



Implementing Multi-scale AGricultural Indicators Exploiting Sentinels

ATBD FOR LAND DATA ASSIMILATION SYSTEM

IMAGINES_RP3.1_ATBD-LDAS

ISSUE I2.10

EC Proposal Reference N° FP7-311766

Due date of deliverable: 30.06.2015

Actual submission date: 16.07.2015

Start date of project: 01.11.2012

Duration : 40 months

Name of lead partner for this deliverable: METEO FRANCE

Book Captain: Jean-Christophe Calvet (Meteo-France)
Contributing Authors: Alina L. Barbu (Meteo-France)
David Fairbairn (Meteo-France)
Emiliano Gelati (Meteo-France)
Helga Toth (OMSZ)
Souhail Boussetta (ECMWF)

Project co-funded by the European Commission within the Seventh Framework Program (2007-2013)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

DOCUMENT RELEASE SHEET

Book Captain:	J.C. Calvet	Date: 19.05.2016	Sign. 
Approval:	R. Lacaze	Date: 23.05.2016	Sign. 
Endorsement:	M. Koleva	Date:	Sign.
Distribution:	Public		

CHANGE RECORD

Issue/Revision	Date	Page(s)	Description of Change	Release
	28.02.2014	All	First Draft	D1.00
D1.00	09.09.2014	14	More detailed description of the generation of CO2 fluxes	I1.00
I1.00	21.11.2014	15-34 54	Add description of Hungarian SURFEX LDAS and of ECMWF global LDAS	I1.10
I1.10	16.07.2015	38-70	Description of the SURFEX LDAS-Monde, EnKF, assimilation of FAPAR, python chain, and hydro-validation	I2.00
I2.00	19.05.2016	61-62	CDF matching of ASCAT SSM data in LDAS-Hungary	I2.10

TABLE OF CONTENTS

1.	<i>Background of the Document</i>	11
1.1.	Executive Summary	11
1.2.	Scope and Objectives	12
1.3.	Content of the Document	13
1.4.	Related Documents	13
1.4.1.	Inputs.....	13
1.4.2.	Output	13
2.	<i>ECMWF Global LDAS</i>	15
2.1.	Theoretical Framework	15
2.1.1.	Overview.....	15
2.1.2.	The land surface model (CTESSEL).....	16
2.1.2.1.	The Canopy conductance formulation.....	16
2.1.2.1.1.	The Jarvis-Type approach	16
2.1.2.1.2.	The photosynthesis based approach (A-gs).....	17
2.1.2.1.2.1.	Soil moisture stress response	18
2.1.2.1.2.2.	CO ₂ concentration and water vapour deficit response	20
2.1.2.1.2.3.	Temperature responses	20
2.1.2.1.2.4.	Radiation response.....	21
2.1.2.1.2.5.	Vertical integration from leaf to canopy	22
2.1.2.2.	The vegetation growth model.....	24
2.1.2.3.	Soil respiration and ecosystem exchanges parameterization.....	25
2.2.	Global LDAS Architecture	26
2.2.1.	Inputs/Outputs	26
2.2.1.1.	2m temperature and relative humidity	27
2.2.1.2.	Soil and snow temperature.....	28
2.2.1.3.	Soil moisture	28
2.2.1.4.	Snow depth.....	28
2.2.1.5.	Analysis of albedo and LAI	29
2.2.1.5.1.	Derivation of climatological series.....	29
2.2.1.5.2.	Assimilation method.....	29
2.2.1.5.3.	The Copernicus Global Land products (GEOV1)	33
2.2.2.	Sequence of operations, data streams, interfaces	33
2.2.3.	Capacity requirements (computing time, timeliness, storage)	34
2.2.4.	Summary of the Global LDAS characteristics	34

3.	<i>SURFEX REGIONAL LDAS</i>	35
3.1.	Theoretical framework	35
3.1.1.	Basic underlying assumptions	35
3.1.2.	Land Surface Model.....	36
3.1.3.	Remote sensing data sets.....	37
3.1.4.	Land Data Assimilation System.....	39
3.1.4.1.	Generic algorithm	39
3.1.4.2.	SEKF.....	40
3.1.4.3.	EnKF	41
3.1.4.4.	Multi-patch representation.....	42
3.1.4.5.	State vector and resolution.....	44
3.1.4.6.	Rescaling and bias correction.....	45
3.1.4.7.	Background and observation errors.....	47
3.1.5.	Limitations of the method	48
3.2.	Regional LDAS architecture	49
3.2.1.	SEKF	49
3.2.2.	EnKF.....	50
3.2.3.	Inputs.....	51
3.2.3.1.	Physiographic fields:	51
3.2.3.2.	Meteorological inputs:.....	51
3.2.4.	Baseline processing	52
3.2.4.1.	Remote sensing observations for the assimilation	52
3.2.4.2.	The EKF namelist	53
3.2.4.3.	Sequence of operations, Data streams, Interfaces	53
3.2.5.	Output	55
3.2.6.	Python LDAS chain.....	55
3.2.7.	Hydro-validation over France	59
3.2.8.	Summary of the Regional LDAS-France characteristics	61
3.2.9.	Regional LDAS in Hungary	61
3.3.	Conclusions and prospects	63
4.	<i>SURFEX LDAS-Monde</i>	64
5.	<i>References</i>	65

LIST OF FIGURES

Figure 1: Flowchart of the albedo and LAI pre-processing, climatology derivation and assimilation....	31
Figure 2: Differences between the analysis and the observed broadband diffuse albedo for: a) 25 January 2006, b) 25 April 2006, c) 25 July2006, d) 25 October 2006, and Analysis broad band diffuse albedo for: e) 25 January 2006, f) 25 April 2006, g) 25 July2006, h) 25 October 2006.....	32
Figure 3: Flow chart of the land data assimilation algorithm	43
Figure 4: The multi-patch data assimilation design for a grid cell split into three patches, each having their variables x_p . The predicted observations y_p are weighted with their fraction denoted by α_p , where $p = 1, 2, 3$. y^o represents the grid-scale observation. The observation operator H aggregates the predicted observations at the grid scale. The data information is split among patches via the Kalman gain K	44
Figure 5: Surface soil moisture evolutions for 2009 at two location in North East (top) and in South West (bottom) of France for model (blue), ASCAT CDF rescaled (red) and ASCAT seasonal CDF rescaled (green) observations. SWI-001 observations ranging between 0 and 1 are depicted by black stars.	46
Figure 6: General code structure of the Python LDAS chain	56
Figure 7: Surface soil moisture evolutions for 2008 at Hegyhatsal (West-Hungary) for model (red), raw ASCAT satellite observations (purple) and ASCAT seasonal CDF rescaled (blue) observations.	62

LIST OF TABLES

<i>Table 1: List of land surface variables in CTESSEL (in bold italics those currently included in the land surface analysis). * developed within ImagineS</i>	<i>27</i>
<i>Table 2: Summary of ECMWF global LDAS characteristics</i>	<i>34</i>
<i>Table 3: Summary of SURFEX regional LDAS characteristics</i>	<i>61</i>
<i>Table 4: Summary of LDAS-Monde characteristics</i>	<i>64</i>

ACRONYMS

AMSR-E (Advanced Microwave Scanning Radiometer)
ASCAT (Advanced Scatterometer)
BDGSF (Base de Données Géographique des Sols de France)
CDF (Cumulative Distribution Function)
CTESSEL (Carbon Tiled ECMWF Scheme for Surface Exchanges over Land)
CHTESSEL (Carbon and Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land)
CYCLOPES (Carbon cYcle and Change in Land Observational Products from an Ensemble of Satellites)
ECMWF (European Centre for Medium-range Weather Forecast)
ECOCLIMAP (ECO-CLIMatic MAP)
EKF (Extended Kalman filter)
EnKF (Ensemble Kalman filter)
EnSRF (Ensemble Square Root Filter)
EO (Earth Observations)
ET (evapo-transpiration)
FAPAR (fraction of absorbed photosynthetically active radiation)
GEO (Group on Earth Observations)
GEOGLAM (Global Agricultural Geo-Monitoring Initiative)
GEOV1 (Version 1 of LAI, FAPAR, FCover products)
GPP (Gross Primary Production)
HTESSEL (Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land)
IFS (Integrated Forecast System)
IPCC (Intergovernmental Panel on Climate Change)
IPMA (Instituto Português do Mar e da Atmosfera)
ISBA-A-gs (Interactions between Soil Biosphere Atmosphere)
LAI (Leaf Area Index)
LAR (Leaf Area Ratio)
LDAS (Land data assimilation system)
LSM (Land Surface Model)
LST (Land surface temperature)

METOP (Meteorological Operational satellite)
MODCOU (MODélisation COUplée nappe surface)
MODIS (MODerate Imaging Spectrometer)
NEE (net ecosystem CO₂ exchange)
NWP (Numerical Weather Prediction)
OI (Optimal Interpolation)
PROBA-V
RECO (Ecosystem respiration)
SAFRAN (Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige)
SEKF (Simplified Extended Kalman Filter)
SIM (SAFRAN-ISBA-MODCOU)
SPOT (Satellite Probatoire d'Observation de la Terre)
SSM (Surface soil moisture)
SURFEX (Externalized surface)
SWI (Soil Wetness Index)
TU Wien (Technische Universität Wien)
VEGETATION (The medium resolution sensor onboard SPOT4 and SPOT5)
WG2 (root-zone soil moisture)

1. BACKGROUND OF THE DOCUMENT

1.1. EXECUTIVE SUMMARY

The Copernicus program is the EU response to the increasing demand for reliable environmental data. The objective of the Copernicus Land Service is to continuously monitor and forecast the status of land territories and to supply reliable geo-information to decision makers, businesses and citizens to define environmental policies and take right actions. ImagineS intends to continue the innovation and development activities to support the operations of the Copernicus Global Land service, preparing the use of the new Earth Observation data, including Sentinels missions data, in an operational context. Moreover, ImagineS aims to favor the emergence of downstream activities dedicated to the monitoring of crop and fodder production, that are key for the implementation of the EU Common Agricultural Policy, of the food security policy, and could contribute to the Global Agricultural Geo-Monitoring Initiative (GEOGLAM) coordinated by the intergovernmental Group on Earth Observations (GEO).

The main objectives of IMAGINES are to (i) improve the retrieval of basic biophysical variables, mainly LAI, FAPAR and the surface albedo, identified as Terrestrial Essential Climate Variables, by merging the information coming from different sensors (PROBA-V and Landsat-8) in view to prepare the use of Sentinel missions data; (ii) develop qualified software able to process multi-sensor data at the global scale on a fully automatic basis; (iii) complement and contribute to the existing or future agricultural services by providing new data streams relying upon an original method to assess the above-ground biomass, based on the assimilation of satellite products in a Land Data Assimilation System (LDAS) in order to monitor the crop/fodder biomass production together with the carbon and water fluxes; (iv) demonstrate the added value of this contribution for a community of users acting at global, European, national, and regional scales.

Two LDAS platforms are considered in ImagineS. The first one has a global coverage and is based on the ECMWF CTESSEL Land Surface Model (LSM) (Balsamo et al. 2009, Balsamo et al. 2011, Boussetta et al. 2013a, Boussetta et al. 2013b). Its carbon module component is developed with the active research support of ImagineS partners and in particular Meteo-France from which the A-gs model code originated. The CTESSEL model is implemented in the Integrated Forecast System (IFS) and is operational at ECMWF. This implies that the entire suites of forecasting products can output Net Ecosystem Exchange of CO₂ (NEE), Gross Primary Production (GPP) and Ecosystem Respiration (Reco). For the purpose of the ImagineS project, CTESSEL was upgraded in research mode to consider interactive vegetation and biomass output.

The second one is a LDAS developed in the SURFEX modeling platform, which includes the ability of performing regional experiments. It permits the joint assimilation of remotely sensed Surface Soil Moisture (SSM) derived from ASCAT backscatter data and the GEOV1 satellite-based LAI into the ISBA-A-gs land surface model. The ASCAT data are bias corrected with respect to the model climatology by using a seasonal-based CDF (Cumulative Distribution Function) matching technique. A multivariate multi-scale LDAS based on the Extended Kalman Filter (EKF) technique is used for monitoring soil moisture, vegetation, and terrestrial surface carbon and energy fluxes across France and Hungary at a spatial resolution of 8 km. Each model grid box is divided in a number of land covers, each having its own set of prognostic variables. The filter algorithm is designed to provide a distinct analysis for each land cover while using one observation per grid box. The updated values are aggregated by computing a weighted average.

1.2. SCOPE AND OBJECTIVES

The possibility of improving the performance of land surface models (LSMs) using remotely sensed observations is a field of active research. The mechanism of integrating observations, in a statistically optimal way, into a numerical model is called “data assimilation”. The latter permits improving the representation of the dynamical behavior of a bio-geophysical system. Land data assimilation systems (LDAS) are needed to integrate satellite data providing information about land state variables such as the surface soil moisture (SSM) and leaf area index (LAI) into LSMs.

Soil moisture is a key factor controlling both the water and energy cycles (through its impact on the fluxes partitioning at the surface). In addition, it is linked to the carbon cycle through the coupling between plant transpiration and photosynthesis. A number of studies have discussed the importance of soil moisture in the description of the carbon cycle whose connexions with the hydrological cycle are largely unknown (van der Molen et al., 2011). Assimilating remotely sensed SSM data into a LSM has proved, in a large number of papers, to be effective in estimating deeper soil moisture in various contexts, such as hydrology (Houser et al., 1998; Reichle et al., 2002a; Draper et al., 2011), numerical weather prediction (NWP) (Mahfouf et al., 2010; Dharssi et al., 2011; De Rosnay et al., 2013) and agricultural studies (Bolten and Crow, 2012).

Also, LAI impacts the exchanges of water vapor and CO₂ between the vegetation canopy and the atmosphere. A number of studies (Jarlan et al., 2008), Gu et al., 2006, Demarty et al., 2007) have shown the potential of assimilating LAI observations to correct vegetation model states.

Recognizing the importance of better exploiting the close link between soil moisture and vegetation variables, efforts were made to implement data assimilation schemes into complex models such as coupled hydrological and crop models or physiologically-based LSM. The possibility of combining these two data streams within such models has been explored in several data assimilation applications either by setting observing system simulation experiments (Pauwels et al., 2007; Nearing et al., 2012) or by monitoring real environments (Sabater et al., 2008; Barbu et al., 2011).

Barbu et al. (2014) have shown that the LDAS is able to: (1) simultaneously ingest EO satellite data providing mixed signals at a grid-scale into the mosaic structure of the ISBA-A-gs LSM; (2) propagate information from the surface into the root-zone soil layer; (3) consistently impact the water and carbon fluxes; (4) improve the short-term vegetation response to drought conditions.

This document provides a description of how the global and regional LDAS are implemented and operated and how the products are obtained.

1.3. CONTENT OF THE DOCUMENT

Chapter 2 presents the ECMWF global LDAS including a description of the underlying LSM and Chapter 3 describes the SURFEX LDAS, its theoretical framework, its configuration and a description of its underlying LSM.

1.4. RELATED DOCUMENTS

1.4.1. Inputs

Overview of deliverables acting as inputs to this document.

Document ID	Descriptor
ImagineS_RP1.1	Users Requirements Document
ImagineS_RP1.2	Service Specifications Document

1.4.2. Output

Overview of other deliverables for which this document is an input:

Document ID	Descriptor
ImagineS_RP6.3	Product User Manual of LDAS output products

2. ECMWF GLOBAL LDAS

Within the Copernicus Global Land service, coordinated efforts are made to produce biophysical variables that describe the continental vegetation state, radiation budget and water cycle with the objective of developing and validating operationally oriented land information services. In particular, satellite-derived products of soil moisture (Soil Water Index), Leaf Area Index (LAI) and albedo are being produced. Including this new information in a Global Land Data Assimilation System (LDAS) and assessing its impact contributes to a better characterization of the vegetation state, permits the monitoring of the surface fluxes (carbon and water) and the associated root-zone soil moisture at the global scale (spatial resolution of 16km x 16km).

The global LDAS is developed within the ECMWF system up to pre-operational phase. The CTESSEL model is operationally implemented in the Integrated Forecast System (IFS). This implies that the entire suites of forecasting products can output Net Ecosystem Exchange of CO₂ (NEE), Gross Primary Production (GPP) and Ecosystem Respiration (Reco).

Apart from this online real-time production chain, an offline LDAS chain is able to assimilate satellite-derived LAI and albedo products and can be attached to the ECMWF reanalysis depending on favourable assessment.

2.1. THEORETICAL FRAMEWORK

2.1.1. Overview

The product algorithm is based on a full land-atmosphere model described in peer-reviewed articles and in a technical documentation. The land surface model CTESSEL is implemented operationally at ECMWF. This scheme is integral part of the IFS (<http://old.ecmwf.int/research/ifsdocs/CY40r1/index.html>). The LDAS for the global component at ECMWF is based on Drusch et al. (2009), de Rosnay et al. (2011), for the land surface soil moisture, on a recently revised version of Drusch et al. (2004) and De Rosnay et al. (2014) for the snow depth, and on Mahfouf (1999) for the soil temperature. The LAI and albedo assimilation are based on Gu et al. (2006) and Boussetta et al. (2014). The system is composed of modelling and data assimilation components, described in the following sections.

2.1.2. The land surface model (CTESSEL)

The CTESSEL model results from the coupling between the HTESSEL model (Balsamo et al. 2009) and the A-gs photosynthesis model. This model is a modified version of the Jacobs's model (Jacobs et al., 1996) taking into account the effects of soil water stress on the photosynthesis and canopy resistance estimation (Calvet et al. 1998, Calvet, 2000, Calvet et al. 2004).

The A-gs model is a semi-empirical physiological model linking the leaf photosynthesis rate, the net assimilation of CO₂ and the leaf conductance to external surface and atmospheric factors (soil moisture, incoming radiation ...).

The photosynthesis module within the A-gs model is coupled with an ecosystem respiration module based on soil moisture and temperature dependency, which is modified to take into account the cold region and snow pack effect on soil respiration emission.

The stomatal behaviour is the main unit driver of the fluxes exchanges between the plant and its surrounding environment. It is therefore directly affected by the plants condition and the environmental factors. The stomatal conductance is related to photosynthesis, through the net CO₂ assimilation (A_n) by the canopy. A_n is inhibited by different environmental factors interacting in a synergistic way. The CO₂ assimilation responds to leaf temperature, to the soil moisture stress, to solar radiation, and is limited by the air CO₂ concentration and water vapour availability.

In order to be used globally within Numerical Weather Prediction (NWP) and climate models, the soil respiration part of the model is parameterized in a numerically simple way, avoiding long spin-up runs that are necessary when carbon pool parameterisations are used. In CTESSEL, the CO₂ ecosystem respiration is split into two terms. The first is the dark respiration parameterised as a fraction of the gross assimilation. The second represents both heterotrophic respiration from the soil and autotrophic respiration from the above and below ground structural biomass. It is parameterized as a function of soil temperature, soil moisture, snow depth and vegetation.

2.1.2.1. *The Canopy conductance formulation*

Two main approaches are commonly used to formulate the canopy conductance:

2.1.2.1.1. *The Jarvis-Type approach*

Given its simple formulation, the Jarvis-type approach is the most used one in land surface models which are adopted by NWP systems. It is based on empirical stress functions related to the environment conditions and combined in multiplicative way meant to limit the

vegetation-dependant maximum value of the stomatal conductance. The main hypothesis behind this formulation is that these stress functions are assumed to be independent from each other.

Within the HTESSEL, the following formulation of the canopy conductance g_s is adopted:

$$g_s = g_{s,\max} LAI [f_1(R_s) f_2(\bar{\theta}) f_3(D_a)] \quad (2.1)$$

with $g_{s,\max}$ being the vegetation type-dependant maximum stomatal conductance and f_1, f_2, f_3 , three inhibition functions expressing respectively shortwave radiation deficit, soil moisture stress and atmospheric humidity deficit:

$$f_1(R_s) = \min \left[1, \frac{bR_s + c}{a(bR_s + 1)} \right] \quad (2.2)$$

where R_s is the downward short-wave radiation and a, b, c are empirical constants,

$$f_2(\bar{\theta}) = \begin{cases} 0 & , \quad \bar{\theta} < \theta_{pwp} \\ \frac{\bar{\theta} - \theta_{pwp}}{\theta_{cap} - \theta_{pwp}} & , \quad \theta_{pwp} \leq \bar{\theta} \leq \theta_{cap} \\ 0 & , \quad \bar{\theta} > \theta_{cap} \end{cases} \quad (2.3)$$

where θ_{pwp} and θ_{cap} are the soil moisture at permanent wilting point and at field capacity respectively and $\bar{\theta}$ is a root distribution-based weighted average over the different soil layer of the unfrozen soil water.

$$f_3(D_a) = e^{-g_D D_a} \quad (2.4)$$

where D_a is the atmospheric humidity deficit, and g_D is a vegetation type dependant coefficient.

2.1.2.1.2. The photosynthesis based approach (A-gs)

In this approach, the canopy conductance is calculated from photosynthesis, which is the net CO₂ assimilation A_n by the canopy. Similarly to the evaporation and using the Kirchoff

analogy, the carbon dioxide flux which results from the difference between the gross assimilation A_g and the dark respiration R_d is given by:

$$A_n = g_{sc}(C_s - C_i) = A_g - R_d \quad (2.5)$$

where g_{sc} is the canopy conductance to carbon dioxide, C_s is the CO₂ concentration at the leaf surface and C_i is the intercellular CO₂ concentration.

A_n is calculated as a function of different environmental factors having different responses and mutually interacting.

2.1.2.1.2.1. Soil moisture stress response

Unlike other A-gs formulation for which the soil moisture stress response is directly applied to the gross assimilation A_m (Ronda et al., 2001) or the net assimilation A_n (Sala and Tenhunen, 1996), Calvet (2000) found that the soil moisture stress response is driven in a more complex way through the coupled space of the mesophyll conductance g_m and the maximum specific humidity deficit tolerated by the vegetation D_{max} , and behave differently between high and low vegetation. In CTESSEL the adopted soil moisture stress response is the drought tolerant strategy described in Calvet et al. (2000, 2004) and based on a meta-analysis of several herbaceous and woody vegetation types. The above-mentioned meta-analysis also showed that under unstressed condition, g_m and D_{max} are well correlated for low vegetation:

$$\ln(g_m^*) = a - b \ln(D_{max}^*) \quad (2.6)$$

while for high vegetation g_m is well correlated with a vegetation dependant coupling factor f_0 :

$$\ln(g_m^*) = c - d f_0^* \quad (2.7)$$

where, X^* denote the value of X under unstressed condition, and a , b , c and d are derived constants from the meta-analysis.

Considering the soil moisture stress function $f_2(\bar{\theta})$ (similarly formulated as in the Jarvis-type approach), a critical value f_{2c} is defined to differentiate between moderate ($f_2 \geq f_{2c}$) and severe ($f_2 < f_{2c}$) stress situations.

a) For low vegetation:

Under moderate conditions ($f_2 \geq f_{2c}$):

$$D_{\max} = D_{\max}^X + (D_{\max}^* - D_{\max}^X) \frac{f_2 - f_{2c}}{1 - f_{2c}} \quad (2.8)$$

Under severe stress conditions ($f_2 < f_{2c}$):

$$D_{\max} = D_{\max}^X \frac{f_2}{f_{2c}} \quad (2.9)$$

where D_{\max}^X is the maximum value of D_{\max} corresponding to f_{2c}

b) For high vegetation:

Under moderate conditions ($f_2 \geq f_{2c}$):

$$g_m = g_m^N + (g_m^* - g_m^N) \frac{f_2 - f_{2c}}{1 - f_{2c}} \quad (2.10)$$

where g_m^N is the stressed value of g_m corresponding to the constant unstressed coupling factor f_0^* .

Under severe stress conditions ($f_2 < f_{2c}$):

$$g_m = g_m^N \frac{f_2}{f_{2c}} \quad (2.11)$$

Further details on the soil stress parameterization can be found in Calvet et al. (2000, 2004) and Voogt et al. (2007).

2.1.2.1.2.2. CO₂ concentration and water vapour deficit response

Further, the photosynthesis assimilation in light-saturating conditions is also limited by the air CO₂ concentration and determined via a saturation equation:

$$A_m = A_{m,\max} \left(1 - e^{(-g_m (C_i - \Gamma)/A_{m,\max})}\right) \quad (2.12)$$

where $A_{m,\max}$ is the maximum net CO₂ assimilation and Γ is the CO₂ concentration at which assimilation compensates respiration, called CO₂ compensation point. The internal CO₂ concentration C_i , is directly derived from a closure equation linking the internal and external concentration C_s to the water vapour deficit through:

$$f = \frac{C_i - \Gamma}{C_s - \Gamma} = f_0 \left(1 - \frac{D_s}{D_{\max}}\right) + \left(\frac{g_c}{g_c + g_m}\right) \frac{D_s}{D_{\max}} \quad (2.13)$$

where the coupling factor f is sensitive to air humidity and depends on the cuticular conductance g_c , the mesophyll conductance g_m , the maximum deficit tolerated by the vegetation D_{\max} and f_0 (the value of f when the specific humidity deficit at the leaf surface $D_s = 0$)

C_i is therefore controlled by the air humidity via D_s , if the deficit exceeds the maximum tolerated by the vegetation (D_{\max}), the plant closes its stomata.

2.1.2.1.2.3. Temperature responses

The temperature dependency of the CO₂ flux is expressed through Q₁₀-type functions applied to the CO₂ compensation point Γ , the mesophyll conductance g_m and the maximum assimilation $A_{m,\max}$. Q₁₀ is defined as the proportional increase of a parameter value for a 10 degree increase in temperature (Berry and Raison, 1982). For the compensation point, it is formulated as:

$$\Gamma(T_s) = \Gamma(25^\circ) Q_{10}^{(T_s - 25)/10} \quad (2.14)$$

For the mesophyll conductance and the maximum assimilation, it is adjusted by the inhibition functions after Collatz et al. (1992) as follows:

$$g_m(T_s) = \frac{g_m(25^\circ)Q_{10}^{(T_s-25)/10}}{(1 + e^{0.3(T_1-T_s)})(1 + e^{0.3(T_s-T_2)})} \quad (2.15)$$

$$A_{m,\max}(T_s) = \frac{A_{m,\max}(25^\circ)Q_{10}^{(T_s-25)/10}}{(1 + e^{0.3(T_1-T_s)})(1 + e^{0.3(T_s-T_2)})} \quad (2.16)$$

where Q_{10} , T_1 and T_2 are different for each parameter and modulate its sensitivity to the plant surface temperature T_s .

2.1.2.1.2.4. Radiation response

The CO_2 assimilation limited by CO_2 concentration and temperature is further limited by radiation through:

$$A_n = (A_m + R_d) \left[1 - e^{\frac{-\varepsilon I_a}{(A_m + R_d)}} \right] - R_d \quad (2.17)$$

where I_a is the photosynthetic active radiation (PAR), R_d is the dark respiration which is simply parameterised as a fraction of A_m . And ε is the initial quantum use efficiency parameterised as:

$$\varepsilon = \varepsilon_0 \frac{C_i - \Gamma}{C_i + 2\Gamma} \quad (2.18)$$

where ε_0 is the maximum quantum use efficiency

The stomatal conductance to CO_2 , g_{sc} , is estimated using the flux-gradient relationship, modified to account for the effect of specific humidity deficit on stomatal aperture given by:

$$g_{sc} = \frac{A_n - A_{\min} \left(\frac{D_s}{D_{\max}} \frac{A_n + R_d}{A_m + R_d} \right) + R_d \left(1 - \frac{A_n + R_d}{A_m + R_d} \right)}{C_s - C_i} \quad (2.19)$$

where A_{\min} represents the residual photosynthesis rate (at full light intensity) associated with cuticular transfers when the stomata are closed because of a high specific humidity deficit.

$$A_{\min} = g_m (C_{\min} - \Gamma) \quad (2.20)$$

where C_{\min} is the maximum value of C_i at maximum specific humidity deficit:

$$C_i = \frac{g_c C_s + g_m \Gamma}{g_c + g_m} \quad (2.21)$$

The diffusion of CO_2 interacts with that of water vapour and the stomatal conductance to CO_2 is corrected for this interaction and iteratively refined by:

$$g_{sc} = g_{sc} + E \frac{M_a}{\rho_a M_v} \frac{C_s + C_i}{2(C_s - C_i)} \quad (2.22)$$

where M_v and M_a are molecular masses of water vapour and air respectively, ρ_a is the air density and E is the leaf transpiration based on the previous guess of the stomatal conductance:

$$E = (1.6g_{sc})D_s\rho_a \quad (2.23)$$

Finally, the stomatal conductance to water vapour g_s is given by:

$$g_s = 1.6g_{sc} + g_c \quad (2.24)$$

where g_c is the cuticular conductance (a vegetation dependent parameter).

2.1.2.1.2.5. Vertical integration from leaf to canopy

The net CO_2 assimilation calculated at the leaf scale is upscaled at the canopy scale assuming that leaf parameters do not vary within the canopy, and the attenuation of the incoming shortwave radiation in the canopy is computed thanks to a simple radiative transfer model.

The incoming shortwave radiation is attenuated within the canopy. At the top of the canopy, the incoming PAR is assumed to be 48% of the incoming shortwave radiation. The PAR extinction is described by Roujean (1996). The PAR at height z in the canopy is given by:

$$I_a(z) = I_a(h)(1 - K(z)) \quad (2.25)$$

where h is the height of the top of the canopy and K is the extinction function given by:

$$K(z) = f(\mu_s)K_{df}(z) + (1 - f(\mu_s))K_{dr}(z) \quad (2.26)$$

where $K_{df}(z)$ and $K_{dr}(z)$ are the extinction coefficients of diffuse and direct light, respectively:

$$K_{df}(z) = 1 - e^{\left(\frac{-0.8bLAI(h-z)}{h}\right)} \quad (2.27)$$

$$K_{dr}(z) = 1 - e^{\left(\frac{G}{\cos(\mu_s)} \frac{bLAI(h-z)}{h}\right)} \quad (2.28)$$

where μ_s is the solar zenith angle and G is a parameter that describes the distribution of leaves (a spherical angular distribution is assumed: $G=0.5$). f is the ratio of diffuse to total downward shortwave radiation at the top of the canopy given by:

$$f(\mu_s) = \frac{0.25}{0.25 + \cos(\mu_s)} \quad (2.29)$$

b is the foliage scattering coefficient given by

$$b = 1 - \frac{1 - \sqrt{1 - \omega}}{1 + \sqrt{1 - \omega}} \quad (2.30)$$

based on the leaf single scattering albedo ω ($=0.2$) for the solar spectrum corresponding to the PAR.

Assuming an homogeneous leaf vertical distribution, the integrated canopy net CO_2 assimilation and conductance can be written as:

$$A_{nI} = LAI \int_0^1 A_n d(z/h) \quad (2.31)$$

$$g_{sI} = LAI \int_0^h g_s d(z/h) \quad (2.32)$$

The integrations are parameterized with a three-point Gauss quadrature method following Jacobs (1996).

2.1.2.2. *The vegetation growth model*

Owing to the photosynthetic activity and based on biomass evolution, CTESSEL is also able to generate its own interactive LAI. This version of the model is the starting point for a near real time assimilation of LAI satellite data. The biomass module simulates growth and mortality of the vegetation. The growth of active biomass B is based on the accumulated net CO_2 assimilation over the previous day $A_{n,day}$ and the LAI is obtained from the biomass following:

$$\alpha_B = \frac{B}{LAI} \quad (2.33)$$

In reality, α_B depends on climate (temperature and CO_2 concentration) and nitrogen fertilisation. In order to account for plant morphology, the nitrogen dilution concept is applied for the biomass evolution. The plant N decline model is a well-established agronomical law relating the plant N in non-limiting N-supply conditions to the accumulated above-ground dry matter. The critical plant N is the value of N maximizing growth, and this value decreases for increasing biomass accumulation following a negative power law. The basis of the model is that the metabolic component B of the plant biomass is related to total biomass B_T through an allometric logarithmic law (Calvet and Soussana, 2001). In CTESSEL, the metabolic biomass component is identified as the active biomass and the relationship between active biomass B and total aboveground biomass B_T is:

$$B_T = \left(\frac{B}{c} \right)^{\frac{1}{1-a}} \quad (2.34)$$

where a and c are constant parameters. The total aboveground biomass consists of the active biomass reservoir and the structural aboveground reservoir B_s , which can be considered as the "living" structural biomass, like the stem. For forests, wood is a dead reservoir and does not contribute to B_s . Within the nitrogen dilution model a relationship between the leaf area ratio (LAR) and the aboveground nitrogen concentration N_T is applied:

$$LAR = \frac{LAI}{B_T} = eN_T + P \quad (2.35)$$

where e and P are called plasticity parameters and are derived per vegetation type. This formulation can be used as a closure equation to estimate α_B :

$$\alpha_B = \frac{1}{eN_a + P/cB_T^{-a}} \quad (2.36)$$

where N_a is the nitrogen concentration in the active biomass. It depends on vegetation type and on the nitrogen fertilisation. For further details and derivations see Calvet and Soussana (2001). In this way, α_b has become a model variable depending on B_T . However, for global simulations, it is desirable to keep α_b as a constant parameter in order to let α_B represent rather intrinsic plant characteristics denoting a biological adaptation to average climate and growing conditions (Calvet and Soussana, 2001). For that purpose, the above Eq.36 can only be solved by iteration. Moreover, LAR and N_T data to derive the plasticity parameters by regression is lacking. However, data is available for leaves in the form of the specific leaf area (SLA) and the nitrogen content in leaves N_L :

$$SLA = \frac{LAI}{B_L} = eN_L + P \quad (2.37)$$

Both the iteration issue and the availability of data to derive e and P give rise to modify the nitrogen dilution module. Eq.37 is simplified by considering α_B as the ratio of the biomass of green leaves to LAI:

$$\alpha_B = \frac{1}{SLA} = \frac{1}{eN_L + P} \quad (2.38)$$

2.1.2.3. Soil respiration and ecosystem exchanges parameterization

In order to be used within a NWP model, soil respiration needs to be parameterized in a numerically simple way, avoiding long spin-up run that are necessary when carbon pool parameterisation is used. In CTESSEL the CO_2 ecosystem respiration R_{eco} is split into two terms. The first is the dark respiration R_d (parameterized from A_m). The second, represents both heterotrophic respiration from the soil and autotrophic respiration from the above and below ground structural biomass and referred hereafter as $R_{soilstr}$, it is parameterized as a function of soil temperature, soil moisture, snow depth and vegetation type as:

$$R_{soilstr} = R_{eco} - R_d = R_0 Q_{10}^{\left(\frac{T_{soil}-25}{10}\right)} f_{sm} f_{sn} \quad (2.39)$$

where f_{sn} and f_{sm} are snow and soil moisture attenuation function respectively defined as:

$$f_{sn} = 1 - C_{vs} (1 - e^{-\alpha z_{snow}}) \quad (2.40)$$

C_{vs} is the surface fraction covered by snow, α is a constant expressing the attenuation of the soil carbon emission within the snow pack and z_{snow} is the snow depth.

The soil moisture stress function for soil respiration is defined following a study by Albergel et al. (2010) as:

$$f_{sm} = \frac{\bar{\theta}}{\theta_{cap}} \quad (2.41)$$

In this case, given its variability with climate regimes, Q_{10} is defined as a function of soil temperature after McGuire et al. (1992)

The vegetation types are affecting the ecosystem respiration through a reference respiration at 25°C (R_0). R_0 is estimated by minimizing the root mean square errors between simulated and observed R_{eco} for each vegetation (section 2.4).

Finally, the following equations describe the relation between the gross primary production GPP , the net ecosystem exchange NEE and the respiration components (R_d , $R_{soilstr}$, R_{eco}):

$$GPP = A_n + R_d \quad (2.42)$$

$$NEE = A_n - R_{soilstr} = GPP - R_{eco} \quad (2.43)$$

2.2. GLOBAL LDAS ARCHITECTURE

2.2.1. Inputs/Outputs

The input of the LDAS is represented by the screen-level parameters (2m temperature and relative humidity) as well as the ASCAT satellite-based soil moisture assimilated in a research version of the LDAS. The use of screen-level parameters is described in Mahfouf (1999) and de Rosnay et al. (2012). The model background for land surface analysis is provided by the CTESSEL land surface scheme, which currently accounts for 15 variables (Table 1).

Table 1: List of land surface variables in CTESSEL (in bold italics those currently included in the land surface analysis). * developed within ImagineS

Code	Name	Abbreviation	Unit
32	Snow albedo	ASN	(0 - 1)
33	Snow density	RSN	kg m ⁻³
39	<i>Volumetric soil water layer 1</i>	<i>SWVL1</i>	<i>m³ m⁻³</i>
40	<i>Volumetric soil water layer 2</i>	<i>SWVL2</i>	<i>m³ m⁻³</i>
41	<i>Volumetric soil water layer 3</i>	<i>SWVL3</i>	<i>m³ m⁻³</i>
42	Volumetric soil water layer 4	SWVL4	m ³ m ⁻³
139	<i>Soil temperature level 1</i>	<i>STL1</i>	<i>K</i>
141	<i>Snow depth</i>	<i>SD</i>	<i>m</i>
170	Soil temperature level 2	STL2	K
183	Soil temperature level 3	STL3	K
236	Soil temperature level 4	STL4	K
238	<i>Temperature of snow layer</i>	<i>TSN</i>	<i>K</i>
66	<i>Leaf Area Index, low vegetation*</i>	<i>LAI_LV</i>	<i>m² m⁻²</i>
67	<i>Leaf Area Index, high vegetation*</i>	<i>LAI_HV</i>	<i>m² m⁻²</i>
174	<i>Albedo*</i>	<i>AL</i>	<i>(0 - 1)</i>

At the moment, four different analysis schemes are active for the surface (and near-surface) variables based, respectively, on spatial Optimum Interpolation (2D-OI, used snow depth and screen-level analyses), the column Optimum Interpolation (1D-OI, used for soil/snow temperature analysis), a Simplified EKF (SEKF, used for soil moisture analysis) and the assimilation of the Leaf Area Index (LAI) and albedo is based on a simple 1D optimal interpolation method (Gu et al., 2006 and Boussetta et al., 2014) which is well adapted to the current global system. All schemes operate independently from the atmospheric analysis.

2.2.1.1. 2m temperature and relative humidity

The screen-level variables are analysed using a spatial 2D-OI algorithm and are used as input to the soil moisture and soil/snow temperature analyses.

The screen-level analyses are performed 4 times per day at synoptic hours (0, 6, 12, 18 UTC). The background field (issued from a 6-hour or 12-hour forecast) is horizontally interpolated to the observation locations using a bilinear interpolation scheme and background increments are calculated. Horizontal structure functions are calculated following a Gaussian formulation and using an e-folding distance of 300 km. The standard deviations of background and observation errors are set to 1.5 K and 2 K for temperature and 5% and 10% for relative humidity, respectively. Those 50 observations closest to a given grid point within a radius of 1000 km are considered for the analysis at that grid point. The analysis is performed over land and ocean but only land (ocean) observations are used for model land (ocean) grid points.

2.2.1.2. Soil and snow temperature

The soil and snow temperature are analysed using a 1D-OI that correlates 2m temperature short-term errors (forecast-analysis) with soil/snow temperature corrections. The corrections are mutually exclusive, i.e. in the presence of snow the soil temperature is not modified.

2.2.1.3. Soil moisture

In the current soil moisture analysis scheme, a SEKF is operated (de Rosnay et al. 2011, Drusch et al., 2009) with the following characteristics:

- the SEKF is run at T1279 resolution over a 12-hour window as is the atmospheric 4D-Var analysis;
- the background errors are kept static and are not evolved, as in the Kalman Filter;
- soil moisture in the top three layers (0-7, 7-28, 28-100 cm) are perturbed individually; atmospheric parameter perturbations are performed globally at the same time (i.e. assuming no correlation) and three atmospheric model integrations are performed; Jacobians are calculated from finite differences;
- the resulting surface analysis is used to initialize the next short forecast which then provides surface conditions to the atmospheric analysis window 12 hours later.

2.2.1.4. Snow depth

The snow depth analysis uses a spatial 2D-OI as well since November 2010. This analysis ingests SYNOP snow-depth reports and IMS satellite-based snow cover data (at 4 km resolution). The structure functions differ from the screen-level analysis and are parameterized following Brasnett (1999). Despite providing a significant improvement of the previous analysis scheme the use of satellite data is still not ideal (since adopted as “ground-

truth” to modify the model background). Ongoing work investigates the use of a SEKF in order to assimilate satellite-based products.

2.2.1.5. Analysis of albedo and LAI

A one dimensional data assimilation method is used to produce analysis of albedo and LAI from the GEOVx data. Following the method of Gu et al. (2006) and Boussetta et al., (2014), the observation and the climatological values are combined with their respective error estimates to produce the analysis product.

2.2.1.5.1. Derivation of climatological series

Before processing the climatological product, unreliable retrievals have been discarded from the analysis thanks to the quality flag (QA). In particular: dead detectors, significant clouds and/or snow contaminated pixels, and failure of the radiative transfer model due to problems other than geometry have been filtered out during this pre-processing. Afterwards, the 1-km products were aggregated to 10 km resolution through a nine-by-nine-point spatial smoothing. The 10 km value is computed when more than 30% of the 1-km products at the grid point scale have not been flagged and a further snow-free screening is performed on the data. Then a first version of the climatological time series is obtained by averaging data from 1999 to 2012 (ALB_cv1/LAI_cv1).

These first versions climatological time-series still contain gaps especially in snow-covered high latitude regions. To overcome this deficiency, a second version of the climatological time series (ALB_cv2/LAI_cv2) is generated by spatially filling the data gaps with values from 36 10-daily “self-derived” look-up tables of LAI and albedos for each vegetation type. The look-up tables are derived through stratification of ALB_cv1/LAI_cv1 by vegetation type based on a 90% vegetation cover threshold for each type. In CHTESSEL, the land use classification follows from the Global Land Cover Characteristics (GLCC) data (Loveland et al., 2000) and use is made of the Biosphere-Atmosphere Transfer Scheme (BATS) classification to assign the vegetation types.

Finally, a three-point temporal smoothing is applied to this second version climatological data to obtain a final version of the climatological time series (ALB_c/LAI_c). These data are then re-projected and interpolated to a target model simulation grid, together with their associated error σ_c in order to be used in the ECMWF model.

2.2.1.5.2. Assimilation method

The main objective of data assimilation is to optimize the use of observational data in order to get the best estimate from all available information. It usually attempts to combine data from different sources in an optimal way to provide the best estimate also called

analysis product. In this system, a one dimensional data assimilation method is used to produce analysis of albedo and LAI from the GEOVx data. Two pieces of information are used to generate the analysis product at a given time t : the observation V_o with its associated error σ_o and the climatological values V_c with their associated errors σ_c , (V being the LAI or the albedo data). The optimal combination of these two pieces of information is obtained by minimizing a quadratic cost function J which corresponds to the least-square estimate V_a :

$$J(V_a) = \left[\frac{V_a - V_o}{\sigma_o} \right]^2 + \left[\frac{V_a - V_c}{\sigma_c} \right]^2 \quad (2.44)$$

Assuming the associated errors are Gaussian, the minimisation of J would lead to the BLUE (best linear unbiased estimate) ensuring a minimum variance for the analysis error, which leads to:

$$V_a = \left[\frac{\sigma_c^2}{\sigma_o^2 + \sigma_c^2} \right] V_o + \left[\frac{\sigma_o^2}{\sigma_o^2 + \sigma_c^2} \right] V_c \quad (2.45)$$

To avoid rough replacement with the climatological data in the case of missing observation and relax non-realistic jumps, a further processing is performed by adding the climatological tendency to the last analysis value assuming the analysis tendency is similar to the climatological one:

$$\frac{V_a(t) - V_a(t-1)}{\Delta t} = \frac{V_c(t) - V_c(t-1)}{\Delta t} \quad (2.46)$$

To summarize, a flowchart of the data processing and analysis procedure is illustrated in Figure 1.

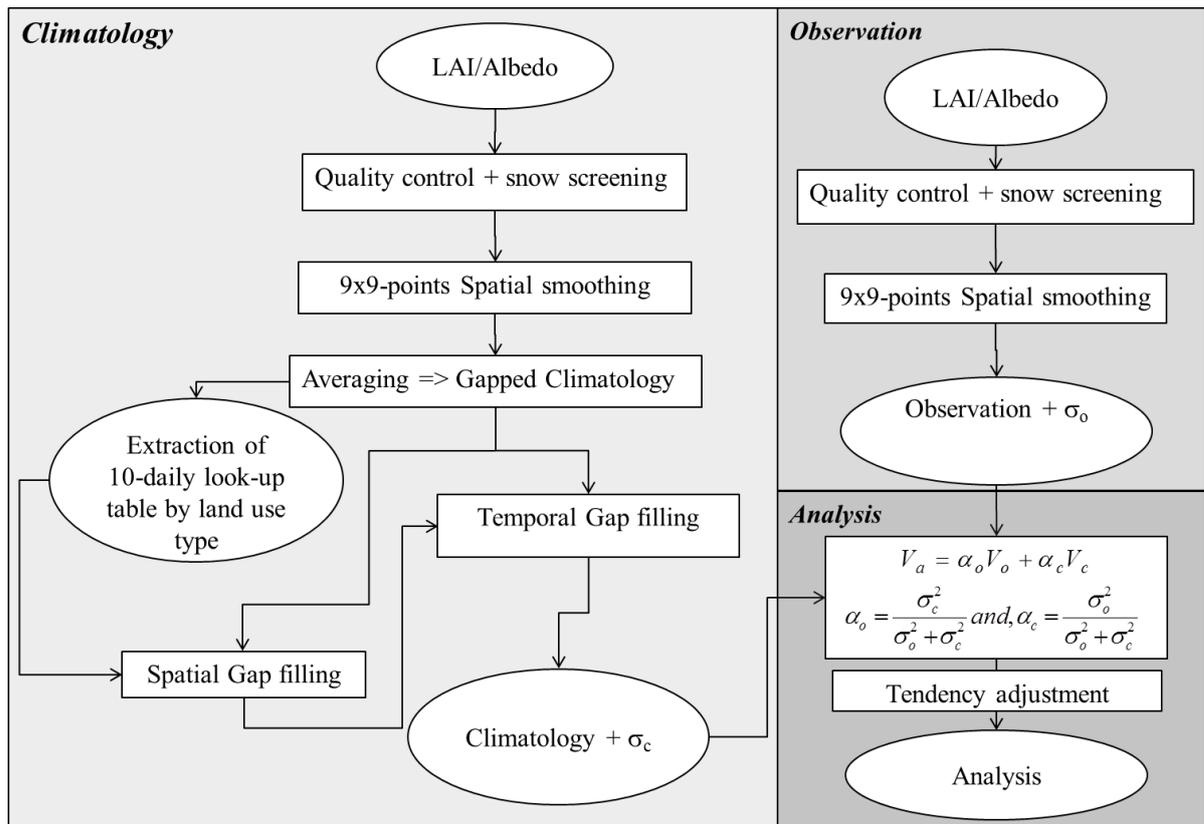


Figure 1: Flowchart of the albedo and LAI pre-processing, climatology derivation and assimilation

The LAI and albedo analysis system is implemented in the framework of the ImagineS project and is initially tested with the available GEOV1 Copernicus Global Land product. The system is implemented in an inter-operable way to be able to ingest forthcoming LAI and albedo products. Figure 2 shows an example of the system outputs for the albedo. It illustrates the analysis of snow-free broadband diffuse albedo and its difference from the observed data for different dates representing the four seasons of year 2006 (January 25th, April 25th, July 25th and October 25th). The PROBA-V demo products are being explored and will be tested in the system when fully stable within the processing chain.

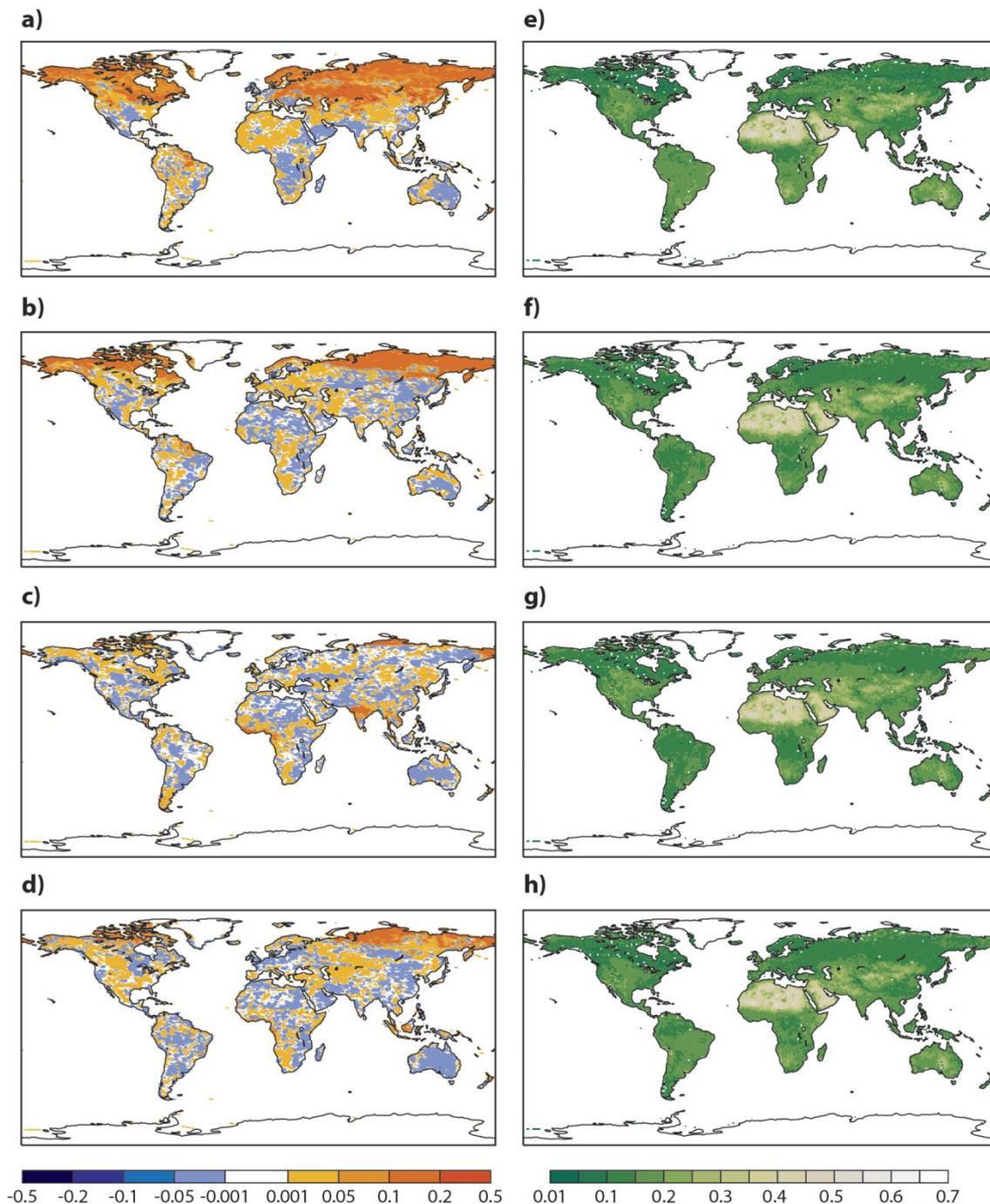


Figure 2: Differences between the analysis and the observed broadband diffuse albedo for: a) 25 January 2006, b) 25 April 2006, c) 25 July 2006, d) 25 October 2006, and Analysis broadband diffuse albedo for: e) 25 January 2006, f) 25 April 2006, g) 25 July 2006, h) 25 October 2006.

2.2.1.5.3. *The Copernicus Global Land products (GEOV1)*

The GEOV1 LAI and albedo products are based on observations from the VEGETATION sensor on board SPOT satellite. They are produced every 10 days using a composite observation from a 30 days moving window at $1/112^\circ$ spatial resolution (about 1 km at the Equator) with a global coverage. Each GEOV1 product is provided with its associated error measure σ_o . To take advantage of previous algorithmic experience and existing LAI products, the “best-performing” LAI data (Garrigues et al., 2008) were combined and then used to train a neural network system. The combined LAI data are the CYCLOPES-V3.1 (Baret et al., 2007) and the collection 5 of MODIS LAI (Myneni et al., 2002). This fusion allows benefiting from the good performance of MODIS LAI for high values and CYCLOPES-V3.1 LAI at low values. After being trained with the fused data, the neural network system is then fed with the atmospherically corrected reflectances in red, near-infrared, and shortwave-infrared bands from VEGETATION as well as the solar zenith angles and the satellite overpass timing which results into the GEOV1 LAI (Baret et al., 2013).

The GEOV1 surface albedo is also based on observations from the VEGETATION sensor and its derivation follows Geiger and Samain (2004). The method includes cloud screening (Hagolle et al., 2004), atmospheric correction (Rahman & Dedieu, 1994), directional reflectance normalization (Roujean et al., 1992), and albedo determination for the different integration angles (direct and diffuse) and different spectral intervals (visible, near-infrared, and broadband).

Under the Copernicus Global Land service, the GEOV1 products were validated by characterising their spatial and temporal continuity and consistency as well as their accuracy at the global and regional scales against other global products and the BELMANIP2 sites network (Benchmark Land Multi-site Analysis and Inter-comparison of Products). The conclusion of this validation was that the GEOV1 products are of good quality, show consistent temporal and spatial distributions and have reasonable accuracy which can meet the requirements for usage within LSMs (Camacho et al., 2013).

2.2.2. **Sequence of operations, data streams, interfaces**

The sequence of operation is documented in the IFS data assimilation manual (see Part II, Chapter 11) available on the web at <http://old.ecmwf.int/research/ifsdocs/CY40r1/ASSIMILATION/IFSPart2.pdf> and in Boussetta et al. (2014).

2.2.3. Capacity requirements (computing time, timeliness, storage)

The computing is performed on the ECMWF supercomputer (CRAY XC30), and the time sequence is in line with the operational production, as described on the ECMWF web site and on the data assimilation suites summarized in Haseler (2004, http://www.ecmwf.int/publications/library/ecpublications/_pdf/tm/401-500/tm454.pdf).

The Storage is managed by the Meteorological Archiving and Retrieval System (MARS) available at ECMWF and that serves the meteorological users and the Member-States.

2.2.4. Summary of the Global LDAS characteristics

Inputs	Screen-level parameters, ASCAT-Soil moisture, GEOVx LAI, GEOVx albedo
Outputs	Analyzed LAI, analyzed albedo, Root-zone soil moisture, NEE, GPP, Evapo-transpiration
Interfaces	IFS (Integrated Forecasting System)
Capacity requirements	CRAY XC30 Supercomputing facility, MARS storage system

Table 2: Summary of ECMWF global LDAS characteristics

3. SURFEX REGIONAL LDAS

3.1. THEORETICAL FRAMEWORK

3.1.1. Basic underlying assumptions

The LDAS is defined as an off-line sequential data assimilation system based on a LSM uncoupled with the atmosphere. The LDAS uses the SURFEX modeling platform (Masson et al., 2013) developed at Meteo-France in collaboration with the HIRLAM and ALADIN meteorological consortia. SURFEX is designed to be coupled to atmospheric and hydrological models. In the LDAS configuration, SURFEX is used offline (i.e. not coupled with the atmosphere) and is driven by gridded atmospheric forcings. For example, over France, The model is driven by observation-based atmospheric forcing data which are derived from the SAFRAN (Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige) meso-scale analysis system at 8-km spatial resolution and hourly temporal sampling (Quintana et al., 2008). Atmospheric variables include precipitation, 2-m air temperature, 2-m specific humidity, wind speed, surface pressure, incoming solar radiation, and incoming long-wave radiation.

The LDAS is able to integrate simultaneously available SSM and LAI observations at a given time step into the ISBA-A-gs LSM, aiming at adjusting the model trajectory at that time. ISBA-A-gs represents the vegetation sub-grid heterogeneity (crops, grasslands, coniferous forests, broadleaf forests) by using a mosaic approach (Koster and Suarez, 1992).

In the absence of a priori knowledge (such as agricultural practices) at the landscape scale, the assimilation approach described in Sect. 3.2.4 used the hypothesis that the distribution of innovations is proportional to the cover area. Useful information can be extracted from the data signal at the grid level and distributed among the patch structure of the model.

An extended Kalman filter (see Sect) is used in order to incorporate the SSM product derived from ASCAT together with the GEOV1 LAI satellite product. An important motivation for combining two sources of information within a land surface model is the expectation that they will contribute in a more coherent manner to reduce model uncertainties.

The patch fraction approach, the Jacobian behavior and the uncertainties in the model and in the observations described in the next sections, offer the possibility to adapt the analysis to each plant functional type. The multi-patch assimilation may be viewed as an

opportunity to include valuable sub-grid information coming from other sources for improving the analysis among patches.

The three main components of the LDAS (land surface model, remote sensing data and analysis scheme) are detailed hereafter.

3.1.2. Land Surface Model

In the SURFEX platform (<http://www.cnrm.meteo.fr/surfex/>), the ISBA LSM (Noilhan and Mahfouf, 1996) describes the exchanges between soil, vegetation and atmosphere. Version 7.3 of SURFEX is used. The force-restore three-layer version of the soil model in ISBA is used (Boone et al., 1999). The soil texture parameters (clay and sand proportions) are taken from the soil geographical database (BDGSF) of the French National Institute of Agronomic Research available at <http://www.gissol.fr/programme/bdgsf/>. For each model grid cell, the modeled soil moisture is partitioned into three variables: the simulated SSM (representative of the first soil centimeter), the volumetric root-zone soil moisture WG2 (defined for rooting depths depending on the vegetation type, with a maximum thickness of 2.5 m) and a volumetric soil moisture value WG3 in the recharge zone below the plant roots (with a maximum thickness of 1 m). Hereafter, the simulated SSM will be referred to as “SSMmod”. In the model, the propagation of surface information to root-zone layer relies on the force-restore dynamics of the model: SSMmod is forced by precipitation and evaporation and restored toward WG2.

The A-gs module of ISBA was developed to allow the simulation of photosynthesis and the growth of vegetation with different biomass reservoirs (Calvet et al., 1998). The vegetation biomass and LAI variables are governed by photosynthesis and evolve dynamically in response to weather and climate conditions. Namely, during the growing phase the net assimilation of CO₂ photosynthesis leads to plant growth from a minimum threshold set to either 1 m² m⁻² for forest or 0.3 m² m⁻² for herbaceous vegetation. The larger threshold for forest was set in order to represent the vegetation layer between the forest canopy and the ground and the evergreen characteristics. A deficit of photosynthesis sets off leaf biomass mortality that exceeds net assimilation. Consequently LAI drops down to its minimum value.

The photosynthetic activity depends on the vegetation types. The input soil and vegetation parameters are provided by the ECOCLIMAP-II global database (Faroux et al., 2013) which describes ecosystem classes and assigns them in twelve elementary land cover types (patches) at 1 km spatial resolution. Three of them represent patches without vegetation (bare soil, permanent snow and rocks). Over the France domain, the dominant ecosystems are grasslands (31 %), C3 croplands (24 %), deciduous forest (20 %), coniferous forests (11

%) and C4 croplands (4 %). Bare soil represents 8 % of the area. The mean root depths are of 1.5 m for herbaceous vegetation and 2 m for forests.

The water and energy budgets are calculated separately for each patch. ISBA-A-gs simulates the aggregation of carbon, water, and energy fluxes from the different patches. The modeled LAI at 8 km resolution is an average value of vegetation types (up to 9 in the current configuration) weighted with their cover fraction. The ISBA A-gs model simulates the interaction between water and carbon cycles. The evapo-transpiration flux (ET) represents the sum of the evaporation of liquid water from the soil surface and from the vegetation, and the sublimation from the snow and soil ice. The net ecosystem CO₂ exchange (NEE) is given by the difference between the ecosystem respiration (RECO) and Gross Primary Production (GPP). The GPP represents the carbon uptake by photosynthesis. The "NIT" version of ISBA-A-gs is used in the LDAS. This version interactively calculates the leaf biomass and LAI, using a plant growth model (Calvet et al., 1998; Calvet and Soussana, 2001) driven by photosynthesis. The vegetation growth and senescence are entirely driven by photosynthesis. The leaf biomass is supplied with the carbon assimilated by photosynthesis, and decreased by a turnover and a respiration term. Turnover is increased by a deficit in photosynthesis. The leaf onset is triggered by sufficient photosynthesis levels and a minimum LAI value is prescribed. The maximum annual value of LAI is prognostic, i.e. it is predicted by the model. The flexibility of the photosynthesis driven vegetation-growth model of ISBA-A-gs facilitates the use of data assimilation techniques. In this version, the RECO value departs from a basal rate as a function of soil temperature and soil moisture (Albergel et al., 2010). Following Lafont et al. (2012) the respiration basal rate was calibrated by assuming near equilibrium between the ecosystem respiration and the vegetation carbon uptake over the period 2007-2012. It is assumed that the accumulated RECO represents 72% of the accumulated GPP, following a study performed over French flux tower data (http://www-lscedods.cea.fr/invSAT/PEYLIN/CARBOFRANCE/rapport_carbofrance_final.pdf). Also, the A-gs module features two different types of the plant response to drought, for both herbaceous vegetation (Calvet, 2000) and forests (Calvet et al., 2004). In the strategy called defensive or drought-avoiding, the plant increases the water use efficiency (WUE) in response to soil water stress, while in the offensive strategy or drought-tolerant, the WUE is stable or even decreases. While C3 crops and coniferous trees are associated to a drought-avoiding behavior, C4 crops, grasslands and broadleaf trees are associated to a drought-tolerant behavior (Calvet et al., 2012, 2004).

3.1.3. Remote sensing data sets

- Satellite-derived SSM

The Advanced Scatterometer (ASCAT) is an active C-band microwave sensor on board the European METOP polar-orbiting satellite. The soil moisture information is derived from ASCAT radar backscatter coefficients delivered at 25 km resolution using a methodology developed at the Vienna University of Technology (TU-Wien). This method is based on a change detection approach originally developed for the active microwave instrument flown on-board the European satellites ERS-1 and ERS-2 (Wagner et al., 1999; Bartalis et al., 2007). An exponential filter in its recursive formulation (Albergel et al., 2008) is applied to this SSM product to estimate the soil water index (SWI) using a time scale parameter T that may vary between 1 day and 100 days. The result for the top soil moisture content (< 5 cm) range between 0 (dry) and 1 (saturated). Hereafter, this quantity will be referred to as "SSMsat". In this study, SSMsat consists of Copernicus Global Land soil water index values with a characteristic time length of one day, denoted by SWI-001 (Kidd et al., 2011). A surface state flag which identifies either frozen conditions, presence of snow cover or temporary melting/water on the surface is provided. The product includes also a quality flag indicating the availability of SSMsat measurements with an acceptable quality. After screening, the remaining data were projected onto the 8 km model grid resolution by assigning each observation to all SAFRAN grid cells within 0.15° and then considering the average of data assigned to each model grid. The model time series (SSMmod) were compared with soil moisture data (SSMsat) to determine their capability to represent the temporal dynamics at a grid scale.

A good agreement between the SSMsat and the SSMmod was found, despite anomalously low values of SSMsat produced in frozen surface conditions. These erroneous values are not adequately identified by the flags. This suggests that an additional frozen surface mask depending on model forecasts of frozen conditions has to be applied to the SSMsat data before being used in a data assimilation application. Similar to Draper et al. (2011), the screening procedure was extended to the use of two additional static masks in order to discard data in urban regions with an urban fraction greater than 15 % in the ECOCLIMAP database, and to remove data with a topographic flag representing mountainous regions with an altitude greater than 1500 m. The topographic information is provided by the GTOPO30, a global digital elevation model (DEM) with a horizontal grid spacing of approximately 1 kilometer.

- Satellite-derived LAI product

The GEOV1 LAI product is derived from the SPOT-VGT satellite observations. Hereafter, this quantity will be referred to as "LAI_{sat}". The LAI_{sat} values are produced by a statistical algorithm, namely a neural network trained using two pre-existing products: the SPOT-VGT CYCLOPES V3.1 product (Baret et al., 2007) and the TERRA/ AQUA MODIS collection 5 product (Myneni et al., 2002). The product is provided globally at a spatial resolution of 1 km and a 10 day sampling time in a Plate Carrée projection. The LAI_{sat} is close to the true LAI

since the saturation effect affecting the CYCLOPES product for large LAI values, corresponding to dense canopies, was reduced by the MODIS contribution in the training process. The retrieval methodology and detailed information about the product are described by Baret et al. (2013). Camacho et al. (2013) performed a validation study by comparing the GEOV1 product with ground measurements and other reference satellite products. They concluded that the GEOV1 is a reliable product and has an important added-value regarding its two precursor products.

A quality check was performed using a number of quality flags provided with the LAIsat. The data are kept only if all the quality flags are set to 0. The 1 km data are aggregated at the model grid at 8 km resolution if at least 32 grid points are present (more than half the maximum amount).

- Satellite-derived FAPAR product

FAPAR corresponds to the fraction of photosynthetically active radiation absorbed by the canopy. The FAPAR value results directly from the radiative transfer model in the canopy which is computed instantaneously. The instantaneous FAPAR value at 10:00 solar time is used under clear sky conditions (equivalent to black-sky conditions as defined also for albedo). A quality check is performed using a number of quality flags provided with the FAPAR product. The data are kept only if all the quality flags are set to 0. The 1 km data are aggregated at the model grid spatial resolution if more than half the model grid-cell surface is covered.

3.1.4. Land Data Assimilation System

3.1.4.1. Generic algorithm

The DA methods employed in this work are derived from the Kalman filter theory.

In sequential data assimilation the system state estimate, given by a solution of the model equations, is updated each time measurements are available. This update is usually referred to as “the analysis”.

The model equations are discretized according to:

$$\boxed{x^b(t_i) = M(x^a(t_{i-1}))}, \quad (\text{Eq. 3.1})$$

Here, the forward operator is the land surface scheme ISBA-A-gs denoted by M . This operator computes the time evolution of the control vector x^b which contains the root-zone soil moisture and the LAI at time t_i given their analyzed values (x^a) at previous time.

The Kalman filter analysis update is:

$$x^a = x^b + K(y^o - y) \quad (\text{Eq. 3.2})$$

where y^o is the assimilated observation and $y=H(x)$ is the model predicted value of the observation.

The model state and the observations are weighted using the Kalman gain K :

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}, \quad (\text{Eq. 3.3})$$

where \mathbf{H} is the linearized observation operator, \mathbf{B} is the background-error covariance and \mathbf{R} is the observation-error covariance.

The SEKF (Section 3.1.4.2) and the ensemble DA methods (Section 3.1.4.3) differ in the way they approximate the background-error covariance and the way they propagate it to the end of the assimilation window.

3.1.4.2. SEKF

The Extended Kalman filter (EKF) uses the full nonlinear model to propagate the state estimate, but uses a local linearization of the dynamics to propagate the state uncertainty, that is the error covariance matrix. A finite difference method is used to linearize the forecast model, as well as the observation operator by performing model integrations with perturbed initial values of the state vector. The simplified Extended Kalman filter (SEKF) (Mahfouf et al., 2009) is based on the EKF. The SEKF simplifies the EKF by using both a diagonal and climatological background-error covariance at the start of the assimilation window. The SEKF implicitly propagates \mathbf{B} to the end of the assimilation window via the Jacobian of the observation operator \mathbf{H} . The Jacobian \mathbf{H}^{kl} for observation k and model point l is calculated by finite differences:

$$H^{kl} = \frac{\partial y^k}{\partial x^l} = \left[\frac{y^k(x + \delta x^l) - y^k(x)}{\delta x^l} \right] = \left[\frac{H^k(M(x^b(t_{i-1}) + \delta x^l)) - H^k(M(x(t_{i-1})))}{\delta x^l(t_{i-1})} \right] \quad (\text{Eq. 3.4})$$

3.1.4.3. EnKF

The EnKF (Evensen, 1994) is a way of representing the uncertainty in the prognostic variables using an ensemble of model trajectories. This circumvents the high computational cost of explicitly storing and propagating the background-error covariance for a large model dimension. Each ensemble member is propagated using the nonlinear model. The deterministic analysis comes from the ensemble mean:

$$\bar{x}^a = \bar{x}^b + K(y^o - H(x^b)). \quad (\text{Eq. 3.5})$$

The Kalman gain is given by:

$$K = (P^b)^T H^T (H P^b H^T + R)^{-1}, \quad (\text{Eq. 3.6})$$

where the ensemble background-error covariance:

$$P^b = \frac{1}{m-1} (X^b)^T X^b. \quad (\text{Eq. 3.7})$$

The ensemble perturbation matrix (of dimension $n \times m$) is defined as:

$$X^b = \frac{1}{\sqrt{m-1}} [\delta x_1^b \quad \dots \quad \delta x_m^b]. \quad (\text{Eq. 3.8})$$

where $\delta x_j^b = x_j^b - \bar{x}^b$ are the perturbations from the ensemble mean.

An additional step is required to avoid ensemble collapse. The traditional EnKF of Burgers et al. (1998) maintains the ensemble spread by perturbing the observations, with perturbations randomly sampled from a zero-mean normal distribution with covariance \mathbf{R} . The serial ensemble square root filter was introduced by Whitaker et al. (2002) as a means of avoiding the sampling error from the perturbed observations. The ensemble perturbations are defined by:

$$\delta x_j^a = \delta x_j^b - \alpha K y_j, \quad (\text{Eq. 3.9})$$

where

$$\alpha = \frac{1.0}{1.0 + \sqrt{\frac{R}{H P^b H^T + R^2}}}. \quad (\text{Eq. 3.10})$$

The added value of EnKF with respect to SEKF is the possibility to account for atmospheric forcing errors and a lower computing time when many variables have to be

analyzed (e.g. using refined versions of the soil hydrology sub-model, including many soil layers). A full description and test of the EnKF can be found in Fairbairn et al. (2015).

3.1.4.4. Multi-patch representation

The model permits a multi-patch representation of land (e.g. grasslands, broadleaf forests, etc.). A model grid box is divided in a number of patches each having its own LAI and soil moisture prognostic variables. With 12 patches, the dimension of the control vector is increased, but the number of observations is the same. The filter algorithm is designed to provide the analysis for each patch independently by using one observation per grid box. Therefore, the model counterpart of the observation is assumed to be the average of the corresponding simulated observation for each patch weighted with the fraction occupied by each patch. The simulated observation over one patch depends only upon the control vector over the same patch. In the SEKF, this approach reduces the number of perturbed runs needed to compute the Jacobian matrix. The Kalman gain is computed independently for each patch. A new value that represents the analysis is obtained as an optimal combination of the observation via the increment and background as in Equations 3.2 and 3.5.

Finally, the updated value is aggregated from the weighted contribution of each patch over the vegetation tile.

An illustration of the land data assimilation algorithm is shown in Figure 3.

Figure 4 illustrates an example of the multi-patch data assimilation scheme for a grid cell split into three patches. The aggregation and disaggregation arrows correspond to the calculation of the model counterparts of a grid-scale observation and to the extraction of the data information distributed among the patches.

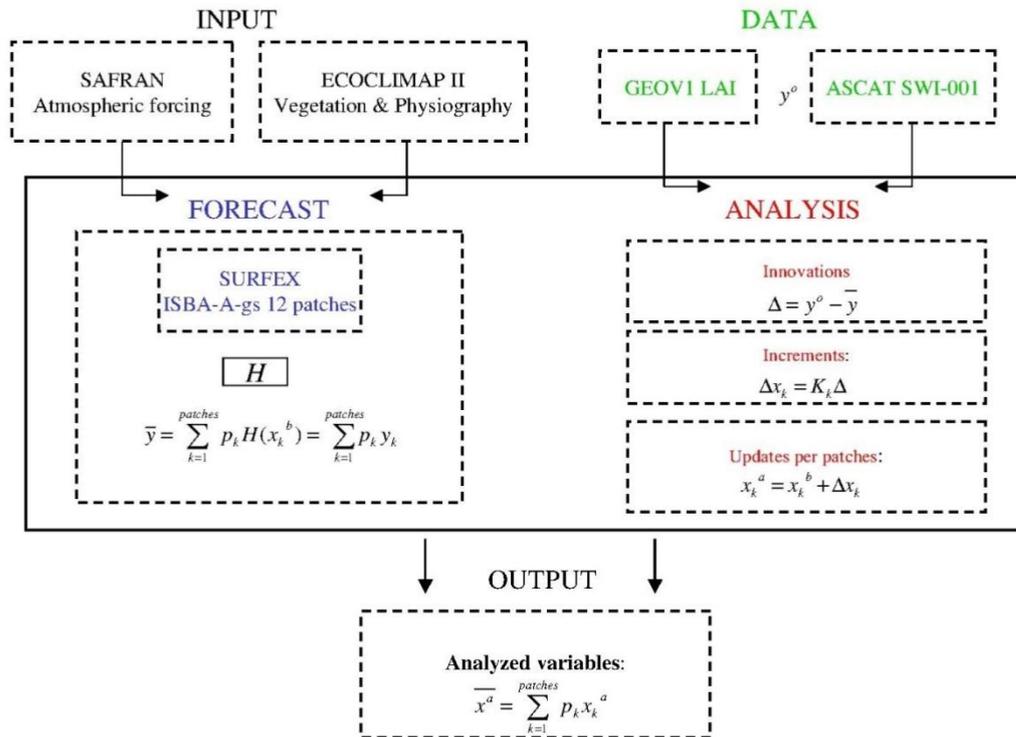


Figure 3: Flow chart of the land data assimilation algorithm

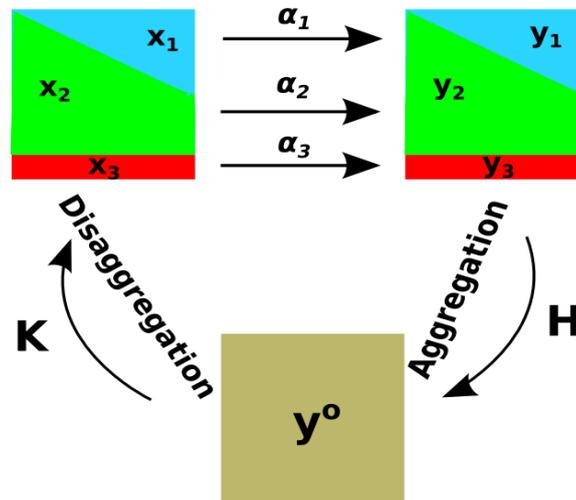


Figure 4: The multi-patch data assimilation design for a grid cell split into three patches, each having their variables x_p . The predicted observations y_p are weighted with their fraction denoted by α_p , where $p = 1, 2, 3$. y^o represents the grid-scale observation. The observation operator H aggregates the predicted observations at the grid scale. The data information is split among patches via the Kalman gain K .

3.1.4.5. State vector and resolution

The vector of observations includes two elements: SSMsat and LAIsat at each grid cell. The control state vector for the analysis consists of two prognostic variables, root-zone soil moisture and LAI, each of them containing 12 values that correspond to the twelve elementary land cover types (patches). The surface soil moisture does not belong to the vector state, but it is a prognostic variable in the ISBA-A-gs LSM. The root-zone soil moisture is of great relevance to this study, as it governs the plant response to drought. Moreover, the Kalman filter is particularly useful for correcting the system variables with a slow temporal evolution such as LAI and WG2. Due to the small capacity of the surface soil water reservoir, the SSM is rapidly influenced by the atmospheric forcing and by the capillarity rises from the deep reservoir. Therefore, a dedicated initialization of the SSM is less important than that of WG2 associated with soil depths up to 2.5 m. Following a recommendation of Draper et al. (2009), the SSM prognostic variable was excluded from the control vector in order to reduce the number of linearization of the SEKF Jacobian. The modelled SSM is used to calculate the innovations and is linked to the control variable WG2 via the prognostic equations of the

ISBA scheme. The SSM variable is indirectly corrected through the changes made by the assimilation in the deep reservoir. The model simulations without data assimilation (prior) start in January 2007. The year 2007 is considered as a spin-up period for the model run in order to obtain an equilibrium state. Then, the assimilation experiment starts on 1 January 2008 and lasts until June 2014. The SEKF assimilates available observations every 24 h at 09:00 UTC, by analysing the initial state via the information provided by an observation at the end of the assimilation window. The SSMsat observations are converted into volumetric water content (see Sect. 3.1.4.6) and assumed to be the observation equivalent of the SSMmod. Possible mismatch between observed and modeled quantities are accounted for in a bias correction scheme described in the next section. Concerning LAI, the satellite-derived product is considered to be the observation equivalent of the simulated LAI. The LAIsat is assimilated at the provided temporal resolution of 10 days. The LDAS products (LAI, root-zone soil moisture, water, carbon and energy fluxes) are provided across the France (world) domain split into 8602 (87612) grid cells of 0.07° (0.50°) at the temporal resolution of one day for the state variables LAI and soil moisture and as cumulated daily outputs for the fluxes. Hereafter, the analyzed variables will be referred to as “posterior”, in opposition to the modeled variables referred to as “prior”.

3.1.4.6. *Rescaling and bias correction*

Prior to assimilation, the SSMsat product has to be transformed into model equivalent volumetric SSM. The discrepancies between the model simulations and the satellite observations are addressed as part of the data assimilation system. The SSMsat data are rescaled by matching its CDF to that of SSMmod. The approach described in Scipal et al. (2008) permits correcting for the differences in the first two moments (mean and variance) of the distribution and can be viewed as a linear transformation. The two parameters of the linear relationship, the intercept a and the slope b vary spatially, but are constant in time:

$$a = \theta_m - b \times \theta_o,$$

$$b = \sigma_m / \sigma_o,$$

where θ_m and θ_o hold for the means of model and observation, respectively, while σ_m and σ_o represent the standard deviation errors for model and observations, respectively. Scipal et al. (2008) noted that the use of a linear transformation produces the bias free observations (with respect to the model) for the entire considered period, but systematic differences related to seasonal or inter-annual variations in bias may remain uncorrected. The importance of accounting for seasonal corrections in the CDF matching was discussed for the AMSR-E

(Advanced Microwave Scanning Radiometer) SSM data and the SSM provided by the NWP system ALADIN of Meteo-France by Draper et al. (2009).

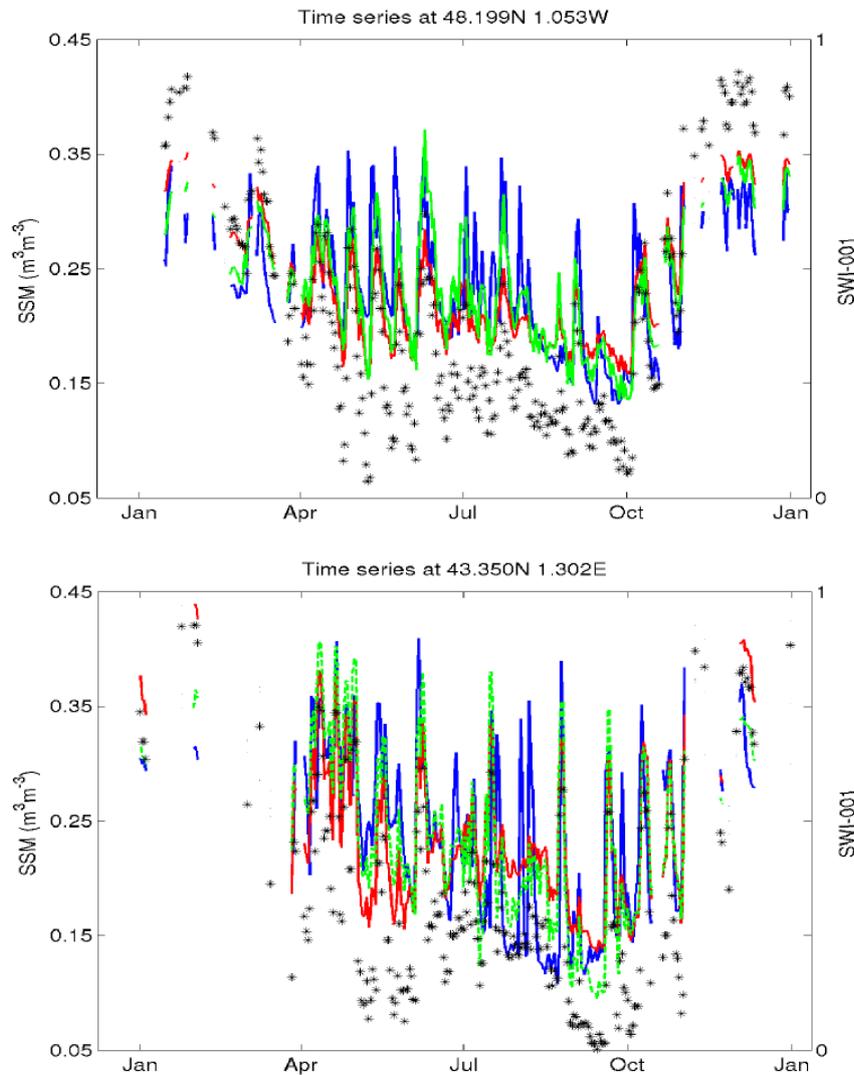


Figure 5: Surface soil moisture evolutions for 2009 at two location in North East (top) and in South West (bottom) of France for model (blue), ASCAT CDF rescaled (red) and ASCAT seasonal CDF rescaled (green) observations. SWI-001 observations ranging between 0 and 1 are depicted by black stars.

Therefore, we have derived the a and b CDF matching parameters on a seasonal basis by using a three-month moving window (De Rosnay et al., 2013) from 2007 to 2014. The two CDF matching parameters are calculated on a monthly basis using a three-month moving window from 2007 to 2014, for each model grid-cell. Therefore a single set of 12 pairs of parameters is obtained for the whole 2007-2014 period. The CDF matching moments are computed based on (1) screened observations with the quality flag provided with the SSMsat and additional aforementioned masks for altitude and urban fractions, and (2) SSMmod values for simulated soil temperature above 0° C.

In Figure 5 the model simulations are compared with two CDF rescaled ASCAT time series (without seasonal corrections and with seasonal corrections) to determine their ability to represent the temporal dynamics at two locations in North-East (48.199 N, 1.053 W) and in South-West (43.350 N, 1.302 E) of France. For both locations, the CDF matching with seasonal correction improves the temporal correlations between the data and the model when compared with the approach without seasonal corrections from 0.72 to 0.79 (location in the North-East) and from 0.70 to 0.79 (location in the South-West). In addition, the seasonal bias correction reduces the standard deviation of the bias by 0.01 m³ m⁻³ for both locations.

3.1.4.7. **Background and observation errors**

The performance of an analysis scheme depends on appropriate statistics for background and observation errors. One source of information relies on the statistics of the innovations (observations minus background). If the background and observation errors are uncorrelated and normal distributed, the variance of the innovations is represented by the sum of observation and background variances (Andersson, 2003).

For simulated root-zone soil moisture, a mean volumetric standard deviation error of 0.02 m³ m⁻³ was chosen as suggested by several authors (Mahfouf et al., 2009; Draper et al., 2011; Barbu et al., 2011). The observational error is set to 0.05 m³ m⁻³ according to the median value of SSMsat ASCAT data error estimates. This value is consistent with errors typically expected for remotely sensed soil moisture (de Jeu et al., 2008; Draper et al., 2011).

Background error is estimated from the innovations (differences between observations and background) statistics using the bias corrected ASCAT data. The distribution is essentially unbiased (by definition after using a bias correction scheme) with a near-Gaussian shape. Assuming that these errors are Gaussian and uncorrelated, the variance of the distribution is $[\sigma_o]^2 + [\sigma_b]^2$. Over the whole period, the standard deviation is 0.07 m³ m⁻³. Moreover, by assuming that the observation error is equal to SSMmod error (considering that neither the ASCAT product nor the model is a better estimate of the true

surface soil moisture) each error has a standard deviation of $0.05 \text{ m}^3 \text{ m}^{-3}$. It is important to keep in mind that this single value is probably not optimal for all soil moisture ranges (probably too high for dry soils and too low for wet soils). Removing the arbitrariness of such assumption will be addressed in future studies with a triple collocation approach.

Similar to Mahfouf et al. (2010) and Dharssi et al. (2011), a background quality control is performed in order to reject SSMSat converted to volumetric soil moisture values that are too far from the model first-guess. The data are discarded if the innovations are larger than $0.21 \text{ m}^3 \text{ m}^{-3}$ (i.e. three times the square root of the sum of the observation and background variances).

Concerning LAI, in a previous study, Barbu et al. (2011) estimated in situ LAI observation errors for a grassland in southwestern France. Generally, the start and the end of the growing season are considered being better represented by the satellite products (Jarlan et al., 2008; Brut et al., 2009; Szczypta et al., 2012). In this version of the algorithm, the standard deviation of errors of LAIsat is assumed to be 20% of LAIsat. The same assumption is made for the standard deviation of errors of LAImod (20% of LAImod) for LAImod values higher than $2 \text{ m}^2 \text{ m}^{-2}$. For LAImod values lower than $2 \text{ m}^2 \text{ m}^{-2}$ a constant error of $0.4 \text{ m}^2 \text{ m}^{-2}$ is assumed (option 3 in Barbu et al., 2011). It has been checked that the observation and model error specifications are consistent with innovation statistics in a similar way as for the soil moisture variable.

3.1.5. Limitations of the method

The assimilation results depend to a large extent upon the quality of the data to be assimilated. The remotely sensed SSM data exhibit a number of non-realistic low values associated with large uncertainties over densely vegetated areas (Kidd et al., 2013). This may be detrimental to analysis by causing a too-large soil moisture depletion. Nevertheless, this effect has been minimized by the seasonal bias removal performed before the assimilation.

A single, thick root-zone soil layer represents the soil hydrology in the model version. Such description increases the time to deplete soil moisture causing this slow response to dry conditions. Also, the propagation of surface soil moisture information within deeper layers may be affected by the lack of vertical resolution of the model. Increasing the number of soil layers will allow an explicit representation of a vertical distribution of the root profile in the soil and, subsequently, a more realistic vegetation response to water stress. In that respect, a multi-layer version of the soil hydrology is expected to improve the overall performance of the system.

Many studies indicate the presence of systematic biases between the observations and the model outputs for soil moisture (Walker et al., 2003; De Lanoy et al., 2007) and LAI (Jarlan et al., 2008; Lafont et al., 2012). These biases have multiple origins related to model inputs, model physics or data retrieval procedures. Correcting only for the existing bias in observations without properly accounting for model uncertainties leads to suboptimal data assimilation results. The characterization of errors associated with model dynamics and parameters is a rather challenging area for land data assimilation as mentioned by Reichle (2008).

Errors affecting model simulations may be related to the distribution of bare soil and vegetation and may depend on the vegetation type. For example, croplands present more heterogeneities than grasslands and many processes of anthropogenic nature are not described for crops in the model. Model (background) errors were set to a single value for all patches. It is important to keep in mind that this single value is probably not optimal for all soil moisture ranges (probably too high for dry soils and too low for wet soils) as well as for LAI ranges. However, even under this simple assumption, the assimilation partly compensates for the lack of description of managed ecosystems by reducing the duration of the crop phenological cycle, which tends to be too long in the model. In the case of highly heterogeneous pixels, one has to be aware that this compensation may be not adequately distributed among the patches, especially in grid cells covering complex crop rotation systems. The choice of model error may be refined by assigning different error statistics to different ecosystems and making use of observed LAI at its original resolution of 1 km. Assimilating original LAI data may help in decreasing the occurrence of sub-grid inconsistencies.

3.2. REGIONAL LDAS ARCHITECTURE

The LDAS can be used a high resolution (0.07°) at a regional scale (e.g. over France, Hungary) and at a lower spatial resolution (0.50°) at a global scale. A low resolution chain working at a global scale (LDAS-Monde) is described in Chapter 4.

The present description of LDAS is based on the offline version of SURFEX v7.3.

3.2.1. SEKF

The VARASSIM (routine name, not an acronym) source package contains a number of specific routines for the assimilation purposes. The main routine performs the various steps

of the assimilation system (forecast, data selection and analysis). A bash script that allows running the EKF over a period of time drives these routines. It operates in several steps:

- First, the initial perturbations are created. The initial background error covariance matrix is defined and stored in a file.
- SURFEX runs with the perturbed initial conditions for each of control variables. The perturbed evolved prognostic variables and the perturbed simulated observations are stored in temporary files.
- SURFEX runs with the reference initial conditions. The reference evolved prognostic variables and the reference simulated observations are stored in temporary files.
- The Jacobian matrix of the forwarded model is computed by using the perturbed and reference runs. The error covariance matrix is evolved in time, and stored in a file to be used further in the analysis.
- The assimilation is prepared for by reading the actual observations, the simulated (reference and perturbed) observations, computing the Jacobian of observation operator and the observation error covariance matrix.
- The analysis is performed by calculating the innovation vector and the Kalman gain and stored in a file ready for the next assimilation cycle. The new analysis error covariance matrix is computed and stored in a file for the next cycle.

3.2.2. EnKF

The EnKF source package is structured in a similar way to the VARASSIM source package. The execution of the EnKF consists of three levels: assimilation parameter prescription, assimilation temporal/spatial organization, elementary assimilation calculations. At the highest level, a python routine specifies the experimental parameters for the EnKF. The version of the EnKF is also specified. There are two versions of the EnKF that use a different calculation of the analysis: The traditional EnKF (referred to as the EnKF) and the Ensemble square root filter (EnSRF). The python script then calls a Bash script, which runs the EnKF over a period of time. The Bash script calls the EnKF routines.

The EnKF operates in several steps:

- For the first assimilation window, the ensemble members are initialized by perturbing the input variables using white noise. Otherwise, the ensemble members are read from the previous analysis.
- SURFEX runs with each ensemble member. The evolved ensemble members are stored in temporary files.
- The assimilation is prepared for by reading the actual observations, the simulated observations for each ensemble member, and the ensemble members for each prognostic variable.
- The analysis is performed for each ensemble member by calculating the innovation vector for each ensemble member, and by calculating the Kalman gain. Depending on the EnKF version (EnKF or EnSRF), the analysis is calculated slightly differently.
- Red noise is then added to the ensemble perturbations. This is necessary to maintain the ensemble spread.
- The analysis ensemble is stored in a file ready for the next assimilation cycle.

3.2.3. Inputs

3.2.3.1. *Physiographic fields:*

- Land cover map: ECOCLIMAP II included in the SURFEX software.
- Soil texture: CLAY and SAND fraction from HWSD
- Orography from GTOPO30 (for pre-processing of the ASCAT remote sensing data by using a topographic mask).

3.2.3.2. *Meteorological inputs:*

Over France (LDAS-France), the meteorological inputs used by the model are provided by the operational SAFRAN (Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige) meso-scale analysis system at 8-km spatial resolution and hourly temporal sampling (Quintana et al., 2008). The SAFRAN data are uploaded monthly on the Météo-France archiving system as direct access binary files. The SAFRAN analysis

covers 9882 grid points on a domain larger than France. The first step of the processing is to select only the grid points located over France (8602 points). In a second step, the atmospheric [CO₂] concentration is added to the forcing files. It evolves yearly according to the IPCC A2 scenario (IPCC, 2001). The SAFRAN data are collected by FTP once per month. On the 22 of month N, the data of month N-1 are available. The forcing is split according to the length of the assimilation window. In this experiment, there is one forcing file per day.

The SAFRAN files include the following fields (which are used by the model):

- Surface air temperature and air humidity,
- Incoming short wave and long wave radiation,
- Atmospheric pressure,
- Wind speed,
- Precipitation.

3.2.4. Baseline processing

3.2.4.1. *Remote sensing observations for the assimilation*

Before being evaluated or assimilated, the remote sensing data should be prepared.

Pre-processing the remote sensing data:

- GEOV1 LAI:

The Copernicus Global Land GEOV1 LAI dataset is collected from the website (<http://land.copernicus.eu/global>) in the HDF5 format. The data are provided at a temporal resolution of 10 days. A quality check based on the Quality Flag field was performed. The data are kept only if all the quality flags are set to 0. Then the 1-km data are aggregated on the model grid cell scale if at least half the surface of the grid cell is covered.

- ASCAT near-surface soil moisture:

The ASCAT-derived SWI-001 (T=1day) product is supplied by IPMA from Jan. 2007 to present in HDF5 format.

Before projecting ASCAT data onto the model grid, the observations are screened in order to remove the observations with a quality flag QF lower than 40% and only the data flagged SSF=0 or SSF=1 are used. After projection, additional masks for urban regions, steep mountainous terrain, and frozen instances indicated by the model simulations but not detected by ASCAT, are applied.

Finally, the CDF matching designed to remove the systematic differences between data and model simulations is performed. The data set is used to compute the seasonal CDF matching parameters as described in Scipal et al., 2008. This technique is improved by computing the statistical parameters based on 3-months moving window (the updated CDF parameters are applied to the second month of the 3-month period).

- Observation files:

All the remote sensing data are gathered in ASCII files. There is one file per assimilation window (one day) that contains the two types of observations located around the analysis time (end of assimilation window). When an observation is missing at a grid cell or at a given time, it is set to 999.0. These daily files integrate all the pre-processing data treatments described here.

3.2.4.2. **The EKF namelist**

The choice of the control variables, observation types to be assimilated, model and observation errors, and size of perturbations is done by setting the corresponding elements in the namelist. The control variables are two prognostic variables: LAI and the volumetric water content WG_2 in the root zone. Two observation types are considered in the assimilation: SWI-001 and LAI.

3.2.4.3. **Sequence of operations, Data streams, Interfaces**

Step 0: Install LDAS

a) Install the last version of SURFEX (see the <http://www.cnrm.meteo.fr/surfex-lab/spip.php?article163>)

- get the latest version from the trunk:

svn co <http://svn.cnrm-game-meteo.fr/projets/surfex/trunk>

- compile the master version of the code: in the src directory, run make, and then make installmaster.

Master executables are created in the directory exe.

b) Download and install de LDAS package (see the SVN branch)

- choose a name for your own source directory in src, for example **MYSRC_name_exp**. Cp the sources from the VARASSIM directory onto **\$\$SRC_SURFEX/src/MYSRC_name_exp**
- modify the compiling list in the Makefile.SURFEX: instead of the list used for compiling the MASTER version, compile the PGD, PREP, OFFLINE & VARASSIM
- go to **\$\$SRC_SURFEX/src** and launch successively export **VER_USER=MYSRC_name_exp**, **./configure**, **./conf/profile_surfex-LXgfortran-SFX-version_SURFEX-MYSRC-MPIAUTO-DEBUG**, **make user** and **make installuser**. New executable files for **MYSRC_name_exp** will be created in the exe directory.

Step1: Download the remote sensing data

a) Copernicus Global Land LAI: <http://land.copernicus.eu/global>

The data are added on an FTP server : *catftp.vgt.vito.be*.

b) Copernicus Global Land SWI : <http://land.copernicus.eu/global>

The SWI data can be downloaded from an ftp server.

The SWI data version 2.0.0 is available from 01/01/2007 to present.

Step 2: Download and pre-processing of the SAFRAN operational analysis

The SAFRAN forcing is generated in an incremental way. Each month, the data are computed between the 1st of August and the 21st of the current month. Once a year (during August) the forcing files are recomputed for the whole year.

Step 3: Running the open loop (offline)

Run the model. The output is in BINARY format at a temporal scale of three hours.

Step 4: Pre-processing of the Copernicus Global Land remote sensing data

extract and screen the LAI and SWI data (see section 3.2.1.),

calibrate (CDF matching) the SWI data using the pre-calculated parameters a and b.

produce the daily observation ASCII files containing the two columns of data for all SAFRAN grid cells (8602 points).

Step 5: Running the assimilation system

Run the assimilation.

3.2.5. Output

The analyzed variables are produced at a temporal scale of three hours. Over a window of 24h, the three-hour output split the window into 8 values. If the assimilation system starts at 9UTC, the first output value in the binary file is three hours later at 12UTC. For fluxes the last output represents the 24h cumulated values. The specific data assimilation tools (increments, Jacobians, gains, innovations) are written in a daily ASCII format.

The following set of products is considered for the ImagineS project:

- LAI, FAPAR and soil moisture
- Total albedo et land surface temperature (LST)
- Evapo-transpiration, drainage, runoff, and carbon fluxes

3.2.6. Python LDAS chain

Here we describe the structure of the python LDAS chain code. The chain implements the following steps:

1. Preparation of forcing files.

2. Pre-processing of remotely sensed observations.

3. Post-processing of the analysis outputs.

The code is available in the zipped archive `ldas_chain.zip`.

For each implemented step of the chain (forcing preparation, observations pre-processing, and outputs post-processing), the code consists of a main program and a set of functions. The main program reads the options, which are defined by the user in a Python script, and calls the functions to perform the chain operations.

The core of the procedure, as well as most pre-processing functions, is variable- and domain-independent. Indeed the code has been designed to be extensible for working with new:

- observation variables;
- model spatial domains;
- observation screening and rescaling methods;
- grid interpolation methods.

The code is written in Python and in Fortran 90. The code makes use of the F2py compiler, which allows calling compiled Fortran 90 subroutines using Numpy arrays for input/output.

A diagram of the general code structure is given below:

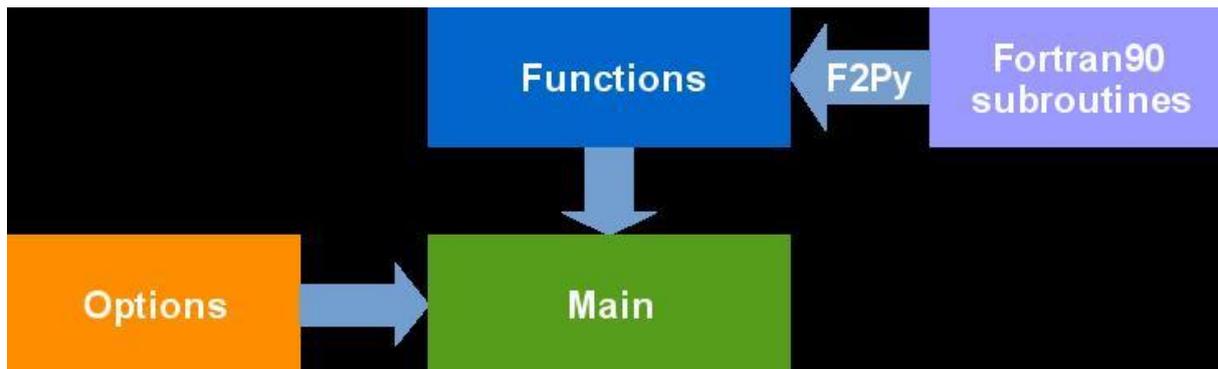


Figure 6: General code structure of the Python LDAS chain

The code is organized in three Python modules, each one implementing a functionality of the chain:

1. `prepareForcing.py`: retrieval and preparation of the forcing files needed to run the model in data assimilation mode.
2. `ldasPre.py`: pre-processing of the remotely sensed observations, which are assimilated by the model and/or used to evaluate the analysis results in the post-processing.
3. `ldasPost.py`: post-processing of the analysis outputs by producing of figures and tables.

The main operations implemented in the 3 modules are briefly described in the following. All modules read and parse user defined options from the same file.

- Forcing preparation (`prepareForcing.py`)
 - Retrieve the raw files containing the forcing data.
 - Write the daily NetCDF forcing files needed to run the analysis.
- Observations pre-processing (`ldasPre.py`) loop over variables and assimilation dates
 - Read observation data and observation flag values.
 - Interpolate observation data onto model grid.
 - Read in model values if rescaling or screening using model output is required (currently for SWI only).
 - Screen observations using surface features, e.g. altitude, town fraction (currently for SWI only).
 - Perform rescaling (currently for SWI only).
 - Write data files (CANARI and Python formats)
- Output post-processing (`ldasPost.py`)
 - Graphical files corresponding to figures and tables are copied to the folder `<graphics_dir>/copernicus/`, where `<graphics_dir>` is specified by the user in the options file.

- Files are named according to the pattern <type>_<number><index>_<figure_name>, where: <type> describes the graphical output type, which can be either 'fig' (for figures) or 'tab' (for tables); <number> is the figure number in the Copernicus report; <index> is the sub-figure/table index; and <figure_name> follows the definitions given in the following Section (2.3.2).
- After each figure/table description, its file name pattern is given between brackets []. The represented time period is indicated in the file name with the strings <start> and <end>, whose format is generally yyyy-mm. For each remotely sensed variable <obs>, the following figures and tables are produced:
 1. Domain averaged time series [<obs>_series_<start>_to_<end>.pdf].
 2. Monthly domain averaged time series [<obs>_series_monthly_<start>_to_<end>.pdf].
 3. Monthly average maps for the current year [<obs>_avg_monthly_maps_<start>_to_<end>.png].
 4. For each score <score> (bias, correlation, RMSD and SDD):
 - 4.1. Monthly score maps for the current year [<obs>_<score>_monthly_maps_<start>_to_<end>.png].
 - 4.2. Annual score maps [<obs>_<score>_annual_maps_<start>_to_<end>.png].
 5. Seasonal cycle of scores computed over the whole domain, comparing the current year with the previous period [<obs>_seasonal_scores_<start>_to_<end>.pdf]. The number of observations are printed in an auxiliary table [<obs>_table_num_obs_seasonal_scores_<start>_to_<end>.pdf].
 6. Annual scores time series computed over the whole domain [<obs>_annual_scores_<start>_to_<end>.pdf]. The number of observations are printed in an auxiliary table [<obs>_table_num_obs_annual_scores_<start>_to_<end>.pdf].
 7. Table of scores computed over the whole domain, comparing the current year with the previous period [<obs>_table_scores_periods_<start>_to_<end>.pdf].
 8. Table of annual scores computed over the whole domain [<obs>_table_scores_annual_<start>_to_<end>.pdf].

Moreover, the following figures are produced:

- a. For each unobserved model output variable `<var>` in `extra_out_names`: domain-average time series plot [`<var>_series_<start>_to_<end>.pdf`].
- b. For each observation variable that is rescaled: raw vs rescaled domain-average monthly time series plot [`<obs>_rescaled_vs_raw_<start>_to_<end>.pdf`].
- c. 3-hourly land surface temperature (LST) bias plots by month `<month>` [`LST_hourly_bias_<month>_<start>_to_<end>.pdf`].
- d. Monthly maps of the differences analysis minus openloop (`<var>` = 'EVAPC_P' and 'DRAIN_C_P') or of the analysis increments (`<var>` = 'LAI' and 'WG2'), comparing the current year with the previous period [`<var>_analysis_increment_monthly_maps_<start>_to_<end>.png`] or [`<var>_analysis_minus_model_monthly_maps_<start>_to_<end>.png`].

3.2.7. Hydro-validation over France

The MODCOU hydrological model (Habets et al., 2008) is used over France to validate the changes in the land surface state triggered by the assimilation.

MODCOU computes the spatial and temporal evolution of the piezometric level of multilayer aquifers, as well as the exchanges between aquifers and rivers, before routing the surface water through the river network. River flows are calculated every 3 hours, while the evolution of the aquifers is computed daily.

Various configurations of the ISBA LSM can be coupled to MODCOU:

- ISBA-A-gs without dynamic evolution of LAI. The annual cycle of LAI is provided by ECOCLIMAP-II as a fixed satellite-derived climatology. This simulation is referred to as "AST" (A-gs and the enhanced soil moisture stress option).
- ISBA-A-gs may simulate daily LAI values. This simulation is referred to as "NIT" (with a nitrogen dilution-based representation of leaf biomass, in addition to the AST capability.)

ISBA-A-gs NIT version may incorporate the data assimilation scheme in conjunction to MODCOU. Two simulations referred to as "LDAS" are provided:

- LDAS1 meaning data assimilation of satellite LAI data only

- LDAS2 meaning data assimilation of both LAI and ASCAT SWI.

The drainage and runoff fluxes generated for each experiment (NIT, AST, LDAS1 and LDAS2) are routed through the surface river network using the MODCOU model.

Comparisons with observed river flow can be performed on a daily or monthly basis at the river gauges located closest to the outlet of the four largest rivers of France (Loire, Seine, Garonne and Rhone). Different statistical scores (bias, correlation coefficient, rmse, Nash score) are calculated.

The MODCOU chain code implements the following steps:

1. Preparation of forcing files consisting of drainage and runoff fluxes provided by the model simulations.
2. Compiling and Running of MODCOU model.
3. Calculate the discharges and their statistical scores.

The code is available at /cnrm/vegeo/barbu/MODCOU/. There are three subdirectories: modcou_src, modcou_forcing and modcou_results.

The modcou_src directory contains several subdirectories. The first one src_init contains the MODCOU sources. The code consists of a main program (hydro_only_jour.F90) and a set of Fortran routines to be compiled. The second directory in the modcou_src called run_init contains the main program (bash script). This program consists of three parts :

- Reading the forcing files and the initial state of the model
- Running the main hydro_only_jour executable
- Collecting output and calculating discharges and statistics using the discharge data.

Several input files (HYDRO_PARAM_*, coord_france, mask, number of days) used for running the modcou executable are located in the same directory.

The third subdirectory (ana_init) contains the main fortran routine lec_debits_v3.f90 that need to be compiled, an option list (OPTIONS_DEBITS.nam) and a list of data stations (STATIONS_ALL).

3.2.8. Summary of the Regional LDAS-France characteristics

Inputs	Meteorological forcing: SAFRAN Physiographic field (from ECOCLIMAP II, SIM) Copernicus Global Land GEOV1 LAI product. Copernicus Global Land ASCAT SWI-001 surface soil moisture.
Outputs	Analyzed LAI, Root-zone soil moisture (WG2), NEE, GPP, Evapo-transpiration
Interfaces	Copernicus Global Land service (http://land.copernicus.eu/global) Météo-France operational SAFRAN production.
Capacity requirements	1 dedicated desktop computer. About 50 Go per year (including output products, inputs and auxiliary variables)

Table 3: Summary of SURFEX regional LDAS characteristics

3.2.9. Regional LDAS in Hungary

The Hungarian LDAS contains many similarities and differences with the French LDAS. In Hungary the applied LDAS is also based on the off-line SURFEX v7_3 model. ISBA-A-gs force-restore three-layer soil model is used in both kinds of experiments (with and without assimilation).

In LDAS-Hungary to analyze the leaf area index (LAI) and the root-zone soil moisture (WG2), we assimilate LAI and ASCAT Soil Water Index (SWI) satellite observations provided by the Copernicus Global Land service (<http://land.copernicus.eu/global>) (Section 3.1.3). Before assimilation, pre-processing of the data is needed. For LAI the same quality check is applied as in France, which is based on the Quality Flag field of the product (Section 3.2.4.1). For SWI, very simple checking option is built in, namely only the missing values are removed from the database. To determine surface soil moisture content from SWI observation we use the following relationship: $SSM = SWI * (w_{max} - w_{min}) + w_{min}$, where w_{max} and w_{min} are the maximal and minimal SSM values that the model can take at a given grid point. These values are determined from a longer (several years) simulation. The derived SSM data are bias corrected with respect to the model climatology by using a seasonal-based CDF (Cumulative

Distribution Function) matching technique described in Scipal et al. (2008). In Figure 7, the model simulation is compared with the raw ASCAT and the CDF rescaled ASCAT time series at Hegyhatsal (located at western part of Hungary). The CDF matching with seasonal correction improves the temporal correlations between the data and the model.

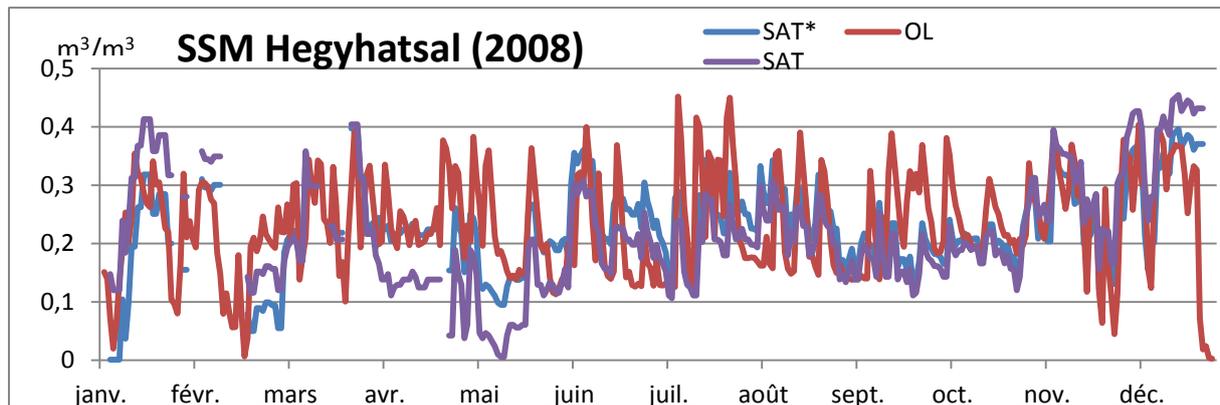


Figure 7: Surface soil moisture evolutions for 2008 at Hegyhatsal (West-Hungary) for model (red), raw ASCAT satellite observations (purple) and ASCAT seasonal CDF rescaled (blue) observations.

The VARASSIM source code originates from Meteo-France which uses EKF method with similar settings as in France: the observation error of LAI is set to $0.5 \text{ m}^2/\text{m}^2$, the model error is $0.2 \text{ m}^2/\text{m}^2$; for WG2, the observation error is $0.04 \text{ m}^3/\text{m}^3$, and the model error equals $0.2 \text{ m}^3/\text{m}^3$.

The applied domain covers Hungary, the resolution of the model is $8 \times 8 \text{ km}$, so the number of grid points is 2698. In SURFEX, each surface grid point is separated into 12 different patches according to the vegetation or surface type. The model calculates the prognostic equations and surface fluxes independently for the different patches.

The input soil and vegetation parameters are derived from ECOCLIMAP II (Faroux et al., 2013). The meteorological inputs are coming from the ALADIN cy36t1 numerical weather prediction (NWP) model forecasts (2 m temperature, pressure, wind speed and rainfall), and LandSAF incoming short and long wave radiation observations are used. The reason for this choice is that radiation influences to a great extent the photosynthesis and the NWP model's fields are not as accurate. Since the forcing fields are available only with 1 hour temporal frequency, but the time step in SURFEX is 5 minutes, the forcing fields are interpolated at the missing time steps.

The model is running in cycling mode which means that one run produces 24 hour forecast (from 6 UTC to the next day 6 UTC) and the next run is started on the next day. The outputs are generated with 6 hours frequency. The products which are evaluated in the frame of ImagineS are: LAI, WG2, NEE, GPP and Evapotranspiration. The outputs are stored in netcdf format.

3.3. CONCLUSIONS AND PROSPECTS

The France domain encompasses a wide variety of soil and vegetation ecosystems. At 8 km pixel scale, there is a high degree of heterogeneity that should be taken into account. Each grid cell is represented as a mosaic of 12 land covers or patches. The ISBA-A-gs LSM provides a detailed computation of the surface fluxes of energy, water and carbon at the sub-grid (patch) level and allows aggregating the information from different ecosystem types. Following this approach, a regional land data assimilation system was designed to produce the updated variables for each land cover by using one grid-scale observation. Taking into account the grid heterogeneity is central to this methodology and represented the main justification for including vegetation patches in the model and in the assimilation scheme.

The regional LDAS system is being extended at a global scale (LDAS-Monde, see below). The assimilation of FAPAR is being tested, together with alternative assimilation techniques (e.g. EnKF).

Next development steps will be to evaluated (1) the assimilation of new satellite products (surface albedo, land surface temperature, ...), (2) the use of new capabilities of the SURFEX modeling platform (e.g. multilayer soil model).

4. SURFEX LDAS-MONDE

Barbu et al. (2014) have implemented the LDAS over France (LDAS-France). Since SURFEX is designed to work at a global scale, the LDAS chain can be extended to larger scale domains (Mediterranean, European or global scale). The ISBA-A_{gs} model has been already used in several applications at large scale domain at resolutions provided by atmospheric forcing fields such as ERA-Interim. For example, Szczypta et al. (2012) used the ERA-Interim atmospheric reanalysis in a study over Europe to drive the ISBA model at a spatial resolution of 0.5°, corresponding to 8142 land grid cells over the considered area.

Therefore application of data assimilation to larger domains is envisaged in the future using the same LDAS configuration as described in this document.

Specific characteristics of LDAS-Monde are related to the atmospheric variables, given by ERA-Interim (Table 4). ERA-Interim products on the MARS archiving system of ECMWF are normally updated once per month, allowing a two-month delay for quality assurance.

Inputs	Meteorological forcing: ERA-Interim Physiographic field (from ECOCLIMAP II, HSWD) Copernicus Global Land GEOV1 LAI product. Copernicus Global Land ASCAT SWI-001 surface soil moisture.
Outputs	Analyzed LAI, Root-zone soil moisture (WG2), NEE, GPP, Evapo-transpiration
Interfaces	Copernicus Global Land service (http://land.copernicus.eu/global) ECMWF operational ERA-Interim production (MARS).
Capacity requirements	Supercomputer. About 500 Go per year (including output products, inputs and auxiliary variables).

Table 4: Summary of LDAS-Monde characteristics

5. REFERENCES

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Pignatelli, B., and Martin, E.: From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations, *Hydrol. Earth Syst. Sci.*, 12, 1323–1337, doi:10.5194/hess-12-1323-2008, 2008

Albergel, C., Calvet, J.-C., Gibelin, A.-L., Lafont, S., Roujean, J.-L., Berne, C., Traullé, O., and Fritz, N.: Observed and modelled ecosystem respiration and gross primary production of a grassland in south-western France, *Biogeosciences*, 7, 1657–1668, doi:10.5194/bg-7-1657-2010, 2010.

Andersson, E.: Modelling of innovation statistics, Pp. 153–164 of *Proceedings of Workshop on recent developments in data assimilation for atmosphere and ocean*, ECMWF, Reading, UK, 2003.

Balsamo, G., P. Viterbo, A. Beljaars, B. van den Hurk, M. Hirschi, A.K. Betts, and K. Scipal, 2009: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. *J. Hydrometeorol.*, 10, 623-643.

Balsamo, G., S. Boussetta, E. Dutra, A. Beljaars, P. Viterbo, B. Van den Hurk, 2011: Evolution of land surface processes in the IFS, *ECMWF Newsletter*, 127, 17-22.

Barbu, A. L., Calvet, J.-C., Mahfouf, J.-F., Albergel, C., and Lafont, S.: Assimilation of Soil Wetness Index and Leaf Area Index into the ISBA-A-gs land surface model: grassland case study, *Biogeosciences*, 8, 1971–1986, doi:10.5194/bg-8-1971-2011, 2011.

Barbu, A. L., Calvet, J.-C., Mahfouf, J.-F., and Lafont, S.: Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: a land data assimilation application over France, *Hydrol. Earth Syst. Sci.*, 18, 173-192, doi:10.5194/hess-18-173-2014, 2014.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Weiss, M., Samain, O., Roujean, J. L., and Leroy, M.: LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. Part 1: Principles of the algorithm, *Remote Sens. Environ.*, 110, 275–286, 2007.

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmarad, H., Pacholczyk, P., and Smetse, B.: GEOV1: LAI, FAPAR essential climate variables and FCOVER global time

series capitalizing over existing products. Part 1: Principles of development and production, *Remote Sens. Environ.*, doi:10.1016/j.rse.2012.12.027, 137, 299-309, 2013.

Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., Figa, J., and Anderson, C.: Initial soil moisture retrievals from the METOP-A advanced Scatterometer (ASCAT), *Geophys. Res. Lett.*, 34, L20401, doi:10.1029/2007GL031088, 2007.

Bolten, J. D. and Crow, W. T.: Improved prediction of quasi-global vegetation conditions using remotely sensed surface soil moisture, *Geophys. Res. Lett.*, 39, L19406, doi:10.1029/2012GL053470, 2012.

Boone, A., Calvet, J.-C., and Noilhan, J.: Inclusion of a third soil layer in a land surface scheme using the force-restore method, *J. Appl. Meteor.*, 38, 1611–1630, 1999.

Boussetta, S., Balsamo, G., Beljaars, A., Kral, T. & Jarlan, L. (2013a). Impact of a satellite-derived Leaf Area Index monthly climatology in a global Numerical Weather Prediction model, *Int. J. Rem. Sens.*, 34 (9-10), 3520-3542.

Boussetta, S. et al., (2013b). Natural land carbon dioxide exchanges in the ECMWF Integrated Forecasting System: Implementation and Offline validation. *J. Geophys. Res.* 118, 1-24. doi:10.1002/jgrd.50488.

Boussetta, S., Balsamo, G., Beljaars, A., Dutra, E., Albergel, C., (2014). Analysis of surface albedo and Leaf Area Index from satellite observations and their impact on numerical weather prediction. *Remote. Sens. Env.*, submitted.

Brut, A., Rudiger, C., Lafont, S., Roujean, J.-L., Calvet, J.-C., Jarlan, L., Gibelin, A.-L., Albergel, C., Le Moigne, P., Soussana, J.-F., Klumpp, K., Guyon, D., Wigneron, J.-P., and Ceschia, E.: Modelling LAI at a regional scale with ISBA-A-gs: comparison with satellite-derived LAI over southwestern France, *Biogeosciences*, 6, 1389–1404, doi:10.5194/bg-6-1389-2009, 2009.

Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, 126, 1719–1724, 1998.

Calvet, J.-C.: Investigating soil and atmospheric plant water stress using physiological and micrometeorological data sets, *Agr. Forest Meteorol.*, 103, 229–247, 2000.

Calvet, J.-C. and Soussana, J.-F.: Modelling CO₂-enrichment effects using an interactive vegetation SVAT scheme, 2001.

Calvet, J.-C., Noilhan, J., Roujean, J.-L., Bessemoulin, P., Cabelguenne, M., Olioso, A., and Wigneron, J.-P.: An interactive vegetation SVAT model tested against data from six contrasting sites *Agr. Forest Meteorol.*, 92, 73–95, 1998.

Calvet, J.-C., Rivalland, V., Picon-Cochard, C., and Guelh, J.-M.: Modelling forest transpiration and CO₂ fluxes-response to soil moisture stress, *Agr. Forest Meteorol.*, 124, 143–156, doi:10.1016/j.agrformet.2004.01.007, 2004.

Calvet, J.-C., Lafont, S., Cloppet, E., Souverain, F., Badeau, V., and Le Bas, C.: Use of agricultural statistics to verify the interannual variability in land surface models: a case study over France with ISBA-A-gs, *Geosci. Model Dev.*, 5, 37–54, doi:10.5194/gmd-5-37-2012, 2012.

Camacho, F., Cernicharo, J., Lacaze, R., Baret, F., and Weiss, M.: GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference product, *Remote Sens. Environ.*, doi:10.1016/j.rse.2013.02.030, 137, 310-329, 2013.

De Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C., and Isaksen, L.: A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF, *Q. J. Roy. Meteorol. Soc.*, doi:10.1002/qj.2023, 2012.

De Rosnay P., G. Balsamo, C. Albergel J. Muñoz-Sabater and L. Isaksen: Initialisation of land surface variables for Numerical Weather Prediction, *Surveys in Geophysics*, 35(3), pp 607-621, 2014 doi: 10.1007/s10712-012-9207-x

De Jeu, R. A. M., Wagner, W., Holmes, T. R. H., Dolman, A. J., van de Giesen, N. C., and Friesen, J.: Global soil moisture patterns observed by space borne microwave radiometers and scatterometers, *Surv. Geophys.*, 29, 399–420, 2008.

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Pauwels, V. R. N., and Verhoest, N. E. C.: Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter, *Water Resour. Res.*, 43, W09410, doi:10.1029/2006WR005449, 2007. Demarty, J., Chevallier, F., Friend, A. D., Viovy, N., Piao, S., and Cias, P.: Assimilation of global MODIS leaf area index retrievals within a terrestrial biosphere model, *Geophys. Res. Lett.*, 34, L15402, doi:10.1029/2007GL030014, 2007.

Dharssi, I., Bovis, K. J., Macpherson, B., and Jones, C. P.: Operational assimilation of ASCAT surface soil wetness at the Met Office, *Hydrol. Earth Syst. Sci.*, 15, 2729–2746, doi:10.5194/hess-15-2729-2011, 2011.

Draper, C. S., Mahfouf, J.-F., and Walker, W.: An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme, *J. Geophys. Res.*, 114, D20104, doi:10.1029/2008JDO/1650, 2009.

Draper, C., Mahfouf, J.-F., Calvet, J.-C., Martin, E., and Wagner, W.: Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France, *Hydrol. Earth Syst. Sci.*, 15, 3829–3841, doi:10.5194/hess-15-3829-2011, 2011.

Drusch, D., D. Vasiljevic, and P. Viterbo, 2004a: ECMWF's global snow analysis: Assessment and revision based on satellite observations. *J. Appl. Meteor.*, 43, 1282–1294.

Drusch, D., K. Scipal, P. de Rosnay, G. Balsamo, E. Andersson, P. Bougeault, P. Viterbo, 2009: Towards a Kalman filter based soil moisture analysis system for the operational ECMWF Integrated Forecast System, *Geophys. Res. Lett.*, 36, L10401, doi:10.1029/2009GL037716.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 10 143–10 162, 1994.

Fang, H., Jiang, C., Li, W., Wei, S., Baret, F., Chen, J. M., Garcia-Haro, J., Liang, S., Liu, R., and Myneni, R., Pinty, B., Xiao, Z., and Zhu, Z.: Characterization and intercomparison of global moderate resolution leaf area index (LAI) products: Analysis of climatologies and theoretical uncertainties, *J. Geophys. Res.-Biogeo.*, 118, 1–20, doi:10.1002/jgrg.20051, 2013.

Fairbairn, D., Barbu, A.L., Mahfouf, J.-F., Calvet, J.-C., and Gelati, E.: Comparing the Ensemble and Extended Kalman filters for in situ soil moisture assimilation with contrasting soil conditions, *Hydrol. Earth Syst. Sci.*, submitted, 2015.

Faroux, S., Kaptue Tchuenta, A. T., Roujean, J.-L., Masson, V., Martin, E., and Le Moigne, P.: ECOCLIMAP-II/Europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models, *Geosci. Model Dev.*, 6, 563–582, doi:10.5194/gmd-6-563-2013, 2013.

Garrigues S., Lacaze, R., Baret, F., Morisette, J. T., Weiss, M., Nickeson, J. E., Fernandes, R., Plummer, S., Shabanov, N. V., Myneni, R. B., Knyazikhin, Y., & Yang, W. (2008). Validation and intercomparison of global Leaf Area Index products derived from remote sensing data. *J. Geophys. Res.*, 113, G02028, doi:10.1029/2007JG000635.

Geiger, B. & Samain, O. (2004). Albedo determination, Algorithm Theoretical Basis Document, of the CYCLOPES project, version 2.0, 25p. Météo-France/CNRM.

Gu, Y., Belair, S., Mahfouf, J. F., and Deblonde, G.: Optimal interpolation analysis of leaf area index using MODIS data, *Remote Sens. Environ.*, 104, 283–296, doi:10.1016/j.rse.2006.04.021, 2006.

Habets, F., Boone, A., Champeaux, J.-L., Etchevers, P., Franchist eguy, L., Leblois, E., Ledoux, E., Le Moigne, P., Martin, E., Morel, S., Noilhan, J., and Quintana-Segui, P.: The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, *J. Geophys. Res.*, 113, D06113, doi:10.1029/2007JD008548, 2008.

Hagolle, O., Lobo, A., Maisongrande, P., Cabot, F., Duchemin, B., de Pereyra, A. (2004). Quality assessment and improvement of temporally composited products of remotely sensed imagery by combination of VEGETATION 1 and 2 images. *Rem. Sens. Envir*, 94, 172-186.

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., and Goodrich, D. C.: Integration of soil moisture remote sensing and hydrologic modelling using data assimilation, *Water Resour. Res.*, 34, 3405–3420, doi:10.1029/1998WR900001, 1998.

IPCC, *Climate Change 2001: The Scientific Basis—Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, edited by J. T. Houghton et al., 881 pp., Cambridge Univ. Press, New York, 2001.

Jarlan, L., Balsamo, G., Lafont, S., Beljaars, A., Calvet, J.-C., and Mougin, E.: Analysis of leaf area index in the ECMWF land surface model and impact on latent heat on carbon fluxes: application to West Africa, *J. Geophys. Res.*, 113, D24117, doi:10.1029/2007JD009370, 2008.

Kidd, R., Makhmara, H., and Paulik, C. : BioPar Product User Manual Soil Water Index (SWI) – Version 2.0, GEOLAND2, 23 pp., available at: <http://web.vgt.vito.be/documents/BioPar/g2-BP-RP-BP053-ProductUserManual-SWI-V2-I2.01.pdf>, 2012.

Koster, R. D. and Suarez, M. J.: Modeling the land surface boundary in climate models as a composite of independent vegetation stands, *J. Geophys. Res.*, 97, 2697–2715, 1992.

Lafont, S., Zhao, Y., Calvet, J.-C., Peylin, P., Ciais, P., Maignan, F., and Weiss, M.: Modelling LAI, surface water and carbon fluxes at high-resolution over France: comparison of ISBA-A-gs and ORCHIDEE, *Biogeosciences*, 9, 439–456, doi:10.5194/bg-9-439-2012, 2012.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Young, L., & Merchant, J. W. (2000). Development of a global land cover characteristics database and IGB6 DISCover from the 1km AVHRR data. *Int. J. Rem. Sens.*, 21, 1303–1330.

Mahfouf, J.-F.: Influence of physical processes on the tangent-linear approximation, *Tellus*, 51, A, 147-166, 1999.

Mahfouf, J.-F.: Assimilation of satellite-derived soil moisture from ASCAT in a limited-area NWP model, *Q. J. Roy. Meteor. Soc.*, 136, 784–798, doi:10.1002/qj.602, 2010.

Mahfouf, J.-F., Bergaoui, K., Draper, C., Bouyssel, F., Taillefer, F., and Taseva, L.: A comparison of two off-line soil analysis schemes for assimilation of screen level observations, *J. Geophys. Res.*, 114, D08105, doi:10.1029/2008JD011077, 2009.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essauouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, *Geosci. Model Dev.*, 6, 929–960, doi:10.5194/gmd-6-929-2013, 2013.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., and Running, S. W.: Global products of vegetation leaf area and fraction of absorbed PAR from one year of MODIS data, *Remote Sens. Environ.*, 83, 214–231, 2002.

Noilhan J. and Mahfouf, J.-F.: The ISBA land surface parameterisation scheme, *Global Planet. Change*, 13, 145–149, 1996.

Pauwels, V. R. N., Verhoest, N. E. C., De Lannoy, G. J. M., Guissard V., Lucau C, and Defourny, P.: Optimization of a coupled hydrology–crop growth model through the assimilation of observed soil moisture and leaf area index values using an ensemble Kalman filter, *Water Resour. Res.*, 43, W04421, doi:10.1029/2006WR004942, 2007.

Quintana-Segui, P., Lemoigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of near surface atmospheric variables: Validation of the SAFRAN analysis over France, *J. Appl. Meteorol. Clim.*, 47, 92–107, 2008.

Rahman, H. & Dedieu, G. (1994). SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum. *Int. J. Rem. Sens.*, 15(1): 123-143.

Reichle, R. H.: Data assimilation methods in the Earth sciences , *Adv. Water Resour.*, 31, 11, 1411-1418, doi:10.1016/j.advwatres.2008.01001, 2008.

Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, 130, 103–144, 2002.

Roujean, J.-L., Leroy, M. & Deschamps, P.-Y. (1992). A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *J. Geophys. Res.*, 97(D18), 20455-20468.

Scipal, K., Drusch, M., and Wagner, W.: Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system, *Adv. Water Resour.*, 31, 1101–1112, doi:10.1016/j.advwatres.2008.04.013, 2008.

Szczypta, C., Decharme, B., Carrer, D., Calvet, J.-C., Lafont, S., Somot, S., Faroux, S., and Martin, E.: Impact of precipitation and land biophysical variables on the simulated discharge of European and Mediterranean rivers, *Hydrol. Earth Syst. Sci.*, 16, 3351–3370, doi:10.5194/hess-16-3351-2012, 2012.

van der Molen, M. K., Dolman A. J., Ciais, P., Eglin, T., Gobron, N., Law, B. E., Meir, P., Peters, W., Phillips, O. L., Reichstein, M., Chen, T., Dekker, S. C., Doubkova, M., Friedl, M. A., Jung, M., van den Hurk, B. J. J. M., de Jeu, R. A. M., Kruijt, B., Ohta, T., Rebel, K. T., Plummer, S., Seneviratne, S. I., Sitch, S., Teuling, A. J., van der Werf, G. R., and Wang, G.: Drought and ecosystem carbon cycling, *Agr. Forest Meteorol.*, 151, 765–773, doi:10.1016/j.agrformet.2011.01.018, 2011. Wagner, W., Lemoine, G., and Rott, H.: A method for estimating soil moisture from ERS scatterometer and soil data, *Remote Sens. Environ.*, 70, 191–207, 1999.

Van der Werf G. R., J. T. Randerson, L. Giglio, G. J. Collatz, M. Mu, P. S. Kasibhatla, D. C. Morton, R. S. DeFries, Y. Jin, and T. T. van Leeuwen.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009)., *Atmos. Chem. Phys.*, 10, 11707-11735, 2010. Whitaker, J. S. and Hamill, T. M.: Ensemble Data Assimilation without Perturbed Observations, *Mon. Weather Rev.*, 130, 1913–1924, 2002.