–1068. DOI: <https://doi.org/10.1002/2017GL076622>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017GL076622>. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL076622>.
- [5] Andrew K Skidmore et al. “Priority list of biodiversity metrics to observe from space”. In: *Nature ecology & evolution* 5.7 (2021), pp. 896–906.
- [6] Lisa Welp et al. “Interannual variability in the oxygen isotopes of atmospheric CO<sub>2</sub> driven by El Niño”. In: *Nature* 477 (Sept. 2011), pp. 579–82. DOI: [10.1038/nature10421](https://doi.org/10.1038/nature10421).
- [7] Markus Reichstein et al. “On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm”. English. In: *Global Change Biology* 11.9 (2005), pp. 1424–1439. ISSN: 1354-1013. DOI: [10.1111/j.1365-2486.2005.001002.x](https://doi.org/10.1111/j.1365-2486.2005.001002.x).
- [8] D. Papale et al. “Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation”. In: *Biogeosciences* 3.4 (2006), pp. 571–583. DOI: [10.5194/bg-3-571-2006](https://doi.org/10.5194/bg-3-571-2006). URL: <https://bg.copernicus.org/articles/3/571/2006/>.
- [9] Eva Falge et al. “Gap filling strategies for long term energy flux data sets”. In: *Agricultural and Forest Meteorology* 107.1 (2001), pp. 71–77. ISSN: 0168-1923. DOI: [https://doi.org/10.1016/S0168-1923\(00\)00235-5](https://doi.org/10.1016/S0168-1923(00)00235-5). URL: <https://www.sciencedirect.com/science/article/pii/S0168192300002355>.
- [10] GITTA LASSLOP et al. “Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation”. In: *Global Change Biology* 16.1 (2010), pp. 187–208. DOI: <https://doi.org/10.1111/j.1365-2486.2009.02041.x>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2486.2009.02041.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2486.2009.02041.x>.
- [11] Dennis Baldocchi et al. “FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities”. In: *Bulletin of the American Meteorological Society* 82.11 (2001), pp. 2415–2434. DOI: [10.1175/1520-0477\(2001\)082<2415:FANTTS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2). URL: [https://journals.ametsoc.org/view/journals/bams/82/11/1520-0477\\_2001\\_082\\_2415\\_fantts\\_2\\_3\\_co\\_2.xml](https://journals.ametsoc.org/view/journals/bams/82/11/1520-0477_2001_082_2415_fantts_2_3_co_2.xml).

- [12] Yelu Zeng et al. “Optical vegetation indices for monitoring terrestrial ecosystems globally”. In: *Nature Reviews Earth & Environment* 3.7 (2022), pp. 477–493.
- [13] National Aeronautics and Space Administration. *MODIS*. <https://aqua.nasa.gov/modis>. Accessed: 10-04-2024.
- [14] Steven W. Running et al. “A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production”. In: *BioScience* 54.6 (June 2004), pp. 547–560. ISSN: 0006-3568. DOI: [10.1641/0006-3568\(2004\)054\[0547:ACSMOG\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2); eprint: <https://academic.oup.com/bioscience/article-pdf/54/6/547/26895742/54-6-547.pdf>. URL: [https://doi.org/10.1641/0006-3568\(2004\)054%5B0547:ACSMOG%5D2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054%5B0547:ACSMOG%5D2.0.CO;2).
- [15] Yanyan Pei et al. “Evolution of light use efficiency models: Improvement, uncertainties, and implications”. In: *Agricultural and Forest Meteorology* 317 (2022), p. 108905. ISSN: 0168-1923. DOI: <https://doi.org/10.1016/j.agrformet.2022.108905>. URL: <https://www.sciencedirect.com/science/article/pii/S0168192322000983>.
- [16] David J. Pannell. “Sensitivity analysis of normative economic models: theoretical framework and practical strategies”. In: *Agricultural Economics* 16.2 (1997), pp. 139–152. ISSN: 0169-5150. DOI: [https://doi.org/10.1016/S0169-5150\(96\)01217-0](https://doi.org/10.1016/S0169-5150(96)01217-0).
- [17] Andrea Saltelli et al. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley, 2004. URL: <https://api.semanticscholar.org/CorpusID:60527419>.
- [18] David M Hamby. “A review of techniques for parameter sensitivity analysis of environmental models”. In: *Environmental monitoring and assessment* 32 (1994), pp. 135–154.
- [19] Anthony V Fiacco. “Introduction to sensitivity and stability analysis in non linear programming”. In: *New York: Academic Press*, (1983).
- [20] James Wise Samantha Fowler Rebecca Roush. *Concepts of Biology*. OpenStax, 2013.
- [21] Yongqiang Zhang et al. “Multi-decadal trends in global terrestrial evapotranspiration and its components”. In: *Scientific reports* 6.1 (2016), p. 19124.
- [22] Rong Gan et al. “Use of satellite leaf area index estimating evapotranspiration and gross assimilation for Australian ecosystems”. In: *Ecohydrology* 11.5 (2018), e1974.
- [23] Yongqiang Zhang et al. “Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017”. In: *Remote Sensing of Environment* 222 (2019), pp. 165–182. ISSN: 0034-4257. DOI: <https://doi.org/10.1016/j.rse.2018.12.031>. URL: <https://www.sciencedirect.com/science/article/pii/S003442571830590X>.
- [24] John HM Thornley et al. *Mathematical models in plant physiology*. Academic Press (Inc.) London, Ltd., 1976.
- [25] G James Collatz et al. “Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer”. In: *Agricultural and Forest meteorology* 54.2-4 (1991), pp. 107–136.
- [26] Gaylon S Campbell and John M Norman. *An introduction to environmental biophysics*. Springer Science & Business Media, 2000.
- [27] John Wilson Rouse et al. “Monitoring vegetation systems in the Great Plains with ERTS”. In: *NASA Spec. Publ* 351.1 (1974), p. 309.

- [28] Bruce K Wylie et al. “Calibration of remotely sensed, coarse resolution NDVI to CO<sub>2</sub> fluxes in a sagebrush–steppe ecosystem”. In: *Remote Sensing of Environment* 85.2 (2003), pp. 243–255.
- [29] Natalie T Boelman et al. “Response of NDVI, biomass, and ecosystem gas exchange to long-term warming and fertilization in wet sedge tundra”. In: *Oecologia* 135 (2003), pp. 414–421.
- [30] A Huete et al. “Overview of the radiometric and biophysical performance of the MODIS vegetation indices”. In: *Remote Sensing of Environment* 83.1 (2002). The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring, pp. 195–213. ISSN: 0034-4257. DOI: [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2). URL: <https://www.sciencedirect.com/science/article/pii/S0034425702000962>.
- [31] Pontus Olofsson et al. “Towards operational remote sensing of forest carbon balance across Northern Europe”. In: *Biogeosciences* 5.3 (2008), pp. 817–832.
- [32] Chi Xu et al. “Evaluating the difference between the normalized difference vegetation index and net primary productivity as the indicators of vegetation vigor assessment at landscape scale”. In: *Environmental monitoring and assessment* 184 (2012), pp. 1275–1286.
- [33] Zhangyan Jiang et al. “Development of a two-band enhanced vegetation index without a blue band”. In: *Remote sensing of Environment* 112.10 (2008), pp. 3833–3845.
- [34] Zhanzhang Cai et al. “Modelling Daily Gross Primary Productivity with Sentinel-2 Data in the Nordic Region—Comparison with Data from MODIS”. In: *Remote Sensing* 13 (Jan. 2021), p. 469. DOI: [10.3390/rs13030469](https://doi.org/10.3390/rs13030469).
- [35] Grayson Badgley, Christopher B. Field, and Joseph A. Berry. “Canopy near-infrared reflectance and terrestrial photosynthesis”. In: *Science Advances* 3.3 (2017), e1602244. DOI: [10.1126/sciadv.1602244](https://doi.org/10.1126/sciadv.1602244). eprint: <https://www.science.org/doi/pdf/10.1126/sciadv.1602244>. URL: <https://www.science.org/doi/abs/10.1126/sciadv.1602244>.
- [36] Shangrong Lin et al. “Evaluating the Effectiveness of Using Vegetation Indices Based on Red-Edge Reflectance from Sentinel-2 to Estimate Gross Primary Productivity”. In: *Remote Sensing* 11.11 (2019). ISSN: 2072-4292. DOI: [10.3390/rs11111303](https://doi.org/10.3390/rs11111303). URL: <https://www.mdpi.com/2072-4292/11/11/1303>.
- [37] Anna Spinosa, Mario Alberto Fuentes-Monjaraz, and Ghada El Serafy. “Assessing the Use of Sentinel-2 Data for Spatio-Temporal Upscaling of Flux Tower Gross Primary Productivity Measurements”. In: *Remote Sensing* 15.3 (2023). ISSN: 2072-4292. DOI: [10.3390/rs15030562](https://doi.org/10.3390/rs15030562). URL: <https://www.mdpi.com/2072-4292/15/3/562>.
- [38] L. Breiman. “Random Forests”. In: *Machine Learning* 45 (2001), pp. 5–32. URL: <https://api.semanticscholar.org/CorpusID:89141>.
- [39] Daniel Clewley et al. “A Method for Upscaling In Situ Soil Moisture Measurements to Satellite Footprint Scale Using Random Forests”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* PP (Apr. 2017), pp. 1–11. DOI: [10.1109/JSTARS.2017.2690220](https://doi.org/10.1109/JSTARS.2017.2690220).
- [40] Bo Yu, Fang Chen, and Hanyue Chen. “NPP estimation using random forest and impact feature variable importance analysis”. In: *Journal of Spatial Science* 64 (Sept. 2017), pp. 1–20. DOI: [10.1080/14498596.2017.1367331](https://doi.org/10.1080/14498596.2017.1367331).

- [41] Daniel E. Pabon-Moreno et al. “On the Potential of Sentinel-2 for Estimating Gross Primary Production”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (Jan. 2022), pp. 1–1. DOI: [10.1109/TGRS.2022.3152272](https://doi.org/10.1109/TGRS.2022.3152272).
- [42] P. Bodesheim et al. “Upscaled diurnal cycles of land–atmosphere fluxes: a new global half-hourly data product”. In: *Earth System Science Data* 10.3 (2018), pp. 1327–1365. DOI: [10.5194/essd-10-1327-2018](https://doi.org/10.5194/essd-10-1327-2018). URL: <https://essd.copernicus.org/articles/10/1327/2018/>.
- [43] Martin Jung et al. “Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations”. In: *Journal of Geophysical Research: Biogeosciences* 116.G3 (2011). DOI: <https://doi.org/10.1029/2010JG001566>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010JG001566>. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JG001566>.
- [44] David Montero et al. “Recurrent Neural Networks for Modelling Gross Primary Production”. In: *arXiv preprint arXiv:2404.12745* (2024).
- [45] Hui Guo et al. “On the use of machine learning methods to improve the estimation of gross primary productivity of maize field with drip irrigation”. In: *Ecological Modelling* 476 (2023), p. 110250.
- [46] Geetika Agarwal et al. “Prediction of Gross Primary Productivity using Hybrid LSTM-GRU Neural Networks and ERA5-Reanalysis Data”. In: *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. IEEE, 2023, pp. 94–99.
- [47] Paul HC Eilers. “A perfect smoother”. In: *Analytical chemistry* 75.14 (2003), pp. 3631–3636.
- [48] David Weisburd et al. “Measuring Association for Scaled Data: Pearson’s Correlation Coefficient”. In: *Basic Statistics in Criminology and Criminal Justice*. Cham: Springer International Publishing, 2020, pp. 479–530. ISBN: 978-3-030-47967-1. DOI: [10.1007/978-3-030-47967-1\\_14](https://doi.org/10.1007/978-3-030-47967-1_14). URL: [https://doi.org/10.1007/978-3-030-47967-1\\_14](https://doi.org/10.1007/978-3-030-47967-1_14).
- [49] Karl Pearson. “VII. Note on regression and inheritance in the case of two parents”. In: *proceedings of the royal society of London* 58.347-352 (1895), pp. 240–242.
- [50] Naomi Altman and Martin Krzywinski. “Points of Significance: Association, correlation and causation.” In: *Nature methods* 12.10 (2015).
- [51] Ilaria Amerise and Agostino Tarsitano. “Correction methods for ties in rank correlations”. In: *Journal of Applied Statistics* 42 (May 2015), pp. 1–13. DOI: [10.1080/02664763.2015.1043870](https://doi.org/10.1080/02664763.2015.1043870).
- [52] Charles Spearman. “The proof and measurement of association between two things.” In: *American Journal of Psychology* (1904).
- [53] Saman Razavi and Hoshin V. Gupta. “What do we mean by sensitivity analysis? The need for comprehensive characterization of “global” sensitivity in Earth and Environmental systems models”. In: *Water Resources Research* 51.5 (2015), pp. 3070–3092. DOI: <https://doi.org/10.1002/2014WR016527>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014WR016527>. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014WR016527>.

- [54] Andrea Saltelli and Paola Annoni. “How to avoid a perfunctory sensitivity analysis”. In: *Environmental Modelling & Software* 25.12 (2010), pp. 1508–1517. ISSN: 1364-8152. DOI: <https://doi.org/10.1016/j.envsoft.2010.04.012>. URL: <https://www.sciencedirect.com/science/article/pii/S1364815210001180>.
- [55] E. Čech. *Point Sets*. Academic Press, 1969. ISBN: 9780121648503.
- [56] C. J. Willmott and K. Matsuura. “On the use of dimensioned measures of error to evaluate the performance of spatial interpolators”. In: *International Journal of Geographical Information Science* 20.1 (2006), pp. 89–102. DOI: [10.1080/13658810500286976](https://doi.org/10.1080/13658810500286976). URL: <https://doi.org/10.1080/13658810500286976>.
- [57] James Bergstra and Yoshua Bengio. “Random search for hyper-parameter optimization.” In: *Journal of machine learning research* 13.2 (2012).
- [58] Petro Liashchynskiy and Pavlo Liashchynskiy. *Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS*. 2019. arXiv: [1912.06059](https://arxiv.org/abs/1912.06059) [cs.LG]. URL: <https://arxiv.org/abs/1912.06059>.
- [59] Alexander Golbraikh et al. “Rational selection of training and test sets for the development of validated QSAR models”. In: *Journal of computer-aided molecular design* 17 (2003), pp. 241–253.
- [60] Trevor Hastie. *The elements of statistical learning: data mining, inference, and prediction*. 2009.
- [61] Julia Slingo and Tim Palmer. “Uncertainty in weather and climate prediction”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369.1956 (2011), pp. 4751–4767.
- [62] Edward Ott et al. “A local ensemble Kalman filter for atmospheric data assimilation”. In: *Tellus A: Dynamic Meteorology and Oceanography* 56.5 (2004), pp. 415–428.
- [63] Tim N Palmer et al. “Stochastic parametrization and model uncertainty”. In: (2009).
- [64] Hannah Mary Christensen, IM Moroz, and TN Palmer. “Stochastic and perturbed parameter representations of model uncertainty in convection parameterization”. In: *Journal of the Atmospheric Sciences* 72.6 (2015), pp. 2525–2544.
- [65] Kenji Ose, Thomas Corpetti, and Laurent Demagistri. “2 - Multispectral Satellite Image Processing”. In: *Optical Remote Sensing of Land Surface*. Ed. by Nicolas Baghdadi and Mehrez Zribi. Elsevier, 2016, pp. 57–124. ISBN: 978-1-78548-102-4. DOI: <https://doi.org/10.1016/B978-1-78548-102-4.50002-8>. URL: <https://www.sciencedirect.com/science/article/pii/B9781785481024500028>.
- [66] European Space Agency. *S2 Mission*. <https://sentiwiki.copernicus.eu/web/s2-mission>. Accessed: 10-04-2024.
- [67] M. Rodell et al. “The Global Land Data Assimilation System”. In: *Bulletin of the American Meteorological Society* 85.3 (2004), pp. 381–394. DOI: [10.1175/BAMS-85-3-381](https://journals.ametsoc.org/view/journals/bams/85/3/bams-85-3-381). URL: <https://journals.ametsoc.org/view/journals/bams/85/3/bams-85-3-381.xml>.
- [68] Eric F. Vermote, J. C. Roger, and J. P. Ray. *MODIS Surface Reflectance User’s Guide Version 1.5*. Apr. 2022.

- [69] MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500m SIN Grid. *MOD17A3HGF: MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500m SIN Grid*. 2006. DOI: [10.5067/MODIS/MOD17A3HGF.061](https://doi.org/10.5067/MODIS/MOD17A3HGF.061). URL: <https://lpdaac.usgs.gov/products/mod17a3hgv061/>.
- [70] Terra and Aqua combined Moderate Resolution Imaging Spectroradiometer (MODIS). *MCD12Q1.061 MODIS Land Cover Type Yearly Global 500m*. 2020. DOI: [10.5067/MODIS/MCD12Q1.061](https://doi.org/10.5067/MODIS/MCD12Q1.061). URL: [https://developers.google.com/earth-engine/datasets/catalog/MODIS\\_061\\_MCD12Q1](https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MCD12Q1).
- [71] *Earth Engine Data Catalog*. <https://developers.google.com/earth-engine/datasets>. Accessed: 29-04-2024.
- [72] Integrated Carbon Observation System. *ICOS Station IT-Tor*. Accessed: 13 September 2024. 2024. URL: [https://meta.icos-cp.eu/resources/stations/ES\\_IT-Tor](https://meta.icos-cp.eu/resources/stations/ES_IT-Tor).
- [73] O. Tetens. “Uber einige meteorologische begriffe”. In: 6 (1930), pp. 297–309.
- [74] David Bolton. “The Computation of Equivalent Potential Temperature”. In: *Monthly Weather Review* 108.7 (1980), pp. 1046–1053. DOI: [10.1175/1520-0493\(1980\)108<1046:TcoEPT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<1046:TcoEPT>2.0.CO;2). URL: [https://journals.ametsoc.org/view/journals/mwre/108/7/1520-0493\\_1980\\_108\\_1046\\_tcoept\\_2\\_0\\_co\\_2.xml](https://journals.ametsoc.org/view/journals/mwre/108/7/1520-0493_1980_108_1046_tcoept_2_0_co_2.xml).
- [75] X. Lan, P. Tans, and K.W. Thoning. *Trends in globally-averaged CO2 determined from NOAA Global Monitoring Laboratory measurements*. Version 2024-08. 2024. DOI: [10.15138/9NOH-ZH07](https://doi.org/10.15138/9NOH-ZH07).
- [76] Jerry L Hintze and Ray D Nelson. “Violin plots: a box plot-density trace synergism”. In: *The American Statistician* 52.2 (1998), pp. 181–184.
- [77] DY Hollinger and AD Richardson. “Uncertainty in eddy covariance measurements and its application to physiological models”. In: *Tree physiology* 25.7 (2005), pp. 873–885.
- [78] Andrew D Richardson et al. “A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes”. In: *Agricultural and Forest Meteorology* 136.1-2 (2006), pp. 1–18.
- [79] Mathias Göckede et al. “Update of a footprint-based approach for the characterisation of complex measurement sites”. In: *Boundary-Layer Meteorology* 118 (2006), pp. 635–655.
- [80] Dennis D Baldocchi. “Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future”. In: *Global change biology* 9.4 (2003), pp. 479–492.
- [81] ML Goulden. “Measurement of carbon storage by long-term eddy correlation: Methods and a critical assessment of accuracy”. In: *Global Change Biol.* 2 (1996), pp. 169–182.
- [82] DD Baldocchi et al. “On measuring net ecosystem carbon exchange in complex terrain over tall vegetation”. In: *Boundary Layer Meteorol* 96 (2000), pp. 257–291.
- [83] Kell B Wilson and Dennis D Baldocchi. “Comparing independent estimates of carbon dioxide exchange over 5 years at a deciduous forest in the southeastern United States”. In: *Journal of Geophysical Research: Atmospheres* 106.D24 (2001), pp. 34167–34178.
- [84] Peter M Anthoni, Beverly E Law, and Michael H Unsworth. “Carbon and water vapor exchange of an open-canopied ponderosa pine ecosystem”. In: *Agricultural and Forest Meteorology* 95.3 (1999), pp. 151–168.

- [85] Shaoyang He et al. “A daily and 500 m coupled evapotranspiration and gross primary production product across China during 2000–2020”. In: *Earth System Science Data Discussions* 2022 (2022), pp. 1–42.
- [86] Dongdong Kong et al. “A robust method for reconstructing global MODIS EVI time series on the Google Earth Engine”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 155 (2019), pp. 13–24.
- [87] Xiaoping Wang et al. “Seasonal variations in leaf maximum photosynthetic capacity and its dependence on climate factors across global FLUXNET sites”. In: *Journal of Geophysical Research: Biogeosciences* 127.5 (2022), e2021JG006709.
- [88] Steven W. Running and Maosheng Zhao. *User’s Guide Daily GPP and Annual NPP (MOD17A2/A3) Products NASA Earth Observing System MODIS Land Algorithm*. Oct. 2015.
- [89] GisGeography. *MODIS: Moderate Resolution Imaging Spectroradiometer*. <https://gisgeography.com/modis-satellite>. Accessed: 25-04-2024.