The revision of the manuscript provides an improvement from the earlier version. However, there remains problems w.r.t. redundancy, sentence structure, etc. The manuscript could benefit from professional editing.

Modelling spatial and temporal dynamics of GPP in the Sahel from earth observation based photosynthetic capacity and quantum efficiency

Torbern Tagesson[1], Jonas Ardö[2], Bernard Cappelaere[3], Laurent Kergoat[4], Abdulhakim Abdi[2], Stéphanie Horion[1], Rasmus Fensholt[1]

[1]Department of Geosciences and Natural Resource Management, University of Copenhagen, Øster Voldgade 10, DK-1350 Copenhagen, Denmark; E-Mails: torbern.tagesson@ign.ku.dk, stephanie.horion@ign.ku.dk, rf@ign.ku.dk

[2]Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, SE- 223 62 Lund, Sweden, E-Mails: jonas.ardo@nateko.lu.se, hakim.abdi@gmail.com

[3]HydroSciences Montpellier, IRD, CNRS, Univ. Montpellier, Montpellier, France, E-Mail: bernard.cappelaere@um2.fr

[4]Geoscience Environnement Toulouse, (CNRS/UPS/IRD), 14 av E Belin, 31400 Toulouse, France, E-Mail: laurent.kergoat@get.obs-mip.fr

Correspondence to: Torbern Tagesson (torbern.tagesson@ign.ku.dk)

Abstract. It has been shown that vegetation growth in semi-arid regions is important ~~for the variability of~~ to the global terrestrial CO_2 sink, which indicates the strong need for improved understanding, and spatially explicit estimates of CO_2 uptake (gross primary production (GPP)) in semi-arid ecosystems. This study has three aims: 1) to evaluate the MOD17A2H GPP (collection 6) product against eddy covariance (EC) based GPP for six sites across the Sahel; 2) to characterise relationships between spatial and temporal variability in EC based photosynthetic capacity (F_{opt}) and quantum efficiency (α) and earth observation (EO) based vegetation indices (normalized difference vegetation index (NDVI); renormalized difference vegetation index (RDVI); enhanced vegetation index (EVI); and shortwave infrared water stress index (SIWSI)); and 3) to study the applicability of EO up-scaled F_{opt} and α for GPP modelling purposes. MOD17A2H GPP (collection 6) underestimated GPP strongly, most likely because maximum light use efficiency is set too low for semi-arid ecosystems in the MODIS algorithm. Intra-annual dynamics in F_{opt} was closely related to SIWSI being sensitive to equivalent water thickness, whereas α was closely related to RDVI affected by chlorophyll abundance. Spatial and inter-annual dynamics in F_{opt} and α were closely coupled to NDVI and RDVI, respectively. Modelled GPP based on F_{opt} and α up-scaled using EO based indices reproduced in situ GPP well for all except a cropped site that was strongly impacted by anthropogenic land use. Up-scaled GPP for Sahel 2001-2014 was 736±39 g C m^{-2} y^{-1}. This study indicates the strong applicability of EO as a tool for spatially explicit estimates of GPP, F_{opt} and α; incorporating EO-based F_{opt} and α in to dynamic global vegetation models could improve global estimates of vegetation production, ecosystem processes and biogeochemical and hydrological cycles.

Keywords: Remote sensing, Gross Primary Productivity, MOD17A2H, light use efficiency, photosynthetic capacity, quantum efficiency

1 Introduction

Vegetation growth in semi-arid regions is an important sink for fossil fuel emissions. Mean carbon dioxide (CO_2) uptake by terrestrial ecosystems is dominated by highly productive lands, mainly tropical forests, whereas semi-arid regions are the main biome driving its inter-annual variability (Ahlström et al., 2015; Poulter et al., 2014). Semi-arid regions even contribute to 60% of the long term trend in the global terrestrial C sink (Ahlström et al., 2015). It is thus important to understand long-term variability of vegetation growth in semi-arid areas and their response to environmental conditions to better quantify and forecast effects of climate change.

Sahel is a semi-arid transition zone between the dry Sahara desert in the North and the humid Sudanian savanna in the South. The region has experienced numerous severe droughts during the last decades that resulted in region-wide famines in 1972-1973 and 1984–1985 and localized food shortages across the region in 1990, 2002, 2004, 2011 and 2012 (Abdi et al., 2014; United Nations, 2013). Vegetation production is thereby an important ecosystem service for livelihood in Sahel, but it is under threat. The region experiences a strong population growth, increasing the demand on ecosystem services due to cropland expansion, increased pasture stocking rates and fuelwood extraction (Abdi et al., 2014).

At the same time as we have reports of declining vegetation production, we have contradicting reports of greening of the Sahel based on earth observation (EO) data (Dardel et al., 2014; Fensholt et al., 2013). The greening of Sahel has mainly been attributed to alleviated drought stress conditions due to increased precipitation since the mid-1990s (Hickler et al., 2005). Climate is thus another important factor regulating vegetation production. Semi-arid regions, such as Sahel, are particularly vulnerable to climate fluctuations due to their dependency on moisture ~~conditions~~:

Estimation of gross primary production (GPP), i.e. uptake of atmospheric CO_2 by vegetation, is still a major challenge within remote sensing of ecosystem services. Gross primary production is a main driver of ecosystem services such as climate regulation, carbon (C) sequestration, C storage, food production, or livestock grassland production. Within EO, spatial quantification of GPP generally involves light use efficiency (LUE), defined as the conversion efficiency of absorbed solar light into CO_2 uptake (Monteith, 1972, 1977). It has been shown that LUE varies in space and time due to factors such as plant functional type, drought and temperature, nutrient levels and physiological limitations of photosynthesis (Garbulsky et al., 2010; Paruelo et al., 2004; Kergoat et al., 2008). The LUE concept has been applied using various methods, either by using a biome-specific LUE constant (Ruimy et al., 1994), or by modifying a maximum LUE using meteorological variables (Running et al., 2004).

An example of an LUE based model is the standard GPP product from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor (MOD17A2). Within the model, absorbed photosynthetically active radiation (PAR) is estimated as a product of the fraction of PAR absorbed by green vegetation (FPAR from MOD15A2) multiplied with daily PAR from the meteorological data of the Global Modeling and Assimilation Office (GMAO). A set of maximum LUE parameters specified for each biome are extracted from a Biome Properties Look-Up Table (BPLUT). Then maximum LUE is modified depending on air temperature (T_{air}) and vapor pressure deficit (VPD) (Running et al., 2004). Sjöström et al. (2013) evaluated the MOD17A2 product (collection 5.1) for Africa, and showed

that it underestimated GPP for semi-arid savannas in Sahel. Explanations for this underestimation were that the assigned maximum LUE from BPLUT was set too low and uncertainties in the FPAR (MOD15A2) product. Recently, a new collection of MOD17A2 at 500 m spatial resolution was released (MOD17A2H; collection 6) with an updated BPLUT, updated GMAO meteorological data, improved quality control and gap filling of the FPAR data from MOD15A2 (Running and Zhao, 2015).

It has been shown that the LUE method does not perform well in arid conditions and at agricultural sites (Turner et al., 2005). Additionally, the linearity assumed by the LUE model is usually not found as the response of GPP to incoming light follows more of an asymptotic curve (Cannell and Thornley, 1998). Investigating other methods for remotely determining GPP is thus of great importance, especially for semi-arid environments. Therefore, instead of LUE we focus on the light response function of GPP at canopy scale, and spatial and temporal variation of its two main parameters: maximum GPP under light saturation (canopy-scale photosynthetic capacity; F_{opt}), and the initial slope of the light response function (canopy-scale quantum efficiency; α) (Falge et al., 2001; Tagesson et al., 2015a). Photosynthetic capacity is a measure of the maximum rate at which the canopy can fix CO_2 during photosynthesis ($\mu mol\ CO_2\ m^{-2}\ s^{-1}$) whereas α is the amount of CO_2 fixed per incoming PAR ($\mu mol\ CO_2\ \mu mol\ PAR^{-1}$). Just to clarify the difference in LUE and α in this study; LUE ($\mu mol\ CO_2\ \mu mol\ APAR^{-1}$) is the slope of a linear fit between CO_2 uptake and absorbed PAR, whereas α ($\mu mol\ CO_2\ \mu mol\ PAR^{-1}$) is the initial slope of an asymptotic curve against incoming PAR.

It has been proven that F_{opt} and α are closely related to chlorophyll abundance due to their coupling with the electron transport rate (Ide et al., 2010). Additionally, in semi-arid ecosystems water availability is generally considered to be the main limiting factor affecting intra-annual dynamics of vegetation growth (Fensholt et al., 2013; Hickler et al., 2005; Tagesson et al., 2015b). Several remote sensing studies have established relationships between remotely sensed vegetation indices and ecosystem properties such as chlorophyll abundance and equivalent water thickness (Yoder and Pettigrew-Crosby, 1995; Fensholt and Sandholt, 2003). In this study we will analyse if EO vegetation indices can be used for up-scaling F_{opt} and α and investigate if this could offer a promising way to map GPP in semi-arid areas. This potential will be analysed by the use of detailed ground observations from six eddy covariance (EC) flux tower sites across Sahel.

The three aims of this study are:

1) To investigate if the recently released MOD17A2H GPP (collection 6) product is better at capturing GPP for Sahel than collection 5.1. We hypothesise that MOD17A2H GPP (collection 6) product will estimate GPP well for the six Sahelian EC sites, because of major changes done in comparison to collection 5.1 (Running and Zhao, 2015).
2) To characterize the relationships between spatial and temporal variability in F_{opt} and α and remotely sensed vegetation indices. We hypothesise that EO vegetation indices that are closely related to chlorophyll abundance will be most strongly coupled with spatial and inter-annual dynamics in F_{opt} and α, whereas vegetation indices closely related to equivalent water thickness will be most strongly coupled with intra-annual dynamics in F_{opt} and α across Sahel.
3) To evaluate the applicability of a GPP model based on the light response function using EO vegetation indices and incoming PAR as input data.

2 Materials and Methods

2.1 Site description

The Sahel stretches from the Atlantic Ocean in the west to the Red Sea in the east. The northern border towards Sahara and the southern border towards the humid Sudanian Savanna are defined by the 150 and 700 mm isohyets, respectively (Fig. 1) (Prince et al., 1995). Tree and shrub canopy cover is now generally low (< 5%) and dominated by species of *Balanites*, *Acacia, Boscia* and *Combretaceae* (Rietkerk et al., 1996). Annual grasses such as *Schoenefeldia gracilis*, *Dactyloctenium aegypticum*, *Aristida mutabilis*, and *Cenchrus biflorus* dominate the herbaceous layer, but perennial grasses such as *Andropogon gayanus*, *Cymbopogon schoenanthus* can also be found (Rietkerk et al., 1996; de Ridder et al., 1982). From the FLUXNET database (Baldocchi et al., 2001) we selected the six available measurement sites with EC based CO_2 flux data from Sahel (Table 1; Fig. 1). The sites represent a variety of ecosystems present in the region, from dry fallow bush savanna to seasonally inundated acacia forest. For a full description of the measurement sites, we refer to Tagesson et al. (2016a) and references in Table 1.

<Table 1>

<Figure 1>

2.2 Data collection

2.2.1 Eddy covariance and hydrometeorological in situ data

Eddy covariance and hydrometeorological data originating from the years between 2005 and 2013 were collected from the principal investigators of the measurement sites (Tagesson et al., 2016a). The EC sensor set-up consisted of open-path CO_2/H_2O infrared gas analysers and 3-axis sonic anemometers. Data were collected at 20 Hz rate and statistics were calculated for 30-min periods. For a full description of sensor set up and post processing of EC data, see references in Table 1. Final fluxes were filtered according to quality flags provided by FLUXNET and outliers were filtered according to Papale et al. (2006). We extracted the original net ecosystem exchange (NEE) data without any gap-filling or partitioning of NEE to GPP and ecosystem respiration. The collected hydrometeorological data were: air temperature (T_{air}; °C), rainfall (P; mm), relative air humidity (Rh; %), soil moisture at 0.1 m depth (SWC; % volumetric water content), incoming global radiation (R_g; W m^{-2}), incoming photosynthetically active radiation (PAR; µmol m^{-2} s^{-1}), VPD (hPa), peak dry weight biomass (g dry weight m^{-2}), C3/C4 species ratio, and soil conditions (nitrogen and C concentration; %). For a full description of collected data and sensor set-up, see Tagesson et al. (2016a).

2.2.2 Earth Observation data and gridded ancillary data

Composite products from MODIS/Terra covering Sahel were acquired at Reverb ECHO (NASA, 2016). Collected products were GPP (MOD17A2H; collection 6), nadir bidirectional reflectance distribution function adjusted reflectance (NBAR) (8-day composites; MCD43A4; collection 5.1) at 500*500 m^2 spatial resolution, the normalized difference vegetation index (NDVI) and the enhanced vegetation index (EVI) (16-day composites; MOD13Q1; collection 6) at 250*250 m^2 spatial resolution. The NBAR product was preferred over the reflectance product (MOD09A1), in order to avoid variability caused by varying sun and sensor viewing geometry (Huber et al., 2014; Tagesson et al., 2015c). We extracted the median of 3x3 pixels centred at the location of each EC tower. Time series of EO products were filtered according to MODIS quality control data; MOD17A2H is a gap-filled and filtered product,

QC data from MCD43A2 were used for filtering of MCD43A4; and bit 2-5 (highest –decreasing quality) was used for MOD13Q1. Finally, data were gap-filled to daily values using linear interpolation.

We downloaded ERA Interim reanalysis PAR at the ground surface (W m^{-2}) with a spatial resolution of 0.25°×0.25° accumulated for each 3-hour period 2000-2015 from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Dee et al., 2011; ECMWF, 2016a).

2.3 Data handling

2.3.1 Intra-annual dynamics in photosynthetic capacity and quantum efficiency

To estimate daily values of EC based F_{opt} and α, the asymptotic Mitscherlich light-response function was fitted between daytime NEE and incoming PAR using a 7-day moving window with a 1-day time step:

$$NEE = -(F_{opt}) \times (1 - e^{\left(\frac{-\alpha \times PAR}{F_{opt}}\right)}) + R_d \quad (1)$$

where F_{opt} is CO_2 uptake at light saturation (photosynthetic capacity; μmol CO_2 m^{-2} s^{-1}), R_d is dark respiration (μmol CO_2 m^{-2} s^{-1}), and α is the initial slope of the light response curve (quantum efficiency; μmol CO_2 μmol PAR^{-1}) (Falge et al., 2001). By subtracting R_d from Eq. 1, the function was forced through zero and GPP was thereby estimated. To assure high quality of fitted parameters, parameters were excluded from the analysis when fitting was insignificant (p-value>0.05), and when they were out of range (F_{opt} and α >peak value of the rainy season times 1.2). Additionally, outliers were filtered following the method by Papale et al. (2006) using a 30-day moving window with a 1-day time step.

2.3.2 Vegetation indices

The maximum absorption in red wavelengths generally occurs at 682 nm as this is the peak absorption for chlorophyll a and b (Thenkabail et al., 2000), which makes vegetation indices that include the red band sensitive to chlorophyll abundance. By far the most common vegetation index is NDVI (Rouse et al., 1974):

$$NDVI = \frac{(\rho_{NIR} - \rho_{red})}{(\rho_{NIR} + \rho_{red})} \quad (2)$$

where ρ_{NIR} is the reflectance factor in near infrared (NIR) band (band 2) and ρ_{red} is the reflectance factor in the red band (band 1). Near infrared radiance is reflected by leaf cells since absorption of these wavelengths would result in overheating of the plant whereas red radiance is absorbed by chlorophyll and its accessory pigments (Gates et al., 1965). Normalization is done to reduce effects of atmospheric errors, solar zenith angles, and sensor viewing geometry, as well as increasing the vegetation signal (Qi et al., 1994; Inoue et al., 2008).

A well-known deficiency of NDVI is problems of index saturation at high biomass because absorption of red light at ~670 nm peaks at higher biomass loads whereas NIR reflectance continues to increase due to multiple scattering effects (Mutanga and Skidmore, 2004; Jin and Eklundh, 2014). By reducing atmospheric and soil background influences, EVI is designed to increase the signal from the vegetation and maintain sensitivity in high biomass regions (Huete et al., 2002).

$$EVI = G\frac{(\rho_{NIR} - \rho_{red})}{(\rho_{NIR} + C_1\rho_{red} - C_2\rho_{blue} + L)} \quad (3)$$

where ρ_{blue} is the reflectance factor in the blue band (band 3). The coefficients C_1=6 and C_2=7.5 correct for atmospheric influences, while L=1 adjust for the canopy background. The factor G=2.5 is a gain factor.

Another attempt to overcome problems of NDVI saturation was proposed by Roujean and Breon (1995) who suggested the renormalized difference vegetation index (RDVI) that combines advantages of DVI (NIR-red) and NDVI for low and high vegetation cover, respectively:

$$RDVI = \frac{(\rho_{NIR} - \rho_{red})}{\sqrt{(\rho_{NIR} + \rho_{red})}} \quad (4)$$

As a non-linear index, RDVI is not only less sensitive to variations in geometrical and optical properties of unknown foliage but also less affected by solar and viewing geometry (Broge and Leblanc, 2001). RDVI was calculated based on NBAR bands 1 and 2.

The NIR and SWIR bands are affected by the same ground properties, except that SWIR bands are also strongly sensitive to equivalent water thickness. Fensholt and Sandholt (2003) proposed a vegetation index, the shortwave infrared water stress index (SIWSI), using NIR and SWIR bands to estimate drought stress for vegetation in semi-arid environments:

$$SIWSI_{12} = \frac{(\rho_{NIR} - \rho_{SWIR12})}{(\rho_{NIR} + \rho_{SWIR12})} \quad (5)$$

$$SIWSI_{16} = \frac{(\rho_{NIR} - \rho_{SWIR16})}{(\rho_{NIR} + \rho_{SWIR16})} \quad (6)$$

where ρ_{swir12} is NBAR band 5 (1230-1250 nm) and ρ_{swir16} is NBAR band 6 (1628-1652 nm). As the vegetation water content increases, reflectance in SWIR decreases indicating that low and high SIWSI values point to sufficient water conditions and drought stress, respectively.

2.3.3 Incoming PAR across Sahel

A modified version of the ERA Interim reanalysis PAR was used in the current study as an error in the code producing these PAR estimates ~~was identified by the data distributor causing PAR values to be~~ too low (ECMWF, 2016b). Accordingly, incoming PAR at the ground surface from ERA Interim was systematically underestimated even though it followed the pattern of PAR measured at the six Sahelian EC sites (Fig. S1 in supplementary material). In order to correct for this error, we fitted and applied an ordinary least square linear regression between in situ PAR and ERA Interim PAR (Fig. S1). The produced PAR from this relationship is at the same level as in situ PAR and should be at a correct level even though the original ERA Interim PAR is actually produced from the red and near infrared part of the spectrum.

2.4 Data analysis

2.4.1 Coupling temporal and spatial dynamics in photosynthetic capacity and quantum efficiency with explanatory variables

The coupling between intra-annual dynamics in F_{opt} and α and the vegetation indices for the different measurement sites were studied using Pearson correlation analysis. As part of the correlation analysis, we used bootstrap simulations with 200 iterations from which mean and standard deviation of the correlation coefficients were calculated (Richter et al.,

2012). Relationships between intra-annual dynamics in F_{opt} and α and the vegetation indices for all sites combined were also analysed. In the analysis for all sites, data were normalised in order to avoid influence of spatial and inter-annual variability. Time series of ratios of F_{opt} and α (F_{opt_frac} and α_{frac}) against the annual peak values (F_{opt_peak} and α_{peak}; see below for calculation of annual peak values) were estimated for all sites:

$$F_{opt\_frac} = \frac{F_{opt}}{F_{opt\_peak}} \quad (7)$$

$$\alpha_{frac} = \frac{\alpha}{\alpha_{peak}} \quad (8)$$

The same standardisation procedure was used for all vegetation indices (VI_{frac}):

$$VI_{frac} = \frac{VI}{VI_{peak}} \quad (9)$$

where VI_{peak} is the annual peak values of the vegetation indices (14 days running mean with highest annual value). The coupling between α_{frac} and F_{opt_frac} and the different VI_{frac} were examined using Pearson correlation analysis for all sites.

redundant see p.6 l. 219-220.

Regression trees were used to fill gaps in the daily estimates of F_{opt} and α. One hundred tree sizes were chosen based on 100 cross validation runs, and these trees were then used for estimating F_{opt} and α following the method in De'ath and Fabricius (2000). We used SWC, VPD, T_{air}, PAR, and the vegetation index with strongest correlation with intra-annual dynamics as explanatory variables in the analysis. In the analysis for all sites, the same standardisation procedure as done for F_{opt}, α, and the vegetation indices was done for the hydrometeorological variables. The 100 F_{opt} and α output subsets from the regression trees were averaged and used for filling gaps in the times series of F_{opt} and α. From these time-series we estimated annual peak values of F_{opt} and α (F_{opt_peak} and α_{peak}) as the 14-day running mean with highest annual value. To investigate spatial and inter-annual variability in F_{opt} and α across the measurement sites of the Sahel, F_{opt_peak} and α_{peak} were correlated with the annual sum of P, yearly means of T_{air}, SWC, RH, VPD, R_g, annual peak values of biomass, soil nitrogen and C concentrations, C3/C4 ratio, and VI_{peak} using Pearson linear correlations.

redundant

2.4.2 Parameterisation and evaluation of the GPP model and evaluation of the MODIS GPP

Based on Eq. 1 and outcome of the statistical analysis previously described under subsection 2.4.1 (for results see subsect. 3.2), a model for estimating GPP across Sahel was created:

$$GPP = -F_{opt} \times \left(1 - e^{\left(\frac{-\alpha \times PAR}{F_{opt}}\right)}\right) \quad (10)$$

Firstly, F_{opt_peak} and α_{peak} were estimated spatially and inter-annually using linear regression functions fitted against the vegetation indices with strongest relationships to spatial and inter-annual variability in F_{opt_peak} and α_{peak} for all sites. Secondly, exponential regression functions were established for F_{opt_frac} and α_{frac} with the vegetation index with the strongest relationships to intra-annual variability of F_{opt_frac} and α_{frac} for all sites. By combining these relationships, F_{opt} and α can be calculated for any day of year and for any point in space across Sahel:

$$F_{opt} = F_{opt\_peak} \times F_{opt\_frac} = \left(k_{Fopt} \times NDVI_{peak} + m_{Fopt}\right)\left(n_{Fopt} \times e^{\left(l_{Fopt} \times RDVI_{frac}\right)}\right) \quad (11)$$

$$\alpha = \alpha_{peak} \times \alpha_{frac} = \left(k_{\alpha} \times RDVI_{peak} + m_{\alpha}\right)\left(n_{\alpha} \times e^{\left(l_{\alpha} \times RDVI_{frac}\right)}\right) \quad (12)$$

where k_{Fopt} and k_{α} are slopes and m_{Fopt} and m_{α} are intercepts of the linear regressions giving F_{opt_peak} and α_{peak}, respectively; l_{Fopt} and l_{α} are coefficients and n_{Fopt} and n_{α} are intercepts of the exponential regressions giving F_{opt_frac} and α_{frac}, respectively. Equation 11 and 12 were inserted into Eq. 10 and GPP were thereby estimated as:

$$GPP = -\left(F_{opt\_peak} \times F_{opt\_frac}\right) \times \left(1 - e^{\left(\frac{-\left(\alpha_{peak} \times \alpha_{frac}\right) \times PAR}{F_{opt\_peak} \times F_{opt\_frac}}\right)}\right) = -\left(\left(k_{Fopt} \times NDVI_{peak} + m_{Fopt}\right)\left(n_{Fopt} \times e^{\left(l_{Fopt} \times RDVI_{frac}\right)}\right)\right) \times \left(1 - e^{\left(\frac{-\left(k_{\alpha} \times RDVI_{peak} + m_{\alpha}\right)\left(n_{\alpha} \times e^{\left(l_{\alpha} \times RDVI_{frac}\right)}\right) \times PAR}{\left(k_{Fopt} \times NDVI_{peak} + m_{Fopt}\right)\left(l_{Fopt} \times RDVI_{frac} + n_{Fopt}\right)}\right)}\right) \quad (13)$$

redundant

A bootstrap simulation methodology with 200 iterations was used when fitting the least-square regression functions for parameterisation of the GPP model (Richter et al., 2012). For each of the iterations, some of the EC sites were included and some were left-out. The bootstrap simulations generated 200 sets of k_{Fopt}, k_{α}, m_{Fopt}, m_{α}, l_{Fopt}, l_{α}, n_{Fopt}, n_{α}, and coefficient of determination (R^2), from which the medians and the standard deviations were estimated. Possible errors (e.g. random sampling errors, aerosols, electrical sensor noise, filtering and gap-filling errors, clouds, and satellite sensor degradation) can be present in both the predictor and the response variables. Hence, we selected reduced major axis regressions to account for errors in both predictor and response variables when fitting the regression functions. The regression models were validated against the left-out sites within the bootstrap simulation methodology by calculating the root-mean-square-error (RMSE), and by fitting an ordinary least squares linear regression between modelled and independent variables.

Similarly, the MODIS GPP product (MOD17A2H, collection 6) was evaluated against independent GPP from the EC sites by calculating RMSE, and by fitting an ordinary least squares linear regression.

3 Results

3.1 Evaluation of the MODIS GPP product

There was a strong linear relationship between the MODIS GPP product (MOD17A2H; collection 6) and independent GPP (slope=0.17; intercept=0.11 g C m^{-2} d^{-1}; R^2=0.69; n=598). However, MOD17A2H strongly underestimated independent GPP (Fig. 2) resulting in high RMSE (2.69 g C m^{-2} d^{-1}). It can be seen that some points for the Kelma site were quite low for MOD17A2H, whereas they were relatively high for the independent GPP (Fig. 2). Kelma is an inundated Acacia forest located in a clay soil depression. These differentiated values were found in the beginning of the dry season, when the depression was still inundated, whereas the larger area was turning dry.

<Figure 2>

3.2 Intra-annual dynamics in photosynthetic capacity and quantum efficiency

Intra-annual dynamics in F_{opt} and α differed in amplitude, but were otherwise similar across the measurement sites in Sahel (Fig. 3). There was no green ground vegetation during the dry season, and the low photosynthetic activity was due to few evergreen trees. This resulted in low values for both F_{opt} and α during the dry season. The vegetation responded strongly to rainfall, and both F_{opt} and α increased during the early phase of the rainy season. Generally, F_{opt} peaked slightly earlier than α (average± 1 standard deviation: 7±10 days) (Fig. 3).

<Figure 3>

All vegetation indices described well intra-annual dynamics in F_{opt} for all sites (Table 2). $SIWSI_{12}$ had the highest correlation for all sites except Wankama Millet, where it was RDVI. When all sites were combined, all indices described well seasonality in F_{opt}, but RDVI had the strongest correlation (Table 2).

Intra-annual dynamics in α were also closely coupled to intra-annual dynamics in the vegetation indices for all sites (Table 2). For α, RDVI was the strongest index describing intra-annual dynamics, except for Wankama Fallow where it was EVI. When all sites were combined all indices described well intra-annual dynamics in α, but RDVI was still the index with strongest relationship (Table 2).

<Table 2>

The regression trees used for gap-filling explained well the intra-annual dynamics in F_{opt} and α for all sites (Table 3; Fig. S2 in Supplementary material). The regression trees explained intra-annual dynamics in F_{opt} better than in α, and multi-year sites were better predicted than single year sites (Fig. S2). The main explanatory variables coupled to intra-annual dynamics in F_{opt} for all sites across Sahel were in the order of RDVI, SWC, VPD, T_{air}, and PAR; and for α they were RDVI, SWC, VPD and T_{air} (Table 3). The strong relationship to SWC and VPD indicates drought stress during periods of low rainfall. For all sites across Sahel, incorporating hydrometeorological variables increased the ability to determine intra-annual dynamics in F_{opt} and α compared to the ordinary least squares linear regressions against vegetation indices (Table 2, data given as *r;* Table 3; Fig. 3 and Fig. S2). For all sites, incorporation of these variables increased R^2 from 0.81 to 0.87 and from 0.74 to 0.84, for F_{opt} and α respectively.

<Table 3>

3.3 Spatial and inter-annual dynamics in photosynthetic capacity and quantum efficiency

Large spatial and inter-annual variability in F_{opt_peak} and α_{peak} were found across the six measurement sites in Sahel; F_{opt_peak} ranged between 10.1 μmol CO_2 m^{-2} s^{-1} (Wankama Millet 2005) and 50.0 μmol CO_2 m^{-2} s^{-1} (Dahra 2010), and α_{peak} ranged between 0.020 μmol CO_2 μmol PAR^{-1} (Demokeya 2007) and 0.064 μmol CO_2 μmol PAR^{-1} (Dahra 2010) (Table 4). The average two week running mean peak values of F_{opt} and α for all sites were 26.4 μmol CO_2 m^{-2} s^{-1} and 0.040 μmol CO_2 μmol PAR^{-1}, respectively. All vegetation indices determined well spatial and inter-annual dynamics in F_{opt_peak} and α_{peak} (Table 5). $NDVI_{peak}$ was most closely coupled with F_{opt_peak} whereas $RDVI_{peak}$ was closest coupled with α_{peak} (Fig. 4). F_{opt_peak} also correlated well with peak dry weight biomass, C content in the soil, and RH, whereas α_{peak} also correlated well with peak dry weight biomass, and C content in the soil (Table 5).

<Table 4>

<Table 5>

<Figure 4>

3.4 Spatially extrapolated photosynthetic capacity, quantum efficiency, and gross primary production across Sahel and evaluation of the GPP model

The spatially extrapolated F_{opt}, α and GPP averaged over Sahel for 2001-2014 were 22.5±1.7 μmol CO_2 m^{-2} s^{-1}, 0.030±0.002 μmol CO_2 μmol PAR^{-1}, and 736±39 g C m^{-2} y^{-1}, respectively. At regional scale it can be seen that F_{opt}, α, and GPP decreased substantially with latitude (Fig. 5). Highest values were found in south-eastern Senegal, western Mali, in parts of southern Sudan and on the border between Sudan and South Sudan. Lowest values were found along the northernmost parts of Sahel on the border to Sahara in Mauritania, in northern Mali, and in northern Niger.

Modelled GPP was similar to independent GPP on average, and there was a strong linear relationship between modelled GPP and independent GPP for all sites (Fig. 6; Table 6). However, when separating the evaluation between measurement sites, it can be seen that the model reproduced some sites better than others (Fig. 7; Table 6). Wankama Millet was generally overestimated, whereas the model worked well on average for Demokeya but underestimated high values (Fig. 7; Table 6). Variability of independent GPP at the other sites was well reproduced by the model (Fig. 7; Table 6). The final parameters of the GPP model (Eq. 13) are given in Table 7.

<Figure 5>

<Figure 6>

<Figure 7>

< Table 6>

< Table 7>

4 Discussions

Our hypothesis that vegetation indices closely related to equivalent water thickness (SIWSI) would be most strongly coupled with intra-annual dynamics in F_{opt} and α was not rejected for F_{opt}, since this was the case for all sites except for Wankama Millet (Table 2). However, our hypothesis was rejected for α, since it was more closely related to vegetation indices related to chlorophyll abundance (RDVI and EVI). In Sahel, soil moisture conditions in the early rainy season are important for vegetation growth and during this phase vegetation is especially vulnerable to drought conditions (Rockström and de Rouw, 1997; Tagesson et al., 2016a; Mbow et al., 2013). Photosynthetic capacity (F_{opt}) peaked earlier in the rainy season than α did (Fig. 3), thereby explaining the close relationship of F_{opt} to SIWSI. Leaf area index increased over the growing season and leaf area index is closely coupled with vegetation indices related to chlorophyll abundance (Tagesson et al., 2009). The increase in leaf area index increased canopy level quantum efficiency (α), which thereby explains the closer relationship of α to RDVI.

Our hypothesis that vegetation indices closely related to chlorophyll abundance would be most strongly coupled with spatial and inter-annual dynamics in F_{opt} and α was not rejected for either F_{opt} or α; NDVI, EVI, and RDVI ~~all had close~~ correlation with spatial and inter-annual dynamics in F_{opt} and α (Table 5). However, it was surprising that $NDVI_{peak}$ had the strongest correlation with spatial and inter-annual variability for F_{opt} (Table 5). Both EVI and RDVI should be less sensitive to saturation effects than NDVI (Huete et al., 2002; Roujean and Breon, 1995), and based on this it can be assumed that peak values of these indices should have stronger relationships to peak values of F_{opt} and α. However, vegetation indices with a high sensitivity to changes in green biomass at high biomass loads, gets less sensitive to green biomass changes at low biomass loads (Huete et al., 2002). Peak leaf area index for ecosystems across Sahel is generally ~2, or less, whereas the saturation issue of NDVI generally starts at an leaf area index of about 2-5 (Haboudane et al., 2004).

The F_{opt_peak} estimates from Agoufou, Demokeya, and the Wankama sites were similar whereas Dahra and Kelma values were high in relation to previously reported canopy-scale F_{opt_peak} from Sahel (~-8 to -23 μmol m^{-2} sec^{-1}) (Hanan et al., 1998; Merbold et al., 2009; Moncrieff et al., 1997; Boulain et al., 2009; Levy et al., 1997; Monteny et al., 1997). These previous studies reported much lower F_{opt} at canopy scale than at leaf scale (e.g. Levy et al. (1997): 10 vs. 44 μmol m^{-2} sec^{-1}; Boulain et al. (2009): 8 vs. 50 μmol m^{-2} sec^{-1}). Leaf area index at Dahra and Kelma peaked at 2.1 and 2.7, respectively (Timouk et al., 2009; Tagesson et al., 2015a), and it was substantially higher than at the above-mentioned

sites. A possible explanation to high F_{opt} estimates at Dahra and Kelma could thereby be the higher leaf area index. Tagesson et al. (2016b) performed a quality check of the EC data due to the high net CO_2 exchange measured at the Dahra field site and explained the high values by a combination of moderately dense herbaceous C4 ground vegetation, high soil nutrient availability, and a grazing pressure resulting in compensatory growth and fertilization effects. Another possible explanation could be that the West African Monsoon bring a humid layer of surface air from the Atlantic, possibly increasing vegetation production for the most western part of Sahel (Tagesson et al., 2016a).

Our model substantially overestimated GPP for Wankama Millet (Fig. 7f). Being a crop field, this site differed from the other studied sites by its species composition, ecosystem structure, as well as land and vegetation management. Crop fields in southwestern Niger are generally characterized by a rather low production resulting from decreased fertility and soil loss caused by intensive land use (Cappelaere et al., 2009). These specifics of the Wankama Millet site may cause the model parameterised with observations from the other study sites without this strong anthropogenic influence to overestimate GPP at this site. Similar results were found by Boulain et al. (2009) when applying an up-scaling model using leaf area index for Wankama Millet and Wankama Fallow. It worked well for Wankama fallow whereas it was less conclusive for Wankama Millet. The main explanation was low leaf area index in millet fields because of a low density of millet stands due to agricultural practice. There is extensive savanna clearing for food production in Sahel (Leblanc et al., 2008; Boulain et al., 2009; Cappelaere et al., 2009). To further understand impacts of this land cover change on vegetation production and land-atmosphere exchange processes, it is of urgent need for more study sites covering cropped areas in this region.

In Demokeya, GPP was slightly underestimated for the year 2008 (Fig. 7c) because modelled F_{opt} was much lower than the actual measured value in 2008 (the thick black line in Fig. 4). An improvement of the model could be to incorporate some parameters that constrain or enhance F_{opt} depending on environmental stress. Indeed, the regression tree analysis indicated that incorporating hydrometeorological variables increased the ability to predict both F_{opt} and α. On the other hand, for spatial upscaling purposes, it has been shown that including modelled hydrometeorological constraints on LUE decreases the ability to predict vegetation production due to the incorporated uncertainty in these modelled variables (Fensholt et al., 2006; Ma et al., 2014). For spatial upscaling to regional scales it is therefore better to simply use relationships to EO data. This is particularly the case for Sahel, one of the largest dryland areas in the world that includes only a few sites of hydrometeorological observations.

The pattern seen in the spatially explicit GPP budgets (Fig. 5c) may be influenced by a range of biophysical and anthropogenic factors. The clear North-South gradient is expected given the strong North-South rainfall gradient in Sahel. The West African Monsoon mentioned above could also be an explanation to high GPP values in the western part of Sahel, where values were relatively high in relation to GPP at similar latitudes in the central and eastern Sahel (Fig. 5c). The areas with highest GPP are sparsely populated woodlands or shrubby savanna with a relatively dense tree cover (Brandt et al., 2016). However, the produced maps should be used with caution as they are based on up-scaling of the only six available EC sites that exist in the region; especially given the issues related to the cropped fields discussed above. Still, the average GPP budget for the entire Sahel 2001-2014 was close to an average annual GPP budget as estimated for these six sites (692±89 g C m^{-2} y^{-1}) (Tagesson et al., 2016a). The range of GPP budgets in Fig. 5c is also similar to previous annual GPP budgets reported from other savanna areas across the world (Veenendaal et al., 2004; Chen et al., 2003; Kanniah et al., 2010; Chen et al., 2016).

Although MOD17A2 GPP has previously been shown to ~~relatively well~~ capture GPP in several ~~different~~ ecosystems (Turner et al., 2006; Turner et al., 2005; Heinsch et al., 2006; Sims et al., 2006; Kanniah et al., 2009), it has been shown to ~~be~~ underestimated ~~for~~ others (Coops et al., 2007; Gebremichael and Barros, 2006; Sjöström et al., 2013). GPP of Sahelian drylands have not been well captured by MOD17A2 (Sjöström et al., 2013; Fensholt et al., 2006), and as we have shown, this underestimation persists in the latest MOD17A2H GPP (collection 6) product (Fig. 2). The main reason for this pronounced underestimation is that maximum LUE is set to 0.84 g C MJ^{-1} (open shrubland; Demokeya) and 0.86 g C MJ^{-1} (grassland; Agoufou, Dahra, Kelma; Wankama Millet and Wankama Fallow) in the BPLUT, i.e. much lower than maximum LUE measured at the Sahelian measurement sites of this study (average: 2.47 g C MJ^{-1}; range: 1.58-3.50 g C MJ^{-1}) (Sjöström et al., 2013; Tagesson et al., 2015a), a global estimate of ~1.5 g C MJ^{-1} (Garbulsky et al., 2010), and a savanna site in Australia (1.26 g C MJ^{-1}) (Kanniah et al., 2009).

Several dynamic global vegetation models have been used for decades to quantify GPP at different spatial and temporal scales (Dickinson, 1983; Sellers et al., 1997). These models are generally based on the photosynthesis model by Farquhar et al. (1980), a model particularly sensitive to uncertainty in photosynthetic capacity (Zhang et al., 2014). This and several previous studies have shown that both photosynthetic capacity and efficiency (both α and LUE) can vary considerably between seasons as well as spatially, and both within and between vegetation types (Eamus et al., 2013; Garbulsky et al., 2010; Ma et al., 2014; Tagesson et al., 2015a). This variability is difficult to estimate using broad values based on land cover classes, yet most models apply a constant value which can cause substantial inaccuracies in the estimates of seasonal and spatial variability in GPP. This is particularly a problem in savannas that ~~comprises~~ consists of several plant functional types (C3 and C4 species, and a large variability in tree/herbaceous vegetation fractions) (Scholes and Archer, 1997). This study indicates the strong applicability of EO as a tool for parameterising spatially explicit estimates of plant physiological variables, which could improve our ability to simulate GPP. Spatially explicit estimates of GPP at a high temporal and spatial resolution are essential for environmental change studies in Sahel and make a major asset for the analysis of changes in GPP, its relationship to climatic change and anthropogenic forcing, and estimations of ecosystem processes and biochemical and hydrological cycles.

Acknowledgements Data is available from Fluxnet (http://fluxnet.ornl.gov) and CarboAfrica (http://www.carboafrica.net/index en.asp). Data for the Mali and Niger sites were made available by the AMMA-CATCH regional observatory (www.amma-catch.org), which is funded by the French Institut de Recherche pour le Développement (IRD) and Institut National des Sciences de l'Univers (INSU). The project was funded by the Danish Council for Independent Research (DFF) Sapere Aude programme. Faculty of Science, Lund University supported the Dahra and Demokeya measurements with an infrastructure grant. Ardö received support from the Swedish National Space Board.

References

Abdi, A., Seaquist, J., Tenenbaum, D., Eklundh, L., and Ardö, J.: The supply and demand of net primary production in the Sahel, Environ. Res. Lett., 9, 094003, doi:10.1088/1748-9326/9/9/094003, 2014.

Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J. G., Friedlingstein, P., Jain, A. K., Kato, E., Poulter, B., Sitch, S., Stocker, B. D., Viovy, N., Wang, Y. P., Wiltshire, A., Zaehle, S., and Zeng, N.: The dominant role of semi-arid ecosystems