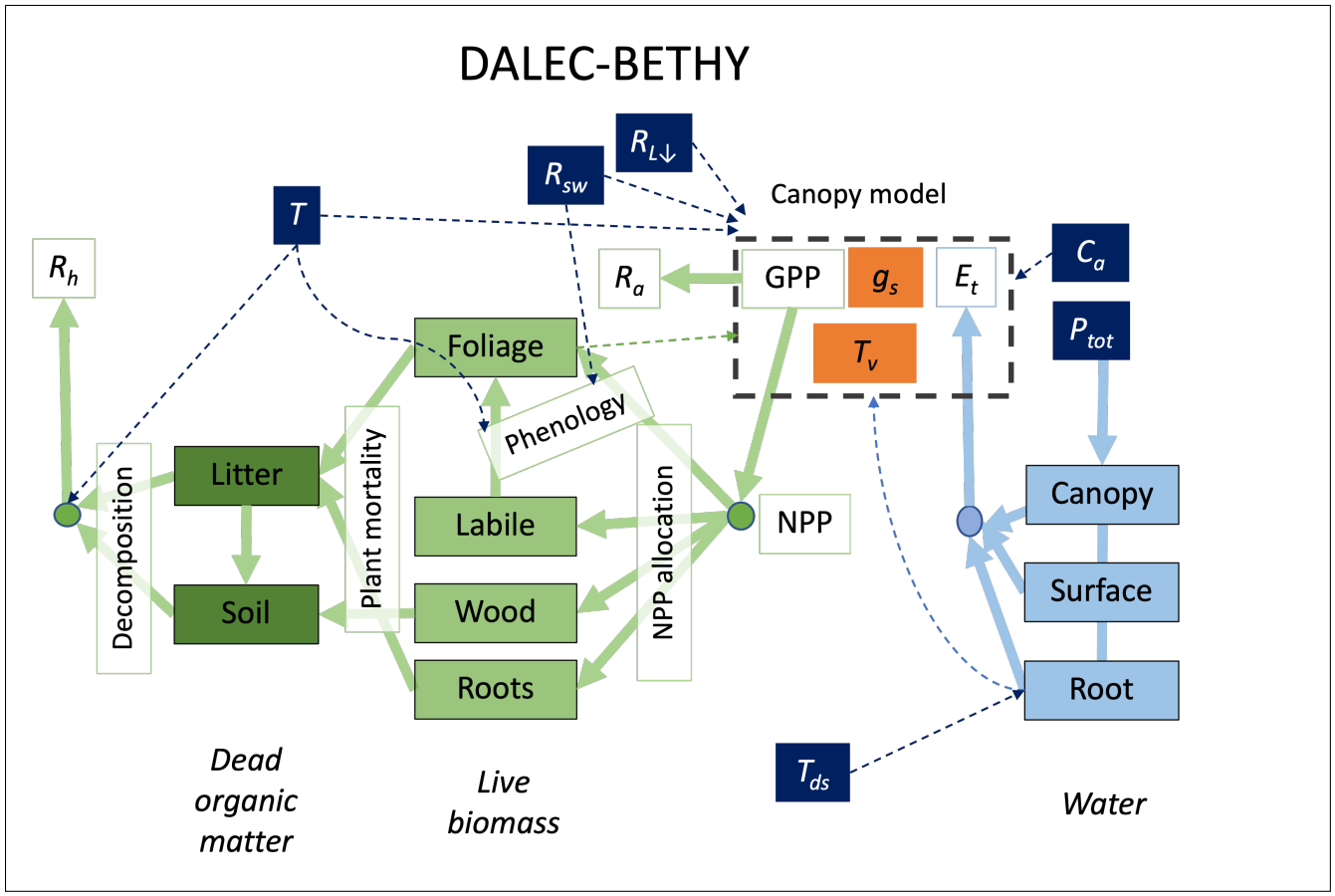
D&B and TCCAS User Manual Version 2.9



T. Kaminski, M. Voßbeck, W. Knorr, T. Quaife, M. Scholze, T.L. Smallman et al.

June 25, 2025

Version	Date	Details	Author
1	June 11, 2024	initial version	all authors
1.1	July 31, 2024	updating observation operator for L-VOD (Section 3.3) and repository and installation Section 8	WK, TK, MV
1.2	August 7, 2024	revised fractional cover concept and introduced hourly time step of water balance	WK, TK, MV
1.3	September 30, 2024	Appending D&B description (Sections 2.1, 2.2, 2.3)	WK
2	October 1, 2024	Updating sample output in Section 9	TK
2.1	November 11, 2024	updates of Section 5.4 and small edits	MV, WK
2.2	November 23, 2024	new version of b_e model, renamed Section to 'Transpiration'	WK
2.3	December 20, 2024	extension of variables list in Section A, instructions on how to prepare static forcing datasets, prior parameters for additional PFS	WK+TK+TLS
2.4	January 6, 2025	change of hourly to general variable names: gpph, neeh, rauth, resh to gpp, nee, raut, rhet	WK + MV
2.5	March 28, 2025	corrected time step and description of nitrogen scaling	WK
2.6	March 31, 2025	corrected confusion between W_s and W_r, $f_{soil}eff$ instead of W_s^{eff}	WK
2.7	April 1, 2025	corrected use of ρ_{bg} instead of ρ_s; adopted radiation docu to better match code	WK
2.8	April 2, 2025	clarified and aligned docu of electron transport, canopy evaporation; included heat and dark inhibition, as in model code	WK

Version	Date	Details	Author
2.9	April 3, 2025	corrected formula and completed docu for atm. vapour pressure; adjusted formulas for pot. soil evaporation, gas exchange to match model	WK

Contents

1 Executive summary

This manual describes the D&B terrestrial biosphere model and the variational Terrestrial Carbon Community Assimilation System (TCCAS) built around that model. It presents the underlying equations and concepts and their implementation. It describes the installation of the code, the operation of D&B and TCCDAS and presents a set of typical use cases with reference output. This document constitutes Deliverable 7 of the TCCAS study. An update of this document is planned as Deliverable 8.

2 D&B

The D&B model is comprised of three interconnected components: (i) photosynthesis and autotrophic respiration, (ii) energy and water balance, and (iii) carbon allocation and cycling, including heterotrophic respiration (Figure 1). The first comprises processes that lead to the uptake of CO_2 via plant photosynthetic activity (gross primary production, GPP), influenced by temperature, light absorption across the canopy, and stomatal control, as well as carbon loss from the respiration of live vegetation (R_a, autotrophic respiration). The remaining carbon flux is then passed as net primary production (NPP $=$ GPP - R_a) into the Carbon Allocation and Cycling component. The Energy and Water Balance determines the energy input to and output from the canopy in the form of radiative heat, latent and sensible heat transport, taking into account the water balance of the canopy and soil, as well as the rate of water uptake from the roots. Historically, components (i) and (ii) are based on the Biosphere Energy-Transfer HYdrology scheme (BETHY, Knorr, 2000), and component (iii) on the Data Assimilation Linked Ecosystem Carbon (DALEC Williams et al., 2005; Bloom and Williams, 2015) model.

Depending on the domain for which the model is set up, D&B in its standard configuration distinguishes 13 Plant Functional Types (PFTs) as shown in Table 1 ("Full List"). Typically, there is a set of parameter values that is assigned per PFT, i.e. there is a single parameter for all plants of a given PFT. The list of PFTs can also be extended, if there is sufficient available map data. Since it is not always possible to distinguish fractions of all PFTs listed in Table 1, a "Reduced List" might be used, also shown in Table 1. For more information regarding use of PFT maps, see Section 5.4. An extended list of 23 PFTs with parameters for BETHY only are also given by Knorr (1997), which includes long vs. short grass, C4, tree and irrigated crops.

All PFTs use the C3 photosynthetic pathway, except for PFT 10, for which a separate module for C4 photosynthesis is used (see Section B.2.1). Management of arable crops, at the moment assumed to be all C3, is represented by appropriate parameters for leaf onset and fall, as well as assumptions about a minimum level to which soil moisture is allowed to fall, as an approximation of irrigation (see Section B.3.5). If phenological type (evergreen or deciduous) is not available from land cover data and the reduced PFT list is used, it may in some cases be possible to simply assume that all trees are deciduous and all shrubs evergreen e.g. in Mediterranean semi-dry biomes.

In reality, however, there is often a continuum between trees and shrubs, and typical Mediterranean shrubs may, if left undisturbed, grow to tree size, or evergreen trees are not natural but planted (e.g. olive groves). Or, such as for the case of the transition between dry tropical woodland and humid tropical forest, there will be a continuum in the drought-deciduous character of the vegetation. In such cases, it will be necessary to make simplifying assumptions, such as ignoring evergreen temperate trees, or derive the degree of deciduousness from observational data. In this latter case, it may be necessary to use the fraction of deciduous vs. evergreen trees as a parameter at every grid cell to be adjusted during data assimilation, while using fixed parameters for each PFT. It is also possible to use a model with the mean climate as input that determines the fraction of deciduous vs. evergreen PFTs and adjust the parameters of this model during assimilation. Alternatively, it may be assumed that the plants' degree of deciduousness varies continuously, and not the PFT fractions. In this case, it will be necessary to adjust parameters relating to leaf growth,

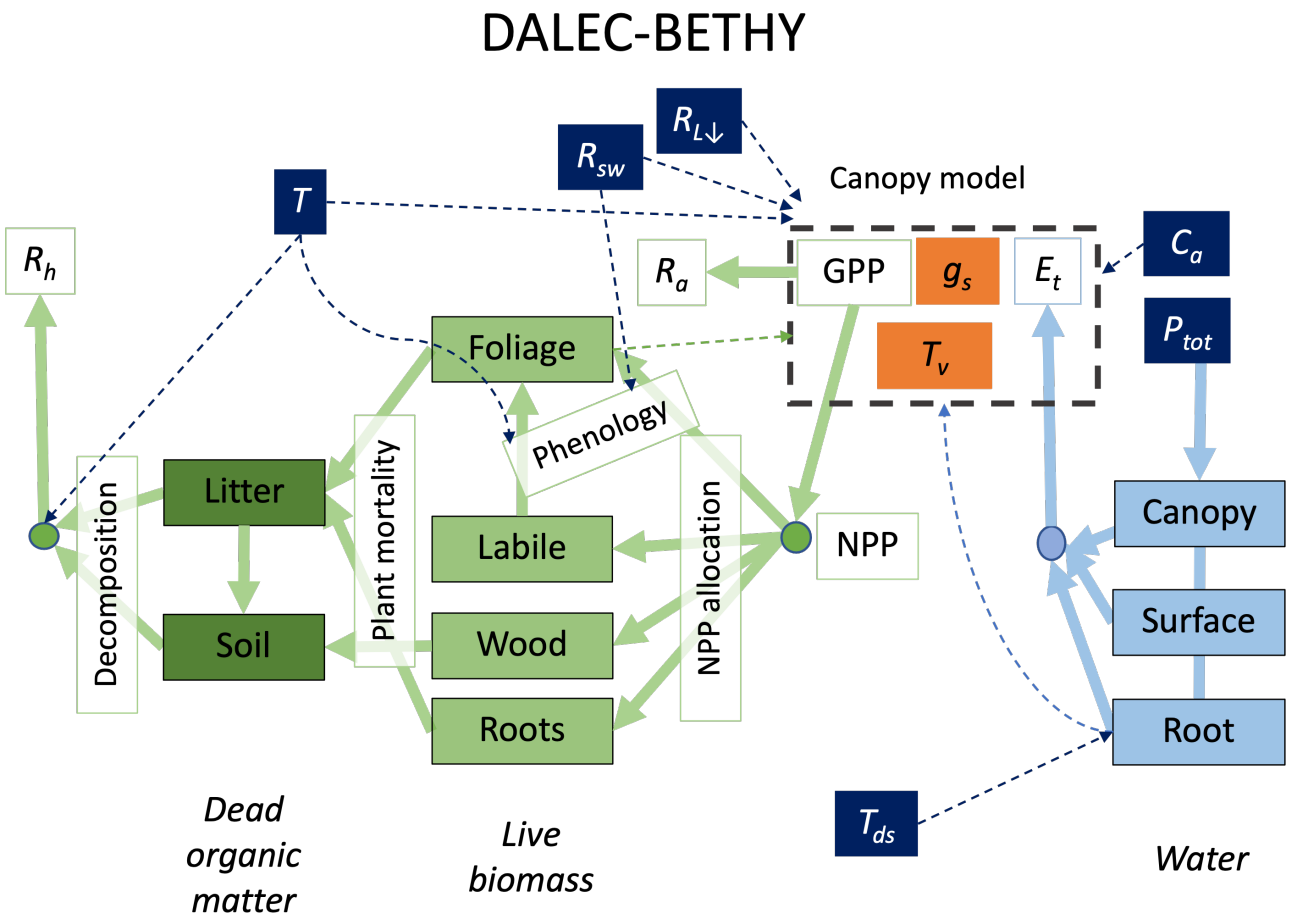


Figure 1: Structure of the DALEC-BETHY coupled model. State variables are in filled boxes; green for carbon, pale blue for water, orange for canopy states of stomatal conductance (g_s) and leaf temperature (T_v). Drivers are shown with white text in dark blue boxes (T: air temperature, T_{ds}: deep-soil temperature, R_{sw}: downwelling shortwave radiation, $R_{L\downarrow}$: downwelling longwave radiation; P_{tot}: total precipitation, C_a: CO_2 concentration in air). Fluxes are shown as solid arrows and annotated by open boxes; coloured green for C fluxes and pale blue for water. GPP is gross primary production; NPP is net primary production; R_a is autotrophic respiration; R_H is heterotrophic respiration; E_t is evapotranspiration. Dashed arrows show influences – for example T and R_{sw} influence the modelling of phenology.

senescence and leaf weight per area during assimilation.

The time step of the model is 1 hour. Depending on availability of driving data, the model can also be run at a different time step, as long as the time step is constant and the period of 24 hours can be expressed as an integer multiple of one time step.

The model simulates several PFTs in sub-grid tiles. Each PFT is simulated separately as if it would cover the full grid cell, with the results re-scaled by multiplying them with the grid-cell fraction

Table 1: Parameter combinations are available for the following plant functional types in D&B:

	Full List		Reduced List	
PFT#	Short name	Description	Short name	Description
1	TrEV	Tropical broadleaf evergreen tree	TrBT	Tropical broadleaf tree
2	TrDD	Tropical broadleaf deciduous tree		
3	TmEv	Temperate broadleaf evergreen tree		
4	TmSg	Temperate broadleaf deciduous tree	TmBT	Temperate broadleaf tree
5	EvCn	Evergreen coniferous tree	CnT	Coniferous tree
6	SgCn	Deciduous coniferous tree		
7	EShr	Evergreen shrub	Shr	Evergreen shrub
8	DShr	Deciduous shrub		
9	C3Gr	C3 grass	C3Gr	C3 grass
10	C4Gr	C4 grass	C4Gr	C4 grass
11	TunV	Tundra	TunV	Tundra
12	WetV	Wetland	WetV	Wetland
13	ArbC	Arable crop (C3)	ArbC	Arable crop (C3)

(f_i) occupied by the specific tile. Inter-PFT competition for light or water are neglected. A given grid cell can thus comprise several PFTs each with its specific grid-cell fraction. We also allow for the possibility of bare soil existing between the vegetation of each PFT. We thus define a "fractional cover" (f_c) for each tile, which is the fraction of the area of this tile that is covered by vegetation as opposed to bare soil. We assume that the fractional cover is the same for each sub-grid tile. Usually, the sum of the grid-cell fraction of all tiles (f_{tot}) equals one. It is, however, also possible that this number is less than one, in which case a part of the overall model grid cell is excluded from the simulation. This can for example be applied to account for water surfaces or built environments. (The fraction of the total grid cell covered by vegetation is thus $f_c \times f_{tot}$.)

2.1 Photosynthesis and Autotrophic Respiration

The photosynthesis and autotrophic respiration part of BETHY comprises the following carbon fluxes (see Figure 2): net rate diffusion of CO_2 through stomatal pores, as equal to net rate of photosynthesis (denoted as GPP, or gross primary productivity, for either C3 or C4 plants), and plant or "autotrophic" respiration, R_a, as the sum of maintenance respiration (as a function of biomass and temperature) plus growth respiration (being proportional to NPP, net primary productivity, with NPP $=$ GPP – R_a). It computes the following radiation fluxes within a two-flux scheme, with (by default) three canopy layers: direct incoming PAR (photosynthetically active radiation), diffuse downwelling PAR, diffuse upwelling PAR.

The C3 photosynthesis module (Section B.2.1) is based on the biochemical model of photosynthesis by Farquhar et al. (1980). It determines light absorption, light-limited electron transport, CO_2-limited carboxylation rate and the resulting gas exchange of CO_2. Light absorption in the photosynthetically active spectrum is calculated within a two-flux approximation (Section B.2.1), following Sellers (1985). D&B divides the canopy into several vertical layers of equal LAI, the sum of which constitutes the total canopy LAI. In the standard configuration, the number of layers is three.

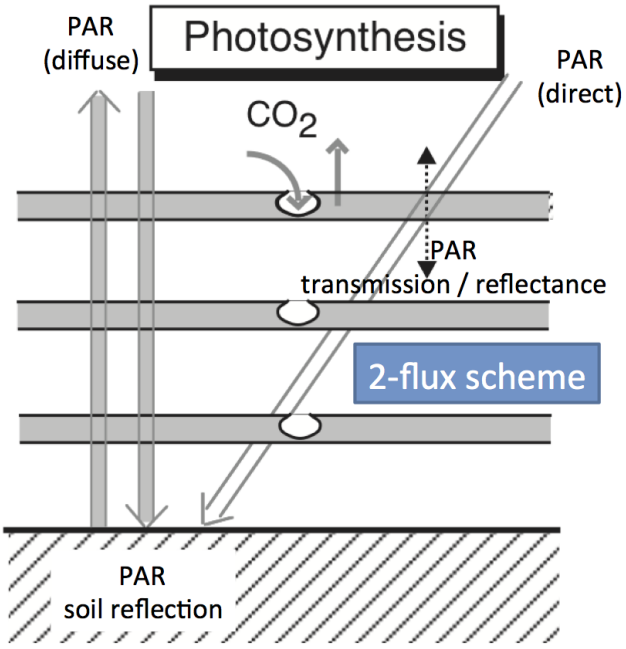


Figure 2: Photosynthesis and autotrophic respiration are computed within a two-flux scheme for PAR (photosynthetically active radiation) absorption within the plant canopy, with three layers in default mode.

The amount of light absorbed and thus available for photosynthesis is dependent on LAI, statistical leaf orientation (assumed to be isotropic) and leaf single-scattering albedo. Photosynthetic capacity decreases from top to bottom of the canopy, assuming that decreasing levels of daily-average solar radiation drives decreases in leaf nitrogen content and maximum rates of light-limited photosynthesis.

The photosynthesis module further divides GPP into NPP and R_a (Section B.2.2) R_a is modelled as the sum of maintenance and growth respiration (Knorr, 1997). Maintenance respiration is proportional to photosynthetic capacity, while growth respiration is proportional to NPP. Growth respiration is zero when NPP is negative. It is calculated via solving an implicit equation of NPP as a function of GPP and R_a as a function of NPP itself. Both continually increase with temperature. Where NPP is estimated to be negative, i.e. maintenance respiration is greater than GPP, the NPP passed to the C allocation and cycling component is assumed zero.

The rate of photosynthesis is first computed under standard conditions without limitation by water availability. This potential photosynthesis rate is translated into an equivalent stomatal conductance, i.e. the stomatal conductance necessary to provide the flow of CO_2 to the leaf interior. This value for stomatal conductance without water limitation is reduced depending on the vapour pressure deficit of the surrounding air, and available soil moisture. This modified stomatal conductance, or "actual stomatal conductance", then determines actual photosynthesis and, using information from the Energy and Water Balance component, the rate of transpiration.

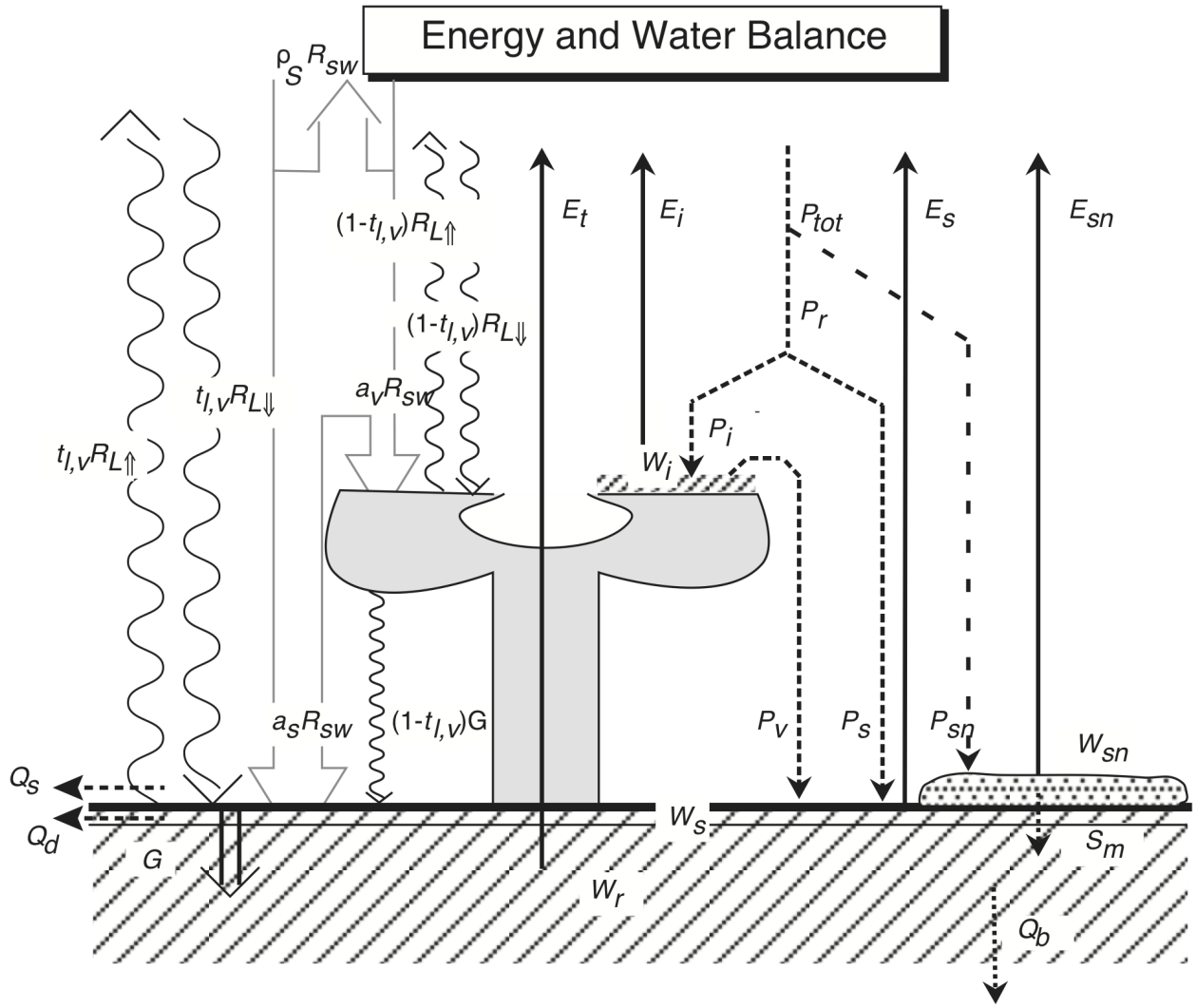


Figure 3: Energy and water balance of the D&B model with symbols for fluxes (normal) and reservoirs (italics). a_s: soil absorption of shortwave radiation, a_v: canopy absorption of shortwave radiation, E_i: intercepted-water (canopy) evaporation, E_s: soil evaporation; E_{sn}: snow evaporation, E_t: transpiration, G: ground heat flux, P_i: intercepted rainfall; P_r: rainfall; P_s rainfall on soil; , P_{tot}: total precipitation; P_v: throughfall, Q_b: baseflow, Q_d: horizontal drainage, Q_s: surface runoff, $R_{L,\uparrow}$: upwelling longwave radiation, $R_{L,\downarrow}$: downwelling longwave radiation, R_{sw}: downwelling shortwave radiation, $t_{l,v}$: longwave canopy transmission, W_i: intercepted water amount, W_r: root-zone soil moisture, W_s: surface-layer soil moisture, W_{sn}: snow amount, ρ_S: surface reflectance.

2.2 Energy and Water Balance

The energy and water balance of D&B is computed separately for the soil and the canopy -– at an hourly time step by default. A summary of fluxes and pools in shown in Figure 3). Fluxes are short

wave (broad arrows) and long wave (curvy arrows) radiation, rainfall and runoff (dotted arrows), snowfall (dashed arrow), evaporative fluxes (sold arrows) and ground heat flux (bold wide arrow). Water pools (striped areas) are two soil moisture layers, one for the root zone and one overlapping for a thin surface layer, intercepted canopy water, and snow. Two further state variables used by the energy and water balance are canopy temperature and stomatal conductance per layer, the latter computed in exchange with the photosynthesis part (see Figure 2).

The Energy and Water Balance component requires the rate of transpiration from the photosynthesis module, due to the tight coupling between water loss through transpiration and CO_2 uptake by leaves. Transpiration (Section B.3.4) is subsequently combined with other evaporative fluxes, namely of intercepted water (evaporation from intercepted skin reservoir, Figure 3, Section B.3.3), and from the soil surface (soil evaporation, Section B.3.5), including snow sublimation (Section B.3.7), to arrive at total evapotranspiration and latent heat flux. Latent heat flux is constrained by the available net radiative energy input, which the model computes separately for the vegetation canopy and the soil (Section B.3.6). Sensible heat flux is computed from the assumption of energy closure from net radiation, latent and soil heat flux. The model uses incoming shortwave (solar) and longwave (thermal) radiation as input, but simulates both outgoing radiation components internally, using information on the albedo of the soil background and vegetation (Section B.3.8).

Soil evaporation proceeds at the equilibrium rate driven by the soil net radiation from a thin surface layer, with depth of 4 cm unless that value exceeds the depth of the soil. This corresponds to a typical depth for which microwave remote sensing can provide soil moisture estimates (Babaeian et al., 2019). Evapotranspiration from the canopy happens as either canopy evaporation from leaf surfaces at the equilibrium rate (E_i, Jarvis and McNaughton, 1986), or as transpiration through leaf pores (E_t). Precipitation enters either the leaf interception pool (W_i), the soil pool (W_r and W_s, see Section B.3.1), or is added to the snow pool. Precipitation happens as either snow (P_{sn}, Section B.3.7), or rainfall (P_r), partitioned into a canopy-interception part (E_i), soil infiltration (P_s), and surface runoff (Q_s). Soil water drains as sub-surface runoff (Q_d), or base flow (Q_b). Infiltration into the soil (Section B.3.2), runoff, drainage and baseflow (Section B.3.5) are simulated following a new implementation of the variable infiltration capacity approach (Wood et al., 1992), where a thin surface layer has been added to a single root-zone layer, (Scholze et al., 2016). The surface soil moisture layer overlaps with the root zone layer, so that the near-surface soil water pool (W_s) forms part of the root-zone soil water pool (W_r, Figure 3). The former has a nominal depth of 4 cm, the latter a depth equal to a PFT-specific root depth, d_r (Table 11). Both depths are limited by depth to bedrock. Soil water exiting the root zone downwards is considered subsurface drainage, while there is no upward water movement from below the root zone. The root zone soil moisture pool contains all simulated soil water, while the surface layer is added in order to be able to calculate soil evaporation, as well as for diagnostic purposes taking account of the impact of surface soil moisture on microwave remote sensing.

Once per day around the time of maximum evaporative demand, assumed to be at the hourly time step closest to 13:00 hours local solar time (Knorr, 1997), the parameters determining actual stomatal conductance are reset to reflect soil water status. To do this, transpiration is simulated as the minimum of a root water supply rate, which increases linearly as soon as soil water exceeds the

permanent wilting point, and the demand for transpiration. This rate of demand is determined by the potential rate of photosynthesis without water stress. Potential photosynthesis is assumed as the rate at a fixed ratio of leaf to atmospheric CO_2 content (0.87 for C3 and 0.67 for C4 photosynthesis). A parameter of the stomatal model is now determined in such a way that stomatal closure limits the rate of transpiration to exactly the rate of root water supply. This hypothetical parameter value is then denoted b_e^{opt}. The actual parameter value, b_e, is not set immediately to b_e^{opt}, but programmed to tend towards that value at every model time step, until a new value for b_e^{opt} is determined. The temporal shape of the function letting b_e relax towards b_e^{opt} is that of an exponential asymptotic, with a time constant τ_b.

Finally, the surface reflectance, or background albedo (ρ_S) is affected by soil brightness, surface soil water content, and the presence of snow. Vegetation albedo as a function of absorption in the photosynthetically active spectrum, computed in the photosynthesis module, and snow albedo is modelled depending on snow age (Loth and Graf, 1996; Knorr, 1997).

2.3 Carbon Allocation and Cycling

The carbon cycle in D&B is expressed as a series of six equations describing the dynamics of six carbon pools. D&B employs an hourly time step for allocation, the same as the time step used by the photosynthesis module. There are three structural live C pools, for foliage (*fol*), wood (*wd*), and fine roots (*fr*), and a labile (*lab*) pool which supports seasonal foliage expansion. There are two dead organic matter pools, namely litter (*lit*) and soil organic matter (*SOM*). The state equations describe the change over time in pool sizes on the basis of C fluxes in and out. NPP is the output from the photosynthesis module into the C allocation and cycling module. NPP is then allocated to each of the four live biomass pools based on parametrisable fixed fractions.

The labile C pool in D&B represents the stored C used to initiate accelerated leaf development at the start of the growing season (Section B.4.1). The phenology scheme parameterises timing of local bud burst via allocation to leaves from the labile pool based on calibrated climate sensitivity. Leaf development is thus dependent on 1) the allocation of labile carbon, replenished from NPP, to the leaf carbon pool and 2) additional direct allocation from NPP. The leaf area index is determined by the conversion of leaf carbon pool size to leaf area using a parameterisable estimate of the leaf mass per area.

The fine root (*fr*) and wood pools (*wd*) are determined by first-order differential equations, using a decay constant combined with an fractional input, again from NPP (Section B.4.2). Thus, a fraction of the fine root pool replenishes the litter pool, added by strongly periodic inputs linked to leaf senescence, while wood directly feeds SOM. The litter pool decays either to CO_2 via heterotrophic respiration, or is transferred to the SOM pool. Mineralisation of both SOM and litter C pools by heterotrophic respiration thus results in further CO_2 fluxes. Total ecosystem respiration (TER) is determined by the sum of autotrophic growth and maintenance respiration, and mineralisation of dead organic matter (*lit* or *som*), creating a flux of heterotrophic respiration. Following the procedure used for DALEC, the prior parameters of the Carbon Allocation and Cycling are set through a regional-scale calibration procedure, as described in Section B.4.3.

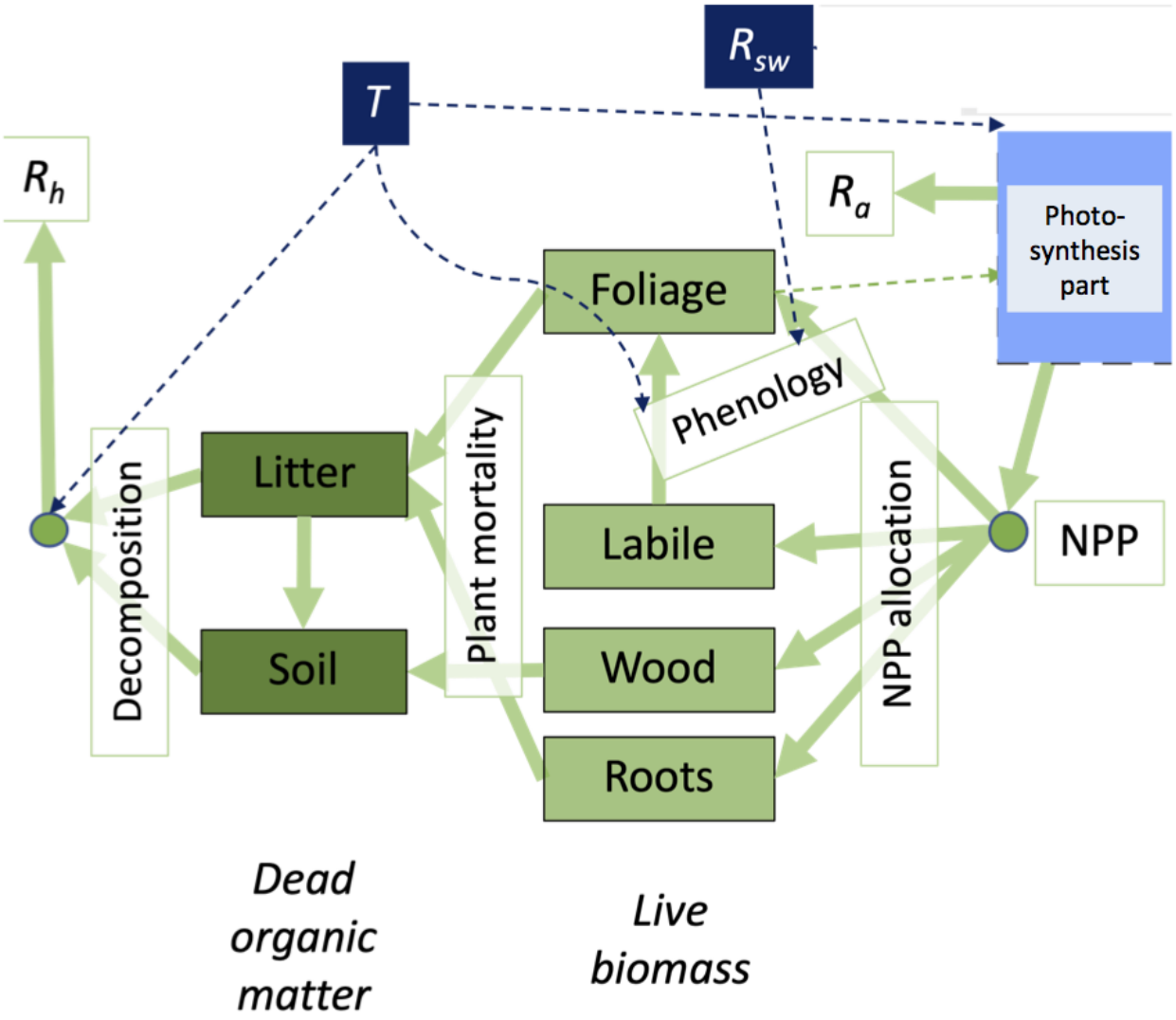


Figure 4: The carbon allocation and cycling part of D&B, taking NPP and autotrophic respiration (R_a) from the photosynthesis part as well as air temperature (T) and short wave incoming radiation (R_{sw}) from input data. Output is heterotrophic respiration (R_h) as well as LAI (leaf area index) from the foliage carbon pool.

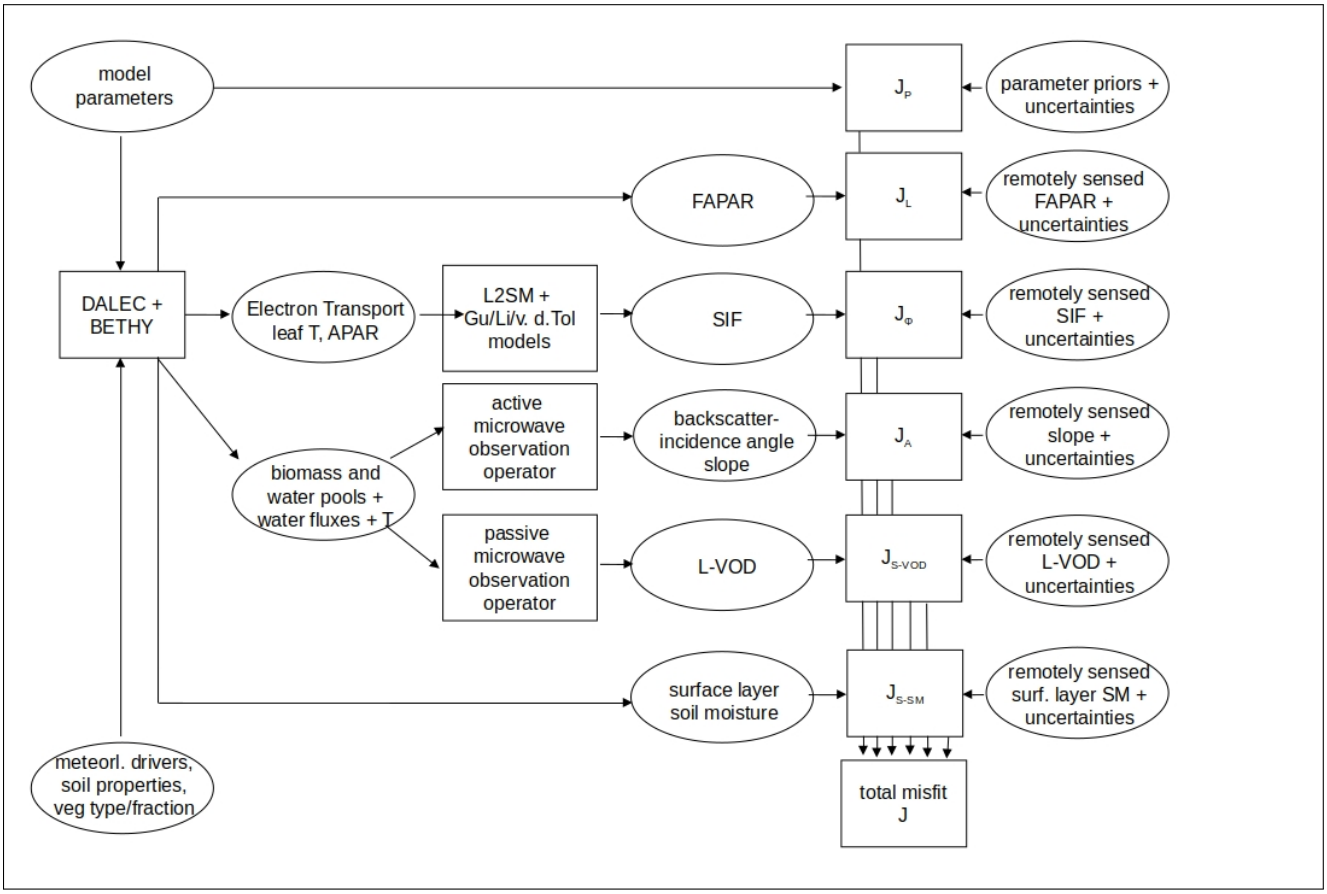


Figure 5: A flow diagram showing the connection between the biogeochemical model and the data streams to be assimilated. The D&B model is forced with meteorological data, and takes soil parameter inputs. Key outputs of the model – the canopy state and the vegetation state – are used as inputs to observation operators. These observation operators generate observation equivalents for five different remote sensing data streams. The data assimilation system determines the parameters of the model and the observation operators that minimise the misfit between observation equivalents and observations at pixel scale over the period of monitoring. fluxes (normal) and reservoirs (italics).

3 Observation Operators

The task of an observation operator is to simulate the equivalent of an observation from the model's state variables. This includes the simulation of the variable that is retrieved at the time when it was observed and over the footprint of the observations (Kaminski and Mathieu, 2017). In the following, we will present the simulation of five data streams, namely FAPAR, SIF, L-band VOD, ASCAT slope, and near-surface soil moisture. Of the five data streams, FAPAR and surface soil moisture are internally calculated. Figure 5 shows the flow of information in the forward sense through D&B and the observation operators.

3.1 Fraction of absorbed photosynthetically active radiation (FAPAR)

FAPAR is a measure of the capacity of terrestrial vegetation to absorb sunlight in the visible spectrum, i.e. that part that can be utilised as photosynthesis. It is defined as the amount of photosynthetically active radiation (PAR) absorbed by green, functioning leaves divided by the total incoming PAR. FAPAR is calculated within the two-flux canopy radiative transfer scheme (Section B.2.1) required for the calculation of GPP (Section 2.1). However, due to the dependence of FAPAR on solar zenith angle, it is necessary to take into account the solar zenith angle at time of observation. Therefore, the observation operator for FAPAR needs to ensure that either a separate calculation of FAPAR at the correct solar zenith angle is performed, or it utilizes FAPAR calculations from the model run at the times and dates where model and observations solar zenith angles match.

3.2 Solar Induced Fluorescence (SIF)

Strictly speaking, the canopy level solar-induced chlorophyll fluorescence, or SIF, is a measure not of the photosynthetic rate as such, but of the amount of radiation absorbed by the leaf and not used for the purpose of photosynthesis. Some of that surplus radiation is re-emitted as fluorescent light as part of a copying mechanism of the photosynthetic system. Under normal field conditions, however, SIF can often be used as an indication of photosynthetic activity, as opposed to FAPAR, which only characterises the photosynthetically active light that is potentially available (Porcar-Castell et al., 2014; Mohammed et al., 2019).

To calculate SIF for individual canopy layers, we offer the formulations of Gu et al. (2019), Van der Tol et al. (2014) and Li et al. (2022). These choices are motivated by the direct link to the photosynthesis routines and the relatively parsimonious implementation, which fits with the modelling strategy adopted here.

3.2.1 Gu model

The SIF for a single canopy layer, S_n, is given by:

$$S_n = s_{SIF} J_n \frac{1 - \psi_{PSIImax}}{q_L \psi_{PSIImax}(1 + k_{DF})}, \tag{1}$$

where J_n is the electron transport in layer n (Eq. (37)), $\psi_{PSIImax}$ is the maximum photochemical quantum yield of photosystem II, q_L is the fraction of open photosystem II reaction centres and k_{DF} the ratio of the first order rate constants for heat dissipation and fluorescence. We take the values prescribed by Gu et al. (2019). Note that the original equation in that paper also has a term for the photon escape probability from the canopy. In D&B, this is calculated explicitly by the layered 2-stream model (Section C) and hence is not required here. As an extension to the model by Gu et al. (2019) in view of the anticipated calibration in a data assimilation scheme, we further introduce the scaling factor s_{SIF}, which compensates for large uncertainties in (1) the values of the three constants $\psi_{PSIImax}$, q_L, and k_{DF} and (2) the spectral conversion that is described below. We set the prior value of s_{SIF} to 1.

3.2.2 Van der Tol model

The emission of SIF from a canopy layer is given by:

$$S_n = s_{SIF} APAR_n \left(1 - \frac{J_n}{APAR_n}\right) \phi'_m \tag{2}$$

where J_n is the electron transport and $APAR_n$ is the absorbed photosynthetic active radiation in layer n. ϕ'_m is the proportion of the remaining photons that are fluoresced:

$$\phi'_m = K_F / \left(K_F + K_D + K_N\right), \tag{3}$$

where K_F, K_D and K_N are rate constants for fluorescence, thermal dissipation and energy-dependent heat dissipation respectively. The value for K_F is set to 0.05 and the value of K_D is 0.95. K_N is given by:

$$K_N = K_{N0} x^{\alpha} \left(1 + \beta\right) / \left(\beta + x^{\alpha}\right), \tag{4}$$

where, $\alpha = 2.83$, $K_{N0} = 2.48$, $\beta = 0.114$ and the non-photochemical quenching (NPQ) parameter x is given by:

$$x = 1 - \frac{J_n}{J_{pot,n}}, \tag{5}$$

where, $J_{pot,n}$ is the potential rate of photosynthesis from the light limited regime.

3.2.3 Li model

The Li model is a variant on the Van der Tol model, but differs in that $K_{N0} = 5.01$, $\beta = 10.$ which results in a more rapid decline ϕ'_m with x. In addition it contains a dynamic formulation for K_D as a function of leaf temperature (T_v):

$$K_D = \max(0.8738, 0.0301 \cdot T_v + 0.0773) \tag{6}$$

3.2.4 Scaling to the canopy

Canopy level SIF is calculated via the layered 2-stream model (L2SM) described in Section C, which is physically consistent with the Sellers two-stream model, but permits internal canopy emissions. When coupled to the above leaf level models, it has native units of $mol \times m^{-2} s^{-1}$. It represents the total flux of photons into the hemisphere above the canopy for all wavelengths. Satellite measurements and *in situ* observations, however, are typically recorded in energy flux units per steradian, per nanometre of the SIF spectra, e.g. $Wm^{-2}s^{-1}nm^{-1}sr^{-1}$. To convert from molar, to energy units we apply the molar form of the Planck equation:

$$e = ahc/\lambda_{\phi}, \tag{7}$$

where a is Avogadro's number (6.023×10^{23}), h is the Planck constant ($6.626 \times 10^{-34}\ m^2 kg \cdot s^{-1}$), c is the speed of light ($3.0 \times 10^8\ m \cdot s^{-1}$), λ_ϕ is the wavelength of the SIF photons in metres and e is the energy per mole of photons at that wavelength.

We convert to steradians by using a constant factor of $\frac{1}{2\pi}$, which assumes that the emittance of SIF from the top of the canopy is isotropic, and finally weight by the relative strength of emissions at λ_ϕ compared to a reference SIF spectrum, i.e.:

$$w = \frac{E(\lambda_\phi)}{\sum_i E(\lambda_{\phi,i})}, \tag{8}$$

where E is the SIF emission spectrum of arbitrary units. Hence:

$$SIF' = SIF \frac{ew}{2\pi}, \tag{9}$$

Where SIF has units of $mol \cdot m^{-2} s^{-1}$ and SIF' has units of $W m^{-2} s^{-1} sr^{-1} nm^{-1}$.

The package contains two different SIF emission spectra, which are both provided by Magney et al. (2019). The first was measured at an oak tree (Quercus robur) in Pasadena and the other one at the Hyytiälä site in Finland at four Scots pine trees at light level of 1200 $\mu mol \cdot m^{-2} s^{-1}$ and then averaged. The respective conversion factors in Eq. (9) to the 743 nm band in which the TROPOMI instrument provides observations are $139.71531 \cdot 10^3\ W \cdot sr^{-1} \cdot nm^{-1} \cdot mol^{-1}$ for oak and $360.528344 \cdot 10^3\ W \cdot sr^{-1} \cdot nm^{-1} \cdot mol^{-1}$ for pine.

3.3 Vegetation optical depth (VOD)

Vegetation optical depth (VOD) is essentially a variable describing the attenuation of microwave radiation at some wavelength due to the presence of vegetation. It depends on the dielectric properties (due to water content, temperature and chemical composition) as well as the structure and geometry of the vegetation, and sensor properties (e.g. wavelength, polarization). Due to the relatively static nature of structure, dynamics of VOD are generally attributed to changes in above ground biomass and water content (Ulaby and Wilson, 1985; Konings et al., 2019). It is measured within the microwave spectrum with passive instruments, using the black body radiation of the surface in the microwave domain, or active instruments such as scatterometers or synthetic aperture radars.

Common retrieval methods may extract both the surface soil moisture and VOD simultaneously from satellite remote sensing data, provided enough measurements are performed. For example the SMOS (Kerr et al., 2010) retrieval algorithm (Kerr et al., 2012) is based on the so-called $\tau - \omega$ formulation for the vegetation contribution (Kirdiashev et al., 1979; Mo et al., 1982) of the microwave signal, where VOD is denoted by τ, the perpendicular vegetation optical depth (Wigneron et al., 2007, 2010).

We use a semi-empirical formulation for L-band VOD, expressed as:

$$\tau_\lambda = f(\bar{T})(l_{wd} C_{wd} + l_{fol} C_{fol})(l_s f_{soil} + l_f f_E + l_0), \tag{10}$$

Table 2: Parameters for empirical observation operator for L-VOD.

Parameter	Unit	Value
l_{wd}	m^2/gC	2.0×10^{-4}
l_{fol}	m^2/gC	2.0×10^{-4}
l_s	-	1.20
l_f	-	4.0
l_0	-	0.4

where the subscript λ denotes its wavelength dependence, C_{fol} and C_{wd} are the leaf and woody biomass pools, respectively (see Section 2.3), f_{soil} (Eq. (126)) is fractional plant-available soil water content, and

$$f_E = E_t / E_t^{pot} \tag{11}$$

i.e. the ratio of actual to potential transpiration (see Eq. (92) and Eq. (97)). f_{soil} describes slow changes in the plant's hydrological status, hence multiplied by parameter l_{hs}, and f_E fast changes, multiplied by parameter l_{hf}. The other empirical parameters are l_0, l_{wd} for dependence on woody biomass, and l_{fol} for dependence on leaf biomass. Following Schwank et al. (2021), we include an explicit temperature dependency in the form of:

$$f(\bar{T}) = 0.25 + 0.75/(1 + e^{-0.5(\bar{T}+3)}), \tag{12}$$

which approximates theoretically derived behaviour around the freezing point. $\bar{T}$ is the daily average 2-m air temperature. This formulation can be used across a range of microwave wavelengths, using different parameter values in each case. The second multiplicative factor in Eq. (10) is an empirical, linear expression using both woody and foliar biomass with the assumption that VOD will be zero if no biomass is present. The third multiplicative factor describes how the water status of the vegetation modifies this expression. This last one also contains a constant factor, $l_0 > 0$, because we expect positive VOD even if vegetation water stress is at its maximum.

In TCCAS, we apply it to passive L-band microwave measurements. The parameters for the empirical VOD observation operator, shown in Table 2, were chosen to reproduce a reasonable fit to L-band observations from SMOS over the Sodankylä and the Majadas del Tietar sites prior to any systematic assimilation.

3.4 Near-surface soil moisture

In the D&B model, near-surface soil moisture is represented by an explicitly modelled thin surface soil moisture layer, with a depth of 4 cm, unless depth-to-bedrock indicates a lower value. It is therefore a state variable in the model's soil water component, and is described in detail in Section 2.2. This surface layer therefore here serves a dual purpose, to diagnose a variable that can potentially be used as an assimilated data stream, and to simulate soil evaporation.

Near-surface soil moisture data is usually available from passive data when retrieved simultaneously with VOD. These retrieval algorithms explicitly separate the contributions to the microwave signal that come either from the vegetation (VOD) or from the soil (surface soil moisture).

3.5 ASCAT slope

The unique viewing geometry of the Advanced Scatterometer (ASCAT) on the Metop series of satellites provides observations of backscatter at a range of incidence angles. The incidence angle (θ) dependence of backscatter can be described with a second order Taylor polynomial as follows:

$$\sigma^{\circ}(\theta) = \sigma^{\circ}(\theta_r) + \sigma'(\theta) \cdot (\theta - \theta_r) + \frac{1}{2} \cdot \sigma''(\theta r) \cdot (\theta - \theta_r)^2 \qquad [dB] \tag{13}$$

where $\sigma^{\circ}(\theta_r)$, $\sigma'(\theta_r)$ and $\sigma''(\theta_r)$ are the normalized backscatter, slope and curvature at some reference angle θ_r. In the TU Wien soil moisture retrieval approach, this expression is used to normalize backscatter acquisitions at different incident angles to a reference angle (θ_r). It is also used to account for the influence of vegetation on backscatter as the incidence angle behaviour of σ° depends on whether total backscatter is dominated by surface scattering from the soil, volume scattering from the vegetation, or multiple scattering (Wagner et al., 1999; Naeimi et al., 2009; Hahn et al., 2017). Vreugdenhil et al. (2016) developed an approach to determine the vegetation optical depth τ from the changing sensitivity to surface soil moisture encapsulated in the slope and curvature. Recent studies have demonstrated that the slope contains information on vegetation phenology and water status, precisely the information required from VOD (Steele-Dunne et al., 2019; Petchiappan et al., 2022). However, using the slope directly avoids the additional assumptions required in the soil moisture and VOD retrieval algorithms (Shan et al., 2022). Given that the slope contains information on both vegetation biomass accumulation and vegetation water status, the observation operator is uses the same input variables as the observation operator for L-VOD (Eq. (10)):

$$\sigma' = a_0 + a_{wd} C_{wd} + a_s f_{soil} + a_{fol} C_{fol} + a_f f_E \tag{14}$$

where C_{fol},C_{wd}, f_{soil}, f_E are the D&B model outputs described in Section 3.3. The coefficients $a_0, a_{wd}, a_s, a_{fol}, a_f$ are empirical constants to be optimized. For Sodankyla the parameters for the empirical observation operator are estimated based on linear regression analysis using a training and testing dataset of varying D&B model outputs and observed slope over the period 2010-2021 (see Table 3). Data was randomly split in 80% training and 20% test data. One important difference between Eq. (14) and Eq. (10) is that the observation operator for ASCAT slope does not require any term to account for temperature effects. Microwave observations, including backscatter, are affected by the dielectric properties, geometry and roughness of the land surface which includes soil, vegetation but also snow. Furthermore, it is reasonable to assume that this sensitivity to snow is particularly relevant when freezing conditions increase the transmissivity of the forest. To avoid any ambiguity, the ASCAT slope data used in the project do not contain any observations acquired that might be affected by frozen soil, or the presence of snow. Therefore the temperature correction is not necessary. The ASCAT slope is obtained by aggregating so-called "local slope" values using an Epanechnikov kernel with a half-width of 14 days (see Hahn et al. (2017) for details). Therefore the ASCAT slope value on a given day is influenced by all local slope values, and therefore all land surface states, within this window. To take this into account, the same Epanechnikov kernel is applied to the D&B model outputs. -0.10708, ahf: 0.0369540, abl: 0.00005786, ahs: -0.0005770, abw: 0.0000007

Table 3: Parameters for empirical observation operator for ASCAT slope.

Parameter	Unit	Value
a_0	dimensionless	-1.0708 E-01
a_{wd}	m^2/gC	7.0 E-07
a_{fol}	m^2/gC	5.786 E-05
a_s	dimensionless	-5.777 E-04
a_f	dimensionless	3.6954 E-02

4 Assimilation Methodology

The method is based on the joint inversion approach of Tarantola (2005) (discussed as Bayesian inversion by Rayner et al. (2019)). It estimates the parameter vector from a given set of observations and the available prior information. The a priori state of information is quantified by a probability density function (PDF) in parameter space, the observational information by a PDF in observation space, and the information from the model by a PDF in the joint space, i.e. the Cartesian product of parameter and observation spaces. The inversion combines all three sources of information and yields a posterior PDF in the joint space.

Prior and observational PDFs are difficult to specify. Here we use Gaussian shapes with respective mean values denoted by x_0 and d and respective covariance matrices denoted by B (prior parameter uncertainty) and R (data uncertainty). The data uncertainty is the sum of uncertainties due to errors in the observational process, R_{obs} and errors in our ability to correctly model the observations, R_{mod}:

$$R = R_{\text{obs}} + R_{\text{mod}} \tag{15}$$

R_{mod} captures all sources of uncertainty that are not explicitly resolved through specification of the parameter vector (parametric uncertainty). Such sources of residual uncertainty include those resulting from uncertainties in driving data, from potential errors in the formulation of the process models (structural uncertainty), and from parameters and initial conditions not included in the parameter vector. For each of the observational data streams, 1-sigma uncertainty ranges were estimated for each observation (Rodríguez-Fernández et al., 2021). The corresponding variances are used to populate the diagonal of R_{obs}; the off-diagonal is set to zero. For the model uncertainty, R_{mod}, we developed an approach that is described in Section 5.1.

By target quantities we understand quantities of particular interest that we simulate with the model. For later use it is convenient to have two separate notations for the model depending on whether we use it to simulate a vector of counterparts of our observations, d_{obs}, or to simulate a vector of target quantities, y, from a given parameter vector x. For simulation of the observation vector we use M and for the simulation of the target vector we use N, i.e. we have

$$d_{\text{obs}} = M(x) \text{ and} \tag{16}$$

$$y = N(x) \ . \tag{17}$$

We approximate the posterior PDF by a Gaussian PDF. The corresponding marginal PDF in

parameter space is thus also Gaussian, with mean value x and covariance A. The mean x_{opt} is approximated by the maximum likelihood point, i.e. the minimum of the misfit function (or cost function):

$$J(x) = \frac{1}{2}[(M(x) - d)^T R^{-1}(M(x) - d) + (x - x_0)^T B^{-1}(x - x_0)] \tag{18}$$

We perform the minimisation of J over a single long assimilation window, to have all observational constraints act simultaneously.

The posterior parameter uncertainty A is approximated by the inverse of the misfit function's Hessian, H, evaluated at the minimum, x_{opt}, i.e.:

$$A \approx H(x_{\text{opt}})^{-1} \quad . \tag{19}$$

To compute H we use the approximation of a linear model (denoted by M'), for which differentiating $J(x)$ (Eq. (18)) twice yields

$$H(x_{\text{opt}}) \approx {M'}^T R^{-1} M' + B^{-1} \quad . \tag{20}$$

The Hessian is the sum of two terms, one reflecting the strength of the constraint from the prior information, and the other reflecting the observational constraint. Typically adding the observational constraint increases the curvature of the cost function which via Eq. (20) translates to a reduction in uncertainty compared to the prior.

From the optimal parameter set we can simulate (see Eq. (17)) a vector target quantities, y. To assess the strength of the observational constraint on a simulated target quantity, we use N', the first derivative of N to propagate the posterior parameter uncertainties forward to the uncertainty in the simulated vector of target quantities, $C(y)$:

$$C(y) = N' A {N'}^T \tag{21}$$

Evaluating Eq. (21) for the prior uncertainty B instead of the posterior uncertainty A, i.e. for a case without observational constraint, yields a prior uncertainty for the target quantities:

$$C(y_0) = N' B {N'}^T \tag{22}$$

For any component of the target vector we can quantify the added value/impact of the observations by the uncertainty reduction (often also termed knowledge gain) relative to the prior.

$$\frac{\sigma(y_{i,0}) - \sigma(y_i)}{\sigma(y_{i,0})} = 1 - \frac{\sigma(y_i)}{\sigma(y_{i,0})}, \tag{23}$$

where, $\sigma(y_i)$ and $\sigma(y_{i,0})$ respectively denote the 1-sigma uncertainty ranges, the squares of which populate the diagonals of $C(y)$ and $C(y_0)$. For example, if $\sigma(y_i)$ is 90 % of $\sigma(y_{i,0})$, then the uncertainty reduction is 10%; i.e. we have increased our knowledge on y by 10 %.

5 Code Components

The main TCCAS directory comprises the following subdirectories, over which the components of the system are distributed:

```
adstack        : sources of utility library
config         : include files for Makefile
diagout        : for model output
driver         : main programs
forcing        : forcing data
LICENSE        : License information
Makefile
mini           : source of minimisation algorithm
mk.compile     : include file to Makefile for local settings
observations   : observational data files for assimilation experiments
obs.nml        : controls the selection of observations
opt.nml        : controls the minimisation
parameters     : parameter files for several configurations
README         : this file
src            : Fortran sources of D&B model, tangent and adjoint
util           : python scripts
```

The following subsections provide a description of the system's components. We note that some of the settings presented here are relatively ad hoc, we invite the user to play with alternative settings.

5.1 Model and Observation Operators

D&B is described in Section 2 here we focus on the handling of the observations and of the parameters.

The task of an observation operator is to simulate the equivalent of an observation from the model's state variables (Kaminski and Mathieu, 2017). This includes the simulation of the variable that is observed (observable) at the time when it was observed and over the footprint of the observations. We have prepared five different EO data streams for assimilation (Figure 5):

- Surface layer soil moisture
- Fraction of Photosynthetically Active Radiation (FAPAR)
- Solar-Induced Fluorescence (SIF)
- L-Band Vegetation Optical Depth (L-VOD)
- Slope of ASCATbackscatter–incidence angle relationship

As described in Section 3, the first two of these observables are available in the core of D&B, more specifically within in the model's water balance (Section 2.2) and photosynthesis (Section 2.1)

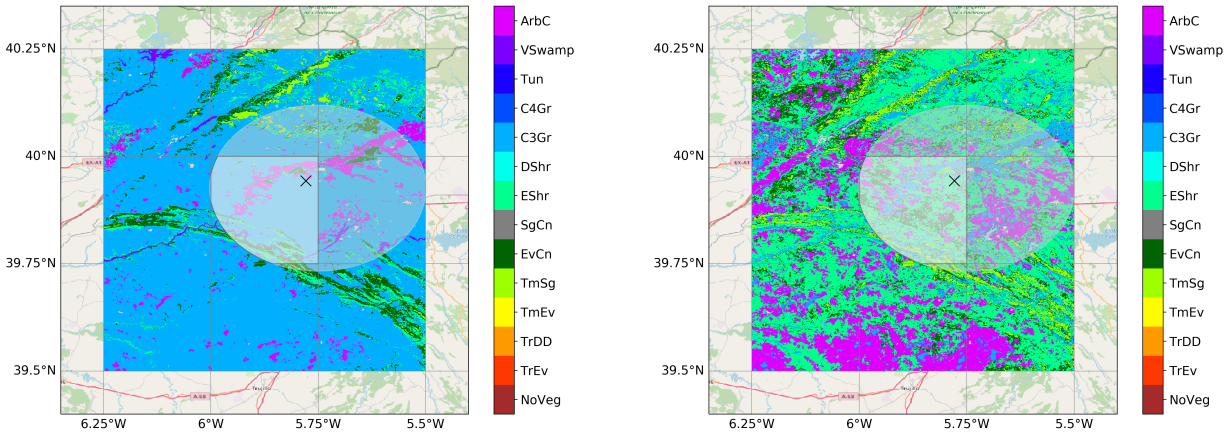


Figure 6: SMOS footprint (ellipse) along with the primary (left) and secondary (right) PFT over the grid defined by the meteorological driving data, with the location of the LM1 site indicated by a cross.

modules. By contrast, to compute the other three variables dedicated observation operators were developed and coupled to the core of D&B (see also Section 3).

5.1.1 Simulation on Footprint

The observation operators for all five variables in addition need to simulate the equivalent of an observation on the observational footprint, if known. For the regional simulations, the model is set up to run with meteorological forcing data from ERA5 on a 0.25 °× 0.25 ° regular grid. At site scale, the model can alternatively be run with locally observed meteorological forcing data.

For each grid cell or site, the model is run for all plant functional types (PFTs) that are present in the respective grid cell or site. Model output is provided separately for each PFT included in the simulation. Using this model output we can provide all simulation results at sub-grid scale by using the appropriate PFT mix at that scale. At the local scale (scale of the field observation footprints) the respective PFT fractions are derived from the actual vegetation cover provided by the field data, while at regional scale we use a high-resolution PFT map.

Matching the footprint of an observation is done by computing the model equivalent of the respective variable based on the PFT distribution present within the observational footprint. This is illustrated in Figure 6. It shows a SMOS footprint (ellipse) along with the primary (left) and secondary (right) PFT over the grid defined by the meteorological driving data around the Majadas de Tietar (LM1) site, with the location of the LM1 site indicated by a cross. The PFT map was derived from the landcover specification provided by the Copernicus Global Land Service at 100 m resolution (Buchhorn et al., 2019).

If a grid cell is covered by n PFT types with fractions f_1 to f_n and the observed variable is, for

example, a pool size or flux, y, then the simulated equivalent is

$$y = \sum_{i=1,n} f_i y_i, \tag{24}$$

where y_i is the simulation for PFT type i. Eq. (24) is also used for the slope of the ASCAT backscatter–incidence angle relationship, implicitly assuming that the footprint-scale slope is a linear weighted combination of contributions from the individual PFTs. If the observed variable is VOD, denoted by τ, we apply Eq. (24) to the transmitted flux

$$\tau = -ln(\sum_{i=1,n} f_i e^{-\tau_i}). \tag{25}$$

5.1.2 Uncertainty contribution reflecting model error

We also included a simple model of the model uncertainty, R_{mod}. The model determines for each of the data streams to be assimilated a model uncertainty component, σ_{mod}, that is specific to that data stream and constant in time:

$$\sigma_{\mathsf{mod}} = f \cdot \bar{d}, \tag{26}$$

where $\bar{d}$ is the average of all observations of that data stream and f a factor that is specific to that data stream. The factor is the sum of two terms, f_{mq} and f_{fp}. f_{mq} is intended to quantify our capability to simulate the observed variable (with calibrated model parameters). We assign a value of 0.1 to our biophysical models for SIF, FAPAR, and surface layer soil moisture and a value of 0.5 for our empirical models for L-VOD and ASCAT slope. The second term, f_{fp}, is intended to capture our uncertainty in simulating the observed footprint with correct driving data, including the PFT map. For the simulation at site level with the site-specific driving data we use 0.1 for the optical observations (which have small footprints), 0.5 for the two data streams derived from SMOS (surface layer soil moisture and L-VOD) with its large footprint, and 0.25 for ASCAT with its footprint size in between that of the optical instruments and that of SMOS. We use the square of the respective σ_{mod} to populate the diagonal of R_{mod} and set the off-diagonal to zero. Table 4 provides the values of f_{mq} and f_{fp} we specified for our site-scale experiment that use a single set of meteorological driving data and a single PFT distribution for simulation of all five data streams. These values can be adapted in the file `src/obs.f90`. In the file `src/observables.f90` the user can specify (parameter `nhourspin`) the length of the spinup period. The spinup period precedes the assimilation window and allows the model to move towards a balanced state, in particular for pools of which the initial states are not included in the parameter vector (soil moisture). The parameter `mmax` in file `src/obs.f90` limits the length of the observation vector.

5.1.3 Parameter Vector

The parameter vector, x, is specified through a set of parameter tables, which reside in subdirectory `parameters` and include:

Table 4: Values of the individual factors for determining the model uncertainty per observable.

Observable	f_{mq}	f_{fp}
ASCAT slope	0.5	0.25
SMOS L-VOD	0.5	0.5
SMOS SM	0.1	0.5
SIF	0.1	0.1
FAPAR	0.1	0.1

1. parameter number,
2. the name of the variable,
3. the PFT it applies to (a value of "-1" indicates that it applies to all PFTs),
4. whether the parameter is active, i.e. whether it should be adapted by the optimisation (value of "1") or not (value of "0"), currently this column takes no effect, i.e. all parameters are active,
5. the prior value,
6. the 1-sigma range corresponding to its prior uncertainty,
7. its minimum reasonable value,
8. its maximum reasonable value,
9. whether these minimum (value of "1"), maximum (value of "3"), or both (value of "2") values should be used as bounds or not (value of "0"),
10. the process the parameter belongs to,
11. a description of the parameter's role

As examples, Table 5 shows the parameters of the core model for the Sodankylä site (file `parameters/FI-Sod-core-params.csv`) and Table 6 for Majadas de Tietar (file `parameters/ES-LM1-core-params.csv`). We note that there are four parameters that have no effect on the simulation. These are the DALEC parameters 2 and 11 for both PFTs, i.e. parameters 6, 15, 23, 32, which relate to autotrophic respiration and photosynthesis, two processes now modelled outside the carbon balance component. They will be removed from the parameter list. Parameters for the observation operators for L-VOD, SIF, and ASCAT slope are specified in separate files, to keep the system modular. These are shown in Table 7 (`ES-LM1-lvod-params.csv`), Table 8 (`ES-LM1-slope-params.csv`), and Table 9 (`ES-LM1-sif-params.csv`). The role of the individual parameters is described in Section 2.

Table 5: Core model parameters for Sodankylä.

#	varname	PFT	active	prior	sigma	min	max	nbound	process	description (don't use commas!)
1	vmax	5	1	0.000029	0.0000058	0.00001	0.00008	2	photosynthesis	maximum carboxylation rate at ref. Temperature [mol(CO2) / m^2 s]
2	vmax	7	1	0.000052	0.0000104	0.00001	0.0001	2	photosynthesis	maximum carboxylation rate at ref. Temperature [mol(CO2) / m^2 s]
3	root_depth_scale	-1	1	1.	0.2	0.	5.	1	soilwater	"root_depth_scale"
4	soil_theta_scale	-1	1	1.	0.2	0.	5.	1	soilwater	"soil_theta_scale"
5	DALEC P1	5	1	0.0004564147	0.00009128294	0.00001	0.01	2	cbalance+pheno	fractional turnover rate of litter to soil organic matter at 0oC
6	DALEC P2	5	1	0.4198307	0.08396614	0.2	0.8	2	cbalance+pheno	fraction of gross primary productivity allocated to autotrophic respiration
7	DALEC P3	5	1	0.1180979	0.02361958	0.01	0.5	2	cbalance+pheno	fraction of gpp-ra allocated to leaves
8	DALEC P4	5	1	0.2770209	0.05540418	0.1	0.8	2	cbalance+pheno	fraction of gpp-ra-alloc_fol-alloc_lab allocated to fine roots
9	DALEC P5	5	1	1.189258	0.2378516	1.001	8	2	cbalance+pheno	maximum leaf lifespan
10	DALEC P6	5	1	0.0001252219	0.00002504438	0.000009	0.001	2	cbalance+pheno	fractional daily turnover of the wood pool
11	DALEC P7	5	1	0.007202744	0.0014405488	0.001368925	0.02	2	cbalance+pheno	fractional daily turnover of the fine root pool
12	DALEC P8	5	1	0.005906897	0.0011813794	0.0001141	0.02	2	cbalance+pheno	fractional daily turnover of the litter pool to heterotrophic respiration
13	DALEC P9	5	1	0.00001565383	0.000003130766	1.37E-06	9.13E-05	2	cbalance+pheno	fractional daily turnover of the soil organic matter pool to heterotrophic respiration
14	DALEC P10	5	1	0.04813463	0.009626926	0.019	0.08	2	cbalance+pheno	coefficient for exponential temperature sensitivity for litter decomposition and litter and som turnover to heterotrophic respiration
15	DALEC P11	5	1	20.83379	4.166758	1.64	42	2	cbalance+pheno	Canopy photosynthetic efficiency parameter
16	DALEC P12	5	1	156.1256	31.22512	10	350	2	cbalance+pheno	Day of year for maximum labile turnover to foliage (i.e. bud burst)
17	DALEC P13	5	1	0.135938	0.0271876	0.01	0.5	2	cbalance+pheno	fraction of gpp-ra-alloc_fol allocated to labile pool (which supports seasonal leaf growth)
18	DALEC P14	5	1	29.27762	5.855524	10	100	2	cbalance+pheno	Number of days over which labile turnover to leaves occurs
19	DALEC P15	5	1	230.3232	46.06464	10	350	2	cbalance+pheno	Day of year for maximum leaf turnover to litter (i.e. leaf senesence)
20	DALEC P16	5	1	50.85548	10.171096	20	150	2	cbalance+pheno	Number of days over which leaf turnover to litter occurs
21	DALEC P17	5	1	36.11364	7.222728	20	180	2	cbalance+pheno	Leaf carbon per unit leaf area
22	DALEC P1	7	1	0.0004564147	0.00009128294	0.00001	0.01	2	cbalance+pheno	fractional turnover rate of litter to soil organic matter at 0oC
23	DALEC P2	7	1	0.4198307	0.08396614	0.2	0.8	2	cbalance+pheno	fraction of gross primary productivity allocated to autotrophic respiration
24	DALEC P3	7	1	0.1180979	0.02361958	0.01	0.5	2	cbalance+pheno	fraction of gpp-ra allocated to leaves
25	DALEC P4	7	1	0.2770209	0.05540418	0.1	0.8	2	cbalance+pheno	fraction of gpp-ra-alloc_fol-alloc_lab allocated to fine roots
26	DALEC P5	7	1	1.189258	0.2378516	1.001	8	2	cbalance+pheno	maximum leaf lifespan
27	DALEC P6	7	1	0.0001252219	0.00002504438	0.000009	0.001	2	cbalance+pheno	fractional daily turnover of the wood pool
28	DALEC P7	7	1	0.007202744	0.0014405488	0.001368925	0.02	2	cbalance+pheno	fractional daily turnover of the fine root pool
29	DALEC P8	7	1	0.005906897	0.0011813794	0.0001141	0.02	2	cbalance+pheno	fractional daily turnover of the litter pool to heterotrophic respiration
30	DALEC P9	7	1	0.00001565383	0.000003130766	1.37E-06	9.13E-05	2	cbalance+pheno	fractional daily turnover of the soil organic matter pool to heterotrophic respiration
31	DALEC P10	7	1	0.04813463	0.009626926	0.019	0.08	2	cbalance+pheno	coefficient for exponential temperature sensitivity for litter decomposition and litter and som turnover to heterotrophic respiration
32	DALEC P11	7	1	20.83379	4.166758	1.64	42	2	cbalance+pheno	Canopy photosynthetic efficiency parameter
33	DALEC P12	7	1	156.1256	31.22512	10	350	2	cbalance+pheno	Day of year for maximum labile turnover to foliage (i.e. bud burst)
34	DALEC P13	7	1	0.135938	0.0271876	0.01	0.5	2	cbalance+pheno	fraction of gpp-ra-alloc_fol allocated to labile pool (which supports seasonal leaf growth)
35	DALEC P14	7	1	29.27762	5.855524	10	100	2	cbalance+pheno	Number of days over which labile turnover to leaves occurs
36	DALEC P15	7	1	230.3232	46.06464	10	350	2	cbalance+pheno	Day of year for maximum leaf turnover to litter (i.e. leaf senesence)
37	DALEC P16	7	1	50.85548	10.171096	20	150	2	cbalance+pheno	Number of days over which leaf turnover to litter occurs
38	DALEC P17	7	1	36.11364	7.222728	20	180	2	cbalance+pheno	Leaf carbon per unit leaf area
39	DALEC IC P18	5	1	30.47807	6.095614	1	2000	2	cbalance+pheno	Initial size of the labile pool
40	DALEC IC P19	5	1	29.29126	5.858252	1	2000	2	cbalance+pheno	Initial size of the foliage pool
41	DALEC IC P20	5	1	17.85931	3.571862	1	2000	2	cbalance+pheno	Initial size of the fine root pool
42	DALEC IC P21	5	1	3072.245	614.449	1	30000	2	cbalance+pheno	Initial size of the wood pool
43	DALEC IC P22	5	1	60.23932	12.047864	1	2000	2	cbalance+pheno	Initial size of the litter pool (NOTE: this is for foliage and fine root only)
44	DALEC IC P23	5	1	40910.35	8182.07	200	250000	2	cbalance+pheno	Initial size of the soil organic matter pool (NOTE: this is the soil plus wood litter)
45	DALEC IC P18	7	1	30.47807	6.095614	1	2000	2	cbalance+pheno	Initial size of the labile pool
46	DALEC IC P19	7	1	29.29126	5.858252	1	2000	2	cbalance+pheno	Initial size of the foliage pool
47	DALEC IC P20	7	1	17.85931	3.571862	1	2000	2	cbalance+pheno	Initial size of the fine root pool
48	DALEC IC P21	7	1	3072.245	614.449	1	30000	2	cbalance+pheno	Initial size of the wood pool
49	DALEC IC P22	7	1	60.23932	12.047864	1	2000	2	cbalance+pheno	Initial size of the litter pool (NOTE: this is for foliage and fine root only)
50	DALEC IC P23	7	1	40910.35	8182.07	200	250000	2	cbalance+pheno	Initial size of the soil organic matter pool (NOTE: this is the soil plus wood litter)

5.1.4 Model output

The output of a model run is stored in two NetCDF files, which are placed in subdirectory `diagout` and with names like

- `diagout/dalec-bethy_daily-output_20150101-20211231.nc`
- `diagout/dalec-bethy_hourly-output_20150101-20211231.nc`

The first one contains daily averages of several variables while the second file contains variables which are written at the hourly time step of the model. Both file names include start and end dates of the simulation. The output variables include carbon and water pools and fluxes and the five observational data streams. They are provided for the combined space-pft dimension `nsp`, which counts all active PFT over all grid cells. During the development phase, we were rather generous in the selection of output variables needed for various tests.

Table 6: Core model parameters for Majadas de Tietar.

#	varname	PFT	active	prior	sigma	min	max	nbound	process	description (don't use commas!)
1	vmax	3	1	0.000041	0.0000082	0.00001	0.00008	2	photosynthesis	maximum carboxylation rate at ref. Temperature [mol(CO2) / m^2 s]
2	vmax	9	1	0.000042	0.0000084	0.00001	0.00008	2	photosynthesis	maximum carboxylation rate at ref. Temperature [mol(CO2) / m^2 s]
3	root_depth_scale	-1	1	1.	0.2	0.	5.	1	soilwater	"root_depth_scale"
4	soil_theta_scale	-1	1	1.	0.2	0.	5.	1	soilwater	"soil_theta_scale"
5	DALEC P1	3	1	0.0003038811	0.00006077622	0.00001	0.01	2	cbalance+pheno	fractional turnover rate of litter to soil organic matter at 0oC
6	DALEC P2	3	1	0.4016085	0.0803217	0.2	0.8	2	cbalance+pheno	fraction of gross primary productivity allocated to autotrophic respiration
7	DALEC P3	3	1	0.04980515	0.00996103	0.01	0.5	2	cbalance+pheno	fraction of gpp-ra allocated to leaves
8	DALEC P4	3	1	0.3553572	0.07107144	0.1	0.8	2	cbalance+pheno	fraction of gpp-ra-alloc_fol-alloc_lab allocated to fine roots
9	DALEC P5	3	1	1.104317	0.2208634	1.001	8	2	cbalance+pheno	maximum leaf lifespan
10	DALEC P6	3	1	0.00006646695	0.00001329339	0.000009	0.001	2	cbalance+pheno	fractional daily turnover of the wood pool
11	DALEC P7	3	1	0.005876034	0.0011752068	0.001368925	0.02	2	cbalance+pheno	fractional daily turnover of the fine root pool
12	DALEC P8	3	1	0.006875432	0.0013750864	0.0001141	0.02	2	cbalance+pheno	fractional daily turnover of the litter pool to heterotrophic respiration
13	DALEC P9	3	1	0.00002478762	0.000004957524	1.37E-06	9.13E-05	2	cbalance+pheno	fractional daily turnover of the soil organic matter pool to heterotrophic respiration
14	DALEC P10	3	1	0.05102073	0.010204146	0.019	0.08	2	cbalance+pheno	coefficient for exponential temperature sensitivity for litter decomposition and litter and som turnover to heterotrophic respiration
15	DALEC P11	3	1	22.29913	4.459826	1.64	42	2	cbalance+pheno	Canopy photosynthetic efficiency parameter
16	DALEC P12	3	1	74.3057	14.86114	10	350	2	cbalance+pheno	Day of year for maximum labile turnover to foliage (i.e. bud burst)
17	DALEC P13	3	1	0.2200071	0.04400142	0.01	0.5	2	cbalance+pheno	fraction of gpp-ra-alloc_fol allocated to labile pool (which supports seasonal leaf growth)
18	DALEC P14	3	1	74.18161	14.836322	10	100	2	cbalance+pheno	Number of days over which labile turnover to leaves occurs
19	DALEC P15	3	1	157.3799	31.47598	10	350	2	cbalance+pheno	Day of year for maximum leaf turnover to litter (i.e. leaf senesence)
20	DALEC P16	3	1	54.62658	10.925316	20	150	2	cbalance+pheno	Number of days over which leaf turnover to litter occurs
21	DALEC P17	3	1	67.66937	13.533874	20	180	2	cbalance+pheno	Leaf carbon per unit leaf area
22	DALEC P1	9	1	0.002584829	0.0005169658	0.00001	0.01	2	cbalance+pheno	fractional turnover rate of litter to soil organic matter at 0oC
23	DALEC P2	9	1	0.5392843	0.10785686	0.2	0.8	2	cbalance+pheno	fraction of gross primary productivity allocated to autotrophic respiration
24	DALEC P3	9	1	0.3613487	0.07226974	0.01	0.5	2	cbalance+pheno	fraction of gpp-ra allocated to leaves
25	DALEC P4	9	1	0.7058508	0.14117016	0.1	0.8	2	cbalance+pheno	fraction of gpp-ra-alloc_fol-alloc_lab allocated to fine roots
26	DALEC P5	9	1	1.008241	0.2016482	1.001	8	2	cbalance+pheno	maximum leaf lifespan
27	DALEC P6	9	1	0.0007320861	0.00014641722	0.000009	0.001	2	cbalance+pheno	fractional daily turnover of the wood pool
28	DALEC P7	9	1	0.008705515	0.001741103	0.001368925	0.02	2	cbalance+pheno	fractional daily turnover of the fine root pool
29	DALEC P8	9	1	0.0005506519	0.00011013038	0.0001141	0.02	2	cbalance+pheno	fractional daily turnover of the litter pool to heterotrophic respiration
30	DALEC P9	9	1	0.00003957299	0.000007914598	1.37E-06	9.13E-05	2	cbalance+pheno	fractional daily turnover of the soil organic matter pool to heterotrophic respiration
31	DALEC P10	9	1	0.03424815	0.00684963	0.019	0.08	2	cbalance+pheno	coefficient for exponential temperature sensitivity for litter decomposition and litter and som turnover to heterotrophic respiration
32	DALEC P11	9	1	14.25688	2.851376	1.64	42	2	cbalance+pheno	Canopy photosynthetic efficiency parameter
33	DALEC P12	9	1	77.34133	15.468266	10	350	2	cbalance+pheno	Day of year for maximum labile turnover to foliage (i.e. bud burst)
34	DALEC P13	9	1	0.4192669	0.08385338	0.01	0.5	2	cbalance+pheno	fraction of gpp-ra-alloc_fol allocated to labile pool (which supports seasonal leaf growth)
35	DALEC P14	9	1	28.08799	5.617598	10	100	2	cbalance+pheno	Number of days over which labile turnover to leaves occurs
36	DALEC P15	9	1	121.9728	24.39456	10	350	2	cbalance+pheno	Day of year for maximum leaf turnover to litter (i.e. leaf senesence)
37	DALEC P16	9	1	65.04758	13.009516	20	150	2	cbalance+pheno	Number of days over which leaf turnover to litter occurs
38	DALEC P17	9	1	46.94364	9.388728	20	180	2	cbalance+pheno	Leaf carbon per unit leaf area
39	DALEC IC P18	3	1	34.5735	6.9147	1	2000	2	cbalance+pheno	Initial size of the labile pool
40	DALEC IC P19	3	1	36.33911	7.267822	1	2000	2	cbalance+pheno	Initial size of the foliage pool
41	DALEC IC P20	3	1	34.91805	6.98361	1	2000	2	cbalance+pheno	Initial size of the fine root pool
42	DALEC IC P21	3	1	6737.396	1347.4792	1	30000	2	cbalance+pheno	Initial size of the wood pool
43	DALEC IC P22	3	1	12.89959	2.579918	1	2000	2	cbalance+pheno	Initial size of the litter pool (NOTE: this is for foliage and fine root only)
44	DALEC IC P23	3	1	11818.84	2363.768	200	2500000	2	cbalance+pheno	Initial size of the soil organic matter pool (NOTE: this is the soil plus wood litter)
45	DALEC IC P18	9	1	60.98822	12.197644	1	2000	2	cbalance+pheno	Initial size of the labile pool
46	DALEC IC P19	9	1	24.8358	4.96716	1	2000	2	cbalance+pheno	Initial size of the foliage pool
47	DALEC IC P20	9	1	10.63674	2.127348	1	2000	2	cbalance+pheno	Initial size of the fine root pool
48	DALEC IC P21	9	1	89.56923	17.913846	1	30000	2	cbalance+pheno	Initial size of the wood pool
49	DALEC IC P22	9	1	16.35128	3.270256	1	2000	2	cbalance+pheno	Initial size of the litter pool (NOTE: this is for foliage and fine root only)
50	DALEC IC P23	9	1	15469.69	3093.938	200	2500000	2	cbalance+pheno	Initial size of the soil organic matter pool (NOTE: this is the soil plus wood litter)

Table 7: Parameters of L-VOD observation operator.

#	varname	PFT	active	prior	sigma	min	max	nbound	process	description (don't use commas!)
1	cbw	-1	1	0.0002	0.0001	0.	5	1	lvodemp	multiplier of woody carbon pool
2	chs	-1	1	1.2	0.6	0.	10	1	lvodemp	multiplier of slowly changing wetting status
3	cbl	-1	1	0.0002	0.0001	0.	1	1	lvodemp	multiplier of leaf carbon pool
4	chf	-1	1	4	2	0.	100	1	lvodemp	multiplier of rapidly changing wetting status
5	c0	-1	1	0.4	0.2	0.	5.	1	lvodemp	offset

The Fortran code of D&B and its observation operators is distributed over a number of Fortran subroutines provided in subdirectory `src` and a main program provided in subdirectory `driver`.

Table 8: Parameters of ASCAT slope observation operator.

#	varname	PFT	active	prior	sigma	min	max	nbound	process	description (don't use commas!)
1	abw	-1	1	7.00E-07	0.00000014	0	1	2	ascatemp	multiplier of woody carbon pool
2	ahs	-1	1	-5.77E-04	0.0001154	-1	1	2	ascatemp	multiplier of slowly changing wetting status
3	abl	-1	1	5.79E-05	0.000011572	0	1	2	ascatemp	multiplier of leaf carbon pool
4	ahf	-1	1	0.03695	0.00739	0	1	2	ascatemp	multiplier of rapidly changing wetting status
5	a0	-1	1	-0.10708	0.021416	-1	1	2	ascatemp	offset

Table 9: Parameters of SIF observation operator.

#	varname	PFT	active	prior	sigma	min	max	nbound	process	description (don't use commas!)
1	sif_scale	-1	1	1.	100.	0.	1000.	0	l2sm SIF	"Scaling factor of leaf level SIF"

5.2 Minimisation algorithm

The minimisation of Eq. (18) uses a numerical library routine (Zhu et al., 1997) that efficiently searches the parameter space and relies on the capability to evaluate J and its gradient for any given parameter vector. This library is written in Fortran and resides in subdirectory `mini`. The algorithm can be further guided by restricting the parameter space through specification of upper and lower bounds for each parameter, see Section 5.1.

5.3 Derivative code

The code for evaluation of the gradient of J is provided by a procedure called automatic differentiation (AD Griewank, 1989) from the code for evaluation of J. In essence, the code that evaluates J is decomposed into elementary functions (such as $+, -, \sin(\cdot)$), for which the derivative (local Jacobian) is straightforward to derive. The derivative of the composite function is then constructed via the chain rule as the product of all local Jacobians. According to the associative law, this multiple matrix product can be evaluated in arbitrary order without changing the result. The *tangent linear* code (or just *tangent* code) does this evaluation in the same order as the function is evaluated, which is called forward mode of AD. The *adjoint* code uses exactly the opposite order, which is called reverse mode of AD. Even though both modes yield the same derivative, depending on the dimensions of the function to be differentiated, there may be large differences in their computational efficiency: The CPU time required by tangent code is proportional to the number of the function's input variables (in our case the dimension of the parameter space) but independent of the number of output variables. By contrast, the CPU time required by the adjoint code is proportional to the number of output variables and independent of the number of input variables. Being an application of the chain rule, AD provides derivatives that are accurate up to rounding error. Tangent and adjoint versions of D&B and of the observation operators were generated by the automatic differentiation tool TAPENADE (Hascoët and Pascual, 2013) in Fortran. They are provided in files `src/cost_tl.f90` and `src/cost_ad.f90`, respectively. The derivative code is modular. Each derivative routine belongs to a specific routine in D&B including its observation operators. Subdirectory `adstack` provides code for a library that arranges storing of intermediate values required for execution of the adjoint.

5.4 Pre-processing

A number of pre-processing steps are performed by python scripts, because they are easier to implement in python than in Fortran and computational speed is not critical for these steps.

They are:

- preparation of driving data
- configuration of model setup
- preparation of observational data

In the following we briefly describe each of these tasks.

As outlined in the D&B model description Section 2, running the model requires surface temperature, incoming solar short-wave radiation, incoming long-wave radiation, precipitation, and soil temperature as time-dependent meteorological forcing data as well as some static fields (map of PFT fractions, elevation, soil texture class, soil depth, and soil brightness class). These input data must be provided in NetCDF format where the names of the datasets and the spatio-temporal structure must follow the conventions shown in Section D (static forcing) and Section E (dynamic forcing). D&B reads the temporal information for a simulation from the variable `yyyymmddhh` in the dynamic forcing file, which provides year, month, day, and hour for each time step of the simulation. Complete forcing data for site-level simulations are provided for two study sites (Sodankylä and the Majadas del Tietar) in subdirectory `forcing` of the TCCAS package:

```
ES-LM1_dynforcing-era5_20090101-20211231_with-lwdown.nc
ES-LM1_dynforcing-insitu_20140401-20220930_with-insitu-lwdown.nc
ES-LM1_staticforcing.nc
FI-Sod_dynforcing-era5_20090101-20211231_with-lwdown.nc
FI-Sod_dynforcing-insitu_20090101-20211231_with-insitu-lwdown.nc
FI-Sod_staticforcing.nc
```

At site level a user may switch between meteorological forcing data extracted from ERA5 or extracted from in situ data collected as part of the LCC study.

For the static data at site level, we recommend using a combination of local observations and expert knowledge. The user needs to select from Table 1 the PFT(s) that cover(s) the simulated area and to specify their respective fractions of the total vegetation over that area. The role of bare soil and the vegetation fraction are explained in Section 2. Further requirements are elevation, soil depth, and assignment to one of three soil brightness classes shown in Table 15 and one of several soil texture classes as shown in Table 14. We recommend examining the static data files provided with TCCAS for selected sites.

For the static data at regional to global scales, one would typically derive maps of PFT fractional cover from global land cover data sets by mapping the available land cover data set to the suitable subset of PFTs (see Table 1) that cover the domain under investigation. We recommend those provided by the Copernicus Global Land Service at 100 m resolution as "Dynamic Land Cover" (Buchhorn et al., 2019, URL: https://land.copernicus.eu/en/map-viewer), as it gives fractions of

land cover classes at the full resolution, making it easy to aggregate to lower resolution. These are in particular fractions of trees, shrubs, grasses (or herbs), cropland, herbaceous wetlands as well as mosses/lichen. For the tree fraction, there is also information available as to whether trees are broadleaf or coniferous.

This land classification can be mapped onto the reduced PFT list (Table 1) by mapping "mosses/lichen" on the Tundra PFT (PFT 11) as well as "herbaceous wetland" onto the Wetland PFT (PFT 12), but also requires dividing grasses into C3 and C4 vegetation, as well as trees into tropical and temperate. This can be done either based on available regional information (e.g. if C4 grasses are negligible in the area), or according to the mean climate of the location. For example, the mean temperature of the coldest month has been used to distinguish between tropical and temperate trees PFTs, with a threshold value of 15.5^{o}C for the occurrence of tropical tree species (Prentice et al., 1992). Further, a global data set of C4 vegetation fraction by Still et al. (2003) can be used to partition grasses into C3 and C4 types (URL: https://daac.ornl.gov/ISLSCP_II/guides/c4_percent_1deg.html). In this case, it is a reasonable approximation to assume that the C4 fraction of the data sets is equal to the C4 fraction of grasses (partitioning between PFTs 9 and 10). There is also a climate based method to determine C4 fraction of the grass vegetation described in Section 2.4 of Knorr (1997), as well as methods to derive other PFTs and their natural occurrence from climate data. If in addition to the reduced list one wants also to represent deciduous conifers (i.e. larch), it is reasonable to assume that this PFT (PFT 6) occurs when the mean temperature of the coldest month is below -35^{o}C (Prentice et al., 1992).

Most other land cover products provide a rough idea of the PFT distribution of each land cover class, but no specific fractions, such as those from the ESA CCI Land Cover project (ESA, 2017).

In such a case, it is up to the user to make assumptions about the typical PFT fractions of each of those land cover categories, based on the description of those categories provided in the literature, as well as local or regional expert knowledge.

Maps of the soil properties can, for example, be taken from the SoilGrids data sets available via https://www.isric.org/explore/soilgrids. Alternatively, soil depth and texture maps provided by Koster et al. (1995); Zobler (1986) are available from https://iridl.ldeo.columbia.edu/SOURCES/.NASA/.ISLSCP/.GDSLAM/.Hydrology-Soils/.soils. A soil brightness data set by Wilson and Henderson-Sellers (1985) is available via https://oidc.rda.ucar.edu/datasets/d767000. Elevation maps are available, e.g., from https://www.usgs.gov/faqs/where-can-i-get-global-elevation-data. We note that each of these input data sets needs careful inspection. For example, after identifying in a previously used map of soil depth, very shallow values of the Iberian peninsula, we decided to use the soil depth at the Majadas de Tietar site for the LCC study region in the Iberian Peninsula. Forcing data for regional simulations of the LCC study regions is available on the TCCAS website `https://tccas.inversion-lab.com/database.html`.

The D&B model is designed to run on different spatial scales and on temporal periods selected by the user, which means that the spatio-temporal dimensions are in principle known only at runtime. The respective configuration is performed in a pre-processing step by the python script

`util/model_setup.py`, which generates the (Fortran) module file `dimensions.f90`. Its arguments can be listed with the command:

```
util/model_setup.py dimensions_setup -h
```

The static and dynamic forcing files for the simulation are mandatory input to this script; and the first and last day of the simulation period must also be provided to the script but are preset to default values of 20150101 and 20211231 respectively. Optionally the user may specify whether D&B shall be run in time with the standard Gregorian calendar or with fixed 365 days per year. Note that the user needs to ensure that the meteorological forcing data for the simulation covers the selected temporal range. Internally the script uses a template (Section F), into which the case-dependent dimensions are then inserted according to the user input.
The TCCAS package contains EO data files with observations covering the location of the two study sites and the following data streams:

- TROPOMI SIF at 743nm wavelength
- SMOS-L2 L-VOD
- SMOS-L2 soil moisture
- ASCAT slope
- FAPAR derived from JRC-TIP

All data files are in ASCII format and reside in the subdirectory `observations`. They are suitable as input for data assimilation experiments and allow easy and fast reading into the Fortran system using the fixed format `'(i10,2e16.8,2(i16,e16.8))'` while the first line denotes the 7 columns:

(1) index of hour of observation in the D&B time loop

(2) observational data

(3) observational data uncertainty

(4) integer number of first PFT

(5) fraction of first PFT

(6) integer number of second PFT

(7) fraction of second PFT

An exemplary extract of such a data file is shown in Section G. For each combination of site and data stream the actual observed value and its uncertainty have been extracted from the regional EO data files in the data base produced by the LCC study, taking into account only the footprint around the site with centre closest to the site for each particular point in time. For all data streams filtering as recommended by the EO team of the LCC study has been applied (Rodríguez-Fernández et al., 2021). The index of the hour of observation in column 1 is suitable for data assimilation experiments with a simulation starting on Januray 1, 2015.

6 Controlling TCCAS

The operation of the system is controlled via a `Makefile` and two Fortran namelist files, `obs.nml` and `opt.nml`. While the first specifies the observational configuration and the formulation used for the SIF source (Section 3.2), the latter allows to perturb the prior parameter vector, a feature that we use for assimilation of synthetic observations (see Section 9.3. The file `mk.compile` contains settings that are specific to the user's computing platform and is included by the main Makefile. Subdirectory `mk.compile` contains a template for such an include file.

7 TCCAS Output

The essential output of a minimisation is the posterior parameter vector. It is used as input to a subsequent run of D&B, which produces from the posterior parameter vector the regular set of D&B outputs.

8 Repository and Installation

8.1 Requirements

For the operation of the model on a given computing platform a basic environment of standard software needs to be available:

- Operating system: A Unix-type environment, including the standard utilities 'make' and 'git'
- Fortran compiler: The kernel of the model is implemented in Fortran, and was tested with the gfortran compiler (versions 9.3.1, 13.2.1), which is available under the GNU license and usually provided by default by every Linux distribution. The model kernel was also tested with the Intel® oneAPI Fortran compiler (versions 2021.7.0, 2021.10.0, 2021.11.1).
- Python: The preprocessor of the model and a basic postprocessing environment are implemented in python. Development and testing was performed with python version 3.6.15, and have subsequently also tested with python version 3.11.8. In addition to libraries already provided by a standard python installation the pre- and postprocessing software requires further libraries, all of which are freely available:
  - netcdf4-python (`https://github.com/Unidata/netcdf4-python`)
  - numpy (`https://numpy.org`)
  - xarray (`https://xarray.dev`)
  - pandas (`https://pandas.pydata.org`)
  - pyproj (`https://github.com/pyproj4/pyproj`)
  - matplotlib (`https://matplotlib.org`)

  - cartopy (`https://scitools.org.uk/cartopy/docs/latest`)

- NetCDF: The I/O of the model core makes use of the NetCDF library (`https://www.unidata.ucar.edu/downloads/netcdf/index.jsp`). It was developed with NetCDF-C version 4.7.4 and NetCDF-Fortran versions 4.5.2. and 4.6.1. (Note that, since major version number 4, the NetCDF library in turn requires the HDF5 library to be installed on the system.)

8.2 Repository

The TCCAS code is hosted by the GitLab repository server of the GWDG, which is the computing and IT competence centre for the Max Planck Society and the university computing centre for the Georg-August-Universität Göttingen.

8.3 Installation

The system code can be cloned from the GitLab repository server, which is typically achieved by the command below:
`git clone git@gitlab.gwdg.de:tccas-team/TCCAS.git [TARGETDIR]`
Specifying a (local) target directory is optional, by default a directory `TCCAS` will be created on your disk. The GWDG GitLab system alternatively offers cloning via http
`https://gitlab.gwdg.de/tccas-team/TCCAS.git`
or the download as an archive (tar or zip format). The system does not require any installation, but compiling and running the system requires an initial adaptation to the new platform, described in the steps outlined below.

1. Change to the generated directory: `cd TCCAS`

2. A file `mk.compile` with the user and platform specific compiler and library settings must be created first. The package contains the file `config/mk.compile-gfortran` as an example. Since gfortran is generally available on any Linux distribution activating this template should be a good starting point which can be done using the command
   `ln -fs config/mk.compile-gfortran mk.compile`.
   This template file makes use of the `nf-config` utility shipped with the NetCDF Fortran library to determine the settings for compilation and linking. However, since the NetCDF Fortran library is not a standard package in many Linux distributions it might be necessary to adapt the macro `NFCONFIG` in this file. When `<NETCDF-FORTRAN-INSTALLDIR>` points to the toplevel directory of a custom NetCDF Fortran library installation, setting
   `NFCONFIG = <NETCDF-FORTRAN-INSTALLDIR>/bin/nf-config`
   should normally be the proper workaround. To also assure that the library is being found at run-time the options for the linker which are defined with macro `LDFLAGS` in the template likely need to be extended as well in this case. This should usually be achieved by adding the flag
   `-Wl,-rpath=<NETCDF-FORTRAN-INSTALLDIR>/lib64`.

3. Check that the model driver can be built successfully. The command `make runmodel.x` must have generated an executable file `runmodel.x` in the current work directory.

4. If the model driver was built successfully the system is prepared to run test experiments as described in Section 9.

9 Preconfigured Examples

This section explains how to run the system. We focus on the demonstration of the technical functionality and deliberately refrain from a scientific discussion of the exemplary results shown. We also note that we may refine several details in the configuration of the experiments, for example the prior uncertainty ranges or the input to the calculation of model uncertainty. As a consequence, the preliminary results and outputs presented here are likely to change as the model develops.

9.1 Makefile

The Makefile provides initial guidance on the main things to do with the system. After successfull installation (Section 8.3) running `make` from the command line will list the targets for some preconfigured examples and give an output similar to the one listed below.

```
================================================
List of Makefile targets:

 Execution targets:

  make xmodel              : Perform model forward run and generate model output.

  make xcost               : Perform model forward run and evaluate cost function.

  make xassi               : Perform assimilation experiment.

 Miscellaneous targets:

  make clean               : remove many reconstructable files (e.g. object files,...)
  make scratch             : remove (hopefully) all reconstructable files


  To indicate the site you want to run at, use the macro DOMAIN (default: ES-LM1),
  e.g. to run at Sodankyla site the commandline would be make xmodel DOMAIN=FI-Sod
  Before switching to another domain use 'make scratch' to delete the previous configuration.
================================================
```

9.2 A forward run

We first perform a forward run at ES-LM1, using the following settings in the namelist file `obs.nml` that controls the observational configuration and selects the SIF source model and the spectral conversion factor (see Section 3.2):

```
&obs
 asssif=.true.
 asslvod=.true.
 asssm=.true.
 assslope=.false.
 assfapar=.true.
 synthetic=.true.
 sifsrc='gu '
 conv743=360.528344
/
```

With these settings the forward run produces synthetic observations of four data streams, at all points in time where there are real observations. At this point, the value of the variable `synthetic` has no effect. The observation operator for SIF uses the source term according to Gu et al. (2019), see 3.2.1.

We now execute the command:

```
make scratch xmodel DOMAIN=ES-LM1
```

where the scratch target makes sure that the configuration previously used is removed and the target xmodel produces the following output:

```
 Reading observational configuration from obs.nml
&OBS
 ASSSIF=T,
 ASSLVOD=T,
 ASSSM=T,
 ASSSLOPE=F,
 ASSFAPAR=T,
 SYNTHETIC=T,
 SIFSRC="gu ",
 CONV743=  360.52834400000000     ,
 /
 nmaxobssif, nmaxobsfapar, nmaxobssm, nmaxobslvod =     4835057    65998427     1858085     2284136
 nobssif, nobsfapar, nobssm, nobslvod =         726         179        1397        1327
 spinup period [hours]:    17520
 reading input/obssif.dat
 observational records read                  :     726
 #obs                                        :     726
 average relative obs. uncertainty [%]       :  51.413
 average relative total data uncertainty [%] :  55.227
 reading input/obslvod.dat
 observational records read                  :    1327
 #obs                                        :    1327
 average relative obs. uncertainty [%]       :  14.611
 average relative total data uncertainty [%] : 101.318
 reading input/obssm.dat
 observational records read                  :    1397
 #obs                                        :    1397
 average relative obs. uncertainty [%]       :  10.068
 average relative total data uncertainty [%] :  61.299
```

```
reading input/obsfapar.dat
there is data in hour        61354  beyond end of assimilation window        61344
observational records read                  :     179
#obs                                        :     179
average relative obs. uncertainty [%]       :  57.727
average relative total data uncertainty [%] :  61.168
filename, exist = obs.b F
Simulation start         : 20150101
Simulation end           : 20211231
Simulation hours         :    61368
Active PFTs              :        2
Number of sample points  :        2
Control vector length    :       56
Simulation vector length :     3629
wrote         3629  synthetic observations to file obs.b
```

The system reports how many data points it reads in for each data stream, and what the average relative observational uncertainty R_{obs} (Eq. (15)) is for each data stream. It further shows the average total data uncertainty R for each data stream, which also takes into account the model uncertainty R_{mod} that is derived as described in Section 5.1.

The system has also stored a vector of synthetic observations in binary format in the file `obs.b` that it placed in the main directory. Section 9.3 demonstrates its use in the assimilation system. In subdirectory `diagout` it has produced the files with output of daily and hourly variables (Section 5.1.4).

9.3 Assimilation of Synthetic Observations

The forward run described in Section 9.2 has written the file `obs.b`, which contains synthetic observations that have been generated from prior parameter values. Our task now is to recover this prior parameter vector from the synthetic observations. The assimilation starts from a first guess of the parameter vector. It then iteratively varies that parameter vector to minimise the misfit to the synthetic data. As first guess we add a perturbation of 1 sigma to each of the parameters. We specify this through the variable `pert` in the namelist file `opt.nml`:

```
&config
! comment
 pert = 10.e-1
!  pert = 0
 itmax = 1000
/
```

`itmax` specifies the maximum number of iterations allowed.

To run the assimilation we execute the command:

```
make xassi DOMAIN=ES-LM1
```

where the clean target removes object files produced from the previous run and the target xassi produces the following output:

```
...
&CONFIG
 PERT=  1.0000000000000000     ,
 ITMAX=1000       ,
 IFIX= 56*0          ,
 /
RUNNING THE L-BFGS-B CODE

           * * *

Machine precision = 2.220D-16
 N =           56     M =           10
 The initial X is infeasible.  Restart with its projection.

At X0         2 variables are exactly at the bounds

At iterate    0    f=  3.96998D+05    |proj g|=  7.88105D+05

At iterate    1    f=  2.04281D+02    |proj g|=  6.38759D+03

At iterate    2    f=  1.78156D+02    |proj g|=  1.31963D+02

...

At iterate  195    f=  2.97229D-03    |proj g|=  6.56580D-01

At iterate  196    f=  2.20345D-03    |proj g|=  8.00452D-01

At iterate  197    f=  2.20181D-03    |proj g|=  1.06322D+00

           * * *

Tit   = total number of iterations
Tnf   = total number of function evaluations
Tnint = total number of segments explored during Cauchy searches
Skip  = number of BFGS updates skipped
Nact  = number of active bounds at final generalized Cauchy point
Projg = norm of the final projected gradient
F     = final function value

           * * *

   N    Tit     Tnf  Tnint  Skip  Nact     Projg        F
   56    197    280    221     0     0   1.063D+00   2.202D-03
  F =   2.2018097965642242E-003
...
```

We have omitted the top lines, because they are identical to those from the model run described in Section 9.2, and also the summary output of iterations 3 to 194. In essence we see that the value

of the cost function was reduced by about 8 orders of magnitude and that of its gradient (projected into the parameter domain inside the bounds) by about 5.

Finally it also returns the final parameter set that it identified as the cost function function minimum in binary (`xopt.b`) and ascii formats:

i	Prior	Posterior	Chg [% of x0]	Chg [% of sx0]
1	0.410000E-04	0.409273E-04	-0.17741	-0.88707
2	0.420000E-04	0.421137E-04	0.27073	1.35364
3	0.100000E+01	0.100097E+01	0.09653	0.48266
4	0.100000E+01	0.100187E+01	0.18669	0.93346
5	0.303881E-03	0.303828E-03	-0.01762	-0.08810
6	0.401608E+00	0.401538E+00	-0.01762	-0.08810
7	0.498051E-01	0.497591E-01	-0.09253	-0.46264
8	0.355357E+00	0.355289E+00	-0.01918	-0.09589
9	0.110432E+01	0.110678E+01	0.22311	1.11554
10	0.664669E-04	0.664549E-04	-0.01805	-0.09027
11	0.587603E-02	0.587500E-02	-0.01762	-0.08810
12	0.687543E-02	0.687422E-02	-0.01762	-0.08810
13	0.247876E-04	0.247833E-04	-0.01762	-0.08810
14	0.510207E-01	0.510117E-01	-0.01762	-0.08810
15	0.222991E+02	0.222952E+02	-0.01762	-0.08810
16	0.743057E+02	0.742852E+02	-0.02757	-0.13787
17	0.220007E+00	0.219191E+00	-0.37078	-1.85391
18	0.741816E+02	0.740864E+02	-0.12841	-0.64203
19	0.157380E+03	0.157294E+03	-0.05478	-0.27389
20	0.546266E+02	0.545861E+02	-0.07415	-0.37073
21	0.676694E+02	0.679545E+02	0.42137	2.10683
22	0.258483E-02	0.258437E-02	-0.01762	-0.08810
23	0.539284E+00	0.539189E+00	-0.01762	-0.08810
24	0.361349E+00	0.361991E+00	0.17787	0.88935
25	0.705851E+00	0.705997E+00	0.02072	0.10361
26	0.100824E+01	0.100817E+01	-0.00751	-0.03754
27	0.732086E-03	0.733345E-03	0.17193	0.85966
28	0.870552E-02	0.870398E-02	-0.01762	-0.08810
29	0.550652E-03	0.550555E-03	-0.01762	-0.08810
30	0.395730E-04	0.395660E-04	-0.01762	-0.08810
31	0.342481E-01	0.342421E-01	-0.01762	-0.08810
32	0.142569E+02	0.142544E+02	-0.01762	-0.08810
33	0.773413E+02	0.775950E+02	0.32799	1.63994
34	0.419267E+00	0.419585E+00	0.07580	0.37898
35	0.280880E+02	0.281060E+02	0.06409	0.32047
36	0.121973E+03	0.122066E+03	0.07663	0.38314
37	0.650476E+02	0.650757E+02	0.04326	0.21630
38	0.469436E+02	0.469360E+02	-0.01625	-0.08123
39	0.345735E+02	0.345678E+02	-0.01646	-0.08231
40	0.363391E+02	0.363332E+02	-0.01631	-0.08157
41	0.349181E+02	0.349119E+02	-0.01762	-0.08810
42	0.673740E+04	0.673715E+04	-0.00362	-0.01811
43	0.128996E+02	0.128973E+02	-0.01762	-0.08810
44	0.118188E+05	0.118168E+05	-0.01762	-0.08810
45	0.609882E+02	0.609800E+02	-0.01345	-0.06727
46	0.248358E+02	0.248315E+02	-0.01751	-0.08757

```
      47     0.106367E+02    0.106349E+02       -0.01762       -0.08810
      48     0.895692E+02    0.895041E+02       -0.07276       -0.36382
      49     0.163513E+02    0.163484E+02       -0.01762       -0.08810
      50     0.154697E+05    0.154670E+05       -0.01762       -0.08810
      51     0.200000E-03    0.200710E-03        0.35519        0.71038
      52     0.120000E+01    0.119577E+01       -0.35276       -0.70552
      53     0.200000E-03    0.200644E-03        0.32221        0.64442
      54     0.200000E+01    0.198734E+01       -0.63322       -1.26644
      55     0.400000E+00    0.401225E+00        0.30622        0.61245
      56     0.100000E+01    0.997139E+00       -0.28612       -0.00286
 writing final control vector to xopt.b
 writing prior uncertainty to sx.b
```

The order of the parameters is:

- D&B core parameters
- L-VOD observation operator parameters (included if `asslvod=.true.`)
- ASCAT slope observation operator parameters (included if `assslope=.true.`)
- SIF observation operator parameters (included if `asssif=.true.`)

Here we run with L-VOD and SIF observation operators. In the above list, parameters 1-50 belong to the core model, parameters 51-55 to the L-VOD observation operator, and parameter 56 to the SIF observation operator.

We see that the experiment was successful in that the final parameter value differs from the true minimum (x_0) by less than 1% for each parameter (last but one column). We note that the exact path of the iterative minimisation procedure through parameter space is impacted by the interplay of the algorithm's branches with the program's rounding properties, so it will depend on the selected platform, compiler, and compiler flags

9.4 Assimilation of Real Observations

9.4.1 Assimilation of multiple data streams

The next example demonstrates the assimilation with real rather than synthetic observations. For this experiment, we would like to start with the best possible first guess of the parameter vector, which is the prior. So, our `opt.nml` needs to specify a perturbation of 0. Further, in our `obs.nml` we need to deactivate the use of synthetic observations.

To run the corresponding assimilation experiment we execute the command:

```
make xassi DOMAIN=ES-LM1
```

This produces the following output:

```
 Reading observational configuration from obs.nml
&OBS
```

```
 ASSSIF=T,
 ASSLVOD=T,
 ASSSM=T,
 ASSSLOPE=F,
 ASSFAPAR=T,
 SYNTHETIC=F,
 SIFSRC="gu ",
 CONV743=  360.52834400000000     ,
 /
...
 Reading test configuration from opt.nml
&CONFIG
 PERT=  0.0000000000000000     ,
 ITMAX=1000       ,
 IFIX= 56*0          ,
 /
RUNNING THE L-BFGS-B CODE

           * * *

Machine precision = 2.220D-16
 N =           56     M =           10

At X0         0 variables are exactly at the bounds

At iterate    0    f=  2.17045D+03    |proj g|=  2.08840D+04

At iterate    1    f=  1.42783D+03    |proj g|=  1.54018D+02

At iterate    2    f=  1.39485D+03    |proj g|=  8.44576D+02

...

At iterate   37    f=  1.00804D+03    |proj g|=  8.75868D+01

At iterate   38    f=  1.00804D+03    |proj g|=  8.71914D+01

At iterate   39    f=  1.00804D+03    |proj g|=  8.71818D+01

           * * *

Tit   = total number of iterations
Tnf   = total number of function evaluations
Tnint = total number of segments explored during Cauchy searches
Skip  = number of BFGS updates skipped
Nact  = number of active bounds at final generalized Cauchy point
Projg = norm of the final projected gradient
F     = final function value

           * * *

   N    Tit     Tnf  Tnint  Skip  Nact     Projg        F
   56     39    125     51     0     0   8.718D+01   1.008D+03
```

```
  F =   1008.0406610660773
```

...

Again, we have omitted the top lines, because they are identical to those from the model run described in Section 9.2, and also the summary output of iterations 3 to 46. The value of the cost function is reduced by about a factor of two and that of its gradient by about a factor of about 200 in 29 iterations with 125 model runs.

The changes of the D&B core parameters (number 1 to 50) are generally moderate. Only parameter 36, which indicates the day of leaf senescence for the grass PFT, experiences a somewhat stronger change of $\sim$ -31%:

i	Prior	Posterior	Chg [% of x0]	Chg [% of sx0]
1	0.410000E-04	0.405216E-04	-1.16694	-5.83470
2	0.420000E-04	0.403398E-04	-3.95297	-19.76485
3	0.100000E+01	0.989673E+00	-1.03268	-5.16340
4	0.100000E+01	0.119076E+01	19.07564	95.37821
5	0.303881E-03	0.303881E-03	0.00000	0.00000
6	0.401608E+00	0.401608E+00	0.00000	0.00000
7	0.498051E-01	0.495394E-01	-0.53360	-2.66800
8	0.355357E+00	0.353484E+00	-0.52708	-2.63539
9	0.110432E+01	0.103983E+01	-5.83927	-29.19636
10	0.664669E-04	0.660880E-04	-0.57017	-2.85084
11	0.587603E-02	0.587603E-02	0.00000	0.00000
12	0.687543E-02	0.687543E-02	0.00000	0.00000
13	0.247876E-04	0.247876E-04	0.00000	0.00000
14	0.510207E-01	0.510207E-01	0.00000	0.00000
15	0.222991E+02	0.222991E+02	0.00000	0.00000
16	0.743057E+02	0.747648E+02	0.61786	3.08928
17	0.220007E+00	0.215047E+00	-2.25436	-11.27182
18	0.741816E+02	0.739571E+02	-0.30271	-1.51357
19	0.157380E+03	0.151876E+03	-3.49723	-17.48614
20	0.546266E+02	0.546517E+02	0.04596	0.22981
21	0.676694E+02	0.694158E+02	2.58076	12.90378
22	0.258483E-02	0.258483E-02	0.00000	0.00000
23	0.539284E+00	0.539284E+00	0.00000	0.00000
24	0.361349E+00	0.349215E+00	-3.35778	-16.78892
25	0.705851E+00	0.700053E+00	-0.82140	-4.10699
26	0.100824E+01	0.101088E+01	0.26165	1.30825
27	0.732086E-03	0.730392E-03	-0.23145	-1.15727
28	0.870552E-02	0.870552E-02	0.00000	0.00000
29	0.550652E-03	0.550652E-03	0.00000	0.00000
30	0.395730E-04	0.395730E-04	0.00000	0.00000
31	0.342481E-01	0.342481E-01	0.00000	0.00000
32	0.142569E+02	0.142569E+02	0.00000	0.00000
33	0.773413E+02	0.763865E+02	-1.23462	-6.17310
34	0.419267E+00	0.405601E+00	-3.25945	-16.29727
35	0.280880E+02	0.278264E+02	-0.93137	-4.65683
36	0.121973E+03	0.836148E+02	-31.44802	-157.24011
37	0.650476E+02	0.557107E+02	-14.35393	-71.76964

```
      38     0.469436E+02    0.504263E+02        7.41890        37.09449
      39     0.345735E+02    0.345721E+02       -0.00396        -0.01979
      40     0.363391E+02    0.363429E+02        0.01054         0.05268
      41     0.349181E+02    0.349181E+02        0.00000         0.00000
      42     0.673740E+04    0.708583E+04        5.17158        25.85788
      43     0.128996E+02    0.128996E+02        0.00000         0.00000
      44     0.118188E+05    0.118188E+05        0.00000         0.00000
      45     0.609882E+02    0.609901E+02        0.00306         0.01530
      46     0.248358E+02    0.248372E+02        0.00551         0.02754
      47     0.106367E+02    0.106367E+02        0.00000         0.00000
      48     0.895692E+02    0.895941E+02        0.02776         0.13878
      49     0.163513E+02    0.163513E+02        0.00000         0.00000
      50     0.154697E+05    0.154697E+05        0.00000         0.00000
      51     0.200000E-03    0.277075E-03       38.53744        77.07489
      52     0.120000E+01    0.135698E+01       13.08144        26.16289
      53     0.200000E-03    0.196761E-03       -1.61933        -3.23866
      54     0.200000E+01    0.180377E+01       -9.81127       -19.62254
      55     0.400000E+00    0.533821E+00       33.45520        66.91040
      56     0.100000E+01    0.877769E+01      777.76940         7.77769
 writing final control vector to xopt.b
 writing prior uncertainty to sx.b
```

Before further discussion of parameter changes, we perform a forward run from the posterior parameter vector in order to analyse the changes in simulated variables. We do this by

1. copying the posterior vector `xopt.b` to the file `x.b` that, if present, the model uses to overwrite the prior parameter settings and
2. executing the command:

```
make clean xmodel DOMAIN=ES-LM1
```

Figure 7 shows prior (red) and posterior (blue) model results and observations (green) at Majadas de Tietar for the four EO data streams that were assimilated. The fit to SIF has very much improved through adjustment of a single scaling factor (parameter 61) which has been increased by a factor of 8.8. Changes to FAPAR and surface layer soil moisture were due to changes in the D&B core parameters: FAPAR was improved by reducing the amplitude of the seasonal cycle. Changes to surface layer soil moisture were small, the RMSE improved by $\sim$ 12%. The calibration of the of the observation operator for L-VOD changed the values of the five parameters in the observation operator (numbers 51 to 55) by up to 39%. This latter parameter quantifies the sensitivity of L-VOD with respect to woody biomass (Section 3.3). With the posterior parameter values (of the core model and the L-VOD observation operator) the fit is strongly improved.

Figure 7 shows prior (red) and posterior (blue) model runs and observations (green) at Majadas de Tietar for two independent data streams: daily NEE and GPP. We note that the NEE data set is based on eddy covariance observations, while the GPP data set is derived from these NEE observations. The fit to NEE is considerably improved through the assimilation. We note, however, that there are parameters affecting heterotrophic respiration (e.g. the initial sizes of the soil organic

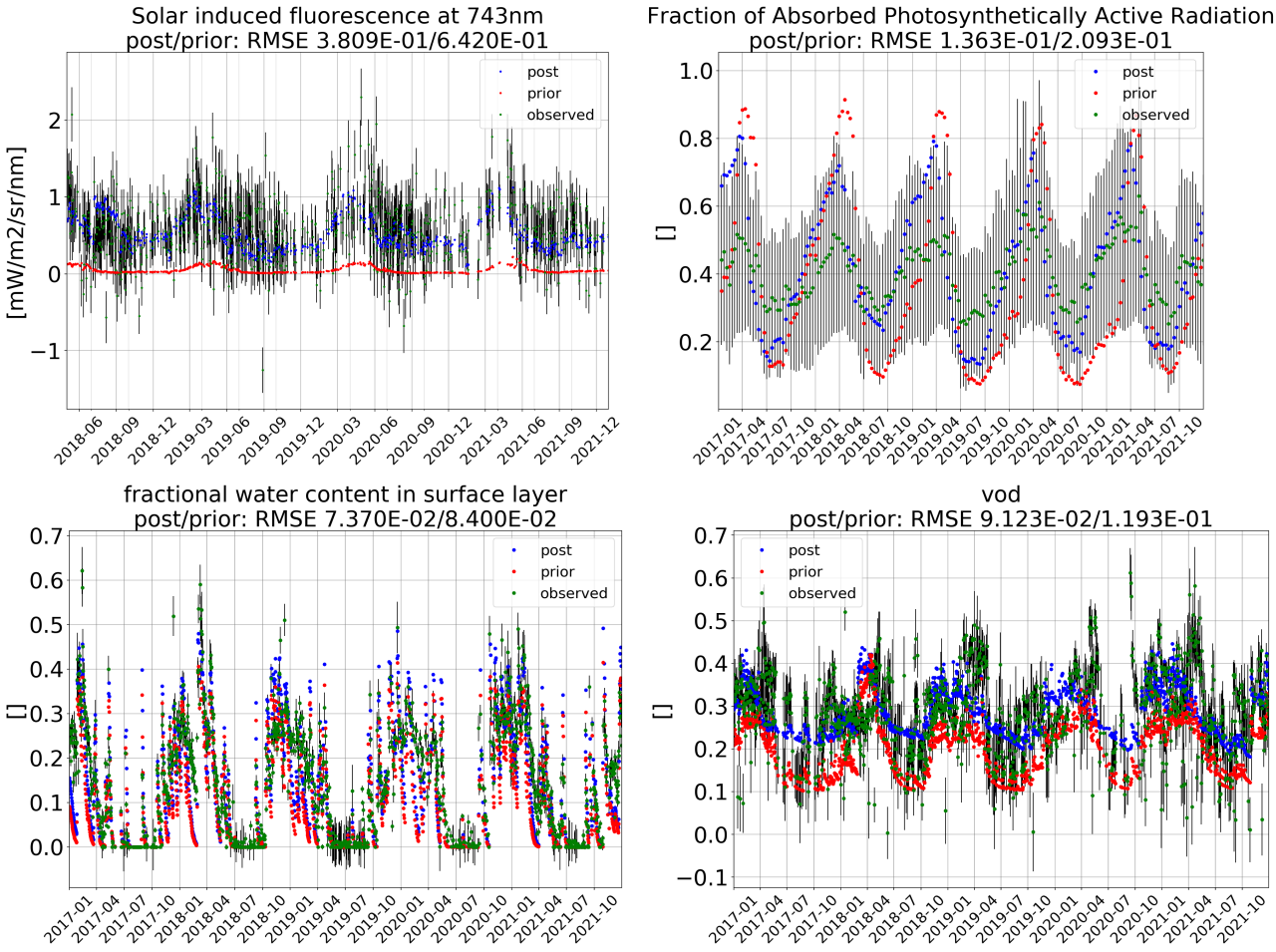


Figure 7: Prior (red) and posterior (blue) model runs and observations (green) at Majadas de Tietar for the four EO data streams that were assimilated: SIF (top left), FAPAR (top right), soil moisture (bottom left), L-VOD (bottom right).

matter pools), which are not constrained by the data streams we assimilated. The match to GPP is also considerably improved.

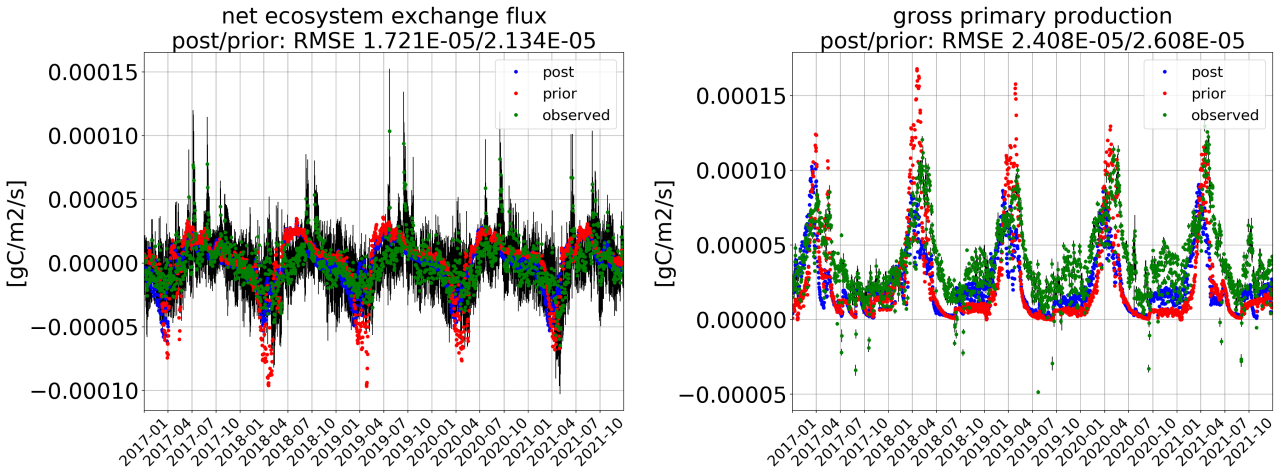


Figure 8: Prior (red) and posterior (blue) simulated and observed (green) values at Majadas de Tietar: Daily NEE (left) and GPP (right). .

9.4.2 Changing the observation operator for SIF

Now we repeat the same experiment, but use a different model for the SIF leaf level source and change to the spectral conversion factor for oak. We do this by specifying `sifsrc='vdt'` instead of `sifsrc='gu '` and `conv743=139.71531` instead of `conv743=360.528344` in the namelist file `obs.nml`:

```
&obs
 asssif=.true.
 asslvod=.true.
 asssm=.true.
 assslope=.false.
 assfapar=.true.
 synthetic=.false.
 sifsrc='vdt'
 conv743=139.71531
/
```

Executing the command:

```
make xassi DOMAIN=ES-LM1
```

produces the following output:

```
 Reading observational configuration from obs.nml
&OBS
 ASSSIF=T,
 ASSLVOD=T,
 ASSSM=T,
 ASSSLOPE=F,
 ASSFAPAR=T,
 SYNTHETIC=F,
 SIFSRC="vdt",
 CONV743=  139.71530999999999     ,
 /

...

At X0         0 variables are exactly at the bounds

At iterate    0    f=  2.38149D+03    |proj g|=  4.31692D+05

At iterate    1    f=  1.30537D+03    |proj g|=  1.63010D+02

At iterate    2    f=  1.29895D+03    |proj g|=  3.46549D+03

...

At iterate   30    f=  1.02569D+03    |proj g|=  4.10907D+02

At iterate   31    f=  1.02568D+03    |proj g|=  3.77740D+02
```

```
At iterate   32    f=  1.02568D+03    |proj g|=  3.76929D+02

          * * *

Tit   = total number of iterations
Tnf   = total number of function evaluations
Tnint = total number of segments explored during Cauchy searches
Skip  = number of BFGS updates skipped
Nact  = number of active bounds at final generalized Cauchy point
Projg = norm of the final projected gradient
F     = final function value

          * * *

   N    Tit     Tnf  Tnint  Skip  Nact     Projg        F
   56     32     82     50     0     0   3.769D+02   1.026D+03
  F =   1025.6797612362052

...
```

i	Prior	Posterior	Chg [% of x0]	Chg [% of sx0]
1	0.410000E-04	0.407312E-04	-0.65556	-3.27778
2	0.420000E-04	0.413499E-04	-1.54778	-7.73892
3	0.100000E+01	0.104863E+01	4.86294	24.31471
4	0.100000E+01	0.119081E+01	19.08088	95.40438
5	0.303881E-03	0.303881E-03	0.00000	0.00000
6	0.401608E+00	0.401608E+00	0.00000	0.00000
7	0.498051E-01	0.496900E-01	-0.23111	-1.15554
8	0.355357E+00	0.354583E+00	-0.21779	-1.08894
9	0.110432E+01	0.102002E+01	-7.63382	-38.16908
10	0.664669E-04	0.663154E-04	-0.22797	-1.13984
11	0.587603E-02	0.587603E-02	0.00000	0.00000
12	0.687543E-02	0.687543E-02	0.00000	0.00000
13	0.247876E-04	0.247876E-04	0.00000	0.00000
14	0.510207E-01	0.510207E-01	0.00000	0.00000
15	0.222991E+02	0.222991E+02	0.00000	0.00000
16	0.743057E+02	0.744984E+02	0.25936	1.29679
17	0.220007E+00	0.218047E+00	-0.89109	-4.45545
18	0.741816E+02	0.738660E+02	-0.42546	-2.12731
19	0.157380E+03	0.157801E+03	0.26756	1.33781
20	0.546266E+02	0.547463E+02	0.21916	1.09581
21	0.676694E+02	0.683649E+02	1.02786	5.13932
22	0.258483E-02	0.258483E-02	0.00000	0.00000
23	0.539284E+00	0.539284E+00	0.00000	0.00000
24	0.361349E+00	0.354252E+00	-1.96406	-9.82030
25	0.705851E+00	0.699077E+00	-0.95960	-4.79798
26	0.100824E+01	0.103240E+01	2.39648	11.98240
27	0.732086E-03	0.729981E-03	-0.28749	-1.43745
28	0.870552E-02	0.870552E-02	0.00000	0.00000
29	0.550652E-03	0.550652E-03	0.00000	0.00000
30	0.395730E-04	0.395730E-04	0.00000	0.00000

```
      31     0.342481E-01    0.342481E-01        0.00000        0.00000
      32     0.142569E+02    0.142569E+02        0.00000        0.00000
      33     0.773413E+02    0.776664E+02        0.42030        2.10150
      34     0.419267E+00    0.415663E+00       -0.85964       -4.29819
      35     0.280880E+02    0.280666E+02       -0.07627       -0.38137
      36     0.121973E+03    0.114618E+03       -6.03008      -30.15039
      37     0.650476E+02    0.639190E+02       -1.73499       -8.67496
      38     0.469436E+02    0.481962E+02        2.66819       13.34096
      39     0.345735E+02    0.345759E+02        0.00701        0.03507
      40     0.363391E+02    0.363442E+02        0.01406        0.07030
      41     0.349181E+02    0.349181E+02        0.00000        0.00000
      42     0.673740E+04    0.687492E+04        2.04116       10.20582
      43     0.128996E+02    0.128996E+02        0.00000        0.00000
      44     0.118188E+05    0.118188E+05        0.00000        0.00000
      45     0.609882E+02    0.609901E+02        0.00302        0.01509
      46     0.248358E+02    0.248368E+02        0.00413        0.02065
      47     0.106367E+02    0.106367E+02        0.00000        0.00000
      48     0.895692E+02    0.896297E+02        0.06753        0.33766
      49     0.163513E+02    0.163513E+02        0.00000        0.00000
      50     0.154697E+05    0.154697E+05        0.00000        0.00000
      51     0.200000E-03    0.236212E-03       18.10601       36.21202
      52     0.120000E+01    0.128448E+01        7.03961       14.07922
      53     0.200000E-03    0.200764E-03        0.38188        0.76375
      54     0.200000E+01    0.196942E+01       -1.52897       -3.05794
      55     0.400000E+00    0.451980E+00       12.99494       25.98988
      56     0.100000E+01    0.486077E+00      -51.39227       -0.51392
 writing final control vector to xopt.b
 writing prior uncertainty to sx.b
```

Here we left out the bits that are identical to the previous experiment and the summary output of iterations 3 to 29. The most remarkable difference to the previous experiment that used the Gu et al. (2019) formulation for the SIF source term is in the posterior value of the scaling factor (parameter 56). While it was 8.8 for the Gu et al. (2019) formulation with the spectral conversion factor for pine, it is now 0.49 for the Van der Tol et al. (2014) formulation and the spectral conversion factor for oak. Some other posterior parameter values changed as well with the formulation of the of the SIF source term and the spectral conversion factor. Overall the formulation of Gu et al. (2019) achieved a slightly lower value of the final cost function than the one by Van der Tol et al. (2014). The fit to the observations that were assimilated (Figure 9) is still much better than for the prior:

Regarding the fit to observations used for validation (Figure 10), the fit to NEE and GPP is improved with respect to the prior but not as much as in the previous experiment that used the Gu model.

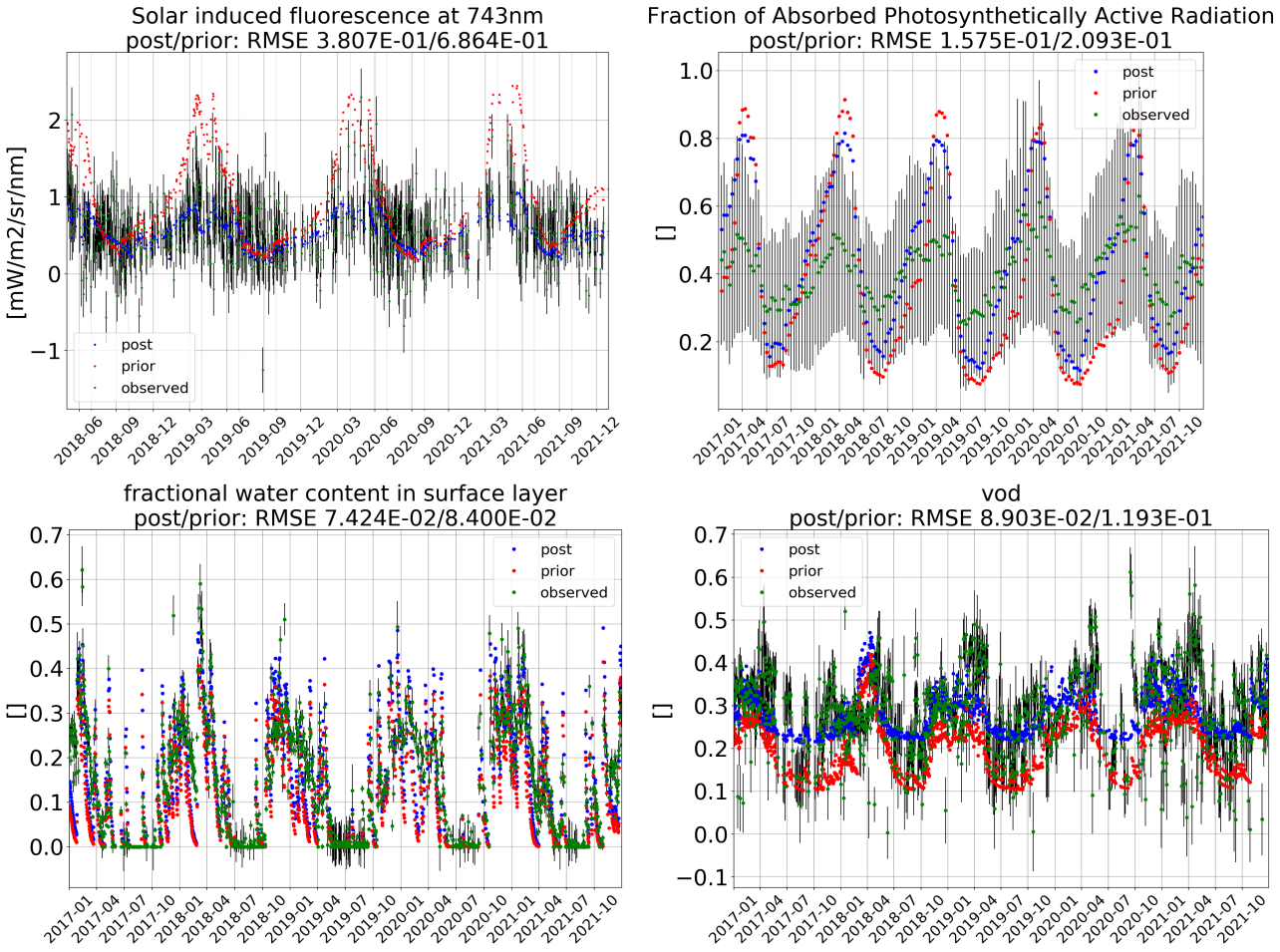


Figure 9: Prior (red) and posterior (blue) model runs and observations (green) at Majadas de Tietar the four EO data streams that were assimilated: SIF (top left), FAPAR (top right), soil moisture (bottom left), L-VOD (bottom right), using the van der Tol et al. (2014) formulation for the SIF source and the spectral conversion for oak.

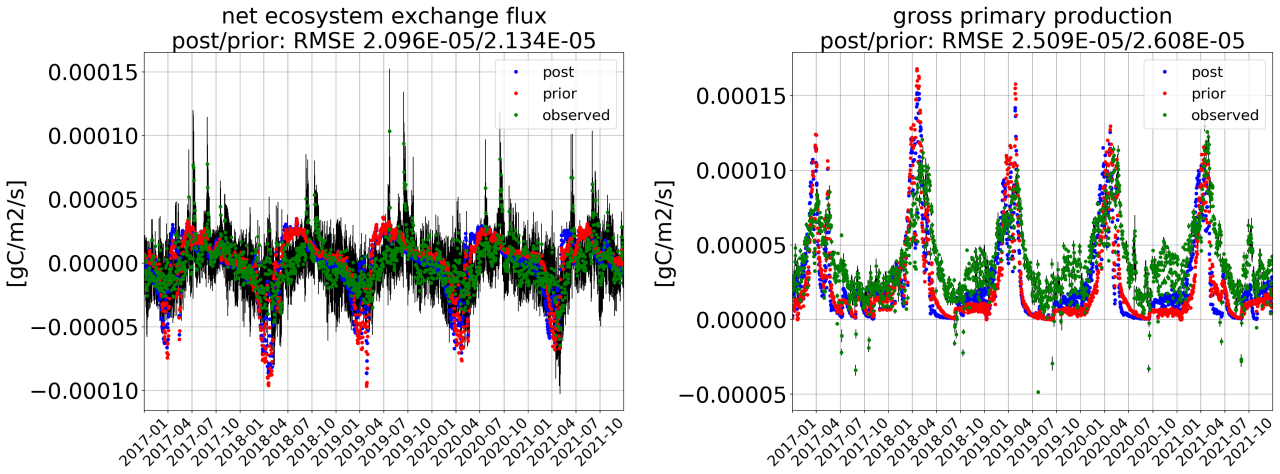


Figure 10: Prior (red) and posterior (blue) simulated and observed (green) values at Majadas de Tietar: Daily NEE (left) and GPP (right). .

9.4.3 Changing the site

Now we conduct the same experiment at Sodankylä. We use the same settings in the namelist files `opt.nml` and `obs.nml` as at Majadas de Tietar (with the initially chose settings for SIF source and spectral conversion, see Section 9.4.1) and execute the command:

```
make scratch xassi DOMAIN=FI-Sod
```

This produces the following output:

```
 Reading observational configuration from obs.nml
&OBS
 ASSSIF=T,
 ASSLVOD=T,
 ASSSM=T,
 ASSSLOPE=F,
 ASSFAPAR=T,
 SYNTHETIC=F,
 SIFSRC="gu ",
 CONV743=  360.52834400000000     ,
 /
 nmaxobssif, nmaxobsfapar, nmaxobssm, nmaxobslvod =     4835057    65998427     1858085     2284136
 nobssif, nobsfapar, nobssm, nobslvod =         445         121         986        1698
 spinup period [hours]:    17520
 reading input/obssif.dat
 observational records read                  :     445
 #obs                                        :     445
 average relative obs. uncertainty [%]       : 130.066
 average relative total data uncertainty [%] : 131.639
 reading input/obslvod.dat
 observational records read                  :    1698
 #obs                                        :    1698
 average relative obs. uncertainty [%]       :  11.611
 average relative total data uncertainty [%] : 100.745
 reading input/obssm.dat
 observational records read                  :     986
 #obs                                        :     986
 average relative obs. uncertainty [%]       :  12.763
 average relative total data uncertainty [%] :  61.539
 reading input/obsfapar.dat
 observational records read                  :     121
 #obs                                        :     121
 average relative obs. uncertainty [%]       :  91.790
 average relative total data uncertainty [%] :  94.053
 filename, exist = obs.b F
 Simulation start         : 20150101
 Simulation end           : 20211231
 Simulation hours         :    61368
 Active PFTs              :        2
 Number of sample points  :        2
 Control vector length    :       56
```

```
 Simulation vector length :     3250
 Reading test configuration from opt.nml
&CONFIG
 PERT=  0.0000000000000000     ,
 ITMAX=1000       ,
 IFIX= 56*0          ,
 /
RUNNING THE L-BFGS-B CODE

           * * *

Machine precision = 2.220D-16
 N =           56     M =           10

At X0         0 variables are exactly at the bounds

At iterate    0    f=  2.03667D+03    |proj g|=  9.66465D+03

At iterate    1    f=  1.86396D+03    |proj g|=  2.29276D+02

At iterate    2    f=  1.77941D+03    |proj g|=  2.09467D+02


...

At iterate   54    f=  7.66548D+02    |proj g|=  4.52700D+01

At iterate   55    f=  7.66520D+02    |proj g|=  2.99949D+01

At iterate   56    f=  7.66519D+02    |proj g|=  1.26622D+01

           * * *

Tit   = total number of iterations
Tnf   = total number of function evaluations
Tnint = total number of segments explored during Cauchy searches
Skip  = number of BFGS updates skipped
Nact  = number of active bounds at final generalized Cauchy point
Projg = norm of the final projected gradient
F     = final function value

           * * *

   N    Tit     Tnf  Tnint  Skip  Nact     Projg        F
   56     56     77     79     1     1   1.266D+01   7.665D+02
  F =   766.51885427316870
```

Here we left in the top lines reporting the properties of the observational data, but again left out the summary output of iterations 3 to 53. The value of the cost function is reduced by $\sim 70\%$ and that of its gradient by roughly 3 orders of magnitude in 56 iterations with 77 model runs.

The changes of the D&B core parameters (number 1 to 50) are moderate, as for Majadas de

Tietar, with the exception of parameter 4 (a simultaneous scaling coefficient for volumetric soil moisture at saturation, field capacity, and wilting point):

i	Prior	Posterior	Chg [% of x0]	Chg [% of sx0]
1	0.290000E-04	0.292791E-04	0.96256	4.81281
2	0.520000E-04	0.444327E-04	-14.55255	-72.76273
3	0.100000E+01	0.110797E+01	10.79727	53.98634
4	0.100000E+01	0.175652E+01	75.65237	378.26183
5	0.456415E-03	0.456415E-03	0.00000	0.00000
6	0.419831E+00	0.419831E+00	0.00000	0.00000
7	0.118098E+00	0.114254E+00	-3.25453	-16.27264
8	0.277021E+00	0.278666E+00	0.59389	2.96943
9	0.118926E+01	0.100333E+01	-15.63433	-78.17166
10	0.125222E-03	0.126091E-03	0.69441	3.47207
11	0.720274E-02	0.720274E-02	0.00000	0.00000
12	0.590690E-02	0.590690E-02	0.00000	0.00000
13	0.156538E-04	0.156538E-04	0.00000	0.00000
14	0.481346E-01	0.481346E-01	0.00000	0.00000
15	0.208338E+02	0.208338E+02	0.00000	0.00000
16	0.156126E+03	0.156580E+03	0.29119	1.45597
17	0.135938E+00	0.132412E+00	-2.59384	-12.96919
18	0.292776E+02	0.286523E+02	-2.13570	-10.67848
19	0.230323E+03	0.245556E+03	6.61384	33.06921
20	0.508555E+02	0.517643E+02	1.78698	8.93491
21	0.361136E+02	0.385099E+02	6.63534	33.17671
22	0.331626E-03	0.331626E-03	0.00000	0.00000
23	0.455470E+00	0.455470E+00	0.00000	0.00000
24	0.166584E+00	0.157846E+00	-5.24534	-26.22669
25	0.320516E+00	0.329425E+00	2.77982	13.89910
26	0.151907E+01	0.134840E+01	-11.23549	-56.17745
27	0.131774E-03	0.133096E-03	1.00256	5.01279
28	0.426638E-02	0.426638E-02	0.00000	0.00000
29	0.131967E-02	0.131967E-02	0.00000	0.00000
30	0.361000E-05	0.361000E-05	0.00000	0.00000
31	0.404480E-01	0.404480E-01	0.00000	0.00000
32	0.209081E+02	0.209081E+02	0.00000	0.00000
33	0.691822E+02	0.700095E+02	1.19579	5.97895
34	0.931292E-01	0.910104E-01	-2.27515	-11.37576
35	0.332853E+02	0.332008E+02	-0.25396	-1.26980
36	0.258468E+03	0.240114E+03	-7.10132	-35.50660
37	0.405744E+02	0.407093E+02	0.33237	1.66185
38	0.698551E+02	0.774492E+02	10.87128	54.35641
39	0.304781E+02	0.304608E+02	-0.05660	-0.28300
40	0.292913E+02	0.292651E+02	-0.08916	-0.44579
41	0.178593E+02	0.178593E+02	0.00000	0.00000
42	0.307224E+04	0.301326E+04	-1.91983	-9.59917
43	0.602393E+02	0.602393E+02	0.00000	0.00000
44	0.409103E+05	0.409103E+05	0.00000	0.00000
45	0.375086E+01	0.373205E+01	-0.50142	-2.50708
46	0.161639E+02	0.158215E+02	-2.11840	-10.59200
47	0.772458E+01	0.772458E+01	0.00000	0.00000
48	0.500704E+03	0.490995E+03	-1.93911	-9.69555

```
      49     0.410396E+02    0.410396E+02         0.00000         0.00000
      50     0.348886E+05    0.348886E+05         0.00000         0.00000
      51     0.200000E-03    0.125612E-03       -37.19383       -74.38765
      52     0.120000E+01    0.144108E+01        20.09002        40.18004
      53     0.200000E-03    0.184859E-03        -7.57046       -15.14092
      54     0.200000E+01    0.000000E+00      -100.00000      -200.00000
      55     0.400000E+00    0.437802E+00         9.45057        18.90113
      56     0.100000E+01    0.523309E+01       423.30898         4.23309
 writing final control vector to xopt.b
 writing prior uncertainty to sx.b
```

Before we continue the discussion of parameter changes, we perform again a forward run from the posterior parameter vector in order to analyse the changes in simulated variables. We do this again by

1. copying the posterior vector `xopt.b` to the file `x.b` that, if present, the model uses to overwrite the prior parameter settings and
2. executing the command:

   ```
   make clean xmodel DOMAIN=FI-Sod
   ```

Figure 11 shows prior (red) and posterior (blue) model results and observations (green) at Sodankylä for the four EO data streams that were assimilated. The large change in the soil properties (together with changes in the other parameters of the core of D&B) moved the simulated surface layer soil moisture much closer to the observations. As for Majadas de Tietar, the fit to SIF has much improved. This improvement was achieved not only through adjustment of the single scaling factor in the observation operator (parameter 61, now increased by a factor of about 5.2) but possibly also through changes in parameters of the core model. FAPAR was reduced, which improved the fit by more than a factor of three. Out of the parameters of the observation operator for L-VOD (number 51 to 55) the largest change was for parameter 54, the scaling factor l_f for the ratio of actual to potential transpiration. The posterior value of l_f is 0, i.e. at the lower boundary that we had specified. This means the assimilation eliminated the impact of the ratio of actual to potential transpiration on simulated L-VOD.

Figure 12 shows prior (red) and posterior (blue) model runs and observations (green) at Sodankylä for the same two independent data streams as for Majadas de Tietar: Daily NEE and GPP. The fit to NEE and to GPP has improved considerably.

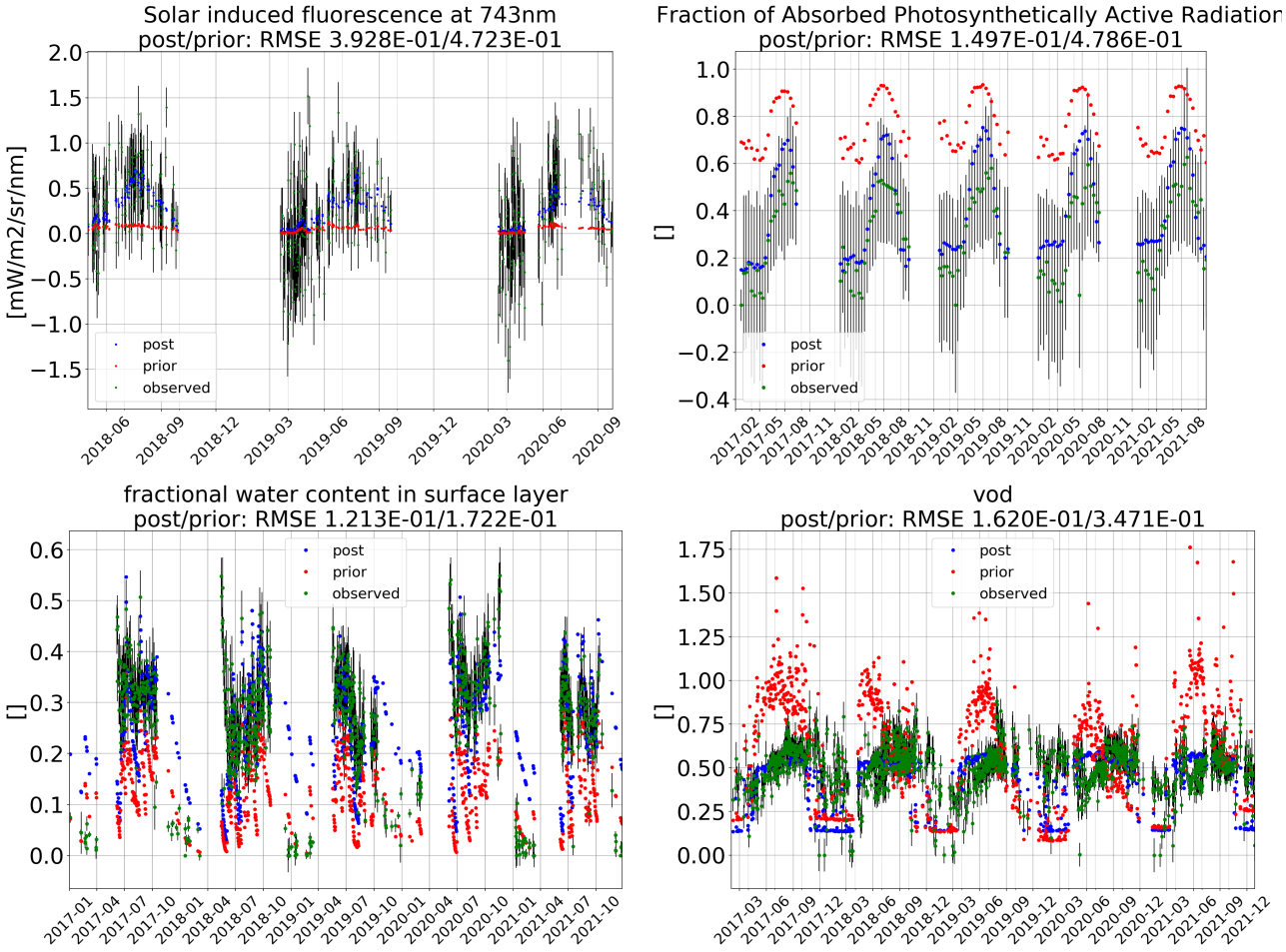


Figure 11: Prior (red) and posterior (blue) model runs and observations (green) at Sodankylä for the four EO data streams that were assimilated: SIF (top left), FAPAR (top right), soil moisture (bottom left), L-VOD (bottom right).

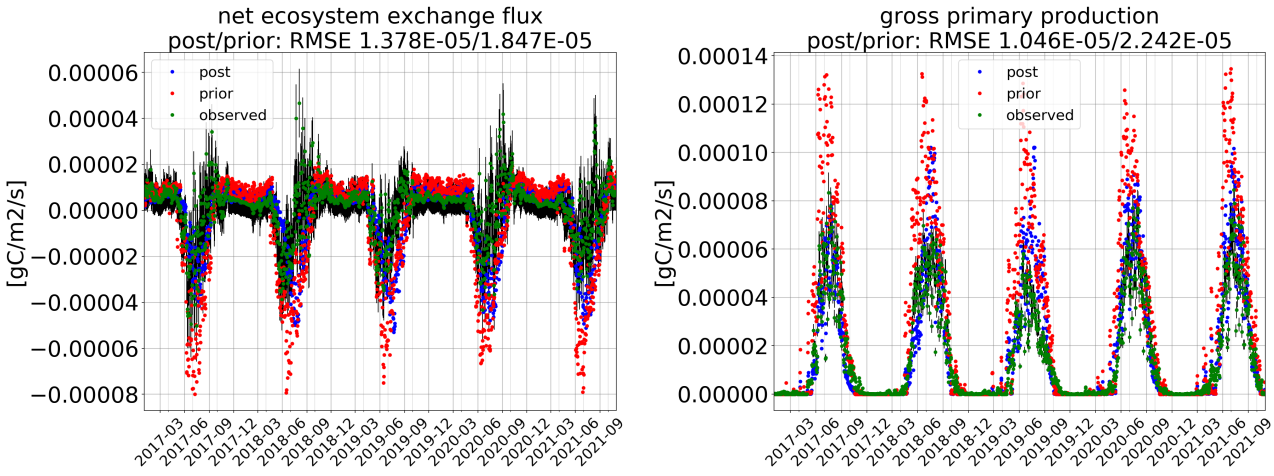


Figure 12: Prior (red) and posterior (blue) simulated and observed (green) values at Sodankylä: Daily NEE (left) and GPP (right).

10 Code Size and Computational Performance

D&B (including observation operators), its tangent and adjoint codes respectively comprise 8,600, 8,500, and 9,500 lines of Fortran code of which are 6,100, 5,500, and 6,500 without comments.

The computational performance of the system has not been systematically assessed, but was tested exemplarily. As an example we provide timing information for a 7 year (2015 to 2021) run of the model and its adjoint (derivative) at one of the study sites (Sodankylä) with two PFTs on a hardware platform equipped with Intel® Core™ i9-7940X processor (19.25M Cache, up to 4.30 GHz), 128GB RAM, and NVMe SSD using a Linux operating system. The model code was compiled with the gfortran compiler (GNU Fortran (SUSE Linux) 13.2.1 20240206 [revision 67ac78caf31f7cb3202177e6428a46d829b70f23]) using some optimising options (`-O3 -funroll-loops -floop-block -ftree-loop-distribution`):

```
CPU time [s]
derivative:                1.49
model     :                0.24
ratio     :                6.11
```

To estimate how the code performance scales when running on larger domains a similar test for the Iberian region (432 grid cells with up to 6 PFTs) was conducted which confirms the good efficiency of the model and its adjoint:

```
CPU time [s]
derivative:             1084.08
model     :              167.69
ratio     :                6.46
```

As another example, an assimilation run at Sodankylä of all five data streams with 19 iterations and 70 model evaluations took 160.54 CPUs on a notebook with Intel® Core™ i7-8550U processor (base frequency 1.8GHz, maximal turbo frequency 4GHz, and 8MB cache), 16GB RAM, and NVMe SSD.

11 Acknowledgements

Sections 2, B, C and much of 3 are taken from Knorr et al. (2025), to which many more authors contributed than those listed as authors of the present document. Development and D&B and TCCAS is funded by ESA through contracts 4000131497 (LCC study) and 4000141232 (TCCAS study) within the Carbon Science Cluster.

References

Anderson, E. A.: A point energy and mass balance model of a snow cover., NOAA Tec. Rep. NWS 19, US Dept. Commerce, Washington D.C., 150 pp, 1976.

Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H., and Tuller, M.: Ground, proximal, and satellite remote sensing of soil moisture, Reviews of Geophysics, 57, 530–616, 2019.

Beerling, D. J. and Quick, W. P.: A new technique for estimating rates of carboxylation and electron transport in leaves of C3 plants for use in dynamic global vegetation models., Global Change Biology, 1, 289–294, 1995.

Bloom, A. A. and Williams, M.: Constraining ecosystem carbon dynamics in a data-limited world: Integrating ecological "common sense" in a model-data fusion framework, Biogeosciences, 12, 1299–1315, doi:10.5194/bg-12-1299-2015, 2015.

Bloom, A. A., Exbrayat, J. F., Van Der Velde, I. R., Feng, L., and Williams, M.: 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 of the United States of America, 113, 1285–1290, doi:10.1073/pnas.1515160113, 2016.

Brooks, A. and Farquhar, G.: Effect of temperature on the CO2/O2 specifity on RBisCO and the rate of respiration in the light, Planta, 165, 397–406, 1985.

Brutsaert, W.: Evaporation into the atmosphere, Reidel Publishing, Dordrecht, Netherlands, 299 pp, 1982.

Buchhorn, M., Smets, B., Bertels, L., Lesiv, M., Tsendbazar, N.-E., Herold, M., and Fritz, S.: Copernicus Global Land Service: Land Cover 100m: collection 2: epoch 2015: Globe, doi:10.5281/zenodo.3243509, URL https://doi.org/10.5281/zenodo.3243509, 2019.

Collatz, G., J.A., B., G.D., F., and J., P.: The relationship between the Rubisco reaction mechanism and models of photosynthesis*, Plant, Cell & Environment, 13, 219–225, doi:https://doi.org/10.1111/j.1365-3040.1990.tb01306.x, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3040.1990.tb01306.x, 1990.

Collatz, G., Ball, J., Grivet, C., and Berry, J. A.: Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer, Agricultural and Forest Meteorology, 54, 107–136, doi:https://doi.org/10.1016/0168-1923(91)90002-8, URL https://www.sciencedirect.com/science/article/pii/0168192391900028, 1991.

Collatz, G. J., Ribascarbo, M., and Berry, J. A.: Coupled Photosynthesis-Stomatal Conductance Model For Leaves Of C_4 Plants, Aust. J. Plant Physiol., 19, 519–538, 1992.

Dickinson, R. E., Henderson-Sellers, A., and Kennedy, P. J.: Biosphere-atmosphere transfer scheme (BATS) version 1e as coupled to the NCAR community climate model, NCAR Technical Note NCAR/TN-387+STR., National Center for Atmospheric Research, Boulder, Colorado, 1993.

ESA: Land Cover CCI, Product User Guide Version 2, ESA, URL https://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf, 2017.

Farquhar, G., von Caemmerer, S. v., and Berry, J.: A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species, Planta, 149, 78–90, 1980.

Farquhar, G. D.: Models relating subcellular effects of temperature to whole plant response, Symposium of the society for Experimental Biology, 42, 395–409, 1988.

Farquhar, G. D. and Sharkey, T. D.: Stomatal conductance and photosynthesis., Ann. Rev. Plant Physiol, 33, 317–345, 1982.

Federer, C.: A soil-plant-atmosphere model for transpiration and availability of soil water, Water Resour. Res., 15, 555–562, 1979.

Federer, C. A.: Transpirational supply and demand: plant, soil, and atmospheric effects evaluated by simulation., Water Resour. Res, 18, 355–362, 1982.

Field, C. and Mooney, H.: The photosynthesis-nitrogen relationship in wild plants, in: On the Economy of Plant form and Function, edited by Givnish, T., pp. 25–55, Cambridge University Press, Cambridge, 1986.

Fischer, R. A. and Turner, N. C.: Plant productivity in the arid and semiarid zones., Ann. Rev. Plant Physiol, 29, 277–317, 1978.

Friend, A. D.: PGEN: an integrated model of leaf photosynthesis, transpiration and conductance., Ecol. Modelling, 77, 233–255, 1995.

Friend, A. D.: Parameterisation of a global daily weather generator for terrestrial ecosystem and biogeochemical modelling., Ecol. Modelling, 109, 121–140, 1998.

Gao, H., Tang, Q., Shi, X., Zhu, C., Bohn, T., Su, F., Sheffield, J., Pan, M., Lettenmaier, D., and Wood, E.: Water budget record form the variable infiltration capcity (VIC) model, algorithm theoretical basis document, University of Washington, USA, 2009.

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., and Justice, C. O.: The Collection 6 MODIS burned area mapping algorithm and product, Remote Sens. Environ., 217, 72–85, doi:10.1016/j.rse.2018.08.005, 2018.

Greenland, D. J. and Kowal, J. M.: Nutrient content of the moist tropical forest of Ghana., Plant and Soil, 12, 154–174, 1960.

Griewank, A.: On Automatic Differentiation, in: Mathematical Programming: Recent Developments and Applications, edited by Iri, M. and Tanabe, K., pp. 83–108, Kluwer Academic Publishers, Dordrecht, 1989.

Gu, L., Han, J., Wood, J. D., Chang, C. Y.-Y., and Sun, Y.: Sun-induced Chl fluorescence and its importance for biophysical modeling of photosynthesis based on light reactions, New Phytologist, 223, 1179–1191, 2019.

Hagemann, S. and Dümenil, L.: Development of a parameterization of lateral discharge for the global scale, Technical Report No. 219, Max-Planck-Institut für Meteorologie, Hamburg, 1996.

Hahn, S., Reimer, C., Vreugdenhil, M., Melzer, T., and Wagner, W.: Dynamic characterization of the incidence angle dependence of backscatter using metop ASCAT, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10, 2348–2359, 2017.

Hansen, M., Potapov, P., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S., Goetz, S., Loveland, T., and et al.: High-Resolution Global Maps of 21st-Century Forest Cover Change, Science, 342, 850–853, 2013.

Harris, I.: A forcings dataset of gridded land surface blend of Climatic Research Unit (CRU) and Japanese reanalysis (JRA) data; Jan.1901 - Dec.2017, Dataset, University of East Anglia Climatic Research Unit, "Centre for Environmental Data Analysis",, doi:10.5285/13f3635174794bb98cf8ac4b0ee8f4ed, URL http://dx.doi.org/10.5285/13f3635174794bb98cf8ac4b0ee8f4ed, 2019.

Hascoët, L. and Pascual, V.: The Tapenade Automatic Differentiation tool: Principles, Model, and Specification, ACM Transactions On Mathematical Software, 39, URL http://dx.doi.org/10.1145/2450153.2450158, 2013.

Heiskanen, J., Rautiainen, M., Stenberg, P., Mõttus, M., Vesanto, V.-H., Korhonen, L., and Majasalmi, T.: Seasonal variation in MODIS LAI for a boreal forest area in Finland, Remote Sensing of Environment, 126, 104–115, doi:https://doi.org/10.1016/j.rse.2012.08.001, 2012.

Hengl, T., Mendes de Jesus, J., Heuvelink, G., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M., Geng, X., Bauer-Marschallinger, B., and et al.: SoilGrids250m: Global gridded soil information based on machine learning, PLoS ONE, 12, 2017.

Houghton, J. T.: The Physics of Atmospheres, Cambridge University Press, Cambridge, 271 pp, 2nd edn., 1986.

Jarvis, P. G.: The interpretation of variations in leaf water potential and stomatal conductance found in canopies in the field., Philos. Trans. R. Soc. London, pp. 593–610, 1976.

Jarvis, P. G. and McNaughton, K.: Stomatal control of transpiration: scaling up from leaf to region, Advances in Ecological Research, 15, 1–49, 1986.

Jones, H. G.: Plants and Microclimate, Cambridge University Press, Cambridge, 323 pp, 1983.

Kaminski, T. and Mathieu, P.-P.: Reviews and syntheses: Flying the satellite into your model: on the role of observation operators in constraining models of the Earth system and the carbon cycle, Biogeosciences, 14, 2343–2357, doi:10.5194/bg-14-2343-2017, URL http://www.biogeosciences.net/14/2343/2017/, 2017.

Kattge, J. and Knorr, W.: Temperature acclimation in a biochemical model of photosynthesis: a reanalysis of data from 36 species, Plant, Cell & Environment, 30, 1176–1190, doi:https://doi.

org/10.1111/j.1365-3040.2007.01690.x, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3040.2007.01690.x, 2007.

Kelliher, F. M., Leuning, R., and Schulze, E.-D.: Evaporation and canopy characteristics of coniferous forests and grasslands., Oecologia, 95, 152–163, 1993.

Kelliher, F. M., Leuning, R. R., R., M., and E.-D., S.: Maximum conductances for evaporation from global vegetation types., Agric. For. Meteorol, 73, 1–16, 1995.

Kerr, Y., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., Gruhier, C., Juglea, S., Drinkwater, M., Hahne, A., Martin-Neira, M., and Mecklenburg, S.: The SMOS Mission: New Tool for Monitoring Key Elements ofthe Global Water Cycle, Proceedings of the IEEE, 98, 666–687, doi:10.1109/JPROC.2010.2043032, 2010.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., Juglea, S. E., et al.: The SMOS soil moisture retrieval algorithm, IEEE transactions on geoscience and remote sensing, 50, 1384–1403, 2012.

Kirdiashev, K., Chukhlantsev, A., and Shutko, A.: Microwave radiation of the earth's surface in the presence of vegetation cover, Radiotekhnika i Elektronika, 24, 256–264, 1979.

Knorr, W.: Satellitengestützte Fernerkundung und Modellierung des Globalen CO_2 -Austauschs der Landvegetation: Eine Synthese, Ph.D. thesis, Max-Planck-Institut für Meteorologie, Hamburg, Germany, 1997.

Knorr, W.: Annual And Interannual CO_2 Exchanges Of The Terrestrial Biosphere: Process-Based Simulations And Uncertainties, Glob. Ecol. Biogeogr., 9, 225–252, 2000.

Knorr, W., Williams, M., Thum, T., Kaminski, T., Voßbeck, M., Scholze, M., Quaife, T., Smallman, T. L., Steele-Dunne, S. C., Vreugdenhil, M., Green, T., Zaehle, S., Aurela, M., Bouvet, A., Bueechi, E., Dorigo, W., El-Madany, T. S., Migliavacca, M., Honkanen, M., Kerr, Y. H., Kontu, A., Lemmetyinen, J., Lindqvist, H., Mialon, A., Miinalainen, T., Pique, G., Ojasalo, A., Quegan, S., Rayner, P. J., Reyes-Muñoz, P., Rodríguez-Fernández, N., Schwank, M., Verrelst, J., Zhu, S., Schüttemeyer, D., and Drusch, M.: A comprehensive land-surface vegetation model for multi-stream data assimilation, D&B v1.0, Geoscientific Model Development, 18, 2137–2159, doi:10.5194/gmd-18-2137-2025, URL https://gmd.copernicus.org/articles/18/2137/2025/, 2025.

Konings, A. G., Rao, K., and Steele-Dunne, S. C.: Macro to micro: microwave remote sensing of plant water content for physiology and ecology, New Phytologist, 223, 1166–1172, doi:https://doi.org/10.1111/nph.15808, URL https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.15808, _eprint: https://nph.onlinelibrary.wiley.com/doi/pdf/10.1111/nph.15808, 2019.

Koster, R. D., Bliss, N. B., Saud, A., and Srooshian, S.: Global soil properties, NASA Global Data Sets for Land-Atmosphere Models, data product accessed at http://iridl.ldeo.columbia.edu/SOURCES/.NASA/.ISLSCP/.GDSLAM/.Hydrology-Soils/.soils/.dataset_documentation.html, 1995.

Li, R., Lombardozzi, D., Shi, M., Frankenberg, C., Parazoo, N. C., Köhler, P., Yi, K., Guan, K., and Yang, X.: Representation of leaf-to-canopy radiative transfer processes improves simulation of far-red solar-induced chlorophyll fluorescence in the community land model version 5, Journal of advances in modeling earth systems, 14, e2021MS002 747, 2022.

Liang, X., Wood, E., and Lettenmaier, D.: Surface soil moisture parameterization of the VIC-2L model: evaluation and modification, Glob. Planet. Change, 13, 195–206, 1996.

Lindroth, A. and Halldin, S.: Numerical analysis of pine forest evaporation and surface resistance., Agric. For. Meteorol, 38, 59–79, 1986.

Loth, B. and Graf, H.: Modelling the snow cover for climate studies, Tech. Rep. 190, Max-Planck-Institut für Meteorologie, Hamburg, Germany, 1996.

Lugo, A. E. and Murphy, P. G.: Nutrient dynamics of a Puerto Rican subtropical dry forest, Journal of Tropical Ecology, 2, 55–72, 1986.

Magney, T. S., Frankenberg, C., Köhler, P., North, G., Davis, T. S., Dold, C., Dutta, D., Fisher, J. B., Grossmann, K., Harrington, A., Hatfield, J., Stutz, J., Sun, Y., and Porcar-Castell, A.: Disentangling Changes in the Spectral Shape of Chlorophyll Fluorescence: Implications for Remote Sensing of Photosynthesis, Journal of Geophysical Research: Biogeosciences, 124, 1491–1507, doi:https://doi.org/10.1029/2019JG005029, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JG005029, 2019.

Meador, W. and Weaver, W.: Two-stream approximations to radiative transfer in planetary atmospheres: A unified description of existing methods and a new improvement, Journal of Atmospheric Sciences, 37, 630–643, 1980.

Mo, T., Choudhury, B. J., Schmugge, T. J., Wang, J. R., and Jackson, T. J.: A model for microwave emission from vegetation-covered fields, Journal of Geophysical Research: Oceans, 87, 11 229–11 237, doi:https://doi.org/10.1029/JC087iC13p11229, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JC087iC13p11229, 1982.

Mohammed, G. H., Colombo, R., Middleton, E. M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., et al.: Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress, Remote sensing of environment, 231, 111 177, 2019.

Monteith, J. L.: Evaporation and environment., Symposium of the Society for Experimental Biology, 19, 205–234, 1965.

Morison, J.: Intercellular CO_2 concentration and stomatal response to CO_2, in: Stomatal Function, edited by Zeiger, E., Farquhar, G. D., and Cowan, I. R., pp. 229–251, Stanford University Press, Stanford, U.S.A, 1987.

Müller, M. J.: Selected climatic data for a global set of standard stations for vegetation science, Junk, Den Haag, Netherlands, 1982.

Murray, F. W.: On the computation of saturation vapour pressure., J. Appl. Meteorol, 6, 203–204, 1967.

Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S., and Wagner, W.: An Improved Soil Moisture Retrieval Algorithm for ERS and METOP Scatterometer Observations, IEEE Transactions on Geoscience and Remote Sensing, 47, 1999–2013, doi:10.1109/TGRS.2008.2011617, 2009.

Paltridge, G. W. and Platt, C. M.: Radiative Processes in Meteorology and Climatology, Elsevier Publishing, New York, 1976.

Paruelo, J. M. and Lauenroth, W. K.: Relative abundance of plant functional types in grasslands and shrublands of North America., Ecol. Appl, 6, 1212–1224, 1996.

Petchiappan, A., Steele-Dunne, S. C., Vreugdenhil, M., Hahn, S., Wagner, W., and Oliveira, R.: The influence of vegetation water dynamics on the ASCAT backscatter–incidence angle relationship in the Amazon, Hydrology and Earth System Sciences, 26, 2997–3019, 2022.

Porcar-Castell, A., Tyystjärvi, E., Atherton, J., Van der Tol, C., Flexas, J., Pfündel, E. E., Moreno, J., Frankenberg, C., and Berry, J. A.: Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: mechanisms and challenges, Journal of experimental botany, 65, 4065–4095, 2014.

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A., and Solomon, A. M.: Special paper: a global biome model based on plant physiology and dominance, soil properties and climate, Journal of biogeography, pp. 117–134, 1992.

Price, J. C. and Bausch, W. C.: Leaf area index estimation from visible and near-infrared reflectance data., Remote Sens. Environ, 52, 55–63, 1995.

Rayner, P. J., Michalak, A. M., and Chevallier, F.: Fundamentals of data assimilation applied to biogeochemistry, Atmospheric Chemistry and Physics, 19, 13 911–13 932, doi:10.5194/acp-19-13911-2019, URL https://www.atmos-chem-phys.net/19/13911/2019/, 2019.

Rodríguez-Fernández, N., Verrelst, J., Lindqvist, H., Barbier, M., Bueechi, E., Munoz, P. R., Mialon, A., Dorigo, W., Bouvet, A., Kaminski, T., and Scholze, M.: Satellite Data Uncertainty analysis (Deliverable 7), Land Surface Carbon Constellation Study Deliverable Report 7, Lund University, Sölvegatan 12, SE-223 62 Lund, 2021.

Rosenberg, N. J.: Microclimate: The Biological Environment, Wiley, New York, 315 pp, 1974.

Running, S., Nemani, R., and R., H.: Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis., Canadian J. Forestry Res, 17, 472–483, 1987.

Ryan, M. G.: Effects of climate change on plant respiration., Ecol. Appl, 1, 157–167, 1991a.

Ryan, M. G.: A simple method for estimating gross carbon budgets of vegetation in forest ecosystems., Tree Physiol, 9, 255–266, 1991b.

Santoro, M.; Cartus, O.: ESA Biomass Climate Change Initiative (Biomass_cci): Global datasets of forest above-ground biomass for the years 2010, 2017 and 2018, v2., Dataset, University of East Anglia Climatic Research Unit, "Centre for Environmental Data Analysis",, 2021.

Scholze, M., Kaminski, T., Knorr, W., Blessing, S., Vossbeck, M., Grant, J., and Scipal, K.: Simultaneous assimilation of SMOS soil moisture and atmospheric CO_2 in-situ observations to constrain the global terrestrial carbon cycle, Remote Sensing of Environment, 180, 334–345, 2016.

Schulze, E.-D.: Carbon dioxide and water exchange in response to drought in the atmosphere and in the soil., Ann. Rev. Plant Physiol, 13, 127–141, 1986.

Schulze, E.-D., Turner, N. C., Gollan, T., and Shackel, K. A.: Stomatal Response to air humidity and to soil drought, in: Stomatal Function, edited by Zeiger, E., Farquhar, G. D., and Cowan, I. R., pp. 311–321, Stanford University Press, Stanford, U.S.A, 1987.

Schulze, E.-D., Kelliher, F. M., Körner, C., Lloyd, J., and Leuning, R.: Relationships among maximum stomatal conductance, ecosystem surface conductance, carbon assimilation rate, and plant nitrogen nutrition: a global ecology scaling exercise., Ann. Rev. Ecol. Syst, 25, 629–660, 1994.

Schwank, M., Kontu, A., Mialon, A., Naderpour, R., Houtz, D., Lemmetyinen, J., Rautiainen, K., Li, Q., Richaume, P., Kerr, Y., and Mätzler, C.: Temperature effects on L-band vegetation optical depth of a boreal forest, Remote Sensing of Environsment, 263, 112 542, doi:https://doi.org/10.1016/j.rse.2021.112542, URL https://www.sciencedirect.com/science/article/pii/S0034425721002625, 2021.

Sellers, P., Randall, D., Collatz, G., Berry, J., Field, C., Dazlich, D., Zhang, C., Collelo, G., and Bounoua, L.: A Revised Land Surface Parameterization (SiB2) for Atmospheric GCMS. Part I: Model Formulation, Journal of Climate, 9, 676 – 705, doi:10.1175/1520-0442(1996)009⟨0676:ARLSPF⟩2.0.CO;2, URL https://journals.ametsoc.org/view/journals/clim/9/4/1520-0442_1996_009_0676_arlspf_2_0_co_2.xml, 1996.

Sellers, P. J.: Canopy reflectance, photosynthesis and transpiration, International Journal of Remote Sensing, 6, 1335–1372, doi:10.1080/01431168508948283, URL https://doi.org/10.1080/01431168508948283, 1985.

Shan, X., Steele-Dunne, S., Huber, M., Hahn, S., Wagner, W., Bonan, B., Albergel, C., Calvet, J.-C., Ku, O., and Georgievska, S.: Towards constraining soil and vegetation dynamics in land surface models: Modeling ASCAT backscatter incidence-angle dependence with a Deep Neural Network, Remote Sensing of Environment, 279, 113 116, 2022.

Sharkey, D. T. and Ogawa, T.: Stomatal Responses to Light, in: Stomatal Function, edited by Zeiger, E., Farquhar, G. D., and Cowan, I. R., pp. 195–208, Stanford University Press, Stanford, U.S.A, 1987.

Smallman, T. L. and Williams, M.: Description and validation of an intermediate complexity model for ecosystem photosynthesis and evapotranspiration: ACM-GPP-ETv1, Geoscientific Model Development, 12, 2227–2253, doi:10.5194/gmd-12-2227-2019, URL https://gmd.copernicus.org/articles/12/2227/2019/, 2019.

Steele-Dunne, S. C., Hahn, S., Wagner, W., and Vreugdenhil, M.: Investigating vegetation water dynamics and drought using Metop ASCAT over the North American Grasslands, Remote Sensing of Environment, 224, 219–235, 2019.

Still, C. J., Berry, J. A., Collatz, G. J., and DeFries, R. S.: Global distribution of C3 and C4 vegetation: carbon cycle implications, Global biogeochemical cycles, 17, 6–1, 2003.

Tarantola, A.: Inverse Problem Theory and methods for model parameter estimation, SIAM, Philadelphia, 2005.

Turner, N. C.: Adaptation to water deficits: a changing perspective., Aust. J. Plant Physiol, 13, 175–190, 1986.

Turner, N. C., Schulze, E.-D., and Gollan, T.: The responses of stomata to vapour pressure deficits and soil water content, I, pp. 338–342, 1984.

Ulaby, F. T. and Wilson, E. A.: Microwave attenuation properties of vegetation canopies, IEEE Transactions on Geoscience and Remote Sensing, pp. 746–753, 1985.

Van der Tol, C., Berry, J., Campbell, P., and Rascher, U.: Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence, Journal of geophysical research: Biogeosciences, 119, 2312–2327, 2014.

van der Tol, C., Berry, J. A., Campbell, P. K. E., and Rascher, U.: Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence, Journal of Geophysical Research: Biogeosciences, 119, 2312–2327, doi:10.1002/2014JG002713, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/2014JG002713, 2014.

Verma, S. B., Baldocchi, D. D., Anderson, D. E., Matt, D. R., and Clement, R. J.: Eddy fluxes of CO_2, water vapor and sensible heat over a deciduous forest., Boundary Layer Meteorol, 36, 71–91, 1986.

Vreugdenhil, M., Dorigo, W. A., Wagner, W., De Jeu, R. A., Hahn, S., and Van Marle, M. J.: Analyzing the vegetation parameterization in the TU-Wien ASCAT soil moisture retrieval, IEEE Transactions on Geoscience and Remote Sensing, 54, 3513–3531, 2016.

Wagner, W., Lemoine, G., Borgeaud, M., and Rott, H.: A study of vegetation cover effects on ERS scatterometer data, IEEE Transactions on Geoscience and Remote Sensing, 37, 938–948, 1999.

Weiss, A. and Norman, J. A.: Partitioning solar radiation into direct and diffuse, visible and near-infrared components., Agric. For. Meteorol, 34, 205–213, 1985.

Wigmosta, M. S., Vail, L., and Lettenmaier, D. P.: A distributed hydrology-vegetation model for complex terrain., Water Resour. Res, 30, 1665–1679, 1994.

Wigneron, J.-P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M.-J., Richaume, P., Ferrazzoli, P., de Rosnay, P., Gurney, R., Calvet, J.-C., Grant, J., Guglielmetti, M., Hornbuckle, B., Mätzler, C., Pellarin, T., and Schwank, M.: L-band Microwave Emission of the Biosphere (L-MEB) Model: Description and calibration against experimental data sets over crop fields, Remote Sensing of Environment, 107, 639–655, doi:https://doi.org/10.1016/j.rse.2006.10.014, URL https://www.sciencedirect.com/science/article/pii/S0034425706004238, 2007.

Wigneron, J.-P., Chanzy, A., Kerr, Y. H., Lawrence, H., Shi, J., Escorihuela, M. J., Mironov, V., Mialon, A., Demontoux, F., De Rosnay, P., et al.: Evaluating an improved parameterization of the soil emission in L-MEB, IEEE Transactions on Geoscience and Remote Sensing, 49, 1177–1189, 2010.

Williams, M., Rastetter, E., Fernandes, D., Goulden, M., Shaver, G., and Johnson, L.: Predicting gross primary productivity in terrestrial ecosystems, Ecological Applications, 7, 882–894, 1997.

Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., and Kurpius, M. R.: An improved analysis of forest carbon dynamics using data assimilation, Global Change Biology, 11, 89–105, 2005.

Wilson, M. F. and Henderson-Sellers, A.: A Global Archive Of Land Cover And Soils Data For Use In General-Circulation Climate Models, Journal of Climatology, 5, 119–143, 1985.

Wood, E., Lettenmaier, D., and Zartarian, V.: A land-surface hydrology parameterization with subgrid variability for general circulation models, J. Geophys. Res., 97, 2717–2728, 1992.

Woodward, F. I.: Climate and Plant Distribution, Cambridge University Press, Cambridge, 1987.

Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-scale Bound-constrained Optimization, ACM Trans. Math. Softw., 23, 550–560, doi:10.1145/279232.279236, URL http://doi.acm.org/10.1145/279232.279236, 1997.

Zobler, L.: A world soil file grobal climate modeling, NASA Tech. memo, 32, 1986.

Appendix

A Variables

Table 10: List of variable names in code and references to model description.

Code	Symbol	Reference	Description
Setup			
Dimensions and resolution			
deltat	Δt	Section B.1	time step
ntstepd	–	–	time steps per day
nsp	–	–	number of subgrid cells
nlayer	N_l	Eq. (79)	number of canopy layers
Site characteristics			
elev	h	Eq. (157)	elevation
fracv	–	Section B.5	PFT area incl. bare soil / grid cell area (sum¡=1)
fc	f_c	Section B.5	fractional ground cover of vegetation
soil_depth	–	Section B.3.5	soil depth to bedrock
Forcing			
Instantaneous meteorological forcing fields			
temp	T	Section B.5	2m air temperature
precip	P_{tot}	Section B.5	precipitation
inrad	R_{sw}	Section B.5	solar incoming radiation
lwdown	$R_{L\downarrow}$	Section B.5	incoming logwave (thermal) radiation
Daily meteorological forcing fields			
dsoiltemp	$\bar{T}_{ds}$	Section B.5	daily mean soil temperature
dtemp	$\bar{T}$	Section B.1	daily mean air temperature
dtempmin	T_{min}	Section B.1	daily minimum air temperature
Carbon			
Canopy characteristics			
vegh	h_v	Table 11	height of canopy
lai	Λ	Eq. (187)	leaf area / PFT area w/ bare soil
dl	l_k/Λ	Eq. (52)	canopy layer partitioning
llai	$l_k - l_{k-1}$	Eq. (79)	LAI by layer
wmax_c	$w_{i,max}$	Eq. (89)	canopy water capacity
Carbon pools			
cpools(:,1)	C_{lab}	Eq. (174)	labile C pool
cpools(:,2)	C_{fol}	Eq. (181)	foliar C pool
cpools(:,3)	C_{fr}	Eq. (188)	fine root C pool
cpools(:,4)	C_{wd}	Eq. (191)	woody C pool
cpools(:,5)	C_{lit}	Eq. (194)	litter C pool
cpools(:,6)	C_{SOM}	Eq. (198)	soil organic carbon (SOC) pool
cpools0	–	Section B.4.2	cpools of previous time step
cgain	–	Section B.4.2	cpools increase after time step
pools, uppools			see cpools
Carbon fluxes (DALEC)			
cfluxes(:,1)	–	–	converted from gpp
cfluxes(:,2)	f_R	Eq. (197)	temperature-dependent rate modifier
cfluxes(:,3)	–	–	converted from raut
–	F_{NPP}	Eq. (64)	cfluxes(:,1) − cfluxes(:,3)
cfluxes(:,4)	P_{fol}	Eq. (182)	leaf C production
cfluxes(:,5)	P_{lab}	Eq. (175)	labile C production
cfluxes(:,6)	P_{fr}	Eq. (189)	root C production
cfluxes(:,7)	P_{wd}	Eq. (192)	wood C production
cfluxes(:,8)	R_{lab}	Eq. (176)	labile C release
cfluxes(:,9)	Φ_{fall}	Eq. (184)	leaf fall factor
cfluxes(:,10)	R_{fol}	Eq. (183)	leaf litter C production

Table 10 – continued from previous page

Code	Symbol	Reference	Description
cfluxes(:,11)	R_{wd}	Eq. (193)	wood litter C production
cfluxes(:,12)	R_{fr}	Eq. (190)	root litter C production
cfluxes(:,13)	R_{lit}	Eq. (195)	litter C respiration
cfluxes(:,14)	R_{SOM}	Eq. (199)	soil organic C respiration
cfluxes(:,15)	F_{decomp}	Eq. (196)	litter to SOM C rate
cfluxes(:,16)	Φ_{onset}	Eq. (177)	labile C release factor
fluxes	–	–	see cfluxes
dcfluxes	–	–	daily sums of C fluxes
Parameters related to carbon turnover (DALEC)			
cpars(:,1)	θ_{SOM}	Eq. (196)	Litter to SOM conversion at 0^oC
cpars(:,2)	–	–	not used
cpars(:,3)	f_{fol}		Photosynthate allocated to foliage
cpars(:,4)	f_{fr}		Photosynthate allocated to fine roots
cpars(:,5)	c_{lf}	Eq. (184)	Inverse of annual fractional leaf loss
cpars(:,6)	θ_{wd}		Turnover rate of wood
cpars(:,7)	θ_{fr}		Turnover rate of fine roots
cpars(:,8)	θ_{lit}		Turnover rate of litter at 0^oC
cpars(:,9)	θ_{SOM}		Turnover rate of SOM at 0^oC
cpars(:,10)	Θ	Eq. (197)	Coefficient on temperature response
cpars(:,11)	–	–	not used
cpars(:,12)	d_{onset}	Eq. (177)	Julian day of peak labile release
cpars(:,13)	f_{lab}	Eq. (175)	Photosynthate allocated to labile
cpars(:,14)	c_{ronset}	Eq. (177)	Labile release period
cpars(:,15)	c_{rfall}	Eq. (184)	Julian day of peak leaf fall
cpars(:,16)	c_{rfall}	Eq. (184)	Leaf fall period
cpars(:,17)	c_{LMA}	Eq. (187)	Leaf C per area
pars			Same as cpars
ff	$[\ln(c_{lf}) - \ln(c_{lf} - 1)]/2$	Eq. (184)	$[\ln(\texttt{cpars}(:,5)) - \ln(\texttt{cpars}(:,5) - 1)]/2$
wf	θ_{fall}	Eq. (184)	$c_{rfall}/\sqrt{2}$
osf	$\psi_c(\ln(c_{ll} - 1), \theta_{res})$	Eq. (177)	correction factor
fl	κ_{leafon}	Eq. (177)	$[\ln(1.001) - \ln(0.001)]/2$
wl	θ_{res}	Eq. (177)	$c_{ronset}/\sqrt{2}$
ml	c_{ll}		minimum labile life span
osl	$\psi_c(\ln(c_{ll} - 1), \theta_{res})$	Eq. (177)	correction factor
sf	σ	Eq. (177)	$365.25/\pi$
Instantaneous C fluxes (not DALEC)			
nee	NEE	Eq. (68)	net ecosystem exchange
gpp	GPP	Eq. (64)	gross primary productivity
lgpp	–	Eq. (64)	GPP by layer
–	A	Eq. (27)	GPP by layer
raut	R_A	Eq. (60)	autotrophic respiration
rhet	R_H	Eq. (201)	soil or heterotrophic respiration
lrleaf	–		$R_d \times (l_k - l_{k-1}) \times M_c$
rleaf	$R_{d,c}$		total leaf respiration
Daily C fluxes (not DALEC)			
dgpp	–	Eq. (68)	daily GPP
draut	–	Eq. (68)	daily autotrophic resp.
drhet	–	Eq. (68)	daily heterotrophic resp.
dnee	–	Eq. (68)	daily NEE
Water			
Water pools			
w_r	W_r	Eq. (85)	root-zone or total soil moisture
w_s	W_s	Eq. (121)	surface soil moisture
w_c	W_i	Eq. (86)	amount of canopy-intercepted water
snow	W_{sn} Eq. (132)		snow mass
Instantaneous water fluxes			
trans	E_t	Eq. (92)	transpiration

Table 10 – continued from previous page

Code	Symbol	Reference	Description
ptrans	E_t^{pot}	Eq. (97)	potential transpiration
evap_s	E_s	Eq. (116)	soil evaporation
evap_sn	E_{sn}	Eq. (136)	snow sublimation rate
pevap_s	E_s^{pot}	Eq. (117)	potential soil evaporation
pevap_sn	E_{sn}^{pot}	Section B.3.7	potential snow sublimation
evap_c	E_i	Eq. (90)	canopy evaporation rate
pevap_c	$E_{v,max}$		potential canopy evaporation
thruf	P_v	Eq. (88)	throughfall of canopy water to ground
rain	P_r	Eq. (82)	rainfall rate
intcp	P_i Eq. (87)		canopy intercepted rainfall
snowf	P_{sn}	Eq. (133)	snowfall rate
smelt	S_m	Eq. (134)	snow melt rate
Daily sums of water fluxes (diagnostic)			
devap_s	–	–	daily soil evaporation, E_s
devap_c	–	–	daily canopy evaporation, E_i
devap_sn	–	–	daily snow sublim., E_{sn}
dpevap_s	–	–	daily pot. soil evap., E_s
dpevap_c	–	–	daily pot. canopy evap., $E_{v,max}$
drain	–	–	daily rainfall, P_r
dsnowf	–	–	daily snowfall, P_{sn}
dtrans	–	–	daily transpiration, E
dthruf	–	–	daily throughfall, P_v
dptrans	–	–	daily pot. transpiration, $E_{t,max}$
dsmelt	–	–	daily snow melt, S_m
Snow model			
snownew	$P_{sn} - E_{sn}$	Eq. (137)	fresh snow
snowh	h_{sn}	Eq. (137)	snow height
dens_sn	ξ_{sn} snow density	kg/m^3	
dens_snnew	ξ_{sn}^n fresh-snow density	kg/m^3	
f_sn	f_{sn}	Eq. (167)	surface snow cover
Derived hydrological variables			
fpasm	f_{soil}^{eff}	Eq. (101)	eff. plant available soil moisture
fwvic	f_{soil}	Eq. (126)	plant available soil moisture
Soil water balance			
root_depth0	–	Table 11	default rooting-depth per PFT
d_r	d_r	Table 11	depth of root zone layer
d_s	d_s	Section B.3.5	depth of surface layer
dsmax	–	Section B.3.5	max. surface layer depth
d_ds	d_{ds}	Eq. (102)	reference depth for dsoiltemp
supp	S	Eq. (100)	root-zone water supply rate
cw0	c_W	Eq. (100)	maximum value for supp
soil_Thetas	Θ_s	Table 14	vol. soil moist. at saturation
soil_Thetaf	Θ_r	Table 14	vol. soil moist, at field capacity
soil_Thetaw	Θ_w	Table 14	vol. soil moist. at wilting point
B_r	B_r		shape param. root-zone
B_s	B_s		shape param. surface layer
kb	k_k		base flow rate (0.2/day)
ks	k_s	Eq. (125)	surface drainage rate
kb_150	$k_{b,150}$	Eq. (125)	parameter for ks (0.095/day)
kb_320	$k_{b,320}$	Eq. (125)	parameter for ks (0.032/day)
Wcr	W_c	Eq. (123)	root-zone moisture storage cap.
Wcs	$W_{c,s}$	Eq. (123)	surface moisture storage cap.
Wfr	W_f	Eq. (124)	root-zone field capacity
Wfs	$W_{f,s}$	Eq. (124)	surface-layer field capacity
Wwr	$W_{w,r}$	Eq. (127)	root-zone wilting point
Qb	Q_b	Eq. (115)	base flow
Qd	Q_d	Eq. (111)	drainage from surf. layer

Table 10 – continued from previous page

Code	Symbol	Reference	Description
Qs	Q_s	Eq. (120)	surface runoff
Processes			
Photosynthesis			
vmax	V_m	Table 11	maximum carboxylation rate at 25^{o}C
be	b_e	Eq. (98)	stomatal-conductance parameter
beopt	$b_{e,opt}$	Eq. (99)	target value of be
x	$e^{-\Delta t/\tau_b}$	Eq. (99)	relaxation par. for b_e
hrmax	$-$	Section B.3.4	start of time step closest to 13:00h
ci	C_i	Section B.2.1	leaf internal CO_2
ci0	$C_{i,0}$	Eq. (47)	default leaf internal CO_2
fci1c3	f_{Ci}	non-stressed leaf internal CO_2 ratio (C3)	0.87
fci1c4	f_{Ci}	non-stressed leaf internal CO_2 ratio (C4)	0.67
nscl	$e^{K_{12}l}$	Eq. (39)	canopy nitrogen scalar
lailim	Λ_{lim}	Eq. (41)	LAI threshold use for nscl
k12	K_{12}	Eq. (41)	modified light extinction at noon
lgpp	$(A+R_d)(l_k-l_{k-1})$	Eq. (62)	actual photosynthesis per layer
lrleaf	$R_d(l_k-l_{k-1})$	Eq. (61)	leaf respiration per layer
ncass	A_c	Eq. (53)	net canopy assimilation
lj	$J_n(l_k-l_{k-1})$	Eq. (37)	electron transport by layer
ljpot	$J(l_k-l_{k-1})$ Eq. (31)		potential el. transp. by layer
ljlim	$-$	Eq. (37)	! carboxylation limited el. transp. rate
a	$A+R_d$	Eq. (54)	gross photosynthesis
rd	R_d	Eq. (59)	dark respiration
j1	J	Eq. (31)	potential electron transport rate
je	J_E or J_e		electron transport limit of a
jc	J_C or J_c		carboxylation limit of a
kc	K_C	Table 12	Michaelis-M. const. for CO_2
ko	K_O	Table 12	Michaelis-M. const. for O_2
gam	$\Gamma_\star$	Eq. (34)	CO_2 compensation point
gamt	γ_T	Eq. (34)	1.7^oC^{-1}
ox	O_x	Eq. (28)	leaf-internal O2 concentration
heatinhib			photosynth. heat inhibition
darkinhib	$\mathcal{D}$	Eq. (36)	photosynth. dark inhibition
darkinhib_c	$1/I_{inh}$	Eq. (36)	dark inhibition param.
thetac4	θ		curve parameter for "je" (C4)
eta	η	Eq. (27)	convexity parameter for je-jc transition
eta2	η_2	Eq. (37)	convexity parameter for electron transport
Radiation			
omega			leaf single-scattering albedo
rho_s	ρ_s	Eq. (167)	soil albedo
rho_s_wet	$\rho_{s,w}$	Eq. (167)	soil albedo
rho_s_dry	$\rho_{s,d}$	Eq. (167)	soil albedo
ff_s	x_w	Eq. (168)	soil dryness factor
rho_sn	ρ_{sn}	Section B.3.7	snow albedo
rho_sn_max	$\rho_{sn,max}$	Section B.3.7	max. snow albedo
rho_bg	ρ_{bg}	Eq. (169)	surface background albedo
rhopar_s	ρ_s^{PAR}	Eq. (77)	PAR surface reflectance
rho_tot	ρ_S	Eq. (166)	albedo (vegetation, soil, snow)
abs_c	a_v	Eq. (165)	solar radiation absorption, canopy
abs_s	a_s	Eq. (164)	solar radiation absorption, soil
abs_s0	$a_{s,0}$	Eq. (164)	abs_s for closed canopy
alb_c0	ρ_v	Eq. (165)	albedo of closed canopy
coszen	μ	Eq. (142)	cosine of solar zenith angle
coszenlim	μ_{lim}	Section B.3.8	coszen lower limit for daytime
relrad	r_{sw}	Eq. (156)	actual / potential solar incoming rad.
pinrad	$R_{PAR}^{pot}+R_{NIR}^{pot}$	$-$	potential incoming solar rad.
par	R_{PAR}	Eq. (144)	incoming PAR

Table 10 – continued from previous page

Code	Symbol	Reference	Description
ppot	R_{PAR}^{pot}	Eq. (145)	potential incoming PAR
ttot	t_{PAR}	Eq. (146)	total (dir.+diff.) PAR transmittance
tdir	$t_{PAR,D}$	Eq. (147)	clear-sky PAR transmittance
tdifmax	$t_{PAR,d}^{max}$	Eq. (146)	PAR max. diffuse transmittance
paec	k_{PAR}	Eq. (147)	PAR atm. extinction coeff.
npot	R_{NIR}^{pot}	Eq. (149)	potential incoming NIR
nttot	t_{NIR}	Eq. (150)	total (dir.+diff.) NIR transmittance
ntdir	$t_{NIR,D}$	Eq. (151)	clear-sky NIR direct transmittance
ntdifmax	$t_{NIR,d}^{max}$	Eq. (150)	NIR max. diffuse transmittance
nawc	w_{NIR}	Eq. (153)	water vapour factor for NIR
ps0	$S_{0,PAR}$		solar constant for PAR
nps0	$S_{0,NIR}$		solar constant for NIR
naec	k_{NIR}		NIR atm. extinction coeff.
pdir	d_{PAR}	Eq. (155)	direct fraction of PAR
apar	I_k^{PAR}	Eq. (78)	absorbed PAR per area by layer
fapar	f_{PAR}	Eq. (80)	fraction of canopy-absorbed PAR (incl. bare soil)
lfapar	I_k^{PAR}/R_{PAR}	Eq. (78)	fapar by layer k
seidistsq	$r_\odot^{-2}$	Eq. (141)	Sun-Earth inv. squ. dist.
latitude	ϕ	Eq. (142)	latitude N
delta	δ	Eq. (143)	solar ecliptic
ih	t_{GMT}		GMT time of day
rn_c	$R_{n,v}$	Eq. (159)	canopy net radiation
rn_s	$R_{n,s}$	Eq. (160)	soil net radiation
rlu	$R_{L\uparrow}$	Eq. (162)	surface thermal upwelling rad.
rln	$R_{L\downarrow} - R_{L\uparrow}$	Eq. (158)	surface thermal net radiation
tl_c	$t_{l,v}$	Eq. (163)	longwave transmissivity of canopy
Micrometeorology			
p_air	p	Eq. (157)	atmospheric pressure
ea	e_a	Eq. (170)	vapour pressure in air
ea0	e_{a0}	Eq. (171)	atm. vapour pressure at sunrise
esta	e_s	Eq. (172)	saturated vapour pressure of air
esta_pos_a	s_1^+	Eq. (172)	constant for esta
esta_neg_a	s_1^-	Eq. (172)	constant for esta
esta_pos_b	s_2^+	Eq. (172)	constant for esta
esta_neg_b	s_2^-	Eq. (172)	constant for esta
esta_c	e_{s0}		constant for esta
reld0	$\hat{h}$	Eq. (170)	parameter for ea
relh0	h_0	Eq. (171)	parameter for ea0
s	s	Eq. (172)	saturated vapour pressure, slope
vpd			vapour pressure deficit in air
temp_c	T_v	Eq. (50)	canopy temperature
temp_sd	–	–	temperature of dry soil surface
temp_sw	–	–	temperature of wet soil surface
ga		Eq. (95)	aerodynamic conductance canopy to air
ga_s		Eq. (118)	aerodynamic conductance soil to air
karman	κ		von Karman (0.41)
href	h_{ref}	Eq. (95)	reference height, surface-air exchange
u0	u	Eq. (95)	wind speed at ref. height
az0	a_{z0}	Eq. (95)	parameter for surface-air exch.
rz0	r_{z0}	Eq. (95)	parameter for surface-air exch.
rough_s		Eq. (118)	surface roughness length of soil
gs	$g_{s,0}(l)$	Eq. (44)	stomatal conductance for vapour
gs	$g_s(l)$	Eq. (48)	stomatal conductance for vapour, no stress
g_c0	$G_{c,0}$	Eq. (45)	canopy conductance, no water stress
g_c	G_c	Eq. (49)	canopy conductance for vapour
hf_s	G	Eq. (161)	soil heat flux
–	H_v	Eq. (130)	sensible heat flux, canopy (not computed)

Table 10 – continued from previous page

Code	Symbol	Reference	Description
–	H_s	Eq. (131)	sensible heat flux, soil (not computed)
Observation operators			
VOD and ASCAT slope observation operators			
vod	τ_λ	Eq. (10)	vegetation optical depth
dvod	–	–	daily average VOD
pool_w	C_{wd}	Eq. (191)	woody biomass (from cpools(:,4))
pool_l	C_{fol}	Eq. (181)	leaf biomass
h_f	f_E	Eq. (11)	plant wetting status
fT	$f(\bar{T})$	Eq. (12)	temperature factor
cbw	l_{wd}	Table 2	woody carbon mass
cbl	l_{fol}	Table 2	foliar carbon mass coeff.
chs	l_s	Table 2	plant available soil moisture coeff.
chf	l_f	Table 2	plant hydrological status coeff.
c0	l_0	Table 2	offset
slope	σ'		slope of ASCAT backscatter-incidence angle
dslope	–	–	daily average of slope
abw	a_{wd}	Table 3	woody carbon mass coff.
abl	a_{fol}	Table 3	foliar carbon biomass coff.
ahs	a_s	Table 3	plant available soil moisture coeff.
ahf	a_f	Table 3	plant hydrological status coeff.
a0	a_0	Table 3	offset
SIF observation operator			
sif	S_n	Eq. (1)	SIF per layer
sif_scale	s_{SIF}	Section 3.2	Scaling factor of leaf-level SIF
J(:,l)	J_n	Eq. (37)	electron transport rate per layer (see lj)
psi_PSIImax	$\psi_{PSIImax}$	Eq. (1)	maximum photochemical quantum yield of photosystem II
q_L	q_L	Eq. (1)	fraction of open photosystem II reaction centres
k_DF	k_{DF}	Eq. (1)	ratio 1st order rate constants heat dissipation / fluorescenc
sif743	–	Section 3.2.4	SIF in 743nm band
leaf_r	$\mathcal{R}_d$	Eq. (202)	leaf reflectance (SIF)
leaf_t	$\mathcal{T}_d$	Eq. (203)	leaf transmittance (SIF)
Others			
Daily diagnostics			
dw_s	–	Section B.3.5	w_s daily mean
dw_r	–	Section B.3.2	w_r daily mean
dfpasm	–	Section B.3.4	fpasm daily mean
dfwvic	–	Section B.3.5	fwvic daily mean
dfapar	–	Section B.2.3	daily mean fapar
Physical constant			
gamma	γ	Eq. (93)	psychrometric constant
lambda	λ	Eq. (96)	latent heat of vaporisation
p0	p_0	Eq. (157)	standard atm. pressure
laps	$\mathcal{L}$	Eq. (157)	atm. lapse rate
gg	g	Eq. (157)	earth gravity constant
sig	σ	Eq. (162)	Stefan-Boltzmann constant
r	R	Eq. (43)	ideal gas constant
cp			dry air specif. heat, constant pressure
mma	M_a	Eq. (51)	molar mass of air
mmc	M_c	Eq. (63)	molar mass of carbon
epar	E_{PAR}		energy content of PAR quanta

B Detailed Process Description of D&B Model

B.1 Time Step

Due its original design to run predominantly on monthly input data, BETHY uses daily time steps for several processes. After integration into D&B, those processes now all run with a single time step, which is one hour by default. However, some legacy model components still require daily variables. These are:

- Computation of (hourly) vapour pressure deficit requires daily minimum temperature, T_{min}, to set vapour press at sunrise (Eq. (171)).
- The snow module, while executed at an hourly time step, uses daily average temperature, $\bar{T}$ (Section B.3.7).
- The VOD and ASCAT slope observation operators likewise use daily average temperature (Section 3.3, Eq. (12)).
- Surface air pressure is computed assuming an ideal constant lapse rate from elevation and daily average temperature (Eq. (157)).

B.2 Photosynthesis and Autotrophic Respiration

B.2.1 The photosynthesis model

Photosynthesis is computed in two steps. The first is the calculation of the non-water-limited net photosynthetic rate, $A_{c,0}$, at a fixed intracellular CO_2 concentration, $C_{i,0}$. Calculation of the actual assimilation rate, A, at the leaf level as a function of the actual intracellular CO_2 concentration, C_i, is here explained for both the C3 and C4 pathway. For C3 plants, following Farquhar et al. (1980), A is formulated as the minimum of an electron transport limited rate, J_E, and a rate, J_C, limited by the carboxylating enzyme Rubisco (in molar units, i.e. mol(CO_2)$m^{-2}s^{-1}$):

$$A(l) = \hat{m}(J_C(l), J_E(l); \eta)\mathcal{H}(T_v) - R_d(l)\mathcal{H}(T_v)\mathcal{D}(I_{PAR}(l)) \tag{27}$$

where l is fractional LAI used as a vertical axis, and

$$J_C(l) = V_m(l)\frac{C_i(l) - \Gamma_\star}{C_i(l) + K_C(1 + O_x/K_O)} \tag{28}$$

$$J_E(l) = J(l)\frac{C_i(l) - \Gamma_\star}{4(C_i(l) - 2\Gamma_\star)} \tag{29}$$

R_d is the leaf or "dark" respiration (see Section B.2.2), and J is the potential electron transport rate. See Table 12 for oxygen concentration, O_x, and the Michaelis-Mention constants K_C and K_O. $\hat{m}(x, y; \eta)$ is a parameterised quasi-minimum function defined as

$$\hat{m}(x, y; \eta) = \frac{x + y - \sqrt{(x + y)^2 - 4\eta xy}}{2\eta} \tag{30}$$

In the first step, at fixed $C_i = C_{i,0}$, we have $\eta = \eta_1 = 0.99$, whereas in the second step step, $\eta = 1$ used, in order to facilitate an analytical solution to computing A at fixed stomatal conductance (see see end of this sub-section).

J is the potential (not limited by carboxylation rate) electron transport rate and depends on the rate of PAR absorption, I, in mol photons m^{-2} s^{-1} in the following way:

$$J = \frac{\alpha I J_m}{\sqrt{J_m^2 + \alpha^2 I^2}} \tag{31}$$

where $I = I_{PAR}/E_{PAR}$, with the PAR absorption rate I_{PAR} in W m^{-2} and E_{PAR} the energy content of PAR quanta, (220 kJ/mol, Jones, 1983, p. 160). The temperature dependence of the maximum electron transport rate, J_m, is calculated according to Farquhar (1988) from the canopy temperature, T_v, in ^{o}C and the respective rate at 25^{o}C:

$$J_m(T_v) = J_m(25^o\text{C}) \times T_v/25 \tag{32}$$

The latter is computed from a PFT-specific $V_m(25^o$C provided in Table 11 and the average of the ratios provided by Kattge and Knorr (2007) for a range of plants:

$$J_m(25^o\text{C}) = 1.97 \cdot V_m(25^o\text{C}). \tag{33}$$

For the CO_2 compensation point without leaf respiration, $\Gamma_\star$ (in μmol(CO_2)mol(air)$^{-1}$), a linear dependence on vegetation temperature (in ^{o}C) is assumed, again following Farquhar (1988):

$$\Gamma_\star = \gamma_T T_v \tag{34}$$

with $\gamma_T = 1.7^o\text{C}^{-1}$.

$\mathcal{H}$ is a function which inhibits assimilation and respiration at temperatures above 55^{o}C taken from Collatz et al. (1991):

$$\mathcal{H}(I) = \frac{1}{1 + e^{1.3(T_v - T_{heat})}} \tag{35}$$

with $T_{heat} = 55^o$C. (D) is a function which inhibits dark respiration in light after Brooks and Farquhar (1985), fitted to inhibit the dark-respiration to 50% of it's uninhibited value up from 50 μmol $m^{-2}s^{-1}$:

$$\mathcal{D}(I) = (1 + e^{-I/I_{inh}})/2 \tag{36}$$

with $I_{inh} = 10^5$ mol m^2 s^{-1}.

The actual electron transport rate, J_n, required for the SIF observation operator, is limited by J_C, and is computed from

$$J_n = \hat{m}(V_m \frac{4(C_i - \Gamma_\star)}{C_i + K_C(1 + O_x/K_O)}, J; \eta_2)\mathcal{D}(T_v) \tag{37}$$

with $\eta_2 = 0.99999$ instead of 1, for computational reasons.

At sufficient light, the rate of photosynthesis is limited by V_m, the maximum turnover rate of the primary CO_2 fixating enzyme, Rubisco, while at low light levels, J_m is limiting the assimilation rate.

Table 11: D&B vegetation parameters. d_r maximum rooting depth, h_v vegetation height, both in m, V_m maximum carboxylation rate in μmol(CO2)/m^{2}s.

PFT	C3/C4	d_r	h_v	$V_m(25^o\text{C})$
Tropical broadleaf evergreen tree	C3	3.0	30	60
Tropical broadleaf deciduous tree	C3	3.0	15	90
Temperate broadleaf evergreen tree	C3	1.5	15	41
Temperate broadleaf deciduous tree	C3	1.5	15	35
Evergreen coniferous tree	C3	1.0	15	29
Deciduous coniferous tree	C3	1.0	15	53
Evergreen shrub	C3	1.5	1	52
Deciduous shrub	C3	1.5	1	160
C3 grass	C3	0.5	1	42
C4 grass	C4	0.5	1	8
Tundra	C3	0.3	0.3	20
Wetland	C3	0.3	0.3	20
Arable crop	C3	0.3	0.6	117

Following Farquhar et al. (1980) and in accordance with Beerling and Quick (1995), a sharp transition from Rubisco to light limited photosynthesis is assumed. In nature, this so-called 'Blackman' curve is observed with a certain transition zone, where both rates are simultaneously limiting. The argument used by Farquhar et al. (1980) is that this co-limitation is a suboptimal behaviour that tends to be minimised (Collatz et al., 1990). Nonetheless, a certain co-limitation is often introduced by a curve parameter (e.g Farquhar et al., 1980; Collatz et al., 1991). Because its actual size is difficult to determine for a global simulation, and because the result does not deviate much from that according to Eq. (27), this parameter is not used here.

The values for V_m and J_m are PFT specific for the standard temperature of 25^{o}C (Table 11). α is assigned a value related to incoming light taken from Beerling and Quick (1995) divided by a leaf absorption of 0.86 according to Collatz et al. (1991). The temperature dependence of V_m and all other rates with an activation energy given in Table 12 is computed from the following equation (with T_v in ^{o}C):

$$k(T_v) = k(25^o\text{C}) \exp\left\{\frac{(T_v - 25)E}{298R(T_v + 273)}\right\} \tag{38}$$

where k stands for the rate in question. Rates and constants that do not depend on vegetation type are also listed in Table 12.

In closed canopies, the light-saturated assimilation rate is normally differentiated according to light availability. Therefore, in accordance with Sellers et al. (1996), an exponential reduction in V_m and J_m from top to bottom is assumed for a leaf area index (LAI) greater than a value of $\Lambda_{lim} = 3$:

$$V_m(l) = V_m \times K_{12} e^{-K_{12} l} \tag{39}$$

$$J_m(l) = J_m \times K_{12} e^{-K_{12} l} \tag{40}$$

where l is the fraction of full LAI used as a vertical coordinate, and V_m and J_m are the temperature dependent values according to Eq. (38) and Eq. (32). K_{12} the extinction coefficient at noon, modified such that the K_{12} approaches 0 for values of the LAI, Λ, smaller than a threshold value of

Table 12: Values of the kinetic parameters and constants for the calculation of the C3 photosynthetic rate with the Farquhar model; lower part: additional constants for C4 photosynthesis. E is the activation energy in J/mol; s.t.: see text, const.: no temperature dependence, F&a: Farquhar et al. (1980), C: Collatz et al. (1992), F: Farquhar (1988), BQ: Beerling and Quick (1995)

symbol	description	value at 25°C	unit	E	reference
α	efficiency of photon capture	0.28	-	(const.)	BQ
$\Gamma_\star$	CO_2 compensation point	42.5	μmol m^{-2} s^{-1}	s.t.	F
O_x	O_2 partial pressure	0.21	mol/mol	35948	F&a
K_C	Michaelis-Menten constant for CO_2	460	μmol/mol	59356	F&a
K_O	Michaelis-Menten constant for O_2	0.33	mol/mol	35948	F&a
V_m	carboxylation capacity	s.t.	μmol m^{-2} s^{-1}	see Table 11	BQ F&a
J_m	electron transport capacity	s.t.	μmol m^{-2} s^{-1}	see Eq. (33)	BQ F
R_d	leaf or dark respiration	s.t.	μmol m^{-2} s^{-1}	50967	C
α_i	integrated C4 quantum efficiency	0.04	mol/mol	(const.)	C
k	PEPcase CO_2 specificity	s.t.	mmol m^{-2} s^{-1}	50967	C
θ	curve parameter for J_e	0.83	-	-	C

$\Lambda_{lim} = 3$: according to

$$K_{12} = \frac{1}{2\mu} \frac{1}{1 + e^{-2(\Lambda - \Lambda_{lim})}} \tag{41}$$

with $\mu = \mu$(12:00h) according to Eq. (142). This scaling of photosynthetic capacity is applied to trees and shrubs but not to the remaining PFTs.

The non-limited or potential photosynthesis rate, A_0, is first calculated from Eq. (27) to Eq. (29) with $C_i = C_{i,0}$; then the canopy rate is formed as an integral over the leaf area (cf. Eq. (45)):

$$A_{c,0} = \int_0^\Lambda A_0 \left(I_{PAR}(l)\right) dl, \tag{42}$$

where Λ is the leaf area index. This integral is solved approximately by summing over N_l (usually 3) layers, each with PAR absorption, I_{PAR}, calculated separately.

The photosynthetic rate, A, can also be expressed by the following diffusion equation (in mol(CO_2)m^{-2}s^{-1}, e.g. Jones, 1983):

$$A = 0.625 g_s (C_a - C_i) \frac{p}{R T_K}, \tag{43}$$

with air pressure p (from Eq. (157), Section B.3.8), the ideal gas constant R (8.314 J K^{-1}mol^{-1}), and air temperature in Kelvin T_K ($T_K = T + 273$). The factor 0.625 takes the lower diffusivity of CO_2 against water vapour into account. C_a and C_i are the CO_2 concentrations of free air and of air within the intracellular air spaces in mol(CO_2)/mol(air). The factor behind the brackets is given for the conversion into units of mol(CO_2)/m^3(air) (= 40.9 mol(air)/m^3(air) at 25°C and standard pressure). This equation can now be used to derive the value of stomatal conductance under conditions without moisture stress.

Assuming a typical value for the intracellular CO_2 concentration under such conditions, $C_{i,0}$, the

non-water-limited stomatal conductance yields (assuming $g_s = g_{s,0}$):

$$g_{s,0} = \frac{1.6A_0}{C_a - C_{i,0}} \frac{RT_K}{p}, \tag{44}$$

and after integration across the canopy to obtain canopy conductance:

$$\begin{aligned} G_{c,0} &= \frac{1.6}{C_a - C_{i,0}} \int_0^{\Lambda} A_0(l) dl \frac{RT_K}{p} \\ &= \frac{1.6A_{c,0}}{C_a - C_{i,0}} \frac{RT_K}{p} \end{aligned} \tag{45}$$

A dependence of g_s on the following external factors is known (Farquhar and Sharkey, 1982): Light (Sharkey and Ogawa, 1987), intracellular CO_2 concentration (Morison, 1987), potential evapotranspiration, i.e. leaf-to-air gradient of vapour pressure (Fischer and Turner, 1978; Schulze, 1986; Schulze et al., 1987) and soil water content (Schulze, 1986; Turner, 1986; Schulze et al., 1987). According to Schulze et al. (1994), its maximum value, $g_{s,max}$, i.e. the value at sufficient light and water supply, increases with leaf nitrogen content. Since maximum photosynthetic rate, A_{max}, also increases with leaf nitrogen (Field and Mooney, 1986), there is a close relationship between $g_{s,max}$ and A_{max}.

There are two important assumptions contained in this last equation: First, the integrated conductance of a canopy does not, as assumed by Woodward (1987) in a modelling study on water limited LAI, increase linearly with the leaf area index, Λ. In fact, like $A_{c,0}$, $G_{c,0}$ saturates at high values of LAI, reaching approximately three times the maximum stomatal conductance, $g_{s,0}$. This is the result of a literature review on field measurements by Kelliher et al. (1995). If soil evaporation is also included, the total surface conductance appears to be largely independent of the LAI. The consequence is, that evapotranspiration from vegetated areas is not controlled by LAI, but by the net radiation, R_n. Therefore, a comprehensive description of the energy balance is a prerequisite for mechanistic modelling of the coupled system of evapotranspiration and photosynthesis.

Second, the Eq. (45) suggests a linear relationship between maximum photosynthetic rate and maximum canopy conductance. Such a relationship is the result of an overview by Schulze et al. (1994). Equating the terms "maximum surface conductance" and "maximum canopy CO_2 assimilation" with $G_{c,0}$ and $A_{c,0}$, respectively, the values for C3 plants cited in Schulze et al. (1994) yield

$$G_{c,0} = 0.883 A_{c,0} \tag{46}$$

with $G_{c,0}$ in mm/s and $A_{c,0}$ in μmol(CO_2)m^{-2}s^{-1}. Inserting this into Eq. (45) yields $C_a - C_{i,0} = 45$ ppm or $C_{i,0}/C_a = 0.87$ (at $C_a = 355$ ppm, 25^{o}C and standard pressure), as assumed by the photosynthesis model for conditions without water stress. This is also supported by the common observation that the ratio of leaf internal to external CO_2 concentration stays nearly constant when incident light intensity or external CO_2 concentration changes (Morison, 1987). Therefore, we set $C_{i,0}$ to a value proportional to the atmospheric CO_2 concentration, C_a:

$$C_{i,0} = f_{Ci} C_a \tag{47}$$

with $f_{Ci} = 0.87$ for all C3 plants. For C4 plants, this ratio is set to $f_{Ci} = 0.67$, based on the relationship for tropical grasses. A lower value for C4 against C3 plants also follows from the different enzyme kinetics of CO_2 uptake, and has been measured regularly, cf. Morison (1987).

Since the leaf or canopy temperature, T_v, depends on the actual rate of transpiration, which in this step has not yet been determined, calculation of $A_{c,0}$ and $G_{c,0}$ up to this point uses the approximation $T_v = T$. A self-consistent calculation, where T_v has to be reinserted iteratively into the equation for $A_{c,0}$, is not carried out here. Since the calculation of $A_{c,0}$ and $G_{c,0}$ at a uniform $C_{i,0}$ has mainly the purpose of formulating a generally applicable dependence of the stomatal conductance on incident light, the error involved is comparatively small, except under significant water stress.

For the second step, the stomatal conductance is modified using the empirical multiplier $1/(1 + b_e\Delta e)$, which is constant across the canopy and therefore applies equally also to $G_{c,0}$ (Eq. (45)). Δe is the vapour pressure deficit in the surrounding air (see Section B.3.9), and b_e an empirical parameter that is set such that the rate of transpiration does not exceed the water supply rate from the root zone, as explained in Section B.3.4. Using this multiplier and the inversion of the diffusion equation for CO_2, we can now determine actual stomatal conductance as

$$\begin{aligned} g_s(l) &= g_{s,0}(l)\frac{1}{1+b_e\Delta e} \\ &= \frac{1.6A_0(l)}{C_a - C_{i,0}}\frac{RT_K}{p}\frac{1}{1+b_e\Delta e} \end{aligned} \tag{48}$$

and actual canopy conductance as:

$$G_c = G_{c,0}\frac{1}{1+b_e\Delta e}. \tag{49}$$

Once this is known, the water balance of the soil and canopy can be computed as described in Section B.3.4, and with it the rate of transpiration, E_t (according to Eq. (92)) and also the canopy temperature, T_v, as

$$T_v = T + \frac{R_{n,v} - \lambda E_t}{\rho_a c_p G_a}, \tag{50}$$

where $R_{n,v}$ is the net radiation of the canopy (Eq. (159)), λ the latent heat of evaporation (Eq. (96)), G_a aerodynamic conductance (Eq. (95)), and c_p the specific heat of air at constant pressure (≈ 1005 J kg^{-1}K^{-1}). The density of air, ρ_a (≈ 1.29 kg m^{-3}), follows

$$\rho_a = \frac{M_a p}{RT_K} \tag{51}$$

with the molar mass of air ($M_a = 28.964 \times 10^{-3}$ kg mol^{-1}). The influence of skin reservoir evaporation (E_i) on T_v is neglected here (i.e. $E_v = E_t$).

The actual assimilation rate, A, is eventually computed at a fixed stomatal resistance, g_s, and at a canopy temperature, T_v, computed as described above. The relevant equations are on the one hand Eq. (27) to Eq. (29) of the Farquhar model, on the other hand the diffusion equation for CO_2:

$$A(l) = 0.625 g_s(l)(C_a - C_i(l)) \frac{p}{RT_{K,v}}, \tag{52}$$

where $T_{K,v}$ is canopy temperature in Kelvin.

Besides A, the undetermined variable is C_i. Eq. (52) is therefore solved for C_i and inserted into Eq. (28) and Eq. (29), with the necessary simplifying assumption of $\eta = 1$. The results are quadratic equations for J_C and J_E; the minimum of the respective lesser solutions of those quadratic equations yields $A(l)$.

Finally, the canopy photosynthesis, A_c in mol(CO_2)$m^{-2}s^{-1}$, is taken as the integral over the leaf area:

$$A_c = \int_0^\Lambda A(l) dl = \frac{0.625p}{RT_K} \int_0^\Lambda g_s(l) \left(C_a - C_i(l)\right) dl \tag{53}$$

For C4 photosynthesis, Eq. (27) to Eq. (29) are replaced according to Collatz et al. (1992) by the following:

$$\begin{aligned}
A(l) &= \min\{J_c(l); J_e(l)\} - R_d(l) && (54)\\
J_c(l) &= kC_i(l) && (55)\\
J_e(l) &= \frac{1}{2\theta}\left[V_m(l) + J_i(l) - \sqrt{(V_m(l) + J_i(l))^2 - 4\theta V_m(l) J_i(l)}\right] && (56)\\
J_i(l) &= \alpha_i \frac{I_{PAR}(l)}{E_{PAR}} && (57)
\end{aligned}$$

Again, photosynthesis rates vary across the canopy due to varying light, C_i, (due to different stomatal conductance, $g_s(l)$, and V_m. As with C3 photosynthesis, a gradual onset of light limitation is assumed, with a sudden transition from J_c to J_e limitation at rising C_i. Another reason for not using a curve parameter for this transition is mathematical: thus, after calculation of $g_s(l)$ as above, $A(l)$ can be derived from:

$$A(l) = \min\left\{J_e; \quad \frac{C_a + R_d/g_s'(l)}{1/k + 1/g_s'(l)}\right\} - R_d \tag{58}$$

with $g_s' = 0.625 g_s p/(RT_{K,v})$.

B.2.2 Autotrophic respiration

Following Farquhar et al. (1980), leaf or dark respiration, R_d, per leaf area at 25^{o}C is assumed proportional to V_m, also at 25^{o}C. The constant of proportionality depends on photosynthetic pathway and is 0.011 for C3 plants (Farquhar et al., 1980; Collatz et al., 1991). For C4 plants we expect a higher value owing to the more complex two-step photosynthetic system. Therefore, the value for C4 grasses is chosen such as that simulated GPP and NPP for C3 and C4 grasses are equal in areas where they were found to be equally abundant. C3 and C4 grass GPP, NPP and abundance were calculated by Knorr (1997), with abundances based on Paruelo and Lauenroth (1996).

The resulting formulation for leaf respiration used here is:

$$R_d(25^o\text{C}) = \begin{cases} 0.011V_m(25^o\text{C}) & \text{(C3)} \\ 0.042V_m(25^o\text{C}) & \text{(C4)} \end{cases} \tag{59}$$

The temperature dependence of R_d is again given by Eq. (38) with an activation energy according to Table 12.

Following Ryan (1991a), total plant or autotrophic respiration, R_A, is divided into two parts, maintenance (R_M) and growth respiration (R_G):

$$R_A = R_M + R_G. \tag{60}$$

The difference between the two is that the so-called growth respiration occurs only when NPP > 0. $R_{d,c}$ is the integral of R_d over the differential leaf area index, l:

$$R_{d,c} = \int_0^\Lambda R_d(l)dl \tag{61}$$

and is assumed to constitute a large part of R_M. Λ is the total leaf area index (LAI). As explained in Section B.2.3, the integral is approximated by a fixed number of discrete layers. A note on definitions: we define the gross primary productivity (GPP) as the gross photosynthesis, excluding leaf respiration, so that total GPP of a given area is calculated as

$$\text{GPP} = A_c + R_{d,c} = \int_0^\Lambda (A(l) + R_d(l))dl. \tag{62}$$

Ryan (1991a) stresses that R_M and the nitrogen content of vegetation, N_{tot}, are usually much better correlated than R_M and biomass, with R_M approximately 0.30μmol(CO_2)m^{-2}s^{-1} per gN at 25^{o}C, if we assume the same temperature dependence as for R_d. Since V_m is nearly proportional to the nitrogen content of leaves (Farquhar et al., 1980) with around 45μmol(CO_2)m^{-2}s^{-1} per gN (at 25 ^{o}C, with 20% of N in Rubisco), it follows from Eq. (59) for C3 plants that R_d is approximately 0.5μmol(CO_2)m^{-2}s^{-1} per gN, somewhat more than R_M. It seems that leaves, in terms of their nitrogen content, take up a higher proportion of total plant respiration than the remaining plant parts. Further, from the data by Ryan (1991b) it follows that typically 40% of maintenance respiration takes place in the leaves. For this reason, the following equation is used:

$$R_M = M_c R_{d,c}/f_{R,leaf} = 1.67 M_c R_{d,c}/f_{N,leaf} \tag{63}$$

$f_{R,leaf}$ is the leaf fraction of the plant-total maintenance respiration, and $f_{N,leaf}$ the leaf fraction of total nitrogen. The factor 1.67 accounts for the higher respiration rates per N in leaves, while the factor $M_c = 12$ gC/mol(CO_2) is used because photosynthesis is expressed in molar units of carbon.

In a subtropical dry forest in Puerto Rico (Lugo and Murphy, 1986), leaf nitrogen was 189 kg/ha of a total of 916 kg/ha, i.e. 21%, whereas in an equatorial moist forest in Zaïre (a 28-year-old secondary forest) 143 kg/ha were found in leaves for a 593 kg/ha total (Greenland and Kowal, 1960), this is a portion of 24%. (With Eq. (63), this would mean $f_{R,leaf} \approx 0.4$). The similarity

contrasts with the fact that with 25%, the root fraction of total nitrogen was significantly lower in the moist than in the dry forest, where it amounted to as much as 60%. It appears that the value of $f_{N,leaf}$ is a conservative quantity compared to the distribution of biomass between leaves, stem and roots. A possible explanation is that trees, because of competition with other trees, accumulate woody biomass until a certain critical value of $f_{N,leaf}$ around 20 to 25% is reached, from where on a reduction in NPP prevents further reduction of the relative leaf fraction. This might also be true for grasses, which increase root biomass under arid conditions until a similar value of $f_{N,leaf}$ is reached. This would explain why respiration costs relative to GPP are remarkably similar for grasses and trees (Ryan, 1991a). Therefore, in this study a uniform value of $f_{N,leaf} = 0.14$ will be used, or $f_{R,leaf} = 0.40$.

A mean value for growth respiration according to Ryan (1991a), which will be used here, is 0.25 gC per gC biomass produced, or $f_{R,G} = 0.25$. Using

$$\mathsf{NPP} = \mathsf{GPP} - R_M - R_G \tag{64}$$

We can thus derive an implicit equation for

$$R_G = f_{R,G} \cdot \mathsf{NPP} = f_{R,G}(\mathsf{GPP} - R_M - R_G), \tag{65}$$

which yields the following explicit form:

$$R_G = \frac{f_{R,G}}{1 + f_{R,G}}(\mathsf{GPP} - R_M) \tag{66}$$

The equation for net primary productivity thus becomes

$$\mathsf{NPP} = \frac{1}{1 + f_{R,G}}(\mathsf{GPP} - R_M). \tag{67}$$

We finally define the net ecosystem exchange as

$$\mathsf{NEE} = R_H + R_A - \mathsf{GPP}. \tag{68}$$

B.2.3 Light absorption

Light absorption in the photosynthetically active spectrum is calculated within the two-flux approximation, following Sellers (1985), expressed by the following equations (with the cumulative leaf area index, l, as the vertical coordinate, where $l = 0$ at the top, and $l = \Lambda$ at the bottom of the canopy):

$$\begin{aligned} \bar{\mu}\frac{dR\downarrow}{dl} + [1 - (1-\beta)\omega]R\downarrow - \omega\beta R\uparrow &= \omega\bar{\mu}K(1-\beta_0)R(l) \\ -\bar{\mu}\frac{dR\uparrow}{dl} + [1 - (1-\beta)\omega]R\uparrow - \omega\beta R\downarrow &= \omega\bar{\mu}K\beta_0 R(l) \end{aligned} \tag{69}$$

$R\downarrow$ and $R\uparrow$ are the diffuse fluxes downward and upward, respectively, and $R(l)$ is the direct flux with $R(0) = d_{PAR}R_{PAR}$ (Eq. (144) and Eq. (155)):

$$R(l) = R(0)e^{-Kl} \tag{70}$$

Further, ω is the leaf single scattering albedo, β the forward scatter fraction for the diffuse flux, β_0 the same for the direct flux, K the extinction coefficient for the direct flux and $\bar{\mu}$ the mean of K^{-1} over the downward hemisphere ($\int_0^1 K^{-1}(\mu)d\mu$).

The following simplifications are used: (1) There is no preferred leaf orientation, i.e. distribution of leaf angles is isotropic; (2) leaf reflectivity and transmissivity are equal (the sum of the two is ω). These assumptions yield:

$$\beta = \frac{1}{2} \tag{71}$$

$$\beta_0 = \frac{1}{2} \tag{72}$$

$$K = \frac{1}{2\mu} \tag{73}$$

$$\bar{\mu} = 1 \tag{74}$$

where μ is the cosine of the solar zenith angle (cf. Eq. (142)).

The following boundary conditions are also needed for the solution of Eq. (69): (1) $R\downarrow(0)$ equals diffuse incoming radiation; (2) the reflection at the lower boundary is given by the soil reflectivity in the PAR region, ρ_s^{PAR}:

$$R\downarrow(0) = (1-d_{PAR})R_{PAR} \tag{75}$$

$$R\uparrow(\Lambda) = \rho_s^{PAR}\left(R(\Lambda)+R\downarrow(\Lambda)\right) \tag{76}$$

The standard value for the single scattering albedo for PAR is set to $\omega = 0.12$, while ρ_s^{PAR} is computed from the background albedo in the total solar spectrum, ρ_{bg}:

$$\rho_s^{PAR} = 0.92\rho_{bg} - 0.015 \tag{77}$$

This dependence has been derived from the "soil line" by Price and Bausch (1995), a linear relationship between PAR and NIR (near infrared) reflectivity for moist soils, with the assumption $\rho_{bg} = (\rho_s^{PAR} + \rho_s^{NIR})/2$ (cf. Section 4.5).

The rate of PAR absorption per leaf area of layer k, I_k^{PAR}, is calculated, to be used for the light limited photosynthetic rate (Eq. (29)), as the sum over N_l layers of equal distance in l-coordinates, going from $l = l_{k-1}$ to $l = l_k$ with $l_0 = 0$ and $l_{N_l} = \Lambda$:

$$\begin{aligned} I_k^{PAR} &= \frac{1}{l_k - l_{k-1}} \int_{l_{k-1}}^{l_k} \left\{ -\frac{d}{dl}(R + R\downarrow + R\uparrow) \right\} dl \\ &= \{[R(l_{k-1}) + R\downarrow(l_{k-1}) - R\uparrow(l_{k-1})] \ldots \\ &\quad \cdots - [R(l_k) + R\downarrow(l_k) - R\uparrow(l_k)]\}/(l_k - l_{k-1})54 \end{aligned} \tag{78}$$

where R, $R\downarrow$ and $R\uparrow$ are the solutions to the Eq. (69), Eq. (75) and Eq. (76) with LAI Λ. The canopy photosynthesis is then

$$A_c = f_c \sum_{k=0}^{N_l} A(I_k^{PAR}) \quad (l_k - l_{k-1}) \tag{79}$$

and FAPAR can be calculated from

$$\begin{aligned} f_{PAR} &= \{[R(0) + R\downarrow(0) - R\uparrow(0)] - [R(\Lambda) + R\downarrow(\Lambda) - R\uparrow(\Lambda)]\}/(R(0) + R\downarrow(0)) \\ &= 1 - \frac{R\uparrow(0) + (1 - \rho_s^{PAR})(R(\Lambda) + R\downarrow(\Lambda))}{R(0) + R\downarrow(0)} \end{aligned} \tag{80}$$

Eq. (79) is the approximation used for the integrals of Eq. (42) and Eq. (53), and fC is the fractional vegetation cover. Following the arguments by Sellers (1985), no separate layer for sunlit leaves is introduced.

B.3 Energy and Water Cycle

B.3.1 Water balance overview

The model considers a total of three water pools: total soil water, also described as root-zone soil moisture (W_r), a skin or intercepted reservoir on leaves and other plant parts (W_i) and snow (W_{sn}). The total amount of water at the surface, W_{tot}, from Eq. (84), is thus partitioned into three components (cf. Eq. (85), Eq. (86) and Eq. (132)):

$$W_{tot} = W_r + W_i + W_{sn} \tag{81}$$

The largest part by far is W_r, the root-zone soil moisture. There is also an additional thin soil layer at the surface, W_s, which, however, overlaps with the root-zone soil moisture layer. The precipitation rate, P_{tot}, and the rate of evapotranspiration, E_{tot}, are also divided into further quantities:

$$P_{tot} = P_r + P_{sn} = P_s + P_i + P_{sn} \tag{82}$$

and

$$E_{tot} = E_s + E_{sn} + E_i + E_t \tag{83}$$

Here, P_r is rainfall and P_{sn} the snowfall rate, P_s the amount of rain falling directly on the soil, and P_i the rain intercepted by vegetation. The partitioning of total precipitation, P_{tot}, into snow and rain is calculated from Eq. (133) (see Section B.3.7) and the interception rate, P_i, according to Eq. (87) (see Section B.3.3). There is also an indirect "rain" rate, P_v, of water that overflows from the leaf surfaces and drips down to the floor (Eq. (88)). The components of total evapotranspiration are E_s (soil evaporation, Eq. (116)), E_{sn} (snow evaporation, Eq. (136)), E_i (evaporation from the intercepted skin reservoir, Eq. (90)), and E_t (transpiration, Eq. (92)).

The water balance can be described by the following equation:

$$P_{tot} - E_{tot} - Q_d - Q_b = \dot{W}_{tot} \tag{84}$$

P_{tot} is the total precipitation rate (rain and snow), E_{tot} the total evapotranspiration from soil, leaf surfaces, through leaf pores and from snow (including sublimation), W_{tot} the total amount of water stored in the soil, on the vegetation and on the ground as snow, Q_d is direct runoff, and Q_b base flow.

B.3.2 Root-zone soil water balance

The root-zone soil water balance can be expressed as

$$W_r(t) - W_r(t - \Delta t) = (P_s(t) + P_v(t) + S_m(t) - E_t(t) - E_s(t) - Q_b(t))\Delta t. \quad (85)$$

The flux terms on the right hand side of the equation are therefore daily averages in kg m^{-2}s^{-1}. P_s and P_v are, respectively, the precipitation rates arriving at the soil directly, or being intercepted by vegetation first and then dripping through to the ground. P_s is calculated from $P_s = P_{tot} - P_{sn} - P_i$ with P_{sn} from Eq. (133) and P_i from Eq. (87); P_v is given by Eq. (88) (Section B.3.3), S_m, the rate of snow melt, by Eq. (134) (Section B.3.7), E_t by Eq. (92) (transpiration, Section B.3.4) and E_s by Eq. (116) (soil evaporation, Section B.3.5). $Q_b(t)$ is the base flow, which is proportional to the root-zone soil moisture and capped to zero when the root-zone soil moisture is less than the field capacity.

B.3.3 Skin water balance

The balance equation of the skin or intercepted reservoir, W_i, is

$$W_i(t) - W_i(t - \Delta t) = (P_i(t) - E_i(t) - P_v(t))\Delta t \quad (86)$$

If rain falls on dense vegetation (approx. LAI > 3), most of it falls on leaves and branches. A considerable fraction initially remains as a thin film on the vegetation, while another fraction, depending on the size of raindrops and the interception capacity of the canopy, reaches the ground. While rainfall continues, the skin reservoir, W_i, reaches a maximum and the water begins to drip through to the ground. A good approximation for the LAI dependence of interception is the vertical projection of the leaf area divided by the ground area. When this value approaches 1, it is assumed that 100% of rainfall is intercepted by the vegetation canopy:

$$P_i = f_c \left(1 - e^{-0.5\Lambda/f_c}\right) P_r \quad (87)$$

f_c is the fractional cover of vegetation and Λ the LAI of the total area (vegetated and non-vegetated). Finally, the throughfall of rain through the canopy, $P_v\Delta t$, is calculated according to:

$$P_v\Delta t = \max\left\{W_i(t - \Delta t) + P_i\Delta t - W_{i,max}; \quad 0\right\} \quad (88)$$

Thereby it is assumed that the skin reservoir fills up to a capacity of $W_{i,max}$ and that no evaporation happens during rain. The interception capacity of the vegetation, $W_{i,max}$, is assumed proportional to LAI as in the BATS model (Dickinson et al., 1993)

$$W_{i,max} = w_{i,max}\Lambda \quad (89)$$

with an area-specific capacity, $w_{i,max}$, of 0.1 kg m^{-2}. Finally, the evaporation rate of the canopy-intercepted water happens at the potential rate, $E_{v,max}$ (Eq. (91)), but is limited by the amount of available canopy water W_i and the input of water as intercepted rainfall, P_i, minus throughfall, P_v:

$$E_i(t) = \min\{E_{v,max}; W_i(t - \Delta t) + (P_i(t) - P_v(t))\Delta t\}. \quad (90)$$

B.3.4 Transpiration

Total evapotranspiration from vegetation ($E_v = E_t + E_i$) is primarily driven by the net radiative balance of the vegetation ($R_{n,v}$, Eq. (159)) and is limited by the available amount of soil (W_s, Eq. (85)) and skin water (W_i, Eq. (86)). If the vegetation surfaces are wet ($W_i > 0$), the canopy conductance is infinite ($G_c \to \infty$) so that evaporation follows its maximum rate, $E_{v,max}$, with the evaporated water coming from the skin reservoir:

$$E_v = E_i = E_{v,max} = \frac{sR_{n,v} + \rho_a c_p (e_s(T) - e_a) G_a}{s + \gamma} \qquad (W_i > 0) \tag{91}$$

Here, E_i is allowed to assume negative values during dew formation. When the vegetation is dry ($W_i = 0$), evapotranspiration is determined by G_c, which is the combined conductance of all stomata within the plant canopy. The transpiration rate is then calculated from the Penman-Monteith formula (Monteith, 1965):

$$E_t = \frac{1}{\lambda} \frac{sR_{n,v} + \rho_a c_p (e_s(T) - e_a) G_a}{s + \gamma(1 + G_a/G_c)} \qquad (W_i = 0) \tag{92}$$

An additional condition is given by $E_t \geq 0$, i.e. transpiration is only allowed from the vegetation to the atmosphere.

Here, γ is the psychrometric constant (≈ 65 Pa K^{-1}), defined as

$$\gamma = \frac{p c_p}{0.622 \lambda}. \tag{93}$$

e_a is vapour pressure of air (see Eq. (170)), e_s saturated vapour pressure, and s the strongly temperature dependent slope of the vapour pressure curve, $\partial e_s(T)/\partial T$ (see Eq. (172), for rho_a see Eq. (51)).

According to Eq. (92), transpiration can only happen when the vegetation is dry. To account for this fact, the sum of transpiration and canopy evaporation are limited to the potential canopy evaporation rate by subtracting the surplus from the rate of transpiration, i.e. the following condition is enforced:

$$E_t \leq E_{v,max} - E_i \tag{94}$$

The canopy conductance, G_c, is computed from Equs. Eq. (45) and Eq. (49) as described in Section B.2.1. The aerodynamic conductance, G_a, between the canopy and a reference height of 10 m is estimated from vegetation height, h_v (from Table 11):

$$G_a = \frac{\kappa^2 u}{\left[\ln(\frac{h_{ref}}{r_{z0} h_v} + a_{z0})\right]^2} \tag{95}$$

with h_v in metres, a roughness length of $r_{z0} h_v$ and a wind speed, u at a height of $h_{ref} = 10$ m above the canopy. u is set to a uniform value of 3 m/s, and κ is the von Karman constant (0.41). The parameters are $r_{z0} = 0.0999995$ and $a_{z0} = 4.3568$. For temperate broadleaf and coniferous forests ($h_v = 15$ m), this amounts to a value for G_a of 0.198 m/s, while for short grass ($h_v = 0.3$ m,

Table 13: Results of a curve fit of $g_s = g_0/(1 + b_e\Delta e)$ to measurements by Turner et al. (1984), after minimising the root mean squared (r.m.s.) difference between curve and measurements. g_0 is in mmol m^{-2} s^{-1}, b_e in kPa^{-1}.

Species	g_0	b_e	r.m.s.	n
Helianthus annus	707	0.35	9.5	5
Vigna unguiculata	2161	3.23	13.7	4
Pistacia vera	606	0.85	11.7	5
Nerium oleander	344	0.39	12.1	5
Prunus dulcis	209	0.39	6.4	5

Table 11) it is 0.024 m/s, in good agreement with the average observed values in Kelliher et al. (1993).

Further, , and λ the latent heat of evaporation (2.45 MJ kg^{-1}at 20^{o}C). Since snow sublimation is also modelled, e_s, s and λ are calculated differently for temperatures above or below 0^{o}C, e_s and s according to Eq. (172), and λ (in J kg^{-1}, with T in ^{o}C) from

$$\lambda = \begin{cases} 2.501 \times 10^6 - 2.38 \times 10^3 T & \text{for} \quad T > 0^o\text{C} \\ 2.834 \times 10^6 & \text{for} \quad T < 0^o\text{C} \end{cases} \tag{96}$$

with a slightly temperature dependent evaporation heat (Jones, 1983) and a fixed sublimation heat (Anderson, 1976).

With Eq. (45), the potential transpiration rate is now defined:

$$\lambda E_t^{pot} = \frac{sR_{n,v} + \rho_a c_p \Delta e G_a}{s + \gamma(1 + G_a/G_{c,0})} \tag{97}$$

This value is called "demand" by Federer (1979) and is equal to the transpiration rate without water limitation (see above, Eq. (97)).

Calculation of the actual canopy conductance, G_c, and the actual transpiration rate, E_t, follows a combination of the approaches by Jarvis (1976) and Federer (1982). There is only one empirical multiplicative factor depending on vapour pressure deficit, $\Delta e = e_s(T) - e_a$, and with a functional form proposed by Lindroth and Halldin (1986) (see also Section B.2.1):

$$G_c = G_{c,0} \frac{1}{1 + b_e \Delta e} \tag{98}$$

A test of this form with measurements by Turner et al. (1984) is shown in Table 13. A curve described by $g_s = g_0/(1 + b_e\Delta e)$ with two free parameters is fitted by minimising the root mean squared (r.m.s.) deviation from the measurement. The success of the test is documented by the fact that the deviation is always much smaller than g_0.

Studies by Turner et al. (1984) and Schulze et al. (1987) suggest that a signal transmitted by hormones from the roots is responsible for the closing of stomata under water limitation. To account for this effect, another multiplicative factor could be added to Eq. (98), as in the Jarvis approach. This would mean that stomata close even under sufficient water supply as a reaction to rising vapour

pressure deficit. However, stomata do not react directly to vapour pressure deficit of the surrounding air, but to a rise in the evaporative demand (Schulze et al., 1987); as long as the supply of soil water is sufficient, stomata tend to remain open even under high atmospheric drought.

A different approach is therefore chosen here, following Federer's model: The empirical constant b_e is adjusted at each daily integration by computing an optimal value for b_e, $b_{e,opt}$, at the time of the highest transpirational demand, E_t^{pot}, assumed at 13:00 hours solar time. $b_{e,opt}$ is set at the time step closest to 13:00 solar time such that the transpiration rate from Eq. (92) with $b_e = b_{e,opt}$ is less or equal to a supply rate, S. Further, at each time step, b_e is "relaxed" towards $b_{e,opt}$ via a time constant τ_b:

$$b_e(t + \Delta t) = b_e(t)e^{-\Delta t/\tau_b} + b_{e,opt}(1 - e^{-\Delta t/\tau_b}), \tag{99}$$

with τ_b =12 hours. Without detailed knowledge of soil hydrology and root distribution, a reasonable approximation according to Federer (1982) is

$$S = c_w f_{soil}^{eff} \tag{100}$$

with values for c_w in the range of 0.5 to 2.0 mm/hour and an effective soil water fraction, f_{soil}^{eff} (Eq. (101)). In particular, $b_e \to 0$ if demand, E_t^{pot}, is less than S.

Thus, the only remaining free parameter of the stomatal model is the rate c_w. This is justified by the fact that this parameter represents the root system that is not described explicitly. All other elements of the model are based on general principles of stomatal control or on empirical findings. An empirical approach is chosen in Eq. (98) instead of a mechanistic description of stomatal response to atmospheric demand (Schulze et al., 1987; Friend, 1995), because designing a global model of detailed soil hydraulic processes appears to be unfeasible. Sensitivity tests by Knorr (1997) have shown that global carbon and water fluxes are rather insensitive to the choice of c_w, which are mostly contrained by the available soil water and atmospheric CO_2 content.

The last point described in this subsection is the dependence of stomatal conductance on soil and air temperature as assumed in the model: On the one hand, the temperature dependence of G_c is given by the temperature dependence of $A_{c,0}$ via Eq. (45). Also, $A_{c,0} = 0$ if the daily average of air temperature, $\bar{T}$, is zero degrees or less, and thus $G_c = 0$ and $E_t = 0$. On the other hand, if the soil is partly frozen while the air is already warmer than 0^{o}C, there is a dependence on soil temperature through the effective plant available soil water fraction, f_{soil}^{eff} (cf. f_{soil} Eq. (126))

$$f_{soil}^{eff} = \begin{cases} \max\left\{f_{soil}\left(1 - \frac{d_{fr}}{d_r}\right); 0\right\} & \text{if} \quad \bar{T} > 0 \quad \text{and} \quad \bar{T} > \bar{T}_{ds} \\ 0 & \text{if} \quad \bar{T} < 0 \\ f_{soil} & \text{else.} \end{cases} \tag{101}$$

d_r is the rooting depth (Table 11), $\bar{T}_{ds}$ the daily mean soil temperature (subscript "ds" for deep soil) at depth d_{ds}, and d_{fr} the frost penetration depth.

To determine the frost depth, d_{fr}, a linear temperature course is assumed up to the surface. With this assumption, d_{fr} can be computed for the first case in Eq. (101):

$$d_{fr} = \frac{d_{ds}\bar{T}}{\bar{T} - \bar{T}_{ds}} \tag{102}$$

$\bar{T}_{ds}$ is one of the driving variables, see Section B.5. The depth for which soil temperatures are available varies by driving data set, but should be similar to rooting depth. In the current model implementation we assume d_{ds}=1.50 m.

The following effect is neglected in the model described here: When calculating net radiation of vegetation, $R_{n,v}$, or soil, $R_{n,s}$, longwave upward radiation after Eq. (162) is calculated from air instead of skin temperature. For example, the skin temperature of vegetation, T_v, depends itself on the sensible heat flux $H_v = R_{n,v} - \lambda E_v$. To a linear approximation, this can be accounted for by the concept of isothermal conductivity, G_i (e.g. Jones, 1983), that has to be added to the surface conductivity. Its value can be calculated from $G_i = 4\sigma T^4 t_{l,v}/(pc_p M_a)$, which is 3.5 mm/s at 0^{o}C and 6.0 mm/s at 40^{o}C, $p = p_0$ and $t_{l,v} = 1$. This is an order of magnitude smaller than G_a for a typical grassland and even two orders of magnitude for coniferous forests (Kelliher et al., 1993), so that G_i can generally be neglect in large-scale studies.

To close the energy balance given by Eq. (130) and Eq. (131), the sensible heat flux, H, can be calculated as follows:

$$\begin{aligned} H &= H_v + H_s \\ &= R_n - G - \lambda E_{tot} \\ &= R_{n,v} + R_{n,s} - G - \lambda(E_t + E_i + E_s + E_{sn}) \end{aligned} \tag{103}$$

B.3.5 Soil water module

The following description of the soil water balance is taken to a large extent from Scholze et al. (2016). Soil evaporative demand is calculated using the net radiation of the soil (Eq. (131)). Actual soil evaporation and the balance of soil water inputs and outputs follow the 1-layer version of the Variable Infiltration Capacity (VIC-1L) model (Wood et al., 1992).

The assumptions made by VIC-1L are as follows:

- The total soil water column that can be stored at a given place within a grid cell until the soil is saturated, denoted i, has some statistical distribution $f(i)$ between the values 0 and i_m.
- The present soil water content at some location within the grid cell is either at saturation, or assumes a value i_0 that is uniform for all locations accept those that are saturated, i.e. where i_m is reached.
- Rain falls uniformly onto the grid cell.

We denote by A the fraction of the grid cell that is saturated before rainfall begins. The dependence of A on i_0 can be computed from the statistical distribution $f(i)$ via:

$$A(i_0) = \int_0^{i_0} f(i)\, di. \tag{104}$$

Based on these assumptions, for an infinitesimal rain input dP, the state variable i_0 increases by exactly the amount dP, unless $i_0 + dP > i_m$, in which case i_0 increases to its maximum i_m. The

increase in total soil water content after the infinitesimal rainfall is

$$dW = [1 - A(i_0)]dP, \tag{105}$$

because only the unsaturated fraction of the grid cell stores additional water. The saturated fraction A will create surface runoff of the amount $dP \times A$. Presuming that this runoff is completely absorbed by the unsaturated fraction of the same grid cell, increases A by the amount

$$dA = \frac{dA}{di_0} dP = f(i_0)\, dP. \tag{106}$$

Integration of Eq. (105) yields an expression of the more commonly used state variable total soil water content, W_0, which can be used to substitute i_0:

$$W_0(i_0) = i_0 - \int_0^{i_0} A(i)\, di. \tag{107}$$

In this integration, the soil is filled from fully dry by a series of infinitesimal amounts di until the amount stored locally in the soil reaches either i_0 or its local saturation level. The soil is filled up in the same way as when adding rain, so that dP in Eq. (106) can be substituted by di , and the modified equation be integrated from 0 to i_0. Eq. (107) is valid for $i_0 < i_m$, and i_0 the total amount of water added. The integral over $A(i)$ on the r.h.s. of Eq. (107) is the amount of water lossed when local parts of the soil become saturated.

Similar to the use of $W_0(i_0)$ instead of i_0, i_m does not have to be defined explicitly as a model parameter, but can be derived using the definition $W_c = W_0(i_m)$, with W_c, the moisture storage capacity of the soil at the grid-cell scale, used as the primary model parameter.

Using Eq. (107), we derive direct runoff Q_d for a finite amount of rainfall, P_s, as long as the sum of i_0 and this value is not more than i_m, considering that i_0 increases to $i_0 + P_s$:

$$Q_d = P_s - W_0(i_0 + P_s) + W_0(i_0) = \int_{i_0}^{i_0+P_s} A(i)\, di. \tag{108}$$

Otherwise, the maximum infiltration of the grid cell determines direct runoff:

$$Q_d = P_s - W_0(i_m) + W_0(i_0) = P_s - W_c + W_0. \tag{109}$$

VIC-1L uses the following functional form of the cumulative sub-gridscale distribution of infiltration capacity:

$$\mathcal{A}(i) = 1 - (1 - i/i_m)^B \tag{110}$$

With this, we obtain

$$Q_d = P_s - W_c + W_0 + W_c \max\left\{1 - \frac{i_0 + P_s}{i_m}; 0\right\}^{1+B} \tag{111}$$

with

$$i_m = (1 + B)W_c \tag{112}$$

and

$$i_0 = i_m \left[1 - \left(1 - \frac{W_0}{W_c} \right)^{\frac{1}{1+B}} \right]. \tag{113}$$

This yields the following expression for infiltration ($I = P_s - Q_d$):

$$I = W_c \left(1 - \max \left\{ 1 - \frac{i_0 + P_s}{i_m}; 0 \right\}^{1+B} \right) - W_0. \tag{114}$$

The VIC-1L model is applied to W_r, i.e. the entirety of the plant-available soil moisture, with the following exceptions:

- Soil evaporation is computed similar to VIC-3L (Liang et al., 1996), where the saturated fraction of the grid cell, $\mathcal{A}(i_0)$, evaporates at the potential rate, and the remaining fraction as a function of the surface layer soil moisture. Different to VIC, the surface layer soil moisture of the unsaturated fraction is represented by its average volumetric soil moisture.
- Base flow is computed such that the soil can only drain to W_f instead of 0.
- There is a further thin surface layer containing W_s of water, and with a maximum water storage capacity of $W_{s,c}$.

The base flow equation used by D&B can therefore be written as

$$Q_b = k_b \max\{W_r - W_{r,f}; 0\}, \tag{115}$$

where W_r is the root-zone soil moisture, and $W_{r,f}$ its value at field capacity. $W_{s,c}$ is computed based on the depth of the surface soil layer, d_s, and volumetric soil moisture at saturation (Table 14) via $W_{s,c} = d_s \Theta_s$. d_s is set to 4 cm, but limited (in rare cases) to the depth of the soil to bedrock.

It is further assumed that the infiltration capacity of the surface layer varies spatially following Eq. (110), but with B replaced by B_s, i.e. has a different statistical distribution from the root-zone. However, an important, simplifying assumption is that this variation is statistically independent of the spatial variation of infiltration capacity of the root-zone layer. It is thus assumed that if $\mathcal{A}(i_0)$ is fractional grid cell area where the soil is saturated, the same fraction $\mathcal{A}(i_0)$ of the surface layer is also saturated, but the remaining fraction always has the same statistical distribution, and its average saturated per-area water content is always $W_{s,c}$, independent of $\mathcal{A}(i_0)$.

Soil evaporation, E_s, is computed such the actual rate is equal to the potential rate times the ratio between actual and saturated surface-layer soil moisture. Different to VIC-3L, this ratio is based not on the point-wise soil water content, but on the average per-area soil water content of the unsaturated grid cell fraction, i.e. $W_{s,u} = \frac{W_s - \mathcal{A}(i_0) W_{s,c}}{1 - \mathcal{A}(i_0)}$. This assumption leads to the following simple expression:

$$E_s / E_s^{pot} = \mathcal{A}(i_0) + [1 - \mathcal{A}(i_0)] \frac{W_{s,u}}{W_{s,c}} = \frac{W_s}{W_{s,c}}. \tag{116}$$

where the potential soil evaporation rate is taken as

$$\lambda E_s^{pot} = \frac{sR_{n,s} + \rho_a c_p (e_s(T) - e_a) G_{a,s}}{s + \gamma} \tag{117}$$

with s from Eq. (172) with $s(T) = \partial e_s / \partial T$, γ from Eq. (93), and the aerodynamic conductance for the soil surface, $G_{a,s}$, computed in an analogous way to Eq. (95), but with a soil roughness length of $\xi_s = r_{z0} \times 0.1$m:

$$G_{a,s} = \frac{\kappa^2 u}{\left[\ln(\frac{h_{ref}}{\xi_s} + a_{z0})\right]^2} \tag{118}$$

Infiltration, I, is first computed for the root-zone layer, which overlaps with the surface layer and is therefore the infiltration for the entire simulated soil water pool. The same concept of variable infiltration capacity is used to determine how much of I remains in the surface layer, building on the equivalent expression for the root-zone layer (Eq. (114)), but where P_s is replaced by I, B by the specific parameter for the surface layer, B_s, and the resultant values yields the part of infiltration that remains in the surface layer.

Surface soil moisture is updated to the next time step, following the same procedure as for the root-zone layer, which is executed first to compute infiltration. After that, W_s is updated to an intermediate value W_s^+ by subtracting soil evaporation (Eq. (116)) and a term for drainage to below the surface layer, formulated in a similar way to the base flow:

$$W_s^+ = W_s - E_s - Q_s \tag{119}$$

with

$$Q_s = k_s \max\{W_s^+ - W_{s,f}; 0\}. \tag{120}$$

After that, infiltration into the surface layer is calculated based on W_s^+, which is then added to W_s^+ to obtain surface soil moisture at the beginning of the following time step:

$$W_s(t + \Delta t) = W_s^+ + I_s = W_{s,c} \left(1 - \max\left\{1 - \frac{i_{s,0}^+ + I}{(1 + B_s) W_{s,c}}; 0\right\}^{1+B_s}\right), \tag{121}$$

with

$$i_{s,0}^+ = (1 + B_s) W_{s,c} [1 - (1 - W_s^+ / W_{s,c})^{1/(1+B_s)}]. \tag{122}$$

This hydrology scheme has two state variables – W_s and W_r – and the following parameters: volumetric soil moisture at saturation, Θ_s and field capacity Θ_f differentiated by soil texture class (Table 14, the same texture class is assumed for both layers), the base flow parameter k_b, the surface drainage parameter k_s, and the shape parameters for the root-zone, B, and for the surface layer, B_s.

Using the depth of the corresponding layer, d_i (d_s for surface, d_r for root-zone layer), $W_{c,i}$ and $W_{f,i}$ of layer i is determined from:

$$W_{c,i} = \Theta_s d_i, \tag{123}$$

Table 14: Assumed volumetric soil moisture at saturation (Θ_s), field capacity (Θ_f) and wilting point (Θ_w) differentiated by texture class.

	Coarse	Medium/Coarse	Medium	Fine/Medium	Fine	Organic
Θ_s	0.410000	0.435000	0.451000	0.420000	0.476000	0.451000
Θ_f	0.193706	0.245704	0.298119	0.303402	0.377204	0.298119
Θ_w	0.071982	0.110032	0.149533	0.170485	0.244554	0.149533

for the moisture storage capacity, and

$$W_{f,i} = \Theta_f d_i. \tag{124}$$

for the field capacity.

k_s is simulated from spatially constant base flow drainage rates for 150 mm and 320 mm of $kb_{150} = 0.095$/day and $kb_{320} = 0.032$/day taken from Wood et al. (1992):

$$k_s = (kb_{150} - (150. - W_{c,s}) * (k_{b,320} - k_{b,150})/(320. - 150.)) * 2.1 \tag{125}$$

We use $k_b = 0.2$/day, $B_s = 10$, and $B_r = 0.25$. Gao et al. (2009) indicate a reasonable range of B of (0,4]. A higher value of B gives lower infiltration and yields higher surface runoff. The parameter summarises the effect of the spatial variability of rainfall over the grid cell, the soil texture, and the slope of the terrain.

A time step for the entire scheme is executed following the procedure described in Wood et al. (1992), first for the root-zone, then for the surface layer:

1. Base flow from the entire soil moisture store is computed according to Eq. (115).
2. Soil evaporation, E_s, is computed from Eq. (116), based on W_s of the previous time step.
3. From W_r, base flow, soil evaporation and transpiration are subtracted.
4. Direct runoff, Q_d, is computed from Eq. (111) using the intermediate value of W_r from the previous sub-step replacing W_0. The value of P_{tot} (Eq. (84)) is computed as the sum of non-intercepted precipitation, throughfall from the canopy water pool, and snow melt.
5. The value of W_r from the intermediate step is updated by adding P_s and subtracting Q_d to obtain the value of the surface soil moisture at the end of the time step.
6. The value of W_s from the previous step is updated to an the intermediate value W_s^+ (Eq. (119)) by subtracting soil evaporation, E_s, and drainage from the surface layer (Q_s, Eq. (120)).
7. The part of infiltration remaining in the surface layer is is added to W_s^+ to obtain the surface soil moisture at the next time step (Eq. (121)).

The information used by the model's photosynthesis module is the ratio between the total plant-available soil moisture and its value at field capacity, $W_{r,f}$, where plant-available soil moisture equals

actual root-zone soil moisture, W_r minus root-zone soil moisture at wilting point, $W_{r,w}$. This value is cut off at both 0 and 1 according to:

$$f_{soil} = \begin{cases} 0 & \text{if} \quad W_r \leq W_{r,w} \\ \frac{W_r - W_{r,w}}{W_{r,f} - W_{r,w}} & \text{if} \quad W_{r,w} < W_r < W_{r,f} \\ 1 & \text{if} \quad W_r \geq W_{r,f} \end{cases} \tag{126}$$

The wilting point soil moisture content is computed using Θ_w from Table 14:

$$W_{w,r} = \Theta_w d_r. \tag{127}$$

In order to simulate the effect of irrigation on cropland soil water balance, we applied a lower threshold for W_r according to of 0.3 to

$$W_r \geq 0.3(W_{r,f} - W_{r,w}) + W_{r,w}. \tag{128}$$

Similarly, for surface layer soil moisture, W_s we applied a threshold of 30% of the field capacity:

$$W_s \geq 0.3 W_{s,f}. \tag{129}$$

B.3.6 Energy balance

Through the energy balance (Eq. (130) and Eq. (131)), surface temperature and thereby evaporation and transpiration rates are highly dependent upon net radiation, R_n. Consequently, it is in the interest of the plant to absorb sunlight as effectively as possible, while at the same time keeping the absorbed radiative energy, R_n, as small as possible. In fact, plants absorb photosynthetically active light to almost 90% while they reflect or scatter around 90% of the light in the near infrared, which is of no value for photosynthesis. Besides this, through "clumping" and keeping leaves in an erect position, vegetation can reduce light absorption during midday, while increasing it in the morning and the evening when atmospheric demand for transpiration is lowest. To simulate such effects, net radiation and with it the entire energy balance is split into a vegetation, $R_{n,v}$, and a soil part, $R_{n,s}$. The energy balance for both parts can be written as:

$$R_{n,v} = H_v + \lambda(E_t + E_i) \tag{130}$$

and

$$R_{n,s} - G = H_s + \lambda(E_s + E_{sn}) \tag{131}$$

where H_v and H_s are the sensible heat flux for the soil and the canopy, respectively, and λ the latent heat of evaporation (see Eq. (96)). There are two evaporation fluxes controlled by $R_{n,v}$ (transpiration and skin reservoir evaporation) and two controlled by $R_{n,s}$ (soil and snow evaporation).

B.3.7 Snow module

The effect of snowfall on the energy and water balance is twofold: On the one hand, snow increases the maximum amount of water that can be stored at the land surfaces; on the other hand, snow, and

fresh snow in particular, has a very low albedo, decreasing net radiation at the surface (Eq. (158)) and thereby evaporation (Eq. (136)). Since snow height enters the calculation of the snow albedo, a simple formulation is included for this variable.

The snow balance is described by the following equation:

$$W_{sn}(t) - W_{sn}(t - \Delta t) = (P_{sn}(t) - S_m(t) - E_{sn}(t))\Delta t \tag{132}$$

The snowfall rate, P_{sn}, depends on the daily average temperature, $\bar{T}$, and the precipitation rate, P_{tot} (Wigmosta et al., 1994):

$$P_{sn} = \begin{cases} P_{tot} & \text{for} \quad \bar{T} \leq -1.1^o\text{C} \\ \frac{3.3-\bar{T}}{4.4} P_{tot} & \text{for} \quad -1.1^o\text{C} < \bar{T} < 3.3^o\text{C} \\ 0 & \text{for} \quad 3.3^o\text{C} \leq \bar{T} \end{cases} \tag{133}$$

For snow melt, a simple function of temperature is chosen (in kg m^{-2}day^{-1} Hagemann and Dümenil, 1996):

$$S_m = 3.22 \max\{\bar{T}; 0\} \text{ kg m}^{-2}\text{day}^{-1} \tag{134}$$

with $\bar{T}$ in ^{o}C. Since $W_{sn}(t)$ in Eq. (132) is not allowed to assume negative values, the maximum for S_m is given by

$$S_m(t)\Delta t \leq W_{sn}(t - \Delta t) + (P_{sn}(t) - E_{sn}(t))\Delta t \tag{135}$$

The potential snow sublimation rate, E_{sn}^{pot}, is derived from the energy input to the ground, $R_{n,s} - G$ (Eq. (160) and Eq. (161), cf. Eq. (131)), in the same way as potential soil evaporation, E_s^{pot} (see Eq. (117)). The difference is only that for temperatures below freezing, λ is the latent heat of sublimation, not evaporation (see Eq. (96)). E_{sn}^{pot} is thus determined primarily by net radiation, $R_{n,s}$, and depends to a large extent on snow albedo. The daily snow evaporation cannot exceed the amount of snow on the ground plus the snowfall:

$$E_{sn}(t)\Delta t = \min\{E_{sn}^{pot}(t)\Delta t; \quad W_{sn}(t - \Delta t) + P_{sn}(t)\Delta t\}. \tag{136}$$

We assume snow sublimation at the equilibrium rate, which results from the Penman-Monteith equation (Eq. (92)) in the limit $G_a \to 0$.

When calculating the snow balance, the snowfall rate is first determined (P_{sn}, Eq. (133)), then the evaporation rate E_{sn} from Eq. (136). Once these two variables are known, the limitation of the rate of snow melt, S_m (Eq. (134)), can be determined from the relationship of Eq. (135). Finally, the updated value for the snow pool size, W_{sn}, is calculated from Eq. (132).

Snow height, h_{sn}, is calculated from old-snow, ξ_{sn}, and fresh-snow density, ξ_{sn}^n:

$$h_{sn} = \frac{max\{W_{sn} - P_{sn} + E_{sn}, 0\}}{\xi_{sn}} + \frac{max\{P_{sn} - E_{sn}, 0\}}{\xi_{sn}^n} \tag{137}$$

with ξ_{sn}^n in kg m^{-3} (Loth and Graf, 1996):

$$\xi_{sn}^n = \begin{cases} 30 & \text{for} \quad \bar{T} \leq -22.5^o\text{C} \\ 10 + \frac{8}{3}(\bar{T} + 30) & \text{for} \quad -22.5^o\text{C} < \bar{T} \leq -15^o\text{C} \\ 50 + 1.7(\bar{T} + 15)^{1.5} & \text{for} \quad \bar{T} > -15^o\text{C} \end{cases} \tag{138}$$

and daily average temperature, $\bar{T}$, in ^{o}C.

Assuming vertically uniform density, i.e.

$$\xi_{sn}(t-\Delta t) = \frac{W_{sn}(t-\Delta t)}{h_{sn}(t-\Delta t)} \tag{139}$$

density of old snow is computed with a compaction rate by Anderson (1976) as follows:

$$\xi_{sn}(t) = \xi_{sn}(t-\Delta t)\left(1+\frac{g}{\eta_c^0}\exp\left[-a_c\xi_{sn}(t-\Delta t)+b_c\bar{T}\right]\frac{W_{sn}(t-\Delta t)\Delta t}{2}\right) \tag{140}$$

The constants are set according to the recommendations of Anderson (1976): $\eta_c^0 = 3.7{\times}10^7$ kg m^{-1} s^{-1}, $a_c = 2.1{\times}10^{-2}$m ^{3} kg^{-1} and $b_c = 8{\times}10^{-2}$ K^{-1}. In Eq. (137) a floor value of 10^{-9} kg m^{-3} is used for the densities of old and new snow.

B.3.8 Radiation balance

The radiative balance at the surface is computed in five steps:

1. Actual shortwave (solar) incoming radiation, R_{sw} is used as input.
2. Solar elevation, earth-sun distance, solar flux and height above sea level are computed from geographical position, Julian day and hour (UTC) and taken as input to the computation of potential solar incoming radiation in both the photosynthetically active (PAR) and near-infrared (NIR) domains.
3. Potential NIR (R_{NIR}^{pot}) and photosynthetically active incoming radiation (R_{PAR}^{pot}) in combination with actual solar incoming radiation are used to calculate actual incoming PAR, R_{PAR}.
4. The ratio r_{sw} of potential to actual incoming solar radiation is used to calculate the direct fraction of incoming PAR, as opposed to diffuse incoming PAR clouds and atmospheric scattering.
5. The radiative balance at both vegetated and bare-soil surfaces are calculated using incoming shortwave radiation (R_{sw}), LAI (Λ), fractional cover (f_c) and air temperature (T).

The second step begins with the computation of the inverse squared earth-sun distance in astronomical units, $r_\odot^{-2}$ (ie. in units of the average distance) according to Paltridge and Platt (1976):

$$\begin{aligned} r_\odot^{-2} &= 1.00011 + 0.034221\cos(\alpha_0) + \ldots \\ &\quad 0.0128\sin(\alpha_0)0.000719\cos(2\alpha_0) + \ldots \\ &\quad 0.000077\sin(2\alpha_0) \end{aligned} \tag{141}$$

with $\alpha_0 = 2\pi(d-1)/365$ and the Julian day d (1: January 1st; 365: December 31st). Solar elevation, μ, defined as the cosine of the angle between zenith and the position of the sun (i.e. $\mu = 1$ if the sun stands at zenith and $\mu = 0$ if the sun is at the horizon), is computed in the following way:

$$\mu = \sin(\phi)\sin(\delta) - \cos(\phi)\cos(\delta)\cos(t_{sol}\pi/12) \tag{142}$$

ϕ is the latitude and

$$\delta = -23.4^{\circ}\frac{\pi}{180}\cos\left(\frac{2\pi(d+10)}{365}\right) \tag{143}$$

is the position of the sun within the ecliptic, i.e. the latitude where the sun is at zenith at noon. d is Julian day and t_{sol} local solar time of day in hours. Computations are carried out in Greenwich mean time hours, t_{GMT}, which is local solar time at the zero degrees meridian. We convert t_{MGT} to t_{sol} by adding longitude of the site or the grid cell centre divided 15^{o}.

If $\mu < \mu_{lim} = 10^{-3}$, incoming radiation is neglected and treated as zero.

In the second step, PAR at the surface, R_{PAR}, is first calculated following Weiss and Norman (1985):

$$R_{PAR} = R_{sw}\frac{R^{pot}_{PAR}}{R^{pot}_{PAR} + R^{pot}_{NIR}} \tag{144}$$

with R_{sw} incoming shortwave radiation (a driving variable). Potential PAR is computed as

$$R^{pot}_{PAR} = t_{PAR}S_{0,PAR}r_{\odot}^{-2}\mu \tag{145}$$

with the total PAR transmittance

$$t_{PAR} = t^{max}_{PAR,d} + (1 - t^{max}_{PAR,d}\mu)t_{PAR,D} \tag{146}$$

with $t_{PAR,min} = 0.4$ and the PAR transmittance for direct radiation under clear skies

$$t_{PAR,D} = \exp\left(\frac{k_{PAR}}{\mu}\frac{p}{p_0}\right) \tag{147}$$

with the PAR atmospheric extinction coefficient $k_{PAR} = 0.185$. The transmittance for diffuse incoming PAR is

$$t_{PAR,d} = t_{PAR} - t_{PAR,D} = t^{max}_{PAR,d}(1 - \mu t_{PAR,D}). \tag{148}$$

Potential near infrared radiation (NIR) is computed as

$$R^{pot}_{NIR} = t_{NIR}S_{0,NIR}r_{\odot}^{-2}\mu \tag{149}$$

with the total NIR transmittance

$$t_{NIR} = t_{NIR,D} + t^{max}_{NIR,d}\left(1 - t_{NIR,D}\mu - \frac{w_{NIR}}{S_{0,NIR}}\right), \tag{150}$$

and the direction NIR transmittance

$$t_{NIR,D} = \exp\left(-\frac{k_{NIR}}{\mu}\frac{p}{p_0}\right) - \frac{w_{NIR}}{S_{0,NIR}} \tag{151}$$

with the NIR atmospheric extinction coefficient $k_{NIR} = 0.06$. Diffuse NIR transmittance is again defined as the remainder:

$$t_{NIR,d} = t_{NIR} - t_{NIR,D} = t^{max}_{NIR,d}\left(1 - t_{NIR,D}\mu - \frac{w_{NIR}}{S_{0,NIR}}\right), \tag{152}$$

and w_{NIR} is the water absorption in the NIR for 10 mm of precipitable water:

$$w_{NIR} = S_0 \times 10^{-1.1950+0.4459\log_{10}(1/\mu)-0.0345[\log_{10}(1/\mu)]^2}, \tag{153}$$

where $S_{0,PAR} = 600\text{Wm}^{-2}$, $S_{0,NIR} = 720\text{Wm}^{-2}$ and $S_0 = S_{0,PAR} + S_{0,NIR} = 1360\text{Wm}^{-2}$ is the solar constant. w_{NIR} is further limited by

$$w_{NIR} \leq S_{0,NIR} \exp\left(-\frac{0.06}{\mu}\frac{p}{p_0}\right). \tag{154}$$

The fraction of direct radiation in PAR, d_{PAR}, is also calculated according to Weiss and Norman (1985):

$$d_{PAR} = \left\{ \begin{array}{ll} 0 & \text{for} \quad r_{sw} < 0.2 \\ \left(1 - \left(\frac{0.9-r_{sw}}{0.7}\right)^{2/3}\right)\frac{t_{PAR,D}}{t_{PAR}} & \text{for} \quad 0.2 < r_{sw} < 0.9 \\ 1 & \text{for} \quad r_{sw} > 0.9 \end{array} \right\}, \tag{155}$$

where

$$r_{sw} = \frac{R_{sw}}{R^{pot}_{PAR} + R^{pot}_{NIR}} \tag{156}$$

is the ratio of actual to potential incoming shortwave radiation.

Further, we have p, the surface air pressure and $p_0 = 1.01325 \times 10^5$ Pa, the reference surface air pressure. p is computed from elevation, h, in m and daily mean temperature, $\bar{T}$, in K with the lapse rate of the standard atmosphere, $\mathcal{L}$ (6×10^{-3} K/m, Houghton, 1986):

$$p = p_0 \left(\frac{1}{1 + hl/\bar{T}}\right)^{gM_a/(R\mathcal{L})} \tag{157}$$

with the standard surface gravity, g (9.81 m s^{-1}).

The final step consists of calculating the radiative balance at the surface:

$$R_n = R_{L\downarrow} - R_{L\uparrow} + (1 - \rho_S)R_{sw} \tag{158}$$

where ρ_S stands for surface albedo, $R_{L\downarrow}$ for longwave thermal radiation from the atmosphere to the surface and $R_{L\uparrow}$ for long wave radiation back from the surface. As explained above, net radiation is divided into a vegetation and a soil part ($R_n = R_{n,v} + R_{n,s}$):

$$R_{n,v} = (1 - t_{l,v})(R_{L\downarrow} - R_{L\uparrow} - G) + a_v R_{sw} \tag{159}$$

and

$$R_{n,s} = t_{l,v}(R_{L\downarrow} - R_{L\uparrow}) + (1 - t_{l,v})G + a_s R_{sw} \tag{160}$$

$t_{l,v}$ is the longwave (thermal) transmissivity of vegetation, a_v and a_s the shortwave (solar) effective absorptivity of vegetation and soil, respectively (see Eq. (166)), and G the soil heat flux. G must be subtracted from $R_{n,s}$ to obtain the total available energy for evapotranspiration and latent heat

loss from the soil (see Eq. (131)). According to Verma et al. (1986), G is assumed to be a constant fraction of net radiation (cf. also Rosenberg, 1974, p. 179ff):

$$G = 0.036 R_n \tag{161}$$

It is assumed that the fraction $(1 - t_{l,v})$ of the soil heat flux is equal to the thermal radiation from the vegetation to the soil, so that this amount enters the radiative balances of both vegetation and soil.

Thermal upward radiation from the surface, $R_{L\uparrow}$, is computed from air temperature with a single value for surface emissivity, ε_O, of 0.97 (average for land surfaces, Brutsaert, 1982, p. 137):

$$R_{L\uparrow} = \varepsilon_O \sigma T_K^4 \tag{162}$$

with the Stefan-Boltzmann constant σ (5.6703$\times 10^{-8}$ Wm^{-2}K^{-4}) an the air temperature in Kelvin ($T_K = T + 273.16$), where T is air temperature in degrees Celsius.

Since thermal upwelling radiation is computed from air temperature, T, the effect of warming or cooling of both vegetation and soil surfaces is neglected. This effect can be accounted for by the concept of isothermal conductivity, which is discussed below (Section B.3.4).

Transmission of radiation through the vegetation canopy is computed from the two-flux equation with zero single-scattering albedo ($\omega = 0$, cf. Section B.2.3), which is equivalent to Beer's Law of radiation absorption:

$$t_{l,v} = f_c \exp(-\bar{\mu}\Lambda / f_c) + (1 - f_c) \tag{163}$$

with $\bar{\mu} = 1$ and f_c the fractional vegetation cover. In order to insure radiative balance between vegetation and soil, it is further assumed that the fraction $(1 - t_{l,v})$ of the soil heat flux comes from the net radiation of the vegetation canopy (see above).

Absorptivity of vegetation and soil, a_v and a_s, depends in a complex fashion on structure and distribution of the leaves, and on the optical properties of leaves and the soil. Here, both values are estimated on the basis of f_{PAR}, the fraction of absorbed PAR computed in the photosynthesis part of the model (see Eq. (80)):

$$a_s = (1 - \rho_{bg}) - (1 - \rho_{bg} - a_{s,0}) f_{PAR} \tag{164}$$

where $a_{s,0} = 0.05$ is the fraction absorbed by the soil under a closed canopy, and

$$a_v = (1 - \rho_v - a_{s,0}) f_{PAR} \tag{165}$$

ρ_v is the albedo of dense vegetation (standard value: 0.15, Brutsaert 1982, p. 136). With these two equations, surface albedo, ρ_S, can be expressed as

$$\rho_S = 1 - a_v - a_s = \rho_{bg} + (\rho_v b - \rho_{bg}) f_{PAR} \tag{166}$$

The given value for $a_{s,0}$ has been found with the two-flux equations for PAR and NIR at medium soil brightness (0.15 and 0.25) and a LAI of 3 (cf. Section B.2.3). Since the required accuracy for net radiation is lower than for absorption of PAR for photosynthesis calculations, this mode of

Table 15: Values for the soil albedo of three different brightness classes according to Wilson and Henderson-Sellers (1985), for both wet ($\rho_{s,w}$) and dry ($\rho_{s,d}$) soils.

brightness class	$\rho_{s,w}$	$\rho_{s,d}$
light	0.18	0.35
medium	0.10	0.20
dark	0.07	0.15

estimate should be sufficient for the complete range of f_{PAR}. The necessity to solve the two-flux equations for both PAR and NIR is thus avoided.

The value for the background albedo, ρ_{bg}, depends either on soil water content of the surface layer, W_s, and the brightness class of the soil, or, in the presence of snow (h_{sn}>0, Eq. (137)), on snow albedo, ρ_{sn}. In the absence of snow, the background albedo is the same as the soil albedo, i.e.

$$\rho_s = x_w \rho_{s,w} + (1 - x_w)\rho_{s,d} \tag{167}$$

where x_w reflects the wetting status of the surface soil layer. If surface-layer water content is at field capacity or above (i.e. $W_s \geq W_{s,f}$), the wet-soil value is assumed ($x_w = 1$). If the surface soil has dried out completely, the value for dry soils is taken ($x_w = 0$), while for intermediate values of the soil water content, the albedo is linearly interpolated between the two values, with weight

$$x_w = \min\{W_s/W_{s,f}, 1\}. \tag{168}$$

The albedo for wet and dry soils, $\rho_{s,w}$ and $\rho_{s,d}$, listed in Table 15, is determined by the brightness classification by Wilson and Henderson-Sellers (1985), which is part of the input data (see Section B.5.

If snow is present ($h_{sn} > 0$), ρ_{bg} follows the snow albedo, ρ_{sn}. In this case, we have

$$\rho_{bg} = (1 - f_{sn})\rho_s + f_{sn} * \rho_{sn} \tag{169}$$

with $f_{sn} = \min(h_{sn}/0.1, 1)$, the snow albedo ρ_{sn}, and snow height h_{sn}.

B.3.9 Atmospheric humidity

Since no reliable data of near-surface air humidity exist for purposes of global modelling, this quantity has to be estimated. In such cases (e.g. Running et al., 1987) it is often assumed that the daily mean of the vapour pressure is equal to the saturation vapour pressure at the daily minimum temperature. Friend (1998) has investigated this assumption with climate data by Müller (1982) and has found a good agreement for Europe and North America. For weather stations in arid regions, however, agreement is much less satisfactory, resulting in an overall correlation coefficient of $r^2 = 0.87$ for 805 stations. An overestimate of the vapour pressure occurs, when the air is not saturated at the minimum temperature, T_{min}, as under severe drought (Running et al., 1987), and an underestimate, when the vapour pressure rises during the day because of evapotranspiration (Rosenberg, 1974, p. 132ff).

In order to account for such findings, the daily course of the vapour pressure, $e_a(t)$, is calculated from instantaneous saturation vapour pressure, $e_s(T)$, saturation vapour pressure at sunrise, $e_s(T_{min})$, and the ratio f_e of daily mean evapotranspiration and daily mean evaporative demand. Variation of the dependence is achieved through the parameters h_0 (relative humidity at sunrise, when $T = T_{min}$, and total drought, i.e. $f_e = 0$) and $\hat{h}$ (daily amplitude of the vapour pressure under moist conditions, i.e. $f_e = 1$, as a fraction of the amplitude at constant saturation):

$$e_a = e_{a0} + f_{soil}\hat{h}\,(e_s(T) - e_{a0}) \tag{170}$$

where

$$e_{a0} = (h_0 + (1 - h_0)f_{soil})\,e_s(T_{min}) \tag{171}$$

with f_{soil} from Eq. (126). For the computation of e_a, we use the minimum temperature of the previous day for T_{min}. The saturation vapour pressure over water or ice, $e_s(T)$, is calculated from Murray (1967). This results in the following equation:

$$e_s(T) = \begin{cases} e_{s0}\exp\left(s_1^+ T/(s_2^+ + T)\right) & \text{for} \quad T > 0 \\ e_{s0}\exp\left(s_1^- T/(s_2^- + T)\right) & \text{for} \quad T < 0 \end{cases} \tag{172}$$

with $e_{s0} = e_S(0) = 610.78$ Pa, $s_1^+ = 17.269^o\mathsf{C}^{-1}$, $s_1^- = 22.33^o\mathsf{C}^{-1}$, $s_2^+ = 237.3^o\mathsf{C}$ and $s_2^+ = 271.15^o\mathsf{C}$. The parameters are set to $h_0 = 0.96$ and $\hat{h} = 0.49$ such that the results with the parametrisation of this model agree with the formulations found by Friend (1998), see Knorr (1997).

B.4 Carbon Allocation and Cycling

The overall driver of the carbon cycling module is $F_{NPP}(t)$, the NPP flux into the system at time t, which is partitioned into fraction described by factor f_i, where i is the corresponding pool.

B.4.1 Leaf dynamics and phenology

Leaf dynamics is controlled by inputs to the leafs and their losses. Inputs are the combination of direct allocation of NPP ($f_{fol}F_{NPP}(t)$) and seasonal inputs ($R_{lab}(t)$) from the labile pool controlled by the labile release factor $\tilde{\Phi}_{onset}$. Output from the leaves is controlled by a related leaf fall function $\tilde{\Phi}_{fall}$. $\tilde{\Phi}_{onset}$ and $\tilde{\Phi}_{fall}$ are based on the original DALEC labile release (Φ_{onset}) and leaf fall (Φ_{fall}) factors but temporally downscaled from daily to the sub-daily D&B time step.

Time step adjustment from DALEC's native daily time step to D&B's sub-daily step is computed following

$$\tilde{X} = [1 - (1 - X)]^{\Delta t/24\mathsf{h}}/(\Delta t/24\mathsf{h}) \tag{173}$$

where X is the respective factor of interest (Φ_{onset}, Φ_{fall}, θ_{fr}, θ_{wd}, θ_{lit}, θ_{decomp} and θ_{SOM}).

The dynamics of the labile pool are according to:

$$C_{lab}(t + \Delta t) = C_{lab}(t) + [P_{lab}(t) - R_{lab}(t)]\Delta t \tag{174}$$

with the labile carbon production (i.e. allocation from NPP)

$$P_{lab}(t) = (1 - f_{fol}) f_{lab} F_{NPP}(t) \tag{175}$$

and the labile carbon release to foliage

$$R_{lab}(t) = \tilde{\Phi}_{onset} C_{lab}(t). \tag{176}$$

Φ_{onset} is controlled by the day of year (doy, 1-365(6)), and parameters which define the day of year at which labile turnover is its greatest (d_{onset}) and the period over which turnover occurs (c_{ronset}). The residence time of the labile pool assumed to be 1 year ($c_{ll} = 1.001$).

$$\Phi_{onset}(doy, d_{onset}, c_{ronset}) = \frac{2}{\sqrt{\pi}} \cdot \frac{\kappa_{res}}{\theta_{res}} \cdot e^{-(\sin \frac{doy - d_{onset} + \psi_c(\ln(c_{ll}-1), \theta_{res})}{\sigma} \cdot \frac{\sigma}{\theta_{res}})^2} \tag{177}$$

where

$$\kappa_{res} = 0.5(\ln(c_{ll}) - \ln(c_{ll} - 1)) \tag{178}$$

and

$$\theta_{res} = 0.5\sqrt{2} c_{ronset} \tag{179}$$

$\sigma = 365.25/\pi$ and $\psi_c(\alpha, \beta)$ is an empirical correction for an offset in the sine function aligning with the number of days in a year. An equivalent correction is used in the leaf fall calculation. Therefore, $\alpha = \ln(c_{ll} - 1)$ for labile turnover and $\alpha = \ln(c_{lf} - 1)$ for leaf turnover. Similarly $\beta = \theta_{res}$ and θ_{fall} respectively.

$$\begin{aligned}\psi_c(\alpha, \beta) = (&2.35998 \times 10^{-5}\alpha^6 + 3.3273 \times 10^{-4}\alpha^5 + 9.0186526 \times 10^{-4}\alpha^4 - 5.5437737 \times 10^{-3}\alpha^3 \\ &- 0.020836\alpha^2 + 0.126972\alpha - 0.18846) \cdot \beta\end{aligned} \tag{180}$$

For more details on the phenology equations see Bloom and Williams (2015) Appendix A.

The dynamics of leaf area are determined in D&B from the change in leaf mass (C_{fol}). Leaf mass change is a dynamic outcome of allocation to leaves from the labile C pool (described above) and direct allocation from a fraction of current NPP, and timed losses from leaf senescence using the leaf fall function Φ_{fall}.

$$C_{fol}(t + \Delta t) = C_{fol}(t) + [P_{fol}(t) - R_{fol}(t) + R_{lab}(t)]\Delta t \tag{181}$$

with the foliage carbon production

$$P_{fol}(t) = f_{fol} F_{NPP}(t) \tag{182}$$

and the foliar-carbon release, or leaf litter production

$$R_{fol}(t) = \tilde{\Phi}_{fall}(doy, d_{fall}, c_{rfall}, c_{lf}) C_{fol}(t). \tag{183}$$

Losses from the foliar pool are linked to specific periods in the annual cycle through parameters for the day of leaf fall (d_{fall}) and period of labile release (c_{rfall}):

$$\Phi_{fall}(doy, d_{fall}, c_{rfall}) = \frac{2}{\sqrt{\pi}} \cdot \frac{\kappa_{fall}}{\theta_{fall}} \cdot e^{-(\sin \frac{doy - d_{fall} + \psi_c(\ln(c_{lf}-1), \theta_{res})}{\sigma} \cdot \frac{\sigma}{\theta_{fall}})^2} \tag{184}$$

where

$$\kappa_{fall} = 0.5(log(c_{lf}) - log(c_{lf} - 1)) \tag{185}$$

and

$$\theta_{fall} = 0.5\sqrt{2}c_{rfall} \tag{186}$$

c_{lf} is the inverse of the annual leaf loss fraction, commonly refered to as the leaf life span. ψ_f is an empirical offset term a function of c_{lf} described above but also see Bloom and Williams (2015) Appendix A.

Leaf area index is determined from foliar C and a parameter for leaf mass per area in C units (c_{LMA}, gCm^{-2} leaf area)

$$\Lambda = C_{fol}/c_{LMA} \tag{187}$$

B.4.2 Plant and soil carbon turnover

Dynamics of the fine root (fr) and wood pools (wd) are similarly determined by allocation of NPP and by first order turnover:

$$C_{fr}(t + \Delta t) = C_{fr}(t) + [P_{fr}(t) - R_{fr}(t)]\Delta t \tag{188}$$

with fine root carbon production

$$P_{fr}(t) = (1 - f_{fol} - (1 - f_{fol})f_{lab})f_{fr}F_{NPP}(t) \tag{189}$$

and fine root carbon release, or fine root litter production:

$$R_{fr}(t) = \tilde{\theta}_{fr}C_{fr}(t); \tag{190}$$

as well as

$$C_{wd}(t + \Delta t) = C_{wd}(t) + [P_{wd}(t) - R_{wd}(t)]\Delta t \tag{191}$$

with wood carbon production

$$P_{wd}(t) = F_{NPP}(t) - F_{fol}(t) - F_{lab}(t) - F_{fr}(t) \tag{192}$$

and wood carbon release, or wood litter production

$$R_{wd}(t) = \tilde{\theta}_{wd}C_{wd}(t). \tag{193}$$

Litter turnover is driven by both mineralisation to CO_2 and conversion to SOM by decomposition:

$$C_{lit}(t + \Delta t) = C_{lit}(t) + [R_{fol}(t) + R_{fr}(t) - R_{lit}(t) - F_{decomp}(t)]\Delta t \tag{194}$$

with the litter carbon respiration rate

$$R_{lit}(t) = \tilde{\theta}_{lit}f_R(T(t))C_{lit}(t) \tag{195}$$

and the litter to SOM carbon decomposition rate

$$F_{decomp}(t) = \tilde{\theta}_{decomp}f_R(T(t))C_{lit}(t). \tag{196}$$

$T(t)$ is air temperature at time t, and f_R the temperature factor of carbon turnover

$$f_R(T) = e^{\Theta T} \tag{197}$$

Litter fall from the foliar and fine root C pools are input to the litter pool, which has a faster turnover than the SOM pool. The litter pool has strongly periodic inputs linked to leaf senescence plus a continuous input from fine root mortality.

Dynamics of SOM are determined by inputs from turnover of the woody and litter pools, and by mineralisation linked to air temperature:

$$C_{SOM}(t + \Delta t) = C_{SOM}(t) + [R_{wd}(t) + F_{decomp}(t) - \theta_{SOM} f_R(T(t)))C_{SOM}(t)]\Delta t \tag{198}$$

with the SOM carbon respiration rate

$$R_{SOM}(t) = \tilde{\theta}_{SOM} f_R(T(t))C_{SOM}(t). \tag{199}$$

Total ecosystem respiration defined as the sum of autotrophic respiration, R_A, using inputs from Eq. (60), and heterotrophic respiration, R_H, which depends on the rate of mineralisation of dead organic matter, litter and soil organic material:

$$R_{reco}(t) = R_A(t) + R_H(t) \tag{200}$$

with

$$R_H(t) = (\theta_{lit} C_{lit}(t) + \theta_{SOM} C_{SOM}(t)) f_R(T(t)). \tag{201}$$

B.4.3 Parameter priors

Prior values for the PFT specific parameters related to carbon allocation and cycling were created using the CARDAMOM model-data fusion system (Bloom et al., 2016) for the Finland (Table 16) and Spain sites (Table 17). Further PFTs of Table 1 that are not listed in these two tables, are provided in Table 18. For PFT 10, C4 grass, we recommend to use the prior values of PFT 9, C3 grass. CARDAMOM was used to calibrate a previous version of DALEC coupled to a water cycle model (Bloom and Williams, 2015; Smallman and Williams, 2019) at the site scale and across two domains of around 500 by 500 km surrounding those sites, at 0.25×0.25 degrees spatial resolution and with a monthly time step for the years 2001 to 2018, inclusive.

This version of DALEC used the Aggregrated Canopy Model (ACM (Williams et al., 1997)) to provide canopy flux inputs rather than BETHY. The analyses were driven with CRU-JRAv2.1 meteorology (Harris, 2019), MODIS burned area (Giglio et al., 2018) and Global Forest Watch forest cover loss (Hansen et al., 2013). The parameters were retrieved based on calibration against time series of leaf area index (Copernicus Service Information (2020)), woody biomass stock information (Santoro, 2021) and an initial soil C stock (Hengl et al., 2017). Due to the availability of field data for specific PFTs, three site-level analyses were further carried out to provide more constrained diagnostics. These sites included an evergreen coniferous (EvCn) forest using field observations from Hyytiälä, Finland (FLUXNET2015 database (`http://fluxnet.fluxdata.org/`, accessed 01/11/2016); Heiskanen et al., 2012), and both C3 grasslands (C3Gr) and evergreen coniferous

Table 16: D&B parameters relating to carbon balance and phenology used for Sodankylä site, by plant functional type (PFT; Columns 2, 3). Columns 3-8: same D&B parameters, but applicable for the entire northern Scandinavian study region (18^{o}E - 32^{o}E, 65^{o}N - 69^{o}N).

parameter / PFT (#)	EvCn (5)	EShr (7)	TmSg (4)	EvCn (5)	C3Gr (9)	Tundra (11)	WetV (12)
θ_{decomp}	4.60×10^{-4}	3.30×10^{-4}	5.20×10^{-4}	6.80×10^{-4}	2.65×10^{-4}	2.57×10^{-4}	3.20×10^{-4}
f_{fol}	0.118	0.167	0.184	0.090	0.171	0.181	0.156
f_{fr}	0.277	0.32	0.307	0.276	0.44	0.40	0.34
c_{lf}	1.19	1.52	1.09	1.17	1.21	1.19	1.03
θ_{wd}	1.25×10^{-4}	1.32×10^{-4}	1.81×10^{-4}	9.70×10^{-5}	3.03×10^{-4}	3.40×10^{-4}	1.78×10^{-4}
θ_{fr}	0.0072	0.0043	0.0091	0.0064	0.0067	0.0062	0.0073
θ_{lit}	0.0059	0.00132	0.0039	0.0045	0.00126	0.00123	0.00087
θ_{SOM}	1.57×10^{-5}	3.60×10^{-6}	2.33×10^{-5}	2.08×10^{-5}	7.70×10^{-6}	6.90×10^{-6}	4.00×10^{-6}
Θ	0.048	0.040	0.041	0.042	0.040	0.040	0.039
d_{onset}	156.13	69.18	162.31	136.88	178.69	177.78	149.81
f_{lab}	0.136	0.093	0.129	0.146	0.187	0.151	0.258
c_{ronset}	29.28	33.29	31.04	43.76	20.53	23.00	22.28
d_{fall}	230.32	258.47	239.75	232.19	252.41	253.59	249.68
c_{rfall}	50.86	40.57	48.55	58.18	34.74	34.54	45.88
c_{LMA}	36.11	69.86	45.80	41.99	57.28	50.24	62.69
Initial C_{lab}	30.48	3.75	41.87	37.89	27.32	17.06	22.60
Initial C_{fol}	29.29	16.16	34.72	27.04	27.12	20.66	2.82
Initial C_{fr}	17.86	7.72	29.09	31.40	16.85	14.89	8.09
Initial C_{wd}	3072.24	500.70	3992.10	4689.19	613.07	371.10	743.32
Initial C_{lit}	60.24	41.04	153.42	76.40	158.34	101.35	92.53
Initial C_{SOM}	40910.35	34888.61	37511.67	34302.16	44359.82	35311.00	40989.89

forest (EvCn) using observations at Majadas de Tietar. These three analyses took advantage of in-situ estimates of NEE data from eddy covariance and field inventories of LAI and biomass. All other PFT parameter priors were selected based on CARDAMOM grid points which had the largest coverage of each target PFT and realistic parameter retrievals (e.g. evergreen PFTs had leaf life spans, c_{lf}, $>$1 year).

B.5 Model setup

In the current setup ($\Delta t = 1$hour, Section B.1), the model requires the following driving variables at an hourly time step over the integration period:

- 2m air temperature (T)
- precipitation (P_{tot})
- incoming shortwave (solar) radiation (R_{sw})
- incoming longwave (thermal) radiation ($R_{L\downarrow}$)

Additional inputs are:

- soil temperature, used as daily mean ($\bar{T}_{ds}$)
- molar ratio of CO_2 in air (C_a)
- soil texture class

Table 17: D&B parameters relating to carbon balance and phenology used for Majadas del Tietar site, by plant functional type (PFT; Columns 2, 3). Columns 4-9: same D&B parameters, but applicable for the entire Iberian study region (8.5^{o}W - 2.5^{o}W, 38.5^{o}N - 43^{o}N).

parameter / PFT (#)	TmEv (3)	C3Gr (9)	TmSg (4)	EvCn (5)	EShr (7)	C3Gr (9)	WetV (12)	ArbC (13)
θ_{decomp}	3.04×10^{-4}	2.58×10^{-3}	5.30×10^{-4}	4.80×10^{-4}	5.90×10^{-4}	9.60×10^{-4}	9.10×10^{-4}	7.40×10^{-4}
f_{fol}	0.050	0.36	0.163	0.139	0.193	0.45	0.163	0.278
f_{fr}	0.36	0.71	0.35	0.51	0.62	0.79	0.53	0.74
c_{lf}	1.10	1.01	1.18	1.79	1.23	1.00	1.17	1.00
θ_{wd}	6.60×10^{-5}	7.30×10^{-4}	1.94×10^{-4}	2.08×10^{-4}	4.90×10^{-4}	9.60×10^{-4}	5.00×10^{-4}	7.80×10^{-4}
θ_{fr}	0.0059	0.0087	0.0076	0.0050	0.0099	0.0126	0.0084	0.0117
θ_{lit}	0.0069	0.00055	0.0043	0.0040	0.0042	0.0058	0.0038	0.0042
θ_{SOM}	2.48×10^{-5}	4.00×10^{-5}	3.04×10^{-5}	2.55×10^{-5}	1.83×10^{-5}	1.04×10^{-5}	2.03×10^{-5}	1.55×10^{-5}
Θ	0.051	0.034	0.043	0.042	0.042	0.060	0.045	0.042
d_{onset}	74.31	77.34	139.84	145.00	82.33	104.50	329.86	106.80
f_{lab}	0.220	0.42	0.148	0.107	0.139	0.0308	0.188	0.215
c_{ronset}	74.18	28.09	20.33	20.64	44.56	26.08	38.56	30.95
d_{fall}	157.38	121.97	260.15	268.05	176.29	129.66	151.99	163.70
c_{rfall}	54.63	65.05	96.93	113.03	67.99	63.26	41.56	71.49
c_{LMA}	67.67	46.94	60.25	104.96	104.66	56.52	107.23	45.11
Initial C_{lab}	34.57	60.99	80.62	30.77	32.41	5.16	5.51	40.37
Initial C_{fol}	36.34	24.84	64.07	139.71	73.05	57.56	128.16	11.35
Initial C_{fr}	34.92	10.64	67.02	97.49	30.30	49.35	29.74	18.86
Initial C_{wd}	6737.40	89.57	5962.44	2227.35	415.67	140.53	874.43	127.46
Initial C_{lit}	12.90	16.35	205.51	146.72	51.40	4.17	21.50	49.60
Initial C_{SOM}	11818.84	15469.69	26847.23	19030.57	13952.27	13573.13	12117.49	12847.95

Table 18: D&B parameters relating to carbon balance and phenology and not covered in tables 16 and 17, by plant functional type (PFT; Columns 2-5).

parameter / PFT (#)	TrEV	TrDD	DSrh	SgCn
θ_{decomp}	1.11E-03	1.48E-03	1.04E-03	6.65E-04
f_{fol}	1.18E-01	1.27E-01	2.65E-01	1.60E-01
f_{fr}	3.46E-01	5.05E-01	6.28E-01	5.10E-01
c_{lf}	4.23E+00	1.53E+00	1.35E+00	1.10E+00
θ_{wd}	5.70E-05	2.12E-04	5.06E-05	3.88E-04
θ_{fr}	3.31E-03	2.27E-03	4.66E-03	3.08E-03
θ_{lit}	6.73E-03	5.92E-03	5.85E-03	5.39E-03
θ_{SOM}	1.53E-05	1.24E-05	7.47E-06	3.19E-05
Θ	4.83E-02	5.33E-02	5.20E-02	3.65E-02
d_{onset}	9.55E+01	3.04E+02	3.32E+02	1.62E+02
f_{lab}	4.77E-02	4.48E-02	5.43E-02	1.91E-01
c_{ronset}	2.96E+01	3.55E+01	3.55E+01	2.42E+01
d_{fall}	1.72E+02	1.61E+02	3.04E+02	2.36E+02
c_{rfall}	3.34E+01	8.66E+01	9.62E+01	3.89E+01
c_{LMA}	8.58E+01	5.20E+01	6.71E+01	5.45E+01
Initial C_{lab}	1.68E+01	3.42E+00	7.47E+00	2.93E+01
Initial C_{fol}	4.34E+02	1.05E+02	2.67E+02	1.69E+01
Initial C_{fr}	1.58E+02	3.03E+02	3.22E+02	6.84E+01
Initial C_{wd}	1.81E+04	2.74E+03	1.53E+04	5.26E+02
Initial C_{lit}	1.96E+01	2.35E+01	2.01E+02	9.66E+01
Initial C_{SOM}	2.97E+04	1.99E+04	4.42E+04	1.45E+04

- subgrid cell fractions
- PFTs for each subgrid cell
- fractional vegetation cover (f_c)
- elevation (h, to computer air pressure, Eq. (157)).

C The Layered 2-Stream Model

The layered 2-stream model runs in parallel to the two-flux scheme of the photosynthesis part, but being based on the same equations reproduces identical results. It is part of the observation operator for solar-induced fluorescence (SIF).

To simulate the SIF leaving the top of the canopy, we take the Meador and Weaver (1980) solutions to the radiative transfer problem given for the reflectance ($\mathcal{R}_d$) and transmittance ($\mathcal{T}_d$) of discrete canopy layers given diffuse incident light, which in this case, will be the SIF emission from within the layer, to give:

$$\mathcal{R}_d = \frac{\gamma_2[1 - e^{-2k\tau}]}{\tilde{k} + \gamma_1 + (\tilde{k} - \gamma_1)e^{-2k\tau}} \tag{202}$$

$$\mathcal{T}_d = \frac{2e^{-2k\tau}}{\tilde{k} + \gamma_1 + (\tilde{k} - \gamma_1)e^{-2k\tau}} \tag{203}$$

with:

$$\tau = \frac{LAI}{2} \tag{204}$$

$$\tilde{k} = \sqrt{\gamma_1^2 - \gamma_2^2} \tag{205}$$

$$\gamma_1 = 2\left(1 - \omega(1 - \beta)\right), \tag{206}$$

$$\gamma_2 = 2\omega\beta. \tag{207}$$

where β in D&B is always 0.5 (see Eq. (71)). These equations are physically consistent with the canopy radiative transfer elsewhere in D&B, but allow for the inclusion of arbitrary emission sources from within canopy layers.

To test the physical consistency we compare the FAPAR calculated by D&B with the same quantity constructed from the L2SM equations. For the case of a black soil the FAPAR is given, trivially, by:

$$\mathrm{FAPAR}_{bs} = 1 - \mathcal{R}_d - \mathcal{T}_d, \tag{208}$$

and for a soil with reflectance ρ_s:

$$\mathcal{R}_d(\rho_s) = \mathcal{R}_d + \frac{\mathcal{T}_d^2 \rho_s}{1 - \mathcal{R}_d \rho_s}, \tag{209}$$

$$\mathcal{T}_d(\rho_s) = \frac{\mathcal{T}_d}{1 - \mathcal{R}_d \rho_s}, \tag{210}$$

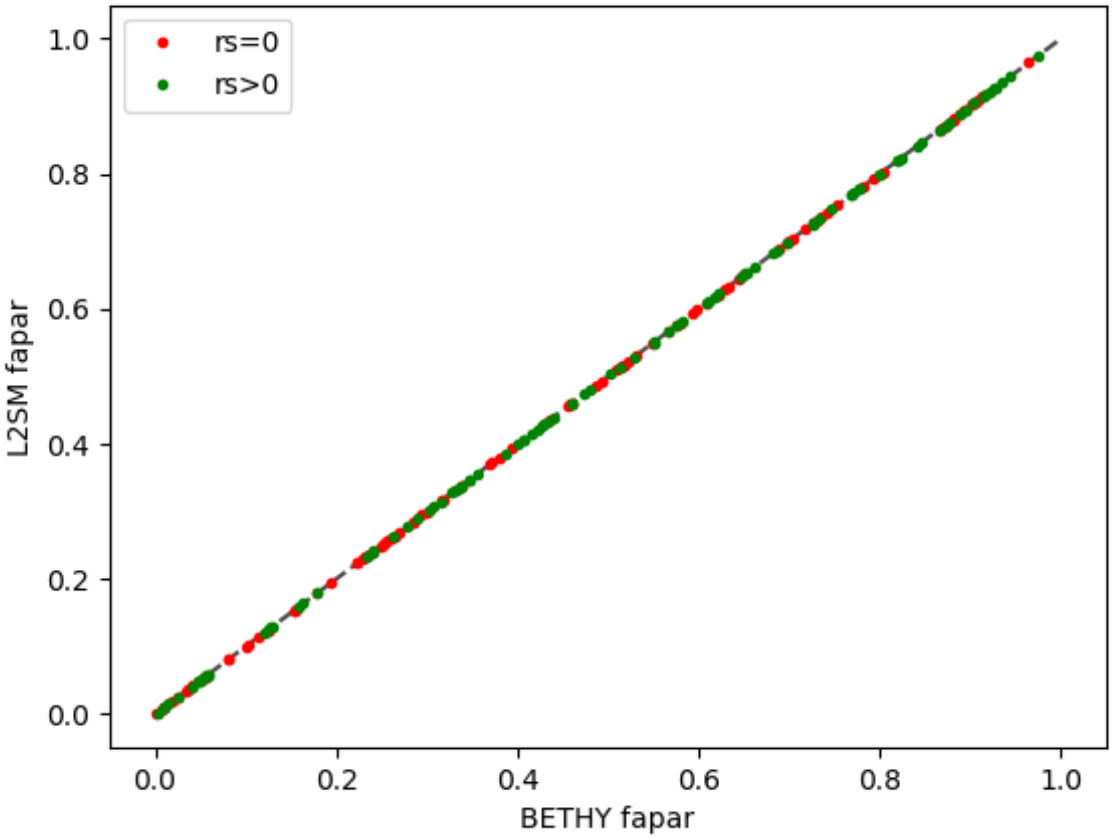


Figure 13: FAPAR predicted by the L2SM and the twoflux scheme of D&B for the case of $\rho_s = 0$ and $\rho_s > 0$. In both cases 100 simulations have been used.

$$\text{FAPAR} = 1 - \mathcal{R}_d(\rho_s) - \mathcal{T}_d(\rho_s)(1 - \rho_s), \tag{211}$$

We can then compare the FAPAR predicted by the two-flux scheme of D&B to these equations. While D&B uses a value of 0.12 for ω, we use a random sample of values between zero and one, as shown in Table 19, 100 times for both soil cases, leads to the scatter plot given in Figure 13. The near-perfect straight line relationship between the two models indicates that the physics and assumptions in the two models is, for all purposes, the same. We chose a log transform of the uniform distribution (U) to sample LAI as this produces an approximately linear distribution in the resulting FAPAR.

The SIF emitted from each layer that escapes the top of the canopy can be computed as:

Table 19: Parameter space for FAPAR tests. $U(a-b)$ stands for the uniform distribution with boundaries a and b.

Parameter	Distribution
Λ	$-2\ln(1-U(0-1))$
ω	$U(0-1)$
ρ_s	$U(0-1)$

$$\mathcal{E}_n = \frac{S_n \mathcal{T}^+_{d,n}\left(1+\mathcal{R}'^-_{d,n}\right)}{1-\mathcal{R}'^-_{d,n}\mathcal{R}^+_{d,n}}, \tag{212}$$

where S_n is the SIF emitted from layer n, assuming it occurs at the mid–point of the layer. The superscript $+$ and $-$ are used to refer the optical properties of the canopy level above $(+)$ and below $(-)$ the mid-point of layer n. For example, $\mathcal{T}^+_{d,n}$ is the transmittance of the canopy above the middle of layer n, $\mathcal{R}^-_{d,n}$ is the reflectance of the canopy below the middle of layer n, and includes the reflectance of the soil. The total canopy leaving SIF is then given by:

$$\mathcal{E} = \sum_{n=1}^{N} \mathcal{E}_n \tag{213}$$

where N is the number of layers. This formulation explicitly accounts for all levels of SIF photon scattering between the canopy layers and the soil below the canopy, as well as re-absorption within the canopy.

The assumption that the emitted SIF comes from the middle of each layer is valid if the layers are optically thin (i.e. optical properties change proportionally with changes in optical depth) and that SIF is generated uniformly throughout the layer (which is consistent with the photosynthesis routines in D&B, which assume all leaves in a layer are photosynthesising at the same rate). To test the impact of the optically thin assumption we generate 50 different canopies with optical properties sampled according to Table 19 and calculated the total canopy leaving emissions for each canopy dividing it into all possible numbers of layers from 1 to 50. Results are shown in Figure 14 and show the percentage difference from the simulation with 50 layers. Even in the most extreme case, with a single layer, the difference is less that 1% from the 50 layer calculation. For three layers (as in D&B) the difference is always less than 0.5% and almost always less the 0.05%. Consequently we can be confident that only very small levels of uncertainty are introduced by this specific assumption.

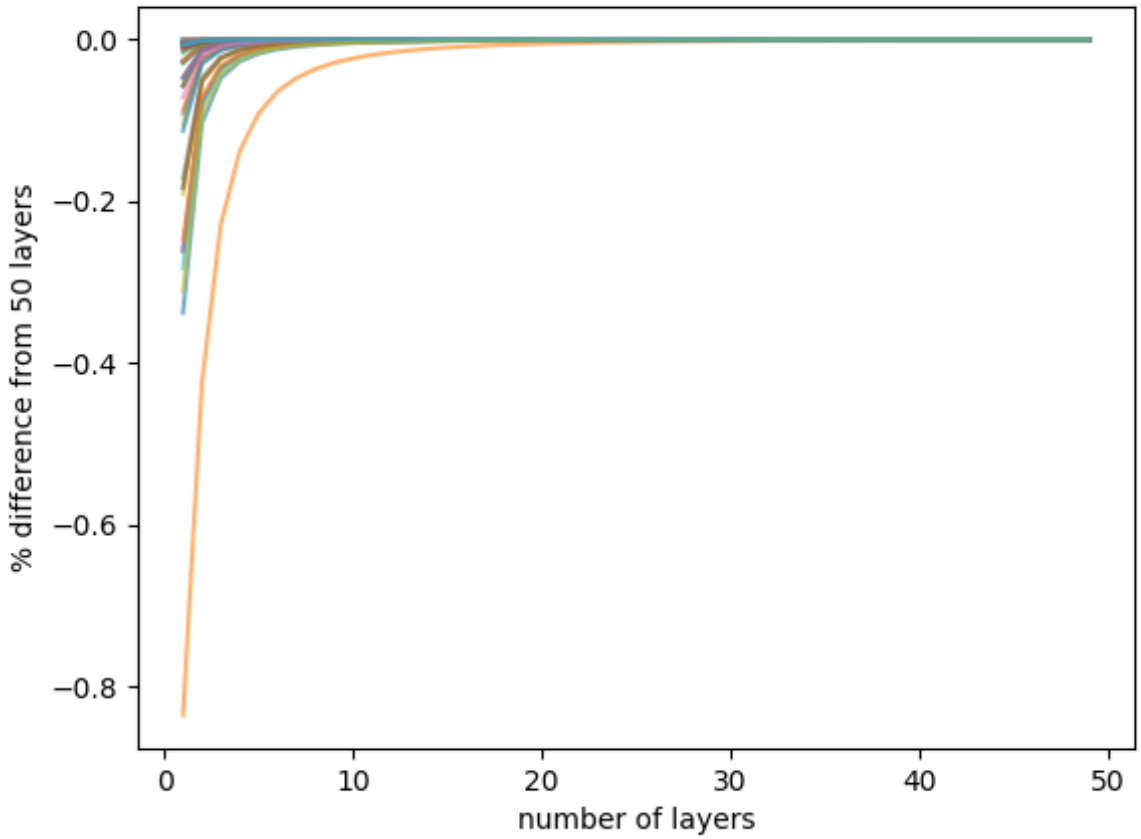


Figure 14: % difference in emissions from 50 random canopies with 1 to 50 layers.

D Static forcing for Majadas de Tietar at site-level

```
netcdf ES-LM1_staticforcing {
dimensions:
  ng = 1 ;
  nmaxpft = 14 ;
variables:
  double lon(ng) ;
    lon:long_name = "longitude" ;
    lon:units = "degrees_east" ;
    lon:comment = "longitude of selected pixel" ;
  double lat(ng) ;
    lat:long_name = "latitude" ;
    lat:units = "degrees_north" ;
    lat:comment = "latitude of selected pixel" ;
  int pft(ng, nmaxpft) ;
    pft:_FillValue = -1 ;
    pft:long_name = "Plant Functional Type" ;
    pft:units = "" ;
    pft:vmin = 0 ;
    pft:vmax = 13 ;
    pft:values = "0,1,2,3,4,5,6,7,8,9,10,11,12,13" ;
    pft:meanings = "NoVeg,TrEv,TrDD,TmEv,TmSg,EvCn,SgCn,EShr,DShr,C3Gr,C4Gr,Tun,VSwamp,ArbC" ;
    pft:info = "value(s) provided by local information/measurement." ;
  double pft_fraction(ng, nmaxpft) ;
    pft_fraction:_FillValue = -1. ;
    pft_fraction:long_name = "Plant Functional Type area fraction" ;
    pft_fraction:units = "" ;
    pft_fraction:vmin = 0. ;
    pft_fraction:vmax = 1. ;
    pft_fraction:comment = "area fraction is the fraction of the vegetated part of the location/grid-cell covered by the specific PFT. The sum of the PFT fractions equals 1." ;
    pft_fraction:info = "value(s) provided by local information/measurement." ;
  double vegetation_fraction(ng) ;
    vegetation_fraction:_FillValue = -1. ;
    vegetation_fraction:long_name = "fraction of gridcell covered by vegetation" ;
    vegetation_fraction:units = "" ;
    vegetation_fraction:vmin = 0. ;
    vegetation_fraction:vmax = 1. ;
    vegetation_fraction:info = "value(s) provided by local information/measurement." ;
  double soil_depth(ng) ;
    soil_depth:_FillValue = -99999. ;
    soil_depth:long_name = "soil profile depth" ;
    soil_depth:units = "mm" ;
    soil_depth:info = "value(s) provided by local information/measurement." ;
  int soil_texture_class(ng) ;
    soil_texture_class:_FillValue = -1 ;
    soil_texture_class:long_name = "Dominant_Soil_Texture" ;
    soil_texture_class:units = "" ;
    soil_texture_class:vmin = 0 ;
    soil_texture_class:vmax = 7 ;
    soil_texture_class:values = "0,1,2,3,4,5,6,7" ;
    soil_texture_class:meanings = "Ocean/Water,Coarse,Medium/Coarse,Medium,Fine/Medium,Fine,Organic,Ice" ;
  int soil_brightness_class(ng) ;
    soil_brightness_class:_FillValue = -9999 ;
    soil_brightness_class:units = "-" ;
    soil_brightness_class:vmin = 1 ;
    soil_brightness_class:vmax = 5 ;
    soil_brightness_class:values = "1,2,3,4,5" ;
    soil_brightness_class:meanings = "dark,medium,light,ice,water" ;
  int elevation(ng) ;
    elevation:_FillValue = -9999 ;
    elevation:long_name = "elevation_above_sea_level" ;
    elevation:units = "m" ;
    elevation:info = "value(s) provided by local information/measurement." ;
  double B_r(ng) ;
    B_r:_FillValue = -1. ;
    B_r:long_name = "shape parameter of root-zone infiltration subgrid distribution" ;
    B_r:units = "" ;
  byte condition_simulate(ng) ;
    condition_simulate:_FillValue = 127b ;
    condition_simulate:long_name = "Flag indicating whether location is suitable for simulation." ;
    condition_simulate:units = "" ;
    condition_simulate:vmin = 0b ;
    condition_simulate:vmax = 1b ;
    condition_simulate:comment = "primary-pft!=NoVeg, 1<=soil_texture_class<=6, soil_depth!=fill-value, 1<=soil_brightness_class<=4, elevation!=fill-value, B_r!=fill-value" ;

// global attributes:
    :creator_email = "Michael.Vossbeck(at)Inversion-Lab.com" ;
    :region = "majadas" ;
    :nsp = 2 ;
    :location_name = "ES-LM1" ;
    :comment = "static forcing for single location/site. Values for the data sets pft, pft_fraction, fc, soil_depth, elevation are coming from in-situ/local information or measurements. Values of all further data sets
    :geospatial_lon_min = -6. ;
    :geospatial_lon_max = -5.75 ;
    :geospatial_lon_resolution = 0.25 ;
    :geospatial_lon_units = "degrees_east" ;
    :geospatial_lat_min = 39.75 ;
    :geospatial_lat_max = 40. ;
    :geospatial_lat_resolution = 0.25 ;
    :geospatial_lat_units = "degrees_north" ;
    :input_file_1 = "/srv/data/lcc/model_input/staticinput_qd_majadas.nc" ;
    :input_file_1_variables = "soil_texture,brightness_class,b" ;
    :input_file_1_sha512 = "60f20723fbc39de28eeba97ee602345e23c55a8d250fceb51b3fa9f1bfa5119693b4f0f646a11a92489fb33d233d27d139108e4577c032dd904810af38aab5dd" ;
    :soil_texture_class___reference = "Randal D. Koster NASA/GSFC, Norman B. Bliss Principal Scientist Science and Applications Branch, Saud A. Amer ECS DAAC Scientist EROS Data Center, Soroosh Sorooshian Professor and
    :soil_texture_class___url = "https://iridl.ldeo.columbia.edu/SOURCES/.NASA/.ISLSCP/.GDSLAM/.Hydrology-Soils/.soils/.dataset_documentation.html" ;
```

```
    :soil_texture_class___version = "April 5, 1995" ;
    :history = "util/lcc_prepare.py forcing4model_static --site=ES-LM1 --outname forcing/ES-LM1_staticforcing.nc" ;
    :date_created = "2024-02-26T11:29:20.943679" ;
}
```

E Meteorological forcing for Majadas de Tietar at site-level derived from in-situ observations

```
netcdf ES-LM1_dynforcing-insitu_20140401-20220930_with-insitu-lwdown {
dimensions:
        ng = 1 ;
        time = 74520 ;
        ntc = 4 ;
variables:
  double time(time) ;
    time:standard_name = "time" ;
    time:long_name = "time" ;
    time:units = "hours since 1900-01-01 00:00:00.0" ;
  int yyyymmddhh(ntc, time) ;
    yyyymmddhh:_FillValue = -1 ;
    yyyymmddhh:long_name = "year/mon/day/hour_as_integer_values" ;
    yyyymmddhh:units = "" ;
    yyyymmddhh:comment = "time values are provided in UTC time, 1<=hour<=24" ;
  double lon(ng) ;
    lon:standard_name = "longitude" ;
    lon:long_name = "longitude" ;
    lon:units = "degrees_east" ;
    lon:axis = "X" ;
    lon:comment = "longitude of single point" ;
  double lat(ng) ;
    lat:standard_name = "latitude" ;
    lat:long_name = "latitude" ;
    lat:units = "degrees_north" ;
    lat:axis = "Y" ;
    lat:comment = "latitude of single point" ;
  double temperature(ng, time) ;
    temperature:_FillValue = -99999. ;
    temperature:long_name = "2meter air temperature" ;
    temperature:units = "Celsius" ;
    temperature:comment = "averaged in-situ temperature from values on hour ending at actual time-stamp and the hour thereafter." ;
  double swrad(ng, time) ;
    swrad:_FillValue = -99999. ;
    swrad:long_name = "incoming solar radiation" ;
    swrad:units = "W/m2" ;
    swrad:comment = "averaged in-situ radiation from values on hour ending at actual time-stamp and the hour thereafter." ;
  double lwdown(ng, time) ;
    lwdown:_FillValue = -99999. ;
    lwdown:long_name = "downwelling long-wave radiation" ;
    lwdown:units = "W/m2" ;
    lwdown:comment = "averaged in-situ long-wave radiation from values on hour ending at actual time-stamp and the hour thereafter." ;
  double precipitation(ng, time) ;
    precipitation:_FillValue = -99999. ;
    precipitation:long_name = "precipitation" ;
    precipitation:units = "mm/h" ;
    precipitation:comment = "averaged in-situ precipitation from values on hour ending at actual time-stamp and the hour thereafter." ;
  double soil_temperature(ng, time) ;
    soil_temperature:_FillValue = -99999. ;
    soil_temperature:long_name = "soil temperature" ;
    soil_temperature:units = "Celsius" ;
    soil_temperature:comment = "soil-temperature averaged from STa080, STb080, STc080, STd080, averaged in-situ soil temperature from values on hour ending at actual time-stamp and the hour thereafter." ;

// global attributes:
    :comment = "forcing data prepared at pixel level" ;
    :pixel_longitude = -5.778625 ;
    :pixel_latitude = 39.942628 ;
    :pixel_identifier = "ES-LM1" ;
    :insitu_input_file = "/srv/data/lcc/insitu/majadas-20221109/ES-LM1_LCC_2014_2022_60min_20221109.csv" ;
    :insitu_input_file_sha512 = "08da52912d07a41105d1028a033e852c204bf374dc1410c96a48ece937652d49e3d8c72b9eec0d56324c6be15b031bf903e79a95f48dfe38d110afbda667b300" ;
    :history = "util/lcc_prepare.py dynforcing4model_insitu /srv/data/lcc/insitu/majadas-20221109/ES-LM1_LCC_2014_2022_60min_20221109.csv --dates 20140401 20220930 --add_longwave_radiation --outname forcing/ES-LM1_dynfor
    :date_created = "2023-12-07T12:19:00.354318" ;
}
```

F Template file used in preprocessing step to generate `dimensions.f90`

```
! D&B spatio-temporal settings
! 09/20 - 04/23 Vossbeck, Knorr, Kaminski (all ilab)
!
!
! This program is free software: you can redistribute it and/or modify
! it under the terms of the GNU Affero General Public License as
! published by the Free Software Foundation, either version 3 of the
! License, or (at your option) any later version.
!
! This program is distributed in the hope that it will be useful,
! but WITHOUT ANY WARRANTY; without even the implied warranty of
! MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
! GNU Affero General Public License for more details.

! You should have received a copy of the GNU Affero General Public License
! along with this program.  If not, see <http://www.gnu.org/licenses/>.
!
! Contact: Thomas.Kaminski@Inversion-Lab.com
!
module dimensions
  implicit none
  !-- spatial settings
  integer, parameter :: nsp     =  NSP  !-- number of sample points (differentiates by combinations of spatial locations ar
  integer, parameter :: ng      =  NG   !-- number of spatially different grid cells or simulation locations
  character(len=*), parameter :: domain = DOMAIN !-- domain descriptor
  !-- temporal settings
  !   - D&B currently does not support simulations on sub-daily time intervals,
  !     but runs from selected day of start to day of end.
  !   - the simulation period ranges from yyyymmdd_startT01:00:00 to yyyymmdd_endT24:00:00
  !     where time is specified in UTC.
  character(len=8), parameter :: yyyymmdd_start = YYYYMMDD_START
  character(len=8), parameter :: yyyymmdd_end   = YYYYMMDD_END
  character(len=*), parameter :: calendar_type  = CALENDAR !-- 'gregorian' or '365_days'
  integer, parameter :: nmon    = NMON
  integer, parameter :: nday    = NDAY
  integer, parameter :: nhour   = nday*24
  integer, parameter :: npft    = 13  ! number of plant functional types (PFTs) available to the model
  integer, parameter :: napft       = NAPFT  ! number of active biomes (PFT types)
  integer, parameter :: apft(napft) = (/APFT_LST/) ! list of active biome identifiers in ascending order
  integer, parameter :: ntxture = 6 ! number of texture classes
  integer, parameter :: nlayer  = 3 ! number of canopy layers
end module dimensions
```

G Initial lines of observational data file for assimilation (SIF, Sodankylä)

ihour	SIF_743NM	sigma SIF_743NM	ipft 1	fcov ipft 1	ipft 2	fcov ipft 2
29194	1.03807725E-01	3.52221072E-01	5	6.70000000E-01	7	3.30000000E-01
29361	6.91338003E-01	2.94855714E-01	5	6.70000000E-01	7	3.30000000E-01
29363	-1.39694616E-01	2.87125349E-01	5	6.70000000E-01	7	3.30000000E-01
29385	9.95292440E-02	2.91882455E-01	5	6.70000000E-01	7	3.30000000E-01
29386	3.08853477E-01	2.65680939E-01	5	6.70000000E-01	7	3.30000000E-01
29408	6.37560010E-01	2.45790735E-01	5	6.70000000E-01	7	3.30000000E-01
29410	-3.01566690E-01	2.48386264E-01	5	6.70000000E-01	7	3.30000000E-01
29412	-1.56198338E-01	2.54602224E-01	5	6.70000000E-01	7	3.30000000E-01
29432	2.90916353E-01	2.33601242E-01	5	6.70000000E-01	7	3.30000000E-01
29434	-1.70324162E-01	2.29710072E-01	5	6.70000000E-01	7	3.30000000E-01
29435	-9.67503712E-02	2.39188030E-01	5	6.70000000E-01	7	3.30000000E-01
29459	5.00491261E-01	3.02187860E-01	5	6.70000000E-01	7	3.30000000E-01
29481	6.19945228E-01	2.27921784E-01	5	6.70000000E-01	7	3.30000000E-01
29483	1.44792721E-04	2.29299068E-01	5	6.70000000E-01	7	3.30000000E-01