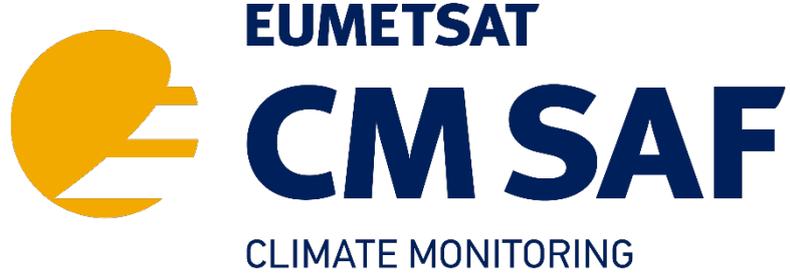


EUMETSAT Satellite Application Facility on Climate Monitoring



**Validation Report
CM SAF Latent and Sensible Heat Flux
Climate Data Record**

**DOI: 10.5676/EUM_SAF_CM/SLF_METEOSAT/V001
Meteosat Latent and Sensible Heat Flux**

CM-23811

Reference Number:
Issue/Revision Index:
Date:

SAF/CM/RMIB/VAL/LEH
1.1
30.05.2023

Document Signature Table

	Name	Function	Signature	Date
Authors	William Moutier	CM SAF scientist		30.05.2023
	Nicolas Clerbaux	CM SAF scientist		
	Françoise Gellens-Meulenberghs	CM SAF scientist		
	Alirio Arboleda	CM SAF scientist		
	Miguel Barrios	CM SAF scientist		
Editor	Marc Schröder	CM SAF Science Coordinator		30.05.2023
Approval	CM SAF Steering Group			
Release	Rainer Hollmann	CM SAF Project Manager		xx.xx.2023

Distribution List

Internal Distribution	
Name	No. Copies
DWD / Archive	1
CM SAF Team	1

External Distribution		
Company	Name	No. Copies
Public		1

Document Change Record

Issue/ Revision	Date	DCN No.	Changed Pages/Paragraphs
1.0	09.05.2023		Initial version for joint DRR3.9
1.1	30.05.2023		Included suggestions from reviewers during DRR3.9

	Validation Report Meteosat Latent and Sensible heat fluxes - Edition 1	Doc.No: SAF/CM/RMIB/VAL/MET/LEH Issue: 1.1 Date: 30.05.2023
---	---	---

Applicable Documents

The following documents, of the exact issue shown, form part of this document to the extent specified herein. Applicable documents are those referenced in the Contract or approved by the Approval Authority. They are referenced in this document in the form [AD X].

Reference	Title	Code, Version, Date
AD 1	CM SAF Product Requirements Document	SAF/CM/DWD/PRD/3.9

Reference Documents

The reference documents contain useful information related to the subject of the project. These reference documents complement the applicable ones, and can be looked up to enhance the information included in this document if it is desired. They are referenced in this document in the form [RD X].

Reference	Title	Code, Version, Date
RD 1	CM SAF Algorithm Theoretical Basis Document. Meteosat Latent and Sensible heat fluxes - Edition 1	SAF/CM/RMIB/ATBD/MET/LEH/1.1
RD 2	CM SAF Validation Report Meteosat Radiative Balance -Edition 1	SAF/CM/MET/VAL/SRB/1.0

Table of Contents

Executive Summary.....	10
1 The EUMETSAT SAF on Climate Monitoring	12
2 Introduction	14
3 Validation strategy.....	16
3.1 <i>In situ</i> datasets (FLUXNET2015 / ICOS).....	19
3.2 ERA5 dataset	20
3.3 GLDAS dataset.....	20
3.4 LSA SAF v3 dataset	21
3.5 GLEAM dataset	21
3.6 Comparison with FLUXNET2015 / ICOS.....	22
3.6.1 Surface latent heat flux	22
3.6.2 Surface sensible heat flux.....	26
3.6.3 Monthly mean diurnal cycle.....	30
3.7 Product inter-comparisons	32
3.7.1 Latent heat flux	32
3.7.2 Evapotranspiration.....	35
3.7.3 Sensible heat flux	38
3.8 Stability.....	41
3.8.1 Latent heat flux	41
3.8.2 Sensible heat flux	42
3.9 Comparison of Meteosat First and Second Generation.....	46
3.9.1 Surface latent heat flux	46
3.9.2 Surface sensible heat flux.....	50
4 Concluding remarks	53
5 References.....	54
6 Appendix.....	60
6.1 Statistical metrics.....	60
6.2 FLUXNET2015/ICOS sites.....	61
6.2.1 In situ eddy-covariance sites.....	61
6.2.2 Potential error at Skukuza station (South Africa).....	62
6.3 Land cover.....	63

List of Tables

Table 0-1: Summary of requirement compliance for the latent heat flux.	10
Table 0-2: Summary of requirement compliance for the sensible heat flux products.....	11
Table 2-1: Main features of the CM-23811 data records.	15
Table 2-2: Main inputs variables used in the CM SAF algorithm.....	15
Table 3-1: Main characteristics of datasets used for the validation.....	17
Table 3-2: Product requirements for evapotranspiration and surface heat fluxes (CM-23811). LE_{obs} , H_{obs} and ET_{obs} refer to the mean absolute value of the reference dataset over the time period analyzed.	18
Table 3-3: Input sources used in the LSA SAF v3 and the CM SAF approach.....	21
Table 3-4: Overview of forcing dataset use in GLEAM V3.5.....	22
Table 4-1: Performance statistics of hourly, daily and monthly CM SAF latent heat flux dataset as compared to FLUXNET2015/ICOS observations at 30 stations for different seasons and day/night conditions.	25
Table 4-2: Performance statistics of hourly, daily and monthly CM SAF sensible heat flux dataset as compared to FLUXNET2015/ICOS observations at 30 stations for different seasons and day/night conditions.	29
Table 4-3: Performance statistics of daily CM SAF latent heat flux dataset as compared to ERA5, LSA SAF and GLDAS datasets averaged over the disk and the considered period.	33
Table 4-4: Performance statistics of daily CM SAF Evapotranspiration dataset as compared to GLEAM, LSA SAF and GLDAS datasets averaged over the disk and the covered period.....	36
Table 4-5: Performance statistics of daily CM SAF sensible heat flux dataset as compared to ERA5, LSA SAF and GLDAS datasets averaged over the disk and the covered period.....	39
Table 4-6: Deseasonalized bias ($W m^{-2}$) of the CM SAF monthly latent heat flux as compared with ERA5, LSA SAF, GLDAS and GLEAM averaged over the full disk and over the time period covered by each satellite.....	42
Table 4-7: Same as Table 4-6 but for sensible heat flux ($W m^{-2}$).	44
Table 7-1: Information about <i>in situ</i> selected stations used to validate the CM SAF product.....	61

List of Figures

Figure 1: Overview of Meteosat satellites used as input for the generation of the land surface flux CDR.....	14
Figure 2: Location of eddy covariance stations used to evaluate the CM SAF dataset. Black frames delineate: Europe (Euro; lat:[34.49,60], lon:[-46.1, 60]), North-Africa (NAfr; lat:[0.2, 39.4], lon:[-21.7, 60]), South-Africa (SAfr; lat:[-40.5, 0.2], lon:[7.7, 60]) and South-America (SAm; lat:[-37.6,12.6], lon:[-60, -32.8]) areas (LSA SAF regions).	20
Figure 3: Bar plot showing the hourly bias ($W m^{-2}$; top) and the hourly unbiased root mean square error ($W m^{-2}$; bottom) of CM SAF (blue), ERA5 (green) and LSA SAF (violet) datasets as compared to FLUXNET2015/ICOS dataset at stations for the latent heat flux. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements.	24
Figure 4: Bar plot showing the daily bias ($W m^{-2}$; top) and the daily unbiased root mean square error ($W m^{-2}$; bottom) of CM SAF (blue), ERA5 (green), GLDAS (orange) and LSA SAF (violet) datasets as compared to FLUXNET2015/ICOS dataset at stations for the latent heat flux. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements.	24
Figure 5: Same as Figure 4 but for monthly latent heat flux.	25
Figure 6: Same as Figure 3 but for sensible heat flux.	27
Figure 7: Same as Figure 4 but for sensible heat flux.	28
Figure 8: Same as Figure 5 but for sensible heat flux.	28
Figure 9: Box plots displaying the monthly mean diurnal cycle bias ($W m^{-2}$; top panel) and uRMSD ($W m^{-2}$; bottom panel) for the latent heat flux at FLUXNET2015/ICOS stations. The height of the box indicates the Interquartile range (IQR), the horizontal line inside the box indicates the median value, the upper box level indicates the upper quartile (75th percentile; Q_3), the lower box level indicates the lower quartile (25th percentile; Q_1), bars (whiskers) indicate the minimum and maximum values, dots indicate the outliers (higher	

than $Q_3 + 1.5 \cdot IQR$ or lower than $Q_1 - 1.5 \cdot IQR$. The shaded areas represent the optimal (green), target (blue) and threshold (red) requirements. 31

Figure 10: Same as Figure 8 but for the sensible heat flux. 31

Figure 11: Time series of latent heat flux disk averaged daily bias ($W m^{-2}$) of daily image of CM SAF dataset as compared to daily images of ERA5 (green, top), LSA SAF (violet, middle) and GLDAS (orange, bottom) datasets. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements. 33

Figure 12: Time series of latent heat flux disk averaged daily unbiased root mean square error ($W m^{-2}$) of daily CM SAF dataset as compared to daily images of ERA5 (green, top), LSA SAF (violet, middle) and GLDAS (orange, bottom) datasets. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements. 34

Figure 13: Maps of averaged bias (top) and unbiased root mean square difference (bottom) of the surface latent heat flux ($W m^{-2}$) over the year 2005 of CM SAF dataset as compared to ERA5 (a-d) LSA SAF (b-e) and GLDAS (c-f) datasets at 0.25° . Average value over the map is indicated in blue over each map. 35

Figure 14: Time series of disk averaged daily evapotranspiration bias ($mm day^{-1}$) of daily images of daily CM SAF dataset as compared to GLEAM (green, top), LSA SAF (violet, middle), and GLDAS (orange, bottom) datasets. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements. 36

Figure 15: Time series of disk averaged daily evapotranspiration unbiased root mean square error ($mm day^{-1}$) of daily CM SAF dataset as compared to daily images of GLEAM (green, top), LSA SAF (violet, middle), the GLDAS (orange, bottom) datasets. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements. 37

Figure 16: Maps of averaged daily bias (top) and unbiased root mean square difference (bottom) of the evapotranspiration ($mm day^{-1}$) over the year 2005 of CM SAF dataset as compared to GLEAM (a-d), LSA SAF (b-e) and GLDAS (c-f) datasets at 0.25° . Average value over the map is indicated in blue over each map. 38

Figure 17: Same as Figure 11 but for sensible heat flux. 39

Figure 18: Same as Figure 12 but for sensible heat flux. 40

Figure 19: Same as Figure 13 but for sensible heat flux. 40

Figure 20: Time series of deseasonalized bias ($W m^{-2}$) of monthly global latent heat flux compared to ERA5 (green), LSA SAF (violet), GLDAS (orange) and GLEAM (blue). Right axis indicates which Meteosat satellite is used (dark-red line). Trends ($W m^{-2} dec^{-1}$) are indicated in the legend and an (*) is added if the slope to calculate the trend is not significant ($p\text{-value} > 0.05$). 41

Figure 21: Same as Figure 20 but for sensible heat flux. 43

Figure 22: Time series of deseasonalized bias ($W m^{-2}$) of the monthly sensible heat flux compared to ERA5 averaged over North Africa (yellow line) and South Africa (brown line) regions. Right axis indicates which Meteosat satellite is used (dark-red line). Trends values ($W m^{-2} dec^{-1}$) are indicated in the legend. 44

Figure 23: Time series of deseasonalized bias ($W m^{-2}$) of the monthly sensible heat flux compared to ERA5 averaged over Europe (blue) and South America (green) regions. Right axis indicates which Meteosat satellite is used (dark-red). Trends values ($W m^{-2} dec^{-1}$) are indicated in the legend. 45

Figure 24: Time series of deseasonalized bias ($W m^{-2}$) of the CM SAF monthly albedo as compared to ERA5. Right axis indicates which Meteosat satellite is used (dark-red line). The trend values ($W m^{-2} dec^{-1}$) of the time series is indicated in the legend. 45

Figure 25: Same as Figure 24 but for the surface incoming shortwave radiation. 46

Figure 26: Bar plot showing the hourly bias ($W m^{-2}$; top) and the hourly unbiased root mean square error ($W m^{-2}$; bottom) of CM SAF-MSG (blue) and CM SAF-MSG (light-blue) datasets as compared to FLUXNET2015/ICOS dataset at stations for the latent heat flux for the year 2005. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements. 47

Figure 27: Bar plot showing the daily bias ($W m^{-2}$; top) and the daily unbiased root mean square error ($W m^{-2}$; bottom) of CM SAF-MSG (blue) and CM SAF-MSG (light-blue) datasets as compared to FLUXNET2015/ICOS dataset at stations for the latent heat flux for the year 2005. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements. 47

Figure 28: Bar plot showing the monthly bias ($W m^{-2}$; top) and the monthly unbiased root mean square error ($W m^{-2}$; bottom) of CM SAF-MSG (blue) and CM SAF-MSG (light-blue) datasets as compared to FLUXNET2015/ICOS dataset at stations for the latent heat flux for the year 2005. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements. 48

Figure 29: Maps (0.05°) of daily latent heat flux ($W m^{-2}$) averaged over 2005 from MFG (a) and MSG (b); (c) difference between MFG and MSG based products; (d) unbiased root mean square difference between MFG and MSG based products ($W m^{-2}$); (e) bias ($W m^{-2}$) between MFG and MSG based products in winter (e) and summer (f) season. Average value over the map is indicated in blue over each map. 49

Figure 30: Same as Figure 26 but for sensible heat flux. 50

Figure 31: Same as Figure 27 but for sensible heat flux. 51

Figure 32: Same as Figure 28 but for sensible heat flux. 51

Figure 33: Same as Figure 29 but for sensible heat flux. 52

Figure 34: Scatter plot of CM SAF latent heat flux bias ($W m^{-2}$) as compared to *in-situ* data for the year 2005 at “Skukusa” station vs. the difference between gap-filled (MDS method) energy balance corrected latent heat flux (LE_{CORR} ; $W m^{-2}$) and the uncorrected gap-filled (MDS method) latent heat flux (LE_{F_MDS} ; $W m^{-2}$). Solid red line is linear fit. 62

Figure 35: Temporal course of gap-filled (MDS method) energy balance corrected latent heat flux (LE_{CORR} ; $W m^{-2}$) from FLUXNET2015 dataset at Skukuza station. 63

Figure 36: Land cover map used as the main tile input for the year 2005 for MFG (left) and MSG (right). 63

Figure 37: Difference between main land cover tile used as input for the year 2005 from MFG and MSG. 64

Figure 38: Land cover map used as the main tile input in the LSA SAF dataset for 2005-03-15. 64

List of acronyms

CF	Climate and Forecast
CDOP	Continuous Development and Operations
CDR	Climate Data Record
CM SAF	Satellite Application Facility on Climate Monitoring
DJF	December January February
DWD	Deutscher Wetterdienst (German MetService)
EBT	Evergreen broadleaved trees
ECMWF	European Center for Medium-Range Weather Forecasts
ECV	Essential Climate Variables
ERA	ECMWF Re-Analysis
ESA-CCI	European Space Agency (ESA) Climate Change Initiative (CCI)
ET	EvapoTranspiration
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
Euro	Europe
FCDR	Fundamental Climate Data Record
FMI	Finnish Meteorological Institute
GCOS	Global Climate Observing System
GLDAS	Global Land Data Assimilation System
GLEAM	Global Land Evaporation Amsterdam Model
GLOBMAP	Global Mapping
GLOBSNOW	Global Snow Monitoring for Climate Research
GSFC	Goddard Space Flight Center
H	Sensible heat flux
H SAF	Hydrological SAF
HOLAPS	High res-olution Land Atmosphere Parameters from Space
HTESSEL	land Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land
h_{tree}	Tree height
ICOS	Integrated Carbon Observation System
IFS	Integrated Forecast System
IGBP	International Geosphere-Biosphere Programme
IQR	interquartile
JJA	June July August
JPL	Jet Propulsion Laboratory
KNMI	Koninklijk Nederlands Meteorologisch Instituut
LAI	Leaf Area Index
LC	Land Cover
LDAS	Land Data Assimilation System
LE	Latent heat flux

LIS	Land Information System
LSA SAF	Land Surface Analysis Satellite Application Facility
LSM	Land Surface Model
LST	Land Surface Temperature
LUT	LookUp Table
MAD	Mean Absolute Difference
MAM	March April May
MARD	Mean absolute relative difference
MEaSURES	Making Earth System Data Records for Use in Research Environments
Meteosat	Meteorological Satellite(s) series operated by EUMETSAT
MeteoSwiss	Swiss Federal Office of Meteorology and Climatology
MFG	METEOSAT First Generation
MSG	Meteosat Second Generation
MSWEP	Multi-Source Weighted-Ensemble Precipitation
MVIRI	Meteosat Visible and Infrared Imager
N	Number
NAfr	North Africa
NASA	National Aeronautics and Space Administration
NCEP	National Center for Environmental Prediction
NetCDF	Network Common Data Form
NMHSs	National Meteorological and Hydrological Services
NSIDC	National Snow and Ice Data Center
NWP	Numerical Weather Prediction
Opt	Optimal requirement
OSCAR	Observing Systems Capability Analysis and Review Tool
Pa	Air pressure
PRD	Product Requirement Document
PT	Priestley-Taylor
PUM	Product User Manual
Q ₁	25th percentile
Q ₃	75th percentile
QC	Quality control
RMIB	Royal Meteorological Institute of Belgium
RMSD	Root Mean Square Difference
RR	Requirements Review
R _{smin}	Minimal stomatal resistance
SAF	Satellite Application Facility
SAfr	South Africa
SAL	Surface Albedo
SAmE	South America

SCOPE-CM	Sustained COordinated Processing of Environmental satellite data for Climate Monitoring
SD	Standard Deviation
SDL	Surface Downward Longwave radiation
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SIS	Surface Incoming Solar radiation
SMHI	Swedish Meteorological and Hydrological Institute
SON	September October November
SRB	Surface Radiation Budget
Ta	2 meter air temperature
Tar	Target requirement
TCDR	Thematic Climate Data Record
Td	2 meter dew-point temperature
Thr	Threshold requirement
U	U component 10 meter wind
UK MetOffice	Meteorological Service of the United Kingdom
uRMSD	Unbiased root mean square difference
V	V component 10 meter wind
VCF	Vegetation Continuous Fields
VODCA	Vegetation Optical Depth Climate Archive
WCRP	World Climate Research Programme
WMO	World Meteorological Organization
4D-Var	Four-dimensional variational data assimilation

Executive Summary

This CM SAF report presents the evaluation of the sensible (H) and latent (LE) heat fluxes climate data record (CM-23811). The evapotranspiration variable being also provided as ancillary product. Retrievals are obtained thanks to an adapted version of the methodology developed by the Land Surface Analysis Satellite Application Facility (LSA SAF) CDR [RD 1]. While meteorological and surface soil moisture are adapted from ERA5 dataset, radiation components are jointly retrieved using the CM SAF software “GeoSatClim” [RD 2] and based on observations from the Meteosat Visible and InfraRed Imager (MVISR) and the Spinning Enhanced Visible and Infrared Imager (SEVIRI), onboard of, respectively, Meteosat First and Second Generation (MFG and MSG).

The product is provided over the Meteosat disk at hourly, daily, monthly and monthly mean diurnal cycle time-step at a spatial resolution of 0.05 degrees (regular grid). The record cover 38 years: from January 1983 to December 2020.

Both products (LE and H) have been validated against observation sources (30 stations from FLUXNET2015 and ICOS datasets) and well-known data records from reanalysis (ERA5 and GLDAS) and satellite-based products (LSA SAF and GLEAM). All evaluations presented in this report have been done to provide a clear idea of the quality of the product to the user while checking their compliance with the optimal, target and threshold requirements of accuracy, precisions and stability. Tables Table 0-1 and Table 0-2 summarize validation results obtained by comparing estimations with in-situ observations at 30 stations and inter-product comparisons for stability for latent and sensible heat flux, respectively¹. Both products comply with the threshold requirements for all temporal aggregations and for the stability. However, they do not comply with the target requirements for all cases. Worth noting that while the uncertainty target requirements are not always reached, uRMSD values obtained are in the same order of magnitude referenced in the literature. In addition, similar range of uRMSD, as compared to in-situ data, are calculated for other tested products. Grid-based comparisons showed good agreements in the northern part of the Meteosat disk (latitudes higher than 16°N) but systematic differences have been observed in the southern regions.

Finally, decadal stability target requirement is only complied for the latent heat flux and for the sensible heat flux over Europe area.

Table 0-1: Summary of requirement compliance for the latent heat flux.

	Optimal Bias / uRMSD	Target Bias / uRMSD	Threshold Bias / uRMSD
<i>Hourly</i>	X / X	✓ / ✓	✓ / ✓
<i>Daily</i>	X / X	✓ / X*	✓ / ✓
<i>Monthly</i>	X / X	✓ / ✓**	✓ / ✓
<i>Monthly mean diurnal cycle</i>	X / X	✓ / ✓	✓ / ✓

¹ Conclusion for latent heat flux are also valid for the evapotranspiration (ET=LE λ; where λ is the latent heat of vaporization which depend on the temperature)

<i>Stability</i>	X / X	✓ / ✓	✓ / ✓
------------------	--------------	--------------	--------------

(*) **X**: Comply with the target for 60% of the stations (N=30 stations).

(**) **✓**: Comply with the target for 77% of the stations (N=30 stations).

Table 0-2: Summary of requirement compliance for the sensible heat flux products.

	Optimal Bias / uRMSD	Target Bias / uRMSD	Threshold Bias / uRMSD
<i>Hourly</i>	X / X	✓ / ✓	✓ / ✓
<i>Daily</i>	X / X	✓ / X	✓ / ✓
<i>Monthly</i>	X / X	✓ / X*	✓ / ✓
<i>Monthly mean diurnal cycle</i>	X / X	✓ / ✓	✓ / ✓
<i>Stability</i>	X / X	X / ✓	✓ / ✓

(*) **X**: Comply with the target for 63% of the stations (N=30 stations).

1 The EUMETSAT SAF on Climate Monitoring

The importance of climate monitoring with satellites was recognized in 2000 by EUMETSAT Member States when they amended the EUMETSAT Convention to affirm that the EUMETSAT mandate is also to “contribute to the operational monitoring of the climate and the detection of global climatic changes”. Following this, EUMETSAT established within its Satellite Application Facility (SAF) network a dedicated center, the SAF on Climate Monitoring (CM SAF, <http://www.cmsaf.eu>).

The consortium of CM SAF currently comprises the Deutscher Wetterdienst (DWD) as host institute, and the partners from the Royal Meteorological Institute of Belgium (RMIB), the Finnish Meteorological Institute (FMI), the Royal Meteorological Institute of the Netherlands (KNMI), the Swedish Meteorological and Hydrological Institute (SMHI), the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss), and the Meteorological Service of the United Kingdom (UK MetOffice). Since the beginning in 1999, the EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF) has developed and will continue to develop capabilities for a sustained generation and provision of Climate Data Records (CDR’s) derived from operational meteorological satellites.

In particular, the generation of long-term data sets is pursued. The ultimate aim is to make the resulting data sets suitable for the analysis of climate variability and potentially the detection of climate trends. CM SAF works in close collaboration with the EUMETSAT Central Facility and liaises with other satellite operators to advance the availability, quality and usability of Fundamental Climate Data Records (FCDRs) as defined by the Global Climate Observing System (GCOS). As a major task, the CM SAF utilizes FCDRs to produce records of Essential Climate Variables (ECVs) as defined by GCOS. Thematically, the focus of CM SAF is on ECVs associated with the global energy and water cycle.

Another essential task of CM SAF is to produce data sets that can serve applications related to the new Global Framework of Climate Services initiated by the WMO World Climate Conference-3 in 2009. CM SAF is supporting climate services at national meteorological and hydrological services (NMHSs) with long-term data records but also with data sets produced close to real time that can be used to prepare monthly/annual updates of the state of the climate. Both types of products together allow for a consistent description of mean values, anomalies, variability and potential trends for the chosen ECVs. CM SAF ECV data sets also serve the improvement of climate models both at global and regional scale.

As an essential partner in the related international frameworks, in particular WMO Sustained COordinated Processing of Environmental satellite data for Climate Monitoring (SCOPE-CM), the CM SAF - together with the EUMETSAT Central Facility, assumes the role as main implementer of EUMETSAT’s commitments in support to global climate monitoring. This is achieved through:

- Application of highest standards and guidelines as lined out by GCOS for the satellite data processing,
- Processing of satellite data within a true international collaboration benefiting from developments at international level and pollinating the partnership with own ideas and standards,
- Intensive validation and improvement of the CM SAF climate data records,

- Taking a major role in data set assessments performed by research organizations such as WCRP. This role provides the CM SAF with deep contacts to research organizations that form a substantial user group for the CM SAF CDRs,
- Maintaining and providing an operational and sustained infrastructure that can serve the community within the transition of mature CDR products from the research community into operational environments.

2 Introduction

This CM-23811 climate data record includes sensible and latent heat fluxes using input data derived from Meteosat First and Second Generation (MFG/MSG) satellites. The sensible heat flux represents the amount of energy transferred by convection and/or conduction from the surface to the atmosphere (Mito et al., 2012; Pipunic et al., 2008). The amount of energy and water consumed by evaporation and transpiration corresponds to the latent heat flux and the evapotranspiration process (Pipunic et al., 2008; Katul et al., 2012). By materializing the exchange of water and energy from the earth surface to the atmosphere, the latent and sensible heat fluxes control the development of the planetary boundary layer and govern land-atmosphere interaction (Michel et al., 2016; Behrendt et al., 2020). They play a major role in the hydrological cycle (Oki et al., 2006), carbon cycle (Sellers et al., 1997) and surface energy balance (Trenberth et al., 2009). Various applications as water resource management, agricultural planning, weather forecasting, drought/flood detection, etc., are possible thanks to their estimations (Fisher, 2017; Liou et al., 2014 and reference there in). For instance, monitoring of H/LE allows the detection of desertification, monsoon circulation and climate change (e.g., Yang et al., 2009; Wang and Li 2011; Shan et al., 2016).

To create the current CDR, surface latent and sensible heat fluxes, and the evapotranspiration are retrieved using an adapted version of the methodology developed by the Land Surface Analysis Satellite Application Facility (LSA SAF v3) [RD 1]. This latter can be described as a Surface Vegetation-Atmosphere Transfer (SVAT) scheme modified to accept input data from external sources (Gellens-Meulenberghs et al., 2006, 2007). The algorithm has been adapted from the Tiled ECMWF (European Centre for Medium-Range Weather Forecasts) Scheme for Surface Exchanges over Land (TESSEL) model (Van den Hurk et al., 2000; Viterbo and Beljaars, 1995) and H-TESEL (Balsamo et al, 2009) allowing the use of satellite-based data and numerical weather prediction (NWP) models' outputs (ECMWF reanalysis) as forcing. Observations from the Meteosat Visible and InfraRed Imager (MVISIR) and the Spinning Enhanced Visible and Infrared Imager (SEVIRI), onboard of, respectively, Meteosat First and Second Generation (MFG and MSG), are used as inputs for all radiation components - including the Surface Incoming Solar radiation (SIS), the Surface Albedo (SAL) and the Surface Downward Longwave radiation (SDL) - are jointly retrieved using the CM SAF software "GeoSatClim".

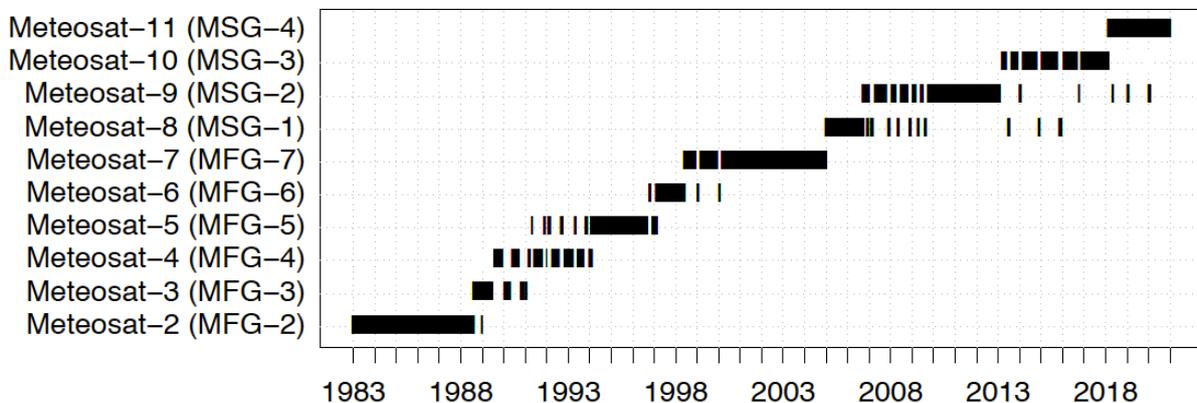


Figure 1: Overview of Meteosat satellites used as input for the generation of the land surface flux CDR.

The covered time period by consecutive Meteosat missions used for SIS, SDL and SAL data record are shown in Figure 1. The data record is based on 21 years (1983–2004) of MFG

	Validation Report Meteosat Latent and Sensible heat fluxes - Edition 1	Doc.No: SAF/CM/RMIB/VAL/MET/LEH Issue: 1.1 Date: 30.05.2023
---	--	---

MVIRI and 15 years (2005–2020) of MSG SEVIRI data, respectively. Data estimated from Meteosat-7 and MSG-1 being available simultaneously in 2004 and 2005, differences in estimations for the year 2005 from MVIRI and SEVIRI-based inputs have been analyzed in the section 0.

The present document aims at evaluating the hourly, daily and monthly mean and the monthly mean diurnal cycle of sensible heat flux (H) and latent heat flux (LE). As ancillary products, the evapotranspiration (ET) is also considered (values are accumulated and not averaged).

The main features and inputs of the CM-23811 data records are summarized in Tables 2-1 and Table 2-2.

Table 2-1: Main features of the CM-23811 data records.

Sensors	MVIRI & SEVIRI.
Methodology	Adaptation from LSA SAF.
Covered period	1983 to 2020.
Area Covered	Meteosat Disk (60°N – 60°S; 60°W-60°E).
Temporal characteristics	hourly mean, daily mean, monthly mean and the monthly mean diurnal cycle.
Spatial resolution	Regular lat-lon grid with a spatial resolution of (0.05°), i.e., about (5.5 km) ² at sub-satellite point.
Output quantities	Latent and sensible heat fluxes (W m ⁻²) and evapotranspiration (mm h ⁻¹ , mm day ⁻¹ or mm month ⁻¹).
Format	NetCDF file following the CF convention.

Table 2-2: Main inputs variables used in the CM SAF algorithm.

Dataset name (short name; unit)	Sources
Surface Incoming Shortwave radiation (<i>SIS</i> ; W m ⁻²) Surface Downward Longwave radiation (<i>SDL</i> ; W m ⁻²) Surface Albedo (<i>SAL</i>)	CM-23271 [RD 2]
Leaf Area Index (<i>LAI</i> ; m ² m ⁻²)	GLOBMAP (<i>Liu et al., 2012, 2017</i>)
Land Cover (<i>LC</i>)	ESA-CCI (<i>ESA 2017; Bontemps et al., 2012</i>)
Tree height (<i>h_{tree}</i> , m)	NASA/JPL (<i>Simard et al., 2011</i>)
Meteorological data (<i>T_a</i> , <i>T_d</i> , <i>U</i> , <i>V</i> , <i>P_a</i>) & Surface soil moisture	ERA5 (<i>Hersbach et al., 2019</i>)

	<p style="text-align: center;">Validation Report Meteosat Latent and Sensible heat fluxes - Edition 1</p>	Doc.No: SAF/CM/RMIB/VAL/MET/LEH Issue: 1.1 Date: 30.05.2023
---	---	---

3 Validation strategy

The purpose of the validation is to characterize the products in terms of their accuracy, precision and stability, thus to give limitations in various aspects and guidance for product applicability. Each product evaluation is confronted to user's requirements defined in [AD 1] which are based on GCOS values, a literature review and user's needs. The Table 3-2 presents the requirements for the hourly, daily and monthly datasets. Note that similar requirements are used for the hourly and the monthly mean diurnal cycle dataset. The requirements are sub-divided in 3 categories: threshold, target and optimal. Those categories refer the one defined by the WMO Observing Systems Capability Analysis and Review Tool (OSCAR, <https://space.oscar.wmo.int/observingrequirements>):

- Threshold: the minimum requirement to be met to ensure that data are useful.
- Optimal: or the "goal" is an ideal requirement above which further improvements are not necessary.
- Target: an intermediate level between "threshold" and "optimal" which, if achieved, would result in a significant improvement for the targeted application.

The accuracy, the precision and the stability are evaluated by calculating, respectively, the bias, the unbiased root mean square difference (uRMSD) and the decades slope (see the Appendix in section 7.1 for calculation details).

The climate data record is validated by comparing with:

- 1) *In situ datasets*: FLUXNET2015 and the ICOS datasets at selected locations in different climatic/vegetation conditions.
- 2) *Reanalyse datasets*: output of models (ingesting satellite data or not) recognized to produce valuable meteorological information in about the same conditions: near-real time and intra-day variations of evapotranspiration and fluxes, e.g., GLDAS Noah and ERA5.
- 3) *Satellite-based products*: GLEAM (ET) and LSA SAF (LE and H) satellite products.

These products have been selected according to their temporal coverage and their spatial and temporal resolutions. The Table 3-1 summaries their characteristics (see details about each product in the following sub-sections).

The validation (section 4) is sub-divided in two main parts: comparisons with in situ data at station points and comparisons with reanalysis and satellite-based products. Comparison with in situ data allows an evaluation of the dataset with independent data. However, while FLUXNET2015/ICOS is considered as the reference dataset, it is subject to error between 10% and 30%; e.g., Wilson et al., 2002, Foken et al., 2008; Franssen et al., 2010; Stoy et al., 2013; Senay et al., 2020.

Table 3-1: Main characteristics of datasets used for the validation.

	FLUXNET2 015/ICOS	ERA5	GLDAS v2.0	GLEAM v3.5a	LSA SAF v3
Variables	LE-H	LE-H	LE-H-ET	ET	LE-H-ET
Spatial resolution	point	0.25°	0.25°	0.25°	0.05°
Temporal resolution used for comparisons	Hourly Daily Monthly	Hourly	Daily Monthly	Daily Monthly	Daily Monthly
Period covered used for comparisons	-	1983-2020	2000-2020	1983-2020	2004-2020
Product type	Eddy-covariance	Reanalyse	Reanalyse	Satellite-based	Satellite-based
References	Pastorello et al., 2020	Hersbach et al., 2019	Rodell et al., 2004	Miralles et al., 2011 ; Martens et al., 2017	Ghilain et al., 2012

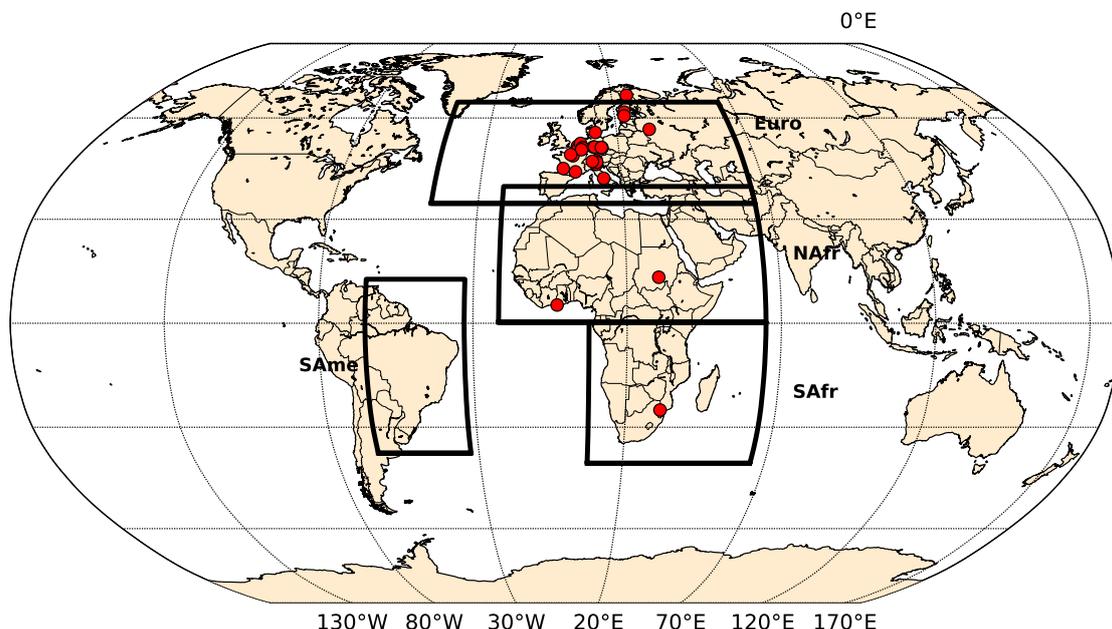
Table 3-2: Product requirements for evapotranspiration and surface heat fluxes (CM-23811). LE_{obs} , H_{obs} and ET_{obs} refer to the mean absolute value of the reference dataset over the time period analyzed.

Product	Metrics	Optimal	Target	Threshold
<i>Evapotranspiration (ET)</i>				
hourly-mean [mm/h]	Bias	$<0.1 \times ET_{Obs}$	$< (0.2 \times ET_{Obs} + 0.02)$	$< (0.4 \times ET_{Obs} + 0.03)$
	uRMSD	$<0.1 \times ET_{Obs}$	$< (0.2 \times ET_{Obs} + 0.06)$	$< (0.4 \times ET_{Obs} + 0.12)$
Daily-accumulated [mm/day]	Bias	$<0.1 \times ET_{Obs}$	$< (0.2 \times ET_{Obs} + 0.35)$	$< (0.4 \times ET_{Obs} + 0.71)$
	uRMSD	$<0.1 \times ET_{Obs}$	$< (0.2 \times ET_{Obs} + 0.53)$	$< (0.4 \times ET_{Obs} + 1.06)$
Monthly-accumulated [mm/mth]	Stability [mm/dec]	<0.1	<4	<8
	Bias	$<0.1 \times ET_{Obs}$	$< (0.2 \times ET_{Obs} + 10.6)$	$< (0.4 \times ET_{Obs} + 21.2)$
	uRMSD	$<0.1 \times ET_{Obs}$	$< (0.2 \times ET_{Obs} + 10.6)$	$< (0.4 \times ET_{Obs} + 21.2)$
<i>Latent heat flux (LE)</i>				
hourly-mean [W/m ²]	Bias	$<0.1 \times LE_{Obs}$	$< (0.2 \times LE_{Obs} + 10)$	$< (0.4 \times LE_{Obs} + 20)$
	uRMSD	$<0.1 \times LE_{Obs}$	$< (0.2 \times LE_{Obs} + 40)$	$< (0.4 \times LE_{Obs} + 80)$
Daily-mean [W/m ²]	Bias	$<0.1 \times LE_{Obs}$	$< (0.2 \times LE_{Obs} + 10)$	$< (0.4 \times LE_{Obs} + 20)$
	uRMSD	$<0.1 \times LE_{Obs}$	$< (0.2 \times LE_{Obs} + 15)$	$< (0.4 \times LE_{Obs} + 30)$
Monthly-mean [W/m ²]	Stability [W/m ² /dec]	<0.1	<4	<8
	Bias	$<0.1 \times LE_{Obs}$	$< (0.2 \times LE_{Obs} + 10)$	$< (0.4 \times LE_{Obs} + 20)$
	uRMSD	$<0.1 \times LE_{Obs}$	$< (0.2 \times LE_{Obs} + 10)$	$< (0.4 \times LE_{Obs} + 20)$
<i>Sensible heat flux (H)</i>				
hourly-mean [W/m ²]	Bias	$<0.1 \times H_{Obs}$	$< (0.3 \times H_{Obs} + 10)$	$< (0.6 \times H_{Obs} + 20)$
	uRMSD	$<0.1 \times H_{Obs}$	$< (0.3 \times H_{Obs} + 50)$	$< (0.6 \times H_{Obs} + 100)$
Daily-mean [W/m ²]	Bias	$<0.1 \times H_{Obs}$	$< (0.3 \times H_{Obs} + 10)$	$< (0.6 \times H_{Obs} + 20)$
	uRMSD	$<0.1 \times H_{Obs}$	$< (0.3 \times H_{Obs} + 15)$	$< (0.6 \times H_{Obs} + 30)$
Monthly-mean [W/m ²]	Stability [W/m ² /dec]	<0.1	<4	<8
	Bias	$<0.1 \times H_{Obs}$	$< (0.3 \times H_{Obs} + 10)$	$< (0.6 \times H_{Obs} + 20)$
	uRMSD	$<0.1 \times H_{Obs}$	$< (0.3 \times H_{Obs} + 10)$	$< (0.6 \times H_{Obs} + 20)$

3.1 *In situ* datasets (FLUXNET2015 / ICOS)

In situ measurements used as reference dataset were extracted from the FLUXNET2015 dataset (<https://fluxnet.fluxdata.org>) and from the Integrated Carbon Observation System (ICOS) international platform (<https://www.icos-cp.eu>). While the FLUXNET2015 and ICOS networks provide data from 212 stations (117 in the field of view of METEOSAT) we selected 30 of them to perform analysis. Four criteria were used to select the stations: i) well established stations and already used datasets by the LSA SAF; ii) data quality and representativeness (Rebmann et al., 2005); iii) ability to sample different climates and biomes; and iv) spatial homogeneity of the station. Localizations of selected sites are shown in Figure 2. An overview of the sites can be found in Table 7-1 (i.e., climate, area, available years etc.).

At each site, turbulent fluxes data are gap-filled using the marginal distribution sampling (MDS) method (Reichstein et al., 2005) and are available at half-hourly (local time), daily and monthly time scales (full description in Pastorello et al., 2020 and <https://fluxnet.org/data/fluxnet2015-dataset/data-processing>). Eddy covariance measurements do not allow an energy balance closure (EBC; the sum of latent and sensible fluxes equal to net radiation minus ground heat flux). The energy imbalance can be between 10 and 30% (average 20%) (e.g., Wilson et al., 2002, Foken et al., 2008; Franssen et al., 2010; Stoy et al., 2013; Senay et al., 2020). At the opposite, the EBC is closed while using CM SAF data. Thus, to do the comparison we use corrected variables, named “LE_CORR” and “H_CORR”, which are calculated thanks to a closure correcting factor differing according to the time resolution and assuming a correct Bowen ratio (Bowen et al., 1926; Foken et al., 2008; Pastorello et al., 2020). For each time-step, different quality control, are provided to ensure the quality level of the gap-filling. For half-hourly data, quality control values of 0 (measured), 1 (good), 2 (medium) or 3 (bad). For daily and monthly data, fractions between 0-1 indicate the percentage of measured and good quality gap-filled data. To minimize potential differences due to poor quality measurements we extracted data with a quality control of 1 (“good quality”) for half-hourly data, and 100 % of good data for daily and monthly data. Quality control limit has been relaxed to 2 (medium) for the monthly mean diurnal cycle (not enough match-up with a quality control of 1).



	Validation Report Meteosat Latent and Sensible heat fluxes - Edition 1	Doc.No: SAF/CM/RMIB/VAL/MET/LEH Issue: 1.1 Date: 30.05.2023
---	---	---

Figure 2: Location of eddy covariance stations used to evaluate the CM SAF dataset. Black frames delineate: Europe (Euro; lat:[34.49,60], lon:[-46.1, 60]), North-Africa (NAfr; lat:[0.2, 39.4], lon:[-21.7, 60]), South-Africa (SAfr; lat:[-40.5, 0.2], lon:[7.7, 60]) and South-America (SAm; lat:[-37.6,12.6], lon:[-60, -32.8]) areas (LSA SAF regions).

The station's half-hourly data are hourly averaged and an offset is applied to get the hour in UTC and not in local time.

It is worth-noting that *in situ* measurements are subject to typical error between 10% and 30% (Chehbouni et al., 1999; Chavez et al., 2009; Watts et al., 2000; Wilson et al., 2002; Drexler et al., 2004; McJannet et al., 2011; Perez et al., 2017; Hirschi et al., 2017; Foken et al., 2008 and reference there in).

Please note that, the evapotranspiration being not directly available in this *in situ* archive, it was not possible to evaluate it but similar conclusions can be made for LE and ET as those two variables are linked via a direct relation ($ET=LE \lambda$; where λ is the latent heat of vaporization).

3.2 ERA5 dataset

ERA5 (Hersbach et al., 2019) is the fifth generation of global atmospheric reanalysis generated by the ECMWF. ERA5 is produced from the ECMWF's Integrated Forecast System (IFS) cycle 41r2 using the four-dimensional variational data assimilation (4D-Var) and the Land Data Assimilation System (LDAS) to assimilate, respectively, atmospheric and land data. The Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (H-TESEL) is the land surface model. This later demonstrated its performance to simulate surface turbulent heat fluxes in offline experiments (Balsamo et al., 2009; Balsamo et al., 2015; Albergel et al., 2012). All global atmospheric, oceanic and land surface fields are available at an hourly time step with a spatial resolution of 0.25° ($\sim 32\text{km}$) covering the period from January 1950 to present, thus the full extend of CM-23811.

Specific analysis of the land-surface energy partitioning in ERA5 has been performed in Martens et al., 2020. Comparison at station showed that ERA5 tends to overestimate LE with Mean difference around 10 W m^{-2} . ERA5 appeared to be almost unbiased for sensible heat flux (minor underestimation).

3.3 GLDAS dataset

Global Land Data Assimilation System version 2 (GLDAS; Rodell et al., 2004) is a new generation of reanalysis developed jointly by the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) and National Center for Environmental Prediction (NCEP). GLDAS, which has been streamlined and parallelized by the Land Information System (LIS; Kumar et al. 2006), generates land surface products by using various offline (not coupled to the atmosphere) land surface models (LSM) and ingesting satellite- and ground-based observational datasets (Rodell et al., 2004). Details about the forcing data and description of the model are available on <http://disc.Sci.GSFC.NASA.Gov/Hydrology>. Currently, GLDAS has three components: GLDAS-2.0 (1948-2014), GLDAS-2.1 (2000- Present), and GLDAS-2.2 (Fev 2003-Present).

Beyond their differences in term of forcing or data assimilation source, the choice of the product has been dictated by the period covered to validate the CM SAF product. GLDAS-2.0,2.1 and 2.2 cover, respectively, 31, 20 and 17 years. Thus, we selected the “open-loop” (i.e., no data assimilation) GLDAS-2.0 product with a daily and monthly (averaged from 3-hourly) temporal resolution. GLDAS-2.0 has been forced entirely with the Princeton meteorological forcing input (Sheffield et al., 2006) and the data has been simulated by the Noah Model 3.3 (Chen et al., 1996; Ek et al., 2003) in Land Information System Version 7 with a spatial resolution of 0.25°.

3.4 LSA SAF v3 dataset

As mentioned in the section 2, the LSA SAF v3² and the CM SAF core model is similar and only the forcing differs. The Table 3-3 summarize input sources used in the LSA SAF code and the current one. For instance, while ERA5 is used as input to get the soil moisture in the CM SAF code, the H SAF dataset is used in the LSA SAF approach. Other difference can be seen in the retrieval of the LAI for each tile. Indeed, an inversion matrix approach is used in the LSA SAF (Ghilain et al., 2012) while an LUT approach is adopted here. All details about CM SAF (pre)processing of the data inputs can be seen in the [RD 1] documents. Finally, the minimal stomatal resistance ($R_{s_{min}}$) value for evergreen broadleaved trees tile is of 140 in the LSA SAF model and of 200 in the CM SAF model.

Table 3-3: Input sources used in the LSA SAF v3 and the CM SAF approach.

Variable name	CM SAF source	LSA SAF source
Surface Incoming shortwave radiation	CM SAF	LSA SAF (CDR)
Surface Downward Longwave radiation	CM SAF	LSA SAF (CDR)
Surface Albedo	CM SAF	LSA SAF (CDR)
Leaf Area Index	GLOBMAP	LSA SAF (CDR)
Land Cover	ESA CCI	ECOCLIMAP
Surface Soil Moisture	ERA5	H SAF

3.5 GLEAM dataset

The Global Land Evaporation Amsterdam Model (GLEAM; Miralles et al., 2011, Martens et al., 2017) is a remote sensing-based model allowing the estimation of the terrestrial evapotranspiration components (transpiration, bare soil evaporation, interception loss, and sublimation) and root-zone soil moisture. To correct random forcing errors, and other potential effects such as irrigation, that are not explicitly modelled in GLEAM, observations of surface soil moisture are also assimilated into the soil profile. Interception loss is calculated separately in GLEAM using a Gash’s analytical model (Miralles et al., 2010 and reference therein). Finally, the potential evapotranspiration is calculated using the Priestley-Taylor (PT) equation (Priestley and Taylor, 1972), and then, the actual evapotranspiration is obtained by including stress factors such as soil moisture state and vegetation physiological characteristics. The key features of this model are the use of microwave-derived soil moisture, land surface temperature and vegetation density, and the detailed estimation of rainfall interception loss.

² The LE and H datasets have been processed off-line for 2004–2022, and the data are available to users in the [LSA SAF Dataserver](#) as a demo with a release by the end of the year.

	Validation Report Meteosat Latent and Sensible heat fluxes - Edition 1	Doc.No: SAF/CM/RMIB/VAL/MET/LEH Issue: 1.1 Date: 30.05.2023
---	--	---

Here, we use GLEAM V3.5 data globally at daily and monthly temporal resolutions with a spatial resolution of 0.25°.

Table 3-4: Overview of forcing dataset use in GLEAM V3.5.

Variable name	Dataset name	Reference
Radiation	ERA5	Hersbach et al., 2019
Air temperature	ERA5	Hersbach et al., 2019
Precipitation	MSWEP v2.8	Beck et al., 2017
Snow water equivalent	GLOBSNOW L3v2 & NSIDC v01	Luojus et al., 2013 & Armstrong et al., 2005
Vegetation optical depth	VODCA	Moesinger et al., 2020
Surface Soil Moisture	ESA-CCI v5.3	Gruber et al., 2017 Dorigo et al., 2017
Vegetation fraction cover	MEaSURES VCF5KYR_001	Hansen et al., 2018

Validation

The main goal of the validation is to characterize the current products in term of accuracy and precision and stability while ensuring that the requirements are reached. For all the validation, the shaded areas in plots represent the optimal (green), target (blue) and threshold (red) requirements. Please note that requirement values would depend on the time series analyzed. To facilitate the distinction between datasets, the following color code is used: CM SAF in blue, ERA5 and GLEAM in green, GLDAS in orange and LSA SAF in violet.

3.6 Comparison with FLUXNET2015 / ICOS

In the following section, matchups are obtained by extracting the closest pixel of each station localization (see Table 7-1) at the desired time. It seems important to remind that most of the stations are mainly located in Europe and so the full Meteosat disk is not evenly covered (see section 4.2 for grid-based analysis).

3.6.1 Surface latent heat flux

Figures 3, 4 and 5 display the bias (top panels) and the uRMSD (bottom panels) of selected products (CM SAF, ERA5, LSA SAF and GLDAS) as compared to *in situ* FLUXNET2015/ICOS data for hourly, daily and monthly latent heat flux datasets, respectively. Performance metrics presented have been calculated from a subset including only observations with common matchups for all products. Thus, performance metrics can be directly compared between products.

Results show that CM SAF values comply with the target accuracy and precision requirements (blue shaded area in the figures) as most of the values (blue bar in the figures) are below the target requirement limit. Concerning the bias, 73% (average of -11 W m⁻²), 73% (average of -13 W m⁻²) and 70% (average of -14 W m⁻²) of the stations comply with the target requirements for hourly, daily and monthly data, respectively. Considering all time-step, CM SAF data have higher biases than other datasets. The mean biases at all stations (N=90) is of -13 W m⁻² for CM SAF while it is of -3 W m⁻² (N=90) 4 W m⁻² (N=90) and 3 W m⁻² (N=60) for LSA SAF, ERA5 and GLDAS, respectively.

Concerning uRMSD, 87% (average of 31 W m^{-2}), 60% (average of 23 W m^{-2}) and 77% (average of 17 W m^{-2}) of the stations comply with the target requirements for hourly, daily and monthly data, respectively. Note that considering all temporal resolution, the Skukuza station appears to be the localization where the worst results are obtained (never below the target requirement). Some doubts about corrections applied to close the budget (see section 3.1) would need to be raised as bias is linearly correlated with the correction applied (see section 7.2.2). Thus, this station has not been considered in the sections 4.1.3 and 4.4.

Worth noting that uRMSD values obtained for the CM SAF dataset are in-line with other dataset performances. Indeed, results obtained from CM SAF and all other products differ, on average, of 1.3% (N=240). In addition, results are consistent with the literature (e.g., Loew et al., 2016, Martens et al., 2017, Zhang et al., 2017, Albergel et al., 2018, Peng et al., 2020, Guo et al., 2022, Xin et al., 2022). For instance, by comparing estimations from 7 land surface models to eddy correlation measurements in China, Guo et al., 2022 calculated bias values from -14.27 to -2.93 W m^{-2} and uRMSD ranged from 29.8 W m^{-2} to 39.3 W m^{-2} .

These various observations are confirmed with the detailed statistics table (Table 4-1) which includes performance statistics for all available matchups between CM SAF and *in situ* data (no restriction with other datasets) for different time resolutions and periods. Globally, similar results are observed for bias and uRMSD complying with the target requirements. As expected, better absolute results are obtained in the winter season (most of the stations being in Europe) and during the night. The bias and the uRMSD are on average 2.7 and 2.2 times higher in summer than in winter. In addition, bias and uRMSD are, respectively, 4.1 and 4.6 times better during the night than the day. Those observations tend to indicate better absolute retrievals for low value cases. Please note that if we look at relative errors, the model performs better for high value cases. For instance, the bias -6.5 W m^{-2} observed in winter represents a relative error of 76% as compared to mean *in situ* value while, the bias of -12.8 W m^{-2} in summer represents only 26% of the mean *in situ* value. It is worth noting that an underestimation is obtained for all considered cases (seasons, day/night and overall). Deeper investigation would be required in the future to find the source of this negative bias and adapt the model or the inputs accordingly.

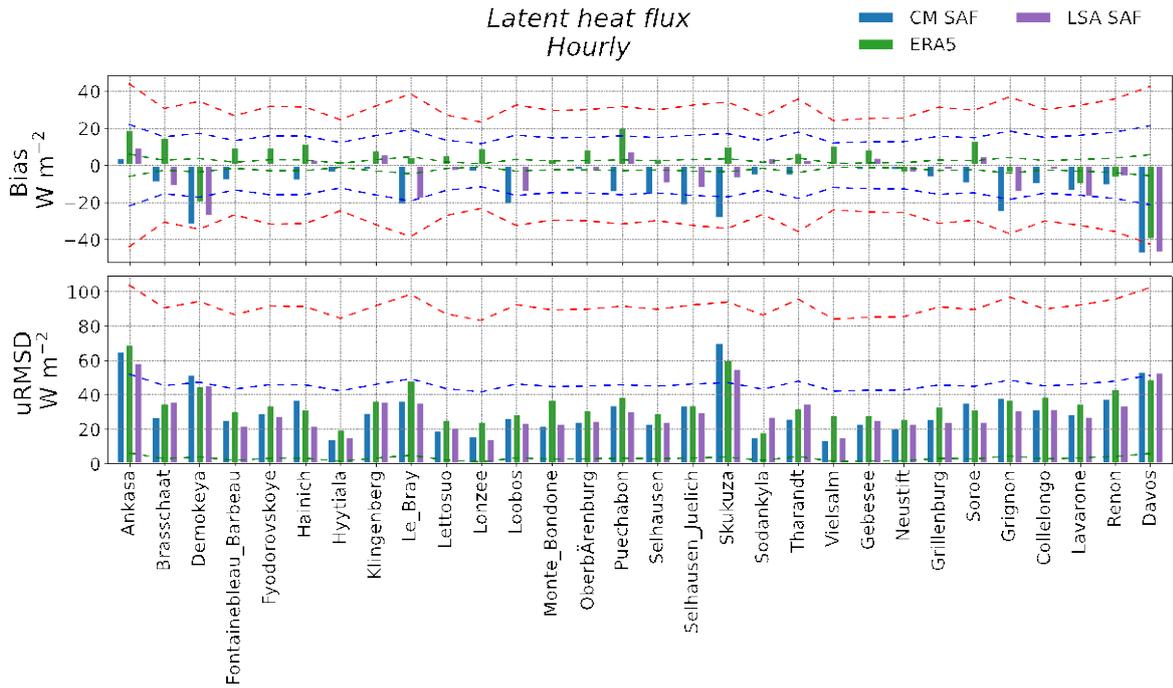


Figure 3: Bar plot showing the hourly bias ($W m^{-2}$; top) and the hourly unbiased root mean square error ($W m^{-2}$; bottom) of CM SAF (blue), ERA5 (green) and LSA SAF (violet) datasets as compared to FLUXNET2015/ICOS dataset at stations for the latent heat flux. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements.

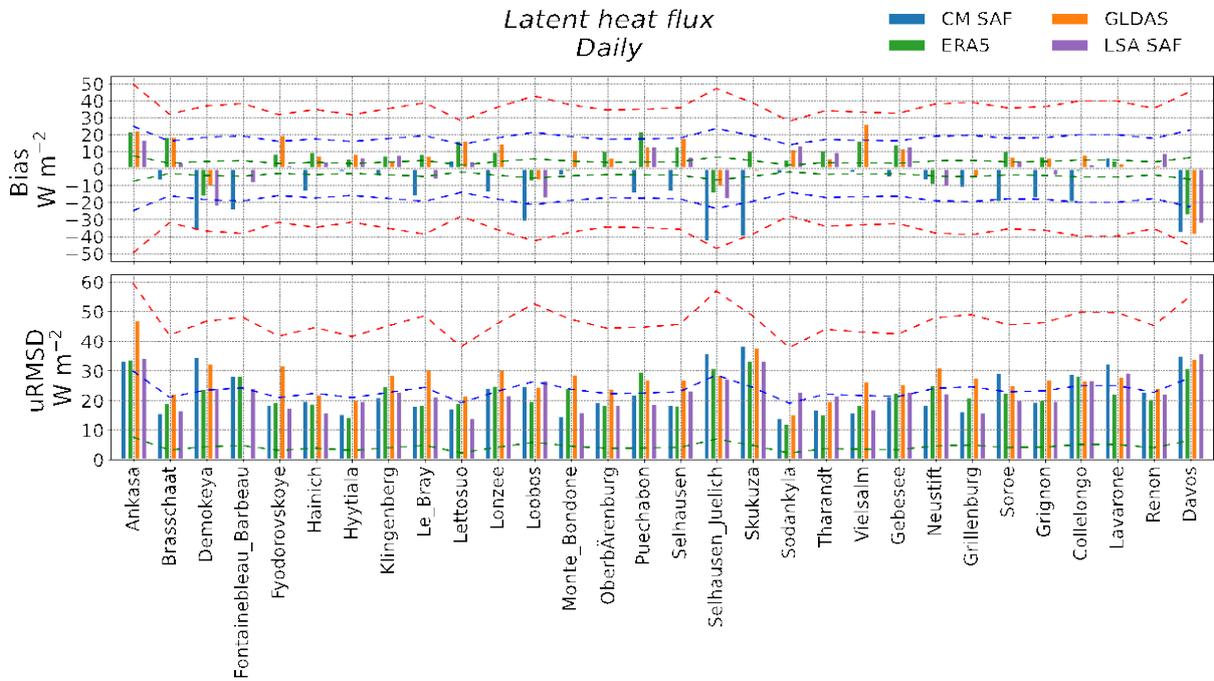


Figure 4: Bar plot showing the daily bias ($W m^{-2}$; top) and the daily unbiased root mean square error ($W m^{-2}$; bottom) of CM SAF (blue), ERA5 (green), GLDAS (orange) and LSA SAF (violet) datasets as compared to FLUXNET2015/ICOS dataset at stations for the latent heat flux. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements.

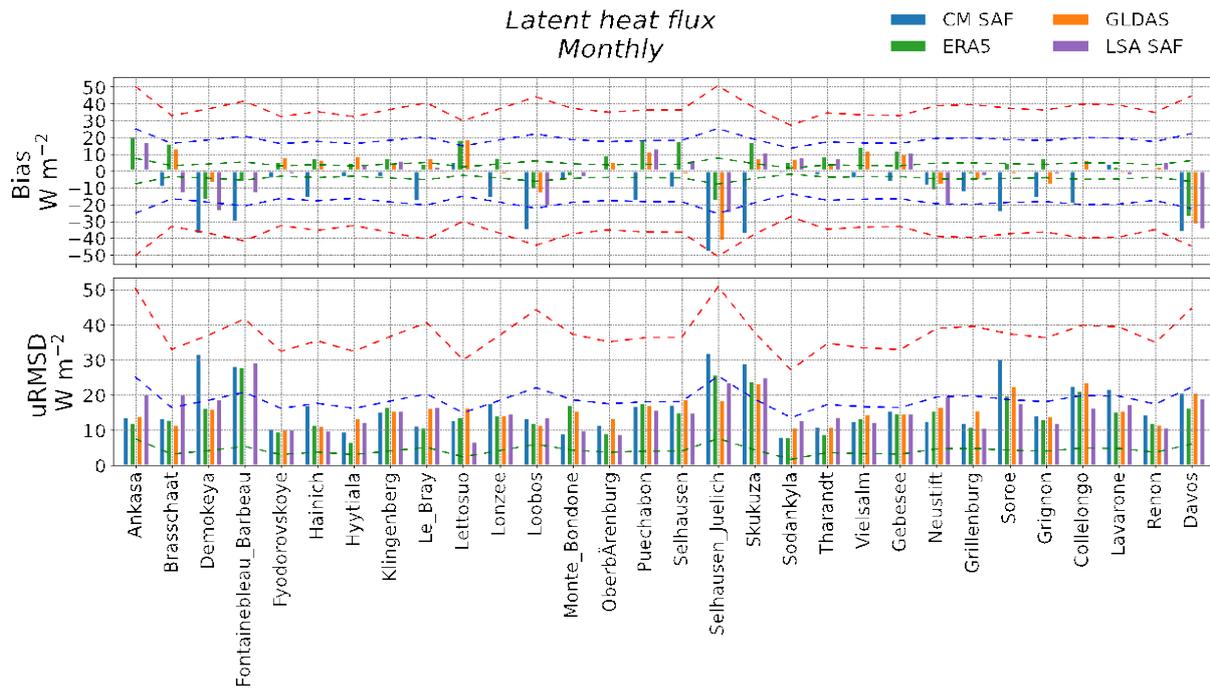


Figure 5: Same as Figure 4 but for monthly latent heat flux.

Table 4-5: Performance statistics of hourly, daily and monthly CM SAF latent heat flux dataset as compared to FLUXNET2015/ICOS observations at 30 stations for different seasons and day/night conditions.

	N	< FLUXNET2015/ICOS > (W m ⁻²)	< CM SAF > (W m ⁻²)	Bias (W m ⁻²) ± (Opt / Tar / Thr)**	uRMSD (W m ⁻²) (Opt / Tar / Thr)**
<i>Hourly</i>					
DJF*	174598	8.6	-2.2	-6.5 (1 / 11.9 / 23.8)	22.9 (1 / 41.9 / 83.8)
MAM*	122072	29.7	18	-11.7 (3 / 16.1 / 32.2)	36.5 (3 / 46.1 / 92.2)
JJA*	120042	49.7	36.9	-12.8 (5 / 20.1 / 40.1)	44.2 (5 / 50.1 / 100.1)
SON*	149168	17.5	8.2	-9.3 (1.8 / 13.7 / 27.4)	26.6 (1.8 / 43.7 / 87.4)
Day	214847	55.3	37.1	-18.2 (5.6 / 21.1 / 42.3)	49.6 (5.6 / 51.1 / 102.3)
Night	323476	5.1	0.7	-4.4 (0.6 / 11.2 / 22.5)	10.9 (0.6 / 41.2 / 82.5)
Overall	565880	24.2	14.5	-9.7 (2.5 / 15 / 30)	32.5 (2.5 / 45 / 90)
<i>Daily</i>					
DJF*	24357	8	3	-5 (0.8 / 11.7 / 23.4)	14.44 (0.84 / 16.7 / 23.4)
MAM*	26138	42.7	32.1	-10.6 (4.3 / 18.6 / 37.1)	25.16 (4.3 / 23.6 / 37.1)
JJA*	27293	77.4	61.5	-15.9	33.9

				(7.7 / 25.5 / 51)	(7.7 / 30.5 / 51)
SON*	27335	26.2	15.3	-10.9 (2.6 / 15.3 / 30.5)	19 (2.6 / 20.3 / 30.5)
Overall	105123	39.4	28.6	-10.8 (4 / 17.9 / 35.8)	24.7 (4 / 22.9 / 35.8)
<i>Monthly</i>					
DJF*	747	9	3	-5.9 (0.9 / 11.9 / 23.7)	12.1 (0.9 / 11.8 / 23.7)
MAM*	825	44.8	32.2	-12.6 (4.5 / 19 / 38)	18.8 (4.5 / 19 / 38)
JJA*	853	79.3	62.3	-17.1 (7.9 / 25.9 / 51.7)	26.4 (7.9 / 25.9 / 51.7)
SON*	870	28.1	15.1	-12.9 (2.8 / 15.6 / 31.3)	16.4 (2.8 / 15.6 / 31.3)
Overall	3295	41.2	28.8	-12.32 (4.1 / 18.3 / 36.5)	19.7 (4.1 / 18.3 / 36.5)

(*) DJF: December January February; MAM: March April May; JJA: June July August; SON: September October November.

(**) Opt: Optimal requirement; Tar: Target requirement; Thr: Threshold requirement.

3.6.2 Surface sensible heat flux

Figures 6, 7 and 8 display the bias (top panels) and the uRMSD (bottom panels) of selected products (CM SAF, ERA5, LSA SAF and GLDAS) as compared to *in situ* FLUXNET2015/ICOS data for hourly, daily and monthly sensible heat flux datasets. Performance metrics presented have been calculated from a subset including only observations with common matchups for all products.

As with the latent heat flux, results show that CM SAF dataset (blue bar in the figures) comply, for most of the stations, the target accuracy and precision requirements (blue shaded area in the figures) for the bias and the uRMSD. Concerning the bias, 90% (average of -2 W m^{-2}), 77% (average of -0.7 W m^{-2}) and 73% (average of -0.2 W m^{-2}) of the stations comply with the target requirements for hourly, daily and monthly data, respectively. CM SAF sensible heat flux is almost unbiased with mean bias, over all time-step and stations, of -1 W m^{-2} (N=90).

Other products slightly underestimate the sensible heat flux with mean biases of -5 W m^{-2} (N=90), -7 W m^{-2} (N=90) and -10 W m^{-2} (N=60) for LSA SAF, ERA5 and GLDAS datasets, respectively.

Concerning the precision (uRMSD), 93% (average of 47 W m^{-2}) and 63% (average of 18 W m^{-2}) of the stations comply with the target requirements for hourly, and monthly data, respectively. For daily data, most of the values are slightly higher than the target requirement value. Indeed, 30% (average of 28 W m^{-2}) of the stations comply with the target requirement (average of 25.2 W m^{-2}) but all stations comply with the threshold requirements.

Please note that CM SAF uRMSD values at stations are in the same order of magnitude than other products. Results obtained from CM SAF and all other products differ, on average, of -3% (N=240). In addition, results for our comparison with FLUXNET2015/ICOS is in the same order of magnitude as reported in previous studies

(e.g., Loew et al., 2016, Zhang et al., 2017, Albergel et al., 2018, Siemann et al., 2018, Peng et al., 2020, Xin et al., 2022). For instance, Loew et al., 2016 obtained, while comparing sensible heat flux (from HOLAPS) with data at 48 Fluxnet stations, RMSDs between 38 and 84 $W m^{-2}$ for hourly and daily data.

The Table 4-2 presents the performance metrics between CM SAF and FLUXNET2015/ICOS data including all possible matchups for all stations for different time resolutions and periods. The bias always complies with the target requirement and even with the optimal requirement in some conditions. Daily and monthly data tend to underestimate the sensible heat flux (bias between $-1.8 W m^{-2}$ and $-6.9 W m^{-2}$) except in autumn ($\sim +4 W m^{-2}$). No clear significant bias is observed for the hourly product. Please note that the low mean values for hourly data ($\sim 6 W m^{-2}$) as compared to daily and monthly data ($\sim 23 W m^{-2}$) is due to the high number of night cases.

Concerning the precision (uRMSD), values are slightly higher than the target requirements for daily and monthly data but hourly results are well below the target. As observed for the LE, the model has a better absolute accuracy for low values cases (night and winter) than high values cases (day and summer). For instance, the uRMSD values in of $69 W m^{-2}$ during the day while it is of $25 W m^{-2}$ during the night.

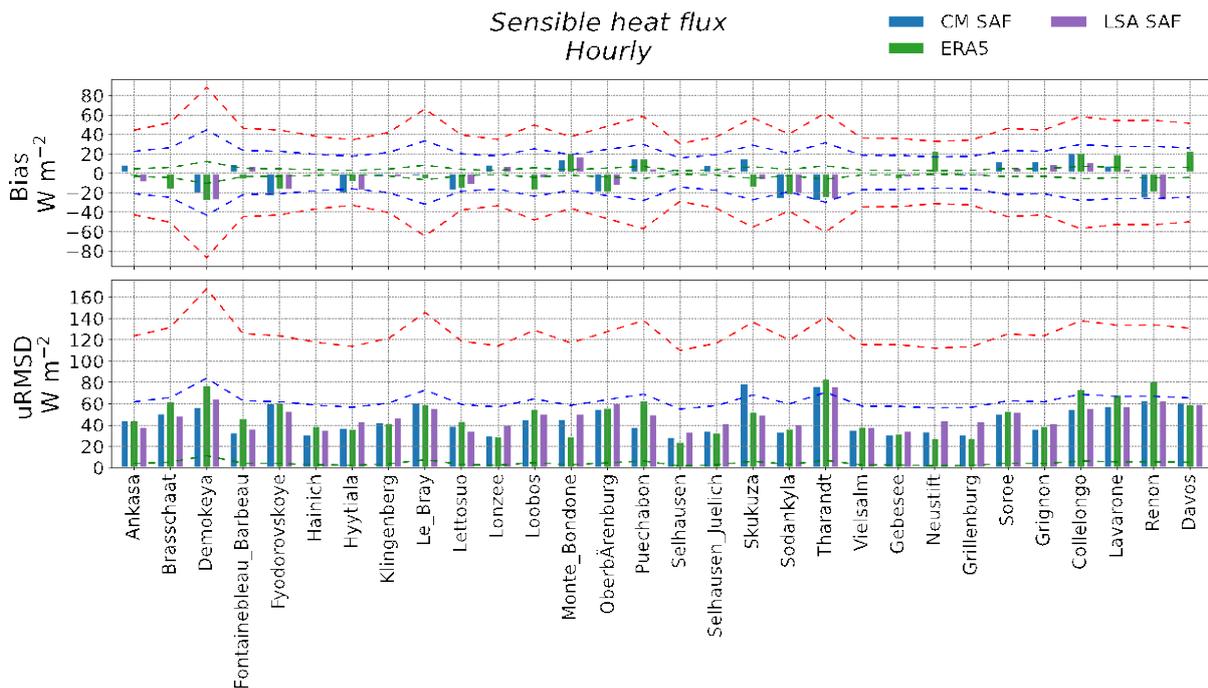


Figure 6: Same as Figure 3 but for sensible heat flux.

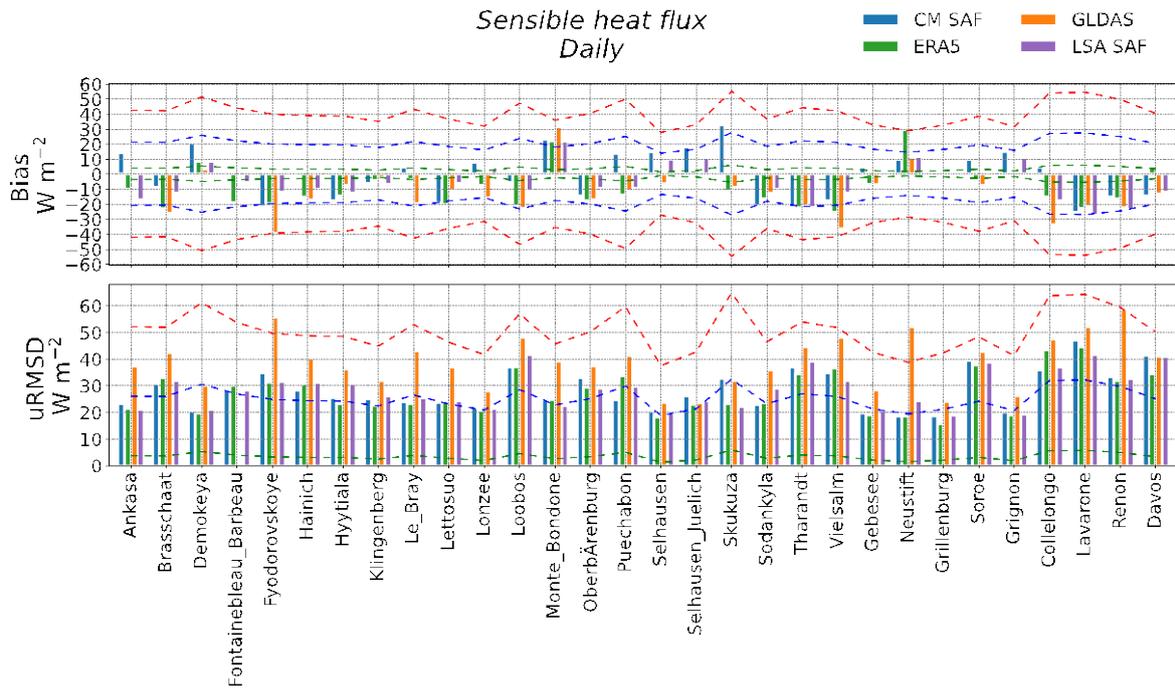


Figure 7: Same as Figure 4 but for sensible heat flux.

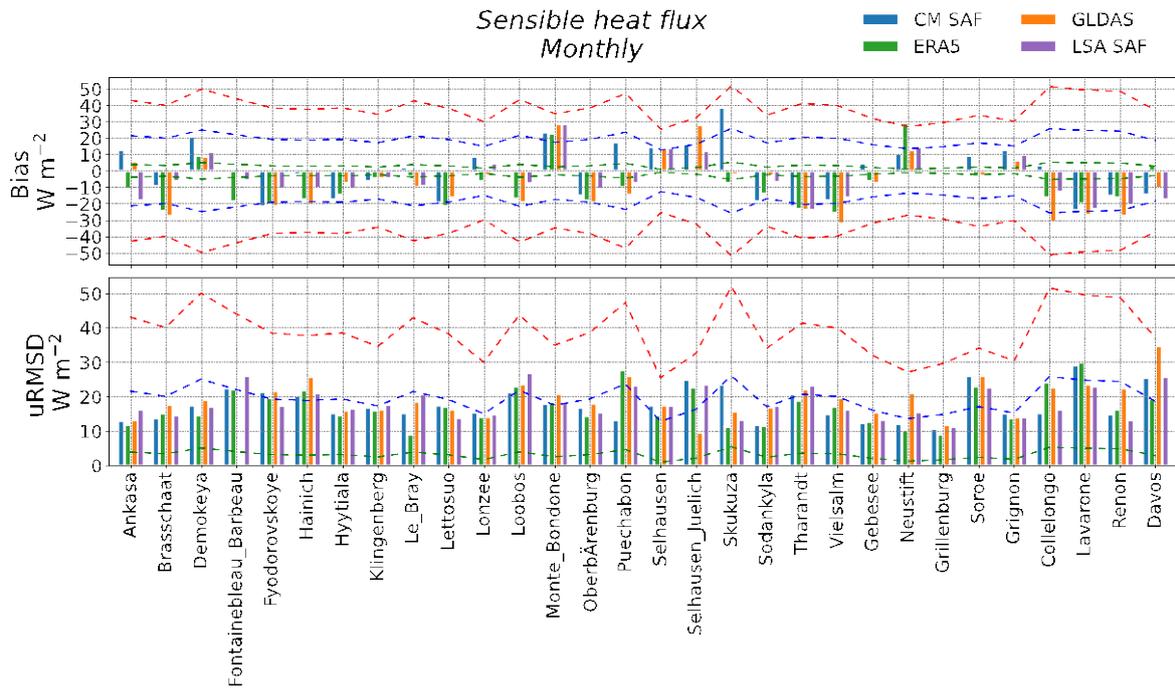


Figure 8: Same as Figure 5 but for sensible heat flux.

Table 4-6: Performance statistics of hourly, daily and monthly CM SAF sensible heat flux dataset as compared to FLUXNET2015/ICOS observations at 30 stations for different seasons and day/night conditions.

	N	< FLUXNET2015/ICOS > (W m ⁻²)	< CM SAF > (W m ⁻²)	Bias (W m ⁻²) ± (Opt / Tar / Thr)**	uRMSD (W m ⁻²) (Opt / Tar / Thr)**
<i>Hourly</i>					
DJF*	133541	-9.0	-10.7	-1.7 (2.5 / 17.5 / 34.9)	37.7 (2.5 / 57.5 / 114.9)
MAM*	88058	17.0	18.5	+1.5 (5.4 / 26.2 / 52.5)	57.7 (5.4 / 66.2 / 132.5)
JJA*	85010	26.4	25.8	-0.6 (5.9 / 27.7 / 55.4)	60.9 (5.9 / 67.69 / 135.4)
SON*	106625	-3.0	2.0	+4.9 (3.3 / 20 / 40)	39.9 (3.3 / 60 / 120)
Day	168175	46.1	53.4	+7.2 (6.4 / 29.3 / 58.6)	69.2 (6.4 / 69.3 / 138.6)
Night	226285	-22.5	-26.0	-3.5 (2.4 / 17.1 / 34.2)	25 (2.4 / 57.1 / 114.2)
Overall	413234	5.4	6.3	+1 (4 / 22.1 / 44.2)	48.5 (4 / 62.1 / 124.2)
<i>Daily</i>					
DJF*	25817	-2.6	-5.3	-2.7 (1.4 / 14.3 / 28.5)	24.1 (1.4 / 19.3 / 38.5)
MAM*	27222	41.2	35.7	-5.5 (4.7 / 24.2 / 48.4)	39 (4.7 / 29.2 / 58.4)
JJA*	28385	48.1	42.0	-6.1 (5.3 / 25.8 / 51.6)	42.9 (5.3 / 30.8 / 61.6)
SON*	28443	8.8	12.4	+3.6 (2.1 / 16.3 / 32.6)	24.5 (2.1 / 21.3 / 42.6)
Overall	109867	24.3	21.7	-2.6 (3.4 / 20.2 / 40.5)	34.1 (3.4 / 25.2 / 50.5)
<i>Monthly</i>					
DJF*	817	-3.9	-5.7	-1.8 (1.3 / 14 / 27.9)	19.4 (1.3 / 14 / 27.9)
MAM*	864	42.8	35.6	-7.2 (4.4 / 23.2 / 46.4)	25.6 (4.4 / 23.2 / 46.4)
JJA*	910	48.5	41.6	-6.9 (4.9 / 24.7 / 49.4)	28.3 (4.9 / 24.7 / 49.4)
SON*	921	6.8	11.5	+4.6 (1.7 / 15.1 / 30.1)	16.9 (1.7 / 15.1 / 30.1)
Overall	3512	24	21.2	-2.8 (3.1 / 19.3 / 38.6)	23.6 (3.1 / 19.3 / 38.6)

(*)DJF: December January February; MAM: March April May; JJA: June July August; SON: September October November.

(**)Opt: Optimal requirement; Tar: Target requirement; Thr: Threshold requirement.

	<p style="text-align: center;">Validation Report Meteosat Latent and Sensible heat fluxes - Edition 1</p>	Doc.No: SAF/CM/RMIB/VAL/MET/LEH Issue: 1.1 Date: 30.05.2023
---	---	---

3.6.3 Monthly mean diurnal cycle

The monthly mean diurnal cycle values have been evaluated by comparing CM SAF with FLUXNET2015/ICOS values. As noticed in the section 3.1, the quality control at FLUXNET2015/ICOS stations has been relaxed to medium (QC value of 1 or 2) due to the miss of match-up.

Figures 9 and 10 display the box plots of the distributions of bias, and uRMSDs calculated at all considered stations of the monthly mean diurnal cycle for the latent and the sensible heat flux. Note that for the validation, the hours are local time.

Bias and uRMSD follow, respectively an inverted bell curve and a bell curve. As observed previously, the latent heat flux is underestimated with mean bias values between -1.7 W m^{-2} at 6 p.m. and -10.9 W m^{-2} at 9 a.m.

Highest discrepancies are obtained from 7 a.m. and 5 p.m. (mainly day time) with most of the values in the target and the threshold regions for bias (around -7.8 W m^{-2}), and uRMSD (around 25.8 W m^{-2}), respectively. Between 5 p.m. and 7 a.m., values are close to the optimal requirement for the bias and in the target area (blue shaded area) for the uRMSD.

For the sensible heat flux (Figure 10), similar trends are observed for the uRMSD with third quartiles forming a bell-shaped curve around noon. Mean values are between 10.7 W m^{-2} (N=10) at 11 p.m. and 39 W m^{-2} (N=10) at 11 a.m. From 8 a.m. to 5 p.m., values are in the threshold requirement region ($\sim 31.2 \text{ W m}^{-2}$) while they are in the target ($\sim 13.1 \text{ W m}^{-2}$) area for the rest of the day. Concerning the bias, values follow a sinusoidal curve ($\sin(\text{Hour}/4)$) with an overestimation during the morning and underestimation in the afternoon. Interquartile range of values (full boxes) are between optimal and target requirement with mean values between -23 W m^{-2} (N=10, SD= 24 W m^{-2}) at 3 p.m. and 10 W m^{-2} (N=10, SD= 26 W m^{-2}) at 9 a.m.

Thus, while requirements are fulfilled, efforts will be made in the future to decrease uncertainty around noon where highest values are observed.

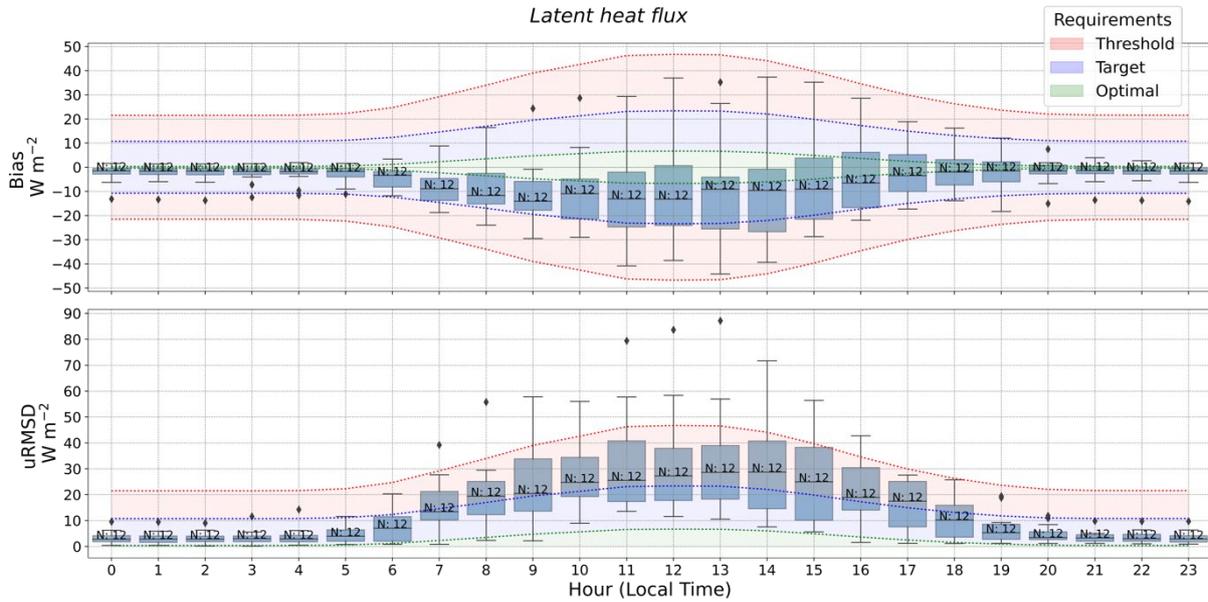


Figure 9: Box plots displaying the monthly mean diurnal cycle bias ($W m^{-2}$; top panel) and uRMSD ($W m^{-2}$; bottom panel) for the latent heat flux at FLUXNET2015/ICOS stations. The height of the box indicates the Interquartile range (IQR), the horizontal line inside the box indicates the median value, the upper box level indicates the upper quartile (75th percentile; Q_3), the lower box level indicates the lower quartile (25th percentile; Q_1), bars (whiskers) indicate the minimum and maximum values, dots indicate the outliers (higher than $Q_3 + 1.5 \cdot IQR$ or lower than $Q_1 - 1.5 \cdot IQR$). The shaded areas represent the optimal (green), target (blue) and threshold (red) requirements.

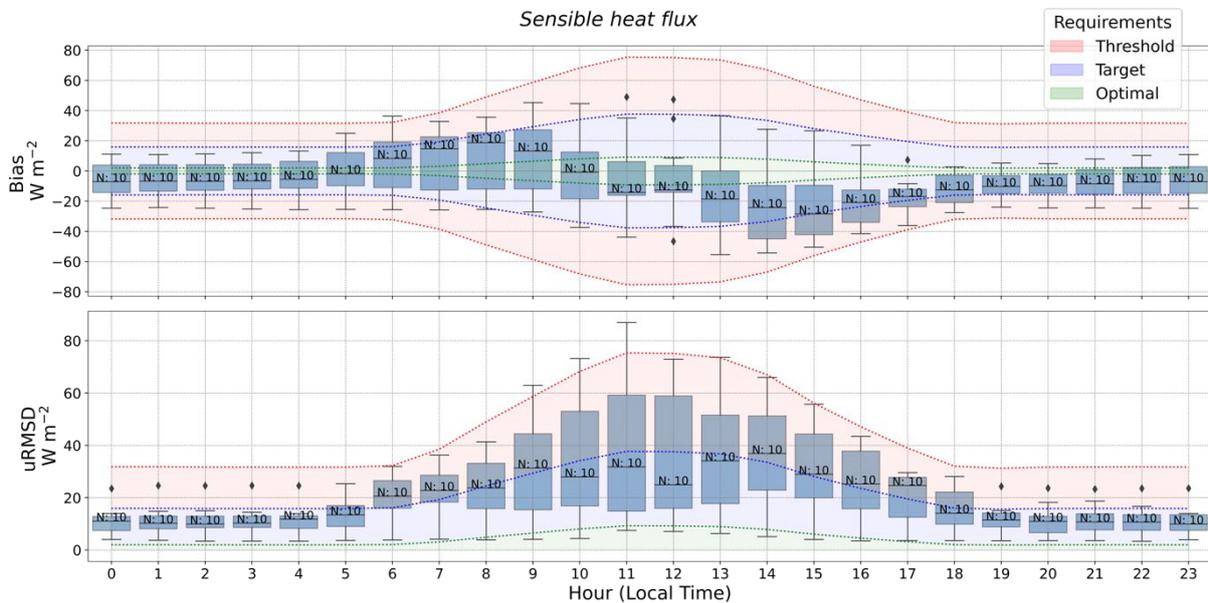


Figure 10: Same as Figure 8 but for the sensible heat flux.

	Validation Report Meteosat Latent and Sensible heat fluxes - Edition 1	Doc.No: SAF/CM/RMIB/VAL/MET/LEH Issue: 1.1 Date: 30.05.2023
---	--	---

3.7 Product inter-comparisons

Based on level 3 daily means, we carried out comparisons between the CM SAF datasets (LE, H and ET) and other reference datasets (ERA5, LSA SAF, GLDAS and GLEAM). As a reminder, CM SAF and LSA SAF have been spatially aggregated from 0.05° to 0.25° to match with ERA5, GLDAS and GLEAM datasets.

For each variable, two approaches have been used to assess the accuracy and the precision of the CM SAF products over the full disk.

- 1) Time series of the daily bias and uRMSD averaged over the full disk.
- 2) Images of the daily bias and uRMSD averaged over the year 2005.

For time series analysis, in order to avoid variation in the performance metric due to missing data, the minimum number of valid pixels has been determined as the first percentile of the corresponding record.

The following performance analysis and comparisons with requirements should be considered carefully. Comparisons performed thereafter involve biased assumptions that inter-compared products have perfect accuracy and precision.

3.7.1 Latent heat flux

Figures 11 and 12 present, respectively, daily bias and uRMSD values averaged over the disk of latent heat flux CM SAF data as compared to ERA5 (green dots), LSA SAF (violet dots) and GLDAS (orange dots) data. On average, the CM SAF dataset underestimates the latent heat flux as compared to other products. This bias is quite constant over the full period (no specific break) with an average value for the 3 product comparisons around -19 W m^{-2} (Table 4-3). Bias values are around the target requirement. Similarly, the uRMSD is quite constant for all datasets with values of 25.7 W m^{-2} , 25.6 W m^{-2} and 31.4 W m^{-2} for ERA5, LSA SAF and GLDAS datasets, respectively. Those values are mainly between the target and the threshold requirements (red areas in the Figures 11 and 12 and Table 4-3).

Grid based comparisons, carried out over the full disk for the year 2005 (MSG; Figure 13), show an underestimation over most of the globe (top panel in Figure 13). Bias is lower for latitudes higher than 16°N (e.g., average values of -11.8 W m^{-2} for ERA5 comparison) than southern latitudes (e.g., average value of -28.2 W m^{-2} for ERA5 comparison). Specific areas around the equator in South America and Africa have bias around 0. They both correspond to evergreen broadleaved trees (EBT) land cover cases (Figure 36). Around the EBT region in Africa, larger underestimations are observed for all products. For LSA SAF comparison, patches of high values correspond to pixels where the deciduous broadleaved trees land cover are used in the processing (Figure 36). The differences are most probably due to differences in land cover type used as input in LSA SAF dataset (Table 3-3 and Figure 38). For instance, according to the land cover, the minimal stomatal resistance would directly impact LE via the canopy resistance factor (see detail in [RD 1]).

Concerning uRMSD, lowest values (dark blue areas in the graph) are found for desert areas (latitudes between 16N° and 35N°) for all the products. For the rest of the map, uRMSD values calculated when comparing CM SAF with GLDAS retrievals differ from comparison results obtained with other datasets. Indeed, uRMSD, averaged over the map, obtained for GLDAS is about 50% higher than the uRMSD calculated for LSA SAF and ERA5 map. In addition, patches of values higher than 40 W m⁻² are only present for GLDAS comparison in South America and in the north-west of Africa regions.

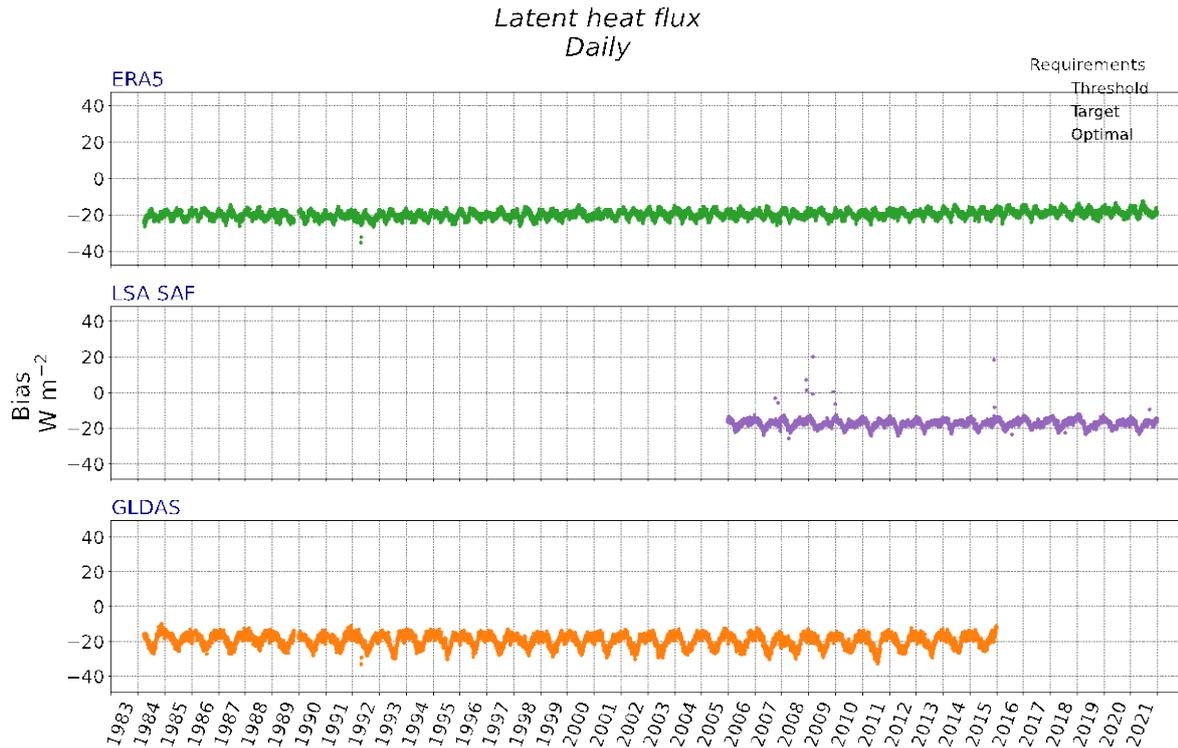


Figure 11: Time series of latent heat flux disk averaged daily bias (W m⁻²) of daily image of CM SAF dataset as compared to daily images of ERA5 (green, top), LSA SAF (violet, middle) and GLDAS (orange, bottom) datasets. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements.

Table 4-7: Performance statistics of daily CM SAF latent heat flux dataset as compared to ERA5, LSA SAF and GLDAS datasets averaged over the disk and the considered period.

	Covered period	N _{Day} *	N _{Pix} **	<CM SAF> (W m ⁻²)	<Ref product> (W m ⁻²)	Bias (W m ⁻²) ± (Opt / Tar / Thr)***	uRMSD (W m ⁻²) (Opt / Tar / Thr)***
ERA5	1983-2020	13214	84379	27.8	47.7	-20.0 (4.8 / 19.6 / 39.1)	25.7 (4.8 / 24.6 / 49.1)
LSA SAF	2005-2020	5708	83660	27.7	45.4	-17.7 (4.5 / 19.1 / 38.2)	25.6 (4.5 / 24.1 / 48.2)
GLDAS	1983-2005	11053	81885	27.3	47.2	-19.9 (4.7 / 19.4 / 38.9)	31.4 (4.7 / 24.4 / 48.9)

(*)N_{Day}: Number of daily images.

(**)N_{Pix}: Number of average pixel by daily image.

(***)Opt: Optimal requirement; Tar: Target requirement; Thr: Threshold requirement.

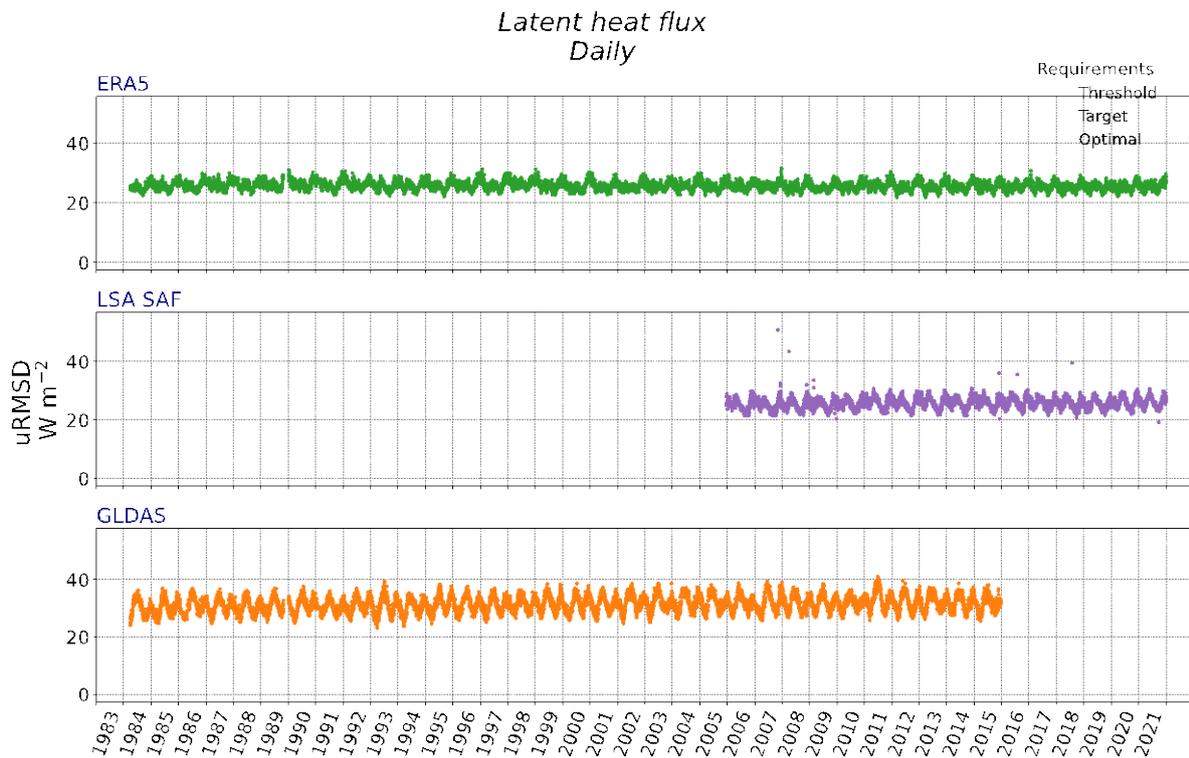


Figure 12: Time series of latent heat flux disk averaged daily unbiased root mean square error ($W\ m^{-2}$) of daily CM SAF dataset as compared to daily images of ERA5 (green, top), LSA SAF (violet, middle) and GLDAS (orange, bottom) datasets. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements.

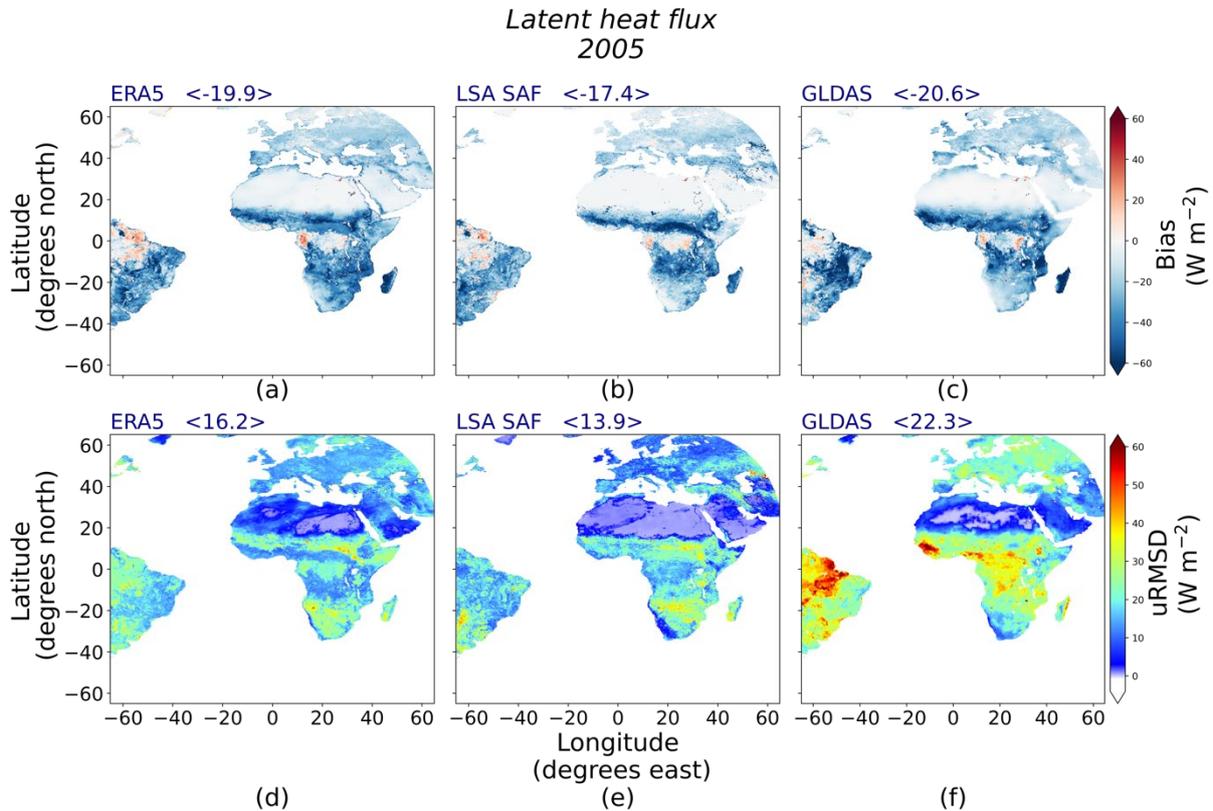


Figure 13: Maps of averaged bias (top) and unbiased root mean square difference (bottom) of the surface latent heat flux (W m^{-2}) over the year 2005 of CM SAF dataset as compared to ERA5 (a-d) LSA SAF (b-e) and GLDAS (c-f) datasets at 0.25° . Average value over the map is indicated in blue over each map.

3.7.2 Evapotranspiration

Figures 14 and 15 present, respectively, daily bias and uRMSD values averaged over the disk of evapotranspiration as compared to GLEAM, LSA SAF and GLDAS data. As expected, evapotranspiration and latent heat flux results are consistent. While the CM SAF dataset complies with the target requirement (except with respect to GLDAS), it always underestimates the ET with a bias of $-0.51 \text{ mm day}^{-1}$, $-0.63 \text{ mm day}^{-1}$ and -0.7 mm day^{-1} for, GLEAM, LSA SAF and GLDAS datasets respectively (red areas in the Figures 14 and 15 and Table 4-4). Better agreement is observed with GLEAM than with other datasets. Same observation can be done with the uRMSD where the target requirement is only reached when the GLEAM dataset is used as reference.

Grid-based analysis for the year 2005 (MSG; Figure 16) shows a global underestimation between $-0.52 \text{ mm day}^{-1}$ (GLEAM) and $-0.73 \text{ mm day}^{-1}$ (GLDAS). Map of GLEAM bias (Figure 16a) is more homogeneous as compared to bias obtained with LSA SAF and GLDAS datasets. Indeed, the interquartile is of 0.65 mm day^{-1} for the GLEAM dataset while, interquartiles of 0.92 mm day^{-1} and 0.85 mm day^{-1} are observed for LSA SAF and GLDAS datasets, respectively. At the opposite, as observed for LE, larger negative bias patches are present for southern regions

(latitude lower than 16°N) except for two evergreen broadleaved trees land cover regions in South America and West Africa. In addition, same patch of low values is observed for deciduous broadleaved trees land cover.

Concerning uRMSD, lowest values (dark blues color) are found for desert areas (latitudes between 16°N and 35°N) for all the products. For the rest of the map, uRMSD values calculated when comparing CM SAF with GLDAS retrievals differ from comparison results obtained with other datasets. Indeed, the map averaged uRMSD of GLDAS is about 50% higher than LSA SAF and ERA5 values. In addition, patches with values around 1.5 mm day⁻¹ in South America and in the north-west of Africa regions are only present in GLDAS comparison.

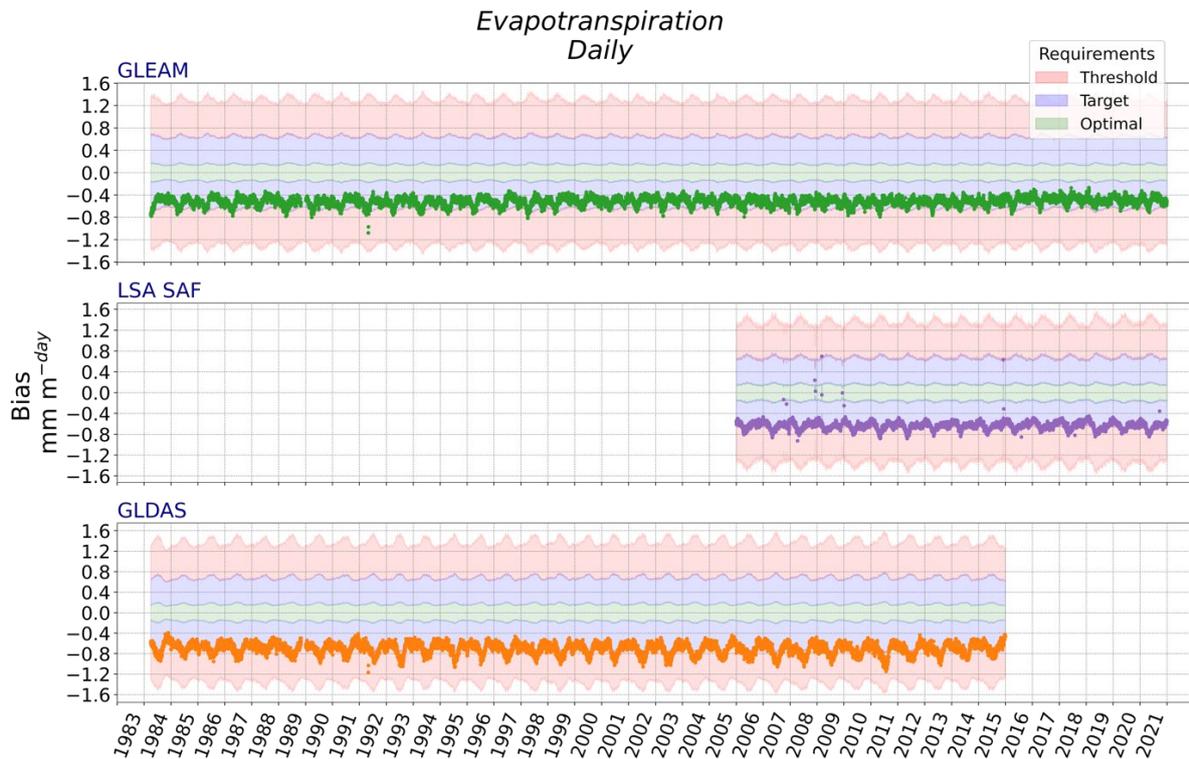


Figure 14: Time series of disk averaged daily evapotranspiration bias (mm day⁻¹) of daily images of daily CM SAF dataset as compared to GLEAM (green, top), LSA SAF (violet, middle), and GLDAS (orange, bottom) datasets. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements.

Table 4-8: Performance statistics of daily CM SAF Evapotranspiration dataset as compared to GLEAM, LSA SAF and GLDAS datasets averaged over the disk and the covered period.

	Covered period	N _{Day} *	N _{Pix} **	<CM SAF> (mm day ⁻¹)	<Ref Product> (mm day ⁻¹)	Bias (mm day ⁻¹) ± (Opt / Tar / Thr)**	uRMSD (mm day ⁻¹) (Opt / Tar / Thr)***
GLEAM	1983-2020	13214	83041	0.98	1.49	-0.51 (0.15 / 0.65 / 1.31)	0.83 (0.15 / 0.83 / 1.66)
LSA SAF	2005-2020	5708	83660	0.98	1.61	-0.63 (0.16 / 0.67 / 1.36)	0.91 (0.16 / 0.85 / 1.71)
GLDAS	1983-2005	11053	81885	0.97	1.67	-0.7 (0.17 / 0.68 / 1.38)	1.1 (0.17 / 0.86 / 1.73)

(*)N_{Day}: Number of daily images.

(**) N_{Pix} : Number of average pixel by daily image.

(***) Opt: Optimal requirement; Tar: Target requirement; Thr: Threshold requirement.

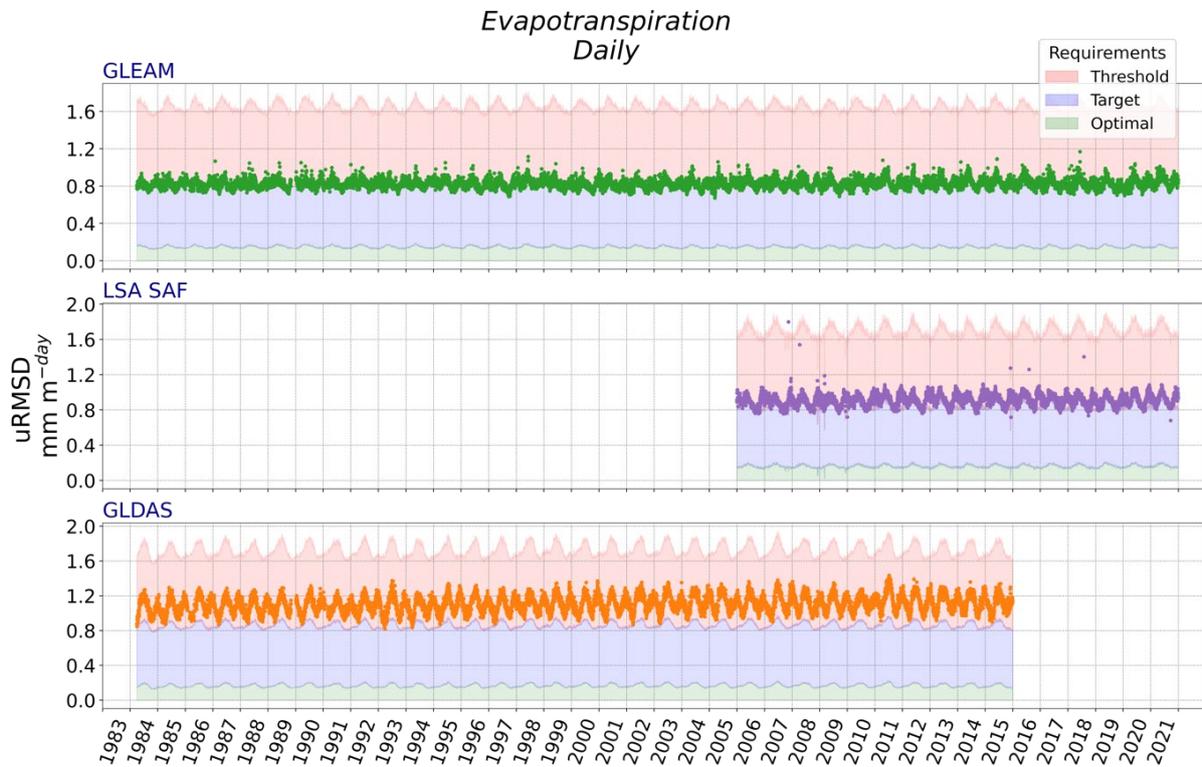


Figure 15: Time series of disk averaged daily evapotranspiration unbiased root mean square error (mm day^{-1}) of daily CM SAF dataset as compared to daily images of GLEAM (green, top), LSA SAF (violet, middle), the GLDAS (orange, bottom) datasets. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements.

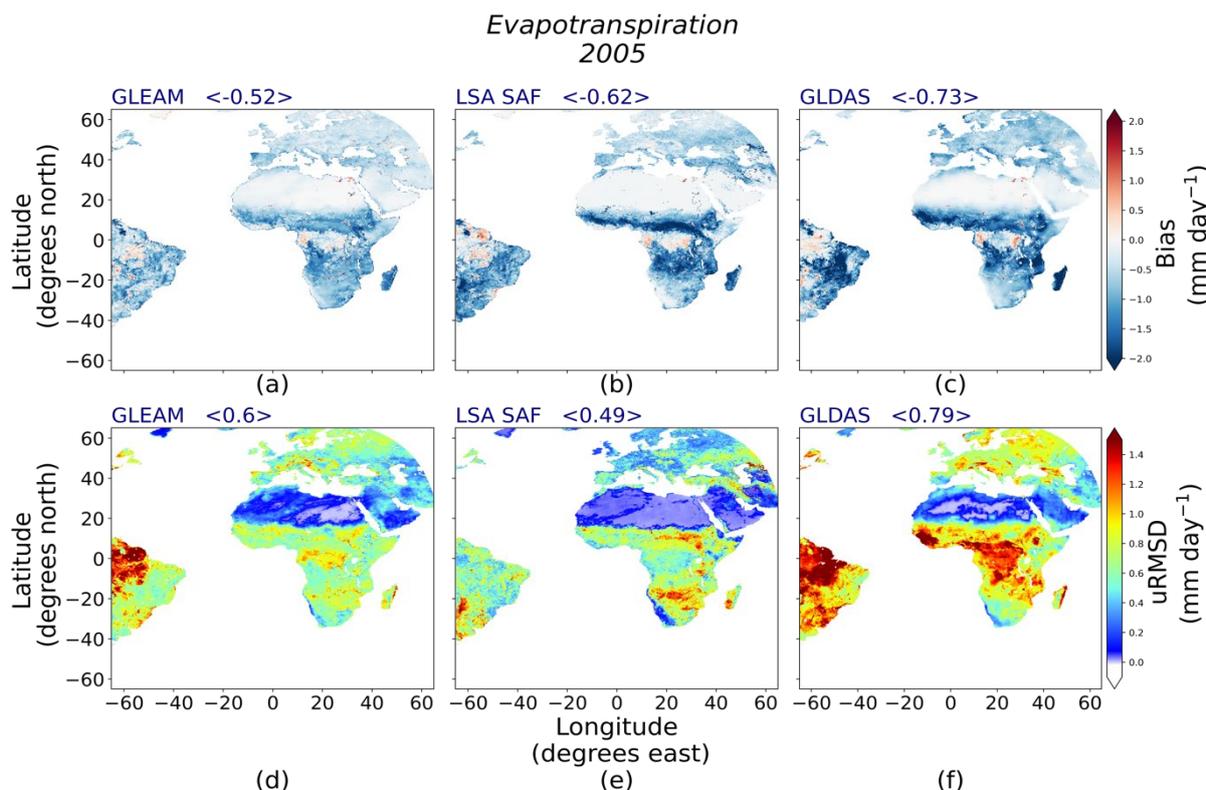


Figure 16: Maps of averaged daily bias (top) and unbiased root mean square difference (bottom) of the evapotranspiration (mm day^{-1}) over the year 2005 of CM SAF dataset as compared to GLEAM (a-d), LSA SAF (b-e) and GLDAS (c-f) datasets at 0.25° . Average value over the map is indicated in blue over each map.

3.7.3 Sensible heat flux

Figures 17 and 18 present daily bias and uRMSD averaged over the disk of CM SAF sensible heat flux data as compared to ERA5, LSA SAF and GLDAS data. Contrary to the latent heat flux, the sensible heat flux is slightly overestimated as compared to other product retrievals. Bias values comply with the target requirements with average values, over the considered periods, between 6.2 W m^{-2} and 7 W m^{-2} (Table 4-5). Distinction can be made between MFG and MSG periods with lower and more stable bias using MSG-based product. Indeed, bias values decrease from 9.8 W m^{-2} ($N=7449$, $SD = 2.7 \text{ W m}^{-2}$) to 3.2 W m^{-2} ($N=5765$, $SD = 2 \text{ W m}^{-2}$) for ERA5 comparison and from 7.7 W m^{-2} ($N=7449$, $SD = 3.4 \text{ W m}^{-2}$) to 3.2 W m^{-2} ($N=3604$, $SD = 2.8 \text{ W m}^{-2}$) for GLDAS comparison (see section 4.3 for specific analysis of the stability). On the contrary, the uRMSD is quite constant for all cases with values between of 24.6 W m^{-2} ($N=13214$, $SD = 1.8 \text{ W m}^{-2}$), 22.7 W m^{-2} ($N=5708$, $SD = 1.4 \text{ W m}^{-2}$) and 32.4 W m^{-2} ($N=11053$, $SD = 1.8 \text{ W m}^{-2}$) for ERA5, LSA SAF and GLDAS datasets, respectively (Table 4-5). These values comply with the target requirements for accuracy and precision, except the uRMSD calculated with respect to the GLDAS dataset which only complies with the threshold requirement.

Grid based comparisons, carried out for the year 2005 (MSG) over the full disk (Figure 19), show average bias of 3.1 W m^{-2} , 6.1 W m^{-2} and 3.0 W m^{-2} for ERA5, LSA SAF and GLDAS

datasets, respectively. Similarly to LE analysis (section 4.2.1), distinction can be made between values obtained for latitude higher and lower than 16°N. For instance, for ERA5 comparison, an average bias value of -1 W m⁻² is calculated for latitudes higher than 16°N while it is of 13 W m⁻² for lower latitudes. In those southern regions, two specific areas around the equator in South America and Africa tend to have negative bias. They both correspond to evergreen broadleaved trees land cover cases.

Concerning uRMSD, as observed for LE, lowest values around 10 W m⁻² (dark blues color) are found for desert areas (latitudes between 16°N and 35°N). Localizations of higher value patches are similar for all products. However, absolute uRMSD values calculated when comparing CM SAF with GLDAS retrievals differ from comparison results obtained with other datasets. Indeed, the average uRMSD is 17.8 W m⁻² and 13.3 W m⁻² for ERA5 and LSA SAF datasets and 24.3 W m⁻² for GLDAS.

Table 4-9: Performance statistics of daily CM SAF sensible heat flux dataset as compared to ERA5, LSA SAF and GLDAS datasets averaged over the disk and the covered period.

	Covered period	N _{Day} (*)	N _{Pix} (**)	<CM SAF> (W m ⁻²)	<Ref product> (W m ⁻²)	Bias (W m ⁻²) ± (Opt / Tar / Thr)***	uRMSD (W m ⁻²) (Opt / Tar / Thr)***
ERA5	1983-2020	13214	84379	46.2	39.3	6.9 (4.2 / 22.5 / 45.0)	24.6 (4.2 / 27.5 / 55.0)
LSA SAF	2005-2020	5708	83660	43.7	36.7	7.0 (3.9 / 21.8 / 43.5)	22.7 (3.9 / 26.8 / 53.5)
GLDAS	1983-2005	11053	81885	47.1	40.8	6.2 (4.7 / 24.0 / 48.0)	32.4 (4.7 / 29.0 / 58.0)

(*)N_{Day}: Number of daily images.

(**)N_{Pix}: Number of average pixel by daily image.

(***)Opt: Optimal requirement; Tar: Target requirement; Thr: Threshold requirement.

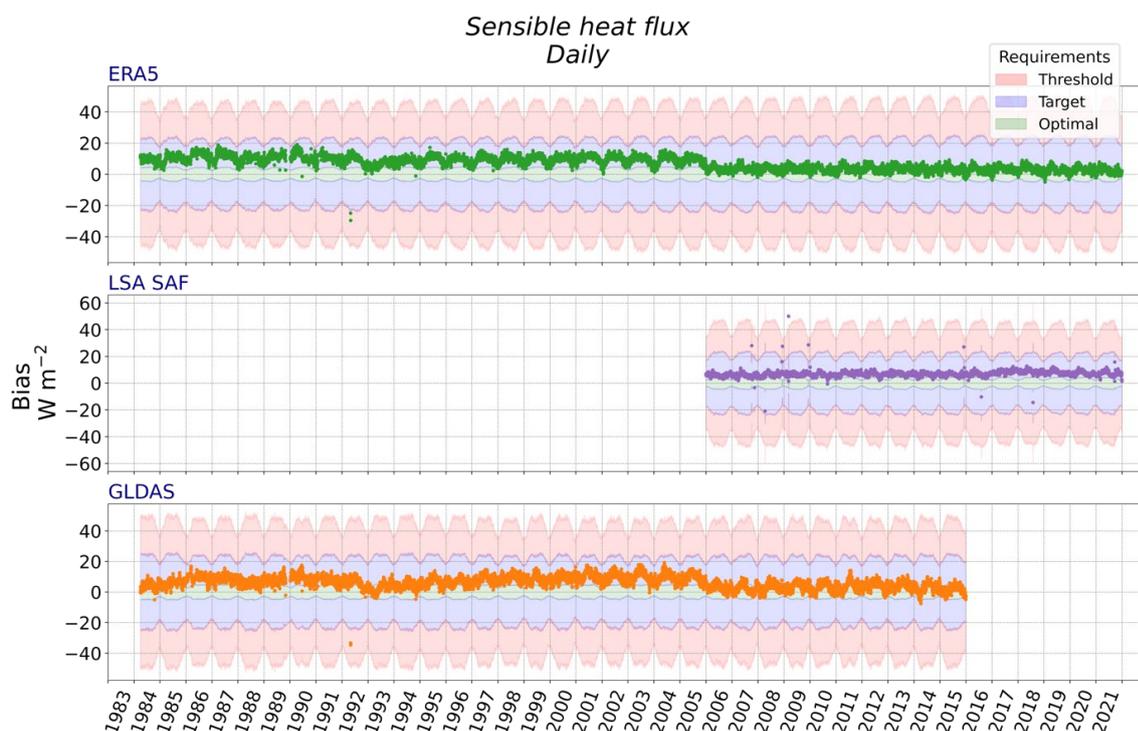


Figure 17: Same as Figure 11 but for sensible heat flux.

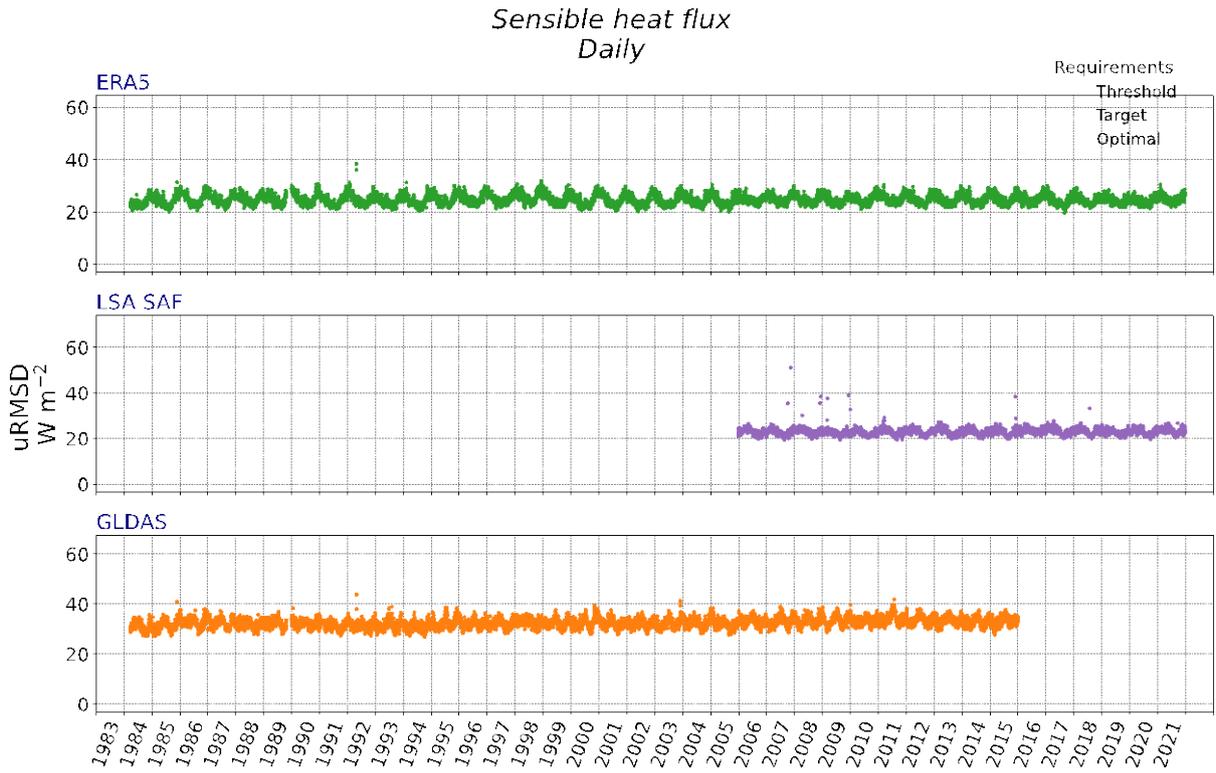


Figure 18: Same as Figure 12 but for sensible heat flux.

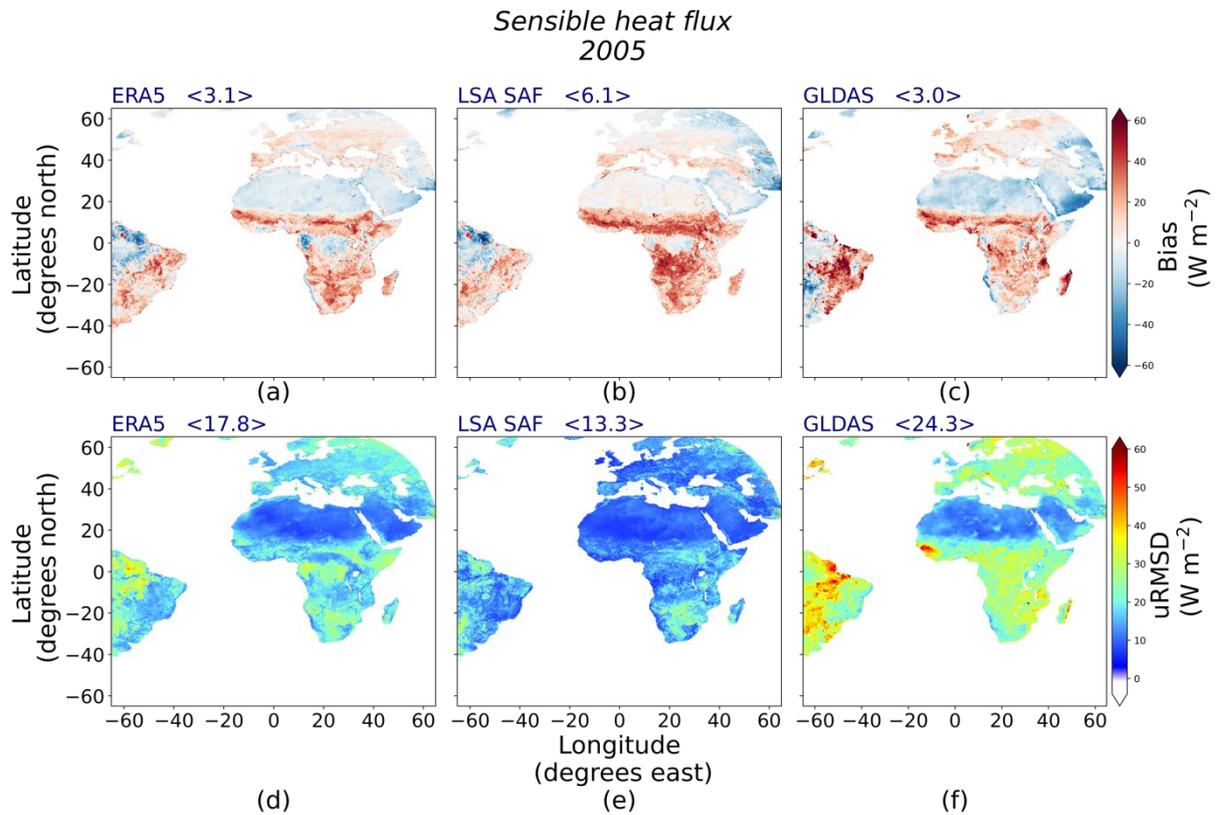


Figure 19: Same as Figure 13 but for sensible heat flux.

3.8 Stability

The Table 3-2 reminds the requirements for decadal stability which are of $0.1 \text{ W m}^{-2} \text{ dec}^{-1}$, $4 \text{ W m}^{-2} \text{ dec}^{-1}$ and $8 \text{ W m}^{-2} \text{ dec}^{-1}$ for optimal, target and threshold, respectively. To assess the stability, we calculated the monthly mean deseasonalized bias, averaged over the full disk, of the latent and sensible heat flux between CM SAF as compared to other products.

In this part, we assumed a temperature of 20°C to convert ET (mm month^{-1}) in LE (W m^{-2}) in order to add the GLEAM dataset in the Latent heat flux stability analysis.

3.8.1 Latent heat flux

Figure 20 shows the monthly mean deseasonalized bias³, averaged over the full disk, between CM SAF and ERA5, LSA SAF, GLDAS and GLEAM latent heat flux values. Over the whole period, the mean deseasonalized bias values are of -19.9 W m^{-2} ($N=450$, $SD=1 \text{ W m}^{-2}$), -17.7 W m^{-2} ($N=190$, $SD=0.7 \text{ W m}^{-2}$), -19.8 W m^{-2} ($N=379$, $SD=1.2 \text{ W m}^{-2}$), -14.6 W m^{-2} ($N=450$, $SD=1 \text{ W m}^{-2}$) and for ERA5, LSA SAF, GLDAS and GLEAM comparison, respectively. Trend values are between $-0.47 \text{ W m}^{-2} \text{ dec}^{-1}$ (GLDAS) and $0.564 \text{ W m}^{-2} \text{ dec}^{-1}$ (ERA5). Slope for ERA5 is higher from 2010 while higher values is observed for GLDAS before 1990.

Excellent stability is observed with LSA SAF (no significant trend) and GLEAM ($0.208 \text{ W m}^{-2} \text{ dec}^{-1}$). No specific break is observed at the transition of each satellite (Table 4-6). The maximum difference between mean values calculated over the time period covered by two successive instruments is 1.1 W m^{-2} . Decadal stability fulfills the target decadal stability requirements for the latent heat flux.

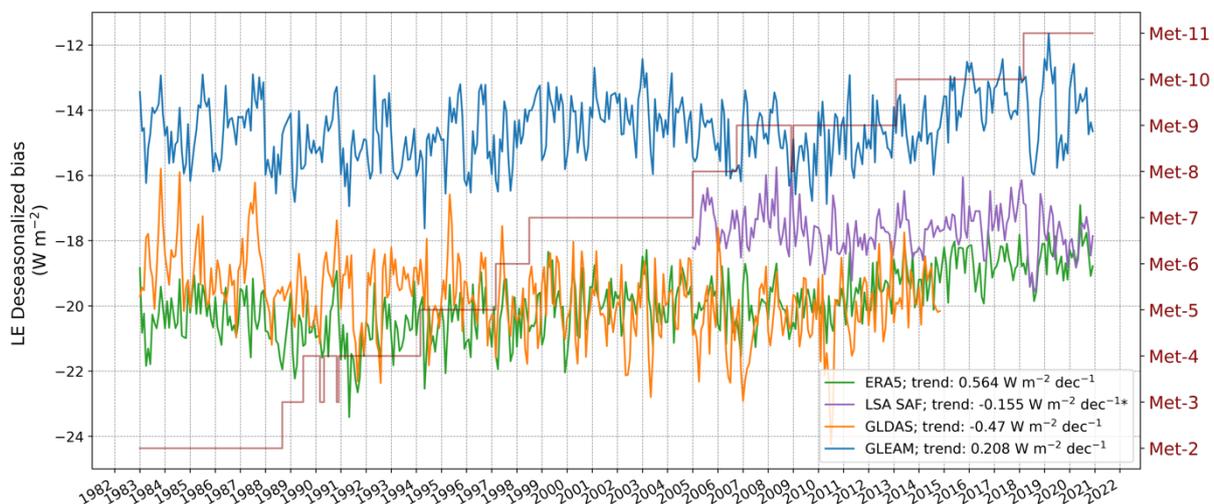


Figure 20: Time series of deseasonalized bias (W m^{-2}) of monthly global latent heat flux compared to ERA5 (green), LSA SAF (violet), GLDAS (orange) and GLEAM (blue). Right axis indicates which Meteosat satellite is used (dark-red line). Trends ($\text{W m}^{-2} \text{ dec}^{-1}$) are indicated in the legend and an (*) is added is the slope to calculate the trend is not significant ($p\text{-value} > 0.05$).

³ The deseasonalization is done by the removal of the corresponding monthly climatology.

Table 4-10: Deseasonalized bias ($W m^{-2}$) of the CM SAF monthly latent heat flux as compared with ERA5, LSA SAF, GLDAS and GLEAM averaged over the full disk and over the time period covered by each satellite.

Satellites	ERA5	LSA SAF	GLDAS	GLEAM
	$W m^{-2}$ (N_{Mth} , CV %)*			
Meteosat-2 (MFG-2)	-20.3 (67, -3)	-	-18.8 (67, -6)	-14.6 (67, -6)
Meteosat-3 (MFG-3)	-21 (11, -5)	-	-19.5 (11, -5)	-15.1 (11, -7)
Meteosat-4 (MFG-4)	-20.9 (52, -4)	-	-19.6 (52, -5)	-15.1 (52, -5)
Meteosat-5 (MFG-5)	-20.5 (36, -4)	-	-19.6 (36, -5)	-14.9 (36, -7)
Meteosat-6 (MFG-6)	-20.5 (16, -4)	-	-20.2 (16, -5)	-15 (16, -6)
Meteosat-7 (MFG-7)	-19.9 (78, -4)	-	-20 (78, -5)	-14.2 (78, -5)
Meteosat-8 (MSG-1)	-20 (22, -3)	-17.4 (22, -4)	-20.5 (22, -6)	-15 (22, -5)
Meteosat-9 (MSG-2)	-19.9 (75, -3)	-17.7 (75, -4)	-20.4 (75, -6)	-14.9 (75, -6)
Meteosat-10 (MSG-3)	-18.8 (59, -3)	-17.5 (59, -3)	-19.5 (22, -4)	-14 (59, -6)
Meteosat-11 (MSG-4)	-18.6 (34, -3)	-18 (34, -4)	-	-14 (34, -8)

* N_{Mth} : Number of months; CV: Coefficient of variation

3.8.2 Sensible heat flux

Figure 21 displays the monthly mean deseasonalized bias, averaged over the full disk, between CM SAF and ERA5, GLDAS and LSA SAF sensible heat flux. While trends comply with the target decadal stability requirement with values between $-2.65 W m^{-2} dec^{-1}$ for ERA5 comparison and $1.22 W m^{-2} dec^{-1}$ for LSA SAF comparison, it seems clear that instabilities higher than $4 W m^{-2} dec^{-1}$ (target requirement value) occur while analyzing variations according to time period covered by each instrument. Values estimated over the MFG period are higher and less stable than estimations over the MSG period (Figure 21 and Table 4-7). For instance, for comparison with ERA5, which covers the longest period, the mean value is of $9.7 W m^{-2}$ ($N=260$, $SD=1.7 W m^{-2}$) for MFG period while it is of $3.2 W m^{-2}$ ($N=190$, $SD=1.1 W m^{-2}$) for MSG period.

The maximum difference between mean values calculated over time period covered by two successive instruments is of $3.8 W m^{-2}$ during MFG period (between Met-3 and Met-4) and $0.5 W m^{-2}$ for MSG period (Table 4-7). In addition, a decrease of $6.4 W m^{-2}$ at MFG/MSG transition (between Met-7 and Met-8) is observed.

Stability clearly differ according to Europe, North Africa, South Africa and South America regions (Figures 22 and 23). First, over the full period, deseasonalized bias is higher in the South Africa region (on average $18 W m^{-2}$) than the rest of the world (average around $6 W m^{-2}$). Second, an important distinction can be made between MFG and MSG for all regions except

Europe. Indeed, difference between averaged values over MFG and MSG period is of 9.5 W m^{-2} , 5.6 W m^{-2} , 7.4 W m^{-2} over North Africa, South Africa and South America regions, respectively, while it is of 2 W m^{-2} over Europe. Specific pattern can be noticed during Met-2 over the South America region (Figure 22) with peaks of high values ($> 20 \text{ W m}^{-2}$) as compared to mean value (12.7 W m^{-2} , SD of 3.4 W m^{-2}). Finally, considering MFG and MSG period separately, values for Africa regions are more stable than values obtained for the Europe and South America regions. For MFG period, the standard deviation is on average 2.2 W.m^{-2} for Africa regions while it is 2.3 W.m^{-2} for Euro/SAmE. Similarly, for the MSG period, the standard deviation is on average 1.9 W.m^{-2} for Africa regions while it is 2.8 W.m^{-2} for Europe and South America.

Thus, if we consider the full disk, the sensible heat flux only complies the decadal stability threshold requirement but, instability differ according the region and the target requirement is complied over Europe. A specific analyze for the transition period is made in the following section.

The variations in the sensible heat flux are mainly due to instability in the surface incoming shortwave and the surface albedo which are, respectively, negatively and positively correlated to the sensible heat flux. Similar breaks are observed in the monthly mean deseasonalized bias as compared to ERA5 for those two variables (Figures 24 and 25). In addition, analysis performed over different regions showed similar trends than the ones observed for the sensible heat flux (results not show here).

Variations between MFG and MSG periods can be related to the narrow band to broad band relation used to derive variables over sand areas (see [RD 2]). This latter would explain why the latent heat flux is not impacted as almost all energy goes to the sensible heat flux in such arid regions. Of course, sensor degradations and/or inter-calibration can still impact the results.

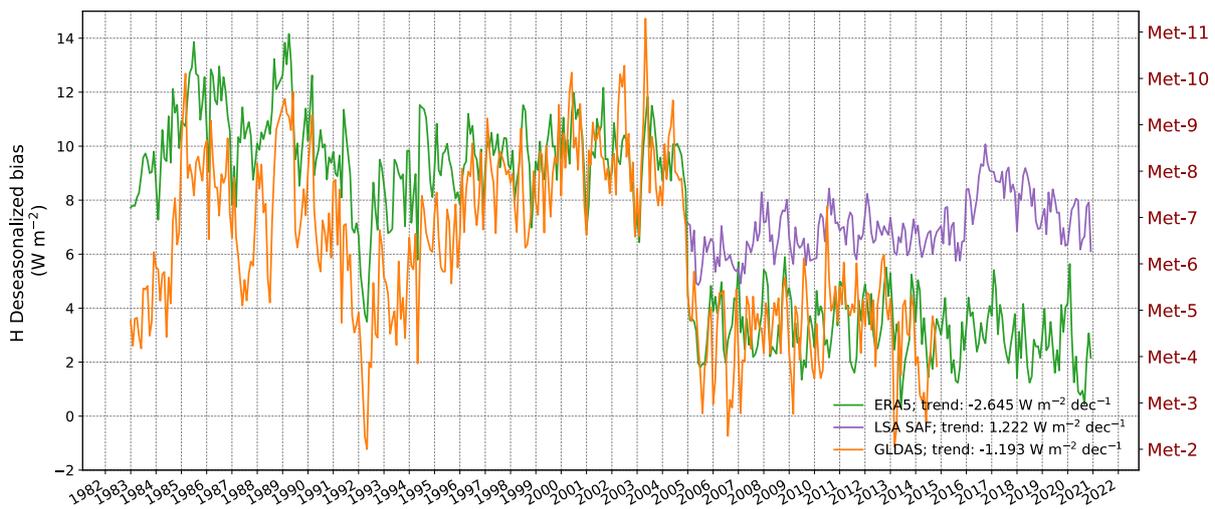


Figure 21: Same as Figure 20 but for sensible heat flux.

Table 4-11: Same as Table 4-6 but for sensible heat flux (W m^{-2}).

Satellites	ERA5	LSA SAF	GLDAS
	W m^{-2} (N_{Mth} , CV %)*	W m^{-2} (N_{Mth} , CV %)*	W m^{-2} (N_{Mth} , CV %)*
Meteosat-2 (MFG-2)	10.4 (67, 15)	-	6.7 (67, 35)
Meteosat-3 (MFG-3)	12.2 (11, 13)	-	10.4 (11, 15)
Meteosat-4 (MFG-4)	8.4 (52, 22)	-	5 (52, 46)
Meteosat-5 (MFG-5)	9.4 (36, 13)	-	7.2 (36, 24)
Meteosat-6 (MFG-6)	9.7 (16, 7)	-	8.9 (16, 11)
Meteosat-7 (MFG-7)	9.7 (78, 12)	-	9.4 (78, 18)
Meteosat-8 (MSG-1)	3.3 (22, 30)	6.2 (22, 13)	2.8 (22, 59)
Meteosat-9 (MSG-2)	3.5 (75, 31)	6.7 (75, 11)	3.6 (75, 40)
Meteosat-10 (MSG-3)	3 (59, 35)	7.5 (59, 16)	2.2 (22, 72)
Meteosat-11 (MSG-4)	2.5 (34, 45)	7.5 (34, 11)	-

* N_{Mth} : Number of months; CV: Coefficient of variation

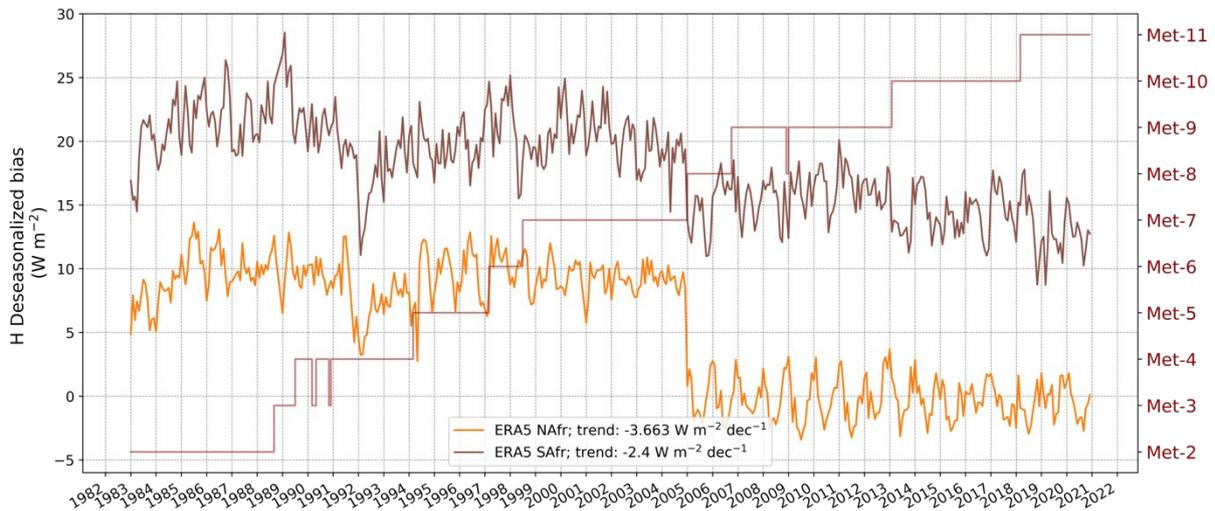


Figure 22: Time series of deseasonalized bias (W m^{-2}) of the monthly sensible heat flux compared to ERA5 averaged over North Africa (yellow line) and South Africa (brown line) regions. Right axis indicates which Meteosat satellite is used (dark-red line). Trends values ($\text{W m}^{-2} \text{dec}^{-1}$) are indicated in the legend.

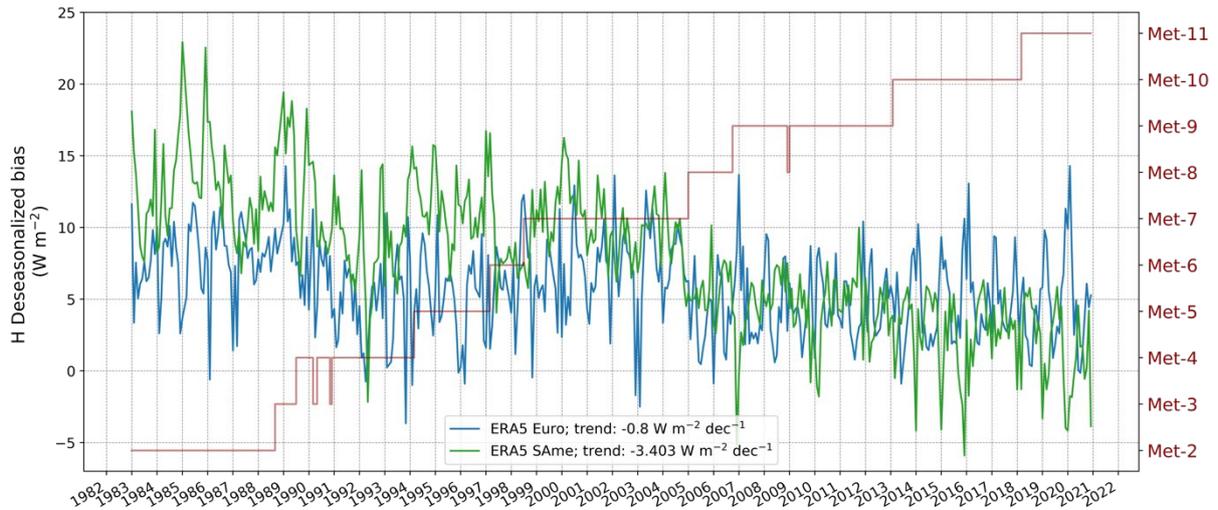


Figure 23: Time series of deseasonalized bias ($W m^{-2}$) of the monthly sensible heat flux compared to ERA5 averaged over Europe (blue) and South America (green) regions. Right axis indicates which Meteosat satellite is used (dark-red). Trends values ($W m^{-2} dec^{-1}$) are indicated in the legend.

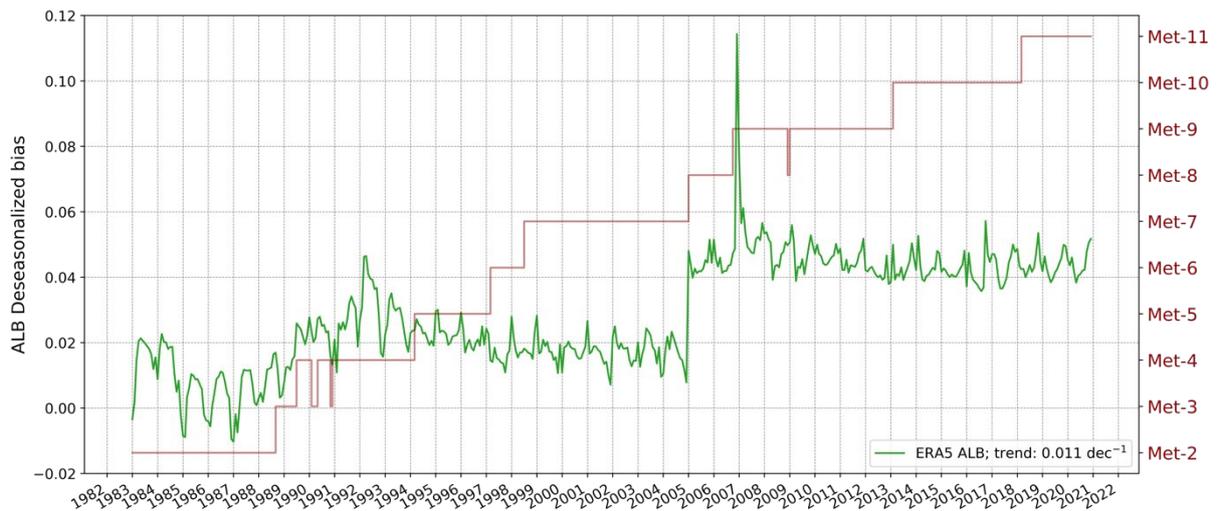


Figure 24: Time series of deseasonalized bias ($W m^{-2}$) of the CM SAF monthly albedo as compared to ERA5. Right axis indicates which Meteosat satellite is used (dark-red line). The trend values ($W m^{-2} dec^{-1}$) of the time series is indicated in the legend.

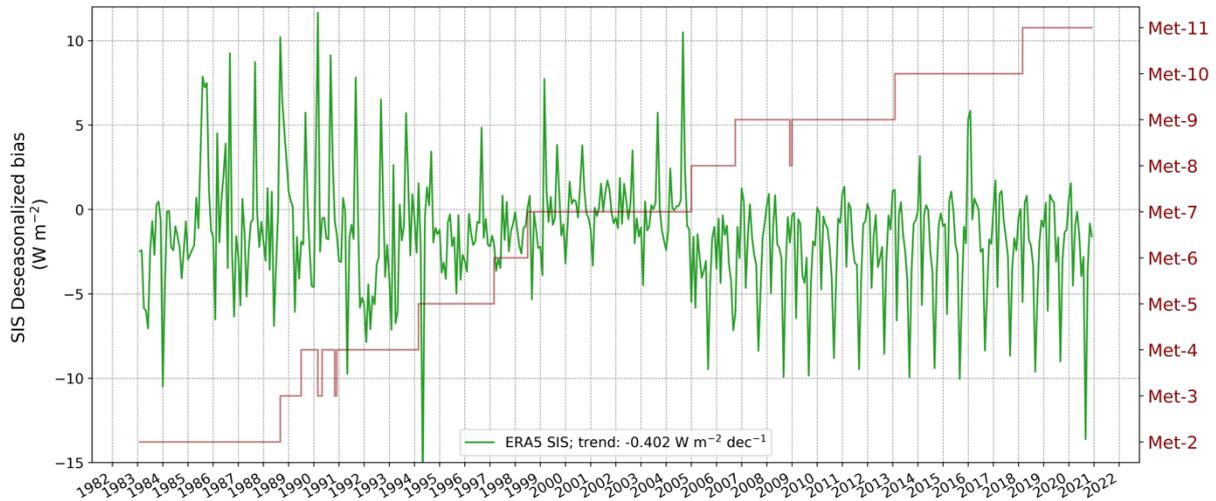


Figure 25: Same as Figure 24 but for the surface incoming shortwave radiation.

3.9 Comparison of Meteosat First and Second Generation

As showed in the previous section, results from MVIRI and SEVIRI-based inputs seem to differ. Thus, in the following subsections, differences for the year 2005 from MVIRI and SEVIRI-based inputs have been analyzed.

It is worth noting that “Skukuza” station has been removed from analysis in the following subsections as 2005 is an exceptional year with extreme/doubtful *in situ* values (higher than 800 W m^{-2}). Results at this station would not be relevant for the purpose of this section. However, specific analyze of the data for this station can be found in the section 7.2.

3.9.1 Surface latent heat flux

Performance metrics of estimations from MVIRI (MFG) and MSG-based inputs (named CM SAF -MFG and CM SAF -MSG) as compared to *in situ* FLUXNET2015/ICOS data have been calculated for hourly, daily and monthly timestep for surface latent heat flux (Figures 26 to 28). No significative difference is observed between CM SAF -MFG and CM SAF -MSG results at stations. On average, MAD between CM SAF -MFG and CM SAF -MSG results as compared to Fluxnet data is of 1.5 W m^{-2} (N=68, SD= 2 W m^{-2}) and of 0.9 W m^{-2} (N=68, SD= 1.1 W m^{-2}) for bias and uRMSD metrics, respectively.

Grid based comparisons of yearly and monthly mean MVIRI- and SEVIRI-based inputs surface latent heat flux is carried out for different seasons of the year 2005 at native resolution 0.05° (Figure 29). Yearly mean comparisons show average mean bias and uRMSD values of 0.5 W m^{-2} and 2 W m^{-2} over the disk. CM SAF -MFG values are slightly higher (around 2 W m^{-2}) in South America, central West coast of Africa and East coast of Madagascar. Similar patterns are only observed in summer time (Figure 29f). Note that patches of similar values are present in Europe. At the opposite, South America region is the most impacted by the source on inputs in Winter with mix of over- and underestimation.

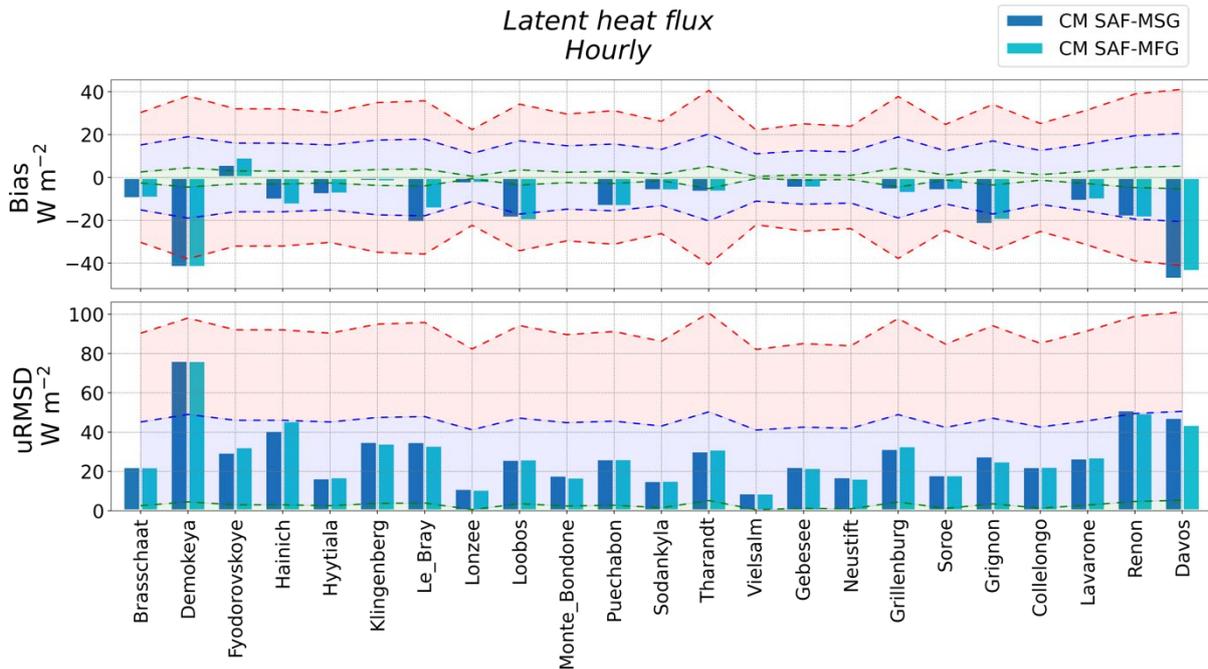


Figure 26: Bar plot showing the hourly bias ($W m^{-2}$; top) and the hourly unbiased root mean square error ($W m^{-2}$; bottom) of CM SAF-MSG (blue) and CM SAF-MFG (light-blue) datasets as compared to FLUXNET2015/ICOS dataset at stations for the latent heat flux for the year 2005. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements.

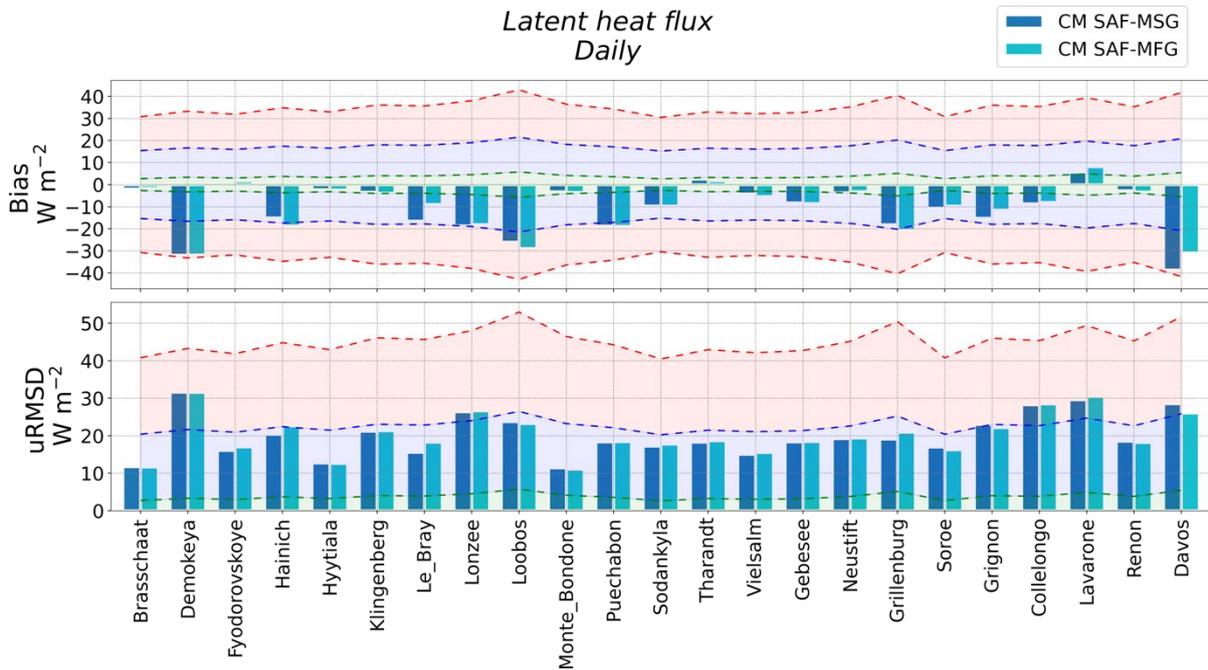


Figure 27: Bar plot showing the daily bias ($W m^{-2}$; top) and the daily unbiased root mean square error ($W m^{-2}$; bottom) of CM SAF-MSG (blue) and CM SAF-MFG (light-blue) datasets as compared to FLUXNET2015/ICOS dataset at stations for the latent heat flux for the year 2005. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements.

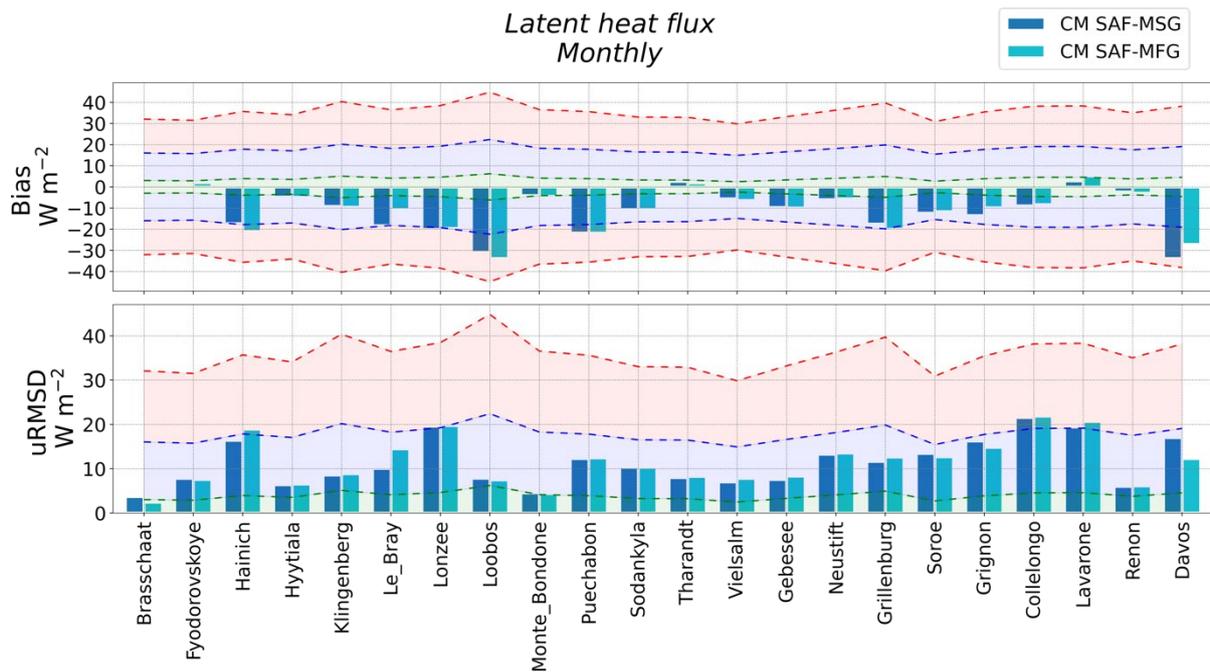


Figure 28: Bar plot showing the monthly bias ($W m^{-2}$; top) and the monthly unbiased root mean square error ($W m^{-2}$; bottom) of CM SAF-MSG (blue) and CM SAF-MFG (light-blue) datasets as compared to FLUXNET2015/ICOS dataset at stations for the latent heat flux for the year 2005. The shaded areas represent the optimal (green), target (blue) and threshold (red) accuracy requirements.

Latent heat flux
 2005

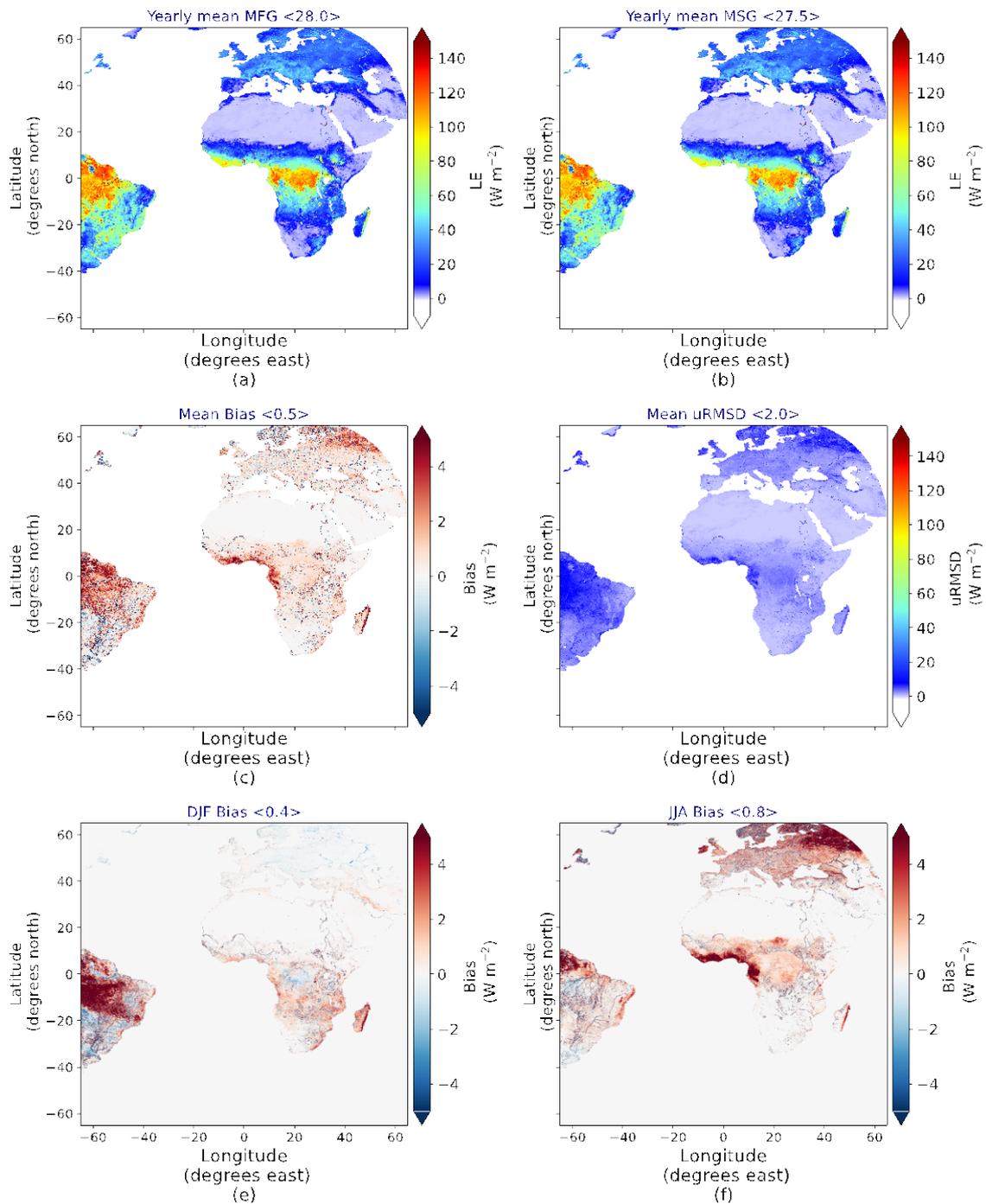


Figure 29: Maps (0.05°) of daily latent heat flux ($W m^{-2}$) averaged over 2005 from MFG (a) and MSG (b); (c) difference between MFG and MSG based products; (d) unbiased root mean square difference between MFG and MSG based products ($W m^{-2}$); (e) bias ($W m^{-2}$) between MFG and MSG based products in winter (e) and summer (f) season. Average value over the map is indicated in blue over each map.

3.9.2 Surface sensible heat flux

Performance metrics of estimations from MFG and MSG-based inputs as compared to *in situ* FLUXNET2015/ICOS data have been calculated for hourly, daily and monthly timestep at station for surface sensible heat flux (Figures 30 to 32). As observed for the latent heat flux, no significant difference is observed in between CM SAF -MFG and CM SAF -MSG results at stations. On average, CM SAF -MFG and CM SAF -MSG results as compared to Fluxnet data are similar.

Grid based comparison of mean MVIRI- and SEVIRI-based surface sensible heat flux is carried out for the year 2005 (Figure 33). It shows higher differences in the results than for the latent heat flux. On average, the difference is of 5.7 W m^{-2} and the uRMSD is of 5.4 W m^{-2} . As observed in the previous section (section 4.3.2), sand areas are subject to highest differences between results based on MFG and MSG inputs (Figure 33 c,e and f). This confirms the hypothesis of an issue in the narrow band to broad-band conversion over sand areas which would impact inputs (SIS and ALB) and thus the sensible heat flux. As mentioned previously, the latent heat flux is less impacted over very dry areas. The CM SAF -MFG and CM SAF -MSG difference is slightly lower in winter than in summer with bias of 3.8 W m^{-2} and 7.1 W m^{-2} , respectively.

The differences between MFG and MSG-based products would deserve further investigations in particular over sandy areas and central-East Europe.

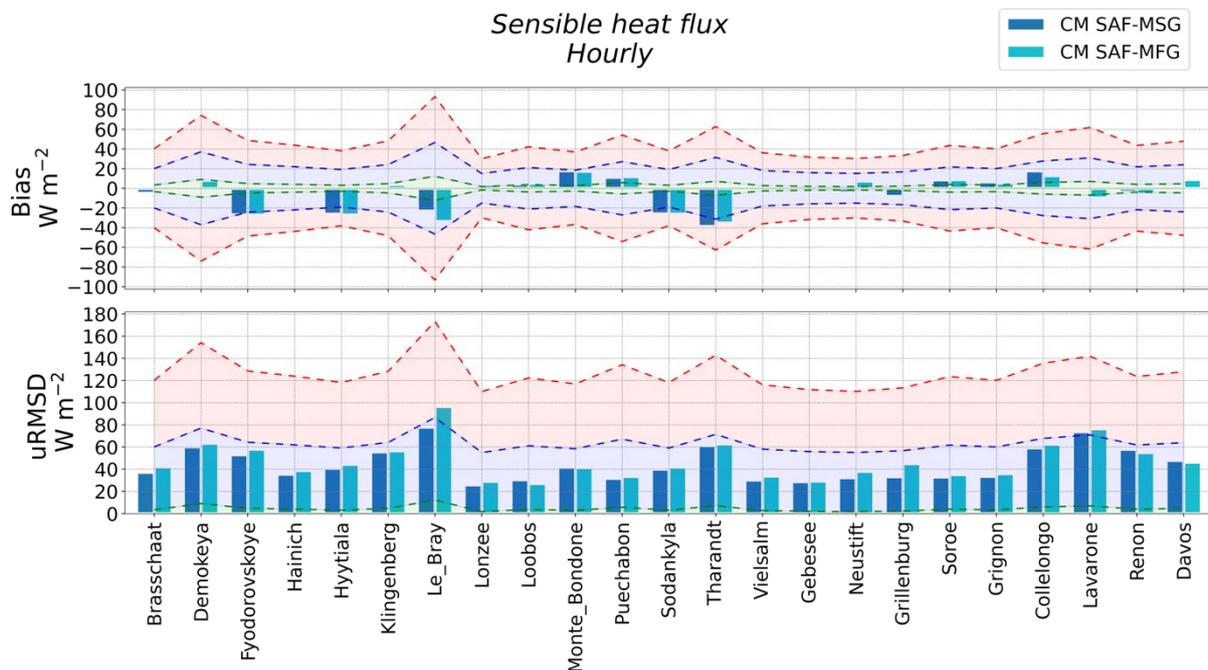


Figure 30: Same as Figure 26 but for sensible heat flux.

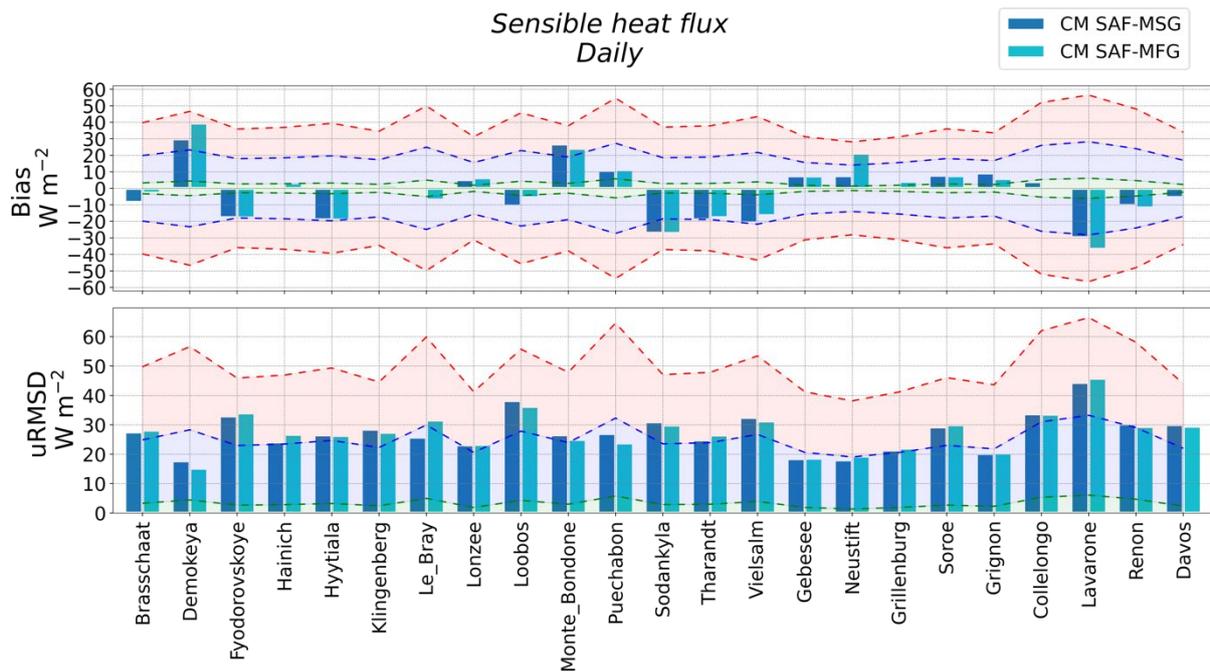


Figure 31: Same as Figure 27 but for sensible heat flux.

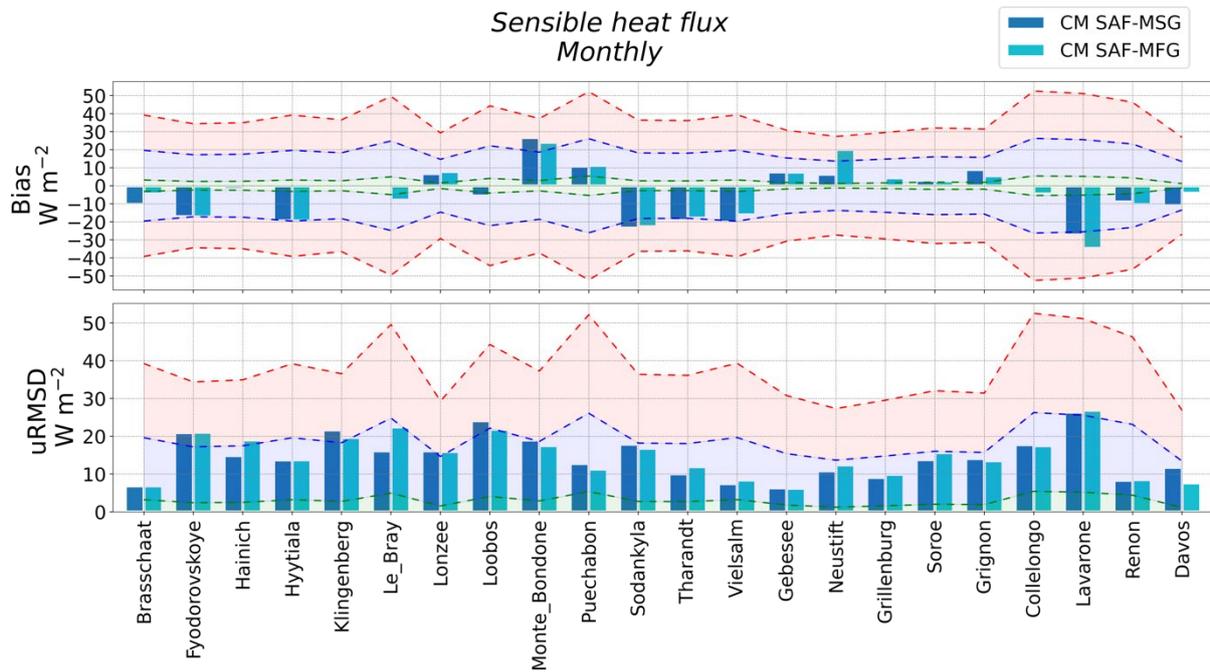


Figure 32: Same as Figure 28 but for sensible heat flux.

Sensible heat flux
 2005

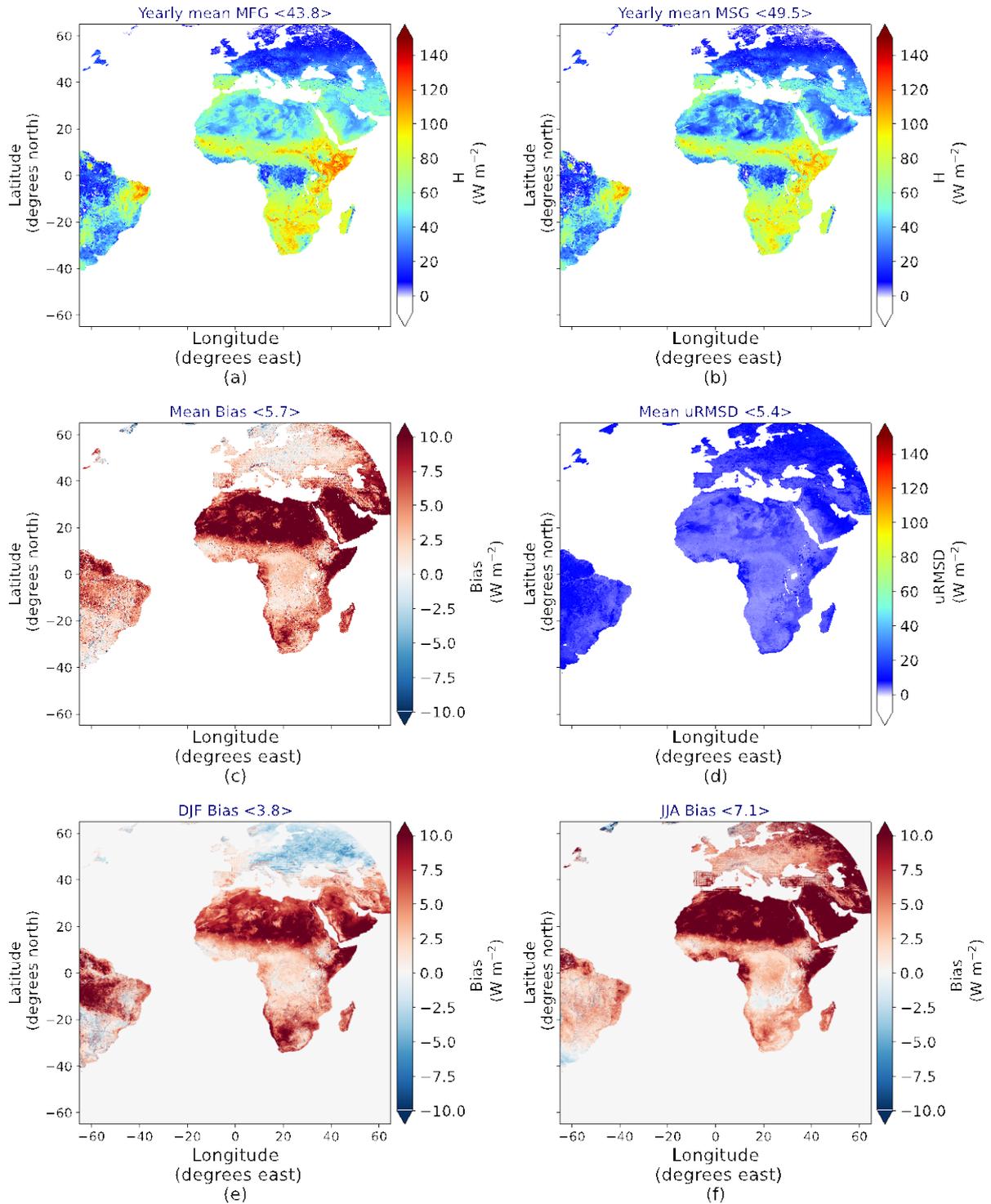


Figure 33: Same as Figure 29 but for sensible heat flux.

	<p style="text-align: center;">Validation Report Meteosat Latent and Sensible heat fluxes - Edition 1</p>	Doc.No: SAF/CM/RMIB/VAL/MET/LEH Issue: 1.1 Date: 30.05.2023
---	---	---

4 Concluding remarks

This report presents the validation of the CM SAF latent and sensible heat flux climate data record at hourly, daily, monthly scale and the monthly diurnal cycle. To this end, we used reference ground observations from: FLUXNET2015 and ICOS networks. The dataset has been compared with other products: ERA5, GLDAS, LSA SAF and GLEAM. and state-of-the-art literature. Furthermore, inter-comparisons allowed an analysis of the stability over long time period. Finally, MFG/MSG transition have been studied at 0.05° by estimating LE and H based on MFG/MSG data for the year 2005.

All together, accuracy, precision and stability target/threshold requirements are met. The target uncertainty requirements for LE and H are not always reached for daily and monthly timestep at stations but similar uncertainties values are observed in the literature or by comparing *in-situ* values with other datasets.

The following issues will need further attention:

- Comparisons with *in-situ* observations at stations and with other products reveal an underestimation of the latent heat flux. Source of soil moisture, adaptation of the minimum stomatal resistance parameter and/or source surface land cover map, could explain this bias.
- Comparison at stations for the latent heat flux showed better performance for low value cases e.g. winter season and during the night. Further investigation would be led to improve the repartition of energy between LE, H and G for high values cases.
- Temporal variation of the bias is in particular during the MFG period and at the MFG/MSG transition as well. Those variations are much small while considering Europe area exclusively. A likely source of error at the transition could be the narrow-band to broad-band conversion over sand area. It can be also however related land cover map which does not vary before 1992 (for instability during the MFG period), to sensor degradation and/or intercalibration issues.

With this report we tried to describe positive features as well as limitations for each variable. We emphasize that LE/H CDR is the longest high temporal (hourly) and high spatial (0.05°) resolution dataset with 38 years length. Users are encouraged to report back findings on applicability, stability and potential deficiencies to CM SAF (email: cmsaf.contact@dwd.de).

5 References

Albergel, C., Balsamo, G., De Rosnay, P., Muñoz-Sabater, J., & Boussetta, S. (2012). A bare ground evaporation revision in the ECMWF land-surface scheme: evaluation of its impact using ground soil moisture and satellite microwave data. *Hydrology and Earth System Sciences*, 16(10), 3607-3620, ISO 690; doi:10.5194/hess-16-3607-2012.

Albergel, C., Munier, S., Bocher, A., Bonan, B., Zheng, Y., Draper, C., ... & Calvet, J. C. (2018). LDAS-Monde sequential assimilation of satellite derived observations applied to the contiguous US: An ERA5 driven reanalysis of the land surface variables. *Remote Sensing*, 10(10), 1627.

Armstrong, R. L., Brodzik, M. J., Knowles, K., & Savoie, M. (2005). Global monthly EASE-Grid snow water equivalent climatology. *Boulder, CO: National Snow and Ice Data Center, Digital media*.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., & Betts, A. K. (2009). A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *Journal of hydrometeorology*, 10(3), 623-643.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., ... & Vitart, F. (2015). ERA-Interim/Land: a global land surface reanalysis data set. *Hydrology and Earth System Sciences*, 19(1), 389-407.

Beck, H. E., Van Dijk, A. I., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & De Roo, A. (2017). MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, 21(1), 589-615.

Behrendt, A., Wulfmeyer, V., Senff, C., Muppa, S. K., Späth, F., Lange, D., ... & Wieser, A. (2020). Observation of sensible and latent heat flux profiles with lidar. *Atmospheric Measurement Techniques*, 13(6), 3221-3233.

Bontemps, S., Herold, M., Kooistra, L., Van Groenestijn, A., Hartley, A., Arino, O., Moreau, I., & Defourny, P. (2012). Revisiting land cover observation to address the needs of the climate modeling community. *Biogeosciences*, 9(6), doi:10.5194/bg-9-2145-2012

Bowen, I. S. (1926). The ratio of heat losses by conduction and by evaporation from any water surface. *Physical review*, 27(6), 779.

Chávez, J. L., Howell, T. A., & Copeland, K. S. (2009). Evaluating eddy covariance cotton ET measurements in an advective environment with large weighing lysimeters. *Irrigation Science*, 28(1), 35-50.

Chehbouni, A., Kerr, Y. H., Watts, C., Hartogensis, O., Goodrich, D., Scott, R., ... & De Bruin, H. A. R. (1999). Estimation of area-average sensible heat flux using a large-aperture

scintillometer during the Semi-Arid Land-Surface-Atmosphere (SALSA) Experiment. *Water Resources Research*, 35(8), 2505-2511.

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H. L., Koren, V., ... & Betts, A. (1996). Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research: Atmospheres*, 101(D3), 7251-7268.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., ... & Haas, E. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, 203, 185-215, doi:10.1016/j.rse.2017.07.001

Drexler, J. Z., Snyder, R. L., Spano, D., & Paw U, K. T. (2004). A review of models and micrometeorological methods used to estimate wetland evapotranspiration. *Hydrological Processes*, 18(11), 2071-2101.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., ... & Tarpley, J. D. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, 108(D22).

ESA. Land Cover CCI Product User Guide Version 2. Tech. Rep. (2017). Available at: maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., ... & Kilic, A. (2017). The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. *Water Resources Research*, 53(4), 2618-2626.

Foken, T. (2008). The energy balance closure problem: an overview. *Ecological Applications*, 18(6), 1351-1367.

Franssen, H. H., Stöckli, R., Lehner, I., Rotenberg, E., & Seneviratne, S. I. (2010). Energy balance closure of eddy-covariance data: A multisite analysis for European FLUXNET stations. *Agricultural and Forest Meteorology*, 150(12), 1553-1567.

Gellens-Meulenberghs, F., Arboleda, A., & Ghilain, N. (2006, March). Status of development of the LSA-SAF evapotranspiration product. In Proc. 2nd LSA-SAF Training Workshop (pp. 8-10).

Gellens-Meulenberghs, F., Arboleda, A., & Ghilain, N. (2007). Towards a continuous monitoring of evapotranspiration based on MSG data. *IAHS PUBLICATION*, 316, 228.

Ghilain, N., Arboleda, A., Sepulcre-Canto, G., Batelaan, O., Ardo, J., & Gellens-Meulenberghs, F. (2012). Improving evapotranspiration in a land surface model using biophysical variables derived from MSG/SEVIRI satellite, doi:10.5194/hess-16-2567-2012

Gruber, A., Dorigo, W. A., Crow, W., & Wagner, W. (2017). Triple collocation-based merging of satellite soil moisture retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12), 6780-6792.

Guo, X., Meng, D., Chen, X., & Li, X. (2022). Validation and Comparison of Seven Land Surface Evapotranspiration Products in the Haihe River Basin, China. *Remote Sensing*, 14(17), 4308.

Hansen, M., & Song, X. (2018). Vegetation Continuous Fields (VCF) Yearly Global 0.05 Deg [Data set], NASA EOSDIS Land Processes DAAC. doi: 10.5067/MEaSURES/VCF/VCF5KYR.001

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Dee, D., Horányi, A., ... & Vamborg, F. (2019, January). The ERA5 Global Atmospheric Reanalysis at ECMWF as a comprehensive dataset for climate data homogenization, climate variability, trends and extremes. In *Geophysical Research Abstracts* (Vol. 21).

Hirschi, M., Michel, D., Lehner, I., & Seneviratne, S. I. (2017). A site-level comparison of lysimeter and eddy covariance flux measurements of evapotranspiration. *Hydrology and Earth System Sciences*, 21(3), 1809-1825.

Katul, G. G., Oren, R., Manzoni, S., Higgins, C., & Parlange, M. B. (2012). Evapotranspiration: a process driving mass transport and energy exchange in the soil-plant-atmosphere-climate system. *Reviews of Geophysics*, 50(3).

Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Houser, P. R., Geiger, J., Olden, S., ... & Sheffield, J. (2006). Land information system: An interoperable framework for high resolution land surface modeling. *Environmental Modelling & Software*, 21(10), 1402-1415.

Liou, Y. A., & Kar, S. K. (2014). Evapotranspiration estimation with remote sensing and various surface energy balance algorithms—A review. *Energies*, 7(5), 2821-2849.

Liu, Y., Liu, R., & Chen, J. M. (2012). Retrospective retrieval of long-term consistent global leaf area index (1981–2011) from combined AVHRR and MODIS data. *Journal of Geophysical Research: Biogeosciences*, 117(G4), doi:10.1029/2012JG002084

Liu, Y., Liu, R., Pisek, J., & Chen, J. M. (2017). Separating overstory and understory leaf area indices for global needleleaf and deciduous broadleaf forests by fusion of MODIS and MISR data. *Biogeosciences*, 14(5), 1093, doi:10.5194/bg-14-1093-2017

Loew, A., Peng, J., & Borsche, M. (2016). High-resolution land surface fluxes from satellite and reanalysis data (HOLAPS v1. 0): evaluation and uncertainty assessment. *Geoscientific Model Development*, 9(7), 2499-2532.

Luojus, K., Pullianinen, J., Takala, M., Lemmetyinen, J., Kangwa, M., Smolander, T., ... & Pinnock, S. (2013). ESA GlobSnow: Algorithm Theoretical Basis Document-SWE-algorithm.

Martens, B., Miralles, D.G., Lievens, H., van der Schalie, R., de Jeu, R.A.M., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A., & Verhoest, N.E.C. (2017). GLEAM v3: satellite-based land evaporation and root-zone soil moisture, *Geoscientific Model Development*, 10, 1903–1925.

Martens, B., Schumacher, D. L., Wouters, H., Muñoz-Sabater, J., Verhoest, N. E., & Miralles, D. G. (2020). Evaluating the land-surface energy partitioning in ERA5. *Geoscientific Model Development*, 13(9), 4159-4181.

McJannet, D. L., Cook, F. J., McGloin, R. P., McGowan, H. A., & Burn, S. (2011). Estimation of evaporation and sensible heat flux from open water using a large-aperture scintillometer. *Water Resources Research*, 47(5).

Michel, D., Jiménez, C., Miralles, D. G., Jung, M., Hirschi, M., Ershadi, A., ... & Seneviratne, S. I. (2016). The WACMOS-ET project—Part 1: Tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms. *Hydrology and Earth System Sciences*, 20(2), 803-822.

Miralles, D. G., Gash, J. H., Holmes, T. R., de Jeu, R. A., & Dolman, A. J. (2010). Global canopy interception from satellite observations. *Journal of Geophysical Research: Atmospheres*, 115(D16).

Miralles, D.G., Holmes, T.R.H., de Jeu, R.A.M., Gash, J.H., Meesters, A.G.C.A., Dolman, A.J. (2011). Global land-surface evaporation estimated from satellite-based observations, *Hydrology and Earth System Sciences*, 15, 453–469.

Mito, C. O., Boiyo, R. K., & Laneve, G. (2012). A simple algorithm to estimate sensible heat flux from remotely sensed MODIS data. *International journal of remote sensing*, 33(19), 6109-6121.

Moesinger, L., Dorigo, W., de Jeu, R., van der Schalie, R., Scanlon, T., Teubner, I., & Forkel, M. (2020). The global long-term microwave vegetation optical depth climate archive (VODCA). *Earth System Science Data*, 12(1), 177-196.

Oki, T., & Kanae, S. (2006). Global hydrological cycles and world water resources. *science*, 313(5790), 1068-1072.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y. W., ... & Law, B. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific data*, 7(1), 1-27. <https://doi.org/10.1038/s41597-020-0534-3>

Peng, J., Kharbouche, S., Muller, J. P., Danne, O., Blessing, S., Giering, R., ... & Dadson, S. (2020). Influences of leaf area index and albedo on estimating energy fluxes with HOLAPS framework. *Journal of Hydrology*, 580, 124245.

Perez-Priego, O., El-Madany, T. S., Migliavacca, M., Kowalski, A. S., Jung, M., Carrara, A., ... & Reichstein, M. (2017). Evaluation of eddy covariance latent heat fluxes with independent lysimeter and sapflow estimates in a Mediterranean savannah ecosystem. *Agricultural and forest meteorology*, 236, 87-99.

Pipunic, R. C., Walker, J. P., & Western, A. (2008). Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: A comparative synthetic study. *Remote Sensing of Environment*, 112(4), 1295-1305.

Priestley, C. H. B., & Taylor, R. J. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly weather review*, 100(2), 81-92.

Rebmann, C., Göckede, M., Foken, T., Aubinet, M., Aurela, M., Berbigier, P., ... & Yakir, D. (2005). Quality analysis applied on eddy covariance measurements at complex forest sites using footprint modelling. *Theoretical and Applied Climatology*, 80(2), 121-141.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., ... & Valentini, R. (2005). On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global change biology*, 11(9), 1424-1439.

Rodell, M., Houser, P. R., Jambor, U. E. A., Gottschalck, J., Mitchell, K., Meng, C. J., ... & Toll, D. (2004). The global land data assimilation system. *Bulletin of the American Meteorological society*, 85(3), 381-394.

Sellers, P. J., Dickinson, R. E., Randall, D. A., Betts, A. K., Hall, F. G., Berry, J. A., ... & Sato, N. (1997). Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science*, 275(5299), 502-509.

Senay, G. B., Kagone, S., & Velpuri, N. M. (2020). Operational global actual evapotranspiration: Development, evaluation, and dissemination. *Sensors*, 20(7), 1915.

Shan, N., Shi, Z., Yang, X., Zhang, X., Guo, H., Zhang, B., & Zhang, Z. (2016). Trends in potential evapotranspiration from 1960 to 2013 for a desertification-prone region of China. *International Journal of Climatology*, 36(10), 3434-3445.

Sheffield, J., Goteti, G., & Wood, E. F. (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of climate*, 19(13), 3088-3111.

Siemann, A. L., Chaney, N., & Wood, E. F. (2018). Development and validation of a long-term, global, terrestrial sensible heat flux dataset. *Journal of Climate*, 31(15), 6073-6095.

Simard, M., Pinto, N., Fisher, J. B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research: Biogeosciences*, 116(G4), doi: [10.1029/2011JG001708](https://doi.org/10.1029/2011JG001708)

Stoy, P. C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., ... & Varlagin, A. (2013). A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity. *Agricultural and forest meteorology*, 171, 137-152.

Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's global energy budget. *Bulletin of the American Meteorological Society*, 90(3), 311-324.

Van den Hurk, B. J. J. M., Viterbo, P., Beljaars, A. C. M., & Betts, A. K. (2000). Offline validation of the ERA40 surface scheme, ECMWF TechMemo 295. Reading, UK.

Viterbo, P., & Beljaars, A. C. (1995). An improved land surface parameterization scheme in the ECMWF model and its validation. *Journal of climate*, 8(11), 2716-2748, 10.1175/1520-0442(1995)008<2716:AILSPS>2.0.CO;2

Wang, H., & Li, D. (2011). Correlation of surface sensible heat flux in the arid region of northwestern China with the northern boundary of the East Asian summer monsoon and Chinese summer precipitation. *Journal of Geophysical Research: Atmospheres*, 116(D19).

Watts, C. J., Chehbouni, A., Rodriguez, J. C., Kerr, Y. H., Hartogensis, O., & De Bruin, H. A. R. (2000). Comparison of sensible heat flux estimates using AVHRR with scintillometer measurements over semi-arid grassland in northwest Mexico. *Agricultural and Forest Meteorology*, 105(1-3), 81-89.

Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P., ... & Verma, S. (2002). Energy balance closure at FLUXNET sites. *Agricultural and Forest Meteorology*, 113(1-4), 223-243.

Xin, Y., Liu, J., Liu, X., Liu, G., Cheng, X., & Chen, Y. (2022). Reduction of uncertainties in surface heat flux over the Tibetan Plateau from ERA-Interim to ERA5. *International Journal of Climatology*.

Yang, K., Qin, J., Guo, X., Zhou, D., & Ma, Y. (2009). Method development for estimating sensible heat flux over the Tibetan Plateau from CMA data. *Journal of applied meteorology and climatology*, 48(12), 2474-2486.

Zhang, X., Dai, Y., Cui, H., Dickinson, R. E., Zhu, S., Wei, N., ... & Fu, W. (2017). Evaluating common land model energy fluxes using FLUXNET data. *Advances in atmospheric Sciences*, 34(9), 1035-1046.

6 Appendix

6.1 Statistical metrics

The bias, the root mean square difference (RMSD), the unbiased root mean square difference (uRMSD), the mean absolute relative difference (MARD) and the mean absolute difference (MAD), are defined as:

$$\text{Bias} = \frac{1}{N} \times \sum_{i=1}^N (E_i - M_i) \quad (1)$$

$$\text{RMSD} = \left[\frac{\sum_{i=1}^N (E_i - M_i)^2}{N} \right]^{1/2} \quad (2)$$

$$\text{uRMSD} = \left[(\text{RMSD})^2 - (\text{Bias})^2 \right]^{1/2} \quad (3)$$

$$\text{MARD (\%)} = \frac{100}{N} \times \sum_{i=1}^N \frac{|E_i - M_i|}{M_i} \quad (4)$$

$$\text{MAD} = \frac{1}{N} \times \sum_{i=1}^N |E_i - M_i| \quad (5)$$

Where N is the number of estimations/measurements and E_i and M_i represent the estimated and measured (or reference) values, respectively. For uRMSD and bias calculations, reference values correspond to the FLUXNET2015/ICOS data in the section 4.1 while they are other dataset values (LSA SAF, ERA5, GLDAS and GLEAM) in the section 4.2.

6.2 FLUXNET2015/ICOS sites

6.2.1 In situ eddy-covariance sites

Table 7-1 provides information about the 30 selected stations for validations.

Table 7-1: Information about *in situ* selected stations used to validate the CM SAF product.

	SITE_ID	longitude	latitude	Country	Elevation (m)	IGBP	Area	Climate
Ankasa	GH-Ank	-2.69421	5.26854	GH	124.0	EBF	NAfr	Am
Brasschaat	BE-Bra	4.51984	51.30761	BE	16.0	MF	Euro	Cfb
Demokeya	SD-Dem	30.4783	13.2829	SD	500.0	SAV	NAfr	BWh
Fontainebleau u Barbeau	FR-Fon	2.7801	48.47636	FR	103.0	DBF	Euro	Cfb
Fyodorovskoye	RU-Fyo	32.92208	56.46153	RU	265.0	ENF	Euro	Dfb
Hainich	DE-Hai	10.45217	51.07921	DE	430.0	DBF	Euro	Cfb
Hyytiala	FI-Hyy	24.29477	61.84741	FI	181.0	ENF	Euro	Dfc
Klingenberg	DE-Kli	13.52238	50.89306	DE	478.0	CRO	Euro	Cfb
Le_Bray	FR-LBr	-0.7693	44.71711	FR	61.0	ENF	Euro	Cfb
Lettosuo	FI-Let	23.95952	60.64183	FI	111.0	ENF	Euro	Dfb
Lonzee	BE-Lon	4.74623	50.55162	BE	167	CRO	Euro	Cfb
Loobos	NL-Loo	5.74356	52.16658	NL	25	ENF	Euro	Cfb
Monte Bondone	IT-MBo	11.04583	46.01468	IT	1550	GRA	Euro	Dfb
Oberbärenburg	DE-Obe	13.72129	50.78666	DE	734	ENF	Euro	Cfb
Puechabon	FR-Pue	3.5957	43.7413	FR	270	EBF	Euro	Csa
Selhausen	DE-Seh	6.44965	50.87062	DE	103	CRO	Euro	Cfb
Selhausen Juelich	DE-RuS	6.44714	50.86591	DE	102	CRO	Euro	Cfb
Skukuza	ZA-Kru	31.4969	-25.0197	ZA	359	SAV	SAfr	Cwa
Sodankyla	FI-Sod	26.63859	67.36239	FI	180	ENF	Euro	Dfc
Tharandt	DE-Tha	13.56515	50.96256	DE	385	ENF	Euro	Cfb
Vielsalm	BE-Vie	5.99812	50.30493	BE	493	MF	Euro	Cfb
Gebesee	DE-Geb	10.91463	51.09973	DE	161.5	CRO	Euro	Cfb
Neustift	AT-Neu	11.3175	47.11667	AT	970	GRA	Euro	Dfc
Grillenburg	DE-Gri	13.51259	50.95004	DE	385	GRA	Euro	Cfb
Soroe	DK-Sor	11.64464	55.48587	DK	40	DBF	Euro	Cfb
Grignon	FR-Gri	1.95191	48.84422	FR	125	CRO	Euro	Cfb
Collelongo	IT-Col	13.58814	41.84936	IT	1560	DBF	Euro	Cfa
Lavarone	IT-Lav	11.28132	45.9562	IT	1353	ENF	Euro	Cfb
Renon	IT-Ren	11.43369	46.58686	IT	1730	ENF	Euro	Dfc
Davos	CH-Dav	9.843558	46.81297	CH	1639	ENF	Euro	ET

6.2.2 Potential error at Skukuza station (South Africa)

Figure 34 displays, for the year 2005 at Skukuza station, the scatter plot between CM SAF latent heat flux bias as compared to *in-situ* data and the difference between gap-filled (MDS method) energy balance corrected latent heat flux (LE_{CORR} ; $W m^{-2}$) and uncorrected gap-filled (MDS method) latent heat flux (LE_{F_MDS} ; $W m^{-2}$). The strong linear relationship obtained ($R^2=0.98$; $N=950$) was not expected as the bias should not be linked to the correction applied on the data to close the energy budget. Such high correction ($> 200 W m^{-2}$) was not only observed for this specific year. In addition, temporal course of LE_{CORR} at station (Figure 35) point out the year 2005 where extremely high value are observed. Therefore, some doubts might be raised about the quality of the correction for this station.

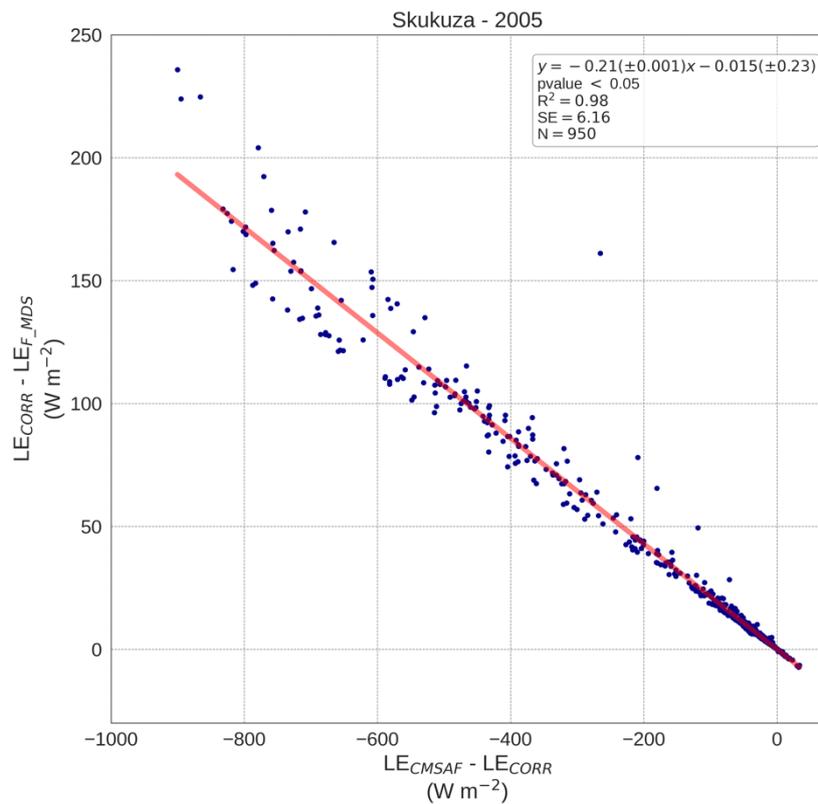


Figure 34: Scatter plot of CM SAF latent heat flux bias ($W m^{-2}$) as compared to *in-situ* data for the year 2005 at “Skukuza” station vs. the difference between gap-filled (MDS method) energy balance corrected latent heat flux (LE_{CORR} ; $W m^{-2}$) and the uncorrected gap-filled (MDS method) latent heat flux (LE_{F_MDS} ; $W m^{-2}$). Solid red line is linear fit.

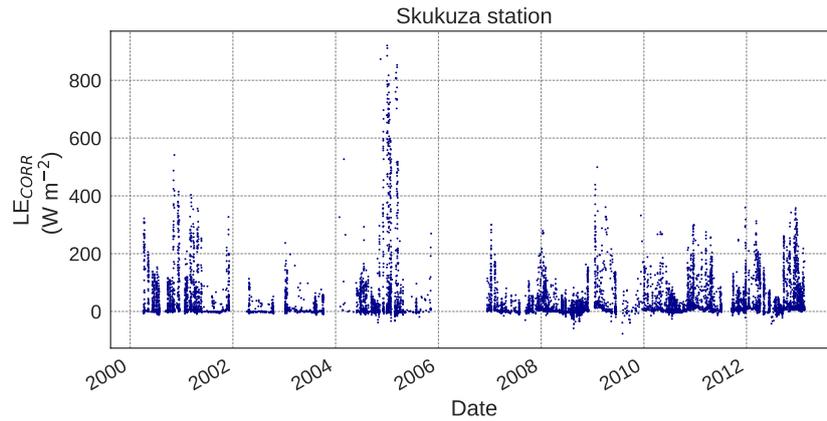


Figure 35: Temporal course of gap-filled (MDS method) energy balance corrected latent heat flux (LE_{CORR} ; $W m^{-2}$) from FLUXNET2015 dataset at Skukuza station.

6.3 Land cover

Figure 36 displays the remapped (0.05° , regular grid) land cover of the main tile used as input for the year 2005 of MFG and MSG. While both maps look similar, the difference in grid size between MFG and MSG maps has an impact on the land cover use (Figure 37).

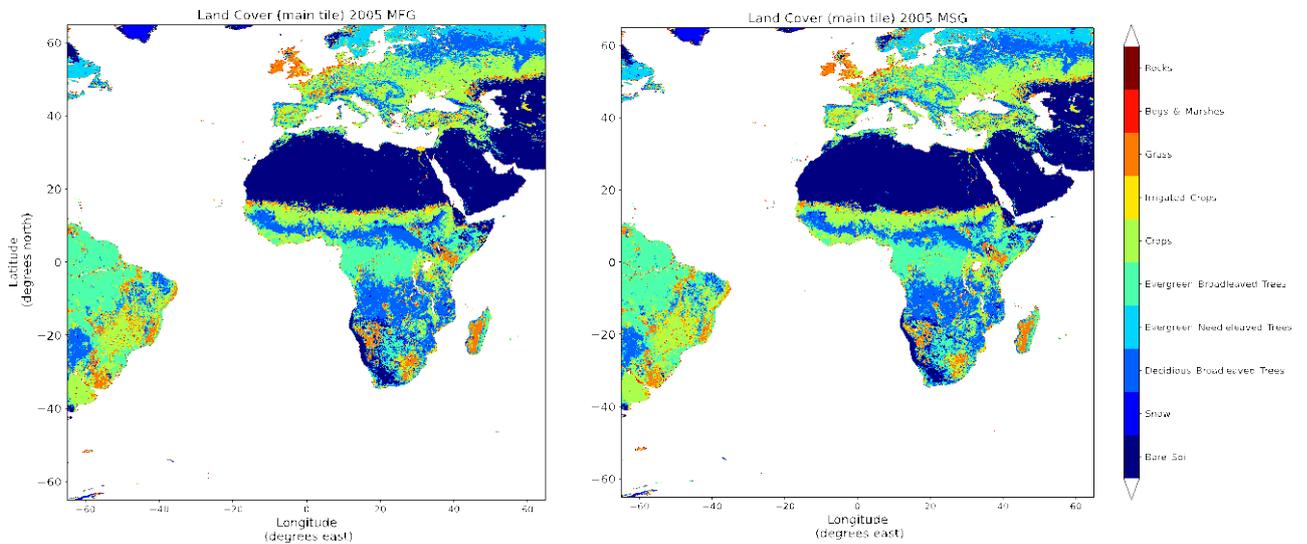


Figure 36: Land cover map used as the main tile input for the year 2005 for MFG (left) and MSG (right).

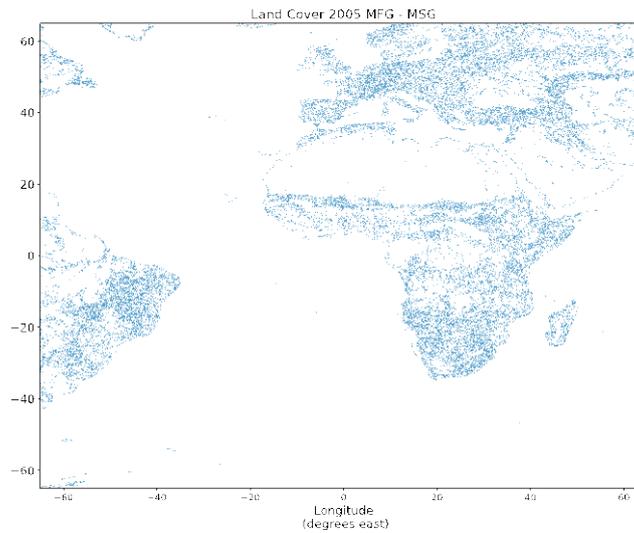


Figure 37: Difference between main land cover tile used as input for the year 2005 from MFG and MSG.

Figure 38 shows the land cover map (0.05°; geostationary view) used as the main tile input for March 2005 in the LSA SAF dataset. Clearly, land cover used as input differ between LSA SAF and CM SAF product and would impact the estimations.

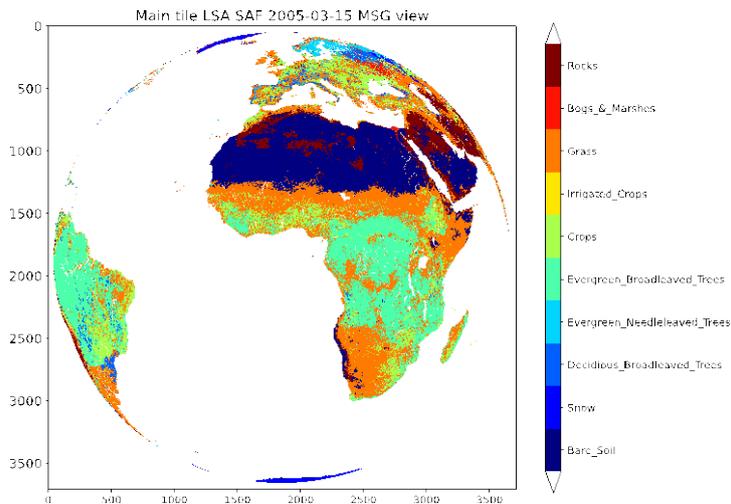


Figure 38: Land cover map used as the main tile input in the LSA SAF dataset for 2005-03-15.