

Evaluation of Satellite-Retrieved Extreme Precipitation over Europe using Gauge Observations

M. LOCKHOFF

Deutscher Wetterdienst, Offenbach, Germany

O. ZOLINA

*Le Laboratoire de Glaciologie et Géophysique de l'Environnement, Grenoble, France,
and P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia*

C. SIMMER

Institute for Meteorology, University of Bonn, Bonn, Germany

J. SCHULZ

European Organisation for the Exploitation of Meteorological Satellites, Darmstadt, Germany

(Manuscript received 26 March 2013, in final form 26 August 2013)

ABSTRACT

Climate change is expected to change precipitation characteristics and particularly the frequency and magnitude of precipitation extremes. Satellite observations form an important part of the observing system necessary to monitor both temporal and spatial patterns of precipitation variability and extremes. As satellite-based precipitation estimates are generally only indirect, however, their reliability has to be verified.

This study evaluates the ability of the satellite-based Global Precipitation Climatology Project One-Degree Daily (GPCP1DD) dataset to reliably reproduce precipitation variability and extremes over Europe compared to the European Daily High-resolution Observational Gridded Dataset (E-OBS). The results show that the two datasets agree reasonably well not only when looking at climatological statistics such as climatological mean, number of wet days (rain rates ≥ 1 mm), and mean intensity (i.e., mean over all wet days) but also with respect to their distributions. The results also reveal a pronounced seasonal cycle in the performance of GPCP1DD that is worse in winter and spring. Both deterministic and fuzzy verification methods are used to assess the ability of the GPCP1DD dataset to capture extremes. Fuzzy methods prove to be the better suited evaluation approach for such a highly variable parameter as precipitation because it compensates for slight spatial and temporal displacements. Whereas the deterministic diagnostics confirm previous findings on the deficiencies of satellite products, the “fuzzy” results show that at larger spatiotemporal scales (e.g., 3°/5 days) GPCP1DD has useful skill and is able to reliably represent the spatial and temporal variability of extremes.

1. Introduction

The Clausius–Clapeyron relationship suggests that the observed warming of the troposphere in response to increasing greenhouse gas concentrations intensifies the hydrological cycle (Trenberth 1999). According to Trenberth et al. (2003) and Allan and Soden (2008)

precipitation will not only increase on average, but more importantly also in intensity, leading to enhanced contributions of heavy and extreme events to total precipitation. Such a change will have large impacts on many societal areas such as water resource management, agriculture, and infrastructure planning, for example, for mitigating flood risk. Owing to the stochastic nature of precipitation long-term observations with high temporal and spatial resolution are required to document, analyze, and improve our understanding of past precipitation variability and changes. This in turn will help to improve

Corresponding author address: M. Lockhoff, DWD, Frankfurter Strasse 135, D-63067 Offenbach, Germany.
E-mail: maarit.lockhoff@dwd.de

model predictions of future precipitation variability and extremes that form the basis for the development of strategies of adaptation and mitigation.

Precipitation estimates over land areas were typically derived from surface rain gauge observations at automated or human-operated sites. The main advantage of these data is their long-term temporal coverage (e.g., Brien et al. 2013). In most parts of the world they extend back to the early decades of the twentieth century or even earlier. However, surface rain gauge observations are very inhomogeneously distributed in space and often suffer from a large fraction of missing data resulting in inadequate temporal and spatial sampling even over relatively densely sampled areas like Europe. Consequently, continental-scale studies on precipitation variability and trends based on rain gauge observations (e.g., Klein Tank and Können 2003; Zolina et al. 2009) seldom provide useful estimates of variability patterns except over areas with high-quality long-term dense observation networks.

Satellite-based precipitation estimates, which are available since the late 1970s, provide spatially homogeneous observations with almost global coverage. These estimates are, however, indirect because they rely on the interpretation of emitted or scattered radiation received by the satellite instruments. Retrieval algorithms can be categorized according to the type of radiation exploited (Kidd and Levizzani 2011): 1) scattered solar [visible (VIS) and near infrared (NIR)] and emitted infrared (IR) radiation, 2) emitted microwave radiation [passive microwave (PMW)] and backscattered radar-emitted microwave radiation [active microwave (AMW)], and 3) multisensor methods using a mixture of 1 and 2. VIS-, NIR-, and/or IR-based algorithms use information on cloud properties as proxy for rainfall, taking advantage of the relationship between rainfall and the visible brightness of clouds, the cloud microphysical properties derived from VIS/NIR, and IR observations (e.g., Lensky and Rosenfeld 2006; Roebeling and Holleman 2009), or the cloud-top temperature obtained from IR observations [e.g., Geostationary Operational Environmental Satellite Precipitation Index (GPI) by Arkin and Meisner (1987)]. VIS-/NIR- and/or IR-based retrievals are particularly utilized for observations from geostationary orbit [e.g., Meteosat and Geostationary Operational Environmental Satellite (GOES)] and the data products therefore benefit from the high temporal resolution.

The PMW techniques have the closest relation to precipitation processes and can be further divided into emission- (used over ocean) and scattering-based (used over land) algorithms depending on the processes exploited for the retrievals. The main challenges for PMW algorithms are the detection and quantification of

precipitation over land, especially over cold, snow, ice, and desert surfaces and the detection of light rain (Ferraro et al. 1998). The relatively long wavelengths still prohibit, however, the deployment of microwave sensors on geostationary orbits, thus instruments are only found on board low-earth-orbiting (LEO) satellites [e.g., Special Sensor Microwave Imager (SSM/I); Advanced Microwave Scanning Radiometer (AMSR-E)], which allow only low temporal resolutions in the range of hours. Examples of algorithms include the SSM/I operational precipitation rate algorithm from Ferraro (1997) and the Bayesian Rain Algorithm including Neural Networks (BRAIN) algorithm from Viltard et al. (2006).

AMW instruments provide the most direct information on precipitation from satellites. So far, the Tropical Rainfall Measuring Mission (TRMM) Precipitation Radar (PR) is the only instrument in orbit specifically designed for retrieving precipitation. PR algorithms exploit the backscatter of microwave radiation from precipitation, which is roughly proportional to the particle density weighted with the particle diameter raised to the power of six. The narrow swath width of the TRMM PR combined with the low temporal sampling characteristics of LEO satellites prohibit, however, the use of TRMM PR stand-alone precipitation products for climatological studies. But, owing to its high-quality rainfall estimates, the PR has been extensively used as part of multisensor retrievals, in particular with PMW instruments [e.g., Bauer 2001; Tropical Amount of Precipitation with an Estimate of Errors (TAPEER)–BRAIN product as described in Chambon et al. (2013)].

Other multisensor methods exploit synergies of geostationary VIS/IR and PMW algorithms to generate estimates with higher temporal and spatial resolution than possible with single instruments [in addition to the algorithms mentioned above, the Climate Prediction Center morphing technique (CMORPH) (Joyce et al. 2004) and Global Precipitation Climatology Project one-degree daily (GPCP1DD) (Huffman et al. 2001)]. For more details on satellite-based rainfall estimation methods, the reader is referred, for example, to Adler et al. (2001), Kidd (2001), Levizzani et al. (2002), and Kidd and Levizzani (2011).

The considerable efforts in algorithm development and the exploitation of sensor synergy has led in the meantime to datasets that are relatively long (covering at least one decade) and mature enough to be used for analyses of the interannual variability of precipitation and even provide the required spatial and temporal resolution to effectively capture precipitation extremes, which is the focus of this study.

To estimate the value of satellite datasets for climatological studies, it is crucial to evaluate the results against

high-quality ground-based observations. Such an evaluation is necessary to reveal both the strengths and weaknesses of the satellite-based estimates and might even lead to further improvements of retrieval algorithms and satellite sensors. This is particularly important assuming that the role of satellite measurements in observing precipitation, which form part of the essential climate variables (ECV) listed by the Global Climate Observation System (GCOS), will increase in the future. Many validation studies for satellite-based precipitation estimates have been carried out using ground-based observations, for example in the framework of the International Precipitation Working Group (IPWG), which was established in 2001 in order to coordinate the improvement of satellite retrieval algorithms (e.g., Ebert et al. 2007; Turk et al. 2008; Sapiano and Arkin 2009). Mid- to high-latitude regions present a special challenge for satellite retrievals owing to the occurrences of snow and low precipitation intensities, as well as cold/frozen surface backgrounds. Recent validation studies carried out for this region include, for example, Bolvin et al. (2009), who compared both monthly and daily estimates of the GPCP1DD dataset against rain gauge data in Finland, and Kidd et al. (2012), who studied the performance of high-resolution precipitation products (HRPP) in northwestern Europe by comparison to radar and rain gauge observations. Both found seasonality in the skill of the satellite retrieval algorithms to retrieve precipitation with a poorer performance during winter. Bolvin et al. found GPCP1DD to underestimate precipitation in summer and to overestimate precipitation in winter compared to gauge data. All four satellite HRPPs (one IR-only and three multisensor retrievals) studied by Kidd et al. underestimate precipitation over northwestern Europe in all seasons (compared to radar).

The aim of this study is to assess the ability of the GPCP1DD rainfall estimates (Huffman et al. 2001) over Europe to accurately replicate precipitation extremes compared to rain-gauge-based estimates from the European Daily High-Resolution Observational Gridded Dataset (E-OBS) (Haylock et al. 2008). GPCP1DD was chosen as one of the few daily datasets available for the midlatitudes covering a time period of more than 10 years. Extremes are defined in this study as rain accumulations exceeding a high (e.g., the 90th or 95th) percentile. By using GPCP1DD and E-OBS over Europe we will assess 1) the comparability of basic climatological statistics of precipitation (means, intensities, and number of wet days), 2) the comparability of the characteristics of the empirical frequency distributions of daily precipitation as revealed by satellite data and rain gauge observations, and 3) the ability of GPCP1DD to replicate precipitation extremes and to effectively

capture their variability. Both deterministic (point by point) and fuzzy verification methods (Ebert 2008) are applied.

The paper is divided into five sections. The GPCP1DD and E-OBS datasets are described in section 2, section 3 gives details on the analysis methodology applied, and in section 4 the comparison results are presented and discussed. A summary is given in section 5 together with some concluding remarks.

2. Data

a. GPCP1DD

The Global Precipitation Climatology Project (GPCP) was established by the World Climate Research Programme (WCRP) in 1986. The GPCP product suite includes the monthly satellite-gauge (SG) (Adler et al. 2003; Huffman et al. 2009), the pentad (Xie et al. 2003), and the One-Degree Daily (1DD) (Huffman et al. 2001) datasets, with the latter being the one used in this study. The GPCP1DD dataset provides daily precipitation estimates at One-Degree Daily latitude/longitude spatial resolution with the GPCP day defined as midnight to midnight UTC. We use the GPCP1DD version 1.1, which covers the period from October 1996 to 2009.

Between 40°N and 40°S, the GPCP1DD algorithm uses SSM/I precipitation estimates together with geosynchronous IR (GEO-IR) and low-earth-orbit IR (LEO-IR) brightness temperatures to derive the so-called threshold-matched precipitation index (TMPI). The TMPI is an adaptation of the GOES Precipitation Index (GPI) (Arkin and Meisner 1987), which assigns a constant conditional rain rate of 3 mm h^{-1} to all pixels that have temperature values lower than 235 K and zero rain rate to all others. For the TMPI the IR temperature threshold is set locally by month using the SSM/I-based precipitation frequency and a single (local) conditional rain rate based on the monthly GPCP SG product (Huffman et al. 2001). The latter is generated by first using the higher accuracy of low-orbit PMW observations to calibrate the more frequent GEO-IR observations and then adjusting in a second step the resulting combined satellite-based product using the Global Precipitation Climatology Centre (GPCC) rain gauge analysis.

Outside the latitudinal band from 40°N to 40°S—and thus over the main part of the area of interest of this study—precipitation estimates are computed based on recalibrated Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) and Atmospheric Infrared Sounder (AIRS) data from polar-orbiting satellites. TOVS data is used for the time span up to March 2005 and is then replaced by AIRS in the

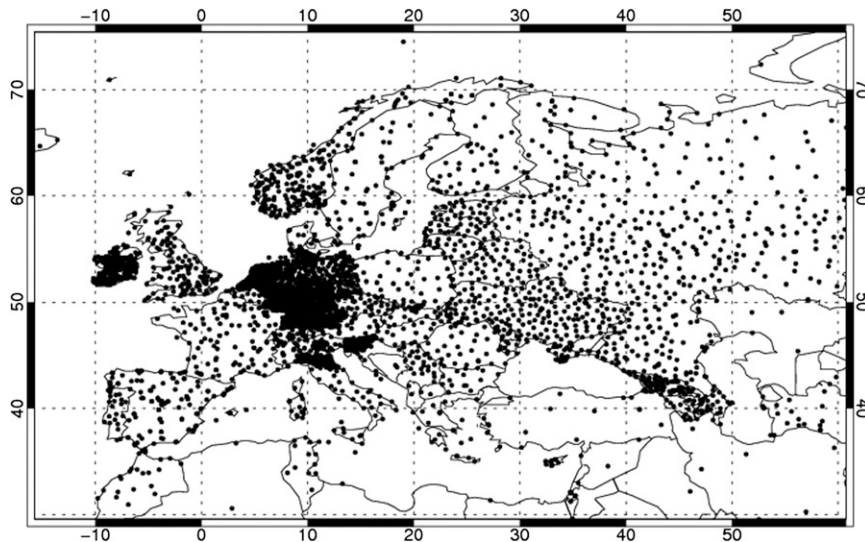


FIG. 1. The complete gridding region (land only), showing the station network for precipitation used for generating E-OBS version 5.

time period from April 2005 onward. The TOVS data up to February 1999 are based on information from two satellites. Afterward the TOVS and AIRS estimates are based on information from one satellite only, namely *NOAA-14* and subsequently *Aqua*. The algorithm uses a multiple regression relationship between collocated rain gauge observations and several TOVS–AIRS-retrieved cloud-volume-related quantities such as cloud-top pressure, fractional cloud cover, and relative humidity profile (Susskind and Pfaendtner 1989; Susskind et al. 1997). These quantities are retrieved using IR information provided by the High-Resolution Infrared Radiation Sounder version 2 instrument (HIRS/2). The regression relationship varies by latitude, month, and surface type (land or ocean). As the number of wet days was found to be systematically high compared to TMP1, the frequency of wet days in the TOVS–AIRS-based product was scaled down to match that of TMP1 at the data region boundaries (Huffman et al. 2001). In both data regions, the GPCP1DD product is scaled to sum up to the monthly accumulation provided by the GPCP SG product. This ensures consistency between the GPCP monthly and daily products and includes wind-loss-adjusted gauge information (Huffman et al. 2001). To prevent abrupt changes in the precipitation fields across the data boundaries at 40°N and 40°S a gradual smoothing has been applied to the TOVS–AIRS product over the latitude band 40°–50° based on the differences computed for the 39°–40°N and 39°–40°S grid boxes.

b. E-OBS

Direct comparison of precipitation products providing area averages such as satellite-based products to

rain gauge data representing point observations in the context of extreme events is problematic. In our study the GPCP1DD precipitation statistics are therefore validated with the E-OBS dataset provided by the Royal Netherlands Meteorological Institute (KNMI) (Haylock et al. 2008). E-OBS provides gridded land-only daily precipitation amounts and minimum, maximum, and mean screen-level (2 m) temperatures over Europe for the period 1950–2010. The dataset covers the land areas within 25°–75°N, 40°W–75°E and is available both on a 0.25° and 0.5° regular latitude–longitude grid and on a 0.22° and 0.44° rotated pole grid. Within this study the version 5 dataset with 0.5° spatial resolution is used.

The E-OBS dataset is based on daily observations from approximately 4500 land stations (see Fig. 1). According to the authors, the dataset provides a best estimate of gridbox averages rather than point values in order to enable direct comparisons with products providing real areal averages such as regional climate model output or satellite-derived datasets. For precipitation, the interpolation applied in E-OBS comprises three steps (Haylock et al. 2008): 1) interpolation of the monthly totals using three-dimensional thin-plate splines, 2) interpolation of the daily anomalies using indicator and universal kriging, and 3) a combination of steps 2 and 3. As an interpolation uncertainty, daily standard errors are provided for every grid. It is important to keep in mind that at gauge stations daily accumulations of precipitation are mostly reported from 0900 to 0900 UTC (Haylock et al. 2008) and are therefore shifted 9 h from the midnight-to-midnight UTC GPCP1DD day. Another important difference between the gauge- and

satellite-based precipitation estimates is that the latter rely on the accumulation of “instantaneous” estimates (in case of the single satellite LEO-based estimates derived from March 1999 onward only two per day), whereas the gauge-based estimates represent true accumulations.

Uncertainties of the E-OBS dataset are mainly associated with three major problems. First is the inaccuracy of daily station data due to instrumental errors and errors associated with observational practices. The latter also includes the underreporting of daily precipitation owing to spurious zeros and incomplete records, which may result in a negative bias if not removed. Second is the inhomogeneity of the station distribution, and third, the temporally varying number of gaps in the observational data. The first problem is tackled within the generation of the E-OBS dataset by applying a series of quality checks on the raw station observations (Haylock et al. 2008). This partly reduces these errors but does not remove them completely. Thus, this uncertainty propagates to the E-OBS grids. The two latter problems additionally affect the gridbox area-average estimates in three main ways (Haylock et al. 2008; Hofstra et al. 2010): representativeness (estimate will not be a “true” areal average), smoothing (contributions by using stations outside the grid box for gridbox estimates), and variable degree of smoothing depending on station density across the grid domain. The representation of extremes is in particular influenced by (over and under) smoothing; therefore, Hofstra et al. (2010) recommended treating estimates in areas of sparse station density with special caution.

c. Data processing

In a first step both datasets—GPCP1DD and E-OBS—were brought to the same resolution. To this end the daily precipitation sums of E-OBS were averaged onto the $1^\circ \times 1^\circ$ grid, that is, the grid at which the GPCP1DD dataset is provided. Furthermore, both datasets were confined to the region of interest extending over 35° – 70° N, 10° W– 40° E that encompasses several climate zones ranging from maritime to continental and semiarid to temperate.

To assure a comparable quality level (with respect to representativeness and minimization of the smoothing effect) between E-OBS grid values and thereby following the recommendations of Hofstra et al. (2010) mentioned above, the study area was further confined to grid boxes including at least one station or having a minimum of four neighboring grid boxes (within the 3×3 gridbox neighborhood around the central grid box) with at least one station available over the whole period of analysis. After applying this quality criterion a broad spread in the station density still remains, but the relative standard error of E-OBS for high rain rates was reduced to 25% (figure not shown here). Our

study is based on 11 years of daily precipitations accumulations ranging from January 1998 to December 2008.

3. Methodology

a. General precipitation evaluation

For the evaluation of precipitation statistics from E-OBS and GPCP1DD we used the threshold of 1 mm day^{-1} to distinguish between wet and dry days; that is, wet days are defined as those days with precipitation totals $\geq 1 \text{ mm}$. This is justified by large uncertainties of estimation of very light precipitation in station data and in satellite measurements (Klein Tank and Können 2003; Zolina et al. 2010). For the general evaluation the following diagnostics are used:

- MEAN (mm day^{-1}): average over all days,
- INT (mm day^{-1}): average over all wet days,
- NWET: number of wet days,
- Q90, Q95, and so on (mm day^{-1}): 90th, 95th, and so on, percentile estimated from the empirical (wet day) distribution functions,
- bias ratio (–): the ratio of average GPCP1DD to average E-OBS,
- quantile–quantile plots (QQ plots; see Wilks 2006),
- cumulative frequency distributions of daily precipitation.

The three first diagnostics provide standard measures of precipitation. Empirically estimated percentiles are used to quantify the skill of representing precipitation extremes. The bias ratio is used to quantitatively compare the results of the two datasets. The QQ plots provide information on the empirical quantiles of the E-OBS- and GPCP1DD-based wet-day time series. They serve to assess the consistency between the ground- and satellite-based distributions of precipitation. Finally, we used the nonparametric Kolmogorov–Smirnov test (KS test; see Wilks 2006) in order to further investigate and quantify the comparability of the statistical structure of the two daily precipitation datasets.

All results have been computed at each grid box so as to avoid the pooling of data over areas belonging to different climate zones. Additionally, results are presented separately per season to account for the changing precipitation type and weather regimes, which influence the performance of the satellite-based estimates but also the uncertainty and representativeness of the in situ measurements. The seasons are defined as winter: December to February (DJF), spring: March to May (MAM), summer: June to August (JJA), and autumn: September to November (SON).

b. Extreme precipitation assessment

Extreme precipitation events are defined as daily totals exceeding high (e.g., 90th and 95th) percentiles. These percentile thresholds were calculated from the sample of all wet days in the analyzed time period (i.e., 1998 to 2008). Percentiles were chosen instead of absolute thresholds because they are more easily comparable between different climate regions (see Groisman et al. 2005). We used two approaches, namely deterministic (point by point) and fuzzy verification to assess how well extreme events are represented by the GPCP1DD dataset compared to station data.

1) DETERMINISTIC APPROACH

Traditionally, gridded precipitation products are compared using deterministic verification methods, which are based on simple (spatial and temporal) point-by-point matching. To assess the agreement between the occurrence of extreme events seen by the satellite dataset on the one hand and the ground-based dataset on the other, the respective rain rates are transformed to binary (yes–no) indicators of extreme events (using a given extreme threshold). These matched indicator grid pairs are then counted to complete the standard contingency table (Wilks 2006) from which common evaluation measures like frequency bias, the probability of detection (POD), false alarm rate (FAR), threat score (TS), and equitable threat score (ETS) can be estimated (e.g., Wilks 2006). We use quantile-specific POD and quantile-specific FAR as proposed by AghaKouchak (2011) for the deterministic verification. The quantile-specific POD and FAR values were calculated based on the time series of the E-OBS and GPCP1DD for each grid box as follows:

- The quantile probability of detection (QPOD) is defined as the POD above a certain percentile (or quantile) threshold (here, e.g., Q90 and Q95 representing the 90th and 95th percentile, respectively) and is equal to the ratio of the number of precipitation events being correctly detected as exceeding a given threshold to the total number of precipitation occurrences above the same threshold in the reference. QPOD ranges between 0 and 1, with 1 indicating the perfect score. Thresholds are calculated separately for each of the datasets.
- The quantile false alarm rate (QFAR) is defined as the FAR above a certain percentile (or quantile) threshold. QFAR is equal to the ratio of the number of precipitation events being falsely indicated as exceeding a given threshold to the total number of correct and false occurrences over the same threshold as indicated by the reference. QFAR ranges between 0 and 1, with 0 indicating the perfect score. As for QPOD,

thresholds are calculated separately for each of the datasets.

2) FUZZY APPROACH

Fuzzy verification or neighborhood methods as described in Ebert (2008) aim at relaxing the requirement for exact matching by allowing slight displacements. The maximum displacement allowed is defined by a local neighborhood (or window) around the grid box of interest. In this study, both spatial and temporal displacements are considered. Following this approach a spatiotemporal neighborhood of grid boxes is defined around each central grid box. For example, for a given spatiotemporal scale of 5° and 3 days the neighborhood encompasses $5 \times 5 \times 3 = 75$ grid boxes. The treatment of the neighborhood data depends on the selected fuzzy method and includes, for example, averaging, thresholding, or the generation of empirical frequency distributions.

From the available fuzzy methods we chose the fractions skill score (FSS) (see also Roberts and Lean 2008). This score directly compares the fractional coverage of events (here extreme events as defined above) in the given spatiotemporal neighborhood defined as the ratio of the number of grid boxes in the spatiotemporal neighborhood where the extreme event occurs to the total number of valid neighborhood grid boxes. A dataset shows useful skill if the fraction of events, as seen by, for example, the satellite product, is similar to the fraction obtained from the ground-based product. The calculation of the FSS encompasses the following steps. 1) For each selected space–time scale pair (1° and 1 day, 1° and 3 days, 3° and 3 days, etc.) and thresholds (e.g., Q90) the daily precipitation accumulations from GPCP1DD and E-OBS were converted to fractions of extreme events. 2) These fractions were then used to compute a fractions Brier score (FBS), which is defined in Ebert (2008):

$$\text{FBS} = \frac{1}{N} \sum_N (\langle P_y \rangle_s - \langle P_x \rangle_s)^2, \quad (1)$$

where $\langle P_x \rangle$ is the fraction of grid boxes in a neighborhood with extreme events observed by GPCP1DD; $\langle P_y \rangle$ is the fraction of the neighborhood with extreme events observed by E-OBS, where $\langle \rangle_s$ indicate that the fractions are calculated based on the neighborhood surrounding the grid box of interest for the indicated spatiotemporal scale; N is the number of neighborhoods in the domain considered. In this study, the FBS is calculated per grid box so that the FBS is calculated for the temporal domain (i.e., the time period covered). Therefore, N is equal to the total number of days per season for

11 years [e.g., 1012 days (and thus also 1012 neighborhoods) for the summer season]. 3) FBS and $\langle P_x \rangle$ and $\langle P_y \rangle$ are then used to calculate the FSS (Ebert 2008):

$$\text{FSS} = 1 - \frac{\text{FBS}}{N^{-1} \left(\sum_N \langle P_y \rangle_s^2 + \sum_N \langle P_x \rangle_s^2 \right)}. \quad (2)$$

The FSS ranges between 0 and 1 with 1 indicating the perfect score. The value of FSS above which the assessed dataset is considered to have useful (better than random) skill is given by

$$\text{FSS}_{\text{usefull}} = 0.5 + \frac{f_y}{2}, \quad (3)$$

where f_y is the domain average fraction observed by the reference dataset (Roberts and Lean 2008); that is, here the average fraction of extreme events observed by E-OBS at a specific grid point over the entire time period.

The values of the extreme thresholds were calculated per grid box at a 1° resolution. With increasing size of the spatial neighborhood, the extreme threshold values were averaged over the spatial neighborhood. With increasing (spatial) size the neighborhood window will cross the boundaries of the study area and, therefore, also include no-data values. So, a neighborhood was only considered when at least 50% of the neighborhood grid boxes provided valid values. This leads to a decrease in the size of the area along the boundaries with increasing spatial scale.

4. Results

a. Climatological statistics

Based upon 11 years of daily rainfall estimates, climatological means of several basic statistics were calculated for both the GPCP1DD and E-OBS datasets. In Figs. 2 and 3, regional maps of mean precipitation (MEAN), mean wet-day intensity (INT), mean number of wet days per year (NWET), and the long-term 90th percentile of wet days (Q90) for E-OBS (left) and GPCP1DD (middle) are presented for winter and summer, respectively. Bias ratio maps (GPCP1DD/E-OBS) are shown on the right-hand side. Table 1 lists the corresponding summery statistics of these four diagnostics (MEAN, INT, NWET, and Q90) for the entire domain during the time period 1998 to 2008. Qualitatively, the spatial patterns compare satisfactorily for all parameters. In general, the typical features of the respective season are captured well. During the winter season, the weather in Europe is characterized by prevailing westerly winds

with associated frontal systems. Therefore, the highest precipitation amounts are found along the coasts of western Europe and the Mediterranean, for example, along the western coast of Norway, the northern part of Great Britain, the northwestern part of Spain, and the east of the Adriatic coast. Over these areas MEAN, INT, and Q90 reach values of >4 , >10 , and $>20 \text{ mm day}^{-1}$, respectively. The lowest values of MEAN, INT, and Q90 are found in eastern Europe (Romania and Hungary), as well as in western Spain. Summertime weather (Fig. 3) is dominated by convection. MEAN and NWET clearly increase from south to north. The highest values for MEAN, INT, and Q90 (>3.5 , >9 , and $>17 \text{ mm day}^{-1}$, respectively) occur over the alpine region (stretching from southeastern France to Romania), which is the region of highest convective activity in Europe. During both seasons, NWET increases from south to north with NWET ranging between 0 and 45 days in summer and between 7 and 50 days in winter.

Good agreement between E-OBS and GPCP1DD should be expected for mean precipitation since daily estimates of GPCP1DD are scaled to match the monthly accumulation provided by the GPCP SG product. Over areas with dense gauge networks, such as Europe, the spatial distribution of precipitation of the GPCP SG product is dominated by the gauge analysis of the GPCC product. One possible reason for differences between the GPCC and E-OBS products (and thus between GPCP1DD and E-OBS for mean precipitation) is the difference in the number of stations used. In general, GPCC includes more stations than E-OBS; for example, in 2008 GPCC and E-OBS include about 5000 and 2700 stations, respectively. Most of the additional GPCC gauges are situated in Germany, France, and Great Britain. Despite the difference in station density, the MEANs of E-OBS and GPCC, in general, agree well with relative differences smaller than 10% both in winter and summer (Fig. 4). Larger differences are only visible at latitude below 40°N and mainly in summer. In contrast to the MEAN, the other three diagnostics (INT, NWET, and Q90) derived from GPCP1DD are mainly based on satellite information. INT (and also Q90) is influenced by MEAN though because daily estimates are scaled to sum up to the monthly means provided by the GPCP SG product.

Despite consistent spatial patterns, absolute values differ by season and region (see bias ratio maps on the right-hand side of Figs. 2 and 3). Except for NWET in winter, GPCP1DD gives higher values than E-OBS for all parameters and seasons over the full spatial domain. Table 2 shows that the overestimation of MEAN, INT, and Q90 by GPCP1DD is highest during winter with 50% to 60% higher values than E-OBS, whereas the values are only 10% to 20% higher during summer.

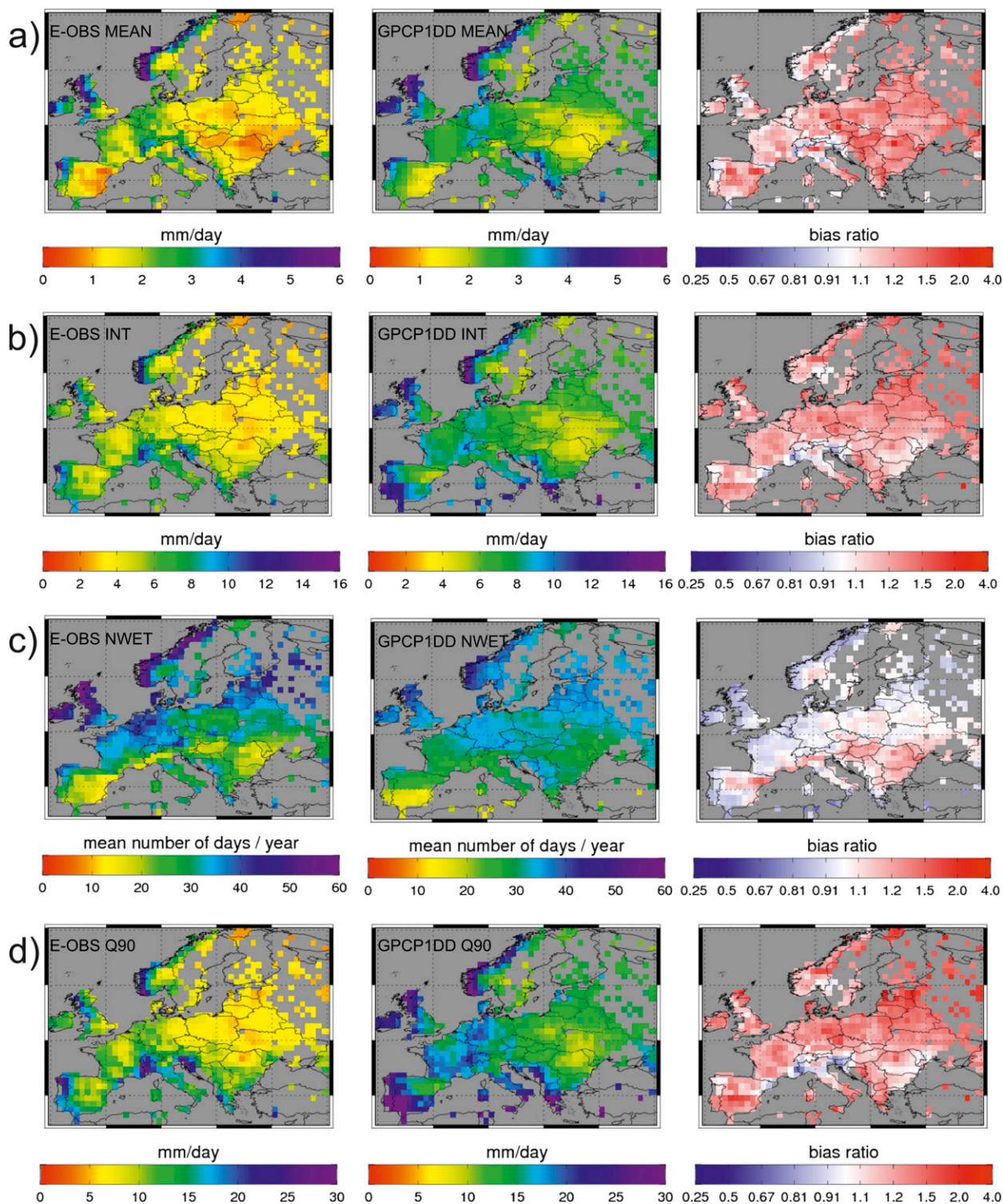


FIG. 2. (a) Climatological mean precipitation (MEAN), (b) mean wet-day intensity (INT), (c) number of wet days (NWET) (i.e., days with precipitation sums to ≥ 1 mm), and (d) 90th percentile of wet days (Q90) based on the winter season (DJF) of the entire time period (1998–2008) for (left) E-OBS results, (middle) GPCP1DD estimates, and the (right) bias ratio (GPCP1DD/E-OBS).

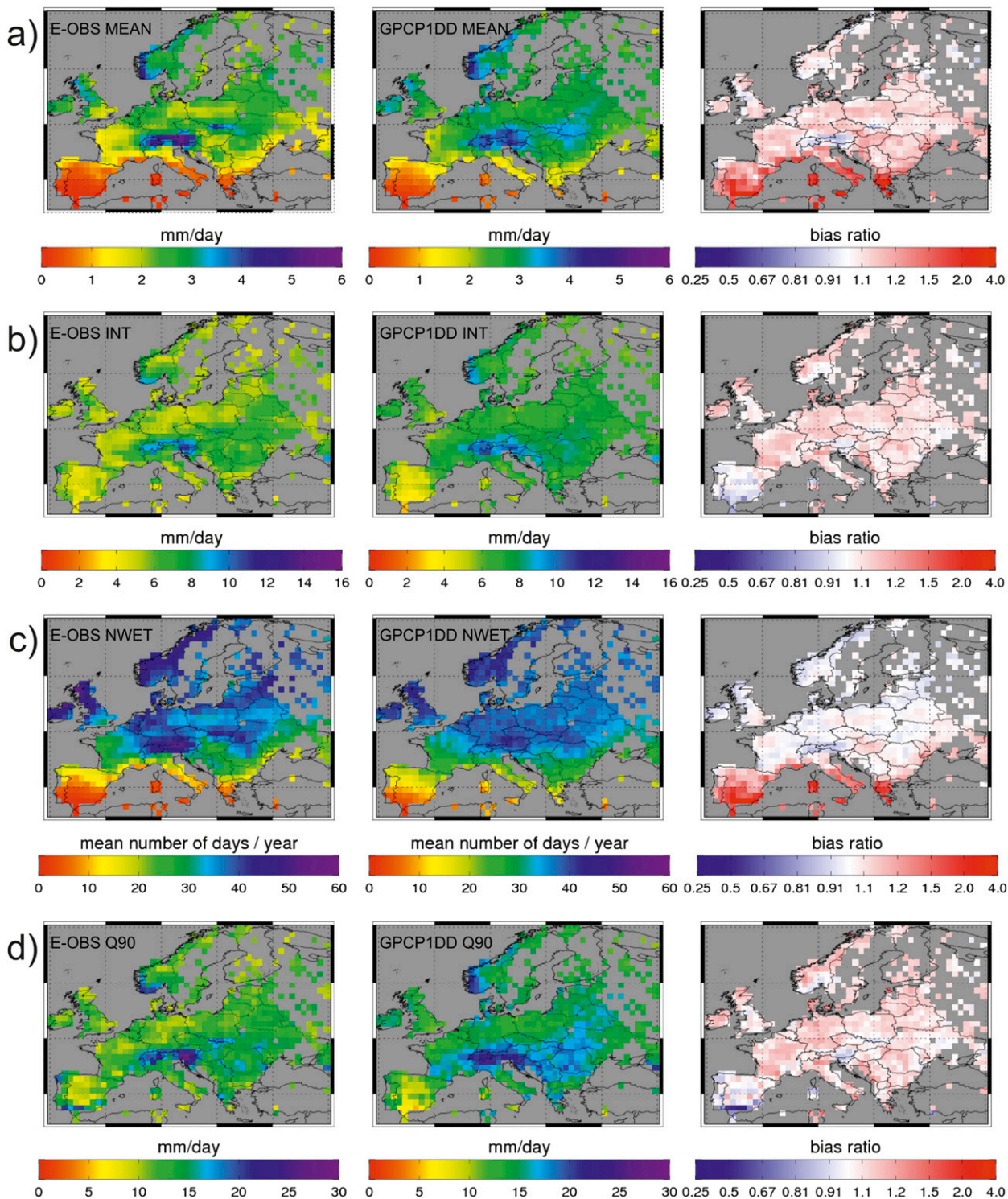


FIG. 3. As in Fig. 2, but based on the summer season (JJA).

In winter, the overestimation of MEAN is largest in the eastern part of the study area. Since both products have similar NWET values, this overestimation translates to higher INT values in GPCP1DD. In the Mediterranean

region GPCP1DD overestimates both MEAN and NWET relative to E-OBS, leading to a closer agreement of INT of both products. The differences due to higher GPCP1DD MEAN values can be explained by the

TABLE 1. Statistics of daily rainfall accumulation (mm day^{-1}) averaged over the entire domain and time period (1998 to 2008) separately per season as DJF (winter), MAM (spring), JJA (summer), SON (autumn): number of grid boxes (N), mean precipitation (MEAN, mm day^{-1}), mean wet-day precipitation intensity (INT, mm day^{-1}) (i.e., on days with rain rate $\geq 1 \text{ mm}$), total number of wet days (NWET), and long-term 90th percentile of wet days (Q90, mm day^{-1}).

Season	N	MEAN		INT		NWET		Q90	
		E-OBS	GPCP1DD	E-OBS	GPCP1DD	E-OBS	GPCP1DD	E-OBS	GPCP1DD
DJF	763 853	1.7	2.5	4.7	7.1	268 642	267 158	10.0	15.5
MAM	772 970	1.6	2.1	4.7	6.0	248 892	263 797	10.2	12.9
JJA	780 625	2.0	2.4	5.9	6.9	260 844	272 651	12.6	14.7
SON	768 453	2.1	2.6	5.8	6.7	264 717	287 394	12.6	14.5

latitude-dependent wind-loss correction applied only to the GPCP1DD product. The differences caused by higher NWET values might relate to unaccounted for effects of surface emissivity in the TOVS–AIRS-based retrieval owing to snow/ice and/or vegetation cover changes (Eyre and Menzel 1989). The performance of the TMPI-based product is not affected by cold/snow surface backgrounds because it is only used up to 40°N , that is, only in regions where snow and ice do not play a role.

In summer, the overestimation of MEAN is much lower and mainly restricted to the drier southern part (Mediterranean region) with only very few wet days ($\text{NWET} < 10$). The spatial patterns of largest overestimation of MEAN mirror the pattern of largest underestimation by E-OBS compared to GPCC shown in Fig. 4 (right). The dry region also shows an overestimation of 50% to 100% for NWET, which only corresponds to a few days. The feature is visible across the data region boundary (40°N) and might be explained by the low temporal resolution of the TOVS–AIRS-based retrieval (thereby missing short-lived events) and by emissivity-related difficulties of both retrievals over dry areas

(Eyre and Menzel 1989; Ferraro et al. 1998). Outside the dry region in the south, relative differences lie below 10%.

The results confirm findings of Bolvin et al. (2009), who found that GPCP1DD tends to overestimate mean precipitation in winter for southern Finland compared to the high-density rain gauge observations provided by the Finnish Meteorological Institute (FMI) with a relative bias of 89% before and still 10% after wind-loss correction of the gauge dataset. In summer, however, Bolvin et al. found GPCP1DD to underestimate precipitation amounts with a relative bias of -20% (after wind-loss correction of the gauges). This discrepancy between these findings and our results may be explained by the station density being higher for the FMI gauge dataset (on average nine stations per $1^\circ \times 1^\circ$ grid box). The latter therefore provides more representative area means, which consequently means that E-OBS underestimates mean precipitation in this area in the summer. Kidd et al. (2012) also found satellite-based products to generally underestimate (instead of overestimate) precipitation over northwest Europe. This can to a large extent also be explained by the references used: Kidd

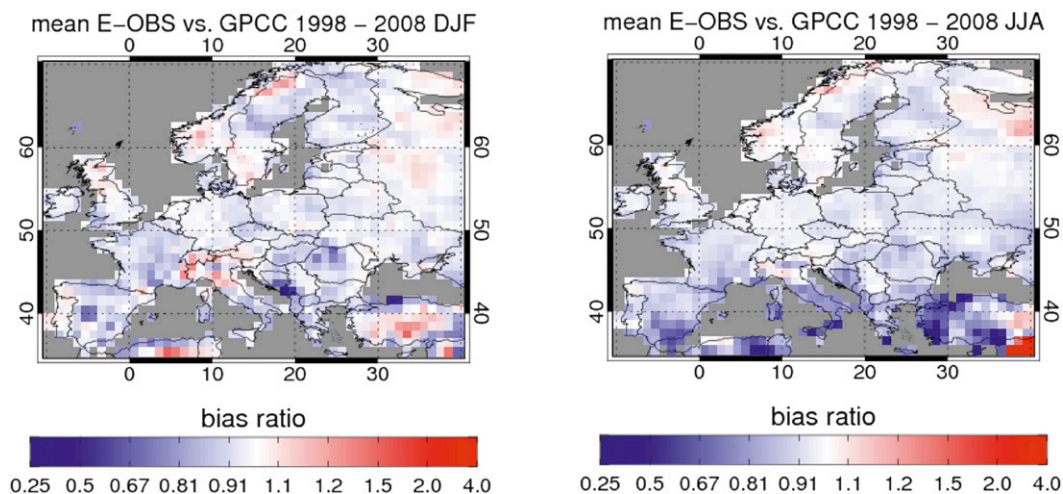


FIG. 4. Bias ratio (GPCC/E-OBS) of mean precipitation for the entire time period (1998–2008) for (left) winter and (right) summer.

TABLE 2. Bias ratio (GPCP1DD-based value divided by E-OBS-based value) for parameters given in Table 1 separately per season.

Season	MEAN	INT	NWET	Q90
DJF	1.46	1.51	0.99	1.55
MAM	1.29	1.23	1.06	1.26
JJA	1.2	1.16	1.05	1.16
SON	1.24	1.15	1.09	1.15

et al. used reference weather radar data that were found to give 65% higher values compared to E-OBS over central Europe (see Roebeling et al. 2012).

Figure 5 displays the QQ plots for E-OBS and GPCP1DD quantiles based on the respective wet-day time series separately for winter (DJF, left) and summer (JJA, right). GPCP1DD overestimates, in general (most in winter and least in summer), the frequency of rain along the entire distribution, which results in an overall overestimation of MEAN, INT, and Q90 as previously seen from the climatological maps (Figs. 2 and 3).

The consistency between the wet-day distributions on the gridbox level was further investigated and quantified using the KS test applied separately by season and year. Based on the seasonal statistics for each year, the fraction of years for which the hypothesis of consistency between the distributions of the two datasets was not rejected at the 0.05% significance level was calculated. Figure 6 shows the regional maps of this fraction of years for the winter and the summer season. In summer,

fractions of 0.8 and higher are found throughout the spatial domain (except for some grid boxes mainly located in the northern part of Great Britain and Norway), which means that for at least 9 out of the overall 11 years the null hypothesis of similar distributions was not rejected. In winter, agreement between the wet-day precipitation distributions is still found in the southern, but not in northern, part of Europe. The latter results from the general overestimation of precipitation rates already seen earlier in the climatological maps in Figs. 2 and 3, mainly stemming from the wind-loss correction applied to the GPCP1DD dataset.

b. Detection of extremes

1) DETERMINISTIC APPROACH

First, we present results of point-by-point comparisons that are traditionally used for the comparison of gridded products. To quantify the ability of GPCP1DD to capture extreme events, scores of QPOD and QFAR were computed for each 1° grid box for individual calendar seasons. Figure 7 shows box-and-whisker plots of QPOD (left) and QFAR (right) for thresholds ranging from Q75 to Q99. For all seasons QPOD decreases with increasing threshold from median values around 0.4 for Q75 to 0.2 for Q95. At the same time QFAR increases with increasing threshold (respective median values increase from 0.6 to 0.8). For all thresholds the QPOD (QFAR) values are highest (lowest) in autumn. These results confirm the findings of AghaKouchak et al. (2011),

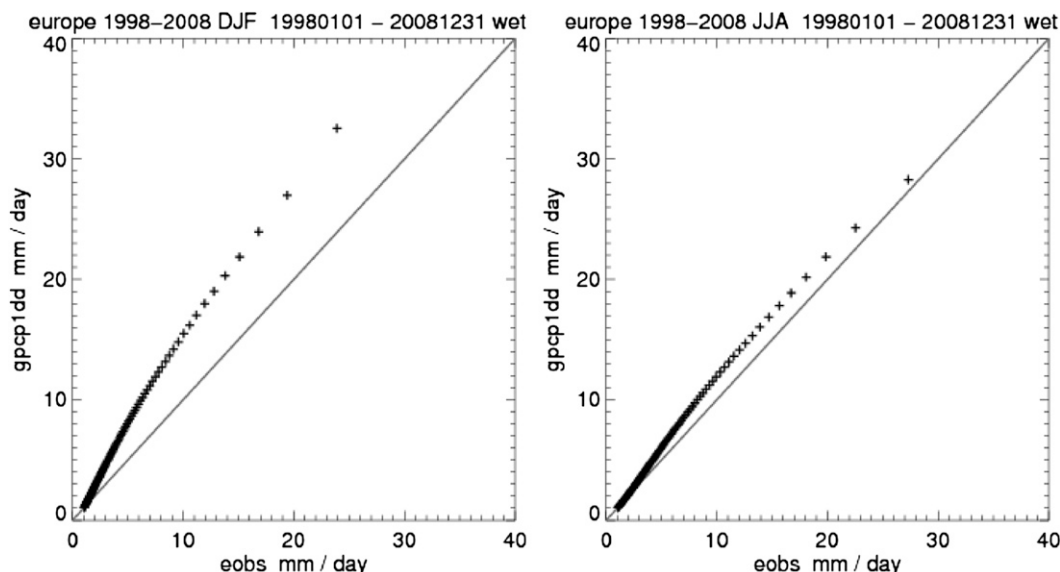


FIG. 5. QQ plots of E-OBS (x axis) and GPCP1DD (y axis) quantiles separately for (left) winter (DJF) and (right) summer (JJA) based on the respective wet-day time series. Points correspond to the respective 1st, 2nd, ..., 99th quantiles.

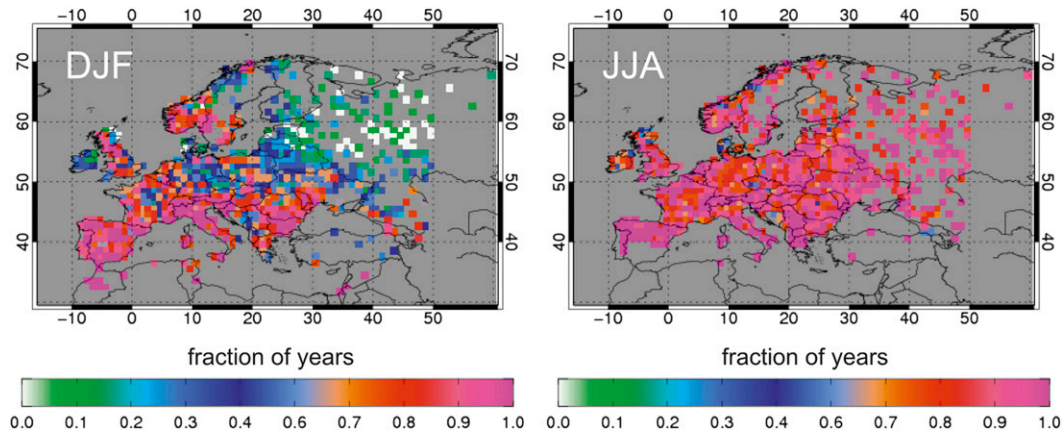


FIG. 6. Regional map of the fraction of years for which the consistency between the wet-day precipitation distributions of the GPCP1DD and E-OBS datasets was not rejected based on the 0.05% significance level (based on annual calculation of the Kolmogorov–Smirnov test statistic and the respective p value), separately per season: (left) DJF and (right) JJA.

who assessed four satellite-retrieved precipitation products (including one microwave-only and three multiple-source products) with respect to their ability to detect extreme precipitation over the United States using radar-based, gauge-adjusted precipitation estimates as a reference. They found (on the basis of a deterministic verification approach) poor scores for all products analyzed so that none of them qualified for having skill for the detection of extremes.

These poor scores are not too surprising considering that precipitation is highly variable in space and time, requiring a measurement system with both high temporal and high spatial resolution. Precipitation products derived from available observing systems fulfill this requirement only to a variable degree. Differences between the products considered in this study stem to a large extent from differences in spatial and temporal sampling:

gauges provide theoretically good temporal sampling (but may suffer from missing values), while their spatial sampling can be considered rather poor. The TMP1-based part of the GPCP1DD product provides both good spatial and temporal sampling owing to the GEO-IR data being used (though it only covers the latitude band between 40°N and 40°S). LEO-based single-sensor products like the TOVS–AIRS-based data product (for the years from 1999 onward), however, can only provide two observations per day. This low temporal sampling results in precipitation fields showing a speckled “salt and pepper” pattern compared to the smooth spatial patterns of the E-OBS product. This speckled pattern also translates into the extreme precipitation frequency patterns and can result in spatial mismatches of extent and/or location of extreme rain events as seen by E-OBS and GPCP1DD. Besides the differences in

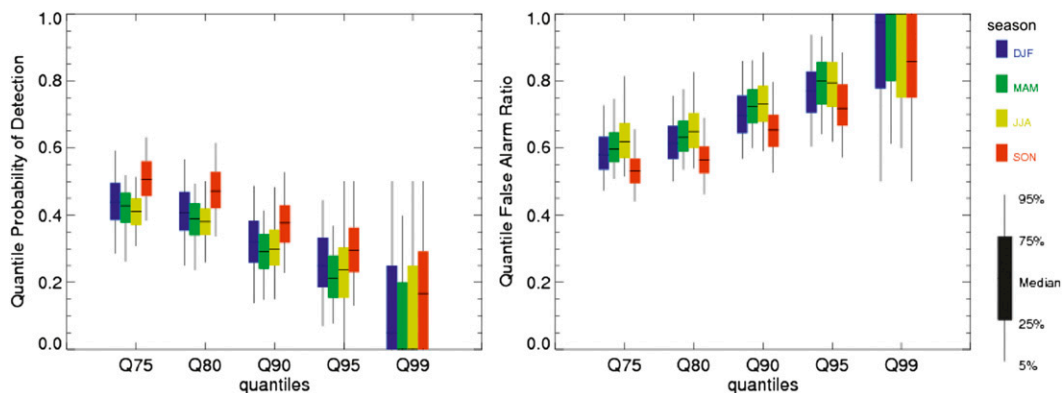


FIG. 7. Box-and-whisker plots of (left) QPOD and (right) QFAR calculated at each grid box with respect to different thresholds (75th, 80th, 90th, 95th, and 99th percentile) and different seasons (DJF: winter, MAM: spring, JJA: summer, and SON: autumn).

temporal and spatial sampling, differences in the generation process of the products, such as the different definition of a day, are another reason why exact matching—as required by point-by-point verification—results in poor scores. Because of the difference in the day definition, heavy precipitation events occurring between 0000 and 0900 UTC are assigned to different days by GPCP1DD and E-OBS. The effect is most pronounced for single satellite LEO-based estimates (TOVS-AIRS-based product from 1999 onward) where always only one of the two overpasses per day falls into the time overlap of the E-OBS and the GPCP1DD day. The resulting poor scores, however, often contradict the impression gained by visual inspection of daily precipitation maps. Therefore, we additionally used fuzzy methods in our study.

2) FUZZY APPROACH

Fuzzy verification methods avoid the so-called “double penalty” (Ebert 2008), describing the fact that non-detection at a given day and grid box is punished as well as its detection at an earlier/later or close-by grid box—a natural consequence of the exact match imposed by traditional verification methods. A variety of fuzzy methods are available that try to avoid this phenomenon (Ebert 2008). We chose the fraction skill score (FSS), which builds upon the estimates of the original 1-day and 1° resolution of the GPCP1DD dataset. Figure 8 depicts the FSS using the 90th percentile threshold to define extremes events as a function of spatial and temporal neighborhood sizes for the summer season. The [1, 1] space–time scale pair (i.e., the first number within brackets indicates the spatial neighborhood sizes in degrees, the second one the temporal neighborhood size in days) in the upper left corner corresponds to the traditional point-by-point verification that results in extremely low FSS values. All grid boxes lie well below the individual threshold of usefulness (FSS_{useful}), which at this scale generally ranges between 0.51 and 0.52 (figure not shown here). With temporally and spatially increasing neighborhood size, the effect of mismatches due to sampling and difference in the day definition is mitigated, and the FSS steadily increases and reaches values well above the local FSS_{useful} values, assigning GPCP1DD a useful skill at these scales. The FSS increases from 0.3 (area average for the [1,1] scale pair) to 0.8 (area average for the [7, 7] scale pair). From the [3, 5], [5, 3], and [7, 1] space–time scale pairs onward, the GPCP1DD shows useful skill over almost the entire spatial domain (more than 95% of the grid boxes, as indicated by the numbers shown in the upper left corner of the individual scale pair maps in Fig. 8) except for the southern part of the Iberian Peninsula. This area continues showing lowest

skill with further increasing scales. A behavior that can most likely be attributed to the very few wet days (~ 10 per season) and extreme events (one per season on average when using Q90 as threshold) occurring in the dry conditions dominating in summer here. The [3, 7] scale provides the maximum spatial coverage of grid boxes with useful skill. Concerning the other seasons, useful skill over more than 95% of the spatial domain was found as well from [3, 5], [5, 3], and [7, 1] space–time scales onward for autumn, for spring from [5, 5] and [7, 3], and for winter from [3, 7] and [5, 3] onward (figures not shown here). Figure 9 exemplarily shows a box-and-whisker plot of FSS values based on the [3, 7] space–time neighborhood for increasing extreme thresholds and different seasons. Highest scores are reached in summer and autumn for all extreme thresholds, whereas FSS values are consistently lowest for the spring season. For the 90th percentile threshold, all seasons except spring show useful skill over the entire area with area median FSS values higher than 0.65. The FSS decreases with increasing extreme threshold, such that for the 95th percentile threshold useful skill is still assigned to almost the entire area for summer and autumn but only to 75% and 50% of the area in winter and spring, respectively. For the 99th percentile threshold useful skill is lost over almost the entire area for all seasons. Seasonal area median FSS values, then, all lie below 0.45.

To demonstrate the assigned skill of the GPCP1DD dataset time series of extreme precipitation frequency (defined with respect to 90th percentile threshold) are exemplarily shown in Fig. 10 for the [3, 7] space–time scale for two locations. The first time series (upper panel) presents the extreme frequency for all summer seasons at the grid box centered at 43.5°N, 25.5°W (Romania). The second one (lower panel) shows the time series of extreme frequencies in summer at the grid box centered at 48.5°N, 14.5°E (Austria). Both time series show very good agreement with E-OBS with correlation coefficients greater than 0.7. The GPCP1DD time series capture very well the interannual variability of the extreme frequency depicted by E-OBS, in particular the major peaks in 2005 and 2007 (upper panel) and in 2002, 2005, and 2006 (lower panel). Especially the peaks in 2002 and 2005 correspond to extreme rainfall events that went along with severe large-scale flooding in the affected regions.

5. Conclusions

Within this study we compared daily precipitation estimates from E-OBS and GPCP1DD over Europe. We addressed 1) the comparability of basic climatological statistics (MEAN, INT, and NWET), 2) the similarity of precipitation distributions derived from satellite

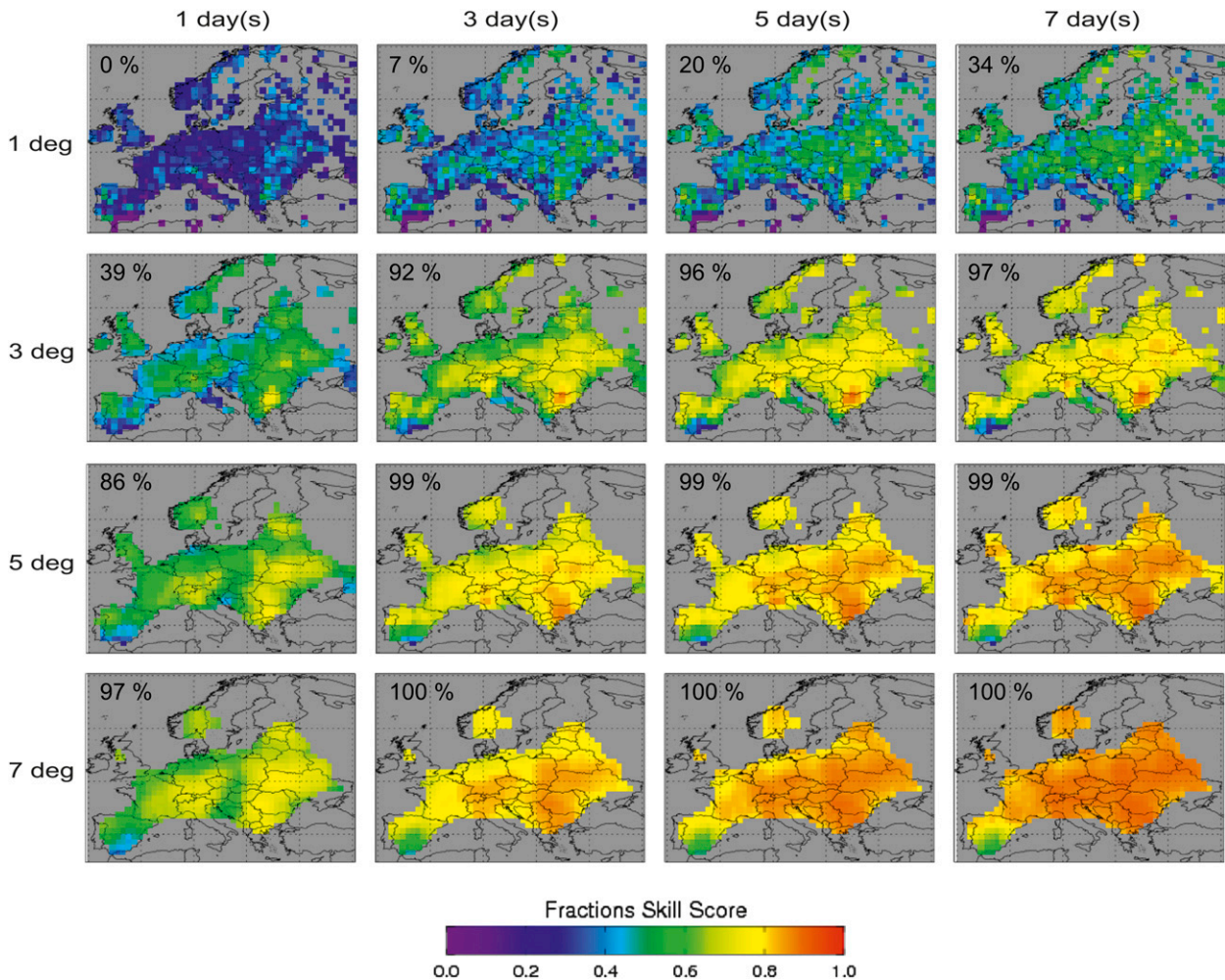


FIG. 8. FSS based on the 90th percentile threshold for the summer season (JJA) as a function of increasing temporal (1, 3, 5, and 7 days) and spatial (1°, 3°, 5°, and 7°) size of the neighborhood. The numbers in the upper left corner of the individual maps indicate the relative number of grid boxes (%) with FSS values exceeding the local FSS_{useful} .

data and rain gauge observations, and 3) the ability of GPCP1DD to replicate daily scale precipitation extremes and to effectively capture their variability.

We found good agreement between the E-OBS and the GPCP1DD dataset not only for the climatological statistics analyzed (MEAN, INT, NWET, and Q90) but also with respect to their distributions. The results revealed pronounced seasonal and regional variations in the performance of GPCP1DD. Larger differences were found in winter and dry regions/seasons where MEAN, INT, and Q90 (winter) and MEAN and NWET (dry regions) are generally overestimated by more than 50%. This may, to a large extent, be explained by the problems of the satellite retrievals used within GPCP1DD to correctly detect precipitation areas over cold, snow, and ice surfaces and in dry regions (due to variations in surface emissivity; e.g., Eyre and Menzel 1989; Ferraro

et al. 1998) as well as their inability to detect light or small areas of precipitation and short-lived events. In winter the larger differences may also be attributed to the wind-loss correction applied to the GPCP1DD dataset (but not to the E-OBS dataset), resulting in a higher MEAN value compared to E-OBS. In summer (outside the dry region in southern Europe) GPCP1DD and E-OBS show good agreement for all parameters with respective relative differences well below 10%. These results clearly demonstrate that satellite retrievals should in any case be analyzed both separately by season and climate region.

Both traditional deterministic and fuzzy verification methods were used to assess the skill of the GPCP1DD dataset to detect extreme precipitation events. Deterministic verification methods showed that with increasing threshold seasonal area median QPOD values drop from

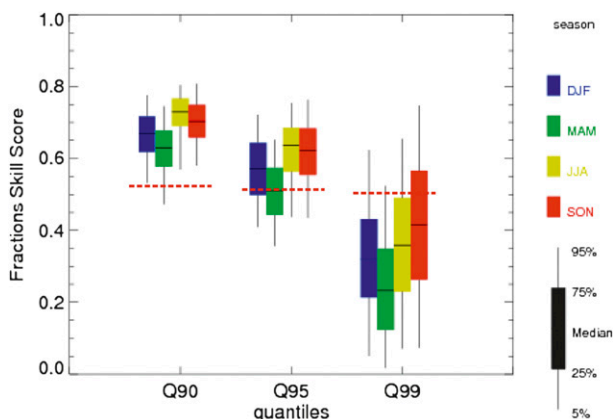


FIG. 9. Box-and-whisker plots of FSS for the 3°, 7-day neighborhoods calculated at each grid box with respect to different thresholds (90th, 95th, and 99th percentile) and different seasons (DJF: winter, MAM: spring, JJA: summer, and SON: autumn) over the whole region. The red dotted lines indicate the mean FSS_{useful} values for the three quantile thresholds: 0.52 (Q90), 0.51 (Q95), and 0.5 (Q99).

0.4–0.5 for the 75th to 0.2–0.3 for the 95th percentile, while at the same time seasonal area median QFAR values increase from 0.5–0.6 to 0.7–0.8, thereby assigning GPCP1DD only poor skill for the detection of extreme events. These results reinforce findings by AghaKouchak et al. (2011).

The fuzzy method approach proved to be better suited for the dataset comparison of such a highly variable parameter as precipitation because it compensates for any slight displacement by taking neighborhoods (defined by space–time scale pairs [degree, days]) around the individual grid boxes into account. From the fuzzy methods available (see Ebert 2008), the fraction skill

score (FSS) was used and increasing skill was found at larger spatial and temporal scales with area mean FSS values increasing for the summer season from 0.3 for the [1, 1] space–time scale to 0.8 for the [7, 7] space–time scale. GPCP1DD was found to have useful (better than random) skill in detecting extremes over the entire spatial domain (at least 95% of the grid boxes) from [3, 5], [5, 3], and [7, 1] space–time scale pairs onward in summer and autumn, in spring from [5, 5] and [7, 3], and in winter from [3, 7] and [5, 3] onward. On these scales (and higher) the GPCP1DD is able to represent variability of the frequency of extreme events as depicted by E-OBS. This was exemplarily demonstrated with extreme frequency time series of E-OBS and GPCP1DD at two locations.

The question of which scale to choose from those identified as skillful will finally depend on individual application-specific requirements with respect to temporal or spatial resolution, quality (FSS value), and spatial coverage (which decreases along the area boundaries with increasing spatial neighborhood size). GPCP1DD has limitations with respect to the type of extremes that are represented, especially over mid and high latitudes where precipitation estimates mostly rely on single-sensor LEO-based satellite observations that provide only low temporal sampling. Therefore, GPCP1DD does not capture small-scale events like thunderstorms, but precipitation extremes occurring in the context of mesoscale convective systems (MCS) in summer and in frontal systems in winter can be captured. For many applications those large events (occurring on large time and space scales) are more important than small-scale short-term events because the impact on people and infrastructure is higher. For studying extreme events at smaller scales as well as their changes in intensity and

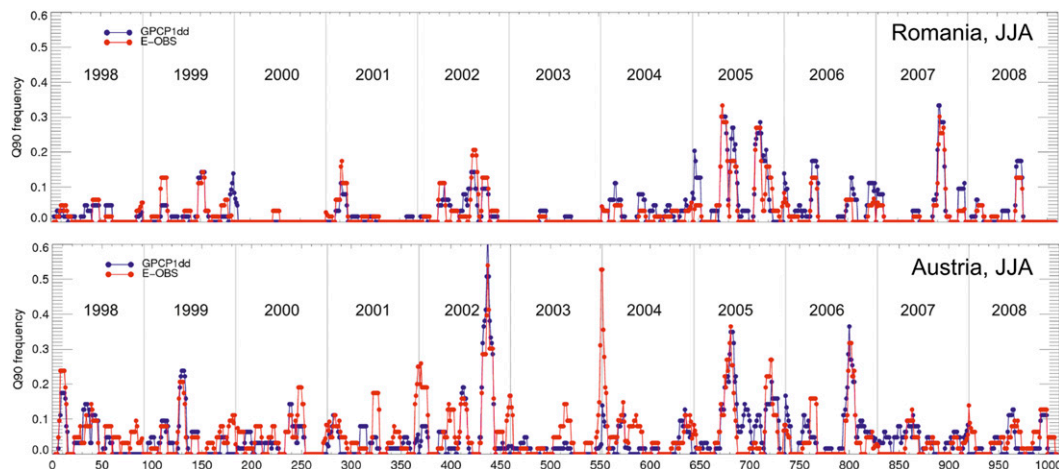


FIG. 10. Time series of extreme precipitation frequency (defined with respect to 90th percentile threshold) within a 3°, 7-day neighborhood for 11 summer seasons at two locations as depicted by E-OBS (red) and GPCP1DD (blue): grid box centered at (top) 43.5°N, 25.5°W (Romania) and (bottom) 48.5°N, 14.5°E (Austria).

frequency, there is a need for products with higher spatial and temporal resolution, both satellite and ground based. Recent efforts exploit regional networks of gauge observations to generate datasets at higher spatial resolutions for Germany (e.g., Brien et al. 2013), Spain, and the Netherlands. Moreover high-resolution precipitation datasets from satellites, such as the TRMM Multisatellite Precipitation Analysis (TMPA) (Huffman et al. 2007) 3B42RT (real time) product, CMORPH, and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) (Sorooshian et al. 2000) provide precipitation estimates at $0.25^\circ \times 0.25^\circ$ and 3-h resolution over latitudes between 60°S and 60°N . Besides, these satellite products include the more direct PMW observations in their retrieval scheme also for latitudes greater than 40°N and are therefore worth being considered for future studies of satellite-based extreme precipitation variability. Finally, spatially homogenized precipitation products from rain radar are being developed for European land areas [e.g., Operational Program for the Exchange of Weather Radar (OPERA); Huuskonen 2006]. Although these products will be made available at superior spatial (1 km) and temporal (1 h) resolution, the spatial homogenization of these products still presents a challenge (see Kidd et al. 2012; Roebeling et al. 2012).

Another limitation affecting the capability of both IR- and PMW-based satellite retrievals in representing extreme precipitation events in general is the saturation of the relationship between precipitation and the satellite measured quantities used to derive the precipitation estimates (IR and PWV brightness temperatures, respectively). In the case of PMW techniques, emission-based algorithms (used over the ocean) are affected by saturation effects, resulting in a maximum detectable rainfall rate that varies according to the depth of the rain layer (Adler et al. 1991). IR techniques rely on the assumption that colder cloud-top temperatures in the IR are always associated with higher rain rates. On the one hand, this causes heavy precipitation events from warm clouds to be missed; on the other, it also imposes a maximum retrievable rain rate by not being able to represent physical processes leading to a further increase in rainfall not related to an increase in cloud-top height. Whereas the GPCP1DD is well affected by saturation effects in the IR (TOVS-AIRS-based product), the PMW saturation effects do not translate into the final TMPA-based precipitation estimates. This is because GEO-IR is calibrated by matching the precipitation frequencies (of IR and PMW); that is, not the rain rates but the precipitation occurrences derived from SSM/I are used.

Further work shall concentrate on 1) applying the methodology introduced over other regions (Africa,

South America), 2) including reanalysis datasets as well as other multisource satellite products in the comparison, and 3) including dry extremes (e.g., Zolina et al. 2013) as well as considering other definitions of extreme wet events such as the duration, that is, number of consecutive days exceeding a certain threshold (e.g., Zolina et al. 2010), in order to take in addition to the isolated extreme rainfalls also the continuous periods of heavy and moderate precipitation into account.

Acknowledgments. We acknowledge the use of the E-OBS dataset provided by the EU-FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>), and the data providers in the ECA&D project (<http://eca.knmi.nl>). The GPCP1DD data were provided by the NASA/Goddard Space Flight Center (GSFC) Laboratory for Atmospheres, which develops and computes the GPCP1DD as a contribution to the GEWEX Global Precipitation Climatology Project. Software used within this study partly builds upon pieces of code provided by Elisabeth Ebert (<http://www.cawcr.gov.au/staff/eee/>). This study was partly supported by the projects 14.B25.31.0026 and 8338 funded by Russian Ministry of Education and Science and by the RFBR project 13-05-00930. Thanks go to DWD for funding a visit of M. Lockhoff at NASA/GSFC during which part of the work presented here was carried out and to George J. Huffman for enabling this stay as well as for valuable discussions and feedback. We also acknowledge three anonymous reviewers for their helpful comments and suggestions.

REFERENCES

- Adler, R. F., H.-Y. M. Yeh, N. Prasad, W.-K. Tao, and J. Simpson, 1991: Microwave simulations of a tropical rainfall system with a three-dimensional cloud model. *J. Appl. Meteor.*, **30**, 924–953.
- , C. Kidd, G. Petty, M. Morissey, and H. Goodman, 2001: Intercomparison of global precipitation products: The third Precipitation Intercomparison Project (PIP-3). *Bull. Amer. Meteor. Soc.*, **82**, 1377–1396.
- , and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167.
- AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai, 2011: Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *J. Geophys. Res.*, **116**, D02115, doi:10.1029/2010JD014741.
- Allan, R. P., and B. J. Soden, 2008: Atmospheric warming and the amplification of precipitation extremes. *Science*, **32**, 1481–1494.
- Arkin, P. A., and B. N. Meisner, 1987: The relationship between large-scale convective rainfall and cold cloud over the Western Hemisphere during 1982–84. *Mon. Wea. Rev.*, **115**, 51–74.
- Bauer, P., 2001: Over-ocean rainfall retrieval from multisensor data of the tropical rainfall measuring mission. Part I: Design and evaluation of inversion databases. *J. Atmos. Oceanic Technol.*, **18**, 1315–1330.
- Bolvin, D. T., R. F. Adler, G. J. Huffman, E. J. Nelkin, and J. P. Poutiainen, 2009: Comparison of GPCP monthly and daily

- precipitation estimates with high-latitude gauge observations. *J. Appl. Meteor. Climatol.*, **48**, 1843–1857.
- Brienen, S., A. Kapala, H. Mächel, and C. Simmer, 2013: Regional centennial precipitation variability over Germany from extended observation records. *Int. J. Climatol.*, **33**, 2167–2184.
- Chambon, P., I. Jobard, R. Roca, and N. Viltard, 2013: An investigation of the error budget of tropical rainfall accumulation derived from merged passive microwave and infrared satellite measurements. *Quart. J. Roy. Meteor. Soc.*, **139**, 879–893, doi:10.1002/qj.1907.
- Ebert, E. E., 2008: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. *Meteor. Appl.*, **15**, 51–64.
- , J. E. Janowiak, and C. Kidd, 2007: Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Amer. Meteor. Soc.*, **88**, 47–64.
- Eyre, J., and W. P. Menzel, 1989: Retrieval of cloud parameters from satellite sounder data: A simulation study. *J. Appl. Meteor.*, **28**, 267–275.
- Ferraro, R. R., 1997: Special sensor microwave imager derived global rainfall estimates for climatological applications. *J. Geophys. Res.*, **102** (D14), 16 715–16 735.
- , E. A. Smith, W. Berg, and G. J. Huffman, 1998: A screening methodology for passive microwave precipitation retrieval algorithms. *J. Atmos. Sci.*, **55**, 1583–1600.
- Groisman, P. Ya., R. W. Knight, D. R. Easterling, T. R. Karl, G. C. Hegerl, and V. N. Razuvaev, 2005: Trends in intense precipitation in the climate record. *J. Climate*, **18**, 1326–1350.
- Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res.*, **113**, D21019, doi:10.1029/2008JD010201.
- Hofstra, N., M. New, and C. McSweeney, 2010: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dyn.*, **35**, 841–858.
- Huffman, G. J., R. F. Adler, M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeorol.*, **2**, 36–50.
- , D. T. Bolvin, G. Gu, E. J. Nelkin, K. P. Bowman, E. F. Stocker, and D. B. Wolff, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scale. *J. Hydrometeorol.*, **8**, 38–55.
- , R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.*, **36**, L17808, doi:10.1029/2009GL040000.
- Huuskonen, A., 2006: EUMETNET OPERA: Operational Programme for the Exchange of Weather Radar information. *Proc. ERAD*, **3**, 371–373.
- Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.*, **5**, 487–503.
- Kidd, C., 2001: Satellite rainfall climatology: A review. *Int. J. Climatol.*, **21**, 1041–1066.
- , and V. Levizzani, 2011: Status of satellite precipitation retrievals. *Hydrol. Earth Syst. Sci.*, **15**, 1109–1116.
- , P. Bauer, J. Turk, G. J. Huffman, R. Joyce, K.-L. Hsu, and D. Braithwaite, 2012: Inter-comparison of high-resolution precipitation products over northwest Europe. *J. Hydrometeorol.*, **13**, 67–83.
- Klein Tank, A. M. G., and G. P. Können, 2003: Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99. *J. Climate*, **16**, 3665–3680.
- Lensky, I. M., and D. Rosenfeld, 2006: The time-space exchangeability of satellite retrieved relations between cloud top temperature and particle effective radius. *Atmos. Chem. Phys.*, **6**, 2887–2894.
- Levizzani, V., R. Amorati, and F. Meneguzzo, 2002: A review of satellite based rainfall estimation methods. European Commission Project MUSIC Rep. EVK1-CT-20000-00058, 70 pp.
- Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97.
- Roebeling, R. A., and I. Holleman, 2009: SEVIRI rainfall retrieval and validation using weather radar observations. *J. Geophys. Res.*, **114**, D21202, doi:10.1029/2009JD012102.
- , E. L. A. Wolters, J. F. Meirink, and H. Leijnse, 2012: Triple collocation of summer precipitation retrievals from SEVIRI over Europe with gridded rain gauge and weather radar data. *J. Hydrometeorol.*, **13**, 1552–1566.
- Sapiano, M., and P. Arkin, 2009: An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *J. Hydrometeorol.*, **10**, 149–166.
- Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046.
- Susskind, J., and J. Pfendner, 1989: Impact of interactive physical retrievals of NWP. *Report on the Joint ECMWF/EUMETSAT Workshop on the Use of Satellite Data in Operational Weather Prediction: 1989–1993*, T. Hollingsworth, Ed., Vol. 1, ECMWF, 245–270.
- , P. Piraino, L. Rokke, T. Iredell, and A. Mehta, 1997: Characteristics of the TOVS Pathfinder Path A dataset. *Bull. Amer. Meteor. Soc.*, **78**, 1449–1472.
- Trenberth, K. E., 1999: Conceptual framework for changes of extremes of the hydrological cycle with climate change. *Climatic Change*, **42**, 327–339.
- , A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, **84**, 1205–1217.
- Turk, F. J., P. Arkin, E. E. Ebert, and M. R. P. Sapiano, 2008: Evaluating high-resolution precipitation products. *Bull. Amer. Meteor. Soc.*, **89**, 1911–1916.
- Viltard, N., C. Burlaud, and C. Kummerow, 2006: Rain retrieval from TMI brightness temperature measurements using a TRMM PR-based database. *J. Appl. Meteor. Climatol.*, **45**, 455–466.
- Wilks, D., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 648 pp.
- Xie, P., J. E. Janowiak, P. A. Arkin, R. F. Adler, A. Gruber, R. Ferraro, G. J. Huffman, and S. Curtis, 2003: GPCP pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. *J. Climate*, **16**, 2197–2214.
- Zolina, O., C. Simmer, K. Belyaev, A. Kapala, and S. Gulev, 2009: Improving estimates of heavy and extreme precipitation using daily records from European rain gauges. *J. Hydrometeorol.*, **10**, 701–716.
- , —, S. Gulev, and S. Kollet, 2010: Changing structure of European precipitation: Longer wet periods leading to more abundant rainfalls. *Geophys. Res. Lett.*, **37**, L06704, doi:10.1029/2010GL042468.
- , —, K. Belyaev, S. K. Gulev, and P. Koltermann, 2013: Changes in the duration of European wet and dry spells during the last 60 years. *J. Climate*, **26**, 2022–2047.