The EUMETSAT Satellite Application Facility on Land Surface Analysis (LSA SAF)

Validation Report

Evapotranspiration (ET)

PRODUCTS: LSA-16 (MET), LSA-17 (DMET)

Reference Number:	SAF/LAND/RMI/VR/0.7
Issue/Revision Index:	Issue: Version 0.7
Last Change:	08/04/2011

DOCUMENT SIGNATURE TABLE

	Name	Date	Signature
Prepared by :	RMI Team	11/03/2010	
Approved by :	Land SAF Project Manager (IM)	11/03/2010	

DOCUMENTATION CHANGE RECORD

Issue / Revision	Date	Description:
Version 0.1	23/07/2007	Version presented at Checkpoint Meeting (04/09/2007)
Version 0.2	11/04/2008	Version presented at ORR3 (14/5/2008)
Version 0.3	10/07/2008	Improved version after ORR3
Version 0.4	01/03/2009	Version presented at OR3 (15/4/2009) and DRI (1/7/2009)
Version 0.5	11/03/2010	Version (0.4) including first validation for DMET; presented at ORR (14 April 2010)
Version 0.6	12/10/2010	Version presented after ORR 04/2010 including DMET validation and incorporating recommendation 5 and answers to RIDS (in order to declare MET operational)
Version 0.7	08/04/2011	Version (0.4) including comparison of daily ET to ground measurements; presented at OR-5 (April 2011)

DISTRIBUTION LIST

Internal Consortium Distribution		
Organisation	**Name**	**No. Copies**
IM	Pedro Viterbo	
IM	Luís Pessanha	
IM	Isabel Trigo	
IDL	Carlos da Camara	
IM	Isabel Monteiro	
IM	Sandra Coelho	
IM	Carla Barroso	
IM	Pedro Diegues	
IM	Teresa Calado	
IM	Benvinda Barbosa	
IM	Ana Veloso	
IMK	Folke-S. Olesen	
IMK	Frank Goettsche	
IMK	Ewa Kabsch	
MF	Jean-Louis Roujean	
MF	Olivier Hautecoeur	
MF	Dominique Carrer	
RMI	Françoise Meulenberghs	
RMI	Arboleda Alirio	
RMI	Nicolas Ghilain	
FMI	Niilo Siljamo	
UV	Joaquin Melia	
UV	F. Javier García Haro	
UV/EOLAB	Fernando Camacho	
UV	Aleixander Verger	

External Distribution		
Organisation	**Name**	**No. Copies**
EUMETSAT	Frédéric Gasiglia	
EUMETSAT	Dominique Faucher	
EUMETSAT	Lorenzo Sarlo	
EUMETSAT	Lothar Schueller	
EDISOFT	Teresa Cardoso	
EDISOFT	Carlos Vicente	
EDISOFT	Cleber Balan	
SKYSOFT	Rui Alves	
SKYSOFT	João Canário	

Steering Group Distribution		
Nominated by:	**Name**	**No. Copies**
IM	Carlos Direitinho Tavares	
EUMETSAT	Lorenzo Sarlo	
EUMETSAT	Yves Govaerts	
EUMETSAT	François Montagner	
STG/AFG (UGM)	Luigi de Leonibus	
MF	François Bouyssel	
RMI	Steven Dewitte	
FMI	Tapio Tuomi	

TABLE OF CONTENTS

Summary

The objective of VR is threefold: at first to assess the performances of proposed algorithm, secondly to evaluate the extent to which algorithm performances conforms with requirements targeted in Product Requirement Document (PRD) and finally, to identify error sources to open the way towards further improvements in next versions. This report summarizes the findings obtained during the validation of LSA-16 instantaneous MSG EvapoTranspiration (MET) and LSA-17 Daily-accumulated MSG EvapoTranspiration (DMET) products. Three validation approaches were adopted:

a) *In-situ validation*. This validation is performed by comparing the algorithm output (instantaneous and cumulated evapotranspiration) to evapotranspiration derived from measurements made at selected locations in different climatic/vegetation conditions.
b) *Models inter-comparison*. In this approach, the algorithm output is compared to the output of models recognized to produce valuable meteorological information. In the current validation report, data from ECMWF model and GLDAS was used to perform models inter-comparison.
c) *Consistencies check* The objective is to quantify the model performances based on the knowledge of intrinsic sources of error and to check the consistency with the other LSA-SAF products some of which are used as input to the MET algorithm.

Based on the results of the validation tests reported in this document and its annexes, it is concluded that overall algorithm performances are high and that LSA-SAF MET/DMET algorithms are able to reproduce the temporal evolution of evapotranspiration (ET), with values equivalent to observations. For estimates flagged 'Nominal' or 'Below Nominal', the PRD quality criterion is satisfied to a rate higher than 70%. Good agreement is found for stations at which close correspondence exist between ECOCLIMAP and station land cover, with the best scores for stations over grass and mixed forests.

From the inter-comparison exercise it is concluded that MET estimates are in agreement with ECMWF and GLDAS ET estimates with a spatial correlation ranging between 85% and 95% for midday images through the studied period. For high co-zenithal angles better correlation is found with ECMWF while for low angles (spring/late autumn and morning/evening) with GLDAS. Observed discrepancies between models estimates are not systematic and can be explained in terms of differences in input variables and model parameterisation. The comparison of Tskin/LST morning heating rates highlighted regions where improvements related to soil moisture and/or vegetation parameterisation are still possible. Concerning the validation of the Daily ET (DMET) product, a first attempt to assess the product accuracy confirmed the results obtained during the validation of instantaneous product; it is that spatial correlation between images remains high (between 85 and 95%), with the highest scores over Europe.

It has been identified that the main sources of differences for the in-situ validation as well as for models inter-comparisons are related to differences in input variables, land cover definition and models parameterisation. Any further improvements of the MET algorithm will have positive impact on both MET and DMET products.

Introduction

The purpose of this document is to provide to the reader detailed information about the validation and capabilities of the LSA-SAF MET product version 4.0, described in the Product User Manual (PUM), Gellens-Meulenberghs et al. (2007, 2008, 2009) and Ghilain et al. (2011). Version 4.0 of LSA-SAF MET algorithm was integrated in the LSA-SAF operational system in June 2008 to replace version 03 (see annex A for a short algorithm evolution history).

The product characteristics, as listed in the PUM and in the PRD, are:
Variable: ET
Unit : mm/h
Temporal resolution: 30 minutes
Spatial resolution: MSG resolution
Uncertainty:
Optimum: uncertainty = 10%
Nominal (target accuracy): if (ET>0.4mm/h) uncertainty = 25%
if (ET<0.4mm/h) uncertainty = 0.1 mm/h
Minimum: uncertainty = 30%

LSA-SAF MET algorithm is one of the firsts that intent to derive operationally in quasi-real time ET over large area (i.e. Europe, Africa and South America) in the context of remote sensing. Using radiation variables derived from remote sensing, it allows deriving ET at a high temporal scale, with MSG resolution and over a large domain. The methodology proposed in this framework follows a model-based approach, driven by remote sensing derived variables: radiations terms and land-use/vegetation database. On a long-term basis, this vision that takes advantage of both satellite information and modelling improvements insures: 1) applicability of the method at pixel scale over a large domain and in different climatic regions; 2) an improved accuracy for successive versions of the product through progress in physical modelling and parameterization, growing accuracy of remotely sensed products already used and assimilation of new source of information from remote sensing.

The ET algorithm produces instantaneous evapotranspiration estimates (MET) over four regions defined inside the MSG disk. In many research areas like hydrology, agriculture, ecology, climate studies and water management, the main concern is not instantaneous but cumulated values over days, months or longer periods. The objective of developing a daily evapotranspiration product (DMET) is to provide above-mentioned potential users with evapotranspiration estimates on daily basis, which in turn can easily be accumulated for longer periods.

The added-value of the products relies on the principles and vision developed in the previous paragraph: 1) applicability on a large domain, while empirical ET algorithm would not be applicable; 2) output at meso-scale spatial resolution with the same methodology over continents; 3) use of up-to-date quasi-real time remotely sensed products, like radiation terms, to insure a close follow-up of the ET evolution along the day without been affected by forecast errors; 4) an extensively validated product through information available in-situ.

The document is structured in 5 sections:

1. Section 1 focuses on in-situ validation of LSA-SAF MET/DMET estimates;
2. Section 2 deals with the inter-comparison with other sensors/products;
3. Section 3 is dedicated to study the performance and the limitations of the LSA-SAF MET algorithm, intrinsic or able to be improved;
4. Section 4 is a user-oriented summary from an expert point of view on product's maturity and on the use/application domain;
5. Section 5 addresses further work and planned solutions for the problems/limitations raised in Section 3.

The present Validation Report concerns the LSA SAF MET v4.0 and the. LSA SAF DMET v01 algorithms. Images over the full area covered by MSG have been re-processed at RMI for the year 2007 (March to December 2007, from 06:00 UTC to 19:00 UTC), using original forcing files from Land SAF archives.

1. Comparison with ground reference

This validation is performed by comparing the ET algorithm output to evapotranspiration derived from measurements realized at meteorological stations. The selection of validation sites was operated on basis of data availability, representative ness of a particular land cover and/or climatic conditions (references to used stations are included in annex B). It is important to notice that there exist neither a direct measurement method nor universal reference for ET and surface turbulent fluxes. The methods involved for the ground estimation of surface turbulent exchanges rely on theoretical assumptions that are not always verified (for example, homogeneity assumption for Monin-Obukhov theory) and on models with intrinsic limitations. Today some problems remain unsolved (i.e. representativeness of the measurements, closure of the energy balance, etc). This is why, measuring accurately surface turbulent fluxes is still challenging and an active domain of research.

For the In-situ validation exercise, stations from CarboEurope, CarboAfrica, Belgian AWS network and CEOP station Cabauw were used (for a descriptions of measurement networks, stations, and assumed error, refer to annex B). This validation concerns the version 4.0 of the algorithm as implemented in the LSA-SAF system, i.e. the model of energy exchange driven by remotely sensed derived radiation terms, meteorological variables from ECMWF and vegetation characteristics/parameterisation according to ECOCLIMAP land cover database. For a complementary analysis of model performances in stand-alone mode refer to Annex C.

1.1. Methodology

Due to the spatial heterogeneity and difference of scales, the in-situ measurements will often not be representative of the same vegetation types than the MSG pixel estimates. Indeed, while station measurements are representative of a domain varying from few meters to few hundred meters around measurement site, LSA-SAF MET provides estimates averaged for MSG pixels of few kilometres resolution. A way to reduce the "foot-print" effect and to achieve consistent comparison is to retrieve the estimates computed by LSA-SAF MET algorithm at 'tile' level, computed before spatially aggregating to obtain pixel estimate. For this purpose, a validation file, containing 'tiled' computed variables for a set of pixels corresponding to validation stations location, is created simultaneously with the ET image (cfr PUM).

The basic material used includes: A) raw in-situ data, B) LSA-SAF MET/DMET images and C) validation files. Latent heat flux (W/m²) is the variable derived from measurements at stations. A simple transformation, accounting for the air temperature dependence, given in equation 3, is then applied to obtain the 'measured' ET (mm/h).

$$ET = 3600.LE / L_v \quad \text{[mm/h]} \qquad (3)$$

where $L_v = (2.50084 - 0.00234\,T_a).10^6$ is the latent heat of vaporization [J/kg], with T_a expressed in C.

The ECOCLIMAP land cover composition for the pixel containing considered station is presented in Table 1. The tile best fitting the vegetation type of the station is shaded in grey.

Station	G+	DBF+	ENF+	C3 crops	C4 crops	B&M+	BS and Urban+
Amplero	8.4%	91.6%					
Buzenol	75.7%	4.1%		20.1%			
Cabauw	85.1%			11.9%			2.9%
Hesse	29.0%	44.8%		26.1%			
Humain	36.3%	44.7%		18.9%			
Kaamanen			43.4%			45.9%	10.7%
Las Majadas	29.8%	33.4%		36.8%			
Lonzée		7.5%		82.1%			10.3%
Loobos	17.1%		47.2%	35.5%			
Monte Bondone	36.0%	26.0%	38.0%				
Puéchabon	5.3%	94.6%					
Roccarespampini		25.2%		67.6%	7.0%		
Sodankylä	3.9%		88.7%			7.2%	
Tojal		8.3%		85.3%	8.5%		
Vielsalm	41.8%	30.9%	27.2%				
Wetzstein			57.8%	40.4%			1.6%

Table 1 Vegetation composition of the MSG pixel encompassing the stations, as used by LSA-SAF MET algorithm v4.0. The tile most representative of the station surroundings is coloured in grey.

+Grassland (G), Deciduous Broadleaved Forest (DBF), Evergreen Needle Forest (ENF), B&M (Bogs and Marshes), Bare soil (BS).

As an example of differences between observations and model estimates compared at píxel and tile level is illustrated in Figure 1, for the CarboEurope-IP Vielsalm station. This station consists of a meteorological mast over a mixed forest. The measurements have been proven to be representative of the mixed forest for most of meteorological conditions. At MSG scale, the pixel encompassing Vielsalm is composed by 58.2% of forest, and 41.8% of grass. Hence, it implies that the estimate at pixel scale given by LSA SAF MET v4.0 algorithm is not comparable to measurements and that 'tile' estimates compare better to observations.

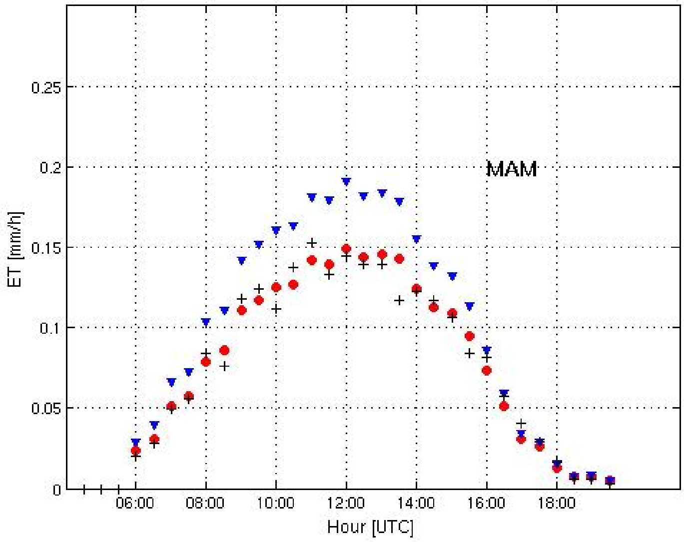


Figure 1 Mean seasonal diurnal cycle of ET for the CarboEurope-IP Vielsalm station for March-April-May. Observation (+), pixel (v) and 'tile' (o) estimate from LSA SAF MET v4.0 algorithm are represented.

1.2. Instantaneous product (MET)

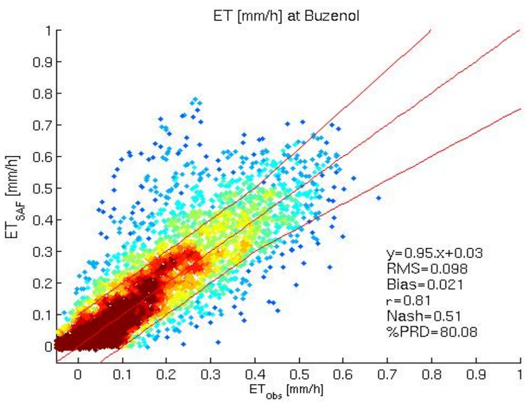
For the sake of clarity, the comparisons have been ordered according to the dominant at station vegetation types (grassland, deciduous/evergreen broad/needle leaved forest, bogs, crops and mixed forest). For each validation site, two types of plots are provided:

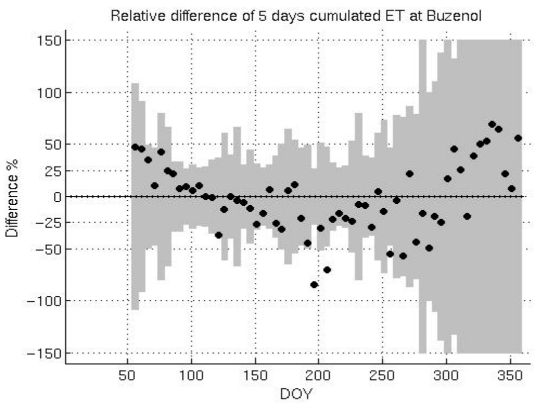
1) A density scatter plot of the 30-minutes daytime "observed" vs estimates ET, including statistical indicators of goodness of fit (regression coefficients, bias, root mean square error (RMS), correlation coefficient (r), Nash index) and percentage of cases for which the PRD quality criterion (%PRD) is satisfied. In this plot, the uncertainty boundaries (read lines at both sides of the 1-1 line) defined in the PRD are also included.
2) Time series of the relative difference between measurements and simulations of cumulated ET over 1 or 5 days depending of the length of the validation dataset. Positive relative error indicates an underestimation of ET. With this convention, highest relative differences, corresponding to low absolute values, are expected. Total error allowed by the PRD is superimposed in shaded grey areas.

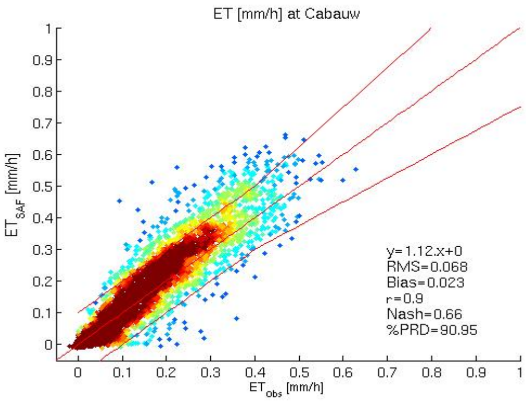
The representation of scatter density by means of colours (red=high density; blue=low density) allows a more clear interpretation of statistical indicators and the spread out of the points perpendicularly to the 1-1 line.

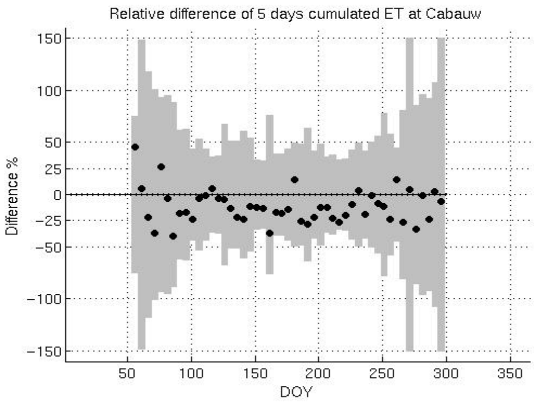
The temporal variability of the shading occurs because the PRD requirement for quality is defined in two parts: if measured ET is less than 0.4 mm/h, the accuracy threshold is an absolute value of 0.1 mm/h. For larger ET estimates, the accuracy threshold is expressed as a relative proportion of the measurement (25%). Since the y-axis of the left hand-side graphs represents the relative difference between observations and model output, the absolute accuracy threshold standing for small estimates had to be translated into relative accuracy threshold. For example, if observed ET gives 0.1 mm/h for some time step, the absolute accuracy threshold of 0.1 mm/h is translated to a relative threshold of 100%. Seasonal variations of the shaded areas are due to the variations of observed ET caused either by annual cycle of solar radiation or soil moisture stress. Temporal variability of the shaded area can also be caused by periods of missing observations (for example, higher limit are obtained when midday estimates are missing).

1. Grassland (G)









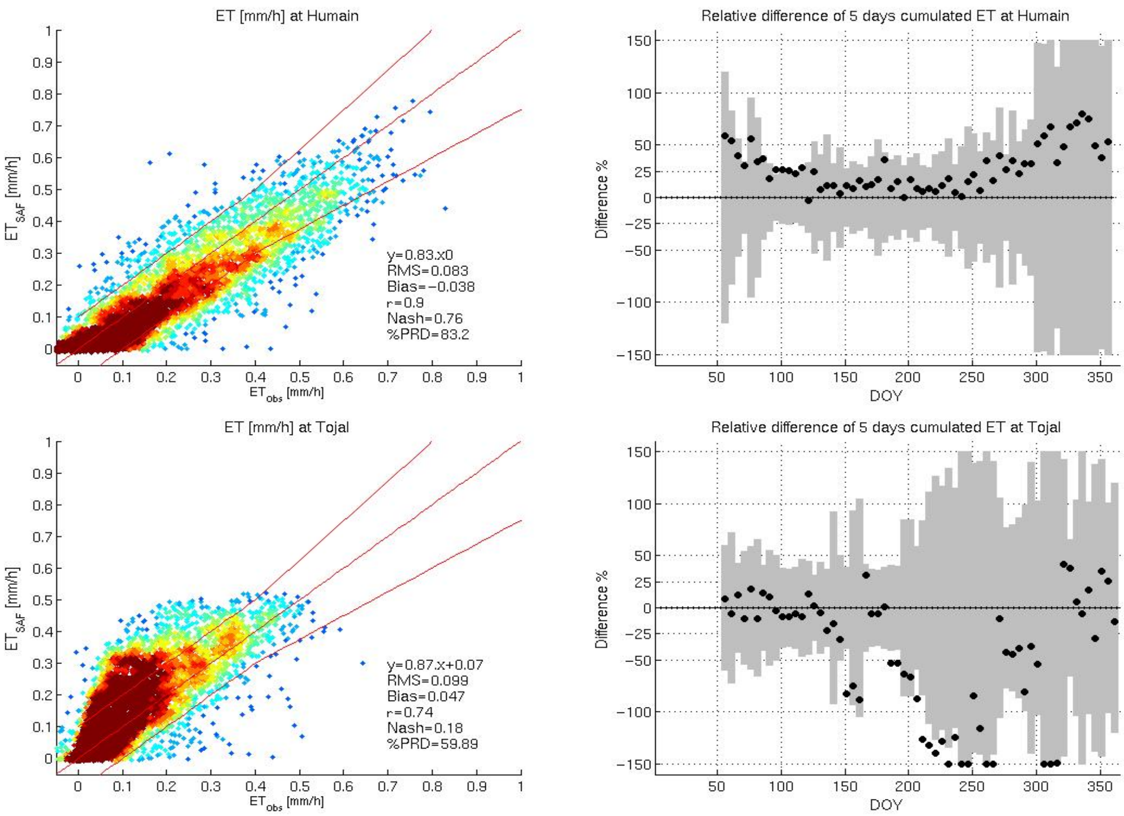


Figure 2 Comparison of LSA-SAF MET tiled estimates with in-situ measurements, converted in evapotranspiration [mm/h], over a period ranging from 1st of March to 31st of December 2007. From top to bottom: AWS Buzenol, CEOP Cabauw, AWS Humain, CarboEuroIP Tojal. Left-hand side figures show scatterplot of the half-hourly estimates, uncertainty bounds given by the PRD and statistical indices. At right, we present comparison of the evolution for the whole period (1st of March to 31st of December 2007) of the relative difference between 5-days cumulates, for which only available estimates are considered in both time series. Cumulated error allowed by the PRD is added (shaded grey area).

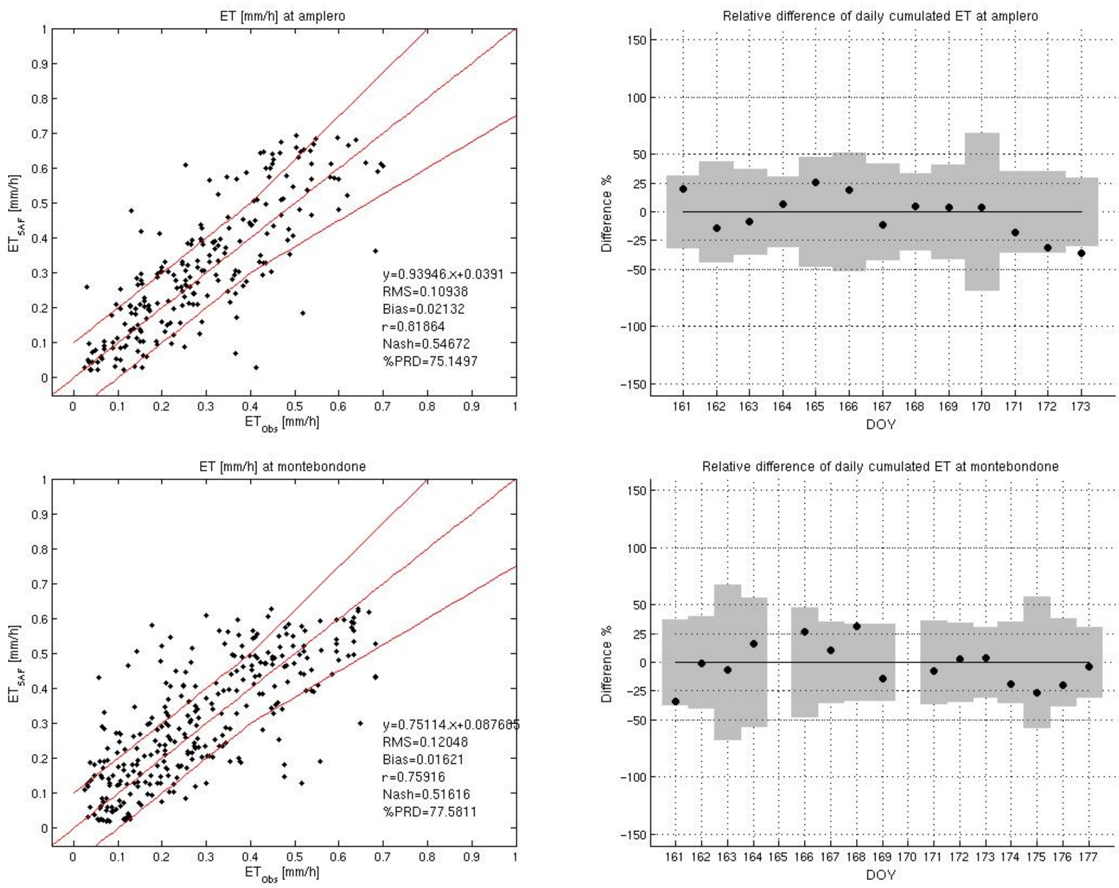


Figure 3 Same as Figure 2 but for CarboEuroIP Amplero (top) and Monte Bondone (bottom). Comparison is performed over a period ranging from 10th to 23th/27th of June 2005.

For the “grassland” ECOCLIMAP/MSG tile, the results obtained by the LSA SAF MET algorithm are close to the measurements for most of the stations considered, i.e. three temperate sites (Buzenol, Cabauw and Humain), two Italian mountainous sites (Amplero and Monte Bondone) and one Portuguese site (Tojal) composed of a mixture of C3 crops and C4 grass. Between 70% and 90% of events satisfy the PRD criterion of quality, apart for the Tojal station. No systematic bias is found when looking at the evolution of the relative error. We notice two particularities of the comparison. 1) At Tojal station, while results are of good quality for spring and autumn, a significant overestimation appears in summer. This bias will be investigated in the subsection E3, where we illustrate the soil moisture impact on ET estimates; 2) At RMI-AWS stations (Buzenol and Humain), the results are quite good, nevertheless, a seasonality is observed in the relative error. Error is smaller in Spring-Summer than in Autumn-Winter.

2. Deciduous Broadleaved Forest (DBF)

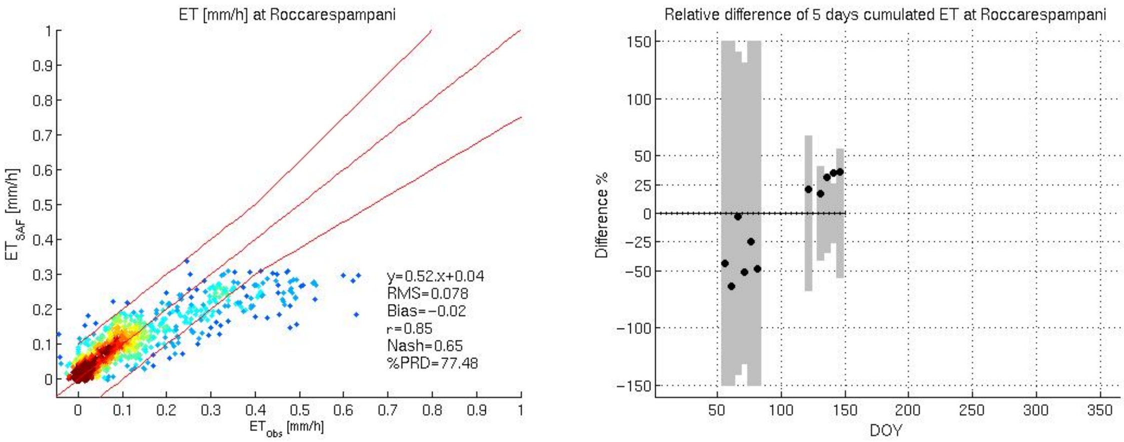


Figure 4 Same as Figure 2 but for CarboEuroIP Roccarespampani. The period of comparison ranges from 1st of March to 31st of May 2007.

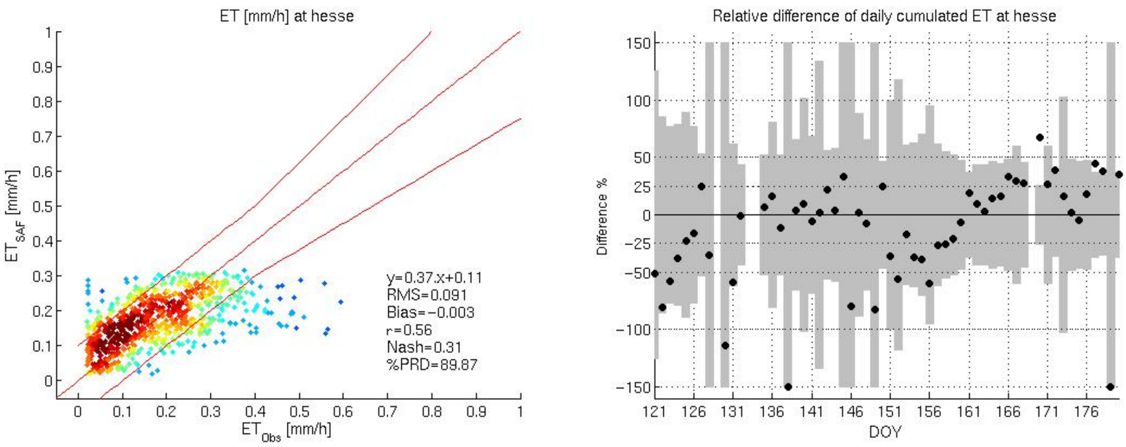


Figure 5 Same as Figure 2 but for CarboEuroIP Hesse for the period ranging from 1st of May to 30th of June 2006.

For the "deciduous broadleaved forest" ECOCLIMAP/MSG tile, the results obtained by the LSA SAF MET algorithm are satisfying the PRD criterion at 89.9% and 77% for the temperate beech forest (Hesse) and the Turkey oak forest (Roccarespampani), no systematic bias was detected. The period covered by the datasets is approximately two months. Therefore, no seasonal effect in the error (if they exist) can be analysed. For Hesse, LSA SAF MET overestimates the daily cumulates between DOY 151 to 160 and DOY 121 to 130, corresponding to the first days of May 2006 and June 2006 respectively. This difference can be attributed to the difference in vegetation description between ECOCLIMAP/MSG and at local, as investigated in the subsection E2. Indeed, while ECOCLIMAP/MSG gives a slow continuous increase of leaf area index during spring, observations shows a sharp increase after bud break generally occurring in mid-May (Granier et al., 2000). It should be also mentioned that Göckede et al. (2007) do not consider these two stations as fully homogeneous, and recommend a footprint analysis prior the use of data. However, this information is not provided with the obtained datasets.

3. Evergreen Needle Forest (ENF)

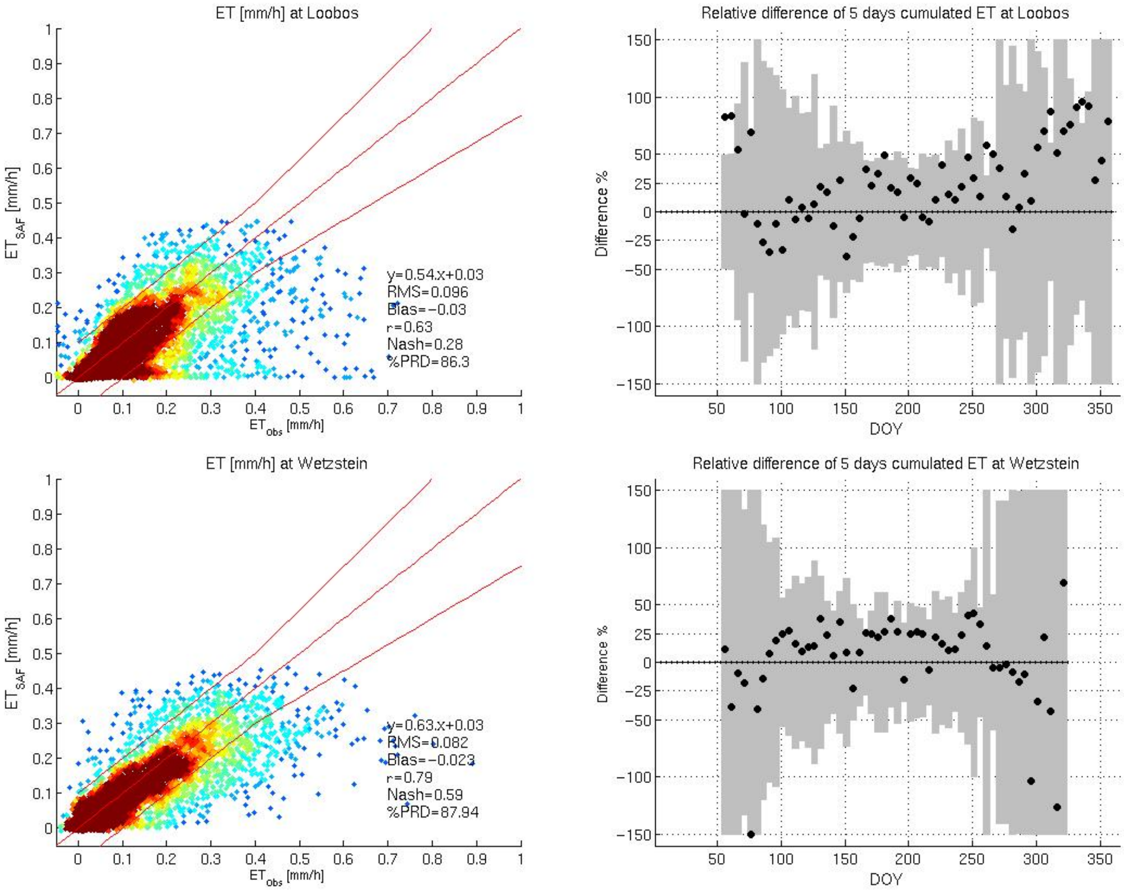


Figure 6 Same as Figure 2 but for CarboEuroIP Loobos (top) and CarboEuroIP Wetzstein (bottom) for a period ranging from 1st of March to 31st of December 2007.

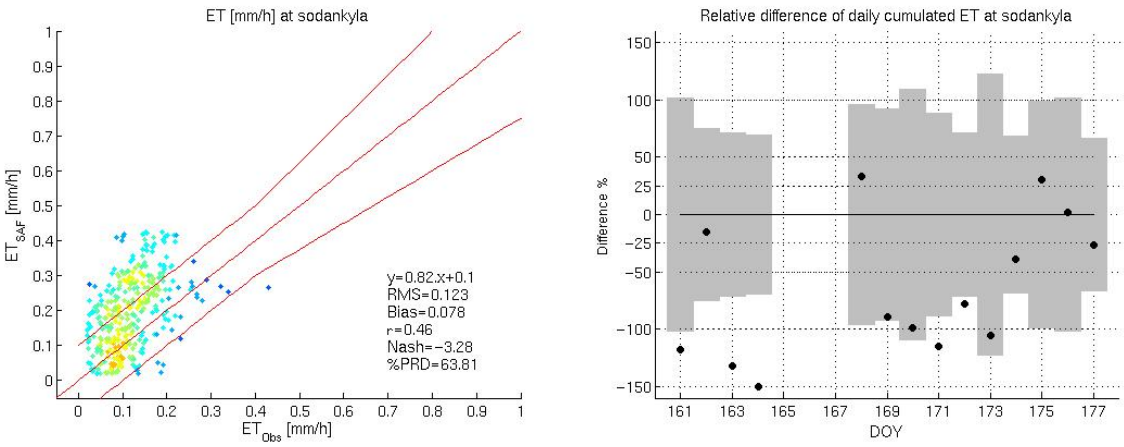


Figure 7 Same as Figure 2 but for CarboEuroIP Sodankylä for a period ranging from 10th to 27th of June 2005.

For the “evergreen broadleaved forest” ECOCLIMAP/MSG tile, the results obtained by the LSA SAF MET algorithm are satisfying the PRD criterion at 86% and 88% for the temperate sites (Loobos and Wetzstein) and 64% for the boreal site (Sodankylä). For the temperate sites, LSA SAF MET is slightly underestimated during summer. Good time consistency is found with Wetzstein and Loobos. However, the results show that LSA-SAF MET algorithm is less capable to reproduce the variability of the measurements over forested sites (RMS~0.06-0.10 mm/h) in contrast with grassland sites (RMS~0.06 mm/h). We also remark from scatter plots that, for some events, measurements give very high estimates in contrast with modelled values. These raw measurements are of dubious quality and would have to not be taken in account. However, no quality flag was attached to data by the experimentalist because of the lack in time. At Sodankylä, results are biased for this period: LSA SAF MET overestimates, but reasons are still to be investigated.

4. Crops C3 (C3)

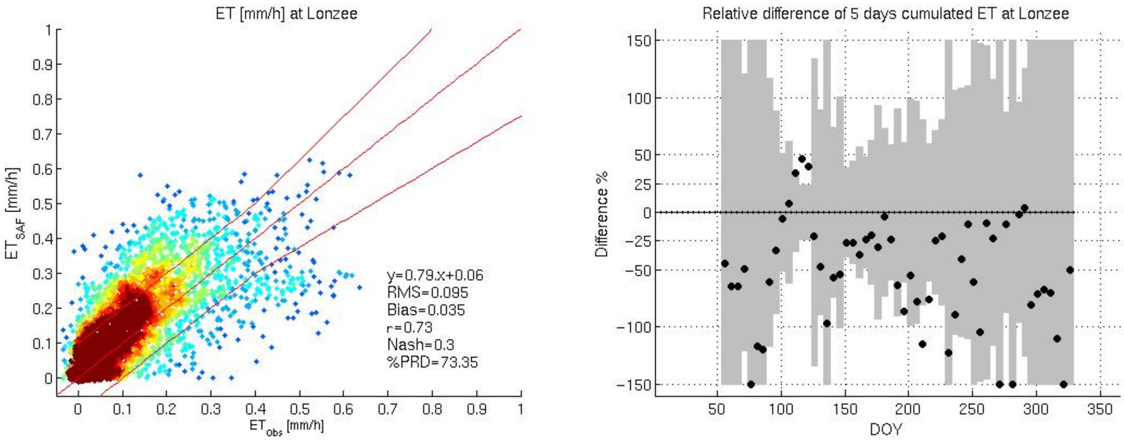


Figure 8 Same as Figure 2 but for CarboEuroIP Lonzée.

For the "C3 crops" ECOCLIMAP/MSG tile, the results obtained are satisfying the PRD criterion at 73%. However, there is a tendency to overestimate the ET (Bias=0.035) and the dispersion is larger than for grassland sites (RMS=0.095). The slope of the linear regression seems to indicate an underestimation (slope=0.79), however a clear bias (independent term of the regression = 0.06) is demonstrating the converse. Results obtained at Lonzée are strongly affected by the differences in spatial scales. It is a typical case for which vegetation description at MSG scale does not match with local vegetation characteristics. Indeed, at this station, we are comparing "C3 temperate crops" estimates with measurements over a winter wheat parcel. Therefore, on the basis of the direct comparison, at this stage, it remains difficult to draw any conclusion. The vegetation characteristic mismatches and the related impacts on ET are investigated in subsection E2.

5. Evergreen Mediterranean Forest

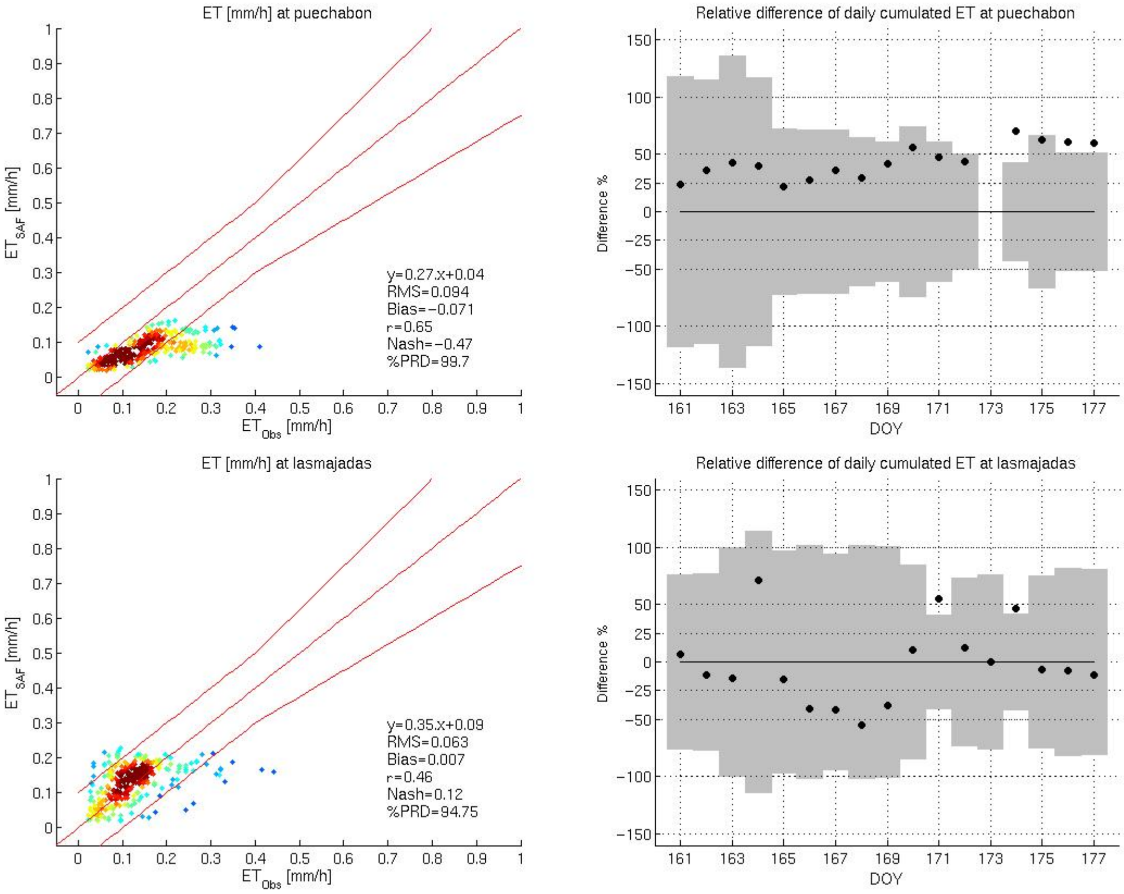


Figure 9 Same as Figure 2 but for CarboEuroIP Puéchabon and Las Majadas del Tietar.

In Western Europe, the "evergreen broadleaved forest" ECOCLIMAP/MSG tile is not existent. We therefore used the tile 'deciduous broadleaved forest' instead. The results obtained are contrasted. PRD quality criterion is fulfilled at 99% for the Mediterranean forest (Puéchabon) and 95% for the Dehesa (Las Majadas del Tietar). Indeed the point density is high around the 1-1 line of the scatter plot. While no systematic bias is found for Las Majadas, an evident underestimation is observed at Puéchabon. However, this bias was not detected when we performed an off-line validation of the algorithm on a long time series of nearly two years (p.68). Longer time series for this station will probably confirm the results obtained in the off-line validation.

6. Bogs

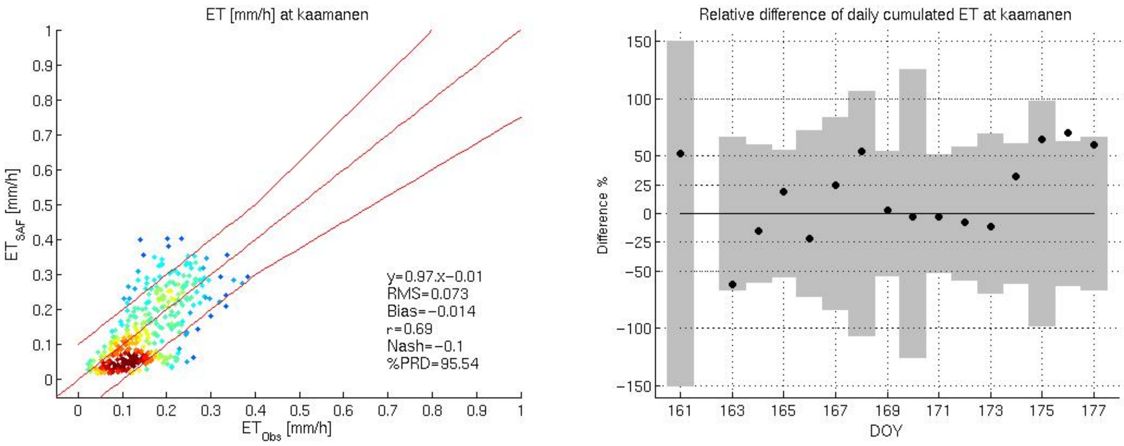


Figure 10 Same as Figure 2 but for CarboEuroIP Kaamanen.

For the “bogs” ECOCLIMAP/MSG tile, the results obtained from the comparison at Kaamanen station, an Aapa mire of Northern Finland, for a time period of 17 days in June 2005 gives satisfactorily results in term of PRD criterion fulfilment (95%) and bias (no bias observed). For DOY 161, 168, 174 to 177, for which larger relative differences are observed, LSA SAF ALBEDO was missing and ECOCLIMAP was used instead, degrading the quality flag to “Medium Quality”.

7.Mixed Forest

Definition from ECOCLIMAP of mixed forest for this pixel: 50% DBF and 50% ENF.

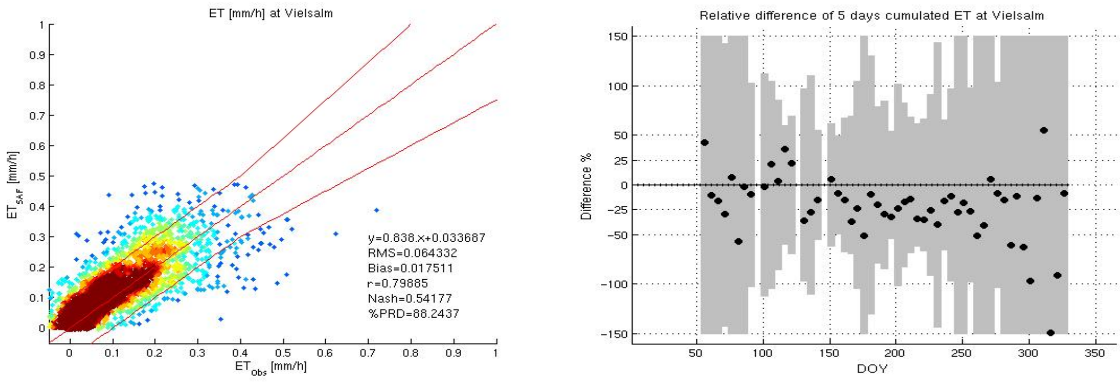


Figure 11 Same as Figure 2 but for CarboEuroIP Vielsalm.

Mixed forest does not correspond to an ECOCLIMAP/MSG tile, but can be composed by adding the contribution of both deciduous broadleaved and evergreen needle leaved trees. Following the definition of mixed forest given in the original ECOCLIMAP database, we attributes equal weight to both contribution. PRD quality criterion is fulfilled at 88%. The dispersion is quite low, indicating that LSA SAF MET is able to reproduce the evolution of ET over this forest. However, the results are slightly overestimating the measurements.

8.Sahelian savanna

One illustration showing the comparison of the total time series of latent heat flux is given in Figure 12. Looking at the comparison, we notice that the two estimates don't have the same magnitude. MET v4.0 estimates are much lower than measurements. In subsection E3, the input soil moisture of the model explains the reason why the algorithm is not able to capture the in-situ dynamics.

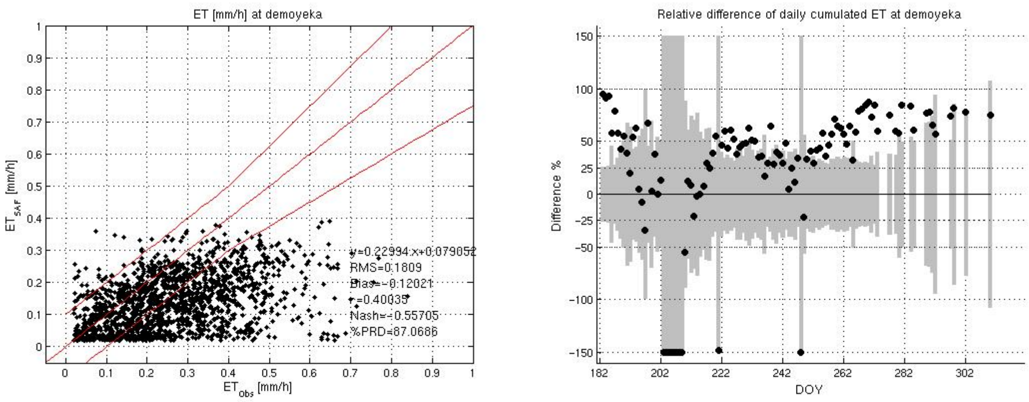


Figure 12 Same as Figure 2 but for CarboAfrica Demokeya for a period ranging from 1st of July to 30th of November 2007.

Explanation of the differences between simulation and measurements

-E1: Error caused by differences in global solar radiation

As a main driver of the MET v4.0 algorithm, global solar radiation differences can have a significant impact on the comparison of evapotranspiration. To quantify the impact, we compute the absolute (Δ) and the relative difference($\Delta_{\%}$) defined in (4), between observed global incoming radiation at the stations and LSA SAF DSSF, retrieved from the LSA-SAF archiving system.

$$\Delta_{\%} = \frac{Rg_{obs} - DSSF_{SAF}}{Rg_{obs}} \qquad (4)$$

Figure 13 illustrates the improvement of PRD criterion by removing from statistics the cases with large relative differences noted between LSA-SAF and observed global radiation ($\Delta_{\%}$). The threshold envisaged is defined by the PRD quality criterion of LSA SAF DSSF: 10% if global radiation is larger than 200 W/m², and 20W/m² if global radiation is less than 200 W/m².

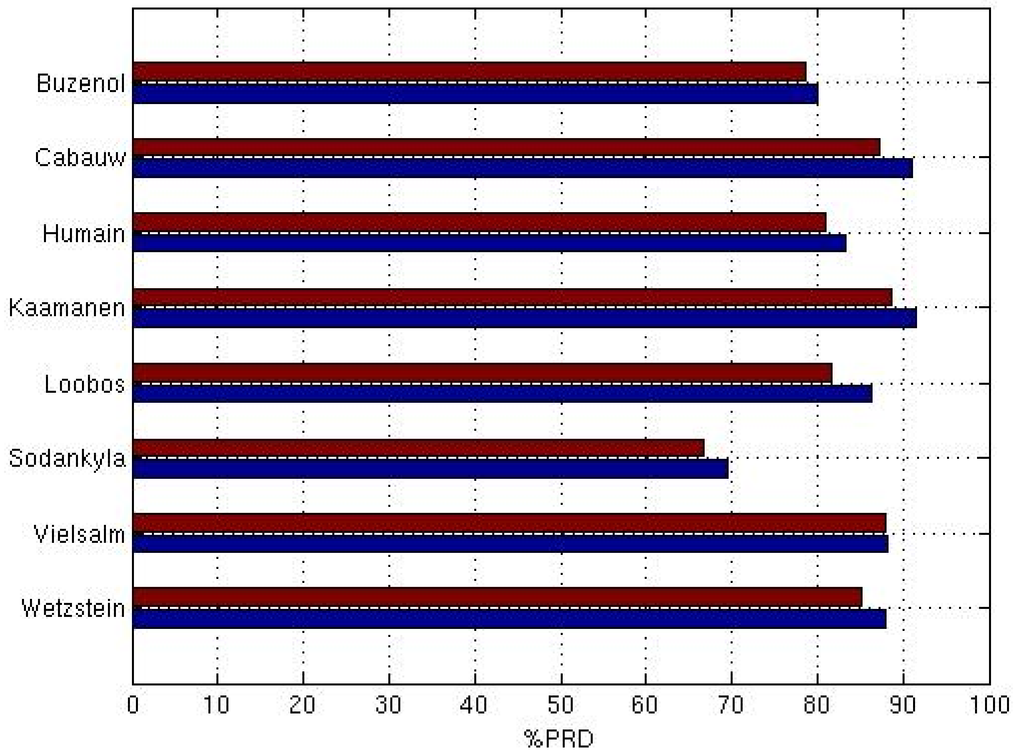


Figure 13 For eight validation stations, fulfilment percentage of the PRD quality criterion is shown for the total sample (bars in red) and for the sub-sample for which PRD criterion for LSA SAF DSSF is satisfied.

An example of improvement of the comparison for the CarboEurope-IP Wetzstein station is given at Figure 14. Most of the points beyond the limits of the PRD quality criterion have been removed, when the cut-off threshold is applied on difference in global radiation.

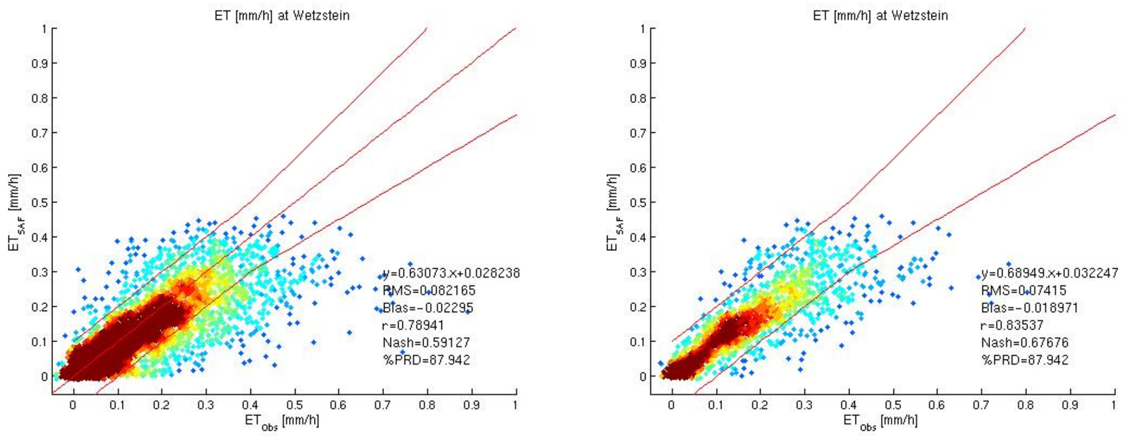


Figure 14 Scatterplot of ET [mm/h] provided by LSA SAF MET v4.0 versus measured at CarboEurope-IP Wetzstein station for the whole dataset (left) and after removing ET estimates for which difference in global radiation between LSA SAF and measurement is larger than foreseen by PRD (10% for global radiation > 200 W/m²; 20W/m² for global radiation<200 W/m²).

-E2: Error caused by differences in vegetation description

We investigate in this subsection the role of the vegetation description mismatches in the differences observed between LSA SAF MET and flux measurements. These mismatches can be of different nature: 1) the ECOCLIMAP/MSG tile does not correspond to the vegetation type of the station; 2) vegetation characteristics, i.e. leaf area index (LAI), fraction of vegetation cover (FVC) and roughness length for momentum (z0), do not correspond.

Lonzée

As explained above, comparison at Lonzée C3 crops station is not straightforward, due to the differences in observations scale. Because the cropland in Belgium is very patchy, looking at a scale of few kilometres will not be representative of the exact type of vegetation at Lonzée station, i.e. winter wheat in 2007. Therefore, part of the comparison is not relevant, because dates of growth and senescence are completely different. To evaluate the relevant periods of comparison, we have represented in Figure 15 for the year 2005 the leaf area index used by LSA SAF MET v4.0 and measured at Lonzée over winter wheat (C. Moureaux, pers. comm.). We can see that only the fully developed canopy stage can be used for validation. For 2005, it ranged from day 120 to 200. Making the hypothesis that evolution of this winter wheat parcel is the same for 2007, we can select this period for direct comparison in Figure 8. Relative difference is of the order of 25% for this period, showing a slight overestimation of the LSA SAF MET estimates.

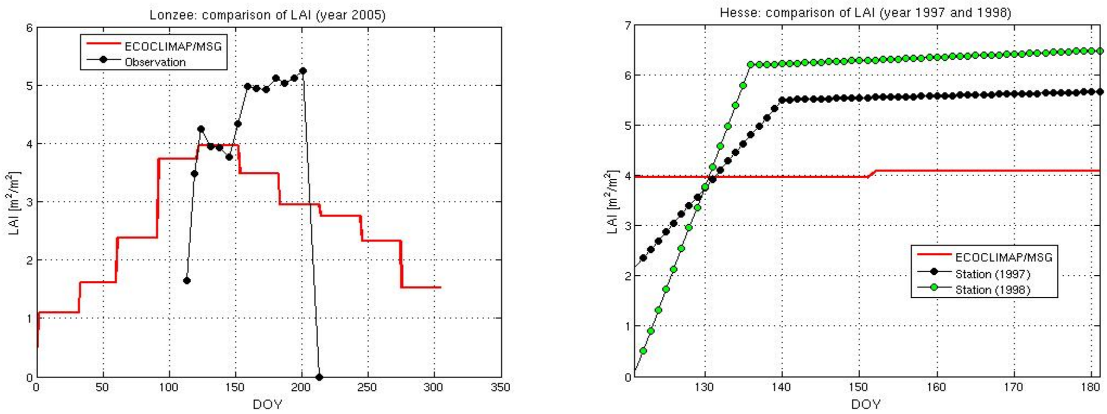


Figure 15 Left: comparison for 2005 of the leaf area index measured at Lonzée station over winter wheat and given by ECOCLIMAP/MSG for the corresponding MSG pixel for the 'C3 crops' tile. Right: comparison for 1997 and 1998 of the leaf area index interpolated from measurements (Rivalland, 2003) at Hesse station over a beech forest and given by ECOCLIMAP/MSG for the corresponding MSG pixel for the 'Deciduous Broadleaved Forest' tile.

Hesse

Comparison performed for 2 months at the Hesse site (figure 5) shows some disagreement at the beginning and at the end of the period that can be partly explained by the difference between ECOCLIMAP/MSG and in-situ vegetation description. In the right side of Figure 15, we show the daily evolution of the leaf area index given by ECOCLIMAP/MSG for the 'Deciduous broadleaved forest' at Hesse location, and the simplified interpolation of measurements for the years 1997 and 1998 (Rivalland, 2003; Granier et al., 2000). During the whole period, LAI given by ECOCLIMAP/MSG is quasi-constant, corresponding to the fully developed canopy stage. However, in-situ measurements show that during the first 20 days, leaf area index is increasing from [0-2 m²/m²] to [5.5-6.2 m²/m²]. Therefore, it is not surprising to observe ET differences between simulation and measurements for this period, for which LSA SAF MET v4.0 gives higher values due to higher foliage coverage. For the period extending from DOY 150, corresponding to June, the origin of the differences noticed between measurements and LSA SAF MET v4.0 estimates in Figure 5 is still under investigation.

-E3: Error caused by differences in soil moisture

For the Tojal station (Portugal), a clear disagreement is seen from Figure 16 during the summer season. While LSA SAF MET v4.0 gives a clear decreasing ET during this season due to the combination of water stress and decrease of vegetation cycle, observations show that the decrease in ET is still stronger. Soil water content available for evapotranspiration is suspected to be the source of the differences observed. Soil moisture is measured at the Tojal station with frequency domain reflectometer probes up to 30 cm deep (Pereira et al., 2007). Therefore, we run MET v4.0 algorithm on Tojal pixel, with the replacement of soil moisture from ECMWF by rescaled soil moisture measured at Tojal station. The results of evapotranspiration obtained are presented in Figure 16 (right and left). Clearly, the simulation of ET based on rescaled measured local soil moisture is in agreement for the whole period. Statistical measures of performance are improved.

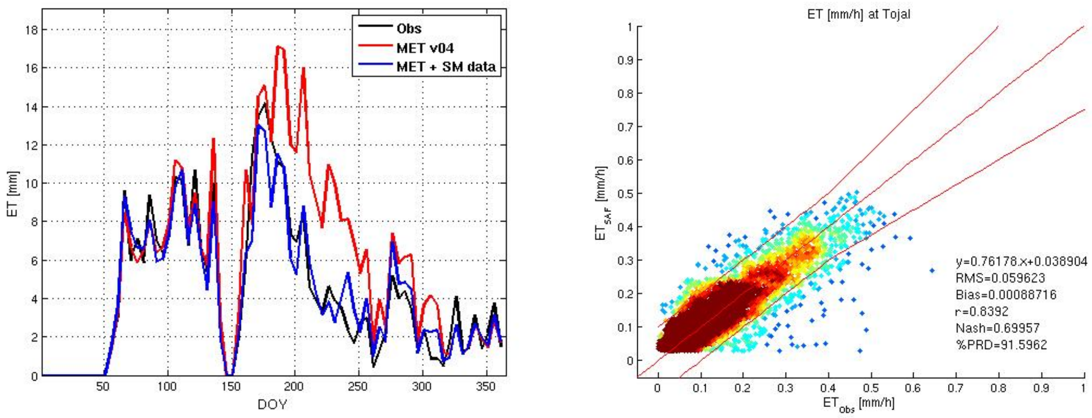


Figure 16 Left: time series of the 5-days cumulated ET measured (black) and modelled (red) for the CarboEurope-IP Tojal station (Portugal) for the period from 1st of March to 30th of November 2007 as provided by LSA SAF MET v4.0. Modelled ET using soil moisture measured at the station (blue) has been superimposed. Right: Modelled ET using soil moisture measured at the station compared to ET measured at the station at a time step of 30 minutes.

For the Demokeya station, in order to explain why the algorithm is not able to capture the in-situ dynamics, we compare in Figure 17 time series of both soil moisture used by MET v4.0, i.e. ECMWF forecasts, and rescaled soil moisture measured at the Demokeya station for the surface layer. It appears that the general dynamics respect the transition between wet and dry season. However, ECMWF soil moisture generally gives lower values and reaches more rapidly severe drought. One of the possible causes is the underestimation of the ECMWF convective rainfall rate compared to the real amount of precipitation measured at the station. That implies that the soil water reservoirs are not sufficiently recharged. This is partly due to the difference in the spatial scale between ECMWF grid and local measurements. Using observed soil moisture rescaled for the use in MET v4.0, we perform a simulation. The results shown in Figure 18 indicate that, with this observed soil moisture input, the MET algorithm is able to reproduce the observed latent heat flux time series.

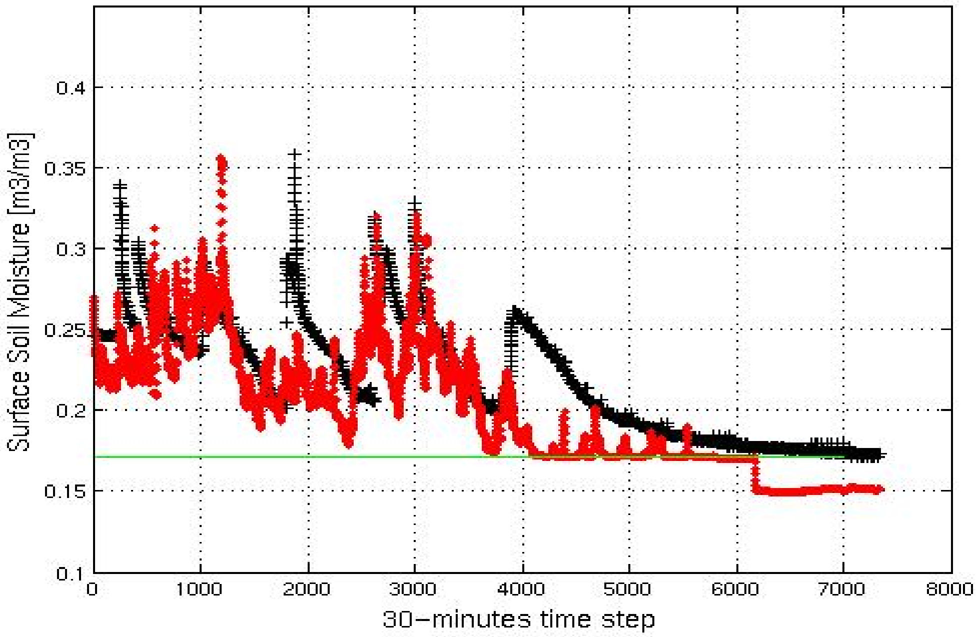


Figure 17 Comparison of surface soil moisture from ECMWF, used by MET v4.0, and rescaled soil moisture observations in the first cm in the soil for the period from July to October 2007. ECMWF soil moisture in m³/m³ is represented in red; observations are represented in black. A green line represents the ECMWF wilting point. Below this treshold no evapotranspiration can occur.

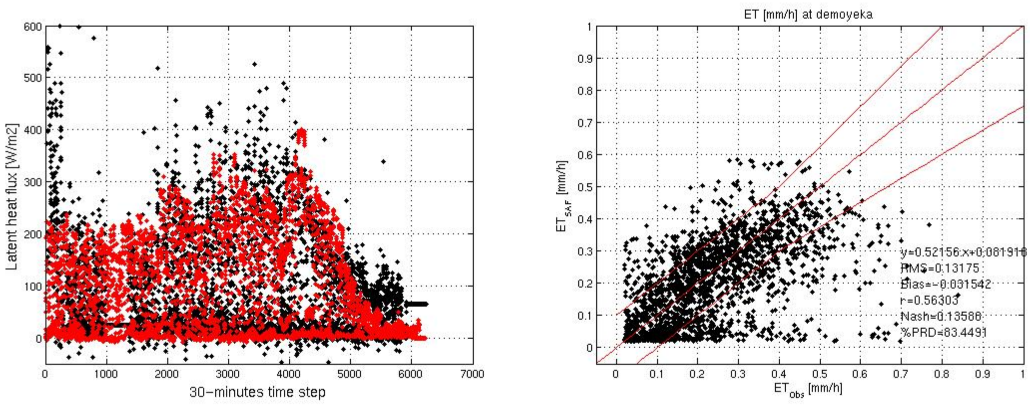


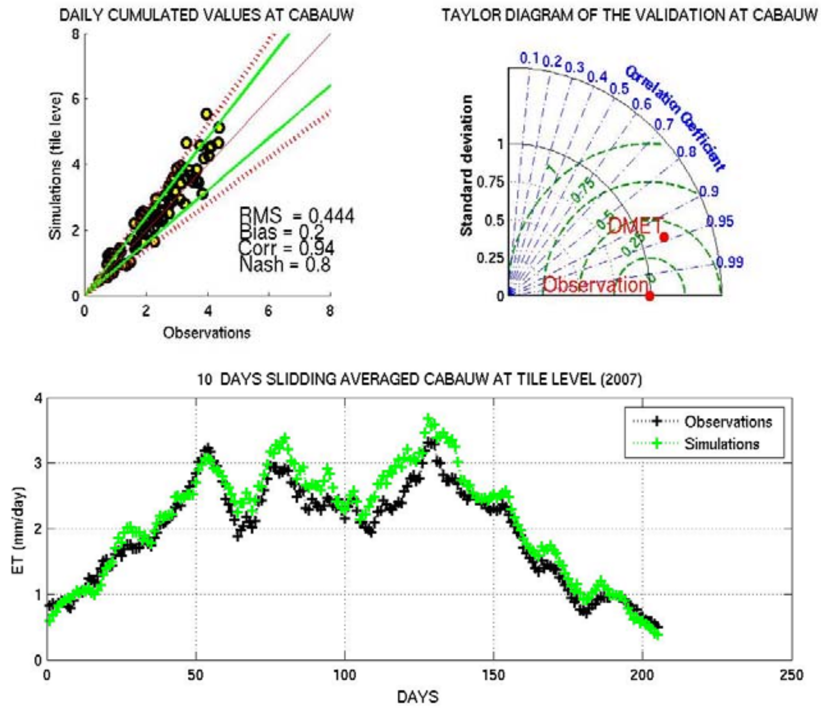
Figure 18 At left, comparison of time series of latent heat flux in W/m² at 30 minutes time step. Observations are represented in black and the simulation of MET algorithm with observed soil moisture is given in red. At right, scatterplot of ET estimates at 30-minute time step, for the same period.

1.3. Daily product (DMET)

The output of the DMET algorithm is compared to daily-cumulated ET values at selected locations that have already been used for the validation of MET product. By using the same validation stations used for MET, it is possible to show the coherence between the two products, which in turn allows the extrapolation of the conclusions from MET to DMET. Based on the results of MET validation presented in previous sections, the DMET product is expected to be as accurate as the instantaneous one. A meaningful comparison of the daily product against observations is achieved only if measurements as well as model simulations exist for the full diurnal cycle in the analyzed period. This strongly limits the number of days available for comparison since 1) instantaneous observations contain many gaps, 2) most of measurements have negative values during night time, 3) MET can be missing due for instance to the lack of an input variable. Given that gaps and negative values are very frequent in available datasets, we have replaced the negative values by 0 (no evapotranspiration assumed) and a daily ET is used for the comparison only if at least 75% of observations (36 out of 48) and simulations exist for a given day. The quality of fit between model simulation and observations has been evaluated by means of classical statistical indicators (bias, root-mean-square (RMS), correlation, Nash index). For visual interpretation, 3 types of plots have been generated (Figure 19):

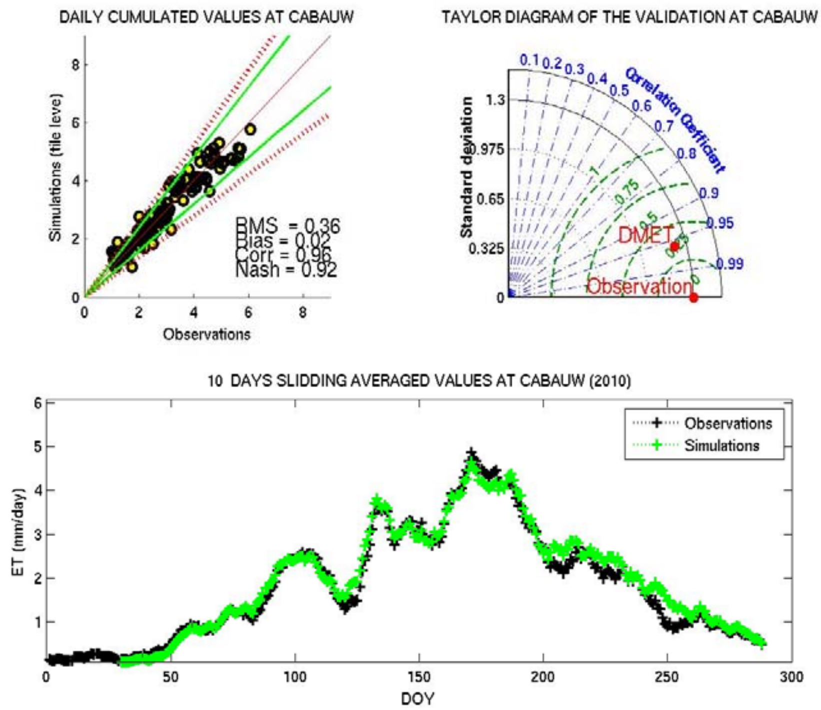
A) Ten days sliding averages, useful for detecting trends on the datasets and deficient model parameterisation on monthly/seasonal basis; it also provides some insight into possible model systematic errors.
B) Scatter plots of simulations vs. observations with superposed accuracy boundaries (target accuracy in green; threshold accuracy in red) plus statistical indicators for every station.
C) Taylor diagrams (Taylor, 2001) to summarise the statistics of the comparison in a single diagram. It is produced from statistic values given by STD (standard deviation), RMS (centred root mean square error) and COR (correlation).

a)



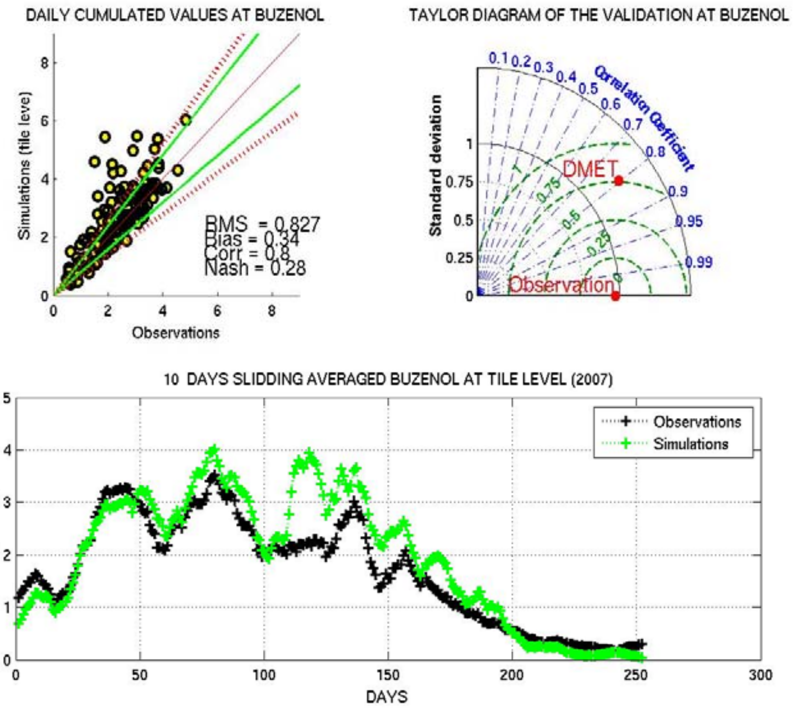
DAILY CUMULATED VALUES AT CABAUW
Simulations (tile leve)
Observations
RMS = 0.444
Bias = 0.2
Corr = 0.94
Nash = 0.8
TAYLOR DIAGRAM OF THE VALIDATION AT CABAUW
Standard deviation
Correlation Coefficient
DMET
Observation
10 DAYS SLIDDING AVERAGED CABAUW AT TILE LEVEL (2007)
ET (mm/day)
DAYS
Observations
Simulations

b)



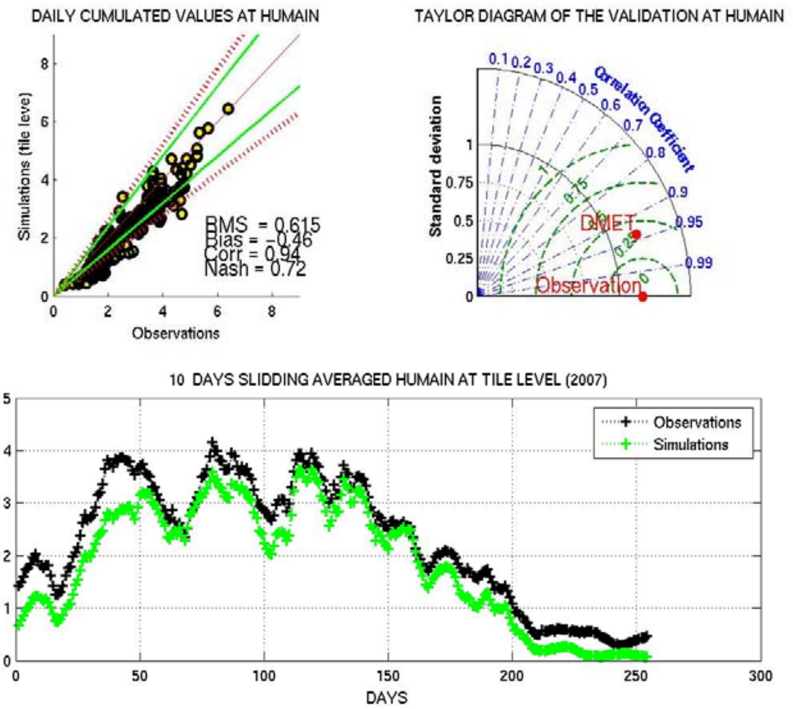
DAILY CUMULATED VALUES AT CABAUW
Simulations (tile leve)
Observations
RMS = 0.36
Bias = 0.02
Corr = 0.96
Nash = 0.92
TAYLOR DIAGRAM OF THE VALIDATION AT CABAUW
Standard deviation
Correlation Coefficient
DMET
Observation
10 DAYS SLIDDING AVERAGED VALUES AT CABAUW (2010)
ET (mm/day)
DOY
Observations
Simulations

c)



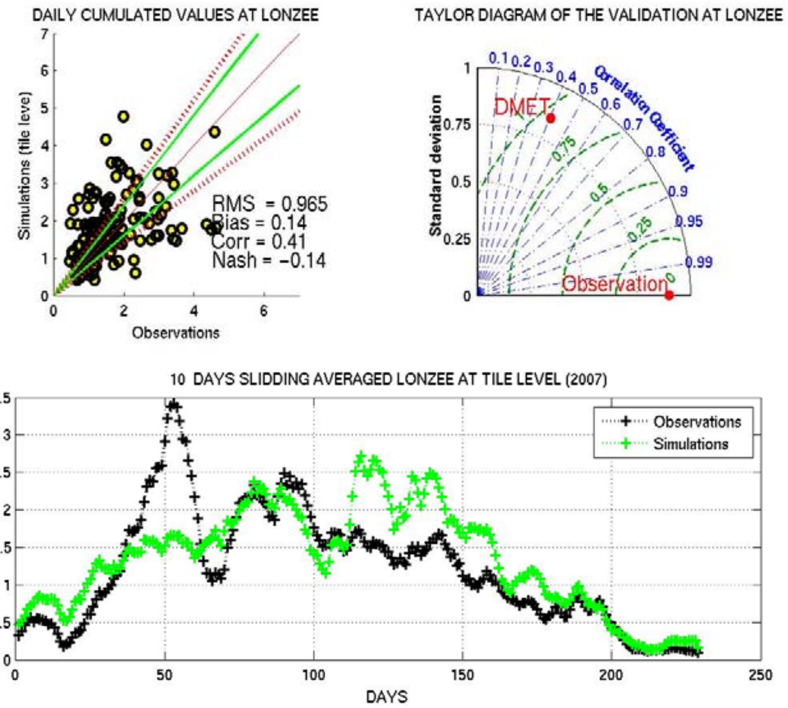
DAILY CUMULATED VALUES AT BUZENOL
Simulations (tile leve)
Observations
RMS = 0.827
Bias = 0.34
Corr = 0.8
Nash = 0.28
TAYLOR DIAGRAM OF THE VALIDATION AT BUZENOL
Standard deviation
Correlation Coefficient
DMET
Observation
10 DAYS SLIDDING AVERAGED BUZENOL AT TILE LEVEL (2007)
Observations
Simulations
ET (mm/day)
DAYS

d)



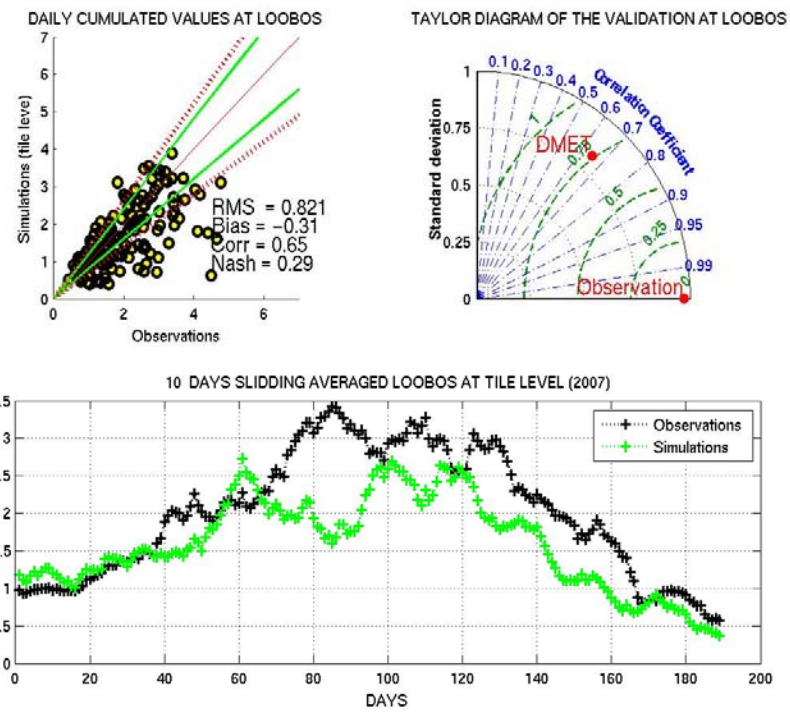
DAILY CUMULATED VALUES AT HUMAIN
Simulations (tile leve)
Observations
RMS = 0.615
Bias = -0.46
Corr = 0.94
Nash = 0.72
TAYLOR DIAGRAM OF THE VALIDATION AT HUMAIN
Standard deviation
Correlation Coefficient
DMET
Observation
10 DAYS SLIDDING AVERAGED HUMAIN AT TILE LEVEL (2007)
Observations
Simulations
ET (mm/day)
DAYS

e)



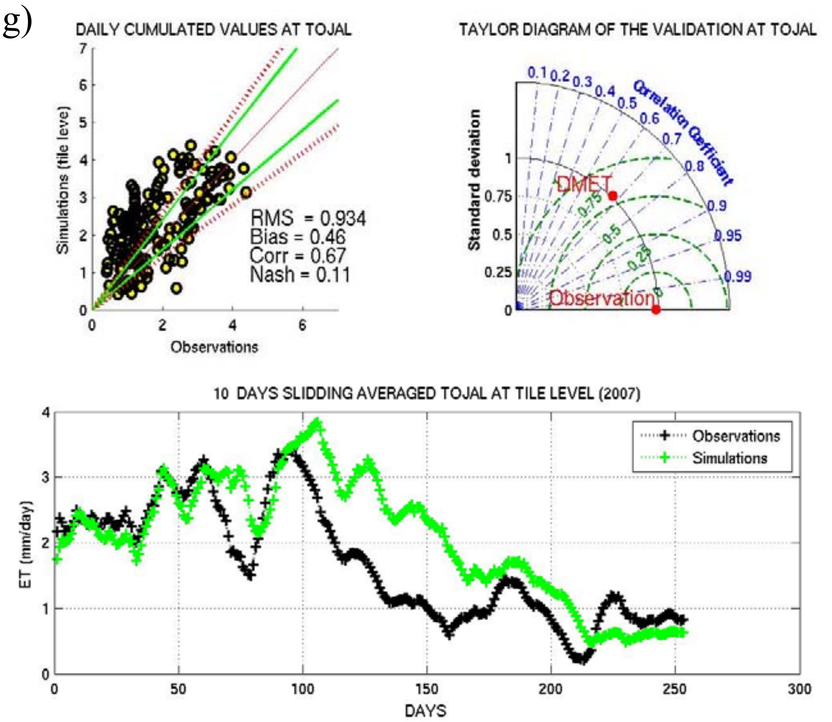
DAILY CUMULATED VALUES AT LONZEE
Simulations (tile leve)
Observations
RMS = 0.965
Bias = 0.14
Corr = 0.41
Nash = −0.14
TAYLOR DIAGRAM OF THE VALIDATION AT LONZEE
Standard deviation
Correlation Coefficient
DMET
Observation
10 DAYS SLIDDING AVERAGED LONZEE AT TILE LEVEL (2007)
ET (mm/day)
DAYS
Observations
Simulations

f)

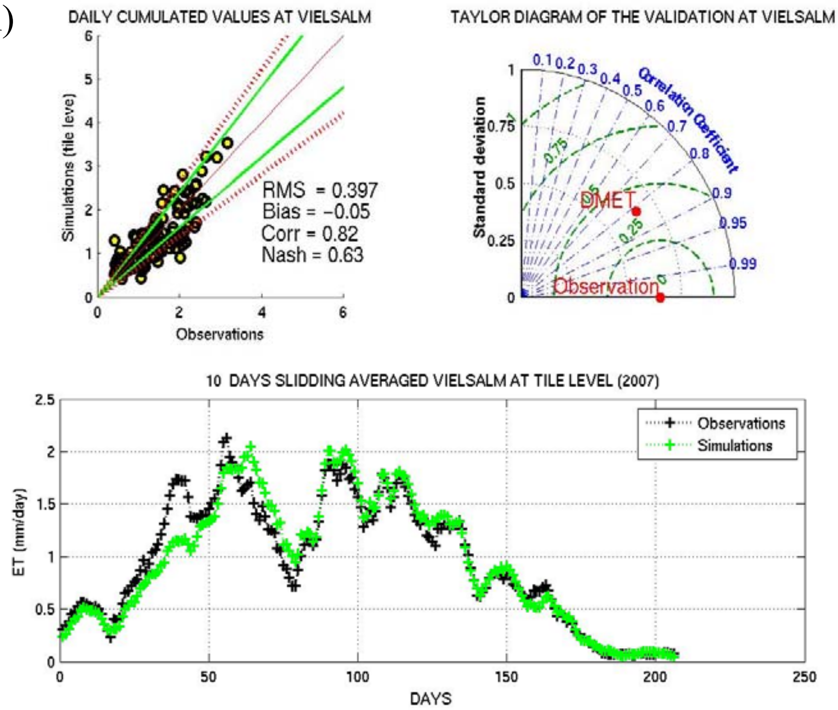


DAILY CUMULATED VALUES AT LOOBOS
Simulations (tile leve)
Observations
RMS = 0.821
Bias = −0.31
Corr = 0.65
Nash = 0.29
TAYLOR DIAGRAM OF THE VALIDATION AT LOOBOS
Standard deviation
Correlation Coefficient
DMET
Observation
10 DAYS SLIDDING AVERAGED LOOBOS AT TILE LEVEL (2007)
ET (mm/day)
DAYS
Observations
Simulations

g)


DAILY CUMULATED VALUES AT TOJAL
Simulations (tile leve)
Observations
RMS = 0.934
Bias = 0.46
Corr = 0.67
Nash = 0.11
TAYLOR DIAGRAM OF THE VALIDATION AT TOJAL
Standard deviation
Correlation Coefficient
DMET
Observation
10 DAYS SLIDDING AVERAGED TOJAL AT TILE LEVEL (2007)
ET (mm/day)
DAYS
Observations
Simulations

h)


DAILY CUMULATED VALUES AT VIELSALM
Simulations (tile leve)
Observations
RMS = 0.397
Bias = −0.05
Corr = 0.82
Nash = 0.63
TAYLOR DIAGRAM OF THE VALIDATION AT VIELSALM
Standard deviation
Correlation Coefficient
DMET
Observation
10 DAYS SLIDDING AVERAGED VIELSALM AT TILE LEVEL (2007)
ET (mm/day)
DAYS
Observations
Simulations

i)

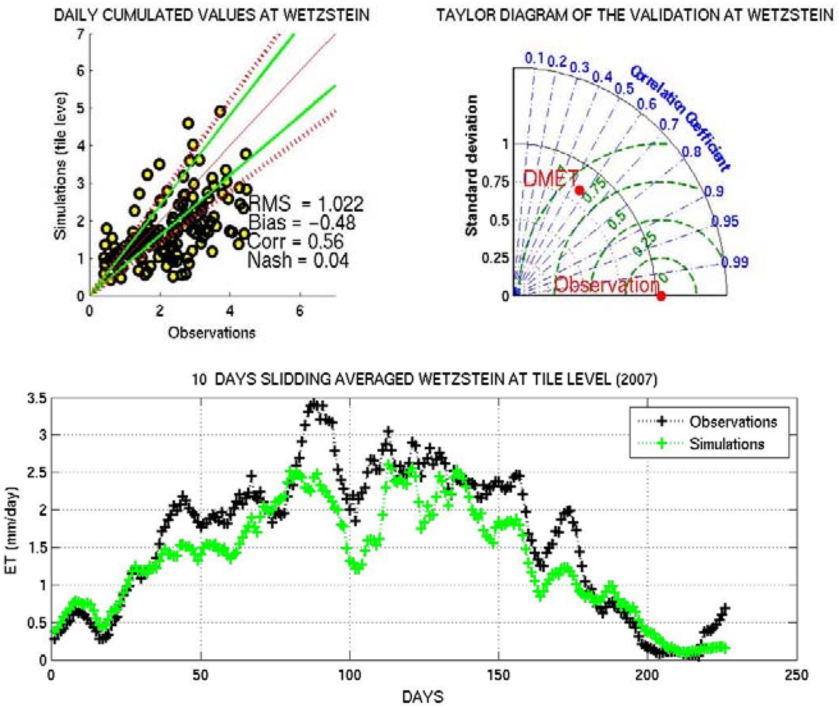


Figure 19 Comparison of DMET product to daily-cumulated ET values at selected measurement sites: Cabauw 2007/2010 (a, b); Buzenol 2007 (c); Humain 2007 (d); at Lonzee 2007 (e); Loobos 2007 (f); Tojal 2007 (g); Vielsalm 2007 (h); and Wetzstein. (i).

1.4. Discussion

When performing comparisons between models estimates and data derived from in-situ observations, special attention needs to be paid to the representativeness and accuracy of the measurements. Most of selected validation stations have proven to be representative of the targeted ecosystem (Göckede et al., 2007; Rebmann et al., 2005). However accuracy of the measurements is not provided, it is generally estimated to 20% based on the energy closure imbalance (Wilson et al., 2002; CarboEurope-IP investigators, pers. Comm.). We performed the comparison of ET model output to ground reference at “tile” level, because it compares better to observations than ET estimates for the whole pixel, which is not necessarily representative of the site (Figure 1).

For most of the comparisons, LSA-SAF MET v4.0 algorithm is able to reproduce the temporal evolution of evapotranspiration with values comparables with observations. No evident systematic bias has been discovered for the different vegetation classes; nevertheless unexpected behaviour was observed for a station (Demokeya) over African dry savannah where the model tends to underestimate ET. Compliance with PRD quality criterion is satisfied to a rate generally higher than 70%, for estimates flagged Nominal or Below Nominal. A very good agreement is found for stations over grassland (Figure 2 and Figure 3) and mixed forest (Figure 11). For stations over other types of land cover, good agreement is found in periods for which no problems are detected in the measurement procedure at the station and a close correspondence exist between the types defined in ECOCLIMAP and the real cover type at the station. However, though the rate of satisfaction is generally high, LSA-SAF MET v4.0 algorithm seems to better capture, at the temporal scale of 30 minutes, the temporal variations of ET over grassland sites than over forested sites.

Sources of difference have been explored to explain the mismatches in the comparison. Difference in solar radiation (-E1: Error caused by differences in global solar radiation), vegetation characteristics (-E2: Error caused by differences in vegetation description) and soil water availability (-E3: Error caused by differences in soil moisture), caused either by model errors or because we work at different spatial scales, can explain differences of evapotranspiration observed.

Comparisons of DMET product to daily-cumulated ET observations at reference sites confirm the main conclusions drawn for MET. For example, results from MET validation for Lonzée (crop site, 2007) and Tojal (semi-arid site, 2007) are extended to DMET. Indeed, at MSG/SEVIRI scale, the vegetation does not fit the specific site vegetation evolution (winter wheat in 2007) at Lonzée, in particular after harvest (Sepulcre-Cantó et al, 2011). For Tojal, ECMWF soil moisture is not appropriate for this site (chapter 2.1, E3). For the Lonzée station Figure 19(e), a difference is observed at the beginning of the period and it can be explained by a high difference between the real soil moisture at the station and the soil moisture used in the model (from ECMWF). Discrepancies are also observed for Loobos station (Figure 19, f) where DMET model underestimates ET for most of the period. This situation has also been observed for other models (e.g. Voogt et al, 2006). Differences observed at Buzenol between days 100 and 120 could be due to spurious, underestimated, observed ET for this period.

Overall, DMET fairly matches the observed variations, showing very good agreement with observations at well-watered sites over grassland (Cabauw (2007, 2010), Humain (2007), Buzenol (2007)) and a good seasonal variation for temperate forested sites (Vielsalm, Wetzstein and Loobos, 2007), with better agreement over the mixed than coniferous forests. Some restrictions apply for arid and semi-arid sites where research need to be pursued to check model capabilities and look for possible improvements.

2. Inter-comparison with other products

Although there are several methodologies developed to derive ET from remote sensing (Courault et al, 2005, for a short review), only one satellite-based method (Rosema, 1993) for retrieving ET runs on a regular operation chain processed by EARS (http://www.ears.nl/; http://www.earlywarning.nl/) and participating to the GEOLAND project (http://www.gmes-geoland.info/). More recently evapotranspiration products based on MODIS imager have been set in operations at NASA (Cleugh et al., 2007; Mu et al., 2007). Comparisons with above product had not been done till now because of the cost related to product acquisition or because the product is still in implementation phase. Another kind of products available for comparison are NWP model/assimilation systems outputs, generally recognized as reference by the scientific community, i.e. ECMWF global model and GLDAS land assimilation system (both participating in the CEOP projects). In this section we present the results of the comparison with ECMWF and GLDAS. Main characteristics of ECMWF and GLDAS are summarized in Table 2 below. For a more complete description please refer to annexe D.

	ECMWF	GLDAS
Versioning	Cycle 31r1	GLDAS/Noah 2.7.1.
Land Surface Scheme	TESSEL	Noah
Coupled to atmosphere	Yes	No
Domain	Global	Global
Horizontal spatial resolution	0.25°+	1°+
Available temporal resolution	3h+	3h+
Land Cover Used	IGBP	University of Maryland
Period covered Begin (aaaa-mm-dd)	2007-03-01 (Euro) 2007-04-01 (Others)	2007-03-01 (Euro) 2007-04-01 (Others)
Period covered End (aaaa-mm-dd)	2007-11-30	2007-11-30

Table 2 Summary of characteristics of the models and outputs used for inter-comparison.

+This is the temporal/spatial resolution available to external user, which does not correspond to time step or grid definition for computation.

2.1. Methodology

To insure a meaningful framework of inter-comparison between the three models, LSA-SAF MET results to be compared must have the same spatial and temporal resolution. All estimates have to be degraded to the coarser resolution (3-hourly averages on a regular 1°x1° latitude-longitude grid).

GLDAS maps are not modified. Only the unit of ET is changed to mm/h. ECMWF maps are aggregated from the 0.25°x0.25° regular grid to 1°x1° grid and expressed in mm/h. LSA-SAF MET maps are cumulated over 3 hours, expressed in mm/h (6 images by estimate equally weighted) and projected on a regular 1°x1° latitude-longitude grid. Truncation of the image occurs where MSG pixel area is approximately equal to 1°x1°. If one of the six estimates to perform the mean is missing, the weights are equally redistributed between the remaining five. If more than one estimate is missing, the mean is not computed.

For the inter-comparison, three types of statistical tests were performed:

1) One-to-one comparison of the images, for direct visualization;
2) Global analysis of the images over the whole period; and,
3) Regionalized statistical test (Spatio-temporal comparison: regional analysis) in view to determine the differences over different biomes.

A last point is dedicated to explain the differences in modelled ET in terms of input variables/parameters differences.

2.2. Instantaneous product (MET)

In Figures 20 and 21, samples of LSA-SAF MET images over Europe are presented for visual interpretation. DSSF is one of the input variable that most influences the output of the ET algorithm. For this reason, some DSSF images are also presented in order to show the spatial coherence between the two variables. In those figures, the images are presented for two different dates (beginning of April and beginning of July 2007).

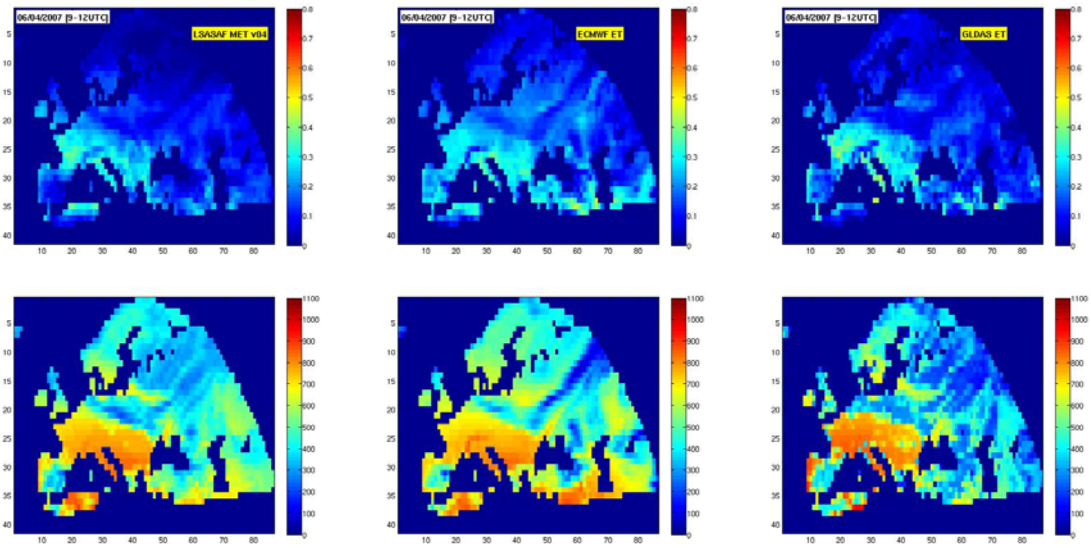


Figure 20 Comparison of LSA-SAF (left), ECMWF (middle) and GLDAS (right) ET (top) and global radiation (bottom), averaged over 3 hours (except GLDAS global radiation, instantaneous 12 UTC), for the day 06/04/2007, 09 UTC to 12 UTC.

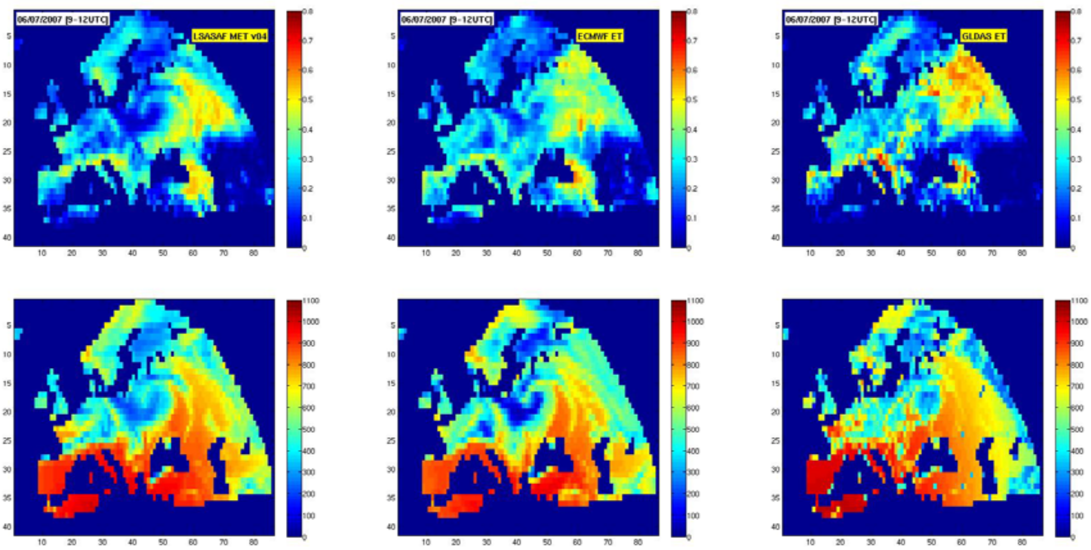


Figure 21 Comparison of LSA-SAF (left), ECMWF (middle) and GLDAS (right) ET (top) and global radiation (bottom), averaged over 3 hours (except GLDAS global radiation, instantaneous 12 UTC), averaged over 3 hours, for the day 06/07/2007, 09 UTC to 12 UTC.

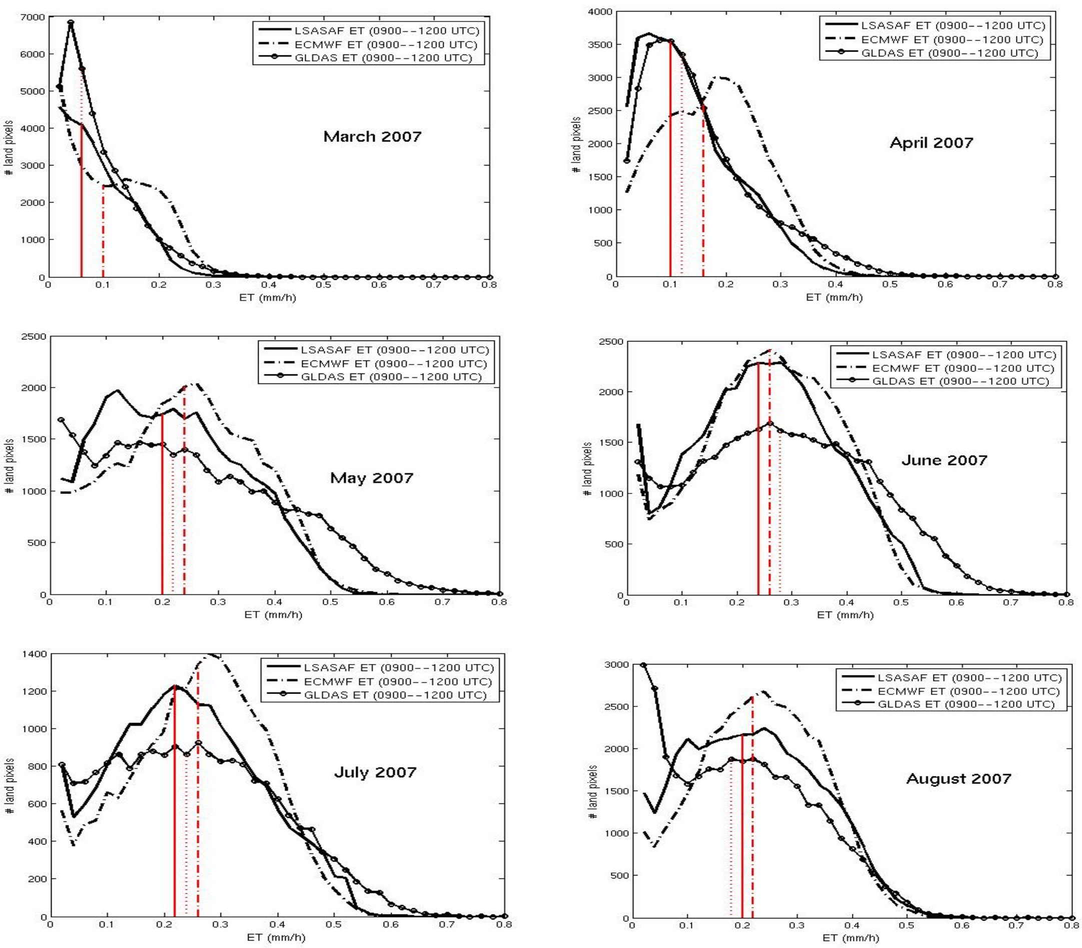
As we can see, the range of ET values is similar for the 3 models in both cases. We observe however some differences in patterns of ET. These differences seem to be correlated with differences in global radiation at surface, which can be large in some regions (example: cloudy sky over Poland in Figure 21). Solar radiation is not the only reason of the differences. For example, while global radiation at surface is quite similar in the Iberian Peninsula in Figure 21, patterns of ET are different.

Global analysis

We perform an analysis over the whole MSG window and the full period between 01/03/2007 and 30/11/2007 for Europe and between 01/04/2007 and 30/11/2007 for Africa and South America. The presentation of the results is structured as follows:

- Comparison of mean distributions of ET, providing quantitative information on temporal distribution of ET over Europe.
- Evolution in time of the global spatial correlation between ECMWF, GLDAS LSA-SAF MET images to provide information on similarities of global spatial patterns on maps, and detect where the differences occur.

In Figure 22, we compare the distribution of ET values in the images averaged by month for one time step of comparison, i.e. 3 hour, for ECMWF, GLDAS and LSA-SAF MET, over Europe. For the sake of clarity, we only represent the mid-day time step, i.e. average between 9:00 and 12:00 UTC.



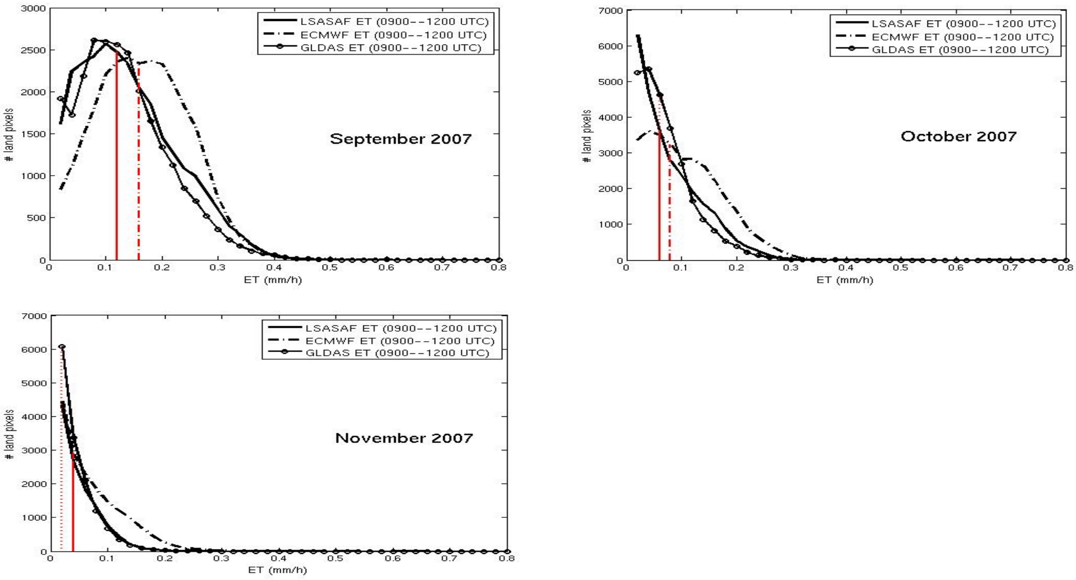
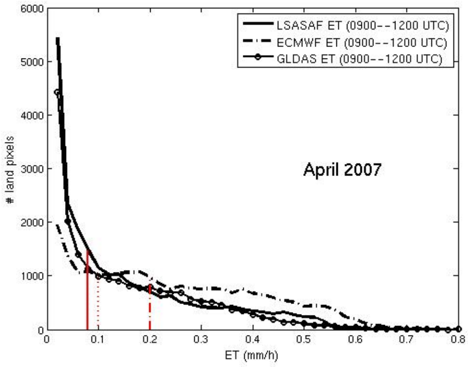


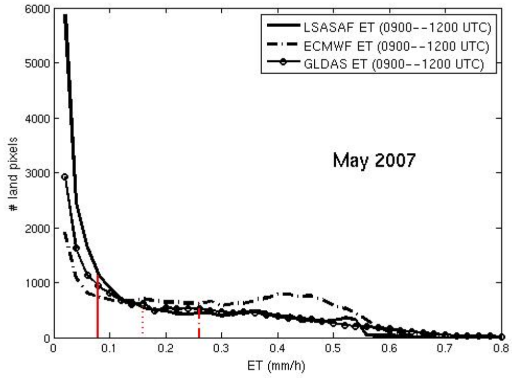
Figure 22 Distributions of ET estimates from LSA-SAF (solid line), ECMWF (dash-dotted line) and GLDAS (solid line and circles). Each figure encompasses the mean distribution of the 3 hourly average ET (09UTC to 12UTC) for a month (from March to November 2007), as well as the mean value of the distribution, for Europe.

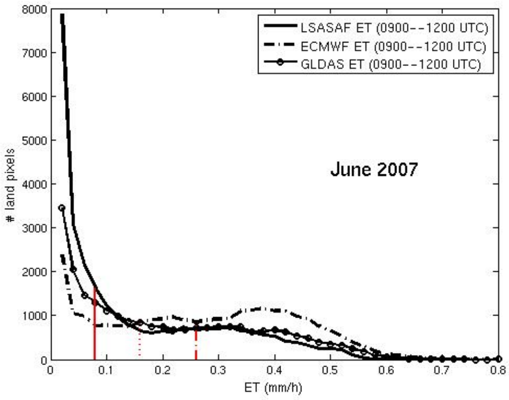
The amplitude and means of distributions of LSA SAF MET over Europe are comparable to ECMWF and GLDAS. However, some differences are detected. While, during spring and autumn, LSA SAF MET distributions are closer to GLDAS, summer distributions exhibit a closer similarity with ECMWF than with GLDAS, for which more spread distributions are found. In combination with the fact that evening estimates correlate better with GLDAS, we can suggest that the evolution of ET for low solar angles is closer with GLDAS and for around-noon estimates the evolution is closer to ECMWF. Another remark is that the LSA SAF MET mean of distributions is systematically lower than that of ECMWF. Inter-comparison of the mean of the distribution with GLDAS shows very good agreement with a maximum of difference of 15% in June.

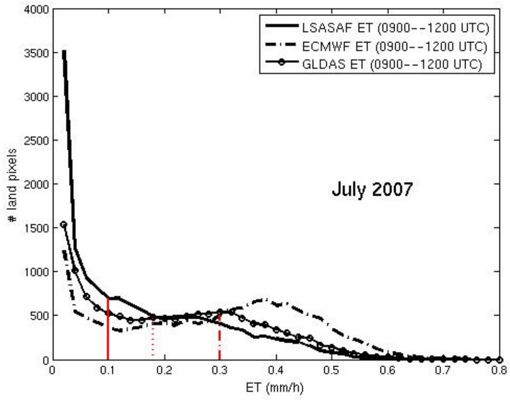
In Figure 23, Figure 24 and Figure 25, we compare the monthly mean distribution of ET values for one time step per day for ECMWF, GLDAS and LSA-SAF MET over African and South American windows. The mid-day time step is chosen: between 9:00 and 12:00 UTC for African windows, and between 15:00 and 18:00 UTC for South American window. For each distribution, the mean is given.

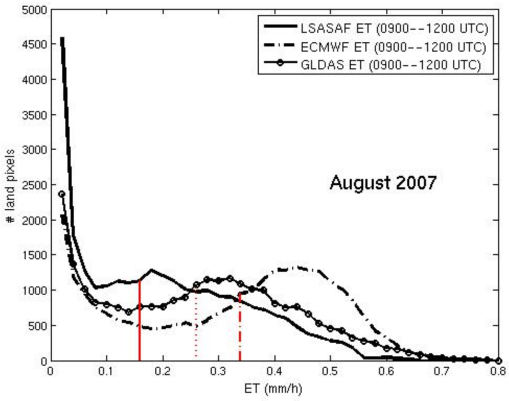
For the northern African window in Figure 23, the three models give different distributions with different mean value. MET v4.0 gives the lower mean value. ECMWF ET gives systematically the highest mean value, often twice the MET v4.0 mean. The distribution of GLDAS ET is in agreement with MET v4.0 for small to medium estimates, while is closer to ECMWF for the tail of the distribution towards the large values, as we can see for May, September and October.

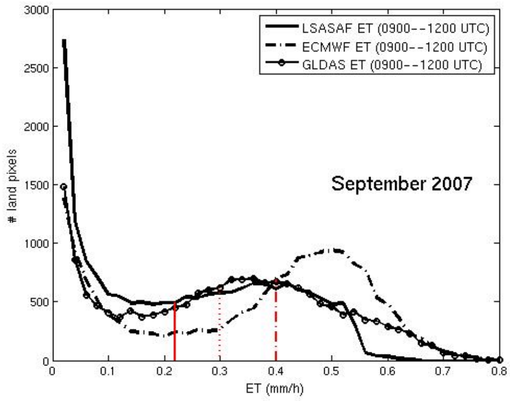

LSASAF ET (0900--1200 UTC)
ECMWF ET (0900--1200 UTC)
GLDAS ET (0900--1200 UTC)
April 2007
land pixels
ET (mm/h)


LSASAF ET (0900--1200 UTC)
ECMWF ET (0900--1200 UTC)
GLDAS ET (0900--1200 UTC)
May 2007
land pixels
ET (mm/h)


LSASAF ET (0900--1200 UTC)
ECMWF ET (0900--1200 UTC)
GLDAS ET (0900--1200 UTC)
June 2007
land pixels
ET (mm/h)


LSASAF ET (0900--1200 UTC)
ECMWF ET (0900--1200 UTC)
GLDAS ET (0900--1200 UTC)
July 2007
land pixels
ET (mm/h)


LSASAF ET (0900--1200 UTC)
ECMWF ET (0900--1200 UTC)
GLDAS ET (0900--1200 UTC)
August 2007
land pixels
ET (mm/h)


LSASAF ET (0900--1200 UTC)
ECMWF ET (0900--1200 UTC)
GLDAS ET (0900--1200 UTC)
September 2007
land pixels
ET (mm/h)

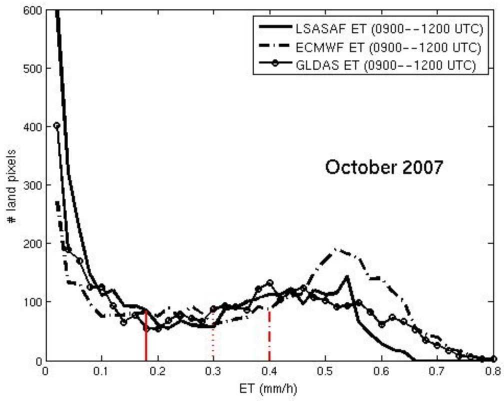
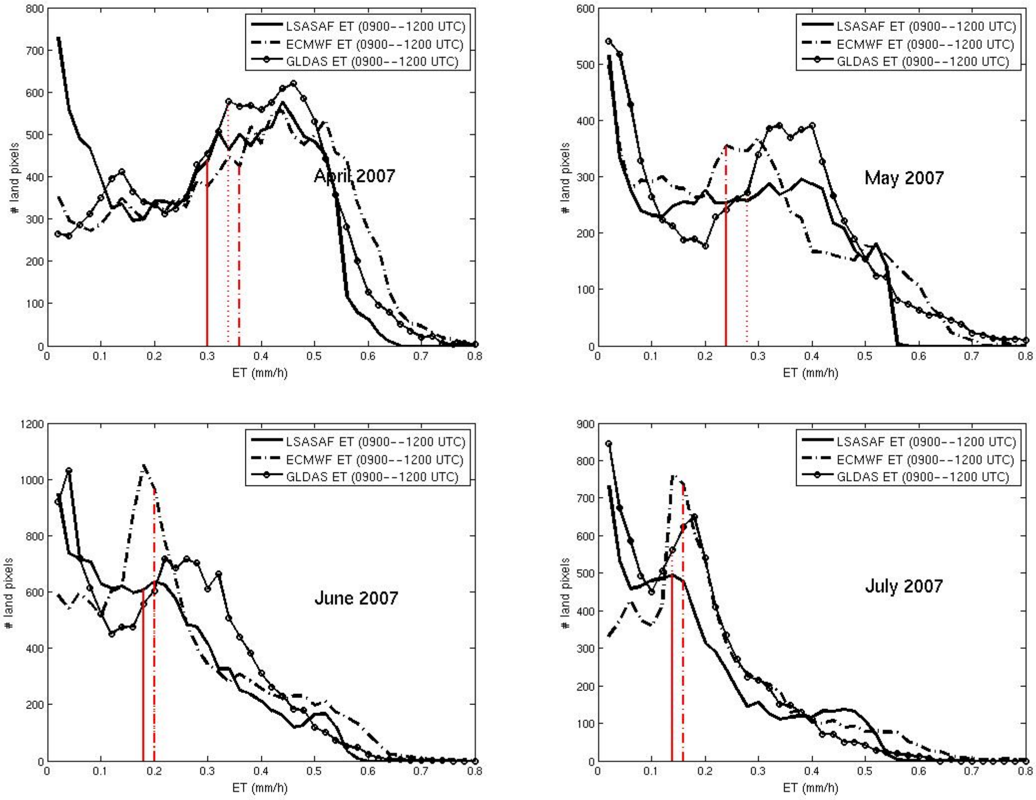

LSASAF ET (0900--1200 UTC)
ECMWF ET (0900--1200 UTC)
GLDAS ET (0900--1200 UTC)
October 2007
land pixels
ET (mm/h)

Figure 23 Distributions of ET estimates from LSA-SAF (solid line), ECMWF (dash-dotted line) and GLDAS (solid line and circles) for North African window. Each figure encompasses the mean distribution of the 3 hourly average ET (09UTC to 12UTC) for a month (from April to October 2007), as well as the mean value of the distribution.

For the South African window, we observe close agreement between the three models, with few differences between the distribution means. The shapes of the distributions are also similar for medium and large values, with the exception of May for which the modes are not located at the same place and the tail is different.



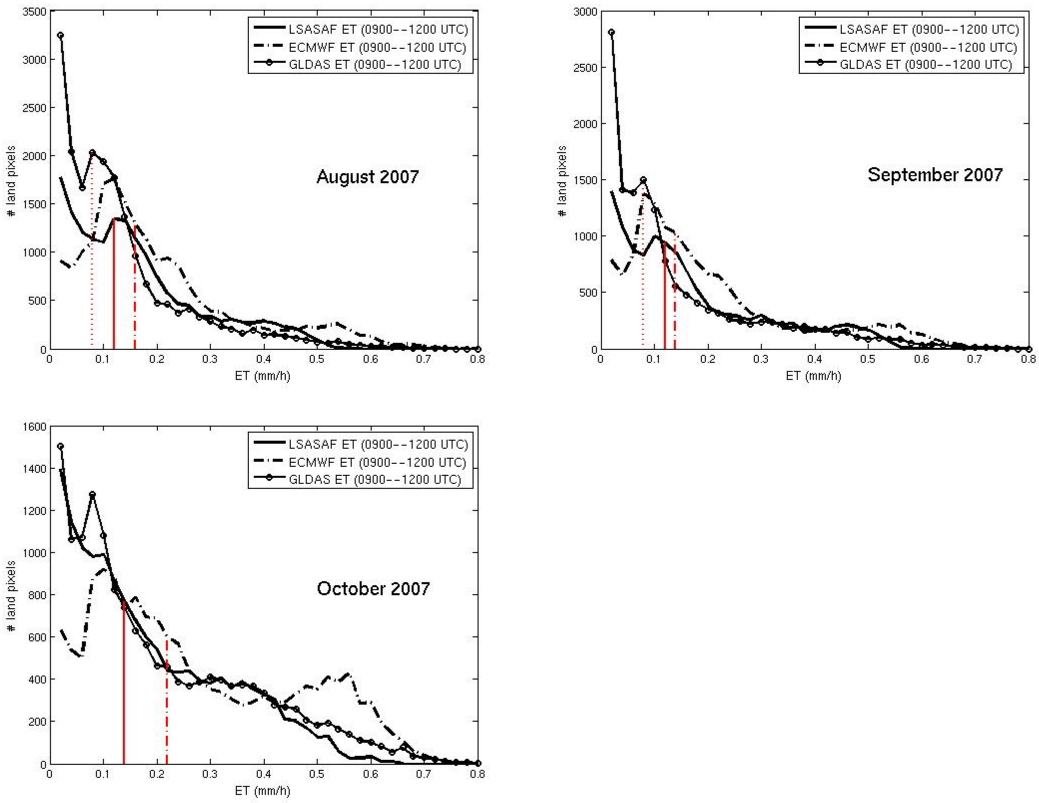


Figure 24 Same as Figure 23, but for South Africa.

For the South American window, ET estimates for ECMWF and GLDAS are systematically larger than MET v4.0 estimates with mean bias between 15% and 35%. Distinct modes of the distributions are easily observed in April and May. Comparing the tails of the distributions, we see that GLDAS and ECMWF tend to larger values than MET v4.0.

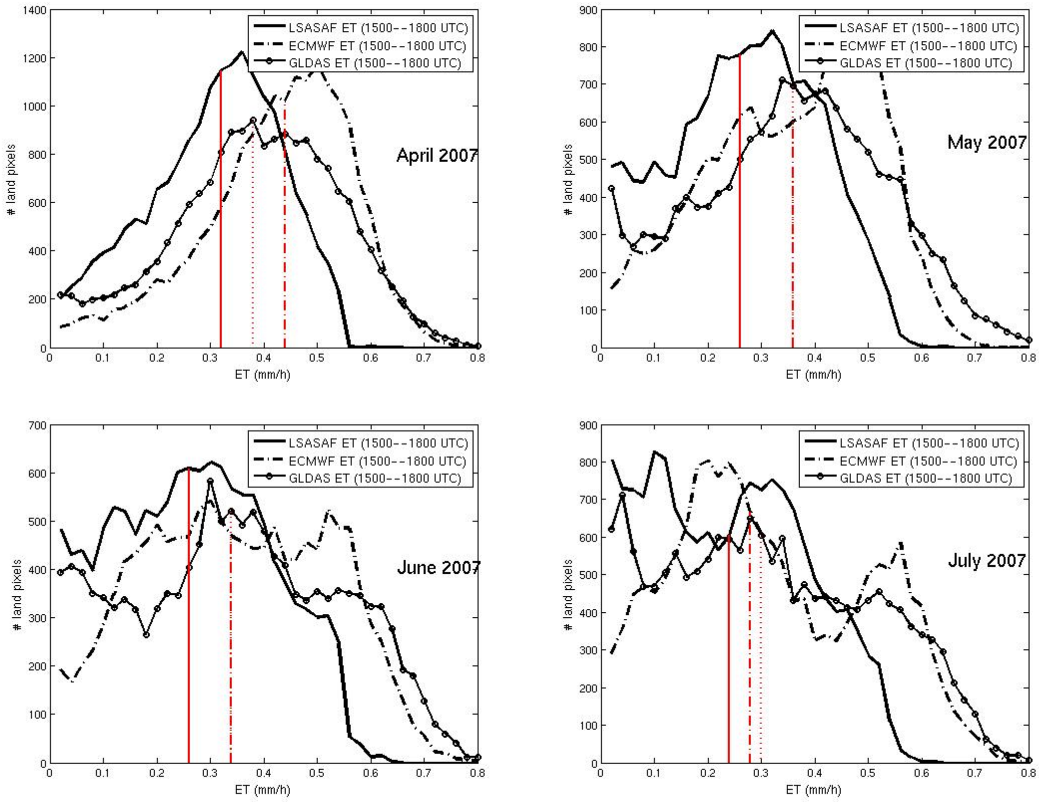
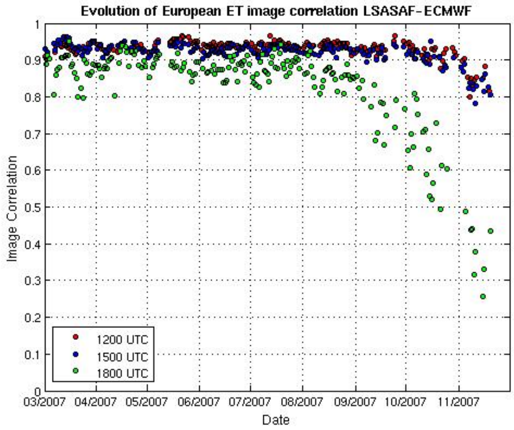


Figure 25 Same as Figure 23, but for South America, for the period from April to July 2007.

A spatial correlation coefficient of each 3-hourly image between 09:00 UTC and 18:00 UTC of ECMWF, LSA-SAF MET and GLDAS is calculated, following formula (1), and the temporal evolution of this index is represented in Figure 26. Each hour of the day considered is represented by a different coloured label.

$$corr_{images} = \frac{\sum_j \sum_i (A_{ij} - \overline{A}).(B_{ij} - \overline{B})}{\sqrt{(\sum_j \sum_i (A_{ij} - \overline{A})^2).(\sum_j \sum_i (B_{ij} - \overline{B})^2)}}$$ where A and B are the images (1)



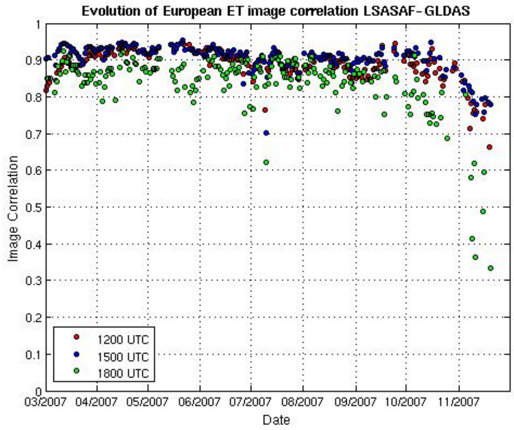


Figure 26 Evolution of 3-hourly mean image correlation over MSG European window between LSA-SAF ET and evapotranspiration forecasts from ECMWF, at left, and with GLDAS ET images, at right, from 1st March to 30th November 2007.

From Figure 26, we can observe that the image correlation is quasi constant for 12:00 and 15:00 UTC, with a correlation generally above 90%. For 18:00 UTC, we observe a seasonal effect with a good correlation for March to September and a decreasing correlation during autumn, due to the fact that image progressively changes from day to night for these hours. Image correlation is slightly better with ECMWF than GLDAS for 12:00 and 15:00 UTC, especially for the summer months. This observation confirms that the general patterns characterizing European ET maps, like meteorological (e.g. solar radiation) and land cover effects, are found in both products and are in agreement with what was already remarked from the monthly mean distributions at noon. However, better correlation with GLDAS is observed for 18:00 UTC at the end of the comparison period. Low correlation for 18:00 UTC images at the end of the period is probably due to the fact that ET is nearly zero for land pixels in both products, inducing very few contrasts on images.

Spatial correlation between MET v4.0 and ECMWF ET and GLDAS ET is assessed for the 3 other geographical windows in Figure 27, Figure 28 and Figure 29, respectively. For Africa, the evolution of the 2-D correlation coefficient is performed for two periods of 3 hours, i.e. 9:00 to 12:00 UTC and 12:00 to 15:00 UTC, while for South America, the periods are 12:00 to 15:00 UTC and 15:00 to 18:00 UTC, because of the longitude difference. Globally, 2-D correlation is higher than 80% in Africa and 90% in South America. Correlation values obtained with ECMWF and GLDAS are similar, indicating that the three outputs agree on the general patterns of ET variations. For Africa, a seasonal fluctuation is visible with a better correlation during autumn and winter and a lower one during spring and summer, with a phase shift of 6 months between both windows. This effect is not observed for South America, mostly because the period considered for analysis is too short.

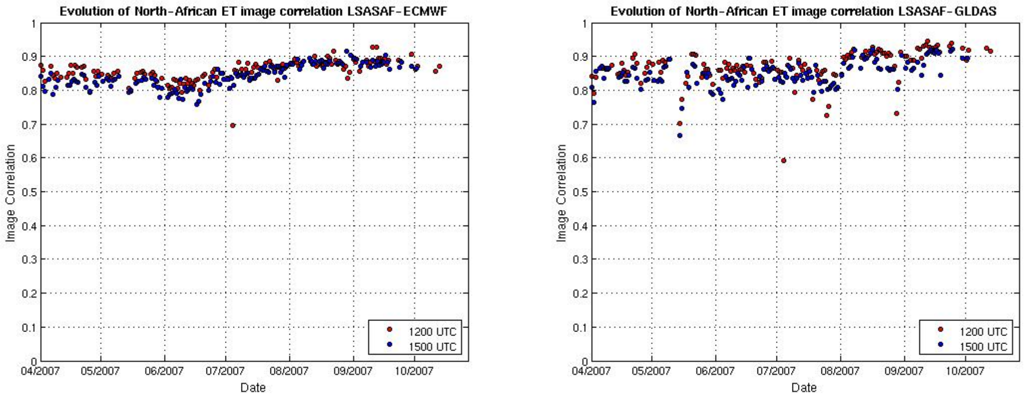


Figure 27 Time series of the spatial correlation of LSA-SAF ET with ECMWF ET images of Northern Africa, at left, and with GLDAS ET images, at right. Spatial correlation are computed from April till October 2007 for two consecutive periods of the day: 09:00 to 12:00 UTC and 12:00 to 15:00 UTC.

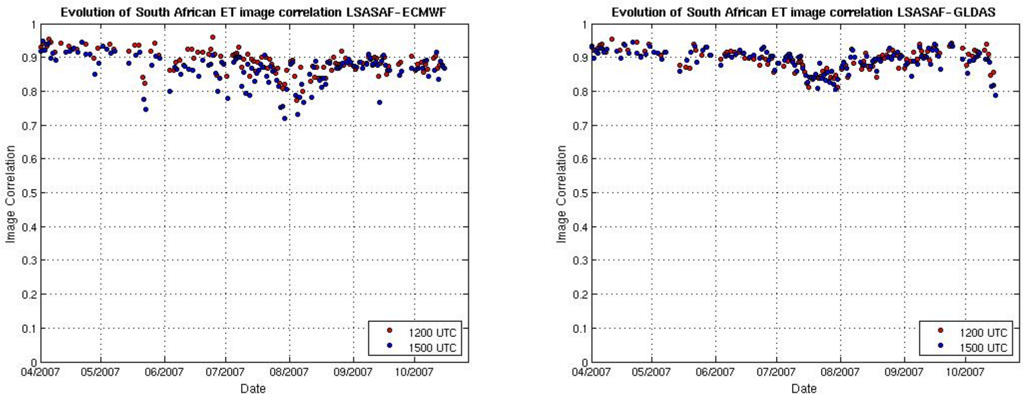


Figure 28 Same as in Figure 27, but for South Africa.

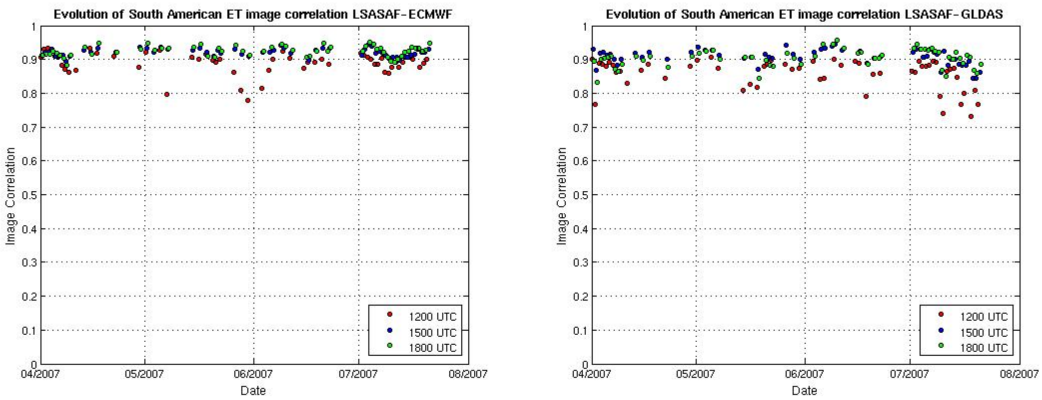


Figure 29 Same as in Figure 27, but for South America. The period considered extends from April to end of July 2007.

Spatio-temporal comparison: regional analysis

In this subsection, we analyze the results of inter-comparison in a spatio-temporal framework. For the whole period considered, we analyze the geographical distribution of temporal inter-comparison statistical indices. Two different indices are calculated: the slope of the regression line and the mean relative bias. The slope of the regression line has been reduced by unity to rapidly identify the sign of the difference: positive corresponds to smaller values for LSA SAF MET, and negative to larger values. The same meaning is found for the mean relative bias: positive correspond to under-estimation of LSA SAF MET compared to ECMWF or GLDAS. For the mean relative bias, the reduction is done by reference to the estimates of LSA SAF MET v4.0. As one can expect, the relative difference will be more meaningful for large than for small ET. Therefore, large relative differences should not be interpreted as bad quality when considering small values. For the sake of clarity of the maps in Figure 30, Figure 31, Figure 32 and Figure 33, all biases greater than 50% were truncated to this threshold and the same kind of truncation was performed for biases with the opposite sign.

The comparison of time series of ET at grid cell level is complementary information to allow localizing the regions of agreement and disagreement among the 3 models. For each grid cell, we build a time series of 3-hourly ET estimates between 9:00 UTC and 12:00 UTC for Europe and Africa, and between 15:00 and 18:00 UTC for South America. For each window and each grid cell, a linear regression is performed between MET v4.0 ET time series and ECMWF ET or GLDAS ET. The slope of the regression line reduced by one and the relative bias compared to MET v4.0 are displayed for each window in Figure 31, Figure 32 and Figure 33. In general, both indicators reflect the differences in term of bias. Where positive bias is observed, the slope of the regression line reduced by one is negative. Positive bias means ECMWF or GLDAS gives larger estimates than MET v4.0.

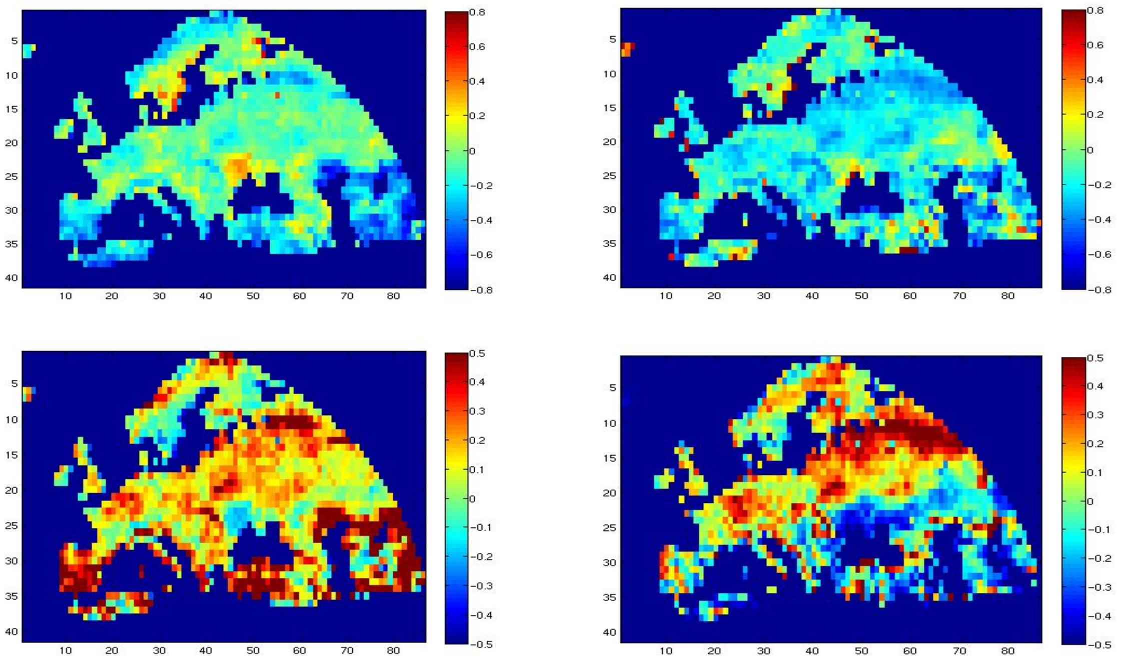


Figure 30 European 1°x1° maps of temporal statistical indices to measure the difference between ECMWF and LSA SAF MET (left) and between GLDAS and LSA SAF MET (right). The slope of the regression line

reduced by one is shown at top and the mean relative bias over the whole period of inter-comparison is given in the bottom images.

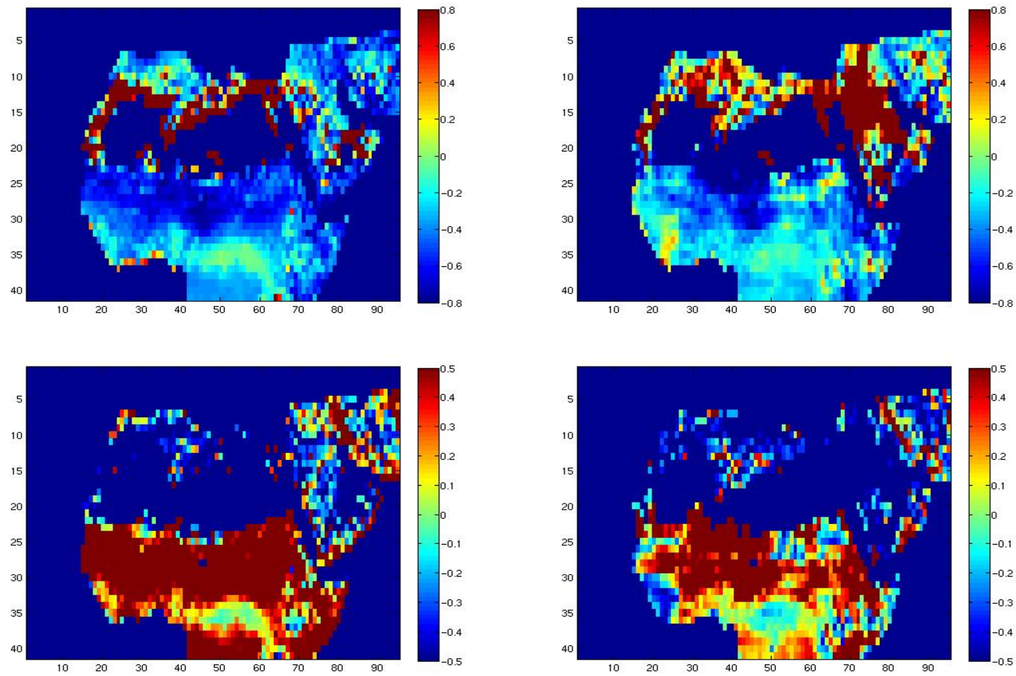


Figure 31 North African 1°x1° maps of temporal statistical indices to measure the difference between LSASAF MET and ECMWF (left) or GLDAS (right). The slope of the regression line reduced by one is shown at top and the mean relative bias at bottom. The statistical indices are based on 3-hourly means between 9:00 and 12:00 UTC from April to October 2007.

In Figure 31, we can see that, in comparison with ECMWF, ET from MET v4.0 is 50% lower in most of the North African window, especially in the Sahel region. That observation corroborates the differences in the monthly mean distribution noticed previously. The differences with GLDAS are not systematic over the whole area. In Arabia and over the Nile delta, ET from MET v4.0 is 50% higher than GLDAS. In the Sahel region, some parts show positive bias and others negative bias. A null bias is found for the region making the transition from Sahel to equatorial forest.

For South Africa, in Figure 32, we observe more differences with ECMWF than with GLDAS. For that region, areas where the bias is higher than 50% correspond to very low absolute ET values. It corresponds to dry pixels, for which evaporation is very small. In term of absolute differences, for these pixels, the evapotranspiration will not be very different. It is probably why these differences are not observed in the monthly mean distributions.

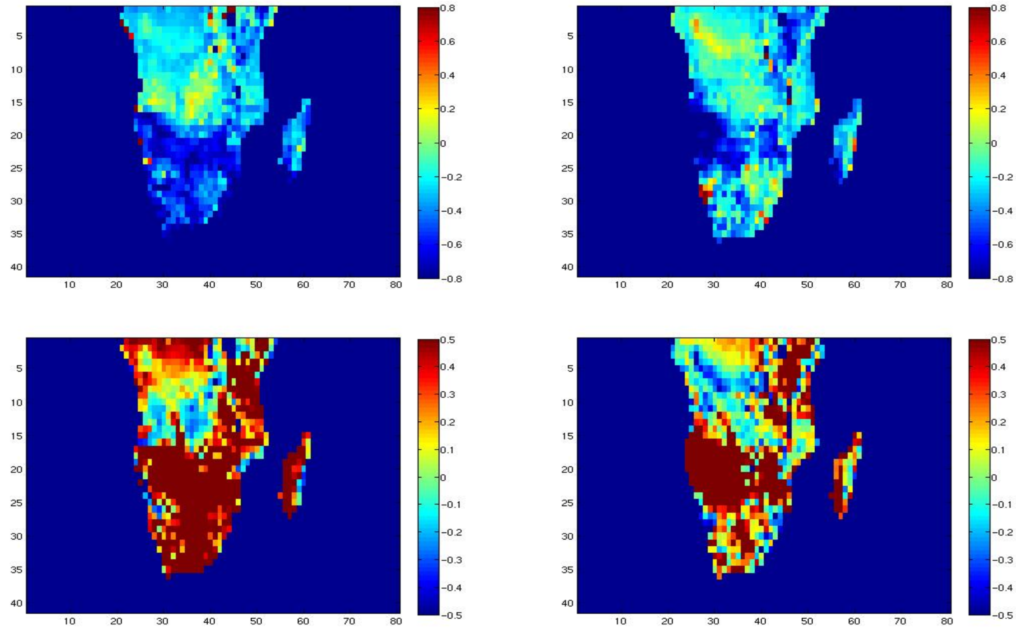


Figure 32 Same as in Figure 31, but for Austral Africa.

For South America, in Figure 33, ET estimates by MET v4.0 are low over most part of the window compared to ECMWF and GLDAS. This systematic difference was already noted in the monthly mean distribution.

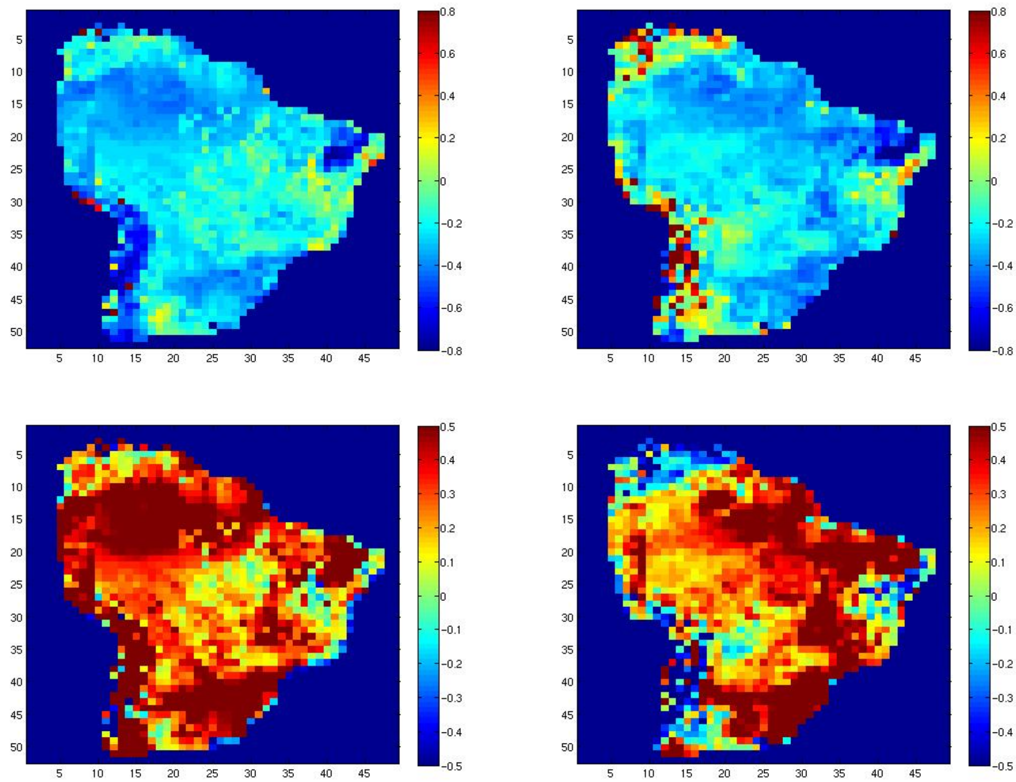


Figure 33 Same as in Figure 31, but for South America, and period from April to July 2007.

Explanation of the differences between the model results

This subsection is dedicated to search the main reasons to explain differences between the ET computed by the three models. Two main sources have been explored: global solar radiation (D1) and vegetation characteristics (D2).

-D1: differences in global solar radiation

European maps of bias between LSA SAF DSSF and ECMWF are established on the basis of the 09:00 to 12:00 UTC images. European maps of bias between LSA SAF DSSF and GLDAS/AFWA are established on the basis of the image at 12:00 UTC. From Figure 34, a clear bias, ranging between 15% and 30%, is observed between GLDAS/AFWA and LSA-SAF global radiation products in the region of the Caucase, Ukraine, East of Turkey and near the Baltic Sea. In comparison with the right part of Figure 30, no clear spatial correlation is found. Indeed, solar radiation impact is combined to other factors like soil moisture state. For high latitudes, like in region around the Baltic Sea, GLDAS ET is enhanced by a higher radiation at the surface. On the contrary, the global radiation difference in the region of the Caspian Sea could explain ET negative bias in this region, by an indirect effect of surface drying. If there is more radiation coming into the surface, in absence of rain, the surface will dry more rapidly and ET rates will be lower. In Figure 35, monthly occurrence histograms for global radiation at surface are presented. Only two months are shown, but conclusions are the same for the non-included figures. A quite good correspondence with ET histograms presented in Figure 22 is observed. In April, while global radiation mean correspond for LSA SAF DSSF and GLDAS/AFWA, ECMWF global radiation is globally overestimated compared to LSA SAF DSSF. This observation results in close mean distribution of ET for GLDAS and LSA SAF MET v4.0, and an overestimation of ECMWF compared to LSA SAF MET. In June, ECMWF global radiation and LSA SAF DSSF agree, but GLDAS/AFWA overestimates a little by comparison with LSA SAF DSSF. A similar bias with same sign on the ET distribution for this month is observed at Figure 22 when comparing GLDAS and LSA SAF MET. However, the implication is not straightforward, since in arid regions the ET is reduced while in well-watered regions the ET is enhanced (cf. Figure 30).

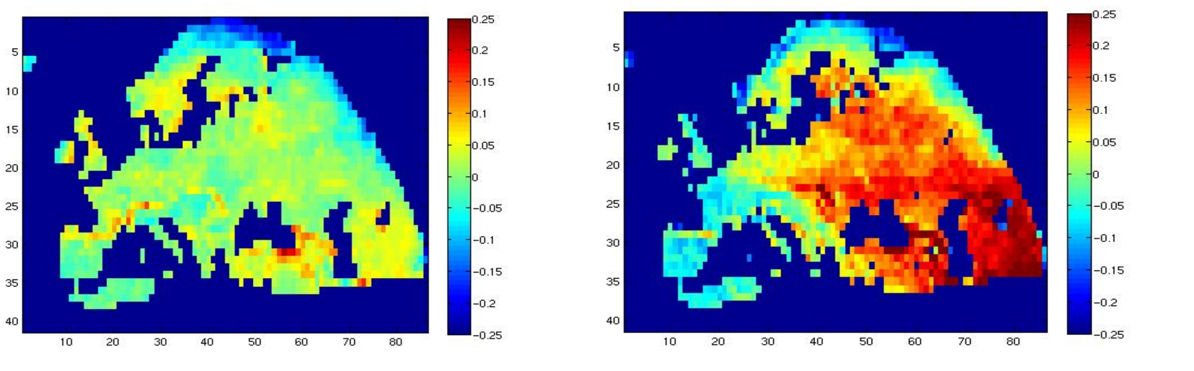
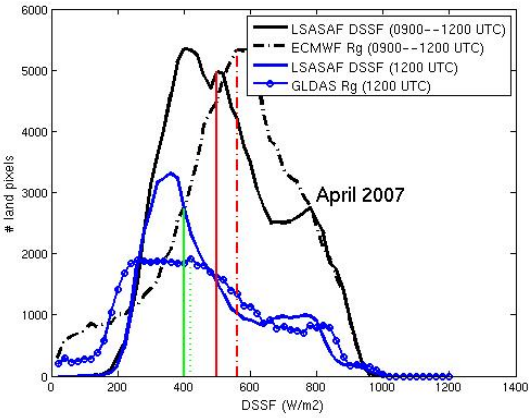


Figure 34 European 1°x1° maps of temporal statistical indices to measure the difference between ECMWF global radiation and LSA SAF DSSF (left) and between GLDAS global radiation and LSA SAF DSSF (right). The mean relative bias over the whole period of inter-comparison is shown.



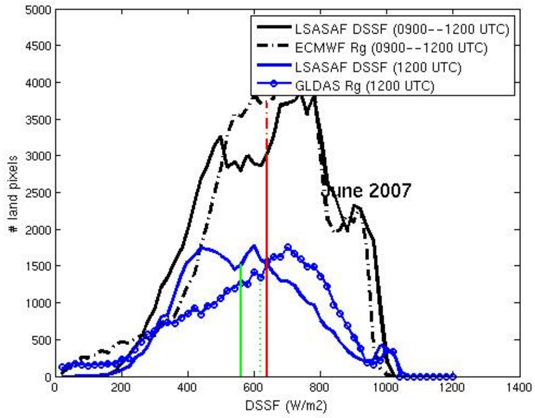


Figure 35 Distributions of global radiation at surface estimates from LSA-SAF (solid line), ECMWF (dash-dotted line) and GLDAS (solid line and circles). Each figure encompasses the mean distribution of the 1) 3 hourly average global radiation (09UTC to 12UTC) for LSA SAF and ECMWF (black); 2) average global radiation at 12 UTC for LSA SAF and GLDAS/AFWA (blue), for two months (April and June 2007), as well as the mean value of the distribution.

-D2: differences in vegetation

Two different aspects are shown: 1) comparison of the land cover maps; 2) comparison of the ratio between leaf area index (LAI) and minimum stomatal resistance (Rsmin).

At first, land cover is an important factor influencing the partition of energy fluxes at the surface, through the model parameters chosen to constrain ET for the different vegetation types. The three models considered are using different land cover maps, with different classifications and from different sources, i.e. ECOCLIMAP, IGBP and UMD. Differences of classification can be translated into differences in evapotranspiration. Therefore, we inter-compare the different land cover used by the models. Since the classifications of basic vegetation types are different, we propose to separate vegetation in fractions of high and low types. The procedure applied is documented in Annex E. In Figure 36, we show the European map of the difference in high vegetation percentage between LSA SAF MET and, respectively, ECMWF, at left, and GLDAS, at right. While no clear conclusion can be extracted from the land cover difference with GLDAS, some useful information is given for ECMWF. Indeed, we can observe regions where differences in high vegetation percentage can explain ET bias pointed out in Figure 30. For example, in the region Northwest of the Black Sea and in Southwest of France, a smaller percentage of forest in ECOCLIMAP/MSG lead to larger ET estimates from LSA SAF MET, compared to ECMWF.

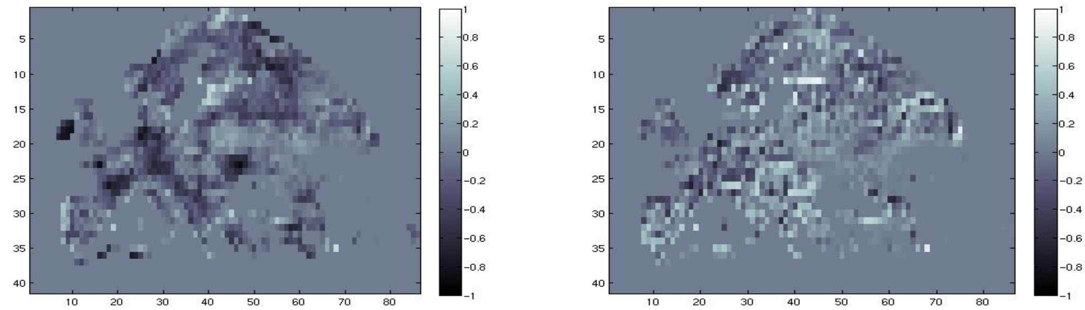


Figure 36 European 1°x1° map of the difference in high vegetation percentage between the land cover used by the 3 models. Differences between LSA-SAF and ECMWF are represented left, and between LSA-SAF and GLDAS at right.

At last, vegetation influence on surface energy partitioning is not limited to land cover. Vegetation characteristics also play a role through different parameters, like Leaf Area Index (LAI), Fraction of Vegetation Cover, roughness length, or modeled stomatal resistance of the canopy to transpiration. In this context, we choose to show the differences of a combination of vegetation parameters. Since in SVAT, LAI is scaled by the minimum stomatal resistance parameter (Rsmin) to compute the total stomatal resistance to transpiration, we compared the European maps of the ratio of LAI and Rsmin averaged over the whole inter-comparison period (Figure 37). Negative differences indicate stronger resistance to ET in the LSA SAF MET model, and therefore imply smaller ET estimates. Some clear patterns are found in the LSA SAF MET-ECMWF differences that correlate with ET biases, for example, Mediterranean basin, North of Caspian Sea and Northeast Europe. More contrast is observed for the comparison with GLDAS. For example, in a large part of the North-East of Europe, the resistance is greater for LSA SAF MET. This implies smaller estimates, as observed at the right part of Figure 30. In reverse, smaller resistance in the region of Aral Sea implies larger LSA SAF MET estimates observed in that region.

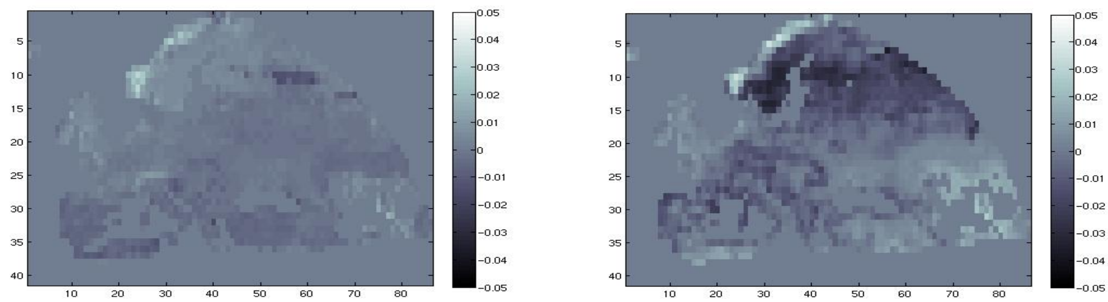


Figure 37 European 1°x1° map of the difference in the ratio between Leaf Area Index (LAI) and Minimum Stomatal Resistance (Rsmin) used by the 3 models. Differences with ECMWF are represented left, and with GLDAS at right.

Finally, we spatially correlate the mean relative bias of ET ($\Delta_{\%}ET$) to the different source differences considered, i.e. the mean relative bias of the global radiation at surface ($\Delta_{\%}DSSF$), the difference in the ratio between LAI and Rsmin (Δ *LAI/Rsmin*) and the difference between the High Vegetation Fractions considered (ΔCVH). The time series of this index based on the 3-hourly averaged European images (9 to 12 UTC) is shown in Figure 38. Solar radiation is the main source of ET difference, as expected, especially during summer. The ratio between LAI and Rsmin is the second main source of ET differences, when observing the magnitude of the 2-D correlation. The time series of this index shows an important seasonal behavior. For ECMWF, it correlates during spring and autumn, and slightly anti-correlates in summer. For GLDAS, the stronger correlation is found during summer. In these periods, the difference in the ratio LAI/Rsmin seems to explain the differences in ET. Land cover differences play a minor role at global scale, but noticeable for the ET difference with ECMWF.

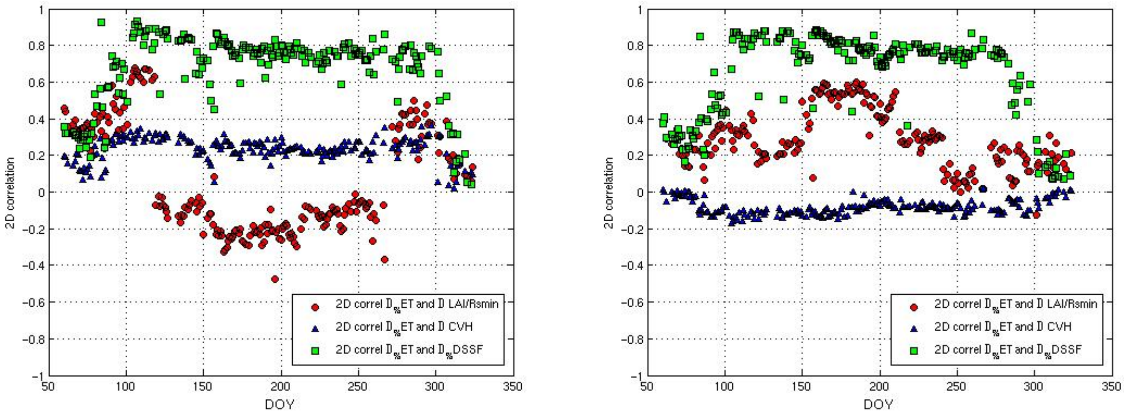


Figure 38 Time series of the 2-D spatial correlation between 3-hourly images for 9:00 to 12:00 UTC of the mean relative bias in ET and differences in LAI/Rsmin (green squares), differences in High vegetation percentage (CVH, blue triangles), and mean relative bias in global radiation at surface. At left: results from the comparison between LSA SAF MET v4.0 and ECMWF. At right: results from the comparison between LSA SAF MET v4.0 and GLDAS.

Differences in global radiation at surface is the main source contributing to differences in ET at each time step, even if there is no systematic bias in global radiation products. It acts as a short-term source of differences. Differences in vegetation characteristics, i.e. ratio between LAI and Rsmin, are the second source of ET differences, especially in spring/autumn for ECMWF and in summer with GLDAS. It acts as a medium-ranged-term source of difference, because it biases ET estimates at a monthly time scale. Land cover differences play a smaller role in the explanation of ET difference, because regions for which differences in land covers appear are quite spatially limited. However, it acts as long-term source of differences.

For the other two regions (North and South Africa), we present only the time series of spatial correlation, as it was shown for Europe. In the following paragraph, $\Delta_{\%}ET$ denotes the mean relative bias in ET, $\Delta\%DSSF$ stands for the mean relative bias in global radiation, ΔCVH is the absolute difference in high vegetation percentage and $\Delta\%LAI/Rsmin$ means the mean relative bias in the ratio LAI/Rsmin. For each variable, positive value indicates lower value in MET v4.0 compared to ECMWF or GLDAS. If the *%* symbol is indicated, the variable is normalized by the MET v4.0 value; otherwise the variable is not normalized.

Global spatial correlation of $\Delta_{\%}ET$ with $\Delta\%DSSF$, ΔCVH and $\Delta\%LAI/Rsmin$ are computed at two different time scales: at daily time scale and over 7 months, using for both the three-hourly means before noon. In Figure 39, the evolution of the 2-D correlations between April and October 2007 is computed. In Table 3, however, the time scale is the entire period of 7 months.

Looking at both time scales must give some hints about main sources of differences acting at short term and at long term.

For the North African window, the two major sources of differences between MET v4.0 and ECMWF ET at both time scales are *Δ%DSSF* and *Δ%LAI/Rsmin*, the latter one having a large influence during spring, but decreasing until the autumn. $\Delta_{\%}ET$ with GLDAS indicates *Δ%DSSF* as the main driver of the differences, followed by *Δ%LAI/Rsmin*, at short time scale, and by *ΔCVH* over the entire period. A seasonal cycle can be observed in top right side of Figure 39 for the curve of the 2-D correlation with *Δ%LAI/Rsmin*, with a positive correlation with ET differences in summer and a negative one during autumn. This cycle in 2-D correlation can explain why *Δ%LAI/Rsmin* does not explain $\Delta_{\%}ET$ at larger time scale. For the South African window at the bottom of Figure 39, best 2-D correlation with $\Delta_{\%}ET$ is observed with *Δ%DSSF*, as for the North African window, at both time scales. Differences appear when comparing with ECMWF or GLDAS. In MET v4.0 and ECMWF comparison, *Δ%LAI/Rsmin* is the second dominant cause of $\Delta_{\%}ET$ at both time scale. Comparison with GLDAS at short time scale in bottom-right Figure 39 shows opposite correlation evolution for *Δ%LAI/Rsmin* and *ΔCVH* with $\Delta_{\%}ET$. The spatial correlation is weak, but still present at longer time scale.

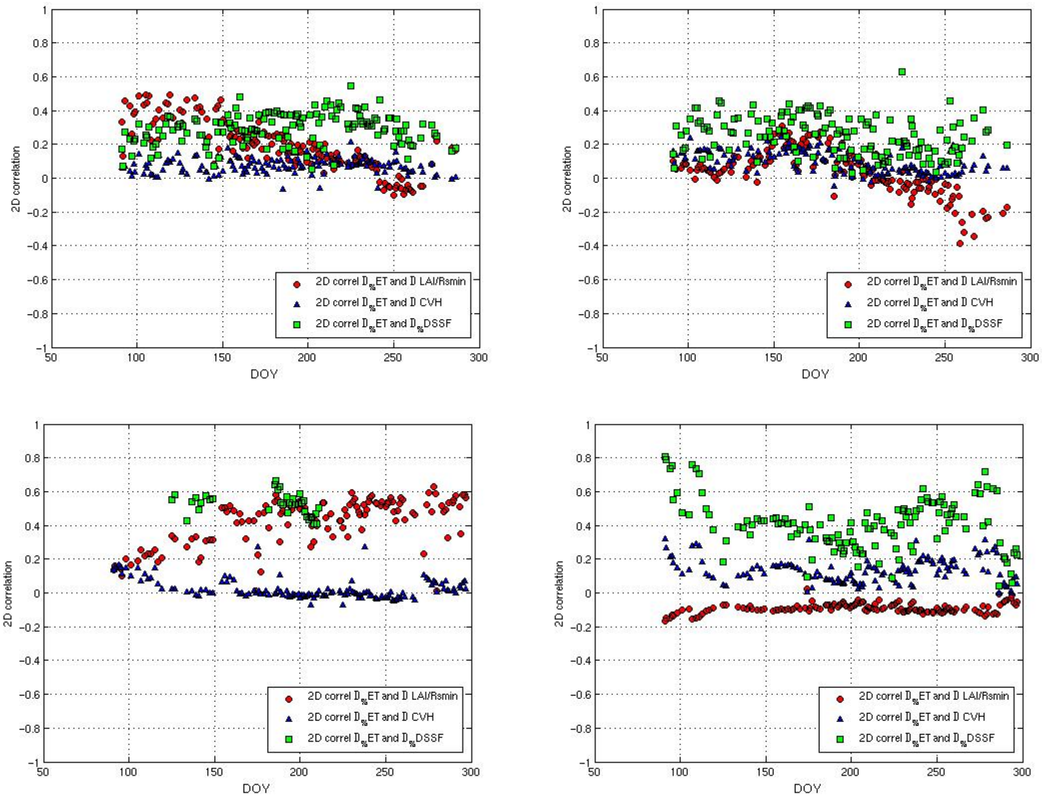


Figure 39 Time series of the 2-D spatial correlation between 3-hourly images for 9:00 to 12:00 UTC of the mean relative bias in ET and differences in LAI/Rsmin (red circles), differences in High vegetation percentage (CVH, blue triangles), mean relative bias in global radiation at surface (green squares) and differences in averaged soil moisture (SM, yellow triangles). At left: results from the comparison between LSASAF MET v4.0 and ECMWF. At right: results from the comparison between LSASAF MET v4.0 and GLDAS. At top, for North African window, at bottom for the south African window.

NAfr

	LAI/Rsmin	%High vegetation	Solar radiation
MET-ECMWF	0.31	0.12	0.42
MET-GLDAS	-0.02	0.19	0.48

SAfr

	LAI/Rsmin	%High vegetation	Global radiation
MET_ECMWF	0.54	0.01	0.65
MET_GLDAS	-0.09	0.10	0.38

Table 3 2-D spatial correlation between mean bias in ET and the mean differences in LAI/Rsmin, High vegetation percentage and in global radiation for North and South African windows.

2.3. Daily product (DMET)

Inter-comparison of the daily product was carried out by comparing the LSA-SAF DMET product to daily-cumulated ET (from 3-hourly forecast) from ECMWF model, for December 2009. In Figure 40, daily DMET estimates are plotted next to ECMWF corresponding images for qualitative interpretation. In figure 41, examples of DMET product over full MSG disk are presented for 12 days in September 2010. Figure 42 represents the distribution of DMET values, over the four MSG windows, for December 2009 and on Figure 43, the spatial (2D) correlation between DMET and ECMWF estimates over the four LSA-SAF windows are presented. On this figure we see that spatial correlation between images is high (between 85% and 95% for the studied period, over the four windows).

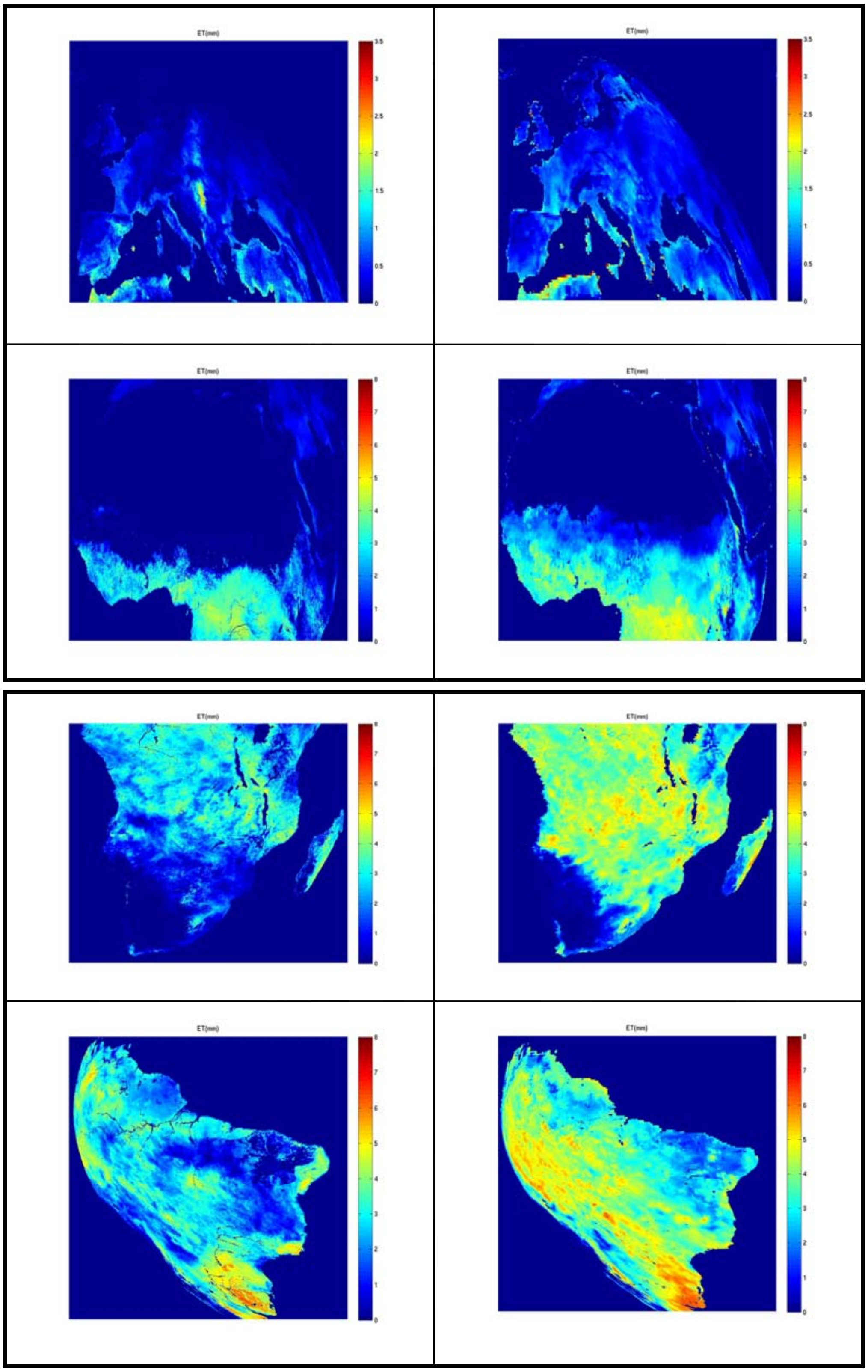


Figure 40 Daily-cumulated evapotranspiration values for the 1st December 2009 over Europe, North Africa South Africa and South America. Images on the left are from LSA SAF DMET product and on the right from ECMWF.

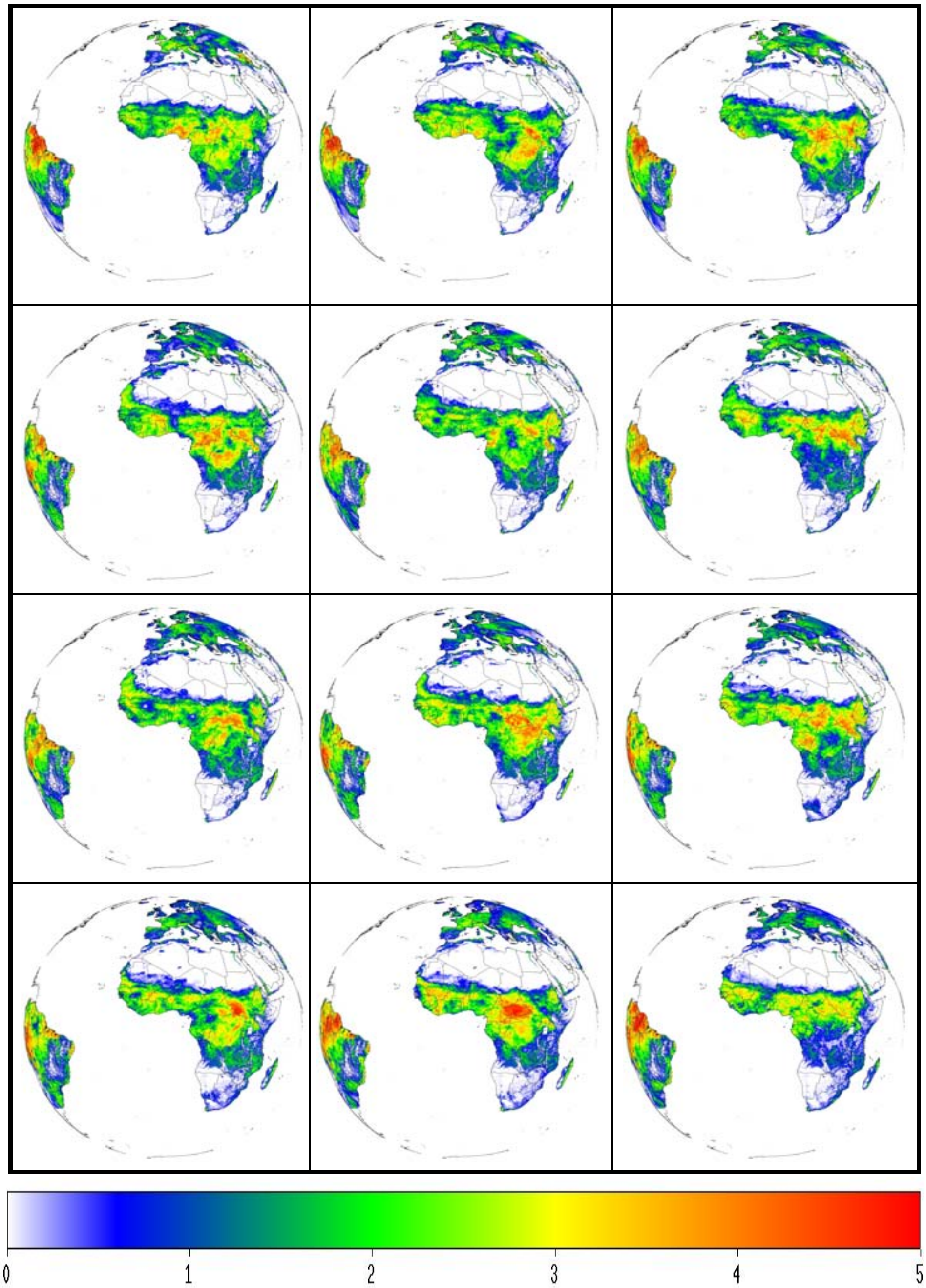


Figure 41 Example of daily evapotranspiration product (DMET) over MSG disk for the period between 1st (top-left) and 12th (bottom-right) September 2010.

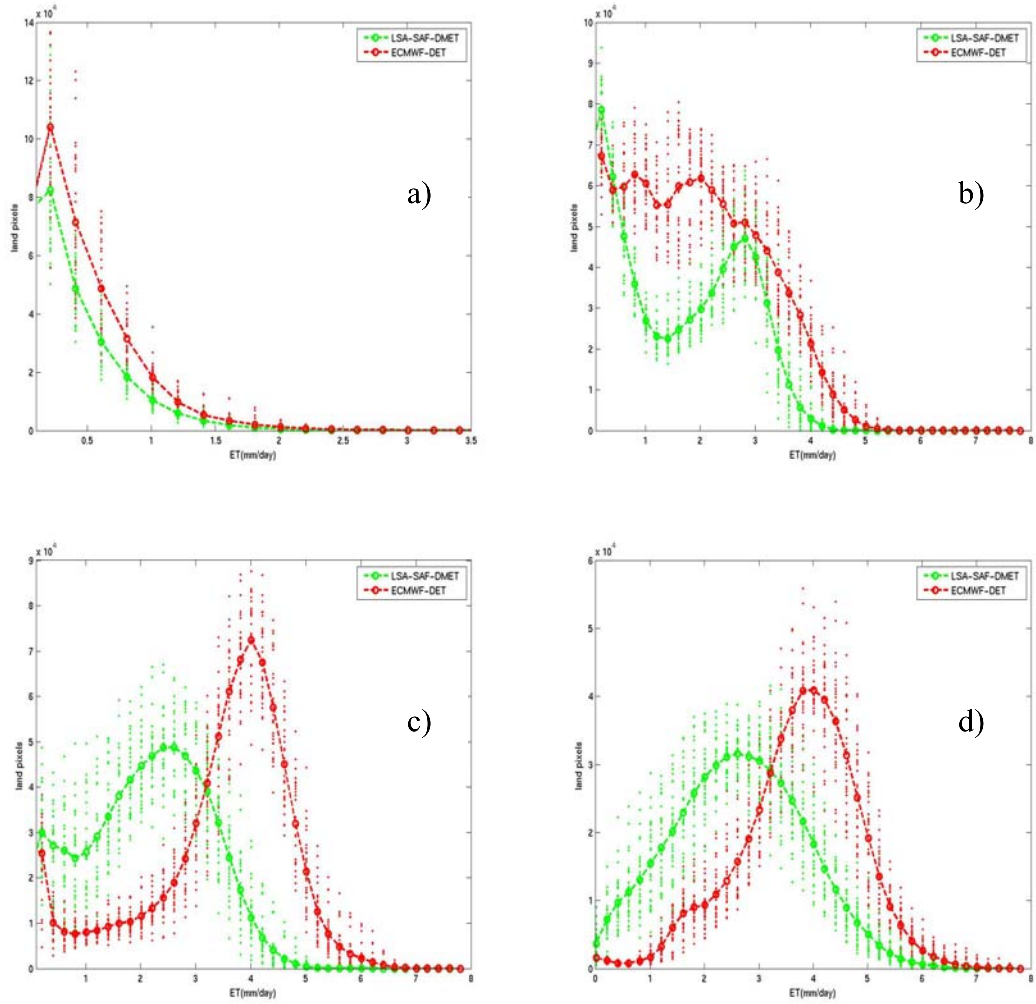


Figure 42 Distribution of average DMET (mm/d) and daily-cumulated ECMWF ET (mm/d) values over the European (a), North African (b), South African (c) and South American (d) windows, for December 2009. In red values from ECMWF, in green LSA-SAF DMET values. Small points represent the distribution of daily values and curves in bold are averages for the whole month.

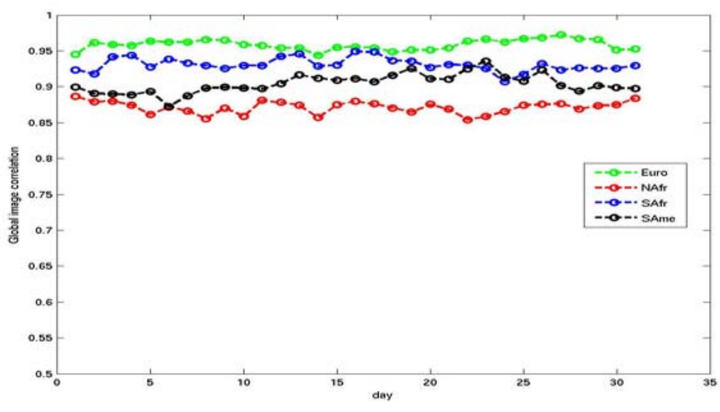


Figure 43 Spatial correlation of daily images from DMET and ECMWF for the month of December 2009 for: Europe (green), North Africa (red), South America (black) and South Africa(blue).

2.4. Discussion

The global image analysis shows that LSA SAF MET estimates are in equivalent range with estimates from the two selected models (ECMWF and GLDAS), with a spatial correlation between 80% and 95% for midday images, throughout the whole period, i.e. 01/03/2007 to 30/11/2007, and for all regions (Figure 26, Figure 27, Figure 28, Figure 29). While similarity with GLDAS is observed in case of low solar co-zenithal angle, i.e. early spring/late autumn and morning/evening, summer estimates correlates better with ECMWF, as suggested from Figures 15 and 16. A slight bias is found with ECMWF, which is not noticeable with GLDAS (Figure 22). The bias with ECMWF is correlated with a slight bias in global radiation at surface (Figure 34 and Figure 35).

From a spatio-temporal analysis, we clustered the different geographical regions where difference in time series is noticeable (Figure 30, Figure 31, Figure 32, Figure 33). Systematic biases were identified for few regions in Europe, but more in Africa where biases are correlated with arid regions. For South America, a global bias is found for the considered period. Most of the differences observed are not systematic, since a large disparity between ECMWF and GLDAS exists. This implies that no further inference on LSA SAF MET v4.0 quality can be performed on the basis of inter-comparison. However, most of the ET differences can be explained in terms of differences of input variables/parameters, i.e. incoming global radiation at surface (Figure 34, Figure 35), land cover (Figure 36) and resistance to transpiration of the canopy (Figure 37), linked to LAI. While global radiation at surface is the main source of difference on short-term basis, vegetation characteristics act on long-term basis and cause major ET biases observed (Figure 38, Figure 39).

The comparisons of daily product (DMET) with ECMWF daily-accumulated data confirm the results obtained during the validation of instantaneous product (MET). I.e., the spatial correlation between DMET/ECMWF images remains high (between 85% and 95%, Figure 43) for the studied period, over the four windows. A higher spatial correlation is observed over Europe, indicating that general patterns of ET are common to the two models. The correlation is quite constant during the whole period as it was already the case for instantaneous product during the year 2007. The mean distributions of DMET values for Europe (figure 42a) is equivalent to the mean distribution of 3 hourly averages presented for instantaneous MET results for end of autumn (figure 22). A negative bias is observed by comparison of DMET to daily ECMWF values, mainly in the Southern hemisphere. Validation is pursued on MET and DMET results. Further improvements of the ET algorithm will have positive impact on both MET and DMET products.

We can conclude that LSA SAF MET European estimates behave in a reasonable range compared to ECMWF and GLDAS. Most of the differences between models output have been attributed to input variables/parameters differences, indicating that models behaviours are globally similar.

3. Intrinsic limitation of the algorithm

3.1. Consistency with used LSA SAF products (DSSF, ALBEDO, DSLF)

Two consistency checks are performed in order to, first, verify that the quality index provided by each of the three products is correctly propagated to MET quality flag, and, secondly, check on example basis the spatial correlation between ET and DSSF, which is one of the main driver of land surface evapotranspiration.

Quality information of the MET algorithm v4.0 reflects the quality indices provided by DSSF, DSLF and ALBEDO products. On one example, we can check if the quality flags provided by MET algorithm v4.0 are consistent with quality flags of the three LSA SAF input products. One case is considered in Figure 44, corresponding to the 15/05/2007 at 12:00 UTC. A zoom over Baltic Sea region allows us to visually better understand the correspondence between the quality flags. Colours codes for each image are listed in Table 4, by quality class of the MET v4.0 algorithm.

ET	DSSF	DSLF	ALBEDO
Dark red (Good quality)	**Medium blue** (good quality)	**Dark orange** (Good quality)	**Light green** (ERR<=50%)
Yellow **Orange** **Light green** (Medium quality)	**Light Blue** (Medium quality) --------------------- ---------------------	--------------------- --------------------- **Green/Dark Blue** (Medium quality)	--------------------- **Dark red** (ERR>50%) ---------------------
Dark blue (Not processed)	**Dark Blue** (not processed)	**Dark Blue** (not processed)	**Dark Blue** (not processed)

Table 4 Codes of colours for reading image of Figure 44. For each quality class of the MET v4.0 algorithm, the conditions in terms of quality in input products are listed accompanied with the colour associated in the corresponding image.

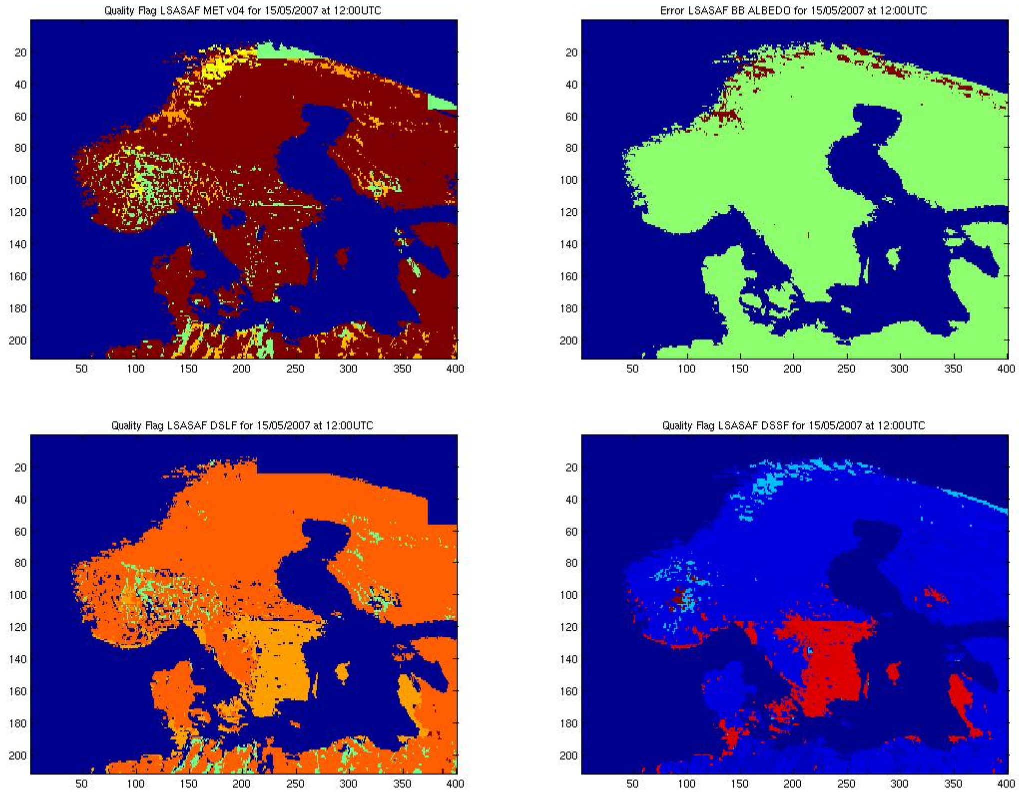


Figure 44 Zoom on the European Baltic Sea region for the date 15/05/2007 at 12:00 UTC of the quality flags and errors information of LSA SAF ET, ALBEDO, DSLF and DSSF products. Top left: quality flag of the MET v4.0 algorithm; Top right: error of the broadband albedo separated into 2 classes (>50% and <=50%); Bottom left: quality flag of the DSLF product; Bootom right: quality flag of the DSSF product. Legend of colours used in each image is described in Table 4.

Examining in details Figure 44, we see that relevant quality information from LSA SAF input products is correctly reflected into MET quality information. Spatial consistency between MET and DSSF, one of the main drivers of the model, is assessed for the day 06/04/2007 at 12UTC. European images are compared in Figure 45, as well as a sub-window for finer inside look in Figure 46. In both figures, we can observe the effect of clouds on DSSF values and consequently on ET estimates.

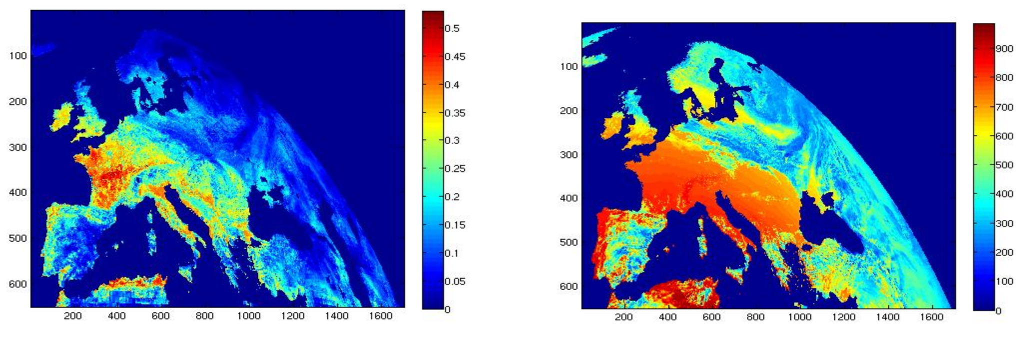


Figure 45 ET [mm/h] (left) and DSSF [W/m²] (right) images over Europe for the day 06/04/2007 at 12 UTC.

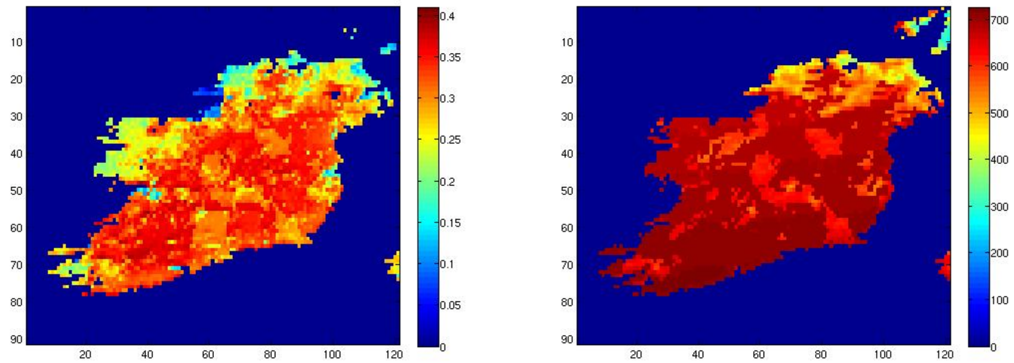


Figure 46 Zoom on ET [mm/h] (left) and DSSF [W/m²] (right) over Ireland, for the day 06/04/2007 at 12 UTC.

Main features of DSSF are correctly represented in generated MET output. Although some minor differences that testify of the impact of remaining forcing variables (DSLF, wind speed, air temperature, air humidity, soil moisture, vegetation) on the ET product can be observed.

3.2. Consistency with unused LSA SAF products (LST, FVC, LAI, FAPAR)

- **LST / MET products Consistency**

LSA-SAF LST is not used in the current version of the LSA-SAF MET algorithm; nevertheless, it should be consistent with the skin temperature (T_{skin},) generated by the MET algorithm. A way to assess this consistency is by comparing LSA-SAF LST (black cross) to the 'skin temperature' (T_{skin}, red triangle) computed by MET algorithm at a selected set of locations. The computed 'skin temperature' is neither radiative temperature nor aerodynamic temperature, but a mixture of both definitions (cf. ATBD). Physical interpretation of LST, derived by split-window technique, is not straightforward. While LST and T_{skin} correspond to different concepts and therefore should not be equal, we expect a coherent evolution between both. Comparisons of time series are shown for the pixels encompassing Carpentras, Evora and Melle (Table 5) in Figure 47. On each figure, RMS is shown, as well as the size of the sample.

Location Name	*Latitude*	*Longitude*	*Tile 1(%)*⁺	*Tile 2(%)*⁺	*Tile 3(%)*⁺
Barrax	39.04°N	2.09°W	C3 (61.3%)	IC (31.8%)	C4 (6.8%)
Carpentras	44.08°N	5.04°E	C3 (87.4%)	BF (10.1%)	R (2.4%)
Evora	38.47°N	8.00°W	C3 (90.0%)	C4 (10.0%)	-
Melle	50.98°N	3.82°E	C3 (71.7%)	BF (15.7%)	R (12.4%)
Valencia	39.56°N	1.27°W	C3 (37.8%)	BF (33.1%)	BS (28.9%)

Table 5 Location of the selected pixels for LSA-SAF products consistency check, with ECOCLIMAP/MSG vegetation partition.

⁺Broadleaved Forest (BF), C3 Crops (C3), C4 Crops (C4), Irrigated Crops (IC), Bare Soil (BS), Rocks and Urban material (R).

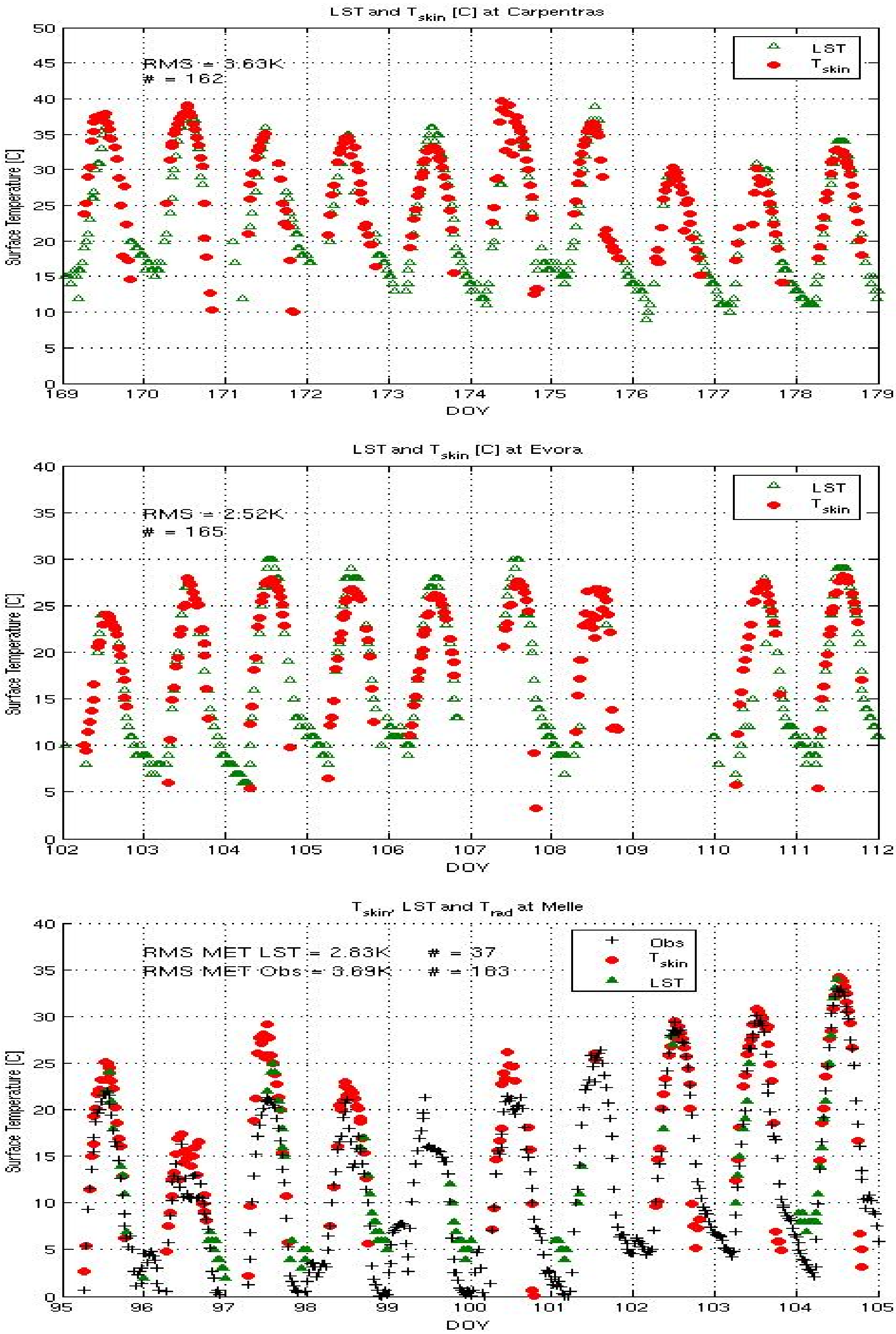


Figure 47 Comparison of LSA-SAF LST (Δ) and 'skin temperature' of the LSA-SAF MET algorithm (o) for Carpentras, Evora and Melle pixels, for the period 05/04/2007 to 15/04/2007. Data from Melle station has been added (+). The root-mean squared differences (*RMS*) and the number of events used for statistics (*#*) are indicated.

Globally, the two variables are relatively well comparable (RMS between 2.5 K and 3.5 K) for the period considered, keeping in mind that error associated to LST is of the order of 2K and that uncertainty of input variables has an impact on T_{skin}. In Evora, LST bias has been evaluated to 3.0 ± 1.5 K during the day (Kabsch et al., 2008). However, since LST and MET algorithms use different land cover maps and vegetation parameters, it is obvious that differences are expected. For the station of Melle (Belgium), radiative temperature is computed from the long-wave upward radiation measured at the station over grassland. Although the tile 'grassland' is not represented in this pixel, we can easily compare with the pixel estimate, dominated by C3 crops, because, for this period, vegetation characteristics behave in a similar way. Most of time, LSA-SAF LST and T_{skin} from LSA-SAF MET agrees very well with the in-situ data. For some days, LST presents a different diurnal evolution. These differences can be due to a different characterization of surface, cloud contamination or viewing angle effect.

In Figure 48, scatter plots and statistics of LST and T_{skin} computed at Carpentras, Evora and Melle for the total validation period, i.e. 01/03/2007 to 30/11/2007 are presented. For Melle pixel, LST and T_{skin} evolves in the same way, and are therefore consistent. In Evora, we observe in the scatter plot that for small values of LST, corresponding to morning/evening, spring and autumn, LST is very close to T_{skin}. However, for summer estimates, there is a evident bias, with increasing differences for the highest absolute values. This behaviour was already observed at Tojal station (Figure 2**)**, What could suggest that soil moisture modelled by ECMWF for this region is probably to high. In Carpentras, no systematic difference can be observed. However, the root-mean squared difference computed is larger than at Melle and several estimates can have quite large differences, up to 20 K, when looking at points that are out of the scatter plot 'cloud'. While this fact can also be noticed for Evora, the effect is particularly pronounced at Carpentras, as shown in Figure 48. These 'spikes' features in the computed skin temperature of LSA-SAF MET for several midday periods are not realistic and are an artefact of the model in case of very low wind speed conditions, i.e. free convection. However, it seems not to affect at a significant level the estimation of the surface latent heat flux, as shown in Figure 49.

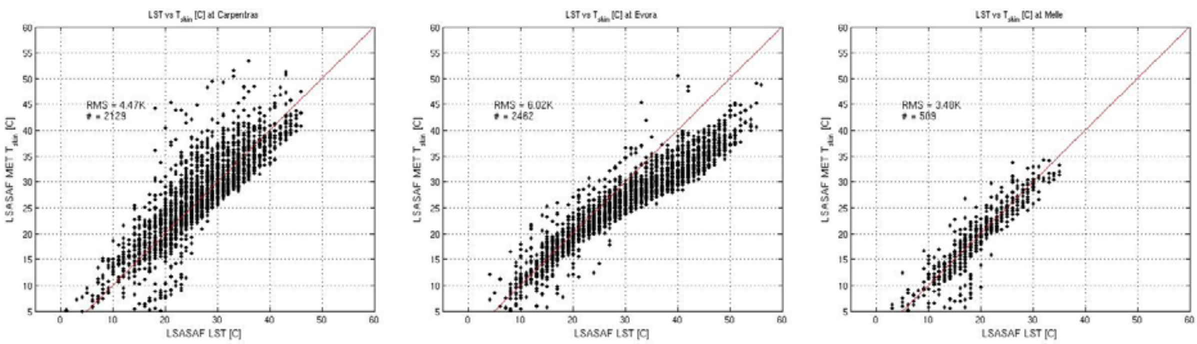


Figure 48 Scatter plot of LSA SAF LST versus LSA SAF MET T_{skin} for the pixels encompassing Carpentras, Evora and Melle. The root-mean squared difference (*RMS*) and the number of time steps used for statistics (*#*) are indicated.

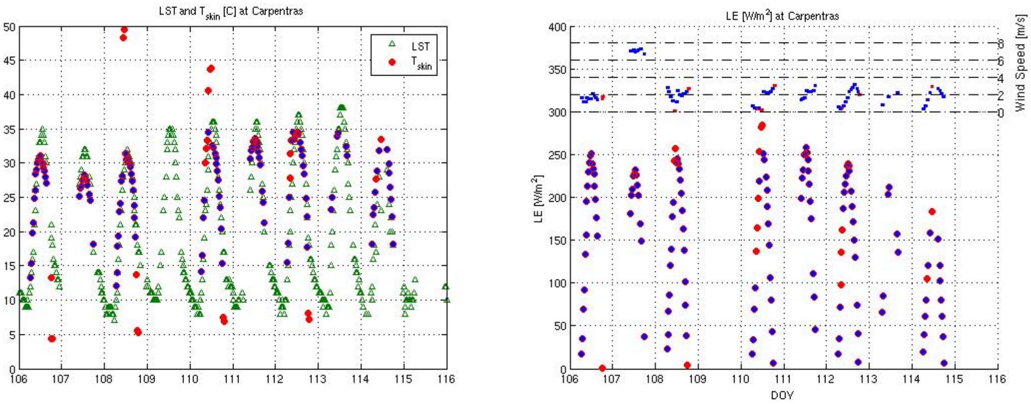


Figure 49 At left: time series comparison between LSA SAF LST ( ) and LSA SAF MET T_{skin} (O) for the period from DOY 106 to DOY 115 at Carpentras. Blue circles indicate difference with LST less than 7 K and, red circles, differences greater than 7 K. At right: time series of the LE calculated by LSA SAF MET v4.0 for the same location and period. Blue circles indicate events for which difference between T_{skin} and LST less than 7 K and, red circles, events for which differences are greater than 7 K. Wind speed is shown on the same figure, with the same colours convention.

In addition, we perform a global comparison between the modelled skin temperature of MET v4.0 and the LSA-SAF LST over the two geographical windows covering Africa. Given that the definition of the two variables is different, a direct comparison on 30 minutes basis is not performed. Factors like wind speed can influence the modelled skin temperature, while LST is less sensitive to it. However, even if the variables are not equal, we can expect a similar general seasonal behaviour, related to vegetation changes and soil moisture status. Due to clouds and Sahara desert, LST is sparsely generated, especially over North Africa, and it is not suitable to perform a global spatial correlation. Heating rates, using images from sunrise to local noon, are calculated by linear least square fit for each pixel of the image, in order to condense the information provided by the two variables and to smooth the finer temporal variations. Heating rates images have been produced for 3 days in each month, the 1st, 10th and 20th, between April and October. Instead of spatial correlation, we perform linear regression on the temporal profile of heating rates for each pixel in the geographical windows. Maps with the residue of the linear regression are given in Figure 50. The residue gives information on the reliability of the linear relation assumption. Smaller it is, more reliable is the assumption. Large residue can correspond to an image of very dispersed points or a non-linear relation. We can therefore detect regions for which the linear relationship between skin temperature and LST heating rates does not hold. For North Africa, the residue is quite large in the Sahel region, in the region of Benin and in Morocco. But, for the rest of the image, excluding the Sahara desert, the residue is small. In South Africa, the image is quasi split into two parts between small and large residues. For the northern half, no deviation from linearity is detected, however, very large discrepancies appear in the southern part and especially at the West coast and also in the Madagascar Island. Since the skin temperature of the model is not conceptually the same as the land surface temperature provided by LSASAF, signal can potentially have different amplitudes or absolute values, it is why only the quality of the linear fit is shown and not the regression coefficients.

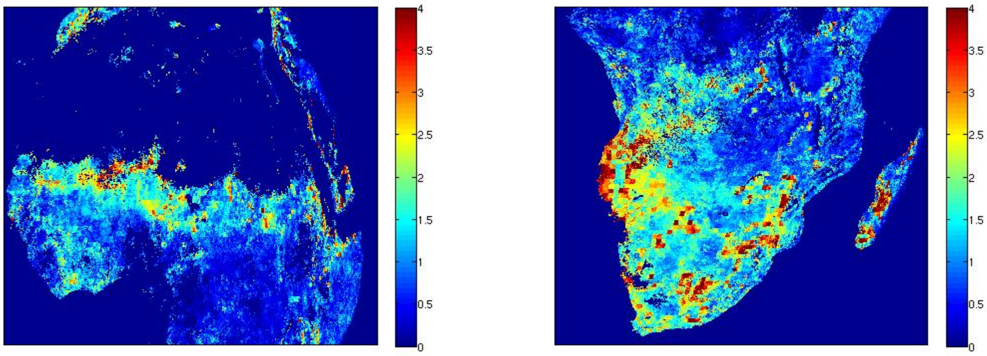


Figure 50 1°x1° maps of the residue of the linear regression, in K.hr^{-1}, between LSASAF MET skin temperature morning heating rate and LSASAF LST morning heating rate, for the two African windows, corresponding to left and right columns (from an images selection between April and October 2007, see text).

In addition to both consistency checks presented, a temporal correlation coefficient is computed for each pixel of the image between heating rates obtained from LST or skin temperature and FVC. The land surface temperature is often viewed as an indicator of soil moisture stress, if vegetation abundance is taken in account (Gillies et al., 1996; Moran et al., 1994; Prigent et al., 2005; Verstraeten et al., 2006), because of the strong negative correlation between both variables. Moreover, the abundance of green vegetation, parameterized here by FVC, is also a direct effect of the soil water available for plant respiration. Therefore, a strong positive or negative correlation is expected between heating rates of the surface and FVC. Figure 51, left image, represents the temporal correlation between heating rates of LST and FVC for the South African window. As expected, the anti-correlation in most of the window is observed. A strong positive correlation is noticeable near the equator, where soil moisture is not the driver of the vegetation evolution. A region with almost no correlation is found in the south-western part of South Africa. The corresponding image using the skin temperature is displayed at right. A large region with anti-correlation is found again. However, the southern region with almost no correlation is quite larger than in the left hand side image. It could possibly indicate that input soil moisture of the algorithm would have to be improved. We can also notice that the region of no correlation seems to be the same as the region found in Figure 32, where the diagnostics of models diverge.

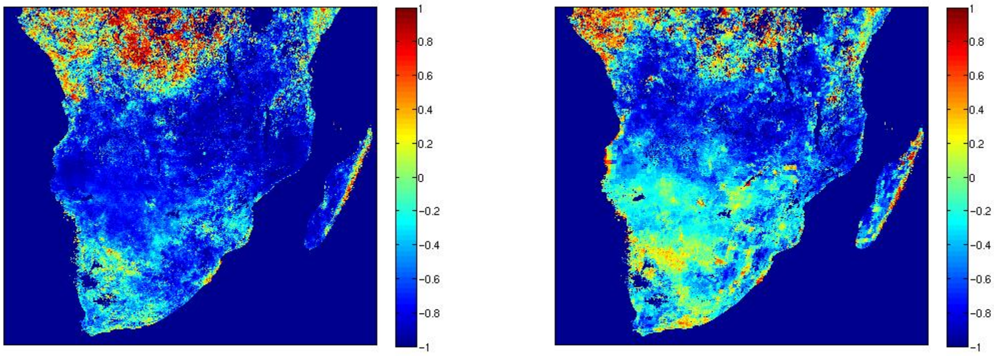


Figure 51 Map over the South African MSG window of the temporal correlation coefficient between LSA-SAF FVC and the morning heating rates of LSA-SAF LST (left) and of MET v4.0 skin temperature (right) (from an images selection between April and October 2007, see text).

- **VEGA / MET products Consistency**

Consistency of the product with LSA-SAF VEGA parameters, not used in the present version, is directly related to the vegetation parameters used in the MET algorithm, i.e. ECOCLIMAP parameters. Although comparisons at point locations have been performed, no results are presented, because this issue is most related to validation of ECOCLIMAP (based on MODIS) and VEGA products.

First, we perform spatial correlation between ET and LSA-SAF FVC. Comparison with LAI would provide similar results because these indices are not derived independently (cf. ATBD LSA-SAF VEGA). To ease the comparison, ET is scaled by the global radiation at the surface to remove any effects coming from clouds. The time series of spatial correlation is shown in Figure 52, for the three windows over Africa and South America. We can see that the spatial correlation is generally high, between 82% and 95%, and has a seasonal signal. It tends to increase until the autumn, as seen from the North African curve, and to decrease from autumn to spring, as seen from the South African curve.

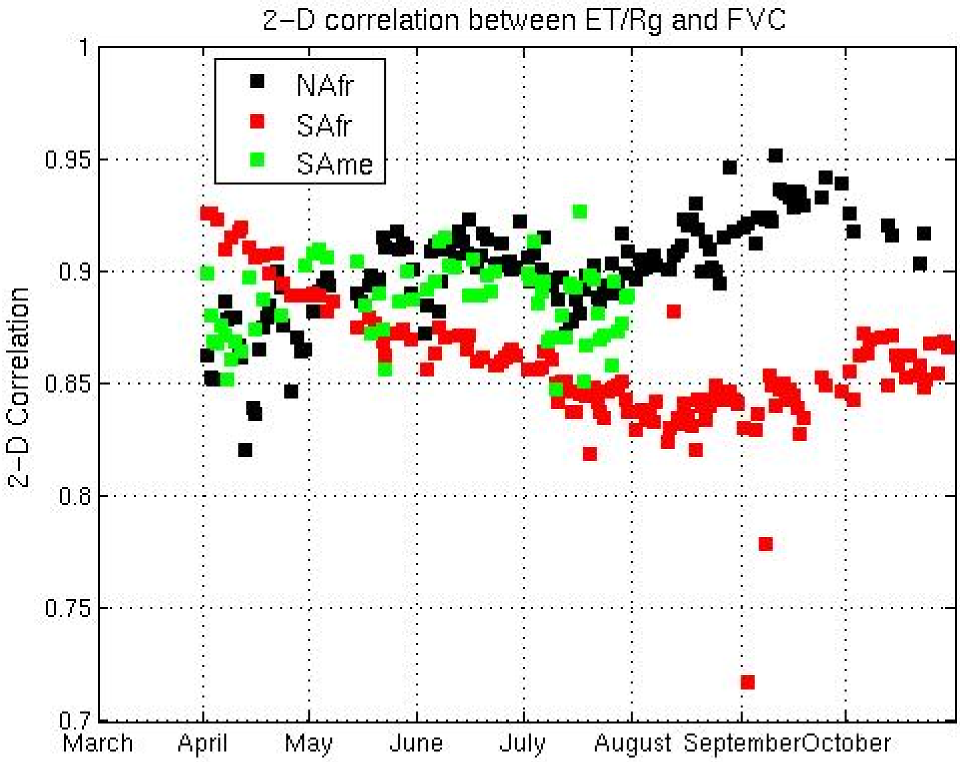


Figure 52 Spatial correlation between ET scaled by solar radiation at 12:00 UTC and LSA-SAF daily FVC for North and South Africa and South America, black, red and green squares respectively.

3.3. Systematic errors reported from the follow up of the product

Systematic gap-filled areas

For the detection of systematic non-processed areas, we analyze all ET Quality Flag images from 01/03/2007 to 30/11/2007, and cumulate the rates of occurrence on the European map in Figure 53 for the ‘Poor Quality’ flag.

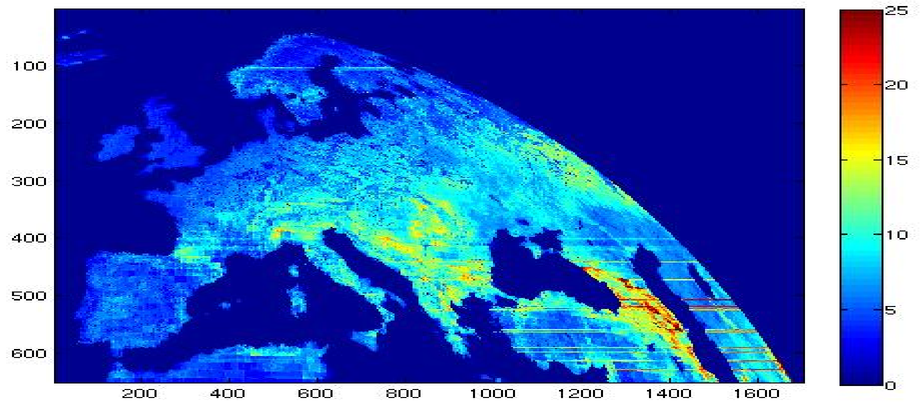


Figure 53 Map of occurrence (0-1) of gap filling “Poor Quality” flag for the period 01/03/2007 to 30/11/2007.

Figure 54 Shows that the relative amount of ‘Poor Quality’ flags by pixel is in average around 10%. The regions most affected by systematic attribution of this quality flag are Central Europe and North-East of Turkey, with a rate of 15% and 20% respectively. However, this quality flag is not distributed uniformly through the whole period and varies within a single day. While, in average over the whole period considered, at 18:00 UTC 22% of land pixels are flagged “Poor Quality”, only 1% of land pixels at 10:00 UTC are labelled with this flag. In Figure 54, we can conclude that Poor Quality Flag is attributed mostly for early morning and late afternoon, for which solar radiation is quite low. We also remark that, on monthly basis, around the half of the ET estimates for which DSSF is zero is flagged “Poor Quality”. This is because there does not exist any discrimination between day and night in the quality flag of the MET v4.0 algorithm.

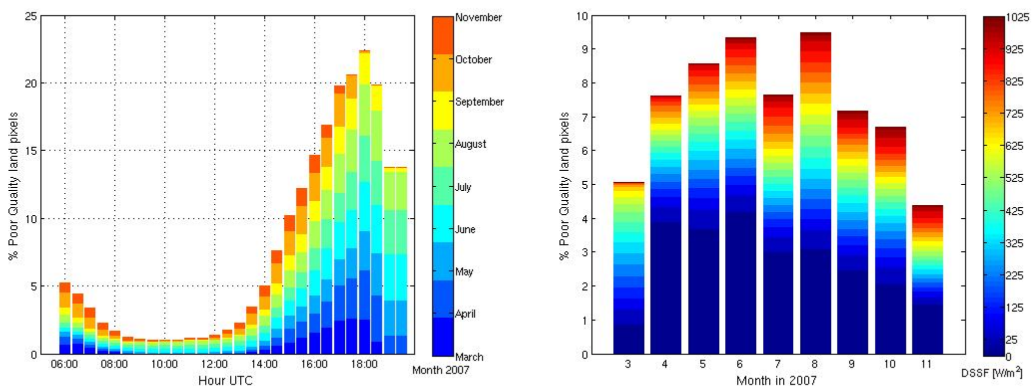


Figure 54 Left: For each hour of the day, percentage of land pixels flagged “Poor Quality”. Contribution of each month is represented. Right: For each month, percentage of land pixels flagged “Poor Quality”. Contribution by slice of DSSF is represented.

Bad/reprocessed input data
Meteorological input for LSA-SAF MET algorithm is processed by the pre-processing package developed at IM for LSA-SAF. 2-m temperature and dew point temperature are linearly interpolated in all directions, but a simple nearest neighbour algorithm is used to project the other ECMWF input variables onto MSG grid (wind speed, …). Spurious features, like big rectangles, are recognizable, in some meteorological conditions, on LSA-SAF MET images mainly due to the fact that wind speed and soil moisture are not interpolated from ECMWF to MSG spatial resolution. Moreover, soil moisture is not available for every land pixels (for instance, at the west part of Sicily, and at the east coast of Spain), or produces not land representative estimates and therefore leads to a non-processed ET estimate for those pixels. This problem is related to a misinterpretation of the 0.25°x0.25° ECMWF product available for users, which is obtained by interpolation from the original Gaussian grid.

3.4. Known caveats of the methodology and evaluation of the impact on the results

1. Approximations related to the vegetation parameterization and land cover
-Intra-class variability of the control stomatal resistance

Version 4.0 of the LSA-SAF MET algorithm uses, for each ECOCLIMAP class, one associated minimum stomatal resistance value. Therefore, it is assumed that there is almost no variation within this particular class concerning minimum stomatal resistance, while studies show that it depends on the species and age of the vegetation. Although standard SVAT models use calibrated parameters to give locally the best results, it cannot be done at large scale and a compromise has to be found. The degree of accuracy of that common assumption of regional SVAT models is evaluated by means of the comparison with in-situ observations. See for instance the case of "grassland" p.12.

-Use of three main vegetation classes per MSG pixel

For each MSG pixel, the computation of ET is restricted to the 3 main vegetation classes from ECOCLIMAP for which the proportion is adapted to cover 100% of the pixel. However, in ECOCLIMAP, the proportion of MSG pixels covered by more than 3 different vegetation classes is small, and generally contains vegetation classes with a very small percentage.

-Monthly evolution of the vegetation

From the 10-days averaged evolution of vegetation characteristics, i.e. leaf area index, fraction of vegetation cover, roughness length, ECOCLIMAP/MSG averaged vegetation fields to be monthly varying. In average, this simplification is sufficient to reach the quality criterion proposed, and to be competitive with other ET products. However, in attempting to validate at pixel scale, we have shown the limitation of this approximation for the vegetation types with a well-marked phenological cycle, as for deciduous forest (Hesse). Future research is foreseen on this question (see section 5).
-Averaging of the vegetation characteristics of the same elementary class

The original ECOCLIMAP land cover contains a classification in ecosystems, like woody savannah, agro-forestry areas, etc. The methodology proposed by Masson et al. (2003) aims to separate these ecosystems into a small number of elementary vegetation types, for example grass, deciduous broadleaved forest, etc. Therefore, a vegetation type belonging to the same elementary class can be issued from different ecosystems, with different vegetation characteristics. When ECOCLIMAP was adapted for LSA SAF MET algorithm, vegetation characteristics for each elementary class were averaged inside a MSG pixel. This introduces a small error for the MSG pixels where it occurs, i.e. at transitions between composite ecosystems.

-Snow cover

In the current version, snow sublimation is not modelled for permanent snow: the pixel is labelled as not processed. For snow events, snow sublimation is not modelled, but evapotranspiration from the vegetation is considered instead. However, the error that is induced by this modelling effect is relatively low for vegetated areas. Indeed, when persistent snow occurs, soil temperature is generally less than 273.15K, and soil moisture available for transpiration of vegetation reduces to zero. The remaining error consists of the difference between occurring snow sublimation and computed evaporation rate from bare soil.

2. Pre- and post- processing assumptions

-Pre-processing

Since estimation of ET is conditioned by the existence of all the needed input, we consider filling non-processed pixels within DSLF images over Europe, occurring mainly at the border of clouds for partially covered pixels, due to missing information from the cloud mask used by LSA-SAF. The iterative filling procedure followed is described in the ATBD. Averaging DSLF between cloudy and clear skies therefore corresponds to consider a mix between clouds and clear sky over the partially covered pixel. Although the proportion of clouds in this pixel is not known for averaging, the error introduced by the 'equality' assumption will not be sufficient to have a large impact on ET estimates, as studied in the sensitivity analysis tests, but degrades the quality of ET estimates (Nominal to Below Nominal).

-Post-processing

A final gap filling procedure is applied on MET images to fill the non-processed pixels by an approximate estimate, flagged with poor quality, because no direct information is used for the filled pixel. While this procedure is suitable to fill the non-processed pixels with the approximately same vegetation composition than the surrounding pixels for small regions with the same meteorological conditions, this method fails to give estimates with the right order of magnitude when dealing other cases.

3. Choice of the algorithm input

Soil moisture is not computed by the LSA-SAF MET algorithm v4.0. Instead, 4-layered soil water content is taken from ECMWF model short-range predictions. While soil water content is not a very sensitive input for North of Europe, it becomes one of the main limiting factors to evapotranspiration for Southern part of Europe in summer. Recent studies have shown the deficiencies of soil moisture forecasts by the ECMWF model due to the 4D-var assimilation processes. A better and independent estimate of soil moisture is therefore preferable to assess a correct evaluation of ET in dry regions. We intend to explore this issue for a separate product, i.e. LSA SAF MEET (cfr. PRD).

4. Maturity, limitations of the product and domain of use/application

4.1. Maturity

Validation performed and reported in the previous chapters shows that good quality results are obtained over the European MSG window. For Africa and South America, the results are more mitigated, partly because of the lack of global knowledge on the processes occurring in these regions. In such regions, comparing with other models is not an effective way to validate, but only a hint, if it is assumed that the other accepted models rely on the most up-to-date information on land surface processes in these regions. In-situ validation data in Africa and South America is quite scarce, however, with the growing number of initiatives in Africa, like AMMA, GLOWA-Volta and CarboAfrica, new possibilities are offered for the future validation exercises.

4.2. Domains of use/application

Applications based on LSA-SAF MET/DMET products are expected for:

- Regional ET estimation
- Hydrological applications
- Environmental monitoring purposes
- Assimilation in hydrological and crop growth models
- Long-term studies on evapotranspiration evolution related to climate.

4.3. Added value

The added-value of the product relies on: 1) applicability on a large domain, while empirical ET algorithm would not be applicable; 2) use of up-to-date quasi-real time remotely sensed products, like radiation, to insure a continuous follow-up of the ET evolution through days for all nebulosity conditions (clear, cloudy or overcast cases) with low impact by forecast errors; 3) an extensively validated product through information available in-situ, determining exact capabilities of the proposed method; 4) a uniform methodology applied on the full MSG disk, at a temporal resolution of 30 minutes and a spatial resolution comparable to LAM output.

5. Planned science activities

Based on the findings during the validation exercise, two main axes will be followed to improve the general quality and reliability of MET/DMET products. The planned science activities focus mainly on instantaneous (MET) products, assuming that any improvement on this product will automatically improve the daily product.

The first axe will concern the improvement of the knowledge on vegetation state. For this purpose, LSA-SAF LAI and FVC products are used in combination to the ECOCLIMAP (Masson et al., 2003) ecosystem land cover map. Benefits of using LSA-SAF VEGA products for assessing the evapotranspiration are directly related to the close monitoring of vegetation using remote sensing techniques applied for geostationary satellites. It allows, for example, taking into account daily variations in vegetation characteristics, at a relatively fine spatial scale. Detection of the inter-annual variability of the signal is also an advantage of using remote sensing technique. As presented in Ghilain et al (2008), using LSA-SAF VEGA products is an improvement compared to the use of ECOCLIMAP vegetation monthly database considering two different aspects. The first one is the finer temporal and spatial resolution of LSA-SAF VEGA products compared to ECOCLIMAP that allows a better spatial distribution of the evapotranspiration. The second one is the reduction of the uncertainty of the vegetation parameters and consequently the reduction of the uncertainty of evapotranspiration. A new version of MET algorithm, implementing the use of LSA-SAF VEGA products is under test.

The second axe of development is dedicated to improvement of the knowledge on soil moisture status. For this purpose, a full Soil-Vegetation-Atmosphere Transfer model has been developed. In practice, soil water and temperature evolution equations have been added to MET v4.0 algorithm, as well as rain interception and snow depth evolution. We have develop two versions, called RMI-2 and RMI-3, based on Viterbo et al. (1995) and Balsamo et al. (2008) respectively. By these developments we contribute to assess evapotranspiration in a more consistent way using self-computed soil moisture. However, since the model accumulates errors on the long term, because of modelling inaccuracy or biased input rainfall rates, it is intended to use METOP-ASCAT surface soil moisture to adapt on a regular basis the modelled soil moisture. This latter part is developed in collaboration with the Technical University of Vienna (Inter-SAF activity between LSA-SAF and H-SAF; de Crane et al., 2009).

An illustration of the new developments for improvement of soil moisture is given in Figure 55. For the African station Demokeya, presented in the first section of this report, RMI-3 is forced by ECMWF forecasts to produce time series of ET and soil moisture. These results, seen as the control run, are compared to local observations at the station. A second simulation using RMI-3 is performed for which soil moisture profile is re-initialized on a regular basis of 3 days using in-situ measurements. As expected, the regular re-initializations of the soil moisture with observations allow getting a better-simulated ET time series.

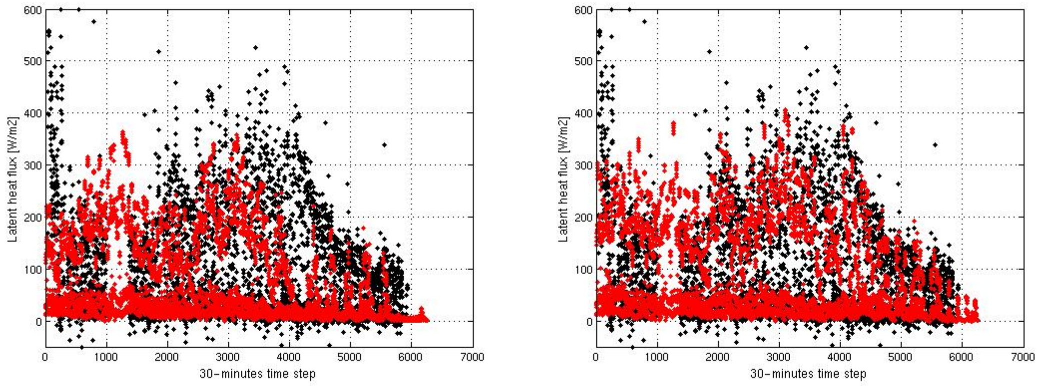


Figure 55 Comparaison of RMI-3 latent heat flux estimates at 30 minutes time step (in red) with in-situ measurements (in black) [W/m2] over the period ranging from 1st July to 30th October 2007, for the CarboAfrica Demokeya station (Sudan), at left. RMI-3 latent heat flux estimates where the model is re-initialized with soil moisture observation every 3 days is considered at right.

The use of the land surface temperature product, LSA-SAF LST, is also investigated. Search of the informative content of LST on the soil moisture state is performed. A first study of the retrieval of soil moisture content from LST and FVC observations has been presented at EGU 2009 (Ghilain et al, 2009).

At last, validation activities on Africa continue, especially in-situ validation, with a strong interaction with the AMMA and CarboAfrica programmes. For South America, we intent to establish contacts with the measurement network LBA (http://lba.cptec.inpe.br/lba/site/).

6. Conclusions

The validation of LSA-16 (MET) and LSA-17 (DMET) products was achieved by comparing the algorithm output (instantaneous and cumulated evapotranspiration) to evapotranspiration derived from measurements made at selected locations and by comparing the algorithm output to the output of models recognized to produce valuable meteorological information. Output from ECMWF model and GLDAS was used for the models inter-comparison. Table 1, summarizes the results of the in-situ validation, providing statistical indicators of the comparisons and percentage of cases where PRD criterion is satisfied. Points 1 to 6 resume the main conclusions of the validation report.

Station	Vegetation Type *	Bias	RMS	Corr	% PRD
Amplero	G	0.02	0.11	0.82	75.1
Buzenol	G	0.02	0.10	0.81	80.1
Cabauw	G	0.02	0.07	0.90	90.1
Humain	G	-0.04	0.08	0.90	83.2
Monte Bondone	G	0.02	0.12	0.76	77.6
Tojal	G	0.05	0.10	0.74	59.9
Hesse	DBF	0.00	0.09	0.56	89.9
Roccarespampani	DBF	-0.02	0.08	0.85	77.5
Loobos	ENF	-0.03	0.10	0.63	86.3
Wetzstein	ENF	-0.02	0.08	0.79	87.9
Sodankylä	ENF	0.08	0.12	0.46	63.8
Lonzée	C	0.03	0.09	0.73	73.4
Las Majadas	EMF	0.01	0.06	0.46	94.7
Puéchabon	EMF	-0.07	0.09	0.65	99.7
Kaamanen	B	-0.01	0.07	0.69	95.5
Vielsalm	MF	0.02	0.06	0.80	88.2
Demokeya	Ss	-0.12	0.18	0.40	87.6

Table 8. Summary of the comparison between output from MET algorithm and ET derived from measurements. Column 1 is the name of station; column 2, the vegetation type considered at station; column 3 the root mean square of the comparison; column 4 the correlation coefficient; column 5 the bias ; column 6 the percentage of steps for which PRD requirements are met.
** Grassland (G), Deciduous Broadleaved Forest (DBF), Evergreen Needle Forest (ENF), Crops (C), Mixed Forest (MF), Evergreen Mediterranean Forest (EMF), Ss (Sahelian savannah)*

1) Results of the in-situ validation summarized in table 8 show that for estimates flagged 'Nominal' or 'Below Nominal', the PRD quality criterion is satisfied to a rate higher than 70%. Globally, good agreement is found for stations at which close correspondence exist between land cover defined in ECOCLIMAP and the effective cover at the station; with the best agreement for stations over grassland and mixed forests. The model does not present systematic bias; nevertheless modelled ET is underestimated at a station (Demokeya) over African dry savannah. In general, LSA-SAF MET algorithm is able to reproduce the temporal evolution of evapotranspiration with values equivalent to observations.

2) From the models inter-comparison, it is concluded that ET estimates provided by the MET algorithm are equivalent to estimates provided by ECMWF and GLDAS, with spatial correlation between 85% and 95% for midday images. For high co-zenithal angles better correlation is found with ECMWF while for low angles (spring/late autumn and morning/evening) with GLDAS. Observed discrepancies between models estimates are explained by differences in models parameterization, radiation, land cover information and soil water content.

3) From the consistency check it is observed that uncertainties on AL, DSSF and DSLF used as input to MET are correctly reflected on MET quality flag. The comparison of morning heating rates from Tskin and LST highlighted regions of low /high correlations which correspond roughly to areas of large relative bias between MET and GLDAS ET. Given that relationships between FVC and LST/Tskin can provide some insight to soil water content, regions of low correlation indicate that there is still place for improvements related to soil moisture and/or vegetation parameterization.

4) For the validations of the daily product, the output of the DMET algorithm has been compared to daily-cumulated ET values at selected locations. Overall, DMET fairly matches the observed variations, with a very good agreement with observations for well-watered sites and a good seasonal variation for temperate forests.

5) The comparisons of DMET product to ECMWF daily-accumulated ET for December 2009, over the four windows defined inside the MSG FOV, confirm the results obtained during the validation of instantaneous product. I.e., the spatial correlation between the models estimates is high (85% to 95%) and remains quite constant during the analysed period.

6) The validation of the instantaneous and daily products (in-situ and models inter-comparison) provides higher scores for comparisons over Europe. For Africa and South America, the work must be continued in order to check and/or improve the quality of modelled estimations over areas affected by strong soil water stress. Further improvements of the MET algorithm will have positive impact on both MET and DMET products.

Acknowledgments

RMI team acknowledges all the data providers and their associates for their interest in this validation work and for their collaboration. We would like in particular to thank Nicola Arriga, Marc Aubinet, João Banza, Fred Bosveld, Bernard Heinesch, Eddy Moors, Christine Moureaux, Dario Papale, João S. Pereira, Casimiro A. Pio, Corinna Rebmann, Bernard Longdoz, Arnaud Carrara, Jan Elbers, Jonas Ardo and all the principal investigators and collaborators involved in the preparation and sending of the ground measured fluxes. We are also grateful to ECMWF and GLDAS centre for their model output used in this report. We thank in particular Matthew Rodell and Hiroko Kato. We also thank Gianpaolo Balsamo for providing the last version of the codes of TESSEL and H-TESSEL.

The research and developments achieved by the RMI team are funded by EUMETSAT in the framework of the SAF on Land Surface Analysis and by the PRODEX programme managed by the European Space Agency in collaboration with the Belgian Federal Science Policy Office (ESA/Contract Nr. 15066/01/NL/Sfe(IC)). Recent developments made about SM have been achieved in an Inter-SAF activity between H and LSA-SAF funded by EUMETSAT.

References

Aubinet M., A. Grelle, A. Ibrom, U. Rannik, J. Moncrieff, T. Foken, A.S. Kowalski, P.H. Martin, P. Berbigier, Ch. Bernhofer, R. Clement, J. Elbers, A. Granier, T. Grünwald, K. Morgenstern, K. Pilegaard, C. Rebmann, W. Snijders, R. Valentini and T. Vesala, 2000: Estimates of the Annual Net Carbon and Water Exchange of Forests: The EUROFLUX Methodology, *Advances in Ecological Research Vol. 30, pp. 113-175*.

Balsamo G., Viterbo P., Beljaars A., van den Hurk B., Hirschi M., Betts A.K. and Scipal K., (2008) A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System, ECMWF Technical Memorandum N°563, 30 pp.

Beljaars, A. C. M., and F. C. Bosveld, 1997: Cabauw data for the validation of land surface parameterization schemes, *J. Climate Vol. 10, pp. 1172-1193.*

Beljaars A.C.M., Viterbo P., 1994: The sensitivity of winter evaporation to the formulation of aerodynamic resistance in the ECMWF model, *Boundary-Layer Meteorol. Vol. 71, pp. 135-149.*

Blondin C., 1991: Prameterization of Land-Surface Processes in Numerical Weather Prediction. *In: T.J. Schmugge and J.C. André, Editors, Springer-Verlag, pp. 31-54.*

Cleugh H.A., Leuning R., Mu Q.Z. and Running S.W., 2007: Regional evaporation estimate from flux tower and MODIS satellite data, *Rem. Sens. Env., 106, pp.285-304*.

Courault D., Seguin B. and A. Olioso, 2005: Review on estimation of evapotranspiration from remote sensing data: from empirical to numerical modelling approaches, *Irrigation and Drainage Systems Vol. 19, pp. 223-249.*

de Crane, F., Hasenauer, S., Arboleda, A., Ghilain, N. and Gellens-Meulenberghs, F., 2009: Investigation of the relationship between ERS scatterometer derived superficial SM (H-SAF) and both LSA-SAF ET and ECMWF upper layer SM. Report of Inter-SAF activity between the LSA-SAF and the H-SAF, 40 pp.

Drusch M., 2007: Initializing numerical weather prediction models with satellite derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System, *J. Geophys. Res. Vol. 112, D03102, doi:10.1029/2006JD007478.*

Drusch M. and P. Viterbo, 2007: Assimilation of screen-level variables in ECMWF's Integrated Forecast System: A study on the impact on the forecast quality and analysed soil moisture, *Mon. Wea. Rev. Vol. 135, pp.300-314.*

Foken T. and B. Wichura, 1996: Tools for quality assessment of surface-based flux measurements, *Agric. For. Meteorol. Vol. 78, pp. 83–105*.

Foken T., M. Göckede, M. Mauder, L. Mahrt, B. Amiro and W. Munger, Post-field data qualtiy control. *In: X. Lee, W. Massman and B.E. Law, Editors, Handbook of Micrometeorology, Kluwer, Dordrecht (2004), pp. 181–208.*

Gellens-Meulenberghs F., 2005: Sensitivity Tests of an Energy Balance Model to Choice of Stability Functions and Measurement Accuracy, *Boundary Layer Meteorology, Vol 115 (3), pp. 453-471*.

Gellens-Meulenberghs F., Arboleda A. and Ghilain N., 2007: Towards a continuous monitoring of evapotranspiration based on MSG data. In "Remote Sensing for Environmental Monitoring and Change Detection", M. Owe and Ch. Neale eds., IAHS Publ. 316, 228-234.

Gellens-Meulenberghs, F., Arboleda, A.., Ghilain, N., 2008: Evapotranspiration assessment by LSA-SAF: methodology, status of validation and plans for the near future. Proceedings of the 2008 EUMETSAT meteorological satellite data user's conference, Darmstadt, Germany, 8th-12th September, 7 pp.

Gellens-Meulenberghs, F., Arboleda, A., Ghilain, N., de Crane, F. and Sepulcre Canto, G., 2009: Large scale evapotranspiration modeling using MSG SEVIRI derived data. Proceedings of the conference on Earth Observation and Water Cycle Science (ESA SP-674). Frascati, Italy, 18-20 November 2009, 5 pp.

Ghilain, N., Arboleda, A. and Gellens-Meulenberghs, F., (2008) Improvement of a surface energy balance model by the use of MSG-SEVIRI derived vegetation parameters. Proceedings of the 2008 EUMETSAT meteorological satellite data user's conference, Darmstadt, Germany, 8th-12th September, 7 pp.

Ghilain, N., Arboleda, A., Gellens-Meulenberghs, F., 2009: Use of geostationary satellite imagery in optical and thermal bands for the estimation of soil moisture status and land evapotranspiration. Poster presented at the EGU assembly, Vienna, 20th-24th April 2009, summary in Geophysical Research Abstracts (10) EGU2009-0, 1 pp.

Ghilain, N., Arboleda, A. and Gellens-Meulenberghs F., 2011: Evapotranspiration modelling at large scale using near-real time MSG SEVIRI derived data. Hydrol. Earth Syst. Sci., doi:10.5194/hess-15-771-2011, 15, 771–786.

Gillies R.R., Carlson T.N., Cui J., Kustas W.P. and Humes K. (1997), Verification of the triangle method for obtaining surface soil water content and energy fluxes from remote measurements of Normalized Difference Vegetation Index (NDVI) and surface radiant temperature, International Journal of Remote Sensing, 18, pp. 3145-3166.

Göckede M., Rebmann C. and T. Foken, 2004: A combination of quality assessment tools for eddy covariance measurements with footprint modelling for the characterisation of complex sites, *Agr. For. Meteor. Vol. 127, pp. 175-188.*

Göckede M., Markkanen T., Hasager C.B. and Foken T., 2006: Update of footprint-based approach for the characterisation of complex measurement sites, *Boundary-Layer Meteorology 118(3), pp. 635-655.*

Göckede M., Foken T., Aubinet M., Aurela M., Banza J., Bernhofer C., Bonnefond J.M., Brunet Y., Carrara A., Clement R., Dellwik E., Elbers J., Eugster W., Fuhrer J., Granier A., GrünwaldT., Heinesch B., Janssens I.A., Knohl A., Koeble R., Laurila T., Longdoz B., Manca G., Marek M., Markkanen T., Mateus J., Mateucci G., Mauder M., Migliavacca M., Minerbi S., Moncrieff J., Montagnani L., Moors E., Ourcival J.-M., Papale D., Pereira J., Pilegaard K., Pita G., Rambal S., Rebmann C., Rodrigues A., Rotenberg E., Sanz M.J., Sedlak P., Seufert G., Siebicke L., Soussana J.F., Valentini R., Vesala T., Verbeeck H., and Yakir D., 2007: Quality control of CarboEurope flux data-Part I: Footprint analyses to evaluate sites in forest ecosystems, *Biogeosciences Discuss., Vol. 4, 4025-4066*.

GouldenM.L., Munger J.W., Fan S.M., Daube B.C., and Wofsy S.C., 1996a: Measurements of carbon sequestration by long-term eddy covariance: methods and a critical evaluation of accuracy, *Global Change Biology, Vol. 2, 159-168.*

Granier A., Biron P., and Lemoine D., 2000: *Water balance, transpiration and canopy conductance in two beech stands, Agr. For. Meteor.,Vol.100, pp. 291-308*.

Idso S., 1981: A set of equations for the full spectrum and 8- and 14- micron and 10.5- to 12.5 thermal radiation from cloudless skies, *Wat. Resour. Res. Vol. 17, pp. 295-304.*

Ijpelaar R.J.M., 2000: Evaluation of modified soil parameterization in the ECMWF land surface scheme, *KNMI Tech. Rep. 228, 70pp*.

Kabsch E., Olesen F.S. and Prata F., 2008, Initial results of the land surface temperature (LST) validation with the Evora, Portugal ground-truth station measurements, *Int. J. Remote Sensing., 29(17-18), 5329-5345*.

Kato H., M. Rodell, F. Beyrich, H. Cleugh, E. van Gorsel, H. Liu and T.P. Meyers, 2006: Sensitivity of Land Surface Simulations to Model Physics, Parameters, and Forcings, at Four CEOP Sites, *submitted to J. Meteor. Soc. Japan.*

Kottek M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel, 2006: World Map of the Köppen-Geiger climate classification updated, *Meteorol. Z.*, *Vol. 15, pp. 259-263.*

Kroon P.S., 2004: De Sluiting van de oppervlakte energiebalans in Cabauw gedurende TEBEX (1995-1996), *KNMI Technical report; TR-261, 63 pp.*

Luo L.F., Robock A., Mitchell K.E., Houser P.R., Wood E.F., Schaake J.C., Lohmann D., Cosgrove B., Wen F.H., Sheffield J., Duan Q.Y., Higgins R.W., Pinker R.T. and Tarpley J.D., 2003: Validation of the North American Land Data Assimilation System (NLDAS) retrospective forcing over the southern Great Plains, *J. Geophys. Res. Vol. 108(D22), Art. No. 8843.*

Masson, V., Champeaux, J. L., Chauvin, F., Meriguet, Ch. and Lacaze, R. A., 2003. Global Database of Land Surface Parameters at 1-km Resolution in Meteorological and Climate Models. *J. Climate* ***16****(9), 1261-1282*.

L. Merbold, J. Ardö, A. Arneth, R. J. Scholes, Y. Nouvellon, A. de Grandcourt, S. Archibald, J. M. Bonnefond, N. Boulain, C. Bruemmer, N. Brueggemann, B. Cappelaere, E. Ceschia, H. A. M. El-Khidir, B. A. El-Tahir, U. Falk, J. Lloyd, L. Kergoat, V. Le Dantec, E. Mougin, M. Muchinda, M. M. Mukelabai, D. Ramier, O. Roupsard, F. Timouk, E. M., Veenendaal, and W. L. Kutsch (2008), Precipitation as driver of carbon fluxes in 11 African ecosystems, BGD, 2008, Vol.5, pp. 4071-4105.

Mu Q., Heinsch F.A., Zhao M., and Running S.W., 2007: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of the Environment*, *Vol.111(4), pp. 519-536*.

Nash J.E., and Suttcliffe J.V., 1970: River Flow Forecasting through Conceptual Models, Part I-A Discussion of Principles, *J. Hydrol. Vol. 10, pp. 282-290*.

Pereira J.S., Mateus J.A., Aires L.M., Pita G., Pio C., David J.S., Andrade V., Banza J., David T.S., Paçao T.A., and Rodrigues A., 2007: Net ecosystem carbon exchange in three constrasting Mediterranean ecosystems-the effect of drought, *Biogeosciences, Vol.4, pp.791-802.*

Prigent C., Aires F., Rossow W.B. and Robock A. (2005), Sensitivity of satellite microwave and infrared observations to soil moisture at a global scale: Relationship of satellite observations to in situ soil moisture measurements, Journal of Geophysical Research, 10, D7.

Rebmann C., Gockede M., Foken T., Aubinet M., Aurela M., Berbigier P., Bernhofer C., Buchmann N., Carrara A., Cescatti A., Ceulemans R., Clement R., Elbers JE., Granier A., Grünwald T., Guyon D., Havrankova J., Heinesch B., Knohl A., Laurila T., Longdoz B., Marcolla B., Markkanen T., Miglietta F., Moncrieff J., Montagnani L., Moors E.J., Nardino M., Ourcival J.M., Rambal S., Rannik U., Rotenberg E., Sedlak P., Unterhuber G., Vesala T. and Yakir D., 2005: Quality analysis applied on eddy covariance measurements at complex forest sites using footprint modelling, *Theoretical and Applied Climatology Vol. 80, pp. 121-141.*

Reithmaier L. M., M. Göckede, T. Markkanen, A. Knohl, G. Churkina, C. Rebmann, N. Buchmann, and T. Foken, 2006: Use of remotely sensed land use classification for a better evaluation of micrometeorological flux measurement sites, *Theoretical and Applied Climatology Vol. 84, pp. 219-233.*

Rivalland, 2003: Amélioration et validation du modèle de fonctionnement de la vegetation ISBA-A_gs: stress hydrique et flux de CO2, Thèse présentée en vue de l'obtention du titre de docteur de l'Université Toulouse III.

Robock A., Luo L.F., Wood E.F., Wen F.H., Mitchell K.E., Houser P.R., Schaake J.C., Lohmann D., Cosgrove B., Sheffield J., Duan Q.Y., Higgins R.W., Pinker R.T., Tarpley J.D., Basara J.B. and Crawford K.C., 2003: Evaluation of the North American Land Data Assimilation System over the southern Great Plains during the warm season, *J. Geophys. Res. Vol. 108(D22), Art. No. 8846.*

Rodell M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, and D. Toll, 2004a: The Global Land Data Assimilation System, *Bull. Amer. Meteor. Soc. Vol. 85(3)*.

Rodell M., and P. R. Houser, 2004b: Updating a land surface model with MODIS derived snow cover, *J. Hydromet. Vol. 5(6), pp. 1064-1075.*

Rosema A., 1993: Using METEOSAT for Operational Evapotranspiration and Biomass Monitoring in the Sahel region, *Rem. Sens. of Env. Vol. 45, pp. 1-25.*

Sepulcre-Cantó, G., Gellens-Meulenberghs, F., Arboleda, A., Duveiller, G., Dewit, A., Eerens, H., Piccard, I., Bakary D.and Defourny, P.: Estimating crop specific evapotranspiration using remote sensing imagery at various spatial resolutions for improving crop growth modeling. Submitted to Int. J. Rem. Sens.

Seuffert G., H. Wilker, P. Viterbo, J.F. Mahfouf, M. Drusch and J.C. Calvet, 2003: Soil moisture analysis combining screen-level parameters and microwave brightness temperature: A test with field data, *Geophys. Res. Lett. Vol. 30(10).*

Seuffert G., H. Wilker, P. Viterbo, M. Drusch and J.F. Mahfouf, 2004: On the usage of screen level parameters and microwave brightness temperature for soil moisture analysis, *J. Hydromet. Vol.* **5**, *pp. 516-531.*

Shapiro R., 1987: A simple model for the calculation of the flux of direct and diffuse solar radiation through the atmosphere. AFGL-TR-87-0200, Air Force Geophysics Lab, Hanscom AFB, MA.

Sjöström, M., Ardö, J., Eklundh, L., El-Tahir, B. A., El-Khidir, H. A. M., Hellström, M., Pilesjö, P., and Seaquist, J., 2009: Evaluation of satellite based indices for gross primary production estimates in a sparse savanna in the Sudan, Biogeosciences, 6, 129-138.

Taylor, K.E.: Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res., 106, 7183-7192, 2001 (also see PCMDI Report 55, http://www-pcmdi.llnl.gov/publications/ab55.html).

Van den Hurk B.J.J.M., Viterbo, P., A.C.M. Beljaars and Betts, A.K., 2000: Offline validation of the ERA40 surface scheme, *ECMWF Technical Memorandum No. 295, 41 pp.*

Van den Hurk B.J.J.M. and Viterbo, P, 2003: The Torne-Kalix PILPS 2(e) experiment as a test bed for modifications to the ECMWF land surface scheme, *Global and Planetary Change Vol. 38, pp. 165-173.*

Van Ulden and Wieringa, 1996: Atmospheric boundary layer at Cabauw, *Boundary Layer Meteor. Vol. 78, pp. 39-69.*

Verstraten W.W., Veroustraete F., van der Sande C.J., Grootaers I. and Feyen J. (2006), Soil moisture retrieval using thermal inertia, determined with visible and thermal spaceborne data, validated for European forests, Remote Sensing of Environment, 101(3), pp.299-314.

Viterbo P. and Beljaars, A.C.M., 1995, An improved land surface parameterization scheme in the ECMWF model and its validation, *J. Climate Vol. 8, pp. 2716-2748.*

Voogt M.H., B.J.J.M. van den Hurk and C.M.J. Jacobs, 2006: The ECMWF land surface scheme extended with a photosynthesis and LAI module tested for a coniferous forest site, *KNMI Wetenschappelijk rapport, ISSN 0169-1651 ; WR 2006-02, 23 pp.*

Wesely M.L. and Hart R.L., 1985: Variability of short-term eddy correlation estimates of mass exchange*. In: The Forest-Atmosphere Interaction (Ed. By B.A.Hutchison and B.B. Hicks), pp. 591-612. D. Reidel, Dordrecht.*

Wilson K., Goldstein A., 2002: Energy balance closure at FLUXNET sites, *Agr. For. Meteor. Vol.113, pp. 223–243.*

List of acronyms

AFWA	Air Force Weather Agency
AWS	Automatic Weather Station
BALTEX	Baltic Sea Experiment
CAMELS	Carbon Assimilation and Modelling of the European Land Surfaces
CEOP	Coordinated Enhanced Observation Period
CLM	Community Land Model
CMAP	CPC Merged Analysis of Precipitation
DSLF	Downward Surface Long-wave Flux
DSSF	Downward Surface Short-wave Flux
EARS	Environmental Analysis & Remote Sensing
ECMWF	European Center for Medium Range Weather Forecast
ET	Evapotranspiration
GCM	Global Circulation Model
GLDAS	Global Land Data Assimilation System
GDAS	Global Data Assimilation System
GEWEX	Global Energy and Water Experiment
ISBA	Interaction Soil-Biosphere-Atmosphere
LAM	Limited Area Model
LSA-SAF	Land Surface Analysis-Satellite Application Facility
LST	Land Surface Temperature
MET	Meteosat Evapotarnspiration product
MEET	Meteosat EPS EvapoTranspiration product
MSG	Meteosat Second Generation
NOAA	National Oceanic & Atmospheric Administration
NWP	Numerical Weather Prediction
PRD	Product Requirement Document
PUM	Product User Manual
RMI	Royal Meteorological Institute
SPD	Science Plan Document
SVAT	Soil-Vegetation-Atmosphere Transfer
TESSEL	Tiled ECMWF Surface Scheme for Exchange Processes over Land

Annex A: Algorithm evolution history

Prototyping and first developments
-Selection of a method based on a Soil Vegetation Atmosphere Transfer (SVAT) scheme as the most suitable for use in conjunction with Satellite Remote Sensing (SRS): ECMWF TESSEL SVAT code as the basis for prototyping, FORTRAN 90 as the main programming language;
-Implementation at RMI of an offline 0-D version of the energy exchange module of the ECMWF TESSEL code and extension of the 0-D ET prototype to a regular grid;
-Improvement and further tests on the prototype: computation at the tile level versus pixel level;

Versioning and characteristics (initial operations)

Version 00
date of integration into the LSA-SAF system: MTR2
input: ECMWF archives on a regular latitude-longitude grid (5166 grid points);
method: energy balance module from TESSEL; uniform land cover;

Version 01
date of integration into the LSA-SAF system:
input: ECMWF archives
method: energy balance module from TESSEL; IGBP land cover;

Version 02
Date of integration into the LSA-SAF system: February 2005
Input:
LSA-SAF: DSSF, DSLF, ALBEDO, LST
Land Cover: IGBP
Atmospheric and soil forcing: ECMWF operational forecasts;
Method:
energy balance module from TESSEL, soil temperature from the superficial layer follows the temporal evolution of LST;
Improvement from the previous version:
-use of LSA-SAF radiation variables;
Weaknesses:
As LST is not produce for cloudy skies and is limited by the viewing angle of the satellite, ET computation cannot rely diagnostically on it as initially planned.

Version 03
Date of integration into the LSA-SAF system: November 2006
Input:
LSA-SAF: DSSF, DSLF, ALBEDO
Land Cover: ECOCLIMAP + vegetation parameters
Atmospheric and soil forcing: ECMWF operational forecasts;
Method:
energy balance module from TESSEL, solving the energy balance by assuming ground heat flux is proportional to net radiation;
Improvement from the previous version:
-erase dependence into LST and produce ET for all weather conditions;
-uses ECOCLIMAP land cover and monthly varying vegetation parameters;
Weaknesses:
Estimates seemed overestimated; systematic non-processed areas in northern Europe;

Version 04 (4.0)
Date of integration into the LSA-SAF system: June 2008
Input:
LSA-SAF: DSSF, DSLF, ALBEDO

Land Cover: ECOCLIMAP + vegetation parameters
Atmospheric and soil forcing: ECMWF operational forecasts;
Improvement from the previous version:
-correction of instabilities that led to systematic non-processed areas;
-calibration of the internal vegetation resistance to transpiration;
-extension of the methodology to the full MSG image;
Method: same method that in version 03.

Annex B: Flux measurement networks, methodology of measurement, stations and references for used stations

Stations from three measurement networks were used for the comparison with ground reference. Given that the methodology for measuring turbulent heat flux release into the atmosphere may differ from one network to another, we provide information on the characteristics of each network, and then summarize in the final list of the stations with their characteristics.

B-1. CarboEurope-IP network

CarboEurope-IP project (http://gaia.agraria.unitus.it/cpz/index3.asp), inheriting from the projects Euroflux (http://www.unitus.it/dipartimenti/disafri/progetti/eflux/euro.html), CarboEuroFlux, Medeflux, CAMELS (http://camels.metoffice.com/camels.html), coordinates the largest flux measurement stations for continuous monitoring in Europe, with about 50 main sites and 50 associated sites, covering a large panel of the ecosystems existing in Europe. This project is mainly directed towards the observation of carbon fluxes and experimental studies to better understand the mechanisms that play a role in plant respiration process. It is why flux stations are fully equipped for the measurements of all meteorological variables and are measuring the turbulent exchanges between soil-vegetation-atmosphere. CarboEurope-IP is the European part of the global network FLUXNET.

B-1.1. The stations

We selected a subset of stations among the hundred stations of the network. The choice of this subset is based on different criteria: 1) well established stations and already used datasets by the RMI team; 2) data quality and representativity (Rebmann et al, 2005;
Table B-2); 3) ability to sample different climates and biomes. Location of the stations is represented on the European map in Figure B-1. Geographical attributes of the selected stations used for comparison are listed in Table B-1. Different periods are covered depending on the availability of the in-situ data: February to November 2007 (names colored in black in Table), May-June 2006, June 2005 (green).

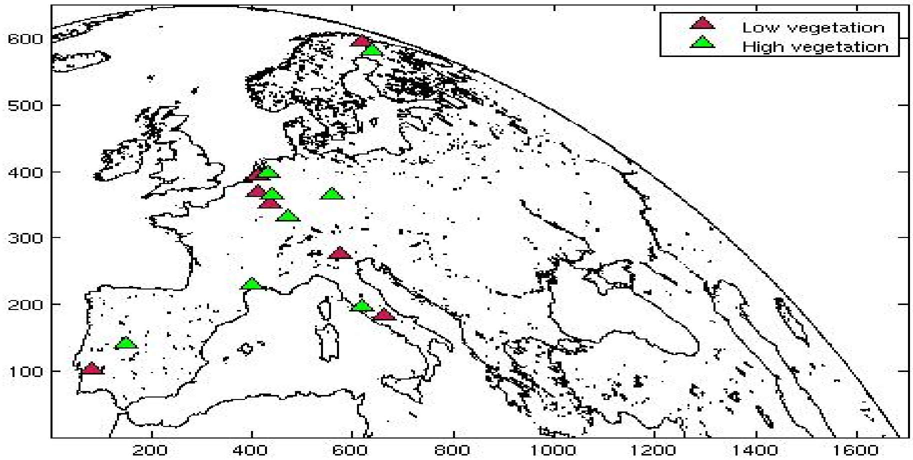


Figure B-1 Localization of the validation stations on the European map as viewed by MSG.

Station (country)+	Latitude	Longitude	Biome++	Climate+++	Altitude	H fetch
Amplero (IT)	41.90 N	13.61 E	G	Csa	884 m	-
Hesse (FR)	48.67 N	7.07 E	BF	Cfb	300 m	0.60
Las Majadas (ES)	39.94 N	-5.77 E	BF	Csa	265 m	1.50
Lonzée (B)	50.55 N	4.74 E	C	Cfb	165 m	0.24
Loobos (NL)	52.17 N	5.74 E	NF	Cfb	25 m	2.00
Kaamanen (FI)	69.14 N	27.29 E	B	Dfc	155 m	0.40
Monte Bondone (IT)	46.02 N	11.04 E	G	Cfa	1550 m	-
Puéchabon (FR)	43.74 N	3.60 E	BF	Csa	270 m	0.30
Roccarespampani (IT)	42.41 N	11.93 E	BF	Csa	234 m	0.60
Sodankylä (FI)	67.36 N	26.64 E	NF	Dfc	180 m	0.30
Tojal (PT)	38.48 N	-8.02 E	G	Csa	190 m	0.25
Vielsalm (B)	50.30 N	6.00 E	MF	Cfb	450 m	1.50
Wetzstein (DE)	50.45 N	11.46 E	NF	Cfb	785 m	1.00

Table B-1 Characteristics of the selected validation stations, including geographical location, the vegetation type intended to be studied, the climate of the surrounding region, altitude of the station, and length of the homogeneous fetch in the prevailing wind direction [km]. Results used for each station are colour coded (black: grid version of MET v4.0; green: 1D version of MET v4.0).

+Belgium(B), Germany (DE), Denmark (DK), Finland (FI), France (FR), Ireland (IE), Italy (IT), The Netherlands (NL), Portugal (PT), Spain (ES).

++Grassland (G), Broadleaved Forest (BF), Needle Forest (NF), Crops (C), Mixed Forest (MF).

+++Climate classification following Kottek et al. (2006): warm temperate (C), Snow (D), fully humid (f), summer dry (s), hot summer (a), warm summer (b).

B-1.2. Measurement methodology

The principles of the measurements methodology are described in details in Aubinet et al. (2000). The method relies on the measurement by a sonic anemometer of wind speed at a high rate, of the order of 20 Hz, installed above the vegetation canopy. These measurements provide quantitative information on turbulent fluctuations, i.e. vertical wind speed fluctuation, w', potential air temperature fluctuation, θ', and specific humidity fluctuation, q'. Mean cross-correlation between these fluctuations allows deriving the surface turbulent sensible (H) and latent (LE) fluxes:

$$H = \rho c_P \overline{w'\theta'} \tag{5}$$

$$LE = \rho L_v \overline{w'q'} \tag{6}$$

where is the air density, c_p the heat capacity constant and L_v the latent heat of vaporization.

B-1.3. Error and representativity

Eddy covariance is the most used method for deriving heat fluxes in the air surface layer. Numerous studies are showing its successes as well as its limits. While error in eddy covariance systems is hard to obtain due to the lack of an absolute reference available for comparison, different criteria can indicate the quality of the measurements: 1) the closure of energy balance; 2) the stationarity of the measurements of fluctuations; 3) the land use representative of the measurements must be known.

At first, a detailed study (Wilson et al., 2002) quantified for 50 contrasting ‘site-years’ of the FLUXNET network the magnitude of the closure of the energy imbalance. It was found that the imbalance ranged between 47% and 1%, with a mean at 17%. Wilson et al (2002) explored the possible causes of imbalance like erroneous measurement of energy storage in the vegetation, advection and biases in instrumentation. While no generalized conclusion has been drawn on the impact on latent heat flux error, it is generally admitted that error on estimates, excluding cases of very low turbulence, ranges between 5% and 20% (Aubinet et al., 2000; Wesely and Hart, 1985; Goulden et al., 1996a).

Concerning the quality of the measurements in term of stationarity, Foken and Wichura (1996) developed a methodology to rank into five quality classes taking in account the deviations in the stationarity and in the integral path of turbulence. Rebmann et al. (2005) show the statistics of the application of this quality test on several sites. In table 2, the percentage of latent heat flux estimates that are of the best quality (QC 1) are listed when available. These statistics have been drawn for periods of one to 4 months in 2000 and 2001, depending of the station.

Representativity of the measurements has a key role in model validation. Without further studies, as it is the case for stations for which the targeted vegetation is low, homogeneous fetch gives an indication of the representativity of the site. For forested stations, a recent study (Göckede et al., 2007) has analyzed quantitatively the measurements representativity of 25 sites by means of a lagrangian footprint model (Göckede et al., 2004; Rebmann et al., 2005; Reithmaier et al., 2006; Göckede et al., 2006).
Table B-2 summarizes the information for the forested sites considered in this report.

Station (country)[+]	% QC 1	Representativity (Göckede et al., 2007)
Hesse (FR)	64	60% to 90% of data of which 80% is emitted by TOI
Las Majadas (ES)	-	Homogeneous
Loobos (NL)	49	Homogeneous
Puéchabon (FR)	82	Homogeneous
Roccarespampini (IT)	-	60% to 90% of data of which 80% is emitted by TOI
Sodankylä (FI)	86	>90% of data of which 80% is emitted by TOI
Vielsalm (B)	54	>90% of data of which 80% is emitted by TOI
Wetzstein (DE)	-	Homogeneous

Table B-2 Summary of studies on quality of eddy covariance estimates at CarboEurope-IP forested sites. Second column lists the percentage of best quality class latent heat fluxes (fulfilment of stationarity hypothesis) as derived in Rebmann et al., 2005. The last column lists the results of the study of Göckede et al. (2007) on the representativity of the Type of Interest vegetation (TOI) in the measured heat fluxes.

As a conclusion, we can see that the choice of validation sites is constrained by several factors. A comprehensive literature is now being developed to guide the choice for validation purposes. While surface heat flux measurement has still a part of un-quantified errors, the stations chosen fulfill with a relative good adequacy the criteria detailed for validation purposes.

B-2. The CarboAfrica network

CarboAfrica project (http://www.carboafrica.net/index_en.asp) coordinates the one of the largest flux measurement stations for continuous monitoring in Africa, with about 25 sites, covering different climate regions in Africa. This project is intended to fill the gap of measurements in Africa. Most of the stations have been recently set up in 2006 and 2007. CarboAfrica is the African part of the global network FLUXNET. Stations operating in the western part of Africa are also used in the framework of the AMMA project (http://www.amma-international.org/).

B-2.1. The stations

For the in-situ validation, one validation dataset was available over Africa at the time of writing this report. Demokeya is situated in Sudan in the Sahel region (13.35°N, 30.47°E). Strong seasonality is observed in the precipitation field, with marked dry and wet seasons. The equipment is installed above a sahelian savannah, mainly composed of rain-fed grass and low percentage of acacia trees (Merbold et al., 2008; Sjöström et al, 2009).

B-2.2. Measurement methodology

The principles of the measurements methodology are the same than in CarboEurope-IP.

B-3.CEOP station: Cabauw

B-3.1. The station

Cabauw is a KNMI site of meteorological measurements (see van Ulden and Wieringa (1996) for a description of the site). Surface turbulent fluxes have been monitored for more than 20 years, through different field campaigns, and now on a routine basis. The system used before 2000 was based on the profile approach, mentioned in section 3.1.3, and widely used for comparison with land surface model (Beljaars and Bosveld, 1997). Different other methods have been tested at Cabauw: Bowen ratio method, eddy covariance method and more recently scintillometry. However, continuous monitoring was only made by means of the eddy covariance method through the last few years. The station is involved in CEOP through the BALTEX experiment.

B-3.2. Measurement methodology

As for CarboEurope-IP network, Cabauw station is equipped with a sonic anemometer/ thermometer (Kroon, 2004) to derive surface heat fluxes by eddy correlation method.

B-3.3. Error and representativiteness

Energy balance of the data obtained during the campaign TEBEX (1995-1996) has been carefully studied by the KNMI (Kroon, 2004). This study shows that energy balance is not closed when using the eddy covariance method. Still now, the reasons why there is non-closure are unknown (Bosveld, pers. comm.). Cabauw site is homogeneous: flat, surrounded by grassland (homogeneous fetch is considered to be 2km in the SE direction and 400m in the other directions).

B-4. AWS Belgium network

B-4.1. The stations

RMIB has developed within the last decade a network of automatic weather stations spread over Belgium and measuring, over short grass environment, standard meteorological variables at two or three levels. The measurements include all components of the radiative balance, wind speed, air temperature and relative humidity. The sites have been chosen to have a homogeneous fetch as large as possible to avoid the contamination of obstacles on measurements. Two of the nine AWS-Belgium stations have been selected for comparison: Buzenol and Humain.

B-4.2. Measurement methodology

The methodology used for deriving surface turbulent fluxes, described in Gellens-Meulenberghs (2005), follows a combination of a profile method and a residual method. Sensible heat flux is deduced from the profile of wind speed and air temperature (two levels at 2 m and 10 m respectively) with careful choice of stability functions and latent heat flux is deduced as the residual of the energy balance, for which each component is measured, except the soil heat flux considered as a fixed proportion of the net surface radiation.

B-4.3. Error and representativity

A careful study of the errors of algorithm has been processed (Gellens-Meulenberghs, 2005), taking into account uncertainty in the input variables and the impact of the hypotheses assumed. This study includes the range of accuracy depending on the quality of the other sensors. During daytime, net radiation is the main error source on LE.

B-5. References of the validation stations used for 'on-line' validation

Station	Dominant species	References	Contact
Amplero	Montaneous grass		N. Arriga, D. Pappale
Buzenol	Grass	Gellens-Meulenberghs, 2005.	F. Gellens_Meulenberghs
Cabauw	Grass	Beljaars and Bosveld, 1997; Van Ulden and Wieringa, 1996.	F. Bosveld
Hesse	Fagus sylvatica	Epron et al., 2001; Granier et al , 2000a; Granier A et al., 2000b; Lebaube et al., 2000.	A. Granier
Humain	Grass	Gellens-Meulenberghs, 2005.	F. Gellens_Meulenberghs
Kaamanen	Aapa mire	Laurila et al., 2001	T. Laurila
Las Majadas	Quercus ilex		A. Carrara
Lonzée	*Triticum æstivum* L.(Winter wheat, 2007)		C. Moureaux, M. Aubinet
Loobos	Pinus sylvestris	http://www.climatexchange.nl/sites/loobos/index.htm	J. Elbers
Puéchabon	Quercus ilex	Joffre R, Rambal S & F Romane, 1996.	S. Rambal
Roccarespampini	Quercus cerris		N. Arriga, D. Pappale
Sodankylä	Pinus sylvestris		T. Laurila
Tojal	C3 annuals ; C4 invasive Cynodon dactylon	J.S. Peireira et al., 2007 ; Aires et al., 2008.	C. Pio
Vielsalm	Fagus sylvatica, Pseudotsuga menziesii	Heinesch et al., 2007; Aubinet et al., 2001.	B. Heinesch, M. Aubinet
Wetzstein	Picea abies		C. Rebmann

Aires L.M, Pio C.A. and J.S. Pereira, 2008: The effect of drought on energy and water vapour exchange above a Mediterranean C3/C4 rassland in Southern Portugal, *Agr. For. Meteorol., Vol 148(4), pp. 565-579.*

Aubinet M., Chermanne B., Vandenhaute M., Longdoz B., Yernaux M. and E. Leitat, 2001: Long term carbon dioxide exchange above a mixed forest in the Belgian Ardennes, *Agr. For. Meteorol., Vol 108(4), pp. 293-315.*

Beljaars, A. C. M., and F. C. Bosveld, 1997: Cabauw data for the validation of land surface parameterization schemes, *J. Climate Vol. 10, pp. 1172-1193.*

Epron D, Farque L, Lucot E, Badot PM, 1999: Soil CO_2 efflux in a beech forest: dependence on soil temperature and soil moisture, *Ann For Sci Vol 56, pp. 221-26.*

Gellens-Meulenberghs F., 2005: Sensitivity Tests of an Energy Balance Model to Choice of Stability Functions and Measurement Accuracy, *Boundary Layer Meteorology, Vol 115 (3), pp. 453-471.*

Granier A., Biron P., and Lemoine D., 2000: Water balance, transpiration and canopy conductance in two beech stands*, Agr. For. Meteor., 100, pp. 291-308*.

Granier A., Ceschia E., Damesin C., Dufr E., Epron D., Gross P., Lebaube S., Le Dantec V., Le Goff N., Lemoine D., Lucot E., Ottorini J.M., Pontailler J.Y., Saugier B., 2000: The carbon balance of a young beech forest, *Functional ecology, Vol 14, pp. 312-325*.

Heinesch, B.; Yernaux, M. and Aubinet, M., 2007: Dependence of CO_2 advection patterns on wind direction on a gentle forested slope, *Biogeosciences Discussions, Vol 4(6), pp.4229-4260*.

Joffre R, Rambal S & F Romane, 1996: Local variations of ecosystem functions in a Mediterranean evergreen oak woodland, *Ann. For. Sci. 53, pp. 561-570*.

Laurila T., Soegaard H., Lloyd C.R., Aurela M., Tuovinen J.P. and C. Nordstroem, 2001: Seasonal variations of net CO2 exchanges in European Artic ecosystems, *Theor. Appl. Climatol., Vol 70(1-4), pp. 183-201*.

Lebaube S., Le Goff N., Ottorini J.-M., Granier A., 2000: Carbon balance and tree growth in a *Fagus sylvatica* stand., *Ann. For. Sci., Vol 57, pp. 49-61*.

Pereira J.S., Mateus J.A., Aires L.M., Pita G., Pio C., David J.S., Andrade V., Banza J., David T.S., Paçao T.A., and Rodrigues A., 2007: Net ecosystem carbon exchange in three constrasting Mediterranean ecosystems-the effect of drought, *Biogeosciences, 4, pp.791-802*.

Van Ulden and Wieringa, 1996: Atmospheric boundary layer at Cabauw, *Boundary Layer Meteor. Vol. 78, pp. 39-69.*

Annex C: Off-line Validation of the 0-D LSA-SAF MET v4.0 algorithm

C-1. Purpose

The purpose of this annex is to provide to the users an insight into preliminary validation of the ET-RMI1 algorithm in off-line idealized conditions. The results can therefore be considered as the best performances of the algorithm at a scale comparable with measurement point-wise stations.

C-2. List of datasets used for validation of the 0-D LSA-SAF MET v4.0 algorithm

Several datasets have been used for off-line validation of the 0-D version of the algorithm. These have been chosen to represent a panel of climate and vegetation as large as possible taken into consideration the different measurements networks and campaigns available.

Station name (Country)	Network/Facility	Vegetation
Bondville (U.S.)	Ameriflux/ARM/CEOP	C3/C4 crops rotation
Cabauw (NL)	KNMI/CEOP	Grass
Fort-Peck (U.S.)	Ameriflux/ARM/CEOP	Grass
Loobos (NL)	CarboEurope-IP	Coniferous Forest
Manaus (BR)	LBA/CEOP	Tropical Forest
Puéchabon (FR)	CarboEurope-IP	Medite. Broadl. Forest
Santarem (BR)	LBA/CEOP	Tropical Forest
Hesse (FR)	CarboEurope-IP	Deciduous broadl. Forest
Hainich (GE)	CarboEurope-IP	Deciduous broadl. Forest
Le Bray (FR)	CarboEurope-IP	Coniferous Forest

C-3. Description of the stations of validation and the datasets

Station name	Latitude-longitude	Dominant species	H fetch [m]	Meas. Height [m]	LAI [m²/m²]	Stem density [trees/ha]
Bondville	40.0061°N, 88.29186°W	Soybean/maize rotation	[300-700]	10	[0 -5]	-
Cabauw	51.97°N, 4.93°E	Grassland			NA	-
Fort-Peck I.R.	48.30768°N, 105.10185°W	Grassland	200	3.5	~2	-
Loobos	52.1679°N, 5.7440°E	Pinus Sylvestris; under:Deschampsia Flexuosa L.	2000	27	[1.7-2.2] ; [0-1.1]	446
Manaus			NA		NA	NA
Puéchabon	43.7414°N, 3.5958°E	Quercus ilex	300	12.2	2.9	8500
Santarem Km87	2.857°S; 54.959°W		NA		NA	NA
Hesse	48°40’ N; 07°04’E	Fagus sylvatica	600	23	[0-6]	3480
Hainich	51°04’N; 10°27’ E	Fagus Sylvatica L. Acer Pseudoplantanus, Frximus excelsior Under: herbs	1000	43.5	[0-5]	334
Le Bray	44°43’N; 00°46’W	Pinus pinaster; under: molignie	693	41	[1.8-4.2]	NA

C-3.2.Forcing:

- Atmospheric and soil forcings:
  - Meteorological and radiation terms measured at the station
  - Soil moisture from the ECMWF operational archives
- Gap filling
  - For some stations the surface longwave incoming radiation is sparsely or not measured, a procedure is applied to fill it from the other radiation terms or from the air temperature (Stöckli et al., 2008).
  - For gaps of few time steps, a linear interpolation is applied. For longer ones, we used a moving average window over 10 daily cycles.
- Vegetation parameters (type, leaf area index, fraction of vegetation cover, roughness length)

Station	Ecoclimap class	Period (begin-end)	Sources vegetation description
Bondville	118-Great Plain Crops	01/01/2002-31/1/2003	T. Meyer
Cabauw	Temperate Grassland	01/01/1995-31/12/1996	-
Fort-Peck I.R.	103-Rockies Grassland	01/01/2001-31/07/2003	T. Meyer
Loobos	211-Temperate Coniferous Forest	01/01/2003-31/12/2003	Rivalland, 2003
Manaus	21-Amazonian Forest		-
Puéchabon	201-Mediteranean Broadleaved Forest	01/01/2002-//2003	http://www.agraria.unitus.it; Joffre et al., 1996
Santarem Km87	21-Amazonian Forest	01/10/2002-30/09/2003	-
Hesse	203-Temperate Broadleaved Forest	01/01/1997-31/12/1998	Rivalland, 2003; Granier et al., 2000b
Hainich	203-Temperate	01/01/2003-31/12/2003	Knohl et al., 2003; Wang et

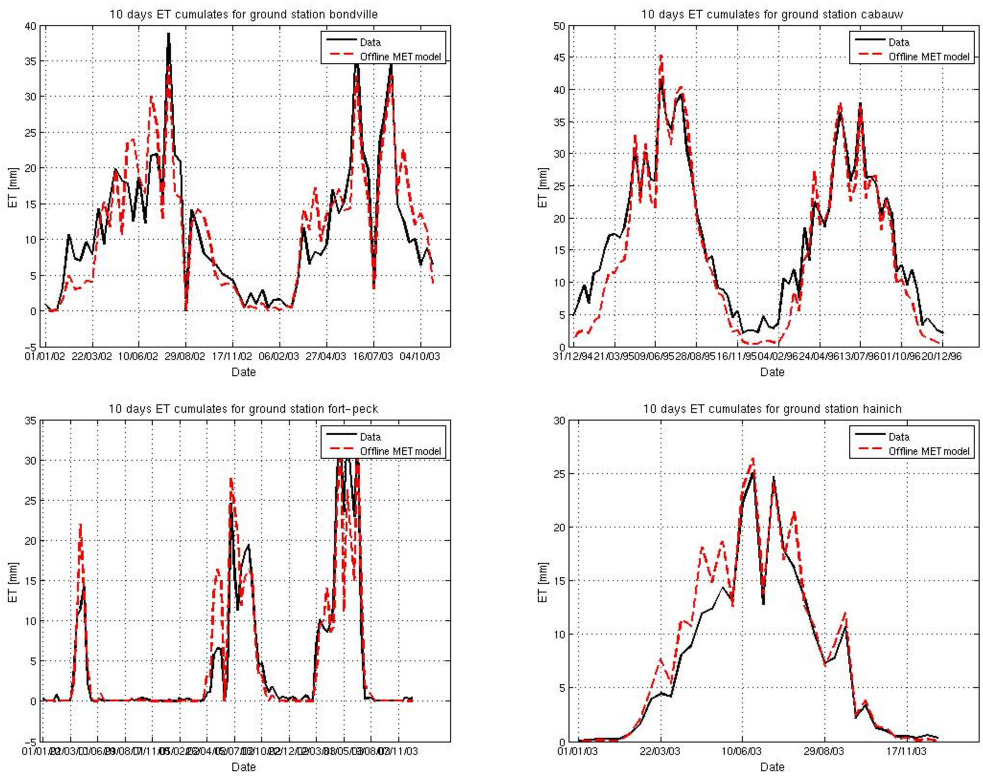
	Broadleaved Forest		al., 2005
Le Bray	209-Landes Forest	01/01/1997-31/07/1998	Rivalland, 2005; Porté, 1999

C-4. Validation results: time series and statistics

For each validation site, we compare simulation of 0-D version of the MET v4.0 algorithm with the in-situ measurements, at a time step of 30 minutes (Table C-1), and also on the basis of 10-days cumulates (Figure C-1). In 10-days cumulates, estimates from both measurement and model are set to zero for time step with non-existing or dubious measurements.

Station	RMSE (mm/h)	Bias (mm/h)	correlation	Nash index
Bondville	0.10	+0.015	0.71	0.38
Cabauw	0.06	-0.014	0.91	0.78
Fort-Peck	0.11	+0.025	0.57	Out of range
Hainich	0.05	+0.019	0.81	0.60
Hesse	0.08	+0.018	0.80	0.60
Le Bray	0.08	-0.024	0.81	0.63
Loobos	0.10	+0.024	0.50	Out of range
Manaus	0.12	+0.063	0.86	0.69
Puéchabon	0.07	+0.012	0.68	0.09
Santarem	0.14	+0.013	0.76	0.56

Table C-1 Statistical indices of the comparison of MET offline model and measurement of ET on half-hourly estimates. Out of range means that the Nash index computed is negative.



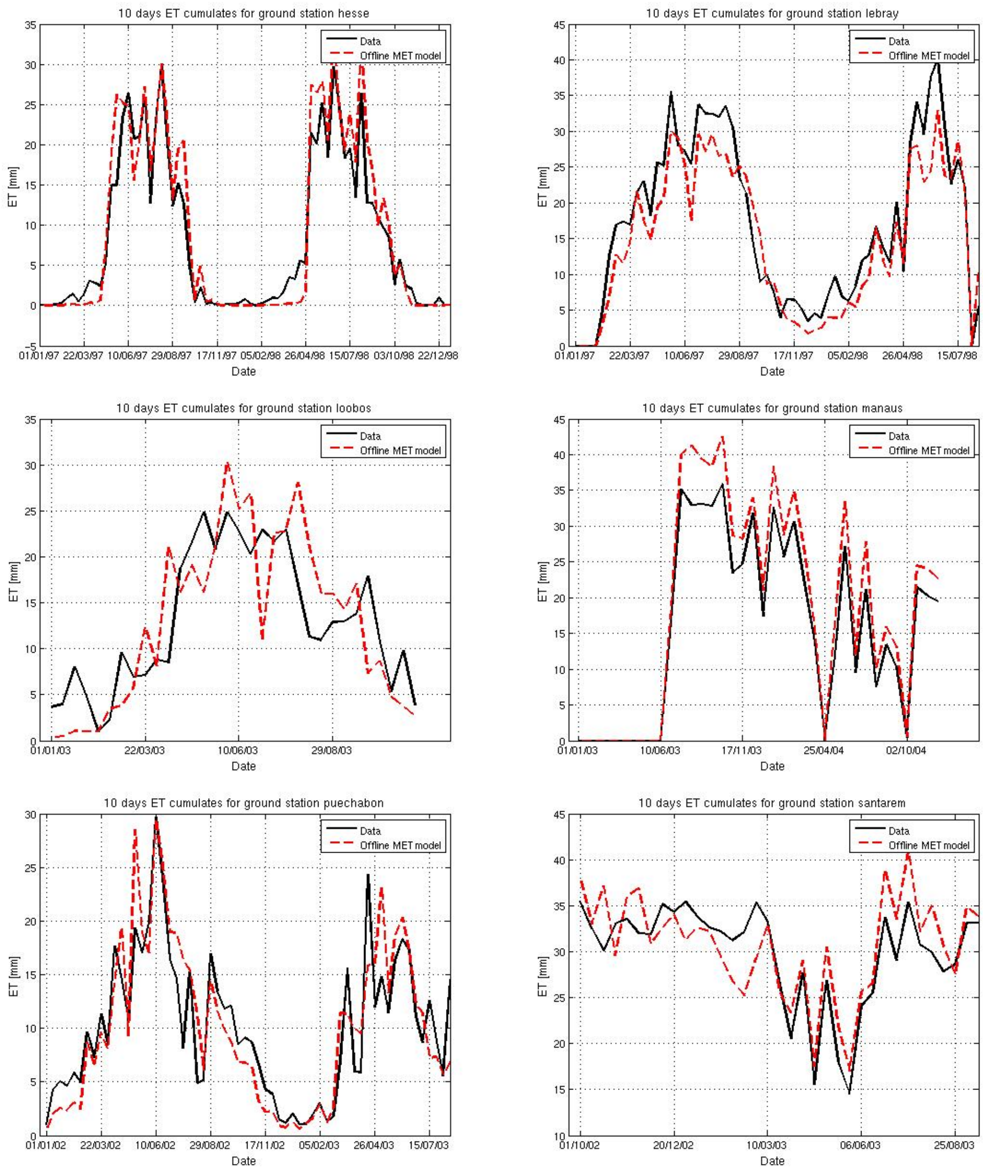


Figure C-1: Comparison of the evolution of 10-days ET cumulates (mm) between MET offline model (red) and in-situ measurements (black). From top-left to bottom-right: Bondville, Cabauw, Fort-Peck I.R., Hainich, Hesse, Le Bray, Loobos, Manaus, Puéchabon, Santarem.

From the results presented, we can see that in overall MET model applied at single point with measured forcing is able to capture quite accurately the evolution of evapotranspiration. For 5 stations on 10, the Nash index is above 0.60, indicating a good similarity between model and measurements. For three others, the results are of reasonable quality, with at least one good statistical indicator. At last, validation at two stations seems problematic, when looking at the Nash index: Loobos and Fort-Peck.

C-5. References

Joffre R, Rambal S & F Romane, 1996: Local variations of ecosystem functions in a Mediterranean evergreen oak woodland, *Ann. For. Sci., Vol 53, pp. 561-570*.

Knohl A., Schulze A.-D., Kolle O. and N. Buchmann, 2003: Large carbon uptake by an unmanaged 250-year-old deciduous forest in Central Germany, *Agr. For. Meteor., Vol. 118, pp. 151-167*.

Porté A., Bosc A., Champion I. and D. Loustau, 2000 : Estimating the foliage area of Maritime pine (Pinus Pinaster Aït) branches and crown with application to modeling the foliage area distribution in the crown, *Ann. For. Sci., Vol. 57, pp. 73-86*.

Rivalland V., 2003: Amélioration et validation du modèle de fonctionnement de la vegetation ISBA-A-gs : stress hydrique et flux de CO2, *thèse soutenue le 27/11/2003 en vue de l'obtention du titre de docteur de l'Université Toulouse III (Paul Sabatier), 232 pp*.

Rivalland V., Calvet J.-Ch., Berbigier P., Brunet Y. and A. Granier, 2005: Transpiration and CO2 fluxes of a pine forest: modeling the undergrowth effect, *Ann. Geophys., Vol. 23(2), pp.291-304*.

Stöckli R., Lawrence D.M., Niu G.-Y., Oleson K.W., Thornton P.E., Yang Z.-L., Bonan G.B., Denning A.S. and Running S.W., 2008: use of FLUXNET in the Community Land Model development, *J. Geophys. Res., Vol.113, 19pp*.

Wang Q., Tenhunen J., Dinh N.Q., Reichstein M., Otienao D., Granier A. and K. Pilegaard, 2005: Evaluation of seasonal variation of MODIS derived leaf area index at two European deciduous broadleaf forest sites, *Rem. Sens. Environ., Vol. 96, pp.475-484*.

Annex D: ECMWF GCM and GLDAS land assimilation systems

D.1. ECMWF

D-1.1. ECMWF atmospheric model

ECMWF (http://www.ecmwf.int/) global circulation model (GCM) is build to run for weather predictions up to 10 days, with a horizontal resolution of 0.25°, since February 2006, and up to 91 vertical pressure levels in the atmosphere. It is run twice a day. Although many studies concern the quality of the medium range prediction, we only focus here on the short-range prediction up to 12h or 24h at maximum. The ECMWF atmospheric model is coupled operationally to a land surface scheme, TESSEL ("Tiled ECMWF Surface Scheme for Exchange Processes over Land").

D-1.2. TESSEL land surface scheme

TESSEL is a SVAT model mainly developed for use into the ECMWF atmospheric model. Successive versions (Blondin, 1991; Beljaars and Viterbo, 1994; Viterbo and Beljaars, 1995; van den Hurk et al., 2000), with improved physical modelling and accuracy lead to the currently used operational version (Cyc31r1) and used for the ERA40 reanalysis. The concept of 'tiling' is used for the computation of release of water and heat fluxes into the atmosphere, vertical movement of water into soil and diffusion of heat into soil are prognosed using differential equations solved over 4 soil layers. Rainfall interception by vegetation canopy is modelled as well as effects of snow and freezing of the soil in the heat and water soil budget.

D-1.3. Overall accuracy and domain of application

The version of TESSEL used for comparison has been validated off-line, before processing the ERA-40 reanalysis. This validation was performed over boreal, tropical and temperate forest, as well as savannah, prairie, crop and grassland, using datasets from long-term surface field campaigns (van den Hurk et al., 2000). The scheme is evaluated further through PILPS experiments for model comparison (van den Hurk et al, 2003).

D-1.4. Impact on land surface community

ECMWF is a well-recognized centre for numerical weather prediction, active in both operational and research activities. The ERA-40 reanalyses are commonly adopted as a reference and largely exploited by many international researches. Moreover, ECMWF is a test-bed for operational research about land surface scheme initialization (Drusch, 2007), new model parameterization (Ijpelaar, 2000; van den Hurk et al, 2003; Voogt et al, 2006), and assimilation schemes (e.g. Seuffert et al., 2004; Drusch and Viterbo, 2007).

D-2. GLDAS

GLDAS ingests satellite- and ground-based observational data products, using land surface modelling and data assimilation techniques, in order to generate optimal fields of land surface states and fluxes (Rodell et al, 2004a). The system drives multiple offline (not coupled to the atmosphere) land surface models, globally at high resolutions (2.5° to 1 km), and produces results in near-real time (http://ldas.gsfc.nasa.gov/).

D-2.1. Models

Noah, CLM and Mosaic

D-2.2. Forcings

Observation-based precipitation (CMAP; Xie and Arkin, 1997) and downward radiation products from AFWA (Idso, 1981; Shapiro, 1987) and the best available analyses from atmospheric data assimilation systems (NOAA's GDAS).

D-2.3. Assimilation

Assimilation of snow cover (Rodell et al, 2004b).

D-2.4. Overall accuracy

Despite the fact that no comprehensive study of the uncertainty in GLDAS inputs or outputs exists, some regional or point assessment have been performed (Luo et al, 2003; Robock et al, 2003; Kato et al, 2006), but especially over North America. It has been also pointed out by the GLDAS team, that results are not reliable over Greenland and Polar regions. (http://www.gewex.org/GLDAS_data_product_4-2006.pdf)

D-2.5. Impact on land surface community

Archive of modelled and observed, global, surface meteorological data, parameter maps, and output which includes 1° and 0.25° resolution 1979-present simulations of the Noah, CLM, and Mosaic land surface models. GLDAS is also involved in GEWEX and CEOP.

Annex E: Definition of High and Low vegetation percentages from the land cover used by LSA SAF MET, ECMWF and GLDAS

	LSA SAF MET	ECMWF	GLDAS
Low vegetation tile	1.Bare Soil 2.Rocks 3.Permanent Snow 7.C3 crops 8.C4 crops 9.Irrigated Crops 10.Grass 11.Bogs/Swamp vegetation/gardens	1.Cropland 2. Short Grass 7.Tall Grass 8.Irrigated Crops 9.Tundra 11.Semi-desert 13.Bogs and Marshes 16.Evergreen Shrubs 17.Deciduous Shrubs	6.Woodland (14.9%) 7.Wooded Grassland (58.5%) 8.Closed Shrubland 9.Open Shrubland 10.Grassland 11.Cropland 12.Bare Ground 13.Urban and built-up
High vegetation tile	4.Deciduous Broadleaved forest 5.Evergreen Broadleaf Forest 6.Evergreen Needleleaf Forest	3.Evergreen Needleleaf Forest 4.Deciduous Needleleaf Forest 5.Deciduous Broadleaved forest 6.Evergreen Broadleaf Forest 18.Mixed Forest 19.Interrupted Forest	1.Evergreen Needleleaf Forest 2.Evergreen Broadleaf Forest 3.Deciduous Needleleaf Forest 4.Deciduous Broadleaved forest 6.Woodland (85.1%) 5.Mixed Cover 7.Wooded Grassland (41.5%)

Table E-1 Classification in High and Low vegetation chosen for inter-comparison of the land cover maps of LSA SAF MET (ECOCLIMAP/MSG), ECMWF and GLDAS (UMD).