Supplement of Biogeosciences, 11, 7025–7050, 2014
http://www.biogeosciences.net/11/7025/2014/
doi:10.5194/bg-11-7025-2014-supplement

Biogeosciences Open Access

Supplement of

Identifying environmental controls on vegetation greenness phenology through model–data integration

M. Forkel et al.

Correspondence to: M. Forkel (mforkel@bgc-jena.mpg.de)

Contents

Figures

Tables

1 LPJmL model details

1.1 Original phenology model (LPJmL-OP)

The phenology model in the original LPJmL formulation has three different routines for summergreen (i.e. temperature-driven deciduous), evergreen (no seasonal variation) and rain-green (i.e. water-driven deciduous) PFTs (Sitch et al., 2003). Evergreen PFTs have a constant phenology status (Phen = 1). The daily phenology status of summergreen PFTs depends on growing degree-days (GDD):

$$\Delta T = T - GDD_{base}$$
$$GDD_t = GDD_{t-1} + \Delta T_t \quad if \quad \Delta T > 0 \tag{S1}$$

Where T is the daily air temperature and GDD_{base} is the minimum temperature threshold to start counting GDDs. Daily GDD is scaled to the phenology status using a parameter ramp which is the amount of GDDs to get full leave cover:

$$Phen_{PFT|\text{summergreen}} = \begin{cases} GDD/ramp & if \quad aphen < aphen_{\max} \\ 0 & if \quad aphen \geq aphen_{\max} \\ 0 & if \quad aphen > aphen_{\min} \; and \; \Delta T < 0 \end{cases} \tag{S2}$$

The daily phenology status is set back to 0 if the accumulated phenology status (aphen) is larger than a parameter $aphen_{max}$ or if aphen is greater than $aphen_{min}$ and the daily temperature is below GDD_{base}. The daily accumulated phenology status is calculated as:

$$aphen_t = aphen_{t-1} + Phen_t \quad \text{(S3)}$$

For rain-green PFTs the daily phenology status is calculated dependent on the daily water availability scaling factor Wscal in LPJmL (Supplement A.2) (Gerten et al., 2004) and a threshold value ($Wscal_{min}$):

$$Phen_{PFT|\text{raingreen}} = \begin{cases} 1 & if\ Wscal \geq Wscal_{\min} \\ 0 & if\ Wscal < Wscal_{\min} \end{cases} \quad \text{(S4)}$$

The phenology of rain-green PFTs has no smooth behaviour but is a binary switch between full leave cover and no leaves according to this formulation. For herbaceous PFTs the same phenology scheme like for summergreen PFTs is used but the phenology status is only set back to 0 at the end of the phenology year (i.e. on the 14th day of the year for the northern hemisphere and on the 195th day of the year for the southern hemisphere).

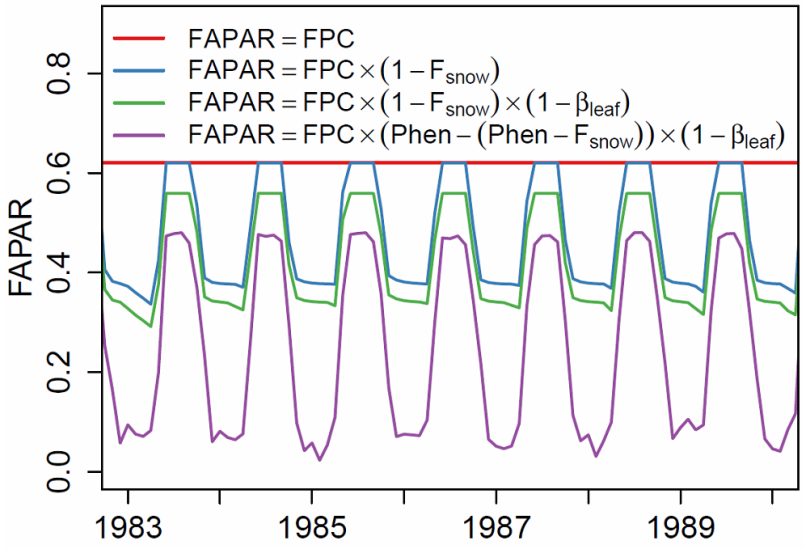

Figure S 1: Effects on FAPAR in LPJmL for an example grid cell in Siberia.
FAPAR in LPJmL is computed from foliar projective cover (FPC), from snow coverage in the green canopy (Fsnow), leaf albedo (βleaf) and phenology status (Phen).

1.2 Water availability scaling factor

The water availability scaling factor Wscal in LPJmL is a ratio between water supply S and atmospheric water demand D for a dry canopy (Gerten et al., 2004):

$$Wscal = \frac{S}{D} \quad \text{(S5)}$$

In the LPJmL-GSI phenology model the water availability scaling factor is expressed as a percentage value:

$$W = Wscal \times 100 \quad \text{(S6)}$$

Water supply is dependent on the maximum transpiration E_{max} under water saturation and relative soil moisture w_r (Gerten et al., 2004):

$$S = E_{\max} \times w_r \quad \text{(S7)}$$

Atmospheric water demand D for a dry canopy is calculated from potential evapotranspiration PET, maximum Priestley-Taylor coefficient α_{max} = 1.391, scaling canopy conductance g_m = 3.26 mm s-1 and potential canopy conductance g_{pot} (Gerten et al., 2004):

$$D = \frac{PET \times \alpha_{\max}}{1 + (g_m / g_{pot})} \quad \text{(S8)}$$

1.3 Albedo

Surface albedo and snow coverage routines have been implemented in LPJmL to use it as a land surface scheme in a coupled vegetation-climate model (Strengers et al., 2010). We used this implementation but made the albedo parameters PFT-dependent as albedo differs between ecosystems (Cescatti et al., 2012). The albedo of a grid cell Alb_{gc} is the area-weighted sum of the vegetation albedo Alb_{veg}, bare-soil albedo Alb_{bare} and snow albedo:

$$Alb_{gc} = Alb_{veg} + F_{bare} \times (F_{snow} \times \beta_{snow} + (1 - F_{snow}) \times \beta_{soil}) \quad \text{(S9)}$$

where F_{bare} and F_{snow} are the coverage of bare soil and snow on top of bare soil in a grid cell and β_{soil} and β_{snow} are the soil and snow albedo parameters, respectively. The parameters β_{soil} = 0.4 and β_{snow} = 0.7 were used as constants (Strengers et al., 2010) and not further considered in this study. Although soil and snow albedo has clear spatial and temporal variations which are due to changing moisture contents, an improvement of these processes is not within the scope of our study. The vegetation albedo is computed as the albedo of each PFT Alb_{PFT} and its corresponding FPC:

$$Alb_{veg} = \sum_{PFT=1}^{PFT=n} Alb_{PFT} \times FPC_{PFT} \quad \text{(S10)}$$

The albedo of a PFT depends on the fraction of the PFT that is completely covered by snow $F_{snow,PFT}$ and the albedo of the PFT without snow coverage ($Alb_{PFT,nosnow}$) (Strengers et al., 2010):

$$Alb_{PFT} = F_{snow,PFT} \times \beta_{snow} + (1 - F_{snow,PFT}) \times Alb_{PFT,nosnow} \quad \text{(S11)}$$

The albedo of a PFT without snow coverage is the sum of leaf, stem/branches and litter (background) albedo:

$$Alb_{PFT,nosnow} = Alb_{leaf,PFT} + Alb_{stem,PFT} + Alb_{litter,PFT} \quad \text{(S12)}$$

The albedo of green leaves depends on the foliar projective cover, the daily phenology status and the PFT-dependent leaf albedo parameter:

$$Alb_{leaf,PFT} = FPC_{PFT} \times Phen_{PFT} \times \beta_{leaf,PFT} \quad \text{(S13)}$$

The albedo of stems and branches depends on the fractional coverage of the ground by stems and branches (*cstem*) and a PFT-dependent stem albedo parameter $\beta_{stem,PFT}$:

$$Alb_{stem,PFT} = FPC_{PFT} \times (1 - Phen_{PFT}) \times cstem \times \beta_{stem,PFT} \quad \text{(S14)}$$

The parameter cstem = 0.7 (Strengers et al., 2010) was used as a constant and not further considered in this study. The background (i.e. litter) albedo of a PFT depends additionally on a PFT-dependent litter albedo parameter $\beta_{litter,PFT}$:

$$Alb_{litter,PFT} = FPC_{PFT} \times (1 - Phen_{PFT}) \times (1 - cstem) \times \beta_{litter,PFT} \quad \text{(S15)}$$

The parameters $\beta_{leaf,PFT}$, $\beta_{stem,PFT}$ and $\beta_{litter,PFT}$ were implemented as PFT-dependent albedo parameters which differs from the previous implementation (Strengers et al., 2010). The fraction of snow in the green part of the canopy that is used to compute FAPAR (equation 3) depends on the daily phenological status and the fraction of the PFT that is covered by snow:

$$F_{snow,gv,PFT} = Phen_{PFT} \times F_{snow,PFT} \quad \text{(S16)}$$

The fraction of the PFT that is covered by snow depends on snow height and the daily calculated snow water equivalent (Strengers et al., 2010).

2 FAPAR datasets

2.1 Comparison of the Geoland2 and GIMMS3g FAPAR datasets

We compared the Geoland2 and GIMMS3g FAPAR datasets to assess 1) the agreement of two newly developed FAPAR products and 2) to evaluate the suitability of these products for the optimization of FAPAR and phenology-related parameters in LPJmL. We found important differences between the Geoland2 and GIMMS3g FAPAR datasets during our analyses. The differences are mostly related to inter-annual variability and trends.

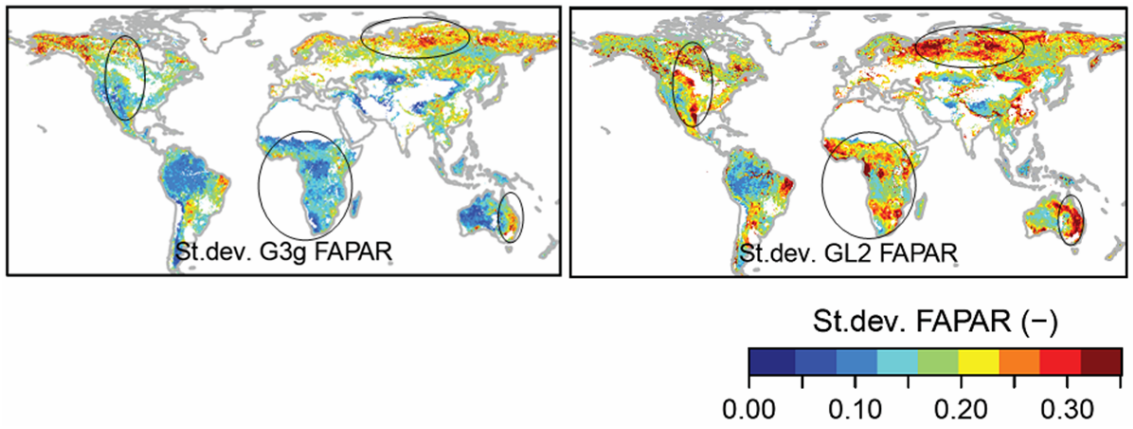

Figure S 2: Standard deviation of mean annual FAPAR from the GIMMS3g and GL2 FAPAR datasets in 1982-2011.
The annual mean FAPAR was calculated for each year from each monthly FAPAR value for months with monthly mean air temperatures > 0°C. Areas with large differences are highlighted with circles.

The GL2 FAPAR dataset had a higher inter-annual variability in most regions especially in northern Russia, central North America, Africa and eastern Australia (Figure S 2). Despite the different amplitudes of inter-annual variability, the temporal dynamic of annual aggregated FAPAR values was well correlated in most regions (Figure S 25). Nevertheless, in some regions like in the North American Tundra, in parts of the Siberian boreal forest and in the tropical forests the inter-annual temporal FAPAR dynamic was weakly or even negatively correlated (Figure S 25).

.

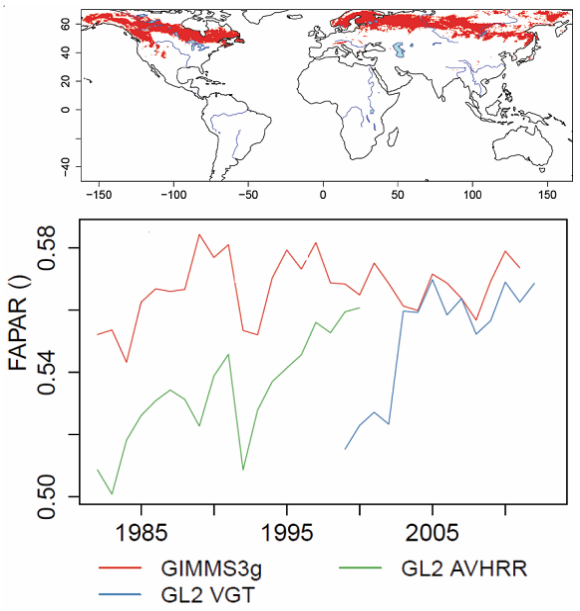

Figure S 3: Comparison of mean annual FAPAR from different datasets averaged for the extent of boreal needle-leaved evergreen forests.

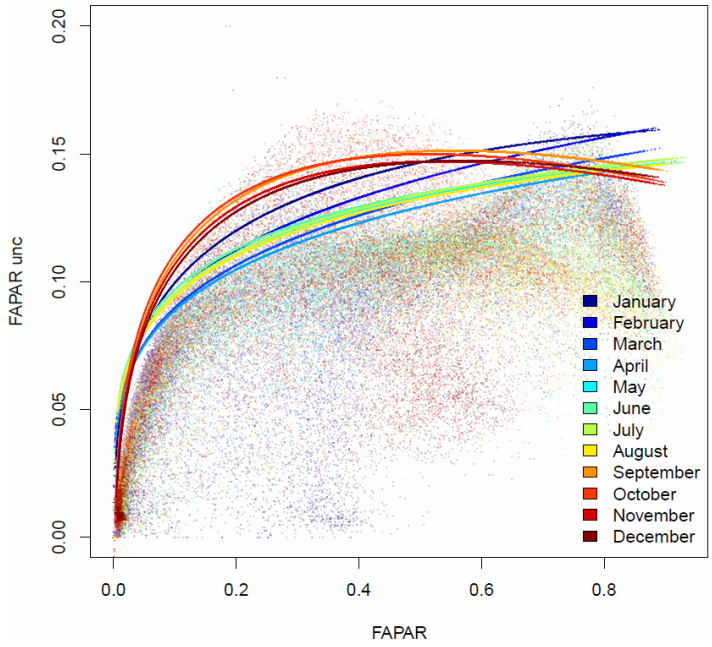

Figure S 4: Monthly quantile regressions between GL2 VGT FAPAR and the GL2 VGT FAPAR fitted to the quantile 0.95.

Each monthly quantile regression was applied to the GIMMS3g FAPAR dataset to estimate uncertainties for this dataset. Using 0.95 quantile regressions provides conservative uncertainty estimates for the GIMMS3g FAPAR dataset.

The temporal dynamics of mean annual FAPAR agreed relatively well between GIMMS3g FAPAR and GL2 FAPAR in the AVHRR period. The temporal dynamic of mean annual FAPAR agreed poorly between GIMMS3g and GL2 FAPAR in the VGT period. Both datasets had higher biases in boreal needle-leaved evergreen forests

(Figure S 3). An offset between the GL2 AVHRR and GL VGT FAPAR time series in the overlapping years 1999 and 2000 is evident in all biomes. Additionally, the GL2 VGT time series shows an abrupt jump from 2002 to 2003 which is probably due to the sensor change from VGT1 to VGT2 (Horion et al., 2014). Because of these reasons, the Geoland2 FAPAR dataset cannot be used for a long-term analysis of FAPAR trends and extremes

2.2 Estimation of uncertainty for the GIMMS3g FAPAR dataset

The GIMMS3g FAPAR dataset was used for parameter optimization. For parameter optimization it is necessary to consider data uncertainty in multiple data stream cost functions. Unfortunately, the GIMMS3g dataset has no uncertainty estimates. On the other hand the GL2 FAPAR dataset has uncertainty estimates but time series are not well harmonized. Thus we were using the GIMMS3g dataset for parameter optimization but estimated uncertainties by using regression to the uncertainty of the GL2 FAPAR dataset (Figure S 4). Therefore we fitted for each month polynomial quantile regressions to the quantile 0.95 between FAPAR and FAPAR uncertainty from the GL2 VGT FAPAR dataset. Then we were using these regressions to estimate uncertainties for the GIMMS3g FAPAR dataset.

3 Land cover

3.1 Creation of an observation-based map of plant functional types

Land cover maps from remote sensing products are not directly comparable with plant functional types in global vegetation models because they are using different legends for the description of vegetation (Jung et al., 2006; Poulter et al., 2011). Land cover classes have to be reclassified into the corresponding PFTs. We were using the SYNMAP land cover map (Jung et al., 2006), the Köppen-Geiger climate classification (Kottek et al., 2006) and tree coverage from MODIS (Townshend et al., 2011). We decided to use the SYNMAP land cover map because it offers fractional land coverage and synergizes already the GLCC, MODIS and GLC2000 land cover maps (Jung et al., 2006). PFTs in LPJmL are defined according to biome (tropical, temperate or boreal), leaf type (needle leaved, broadleaved) and phenology (summergreen, evergreen, rain

green). We extracted the biome information from the Köppen-Geiger climate classification whereas leaf type and phenology were extracted from the SYNMAP land cover map. The FPC of a PFT was derived from MODIS tree cover.

In a first step, we reclassified the Köppen-Geiger climate classification in to bioclimatic zones (biomes) that correspond to the definition used in LPJmL (Figure S 5). This reclassification followed to a large extent the rules of Poulter et al. (2011):

- The climate zone A was reclassified to the tropical biome.
- The climate regions BWh and BSh were reclassified to the tropical biome.
- The climate regions BWk and BSk were reclassified to the temperate biome.
- The climate region Cw was reclassified to the tropical biome.
- The climate regions Cf and Cs were reclassified to the temperate biome.
- The climate regions D and E were reclassified to the boreal biome.

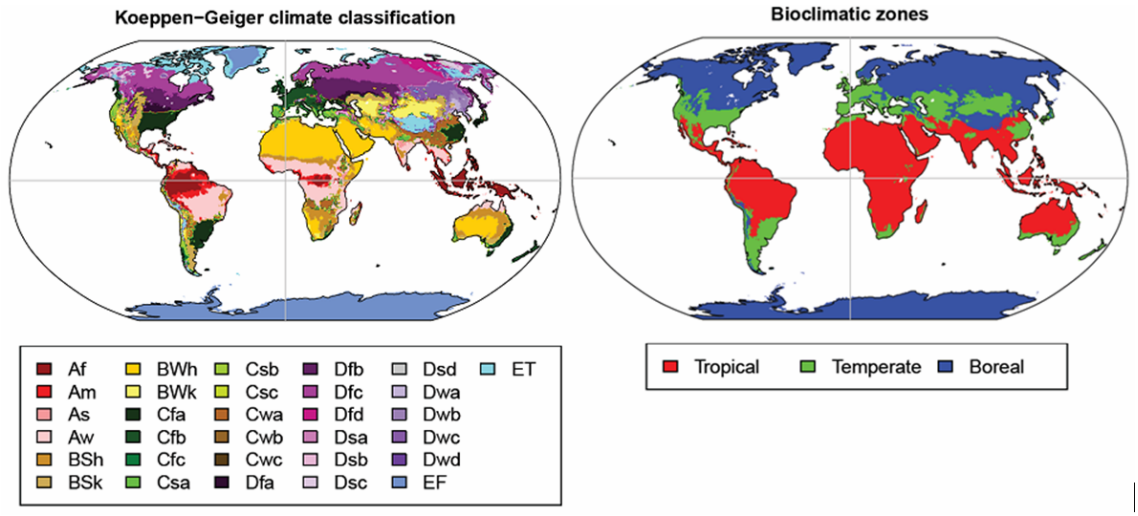

Figure S 5: Reclassification of the Koeppen-Geiger climate classification in bioclimatic zones.

In a second step, we created a land cover map with PFT legend by crossing the land cover information from SYNMAP with the map of biomes following rules for each tree PFT:

- TrBE: EBF (evergreen broadleaved forest) AND tropical biome
- TrBR: DBF (deciduous broadleaved forest) AND tropical biome
- TeNE: ENF (evergreen needleleaved forest) AND temperate biome
- TeBE: EBF (evergreen broadleaved forest) AND temperate biome
- TeBS: DBF (deciduous broadleaved forest) AND temperate biome

– BoNE: ENF (evergreen needleleaved forest) AND boreal biome
– BoBS: DBF (deciduous broadleaved forest) AND boreal biome
– BoNS: DNF (deciduous needleleaved forest) AND boreal biome

Although we translated in this step the land cover classes into PFTs, the fractions represent still fraction of land cover and not FPC. For example, a grid cell can be covered by 100% forest but this forest contains only 70% trees while the rest is covered by herbaceous plants. This difference becomes evident by comparing the total coverage of forest land cover classes from SYNMAP with tree cover from MODIS (Figure S 6). MODIS tree cover is always lower than forest cover but shows more spatial variability.

In a third step, we need to correct the land cover fraction with tree cover to create a map of FPC. Thus, we calculated the FPC of each tree PFT by correcting the land cover fraction of a PFT (LC_{PFT}) with the ratio of fractional tree coverage from MODIS (F_{Tree}) and the total land coverage of all 8 forest PFTs:

$$FPC_{PFT} = LC_{PFT} \times \frac{F_{Tree}}{\sum_{PFT=1}^{PFT=8} LC_{PFT}} \quad \text{(S17)}$$

This calculation of FPC differs from the approach of Poulter et al. (2011) who divided each land cover class in fixed fractions of tree and herbaceous PFTs.

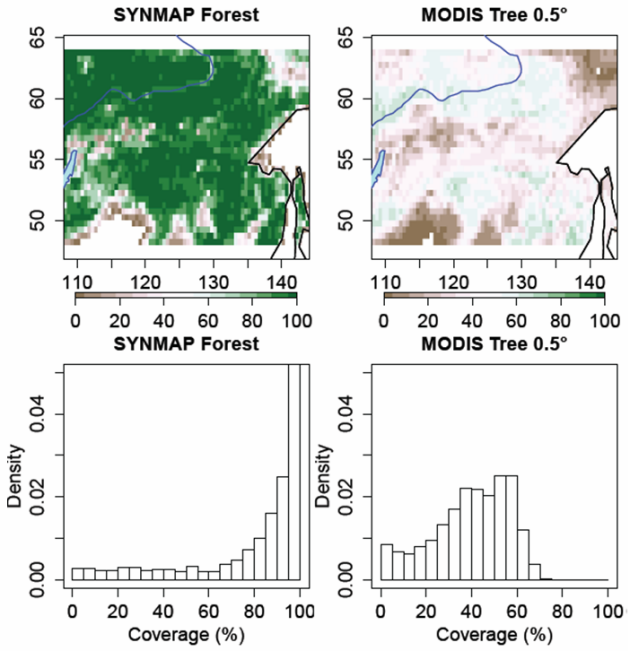

Figure S 6: Comparison of total forest coverage from SYNMAP and MODIS tree coverage for a region in eastern Siberia.

In the last step we need to calculate the FPC of herbaceous PFTs:

$$FPC_{herb} = 1 - F_{Tree} - LC_{Barren} - LC_{Water} - LC_{Snow/Ice} \quad \text{(S18)}$$

which is the residual area by removing the fractional tree coverage from MODIS and the land cover fractions of bare soil and rocks, water and permanent snow and ice from the total grid cell. Thus, grasslands, croplands and shrub lands were assigned to herbaceous vegetation. Then we divided the herbaceous FPC into the TeH, PoH and TrH PFTs according to biomes:

- TrH: FPC_{herb} AND tropical biome
- Old TeH: FPC_{herb} AND temperate OR boreal biome

The TeH was further splitted in a new temperate herbaceous and a polar herbaceous PFT to separate between temperate grasslands and tundra:

- TeH (new): old TeH AND temperate OR boreal biome AND boreal trees < 0.3
- PoH: old TeH AND (boreal biome OR Koeppen-Geiger E climate) AND boreal trees > 0.3

These steps yielded in observation-based maps of foliar projective cover for each PFT (Figure S **7**). As the input data (SYNMAP and MODIS VCF) is based on satellite data from the years 2000/2001 the retrieved maps reflect the distribution of PFTs for the year 2000.

3.2 Comparison of simulated and observed PFT distributions

We compared the observation-based PFT map with the simulated PFT distribution from LPJmL-OP for the year 2000. LPJmL with dynamic vegetation simulated usually too high tree and too low herbaceous cover in all regions (Figure S 8). In the central tropical forests (Amazon, Congo basin) LPJmL simulated too low cover of TrBE but too high cover of TrBR. The coverage of BoNE was too low in some regions in North America and Eastern Siberia. The simulated distribution of BoNS did not agree much with the observed distribution which is almost limited to eastern Siberia. Tree cover was especially overestimated in regions with only sparse tree cover (Savannahs, Steppe/boreal forest transition, eastern Siberia). The extent of boreal forest PFTs (BoNE, BoBS, BoNS) is generally too large with far southward extensions into the Steppe and northward extensions into the Tundra.

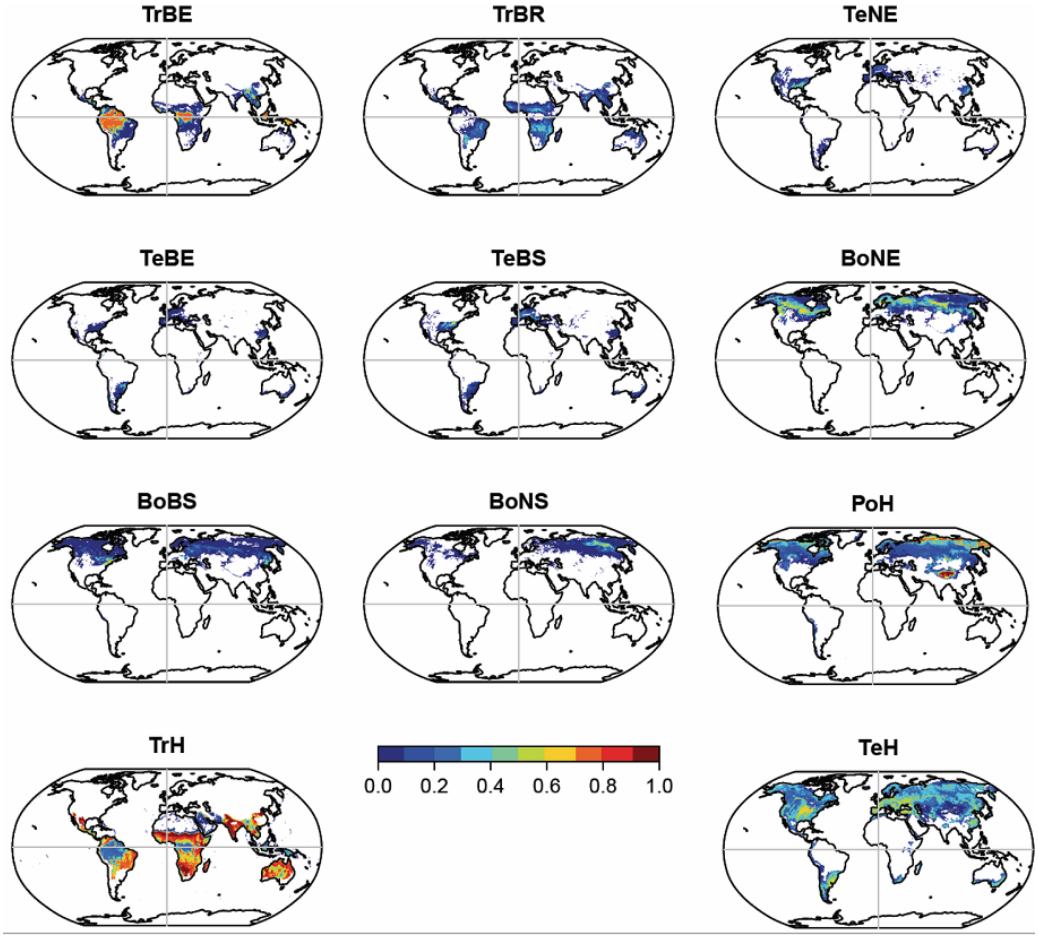

Figure S 7: Observation-based maps of the foliar projective cover of plant functional types. Agricultural areas are included in the TrH and TeH PFTs.

As expected, the prescription of the observed PFT maps into LPJmL generally improved the representation of the observed PFT distributions (Figure S 8). The spatial patterns of PFT distributions were highly correlated and the bias in comparison to the observed distribution was clearly reduced in comparison with the model run with dynamic vegetation. The PFT distribution of the LPJmL model run with prescribed land cover does not perfectly agree with the observed PFT distribution which is due to the applied prescription approach. Tree PFTs can have a lower FPC in LPJmL than the prescribed FPC value because the trees are still growing or because mortality reduced the FPC. This effect especially happened in the BoNE PFT where fire reduced the FPC in large regions in Canada and eastern Siberia (Figure S 8). Herbaceous PFTs can have a higher FPC than the observed FPC value because these PFTs were allowed to establish the entire grid cell (except the fraction that is barren, water or permanent snow/ice in the observations). This happened for example when fires burnt tree PFTs and herbaceous PFTs succeeded afterwards in LPJmL. This is the reason for the

overestimation of herbaceous coverage in large regions in Canada and eastern Siberia where the BoNE PFT was underestimated (Figure S 8). In summary, the prescription of land cover improved the representation of observed spatial patterns of PFTs in LPJmL. Differences to the observed PFT distribution are due to the desired ability of LPJmL to represent important processes of vegetation dynamics like mortality processes.

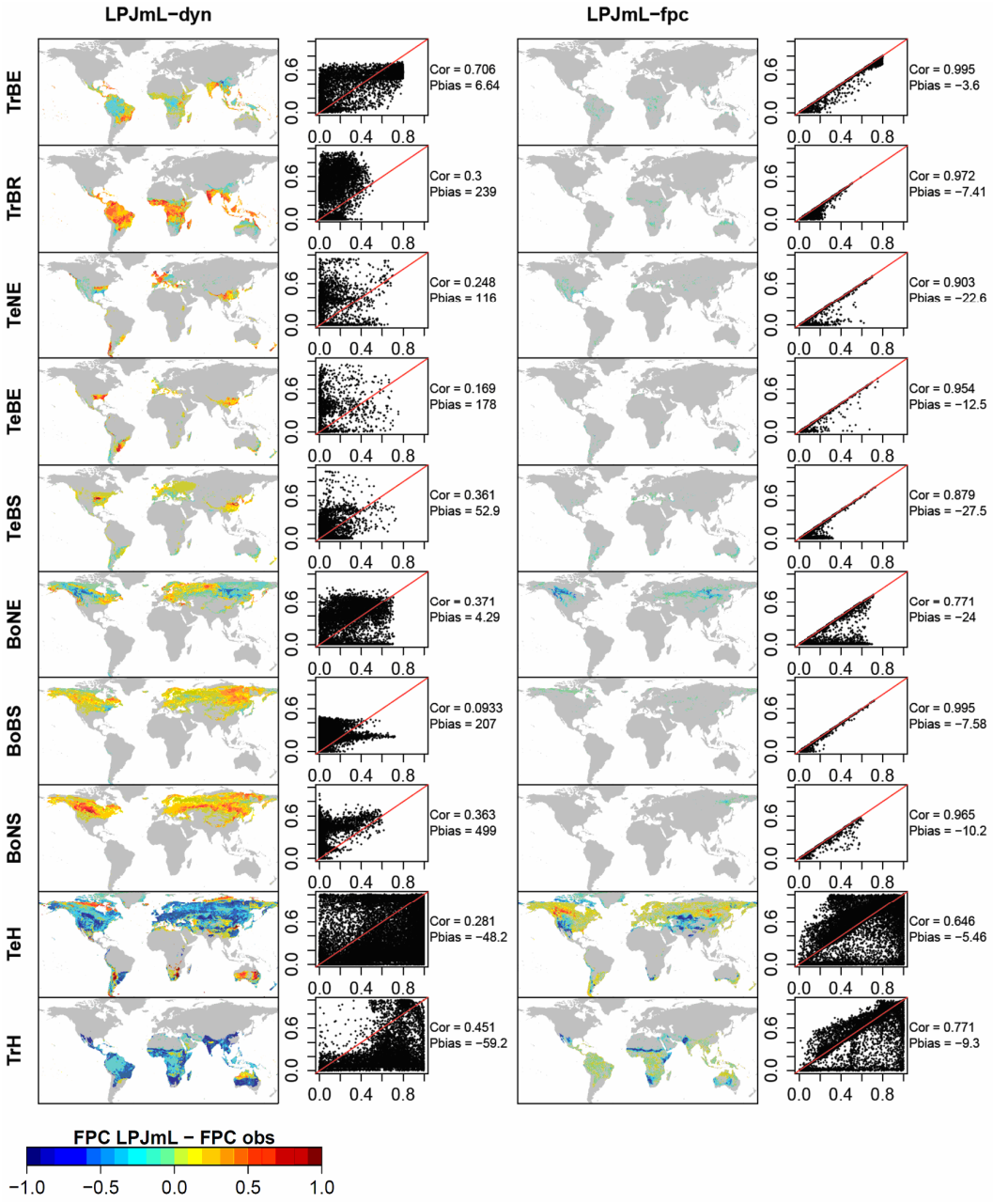

Figure S 8: Comparison between simulated and observed PFT distributions for the year 2000. The maps are difference maps between simulated FPC values from LPJmL-OP and observed FPC values. The scatter plots show observed FPC values on the x-axis and simulated FPC values on the y-axis. Left: LPJmL-OP with dynamic vegetation and prescribed burnt areas. Right: LPJmL-OP with prescribed land cover and prescribed burnt areas.

4 Model parameter optimization

4.1 Parameter definitions and values

This section documents the LPJmL parameters that were addressed in this study. The parameters and their use in the model are described in Table S 1. The information sources from which prior parameter values were extracted for each optimization experiment are shown in Figure S 9. Tables S 2-5 list prior and posterior parameter values of each optimization experiment according to the logical flow of optimization experiments indicated in Figure S 9.

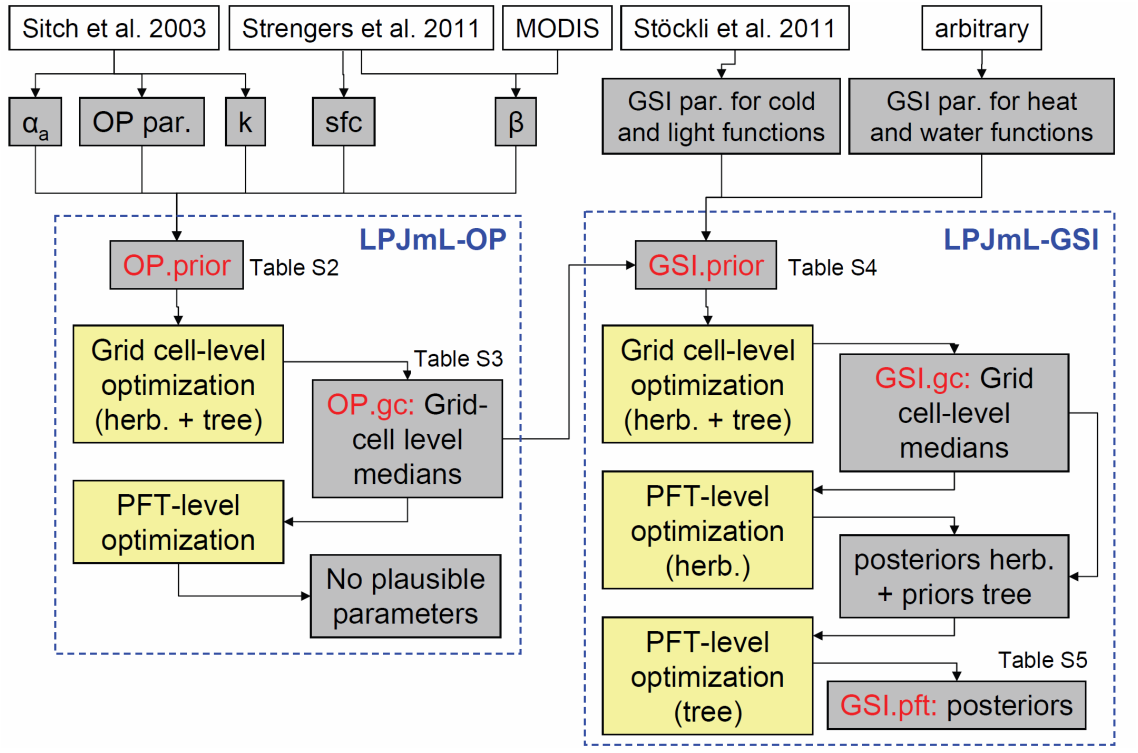

Figure S 9: Information sources for prior and posterior parameter sets and overview of model optimization experiments.
Grey boxes indicate model parameters or parameter sets. White boxes are information sources for parameters. Yellow boxes are optimization experiments.

Table S 1: Description of LPJmL model parameters that were addressed in this study.

Para-meter	Alternative name	Use	Description	Unit
α_a	ALPHAA	Photo-synthesis	Leaf-to-canopy scaling parameter (amount of radiation absorbed at leaf-level in comparison to total canopy)	-
β_{leaf}	ALBEDO_LEAF	Albedo, FAPAR	Albedo of green leaves	-
β_{stem}	ALBEDO_STEM	Albedo	Albedo of stems and branches	-
β_{litter}	ALBEDO_LITTER	Albedo	Albedo of litter	-
k	LIGHTEXTCOEFF	FPC, FAPAR	Light extinction coefficient in Lambert-Beer relationship	-
sfc	SNOWCANOPYFRAC	Albedo, FAPAR	Maximum fraction of snow in the green canopy	-
$Wscal_{min}$	MINWSCAL	Original phenology	Minimum value of the water availability scaling factor for leaf onset in rain green PFTs	-
GDD_{base}	GDDBASE	Original phenology	Minimum daily temperature to start counting growing degree days	°C
ramp	RAMP	Original phenology	Number of growing degree days to reach full leave cover in summergreen PFTs	°C
$aphen_{min}$	APHEN_MIN	Original phenology	Minimum accumulated phenology state to allow senescence if temperature < GDDBASE	-
$aphen_{max}$	APHEN_MAX	Original phenology	Maximum accumulated phenology state. Phenology is set back to 0 if this value is passed.	-
sl_{tmin}	TMIN_SL	GSI phenology	Slope of cold temperature limiting logistic function for phenology	1/°C
$base_{tmin}$	TMIN_BASE	GSI phenology	Inflection point of cold temperature limiting logistic function for phenology	°C
τ_{tmin}	TMIN_TAU	GSI phenology	Change rate of actual to previous day cold temperature limiting function value for phenology	-
sl_{light}	LIGHT_SL	GSI phenology	Slope of light limiting logistic function for phenology	$1/(W/m^2)$
$base_{light}$	LIGHT_BASE	GSI phenology	Inflection point of light limiting logistic function for phenology	W/m^2
τ_{light}	LIGHT_TAU	GSI phenology	Change rate of actual to previous day light limiting function value for phenology	-
sl_{water}	WATER_SL	GSI phenology	Slope of water limiting logistic function for phenology	1/%
$base_{water}$	WATER_BASE	GSI phenology	Inflection point of water limiting logistic function for phenology	%
τ_{water}	WATER_TAU	GSI phenology	Change rate of actual to previous day water limiting function value for phenology	-
sl_{heat}	TMAX_SL	GSI phenology	Slope of heat limiting logistic function for phenology	1/°C
$base_{heat}$	TMAX_BASE	GSI phenology	Inflection point of heat limiting logistic function for phenology	°C
τ_{heat}	TMAX_TAU	GSI phenology	Change rate of actual to previous day heat limiting function value for phenology	-

Table S 2: Prior parameter values of LPJmL-OP (OP.prior).

The values in brackets are ranges of uniform parameter distributions that were used during optimization. Note: * The parameter GDDbase was changed to 0°C. This value gave better agreements between simulated and observed seasonal FAPAR dynamics than the original value of 5°C. Nevertheless, GDDbase was not included in optimization experiments because this parameter is highly correlated with the parameter ramp.

	TrBE	**TrBR**	**TeNE**	**TeBE**	**TeBS**	**BoNE**	**BoBS**	**BoNS**	**TeH**	**TrH**
α_a	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)
β_{leaf}	0.15 (0.1-0.2)	0.15 (0.1-0.2)	0.15 (0.06-0.23)	0.15 (0.09-0.23)	0.16 (0.086-0.23)	0.14 (0.05-0.23)	0.15 (0.09-0.21)	0.12 (0.1-0.15)	0.14 (0.072-0.22)	0.15 (0.09-0.21)
β_{stem}	0.15 (0.018-0.29)	0.15 (0.073-0.23)	0.13 (0-0.31)	0.15 (0.029-0.28)	0.13 (0.038-0.23)	0.14 (0-0.31)	0.14 (0.059-0.23)	0.13 (0.052-0.32)	--	--
β_{litter}	0.15 (0.018-0.29)	0.14 (0.058-0.27)	0.13 (0.047-0.21)	0.15 (0.044-0.29)	0.14 (0.085-0.2)	0.13 (0.035-0.26)	0.14 (0.078-0.22)	0.12 (0.088-0.23)	0.14 (0.027-0.38)	0.13 (0.02-0.28)
sfc	0.4 (0.1-0.9)	0.4 (0.1-0.9)	0.4 (0.1-0.9)	0.4 (0.1-0.9)	0.4 (0.1-0.9)	0.4 (0.1-0.9)	0.4 (0.1-0.9)	0.4 (0.1-0.9)	0.4 (0.1-0.9)	0.4 (0.1-0.9)
k	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)	0.5 (0.1-0.9)
GDD_{base}	--	--	--	--	0*	--	0*	0*	0*	0*
$Wscal_{min}$	--	0.3 (0-1)	--	--	--	--	--	--	--	--
Ramp	--	--	--	--	300 (0-1000)	--	200 (0-1000)	200 (0-1000)	100 (0-1000)	100 (0-1000)
$aphen_{min}$	--	--	--	--	10 (1-600)	--	10 (1-600)	10 (1-600)	--	--
$aphen_{max}$	--	--	--	--	210 (1-600)	--	210 (1-600)	210 (1-600)	--	--

4.2 Genetic optimization algorithm

We were using a genetic optimization algorithm to minimize the cost function J(d) by optimizing the scaled parameter vector d. The GENOUD algorithm (genetic optimization using derivatives) (Mebane and Sekhon, 2011) combines global genetic optimization search with local gradient-based search algorithms. In genetic optimization algorithms, each model parameter is called a gene and each parameter set is called an individual. The fitness of this individual is the cost of the model against the observations. At the beginning of the optimization, a first generation of individuals is initialized by random sampling of parameter sets within the prescribed parameter

ranges. After the calculation of the cost of all individuals of the first generation, a next generation is generated by cloning the best individuals, by mutating the genes or by crossing different individuals (Mebane and Sekhon, 2011). This results after some generations in a set of individuals with highest fitness, i.e. parameter sets with minimized cost. Within the GENOUD algorithm we were using also the BFGS (Broyden-Fletcher-Goldfarb-Shanno) gradient search algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) to find an optimum parameter set. An optimized parameter set of the BFGS algorithm is used as individual in the next generation. The BFGS gradient search algorithm was first applied on the best individual of the second last generation to avoid a too fast convergence of the optimization algorithm towards a local optimum. For grid cell-based optimization experiments we were applying the GENOUD algorithm with at least 20 generations and a population size of 1000 individuals per generation, i.e. at least 20000 single model runs. For PFT-level optimization experiments we were applying the GENOUD algorithm with at least 15 generations and a population size of at least 700 individuals per generation, i.e. at least 10500 single model runs.

Table S 3: Posterior parameter values for LPJmL-OP based on grid cell-level optimization experiments (OP.gc).

Parameters written *in italics* were derived as the median value of the single grid cell optimization experiments whereas all other parameters were derived from prior parameter sources. For the parameter ramp no plausible parameter was found. The parameter GDDbase was changed to 0 but not included in the optimization.

	TrBE	**TrBR**	**TeNE**	**TeBE**	**TeBS**	**BoNE**	**BoBS**	**BoNS**	**TeH**	**TrH**
α_a	*0.6*	*0.56*	*0.38*	*0.41*	*0.38*	*0.28*	*0.34*	*0.27*	*0.32*	*0.39*
β_{leaf}	*0.13*	*0.1*	*0.06*	*0.1*	*0.16*	*0.05*	*0.18*	*0.11*	*0.08*	*0.15*
β_{stem}	0.15	*0.07*	0.13	0.15	*0.04*	0.14	*0.06*	*0.05*	--	--
β_{litter}	0.15	*0.06*	0.13	0.15	*0.09*	0.13	*0.08*	*0.09*	*0.1*	*0.14*
sfc	0.4	0.4	*0.1*	0.4	0.4	0.1	*0.15*	*0.18*	0.4	0.4
k	*0.36*	*0.73*	*0.41*	*0.44*	*0.74*	*0.71*	*0.51*	*0.88*	*0.39*	*0.46*
GDD_{base}	--	--	--	--	0	--	0	0	0	0
$Wscal_{min}$	--	*0.85*	--	--	--	--	--	--	--	--
Ramp	--	--	--	--	300	--	200	200	100	100
$aphen_{min}$	--	--	--	--	10	--	10	10	--	--
$aphen_{max}$	--	--	--	--	*201.97*	--	*181.62*	*105.78*	--	--

Table S 4: Prior parameter values for LPJmL-GSI (GSI.prior).

Parameters marked with * were identified as insensitive and were not included in the optimization. The values in brackets are ranges of uniform parameter distributions that were used during optimization. The values for the first 6 parameters were derived from the single grid-cell optimization experiments of LPJmL-OP (Table S 3).

	TrBE	TrBR	TeNE	TeBE	TeBS	BoNE	BoBS	BoNS	TrH	TeH PoH
α_a	0.6 (0.2-0.8)	0.56 (0.1-0.9)	0.38 (0.23-0.49)	0.41 (0.1-0.9)	0.38 (0.15-0.6)	0.28 (0.16-0.57)	0.34 (0.15-0.61)	0.27 (0.16-0.55)	0.39 (0.21-0.83)	0.32 (0.1-0.83)
β_{leaf}	0.13 (0.1-0.2)	0.1 (0.05-0.2)	0.06 (0.01-0.23)	0.1 (0.09-0.23)	0.16 (0.13-0.19)	0.05 (0.01-0.23)	0.18 (0.09-0.21)	0.11 (0.1-0.14)	0.15 (0.09-0.21)	0.08 (0.072-0.22)
β_{stem}	0.15 (0.018-0.29)	0.07 (0.06-0.23)	0.13 (0-0.31)	0.15 (0.029-0.28)	0.04 (0.038-0.23)	0.14 (0-0.31)	0.06 (0.059-0.23)	0.05 (0.04-0.32)	--	--
β_{litter}	0.15 (0.054-0.29)	0.06 (0.058-0.27)	0.13 (0.047-0.21)	0.15 (0.044-0.29)	0.09 (0.085-0.2)	0.13 (0.035-0.26)	0.08 (0.078-0.22)	0.09 (0.088-0.23)	0.14 (0.02-0.28)	0.1 (0.027-0.38)
sfc	0.4*	0.4*	0.1 (0.01-0.9)	0.4*	0.4 (0.1-0.9)	0.1 (0.01-0.9)	0.15 (0.1-0.9)	0.18*	0.4*	0.4 (0.1-0.9)
k	0.36 (0.2-0.9)	0.73 (0.1-0.9)	0.41 (0.1-0.9)	0.44 (0.1-0.9)	0.74 (0.1-0.9)	0.71 (0.1-0.9)	0.51 (0.1-0.9)	0.88 (0.1-0.9)	0.46 (0.1-0.9)	0.39 (0.1-0.9)
sl_{tmin}	0.24 (0.1-2)	0.24*	0.24 (0.1-2)	0.24 (0.1-2)	0.24 (0.1-2)	0.24 (0.1-2)	0.24 (0.1-2)	0.24 (0.1-2)	0.24 (0.1-2)	0.24 (0.1-2)
$base_{tmin}$	8.8 (0-16)	8.8 (0-16)	-3.3 (-6-6)	-0.6 (-3-1)	7.4 (5-9)	3.7 (-6-6)	2.2 (0-5)	-4 (-6-6)	8.8 (0-16)	0.7 (-3-5)
τ_{tmin}	0.2*	0.2*	0.2*	0.2*	0.2*	0.2*	0.2*	0.2*	0.2*	0.2*
sl_{heat}	0.24 (0.01-3)	0.24 (0.01-3)	0.24 (0.01-3)	0.24 (0.01-3)	0.24 (0.01-3)	0.24*	0.24*	0.24*	0.24 (0.01-3)	0.24*
$base_{heat}$	35 (25-45)	35 (25-45)	35 (25-45)	35 (25-45)	35 (25-45)	35 (25-45)	35 (25-45)	35 (25-45)	35 (25-45)	35 (25-45)
τ_{heat}	0.2 (0.01-0.9)	0.2*	0.2 (0.01-0.9)	0.2*	0.2*	0.2*	0.2 (0.01-0.9)	0.2*	0.2 (0.01-0.9)	0.2*
sl_{light}	57 (0.05-157)	23*	20*	0.2 (0.05-40)	58*	14*	101 (0.05-220)	95*	41 (0.05-130)	23*
$base_{light}$	125 (1-200)	62 (1-200)	73 (1-200)	23 (1-50)	123 (50-200)	57 (1-100)	166 (50-200)	156 (130-180)	104 (1-150)	67 (1-180)
τ_{light}	0.2 (0.01-0.9)	0.2*	0.2*	0.2*	0.2*	0.2*	0.2*	0.2*	0.2 (0.01-0.9)	0.2 (0.01-0.9)
sl_{water}	5 (0.1-10)	5 (0.1-10)	5*	5*	5 (0.1-10)	5*	5 (0.1-10)	5*	5 (0.1-10)	5 (0.1-10)
$base_{water}$	20 (1-99)	20 (1-99)	20 (1-99)	20 (1-99)	20 (1-99)	20 (1-99)	20 (1-99)	20 (1-99)	20 (1-99)	20 (1-99)
τ_{water}	0.8 (0.01-0.99)	0.8 (0.01-0.99)	0.8*	0.8*	0.8*	0.8*	0.8*	0.8*	0.8 (0.01-0.99)	0.8 (0.01-0.99)

Table S 5: Final parameters for LPJmL-GSI.

Parameters written *in italics* were derived from PFT-level optimization experiments (GSI.pft) whereas all other parameters were derived from prior parameter sources as described in Figure S 9.

	TrBE	TrBR	TeNE	TeBE	TeBS	BoNE	BoBS	BoNS	TrH	TeH	PoH
α_a	*0.63*	*0.52*	*0.44*	*0.45*	*0.61*	*0.22*	*0.41*	*0.34*	*0.40*	*0.32*	*0.43*
β_{leaf}	0.13	*0.12*	*0.12*	*0.12*	*0.18*	*0.10*	0.16	*0.12*	*0.24*	*0.18*	0.07
β_{stem}	0.10	0.10	0.04	0.04	0.04	0.06	0.06	0.04	0.15	0.15	0.15
β_{litter}	0.10	0.10	*0.05*	0.10	*0.14*	*0.01*	*0.00*	*0.01*	*0.12*	*0.07*	0.03
k	*0.52*	*0.74*	*0.47*	*0.70*	*0.60*	*0.44*	0.41	*0.66*	0.50	0.50	0.50
sl_{tmin}	1.01	0.24	*0.22*	0.55	*0.26*	*0.10*	0.22	0.15	0.91	0.31	0.13
$base_{tmin}$	8.30	7.66	*-7.81*	*-0.63*	*13.69*	*-7.52*	2.05	*-4.17*	6.42	*4.98*	*2.79*
τ_{tmin}	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	*0.01*	0.20
sl_{heat}	1.86	*1.63*	1.83	0.98	1.74	0.24	1.74	0.24	1.47	0.24	0.24
$base_{heat}$	*38.64*	*38.64*	*35.26*	*41.12*	*41.51*	*27.32*	*41.51*	*44.60*	*29.16*	*32.04*	*26.12*
τ_{heat}	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
sl_{light}	77.17	23.00	20.00	18.83	58.00	14.00	58.00	95.00	64.23	23.00	23.00
$base_{light}$	*55.53*	*13.01*	*4.87*	*39.32*	*59.78*	*3.04*	*59.78*	*130.1*	69.90	*75.94*	*50.00*
τ_{light}	0.52	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.40	0.22	0.38
sl_{water}	5.14	*7.97*	5.00	5.00	*5.24*	5.00	5.24	5.00	0.10	0.52	0.88
$base_{water}$	*5.00*	*22.21*	*8.61*	*8.82*	*20.96*	*0.01*	*20.96*	*2.34*	*41.72*	*53.07*	*1.00*
τ_{water}	0.44	*0.13*	0.80	0.80	0.80	0.80	0.80	0.80	0.17	0.01	0.94

4.3 Parameter sensitivities and uncertainties

To explore the sensitivity and uncertainty of LPJmL-GSI parameters after PFT-level optimizations, we computed the likelihood L and Akaikes Information Criterion AIC from the cost J of each individual (i.e. parameter set d) of the genetic optimization:

$$L = e^{-J(d)} \tag{S19}$$

$$AIC = 2 \times n - 2 \times \log(L) \tag{S20}$$

Where n is the number of parameters. The optimum parameter set has the highest likelihood and the lowest AIC. Then, we selected only these individuals with an AIC difference dAIC of < 2 in comparison to the best parameter set:

$$dAIC = AIC - AIC_{best} \tag{S21}$$

Parameter sets or model formulations with an AIC difference < 2 are usually considered as equally plausible like the best parameter set (Burnham and Anderson, 2002, p.70). The relationship between likelihood and the value of each parameter provides both a qualitative insight in the uncertainty of parameters as expressed by the

parameter range and in the parameter sensitivity as expressed by the maximum likelihood at each parameter value.

4.4 Supplementary results and discussion on optimization performance

The optimization of LPJmL-OP and LPJmL-GSI resulted in a significant reduction of the cost in comparison to the respective prior models although there were differences between plant functional types (Figure S 10). LPJmL-OP with prior parameters had high costs especially in herbaceous PFTs (TrH and TeH) and in the boreal needle-leaved summer green PFT (BoNS). The optimization of single grid cells in LPJmL-OP resulted in a significant reduction of the cost in all PFTs ($p \leq 0.01$, Wilcoxon rank-sum test) despite the polar herbaceous and tropical herbaceous PFTs. The global prior parameter set of LPJmL-GSI resulted in a significant lower cost than the grid cell-level optimized parameter sets of LPJmL-OP in TrH, TeBS, BoNS and PoH PFTs. The optimization of single grid cells in LPJmL-GSI resulted in a significant reduction of the cost in all PFTs except BoNS and PoH. PFT-level optimizations of LPJmL-GSI resulted in a significant lower cost than the LPJmL-GSI prior parameter set in all PFTs except TeBE, BoNS and PoH. PFT-level optimizations of LPJmL-GSI resulted in a significant lower cost than the standard LPJmL-OP prior parameter set in all PFTs except TeNE. These results demonstrate an improved overall performance of optimized model parameter sets over prior model parameter sets and of LPJmL-GSI over LPJmL-OP regarding a cost that is defined based on 30 years of monthly FAPAR, mean annual GPP and 10 years of monthly vegetation albedo.

Model optimization experiments resulted in a significant reduction of the annual GPP bias of LPJmL in comparison to the MTE data-oriented GPP product (Figure S 11). LPJmL-OP with prior parameters underestimated mean annual GPP in the TrBE PFT (median Pbias -13%) and overestimated mean annual GPP in all other PFTs (up to 123% median Pbias in TeH). Grid cell-level optimization experiments of LPJmL-OP resulted in a significant reduction of the GPP bias in all PFTs except in the PoH PFT. Especially in the TrBE, TrBR, TrH, TeNE, TeBE, TeBS and BoBS PFTs the bias of mean annual GPP of LPJmL was removed almost completely (i.e. Pbias within 5%). The

LPJmL-GSI prior parameter set had significant lower biases of mean annual GPP than the prior parameter set of LPJmL-OP. This was because the median of each parameter from the OP.gc experiments was used as prior parameter for LPJmL-GSI. Grid cell-level optimization experiments of LPJmL-GSI resulted in significant reductions of the bias in mean annual GPP in most PFTs despite PFTs where the LPJmL-GSI prior parameter set resulted already in GPP biases close to 0 (i.e. TrH, TeBE and PoH). PFT-level optimization experiments of LPJmL-GSI resulted in significant lower biases of mean annual GPP than the prior parameter set of LPJmL-OP in all PFTs except PoH. These results demonstrate that through the applied model optimization biases in mean annual GPP were significantly reduced in all PFTs (except PoH) in LPJmL-OP as well as in LPJmL-GSI.

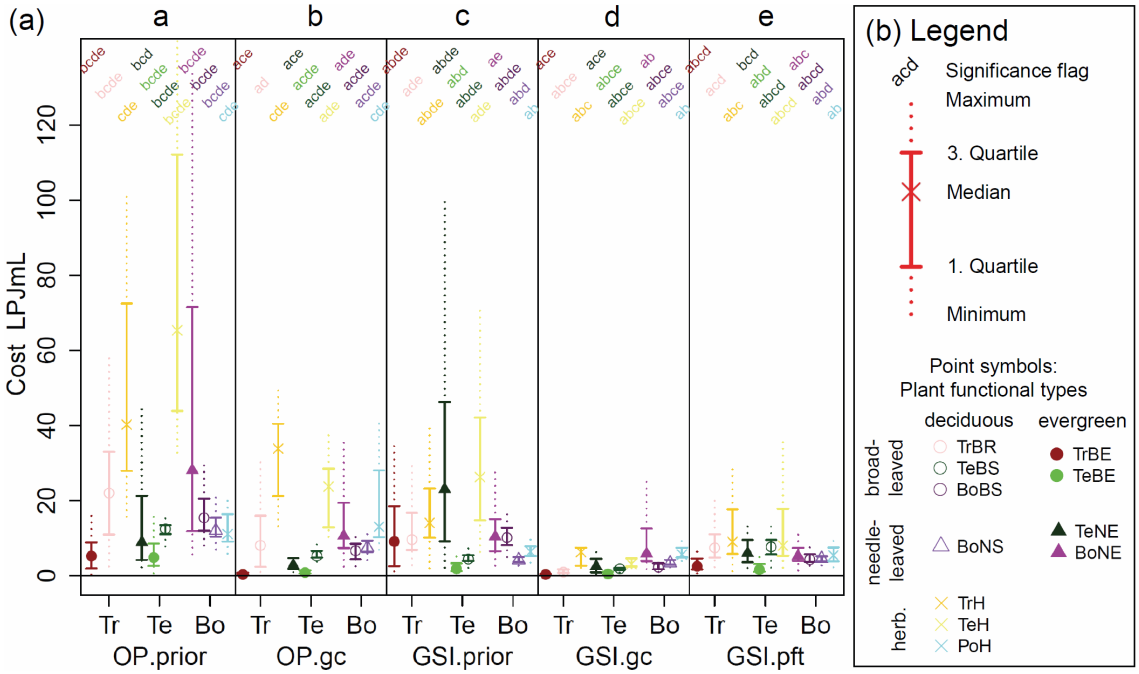

Figure S 10: Distribution of the cost for several grid cells in prior model runs and optimization experiments grouped by plant functional types and biomes.

(a) Cost for LPJmL-OP with default parameters (a, OP.prior), after grid cell-level optimizations (b, OP.gc), cost for LPJmL-GSI with prior parameters (c, GSI.prior), after grid cell-level optimizations (d, GSI.gc) and after PFT-level optimizations (e, GSI.pft). Biomes are Tr (tropical), Te (temperate) and Bo (boreal/polar). (b) Legend for the plot. Each distribution is plotted according to usual boxplot statistics. The point symbols indicate the plant functional type. The significance flag on top of each distribution shows if a distribution is significant different ($p \leq 0.01$) to the corresponding distribution of the same PFT in another optimization experiment. The significance is based on the Wilcoxon rank-sum test. For example "acd" indicates a significant difference to the main categories a (OP.prior), c (GSI.prior) and d (GSI.gc) but no significant difference to b (OP.gc) and e (GSI.pft).

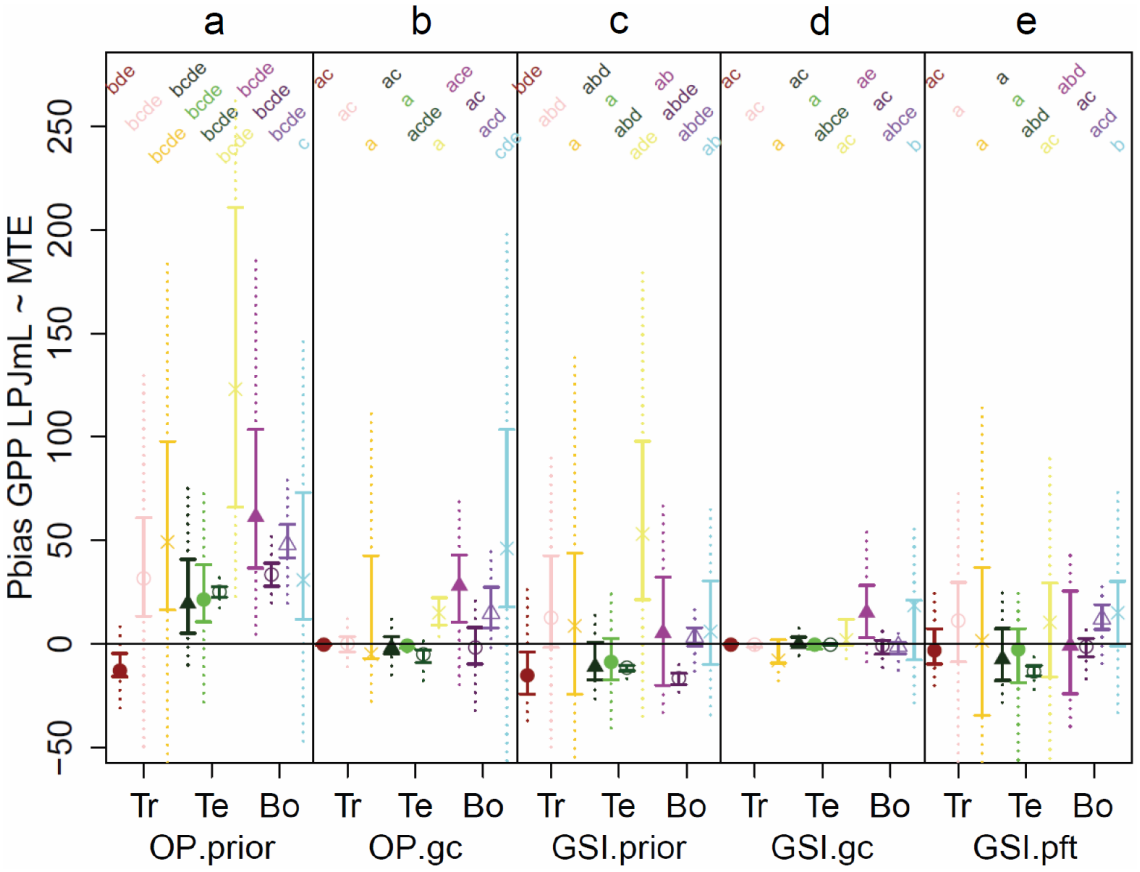

Figure S 11: Distribution of the percent bias between LPJmL and MTE mean annual GPP (1982-2011) for several grid cells in prior model runs and optimization experiments grouped by plant functional types and biomes.
See Figure S 10 for a further explanation of this figure.

We were not able to remove the GPP bias and to reduce the cost of LPJmL-OP and of LPJmL-GSI in the PoH PFT (tundra) in optimization experiments because of inconsistencies between the FAPAR and GPP datasets or in the LPJmL formulation. Although a complete removal of the GPP bias is in principle possible by adjusting the α_a parameter, this would result in a too low FPC of the PoH PFT. Such a low FPC cannot explain the relatively high peak FAPAR values that are seen in the GIMMS3g FAPAR dataset in Tundra regions. It is not possible to explain the low mean annual MTE GPP and the relatively high GIMMS3g peak FAPAR with the current LPJmL model structure in tundra regions. The reasons for this mismatch can be caused by inconsistencies between the GPP and FAPAR datasets or by an insufficient model formulation. The MTE data-oriented GPP product has been upscaled from FLUXNET eddy covariance measurements (Jung et al., 2011). Nevertheless, not many eddy

covariance measurement sites cover tundra regions with mean annual air temperatures < 0°C. Thus, the MTE GPP estimates are not well supported by measurements in tundra regions. But also the FAPAR dataset might be more uncertain in tundra regions than in other parts of the globe. Optical remote sensing in high-latitude regions is usually performed under high-sun zenith angles. Radiation can penetrate deeper into vegetation under high-sun zenith angles which results in higher FAPAR (Tao et al., 2009; Walter-Shea et al., 1998). Thus, the high FAPAR values in the GIMMS3g FAPAR dataset might be caused by satellite observations under high-sun zenith angles. Finally, the inconsistencies between GPP and FAPAR might be also caused by an inappropriate representation of tundra plant communities in LPJmL. The PoH PFT in LPJmL was derived from a grass PFT but does not include shrubs or the large functional diversity of mosses and lichen that are the dominant plant communities in tundra ecosystems (Porada et al., 2013). We currently cannot decide if the inconsistency between FAPAR and GPP in our optimization of productivity and FAPAR parameters in tundra regions is more caused by the specific properties of the datasets or by an insufficient model structure.

All optimization experiments resulted in reasonable albedo biases of LPJmL-OP and LPJmL-GSI in comparison with monthly MODIS albedo time series (Figure S 12). LPJmL-OP with prior parameters overestimated growing season albedo in all PFTs. Grid cell-level optimization experiments of LPJmL-OP resulted in significant reductions of the bias in growing season albedo in TrBE, TeNE, TeBE, TeBS, BoNE, and BoNS PFTs but not in TrBR, TrH, TeH, BoBS and PoH PFTs. The bias in growing season albedo of the latter PFTs was significantly reduced with the LPJmL-GSI prior parameter set. The optimization of LPJmL-GSI for single grid cells significantly reduced the bias in growing season albedo in comparison to the LPJmL-GSI prior parameter set in all PFTs except in the TeH, BoNS and PoH PFTs. These results demonstrate that model optimizations experiments kept growing season albedo within reasonable ranges in comparison to MODIS albedo.

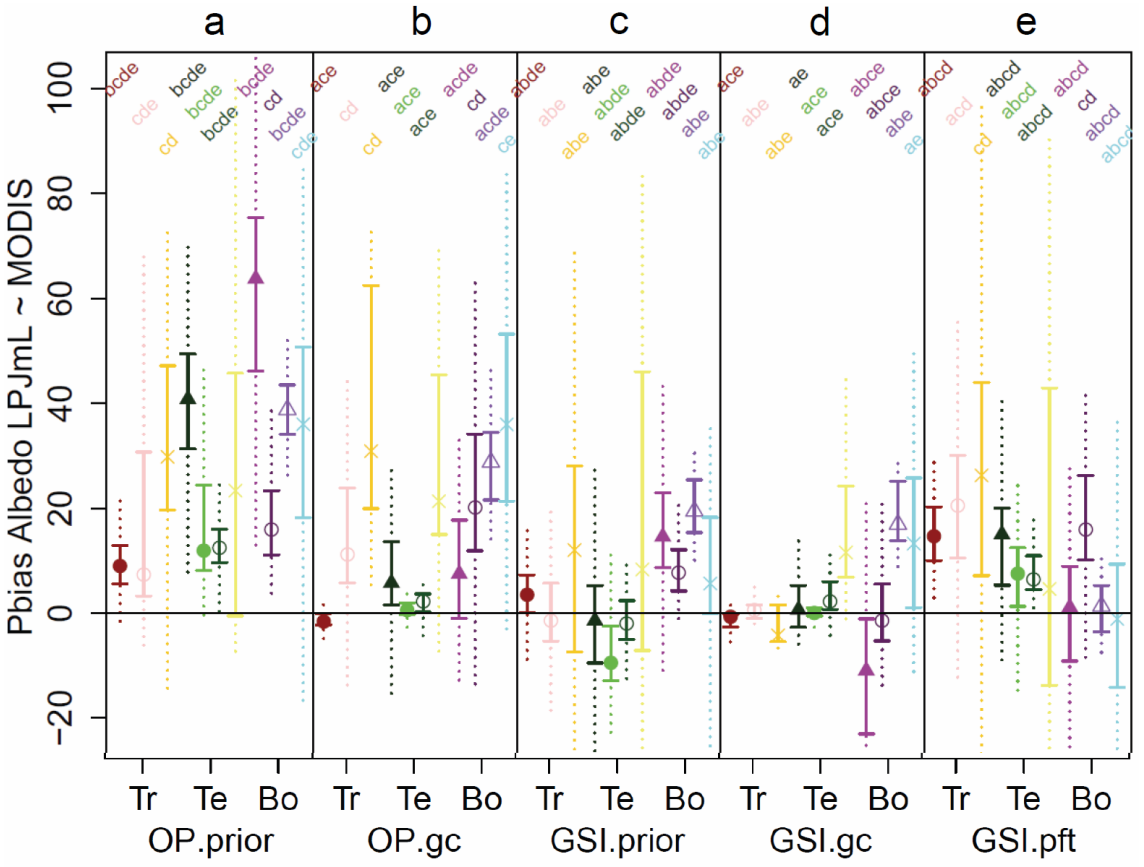

Figure S 12: Distribution of the percent bias between LPJmL and MODIS monthly growing season albedo (2000-2011) for several grid cells in prior model runs and optimization experiments grouped by plant functional types and biomes.
See Figure S 10 for a further explanation of this figure.

4.5 Supplementary results and discussion on parameter variability

The optimization of the leaf albedo parameter β_{leaf} resulted in values that differed especially between broadleaved and needle-leaved evergreen PFTs (Figure S 13). Needle-leaved evergreen PFTs (TeNE and BoNE) had in all optimization experiments the lowest β_{leaf} parameter values while the broad-leaved summergreen PFTs (TeBS and BoBS) had the highest β_{leaf} parameter values. After the PFT-level optimization of LPJmL-GSI herbaceous PFTs had high β_{leaf} parameters. The leaf albedo parameter β_{leaf} was sensitive in all PFTs (Figure 5 of the main text). The optimization resulted in many PFTs in leaf and litter albedo parameters that were close to the boundaries of the prior parameter ranges. This indicates missing environmental controls on surface albedo. The albedo routines of LPJmL need to be further improved to account for moisture-driven changes in surface albedo. Such improved albedo routines would allow a more accurate and constrained estimation of albedo parameters. Because of these current

limitations in the LPJmL albedo routines, albedo simulations in regions or time periods with low vegetation cover need to be assessed with care.

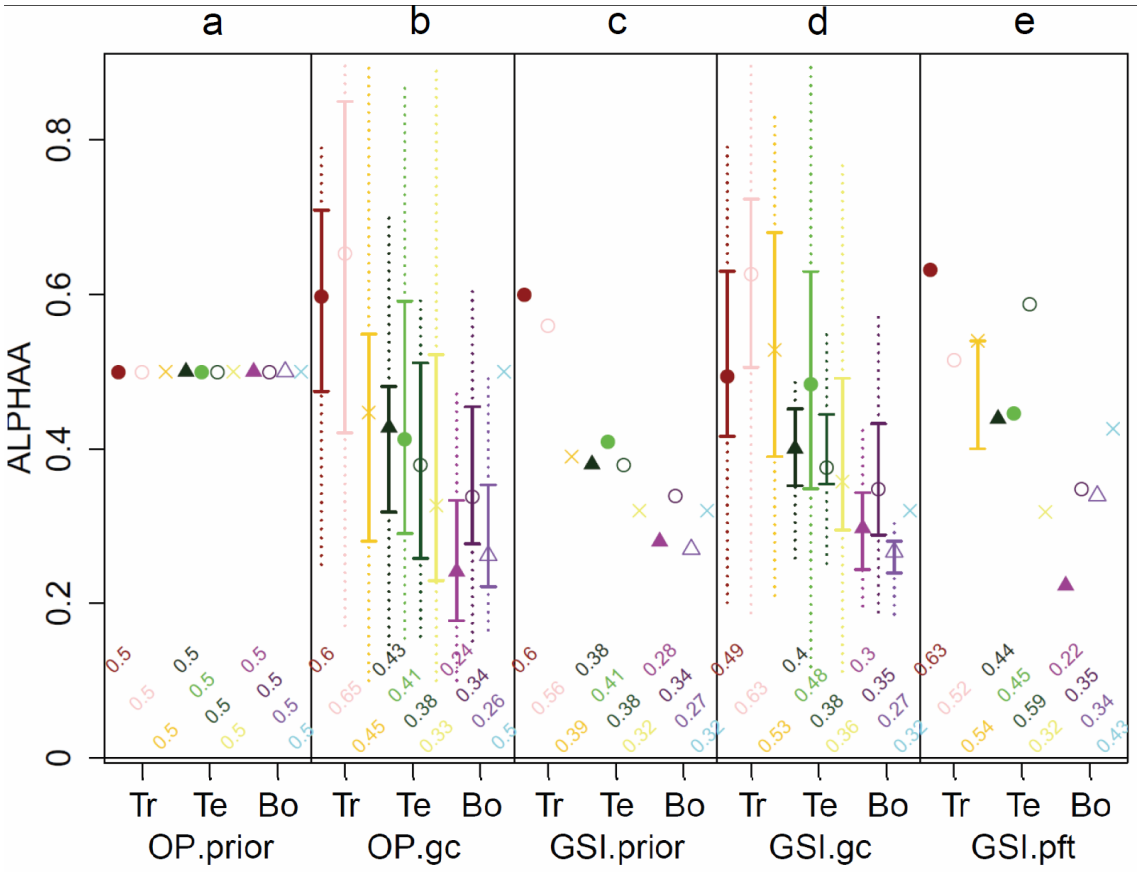

Figure S 13: Prior and optimized values for the parameter αa (fraction of radiation absorbed at leaf level relative to canopy level) grouped by plant functional types and biomes.
The distribution of the parameter in the optimization experiments OP.gc and GSI.gc represents the spatial variability of the parameter from different grid cell-level optimization experiments. See Figure S 10 for a further explanation of this figure.

The light extinction coefficient k had a large spatial variability in all PFTs and in both grid cell-level optimization experiments of LPJmL-OP and LPJmL-GSI (Figure S 15). The spatial variability was lower after grid cell-level optimization experiments of LPJmL-GSI than after grid cell-level optimization experiments of LPJmL-OP. The largest variability was found in evergreen PFTs (TrBE, TeBE, TeNE and BoNE). This result demonstrates that unique or PFT-dependent light extinction coefficient parameter values are not meaningful. Moreover, the spatial variability of the light extinction coefficient needs to be analyzed more detailed and perhaps replaced by a more advanced representation of canopy architecture.

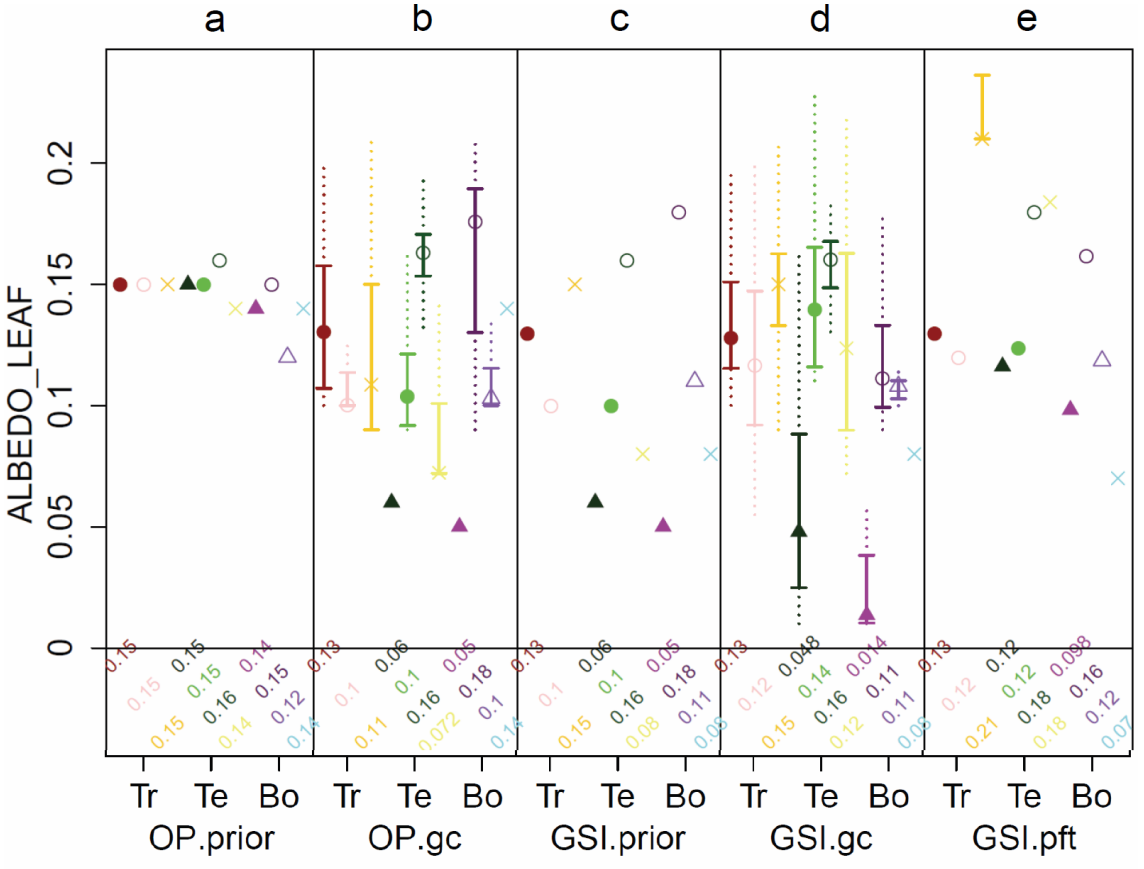

Figure S 14: Prior and optimized values for the parameter βleaf (leaf albedo) grouped by plant functional types and biomes.
The distribution of the parameter in the optimization experiments OP.gc and GSI.gc represents the spatial variability of the parameter from different grid cell-level optimization experiments. See Figure S 10 for a further explanation of this figure.

The highest values of the light extinction coefficient were found in the BoNS PFT. This was caused by an overestimation of tree mortality in years with simulated low productivity. Trees are killed in LPJmL as a result of negative net primary production which reduces FPC and results in a lower peak FAPAR in the following year. Having occurred more often in the simulated time period, it can explain why FAPAR is underestimated in some years. To remove these biases, the light extinction coefficient was optimized towards higher values in the BoNS PFT to reach FAPAR values that are closer to the observed FAPAR values after low-productivity years. However, such high values for the light extinction coefficient would overestimate tree cover and FAPAR under average conditions and when LPJmL is applied with dynamic vegetation. The approach to simulate tree mortality in LPJmL needs further improvement by, e.g., considering for example reserve carbon pools that helps the plants to endure low productivity conditions (Galvez et al., 2011).

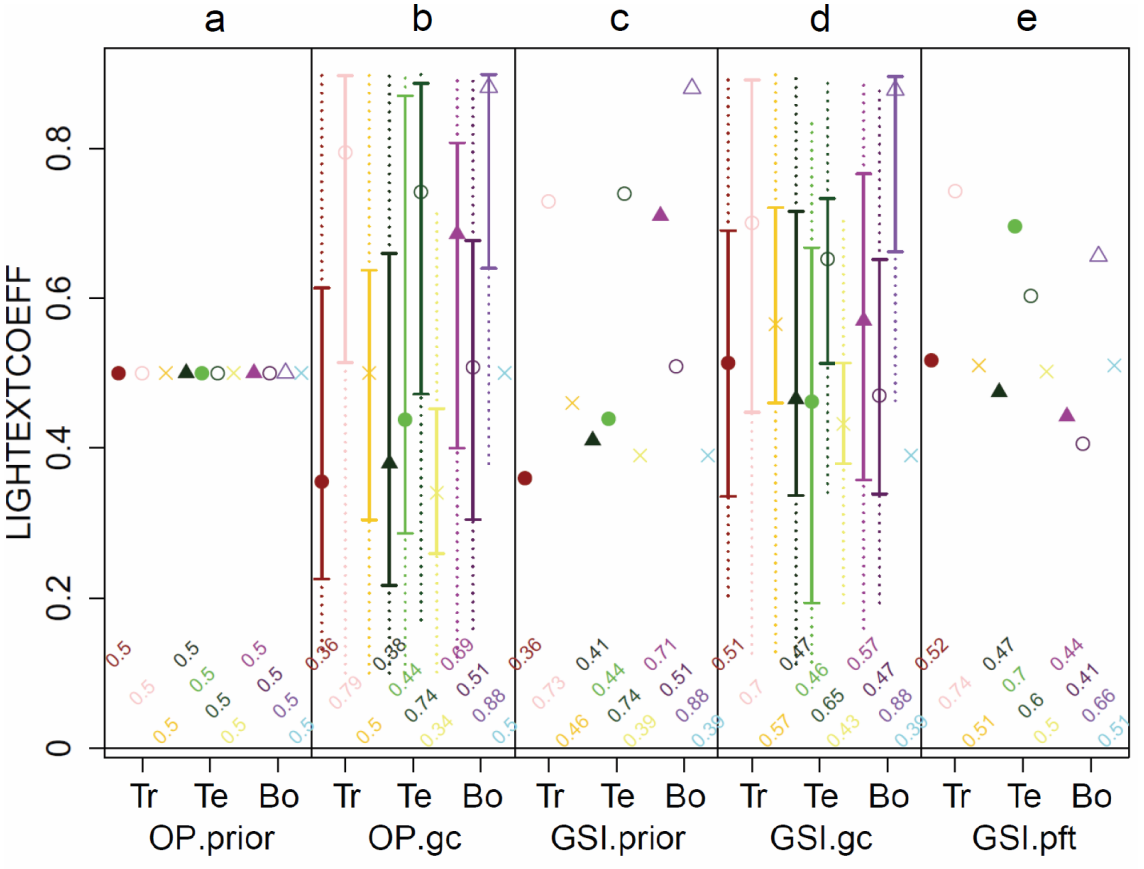

Figure S 15: Prior and optimized values for the parameter k (light extinction coefficient) grouped by plant functional types and biomes.
The distribution of the parameter in the optimization experiments OP.gc and GSI.gc represents the spatial variability of the parameter from different grid cell-level optimization experiments. See Figure S 10 for a further explanation of this figure.

We computed correlations between posterior parameter values for the four most important phenology parameters of LPJmL-GSI (TMIN_BASE, LIGHT_BASE, WATER_BASE and TMAX_BASE) (Figure S 16). Most correlations were low to moderate (maximum r = 0.69). Interestingly, the correlation between the TMIN_BASE and WATER_BASE parameters was low in PFTs that experience strong permafrost dynamics (BoNS r = 0.2, PoH r = -0.28). This indicates that the water and cold temperature limiting in boreal and arctic regions were only weakly correlated. Indeed, our results showed that water availability affected phenology mostly in early spring whereas cold temperature affected phenology during the entire year in boreal and arctic regions (Figure 9 of the main text). These results emphasize the ability to disentangle effects of seasonal air temperature and soil moisture on phenology in boreal and arctic regions.

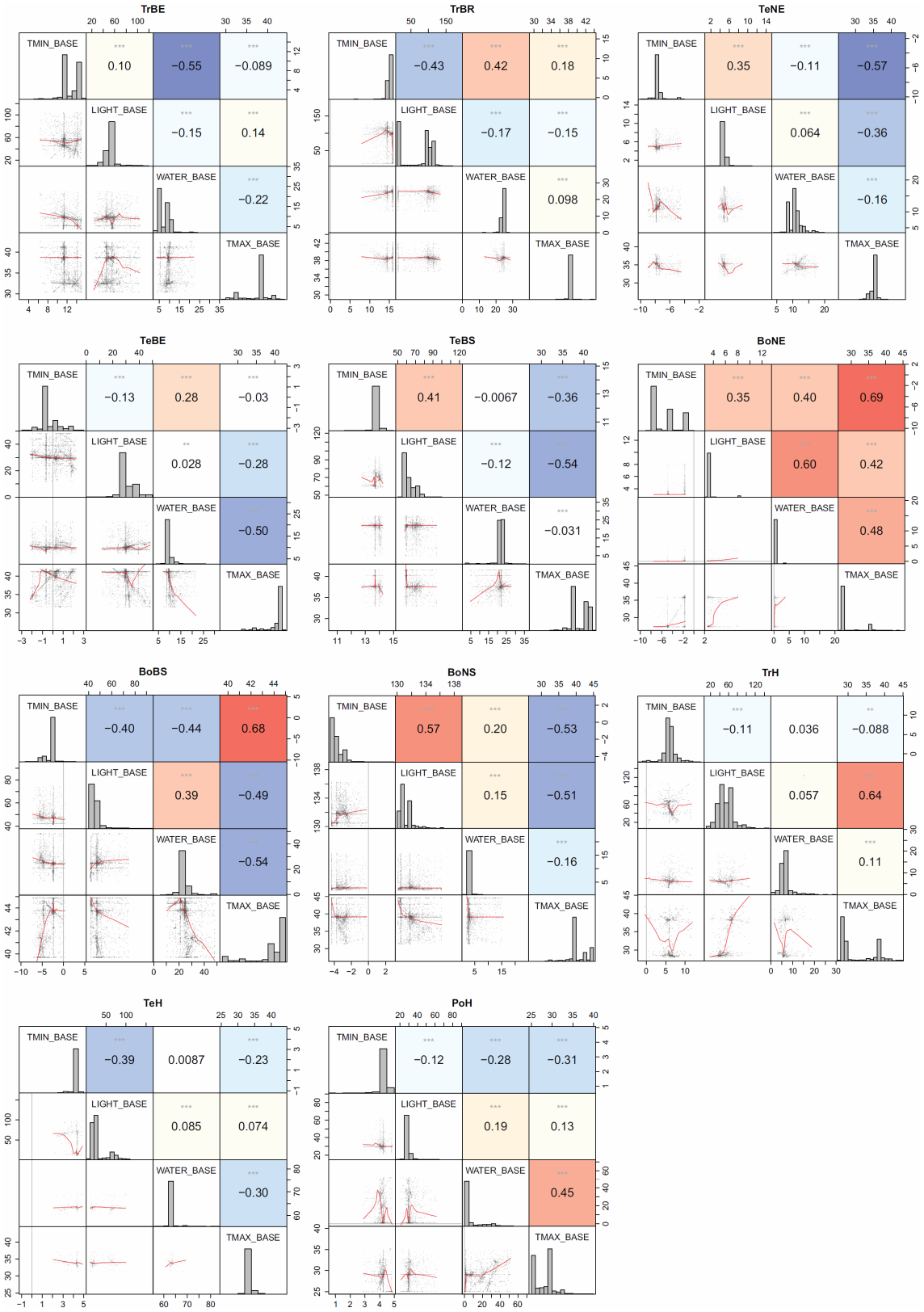

Figure S 16: Correlations between posterior parameters for the four limiting functions for phenology in LPJmL-GSI grouped per PFT.

Correlation matrices were computed based on all "best" individuals (i.e. parameters sets) from PFT-level optimization experiments (GSI.pft). "Best" individuals have an AIC difference of dAIC ≤ 2 in comparison to the individual with the lowest AIC, i.e. they are equally plausible. Numbers are Pearson correlation coefficients. Lines in the scatter plots are LOWESS smoothing lines (locally-weighted polynomial regression).

5 Global model evaluation

5.1 Supplementary results and discussion on carbon stocks and fluxes

LPJmL-GSI estimated global total carbon fluxes and stocks closer to data-oriented estimates than LPJmL-OP-prior and LPJmL-OP-gc (Table S 6). All three LPJmL model versions overestimated global total GPP although LPJmL-GSI was close to the upper uncertainty estimate of the data-oriented GPP estimate. Estimates of ecosystem respiration from LPJmL were clearly larger than the data-oriented estimates. Although LPJmL simulated global total fire carbon emissions within the magnitude of independent estimates (van der Werf et al., 2010), LPJmL-OP-gc had higher and LPJmL-GSI had lower fire carbon emissions despite the use of observed burnt areas in the SPITFIRE fire module. Data-oriented estimates of global total biomass have a large uncertainty. All three version of LPJmL were within these uncertainties. LPJmL-GSI estimated global total biomass the closest to the data-oriented estimates. From Table S 6 it is obvious that LPJmL with the model settings as in (Schaphoff et al., 2013) (i.e. without the BoNS and PoH PFTs and with simulated fire activity) resulted in global total GPP and ecosystem respiration that were even closer to the data-oriented estimates. This is mostly because LPJmL simulates larger burnt areas than seen in the observations and thus higher fire emissions but lower GPP and ecosystem respiration.

Although no information about temporal variations in GPP were used in optimization experiments, the mean seasonal cycle of GPP from LPJmL-GSI and LPJmL-OP-gc agreed better with the MTE data estimate than the mean seasonal GPP cycle from LPJmL-OP-prior especially in temperate and boreal PFTs and tropical grasslands (Figure S 17). GPP simulated by LPJmL-OP-prior increased too early and too fast in spring and decreased too late in autumn in TeNE, TeBS, BoNE, BoBS and TeH PFTs compared to the MTE estimate. These wrong dynamics improved after parameter optimization in both LPJmL-OP-gc and LPJmL-GSI. Additionally, LPJmL-GSI agreed better with the data estimate than LPJmL-OP-gc in TeNE, TeBS, TrH, PoH, TrML and TeML. These results demonstrate that the new GSI-based phenology model improved not only FAPAR seasonality but also GPP seasonality especially in temperate forests and in tropical to polar grasslands.

Table S 6: Global total carbon fluxes and stocks from data-oriented estimates and from LPJmL simulations.

LPJmL-OP-Standard and LPJmL-GSI-Standard are LPJmL model runs with settings as in (Schaphoff et al., 2013), i.e. without the use of the BoNS and PoH PFTs and with using simulated fires instead of prescribed observed burnt areas. Data sources: 1) (Beer et al., 2010; Jung et al., 2011), 2) (van der Werf et al., 2010), 3) (Carvalhais et al., 2014; Saatchi et al., 2011; Thurner et al., 2014), 4) (Carvalhais et al., 2014).

	Gross primary production (PgC a-1)	Ecosystem respiration (PgC a-1)	Fire carbon emissions (PgC a-1)	Biomass (PgC)	Soil organic carbon (PgC)
Data estimate	124.7[1)]	100-110[3)]	2.0[2)]	451.2[3)]	2460[4)]
Data lower uncertainty	110.7[1)]			208.8[3)]	1990[4)]
Data upper uncertainty	138.3[1)]			695.9[3)]	2984[4)]
LPJmL settings as in this study:					
LPJmL-OP-prior	161.3	150.7	1.93	674.1	2723
LPJmL-OP-gc	153.8	143.9	2.45	581.1	2503
LPJmL-GSI	145.8	141.4	1.65	546.4	2508
LPJmL settings as in Schaphoff et al. (2013):					
LPJmL-OP-Standard	138.9	125.8	3.48	597.8	2101
LPJmL-GSI-Standard	120.4	115.1	3.23	582.1	1392

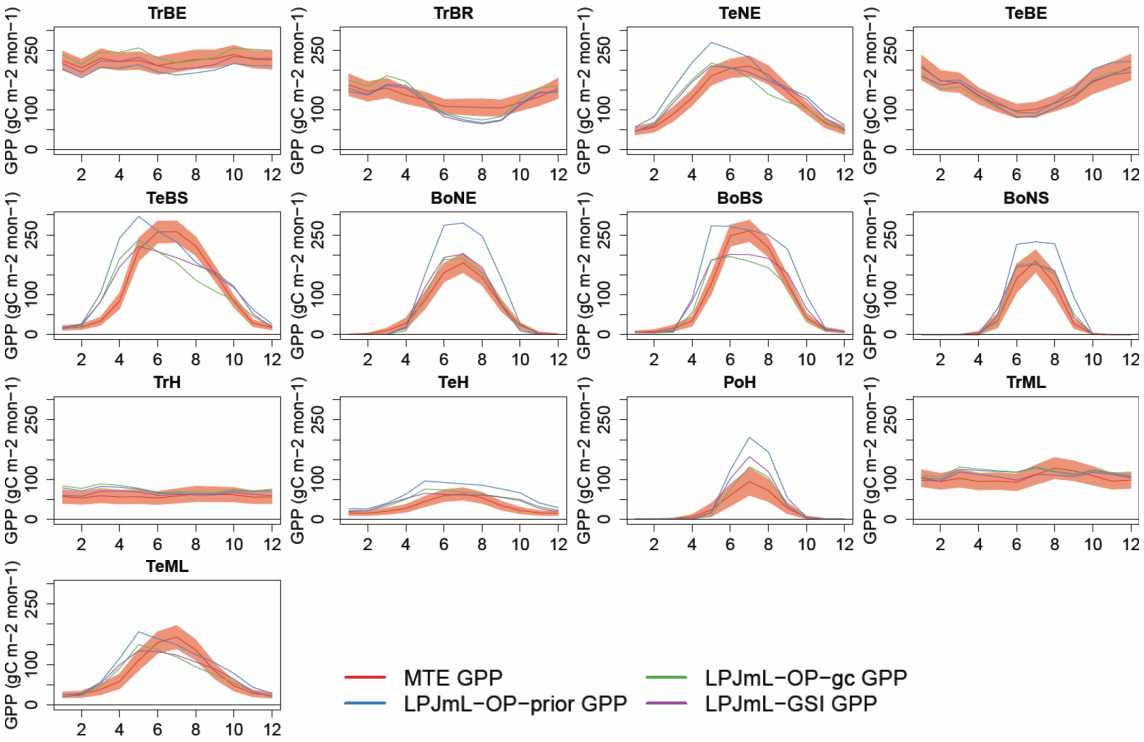

Figure S 17: Comparison of the mean seasonal GPP cycle (averaged over 1982-2011) from MTE and LPJmL spatially averaged for regions with the same dominant PFT.

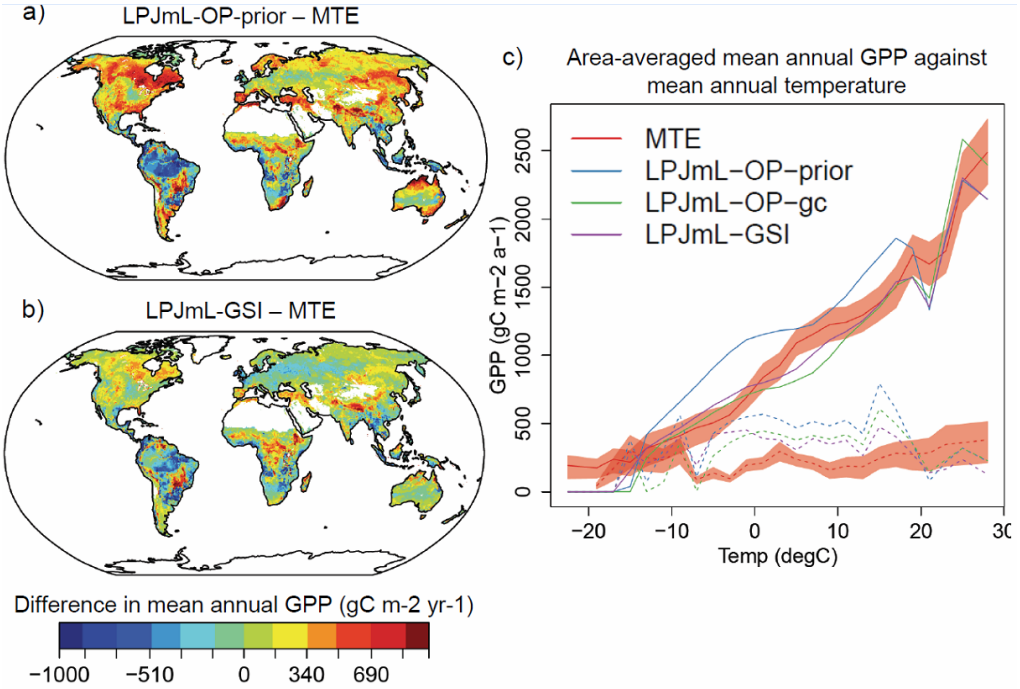

Figure S 18: Comparison of patterns of mean annual total gross primary production from LPJmL and the data-oriented MTE estimate for the period 1982-2011.

(a) Difference in mean annual total GPP between MTE and LPJmL-OP-prior. (b) Difference in mean annual total GPP between MTE and LPJmL-GSI. (c) Global spatial-averaged gradients of mean annual GPP against mean annual temperature. Dashed lines are dry areas with mean annual P/PET < 15 and solid lines are wet areas with mean annual P/PET >= 15. The red area represents the uncertainty of the data-oriented GPP estimate expressed as the inter-quartile range of the MTE ensemble.

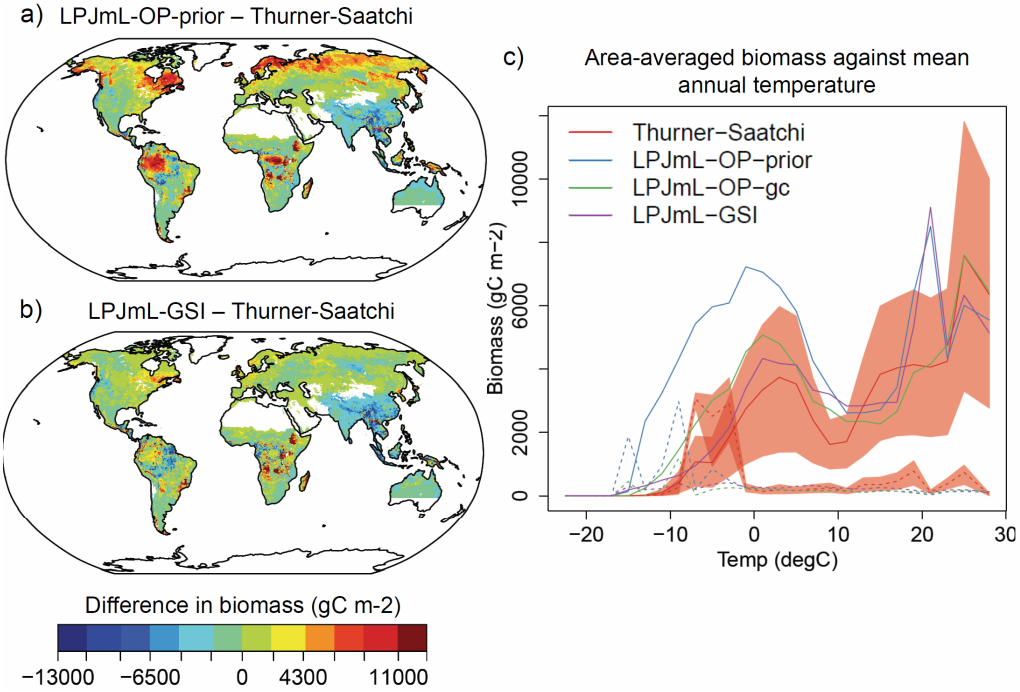

Figure S 19: Comparison of biomass from data-oriented estimates (Thurner and Saatchi datasets) and from LPJmL (averaged 2009-2011). See Figure S 18 for further explanations.

5.2 Supplementary figures on evapotranspiration

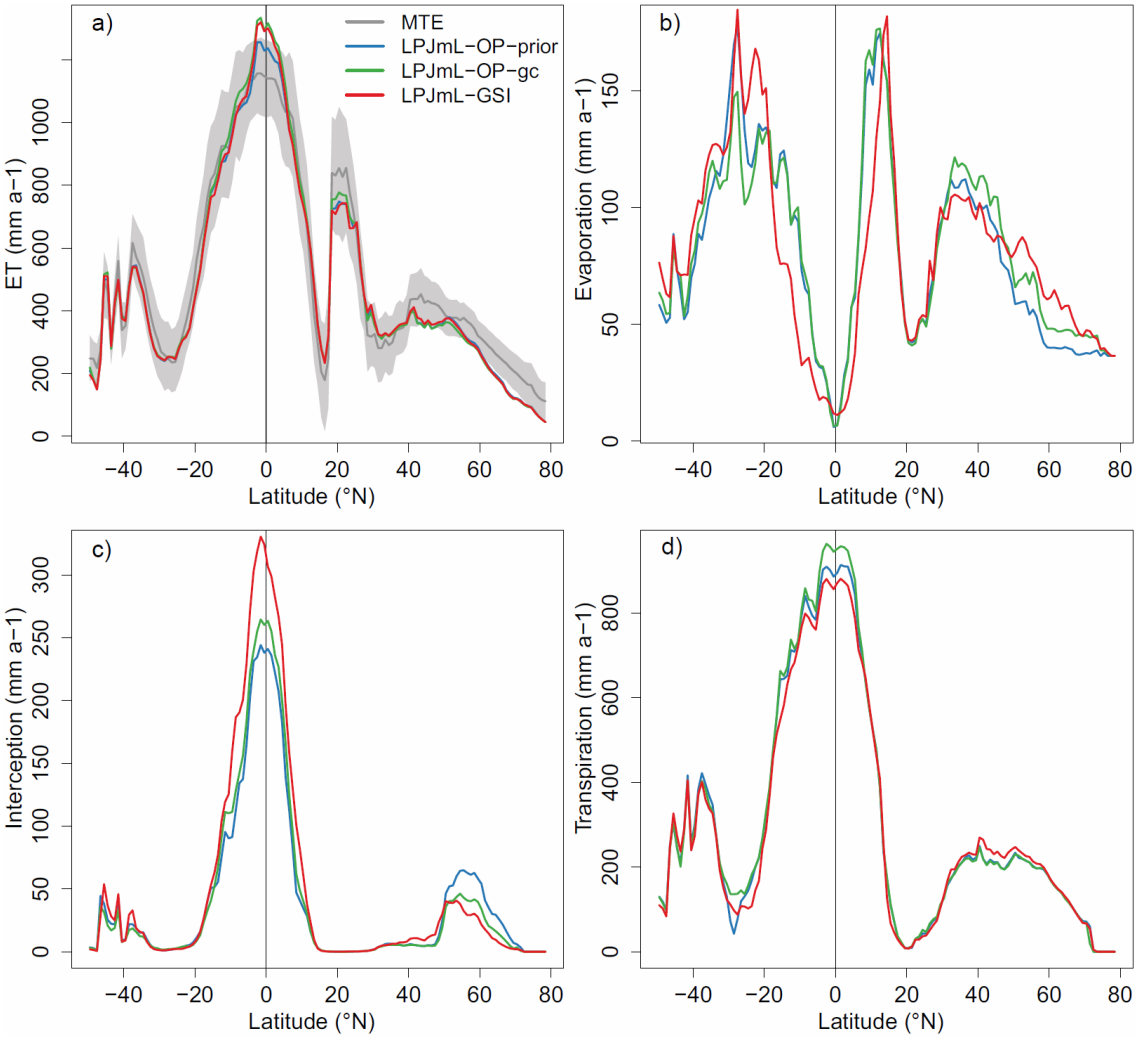

Figure S 20: Latitudinal gradients of evapotranspiration with its components.

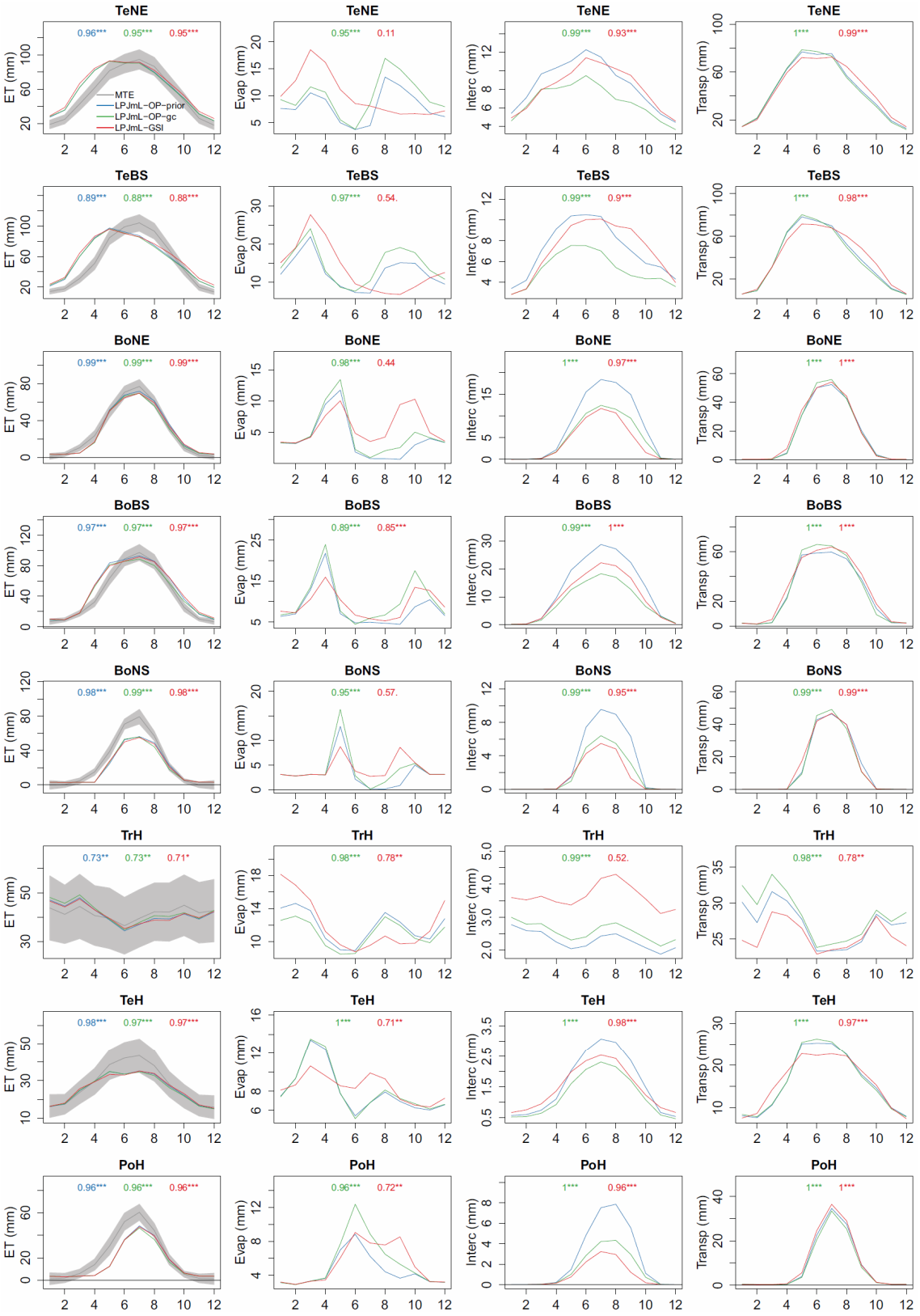

Figure S 21: Mean seasonal cycle (1982-2011) of ET, evaporation, interception and transpiration spatially averaged for PFTs.

Numbers on top of each plot are correlation coefficients between each LPJmL model run and MTE (for ET) and between LPJmL model runs and LPJmL-OP-prior, respectively. The significance of the correlation is indicated as point symbol: *** ($p \leq 0.001$), ** ($p \leq 0.01$), * ($p \leq 0.05$), . ($p \leq 0.1$).

5.3 Supplementary figures on evaluation of FAPAR

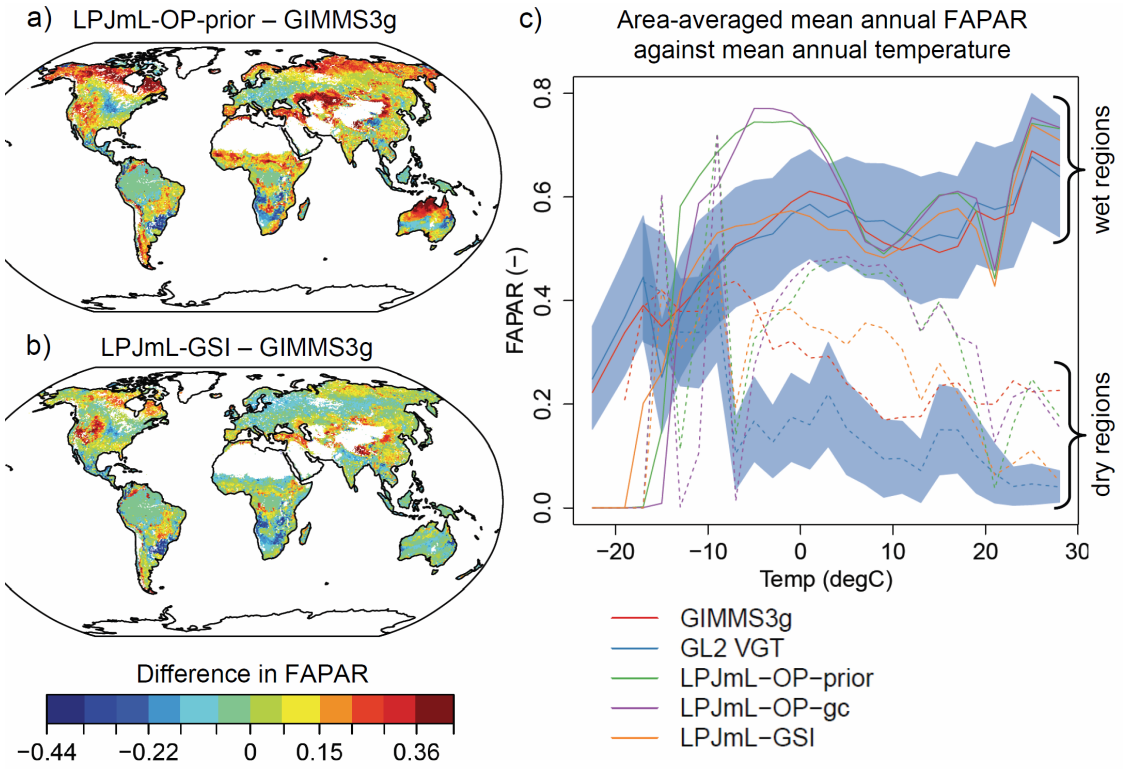

Figure S 22: Comparison of mean annual FAPAR from LPJmL and remote sensing datasets. See Figure S 18 for further explanations.

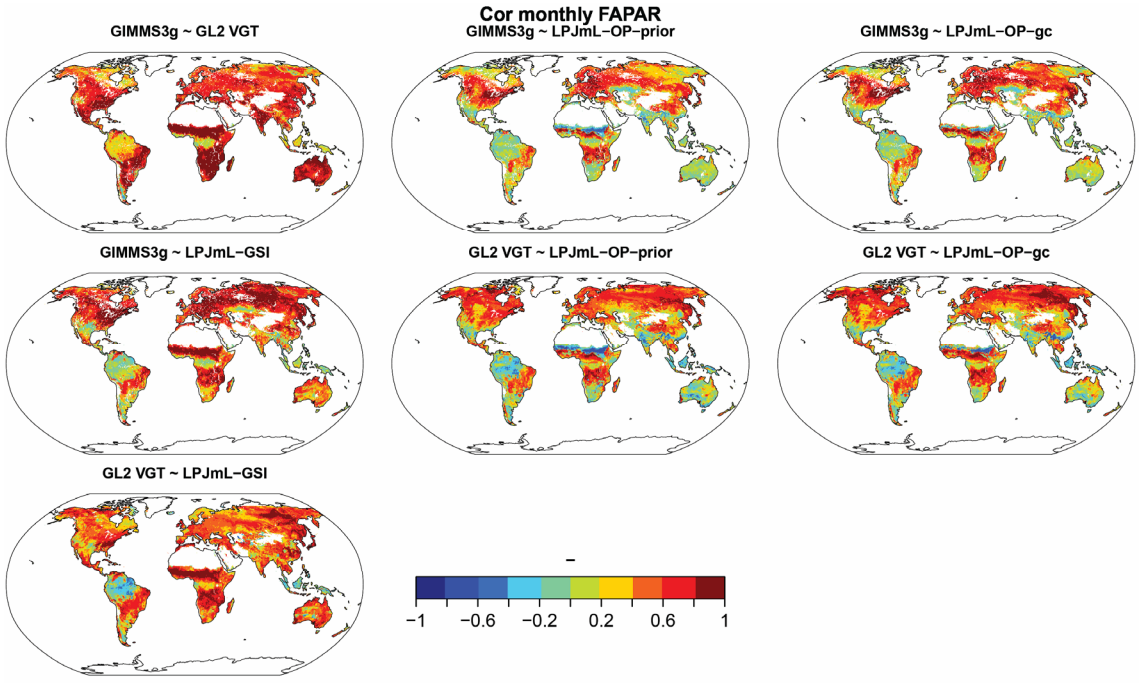

Figure S 23: Correlation coefficients between monthly FAPAR time series from GIMMS3g, GL2 VGT datasets and LPJmL model simulations.

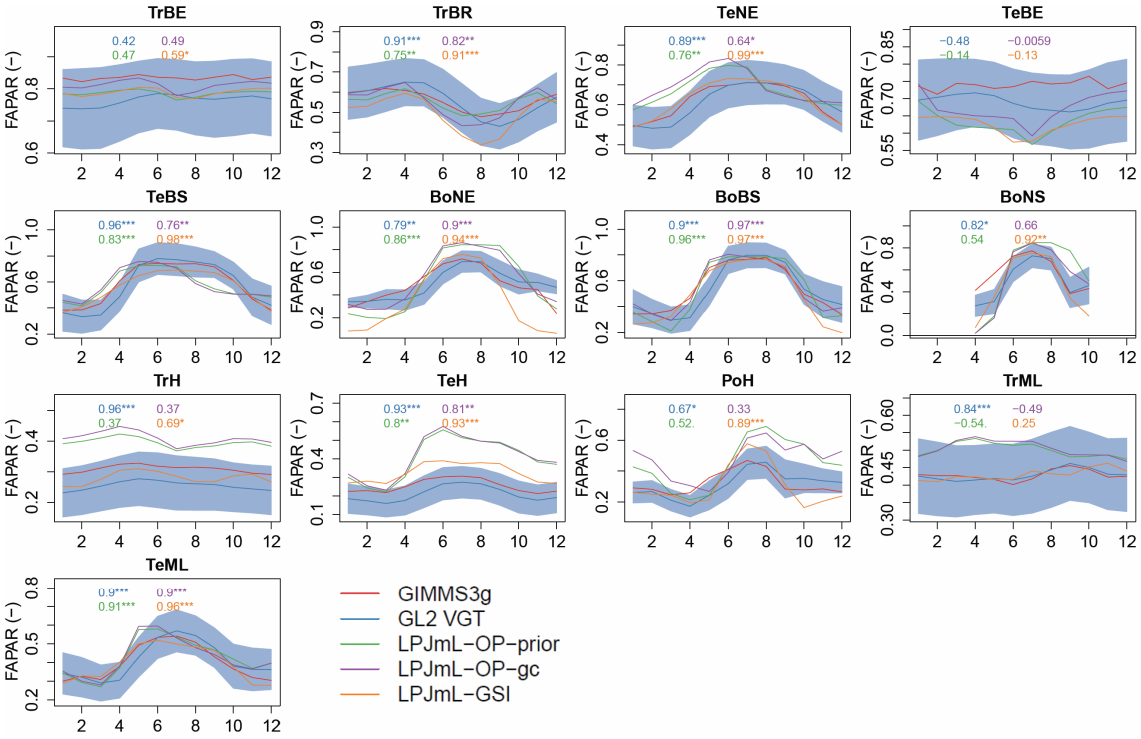

Figure S 24: Comparison of the mean seasonal FAPAR cycle from GIMMS3g, GL2 VGT and LPJmL spatially averaged for regions with the same dominant PFT.

The PFTs for which time series were averaged are shown in Figure 3. Numbers in the figures are correlation coefficients between GIMMS3g and the corresponding time series from GL2 VGT or from LPJmL simulations. The significance of the correlation is indicated as point symbol: *** ($p \leq 0.001$), ** ($p \leq 0.01$), * ($p \leq 0.05$), . ($p \leq 0.1$).

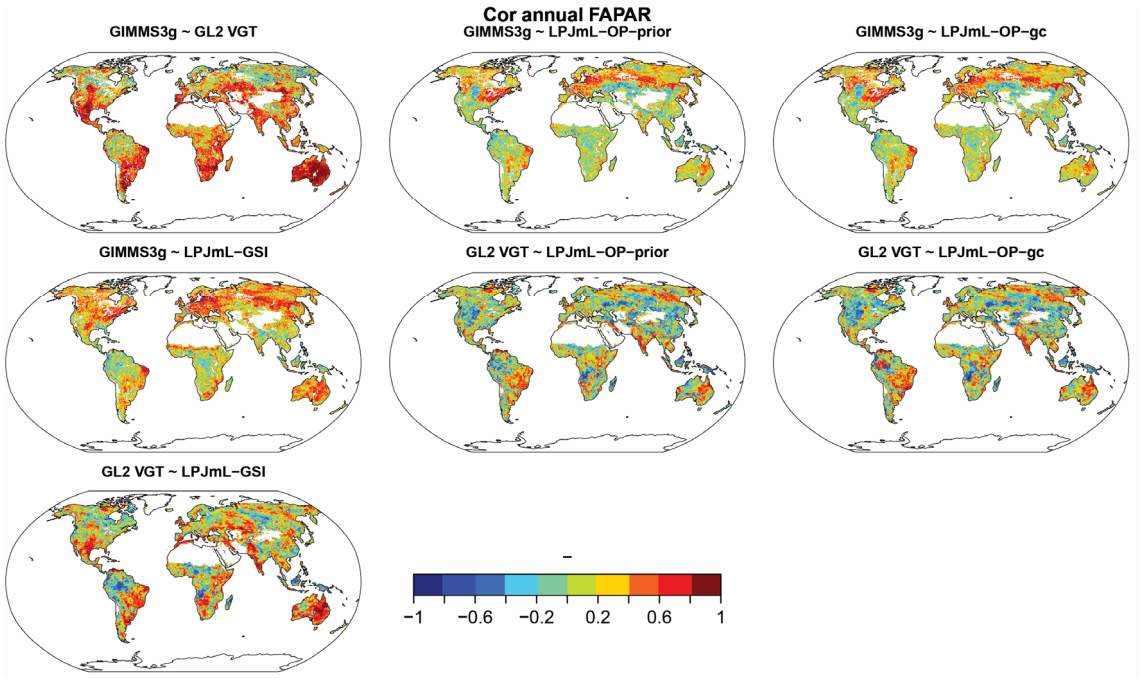

Figure S 25: Correlation coefficients between annual FAPAR time series from GIMMS3g, GL2 VGT datasets and LPJmL model simulations.

Mean annual FAPAR was averaged from monthly FAPAR values with air temperatures > 0°C.

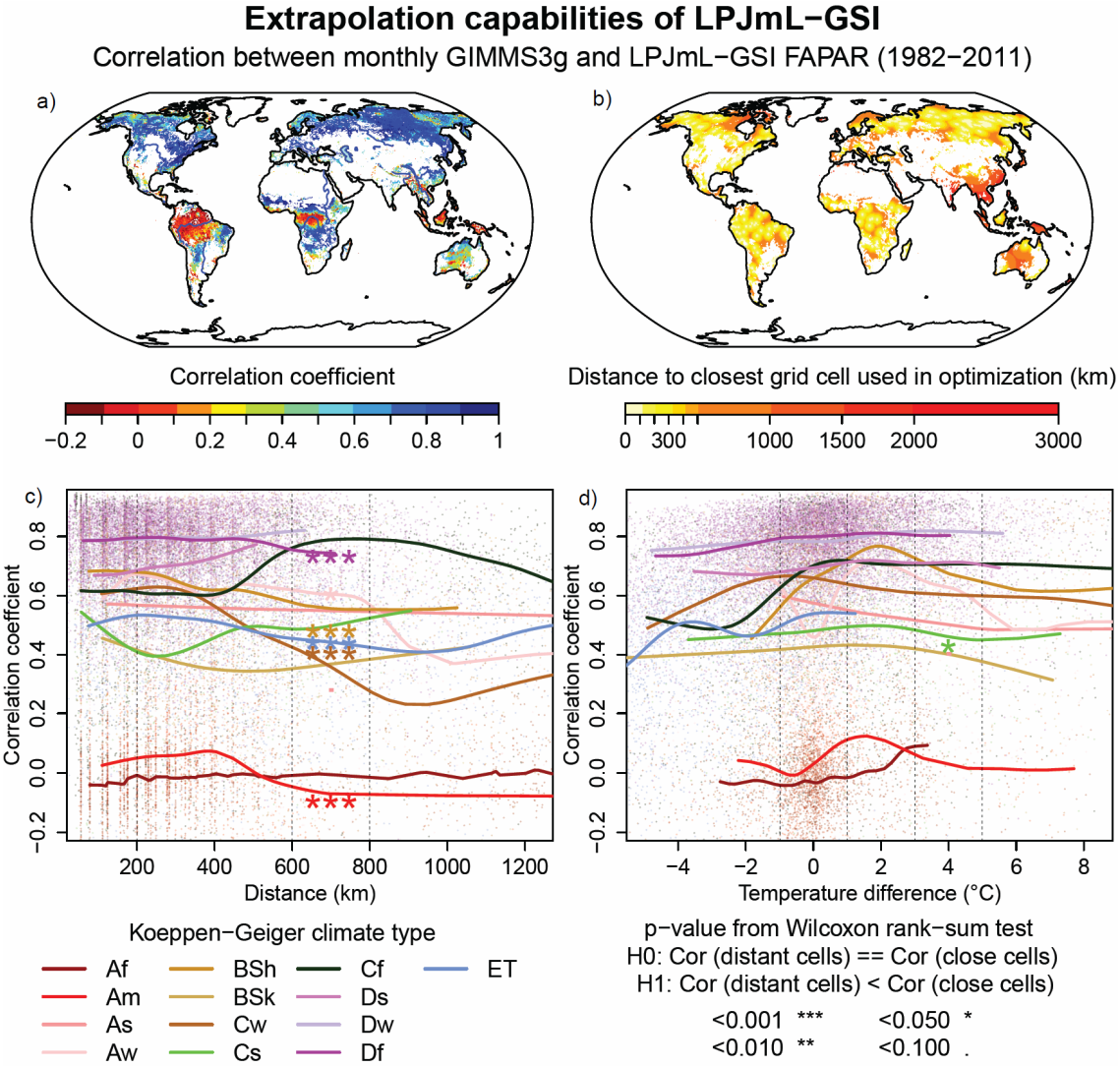

Figure S 26: Extrapolation capabilities of LPJmL-GSI in terms of monthly FAPAR dynamics. (a) Correlation coefficient between monthly FAPAR time series from LPJmL-GSI and GIMMS3g (1982-2011). Areas without vegetation, with more than 50% agricultural use, or without data are excluded (white). (b) The map shows the distance between each 0.5° grid cell and the closest grid cell that was used in a PFT-level optimization experiment of LPJmL-GSI (GSI.pft). (c) Scatterplot between the correlation coefficient from (a) and the distance from (b) coloured by the Köppen-Geiger climate type of each grid cell. Lines are smoothing splines fitted to the quantile 0.5 of the correlation coefficient for each climate type. Star symbols indicate the p-value of a Wilcoxon rank-sum test if the correlation coefficients of distant grid cells (between 600 and 800 km, indicated by vertical dashed lines) are significant lower than of close grid cells (≤ 200 km). (d) Scatterplot between the correlation coefficient from (a) and the difference in mean annual temperature between each grid cell and the corresponding closest grid cell. Star symbols indicate the p-value of a Wilcoxon rank-sum test if the correlation coefficients of warmer grid cells (between +3 and +5°C) are significant lower than of grid cells with similar temperature (± 1°C).

References

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F. I. and Papale, D.: Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate, Science, 329(5993), 834–838, doi:10.1126/science.1184984, 2010.

Broyden, C. G.: The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations, IMA J. Appl. Math., 6(1), 76–90, doi:10.1093/imamat/6.1.76, 1970.

Burnham, K. P. and Anderson, D. R.: Model selection and multimodel inference a practical information-theoretic approach, Springer, New York. [online] Available from: http://site.ebrary.com/id/10047705 (Accessed 17 December 2013), 2002.

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., Mu, M., Saatchi, S., Santoro, M., Thurner, M., Weber, U., Ahrens, B., Beer, C., Cescatti, A., Randerson, J. T. and Reichstein, M.: Global covariation of carbon turnover times with climate in terrestrial ecosystems, Nature, 514(7521), 213–217, doi:10.1038/nature13731, 2014.

Cescatti, A., Marcolla, B., Santhana Vannan, S. K., Pan, J. Y., Román, M. O., Yang, X., Ciais, P., Cook, R. B., Law, B. E., Matteucci, G., Migliavacca, M., Moors, E., Richardson, A. D., Seufert, G. and Schaaf, C. B.: Intercomparison of MODIS albedo retrievals and in situ measurements across the global FLUXNET network, Remote Sens. Environ., 121, 323–334, doi:10.1016/j.rse.2012.02.019, 2012.

Fletcher, R.: A new approach to variable metric algorithms, Comput. J., 13(3), 317–322, doi:10.1093/comjnl/13.3.317, 1970.

Galvez, D. A., Landhäusser, S. m. and Tyree, M. t.: Root carbon reserve dynamics in aspen seedlings: does simulated drought induce reserve limitation?, Tree Physiol., 31(3), 250–257, 2011.

Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W. and Sitch, S.: Terrestrial vegetation and water balance - hydrological evaluation of a dynamic global vegetation model, J. Hydrol., (286), 249–270, 2004.

Goldfarb, D.: A family of variable-metric methods derived by variational means, Math. Comput., 24(109), 23–26, doi:10.1090/S0025-5718-1970-0258249-6, 1970.

Horion, S., Fensholt, R., Tagesson, T. and Ehammer, A.: Using earth observation-based dry season NDVI trends for assessment of changes in tree cover in the Sahel, Int. J. Remote Sens., 35(7), 2493–2515, doi:10.1080/01431161.2014.883104, 2014.

Jung, M., Henkel, K., Herold, M. and Churkina, G.: Exploiting synergies of global land cover products for carbon cycle modeling, Remote Sens. Environ., 101, 534–553, doi:10.1016/j.rse.2006.01.020, 2006.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F. and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, J. Geophys. Res., 116, G00J07–G00J07, doi:10.1029/2010jg001566, 2011.

Kottek, M., Grieser, J., Beck, C., Rudolf, B. and Rubel, F.: World Map of the Köppen-Geiger climate classification updated, Meteorol. Z., 15(3), 259–263, doi:10.1127/0941-2948/2006/0130, 2006.

Mebane, W. R. and Sekhon, J. S.: Genetic Optimization Using Derivatives®: The rgenoud Package for R, J. Stat. Softw., 42(11), 2011.

Porada, P., Weber, B., Elbert, W., Pöschl, U. and Kleidon, A.: Estimating global carbon uptake by lichens and bryophytes with a process-based model, Biogeosciences, 10(11), 6989–7033, doi:10.5194/bg-10-6989-2013, 2013.

Poulter, B., Ciais, P., Hodson, E., Lischke, H., Maignan, F., Plummer, S. and Zimmermann, N. E.: Plant functional type mapping for earth system models, Geosci Model Dev, 4(4), 993–1010, doi:10.5194/gmd-4-993-2011, 2011.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M. and Morel, A.: Benchmark map of forest carbon stocks in tropical regions across three continents, Proc. Natl. Acad. Sci., 108(24), 9899–9904, doi:10.1073/pnas.1019576108, 2011.

Schaphoff, S., Heyder, U., Ostberg, S., Gerten, D., Heinke, J. and Lucht, W.: Contribution of permafrost soils to the global carbon budget, Environ. Res. Lett., 8(1), 014026, 2013.

Shanno, D. F.: Conditioning of quasi-Newton methods for function minimization, Math. Comput., 24(111), 647–656, doi:10.1090/S0025-5718-1970-0274029-X, 1970.

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

Strengers, B. J., Müller, C., Schaeffer, M., Haarsma, R. J., Severijns, C., Gerten, D., Schaphoff, S., van den Houdt, R. and Oostenrijk, R.: Assessing 20th century climate–vegetation feedbacks of land-use change and natural vegetation dynamics in a fully coupled vegetation–climate model, Int. J. Climatol., 30(13), 2055–2065, doi:10.1002/joc.2132, 2010.

Tao, X., Wang, D., Wu, D., Yan, B., Fan, W., Xu, X. and Yao, Y.: A model for instantaneous FAPAR retrieval: Theory and validation, in Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009, vol. 1, pp. I–144–I–147., 2009.

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S. R. and Schmullius, C.: Carbon stock and density of northern boreal and temperate forests, Glob. Ecol. Biogeogr., 23(3), 297–310, doi:10.1111/geb.12125, 2014.

Townshend, J., Carroll, M., Dimiceli, C., Sohlberg, R. a, Hansen, M. C. and DeFries, R. S.: Vegetation Continuous Fields MOD44B. 2000 Percent Tree Cover, Collection 5, [online] Available from: http://glcf.umd.edu/data/vcf/ (Accessed 15 January 2013), 2011.

Walter-Shea, E. A., Blad, B. L., Mesarch, M. A., Hays, C. J., Deering, D. W. and Eck, T. F.: Absorbed photosynthetically active radiation and sun-view geometry effects on remote sensing relationships, Remote Sens. Rev., 17(1-4), 89–102, doi:10.1080/02757259809532365, 1998.

Van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y. and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos Chem Phys, 10(23), 11707–11735, doi:10.5194/acp-10-11707-2010, 2010.