JOURNAL OF GEOPHYSICAL RESEARCH, VOL. ???, XXXX, DOI:10.1029/,

Global inter-comparison of 12 land surface heat flux estimates

C. Jiménez,[1] C. Prigent,[1] B. Mueller,[2] S. I. Seneviratne,[2] M. F. McCabe,[3] E. F. Wood,[4] W. B. Rossow,[5] G. Balsamo,[6] A. K. Betts,[7] P. A. Dirmeyer,[8] J. B. Fisher,[9] M. Jung,[10] M. Kanamitsu,[11] R. H. Reichle,[12] M. Reichstein,[10] M. Rodell,[13] J. Sheffield,[4] K. Tu,[14] K. Wang,[15]

C. Jiménez, Laboratoire d'Etudes du Rayonnement et de la Matière en Astrophysique, Centre National de la Recherche Scientifique, Observatoire de Paris, Paris, France (carlos.jimenez@obspm.fr)

[1]Laboratoire d'Etudes du Rayonnement

Abstract.

A global inter-comparison of 12 monthly mean land surface heat flux products for the period 1993-1995 is presented. The inter-comparison includes some of the first emerging global satellite-based products (developed at Paris Observatory, MPI for Biogeochemistry, University of California Berkeley, University of Maryland, and Princeton University) and examples of fluxes produced by reanalyses (ERA-Interim, MERRA, NCEP-DOE) and off-line land surface models (GSWP-2, GLDAS CLM/Mosaic/Noah). An inter-comparison of the global latent heat flux ($\mathrm{Q_{le}}$) annual means shows a spread of $\sim$20 $W\,m^{-2}$ (all-product global average of $\sim$45 $W\,m^{-2}$). A similar spread is observed for the sensible ($\mathrm{Q_h}$) and net radiative ($\mathrm{R_n}$) fluxes. In general, the products correlate well with each other, helped by the large seasonal variability and common forcing data for some of the products. Expected spatial distributions related to the major climatic regimes and geographical features are reproduced by all products. Nevertheless, large $\mathrm{Q_{le}}$ and $\mathrm{Q_h}$ absolute differences are also observed. The fluxes were spatially averaged for 10 vegetation classes. The larger $\mathrm{Q_{le}}$ differences were observed for the rain forest, but when normalized by mean fluxes the differences were comparable to other classes. In general, the correlations between $\mathrm{Q_{le}}$ and $\mathrm{R_n}$ were higher for the satellite-based products compared with the reanalyses and off-line models. The fluxes were also averaged for 10 selected basins. The seasonality was generally well

et de la Matière en Astrophysique, Centre

captured by all products, but large differences in the flux partitioning were observed for some products and basins.

National de la Recherche Scientifique,

1. Introduction

Land surface heat fluxes are essential components of the water and energy cycles and govern the interactions between the Earth surface and the atmosphere [e.g., *Betts et al.*, 1996]. Variables such as cloud cover, precipitation, surface radiation, or air temperature and humidity, which are related to the atmospheric synoptic patterns and meso-scale structures, strongly influence the fluxes. In turn, the energy balance at the surface and its partitioning between the turbulent sensible (Q_h) and latent (Q_{le}) heat fluxes (here collectively referred as Q) also affect the atmosphere, determining the development of the atmospheric boundary layer [e.g., *Viterbo and Beljaars*, 1995]. Over land, energy balance and flux partitioning are complex mechanisms, with strong variability in both space and time, across climates and ecosystems, and in relation to the physical properties of the surface, especially moisture availability and vegetation. *In situ* measurements of land surface heat fluxes are available from field experiments (e.g., the Boreal Ecosystem-Atmosphere Study (BOREAS) [*Sellers et al.*, 1997]) and from some flux tower networks (e.g. FLUXNET [*Baldocchi et al.*, 2001]), but in order to obtain global, consistent estimates of Q a transition to satellite remote sensing is needed. The challenge is that heat fluxes produce neither absorption nor emission of electromagnetic signals directly. Therefore, observations related to surface temperature, soil moisture, or vegetation have to be combined with an interpretive model to derive the fluxes.

The currently available datasets were grouped based on the degree of complexity of the model used to derive Q. A first group includes the estimates derived from relatively

Observatoire de Paris, Paris, France.

simple models dedicated primarily to deriving the fluxes using remote sensing and meteorological inputs. Different methodologies exist, including empirical models that link the remote sensing observations to measured or modelled fluxes [e.g., *Wang et al.*, 2007b; *Jiménez et al.*, 2009], schemes using remotely sensed land surface temperature as the main boundary condition of a surface energy balance model [e.g., *Su*, 2002; *Anderson and Kustas*, 2008], or algorithms based on the equations predicting the main evapotranspiration processes [e.g., *Nishida et al.*, 2003; *Leuning et al.*, 2008]. Despite a large body of work, it is only recently that this capability has started to be adopted at the global scale. Difficulties arise from the fact that even relatively simple parameterizations may require large amounts of ancillary data that are not available globally (such as surface roughness to characterize heat transfer processes, or surface meteorological data to drive evaporation processes), making it difficult to extend from the local or regional scale to the global scale. In fact, at the moment most methodologies cannot solely rely on remote sensing observations, so that datasets derived from meteorological *in situ* measurements [e.g., *Fisher et al.*, 2008] or analyses [e.g., *Mu et al.*, 2007; *Gellens-Meulenberghs et al.*, 2007] are also needed to provide the required inputs to the models. Nevertheless, clear progress has been made in the recent years, and first global estimates of Q are now available [e.g., *Fisher et al.*, 2008; *Wang and Liang*, 2008; *Jiménez et al.*, 2009] . These estimates are referred to here as satellite-based products, to emphasize the fact that their estimates are derived by relatively simple formulation/models relying to a large extent on diagnostic satellite observations.

---

[2]Institute for Atmospheric and Climate

A second group includes the Q estimates produced by more complex land surface models that are constructed to provide a more complete characterization of surface energy and water budget processes. The land surface model can be coupled with an atmospheric model that assimilates observational data, such as in the weather reanalyses [e.g., *Ek et al.*, 2003; *Balsamo et al.*, 2009], or can be forced off-line by observational or model data [e.g., *Dirmeyer et al.*, 2006; *Boone et al.*, 2009]. There is also work towards the assimilation of surface observations [e.g., *Rodell et al.*, 2004]. Before the emergence of the first global satellite-based products, the only source of Q with adequate time and space samplings came from the land surface models. However, inter-comparisons of the land surface model outputs showed very large differences, due to model parameterizations and forcings (e.g., the Project to Intercompare Land-Surface Parameterization Schemes (PILPS) [*Henderson-Sellers et al.*, 1995] and the Global Soil Wetness Project (GSWP) version 1 and 2 [*Entin et al.*, 1999; *Dirmeyer et al.*, 2006]). Land surface model parameterizations are often developed empirically and tuned to local conditions where the ancillary data needed to estimate the model parameters are measured [e.g., *Wilson et al.*, 2002; *Wright et al.*, 1995]. Some parameters, such as fractional vegetation cover or leaf area index, can be estimated from satellites, but many other parameters are derived from approximate relationships with vegetation, soil type, or climate regime. To aid the discussion, the estimates from the second group are further divided into two sub-groups, referred to here as "reanalyses" (the coupled land surface models) and "off-line models" (the land surface models forced off-line), even if it is clear that the reanalysis estimates also come from a land surface model, and that many off-line forcing datasets are based

Science, ETH Zurich, Zurich, Switzerland.

on reanalysis estimates (which are sometimes corrected towards observations not used in the reanalysis).

Evaluating global Q estimates is difficult. This is not specific for the fluxes, since other major components of the hydrological cycle, such as soil moisture or precipitation, are also difficult to evaluate [e.g., *Grubber and Levizzani*, 2008; *Prigent et al.*, 2005; *Seneviratne et al.*, 2010]. By using tower flux measurements, formulations and models can be evaluated at the tower scale by using a combination of the surface meteorology from the station and, if relevant, the satellite forcing (if the resolution is compatible with the tower measurements) [e.g., *McCabe and Wood*, 2006; *Su et al.*, 2007; *Cleugh et al.*, 2007; *Fisher et al.*, 2008; *Stöckli et al.*, 2008]. The tower data representativity and quality should also be considered [e.g., *Williams et al.*, 2009]. Once the models are driven by global datasets, an evaluation with tower fluxes is more questionable due to the scale miss-match between satellite retrievals, model outputs, and tower observations, and the coverage of the tower network. A qualitative examination of the fluxes, by checking the consistency displayed between the Q estimates and independent but related hydrological observations has also been proposed [e.g., *McCabe et al.*, 2008].

Global inter-comparison of Q between reanalyses [e.g., *Betts et al.*, 2006; *Bosilovich et al.*, 2009], off-line models forced with the same datasets [e.g., *Schlosser and Gao*, 2009], or climate model simulations [e.g., *Lim and Roderick*, 2009] have already been presented. To the best of our knowledge, no systematic inter-comparison that also includes satellite-based products at the global scale has yet been published. In the framework of the Global

[3]School of Civil and Environmental

Energy and Water Cycle Experiment (GEWEX) Radiation Panel (GRP) LandFlux activity, such inter-comparison has been initiated under the dedicated LandFlux-EVAL initiative. LandFlux aims at providing a framework for undertaking coordinated evaluation and assessment of the emerging global flux products, ultimately identifying and delivering a robust procedure for the operational production of a global land surface flux dataset to improve climate scale water and energy cycle characterization. Together with a paper by *Mueller et al.* (in preparation) focusing only on evapotranspiration estimates but for a larger number of products (including estimates from climate models participating in the Intergovernmental Panel on Climate Change Fourth Assessment Report (IPCC-AR4)), this publication presents the first results of the LandFlux-EVAL initiative. The period choosen for this analysis is 1993-1995 (1986-1995 in Mueller et al., in preparation), the final three years of the GSWP-2 exercise and the first three years of the estimates from Paris Observatory. Although analysis of shorter time scales would be desirable, at the moment most of the available global estimates from satellite-based products are limited to monthly averages by the time sampling of the available forcings. As satellite-based products, estimates provided by the University of California, the University of Maryland, Paris Observatory, Princenton University, and the Max Planck Institute for Biochemistry, are included in the inter-comparison. As reanalyses, estimates from the The Modern Era Retrospective-analysis for Research and Applications (MERRA), the National Centers for Environmental Prediction-Department of Energy (NCEP/DOE) reanalysis R-2, and the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis are considered. As off-line models, estimates from the multi-model ensemble

Engineering, University of New South

GSWP-2 and from the land surface models Mosaic, Noah and Community Land Model (CLM) driven by the Global Land Data Assimilation System (GLDAS) are presented. The choice of products is based on a desire to have a representative sample of different approaches. For the off-line models, the GSWP-2 multi-model ensemble is a representative example of multi-model outputs, while the GLDAS runs provide a good example of fluxes from individual models that were forced with the same datasets.

This paper focuses on an inter-comparison of the selected fluxes. There is no attempt to quantify the accuracy of the products, or to claim that one product is superior to the others. The goal is to highlight the differences between the products in order to evaluate the range of the existing global Q estimates. The paper is structured as follows. Section 2 presents the different modeling frameworks. Section 3 explains the spatial and temporal aggregation of the datasets that enables the inter-comparison. Sections 4.1 and 4.2 present the differences in the global yearly and seasonal Q averages. Sections 4.3 inter-compares spatially averaged fluxes for major vegetation classes. Section 4.4 inter-compares spatially averaged fluxes for a group of selected basins. Finally, Section 5 gives the summary and conclusions. The paper is complemented by a collection of additional figures, denoted in the text with a capital S. They can be found in the auxiliary material accompanying the paper.

2. Data

Wales, Sydney, Australia.

2.1. Satellite-based products

The different flux and auxiliary products are described below. The products will be referred to in the text by the first short name given after the product name (e.g., PA-OBS for Paris Observatory). To avoid cluttering in the legends, the products will be referred in the tables and figures as the second (smaller) short name (e.g. PAO for the given example). A summary of the products is given in Table 1.

2.1.1. University of California Berkeley [UCB,UCB]

$\mathrm{Q_{le}}$ is estimated from a bio-meteorological approach that translates Priestley-Taylor estimates of potential evapotranspiration into rates of actual evapotranspiration [*Fisher et al.*, 2008, 2009]. The method was evaluated at the local scale at 36 FLUXNET sites across 2 years, and has been extended to estimate global $\mathrm{Q_{le}}$ by forcing the model with the International Satellite Land Surface Climatology Project, Initiative II (ISLSCP-II) datasets [*Hall et al.*, 2006]. Main inputs are the radiative fluxes ($\mathrm{R_n}$) from the GEWEX Surface Radiation Budget (GEWEX-SRB) [*Stackhouse et al.*, 2004], maximum air temperature and vapor pressure from the Climate Research Unit (CRU) [*New et al.*, 1999, 2000], and a vegetation characterization using the Advance Very High Resolution Radiometer (AVHRR) reflectances [*Gutman*, 1999; *Huete*, 1998] processed as the Fourier-Adjusted, Sensor and Solar zenith angle corrected, Interpolated, Reconstructed (FASIR) Normalized Difference Vegetation Index (NDVI) [*Los et al.*, 2000]. The spatial resolution is 0.5^ox0.5^o, and monthly averaged values in $mm\,month^{-1}$ are available from 1986 to 1995.

2.1.2. University of Maryland [UMD, UMD]

---

[4]Department of Civil and Environmental

Q_{le} is estimated from a statistical approach that locally relates (by linear regression) R_n, near-surface air temperature, surface temperature, and a vegetation index with observed Energy Balance Bowen Ratio (EBBR) Q_{le} at eight sites over the Southern Great Plains [*Wang and Liang*, 2008]. The method was evaluated at local scale at AmeriFlux stations across 4 years, and extended to estimate global Q_{le} by forcing the model with ISLSCP-II datasets. Inputs are the R_n (GEWEX-SRB), daily averaged and diurnal range of the air temperature (CRU), and a vegetation index from AVHRR reflectances. An improved model that explicitly includes the impact of wind speed and water vapor pressure deficit to improve its capability in modeling climate variability of Q_{le} has just been developed [*Wang et al.*, 2010a, b], but the estimates included here correspond to the model presented in *Wang and Liang* [2008]. The spatial resolution is 1^ox1^o, and monthly mean values in $W\,m^{-2}$ from 1986 to 1995 are available.

2.1.3. Paris Observatory [PA-OBS, PAO]

Q_{le} and Q_h are estimated from a statistical approach that globally relates (using non-linear regression) a suite of multi-frequency remote sensing observations with modeled fluxes from the GSWP-2 multi-model ensemble [*Jiménez et al.*, 2009]. The statistical model is driven by the following inputs: reflectances from AVHRR, land surface temperature and its diurnal cycle from the International Satellite Cloud Climatology Project (ISCCP) [*Rossow and Schiffer*, 1999; *Aires et al.*, 2004], active microwave backscatter from the European Remote-sensing Satellite (ERS) scatterometer [*Francis et al.*, 1991; *Frison and Mougin*, 1996], and passive microwave emissivities from the Special Sensor Microwave/Imager (SSM/I) [*Hollinger et al.*, 1987; *Prigent et al.*, 2006] . The approach

Engineering, Princeton University,

was evaluated at local scale at AmeriFlux stations. The spatial resolution is 0.25^ox0.25^o (at the equator), and monthly mean values in $W\,m^{-2}$ of $\mathrm{Q_{le}}$ and $\mathrm{Q_h}$ are currently available from 1993 to 1999.

2.1.4. Princeton University [PRU, PRU]

$\mathrm{Q_{le}}$ is estimated from a modified version of the Penman-Monteith algorithm described in *Sheffield et al.* [2009]. For global application, the formulation is driven by ISCCP $\mathrm{R_n}$ and near surface air and surface temperature, reanalysis wind speed [*Sheffield et al.*, 2006], and vegetation characterization from AVHRR reflectances. The approach was evaluated over Mexico with global forcings down-scaled for this region using data from the North American Regional Analysis (NARR) [*Mesinger et al.*, 2006]. The spatial resolution is 2.5^ox2.5^o, and daily mean values in $mm\,day^{-1}$ are available from 1986 to 2006.

2.1.5. Max Planck Institute for Biogeochemistry [MPI-BGC, MPI]

$\mathrm{Q_{le}}$ and $\mathrm{Q_h}$ are estimated by a global upscaling of eddy covariance (EC) measurements from FLUXNET by a machine learning approach called model tree ensembles (MTE) [*Jung et al.*, 2009]. The EC measurements used are part of the FLUXNET LaThuille synthesis data set, which was established by a standard processing according to *Reichstein et al.* [2005] and *Papale et al.* [2006] and comprises ∼950 years of data from ∼250 sites. The EC measurements are corrected to force energy balance closure on a monthly time scale. The global upscaling is driven by a long-term monthly fraction of absorbed photosynthetically active radiation (fAPAR) dataset (established by harmonizing AVHRR NDVI data [*Vermote and Saleous*, 2005] with fAPAR from SeaWiFS [*Gobron et al.*, 2006] and fAPAR from MERIS [*Gobron et al.*, 2008]), near surface air temperature from CRU,

---
Princeton New Jersey, USA.

precipitation data from the Global Precipitation Climatology Center (GPCC) [*Rudolf and Schneider*, 2005], and an estimation of the top of the atmosphere shortwave radiation. The product was evaluated against river runoff data and the GSWP-2 multi-model ensemble. The spatial resolution is 0.5^ox0.5^o, and monthly mean values in $W\,m^{-2}$ are available from 1982 to 2008. For simplicity this product is included as a satellite-based product, but notice that this product is to a large extent based on in-situ datasets.

2.2. Reanalyses

2.2.1. ERA Interim reanalysis [ERA-INT, ERA]

ERA-Interim is a new global reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF) [*Simmons et al.*, 2006], focusing on the data-rich period since 1989. The ERA-Interim system is based on a recent release of the Integrated Forecasting System (IFS Cy31r2), released operationally in September 2006, containing many improvements both in the forecasting model and analysis methodology. The surface fluxes in ERA-Interim are based on the land surface model TESSEL (Tiled ECMWF Surface Scheme for Exchange over Land, [*van den Hurk et al.*, 2000]) forced by atmospheric analysis and short range forecasts. A land data assimilation constrains the model fields on the basis of short range forecast errors: soil moisture and soil temperature are corrected using air temperature and relative humidity observations from SYNOP stations [*Douville et al.*, 2000]; snow mass errors are constrained by SYNOP snow depth reports and satellite snow cover data [*Drusch et al.*, 2004]. The fluxes were obtained as monthly mean

[5] NOAA-Cooperative Remote Sensing

values in $W\,m^{-2}$ at a resolution of $3/4^o$ x $3/4^o$ (very close to the native ERA-Interim T255 Gaussian reduced grid).

2.2.2. MERRA reanalysis [MERRA, MER]

The Modern Era Retrospective-analysis for Research and Applications (MERRA) is a National Aeronautics and Space Administration (NASA) reanalysis for the satellite era using a major new version of the Goddard Earth Observing System Data Assimilation System Version 5 (GEOS-5) [*Bosilovich*, 2008]. The project focuses on historical analyses of the hydrological cycle on a broad range of weather and climate time scales and places the NASA EOS suite of observations in a climate context. The monthly flux averages in $W\,m^{-2}$ were downloaded from the MERRA data archive at a spatial resolution of $1/2^o$x$2/3^o$, covering 1979 to present. Q and R_n were respectively extracted from the FLX and RAD collections, meaning that fluxes from inland water are also counted in the pixel estimate. Q can also be extracted from the LND collection, where only the fluxes coming from land are counted. For consistency with the other reanalyses estimates used here the FLX fluxes are included.

2.2.3. NCEP-DOE reanalysis (R-2) [NCEP-DOE, NCE]

The National Centers for Environmental Prediction-Department of Energy (NCEP-DOE) Reanalysis 2 is an improved version of the NCEP-National Center for Atmospheric research (NCEP-NCAR) Reanalysis I model [*Kalnay et al.*, 1996] that fixed errors and updated parameterizations of physical processes [*Kanamitsu et al.*, 2002]. Unlike the NCEP/NCAR reanalysis, NCEP/DOE reanalysis utilizes pentad mean observed precipitation to correct model precipitation in driving the soil model, which made the evolution

Science and Technology Center, City

of soil moisture more realistic [*Lu et al.*, 2007]. Users of the NCEP reanalysis are warned that variables such as heat fluxes, humidity, or surface temperature should be interpreted with caution, as there are no assimilated observations to directly affect these variables. NCEP-DOE Reanalysis 2 fluxes were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, as daily averages in $W\,m^{-2}$ at a resolution of $\sim 2.0^o$ x 2.0^o (T62 Gaussian grid, 192x94), and are available from 1979.

2.3. Off-line models

2.3.1. GSWP-2 modeling exercise [GSWP-MMA, GSW]

GSWP is an international modeling research activity with the main goal of producing global datasets of soil moisture, other state variables, and related hydrological quantities using state-of-the-art land surface models. In the second phase of the project (GSWP-2)[*Dirmeyer et al.*, 2006], 15 land surface models driven in off-line mode using global meteorological forcing inputs produced daily land fluxes and related surface variables for 10 years (1986-1995) at a resolution of 1^ox1^o. The model forcing, vegetation, and soil cover were primarily extracted from the ISLSCP-II initiative, though work was undertaken to hybridize the reanalyses data with observational data in order to remove systematic errors [*Zhao and Dirmeyer*, 2003]. The flux estimates compared here are the multi-model ensemble monthly averages in $W\,m^{-2}$ publically available at the GSWP web site. In *Guo and Dirmeyer* [2006], the GSWP-2 multi-model analysis resulting from a simple average across the individual models gave the best overall results when evaluating the modeled soil moisture outputs. This model ensemble is described as an analog to the atmospheric

College of New York, New York, USA.

reanalyses, and judged as the best approach to combine the models, compared with more sophisticated combinations, in the absence of calibration data [*Dirmeyer et al.*, 2006].

2.3.2. GLDAS [GLDAS-Noah/CLM/Mosaic, NOA, CLM, MOS]

The Global Land Data Assimilation System (GLDAS) [*Rodell et al.*, 2004] drives multiple off-line land surface models, integrating a large quantity of observation based data enabled by the Land Information System (LIS) [*Kumar et al.*, 2006]. Currently, GLDAS drives four land surface models (Mosaic, Noah, the Community Land Model (CLM), and the Variable Infiltration Capacity (VIC)), forcing them with satellite derived precipitation and radiation data and atmospheric analysis model outputs. For the inter-comparison the 1^ox1^o monthly averages in $W\,m^{-2}$ from Noah (version 2.7), CLM (version 2.0) and Mosaic were downloaded. The VIC outputs were not included in this analysis as it was run in "water balance mode", without fully solving the surface energy balance, meaning that R_n was not available. The radiative downward forcing for 1993 comes from the ERA-15 reanalysis, but for 1994-1995 from the NCEP/NCAR reanalysis R1, both bias corrected with GEWEX-SRB [*Berg et al.*, 2003].

2.4. Auxiliary products

2.4.1. Precipitation

The Global Precipitation Climatology Project (GPCP) merges rain gauges, satellite geostationary and low-orbit infrared, passive microwave, and sounding observations to estimate monthly rainfall on a 1^ox1^o global grid from 1979 to the present [*Adler et al.*,

[6] European Center for Medium-Range

2003]. Monthly averaged precipitation amount in $cm\,month^{-1}$ (version 2.1) is used in the analysis. These estimates are likely to be different from some of the precipitation generated by the atmospheric reanalyses or prescribed to the off-line models, and most of the satellite-based products do not use precipitation as an observational input. Consequently, the GPCP estimates will only be used to give an approximated idea of the different precipitation regimes, not to compare evaporation/precipitation regimes across the different products.

2.4.2. Snow

A snow mask is obtained from a combination of National Snow and Ice Data Center (NSIDC) data. The snow mask is derived from the NSIDC Northern Hemisphere EASE-Grid Weekly Snow Cover and Sea Ice Extent Version 3 [*Armstrong and Brodzik*, 2005] and the weekly Southern Hemisphere snow flag stored by ISCCP (derived from NSIDC data).

2.4.3. Surface water

A globally applicable remote-sensing technique employing a suite of complementary satellite observations has been developed to estimate spatial and temporal dynamics of surface water extent [*Prigent et al.*, 2001b; *Papa et al.*, 2010]. This dataset has been generated from several satellite instrument types: passive microwave (SSM/I), scatterometer (ERS), and visible and near-IR (AVHRR). It will be used here to identify regions with a likely presence of inland water.

2.4.4. Vegetation

Weather Forecasts, Reading, UK.

The vegetation and land use dataset of *Matthews* [1983] will be used to classify the flux estimates into 10 vegetation classes. The *Matthews* [1983] classification distinguishes 30 classes of natural vegetation, and is associated to a land use dataset that distinguishes 5 levels of cultivation intensity. The version used here is a simplified classification compiled in *Prigent et al.* [2001a], where the original classes are re-grouped into 9 natural vegetation classes and one cultivation class. This classification is likely to differ from the land cover masks employed in some of the data products, and it will be used only for an approximate separation of the estimates into vegetation types.

2.4.5. Basins

The template of the major river basins from TRIP (Total Runoff Integrating Pathways) [*Oki and Su*, 1998] is adopted in this study to delineate the spatial extension of a group of selected basins. The selected basins correspond to the rivers Amazon, Mississippi and Parana (America); Danube, Volga and Yantgze (Eurasia); Nile, Niger and Congo (Africa); and Murray (Australia).

3. Methodology

To make the inter-comparison possible the different products have been aggregated to a common spatial and temporal resolution. First, the spatial resolution of the products has been downgraded to the 2.5^ox2.5^o resolution of the coarser product (PRU) by spatially averaging the original estimates. Next, the products are space-matched, i.e., only pixels having fluxes from all products are retained. Finally, the products are time matched: only

[7]Atmospheric Research, Pittsford,

pixels having fluxes for all months, years and products are kept. This guarantees that differences in the statistics are not due to different spatial coverage or time period. After these operations, for each month ∼2600 pixels at the 2.5^ox2.5^o resolution are compared. This represents ∼70% of the total land surface, with most of the missing pixels over Greenland and Northern Africa. This implies that the reported globally averaged fluxes will not be truly global (although for simplicity they will be refered as global).

During the analysis, estimates of $\mathrm{Q_{le}}$, $\mathrm{Q_h}$, $\mathrm{R_n}$, and the evaporative fraction (EF) are compared. Strictly speaking EF is defined as $\mathrm{Q_{le}}$ / $\mathrm{Q_{le}}$ + $\mathrm{Q_h}$, but the ratio $\mathrm{Q_{le}}$ / $\mathrm{R_n}$ is used here as only $\mathrm{Q_{le}}$ and $\mathrm{R_n}$ is reported by some of the products. Assuming the surface energy balances, i.e., $\mathrm{R_n} = \mathrm{Q_{le}} + \mathrm{Q_h} + \mathrm{Q_g}$, where $\mathrm{Q_g}$ is the ground heat flux, the difference between both expressions depends on the magnitude of $\mathrm{Q_g}$. At monthly time steps $\mathrm{Q_g}$ is generally a small fraction of $\mathrm{R_n}$. However, $\mathrm{Q_g}$ estimates as large as ∼15 $W\,m^{-2}$ are reported at some winter locations by some of the products considered here. This implies that the EF presented here may differ from the EF reported elsewhere for some of the products compared. Notice also that when $\mathrm{Q_{le}}$ and $\mathrm{R_n}$ are very small and/or when they take negative values (e.g., for winter conditions in some regions) the $\mathrm{Q_{le}}$ to $\mathrm{R_n}$ ratio can be well outside the 0 to 1 interval. For those situations the EF will not be reported.

When spatial and/or time averages are required for $\mathrm{Q_{le}}$, $\mathrm{Q_h}$, and $\mathrm{R_n}$, they are estimated by calculating the mean. For the EF, the spatial and/or time $\mathrm{Q_{le}}$ and $\mathrm{R_n}$ means are first calculated, and their ratio given as an estimate of the EF average. For those products where $\mathrm{Q_h}$ is not directly provided (UCB, UMD, PRU), $\mathrm{Q_h}$ is derived by assuming the

Vermont, USA.

surface energy balance. For PRU an estimate of Q_g is given and is included in the energy balance. For UCB and UMD Q_g is not given and assumed here to be zero at the monthly scale. Differences in Q_h between UCB and UMD and the other products could then be related to the fact a zero Q_g is assumed here for UCB and UMD.

For MPI-BGC and PA-OBS an estimate of R_n is not given. Contrary to some of the other satellite-based products where the model partitions R_n into its different flux components, MPI-BGC Q_{le} and Q_h come from a global upscaling of EC measurements that does not require a R_n product. Here the sum $Q_{le} + Q_h$ is used as an approximation of the MPI-BGC R_n. For PA-OBS the situation is different. PA-OBS uses the ISCCP R_n product as an input, but the empirical model is adjusted to reproduce the GSWP Q_{le} and Q_h. This means that on average the PA-OBS Q are consistent with the GSWP R_n, and the ISCCP R_n cannot be used to analyze the partioning of the fluxes (see *Jiménez et al.* [2009] for more details). As for MPI-BGC, the sum $Q_{le} + Q_h$ is used as an approximation of PA-OBS R_n. Some of the differences in R_n between MPI-BGC and PA-OBS and the other products can be related to this approximation.

Most of the Q_{le} estimates were available as monthly averages expressed in $W\,m^{-2}$ and no time-averaging and/or unit-conversion were required. The exceptions were UCB and PRU, where Q_{le} was converted from water depths to $W\,m^{-2}$ mutiplying by the latent heat of vaporization (constant value of 2.45 $MJ\,kg^{-1}$) and dividing by the respective time integration (month and day, respectively). The PRU were further time averaged to get the monthly means. All products were also annually averaged for 1993 and 1994 by calculating

[8]Center for Ocean-Land-Atmosphere

for each geographical pixel the average of the 12 monthly means. For 1995 the products UCB and UMD do not have fluxes for November and December (although a climatological value was used to make the product aggregation for these 2 months possible), and the

Studies, Calverton, Maryland, USA.

[9]Environmental Change Institute, School of Geography and the Environment, Oxford University, Oxford, UK.

[10]Max Planck Institute for Biogeochemistry, Jena, Germany.

[11]Scripps Institution of Oceanography, University of California, San Diego, USA.

[12]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center,Greenbelt, Maryland, USA.

[13]Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

[14]Department of Integrative Biology, University of California, Berkeley, USA.

[15]Department of Geological Science, University of Texas at Austin, Austin, USA.

annual means are not calculated. When plotting monthly time series, the last two months of the year will be left empty for these two products.

During the analysis, an all-product ensemble mean and standard deviation will be displayed together with the individual fluxes in most of the figures. Notice that the objective of this is to highlight the dispersion in the fluxes, not to suggest that an all-product average is a possible outcome of the inter-comparison exercise. The term spread will be used in the text to refer to the difference between the maximum and minimum estimate in the all-product ensemble for a given spatial and/or time average.

4. Analysis

4.1. Comparing annual fluxes

4.1.1. Global fluxes

The 1994 global annual means of $\mathrm{Q_{le}}$, $\mathrm{Q_h}$, and $\mathrm{R_n}$ for the different products are plotted in Figure 1 (left panel). The panel plotting global $\mathrm{Q_{le}}$ versus $\mathrm{R_n}$ shows a spread of $\sim$20 $W\,m^{-2}$ ($\sim$15 $W\,m^{-2}$ if the NCEP-DOE estimate is excluded) for $\mathrm{Q_{le}}$ and $\mathrm{R_n}$, and a larger spread for $\mathrm{Q_h}$. The $\mathrm{Q_{le}}$ ensemble mean and standard deviation of the annual means are $\sim$45 $W\,m^{-2}$ and $\sim$6 $W\,m^{-2}$, respectively. As expected, there is some tendency of higher $\mathrm{Q_{le}}$ for larger $\mathrm{R_n}$, but with a much larger scatter than if all products were similarly partitioning $\mathrm{R_n}$. The reanalyses have the largest $\mathrm{Q_{le}}$ averages (apart from the satellite-based product UCB), but the same does not apply to the $\mathrm{Q_h}$ averages. From the off-line models, GLDAS-Noah and GLDAS-CLM have more similar $\mathrm{Q_{le}}$ and $\mathrm{Q_h}$ averages (compared with GLDAS-Mosaic), coinciding also with more similar $\mathrm{R_n}$. GSWP-MMA, PA-OBS, and MPI-BGC have closer fluxes, compared with the differences with the other products. This is expected for GSWP-MMA and PA-OBS, as PA-OBS fluxes are derived from an empirical model calibrated with GSWP-MMA fluxes. Similar differences are observed for the 1993 annual means (not shown).

The 1994 annual Q_{le} from the different products is plotted in Figure 2 (see Figures S1 to S3 in the auxiliary material for Q_h, R_n, and EF). In broad terms, the expected spatial structures related to the main climate regimes and topographical features are present in all products. Nevertheless, the absolute values of the fluxes can be quite different from one product to the other. In terms of spatial structures, MERRA Q_{le} and Q_h over the Tropics seem different compared to the others, with a sharp flux gradient around 10^{o}S. To highlight the differences, the 1994 all-product ensemble Q_{le}, Q_h, and R_n annual average and the absolute and relative standard deviation (from the 12 products annual means and normalized by the all-product ensemble average) are given in Figure 3. Globally, there is more variability in the derived Q_h than Q_{le}. Compared with Q_{le}, the absolute variability in R_n is larger, but as the absolute R_n values are in general larger than the values for Q_{le}, this results in a smaller relative standard deviation for R_n (i.e., in relative terms there is less variability in R_n). This is expected as some products share common downward or net radiative fluxes. In general the largest relative variability is observed in those regions where the fluxes are smaller (e.g., over deserts and mountainous regions for Q_{le}). Similar statistics are found for 1993 (not shown).

4.1.2. Precipitation regimes

Figure 1 (center and right panels) show the annual fluxes for two different precipitation regimes (using the GPCP estimates): a first one representing regions with high precipitation (P > 1700 $mm\, year^{-1}$), and a second one representative of drier ecosystems (500 > P > 1000 $mm\, year^{-1}$). For the averages when precipitation is high, most of the products have a more constant $\mathrm{Q_{le}}$ to $\mathrm{R_n}$ ratio (i.e., a closer flux partitioning) compared with the global averages, although NCEP-DOE, UMD, and PRU deviate more from the ratio shown by the other products. Notice that even if there is a larger spread of absolute annual $\mathrm{Q_{le}}$ averages for the high precipitation regime, Figure 3 shows that in relative terms the observed variability for some of these regions (e.g, the Amazonia) is comparable to the variability for some more drier regions (e.g. some southern regions in North America). For the drier regions, $\mathrm{Q_{le}}$, $\mathrm{Q_h}$, and $\mathrm{R_n}$ are smaller than for the wetter regions, as expected from the precipitation regime and radiation available at these regions, with the fluxes scattered in ranges similar to the scatter for the global averages.

The UCB and UMD products can be used to illustrate possible factors responsible for flux differences. For instance, for the high precipitation averages, UCB and UMD $\mathrm{Q_{le}}$ differ by $\sim$25 $W\,m^{-2}$. UCB and UMD use the same datasets for $\mathrm{R_n}$ and near-surface air temperature. However, soil moisture is characterized differently (water vapor pressure for UCB, diurnal air temperature range for UMD), and the models are very different (Priestley-Taylor formulation versus an empirical model). Furthermore, the UCB model includes a simple parameterization for evaporation from intercepted rain, which may be of importance for the regions with high canopy density [e.g., *Wang et al.*, 2007a], while it is not clear how the UMD model calibrated on Southern Great Plains EBBR fluxes accounts for the interception in high density canopy areas. All these differences in model and inputs may contribute to the $\mathrm{Q_{le}}$ differences.

4.1.3. Snow-covered regions

To further characterize the fluxes in regions of large variability, an example showing the annual averages of $\mathrm{Q_{le}}$ as a function of $\mathrm{R_n}$ for snow-covered regions in Dec-Jan-Feb (selected by using the NSIDC snow cover mask) and for the same regions in Jul-Aug-Sep is given in Figure 4. Snow-covered regions are difficult to characterize, both from models and observations [e.g., *Boone et al.*, 2004; *Cordisco et al.*, 2006; *Rutter et al.*, 2009]. A large spread (relative to their absolute values) both for $\mathrm{Q_{le}}$ and $\mathrm{R_n}$ can be observed in winter. In the absence of snow, the summer fluxes show expected larger values, with relatively close $\mathrm{Q_{le}}$ to $\mathrm{R_n}$ ratios among the products, apart from PRU, which has the smaller $\mathrm{Q_{le}}$ for the largest $\mathrm{R_n}$.

4.1.4. Impact of data aggregation

The aggregation to a common spatial and temporal resolution of the different products can have an impact on the inter-comparison. To see the impact of the grid selection, the global statistics were re-calculated after re-gridding all products onto the finest product grid (PA-OBS, an equal area grid of ~770 km^2 with a lat-lon box of ~0.25^ox0.25^o at the equator). A simple nearest-neighbor technique was used for the re-gridding in order to keep as much as possible the original spatial structures. Figure 5 (top panels) shows the global differences between the products aggregated into the fine grid. Comparison with similar plots in Figure 1 (estimates aggregated into the 2.5^ox2.5^o) show some differences (e.g., the global $\mathrm{Q_{le}}$ ensemble mean differ by ~2 $W\,m^{-2}$), but the relative differences between individual products are very similar. Another issue is the use of different land masks. For instance, even if only common land surface pixels are compared, for pixels with a mixture or land and water bodies the reanalyses or off-line model fluxes could have been estimated with different land/water partitions. Another problem is that the observational data could already have been integrating the land/water contributions (e.g., if land and water bodies are within the satellite footprint). This is likely to have a direct effect on the satellite-based products that depend more directly on the observational data. To have some idea about how this might be impacting the differences, the surface water product was used to select pixels that are unlikely to have a presence of inland or coastal water. In principle, those are pixels where potential differences related to these issues can be excluded. Figure 5 (bottom panels) shows the new differences. A comparison with Figure 1 shows that although the product averages change (expected as the geographical coverage has changed), the relative differences between products remain quite similar. These examples suggest that although the aggregation into a common spatial and temporal resolution has an effect on the analysis, it is unlikely to be responsible for a large part of the observed differences.

4.2. Comparing seasonal fluxes

4.2.1. Monthly fluxes

An example of monthly $\mathrm{Q_{le}}$ and $\mathrm{R_n}$ (August 1994) for the different products is given in Figures 6 and 7 (see Figures S4 to S10 for $\mathrm{Q_{le}}$, $\mathrm{Q_h}$, $\mathrm{R_n}$, and EF for February and August 1994 in the auxiliary material). As with the annual averages, the main geographical structures related to the main climatic regimes and geographical features are in general present in all products. Nevertheless, the differences in the $\mathrm{Q_{le}}$ absolute values can be large, e.g., the differences between PA-OBS and PRU for Northern Europe, or between ERA-INT and MERRA in South America. In the latter case, some of the differences can be traced back to the MERRA precipitation, which differ from standard gauge- and satellite-based products, and to details in the interception formulation in MERRA. Re-running the MERRA integrations in off-line mode with observation-corrected precipitation narrows the $\mathrm{Q_{le}}$ differences considerably (not shown). The $\mathrm{R_n}$ maps seem to be in better agreement, but large differences can also be observed (e.g., compare PA-OBS and GLDAS-Mosaic). Large differences in the partioning of the fluxes are also evident in the EF maps (e.g., see Figure S10 in the auxiliary material). Maps of the monthly mean $\mathrm{Q_{le}}$ and $\mathrm{R_n}$ differences with the all-product ensemble average for the same month are given in Figures 8 and 9. Large regions with $\mathrm{Q_{le}}$ and $\mathrm{R_n}$ differences (with respect to the ensemble mean) larger than 30 $W\,m^{-2}$ can be observed for some products.

To summarize the monthly mean statistics, Tables 2 and 3 report the mean difference and the root mean square (RMS) difference of the global 1994 product-to-product differences for $\mathrm{Q_{le}}$ and $\mathrm{R_n}$, respectively. The correlations between the different products are also given. The statistics are computed by including for each product the monthly fluxes for all the pixels and months, meaning that the correlations reflect both the spatial and temporal variations between the products. In general, the correlations for $\mathrm{Q_{le}}$ are high (values between 0.72 and 0.95). Some of the lowest correlations relate to the NCEP-DOE and MERRA reanalyses, which can be explained by some of the observed spatial structures. If the products are divided into satellite-based products, reanalyses, and off-line models, Table 2 suggests that the reanalyses $\mathrm{Q_{le}}$ presents the largest RMS difference among them (a high of 30.7 for the pairs NCEP-DOE and GSWP-MMA, and MERRA and GLDAS-CLM). Nevertheless, some of the mean differences for the satellite-based products are larger than the mean differences for the reanalyses (a high of 12.9 for the pair UCB and MPI-BGC). Table S3 (Section S4.2) shows the same statistics for $\mathrm{R_n}$, where the correlations are in generally higher. This is expected as some of the radiative forcings (e.g., the GLDAS off-line models, or UCB and UMD) are common. Nevertheless, some significant differences between some products are observed (e.g., between GSWP-MMA and UCB, or PRU and GLDAS-Mosaic). It should be clear that such differences in R_n limit a possible agreement between estimates of Q. At the same time, the R_n differences cannot be used to completely explain Q differences, since variation in the partitioning of the fluxes was observed in Figure 1.

4.2.2. Annual cycles

The 1994 Q_{le}, R_n, and EF global annual cycles are displayed in Figure 10. The Q_{le} annual cycles have close shapes, with all products having maximum global Q_{le} in July, illustrating the dominance of the Northern Hemisphere land areas. At the cycle maximum, there is a spread of ~25 $W\,m^{-2}$, with an ensemble mean and standard deviation of ~60 and ~10 $W\,m^{-2}$, respectively. For R_n the annual cycles peak between June and August, depending on the product. Some of the products having relatively small amplitude in the R_n cycle also have small amplitudes in the Q_{le} cycle (e.g., GSWP-MMA), but this is not always the case (e.g., PRU, with one of the largest R_n and smaller Q_{le} cycles). For EF the annual cycles are more different from one product to another, though all of them peak between July and September. For the month of highest Q_{le} (July), the EF vary between ~0.4 to ~0.7, suggesting significant differences in the way the different models partition the fluxes. Close annual cycles are found for 1993 (not shown). The more distinctive products are NCEP-DOE and PRU with EF from most months outside the envelope defined by the all-product ensemble mean ± one standard deviation.

4.2.3. Zonal means

Zonal means of $Q_{le}$and R_n, and EF for the months of February and August 1994 can be found in Figure 11. As expected, the seasonal changes in the latitudes of maximum R_n are reflected in the seasonal changes of Q_{le} with latitude. Although the differences in absolute values can be large between products (e.g., at the latitudes of largest Q_{le} in February, a Q_{le} spread of $\sim$40 $W\,m^{-2}$, with a ensemble mean and standard deviation of $\sim$100 $W\,m^{-2}$ and 15 $W\,m^{-2}$, respectively), in broad terms the shapes of the latitudinal distributions seem in general to be consistent from one product to another (e.g., see the triple-peak shape in the August zonal means). The zonal means for August 1993 show similar shapes (not shown).

4.2.4. Monthly anomalies

The global annual correlations are studied further by removing the seasonal component from the flux time series. For each pixel and month the inter-annual mean fluxes are first calculated by averaging the three (1993-1994-1995) monthly values (apart from UCB and UMD November and December, where only the two 1993 and 1994 values are used). The inter-annual mean flux is then subtracted from the original monthly mean fluxes to obtain the monthly anomalies, referred to here as the deseasonalized fluxes. Table 4 gives a summary of the original, inter-annual, and deseasonalized Q_{le}, Q_{h}, and R_{n} global correlations for 1994. The correlations are calculated by first adding together the correlations of each product with all the other products, followed by dividing by the number of products (i.e., by doing a product-average). Table 4 shows that the inter-annual fluxes correlate slightly higher (0.89 to 0.93 for Q_{le}) than the original fluxes (0.86 to 0.91), and much higher than the deseasonalized fluxes (0.12 to 0.45). This confirms that the large seasonal variability of the fluxes (e.g., see Figure 10) is partly responsible for the high correlations between the products. The fact that some products are not completely independent can also be seen in the individual product-to-product correlations for the deseasonalized fluxes. For instance, UCB and UMD models share a large number of forcings, and the Q_{le} correlation is the highest of all products (0.83). The same applies to the GLDAS-Noah, GLDAS-CLM, and GLDAS-Mosaic models, forced with the same datasets, exhibiting higher correlations than other products (0.70 to 0.79). Table 4 also shows that the lowest correlations are for the MPI-BGC deseasonalized fluxes (0.12 for Q_{le}), even if the inter-annual fluxes agree well with the other products (0.91). One might speculate that the use of the in-situ datasets (e.g, the EC measurements and GPCC precipitation) by MPI-BGC (in contrast to some of the other products using more satellite based forcings) may be a factor explaining the low deseasonalized correlations, but this cannot be further tested here.

4.3. Comparing fluxes for different vegetation classes

4.3.1. Annual differences

The vegetation classes are displayed in Figure 12. The class averaged 1994 Q_{le} and R_n annual means for the different classes are presented in Figure 13. The class averaged annual precipitation amount (as estimated from the GPCP data) is also given for each class. The largest spread in Q_{le} are observed for the the rain forest ($\sim$35 $W\,m^{-2}$, with an ensemble mean and standard deviation of $\sim$98 and 25 $W\,m^{-2}$), and for R_n in the desert ($\sim$60 $W\,m^{-2}$, with an ensemble mean and standard deviation of $\sim$75 and 40 $W\,m^{-2}$). Close results are found for 1993 (nor shown).

The large Q_{le} differences in the rain forest (compared with the other vegetation types) may indicate larger observational or modeling difficulties for these regions. Conventional interception-measurement in tropical rain forest sites (e.g, see reported 8% to 40% of total annual precipitation from a compilation in *Czikowsky and Fitzjarrald* [2009]) suggest that canopy evaporation from intercepted rain can be an important component of Q_{le}. Therefore, differences in how interception is modeled may have a larger importance in this region, contributing to some of the observed differences. The reanalyses and off-line models compared here have schemes that account for rain interception, but of the satellite-based products only UMD explicitly accounts for evaporation of intercepted water. Nevertheless, the large differences could also be related to larger absolute fluxes, compared with other regions. If the differences between the products are normalized by the average flux for each product and class, the normalized mean and RMS differences for the rain forest are now comparable to other classes. This is illustrated in Table 5. For instance, the normalized RMS difference for the rain forest takes values between 0.21 and 0.37, while a larger difference between 0.33 and 0.50 is found for the cultivated areas.

4.3.2. Seasonal correlations

The seasonal correlations between Q_{le} and R_n for the different products and vegetation classes are displayed in Table 6 for Dec-Jan-Feb and Jul-Aug-Sep 1994. To have a well defined seasonal cycle, the correlations are calculated only for the classes in Tropical and Northern hemisphere regions (pixels with latitude $< 20^o$S are removed). In general, the correlations are higher for the satellite-based products, compared with the reanalyses and off-line models. This may be related to a more direct dependence of Q_{le} on R_n in the simpler models used by the satellite-based products (in contrast to the more complex parameterizations used in the reanalyses and off-line models). For most of the classes there is consistency in how the correlations for the different products change from winter to summer. For instance, for cultivation, grassland and shrubland there is a clear change in correlations between the winter-dry period (e.g., 0.70 to 0.93 for cultivation) and the summer-wet (0.42 to 0.81), with all products suggesting a larger control of Q_{le} by R_n for the winter-dry conditions. For the evergreen, deciduous forest, and woodlands, the correlation changes between winter and summer are smaller than before, with less variation in the correlations for the satellite-based products than for the reanalyses and off-line models. For the rain forest, all products but NCEP-DOE have smaller correlations for the winter-wet than for the summer-dry season, although the correlation coefficients and seasonal difference vary considerably from one product to another (e.g., 0.82 to 0.88 for UCB, 0.05 to 0.54 for GLDAS-CLM). Correlations for only for the Amazonian rain forest have also been calculated and added to Table 6. Closer correlations between wet and dry season are observed. Although some satellite-based products have slighly larger correlations for the wet season (UCB and MPI-BGC), larger correlations are observed again for the dry season across most of the products. *Hasler and Avissar* [2007] shows evidence of the opposite: larger correlations for the wet season than for the dry season from EC meaurements at a few Amazon rain forest sites, but it is uncertain whether this results holds for the whole Amazon rain forest averaged fluxes compared here.

4.4. Comparing fluxes for different river basins

4.4.1. Annual differences

The geographical location and extent of the 10 selected basins are displayed in Figure 14. They include some of the major tropical and mid-latitudes river systems. The basin averaged $\mathrm{Q_{le}}$ and $\mathrm{R_n}$ 1994 annual means for the different basins are presented in Figure 15. Close differences are observed for 1993 (not shown). The basin averaged annual precipitation amount (as estimated from the GPCP data) is also given for each basin. A larger relative spread (with respect to the all-product ensemble mean) in the annual $\mathrm{Q_{le}}$ is seen for the Danube, Congo, Volga, and Nile basins. For $\mathrm{R_n}$, the larger relative spread is observed for the Yangtze, Danube, Niger and Volga. One could speculate that the large spread in the African basins may be related to difficulties in properly modeling some of the unique features of these regions (e.g., the West African Monsoon), further aggravated by a lack of observations (compared with other better characterized regions). For instance, UMD used EC and EBBR fluxes for the development/validation of its product, but no measurements over Africa are included; UCB uses EC fluxes from 36 stations, but only one is located in Africa. The Danube is the highest latitude basin considered, and the variability may be related to the difficulties in modeling the winter months. In terms of products, the reanalyses NCEP-DOE, MERRA and ERA-INT have in general the largest basin averaged Q_{le}, but this is not followed by the largest basin averaged R_n (e.g., over the Congo basin, where they have some of the smallest R_n). For the GLDAS off-line models, GLDAS-Mosaic has more distinct fluxes than GLDAS-CLM and GLDAS-Noah. For these off-line models a scaling of Q_{le} as a function of R_n is apparent for some basins (e.g., for the Mississippi), but this does not hold for some of the other basins (e.g., the Danube and Yangtze). From the satellite-based products, the plots again show, as expected, similar fluxes for PA-OBS and GSWP-MMA. PRU exhibits large Q_{le} and R_n for the basins with the higest rainfall (e.g., Amazon and Mississippi), while for most of the other basins (e.g., Volga, Murray, Parana, or Danube) it has the smallest average Q_{le} (not necessarily related to a small R_n, as for the Volga and Danube basins).

4.4.2. Monthly time series

Monthly time series of the basin averaged Q_{le}, Q_h, R_n, and EF for the Amazon and Murray rivers are presented in Figures 16 and 17. Figures 16 is representative of a tropical region with large rainfall and relatively small seasonal and inter-annual variability. Figure 17 represents a drier mid-latitude region with large seasonal and inter-annual variability. For the Amazon basin, PA-OBS and GSWP-MMA have relatively close Q_{le}, apart from the last months in 1995, where an anomaly in the GSWP radiative forcing produced large Q_{le}. The PA-OBS model driven by the remote sensing observations modify these fluxes to more expected values (see *Jiménez et al.* [2009] for more details). UMD and MPI-BGC Q_{le} are also close to PA-OBS and GSWP-MMA, while PRU and UCB have relatively higher fluxes.

Large differences in the Amazonian modeled Q_{le} annual cycle have been reported in *Werth and Avissar* [2004], where it was suggested that the differences come from the way the vegetation controls the evapotranspiration in the models. Time series of EC measurements in the Amazon basin (see the complilation reported in *Fisher et al.* [2009]) show that the seasonal changes depend on the location and water conditions at each specific site and year. Large seasonal variarions in strongly water limited regions with long dry seasons are presented in *da Rocha et al.* [2009]. In general, the basin averaged estimates from the satellite-based products presented here do not show much seasonal variability, and some of the changes appear to be related to variations in R_n (the EFs are relatively constant, in agreement with the relatively high Q_{le} to R_n correlations discussed in Section 4.3). For the reanalyses, ERA-INT has a more constant EF (which can be explained by the lack of seasonal cycle in its vegetation scheme), while NCEP-DOE and especially MERRA have more variable fluxes (e.g., MERRA EF changes from $\sim$0.5 to $\sim$1). From the GLDAS off-line models, GLDAS-CLM shows the largest changes in Q_{le} and Q_h, with differences of $\sim$50 $W\,m^{-2}$ between the winter and summer months.

For the Murray basin, all products show much more seasonality than for the Amazon basin. Nevertheless, some of the products exhibit large inter-annual variability than others. For instance, reanalyses and off-line models show significant changes in Q_{le} for January and December during these three years. Some of the satellite-based products also follow these changes (e.g PA-OBS and MPI-BGC), while some show more constant fluxes for these two months (e.g., UCB and UMD, which follow each other closely, or PRU, with smaller fluxes). Close inspection of the mean R_n seems to show that some of these different seasonal values are related to corresponding changes in the R_n forcing the products.

Time series for the remaining basins are presented in the auxiliary material (Figures S11 to S18). Not much inter-annual variability for the 3 years analyzed is evident, and in general all products capture the strong seasonality present in some of the basins.

5. Summary and conclusions

Land surface heat fluxes are essential components of the energy and water cycle. *In situ* measurements of the turbulent land heat fluxes by tower networks exist, but they lack global coverage. For global estimation, the alternative is a range of models forced by global datasets providing information about the physical properties of the surface and/or atmosphere affecting the land surface fluxes.

A global inter-comparison of existing sensible (Q_h) and latent (Q_{le}) heat fluxes (here collectively referrred as Q) datasets for a selected period of time (1993-1995) at monthly time scales is presented here. The inter-comparison includes a representative sample of the first emerging global satellite-based flux products and some examples of estimates produced by reanalyses and off-line forced land surface models (off-line models).

The analysis presented here was conducted by comparing the different estimates Q_{le} and Q_h, the associated net radiative fluxes (R_n), and the evaporative fraction (EF = Q_{le}

/ R_n) after space aggregation of the different products onto a grid of 2.5^{o}x2.5^o (coarsest resolution of the products compared). Comparison of the global Q_{le} annual means shows a spread of ∼20 $W\,m^{-2}$ (∼15 $W\,m^{-2}$ excluding the two products with largest and smallest fluxes) for an all-product ensemble global mean of ∼45 $W\,m^{-2}$. An approximately similar spread is observed in the global annual means of Q_h and R_n (but for R_n with an ensemble mean of ∼90 $W\,m^{-2}$, implying a smaller relative spread). In general, the products correlate well with one another, but it should be noted that the large seasonal variability of the fluxes and the fact that some of the products share forcings are to a large extent responsible for this agreement. Some of the lowest correlations occur with the reanalyses NCEP-DOE and MERRA. Inspection of their global annual mean charts reveal marked difference at some regions (relative to the other products) that could explain the lower correlations.

Inspection of the monthly mean flux distributions for selected months shows that in general main geographical structures related to the principal climatic regimes are present in all products. Nevertheless, large Q_{le} and Q_h differences in the absolute values among some products are observed. Annual cycles for Q_{le} peak for all products in July. The spread in the cycles maximum value is ∼25 $W\,m^{-2}$ (with an ensemble mean of ∼60 $W\,m^{-2}$). For R_n, the annual cycles peak between June and August, depending on the product. For EF, the annual cycles are more different from one product to another, though all of them peak between July and September. For the month of highest Q_{le} (July), the EF vary between ∼0.4 to ∼0.7, suggesting significant differences in the way the different models partition the fluxes.

The fluxes were spatially averaged for 10 major vegetation classes. The larger Q_{le} differences were observed for the rain forest, but in relative terms (differences normalized with the annual class fluxes) the mean difference and root mean square differences were not the largest, compared with the other classes. Q_{le} to R_n seasonal correlations for winter and summer for the different products and classes were calculated. In general, the correlations were higher for the satellite-based products, compared with the reanalyses and off-line models. For most of the classes there is consistency in how the correlations for the different products change from winter to summer. For instance, for cultivation, grassland and shrubland there is a clear change in correlations between the winter-dry period (e.g., 0.70 to 0.93 for cultivation) and the summer-wet (0.42 to 0.81). For the rain forest, all products but NCEP-DOE have smaller correlations for the winter than for the summer season, altough the correlation coefficients and seasonal difference vary considerably from one product to another. For most of the products, correlations recalculated just for the Amazon rain forest showed also a smaller correlation for the wet season than for the dry season.

The fluxes were also spatially averaged for a group of 10 selected basins including some of the major river systems at tropical and mid-latitudes. With respect to the all-product ensemble average, a realtively large spread in Q_{le} was observed for the Danube, Congo, Volga, and Nile basins. For R_n, the largest relative spread is observed for the Yangtze, Danube, Niger and Volga. Monthly time series of basin averaged fluxes were plotted for the three years considered. The seasonality was in general well captured by all products, but some large differences were observed for some products and basins in the partitioning of the fluxes. Apart from the Murray basin, not much inter-annual variability was noticed in these three years.

Despite the existence of a large body of work characterizing Q_{le} and Q_h from the local to the regional scale [e.g., *Verstraeten et al.*, 2008; *Kalma et al.*, 2008], the extension to the global scale requires simplified formulations that are adapted to the existing global datasets and are also robust in the face of the data uncertainties. This inter-comparison highlights the difficulties of producing such global estimates. Some of the satellite-based products are first versions, and improvements in the analysed products are already on their way (e.g., improved UMD estimates [*Wang et al.*, 2010a]), which should result in more consistent fluxes. Nevertheless, the choice of formulation and forcing datasets will always have an effect on the estimated fluxes. For instance, choosing ISCCP or GEWEX-SRB as radiative forcing will immediately have an impact on the fluxes produced. Concerning the atmospheric reanalyses, important differences in some of the surface physical fields has also been noted elsewhere [e.g., *Bosilovich et al.*, 2009], and users are typically advised to use the physical fields (as opposed to the assimilated states) with caution. Regarding the off-line models, the inter-comparison showed that even when forced with the same datasets, their paremeterizations can have a large effect on the partitioning of the fluxes, as has already been shown [e.g., *Schlosser and Gao*, 2009]. Nevertheless, an increasing better understanding of the soil-atmosphere-vegetation transfer processes [e.g., *Betts*, 2009; *Seneviratne et al.*, 2010] is driving the improvement of some of the land surface models considered here [e.g., *Balsamo et al.*, 2009], which should result in a better flux estimation.

This inter-comparison has been made in the framework of the GEWEX LandFlux activity, and it is part of a series of inter-comparison exercises coordinated by the LandFlux-EVAL initiative. This type of exercise will contribute to the objective of identifying and delivering robust procedures for the production of global land surface heat fluxes.

Acknowledgments. The LandFlux-Eval initiative acknowledges suport by GEWEX. C. D. Kummerow, as chair of the GEWEX Radiation Panel, is acknowledged for encouraging the LandFlux activity, and contributing to the scientific discussions. The GLDAS data were acquired as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC) are acknowledged by disseminating the GLDAS data. The Global Modeling and Assimilation Office (GMAO) and the GES DISC are acknowledged for disseminating the MERRA data.

References

Adler, R., et al., The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1970-Present), *J. Hydrol.*, *4*, 1147–1167, 2003.

Aires, F., C. Prigent, and W. Rossow, Temporal interpolation of global surface skin temperature diurnal cycle over land under clear and cloudy conditions, *J. Geophys. Res.*, *109,* D04313, 10.1029/2003JD003,527, 2004.

Anderson, M., and W. Kustas, Thermal Remote Sensing of Drought and Evapotranspiration, *EOS*, *89*, 233–240, 2008.

Armstrong, R., and M. Brodzik, Northern Hemisphere EASE-Grid weekly snow cover and sea ice extent version 3, *Tech. rep.*, Boulder, Colorado USA: National Snow and

Ice Data Center. Digital media., 2005.

Baldocchi, D., et al., FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bull. Amer. Meteor. Soc.*, *82*, 2415–2434, 2001.

Balsamo, G., P. Viterbo, A. Beljaars, B. van den Hurk, M. Hirschi, A. K. Betts, and K. Scipal, A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System, *J. Hydromet.*, pp. 623–643, 2009.

Berg, A., J. Famiglietti, J. Walker, and P. Houser, Impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes, *J. Geophys. Res.*, *108*, 10.1029/2002JD003,334, 2003.

Betts, A., Land-Surface-Atmosphere Coupling in Observations and Models, *J. Adv. Model. Earth Syst.*, *1*, doi:10.3894/JAMES.2009.1.4, 2009.

Betts, A., J. Ball, A. Beljaars, M. Miller, and P. A. Viterbo, The land surface-atmosphere interaction: A review based on observational and global modeling perspectives, *J. Geophys. Res.*, *101*, 7209–7226, 1996.

Betts, A., M. Zhao, P. Dirmeyer, and A. Beljaars, Comparison of ERA40 and NCEP/DOE near-surface data sets with other ISLSCP-II data sets, *J. Geophys. Res.*, *111,* D22S04 , doi:10.1029/2006JD007,174, 2006.

Boone, A., et al., The Rhône-Aggregation Land Surface Scheme intercomparison project: An overview, *J. Clim.*, *17*, 187–208, 2004.

Boone, A., et al., The AMMA Land Surface Model Intercomparison Project, *Bull. Amer. Meteor. Soc.*, *90*, doi:10.1175/2009BAMS2786.1, 2009.

Bosilovich, M., NASA's Modern Era Retrospective-analysis for Research and Applications: Integrating Earth Observations, *EarthZine*, pp. E–Zine article, 2008.

Bosilovich, M., D. Mocko, J. Roads, and A. Ruane, A Multimodel Analysis for the Coordinated Enhanced Observing Period (CEOP), *J. Hydromet.*, *10*, 11.1175/2009JHM1090.1, 2009.

Cleugh, H., R. Leuning, Q. Mu, and S. Running, Regional evaporation estimates from flux tower and MODIS satellite data, *Remote Sens. Envir.*, *106*, 285–304, 2007.

Cordisco, E., C. Prigent, and F. Aires, Snow characterization at a global scale with passive microwave satellite observations, *J. Geophys. Res.*, *111,* D19102, doi:10.1029/2005JD006,773, 2006.

Czikowsky, M., and D. Fitzjarrald, Detecting rainfall interception in an Amazonian rain forest with eddy flux measurements, *J. Hydrol.*, *377*, 92–105, 2009.

da Rocha, H., et al., Patterns of water and heat flux across a biome gradient from tropical forest to savanna in Brazil, *J. Geophys. Res.*, *114,* G00B12, doi:10.1029/2007JG000,640, 2009.

Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki, GSWP-2: Multimodel analysis and implications for our perception of the land surface, *Bull. Amer. Meteor. Soc.*, *87*, 1831–1397, 2006.

Douville, H., P. Viterbo, J.-F. Mahfouf, and A. Beljaars, Evaluation of the optimum interpolation and nudging techniques for soil moisture analysis using FIFE data, *Month. Weath. Rev.*, *128*, 1733–1756, 2000.

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

Ek, M. B., K. E. Mitchell, Y. Lin, P. Grunmann, E. Rogers, G. Gayno, V. Koren, and J. D. Tarpley, Implementation of the upgraded Noah land-surface model in the NCEP operational mesoscale Eta model, *J. Geophys. Res.*, *108 (D22), 8851*, doi:10.1029/2002JD003,296, 2003.

Entin, K. K., A. Robock, K. Y. Vinnikov, V. Zabelin, S. Liu, and A. Namkhai, Evaluation of Global Soil Wetness Project soil moisture simulations, *J. Meteorol. Soc. Jap*, *77*, 183–198, 1999.

Fisher, J. B., K. Tu, and D. D. 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.

Fisher, J. B., et al., The land-atmosphere water flux in the tropics, *Global Change Biology*, *15*, 2694–2714, 2009.

Francis, R., et al., The ERS-1 spacecraft and its payload, *ESA Bulletin*, *65*, 27–48, 1991.

Frison, P. L., and E. Mougin, Monitoring global vegetation dynamics with ERS-1 wind scatterometer data, *Int. J. Remote Sens.*, *17*, 3201–3218, 1996.

Gellens-Meulenberghs, F., A. Arboleda, and N. Ghilain, Towards a continuous monitoring of evapotranspiration based on msg data, in *Symposium on Remote Sensing for Environmental Monitoring and Change Detection*, IAHS Series, IUGG, 2007.

Gobron, N., B. Pinty, O. Aussedat, M. Taberner, O. Faber, F. M. T. Lavergne, M. Robustelli, and P. Snoeij, Uncertainty estimates for the FAPAR operational products derived from MERIS - Impact of top-of-atmosphere radiance uncertainties and validation with field data, *Remote Sens. Envir.*, *112*, 1871–1883, 2008.

Gobron, N., et al., Evaluation FAPAR products for differentcanopy radiation transfer regimes: Methodology and results using JRC products derived from SeaWiFS against ground-based estimations, *J. Geophys. Res.*, *111,* D13110, 10.129/2005JD006,511, 2006.

Grubber, A., and V. Levizzani, Assessment of Global Precipitation Products, *A project of the World Climate Research Programme Global Energy and Water Cycle Experiment (GEWEX) Radiation Panel WCRP-128, WMO/TD-No. 1430*, WCRP, 2008.

Guo, Z., and P. A. Dirmeyer, Evaluation of the second Global Soil Wetness Project soil moisture simulations: 1. Intermodel comparison, *J. Geophys. Res.*, *111,* D22S02, 10.129/2006JD07,233, 2006.

Gutman, G. G., On the use of long-term global data of land reflectances and vegetation indices from the advanced very high resolution radiometer, *J. Geophys. Res.*, *104*, 6241–6255, 1999.

Hall, F., E. B. de Colstoun, G. Collatz, D. Landis, P. Dirmeyer, A. Betts, G. Huffman, L. Bounoua, and B. Meeson, ISLSCP Initiative II global data sets: Surface boundary conditions and atmospheric forcings for land-atmosphere studies, *J. Geophys. Res.*, *111,* D22S01, 10.1029/2006JD007,366, 2006.

Hasler, N., and R. Avissar, What Controls Evapotranspiration in the Amazon Basin, *J. Hydromet.*, *8*, 380–395, 2007.

Henderson-Sellers, A., A. Pitman, P. Love, P. Irannejad, and T. Chen, The project for Intercomparison of Land Surface parameterization Schemes (PILPS): Phase 2 and 3, *Bull. Amer. Meteor. Soc.*, *76*, 489–503, 1995.

Hollinger, J., R. Lo, G. Poe, R. Savage, and J. Pierce, Special Sensor Microwave/Imager user guide, *Tech. rep.*, Nav. Res. Lab., Washington, D.C, 1987.

Huete, A., A soil-adjusted vegetation index (SAVI), *Remote Sens. Envir.*, *25*, 295–309, 1998.

Jiménez, C., C. Prigent, and F. Aires, Toward an estimation of global land surface heat fluxes from multisatellite observations, *J. Geophys. Res.*, *114,* D06305, doi:10.1029/2008JD011,392, 2009.

Jung, M., 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*, 5271–5304, 2009.

Kalma, J., T. McVicar, and M. McCabe, Estimating land surface evaporation: a review of methods using remotely sensing surface temperature data, *Surv. Geophys.*, pp. doi:10.1007/s10,712–008–9037–z, 2008.

Kalnay, E., et al., The NCEP/NCAR 40-year reanalysys project, *Bull. Amer. Meteor. Soc.*, *77*, 437–471, 1996.

Kanamitsu, M., W. Ebisuzaki, J. Woolen, S. K. Yang, J. J. Hnilo, M. Fiorino, and J. Potter, The NCEP-DOE AMIP-II reanalysys (R-2), *Bull. Amer. Meteor. Soc.*, *83*, 1631–1643, 2002.

Kumar, S. V., et al., Land Information System - An interoperable framework for high resolution land surface modeling, *Environ. Modeling and Software*, *21*, 1402–1415, 2006.

Leuning, R., Y. Q. Zhang, A. Rajaud, H. Cleugh, and K. Tu, A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation, *Water Resour. Res.*, *44,* W10419, doi:10.1029/2007WR006,562, 2008.

Lim, W., and M. Roderick, An Atlas of the Global Water Cycle Based on the IPCC AR4 Climate Models, *Tech. rep.*, published by ANU E Press, Canberra, Australia, ISBN 9781921536892, 2009.

Los, S., et al., A global 9-year biophysical land-surface data set from NOAA AVHRR data, *J. Hydromet.*, *1*, 183–199, 2000.

Lu, C.-H., M. Kanamitsu, J. Roads, W. Ebisuzaki, and K. E. Mitchell, Evaluation of Soil Moisture in the NCEP-NCAR and NCEP-DOE Global Reanalyses, *J. Hydromet.*, *6*, 391–408, doi: 10.1175/JHM427.1, 2007.

Matthews, E., Global vegetation and land use: New high-resolution data bases for climate studies, *J. Clim. Appl. Meteor*, *22*, 474–487, 1983.

McCabe, M., and E. Wood, Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors, *Remote Sens. Envir.*, *105*, 271–285, 2006.

McCabe, M., E. Wood, R. Wójcik, M.Pan, J. Sheffield, H. Gao, and H. Su, Hydrological consistency using multi-sensor remote sensing data for water and energy cycle studies, *Remote Sens. Envir.*, *112*, 430–444, 2008.

Mesinger, F., et al., North American Regional Reanalysis, *Bull. Amer. Meteor. Soc.*, *87*, 343–360, 2006.

Mu, A., F. Heinsch, M. Zhao, and S. Running, Development of a global evapotranspiration algorithm based on MODIS and global metereology data, *Remote Sens. Envir.*, *111*, 519–536, 2007.

New, M., M. Hulme, and P. Jones, Representing twentieth-century space-time climate variability. Part I: Development of a 1961-1990 mean monthly terrestrial climatology, *J. Clim.*, *12*, 829–856, 1999.

New, M., M. Hulme, and P. Jones, Representing twentieth-century space-time climate variability. Part II: Development of a 1901-1996 monthly grids of terrestrial surface climate, *J. Clim.*, *13*, 2217–2238, 2000.

Nishida, K., R. Nemani, S. Running, and J. Glassy, An operational remote sensing algorithm of land surface evaporation, *J. Geophys. Res.*, *108*, 10.129/2002JD002,062, 2003.

Oki, T., and Y. C. Su, Design of Total Runoff Integrating Pathways (TRIP): a global river channel network, *Earth Interactions*, *2*, 1–37, 1998.

Papa, F., C. Prigent, C. Jimenez, F. Aires, and W. Rossow, Interannual variability of surface water extent at global scale, 1993-004, *J. Geophys. Res.*, *115,* D12111, doi:10.1029/2009JD012,674, 2010.

Papale, D., et al., Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, *Biogeosciences*, *3*, 571–583, 2006.

Prigent, C., F. Aires, W. Rossow, and E. Matthews, Joint characterization of vegetation by satellite observations from visible to microwave wavelength: A sensitivity analysis, *J. Geophys. Res.*, *106*, 20,665–20,685, 2001a.

Prigent, C., E. Matthews, F. Aires, and W. Rossow, Remote sensing of global wetland dynamics with multiple satellite data sets, *Geophys. Res. Lett.*, *28*, 4631–4634, 2001b.

Prigent, C., F. Aires, W. Rossow, and A. Robock, Sensitivity of satellite microwave and infrared observations to soil moisture at a global scale: Relationship of satellite observations to in situ soil moisture measurements, *J. Geophys. Res.*, *110,* D07110, 10.1029/2004JD005,087, 2005.

Prigent, C., F. Aires, and W. Rossow, Land surface microwave emissivities over the globe for a decade, *Bull. Amer. Meteor. Soc.*, *87*, 1573–1584, 2006.

Reichstein, M., et al., On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global Change Biology*, *11*, 1424–1439, 2005.

Rodell, M., et al., The Global Land Data Assimilation System, *Bull. Amer. Meteor. Soc.*, *85*, 381–394, 2004.

Rossow, W., and R. Schiffer, Advances in understanding clouds from ISCCP, *Bull. Amer. Meteor. Soc.*, *80*, 2261–2287, 1999.

Rudolf, B., and U. Schneider, Calculation of gridded precipitation for the global land-surface using in-situ gauge observations, in *Proceedings of the International Precipitation Working Group. Monterey, California. 25-28 October 2004. ISBN 92-9110-070-6*, pp. 231–247, 2005.

Rutter, N., et al., Evaluation of forest snow processes models (SnowMIP2), *J. Geophys. Res.*, *114,* D06111, doi:10.1029/2008JD011,063, 2009.

Schlosser, C., and X. Gao, Assessing Evapotranspiration Estimates from the Global Soil Wetness Project Phase 2(GSWP-2), *Tech. Rep. 179*, MIT Joint Program on the Science and Policy of Global Change, 2009.

Sellers, P., et al., BOREAS in 1997: Experiment overview, scientific results, and future directions, *J. Geophys. Res.*, *102*, 28,731–28,769, 1997.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, Investigating soil moisture climate interactions in a changing climate: A review, *Earth Science Reviews*, *99*, 125–161, 2010.

Sheffield, J., G. Goteti, and E. Wood, Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling, *J. Clim.*, *19(13)*, 3088–3111, 2006.

Sheffield, J., E. Wood, and F. Munoz-Ariola, Long-term regional estimates of evapotranspiration for Mexico based on downscaled ISCCP data, *in revision*, 2009.

Simmons, A., S. Uppala, D. Dee, and S. Kubayashi, ERA-Interim: New ECMWF reanalysis products from 1989 onwards, *ECMWF Newsletter*, *110*, 25–35, 2006.

Stackhouse, P., S. Gupta, S. Cox, J. Mikovitz, T. Zhang, and M. Chiacchio, 12-year surface radiation budget data set, *GEWEX News*, *14*, 10–12, 2004.

Stöckli, R., D. Lawrence, G.-Y. Niu, K. Oleson, P. Thornton, Z.-L. Yang, and G. Bonan, Use of FLUXNET in the Community Land Model development, *J. Geophys. Res.*, *113,* G01025, doi:10.1029/2007JG000,562, 2008.

Su, H., E. Wood, M. McCabe, and Z. Su, Evaluation of Remotely Sensed Evapotranspiration Over the CEOP EOP-1 Reference Sites, *J. Meteorol. Soc. Jap*, *85A*, 439–459, 2007.

Su, Z., The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, *Hydrol. Earth System Sci.*, *6*, 85–99, 2002.

van den Hurk, B., P. Viterbo, A. Beljaars, and A. Betts, Offline validation of the ERA40 surface scheme, *Tech. Rep. 295*, ECMWF Technical Memorandum, 2000.

Vermote, E. F., and N. E. Saleous, An extended AVHRR 8-km NDVI Data Set Compatible with MODIS and SPOT Vegetation NDVI data, *Int. J. Remote Sens.*, *26*, 4485–5598, 2005.

Verstraeten, W., F. Veroustraete, and J. Feyen, Assessment of Evapotranspiration and Soil Moisture Content Across Different Scales of Observation, *Sensors*, *8*, 70–117, 2008.

Viterbo, P., and A. Beljaars, An improved land surface parameterization scheme in the ECMWF model and its validation, *J. Clim.*, *8*, 2716–2748, 1995.

Wang, D., G. Wang, and E. Anagnostou, Evaluation of canopy interception schemes in land surface models, *J. Hydrol.*, *347(3-4)*, 308–318, 2007a.

Wang, K., and S. Liang, An Improved Method for Estimating Global Evapotranspiration Based on Satellite Determination of Surface Net Radiation, Vegetation Index, Temperature, and Soil Moisture, *J. Hydrol.*, *9*, 712–727, 2008.

Wang, K., P. Wang, Z. Li, M. Cribb, and M. Sparrow, A simple method to estimate actual evapotranspiration from a combination of net radiation, vegetation index, and temperature, *J. Geophys. Res.*, *112,* D15107, 10.1029/2006JD008,351, 2007b.

Wang, K., R. E. Dickinson, M. Wild, and S. Liang, Evidence for Decadal Variation in Global Terrestrial Evapotranspiration between 1982 and 2002, Part 1: Model Development, *J. Geophys. Res.*, *in press*, doi: 10.1029/2009JD013,671, 2010a.

Wang, K., R. E. Dickinson, M. Wild, and S. Liang, Evidence for Decadal Variation in Global Terrestrial Evapotranspiration between 1982 and 2002, Part 2: Results, *J. Geophys. Res.*, *in press*, doi: 10.1029/2010JD013,847, 2010b.

Werth, D., and R. Avissar, The Regional Evapotranspiration of the Amazon, *J. Hydromet.*, *5*, 100–109, 2004.

Williams, M., et al., Improving land surface models with fluxnet data, *Biogeosciences*, *6*, 1341–1359, 2009.

Wilson, K., et al., Energy partitioning between latent and sensible heat flux during the warm season at FLUXNET sites, *Water Resour. Res.*, *38*, 10.1029/2001WR000,989, 2002.

Wright, I., A. Manzi, and H. da Rocha, Surface conductance of Amazonian pasture: model application and calibration for canopy climate, *Agric. For. Meteorol*, *75*, 51–70, 1995.

Zhao, M., and P. Dirmeyer, Production and analysis of GSWP-2 near-surface meteorology data sets, *Tech. Rep. 159*, Center for Ocean-Land-Atmosphric Studies, 2003.

Table 1. Summary of the flux estimates inter-compared. See Section 2 for more details.

	INSTITUTION	Q_{le}	Q_h	R_n	Resolution
SATELLITE-BASED PRODUCTS					
UCB	University of California Berkeley	Physical-biological, Priestley-Taylor, inputs from ISLSCP-II (SRB, CRU, AVHRR)	R_n - Q_{le}	SRB	1986-95 monthly 1^o x 1^o
UMD	University of Maryland	Empirical (linear regression, AmeriFlux Q_{le}), inputs from ISLSCP-II (SRB, CRU, AVHRR)	R_n - Q_{le}	SRB	1986-95 monthly 1^o x 1^o
PRU	Princeton University	Penman-Monteith, inputs from ISCCP, AVHRR, NCEP/NCAR	R_n - Q_{le}	ISCCP-FD	1986-06 daily 2.5^o x 2.5^o
PAO	Paris Observatory	Empirical (neural networks, GSWP modeled Q_{le}), inputs from ISCCP, ERS, SSMI, AVHRR		Q_{le} + Q_h	1992-99 monthly $1/4^o$ x $1/4^o$
MPI	MPI for Biogeochemistry	Empirical (tree ensemble, FluxNet measured Q_{le}) inputs from CRU, GPCC, AVHRR.		Q_{le} + Q_h	1982-08 monthly $1/2^o$ x $1/2^o$
REANALYSIS					
MER	NASA-GMAO	MERRA reanalysis, GEOS-5 atmospheric model coupled with Catchment land model			1979- 1-hourly $1/2^o$ x $2/3^o$
NCE	NCEP/NCAR	NCEP-DOE reanalysis, atmospheric model coupled with OSU land model			1979- 6-hourly 2.5^o x 2.5^o
ERA	ECMWF	ERA Interim reanalysis, atmospheric model coupled with TESSEL land model			1989-98 6-hourly $3/4^o$ x $3/4^o$
OFF-LINE LAND SURFACE MODELS					
GSW	GLASS/ISLSCP	Multi-model ensemble, off-line forced with ISLSCP-II		SRB	1986-95 monthly 1^o x 1^o
NOA	NCAR/OSU/AFWA/HL	Equally off-line forced participating models driven by GLDAS		1993 ERA15 1994/5 NCEP/R1 SRB-bias corrected	1979- 3-hourly 1^o x 1^o
CLM	NCAR +				
MOS	NASA-GSFC				

Table 2. Statistics of the global 1994 Q_{le} monthly mean product differences. The table gives the mean difference (mean) and RMS difference (rmsd) of the monthly means ($W\,m^{-2}$), and the correlation coefficient (r^2) for each pair of products.

	UCB	**UMD**	**PAO**	**PRU**	**MPI**	**NCE**	**ERA**	**MER**	**GSW**	**NOA**	**CLM**	**MOS**
UCB		8.41	11.1	14.2	12.9	-6.79	2.95	1.84	12.9	11.5	13.2	6.26
UMD			2.74	5.78	4.49	-15.2	-5.46	-6.57	4.45	3.06	4.76	-2.15
PAO				3.04	1.75	-17.9	-8.2	-9.3	1.71	0.32	2.02	-4.89
PRU					-1.3	-21.0	-11.2	-12.3	-1.33	-2.73	-1.02	-7.93
MPI						-19.7	-9.95	-11.1	-0.03	-1.43	0.27	-6.64
NCE							9.74	8.64	19.7	18.3	20.0	13.1
ERA								-1.1	9.91	8.52	10.2	3.31
MER									11.0	9.62	11.3	4.42
GSW										-1.4	0.31	-6.6
NOA											1.7	-5.21
CLM												-6.91
MOS	**mean**											
UCB		17.6	19.6	23.2	21.1	24.3	17.2	24.9	23.1	22.3	28.0	21.6
UMD			15.5	21.6	17.3	29.8	18.5	30.0	19.2	19.0	27.0	23.0
PAO				18.0	13.2	28.1	15.9	25.9	12.8	14.6	21.1	19.2
PRU					15.5	32.4	21.2	27.8	21.3	20.0	22.0	23.3
MPI						28.5	17.0	25.8	16.1	13.9	18.9	19.9
NCE							21.8	24.9	28.6	28.0	30.1	25.3
ERA								23.5	17.0	16.7	22.2	17.2
MER									26.4	26.3	26.8	24.6
GSW										12.8	18.3	16.6
NOA											16.5	14.6
CLM												19.1
MOS	**rmsd**											
UCB		0.94	0.93	0.90	0.93	0.85	0.92	0.84	0.89	0.89	0.82	0.88
UMD			0.90	0.84	0.88	0.81	0.89	0.76	0.87	0.86	0.75	0.85
PAO				0.89	0.93	0.87	0.93	0.85	0.94	0.91	0.85	0.91
PRU					0.92	0.83	0.89	0.84	0.84	0.86	0.84	0.86
MPI						0.89	0.93	0.87	0.90	0.92	0.88	0.91
NCE							0.89	0.86	0.88	0.88	0.86	0.88
ERA								0.86	0.93	0.93	0.87	0.92
MER									0.85	0.85	0.85	0.85
GSW										0.94	0.89	0.94
NOA											0.91	0.96
CLM												0.91
MOS	$\mathbf{r^2}$											

Table 3. As Table 2, but for R_n.

	UCB	**UMD**	**PAO**	**PRU**	**MPI**	**NCE**	**ERA**	**MER**	**GSW**	**NOA**	**CLM**	**MOS**
UCB		0.0	22.5	-0.96	13.9	10.7	10.7	2.47	21.3	13.0	8.27	0.65
UMD			22.5	-0.96	13.9	10.7	10.7	2.47	21.3	13.0	8.27	0.65
PAO				-23.4	-8.60	-11.8	-11.8	-20.0	-1.21	-9.45	-14.2	-21.8
PRU					14.8	11.6	11.7	3.43	22.2	14.0	9.23	1.61
MPI						-3.20	-3.15	-11.4	7.39	-0.85	-5.60	-13.2
NCE							0.05	-8.20	10.6	2.35	-2.40	-10.0
ERA								-8.25	10.5	2.30	-2.45	-10.1
MER									18.8	10.6	5.80	-1.82
GSW										-8.24	-13.0	-20.6
NOA											-4.75	-12.4
CLM												-7.62
MOS	**mean**											
UCB		0.0	28.8	27.2	24.7	29.2	21.5	22.9	27.4	27.6	24.7	23.6
UMD			28.8	27.2	24.7	29.2	21.5	22.9	27.4	27.6	24.7	23.6
PAO				31.7	19.4	27.2	21.4	28.1	16.4	27.2	28.8	32.5
PRU					31.5	29.8	27.1	23.3	32.0	33.0	30.7	28.6
MPI						28.6	18.2	23.9	20.9	28.2	26.4	29.6
NCE							20.1	20.3	24.1	29.7	29.7	31.3
ERA								17.8	18.5	25.7	23.9	26.4
MER									25.7	30.5	27.6	27.6
GSW										23.5	24.8	29.8
NOA											12.7	17.4
CLM												15.5
MOS	**rmsd**											
UCB		1.00	0.94	0.92	0.92	0.92	0.94	0.93	0.95	0.92	0.92	0.92
UMD			0.94	0.92	0.92	0.92	0.94	0.93	0.95	0.92	0.92	0.92
PAO				0.96	0.95	0.94	0.95	0.95	0.96	0.91	0.91	0.92
PRU					0.95	0.92	0.94	0.94	0.94	0.90	0.90	0.91
MPI						0.94	0.96	0.96	0.95	0.90	0.92	0.92
NCE							0.96	0.96	0.95	0.90	0.89	0.89
ERA								0.97	0.96	0.91	0.92	0.92
MER									0.96	0.89	0.90	0.90
GSW										0.93	0.94	0.94
NOA											0.98	0.98
CLM												0.98
MOS	$\mathbf{r^2}$											

Table 4. Summary of the 1994 Q_{le} (top), Q_h (middle), and R_n (bottom) correlation coefficients for each product (with respect to all the other products, and then presented here as an average of the individual correlations). The correlations are estimated for three cases: (1) the original monthly fluxes (all); (2) the inter-annual monthly fluxes (int); and (3) the deseasonalized monthly fluxes (des). See the text for more details.

	UCB	**UMD**	**PAO**	**PRU**	**MPI**	**NCE**	**ERA**	**MER**	**GSW**	**NOA**	**CLM**	**MOS**
						$\mathbf{Q_{le}}$						
all	0.90	0.86	0.91	0.87	0.91	0.87	0.91	0.86	0.91	0.91	0.87	0.91
int	0.92	0.89	0.93	0.89	0.93	0.90	0.93	0.88	0.92	0.93	0.89	0.93
des	0.33	0.32	0.35	0.29	0.12	0.31	0.41	0.35	0.45	0.37	0.39	0.39
						$\mathbf{Q_h}$						
all	0.82	0.81	0.88	0.80	0.86	0.81	0.86	0.78	0.85	0.86	0.85	0.83
int	0.83	0.83	0.89	0.82	0.88	0.83	0.88	0.81	0.86	0.88	0.86	0.84
des	0.31	0.28	0.34	0.18	0.11	0.32	0.38	0.35	0.37	0.26	0.33	0.31
						$\mathbf{R_n}$						
all	0.94	0.94	0.94	0.93	0.94	0.93	0.95	0.94	0.95	0.93	0.93	0.93
int	0.96	0.96	0.96	0.95	0.95	0.94	0.96	0.95	0.96	0.94	0.95	0.95
des	0.45	0.45	0.40	0.33	0.12	0.26	0.40	0.32	0.33	0.32	0.34	0.32

Table 5. 1994 monthly mean Q_{le} normalized RMS difference for each vegetation class. For each product the statistics are calculated with respect to all the other products, and then product-averaged to get one estimate per product and class. Normalization is done by dividing the products difference by the average of the product fluxes.

	UCB	**UMD**	**PAO**	**PRU**	**MPI**	**NCE**	**ERA**	**MER**	**GSW**	**NOA**	**CLM**	**MOS**
RaFo	0.22	0.27	0.24	0.24	0.21	0.26	0.20	0.37	0.25	0.24	0.27	0.23
EvFo	0.46	0.43	0.43	0.52	0.41	0.60	0.37	0.49	0.48	0.42	0.54	0.43
DeFo	0.42	0.41	0.40	0.52	0.39	0.53	0.35	0.51	0.41	0.39	0.51	0.42
EvWo	0.59	0.60	0.54	0.68	0.55	0.70	0.49	0.61	0.62	0.55	0.70	0.57
DeWo	0.40	0.43	0.36	0.37	0.35	0.46	0.34	0.43	0.41	0.39	0.53	0.42
Cult	0.39	0.38	0.33	0.50	0.37	0.44	0.33	0.42	0.36	0.34	0.46	0.38
Gras	0.45	0.50	0.42	0.55	0.48	0.51	0.42	0.55	0.43	0.42	0.54	0.46
Tund	0.80	0.68	0.74	0.89	0.66	0.88	0.59	0.74	0.72	0.81	0.84	0.70
Shru	0.61	0.61	0.52	0.71	0.60	0.64	0.53	0.79	0.54	0.54	0.64	0.56
Dese	1.04	1.50	0.86	1.35	1.33	0.92	1.01	0.99	0.90	1.01	1.05	0.91

Table 6. Correlation coefficients between the monthly mean Q_{le} and R_n for the different products and vegetation classes for Dec-Jan-Feb 1994 (DJF)(top) and Jul-Aug-Sep 1994 (JAS)(bottom). Only pixels with latitude > -20^o are considered. AmFo gives the correlations for the Amazonian rain forest pixels. The last column gives the class averaged precipitation in $mm\,day^{-1}$.

	UCB	**UMD**	**PAO**	**PRU**	**MPI**	**NCE**	**ERA**	**MER**	**GSW**	**NOA**	**CLM**	**MOS**	**prec**
						DJF							
RaFo	0.82	0.74	0.50	0.85	0.55	0.72	0.54	0.12	0.42	0.57	0.05	0.47	7.08
EvFo	0.95	0.91	0.98	0.92	0.96	0.89	0.95	0.95	0.94	0.92	0.76	0.91	1.55
DeFo	0.94	0.90	0.97	0.93	0.95	0.91	0.96	0.94	0.91	0.93	0.76	0.93	1.42
EvWo	0.95	0.98	0.94	0.91	0.95	0.87	0.95	0.93	0.91	0.90	0.85	0.89	1.83
DeWo	0.94	0.97	0.93	0.92	0.94	0.89	0.92	0.87	0.87	0.84	0.79	0.85	3.08
Cult	0.92	0.91	0.93	0.92	0.91	0.83	0.92	0.91	0.86	0.85	0.70	0.84	1.73
Gras	0.84	0.91	0.86	0.85	0.86	0.82	0.88	0.82	0.82	0.71	0.66	0.73	1.86
Tund	0.93	0.61	0.91	0.78	0.87	0.03	0.67	0.79	0.48	0.21	0.70	0.60	0.84
Shru	0.90	0.90	0.87	0.87	0.88	0.79	0.85	0.80	0.77	0.72	0.72	0.76	1.19
Dese	0.64	0.51	0.50	0.73	0.40	0.49	0.48	0.49	0.28	0.28	0.28	0.26	0.30
AmFo	0.91	0.78	0.86	0.90	0.78	0.85	0.60	-0.25	0.44	0.64	0.14	0.54	7.75
						JAS							
RaFo	0.88	0.81	0.84	0.91	0.71	0.46	0.67	0.67	0.58	0.71	0.54	0.65	4.74
EvFo	0.96	0.93	0.92	0.90	0.88	0.63	0.78	0.74	0.82	0.83	0.54	0.76	3.37
DeFo	0.96	0.95	0.95	0.93	0.92	0.68	0.87	0.77	0.84	0.89	0.69	0.89	3.41
EvWo	0.86	0.94	0.84	0.88	0.81	0.48	0.84	0.61	0.72	0.68	0.46	0.57	1.89
DeWo	0.51	0.84	0.65	0.78	0.72	0.36	0.50	0.34	0.51	0.42	0.18	0.12	2.40
Cult	0.76	0.75	0.72	0.80	0.81	0.42	0.70	0.61	0.55	0.50	0.49	0.54	3.57
Gras	0.52	0.50	0.61	0.64	0.85	0.41	0.58	0.57	0.51	0.20	0.37	0.34	2.28
Tund	0.98	0.95	0.96	0.92	0.96	0.85	0.96	0.91	0.86	0.77	0.79	0.93	2.05
Shru	0.48	0.49	0.54	0.53	0.77	0.47	0.49	0.58	0.35	0.11	0.18	0.24	1.11
Dese	0.38	0.36	0.61	0.63	0.56	0.54	0.69	0.75	0.31	0.11	0.21	-0.23	0.73
AmFo	0.88	0.82	0.88	0.94	0.75	0.28	0.67	0.78	0.58	0.73	0.54	0.63	4.80

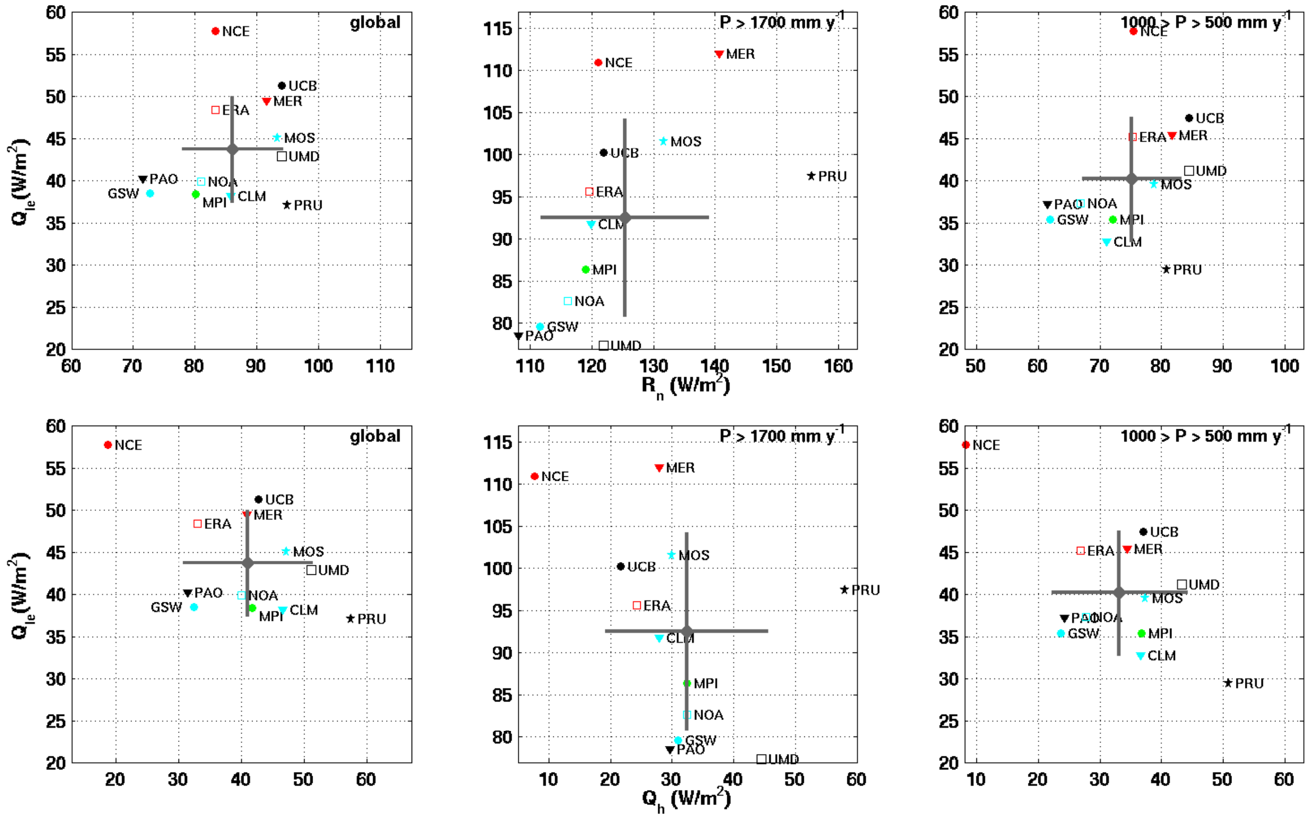

Figure 1. Q_{le} annual means as a function of the R_n (top) and Q_h (bottom) annual means for the year 1994. The averages are plotted for all the globe (left), for the regions where P > 1700 $mm\,year^{-1}$ (middle), and for 500 < P < 1000 $mm\,year^{-1}$(right). The grey dot and lines display respectively the ensemble mean and the standard deviation ($\pm\sigma$) of the individual product annual means around the ensemble mean.

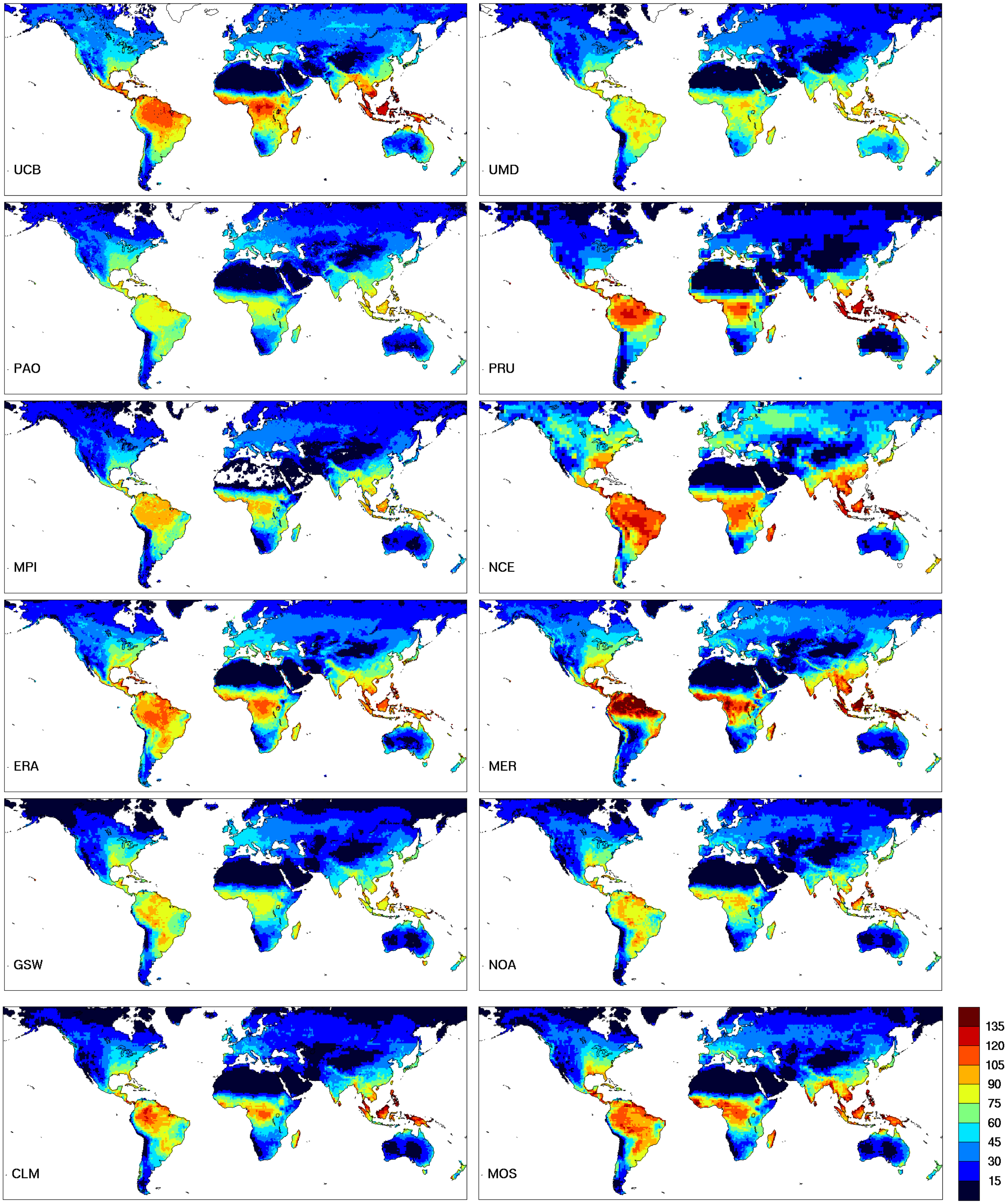

Figure 2. 1994 yearly averaged Q_{le} ($W\,m^{-2}$).

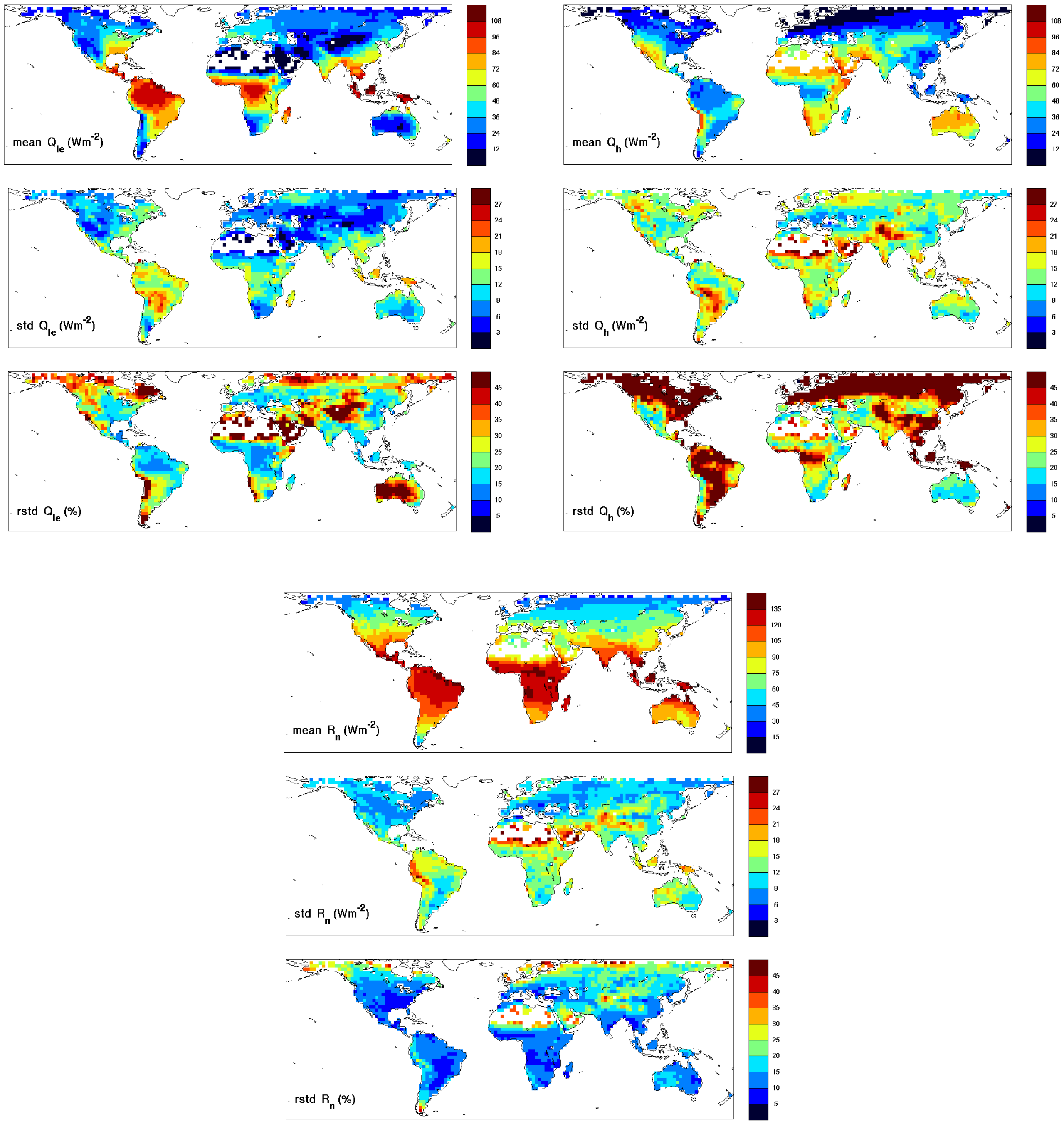

Figure 3. 1994 all-product ensemble mean (mean), standard deviation (std), and relative standard deviation (rstd)(expressed as a percentage of the pixel mean value) for Q_{le} (top-left), Q_h (top-right), and R_n (bottom). Absence of data from some products precludes the computation of the averages at some regions, mainly over Northern Africa.

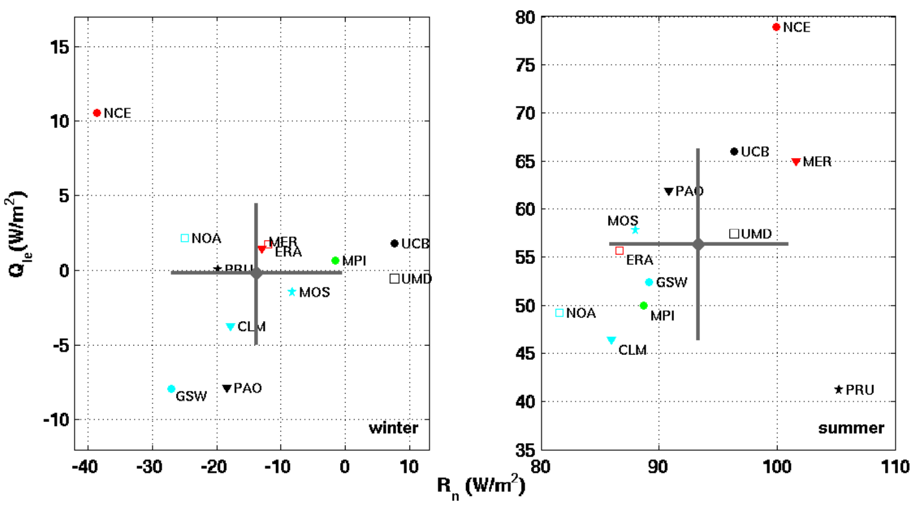

Figure 4. Q_{le} means as a function of the R_n means for snow-covered regions in Dec−Jan−Feb 1994 (left) and for the same regions in Jul−Aug−Sep 1994 (right). The grey dot and lines display respectively the ensemble mean and the standard deviation ($\pm\sigma$) of the individual product annual means around the ensemble mean.

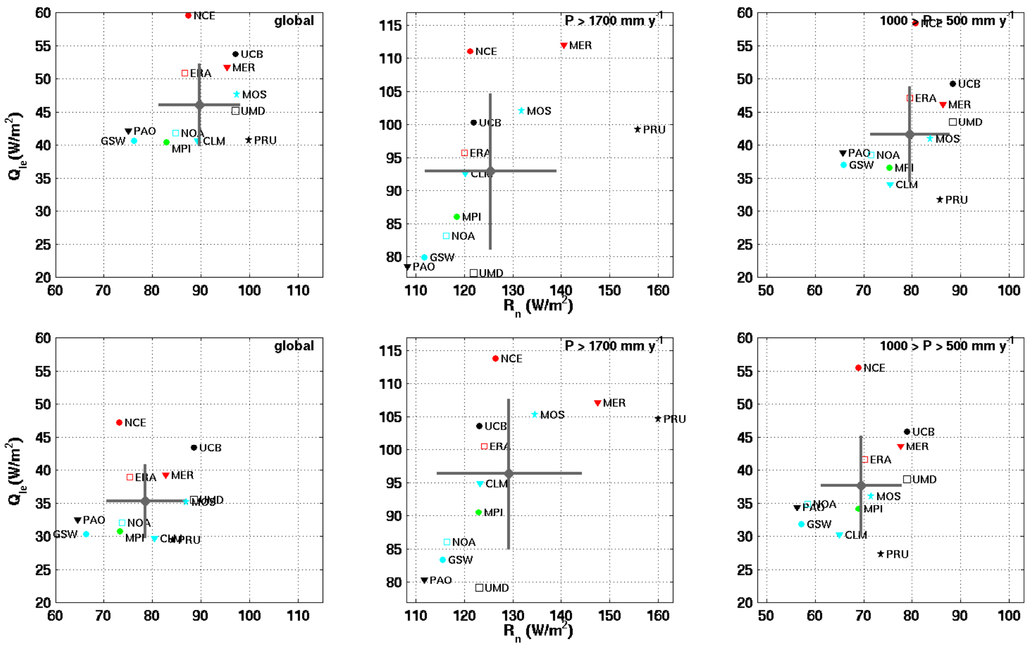

Figure 5. As the Q_{le} versus R_n plots of Figure 1 (products aggreagted onto a 2.5^{o}x2.5^o grid), but here with (top panels) products re-gridded into an equal area grid of $\sim$ 770 km^2 with a lat-lon box of $\sim$ 0.25^{o}x0.25^o at the equator; and (bottom panels) again onto the 2.5^{o}x2.5^o grid, but only for pixels that do not include water bodies, according to a classification derived from a satellite product.

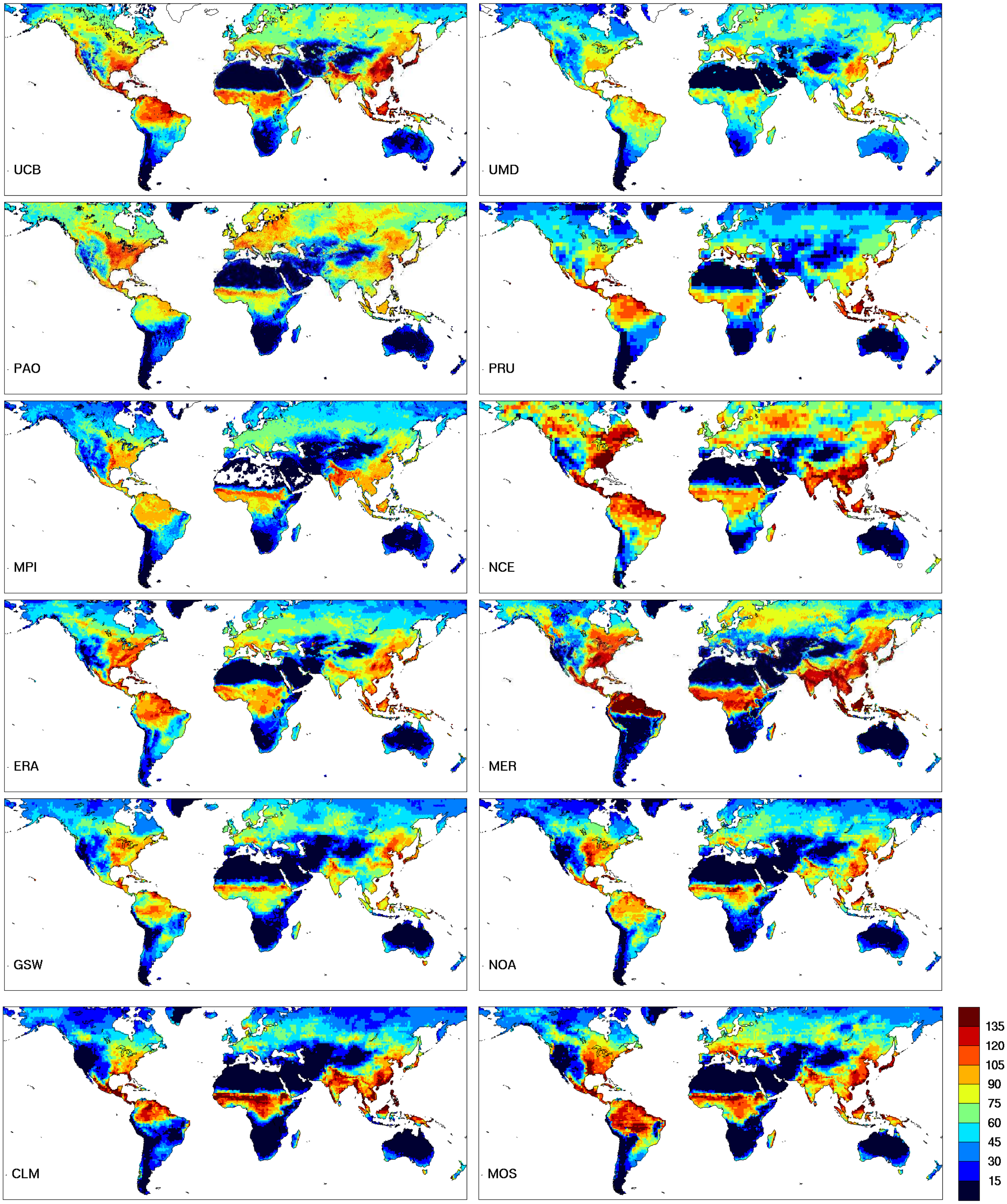

Figure 6. Monthly averaged Q_{le} for August 1994 ($W\,m^{-2}$).

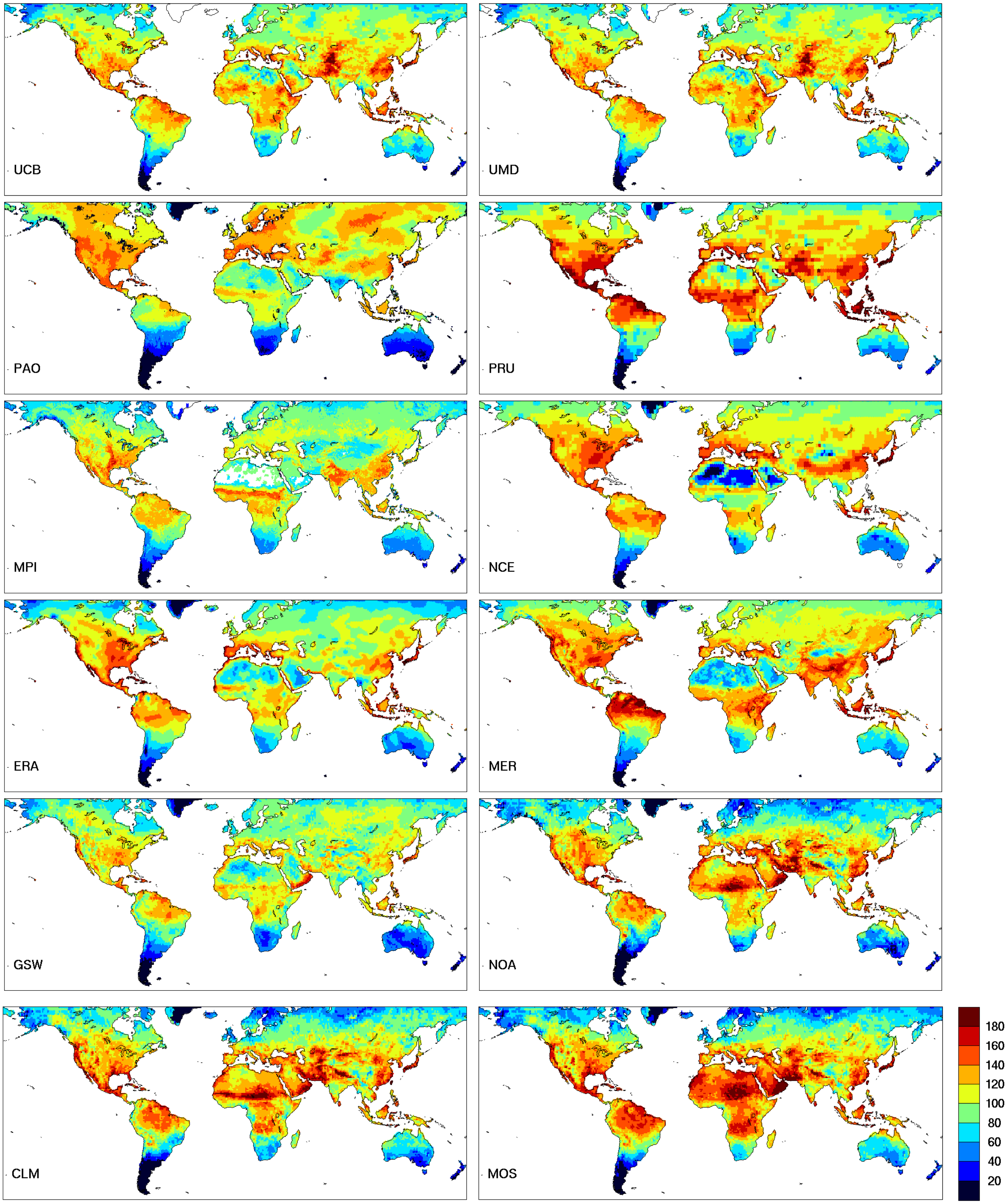

Figure 7. Monthly averaged R_n for August 1994 ($W\,m^{-2}$).

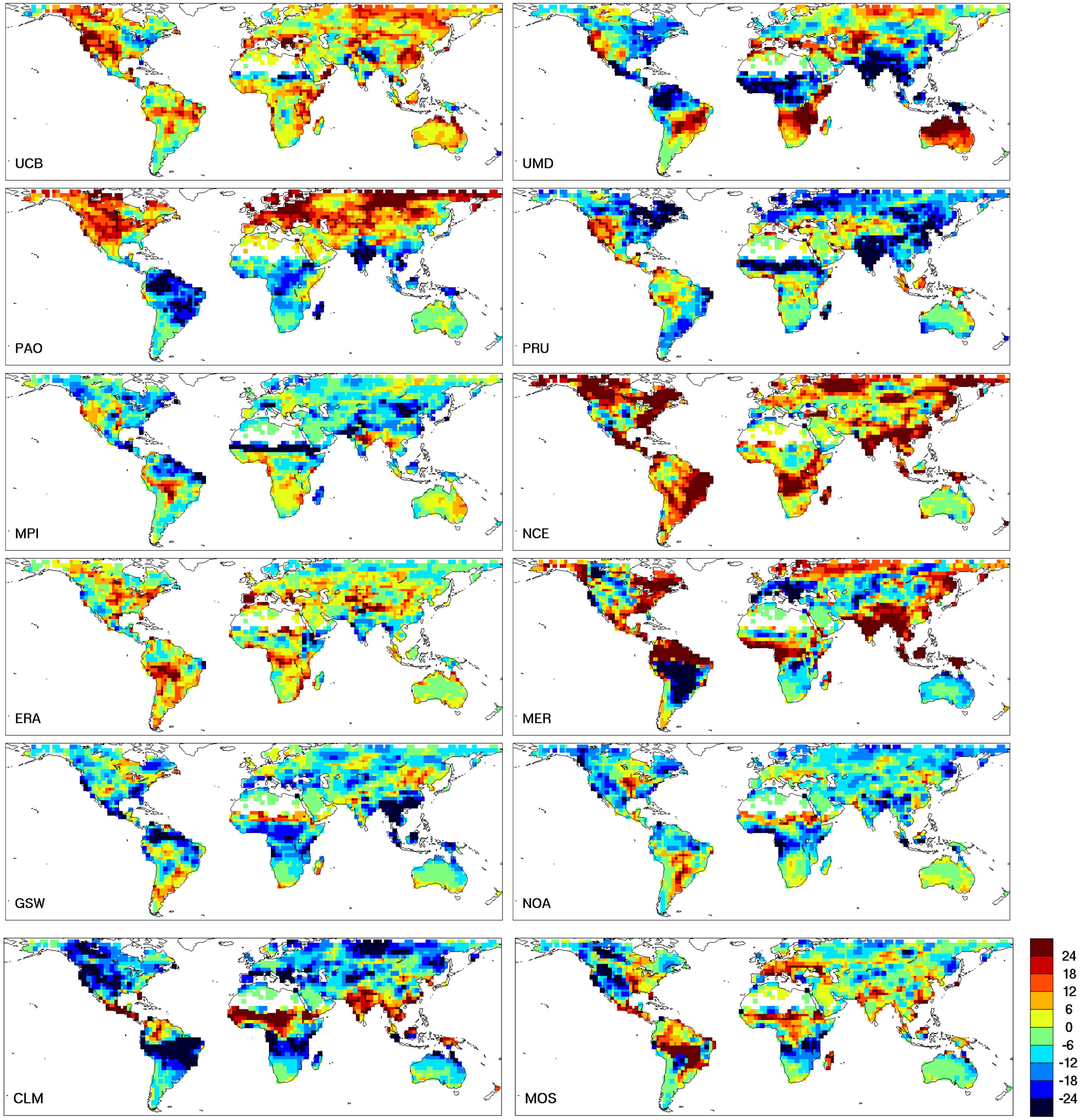

Figure 8. Monthly averaged Q_{le} differences for August 1994 between the products and the all-product ensemble mean ($W\,m^{-2}$).

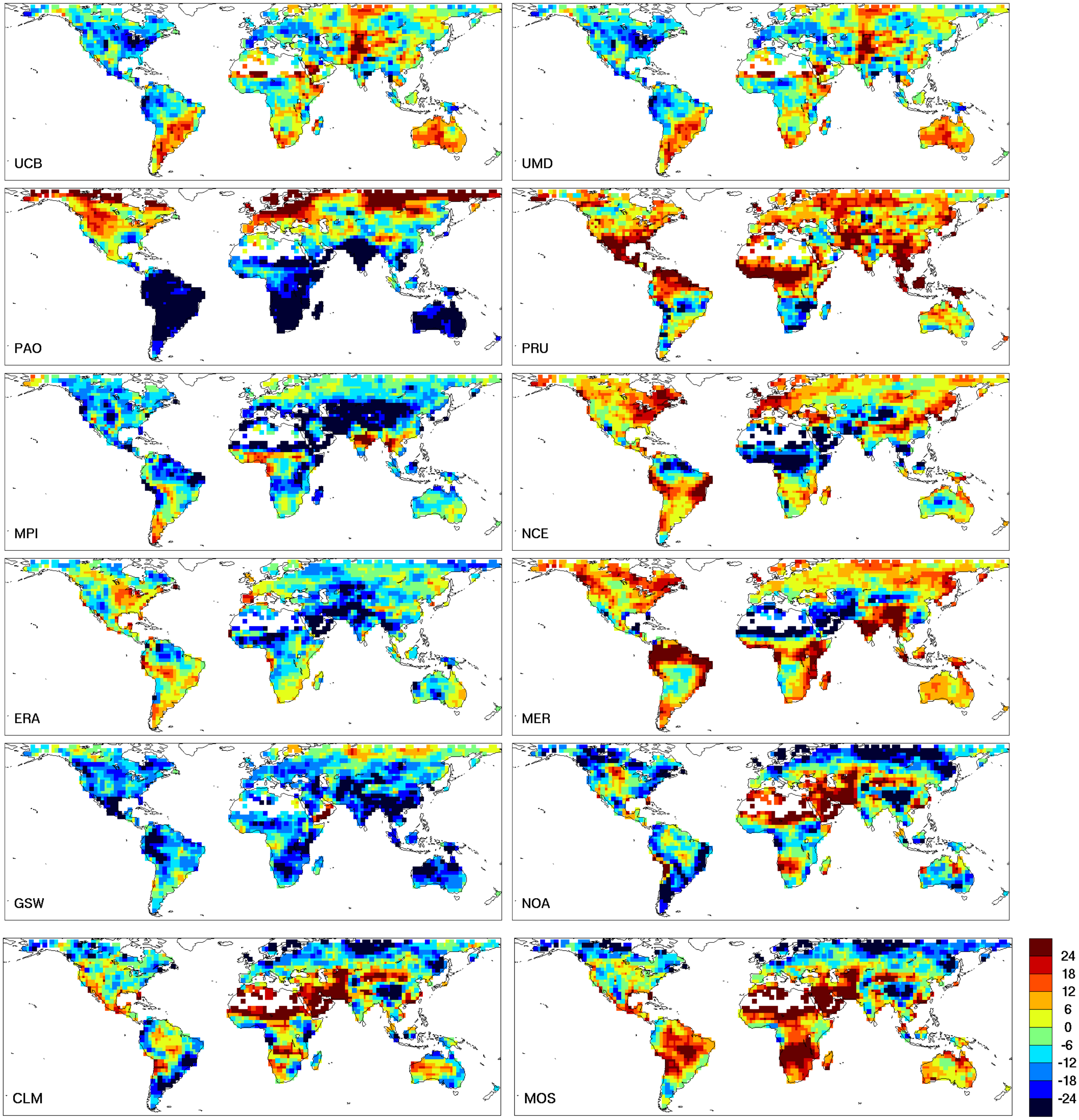

Figure 9. Monthly averaged R_n differences for August 1994 between the products and the all-product ensemble mean ($W\,m^{-2}$).

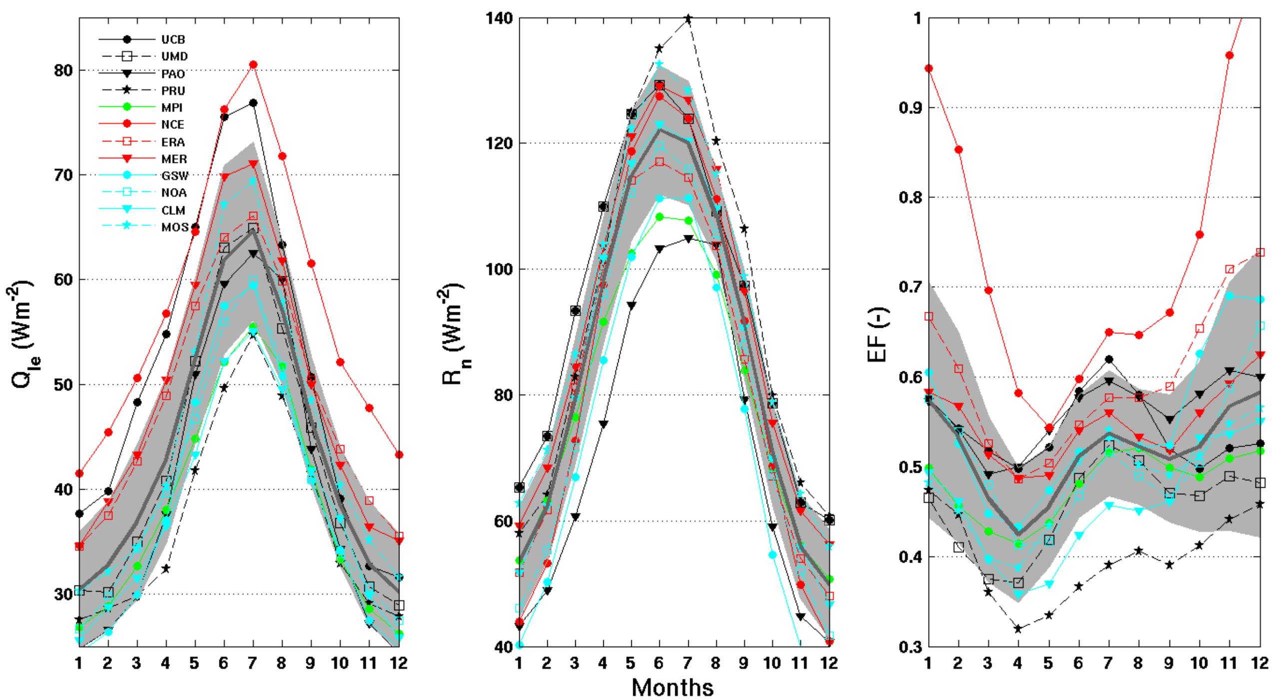

Figure 10. Global 1994 annual cycles of Q_{le} (left), R_n (middle), and EF (right). The grey line and shadow display respectively the ensemble mean and the standard deviation ($\pm\sigma$) of the individual product monthly means around the ensemble mean.

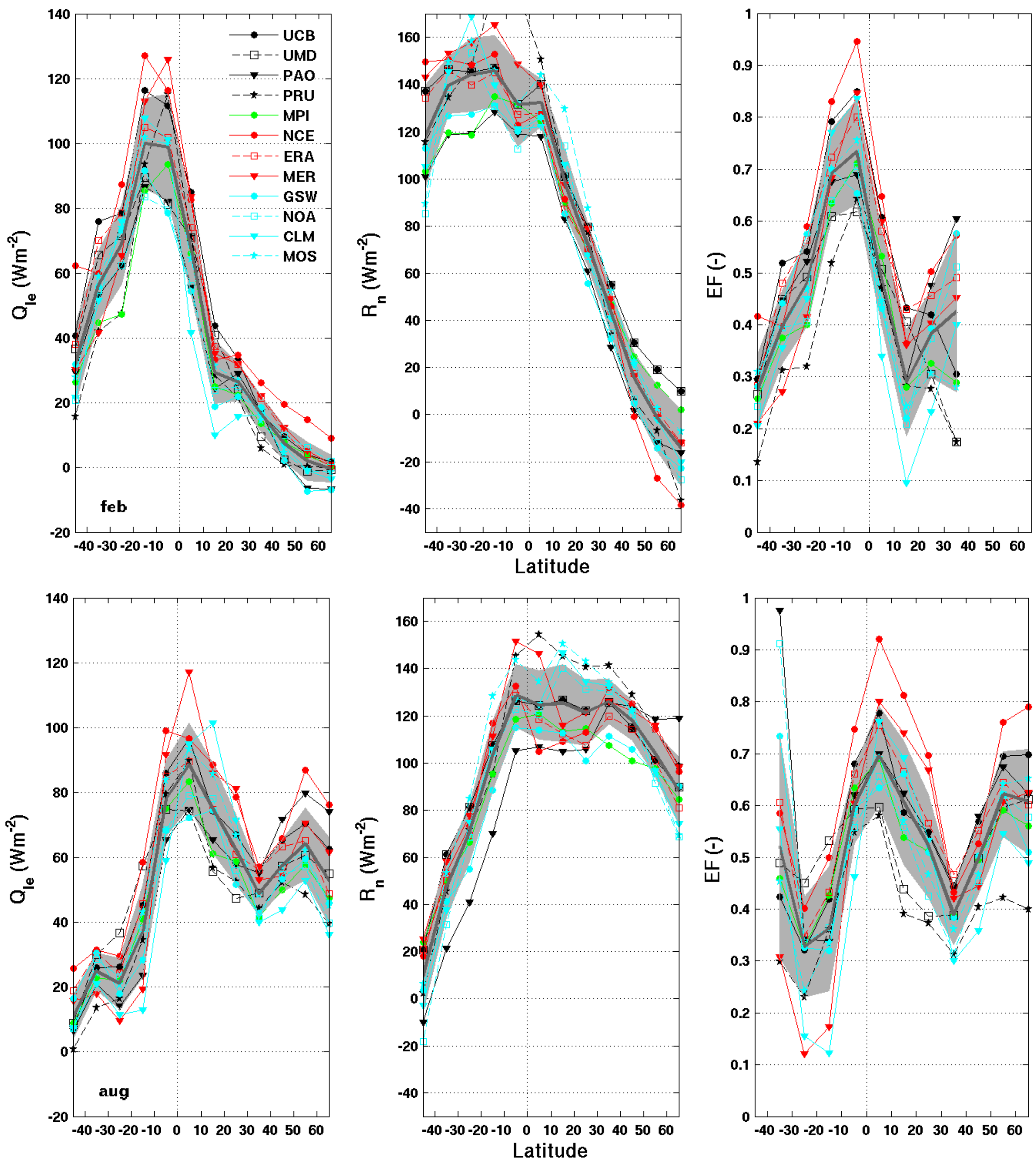

Figure 11. Zonal means of Q_{le} (left), R_n (middle), and EF (right) for February (top) and August (bottom) 1994. The grey line and shadow display respectively the ensemble mean and the standard deviation ($\pm\sigma$) of the individual product annual means around the ensemble mean.

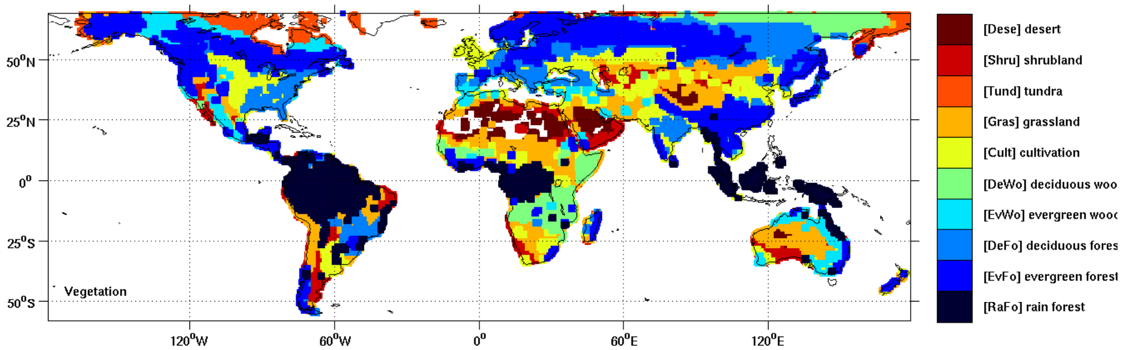

Figure 12. Geographical location of vegetation classes considered in the study.

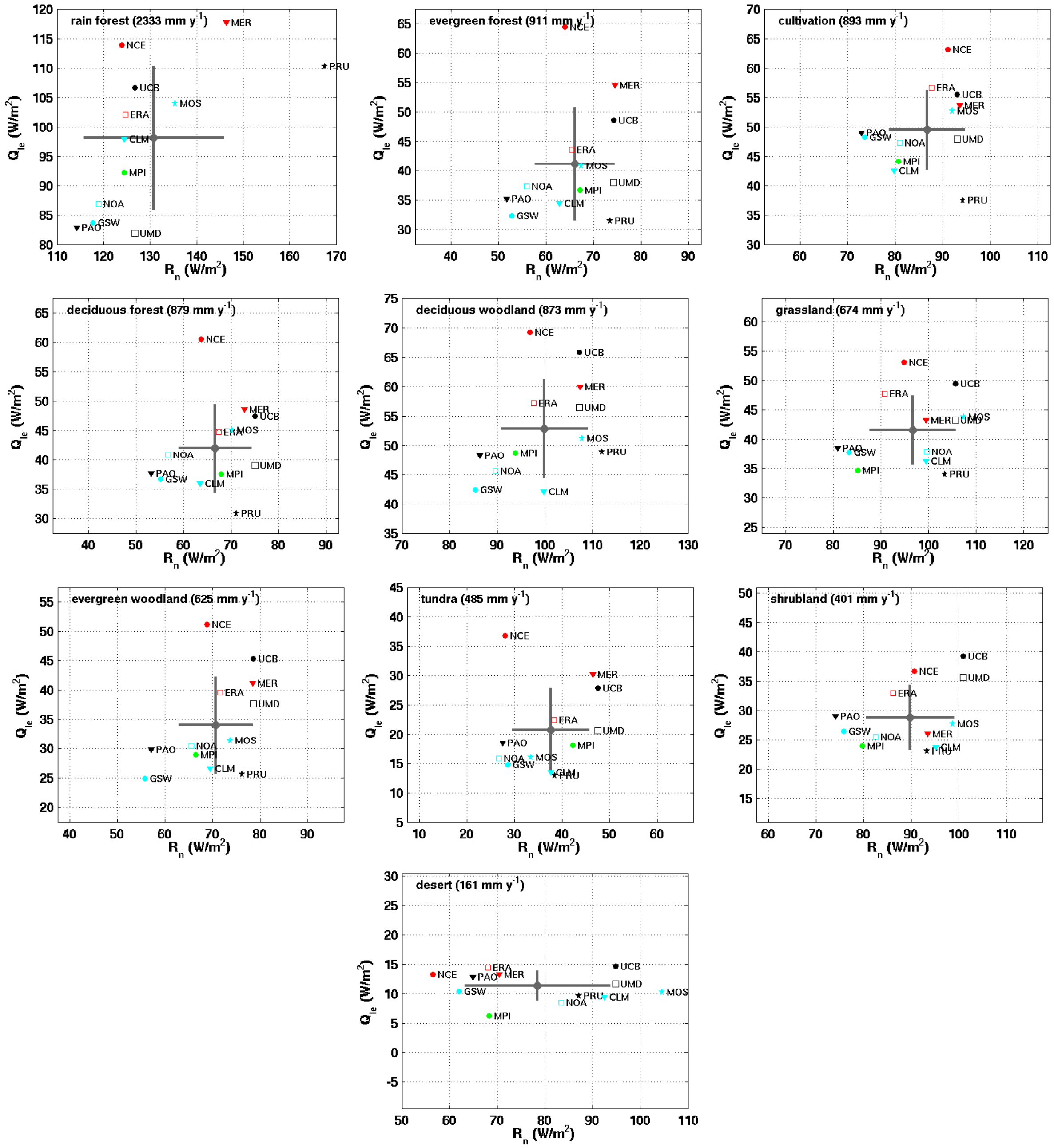

Figure 13. 1994 spatially averaged Q_{le} annual mean as a function of the R_n annual mean for different vegetation classes. The class averaged annual mean precipitation is given close to the class name. The axes scales are different for each class, but they span the same range. The grey dot and lines display respectively the ensemble mean and the standard deviation ($\pm\sigma$) of the individual product annual means around the ensemble mean.

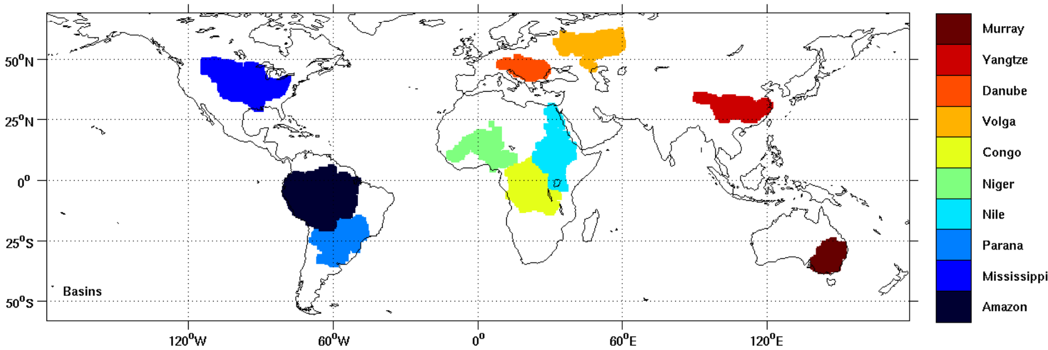

Figure 14. Geographical location of basin areas considered in the study.

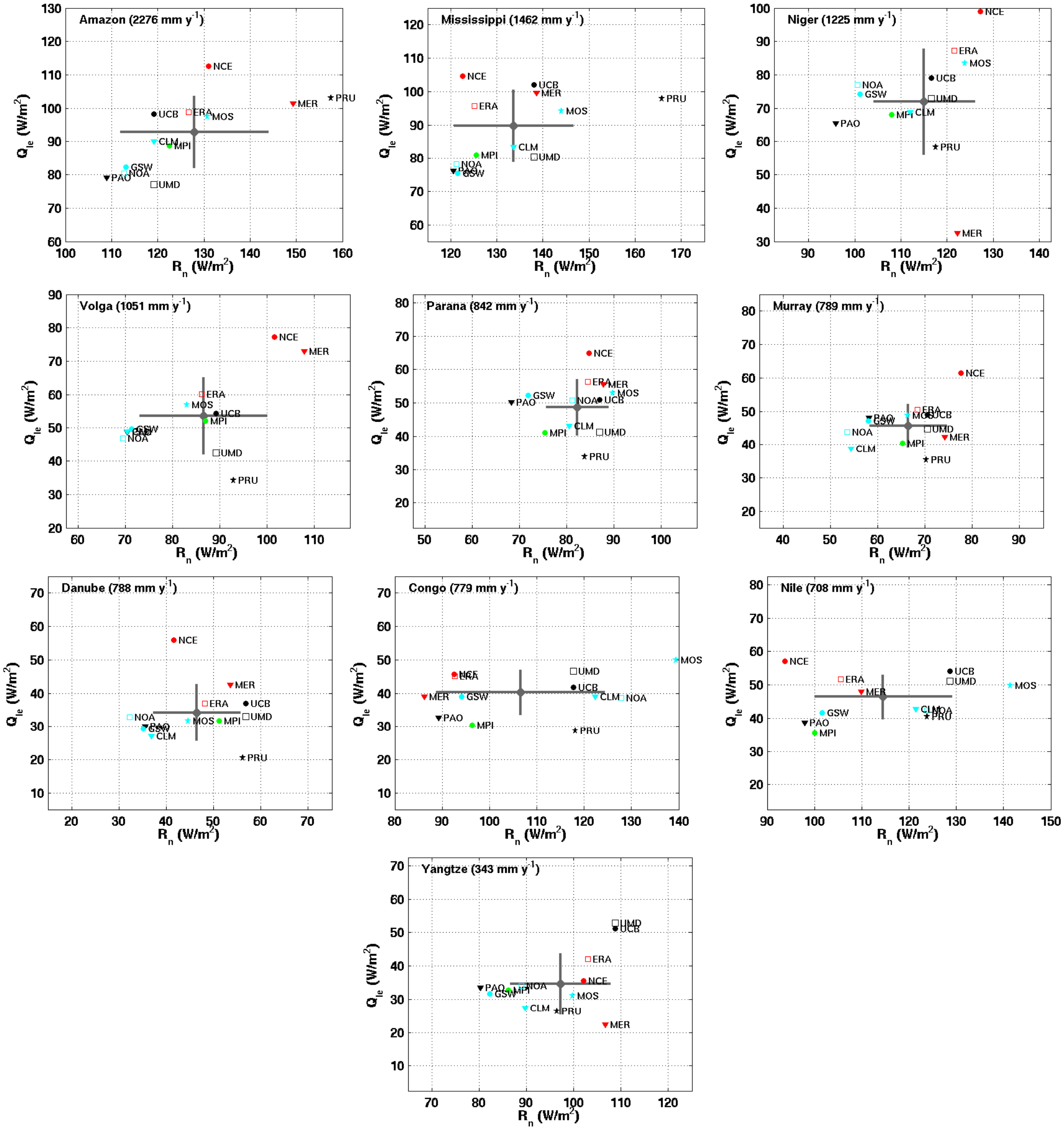

Figure 15. Basin averaged Q_{le} annual mean as a function of the R_n annual mean for the year 1994. The basin averaged annual mean precipitation is given close to the basin name. The axes scales are different for each basin, but they span the same range. The grey dot and lines display respectively the ensemble mean and the standard deviation ($\pm\sigma$) of the individual product annual means around the ensemble mean.

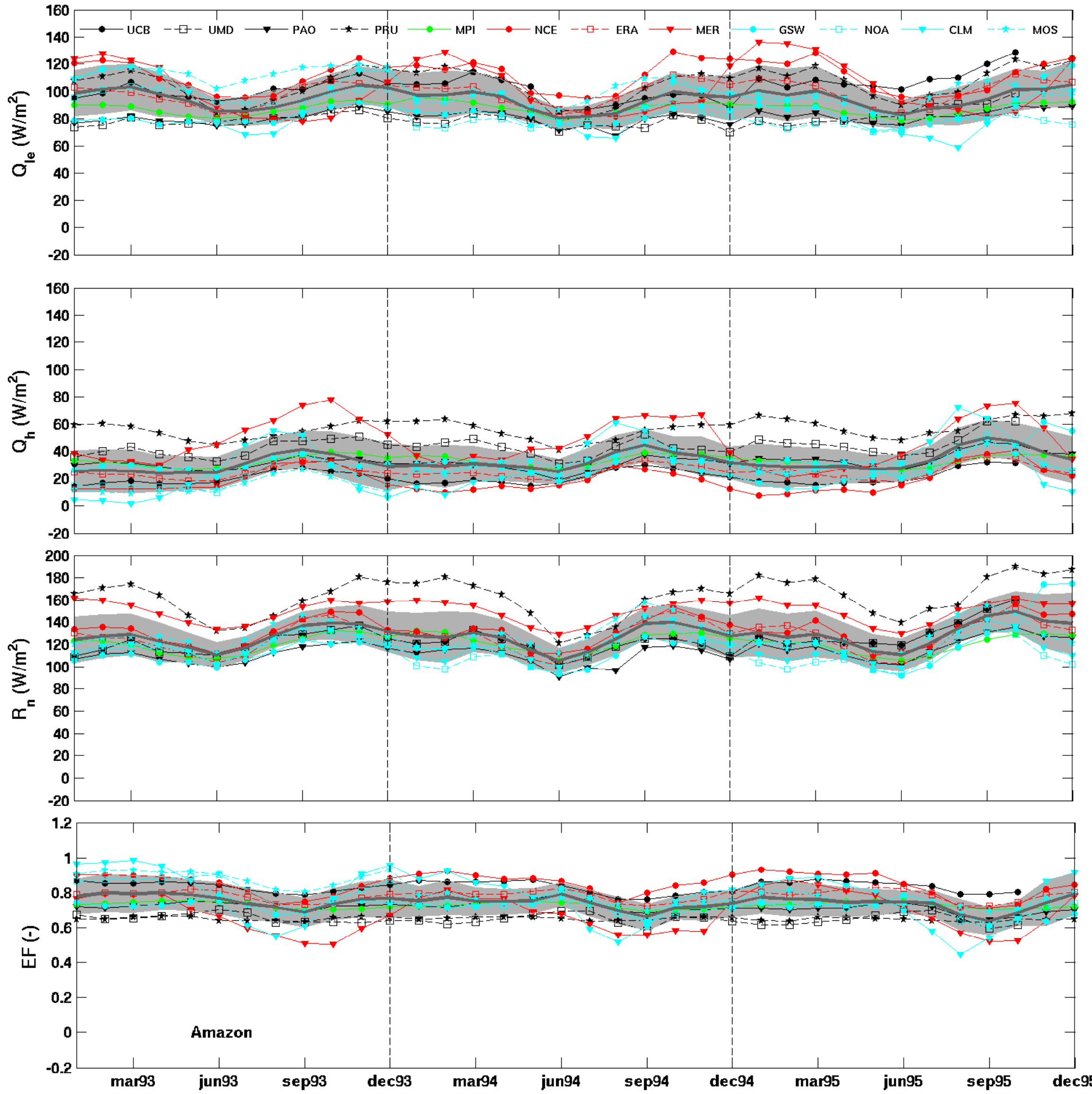

Figure 16. Spatially averaged monthly time series for the Amazon basin. From top to bottom: Q_{le}, Q_h, R_n, and EF.

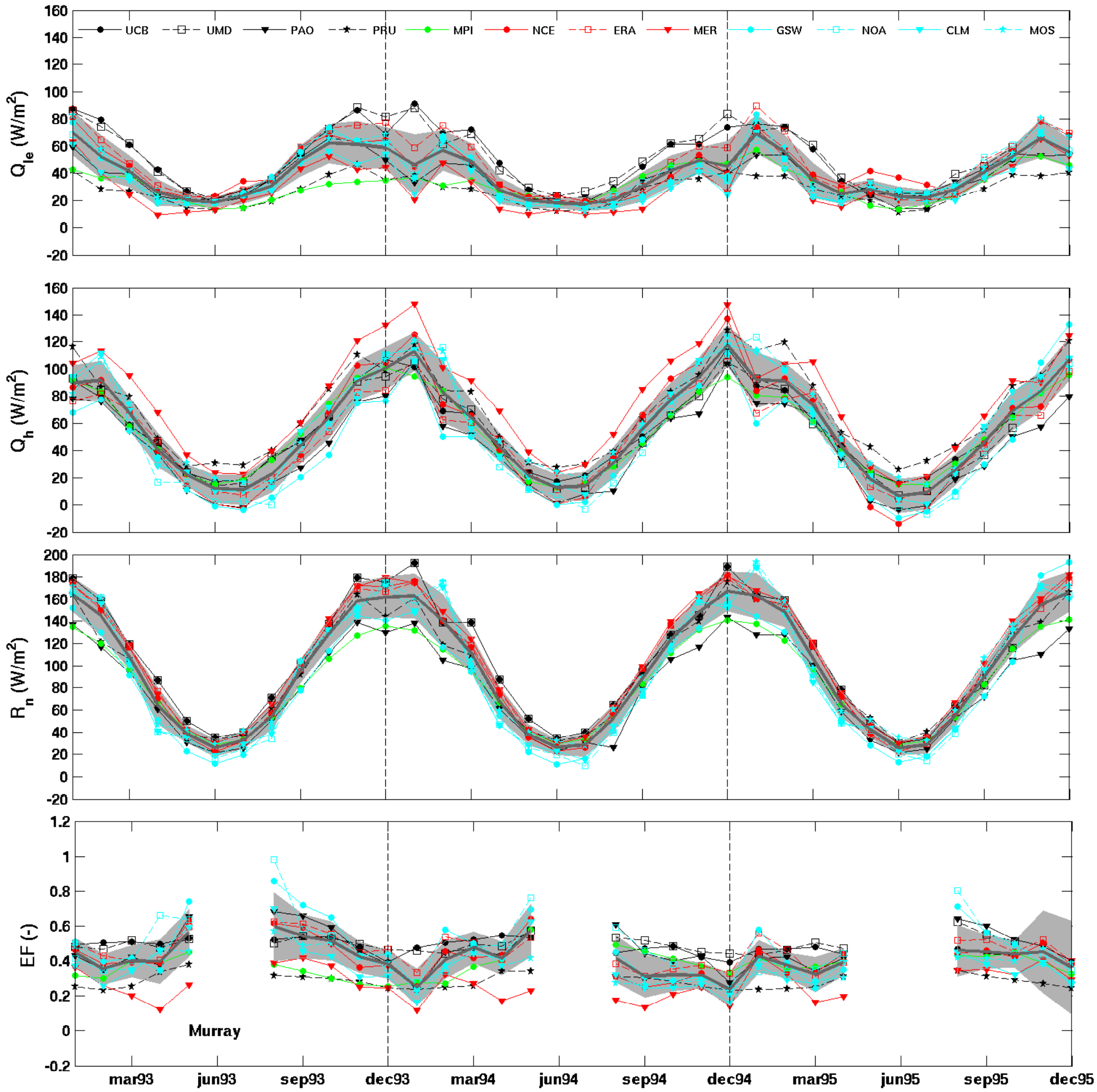

Figure 17. As Figure 16, but or the Murray basin.