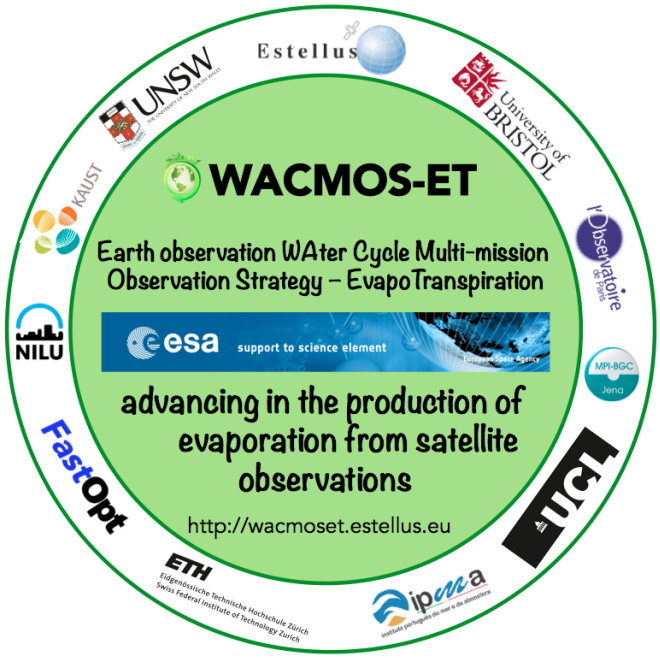

WP3100

ET Experimental Dataset

Prepared by:

Carlos Jimenez
Estellus, France

Diego Miralles
Ghent University, Belgium

Ali Ershadi, Matthew McCabe
KAUST, Saudi Arabia

Dominik Michel
ETHZ, Switzerland

Martin Jung
MPI-Biogeochemistry, Germany

Contents

1 Introduction

This document describes the generation of the WACMOS-ET evapotranspiration (ET) data set and the products used to evaluate the produced ET estimates. The data set consists of a suite of ET products at different spatial and temporal scales generated by running four ET algorithms with the relevant products of the Reference Input Data Set (RIDS). The main driver of the ET generation has been to facilitate the evaluation of the ET algorithms with related data sets, and in that sense the ET generation has been organized around three spatial scales. Firstly, the ET models have been run at a selection of tower flux locations using as much as possible the tower meteorological data. To investigate the degradation in performance when the ET models are run with the spatially coarser data, the ET models have also been run with co-located satellite inputs and inputs from the ERA-Interim reanalysis. Secondly, the ET models have been run at a much larger (compared with the tower) spatial resolution of $\tilde{2}5$ km over all continental surfaces to investigate the model performance for climatological applications. Lastly, one relatively small area in Europe has been selected to run the ET models at a smaller (compared with the continental runs) spatial resolution of $\sim$1 and 5 km to investigate the effects of spatial aggregation in the produced ET. In terms of the temporal scales, the ET models have been run sub-daily at all spatial scales, and also daily at the tower scale to investigate the effect of temporal aggregation.

The document is organized by first presenting the ET models and inputs required to run the models, followed by a description of how the ET estimates at the different spatial resolutions have been produced.

2 ET Modeling

Four models were selected by the project to produce ET estimates at different scales. The rationale for the model selection, detailed model descriptions, and the inputs required can be found at the WP1200 report (Preliminary analyses of ET algorithms). The model names are shortened from here on as:

- the Penman-Monteith algorithm used for the MODIS ET product (MOD16, [Mu et al., 2011]) is referred to as **PM**.
- the Prisley-Taylor algorithm published in Fisher et al. [2008] is referred to as **PT**.
- the Surface Energy Balance System [Su, 2002] is referred to as **SEBS**.
- the Global Land-surface Evaporation: the Amsterdam Methodology of [Miralles et al., 2011b] is referred to as **GLEAM**.

A few aspects of the ET modeling need to be discussed as not all models operate in the same way:

- **Evaporation partitioning.** PM, PT, and GLEAM produces separated ET estimates for the fraction of the land covered by the vegetation, the bare soil, and also an estimate for the evaporation from the rain intercepted by the canopy. This is not the case for SEBS, where only one ET estimate combining soil and vegetation evaporation is available.

- **Night-time evaporation.** PM and GLEAM have dedicated parameterizations to produce night-time ET. This is not the case for PT and SEBS. Given the usual low values of night-time ET, if night-time ET is required (e.g., to estimate daily ET from the 3-hourly ET estimates) it will be set to zero for PT and SEBS.

- **Snow evaporation.** Though all models use the meteorological inputs to adapt soil and plant evaporation rates under very cold conditions, only GLEAM includes an adapted parameterization to capture snow sublimation. The evaporation rates in winter time and snowy conditions are generally low, so even large errors in the ET estimation at this time of the year will not have a large impact when the ET is aggregated over longer time periods. In order to have an all season data record the ET for these conditions will be also considered.

The 4 models have been coded in the interpreted language Matlab.

3 ET inputs

3.1 Spatial resolution

In terms of data types (and somehow the associated spatial resolution) the inputs have been classified as tower, reanalysis, satellite, and gridded:

- The **tower inputs** are variables measured by the instrumentation of the tower, and are used to test the ET models at the tower scale with inputs of comparable spatial and time resolution to the tower ET. There are available with a time resolution of 1/2 hour, but here they are time integrated to 3-hour intervals to evaluate model performance at that time resolution. Their spatial resolution depends on the height of the tower, placement of the instruments within the tower, etc, but in general tower fetchs are of the order of a few hundreds of meters. Notice that the ET models cannot be run solely with tower data, as a few variables are not routinely measured at the towers. This is normally the case for vegetation related variables, such as the Fraction of Vegetation Cover (FVC), or the Leaf Area Index (LAI), or vegetation height. For these variables, satellite products or educated guesses are used.

- The **satellite inputs** are variables measured from space and are used to illustrate ET model performance when the tower inputs are replaced by a given choice of satellite inputs. The term satellite is used here in a very

loose way, as some of the inputs are coming from the ERA-Interim reanalysis where typically a large suite of meteo observations (ground network, radiosondes, satellites, etc) are assimilated into a physical model to derive an estimate of a given variable. Their spatial and time resolutions can differ vastly between products (from the ∼1 km project LAI to the ∼100 km SRB net radiation product), and can be quite different from the tower fetch. Therefore the ET estimated from the models represents an estimate where different spatial and time resolutions are combined, in nearly all cases at poorer resolutions that the tower ET.

- The **reanalysis inputs** are variables derived from the ERA-Interim reanalysis. As indicated above, some of the inputs used here come from this reanalysis, and as turbulent fluxes are also output by ERA-Interim, there may be an interest in comparing our ET models running with the reanalysis inputs with the ERA-Interim ET. Notice that the satellite runs also use ERA-Interim, but only for the meteorology. Here all inputs but the vegetation are from the reanalysis (e.g., also the surface radiation, surface temperature, precipitation, etc). The spatial resolution of the ERA-Interim inputs corresponds to the grid of their physical model (∼75km), which is much larger than the typical tower fetch. Estimates from the forecast exist every 3 hours.

- The **gridded inputs** are the satellite inputs gridded at a common spatial and time resolution to allow running the ET models at continental scales. For instance, the wind used in the satellite inputs was the original ERA-Interim wind at ∼75 km, while now the wind field has been gridded to a resolution of ∼25 km. Another example: LAI at a satellite input has a resolution of ∼1 km, while here we use the ∼25 km LAI also available to the project. The gridded inputs are used to evaluate the model performance from the global runs by comparing the gridded ET estimates from a given gridded cell with the ET fluxes from towers located at that cell. Compared with the satellite inputs, all inputs are now at a common spatial resolution, though the inherent original spatial resolution of the inputs still plays a role in the "true" resolution of the products. The gridding has been done by simple linear interpolations if the original resolution was larger than the gridded resolution, nearest neighbour matching if they were of comparable resolution, or spatial averages if the original resolution was much finer than the gridded resolution. An exception is land cover, where spatial upscaling is done by selecting the dominant cover.

In-situ estimates of ET at large spatial resolutions are not routinely available, so the ET model estimates run with the different inputs are always compared with the same tower ET. The best model performance for an evaluation with the tower ET is then expected to happen when the models are run with the tower inputs, and an expected degradation in model performance is likely to happen when the ET models are run with the other inputs. On the one hand, when the spatial resolution of the inputs starts to differ from the tower fetch, the ET

at the tower will be less representative of the modeled ET. The degradation in model performance (as compared with the tower ET) will be more or less severe depending on surface heterogeneity around the tower. On the other hand, the satellite inputs are likely to have larger errors compared with the in-situ measurements at the tower. Therefore it could be expected that satellite inputs are capturing the physical processes at the tower less accurately than the in-situ observations at the tower, resulting in lower ET model performance due to more uncertain inputs.

3.2 Time resolution

In terms of time resolution, the tower data is available at 1/2 hour intervals and has been time-integrated to 3-hourly in order to run the ET models with 3-hourly inputs. The 3-hourly inputs have also been aggregated to daily values in order to run the models with daily inputs. The remaining data has been time-matched to the 3-hourly or daily resolutions from their native resolution in different ways depending on the type of data and original resolution. For instance, the vegetation products have an 8-day time resolution, and the 8-day value is assigned to all 3-hourly intervals in that period.

It should be noted that the tower data record is not time-continuous as in some occasions there are no measurements. This is not a problem for the PM, PT, and SEBS models because the ET estimation depends only on an instantaneous atmospheric/surface state. When inputs to the models and/or ET for the evaluation are missing those 3 models are not run. GLEAM is different as it operates as a small hydrological model requiring continuous data records. To facilitate running GLEAM the tower inputs have been gap-filled with ERA-Interim data, though as before only the time instants where the tower data existed are used for the model evaluation.

3.3 Tower LST

The tower LST is derived from the longwave radiometer measurements at the tower by inverting the equation relating the upwelling spectral radiance measured by the radiometer and the LST. If LWout and LWin are the upwelling and downwelling longwave radiation, respectively, ϵ is the broadband emissivity of the Earth surface in the spectral range of measurements, and σ is the Stefan Boltzman constant, the equation to derive the LST takes the form:

$$\text{LWout} = (1 - \epsilon)\ \text{LWin} + \epsilon\ \sigma\ \text{LST}^4$$

The broadband emissivity is estimated from the MODIS-based Global Infrared Land Surface Emissivity Database [Seemann et al., 2008] operated by the Cooperative Institute for Meteorological Satellite Studies (CIMSS). The estimates are calculated by following the approach suggested by Wang et al. [2005] using the following linear combination of narrowband emissivities at 8.5, 11, and 12

μm:

$\epsilon = 0.2122\ \epsilon_{8.5} + 0.3859\ \epsilon_{11} + 0.4029\ \epsilon_{12}$

3.4 LAI/FAPAR model adaptation

Given that the PT, PM, and SEBS models have been run in the past with vegetation products derived from MODIS, a direct application of the project TIP-derived LAI and FAPAR is not possible (the reader is referred to the project LAI/FAPAR Validation Report for more details). Therefore we have undertaken a biome-based calibration between the MODIS LAI/FAPAR 8-days 1 km product MOD15A2 and the WACMOS-ET TIP-based LAI/FAPAR in order to keep the dynamics of the WACMOS-ET product but with the absolute values expected from a MODIS-like LAI/FAPAR. The resulting product will be referred to as WACMOS-ET "MODIS-like" LAI/FAPAR.

For the tower runs the calibration is local, i.e., a linear regression between the MOD15A2 and the TIP values co-registered at each tower is carried out, followed by an application of the local regression coefficients to the WACMOS-ET LAI/FAPAR time series to produce the MODIS-like LAI/FAPAR. Figure 1 gives examples of the products at two towers. CA-Qfo is a station in a evergreen needle forest and the MOD15A2 and WACMOS-ET LAI/FAPAR absolute values differ considerably. This is expectd as allowing some form of horizontal clumping (MODIS 3D radiative transfer scheme) or not (TIP 1-D) for these canopies can result in large differences in the estimated LAI/FAPAR. It can be seen that the local calibration of the MODIS-like product retains the dynamics of the WACMOS-ET product while adding absolute values close to the MODIS product. US-Bkg is situated in a cropland area, where the effects of clumping are much less attenuated, and the different LAI/FAPAR values are much closer.

For the global LAI/FAPAR, the calibration is biome/climate based. The minimum and maximum near-surface air temperature and relative humidity (Tmin, Tmax, and RH, respectively) for all the pixels falling in a given biome are collected and a mean Tmin, Tmax, and RH per biome are estimated. These mean values are used to prepare a sub-biome classification of the pixels (a given pixel is associated to a given sub-class depending on the pixel mean Tmin, Tmax, and RH being below or above the biome mean values). For each of the sub-biomes a linear regression between MOD15A2 and WACMOS-ET LAI and FAPAR is calculated and the regression coeffcients applied to each pixel (falling in that sub-biome) time series of WACMOS-ET values to derive the MODIS-like LAI/FAPAR.

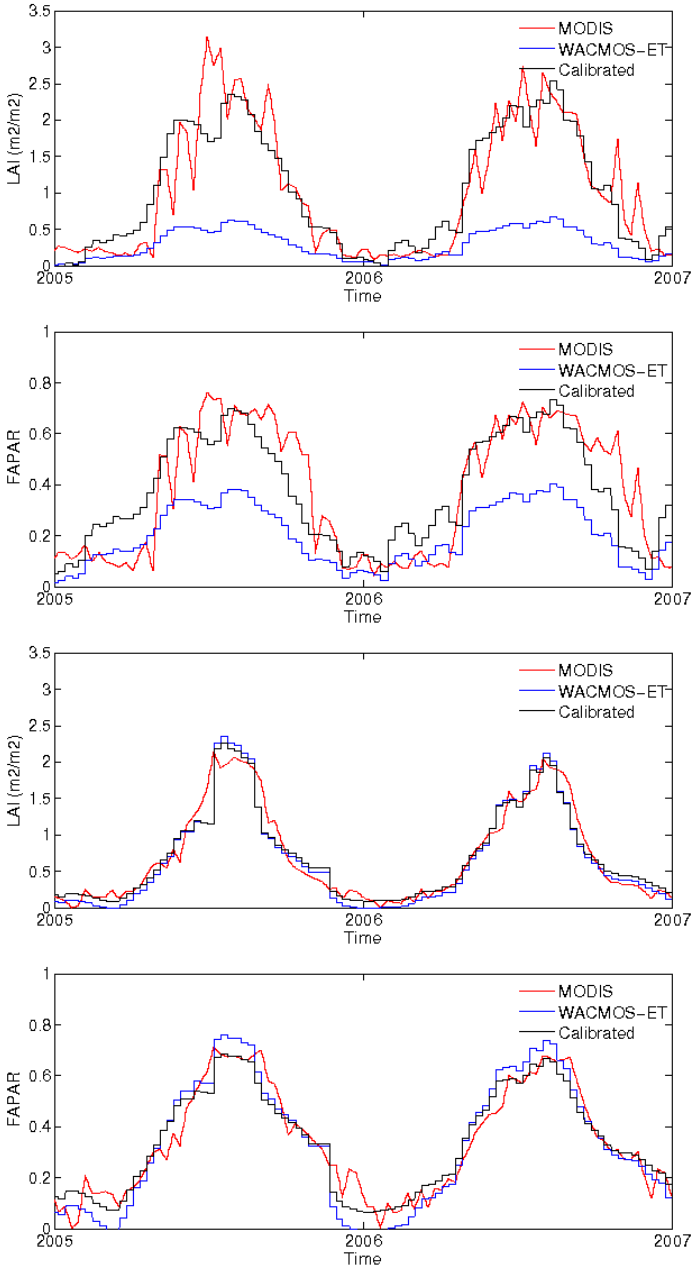

Figure 1: 2005-2006 time series of MODIS MOD15A2 LAI and FAPAR, WACMOS-ET LAI and FAPAR, and the WACMOS-ET MODIS-like LAI and FAPAR (referred to as calibrated in the figures) at the Canadian tower station CA-Qfo (top two panels) and the US tower US-Bkg (bottom two panels).

4 ET generation at tower scale

4.1 Tower selection

Dedicated runs at tower scale are produced by running the four ET models over a sub-set of towers extracted from FLUXNET La Thuille synthesis data set (*http://fluxnet.ornl.gov*). The tower selection was driven by the need of having ET fluxes and relevant tower inputs to run the models in the project period of 2005-2007, and it has resulted in the data set of 24 stations described in Table 1. While some meteorological variables such as near-surface air temperature or humidity are measured at nearly all towers, some other inputs such as the surface net radiation or the ground flux, which are key variables to run the ET models, are measured at a much more limited number of towers. Some stations that were very close to the shore or in places with regular flooding have also been discarded as not being representative of land conditions for the whole year. The final selection certainly does not result in a very large number of towers, but they represent a significant number of biomes and climates, with a reasonable sample of dry-wet conditions.

The length of the data record varies from station to station. For the stations selected only data in 2005-2006 exist. Figure 2 summarizes the existing data by plotting the daily ET derived from the eddy covariance latent flux measurements at the different towers. Notice that the models will be evaluated only for non rainy conditions under the assumption that eddy covariance measurements at the tower are not reliable under rain. Only days with no rain at any of the 3-hour intervals are kept for the evaluation.

Concerning the times selected for the evaluation, in order to sample all tower data at the same times independent of tower location, the 0-3-...21 3-hourly intervals are selected in local time. For the data existing every 3-hours in UTC time, time-matching to the 3-hourly tower intervals has been done by a linear interpolation in time to the 3-hourly local time intervals.

4.2 Tower surface energy balance

In principle the surface energy balance should close at the tower. This is rarely the case: a lack of closure in the surface energy balance of about 10 - 30% is commonly found when using Eddy Covariance (EC) techniques to measure the turbulent fluxes (see e.g., Foken [2006]). In practical terms this means that the uncertainty associated to the absolute ET reported by the EC measurements is relatively large and care should be taken when reporting biases between the tower ET and the ET estimates at the tower.

Given this common lack of energy closure at the tower, the modeled ET is compared with the EC observations but also with the tower ET estimates derived by imposing energy closure at the tower, i.e., the latent flux should be equal to the surface net radiation minus the turbulent sensible and ground fluxes also observed at the tower. These estimates are referred to as Energy Residual (ER) ET.

Table 1: Stations selected to run the models with tower inputs. From left to right the station name; longitude; latitude; Köppen-Geiger Climate Classification (KGCC); International Geosphere-Biosphere International Programme (IGBP) land cover; total number of days with data / no precipitation number of days with data; evaporative fraction for the DJF, MAM, JJA, SON 3-monthly periods.

Name	**Lon**	**Lat**	**KGCC**	**IGBP**	**Days**	**EF**
AU-How	131.15E	12.49S	Aw	SV	114 / 100	0.7/0.7/0.5/0.3
CA-Ojp	104.69W	53.92N	Dfc	ENF	126 / 101	0.2/0.1/0.3/0.5
CA-Qfo	74.34W	49.69N	Dfc	ENF	253 / 166	0.1/0.1/0.4/0.4
DE-Geb	10.91E	51.1N	Cfb	CRO	188 / 113	0.0/0.4/0.5/0.7
DE-Har	7.6E0	47.93N	Cfb	MF	105 / 88	1.0/0.5/0.5/0.7
DE-Kli	13.52E	50.89N	Cfb	CRO	275 / 98	0.0/0.5/0.5/0.0
DE-Meh	13.52E	50.89N	Cfb	CRO	444 / 269	0.0/0.3/0.5/0.5
DE-Wet	11.46E	50.45N	Cfb	ENF	384 / 182	0.1/0.3/0.4/0.6
IT-MBo	11.08E	46.03N	Dfb	MF	149 / 126	0.0/0.6/0.7/0.9
IT-Noe	8.15E	40.6N	Csa	WL	182 / 182	0.7/0.3/0.2/0.3
NL-Ca1	4.93E	51.97N	Cfb	CRO	38 / 22	1.0/0.6/0.6/0.9
PT-Mi2	8.02W	38.48N	Csa	SV	275 / 221	0.5/0.4/0.3/0.4
RU-Fyo	32.92E	56.46N	Dfb	MF	374 / 216	0.0/0.4/0.5/0.4
US-ARM	97.49W	36.61N	Cfa	CRO	159 / 131	0.4/0.5/0.3/0.3
US-Aud	110.51W	31.59N	BSk	OSH	219 / 219	0.5/0.2/0.3/0.5
US-Bkg	96.84W	44.35N	Dfa	CRO	174 / 172	0.6/0.7/0.9/1.0
US-Bo2	88.29W	40.01N	Dfa	CRO	192 / 192	0.3/0.3/0.6/0.3
US-FPe	105.1W	48.31N	BSk	GRA	184 / 184	1.0/0.4/0.6/0.4
US-Goo	89.87W	34.25N	Cfa	CRO	183 / 179	0.7/0.6/0.5/0.6
US-MOz	92.2W	38.74N	Cfa	DBF	252 / 252	0.3/0.4/0.5/0.5
US-SRM	110.87W	31.82N	BSk	OSH	139 / 137	0.2/0.1/0.3/0.4
US-WCr	90.08W	45.81N	Dfb	DBF	338 / 239	0.1/0.2/0.6/0.4
US-Wkg	109.94W	31.74N	BSk	GRA	137 / 137	0.2/0.1/0.2/0.3
US-Wrc	121.95W	45.82N	Csb	ENF	146 / 107	0.4/0.3/0.3/0.5

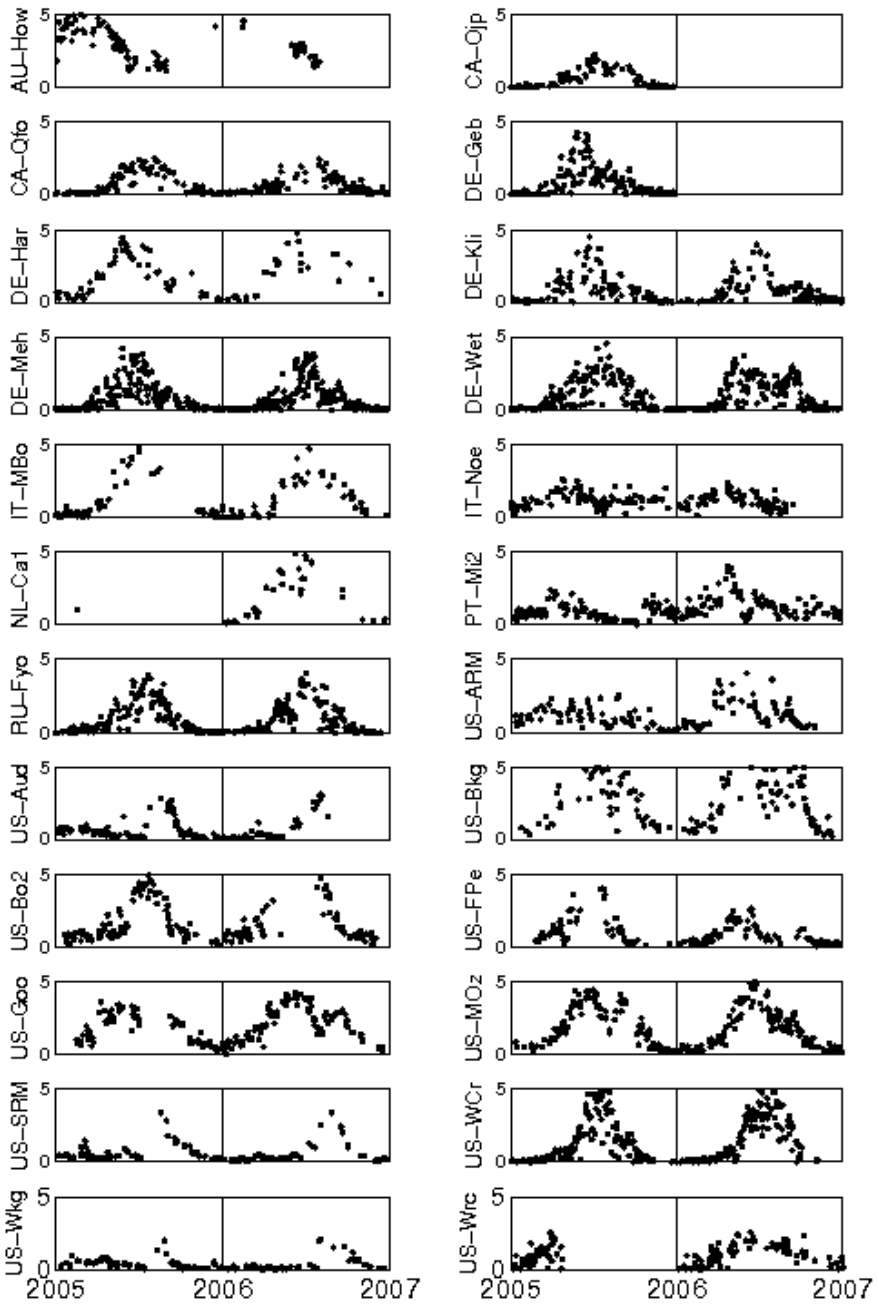

Figure 2: 2005-2006 measured eddy covariance ET in mm/day at the 24 stations where the models are run with tower inputs.

4.3 Model considerations

As discussed in Section 2, only PM and GLEAM specifically deal with nighttime evaporation. Nevertheless, nighttime values are required from all models to integrate the 3-hourly ET estimates to daily values. As indicated in Section 2 the nighttime values are set to zero for SEBS and PT to allow the daily integration. To separate day and night, daylight times are identified by calculating the solar zenith angle, and time intervals where the cosine of the zenith angle is larger than 0.2 are kept as day values.

SEBS is the only model that requires LST as an input. Strictly speaking the tower run with satellite inputs should use a satellite LST. However, due to cloudiness, revisiting time and satellite overpass (for AATSR), and missing data (for the geostationary LST), the number of 3-hourly LSTs at the stations reduces drastically (compared with the tower LST derived from the radiometric observations at the tower). As the general ET evaluation at the tower is done over 3-hourly and daily values where all models output ET, using the satellite LST removes a large number of ET values from the analysis. Therefore we use tower LST for the SEBS satellite runs. However, a special run for SEBS using again the satellite inputs but now with the LST at the tower from AATSR and the geosatellite observations is also produced separately from the satellite run using tower LST.

4.4 ET production at tower scale

Summarising, the different model forcings and time integrations results in the following 3-hourly ET estimates:

- **Tower run.** ET generated by the 4 models run with the tower data (surface radiation; LST, air temperature, humidity, and wind speed; precipitation) complemented with MODIS LAI/FAPAR, and WACMOS soil moisture.
- **ERA run.** ET generated by the 4 models run with the ERA-Interim reanalysis data for all inputs (surface radiation; surface temperature, air temperature and wind speed; precipitation; soil moisture) complemented with MODIS LAI/FAPAR.
- **Sat-MODIS run.** ET generated by the 4 models run with the satellite data (tower LST; SRB surface radiation; ERA-Interim air temperature, humidity, and wind speed; CMORPH precipitation; WACMOS soil moisture) using MODIS products for LAI/FAPAR.
- **Sat-WACMOS run.** As before but with the WACMOS-ET MODIS-like products processed as indicated in Section 3.4. Notice that both runs are identical for GLEAM, as GLEAM does not use LAI/FAPAR.
- **LST runs.** ET generated by SEBS with the satellite inputs as above, with MODIS products for vegetation, and different LST products (from the tower, AATSR, and MSG, GOES, or MTSAT depending on location).

The ET estimates from these runs are available at 3-hourly, as daily values from the time-integration of the 3-hourly ET estimates, and as daily values from running the models with the inputs time-integrated from 3-hourly to daily.

5 ET generation at regional scale

5.1 ET inputs

ET at the finest spatial resolution allowed by the RIDS is produced over an area in Europe of ∼500 x 600 km2 (tile h18v04). Figure 3 shows the location of the area and its IGBP land cover.

ET is generated at 2 spatial resolutions:

- **1 km product**. For these runs the ERA-Interim surface radiation, near-surface air temperature, and humidity are replaced by hourly outputs from the NWP model COSMO, downscaled from the original resolution of ∼7 km to 1 km. The 1 km AATSR product is used for surface temperature, and the 1 km WACMOS-ET MODIS-like LAI/FAPAR characterizes the vegetation.
- **5 km product**. As before, but with the COSMO metorological inputs downscaled to 5 km, the surface temperature from MSG, and the same LAI/FAPAR but from the 5 km product.

5.2 ET models

Only the PT, PM, and SEBS are run at this spatial scale as they can be applied with existing observations from visible and infrared sensors at this fine scale. For GLEAM, key inputs such as the precipitation, soil moisture, and vegetation optical depth, are based on microwave observations with a spatial resolution close to the resolution of the continental runs (∼25 km). In principle, these inputs could be downscaled to 1 km and GLEAM run over the same areas as the other models, but given the difference in the "true" resolution of these inputs and the targeted 1 km resolution, runs of GLEAM at 1 km are not prepared.

5.3 ET evaluation

Evaluation of the regional is mainly done by comparing the runs at different resolutions. The region includes the sub-catchment of Rietholzbach, managed by one of the consortium members. A lysimeter there provides ET measurements representative of the Rietholzbach catchment, and that will be used in the analyses as in situ data.

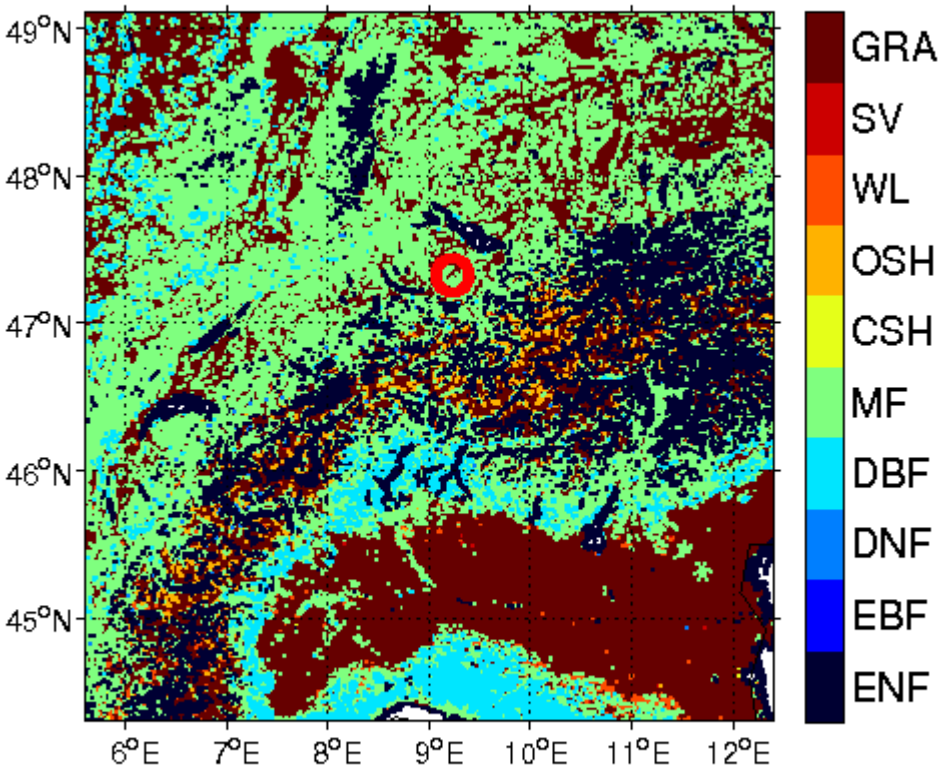

Figure 3: Area selected for the regional runs, displaying the IGBP land cover of the area (ENF: evergreen needleleaf forest; EBF: evergreen broadleaf forest; DNF: deciduous needleleaf forest; DBF: deciduous broadleaf forest; MF: mixed forest; WL: woody savannas; SV: savannas; CSH: closed shrubland; OSH: open shrubland; Grass: grassland, urban and built-up, barren or sparsely vegetated; Crop: cropland). The red mark shows the location of the Rietholzbach catchment.

6 ET generation at continental scale

6.1 ET production

The continental scale runs are produced by running the four models with 3-hourly (UTC) gridded inputs on the project sinusoidal grid of ∼25 km. The 3-hourly ET estimates from PT, PM, and GLEAM, are also integrated to daily, monthly, and yearly values. An example of ET estimates for Day-Of-the-Year (DOY) 125 in 2007 for PT, PM, and GLEAM is given in Figure 4.

To run SEBS over all the continental surfaces AATSR is selected due to their global coverage, with its 1 km LST upscaled to 25 km by simple spatial averages and the ERA-Interim near surface meteorology linearly interpolated in time to the satellite overpass (∼10 am local time) from the closest UTC 3-hourly times. Although some techniques could be used to upscale from local overpass ET to daily ET (e.g., by assuming a day constant evaporative fraction as in Vinukollu et al. [2011]), this has not been pursued here, and no daily, monthly, or yearly ET from SEBS is produced. An example of SEBS ET at AATSR overpass (∼10 am) for the same day of Figure 4 is given in Figure 5.

Summarizing, the following ET estimates are generated:

- 3-hourly ET for all continental surfaces applying the PT, PM, and GLEAM models, also time integrated to daily, monthly, and yearly values.
- ∼10 am SEBS ET for all continental surfaces when AATSR LST observations are available.

6.2 ET for in-situ evaluation

In terms of evaluation, both in situ tower flux observations, global ET estimations from other products, and catchment water balance estimates at major basins were considered. Regarding the tower records, 187 towers from La Thuile data set were found to have ET observations in the 2005-2007 period. From those, only 95 have been retained imposing two conditions (1) the tower was not an area of regular flooding, and was not close to water bodies, and (2) at least 30 non-rainy days existed to evaluate the ET gridded estimates. The GlobCover classification was used to decide about the flooding/water conditions. The final tower selection is listed in Tables 2 and 3.

6.3 ET for global evaluation

For an evaluation at global scale, other ET products at different time and space resolutions have been collected. They include models that have been the basis for the ET algorithms implemented by the project, together with other global data sets build in different ways. There are listed here:

- The **MODIS** ET product (MOD16, Su et al. [2011]), at a resolution of 8 days, ∼1 km. This is the basis of the PM algorithm implemented in the

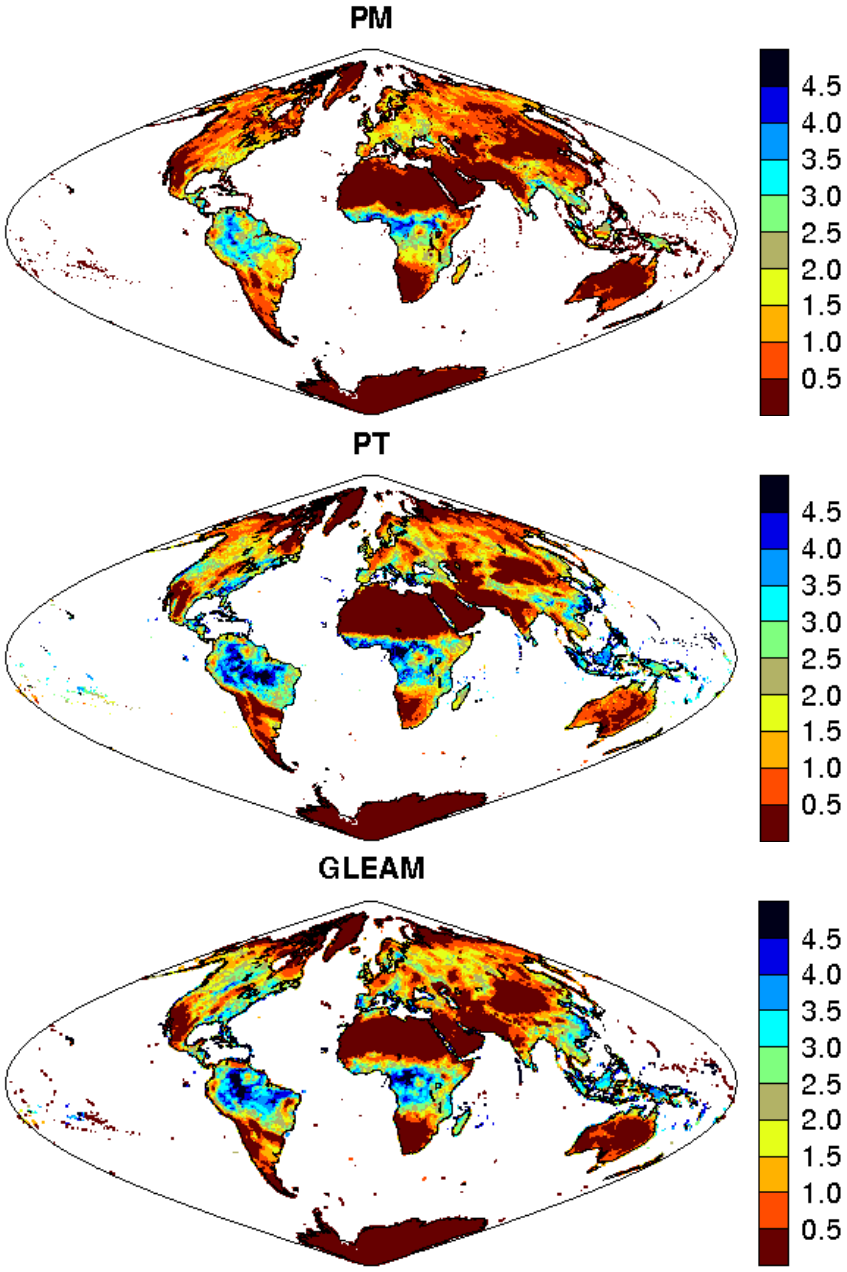

Figure 4: PM, PT, and GLEAM ET in mm/day for DOY 125 in 2007.

Table 2: First 47 stations used to evaluate the gridded ET estimates. From left to right the station name; longitude; latitude; IGBP cover; distance in km to cell center; cell tile; number of days analysed; evaporative fraction for the DJF, MAM, JJA, SON 3-monthly periods.

Name	**Lon**	**Lat**	**km**	**Tile**	**Days**	**IGBP**	**EF**
AU-How	131.15E	12.49S	12.66	h30v10	SV	133	0.6/0.7/0.5/0.3
AU-Wac	145.19E	37.43S	5.18	h29v12	EBF	255	0.5/0.8/1.0/0.6
BE-Bra	4.52E	51.31N	5.14	h18v03	MF	131	0.0/0.3/0.5/0.5
BE-Lon	4.74E	50.55N	3.52	h18v03	CRO	100	0.4/0.6/0.6/0.7
BE-Vie	6E	50.31N	2.85	h18v03	MF	111	0.2/0.2/0.4/0.4
CA-Let	112.94W	49.71N	5.71	h10v04	GRA	161	0.3/0.2/0.7/0.6
CA-NS3	98.38W	55.91N	2.87	h12v03	ENF	107	0.1/0.1/0.4/0.5
CA-NS5	98.49W	55.86N	10.44	h12v03	ENF	67	0.0/0.1/0.4/—
CA-NS6	98.96W	55.92N	10.41	h12v03	ENF	75	0.0/0.1/0.3/—
CA-NS7	99.95W	56.64N	8.81	h12v03	ENF	62	0.1/0.2/0.3/—
CA-Qcu	74.04W	49.27N	8.75	h13v04	MF	139	0.0/0.4/0.4/0.5
CA-Qfo	74.34W	49.69N	8.39	h13v04	ENF	184	0.1/0.1/0.3/0.4
CA-SF1	105.82W	54.49N	9.19	h11v03	ENF	58	0.4/0.3/0.6/1.0
CA-SF2	105.88W	54.25N	6.63	h11v03	MF	75	0.1/0.2/0.6/0.7
CA-SF3	106W0	54.09N	4.42	h11v03	MF	88	0.0/0.2/0.5/1.0
CA-WP1	112.47W	54.95N	6.58	h11v03	MF	145	0.0/0.3/0.5/0.5
CH-Oe1	7.73E	47.29N	12.67	h18v04	MF	56	0.0/0.8/0.8/1.0
DE-Geb	10.91E	51.1N	9.89	h18v03	CRO	124	0.0/0.6/0.4/0.6
DE-Gri	13.51E	50.95N	8.74	h18v03	CRO	48	0.0/0.3/0.3/0.5
DE-Hai	10.45E	51.08N	7.43	h18v03	CRO	72	0.0/0.2/0.3/0.6
DE-Har	7.6E0	47.93N	9.88	h18v04	CRO	88	1.0/0.4/0.5/0.6
DE-Kli	13.52E	50.89N	10.59	h18v03	CRO	105	0.0/0.5/0.5/1.0
DE-Meh	10.66E	51.28N	13.49	h18v03	CRO	285	0.0/0.4/0.5/0.5
DE-Tha	13.57E	50.96N	11.12	h18v03	CRO	159	1.0/0.4/0.4/0.8
DE-Wet	11.46E	50.45N	13.16	h18v03	MF	187	0.1/0.3/0.3/0.6
DK-Lva	12.08E	55.68N	6.31	h18v03	CRO	56	0.0/0.1/0.5/1.0
DK-Sor	11.65E	55.49N	5.06	h18v03	CRO	112	0.0/0.3/0.5/0.7
ES-ES1	0.32W	39.35N	11.11	h17v05	CRO	36	1.0/0.3/0.2/0.6
ES-ES2	0.32W	39.28N	7.64	h17v05	CRO	63	1.0/0.6/0.6/1.0
ES-VDA	1.45E	42.15N	8.99	h18v04	ENF	49	0.0/0.5/0.4/0.4
FI-Hyy	24.29E	61.85N	11.62	h19v02	ENF	251	0.0/0.2/0.4/0.7
FI-Sii	24.19E	61.83N	12.09	h19v02	ENF	83	0.0/0.6/0.5/—
FI-Sod	26.64E	67.36N	9.06	h19v02	WL	245	0.0/0.2/0.3/0.0
FR-Aur	1.11E	43.55N	13.43	h18v04	CRO	53	—/0.6/0.3/—
FR-Fon	2.78E	48.48N	9.26	h18v04	CRO	30	—/0.3/0.5/—
FR-Hes	7.06E	48.67N	3.85	h18v04	MF	247	0.1/0.2/0.4/0.3
FR-LBr	0.77W	44.72N	4.39	h17v04	WL	216	0.8/0.4/0.3/0.5
FR-Lam	1.24E	43.49N	7.59	h18v04	CRO	44	—/0.7/0.3/—
FR-Pue	3.6E0	43.74N	10.64	h18v04	WL	266	0.6/0.3/0.2/0.4
HU-Bug	19.6E	46.69N	8.98	h19v04	CRO	34	1.0/0.4/—/0.6
HU-Mat	19.73E	47.85N	4.32	h19v04	CRO	102	0.0/0.5/0.6/0.7
IE-Ca1	6.92W	52.86N	11.2	h17v03	GRA	30	0.2/0.2/0.1/—
IL-Yat	35.05E	31.34N	4.24	h20v05	GRA	142	0.4/0.2/0.1/0.1
IT-Col	13.59E	41.85N	9.36	h19v04	CRO	97	0.1/0.2/0.4/0.5
IT-LMa	7.15E	45.58N	12.18	h18v04	GRA	51	0.1/0.1/0.5/—
IT-Lec	11.27E	43.3N	8.61	h18v04	CRO	95	0.8/0.4/0.4/0.4
IT-MBo	11.05E	46.02N	12.3	h18v04	MF	121	0.0/1.0/0.7/1.0

Table 3: Remaining 48 stations used to evaluate the gridded ET estimates. From left to right the station name; longitude; latitude; IGBP cover; distance in km to cell center; cell tile; number of days analysed; evaporative fraction for the DJF, MAM, JJA, SON 3-monthly periods.

Name	**Lon**	**Lat**	**km**	**Tile**	**Days**	**IGBP**	**EF**
IT-Non	11.09E	44.69N	7.91	h18v04	CRO	56	0.4/0.5/0.7/—
IT-Ro1	11.93E	42.41N	5.06	h18v04	CRO	260	0.1/0.3/0.4/0.4
IT-Ro2	11.92E	42.39N	5.66	h18v04	CRO	160	0.0/0.3/0.5/0.4
IT-Vig	8.85E	45.32N	8.68	h18v04	CRO	55	1.0/0.5/0.8/0.5
NL-Hor	5.07E	52.03N	12.48	h18v03	GRA	81	0.0/0.5/0.6/1.0
NL-Lan	4.9E0	51.95N	2.85	h18v03	GRA	36	0.3/0.6/0.6/1.0
NL-Loo	5.74E	52.17N	9.60	h18v03	CRO	168	0.5/0.3/0.3/0.4
PT-Esp	8.6W0	38.64N	5.90	h17v05	WL	39	0.4/0.3/0.2/0.2
PT-Mi1	8W	38.54N	15.06	h17v05	CRO	90	0.2/0.2/0.1/0.0
PT-Mi2	8.02W	38.48N	8.79	h17v05	CRO	233	0.8/0.4/0.2/0.3
SE-Faj	13.55E	56.27N	14.45	h18v03	MF	39	0.8/0.3/0.4/—
UK-Ham	0.86W	51.12N	3.51	h17v03	CRO	66	0.2/0.2/0.4/0.3
UK-PL3	1.27W	51.45N	12.69	h17v03	CRO	193	0.0/0.3/0.5/0.9
US-ARM	97.49W	36.61N	9.60	h10v05	CRO	155	0.6/0.5/0.3/0.3
US-Arb	98.04W	35.55N	11.64	h10v05	GRA	45	0.2/0.4/0.7/0.5
US-Aud	110.51W	31.59N	11.53	h08v05	GRA	228	0.8/0.1/0.2/0.7
US-Bkg	96.84W	44.35N	11.26	h11v04	CRO	195	1.0/0.8/0.8/1.0
US-Blo	120.63W	38.9N	10.81	h08v05	ENF	155	0.6/0.4/0.6/0.6
US-FPe	105.1W	48.31N	13.31	h10v04	GRA	188	0.0/0.4/0.3/0.6
US-Fmf	111.73W	35.14N	7.09	h08v05	WL	99	0.1/0.3/0.3/0.3
US-Fuf	111.76W	35.09N	8.98	h08v05	WL	69	0.1/0.2/0.4/0.2
US-Fwf	111.77W	35.45N	9.21	h08v05	OSH	119	0.2/0.2/0.4/0.4
US-Goo	89.87W	34.25N	1.80	h10v05	WL	208	0.9/0.6/0.5/0.7
US-IB1	88.22W	41.86N	12.46	h11v04	GRA	165	1.0/0.4/0.6/0.4
US-IB2	88.24W	41.84N	12.01	h11v04	GRA	218	0.3/0.4/0.7/0.5
US-Los	89.98W	46.08N	9.39	h11v04	MF	83	0.1/0.2/0.5/0.4
US-MOz	92.2W	38.74N	11.47	h10v05	CRO	286	0.3/0.4/0.5/0.5
US-NC1	76.71W	35.81N	13.24	h11v05	CRO	169	0.5/0.5/0.8/0.7
US-NC2	76.67W	35.8N	9.96	h11v05	CRO	240	0.5/0.5/0.8/0.8
US-SO2	116.62W	33.37N	10.99	h08v05	OSH	133	0.7/0.4/0.2/0.2
US-SO3	116.62W	33.38N	10.77	h08v05	OSH	89	0.9/0.4/0.2/0.2
US-SO4	116.64W	33.38N	9.01	h08v05	OSH	153	0.4/0.3/0.2/0.2
US-SP1	82.22W	29.74N	9.85	h10v06	WL	182	0.3/0.3/0.4/0.4
US-SRM	110.87W	31.82N	5.93	h08v05	OSH	159	0.2/0.1/0.1/0.4
US-Syv	89.35W	46.24N	13.7	h11v04	MF	163	0.1/0.2/0.5/—
US-WCr	90.08W	45.81N	11.4	h11v04	MF	251	0.1/0.2/0.6/0.4
US-Wkg	109.94W	31.74N	5.14	h08v05	OSH	154	0.2/0.1/0.1/0.3
US-Wrc	121.95W	45.82N	10.53	h09v04	ENF	74	0.3/0.3/0.3/—
US-NC2	76.67W	35.8N	9.96	h11v05	CRO	240	0.5/0.5/0.8/0.8
US-SO2	116.62W	33.37N	10.99	h08v05	OSH	133	0.7/0.4/0.2/0.2
US-SO3	116.62W	33.38N	10.77	h08v05	OSH	89	0.9/0.4/0.2/0.2
US-SO4	116.64W	33.38N	9.01	h08v05	OSH	153	0.4/0.3/0.2/0.2
US-SP1	82.22W	29.74N	9.85	h10v06	WL	182	0.3/0.3/0.4/0.4
US-SRM	110.87W	31.82N	5.93	h08v05	OSH	159	0.2/0.1/0.1/0.4
US-Syv	89.35W	46.24N	13.7	h11v04	MF	163	0.1/0.2/0.5/—
US-WCr	90.08W	45.81N	11.4	h11v04	MF	251	0.1/0.2/0.6/0.4
US-Wkg	109.94W	31.74N	5.14	h08v05	OSH	154	0.2/0.1/0.1/0.3
US-Wrc	121.95W	45.82N	10.53	h09v04	ENF	74	0.3/0.3/0.3/—

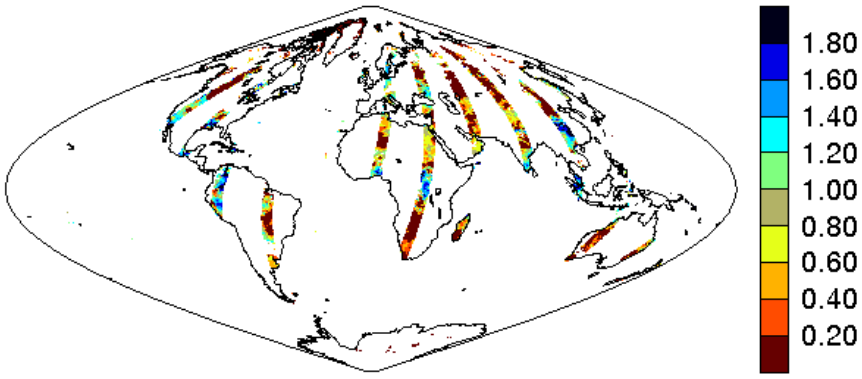

Figure 5: SEBS ET in mm/3-hours at the AATSR overpass of ∼10 am for DOY 125 in 2007.

project, but MOD16 uses different GMAO-MERRA and MODIS based products as model forcings.

- The **PT-JPL** product driven with SRB surface radiation, GIMMS NDVI, and CRU meteorological data [Fisher et al., 2008], produced as monthly estimates at 1^o. This is the basis of the PT model implemented in the project.
- The **GLEAM v2A** product, daily at 0.25^o[Miralles et al., 2011b,a], which is the basis of the 3-hourly GLEAM run by the project. It uses ERA-Interim surface radiation and CPC-unified precipitation, and ISCCP+AIRS air temperature (soil moisture and vegetaion optical depth with the same inputs applied here).
- The **LandFlux-EVAL** benchmark data set [Mueller et al., 2013], a multi-model ensemble including observation-based data sets, land-surface model data and re-analyses data, at $0.5^o$and monthly.
- The **ERA-interim** reanalysis latent fluxes from the forecast fields at 3-hourly and ∼75 km.
- The **MPI-MTE** product, a global empirical upscaling of FLUXNET eddy covariance observations [Jung et al., 2009].

6.4 ET for basin-integrated evaluation

Globally distributed long-term and annually resolved river discharge data for basins larger than 10000 km^2 with corresponding watershed boundaries were obtained from the Global Runoff Data Centre. Watershed boundaries were used

to extract net radiation from CERES, mean annual precipitation for the target period 2005-2007 from the GPCP, and GPCC v6 products. Two versions of GPCC v6 were processed by applying relative gauge correction factors according to Fuchs et al. [2001] and Legates and Willmott [1990] to the native GPCC products as recommended by the producers.

River discharge was converted from m3/s to mm/yr based on the GRDC reported basin area - we excluded basins where the absolute difference between GRDC reported area and area calculated from basin boundaries exceeded 25%. We further discarded basins where the uncertainty of the precipitation products is large by filtering out those with an on average very low density of rain gauges (< 0.1 per 0.5 degree gridcell) or frequent snow fall (> 25 days per year; based on CLOUDSAT). Few basins were further excluded where ETP-Q < 0 or LEP-Q $>$ Rn. Discharge data were often not available for the years 2005-2007. The introduced uncertainty of mean annual watershed ET due to a different reference periods of runoff and precipitation is small in comparison to the uncertainties of precipitation and basin areas. Nevertheless, we used a robust regression approach between GPCC precipitation and measured runoff for the overlapping period (which is always at least 10 years) to estimate runoff for the missing years to account for the possibility of a trend in runoff due to a trend in precipitation.

Few basins were further excluded where ETP-Q < 0 or LEP-Q $>$ Rn. Because few basins showed still implausible runoff values we trained a machine learning algorithm (Random Forests, Breimann 2001)) to predict mean annual runoff from mean annual precipitation and mean annual net radiation, and removed outlier basins. Outliers were identified according to a common procedure by checking if the absolute predicted minus observed runoff value exceeded three time the inter-quartile range of residuals.

The resulting record of 900 basins were clustered in 30 classes based on precipitation (log transformed), net radiation, and evaporative fraction (EF = ET/Rn) to reduce noise and retain clear patterns for the evaluation. The clustering algorithm was k-means with the cityblock distance, and variables were transformed to zero mean and unit variance. Each of the 30 classes was assigned to one of four groups based on the cluster centroids by splitting into extra-tropical and tropical using a threshold of Rn = 2500 MJ/m^2/yr and into moist and dry respectively using a threshold of EF = 0.5.

7 Summary

References

JB Fisher, KP Tu, and DD Baldocchi. Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sens. Envir.*, 112(3):901–919, 2008.

Thomas Foken. 50 Years of the Monin–Obukhov Similarity Theory. *Boundary-Layer Meteorology*, 119(3):431–447, July 2006.

T. Fuchs, J. Rapp, F. Rubel, and B. Rudolf. Correction of Synoptic Precipitation Observations due to Systematic Measuring Errors with Special Regard to Precipitation. *Phases. Phys.Chem. Earth (B)*, 26(9):689–693, 2001.

M. Jung, M. Reichstein, and A. Bondeau. Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences Discussion*, 6(3):5271–5304, 2009.

D.R. Legates and C.J. Willmott. Mean seasonal and spatial variability in gauge-cor- rected, global precipitation. *Int. J. Climatol*, 10:11q–127, 1990.

D.G. Miralles, R.A.M. De Jeu, J.H. Gash, T.R.H. Holmes, and A.J. Dolman. Magnitude and variability of land evaporation and its components at the global scale. *Hydrology and Earth System Sciences*, 15(3):967–981, March 2011a.

D.G. Miralles, T.R.H. Holmes, R.A.M. De Jeu, J.H. Gash, A.G.C. Meesters, and A.J. Dolman. Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth System Sci.*, 15(2):453–469, February 2011b.

Q. Mu, M. Zhao, and S.W. Running. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Envir.*, 115(8):1781–1800, January 2011.

B. Mueller, M. Hirschi, C. Jimenez, P. Ciais, P. A. Dirmeyer, A. J. Dolman, J. B. Fisher, M. Jung, F. Ludwig, F. Maignan, F. Miralles, M. McCabe, M. Reichstein, J. Shefield, K. C. Wang, E. F. Wood, Y. Zhang, , and S.I. Seneviratne. Benchmark products for land evapotranspiration: LandFlux-EVAL multi- dataset synthesis. *HESS*, 10:doi:10.5194/hessd 107692013, 2013.

S. W. Seemann, E. E. Borbas, R. O. Knuteson, G. R. Stephenson, and H.-L. Huang. Development of a global infrared surface emissivity database for application to clear sky retrievals from multispectral satellite radiance measurements. *J. Appl. Meteorol. Climat.*, 47:108–123, 2008.

Z. Su. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrology and Earth System Sciences*, 6(1):85–99, 2002.

Z. Su, R.A. Roebeling, J. Schulz, I. Holleman, V. Levizzani, W.J. Timmermans, H. Rott, N. Mognard-Campbell, R. De Jeu, and W. Wagner. Observation of Hydrological Processes Using Remote Sensing. *Treatise on Water Science, edited by: Wilderer, P., Academic Press, Oxford*, 2:351–399, 2011.

R.K. Vinukollu, E.F. Wood, C.R. Ferguson, and J.B. Fisher. Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches. *Remote Sens. Envir.*, 115(3):801–823, January 2011.

K. Wang, Z. Wan, P. Wang, M. Sparrow, J. Liu, X. Zhou, and S. Haginoya. Estimation of surface long wave radiation and broadband emissiv- ity using Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature/emissivity products. *JGR*, 110(D11):DOI:10.1029/2004JD005566, 2005.