



The University of Edinburgh

School of Geosciences

**COMPARISON OF TOP-DOWN AND BOTTOM-UP APPROACHES
ON SPECIFIC LEAF PATTERNS,
AT GLOBAL, LATITUDINAL AND BIOME SCALES**

By

ANNA CHIRUMBOLO

in partial fulfilment of the requirement
for the degree of BSc with Honours
in Ecological and Environmental Sciences

May 2020

Abstract

A clear understanding of global patterns in plant functionality requires accurate estimation of Specific Leaf Area (SLA; m^2/kg). Over the last decade, two publications pioneered the mapping of SLA at such a large spatial scale, using opposite approaches. Bloom *et al.* (2016) exploited the technological advancement offered by space-borne remote sensing (“top-down” approach), while Butler *et al.* (2017) used a collection of the more traditional ground-based studies (“bottom-up” approach). This study constitutes the first attempt to compare SLA results obtained by the two approaches and assess the degree of consistency between them. I carried out a visual and statistical comparative analysis, on the global, latitudinal, biome, and on the continental scale, for both, SLA mean and uncertainty values. I also tested whether the two sets of SLA mean estimates had similar sensitivities to climate data. My results revealed that correlation is generally poor, and degrees of consistency fluctuate across geographical locations, particularly across continents. These findings suggest that the incompatibility between approaches is largely due to discrepancies in the spatiotemporal scale in which they estimated the trait. The strengths and weaknesses of the top-down and bottom-up approaches are also strongly tied to the vegetation and continent-specific environmental conditions, which affect their measurement quality. In order to maximise the accuracy of global SLA pattern depiction, the methodologies used independently by each approach need to be integrated. This is the only viable solution to truly account for both fine- and broad-scale leaf trait variation. This work has wide-ranging relevance, informing ecological disciplines on the accuracy and reliability of current global SLA estimates to describe patterns in plant functionality.

Contents

1	Introduction	1
1.1	Leaf Economics Spectrum and empirical research on SLA	1
1.2	Approaches to observing SLA spatial patterns	2
1.3	SLA values change with spatial scale	3
1.4	Mapping global trait patterns	3
1.5	Research objectives and questions	4
2	Methods	5
2.1	Datasets	5
2.2	Data manipulation	5
2.3	Correlative analysis	6
3	Results	9
4	Discussion	16
4.1	RQ1: Is there a strong degree of overlap between bottom-up and top-down mean and uncertainty SLA estimates, on global, latitudinal and biome scales?	16
4.2	RQ2: Is there a difference in overlap (mean and uncertainty estimates) between continents with the same biome?	20
4.3	RQ 3: Do the SLA mean estimates have similar sensitivities to climate data, i.e. mean global precipitation and mean temperature?	21
4.4	Study limitations	23
4.5	Implications for ecological disciplines and ESM, and future directions	23
5	Conclusion	24
6	Bibliography	26
7	Appendices	31
7.1	Appendix A: <i>additional figures</i>	31
7.2	Appendix B: <i>additional tables</i>	34
7.3	Appendix C: <i>dissertation code</i>	37

Acknowledgements

I want to thank my supervisor, Matheus Williams, for suggesting the topic which inspired me, and for supporting me throughout the course of this project. I would also like to thank Darren Slevin, for being there to support me at the beginning. I have learned a lot from this project and it has been a very rewarding experience.

Infinite gratitude towards my parents and my brother who, despite the distance, have always supported and encouraged me to put the passion and enthusiasm in everything I do. Thank you Matti for the last minute proof-read.

I would like to thank my “Famiglia Grande”, Dani Gargya, Larisa Alvarez and Giada Barzaghi. You are the best people I know, and placed a smile on my face even during the most challenging times. I would also like to thank Cesare Pertusi, for his great company and for always keeping the energy levels and optimism high. Finally, a big shout out to the troops, an amazing group of people that I feel very lucky to have met, with whom I connected deeply and shared so many amazing adventures. You made my experience at university the best I could ever imagine.

List of Abbreviations

Bottom-up: Approach by Butler *et al.* (2017)

Top-down: Approach by Bloom *et al.* (2015)

SLA: Specific Leaf Area

SD: Standard Deviation

CI: Confidence Interval

RQ: Research Question

R²: R-squared or coefficient of determination

RMSE: Root Mean Square Error

EM: Electromagnetic Spectrum

LCMA: Leaf Carbon-equivalent of Mass per Area

LMA: Leaf Mass per Area

LL: Leaf Lifespan

PFT: Plant Functional Type

ESM: Earth System Modelling

LES: Leaf Economics Spectrum

MODIS: Moderate Resolution Imaging Spectroradiometer

TCF: Temperate Conifer Forests

M: Mangroves

BFT: Boreal Forest/Taiga

TGSS: Temperate Grasslands Savannas and Shrublands

TSGSS: Tropical & Subtropical Grasslands Savannas and Shrublands

DXS: Desert & Xeric Shrublands

TBMF: Temperate Broadleaf & Mixed Forests

TSDBF: Tropical & Subtropical Dry Broadleaf Forests

TSMBF: Tropical & Subtropical Moist Broadleaf Forests

T: Tundra

MGS: Montane Grasslands and Shrublands

M: Mangroves

FGS: Flooded Grasslands and Savannas

MFWS: Mediterranean Forests, Woodlands and Scrublands

TSCF: Tropical and Subtropical Coniferous Forests.

1 Introduction

SLA (m^2/kg) is a plant functional trait highly valued in ecology for its wide-reaching research applications (Heilmeier, 2019). The trait equates to the ratio of light-capturing surface per unit of dry biomass of a leaf (Wellstein *et al.*, 2017). Not only is it used to describe foliar morphology (Diaz *et al.*, 2004), but most importantly to understand whole-plant organismic functioning (Serbin, 2019). SLA is positively related to photosynthetic capacity (as it explains up to 82% of its variation; Ali *et al.*, 2016b), and negatively related to the structural resistance developed by an individual plant (Vile *et al.*, 2005). Therefore, it can explain a plant's adaptation and acclimation to external changes, reflected in its carbon (Asner *et al.*, 2016), nutrient (Hoffmann *et al.*, 2005) and water cycling efficiency (Ceccato *et al.*, 2001), which subsequently affect its growth, production, reproduction and survival rates (Ali *et al.*, 2016b). SLA can be linked to the evolutionary (Crisp *et al.*, 2009) and biogeographical (Mucina, 2019) history of a specific vegetation form, as well as the individual's sensitivity and quick response to short-term changes in its surroundings (Dwyer, Hobbs and Mayfield, 2014). The trait is traditionally studied through field sampling or experimental simulations over one community assemblage (Ruiz-Benito *et al.*, 2020) or a single vegetation form (He *et al.*, 2018). More recently, Earth Observation tools have expanded the potential to observe SLA over regional or continental scales, and over a longer time scale (Ma *et al.*, 2019). In the last decade, two research projects have pioneered the mapping of SLA patterns on a global scale. One of them implemented a 'bottom-up' (Butler *et al.*, 2017) approach of measurement, which relies on ground-based data, the second implement a 'top-down' (Bloom *et al.*, 2016) approach, based on satellite data. Considering the great role that SLA plays in informing a variety of ecological disciplines, namely functional ecology (Worthy and Swenson, 2019) and ESM (Serbin *et al.*, 2019), the reliability of the two global SLA estimates needs to be validated through cross-comparison.

1.1 Leaf Economics Spectrum and empirical research on SLA

SLA was first conceptualised as being very simple to measure (Calow, 1987) and equally relatable to all existing terrestrial plants in describing whole-organism functionality (Reich *et al.*, 1999). This consideration sparked a research interest in observing geographical trait patterns. In 2014, Wright *et al.* marked the global variation in SLA along with a 'LES' – a spectrum which defines the leaf morphological trade-off between resource acquisition and structural resilience (Fig. 1, Wright *et al.*, 2004). The trait is not only the result of leaf-environment interactions (Hulshof *et al.*, 2013). Its variation is also tightly linked with co-occurring plant functional traits (Zeballos *et al.*, 2017), which constrain and determine its value (Díaz *et al.*, 2016). It is therefore very important to observe SLA in conjunction with those traits that strongly correlate to it, for example, nitrogen (Jardine *et al.*, 2020), phosphorous leaf content (Long, Zang and Ding, 2011). High SLA values are related to maximum investment on resource acquisition, defining plants with quick turnover rates and shorter LL, living in resource-rich environments with minimal environmental disturbance. Low SLA values are related to maximum energy investment on structural preservation, representing plants with low turnover rates and longer LL, living in nutrient-poor

environments. Because interactions between functional traits can vary within small spatial scales (Onoda *et al.*, 2011), the same vegetation type can span a wide range of SLA values, often overlapping with other, unrelated vegetation types (Long, Zang and Ding, 2011). This has resulted in inconsistent findings between studies over the same location (Shiklomanov *et al.*, 2020), and has triggered ecologists to include trait covariance in their research, to depict more accurately SLA patterns (Freschet *et al.*, 2011; Yang *et al.*, 2019).

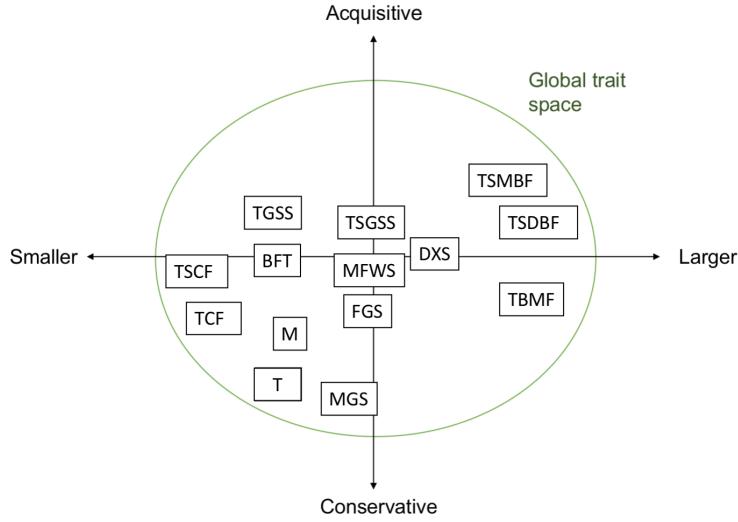


Figure 1: The global LES spectrum for SLA, across which the 14 major are distributed. The horizontal gradient reflects leaf size variation, in correspondence of the resource-acquisition efficiency of a plant. The vertical line reflects foliar structure variation, in correspondence of the plant's energy investment on structural resilience. This diagram is approximative, based on information retrieved from the literature.

1.2 Approaches to observing SLA spatial patterns

Studies have observed SLA mainly with two different approaches, either direct field sampling or remote sensing. The great majority of publications typically concern ground-based measurements on a local scale, assessing SLA change along naturally occurring (Shiklomanov *et al.*, 2020) or simulated (Wellstein *et al.*, 2017) environmental gradients. The sampling procedures and protocols vary between field studies, however, it is common to first measure the size of fresh, fully-opened leaf material, followed by the weighing its dried mass (Long, Zang and Ding, 2011). This process is labour-intensive, time-consuming, and thus financially and humanly expensive. The advent of remote sensing techniques provides a more efficient alternative. The biochemical content within the leaf material determines the effective path length of light that can pass through the medium, controlling for its reflectance and transmittance rates (Jiang *et al.*, 2018). Thanks to this active spectral interaction between the leaf and incident light (Serbin, 2019), remote sensing can comprehensively detect SLA from the measurement of foliar optical properties. Variations in leaf spectral signatures lie within specific, narrow bands of the EM (Ma *et al.*, 2020), and the challenge for remote sensing is to capture such changes. Which of the approaches provides more accurate measurements is currently a matter of debate.

1.3 SLA values change with spatial scale

Comparisons between ground-based and remotely sensed measurements have helped researchers understand that SLA values vary strongly with spatial scale (Ma *et al.*, 2020; Serbin, 2019; Ali *et al.*, 2017). In fact, in each step of the spatial hierarchy SLA is shaped by a different set of environmental variables (Messier *et al.*, 2017; Worthy and Swenson, 2019). Studies on plant communities and individual species carried out on a local scale have demonstrated that SLA variation exists starting from the individual plant (Dwyer, Hobbs and Mayfield, 2014; Goude, Nilsson and Holmström, 2019). Generally, low SLA estimates relate to the sunlit, upper leaves of the canopy, and the value gradually increases moving down towards the shaded, lower leaves (Gara *et al.*, 2019). The vertical heterogeneity of trait values corresponds to the minute, microclimatic factors to which each leaf is differently exposed (Hulshof *et al.*, 2013; Ruiz-Benito *et al.*, 2020). On the other hand, remote sensing has successfully captured how SLA values on a regional and continental scale are shaped by macroclimatic effects, which, over longer periods, have shaped the geographical distribution of plants (Serbin *et al.*, 2019; Ma *et al.*, 2020). Studies have observed a strong correlation between the trait and latitudinal climatic zones (De Frenne *et al.*, 2013), which present steep environmental gradients that feature a natural spectrum of SLA value distribution. Other studies have noted that SLA variations occur within and among biomes (Moncrieff, Bond and Higgins, 2016; Souza *et al.*, 2018), which define biotic communities shaped by climatic factors over ecological periods (Mucina, 2019). The pronounced vertical variation in leaf trait values, from fine- to broad-scale, triggered new research to map SLA patterns that would account for both, and that would be comprehensive of the entire terrestrial vegetated cover.

1.4 Mapping global trait patterns

Bloom *et al.* (2016) and Butler *et al.* (2017) pioneered the mapping of trait patterns on a global scale. The former mapped LCMA (gC/m^2), the carbon-equivalent of the reciprocal of SLA, while the latter mapped SLA. The main scope of their research was to better parametrise ESM, which would provide a more accurate simulation of carbon allocation, pool and residence time within an ecosystem (Bloom *et al.*, 2016). ESM-derived outputs strongly depend on the quality of SLA estimation, for its direct relation to photosynthetic and respiratory processes, that drive global net carbon balance (Bloom *et al.*, 2016; Serbin *et al.*, 2019; Menezes-Silva *et al.*, 2019). Bloom *et al.* (2016) implemented a ‘top-down’ approach, relying on remote sensing: they processed MODIS satellite raw data imagery, fused with a diagnostic ecosystem C balance model (DALEC2). Conversely, Butler *et al.* (2017) opted for a ‘bottom-up’ approach, relying on ground-based measurements: they used raw data from the TRY database (one of the most comprehensive plants functional trait compilations to date; Kattge *et al.*, 2020) in conjunction with remote sensing PFT categorisations and a Bayesian modelling framework. PFT are broad groupings of geographically separate communities that share a similar response to environmental variation, reflected in their leaf ‘habit’: evergreen, deciduous, C3 and C4 types (Shiklomanov *et al.*, 2020). Both publications present innovative solutions, and each has its methodology-related strength in predicting trait patterns. However, they also carry

uncertainties, and for that, they need to be compared with one another.

1.5 Research objectives and questions

In this project, I carry out a comparative analysis on SLA means and uncertainty values (SD) between top-down (Bloom *et al.*, 2016) and bottom-up (Butler *et al.*, 2017) approaches. I first transform LCMA from Bloom *et al.* (2016) to SLA, to allow for the comparison. I correlate the two sets of SLA estimates first on the global scale, then by the 4 climatic zones (latitudinal ranges) and by the 14 major biomes. I also test whether the degree of correlation varies for the same biome when this is compared between continents. Having stressed the importance of SLA-climate relationships, I compare the sensitivity of the two sets of mean estimates against total annual precipitation (mm) and mean annual temperature (°C).

This project is an initial attempt to shed light on the strengths and weaknesses of the two approaches when depicting global SLA patterns. This work is not only critical for the advancement and refinement of ESM, for better-informing carbon balance models, but also for increasing the understanding of SLA variation on a broad spatial scale, which is relevant to ecology and its sub-disciplines.

1. Is there a strong degree of overlap between bottom-up and top-down mean and uncertainty SLA estimates, on global, latitudinal and biome scales?

H0. There is no strong degree of overlap, on global, latitude and biome scales;

H1. The degree of overlap is consistently strong on global, latitude and biome scales;

H2. The degree of overlap varies with scale.

2. Is there a difference in overlap (mean and uncertainty estimates) between continents with the same biome?

H0. There is no difference in overlap between continents with the same biome;

H1. There is a strong difference in overlap between continents with the same biome;

H2. The difference in overlap varies depending on the biome compared between continents.

3. Do the SLA mean estimates have similar sensitivities to climate data, i.e. mean global precipitation and mean temperature?

H0. SLA mean estimates do not have similar sensitivities to climate data;

H1. SLA mean estimates have highly correlated sensitivities to climate data;

H2. SLA mean estimates have poorly correlated sensitivities to climate data.

2 Methods

2.1 Datasets

Main datasets

The top-down approach produced LCMA mean and uncertainty (SD and CI) estimates. LCMA was obtained as a result of the DALEC2 model, based on the fusion between MODIS satellite imagery, 1-2 day sample repetition between 2000 and 2010 at ~500 m resolution, and a Markov Chain Monte Carlo MDF algorithm (Bloom *et al.*, 2016). Single grid cell mean and related uncertainties (SD, CI) were derived from 4,000 samples within each pixel. LCMA variables were mapped at a $1^\circ \times 1^\circ$ (360×180) spatial resolution.

The bottom-up approach produced SLA mean and uncertainty (SD) estimates. SLA was obtained from the fusion of the largest global Categorical Traits database (TRY) - a collection of world-wide field-based data sampling -, with remotely sensed natural PFT categories and a Bayesian framework (Butler *et al.*, 2017). The bottom-up approach used around 32 thousand measurements for SLA, representing almost 3 thousand species. Most measurements were derived from North America, Europe, Australia, China, Japan and Brazil. The variables were mapped at a $0.5^\circ \times 0.5^\circ$ (720×180) spatial resolution.

Both approaches output the two functional traits in only one time dimension, meaning that all pixels are averaged to represent only one value, even if the raw data originally spanned a wider timeframe. I used the estimates from the top-down as the “benchmark” to which I compared the bottom-up.

Other datasets and databases

I used the Ecoregions17 database (Dinerstein *et al.*, 2017) to retrieve the spatial polygons of the 14 major biomes across the world (Fig. 2). I also used *rworldmap* package from R, to obtain the spatial polygons of the 5 continents (Antarctica excluded due to lack of data). Finally, I used historical climate data (i.e. mean annual temperature and total annual precipitation), averaged for the years 1970-2000 at 10 minutes spatial resolution (~ 340 km^2), from WorldClim v2 (Fick and Hijmans, 2017).

2.2 Data manipulation

Estimate standardisation and unit conversion

The first step in data manipulation was the standardisation of the output traits LCMA and SLA, to allow for direct comparison. First, I aggregated the spatial resolution of SLA from the bottom-up, so that it matched the $1^\circ \times 1^\circ$ of the top-down. I converted LCMA (gC/m^2) to LMA, knowing that it measures its carbon equivalent (50% of the dry-biomass). I then

turned LMA to SLA, being the reciprocal ($1/LMA$). Unit conversion from gC/m^2 to m^2/kg was also necessary to ensure standardisation of the two estimates. I carried out this calculation on LCMA mean, SD and its 25th, 75th and 95th CI percentiles. I decided to standardise LCMA from the top-down to SLA and not the opposite, because the latter more generally refers to whole-plant organismic functioning (Onoda *et al.*, 2011), and not specifically to the plant's carbon cycling efficiency like the former. Therefore SLA analysis can be relevant for a wider range of research applications, namely community ecology, organismic ecology, and paleoecology.

$$[1] \text{SLA} = \frac{1}{2 \times LCMA}$$

Manipulation with spatial polygons and climate data

I created subsets of SLA estimates from the two approaches (for both mean and SD) at three spatial extents: by the 4 climatic zone (latitudes), by the 14 major biomes across the world and by 49 biomes split by continent. The sample size varied according to the subset, but it generally decreased as the spatial polygons moved from latitudes to biomes by continent.

I split the sets of data by latitude, taking into account the climatic gradient that characterises vegetation distribution from tropics to polar regions (RQ 1). I followed convention to subset the data into the 4 climatic zones: tropical latitudes between 23.5°S and 23.5°N , sub-tropics between 23.5°S/N and 35°S/N , temperate between 35°S/N and 66.5°S/N and North pole between 66.5°N and 90°N . Splitting the data by biome, I took into account the biogeography and evolutionary history that shaped vegetation dominance distribution across the world. Furthermore, I considered the unique geographic variability between land masses, together with the vegetation dominance and climatic differences, by splitting the 14 biomes by continent (RQ 2). I masked the SLA mean and SD estimates over these spatial polygons, so as to exclude all data points not confined within the specified coordinates. For the climate sensitivity analysis (RQ 3), I prepared the datasets from WorldClim v2 by changing their original spatial resolution ($0,1667^\circ \times 0,1667^\circ$) to match that of the SLA estimates ($1^\circ \times 1^\circ$), and masked their spatial extent against that of SLA.

2.3 Correlative analysis

Statistical analysis

I statistically tested the correlation between approaches, for both mean and SD, calculating the following - R^2 [2], RMSE [3] and bias [4]. R^2 tested what percentage of the variance from the top-down approach could explain the variance of the bottom-up. I extracted the R^2 values from simple linear regression models I ran. I calculated these three values for all the data subsets, namely the total pool of data and data split by latitudes, biomes, and biomes by continent. In the results,

I present data that is only statistically significant ($p < 0.001$). Exhaustive tables of statistical results for both mean and SD values (whether significant or non-significant) are located in Appendix B.

$$[2] R^2 = \frac{\sum_{i=1} (y_{z,i} - \bar{y}_i)^2}{\sum_{i=1} (y_i - \bar{y}_i)^2}$$

I calculated RMSE to assess the average magnitude of error when correlating the top-down to the bottom-up estimates. RMSE was a more appropriate statistical parameter than Mean Absolute Error for the purpose of this analysis, as it poses higher weight onto undesired large errors. I averaged RMSE across pixels and by individual pixel so that I could plot the spatial variation of RMSE values.

$$[3] \text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}}$$

I calculated the bias to estimate the difference between bottom-up and top-down values and highlight the tendency of one to over- or under-estimate values compared to the other. Bias directly indicates the quality of the approaches for collecting and calculating the data. I calculated mean bias including all pixels, and bias by pixel so as to plot the spatial variation of this value.

$$[4] \text{Bias} = \frac{y - x}{2}$$

Where x is the benchmark, y is the model, y_z is the predicted y from the linear regression equation, and n is the count

I executed the entire analysis with R (v. 3.6.2) via RStudio (v. 1.2.5033; RStudio Team, 2019). In the course of this project, I created and modified a variety of functions, in order to automatise the coding process. See Appendix C for an exhaustive list of the packages implemented, the written or customised functions, and relevant code sections for data import, manipulation and analysis.

Visual analysis

I first visualised SLA mean and SD value distributions in space, plotting them on a 1x1 gridded map and on three-way bivariate plot, showing violin (mirrored density), box (summary statistics) and jittering (point scatter) plots. From these initial visualisations, I noticed an outlier from the top-down in both mean and SD distributions, and I removed the pixel from both approaches. See Appendix A to access the figures.

I produced a stippled map showing which mean values from the bottom-up approach fell within the 25pc-95pc CI of the top-down. This map was useful to calculate the percentage of stippling at the global, latitudinal and biome-level. I calculated

the percentage overlap of kernel density distributions for both mean and SD values.

When testing the sensitivity of SLA mean estimates against climate data, I calculated the percentage overlap between 2-dimensional density distributions of each approach. Finally, I calculated the percentage overlap between SLA mean values, after rounding them up to integers, from a pixel-by-pixel correlation. s

3 Results

RQ 1: Is there a strong degree of overlap between bottom-up and top-down mean and uncertainty SLA estimates, on global, latitudinal and biome scales?

Global scale

The heatscatters depict a line-of-best-fit which is non-linear (Fig. 2). In fact, the top-down variance of mean values explains only 0.03% of the model variance, and the correlation is not statistically significant ($p>0.001$). As for SD, the top-down variance explains 0.2% ($p<0.001$) of the bottom-up. The average magnitude error is relatively low ($\text{RMSE} = 8.33 \text{ m}^2/\text{kg}$) compared to the maximum value by pixel ($\text{RMSE} = 30 \text{ m}^2/\text{kg}$). This occurs similarly for SD ($\text{RMSE} = 10 \text{ m}^2/\text{kg}$), compared to the maximum RMSE value by pixel ($60 \text{ m}^2/\text{kg}$).

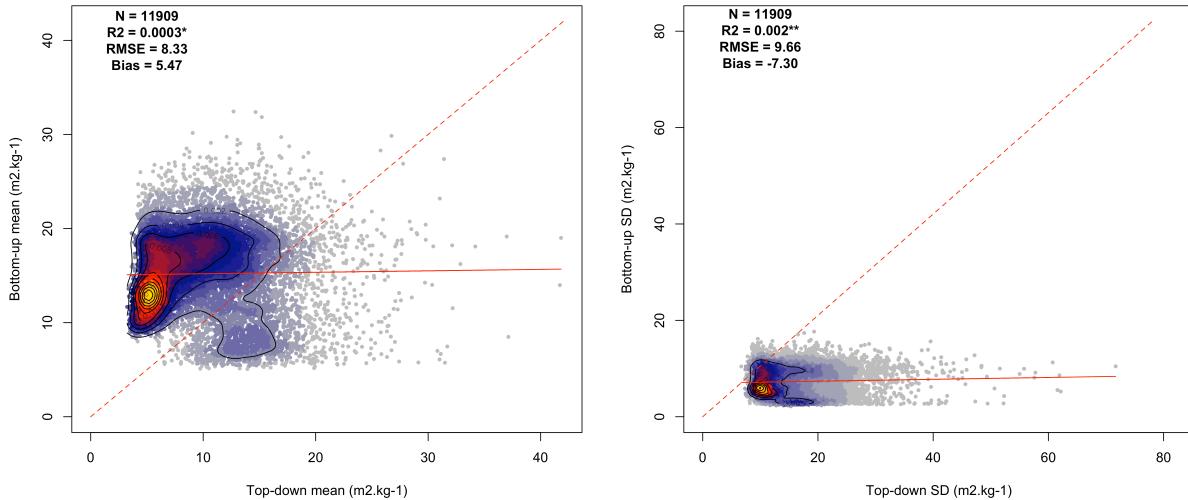


Figure 2: Scatterplot with 2-dimensional kernel density of the model dataset against the benchmark. The dashed line represents the ideal linear relationship between the two parameters, the continuous line represents the actual relationship (obtained from the linear regression equation). On the panels are also reported: sample size (N), coefficient of determination (R²), Root Mean Square Error (RMSE) and Bias. * $p>0.001$; ** $p<0.001$. P-values calculated from linear regression model.

Comparisons between global, latitude and biome scales

Considering the entire pool of data (global), the stippled pixels account for 58% ($N = 11909$; Fig. 3). From that percentage, only 10% is within the 25pc-75pc CI, and it is primarily located along the BFT biome in the northern hemisphere (green stipPLE). The black stipPLEs, counting for the remaining 48%, represent the mean values which lie outside of the 25pc-75pc CI but within the 25pc-95pc CI. Low stippling characterises tropical latitudes (30%), as well as TSDBF/TSMBF (approx. 14%) and M (0%) biomes.

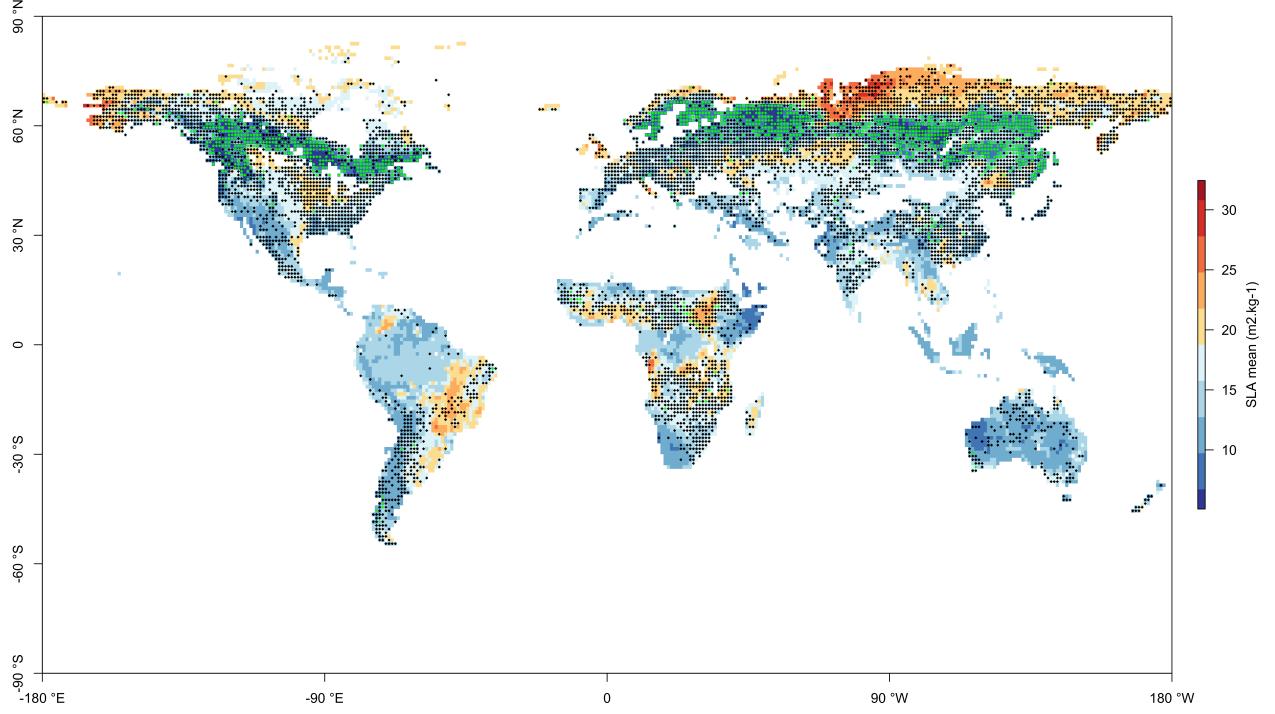


Figure 3: Baseline map shows bottom-up SLA mean values. The stippling shows which values from the bottom-up approach lie within the 25pc-75pc CI (green), and the outside of the 25pc-75pc but within the 25pc-95pc CI (black) of the top-down approach. Pixels without stippling show values that lie outside the CI

Table 1 shows the % overlap between mean and SD density distributions of the two approaches, as well as results for R^2 . Globally, despite the poor correlation, the overlap between mean and SD of the two approaches accounts for around 30% of the total data. Looking at the data by latitude, it is evident that the greatest overlap occurs along temperate latitudes, where more than half (57%) of the mean distributions have the same kernel density. The largest overlap between SD distributions is in the polar and tropical latitudes. The largest correlation between means (R^2) is in the tropics and subtropics, with 13% and 14%. Looking at the biomes, the TCF presents the largest overlap between mean distributions, with approximately 71%, followed by BFT with 45% and TGSS with 25%. In terms of correlation, DXS presents the largest R^2 value, where 32% of the variance in the top-down explains the variance in the bottom-up, followed by MGS (19%). TSGSS and TCF present a large % overlap between SD values. However, only the MGS and DXS present the largest correlation between SD which is around 10%, while all remaining biomes are below that threshold.

Table 1: Results for R^2 and percentage overlap of SLA mean and SD estimates on global, latitudinal and biome-scale analyses. In bold: R^2 values which are statistically significant ($p < 0.001$, from linear regression model), and overlap values higher than 50%.

Scale	% overlap (mean)	R^2 (mean)	R^2 (stdev)	% overlap (stdev)
Global	36.50	0.0003	0.002	26.10
Latitude				
Tropics	9.30	0.13	0.04	24.60
Subtropics	15.10	0.14	0.03	11.10
Temperate	57.10	0.03	0.00002	7.30
Pole	7.70	0.03	0.006	27.00
Biomes				
Boreal Forests/Taiga	45.20	0.02	0.03	4.50
Deserts & Xeric Shrublands	8.40	0.32	0.08	18.20
Mediterranean Forests, Woodlands & Scrub	7.70	0.06	0.01	10.70
Montane Grasslands & Shrublands	20.40	0.19	0.10	18.10
Temperate Conifer Forests	70.70	0.04	0.02	32.40
Temperate Grasslands, Savannas & Shrublands	24.50	0.10	0.06	15.60
Tropical & Subtropical Grasslands, Savannas & Shrublands	13.50	0.05	0.009	44.60
Tropical & Subtropical Moist Broadleaf Forests	2.10	0.04	0.002	8.10
Tundra	10.60	0.01	0.004	28.80

Geographical variation in RMSE and bias values is more varied for the correlation between means than between SDs (Fig. 4). The largest mean values errors (RMSE above 15 m²/kg) feature in the northern hemisphere, concentrated in the Eurasian T (max. RMSE = 21 m²/kg) and sparsely found in the BFT biome (max. RMSE = 29 m²/kg), and along tropical latitudes, in some parts of the TSGSS (max. RMSE = 21 m²/kg) and of the TSDBF (max. RMSE = 17 m²/kg). I also found the largest positive bias in T and tropical latitudes, particularly in TSGSS and TSMBF biomes (approx. 9-20 m²/kg). BFT and TCF present the only negative bias compared to the other subsets, with -0.37 m²/kg and -0.82 m²/kg. These are the only areas where mean top-down values tend to be larger than those found in the bottom-up approach.

Generally, RMSE for SD ranges between 0 and 10 m²/kg, indicating that the approaches have close pixel-by-pixel uncertainties. I found RMSE ranging between 15 m²/kg and 30 m²/kg in BFT. Bias between SDs is largely negative, indicating that top-down uncertainty tends to be larger than bottom-up. The largest negative bias is across BFT, with a median of 13 m²/kg. I found the lowest negative bias in T and tropical/subtropical latitudes and T biome. This does not indicate whether uncertainty is high or low, but only that in these areas the two approaches have similar uncertainty.

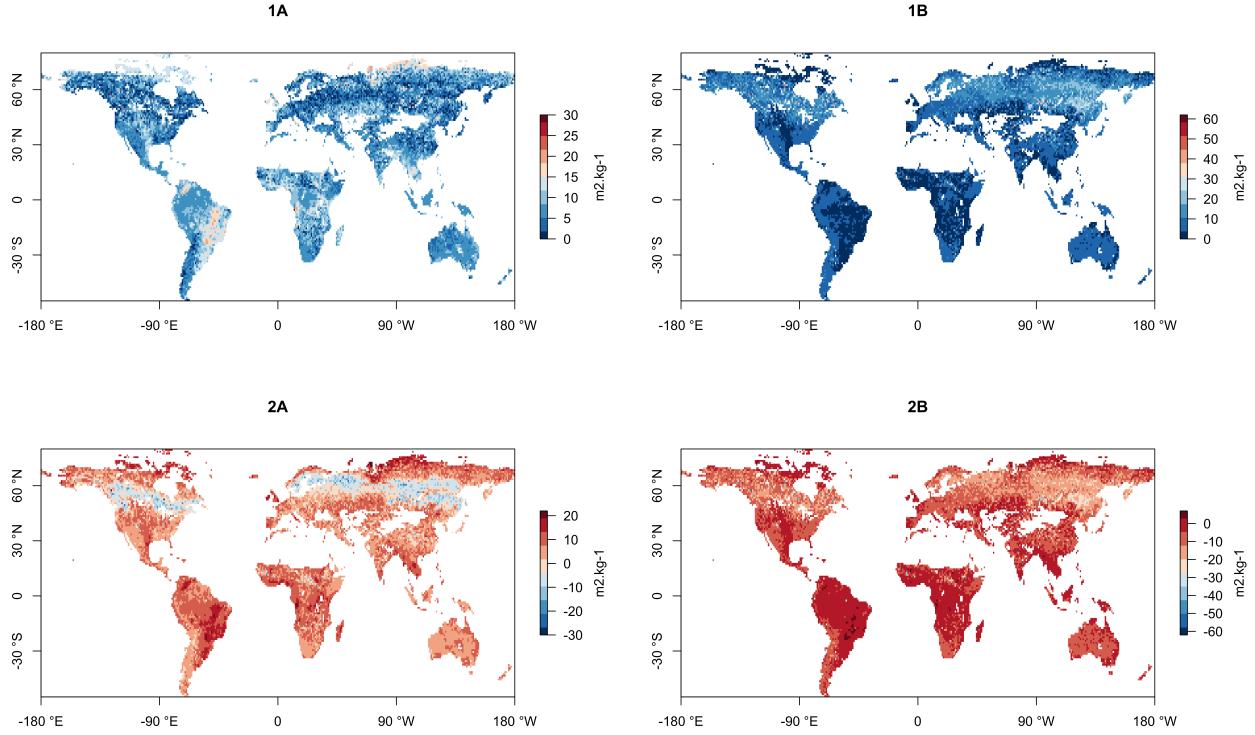


Figure 4: The four panels show geographical variation of the statistical results for SLA mean and SD values. 1A. RMSE mean; 1B: RMSE SD; 2A: bias mean; 2B: bias SD.

RQ 2: Is there a difference in overlap (mean and uncertainty estimates) between continents with the same biome?

Figure 5 shows the statistical outcomes for biomes split by the different continents (only $p < 0.001$). TGSS present the most striking correlation in Oceania between mean values, with $R^2 = 52\%$, followed by DXS with 38% in Africa and 27% in South America, and TSMBF with 23% in Africa. I found the largest error and bias in TSGSS in South America (RMSE = 13 m²/kg; bias = 12 m²/kg) and Africa (RMSE and bias = 8 m²/kg), the latter being also similar to TSMBF. TCF in Eurasia and North America do not have a strong correlation ($R^2=11\%$), but they have one of the lowest errors (RMSE = 7 and 6 m²/kg) and have a low negative bias (-0.83 m²/kg). BFT presents a similar pattern, with a negative bias in Eurasia, the only continent I found it to be significant, with -0.34 m²/kg.

Looking at the results from the SDs, I noticed that the biomes I found to be statistically significant do not match exactly the ones that are statistically significant between the means (Fig. 5). For instance, the correlation between SDs for TGSS is non-significant in Oceania. However, the same biome in Africa and North America present very different correlations, with $R^2 = 4\%$ in the former and $R^2=12\%$ in the latter, but have generally low RMSE (5 m²/kg) and negative bias (-3 m²/kg). Compared to the latter, TSGSS presents similar R^2 values but reversed for the two continents, and a low RMSE and negative bias (3 m²/kg and -3 m²/kg) in both continents. I found TCF to be one of the most correlated, with $R^2 =$

13%, and one with a lower RMSE ($7 \text{ m}^2/\text{kg}$) and negative bias ($-5 \text{ m}^2/\text{kg}$).

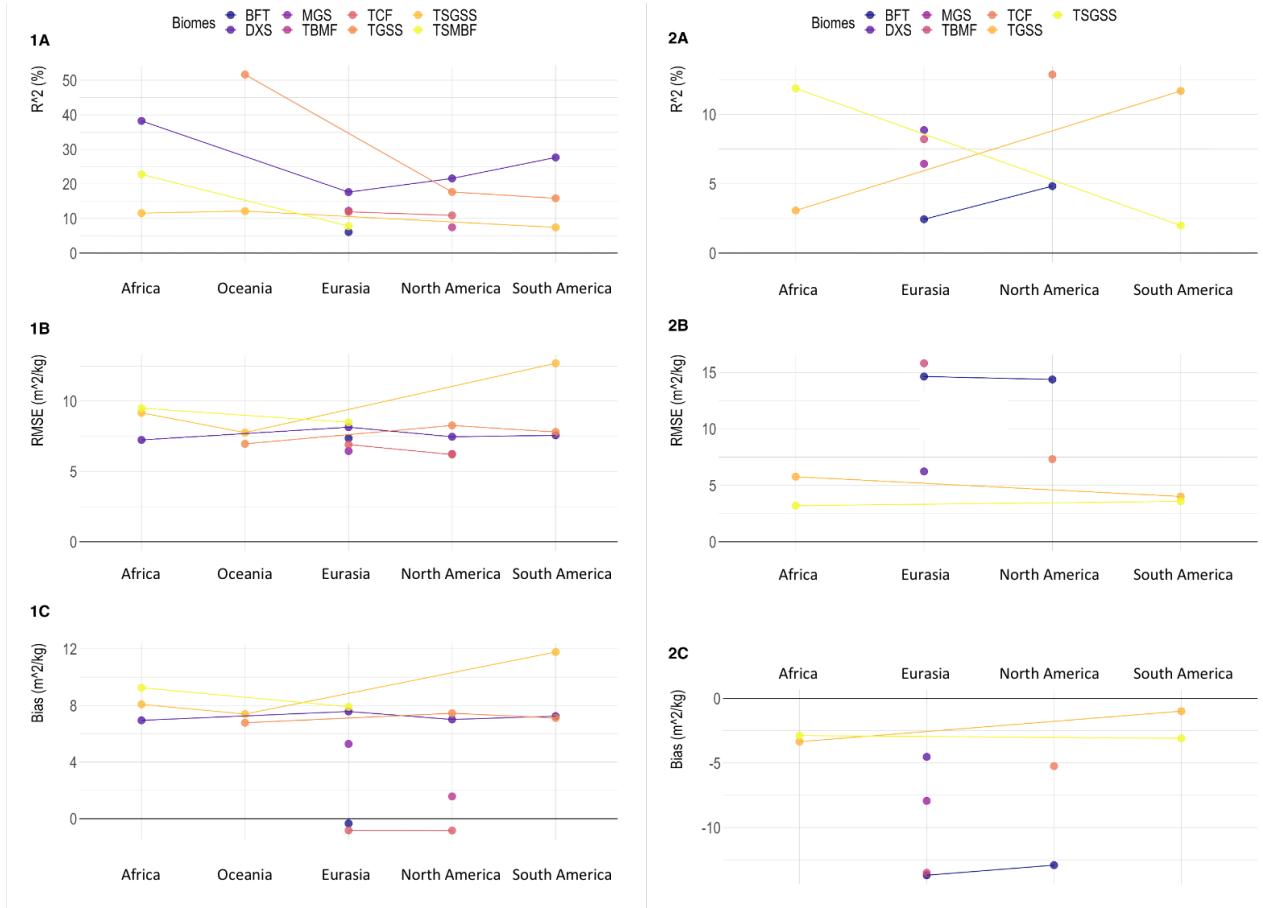


Figure 5: Average R^2 ($p<0.001$), RMSE and bias for biomes divided by continent. From 1A to 1C: R^2 , RMSE and bias results from the correlation between mean values. From 2A to 2C: R^2 , RMSE and bias results from the correlation between SD values. The lines connect the same biomes from different continents (only $p<0.001$, from linear regression). BFT: Boreal Forest Taiga. MGS: Montane Grasslands and Shrublands. TCF: Temperate Conifer Forests. TSGSS: Tropical and Subtropical Grasslands, Savannas and Shrublands. DXS: Desert and Xeric Shrublands. TBMF: Temperate Broadleaf/Mixed Forests. TGSS: Temperate Grasslands, Savannas and Shrublands. TSMBF: Tropical and Subtropical Moist Broadleaf Forests.

RQ 3: Do the SLA mean estimates have similar sensitivities to climate data, i.e. mean global precipitation and mean temperature?

Sensitivity to varying climate data

I found that SLA mean sensitivities to climate data overlap around 50% when plotted graphically through a 2d density distribution, which is concentrated around intermediate climate values (Fig. 6). Generally, the bottom-up approach presents greater variation in its density distribution within the climate space, with multiple density peaks and broader contour areas than those found in the top-down method. For the same level of precipitation at around 300 mm, bottom-up mean values are twice ($10 \text{ m}^2/\text{kg}$) and three times ($20 \text{ m}^2/\text{kg}$) higher than top-down ones ($5 \text{ m}^2/\text{kg}$) (Fig. 6, left). In addition, the ‘tails’ of the density distributions differ substantially from one another, as the top-down has mean values around $5 \text{ m}^2/\text{kg}$

while the bottom-up has around $13 \text{ m}^2/\text{kg}$ in correspondence of larger precipitation values (between 1,500 mm and 2,500 mm). The top-down also does not record data at precipitation levels reaching 3,000 mm. Looking at the interaction with temperature, the bottom-up has a density peak around $19 \text{ m}^2/\text{kg}$ in correspondence of -15°C (Fig. 6, right), while the top-down presents very little if unexistent data around this temperature. At 25°C , bottom-up and top-down present differing mean values, respectively concentrated around $13 \text{ m}^2/\text{kg}$ and $5 \text{ m}^2/\text{kg}$.

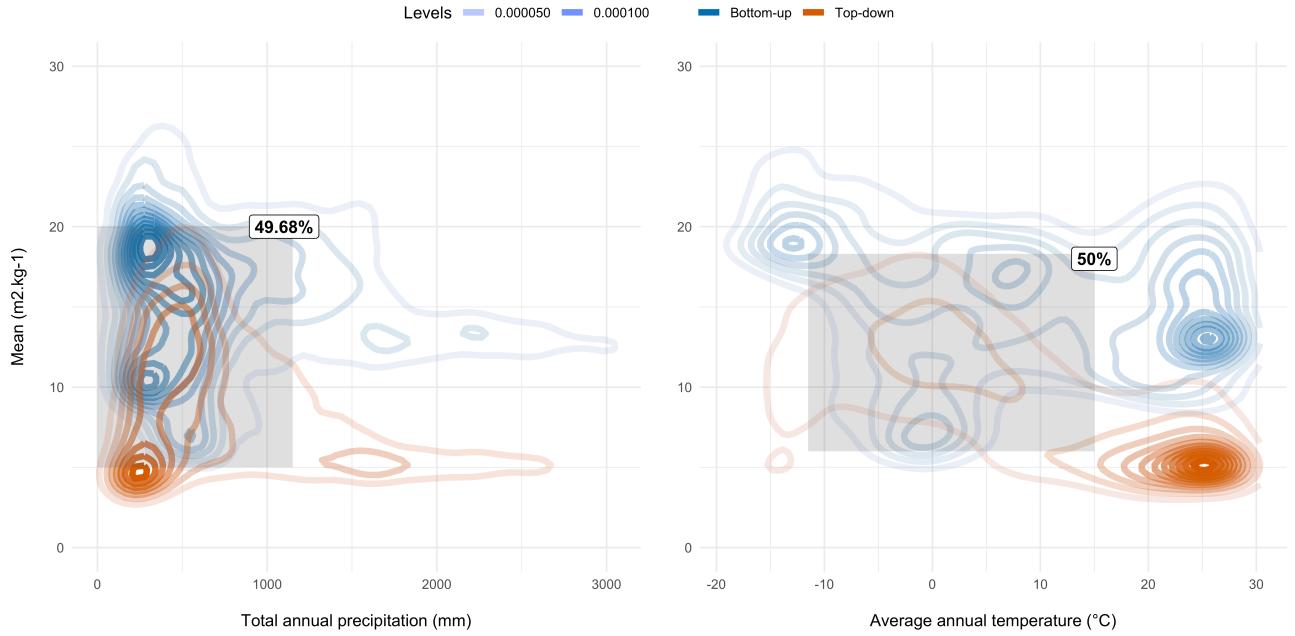


Figure 6: 2-dimensional densities on bivariate plot between SLA means and climate data. The levels represent the kernel density estimates and define the transparency of the contours. The grey boxes are a rough representation of the areas of overlap between densities.

Sensitivity to geographical variation

The bivariate legend allows to depict the spatial variation of the relationships between SLA mean and climate data, pixel-by-pixel (Fig. 7). Without the grouping by kernel density values, the overlap between mean values is very poor (2%). The greater amount of overlap is present primarily in the northern hemisphere, particularly the temperate latitudes. This is consistent with the 2d density plot, in the sense that the greater agreement between sensitivities is for intermediate ranges of climate data, and not for the extremes of the two gradients, located at polar and tropical latitudes.

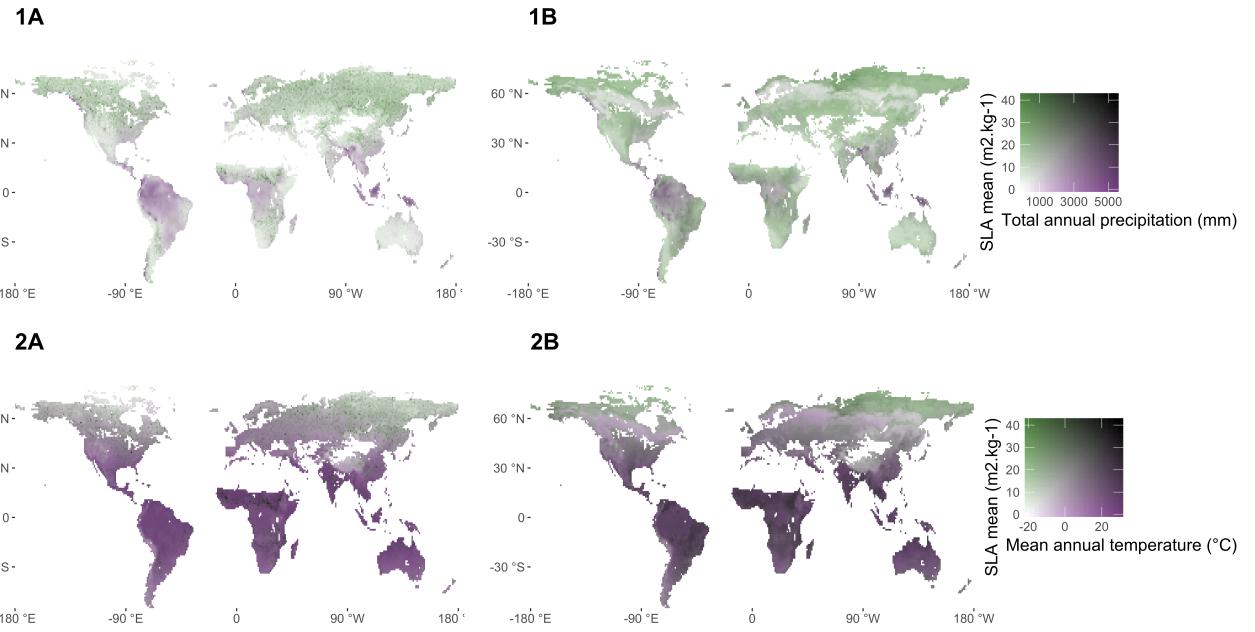


Figure 7: Global SLA-climate patterns. The bivariate legend showing the relationship of SLA against total annual precipitation (mm) and mean annual temperature ($^{\circ}\text{C}$) determines the color of the pixel. 1A/2A: top-down approach; 1B/2B: bottom-up approach.

4 Discussion

In the following section, I provide evidence on the lack of correlation visible across scales, from global to biome (RQ 1, H0). The inconsistency between estimates is undoubtedly linked to the spatial, temporal and protocol-related discrepancies between the two approaches. However, subdividing the total pool of data into latitudinal gradients and biomes helped to obtain a more detailed understanding of how consistency between SLA estimates varies across locations. In particular, the analysis at the latitudinal scale provided an insight into the lack of correlation at both tropical and polar latitudes: their complexity and remoteness make them very challenging environments to study. Unexpectedly, DXS presented the highest correlation, particularly over intermediate values of a wider range of SLA estimates that characterise it. Through the comparative analysis of continents, DXS and TGSS were more strongly correlated in Africa and Oceania than other continents, and compared to the other biomes (RQ 2, H2). In this context, continent-specific environmental characteristics, such as topography and vegetation type distribution, played an important role in favouring the increased consistency between approaches. Lastly, I found a particularly high overlap between SLA mean densities against climate data, though such overlap is not present in the pixel-by-pixel variation of SLA-climate relationships (RQ 3, H2). I provide evidence behind how both approaches show greater overlap at moderate climate ranges, rather than the extremes, typical of polar and tropical latitudes. Lastly, I suggest how the different SLA-climate sensitivity on the map can be related to the spatiotemporal mismatch that affects the two approaches. Ultimately, these findings question the comparability between approaches, as they likely reflect plant functional responses shaped by different sets of environmental variables.

4.1 RQ1: Is there a strong degree of overlap between bottom-up and top-down mean and uncertainty SLA estimates, on global, latitudinal and biome scales?

Inconsistency of results at the global scale

The lack of statistical significance between SLA mean values was surprising, considering that large sample sizes are expected to increase the confidence of predictability between parameters correlated. From the global analysis (Fig. 2), I could not obtain any insight into whether one approach depicts more accurately SLA variation patterns. The total range of SLA values differs substantially between approaches (Appendix A), and neither of the two extremes matches geographically. Both ranges of values are also different compared to another, more recent publication. Moreno-Martínez *et al.* (2018) used the ground-based data from TRY in conjunction with a variety of remote sensing platforms. The more updated version of the TRY database (Kattge *et al.*, 2020) than the one used by Butler *et al.* (2017), together with the integration of Landsat 5 high-spectral resolution onto MODIS spatial resolution, likely reflect the difference in results. From the incompatibility between the two approaches and against a third publication, external to this analysis, it appears that observed differences in global SLA patterns strongly relate to the methodologies used.

I was able to identify a clear spatial mismatch between bottom-up and top-down approaches: the global cover achievable with MODIS outcompetes the costly effort of covering ground with field projects, particularly in remote or less accessible areas. This reality is also sided by a temporal mismatch: the top-down approach has averaged one-pixel value from samples which have been taken every 1-2 days for the length of ten years (Bloom *et al.*, 2016); the bottom-up has aggregated ground-based data from short-term field projects (Butler *et al.*, 2017), which lack a temporal continuum. Lastly, the two methods present inherent measurement biases (Appendix A). The top-down estimates are predominantly skewed towards low values, showing a tendency to underestimate SLA mean values compared to the bottom-up. On the other hand, the bottom-up estimates are predominantly skewed towards high values, showing a tendency to overestimate SLA mean values, compared to the top-down.

The tendency from the top-down to underestimate SLA mean values, and to present larger swaths of uncertainty (SD) than the bottom-up, concerns method-related limitations of capturing ground information from space. Optical sensors possess moderate, but insufficient spatial and spectral sensitivity to depict within-pixel trait differences (Ruiz-Benito *et al.*, 2020), resulting in aggregation and underrepresentation of trait variability. Also, the optical sensor is bound to the observation of object surfaces (Ali *et al.*, 2016a; Gara *et al.*, 2019), capturing reflected radiation exclusively off the top of the canopy and neglecting the vertical trait heterogeneity. Considering that sunlit leaves on the top layer of the canopy typically show lower SLA values (Puglielli *et al.*, 2017), the spaceborne sensor could be limited to capturing primarily these across the globe. Lastly, the larger pixel-by-pixel uncertainty (Fig. 4) is explained by the interference of atmospheric particles, cloud dynamics (Serbin, 2019), topographic features (Wallis *et al.*, 2019), soil background (Ali *et al.*, 2017) and angle to the sun (Roelofsen *et al.*, 2013). These factors typically affect optical sensors, which require pre- and post- data calibration (Ma *et al.*, 2019) to account for them, in sacrifice of their measurement quality.

On the other hand, the tendency from the bottom-up to overestimate SLA mean values, relies on the limitations of the ground-based data used, in terms of lack of vegetation and site representativeness (Kattge *et al.*, 2020), and accuracy of non-standardised sampling protocols (Queenborough and Porras, 2014). In 2016, at the time of Butler *et al.* (2017)'s research, the TRY database hosted around 140 thousand entries related to SLA, which accounted for only 3.5% of the global floral diversity (Sandel *et al.*, 2015). Resource, time and site accessibility strongly limit the spatial representativeness of field sites (Asner *et al.*, 2016), resulting in higher rates of accessible, non-remote study areas. TRY is also biased towards representing growth forms which require relatively easier leaf measurement protocols, namely abundant species over rare, trees over shrubs, and broad-leaved over needle-leaved vegetation forms (Butler *et al.*, 2017; Kattge *et al.*, 2020). Intuitively, the most represented species are primarily found in resource-rich, low-disturbance environments (Ruiz-Benito *et al.*, 2020), therefore possessing higher SLA values. Lastly, the manually sampled leaves tend to be within-reach, in lower, shaded parts of the canopy (particularly if sampling tree species) which typically present higher SLA values (Ma *et al.*, 2020). These considerations confirm the general overestimation detected in SLA mean values from the bottom-up approach.

The degree of overlap strongly varies across geographical locations

The aforementioned biases in the two approaches do not apply to the entire spatial extent under observation. There are evident variations and exceptions, where SLA mean values and SD appear to be closer to one another, or other cases in which the biases are reversed for the two approaches. Methodologies are not the only cause for varying degrees of correlation across geographical locations. Other factors, possibly location-specific biotic and abiotic elements, interact differently with the two methodologies, resulting in more or less correlated SLA values.

Inconsistency in polar and tropical latitudes

Polar and tropical latitudes appear to share similar results of lack of correlation. Correlation between SDs results in low error and bias (Fig. 4), reflecting large uncertainty values being shared by both approaches (averaging $20\text{ m}^2/\text{kg}$). The results are consistent with the general understanding that these areas are one of the most complex and/or understudied (Asner *et al.*, 2016; Pennington and Lavin, 2016). The remoteness of polar latitudes discourages ground-based data collection, and its strong temperature seasonality and interannual variability (De Frenne *et al.*, 2013) likely impair accurate spatial data collection. At the same time, the complexity, high diversity of the tropics provides a variety of confounding effects, which undermine the reliability of ground-based data (Worthy and Swenson, 2019). Also, the multi-layered structure of tropical ecosystems is likely neglected by space-borne measurements (Serbin *et al.*, 2019). Therefore, both approaches present strong weaknesses in estimating SLA values over these areas, because of their complexity and extreme climatic variability.

Temperate latitudes

Temperate zones are considered to be the easiest areas to carry out trait estimation (Seyednasrollah and Clark, 2020), although my findings suggest the opposite. These areas present the largest stippling ratio outside the 25pc-75pc and within the 25pc-95pc (Fig. 3). The range of mean values presented in the outer quartiles of the CI is even wider than that highlighted for BFT, ranging between $15\text{ m}^2/\text{kg}$ and $40\text{ m}^2/\text{kg}$. This emphasises more strongly how the two extremes of the CI represent very different plant functional responses. Previous studies carried out along temperate latitudes have reported mixed results, some consistent with the bottom-up (Seyednasrollah and Clark, 2020), others closer to the top-down (Liu *et al.*, 2017). The bottom-up uses a collection of ground-based studies which cannot comprehensively account for the large variability of vegetation types and biomes that feature these latitudes. Conversely, the top-down likely depicts a more accurate image of SLA patterns along these latitudes, facilitated by the moderate climatic conditions and the clear seasonal leaf variability that characterises most vegetation types (Jiang *et al.*, 2018). Leaf seasonality (LL) is a very good tool for better estimating SLA averages with remote sensing, as it pronounces the temporal differences in trait variability (Dwyer, Hobbs and Mayfield, 2014). It, therefore, appears that the interannual variability of temperate latitudes contributes to the lack of consistency between approaches. This is further supported by the findings over the TCF biome, the only one with evergreen species along these latitudes, with high overlap between means and low bias between means and SDs.

The biome does not follow the seasonality typical of deciduous species, has a simple forest layer structure, made of one understory and one overstory, and has generally low species diversity (Ali *et al.*, 2016b; Serbin *et al.*, 2019). Ultimately, within temperate zones, only evergreen vegetation types from the TCF seem to be depicted more closely by the two approaches.

The Boreal Forest / Taiga (BFT)

The boreal forest biome is the only one to fall within the 25pc-75pc CI (Fig. 3), initially appearing to indicate very strong agreement between approaches. However, this ‘narrow’ CI ranges between approximately 5 m²/kg (25th) and 15 m²/kg (75th), indicating a wide SLA gradient. The two extremes of the CI specify very different vegetation types and evolutionary adaptations. Knowing that BFT does not present such variability in SLA values (Serbin *et al.*, 2019; Boonman *et al.*, 2020), there is a clear lack of accuracy in presenting results by either approach. The bottom-up presents low SLA mean values, skewed towards the 25th percentile of the CI, and consistent with another ground-based study (Goude, Nilsson and Holmström, 2019). These results match with the current understanding of BFT functionality, where evergreen, needle-leaved vegetation is adapted to nutrient-poor soils and low light availability (Kutbay *et al.*, 2016). Considering this, the top-down appears to be weaker at depicting SLA patterns over this biome, presenting greater SLA values and larger uncertainty (Fig. 4). This can be attributed to a misinterpretation of the foliar spectral signature by the sensor, influenced by the darker leaves of trees found in these areas. Darker leaves likely present a similar signature to that of shaded ones (Jiang *et al.*, 2018), which typically feature higher SLA values. Also weather-related interference, mainly snow (De Frenne *et al.*, 2013), likely contributes to the larger swath of uncertainty. Therefore, the original impression of strong agreement between approaches needs to be revised, in consideration of the very different mean values that the two present.

Grassland biomes

All grassland biomes generally show both low correlation and overlap between SLA mean values (Table 1). The fine-scale variation that characterises these biomes makes them one of the hardest to accurately estimate at a broad spatial scale (Pierce *et al.*, 2017). Previous studies have confirmed this, reporting that variation within grassland biomes does not tend to be reflected with global averages (Liu *et al.*, 2017). Within-biome differences are dependent on the dominant habitat – whether grassland, savanna, or shrubland – for the different proportions of graminoids, forbs, C4 and shrub that characterises them (Wellstein *et al.*, 2017). The-bottom-up data reflects the diverging patterns between such habitats types, for both TSGSS and TGSS. On the other hand, these differences are not well captured by the top-down. The shape and planar orientation of grass leaves, typically perpendicular to the ground (Serbin, 2019), are often poorly detected by satellite due to the nadir angle at which it captures images (Roelofsen *et al.*, 2013). Also, the size of graminoids is typically well below the spatial resolution of the space-borne optical sensor, and other elements – such as litter, canopy height and bare ground – likely contribute to the distortion of the perceived foliar spectral signature (Roelofsen *et al.*, 2013; Liu *et al.*, 2017). These biome-related characteristics weaken the accuracy of remotely sensed estimates, resulting in the observed

inconsistency between approaches.

Desert biome

SLA means are the most correlated in the DXS (Table 1). From my observations, SLA means are skewed to the lower value range for top-down, and the higher value range for bottom-up, respectively in agreement (Bodegom, Douma and Verheijen, 2014) and disagreement (Boonman *et al.*, 2020) with other studies. Plants that inhabit these environments can be characterised by very different adaptation mechanisms, to drought, fire and nutrient-poor soils (Onoda *et al.*, 2011; Anacker *et al.*, 2011). These mechanisms do not exclusively relate to low SLA values, but span a wider range, up to approximately 15 m²/kg (Hulshof *et al.*, 2013). The plants that are usually found at the extremes of the trait variation are rare species, primarily succulents. Studies have noted the difficulty of direct ground sampling of succulent plants, for the amount of water content contained in the leaves which challenge accurate measurements (John *et al.*, 2008). The water content also limits remote sensing, as it interacts with the internal leaf structure and dry biomass, confounding the foliar spectral signature (Ceccato *et al.*, 2001). Overall, it appears that the two approaches have consistent results around intermediate SLA values, possibly capturing plants which have not developed strong water retention strategies, unlike the succulents.

4.2 RQ2: Is there a difference in overlap (mean and uncertainty estimates) between continents with the same biome?

The degree of correlation is different between biomes split by continents, and it is not consistent for the same biome between different continents (H2). Grassland and desert biomes present the stronger correlations and TCF and BFT very low correlations. There is a striking difference between temperate and tropical biomes, particularly how these change between three major continents of the Southern Hemisphere: Oceania, Africa and South America.

Grassland biomes

TGSS and TSGSS biomes present variable degrees of correlation when compared across three continents (Fig. 5). Interestingly, these biomes are better correlated than forested biomes. This does not agree with the aforementioned bias from the TRY global trait database, towards better representation of trees over herbaceous species (Kattge *et al.*, 2020). Concerning TGSS, the larger correlation found in Oceania could be attributed to the relatively more homogeneous vegetation type distribution that is continent-specific (Jardine *et al.*, 2020). C3 grasses are known to cover around 90% of Oceania (Woodward, Lomas and Kelly, 2004). On the other hand, in North and South America the distribution of C3 grasses accounts for 40% and 30% of the total cover (Woodward, Lomas and Kelly, 2004), thus increasing the mismatch between SLA values with increased vegetation type variability. The South-American biome is under-represented by the bottom-up, after having classified a large proportion of its land cover as cropland (Butler *et al.*, 2017). Also, TSGSS presents low

correlation, high error and bias in South America, compared to Africa and Oceania. The greater discrepancy between estimates in South America is primarily attributed to the remarkable differences between savannas that distinguish this continent from the others (Ali *et al.*, 2016a). A striking example is the South American Cerrado, the biologically richest savanna in the world, hosting high rates of endemism (Jardine *et al.*, 2020). This savanna habitat is very different compared to African savannas, which present sparse vegetation, and are characterised by lower species richness (Parr *et al.*, 2014; Ringelberg *et al.*, 2020). Such a high diversity of vegetation and vegetation types challenges the accuracy of estimation from both approaches. Overall, the data under-representation of the biomes by the bottom-up, along with the spectral limitations of the optical sensors affecting the top-down, are elements which support the poorer correlation found in South America compared to the other two continents.

Desert biomes

DXS appears to be very strongly correlated, and particularly in Africa (Fig. 5). The topography is an environmental factor which likely contributes to continent-specific differences for this biome. Optical sensors need calibration against shaded slopes, particularly in high elevated regions of a biome, increasing the uncertainty of the trait representation (Serbin, 2019). Previous field-based studies have also focused on the effect of topography on SLA variation, and have confirmed this factor as being compromising for ground-based studies (De Frenne *et al.*, 2013; Hulshof *et al.*, 2013; Wallis *et al.*, 2019). In this context, the larger correlation found within the African desert biome (excluding Sahara for lack of data), the Kalahari and Namib deserts, is attributed to the very flat slope that characterises it, compared to the higher and rougher surfaces typical of North American and Eurasian deserts (John *et al.*, 2008).

4.3 RQ 3: Do the SLA mean estimates have similar sensitivities to climate data, i.e. mean global precipitation and mean temperature?

Considering the strong SLA-environment interactions, it seemed sensible to determine the degree of similarity of SLA mean sensitivities to climate data (Fig. 6). The two sets of mean values do present similarities, particularly at moderate climatic ranges typical of temperate latitudes. The sensitivities were least correlated at the two climatic extremes, typical of tropical and polar latitudes. However, the percentage overlap found is not as strong when sensitivities are correlated pixel-by-pixel (Fig. 7). The geographical inconsistency is likely due to the different spatiotemporal scales at which the two approaches retrieved the baseline data. These findings question more broadly the feasibility of comparing the two approaches.

Sensitivity to precipitation

The amount and seasonal frequency of precipitation strongly drive SLA patterns across the world (Costa-Saura *et al.*, 2017) (Fig. 6, left). This relationship with precipitation is evident in each approach. However, similar sensitivities to annual precipitation occur only in temperate regions. This iterates the challenge of measuring SLA in tropical and polar

regions. The tropics are characterised by high mean annual temperature and high humidity (Dalle Fratte *et al.*, 2019), particularly in TSMBF and the so-called ‘cloud forests’ (Long, Zang and Ding, 2011; Long *et al.*, 2015). T, on the other hand, is characterised by long-lasting frost seasons (De Frenne *et al.*, 2013). These factors prevent constant retrieval of high-quality satellite imagery and make the areas very unpopular to carry out field-work, as reflected in the low number of field sites used by the bottom-up approach.

Sensitivity to temperature

The pattern of SLA-temperature relationships is very similar between approaches, despite SLA values being increasingly different towards the extremes of the latitudinal gradient (Fig. 6, right). Previous research showed that temperature extremes result in plants with similar functionalities (Asner *et al.*, 2016; Borgy *et al.*, 2017). Vegetation types adapted to extreme mean annual temperatures do typically overlap in SLA values, reflecting the shared response to harsh environmental conditions (Asner *et al.*, 2016; Pennington and Lavin, 2016). Tropical vegetation builds structural resilience in surviving from competition for resource acquisition (Duran *et al.*, 2019), while T vegetation develops the structural sturdiness required to cope with frost seasons and poor light availability (Myers□Smith, Thomas and Bjorkman, 2019). Despite the different SLA values, both can detect such SLA-temperature pattern along the latitudinal gradient.

Spatiotemporal mismatch, different climate sensitivities

Observing the spatial variation in SLA-climate relationships, there is little overlap in sensitivity, which must be attributed to the spatiotemporal mismatch between approaches. Research shows how external abiotic factors affecting SLA values change greatly in time and space (Dwyer, Hobbs and Mayfield, 2014; Gillison, 2019). At the local scale and in the short-term, SLA is very sensitive to microclimatic and edaphic factors (Sandel *et al.*, 2015), such as pH, nutrient content and hydraulic conductivity (Díaz *et al.*, 2016), as well as topography-related factors (Zhang *et al.*, 2017), such as wind speed, irradiance, cloud/snow cover and ‘rain shadows’. At the global scale and in the longer-term, macroclimate, soil (i.e. type, composition, erosion rates) and water availability are the stronger environmental drivers of SLA patterns (Mucina, 2019). The bottom-up has predicted global estimates based on local, sparsely distributed, short-term field studies. Therefore, the resulting map represents SLA patterns that are not responsive to global environmental drivers, rather they reflect a static floral response to short-term, local abiotic factors. On the other hand, the top-down has extrapolated estimates based on the continuous sampling of coarse-resolution data, over a decade. The resulting map is an aggregation of the dynamic change in SLA over time, shaped by the large-scale environmental drivers. From these considerations, it is possible to infer that SLA reported by the two approaches reflects sensitivities to climate data which are not comparable, due to the spatiotemporal sampling mismatch.

4.4 Study limitations

Methods

For this project, I converted the variable LCMA to SLA from the top-down, to standardise trait values between approaches. To do so, I followed the conventional assumption that carbon constitutes 50% of the biomass. Despite being widely accepted, this conversion likely underrepresents LCMA, which was originally estimated. I also converted the 25pc, 75pc and 95pc CI values from the top-down originally referring to LCMA, to represent the converted SLA (Appendix A). This also presents inherent uncertainty, as converted CI values do not truly reflect those that would have been calculated from the direct estimation of SLA. To prevent any potential uncertainty derived from these conversions, it would have been preferable to compare SLA values from direct measurements for both approaches.

Lastly, in terms of the environmental data used to assess the overlap in sensitivity of the two approaches, I chose climate data as this is typically correlated with SLA at the global scale. Research findings have also pointed out the relevance of relationships between soil and elevation data with SLA (Asner *et al.*, 2016). I have not tested such sensitivities in this research, however, they can constitute an interesting area to delve into for future research.

Statistics

The observations and inferences made do not entirely rely on the statistical results obtained, but a large part of the research is based on visual analysis. I chose this method as such comparative analysis requires more than a statistical evaluation to infer the state of similarity between approaches. I believe this decision was proven sensible, as often the correlation results and observational outputs portrayed conflicting meanings.

4.5 Implications for ecological disciplines and ESM, and future directions

Implications for ecological sciences

SLA applications are varied in ecology, namely biodiversity monitoring, detection of ecosystem state shifts (Hoffmann *et al.*, 2005), detecting plant acclimation and adaptation to stresses (Pierce *et al.*, 2017; Ruiz-Benito *et al.*, 2020), and understanding local to global patterns of vegetation functional response (Boonman *et al.*, 2020). SLA strongly relates to evolutionary, historical, biogeographic and ecological processes that drive species niche partitioning (Worthy and Swenson, 2019), environmental filtering (Costa-Saura *et al.*, 2017), and establishment (Crisp *et al.*, 2009). Researchers should acknowledge the noted differences derived between approaches that measured this trait at the global scale. The outcome of the comparative analysis cannot confirm whether either approach is more reliable. The strength of the bottom-up approach lies in its ability to retain the fine-scale patterning that is critical for depicting static, intra- and interspecific spatial trait variation. The strength of the top-down lies in the ability to capture broad-scale, coarse spatiotemporal trait variation.

However, some specific areas, because of location-specific biotic and abiotic characteristics, are better measured by both – namely, grasslands and deserts. The greater correlations found in continent-specific biomes should incentivise ecologists to understand what factors facilitate the measurement of SLA with contrasting methodologies within smaller geographical areas.

Implications for ESM

ESM greatly relies on SLA accuracy, to simulate effects of global climate change on plants' ability to cycle carbon, water and energy (Serbin *et al.*, 2019; Menezes-Silva *et al.*, 2019). Bloom *et al.* (2016) and Butler *et al.* (2017) pioneered global mapping of SLA for ESM characterisation, yet they present too-large uncertainty. The outcome of this comparative analysis suggests that using the global map ESM from one approach would lead to very different model results compared to using the other. From the different baseline data used to map the trait, the two approaches reflect SLA sensitivities related to opposed spatiotemporal scales. Considering that ESM aims to predict carbon balance changes, a compromise is necessary to preserve both the spatiotemporal dynamicity and the fine-scale variability of SLA at the global scale. From the evidence discussed above, I support previous statements (Ma *et al.*, 2020) on the absolute necessity to integrate methodologies from such different approaches, to achieve a more accurate estimation of SLA patterns. Recent remote sensing technologies allow for finer spatial and spectral resolution (e.g. hyperspectral, radar and LiDAR) (Asner *et al.*, 2016; Ma *et al.*, 2020), unachievable with spaceborne optical sensors like MODIS. The complementarity between different Earth Observation constellations, together with the intermediate support of airborne platforms (Ruiz-Benito *et al.*, 2020), will result in improved SLA estimation. However, the accuracy of the latter is undoubtedly linked to the information that is available on the ground. It is, in fact, fundamental to increase the number of ground-based data (Bruelheide *et al.*, 2019; Kattge *et al.*, 2020; Gallagher *et al.*, 2020), to use for calibration and validation of remotely sensed products. The suggested interdisciplinary research implies a substantial financial and human challenge, particularly if attempting to uncover the entire globe at once. Taking one step back, to focus research efforts over a smaller geographical area, is the feasible alternative to begin the more accurate estimation of SLA.

5 Conclusion

The study has shed light on the strengths and weaknesses of the top-down and bottom-up approaches when mapping global SLA patterns. The spatial, temporal and protocol-related mismatch strongly affects the lack of correlation found between approaches. However, degrees of correlation fluctuate across geographical locations. In particular, the analysis at the biome-level highlighted novel findings, depicting deserts and grassland biomes with the most consistent results. This should stimulate scientific research to reconsider the conventional belief that only a few biomes and vegetation types, for instance, broad-leaved temperate forests, are the easiest to measure with different methodologies. Considering the

repeating trend of poor correlation and large uncertainty found along tropical and polar regions, new research is urged to focus more over these understudied, and under-represented locations. Also, the different results raised at the continent-specific biome, compared to the larger spatial scales treated beforehand, should offer important insights on the value of analysing regional trait variation that can be later scaled up to the global extent. A profound difference between approaches lies in the opposite spatiotemporal scales at which the initial trait data were extrapolated. This suggests that SLA values output from the two approaches likely reflect different plant sensitivities, and for that, they cannot be compared. However, this inference requires validation: it should be the aim of future research to assess whether time, in particular, affects the incompatibility between the obtained sets of estimates. Ultimately, the combination of broad-scale spatiotemporal dynamicity provided by the top-down approach, with the fundamental fine-scale static variation provided by the bottom-up approach, would result in the most optimal method to accurately map SLA patterns. These findings are not only relevant to inform the current quality of ESM parametrisation, but also to update the current state-of-the-art knowledge on global-scale SLA accuracy, applicable to the variety of ecological disciplines in which the trait is constantly featured.

6 Bibliography

- Ali, A. M., Darvishzadeh, R., Skidmore, A. K. and Duren, I. van. (2016a). Effects of Canopy Structural Variables on Retrieval of Leaf Dry Matter Content and Specific Leaf Area From Remotely Sensed Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9 (2), pp.898–909.
- Ali, A. M., Darvishzadeh, R., Skidmore, A. K. and van Duren, I. (2017). Specific leaf area estimation from leaf and canopy reflectance through optimization and validation of vegetation indices. *Agricultural and Forest Meteorology*, 236, pp.162–174.
- Ali, A. M., Darvishzadeh, R., Skidmore, A. K., Duren, I. van, Heiden, U. and Heurich, M. (2016b). Estimating leaf functional traits by inversion of PROSPECT: Assessing leaf dry matter content and specific leaf area in mixed mountainous forest. *International Journal of Applied Earth Observation and Geoinformation*, 45, pp.66–76.
- Anacker, B., Rajakaruna, N., Ackerly, D., Harrison, S., Keeley, J. and Vasey, M. (2011). Ecological strategies in California chaparral: interacting effects of soils, climate, and fire on specific leaf area. *Plant Ecology & Diversity*, 4 (2–3), pp.179–188.
- Asner, G. P., Knapp, D. E., Anderson, C. B., Martin, R. E. and Vaughn, N. (2016). Large-scale climatic and geophysical controls on the leaf economics spectrum. *Proceedings of the National Academy of Sciences*, 113 (28), pp.E4043–E4051.
- Bloom, A. A., Exbrayat, J.-F., Velde, I. R. van der, Feng, L. and Williams, M. (2016). The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences*, 113 (5), pp.1285–1290.
- Bodegom, P. M. van, Douma, J. C. and Verheijen, L. M. (2014). A fully traits-based approach to modeling global vegetation distribution. *Proceedings of the National Academy of Sciences*, 111 (38), National Academy of Sciences., pp.13733–13738.
- Boonman, C. C. F., Benítez-López, A., Schipper, A. M., Thuiller, W., Anand, M., Cerabolini, B. E. L., Cornelissen, J. H. C., Gonzalez-Melo, A., Hattingh, W. N., Higuchi, P., et al. (2020). Assessing the reliability of predicted plant trait distributions at the global scale. *Global Ecology and Biogeography*, n/a (n/a).
- Borgy, B., Violle, C., Choler, P., Denelle, P., Munoz, F., Kattge, J., Lavorel, S., Loranger, J., Amiaud, B., Bahn, M., et al. (2017). Plant community structure and nitrogen inputs modulate the climate signal on leaf traits. *Global Ecology and Biogeography*, 26 (10), pp.1138–1152.
- Bruelheide, H., Dengler, J., Jiménez-Alfaro, B., Purschke, O., Hennekens, S. M., Chytrý, M., Pillar, V. D., Jansen, F., Kattge, J., Sandel, B., et al. (2019). sPlot – A new tool for global vegetation analyses. *Journal of Vegetation Science*, 30 (2), pp.161–186.
- Butler, E. E., Datta, A., Flores-Moreno, H., Chen, M., Wythers, K. R., Fazayeli, F., Banerjee, A., Atkin, O. K., Kattge, J., Amiaud, B., et al. (2017). Mapping local and global variability in plant trait distributions. *Proceedings of the National Academy of Sciences*, 114 (51), pp.E10937–E10946.
- Calow, P. (1987). Towards a Definition of Functional Ecology. *Functional Ecology*, 1 (1), [British Ecological Society, Wiley]., pp.57–61. JSTOR
- Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S. and Grégoire, J.-M. (2001). Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sensing of Environment*, 77 (1), pp.22–33.
- Costa-Saura, J. M., Trabucco, A., Spano, D. and Mereu, S. (2017). Environmental filtering drives community specific leaf area in Spanish forests and predicts relevant changes under future climatic conditions. *Forest Ecology and Management*, 405, pp.1–8.
- Crisp, M. D., Arroyo, M. T. K., Cook, L. G., Gandolfo, M. A., Jordan, G. J., McGlone, M. S., Weston, P. H., Westoby, M., Wilf, P. and Linder, H. P. (2009). Phylogenetic biome conservatism on a global scale. *Nature*, 458 (7239), Nature Publishing

Group., pp.754–756.

Dalle Fratte, M., Brusa, G., Pierce, S., Zanzottera, M. and Cerabolini, B. E. L. (2019). Plant trait variation along environmental indicators to infer global change impacts. *Flora*, 254, pp.113–121.

De Frenne, P., Graae, B. J., Rodriguez-Sanchez, F., Kolb, A., Chabrierie, O., Decocq, G., De Kort, H., De Schrijver, A., Diekmann, M., Eriksson, O., et al. (2013). Latitudinal gradients as natural laboratories to infer species' responses to temperature. *Journal of Ecology*, 101 (3), pp.784–795.

Diaz, S., Hodgson, J. G., Thompson, K., Cabido, M., Cornelissen, J. H. C., Jalili, A., Montserrat-Marti, G., Grime, J. P., Zarinkamar, F., Asri, Y., et al. (2004). The plant traits that drive ecosystems: Evidence from three continents. *Journal of Vegetation Science*, 15 (3), pp.295–304.

Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Colin Prentice, I., et al. (2016). The global spectrum of plant form and function. *Nature*, 529 (7585), Nature Publishing Group., pp.167–171.

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., et al. (2017). An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience*, 67 (6), Oxford Academic., pp.534–545.

Duran, S. M., Martin, R. E., Diaz, S., Maitner, B. S., Malhi, Y., Salinas, N., Shenkin, A., Silman, M. R., Wieczynski, D. J., Asner, G. P., et al. (2019). Informing trait-based ecology by assessing remotely sensed functional diversity across a broad tropical temperature gradient. *Science Advances*, 5 (12), Washington : Amer Assoc Advancement Science.

Dwyer, J. M., Hobbs, R. J. and Mayfield, M. M. (2014). Specific leaf area responses to environmental gradients through space and time. *Ecology*, 95 (2), pp.399–410.

Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37 (12), pp.4302–4315.

Freschet, G. T., Dias, A. T. C., Ackerly, D. D., Aerts, R., Bodegom, P. M. van, Cornwell, W. K., Dong, M., Kurokawa, H., Liu, G., Onipchenko, V. G., et al. (2011). Global to community scale differences in the prevalence of convergent over divergent leaf trait distributions in plant assemblages. *Global Ecology and Biogeography*, 20 (5), pp.755–765.

Gallagher, R. V., Falster, D. S., Maitner, B. S., Salguero-Gómez, R., Vandvik, V., Pearse, W. D., Schneider, F. D., Kattge, J., Poelen, J. H., Madin, J. S., et al. (2020). Open Science principles for accelerating trait-based science across the Tree of Life. *Nature Ecology & Evolution*, 4 (3), Nature Publishing Group., pp.294–303.

Gara, T. W., Darvishzadeh, R., Skidmore, A. K., Wang, T. and Heurich, M. (2019). Accurate modelling of canopy traits from seasonal Sentinel-2 imagery based on the vertical distribution of leaf traits. *Isprs Journal of Photogrammetry and Remote Sensing*, 157, Amsterdam : Elsevier., pp.108–123.

Gillison, A. N. (2019). Plant functional indicators of vegetation response to climate change, past present and future: I. Trends, emerging hypotheses and plant functional modality. *Flora*, 254, pp.12–30.

Goude, M., Nilsson, U. and Holmström, E. (2019). Comparing direct and indirect leaf area measurements for Scots pine and Norway spruce plantations in Sweden. *European Journal of Forest Research*, 138 (6), pp.1033–1047.

He, D., Chen, Y., Zhao, K., Cornelissen, J. H. C. and Chu, C. (2018). Intra- and interspecific trait variations reveal functional relationships between specific leaf area and soil niche within a subtropical forest. *Annals of Botany*, 121 (6), pp.1173–1182.

Heilmeier, H. (2019). Functional traits explaining plant responses to past and future climate changes. *Flora*, 254, pp.1–11.

Hoffmann, W. A., Franco, A. C., Moreira, M. Z. and Haridasan, M. (2005). Specific leaf area explains differences in leaf traits between congeneric savanna and forest trees. *Functional Ecology*, 19 (6), pp.932–940.

Hulshof, C. M., Violle, C., Spasojevic, M. J., McGill, B., Damschen, E., Harrison, S. and Enquist, B. J. (2013). Intra-specific and inter-specific variation in specific leaf area reveal the importance of abiotic and biotic drivers of species diversity across

- elevation and latitude. *Journal of Vegetation Science*, 24 (5), pp.921–931.
- Jardine, E. C., Thomas, G. H., Forrestel, E. J., Lehmann, C. E. R. and Osborne, C. P. (2020). The global distribution of grass functional traits within grassy biomes. *Journal of Biogeography*, 47 (3), pp.553–565.
- Jiang, J., Comar, A., Burger, P., Bancal, P., Weiss, M. and Baret, F. (2018). Estimation of leaf traits from reflectance measurements: comparison between methods based on vegetation indices and several versions of the PROSPECT model. *Plant Methods*, 14 (1), p.23.
- John, R., Chen, J., Lu, N., Guo, K., Liang, C., Wei, Y., Noormets, A., Ma, K. and Han, X. (2008). Predicting plant diversity based on remote sensing products in the semi-arid region of Inner Mongolia. *Remote Sensing of Environment*, 112 (5), pp.2018–2032.
- Kattge, J., Boenisch, G., Diaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Tautenhahn, S., Werner, G. D. A., Aakala, T., Abedi, M., et al. (2020). TRY plant trait database - enhanced coverage and open access. *Global Change Biology*, 26 (1), Hoboken : Wiley., pp.119–188.
- Kutbay, H. G., Çakmak, A., Yılmaz, H. and Sürmen, B. (2016). Comparison of Leaf Traits (SLA And LMA) on Different Populations of *Alcea* *apterocarpa*. *Hacettepe Journal of Biology and Chemistry*, 2 (44), pp.125–125.
- Liu, M., Wang, Z., Li, S., Lü, X., Wang, X. and Han, X. (2017). Changes in specific leaf area of dominant plants in temperate grasslands along a 2500-km transect in northern China. *Scientific Reports*, 7 (1), pp.1–9.
- Long, W., Schamp, B. S., Zang, R., Ding, Y., Huang, Y. and Xiang, Y. (2015). Community assembly in a tropical cloud forest related to specific leaf area and maximum species height. *Journal of Vegetation Science*, 26 (3), pp.513–523.
- Long, W., Zang, R. and Ding, Y. (2011). Air temperature and soil phosphorus availability correlate with trait differences between two types of tropical cloud forests. *Flora*, 206 (10), pp.896–903.
- Ma, X., Mahecha, M. D., Migliavacca, M., van der Plas, F., Benavides, R., Ratcliffe, S., Kattge, J., Richter, R., Musavi, T., Baeten, L., et al. (2019). Inferring plant functional diversity from space: the potential of Sentinel-2. *Remote Sensing of Environment*, 233, p.111368.
- Ma, X., Migliavacca, M., Wirth, C., Bohn, F. J., Huth, A., Richter, R. and Mahecha, M. D. (2020). Monitoring Plant Functional Diversity Using the Reflectance and Echo from Space. *Remote Sensing*, 12 (8), Multidisciplinary Digital Publishing Institute., p.1248.
- Menezes-Silva, P. E., Loram□Lourenço, L., Alves, R. D. F. B., Sousa, L. F., Almeida, S. E. da S. and Farnese, F. S. (2019). Different ways to die in a changing world: Consequences of climate change for tree species performance and survival through an ecophysiological perspective. *Ecology and Evolution*, 9 (20), pp.11979–11999.
- Messier, J., McGill, B. J., Enquist, B. J. and Lechowicz, M. J. (2017). Trait variation and integration across scales: is the leaf economic spectrum present at local scales? *Ecography*, 40 (6), pp.685–697.
- Moncrieff, G. R., Bond, W. J. and Higgins, S. I. (2016). Revising the biome concept for understanding and predicting global change impacts. *Journal of Biogeography*, 43 (5), pp.863–873.
- Mucina, L. (2019). Biome: evolution of a crucial ecological and biogeographical concept. *New Phytologist*, 222 (1), pp.97–114.
- Myers□Smith, I. H., Thomas, H. J. D. and Bjorkman, A. D. (2019). Plant traits inform predictions of tundra responses to global change. *New Phytologist*, 221 (4), pp.1742–1748. [Online].
- Onoda, Y., Westoby, M., Adler, P. B., Choong, A. M. F., Clissold, F. J., Cornelissen, J. H. C., Díaz, S., Dominy, N. J., Elgart, A., Enrico, L., et al. (2011). Global patterns of leaf mechanical properties. *Ecology Letters*, 14 (3), pp.301–312.
- Parr, C. L., Lehmann, C. E. R., Bond, W. J., Hoffmann, W. A. and Andersen, A. N. (2014). Tropical grassy biomes: misunderstood, neglected, and under threat. *Trends in Ecology & Evolution*, 29 (4), pp.205–213.

- Pennington, R. T. and Lavin, M. (2016). The contrasting nature of woody plant species in different neotropical forest biomes reflects differences in ecological stability. *New Phytologist*, 210 (1), pp.25–37.
- Pierce, S., Negreiros, D., Cerabolini, B. E. L., Kattge, J., Diaz, S., Kleyer, M., Shipley, B., Wright, S. J., Soudzilovskaia, N. A., Onipchenko, V. G., et al. (2017). A global method for calculating plant CSR ecological strategies applied across biomes world-wide. *Functional Ecology*, 31 (2), pp.444–457.
- Puglielli, G., Varone, L., Gratani, L. and Catoni, R. (2017). Specific leaf area variations drive acclimation of *Cistus salvifolius* in different light environments. *Photosynthetica*, 55 (1), pp.31–40.
- Queenborough, S. A. and Porras, C. (2014). Expanding the coverage of plant trait databases – A comparison of specific leaf area derived from fresh and dried leaves. *Plant Ecology & Diversity*, 7 (1–2), pp.383–388.
- Reich, P. B., Ellsworth, D. S., Walters, M. B., Vose, J. M., Gresham, C., Volin, J. C. and Bowman, W. D. (1999). Generality of Leaf Trait Relationships: A Test Across Six Biomes. *Ecology*, 80 (6), pp.1955–1969.
- Ringelberg, J. J., Zimmermann, N. E., Weeks, A., Lavin, M. and Hughes, C. E. (2020). Biomes as evolutionary arenas: Convergence and conservatism in the trans-continental succulent biome. *Global Ecology and Biogeography*, n/a (n/a).
- Roelofsen, H. D., Van Bodegom, P. M., Kooistra, L. and Witte, J.-P. M. (2013). Trait Estimation in Herbaceous Plant Assemblages from in situ Canopy Spectra. *Remote Sensing*, 5 (12), Multidisciplinary Digital Publishing Institute., pp.6323–6345.
- Ruiz-Benito, P., Vacchiano, G., Lines, E. R., Reyer, C. P. O., Ratcliffe, S., Morin, X., Hartig, F., Makela, A., Yousefpour, R., Chaves, J. E., et al. (2020). Available and missing data to model impact of climate change on European forests. *Ecological Modelling*, 416, Amsterdam : Elsevier., p.108870.
- Sandel, B., Gutiérrez, A. G., Reich, P. B., Schrodt, F., Dickie, J. and Kattge, J. (2015). Estimating the missing species bias in plant trait measurements. *Journal of Vegetation Science*, 26 (5), pp.828–838.
- Serbin, S. P. (2019). *Scaling Functional Traits From Leaves to Canopies*. Springer.
- Serbin, S. P., Wu, J., Ely, K. S., Kruger, E. L., Townsend, P. A., Meng, R., Wolfe, B. T., Chlus, A., Wang, Z. and Rogers, A. (2019). From the Arctic to the tropics: multibiome prediction of leaf mass per area using leaf reflectance. *New Phytologist*, 224 (4), pp.1557–1568.
- Seyednasrollah, B. and Clark, J. S. (2020). Where Resource-Acquisitive Species Are Located: The Role of Habitat Heterogeneity. *Geophysical Research Letters*, 47 (8), p.e2020GL087626.
- Shiklomanov, A. N., Cowdery, E. M., Bahn, M., Byun, C., Jansen, S., Kramer, K., Minden, V., Niinemets, Ü., Onoda, Y., Soudzilovskaia, N. A., et al. (2020). Does the leaf economic spectrum hold within plant functional types? A Bayesian multivariate trait meta-analysis. *Ecological Applications*, 30 (3), p.e02064.
- Souza, M. L., Duarte, A. A., Lovato, M. B., Fagundes, M., Valladares, F. and Lemos-Filho, J. P. (2018). Climatic factors shaping intraspecific leaf trait variation of a neotropical tree along a rainfall gradient. *Plos One*, 13 (12), p.e0208512.
- Vile, D., Garnier, É., Shipley, B., Laurent, G., Navas, M.-L., Roumet, C., Lavorel, S., Díaz, S., Hodgson, J. G., Lloret, F., et al. (2005). Specific Leaf Area and Dry Matter Content Estimate Thickness in Laminar Leaves. *Annals of Botany*, 96 (6), pp.1129–1136.
- Wallis, C. I. B., Homeier, J., Peña, J., Brandl, R., Farwig, N. and Bendix, J. (2019). Modeling tropical montane forest biomass, productivity and canopy traits with multispectral remote sensing data. *Remote Sensing of Environment*, 225, pp.77–92.
- Wellstein, C., Poschlod, P., Gohlke, A., Chelli, S., Campetella, G., Rosbakh, S., Canullo, R., Kreyling, J., Jentsch, A. and Beierkuhnlein, C. (2017). Effects of extreme drought on specific leaf area of grassland species: A meta-analysis of experimental studies in temperate and sub-Mediterranean systems. *Global Change Biology*, 23 (6), pp.2473–2481.

Woodward, F. I., Lomas, M. R. and Kelly, C. K. (2004). Global climate and the distribution of plant biomes. Pennington, P. T., Cronk, Q. C. B. and Richardson, J. A. (Eds). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359 (1450), pp.1465–1476.

Worthy, S. J. and Swenson, N. G. (2019). Functional perspectives on tropical tree demography and forest dynamics. *Ecological Processes*, 8 (1), p.1.

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., et al. (2004). The worldwide leaf economics spectrum. *Nature*, 428 (6985), pp.821–827.

Yang, Y., Wang, H., Harrison, S. P., Prentice, I. C., Wright, I. J., Peng, C. and Lin, G. (2019). Quantifying leaf-trait covariation and its controls across climates and biomes. *New Phytologist*, 221 (1), pp.155–168.

Zeballos, S. R., Giorgis, M. A., Cabido, M. and Gurvich, D. E. (2017). Unravelling the coordination between leaf and stem economics spectra through local and global scale approaches. *Austral Ecology*, 42 (4), pp.394–403.

Zhang, B., Lu, X., Jiang, J., DeAngelis, D. L., Fu, Z. and Zhang, J. (2017). Similarity of plant functional traits and aggregation pattern in a subtropical forest. *Ecology and Evolution*, 7 (12), pp.4086–4098.

7 Appendices

7.1 Appendix A: additional figures

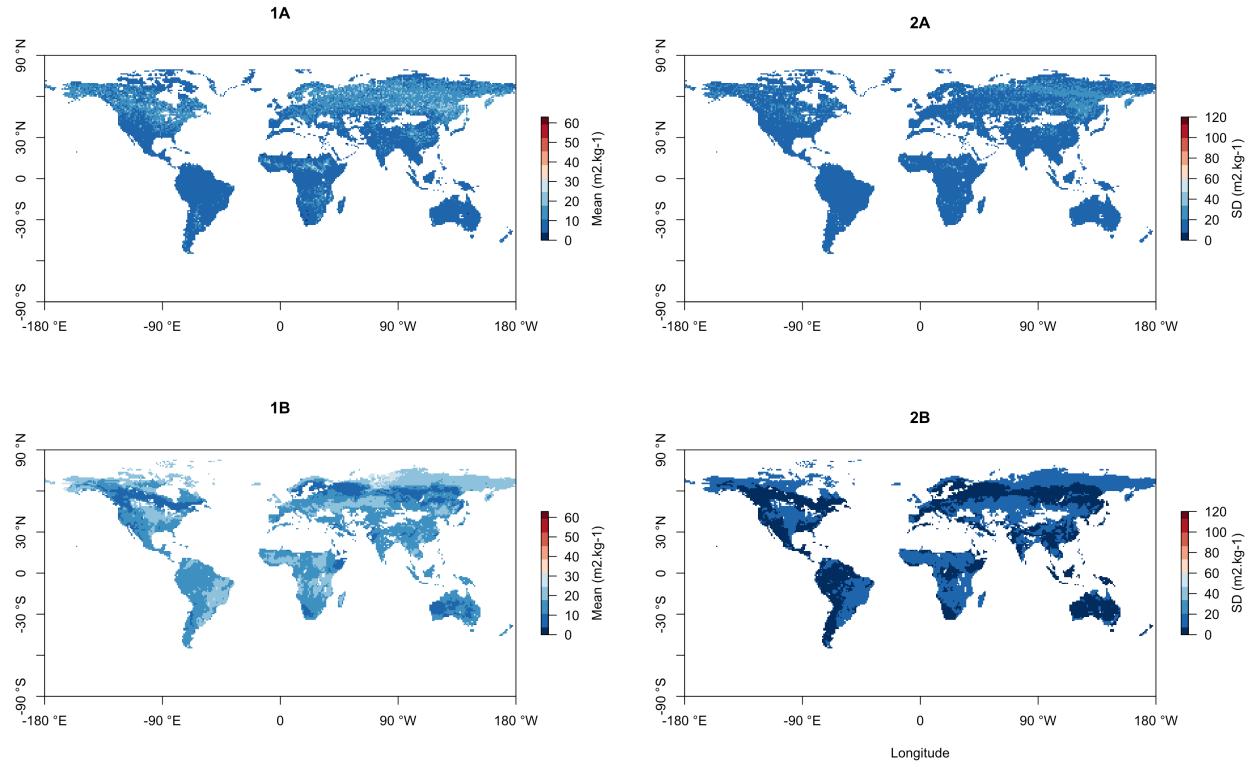


Figure 8: Global patterns in SLA mean and SD values. 1A/2A: SLA mean and SD values from top-down approach; 1B/2B: SLA mean and SD values from bottom-up approach.

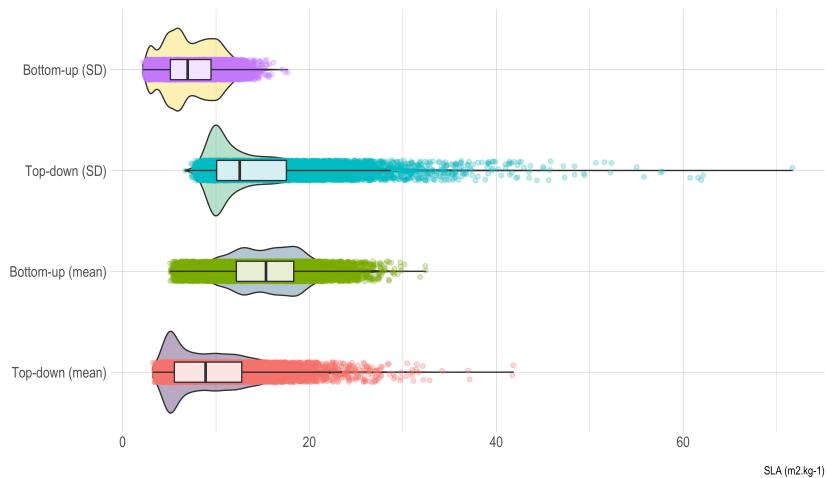


Figure 9: Data distributions for SLA mean and SD values from top-down and bottom-up.

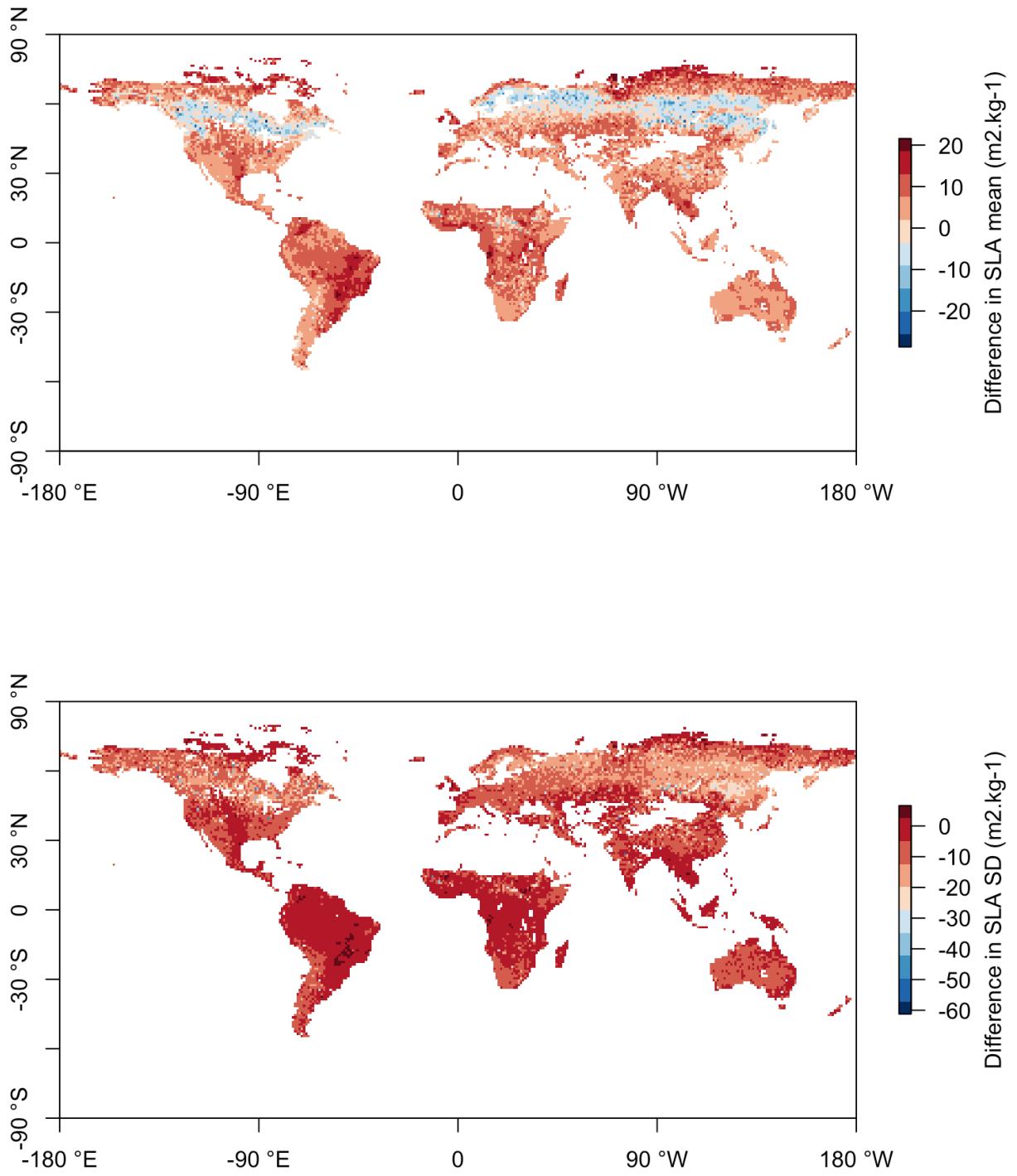


Figure 10: Calculated pixel-by-pixel difference between bottom-up and top-down, for mean and SD values.

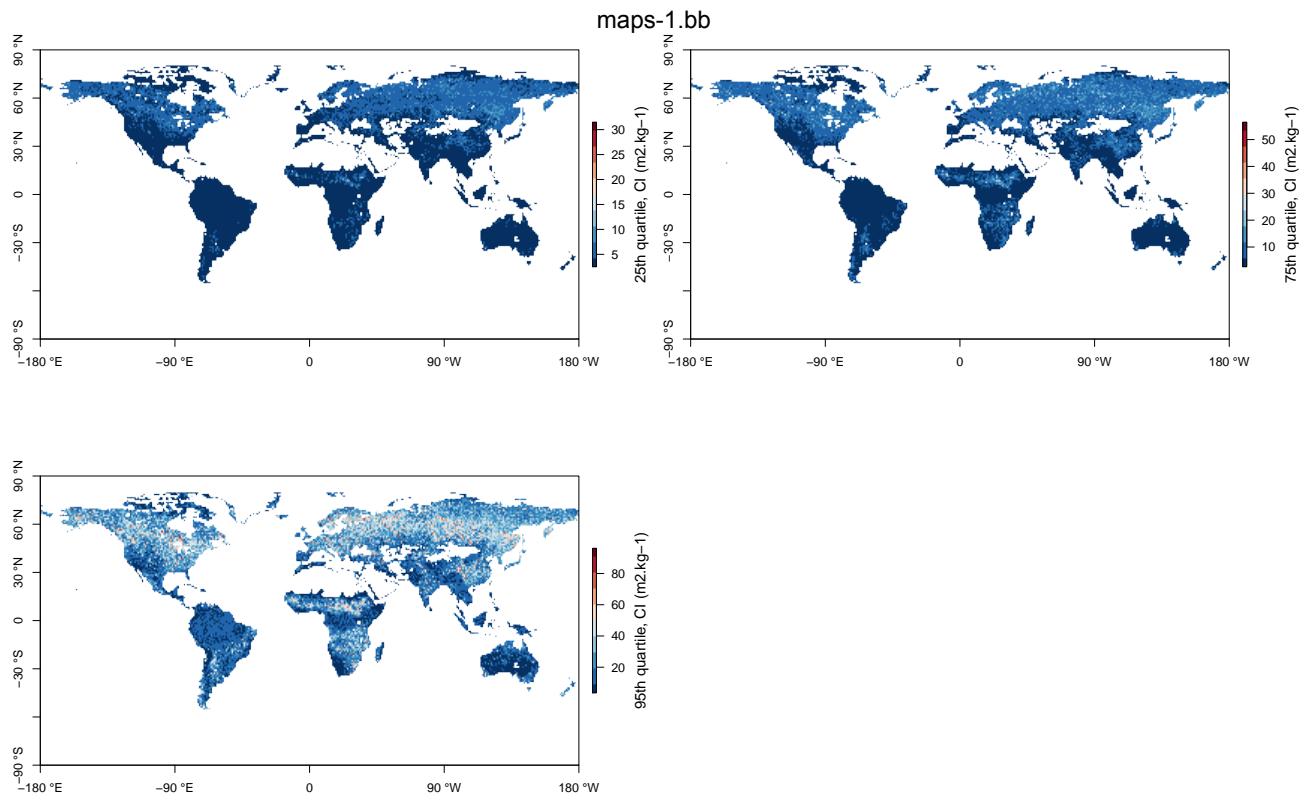


Figure 11: Global patterns of 25th, 75th and 95th CI values from the top-down approach (Bloom *et al.*, 2016).

7.2 Appendix B: additional tables

This appendix contains the exhaustive list of statistical results, which have been omitted in the main text (i.e. including values that have a p-value < 0.001). In bold: significant correlations ($p < 0.001$).

Scale	R^2 (mean)	R^2 (stdev)	RMSE (mean)	RMSE (stdev)	Bias (mean)	Bias (mean)
Global	0.0003	0.002	8.33	9.66	5.47	-7.30
Latitude						
Tropics	0.13	0.04	9.46	4.65	8.74	-3.13
Subtropics	0.14	0.03	7.34	6.69	6.47	-5.52
Temperate	0.03	0.00002	7.12	12.27	2.27	-10.37
Pole	0.02	0.006	10.76	9.54	9.61	-7.01
Biomes						
BorealForests/Taiga	0.02	0.03	7.12	14.56	-0.37	-13.45
Deserts& Xeric Shrublands	0.32	0.08	7.39	6.19	6.96	-5.16
FloodedGrasslands &	0.07	0.03	9.55	8.67	8.17	-4.86
Savannas						
Mangroves	0.34	0.98	12.24	2.36	12.23	-1.76
MediterraneanForests,	0.06	0.01	7.78	5.90	7.08	-4.85
Woodlands & Scrub						
MontaneGrasslands &	0.19	0.10	6.69	9.16	5.69	-7.17
Shrublands						
TemperateBroadleaf & Mixed	0.001	0.006	6.10	10.17	3.38	-8.81
Forests						
TemperateConifer Forests	0.04	0.02	6.54	13.74	-0.82	-11.60
TemperateGrasslands,	0.10	0.06	7.90	8.56	6.645	-6.38
Savannas & Shrublands						
Tropical& Subtropical	0.004	0.002	6.69	5.28	6.22	-4.95
Coniferous Forests						
Tropical& Subtropical Dry	0.0005	0.002	9.98	5.04	9.43	-2.82
Broadleaf Forests						
Tropical& Subtropical	0.05	0.009	9.88	5.43	8.78	-2.99
Grasslands, Savannas &						
Shrublands						
Tropical& Subtropical Moist	0.04	0.002	9.12	3.86	8.67	-3.32
Broadleaf Forests						
Tundra	0.01	0.004	10.48	9.64	9.19	-7.07

Exhaustive table of biomes by continent

Biomes by continent	R^2 (mean)	R^2 (stdev)	RMSE (mean)	RMSE (stdev)	Bias (mean)	Bias (sdtev)
Africa						
Deserts & Xeric Shrublands	0.38	0.01	7.24	6.26	6.95	-5.41
Flooded Grasslands &	0.13	0.02	10.68	4.03	9.95	-1.53
Savannas						
Mediterranean Forests,	0.13	0.09	7.42	5.48	7.31	-5.07
Woodlands & Scrub						

Biomes by continent	R^2 (mean)	R^2 (stdev)	RMSE (mean)	RMSE (stdev)	Bias (mean)	Bias (sdtev)
Montane Grasslands & Shrublands	0.16	0.008	7.34	5.79	6.57	-4.62
Tropical & Subtropical Dry Broadleaf Forests	0.03	0.04	10.17	1.84	10.00	-0.78
Tropical & Subtropical Grasslands, Savannas & Shrublands	0.12	0.03	9.17	5.80	8.08	-3.35
Tropical & Subtropical Moist Broadleaf Forests	0.23	0.12	9.50	3.20	9.25	-2.88
Australia						
Deserts & Xeric Shrublands	0.0002	0.001	6.27	6.24	6.08	-6.07
Mediterranean Forests, Woodlands & Scrub	0.12	0.11	7.52	7.13	6.52	-5.82
Montane Grasslands & Shrublands	0.76	0.64	8.00	4.76	7.12	-4.40
Temperate Broadleaf & Mixed Forests	0.04	0.15	7.52	4.18	7.17	-3.85
Temperate Grasslands, Savannas & Shrublands	0.52	0.06	6.96	5.63	6.78	-5.27
Tropical & Subtropical Grasslands, Savannas & Shrublands	0.12	0.002	7.76	5.66	7.40	-4.63
Tropical & Subtropical Moist Broadleaf Forests	0.20	0.16	7.17	4.26	7.12	-4.24
Eurasia						
Boreal Forests/Taiga	0.06	0.02	7.36	14.64	-0.34	-13.69
Deserts & Xeric Shrublands	0.18	0.09	8.15	6.23	7.58	-4.53
Flooded Grasslands & Savannas	0.23	0.003	5.50	15.13	3.09	-13.00
Mangroves	0.35	0.98	12.24	2.36	12.23	-1.76
Mediterranean Forests, Woodlands & Scrub	0.01	0.002	8.04	5.37	7.24	-4.28
Montane Grasslands & Shrublands	0.12	0.06	6.45	10.11	5.28	-7.93
Temperate Broadleaf & Mixed Forests	0.0003	0.01	6.02	10.09	3.65	-8.76
Temperate Conifer Forests	0.12	0.08	6.91	15.81	-0.83	-13.50
Temperate Grasslands, Savannas & Shrublands	0.02	0.008	7.77	9.97	5.99	-7.61
Tropical & Subtropical Coniferous Forests	0.17	0.04	7.03	5.67	5.96	-5.04
Tropical & Subtropical Dry Broadleaf Forests	0.01	0.001	9.58	6.62	9.00	-3.32
Tropical & Subtropical Grasslands, Savannas & Shrublands	0.10	0.68	8.98	5.02	8.35	-3.85
Tropical & Subtropical Moist Broadleaf Forests	0.08	0.007	8.50	4.52	7.91	-3.87

Biomes by continent	R^2 (mean)	R^2 (stdev)	RMSE (mean)	RMSE (stdev)	Bias (mean)	Bias (sdtev)
Tundra	0.007	0.003	10.67	10.47	9.20	-7.84
North America						
Boreal Forests/Taiga	0.0002	0.05	6.53	14.37	-0.45	-12.91
Deserts & Xeric Shrublands	0.22	0.04	7.47	6.33	7.01	-5.01
Mediterranean Forests, Woodlands & Scrub	0.06	0.02	8.87	3.45	8.70	-3.00
Temperate Broadleaf & Mixed Forests	0.07	0.003	6.25	11.75	1.57	-10.50
Temperate Conifer Forests	0.11	0.04	6.21	11.60	-0.84	-9.93
Temperate Grasslands, Savannas & Shrublands	0.18	0.13	8.27	7.32	7.45	-5.24
Tropical & Subtropical Grasslands, Savannas & Shrublands	0.22	0.34	10.32	3.32	10.12	-2.39
Tundra	0.004	0.0004	10.23	8.50	9.18	-6.09
South America						
Deserts & Xeric Shrublands	0.28	0.002	7.57	5.34	7.25	-4.62
Flooded Grasslands & Savannas	0.02	0.03	10.97	2.35	10.76	-1.49
Mediterranean Forests, Woodlands & Scrub	0.82	0.008	6.31	5.50	6.14	-5.35
Montane Grasslands & Shrublands	0.10	0.02	7.15	6.12	6.93	-5.61
Temperate Broadleaf & Mixed Forests	0.01	0.0001	4.58	7.09	2.77	-6.69
Temperate Grasslands, Savannas & Shrublands	0.16	0.008	7.81	6.35	7.11	-4.96
Tropical & Subtropical Coniferous Forests	0.01	0.00004	6.66	5.16	6.31	-4.89
Tropical & Subtropical Dry Broadleaf Forests	0.002	0.01	10.25	3.74	9.76	-2.63
Tropical & Subtropical Grasslands, Savannas & Shrublands	0.07	0.12	12.69	4.01	11.78	-0.99
Tropical & Subtropical Moist Broadleaf Forests	0.006	0.02	9.44	3.58	9.00	-3.09

7.3 Appendix C: *dissertation code*

For the complete script go to: <https://github.com/AnnaChirumbolo/Dissertation>

Libraries

```
## libraries ####
# spatial analysis
library(ncdf4)
library(raster)
library(sp)
library(sf)
library(cleangeo)
library(tiff)
library(rworldmap)
# data manipulation
library(tidyverse)
library(reshape2)
library(data.table)
# statistics, percentage overlap plots, heatscatters
library(Metrics)
library(overlapping)
library(LSD)
# for plotting
library(MASS)
library(gplots)
library(ggplot2)
library(lattice)
library(ggExtra)
library(ggpubr)
library(rasterVis)
library(grid)
# colour palettes
library(RColorBrewer)
library(viridis)
library(hrbrthemes)
library(colorspace)
library(dichromat)
# latex format tables
library(Hmisc)
library(formattable)
library(kableExtra)
library(stargazer)
```

Custom functions

```

## custom functions ####

# convert raster to data frame
mask.to.df <- function(x){
  new.list <- list()
  for (i in 1:length(x)){
    df <- raster::as.data.frame(x[[i]], xy = TRUE)
    name.df <- paste("df",names(x)[i],sep = ".")
    new.list[[name.df]] <- df
  }
  new.list
}

# joining dataframes
join.f <- function(x,y){
  new.list <- list()
  join <- mapply(left_join, x, y,SIMPLIFY = FALSE)
  join <- lapply(join, na.omit)
  name.df <- names(x)[i]
  new.list[[name.df]] <- join
}

# masking the biomes by continent
mask.biome.f <- function(x,y){
  new.list <- list()
  for (i in 1:length(x)){
    for (j in 1:length(y)){
      mask.biome <- raster::mask(x[[i]],y[[j]])
      if (!is.infinite(mask.biome@data@min)&!is.infinite(mask.biome@data@max)){
        name <- paste(names(x)[i],names(y)[j],sep = " ")
        new.list[[name]] <- mask.biome
      }
      else
        NULL
    }
  }
  new.list
}

# density plots with percentage overlap (modif. from overlapping package)
# for SLA mean
my.final.plot.sla <- function (x, OV = NULL){
  AREA <- NULL
  for (i1 in 1:(length(x) - 1)) {
    for (i2 in (i1 + 1):(length(x))) {
      A <- data.frame(x = x[[i1]], group = names(x)[i1],
                      k = paste(names(x)[i1], names(x)[i2], sep = "-",
                                collapse = ""))
      B <- data.frame(x = x[[i2]], group = names(x)[i2],

```

```

            k = paste(names(x)[i1], names(x)[i2], sep = "-",
                      collapse = ""))
        AREA <- rbind(AREA, rbind(A, B))
    }
}
if (!is.null(OV)) {
    OV <- data.frame(OV = OV, k = names(OV))
    AREA <- merge(AREA, OV, by = "k")
    AREA$k <- paste0(AREA$k, " (ov. perc. ", round(AREA$OV *
                                              100), ")")
}
ggplot(AREA, aes(x = x)) + facet_wrap(~k) +
    geom_density(aes(fill = AREA$group), alpha = 0.35) +
    xlab("\nSLA mean (m2.kg-1)") +
    ylab("") +
    theme_classic() +
    theme(legend.title = element_blank()) +
    scale_color_brewer(palette = "Set1") +
    scale_x_continuous(expand = c(0,0)) +
    scale_y_continuous(expand = c(0,0))
}

my.overlap.sla <- function (x, nbins = 1024, plot = FALSE, partial.plot = FALSE,
                           boundaries = NULL, ...){
  if (is.null(names(x)))
    names(x) <- paste("Y", 1:length(x), sep = "")
  dd <- OV <- FUNC <- DD <- xpoints <- COMPTITLE <- NULL
  for (j in 1:length(x)) {
    if (!is.null(boundaries)) {
      Lbound <- lapply(boundaries, FUN = length)
      if ((Lbound$from == 1) & (Lbound$to == 1)) {
        warning("Boundaries were set all equals")
        boundaries$from <- rep(boundaries$from, length(x))
        boundaries$to <- rep(boundaries$to, length(x))
      }
      else {
        if ((Lbound$from != length(x)) | (Lbound$to !=
                                              length(x))) {
          stop("Boundaries not correctly defined")
        }
      }
      from = boundaries$from[j]
      to = boundaries$to[j]
      dj <- density(x[[j]], n = nbins, from = from, to = to,
                     ...)
    }
    else {
      dj <- density(x[[j]], n = nbins, ...)
    }
  }
}
```

```

ddd <- data.frame(x = dj$x, y = dj$y, j = names(x)[j])
FUNC <- c(FUNC, list(with(ddd, approxfun(x, y))))
dd <- rbind(dd, ddd)
}
for (i1 in 1:(length(x) - 1)) {
  for (i2 in (i1 + 1):(length(x))) {
    comptitle <- paste0(names(x)[i1], "-", names(x)[i2])
    dd2 <- data.frame(x = dd$x, y1 = FUNC[[i1]](dd$x),
                       y2 = FUNC[[i2]](dd$x))
    dd2[is.na(dd2)] <- 0
    dd2$ovy <- apply(dd2[, c("y1", "y2")], 1, min)
    dd2$ally <- apply(dd2[, c("y1", "y2")], 1, max, na.rm = TRUE)
    dd2$dominance <- ifelse(dd2$y1 > dd2$y2, 1, 2)
    dd2$k <- comptitle
    OV <- c(OV, sum(dd2$ovy, na.rm = TRUE)/sum(dd2$ally,
                                                 na.rm = TRUE))
    dd2 <- dd2[order(dd2$x), ]
    CHANGE <- dd2$x[which(dd2$dominance[2:nrow(dd2)] != dd2$dominance[1:(nrow(dd2) - 1)])]
    xpoints <- c(xpoints, list(CHANGE))
    if (partial.plot) {
      gg <- ggplot(dd2, aes(x, dd2$y1)) + theme_bw() +
        geom_vline(xintercept = CHANGE, lty = 2, color = "#cccccc") +
        geom_line() + geom_line(aes(x, dd2$y2)) +
        geom_line(aes(x, dd2$ovy), color = "red") +
        geom_line(aes(x, dd2$ally), color = "blue") +
        ggtitle(comptitle) +
        xlab("") +
        ylab("") +
        theme(plot.title = element_text(hjust = 0.5),
              legend.title = element_blank())
      print(gg)
    }
    DD <- rbind(DD, dd2)
    COMPTITLE <- c(COMPTITLE, comptitle)
  }
}
names(xpoints) <- names(OV) <- COMPTITLE
if (plot)
  print(my.final.plot.sla(x, OV))
return(list(DD = DD, OV = OV, xpoints = xpoints))
}

# for SLA standard deviation
my.final.plot.std <- function (x, OV = NULL){
  AREA <- NULL
  for (i1 in 1:(length(x) - 1)) {
    for (i2 in (i1 + 1):(length(x))) {
      A <- data.frame(x = x[[i1]], group = names(x)[i1],

```

```

        k = paste(names(x)[i1], names(x)[i2], sep = "-",
                   collapse = ""))
B <- data.frame(x = x[[i2]], group = names(x)[i2],
                 k = paste(names(x)[i1], names(x)[i2], sep = "-",
                           collapse = ""))
AREA <- rbind(AREA, rbind(A, B))
}
}
if (!is.null(OV)) {
  OV <- data.frame(OV = OV, k = names(OV))
  AREA <- merge(AREA, OV, by = "k")
  AREA$k <- paste0(AREA$k, " (ov. perc. ", round(AREA$OV *
                                                 100), ")")
}
ggplot(AREA, aes(x = x)) + facet_wrap(~k) +
  geom_density(aes(fill = AREA$group), alpha = 0.35) +
  xlab("\nSLA StDev (m2.kg-1)") +
  ylab("") +
  theme_classic() +
  theme(legend.title = element_blank()) +
  scale_color_brewer(palette = "Set1") +
  scale_x_continuous(expand = c(0,0)) +
  scale_y_continuous(expand = c(0,0))
}
my.overlap.std <- function (x, nbins = 1024, plot = FALSE, partial.plot = FALSE,
                             boundaries = NULL, ...){
  if (is.null(names(x)))
    names(x) <- paste("Y", 1:length(x), sep = "")
  dd <- OV <- FUNC <- DD <- xpoints <- COMPTITLE <- NULL
  for (j in 1:length(x)) {
    if (!is.null(boundaries)) {
      Lbound <- lapply(boundaries, FUN = length)
      if ((Lbound$from == 1) & (Lbound$to == 1)) {
        warning("Boundaries were set all equals")
        boundaries$from <- rep(boundaries$from, length(x))
        boundaries$to <- rep(boundaries$to, length(x))
      }
      else {
        if ((Lbound$from != length(x)) | (Lbound$to !=
                                              length(x))) {
          stop("Boundaries not correctly defined")
        }
      }
      from = boundaries$from[j]
      to = boundaries$to[j]
      dj <- density(x[[j]], n = nbins, from = from, to = to,
                     ...)
    }
    else {

```

```

    dj <- density(x[[j]], n = nbins, ...)
}
ddd <- data.frame(x = dj$x, y = dj$y, j = names(x)[j])
FUNC <- c(FUNC, list(with(ddd, approxfun(x, y))))
dd <- rbind(dd, ddd)
}
for (i1 in 1:(length(x) - 1)) {
  for (i2 in (i1 + 1):(length(x))) {
    comptitle <- paste0(names(x)[i1], "-", names(x)[i2])
    dd2 <- data.frame(x = dd$x, y1 = FUNC[[i1]](dd$x),
                       y2 = FUNC[[i2]](dd$x))
    dd2[is.na(dd2)] <- 0
    dd2$ovy <- apply(dd2[, c("y1", "y2")], 1, min)
    dd2$ally <- apply(dd2[, c("y1", "y2")], 1, max, na.rm = TRUE)
    dd2$dominance <- ifelse(dd2$y1 > dd2$y2, 1, 2)
    dd2$k <- comptitle
    OV <- c(OV, sum(dd2$ovy, na.rm = TRUE)/sum(dd2$ally,
                                                 na.rm = TRUE))
    dd2 <- dd2[order(dd2$x), ]
    CHANGE <- dd2$x[which(dd2$dominance[2:nrow(dd2)] != 
                           dd2$dominance[1:(nrow(dd2) - 1)])]
    xpoints <- c(xpoints, list(CHANGE))
    if (partial.plot) {
      gg <- ggplot(dd2, aes(x, dd2$y1)) + theme_bw() +
        geom_vline(xintercept = CHANGE, lty = 2, color = "#cccccc") +
        geom_line() + geom_line(aes(x, dd2$y2)) +
        geom_line(aes(x, dd2$ovy), color = "red") +
        geom_line(aes(x, dd2$ally), color = "blue") +
        ggtitle(comptitle) +
        xlab("") +
        ylab("") +
        theme(plot.title = element_text(hjust = 0.5),
              legend.title = element_blank())
      print(gg)
    }
    DD <- rbind(DD, dd2)
    COMPTITLE <- c(COMPTITLE, comptitle)
  }
}
names(xpoints) <- names(OV) <- COMPTITLE
if (plot)
  print(my.final.plot.std(x, OV))
return(list(DD = DD, OV = OV, xpoints = xpoints))
}

## extracting p-values from lm summary (credits to Stephen Turner):
#gettinggeneticsdone.blogspot.com/2011/01/
#rstats-function-for-extracting-f-test-p.html
lmp <- function (modelobject) {

```

```

if (class(modelobject) != "lm") stop("Not an object of class 'lm' ")
if (modelobject[["df.residual"]]!=0){ # if the residual is not = 0
  f <- summary(modelobject)$fstatistic
  p <- pf(f[1],f[2],f[3],lower.tail=F)
  attributes(p) <- NULL
  return(p)
}
else
  modelobject <- NULL
}

# perform calculations of statistics iteratively
# linear regression
lm.f <- function(x){
  lm <- lapply(x, function(dat) lm(dat[,4] ~ dat[,3],data=dat))
}

# step-by-step calculation of r-squared, and of RMSE and bias
stats.f <- function(x){
  lapply(x,function(df){
    df$c_mean <- mean(df$cardamom)
    df$b_mean <- mean(df$butler)
    df$diff_butler <- df$butler-df$b_mean
    df$diff_butler2<- df$diff_butler^2
    df$sum_diff_butler2 <-sum(df$diff_butler2)
    df$slope_bf <-sum((df$cardamom-df$c_mean)*(df$butler-df$b_mean))/
      sum((df$cardamom-df$c_mean)^2)
    df$b_intercept <- df$b_mean - (df$slope_bf*df$c_mean)
    df$new_b_val <- df$b_intercept + (df$slope_bf*df$cardamom)
    df$dist_mean_new_b <- df$new_b_val - df$b_mean
    df$sqrddist_b <- df$dist_mean_new_b^2
    df$sum_sqrddist_b <- sum(df$sqrddist_b)
    df$sla_r2 <- df$sum_sqrddist_b / df$sum_diff_butler2
    df$bias_av <- bias(df$butler, df$cardamom)
    df$bias_row <- df$butler - df$cardamom
    df$rmse_av <- rmse(df$butler, df$cardamom)
    df$rmse_row <- sqrt((df$butler - df$cardamom)^2)
    df
  })
}

# function for kableExtra, from Michael Harper (not modified):
#stackoverflow.com/questions/28166168/
#how-to-change-fontface-bold-italics-for-a-cell-in-a-kable-table-in-rmarkdown
format_cells <- function(df, rows ,cols, value = c("italics", "bold", "strikethrough")){

  # select the correct markup
  map <- setNames(c("*", "*", "~"), c("italics", "bold", "strikethrough"))
  markup <- map[value]
}

```

```

for (r in rows){
  for(c in cols){

    # Make sure values are not factors
    df[[c]] <- as.character( df[[c]])

    # Update formatting
    df[r, c] <- paste0(markup, df[r, c], markup)
  }
}

return(df)
}

# function for setting color gradients in bivariate legend
#devtools::install_github("wmurphyrd/colorplaner")
#library(colorplaner)
col_func <- function(x, y){
  x[x == 0] <- 0.000001
  y[y == 0] <- 0.000001
  x[x == 1] <- 0.999999
  y[y == 1] <- 0.999999
  # upper or lower triangle?
  u <- y > x
  # Change me for different hues.
  hue <- ifelse(u, 0.3, 0.8)
  # distance from (0,0) to (x,y)
  hyp <- sqrt(x^2 + y^2)
  # Angle between x axis and line to our point
  theta <- asin(y / hyp)
  # Angle between 45 degree line and (x,y)
  phi <- ifelse(u, theta - pi/4, pi/4 - theta)
  phi <- ifelse(phi < 0, 0, phi)
  # Distance from 45 degree line and (x,y)
  s <- hyp * sin(phi) / sqrt(2)
  # Draw line from (x, y) to 45 degree line that is at right angles.
  # How far along 45 degree line, does that line join.
  v <- 1 - hyp * cos(phi) / sqrt(2)
  # Get hsv values.
  sapply(seq_along(x), function(i) hsv(hue[i], s[i], v[i]))
}

```

Import data

```

# cardamom = top-down approach
# butler = bottom-up approach

# Bloom _et al._ (2015)

```

```

cardamom_lcma <- raster("./DATA/CARDAMOM_2001_2010_LCMA_zeros.nc",
                         varname="lcma")
cardamom_sla <- raster("./DATA/CARDAMOM_2001_2010_LCMA_zeros.nc",
                         varname="sla")
pc75 <- raster("./DATA/CARDAMOM_2001_2010_LCMA_zeros.nc",
                 varname = "75th_percentile")
pc25 <- raster("./DATA/CARDAMOM_2001_2010_LCMA_zeros.nc",
                 varname = "25th_percentile")
pc95 <- raster("./DATA/CARDAMOM_2001_2010_LCMA_zeros.nc",
                 varname = "95th_percentile")
cardamom_sla_std <- raster("./DATA/CARDAMOM_2001_2010_LCMA_zeros.nc",
                            varname= "sla_std")

# Butler _et al._ (2017)
butler_sla <- raster("./DATA/Butler_Leaftraits_Processed_1x1_zeros.nc",
                      varname="sla")
butler_sla_std <- raster("./DATA/Butler_Leaftraits_Processed_1x1_zeros.nc",
                         varname="sla_std")

# Ecoregions17
ecoregions17 <- st_read("./DATA/Ecoregions2017/Ecoregions2017.shp")

# WorldClim data
temp <- raster("./DATA/wc2.1_10m_bio/wc2.1_10m_bio_1.tif")
ppt <- raster("./DATA/wc2.1_10m_bio/wc2.1_10m_bio_12.tif")

```

Data manipulation

```

# Conversions of Confidence Intervals values from LCMA to SLA
lma_75th <- pc75*2 # LCMA to LMA
sla_cardamom_75th <- 1/lma_75th # LMA to SLA
sla_cardamom_75th <- sla_cardamom_75th*1000 # changing units
cardamom_75th <- sla_cardamom_75th

lma_25th <- pc25*2
sla_c_25th <- 1/lma_25th
sla_c_25th <- sla_c_25th*1000
cardamom_95th <- sla_c_25th

lma_95th <- pc95*2
sla_c_95th <- 1/lma_95th
sla_c_95th <-sla_c_95th*1000
cardamom_25th <- sla_c_95th

# removing outlier from top-down
cardamom_sla[cardamom_sla>42] <- NA
cardamom_sla_std[cardamom_sla_std>90]<- NA

```

```

# masking bottom-up so as to match pixels with top-down
butler_sla <- raster::mask(butler_sla, cardamom_sla)
butler_sla_std <- raster::mask(butler_sla_std, cardamom_sla_std)

# making .nc files into data frames
cardamom_sla_df <- raster::as.data.frame(cardamom_sla, xy = TRUE)
cardamom_sla_std_df <- raster::as.data.frame(cardamom_sla_std, xy=TRUE)
butler_sla_df <- raster::as.data.frame(butler_sla, xy = TRUE)
butler_sla_std_df <- raster::as.data.frame(butler_sla_std, xy=TRUE)

# basic data manipulation
cardamom_sla_df <- cardamom_sla_df %>%
  rename("cardamom" = specific.leaf.area)
cardamom_sla_std_df <- cardamom_sla_std_df %>%
  rename("cardamom_std" = sla_std)
butler_sla_df <- butler_sla_df %>%
  rename("butler" = specific.leaf.area)
butler_sla_std_df <- butler_sla_std_df %>%
  rename("butler_std" = specific.leaf.area)
joined_sla <- left_join(cardamom_sla_df, butler_sla_df)
joined_sla_std <- left_join(cardamom_sla_std_df, butler_sla_std_df)

# manipulation of climate data
# checking resolution
res(temp) # res as [1] 0.1666667 0.1666667
# changing res to match that of cardamom and butler (1,1)
temp.1 <- aggregate(temp, fact = 1/res(temp))
res(temp.1) # now is 1,1
res(ppt) # res as [1] 0.1666667 0.1666667
# changing res to match that of cardamom and butler (1,1)
ppt.1 <- aggregate(ppt, fact = 1/res(ppt))
res(ppt.1) # now is 1,1

# turn rasters to data frames, renaming column of values and removing NAs
temp.df <- as.data.frame(temp.1, xy=T) %>%
  rename(temp.C = wc2.1_10m_bio_1) %>%
  filter(temp.C!=0)
ppt.df <- as.data.frame(ppt.1, xy=T) %>%
  rename(ppt.mm = wc2.1_10m_bio_12) %>%
  filter(ppt.mm!=0)

# joining dataframes with cardamom and butler
# joining ppt and temp
climate <- dplyr::left_join(ppt.df,temp.df)
# joining climate data frame with cardamom and butler
# sla mean
cl.sla <- left_join(climate,joined_sla) %>%
  filter(cardamom!=0, butler!=0) %>% # filtered out NAs when joined with clim data
  mutate(`Cardamom`=cardamom, `Butler`=butler) %>%

```

```

gather(key=dataset, value=sla.mean,-x,-y,-ppt.mm,-temp.C,-cardamom,-butler)
# sla stdev
cl.std <- left_join(climate,joined_sla_std) %>%
  filter(cardamom_std !=0, butler_std!=0) %>%# filtered out NAs when joined with clim
  mutate(`Cardamom`=cardamom_std, `Butler`=butler_std) %>%
  gather(key=dataset, value=sla.stdev,-x,-y,-ppt.mm,-temp.C,
         -cardamom_std, -butler_std)

```

Stippling

```

# create the breaks- and label vectors for custom lat and lon axes labels
ewbrks <- seq(-180,180,90)
nsbrks <- seq(-90,90,30)
ewlbls <- unlist(lapply(ewbrks, function(x) ifelse(x < 0, paste(x, "°E"),
                                                    ifelse(x > 0,
                                                       paste(x, "°W"),
                                                       x))))
nslbls <- unlist(lapply(nsbrks, function(x) ifelse(x < 0, paste(x, "°S"),
                                                    ifelse(x > 0, paste(x,
                                                       "°N"),
                                                       x))))
# stippling: finding out which values from bottom-up lie
# within 25pc-75pc and 25pc-95pc of the top-down
gl_stip_locs_75 <- (butler_sla[[1]] >= cardamom_25th[[1]])*
  (butler_sla[[1]] <= cardamom_75th[[1]])
gl_stip_locs_75 <- rasterToPoints(gl_stip_locs_75)
gl_stip_locs_75 <- as.data.frame(gl_stip_locs_75[gl_stip_locs_75[, "layer"] == 1,])
gl_stip_locs_95 <- (butler_sla[[1]] >= cardamom_25th[[1]])*
  (butler_sla[[1]] <= cardamom_95th[[1]])
gl_stip_locs_95 <- rasterToPoints(gl_stip_locs_95)
gl_stip_locs_95 <- as.data.frame(gl_stip_locs_95[gl_stip_locs_95[, "layer"] == 1,])
# difference between 25pc-95pc and 25pc-75pc
diff_pc <- as.matrix(anti_join(gl_stip_locs_95,gl_stip_locs_75))

# plotting the stippling
plot(butler_sla, asp = NA, col = rev(brewer.pal(10, "RdYlBu")),
      xlab=" ", ylab=" ",
      legend.args = list(text="\n\nSLA mean (m2.kg-1)",
                         side=4, font=1, line=2.3),
      axes=F)
axis(1, at=ewbrks, labels = ewlbls)
axis(2, at=nsbrks, labels = nslbls)
points(gl_stip_locs, pch = 18, cex = 0.5, col="green", bg="darkgreen")
points(diff_pc, pch = 18, cex=0.5)

```

SLA sensitivity to climate

```
#### ppt corr #####
# mean
(density.mean <- ggplot(cl.sla,aes(x=ppt.mm,y=sla.mean,fill=dataset)) +
  stat_density2d(geom="tile", aes(fill = dataset, alpha=..density..),
                 contour=FALSE) +
  scale_y_continuous(expand = c(0,0))+ 
  scale_x_continuous(expand = c(0,0))+ 
  scale_fill_manual(values=c("#0072B2", "#D55E00"))+
  theme_minimal())
density.mean.data <- as.data.table(ggplot_build(density.mean)$data[1])
butler.density <- density.mean.data[group==1] %>%
  dplyr::select(density)
cardamom.density <- density.mean.data[group==2] %>%
  dplyr::select(density)
uncommon.mean.dens <- anti_join(cardamom.density, butler.density)
perc.unc.mean <- nrow(uncommon.mean.dens)/nrow(density.mean.data)*100
# percentage uncommon density values 46.06
perc.ov <- 100-perc.unc.mean
# percentage overlap: 49.68%


#### temperature corr #####
# mean
(density.mean.temp <- ggplot(cl.sla,aes(x=temp.C,y=sla.mean, fill=dataset)) +
  stat_density2d(geom="tile", aes(fill=dataset,alpha=..density..),
                 contour=FALSE) +
  scale_y_continuous(expand = c(0,0))+ 
  scale_x_continuous(expand = c(0,0))+ 
  theme_minimal())
density.temp.mean <- as.data.table(ggplot_build(density.mean.temp)$data[1])
butler.temp.mean <- density.temp.mean[group ==1] %>%
  dplyr::select(density)
cardamom.temp.mean <- density.temp.mean[group==2] %>%
  dplyr::select(density)
uncommon.temp.mean <- anti_join(cardamom.temp.mean, butler.temp.mean)
#options(scipen = 999)
perc.unc.temp.mean<- nrow(uncommon.temp.mean)/nrow(density.temp.mean)*100
# percentage not overlap: 41.205
perc.ov.temp.mean <- 100- perc.unc.temp.mean
# percentage overlap 50%


# visualisation of spatial relationship between SLA-climate (one example):
## ppt - top-down (cardamom)
(cardamom.ppt.tile <- ggplot(data=cl.sla,aes(x,y,fill=ppt.mm,fill2=cardamom))+ 
  geom_tile()+
  scale_fill_colourplane(name = " ",
                         na.color = "white",
                         color_projection = col_func,
```

```

axis_title = "Total annual precipitation (mm)",
axis_title_y = "SLA mean (m2.kg-1)",
breaks = c(1000,3000,5000),
limits_y = c(0,42))+
ggtitle("Benchmark\n")+
scale_x_continuous(breaks = ewbrks, labels = ewlbls, expand = c(0, 0)) +
scale_y_continuous(breaks = nsbrks, labels = nslbls, expand = c(0, 0)) +
theme(panel.background = element_rect(fill = "white"),
      panel.grid = element_blank(),
      axis.text=element_text(size=11),
      axis.title=element_text(size=15),
      plot.title = element_text(face = "bold",
                                size=20))+
xlab(" ")+ylab(" ")
ggsave("./figures/global_analysis/cardamom.ppt.tile.png",cardamom.ppt.tile,
       width = 30,height = 15,units = "cm",dpi = 500)

```