Representation of Precipitation Characteristics and Extremes in Regional Reanalyses and Satellite- and Gauge-Based Estimates over Western and Central Europe

M. LOCKHOFF

Deutscher Wetterdienst, Offenbach, Germany

O. ZOLINA

Institut des Géosciences de l'Environnement, Grenoble, France, and P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia

C. SIMMER

Meteorological Institute, University of Bonn, Bonn, Germany

J. SCHULZ

European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), Darmstadt, Germany

(Manuscript received 13 September 2018, in final form 1 February 2019)

ABSTRACT

This paper evaluates several daily precipitation products over western and central Europe, identifies and documents their respective strengths and shortcomings, and relates these to uncertainties associated with each of the products. We analyze one gauge-based, three satellite-based, and two reanalysis-based products using high-density rain gauge observations as reference. First, we assess spatial patterns and frequency distributions using aggregated statistics. Then, we determine the skill of precipitation event detection from these products with a focus on extremes, using temporally and spatially matched pairs of precipitation estimates. The results show that the quality of the datasets largely depends on the region, season, and precipitation characteristic addressed. The satellite and the reanalysis precipitation products are found to have difficulties in accurately representing precipitation frequency with local overestimations of more than 40%, which occur mostly in dry regions (all products) as well as along coastlines and over cold/frozen surfaces (satellite-based products). The frequency distributions of wet-day intensities are generally well reproduced by all products. Concerning the frequency distributions of wet-spell durations, the satellite-based products are found to have clear deficiencies for maritime-influenced precipitation regimes. Moreover, the analysis of the detection of extreme precipitation events reveals that none of the non-station-based datasets shows skill at the shortest temporal and spatial scales (1 day, 0.25°), but at and above the 3-day and 1.25° scale the products start to exhibit skill over large parts of the domain. Added value compared to coarser-resolution global benchmark products is found both for reanalysis and satellite-based products.

1. Introduction

Global warming is expected to change precipitation characteristics, including frequency and intensity of extremes (Trenberth et al. 2003; Allan and Soden 2008; Seneviratne et al. 2012; Collins et al. 2013). Because these changes affect water resource management, agriculture, and infrastructure planning, a better monitoring of precipitation patterns, variability, and changes is required, including a better understanding of the physical mechanisms leading to precipitation extremes (Collins et al. 2018). Especially, the event-like nature of precipitation and its strong clustering in space and time requires high-quality precipitation products with high spatial–temporal resolution.

In situ precipitation data from ground-based gauges often extend back to the beginning of the twentieth century or even earlier (e.g., Brienen et al. 2013), but

Corresponding author: M. Lockhoff, maarit.lockhoff@dwd.de

DOI: 10.1175/JHM-D-18-0200.1

might be inhomogeneous due to station relocations and terminations or change of instrumentation. They are also coarsely distributed in space, and often suffer from missing data even over relatively densely sampled areas like Europe (Zolina et al. 2014). Consequently, continental-scale studies based on gauge observations (e.g., Klein Tank and Können 2003; Zolina et al. 2009) only provide reliable estimates of variability patterns in areas with dense observation networks.

Satellite-based products, available since the late 1970s, have therefore become a complementary and/or alternative data source with spatially homogeneous estimates and almost global coverage. Satellite observations are, however, only indirectly related to surface precipitation by the interaction of radiation with hydrometeors in the whole atmospheric column or with precipitation-generating clouds. Retrieval algorithms can be classified (Kidd and Levizzani 2011) according to wavelength band and its use into 1) scattered solar [visible (VIS) and near-infrared (NIR)] and emitted thermal infrared (IR) radiation, 2) emitted (and scattered) passive microwave (PMW) radiation and backscattered radar-emitted active microwave (AMW) radiation, and 3) multisensor methods. VIS/NIR and/or IR-based retrievals associate cloud albedo and cloud-top temperature statistically with precipitation rate. Accordingly, they are much less accurate for instantaneous estimates than PMW methods, which exploit radiation interacting directly with precipitation-sized hydrometeors. Because VIS and/or IR and/or NIR observations are typically collected from geostationary orbits (GEOs; e.g., Meteosat, GOES), related products benefit from a much higher temporal resolution (30–15 min) compared to the PMW-based products. The latter are only operated on low-Earth orbits [LEOs; e.g., Special Sensor Microwave Imager (SSM/I) and its follow-on instruments; Advanced Microwave Scanning Radiometer (AMSR-E)], which allow only for about three observations per day and satellite. As GEO-VIS/NIR/IR and LEO-PMW algorithms complement each other in terms of temporal resolution and accuracy, they are often combined into merged product like the NOAA Climate Prediction Center Morphing product (CMORPH; Joyce et al. 2004) and the Global Precipitation Climatology Project One-Degree Daily product (GPCP1DD; Huffman et al. 2001). For more details on satellite-based rainfall estimation methods, see, for example, Petty (1995), Adler et al. (2001), Kidd (2001), and Kidd and Levizzani (2011).

Data assimilation–based reanalyses also cover long periods (typically since the 1970s) and provide many climate system state variables including surface precipitation. Reanalyses combine observations and a numerical model using a ''frozen'' (temporally stable) data assimilation system, with the model serving as an intelligent interpolator between sparse observations. Hence, the output may encompass the complete set of (modeled, and thus physically consistent) climate state variables at model resolution and is available over both land and sea (like satellite-based products). Precipitation is mostly not assimilated [still an exception are recent regional high-resolution reanalyses (e.g., Wahl et al. 2017) and, e.g., the global NCEP-2 reanalysis (Kanamitsu et al. 2002)] and the output precipitation fields are therefore only constrained by other observations. The first reanalyses were produced in the early to mid-1990s with global coverage at rather coarse resolutions (e.g., NCEP-1 with ~2° spatial resolution). Subsequent releases have benefited from model improvements and increasing spatial resolution. Recently, a first segment of the ERA5 dataset [which will replace the ECMWF interim reanalysis (ERA-Interim)] became available on a 30-km grid. Regional reanalyses make use of limited-area models with even higher grid resolution, which better represents soil–atmosphere interactions, orographic effects, land-use effects, and land–ocean contrasts and better simulates smaller-scale processes (Bach et al. 2016; Dahlgren et al. 2016).

All products have specific uncertainties and limitations, which should be quantified, for example, via comparisons with high-quality reference data. The International Precipitation Working Group (IPWG) evaluated many satellite-based precipitation products (e.g., Ebert et al. 2007; Turk et al. 2008; Sapiano and Arkin 2009) and quantified their performance depending on season, region, and elevation/topography. The performance of satellite-based products is often found to be affected by both topography and season, but evaluation results also quite strongly reflect the type of observation chosen as reference. For example, Stampoulis and Anagnostou (2012) found PMW-based products to generally overestimate (underestimate) ground-based gauge observations in Europe over low-elevation area during the warm (cold) season. Using radar-based observations over northwestern Europe as a reference, Kidd et al. (2012) found all considered satellite products (both IR- and PMW-based) to underestimate precipitation in all seasons. The quality of both satellite- and model-based products usually deteriorates with increasing rainfall intensity (Ebert et al. 2007; AghaKouchak et al. 2011; Stampoulis and Anagnostou 2012; Lockhoff et al. 2014). However, Lockhoff et al. (2014) demonstrated that for extreme precipitation the detection skill of satellite retrievals considerably increases with increasing temporal and spatial scales. In general, NWP-based products outperform satellite-based

products during winter and in the midlatitudes (Ebert et al. 2007; Kidd et al. 2012).

The global reanalyses generally suffer from a too high occurrence of wet days and too low rainfall intensities, because precipitation generation cannot be resolved by the coarse model and is thus parameterized (Trenberth et al. 2003; Sun et al. 2006). Accordingly, reanalyses have in particular difficulties with small-scale convective precipitation but capture frontal rainfall well (Trenberth et al. 2003; Sun et al. 2006). Gehne et al. (2016) and Trenberth et al. (2017) stress the importance of subdaily time scales for precipitation extremes. Over Europe, global products generally overestimate low and underestimate high precipitation intensities (Zolina et al. 2004). Studies involving regional reanalyses concentrate on individual countries [Bollmeyer et al. (2015), for Germany] or small regions. For example, Isotta et al. (2015) found for Europe that the evaluated regional reanalyses overestimate mean precipitation and the frequency of wet days but underestimated the frequency of heavy precipitation (95th percentile).

Overall, a comprehensive and comparative analysis of continental-scale extreme precipitation from reanalyses, satellite data, and in situ observations is missing. This paper assesses the skill of high-resolution (0.25°, daily) regional reanalyses, recently updated versions of satellite-based precipitation products, and an available ground-based product in representing precipitation including extremes across central and western Europe using high-density gauge-based estimates as reference. We first assess the ability of the different datasets to represent spatial precipitation patterns and frequency distributions by using aggregated (multiyear) statistics. Besides mean, intensity and wet-day frequency we consider wet-spell duration. We also assess the scale-dependent skill of detecting in particular (wet) extremes, based on temporally and spatially matched pairs of precipitation estimates. The datasets to be evaluated and the reference dataset are described in section 2. Section 3 details the analysis methodology, and in section 4 the comparison results are presented. A discussion of the results is provided in section 5, and a summary is given in section 6 together with some concluding remarks.

2. Data

We evaluate six precipitation datasets, hereinafter referred to as candidate datasets: one based on rain gauges, three based on satellite observations, and two based on regional reanalyses for western and central Europe and the years 2007–09. During these years the reference data are available at a grid resolution of about 25 km. We selected daily (or better) precipitation datasets with spatial resolutions of 0.25° or better covering western and central Europe. When we assess the scale dependent skill in detecting precipitation events we also determine the added value of these high-resolution datasets against two global lower-resolved datasets (benchmarks datasets). Table 1 provides an overview of the main characteristics of the datasets (reference, candidate, and benchmark) with references.

a. *Reference dataset*

Our reference—the station dataset of the European Centre for Medium-Range Weather Forecasts (ECMWF; Fig. 1, top, subsequently referred to as ECMWF)—contains daily land precipitation accumulations for the years 2002–09 on a reduced Gaussian grid. Its highest resolution (N400) of ~25 km (0.225°) is available from October 2006 until the end of 2009. The observations cover mostly central, western, and northern Europe with roughly 10 000 stations (Ghelli and Lalaurette 2000), which are gridded by simple area averaging. In a preliminary analysis of the correlations between the daily precipitation time series for each grid box from ECMWF and the E-OBS dataset, we found much lower correlation values than expected ranging between 0.3 and 0.6. Therefore, we checked the effect of shifting the ECMWF data by one day forward or backward (with respect to E-OBS). We found that by shifting the ECMWF data forward in time by one day, we obtained much higher correlations (~0.8) for all countries (except Norway). This revealed the problem generally related to gauge data of whether the 24-h precipitation total recorded, for example, as 1 January 2007 corresponds to the precipitation that fell on that date, or if this was the date the observation was made (often at 0900 hours local time). We addressed this issue by shifting the data according to the shift yielding highest correlation with respect to E-OBS: this resulted in shifting the ECMWF data forward in time by one day for all grid boxes except for Norway. Here shifting the data forward by two days showed the highest correlations.

We did not consider other high-resolution gridded station datasets as, for example, Spatial and Temporal Scales and Mechanisms of Extreme Precipitation Events over Central Europe (STAMMEX; Zolina et al. 2014) due to their limed spatial coverage (e.g., STAMMEX is restricted to Germany).

b. *Ground-based estimates*

The E-OBS dataset (Haylock et al. 2008) provided by the Royal Netherlands Meteorological Institute (KNMI) differs from the reference dataset in terms of spatial and

TABLE 1. Overview of reference, evaluation, and benchmark datasets used in this study.

			Resolution		Coverage			
	Type	Dataset	Space	Time	Space	Time	Method	Producer (reference)
Reference	Ground based	ECMWF	~25 km (T799)	Daily	Western Europe	2006–09	Area averaging	ECMEF (Ghelli and Lalaurette 2000)
Candidate datasets	Ground based	E-OBS (version 10.0)	0.25°	Daily	Europe	1950–2013	Thin-plane splines for means, kriging for daily anomalies	KNMI (Haylock et al. 2008)
	Satellite based	PERSIANN CDR (version 1, rev.1)	0.25°	Daily	60°N/S	1982–2012	GRIDSAT-IRWIN, GPCP monthly precipitation	UC Irvine (Ashouri et al. 2015)
	Satellite based	TMPA3B42RT	0.25°	3 hourly	50°N/S (60°N/S)	2000–present	HQ (TMI, SSM/I, SSMIS, AMSR-E, AMSU, MHS), MW-VAR (IR)	NASA GSFC (Huffman et al. 2007)
	Satellite based	CMORPH (version 1/RAW)	0.25°	Daily	60°N/S	1998–present	TMI, AMSR-E, SSM/I, SSMIS, AMSU, MHS, IR vectors, daily gauge	NOAA/CPC (Joyce et al. 2004)
	Reanalysis	HERZ	6.2 km	6 hourly	Europe (CORDEX-EUR-11)	2007–12	Nudging	University of Bonn (Bollmeyer et al. 2015)
	Reanalysis	HIRLAM	0.20°	6 hourly	Europe	1989–2010	3D-Var data assimilation	SMHI (Dahlgren and Gustafsson 2012)
Benchmark datasets	Satellite based	GPCP1DD (version 1.2)	1.0°	Daily	Global	1996-present	SSM/I- and SSMIS-TMPI (IR), TOVS, AIRS, GPCP monthly	NASA/GSFC (Huffman et al. 2001)
	Reanalysis	ERA-Interim	0.75°	12 hourly	Global	1979–2013	4D-Var data assimilation	ECMWF (Dee et al. 2011)

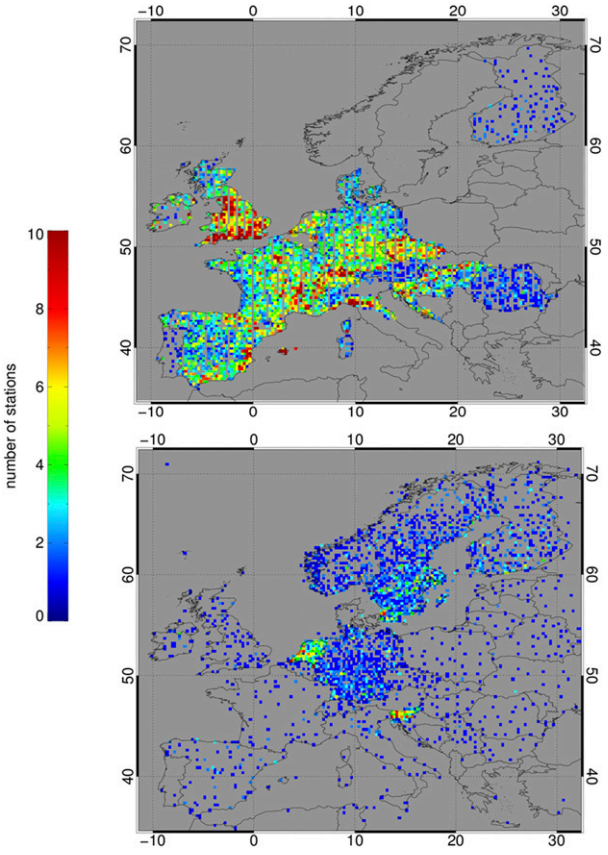

FIG. 1. Station density (number of stations within each 0.25° × 0.25°grid box) of gridded 24-h precipitation accumulations (top) produced by ECMWF and for (bottom) E-OBS for the years 2007–09.

temporal coverage, the number of stations used, and the gridding method. E-OBS covers a larger (than ECMWF) land area between 25° and 75°N and between 40°W and 75°E but only has around 7500 stations; thus, station density (see Fig. 1, bottom) is on average much lower compared to ECMWF (around 1600 stations within our study area compared to ~10 000 for ECMWF. E-OBS station densities are highest over Germany and the Netherlands and lower over France, Spain, and Portugal. We use version 10.0 with a 0.25° spatial resolution. E-OBS interpolates based on ordinary kriging of daily precipitation anomalies that are added to the monthly totals using three-dimensional thin-plate splines (Haylock et al. 2008).

Both ECMWF and E-OBS carry uncertainties due to gauge measurement errors, lacking wind corrections, and possible station relocations. E-OBS suffers from the temporally and spatially inhomogeneous sampling density, which affects area averages estimates in terms of representativeness and oversmoothing effects (Haylock et al. 2008; Hofstra et al. 2010). The latter occurs whenever stations outside of a grid box are used for the estimation of the grid average in case of sparse station densities (Hofstra et al. 2010). Zolina et al. (2014) demonstrated impacts of temporary varying data coverage in E-OBS on long-term tendencies in precipitation characteristics using data from a very dense German network of Deutscher Wetterdienst (DWD; German Weather Service).

c. *Satellite-based estimates*

We evaluate the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Climate Data Record (PERSIANN-CDR), the TRMM Multisatellite Precipitation Analysis (TMPA) 3B42RT (Real Time) dataset (TMPA3B42 RT), and the National Oceanic and Atmospheric Administration's (NOAA) Climate Prediction Centre morphing (CMORPH) dataset, which all have a spatial resolution of 0.25° × 0.25°. The datasets differ with respect to data sources (type of satellite sensors, use or not of rain gauge information) and blending approach.

The PERSIANN-CDR product (Ashouri et al. 2015)—subsequently referred to as PERSIANN—is the only satellite product used, which solely relies on IR input; its operational version additionally considers PMW input. The GPCP monthly 2.5° product, which includes multisatellite (IR and PMW) and land gauge information (version 2.2; Adler et al. 2003), is used for correcting the (monthly) biases. PERSIANN comes as daily accumulation and with a spatial resolution of 0.25° in the latitude band 60°N–60°S covering the years from 1983 to present.

The TMPA3B42RT product (subsequently referred to as TMPA) combines the merged microwave (MW; 3B40) and MW-calibrated IR (3B41) products, and is provided at 0.25° spatial and 3-h temporal resolution (Huffman et al. 2007). TMPA provides instantaneous precipitation estimates in millimeters per hour at 0000, 0300, 0600, 0900, 1200, 1500, 1800, and 2100 UTC and is processed without gauge correction between 60°N and 60°S; estimates north of 50°N and south of 50°S are flagged for lower quality. The calibrated product is restricted to 50°N and 50°S and was therefore not selected although it is considered more accurate (at least in terms of bias). We used version 7, for which the retrieval and analysis system was processed back to 1 March 2000.

CMORPH precipitation estimates are based on PMW-based observations, which are then propagated by motion vectors derived from geostationary satellite IR data (Joyce et al. 2004). We use the reprocessed product (CMORPH version 1) based on a fixed algorithm applied to the same inputs to ensure homogeneity. The original product is created on an 8-km grid at half-hourly time resolution; 3-hourly and daily products with 0.25° spatial resolution are also generated.

CMORPH is available between 60°N and 60°S starting in 1998. In this study, the 0.25°daily raw (satellite-only precipitation estimates, no bias correction applied) dataset was used. The raw version of CMORPH was chosen in order to ensure better comparability to TMPA (for which the noncalibrated RT version was chosen as well). The selection of satellite-only versions has the advantage of making these dataset more independent of the other candidate datasets.

d. Reanalysis-based estimates

We evaluate the Consortium for Small-Scale Modeling (COSMO)-based reanalysis (Bollmeyer et al. 2015) developed by the Climate Monitoring branch of DWD's Hans-Ertel Centre for Weather Research project (HErZ; Simmer et al. 2016)—subsequently referred to as HERZ—and the High Resolution Limited Area Model (HIRLAM) reanalysis (Dahlgren and Gustafsson 2012) provided by the Swedish Meteorological and Hydrological Institute (SMHI) within the framework of the European Reanalysis and Observations for Monitoring (EURO4M) project.

The HERZ reanalysis is based on the limited-area nonhydrostatic NWP model COSMO [formerly known as Lokal Modell (LM); Steppeler et al. 2003; Brdar et al. 2013] and applies the nudging data assimilation technique for conventional observations only without precipitation. The data assimilation also includes external analyses of snow, sea surface temperature, and a variational soil moisture analysis, which uses 7 vertical layers over a depth of approximately 14.5 m. Lateral boundary conditions are taken from ERA-Interim. The model output covers the CORDEX-EUR11 domain (Giorgi et al. 2009), extending from about 22°W to 45°E and from 27° to 72°N at a 6-km spatial resolution and a 3-hourly interval for three-dimensional and 1-hourly interval for two-dimensional output.

The HIRLAM reanalysis is produced using the hydrostatic HIRLAM forecast model and three-dimensional variational data assimilation (3D-Var) system (Dahlgren and Gustafsson 2012). Output is produced at 0.2° (~22 km) grid resolution for 6-hourly intervals over Europe. Again, only conventional in situ observations are assimilated and boundary conditions are taken from ERA-Interim.

Since precipitation observations are not assimilated in the models, large differences between their precipitation fields may occur. Further discrepancies may arise from the differences in the assimilated observation datasets, model resolution and parameterizations, assimilation method (variational or nudging), and the general model setting (hydrostatic versus nonhydrostatic). Blending models with observations by assimilating them is therefore a way to further constrain the models and improve the quality of reanalyses-based precipitation estimates.

e. Benchmark datasets

The higher-resolution regional datasets are compared to two commonly used lower resolution global benchmark datasets in the second part of our analysis, where we focus on the scale dependent skill in detecting precipitation events. These are the GPCP1DD dataset and ERA-Interim.

GPCP1DD (version 1.2; Huffman et al. 2001), available from October 1996 onward, is a global multisatellite product at 1° spatial resolution. Input varies with time and region: the Threshold-Matched Precipitation Index (TMPI) is used for the latitude belt from 40°N to 40°S and the scaled TOVS-AIRS (IR sounders) estimates are used for latitudes higher than 40°N/S. The daily estimates are scaled such that they sum up to the GPCP monthly product. The latter includes gauge information; thus, the daily product over land indirectly includes gauge information.

ERA-Interim is the global 4D-Var-based reanalysis provided by ECMWF (Dee et al. 2011) at 0.7° (~75 km) resolution on 60 vertical levels from the surface up to 0.1 hPa, which makes use of observations from a wide variety of satellite instruments in addition to meteorological observations. As for the regional reanalyses, precipitation observations are not assimilated, but total precipitation forms part of the accumulated forecast parameters, accumulated from the start of the forecast (twice daily starting at 0000 and 1200 UTC). ERA-Interim 12-hourly precipitation accumulations were retrieved from ECMWF's Meteorological Archival and Retrieval System (MARS) archive at 1° resolution.

f. Data processing

Precipitation estimates of the datasets were converted into daily accumulations from 0000 to 2400 UTC and remapped onto a common 0.25° grid, which is the native grid for E-OBS, PERSIANN, TMPA, and CMORPH. Spatial interpolation, which might influence, for example, the number of wet days of neighboring dry grid points, is avoided if possible. The higher-resolved HIRLAM and HERZ data were converted to the 0.25° grid by averaging over those grid points with centers within the coarser grid. For the slightly coarser-resolved ECMWF grid boxes we applied triangulation to convert their reduced Gaussian grid to the regular 0.25° grid. For the coarser-resolution benchmark datasets, ERA-Interim and GPCP1DD, we attributed the value of their 1° grid boxes to the four 0.25 grid boxes contained therein. For gauge-based products a day is typically defined as the period between 0900 and 0900 hours local

time, owing to the reading times of the gauges (Haylock et al. 2008); thus accumulation times for ECMWF and E-OBS are shifted by 9 h. For HERZ, HIRLAM, and ERA-Interim 6-hourly (0000, 0600, 1200, 1800 UTC) and 12-hourly (0000 and 1200 UTC) accumulations, respectively, were aggregated into daily estimates. For TMPA daily estimates were derived assuming that the eight instantaneous rain rates at 0000, 0300, 0600, 0900, 1200, 1500, 1800, and 2100 UTC are representative for 3-h periods centered at the respective synoptic time. Thus, the rain rates at 0300 UTC were attributed to 0200, 0300, and 0400 UTC. The rain rates for 2300 UTC of each day consequently stem from the 0000 UTC rain rates of the next day (which are additionally attributed to 0000 and 0100 UTC).

We restricted the comparison to the region covered by the ECMWF data and south of 60°N, which is the northmost coverage of PERSIANN, TMPA, and CMORPH. Only those grid points were considered, which have at least one station and a maximum of 20% missing days (data gaps) in the ECMWF data record. Note that Zolina et al. (2005) reported that even 30%–40% gaps in daily station records are acceptable for climatological assessments. For some countries there were only precipitation values available but no information on the number of station (e.g., Norway). In these cases, the number of stations per grid box was set to 1, in order to still be able to include the data in the comparison. Finally, the comparison period is restricted to the period from January 2007 to December 2009 for which complete years of the reference data with highest spatial resolution are available.

3. Methodology for comparative assessment

First, we compared spatial precipitation patterns and frequency distributions, and then we analyzed the scale-dependent skills of the datasets to detect and quantify the individual precipitation events. We used a threshold of 1 mm day^{-1} to distinguish between wet and dry days, which excludes very light precipitation from our analysis—a very uncertain parameter in both ground-based datasets (Klein Tank and Können 2003; Zolina et al. 2010) and reanalyses (Haiden et al. 2012). Precipitation statistics are also provided separately for the four seasons defined as follows: winter [December–February (DJF)], spring [March–May (MAM)], summer [June–August (JJA)], and autumn [September–November (SON)].

a. *Spatial precipitation patterns and frequency distributions*

For the assessment of spatial patterns and frequency distributions we used:

- MEAN (mm day^{-1}): average over all days;
- INT (mm day^{-1}): average over all wet days;
- NWET (days): number of wet days;
- Q10, Q90, etc. (mm day^{-1}): 10th, 90th, etc. percentiles estimated from the empirical wet-day distributions;
- WSs and DSs (days): wet-spell and dry-spell length quantified as number of consecutive wet days and dry days, respectively;
- Bias ratio (unitless): average of each candidate dataset divided by the average of the reference (ECMWF);
- PDF (unitless): empirical frequency distributions of daily precipitation; and
- CDF (unitless): cumulative frequency distributions of daily precipitation.

MEAN, INT, and NWET are standard precipitation measures. The empirically estimated percentiles assess the consistency of the precipitation distributions between the candidate and reference data. WS/DS durations (consecutive wet/dry days) are closely related to water availability within a region (Zolina et al. 2010, 2013). The bias ratio quantifies the similarity of the above characteristics between different dataset, and the nonparametric Kolmogorov–Smirnov (KS) test (see Wilks 2006) is used to quantify the agreement between the (cumulative) frequency distributions. Results are computed for each grid box in order to avoid data pooling across climate zones.

WS/DS durations were averaged over areas following the climate zones of the 1° map by Kottek et al. (2006), who generated an updated Köppen–Geiger world map of climate classification based on datasets from the Climatic Research Unit (CRU) of the University of East Anglia and the Global Precipitation Climatology Centre (GPCC) at the German Weather Service; all 0.25° boxes within a 1° grid box were assigned the same 1° value. The resulting climate map distinguishes three main climate types for Europe: the dominant warm temperate type (C), the cold/snow (D), and polar climates (E) with the oceanic subtypes the most dominant. We then selected four countries, which belong to only one of the climate subtypes. By selecting countries—and not the complete climate subtypes—we avoid potential internal spatial inhomogeneities caused by, for example, different national observation strategies:

- Great Britain (BRT): oceanic humid subtype of the temperate climate zone (Cfb);
- Germany (GER): same as BRT, but including a transition zone to continental climate (Cfb);
- Romania (ROM): warm summer humid continental climate (Dfb); and
- Iberian Peninsula (ESP): dry summer subtypes of the temperate climate zone (Csa/Csb).

b. Extreme precipitation events

Lockhoff et al. (2014) showed that gridbox-by-gridbox comparisons are not suitable for evaluating the skill to detect precipitation extremes, as exact matching in terms of location and timing of the event is required. Fuzzy verification methods allow slight (temporal and/or spatial) displacements and thereby avoid the double penalty effect (Ebert 2008). This is done by defining (temporal and/or spatial) neighborhoods around the individual grid boxes and/or single time steps. The number of grid boxes and time steps determine the size of the resulting 2D/3D neighborhood and hence the space–time scale: for example, a neighborhood covering 3 grid boxes (in each direction) and 3 days results in a total of $3 \times 3 \times 3 = 27$ values and hence a 0.75° space and 3-day time scale (when starting from 0.25° spatial and 1-day temporal resolution). Following Lockhoff et al. (2014), we used the fractions skill score (FSS; Roberts and Lean 2008) to assess the skill of the candidate datasets to represent the occurrences of extreme precipitation compared to the reference. An event is defined by the use of a threshold. The FSS score then directly compares the fractional coverage of events in the defined space–time neighborhood. The more similar the frequency of the event, as seen by the candidate dataset, is to the frequency obtained from the reference the larger the FSS score. We analyzed different percentile thresholds of daily precipitation (e.g., 90th, 95th, etc.) estimated from the sample of all wet days for different seasons in the analyzed time period (2007–09). For a continental-scale analysis this approach is more appropriate as compared to using fixed thresholds (see Groisman et al. 2005). Different maximum displacements between the location (0.25°, 0.50°, 0.75°, etc.) and timing (1, 2, 3, etc. days) of precipitation extremes were analyzed. For each selected space–time scale (maximum displacement) and threshold (e.g., Q90) the daily precipitation accumulations from a candidate and the reference dataset were converted to fractions (frequencies of extreme events). For example, 4 grid boxes exceeding the selected threshold within a 3-day and 3×3 (0.25°) gridbox neighborhood results in a fraction of $4/(3 \times 3 \times 3) \sim 0.15$. These fractions were used to calculate the FSS, which ranges between 0 (complete mismatch) to 1 (perfect match). The $\mathrm{FSS_{useful}}$ represents the FSS value above which the assessed candidate dataset is considered to have useful skill. The magnitude of $\mathrm{FSS_{useful}}$ is defined as being halfway between a random skill, represented by the average fraction of extreme events observed by the reference dataset at a specific grid point over the entire time period, and perfect skill, that means FSS equal to 1 (Roberts and Lean 2008). The $\mathrm{FSS_{useful}}$ definition thereby accounts for a change in skill that might be obtained as a result of a change in the event frequency (higher frequencies tend to give greater skill). For details, see Lockhoff et al. (2014, and references therein).

4. Results

a. Precipitation patterns and frequency distributions

The candidate datasets capture the spatial patterns of MEAN, INT, and NWET during the summer months generally well compared to the reference, however, specific differences are also observed (Fig. 2). For example, HIRLAM tends to overestimate MEAN, INT, and NWET, whereas TMPA and CMORPH tend to underestimate MEAN and NWET in certain areas, for instance, over the United Kingdom. Differences in MEAN might result from problems in detecting precipitation (see NWET) and/or from estimating the average precipitation for wet days (see INT), reflected by the bias ratios (BR) for NWET and INT, respectively (Fig. 3). Sign and magnitude of the discrepancies differ largely among the candidate datasets and between different regions. For most datasets, the MEAN-BR maps are similar to the NWET-BR maps; thus, precipitation detection is the main cause for the identified differences in mean precipitation (see also Table 2). Only for E-OBS the BR for INT (domain average 0.91) differs somewhat more from 1 than BR for NWET (domain average 1.06). E-OBS performs best over Germany, where the station density is largest but slightly underestimates MEAN over France and slightly overestimates MEAN over the United Kingdom. HIRLAM and PERSIANN generally overestimate the occurrence of precipitation across the study area (domain average BR of NWET of 1.40 and 1.54, respectively) with highest values over Spain, the Mediterranean coast, and Romania. TMPA and CMORPH underestimate the occurrence of precipitation in northwestern Europe, especially along the coastlines and overestimate its occurrence over southeastern Europe. Underestimation dominates, however (domain average BR of NWET of 0.86 for TMPA and 0.96 for CMORPH). Interestingly, the reanalyses HERZ and HIRLAM differ concerning the detection of precipitation: while HIRLAM overestimates its occurrence (domain average BR of NWET-BR of 1.40), HERZ only slightly underestimates its occurrence (domain average BR of NWET of 0.92). HERZ demonstrates the largest underestimation of precipitation occurrence over the Iberian Peninsula (up to 20%), especially along the southern coast. This region with typically rare precipitation is also characterized by biases in the other datasets with over- or underestimations

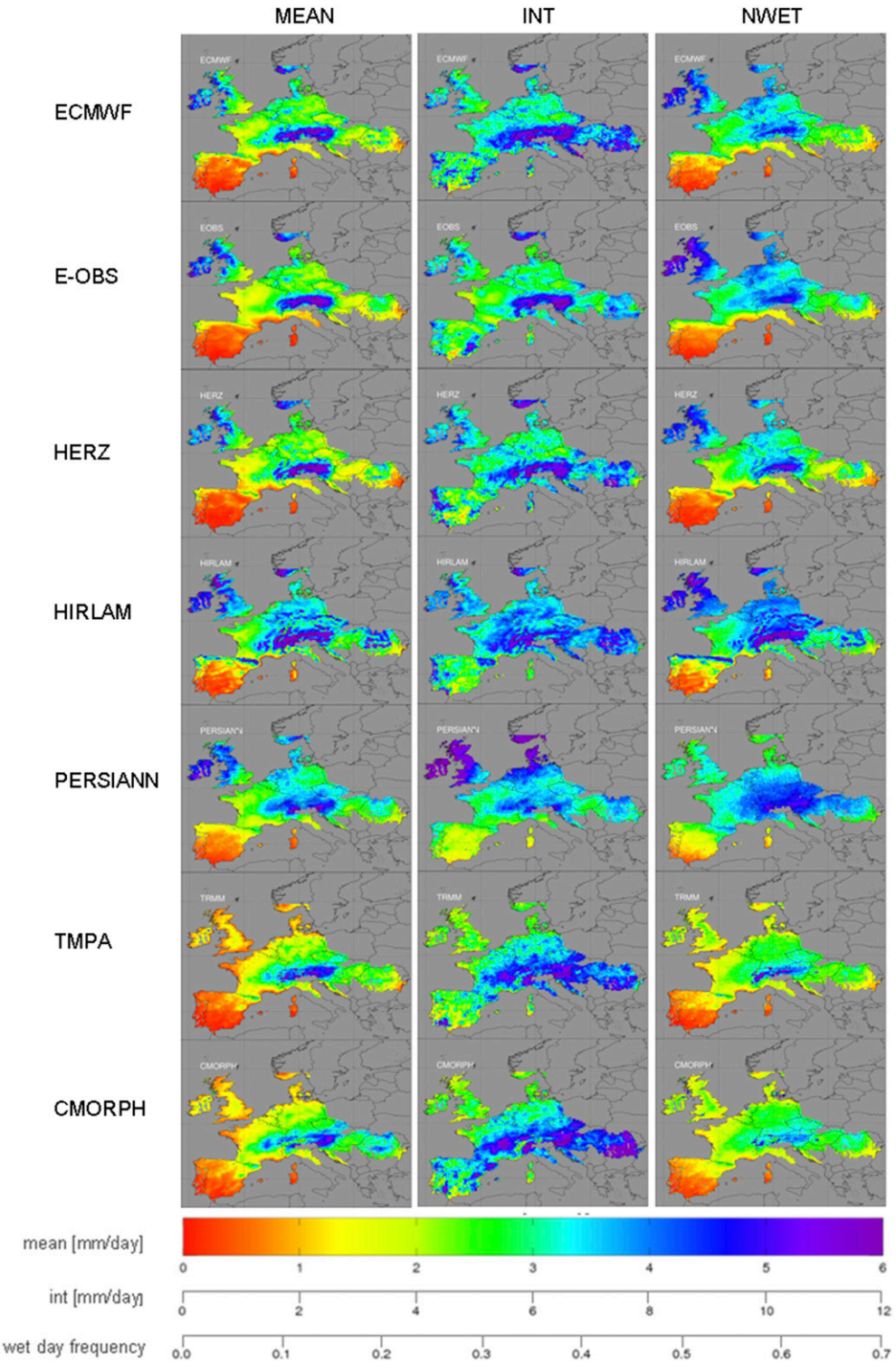

FIG. 2. Multiyear (left) MEAN, (middle) INT, and (right) NWET (here shown as wet-day frequency, thus relative to the total number of wet days) as represented by the reference (ECMWF) and the candidate datasets for the summer seasons (JJA) of the years 2007–09.

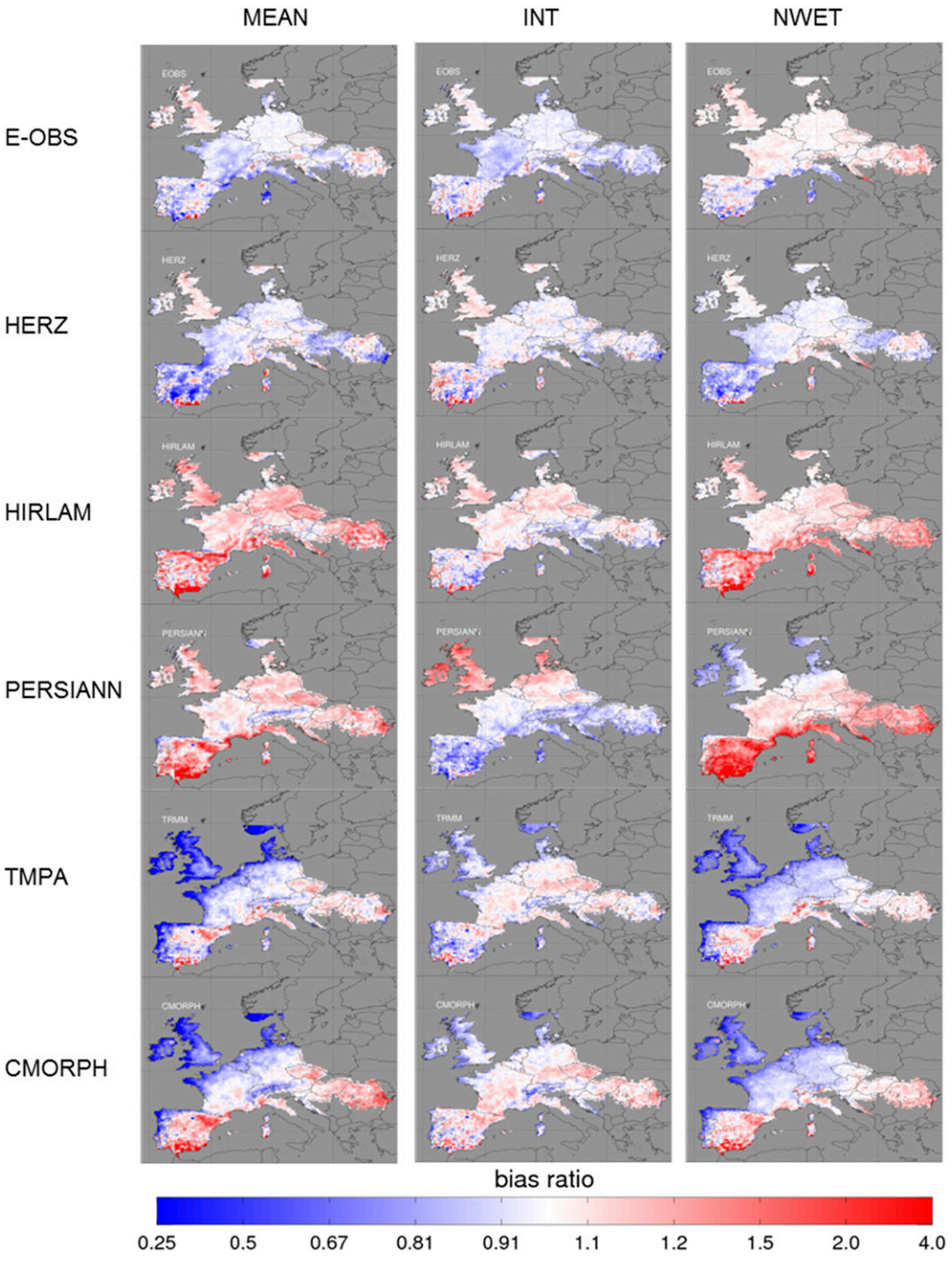

FIG. 3. Bias ratio for (left) MEAN, (middle) INT, and (right) NWET for the candidate datasets for the summer seasons (JJA) of the years 2007–09.

of NWET of up to 50% and larger (especially PERSIANN and HIRLAM). Domain average statistics for daily rainfall statistics for all candidate datasets are summarized in Table 2.

In winter E-OBS and the two regional reanalyses generally perform similar or better than in summer, while for the satellite-based datasets opposite tendencies can be observed (see Table 2). For the at least partially PMW-based products (CMORPH and TMPA) the performance decreases in winter compared to summer with the largest underestimation of precipitation found for CMORPH (up to 80% for MEAN). For the

TABLE 2. Statistics of daily rainfall accumulation (mm day^{-1}) averaged over the entire domain and time period (2007–09) separately per season (summer, winter, and annual values); INT_BIAS = mean bias of wet-day precipitation (mm day^{-1}), i.e., on days with rain rate ≥ 1 mm, WDAYS_BIAS = mean bias in the number of wet days (i.e., day with rain rate ≥ 1 mm day^{-1}), MEAN_BIAS = mean bias of the mean precipitation amount, P10_BIAS = mean bias of the 10th percentile (mm day^{-1}), P90_BIAS = mean bias of the 90th percentile.

	INT_BIAS	WDAYS_BIAS	MEAN_BIAS	P10_BIAS	P90_BIAS
			Summer (JJA)		
E-OBS	0.91 ± 0.25	1.06 ± 0.21	0.93 ± 0.24	1.09 ± 0.33	0.88 ± 0.037
HIRLAM	1.07 ± 0.27	1.40 ± 0.84	1.49 ± 1.27	1.02 ± 0.20	1.08 ± 0.55
HERZ	1.00 ± 0.27	0.92 ± 0.24	0.92 ± 0.39	1.00 ± 0.26	1.05 ± 0.48
TMPA	0.97 ± 0.30	0.86 ± 0.65	0.84 ± 0.53	1.01 ± 0.44	1.02 ± 0.46
CMORPH	1.04 ± 0.30	0.96 ± 0.96	1.03 ± 0.87	0.95 ± 0.19	1.13 ± 0.55
PERSIANN	1.02 ± 0.40	1.54 ± 1.04	1.37 ± 0.85	1.29 ± 0.73	0.98 ± 0.43
			Winter (DJF)		
E-OBS	0.91 ± 0.22	1.07 ± 0.16	0.96 ± 0.29	1.01 ± 0.20	0.89 ± 0.25
HIRLAM	0.97 ± 0.19	1.27 ± 0.28	1.2 ± 0.35	0.99 ± 0.16	0.95 ± 0.24
HERZ	0.99 ± 0.16	1.06 ± 0.20	1.1 ± 0.24	1.00 ± 0.16	0.99 ± 0.21
TMPA	1.0 ± 0.33	0.54 ± 0.30	0.57 ± 0.41	1.04 ± 0.23	1.15 ± 0.48
CMORPH	0.51 ± 0.16	0.29 ± 0.15	0.17 ± 0.091	0.87 ± 0.16	0.52 ± 0.24
PERSIANN	1.0 ± 0.28	1.48 ± 0.41	1.46 ± 0.50	0.97 ± 0.15	1.07 ± 0.37
			Annual values		
E-OBS	0.90 ± 0.15	1.07 ± 0.12	0.94 ± 0.19	1.02 ± 0.12	0.87 ± 0.18
HIRLAM	0.98 ± 0.15	1.27 ± 0.22	1.25 ± 0.28	1.00 ± 0.09	0.98 ± 0.18
HERZ	0.97 ± 0.12	0.98 ± 0.12	0.96 ± 0.17	0.98 ± 0.08	0.98 ± 0.15
TMPA	1.04 ± 0.20	0.70 ± 0.30	0.73 ± 0.35	0.98 ± 0.11	1.09 ± 0.25
CMORPH	0.87 ± 0.21	0.59 ± 0.20	0.54 ± 0.22	0.90 ± 0.10	0.90 ± 0.27
PERSIANN	0.96 ± 0.25	1.40 ± 0.36	1.28 ± 0.30	1.02 ± 0.16	0.97 ± 0.29

IR-based PERSIANN the performance even slightly improves during winter. This is surprising as IR-based retrievals are expected to perform better in summer when most precipitation comes from deep convective clouds that can be easily identified in IR channels (Petty 1995), whereas in winter precipitation comes more often from large-scale frontal systems that are more difficult to identify from IR channels.

We turn now to the analysis of probability distributions of daily precipitation. Figure 4 presents the probability distributions of precipitation intensity during summer for low (up to 25th percentile) and high (larger than 25th percentile) intensities. For low precipitation intensities the datasets are within ±10% of the reference. Both E-OBS and HIRLAM overestimate the intensities while the satellite-based datasets and HERZ have lower intensities. For high and extremely high precipitation intensities (>95th percentile) discrepancies up to 25% are found with E-OBS (25%) and CMORPH (12%) showing largest under- and overestimation, respectively. TMPA and both reanalyses closely agree with the reference. According to the gridbox-based KS test, all products have distributions consistent with the reference at the 5% significance level everywhere (not shown here), except for CMORH during wintertime, when the null hypothesis of the KS test (distribution of the candidate dataset and the reference disagree) is accepted over large parts of the study area. For the other datasets this is only the case for a few isolated grid boxes, for example, for E-OBS over France, HIRLAM over the southern United Kingdom, and PERSIANN over the United Kingdom and Germany.

Following Zolina et al. (2013), we analyzed the probability distributions of WS durations. Figure 5 shows distributions of WS durations normalized to the total number of wet days for the four selected climate regions (BRT, GER, ROM, and ESP; see section 3) for summer and winter. The reference dataset reveals region-characteristic shapes. The distributions for BRT are rather flat without any pronounced peak in both seasons. For GER the distributions are somewhat steeper, especially in winter. The distributions for ESP are steepest and left skewed with pronounced peaks at 1–2-day spell durations (each around 30%) but slightly flatter in winter than in summer. The distributions for ROM are similar to ESP also with a pronounced peak at the 1–2-day spell duration both in summer and in winter. The ability to capture these characteristic shapes varies across the datasets analyzed. Thus, the station-based datasets (E-OBS) and the two regional reanalyses capture the shapes well for both the convective/continental and ocean-influenced regions. The satellite-based datasets reproduce the distribution shapes for convective regimes (ESP, ROM) relatively well, but

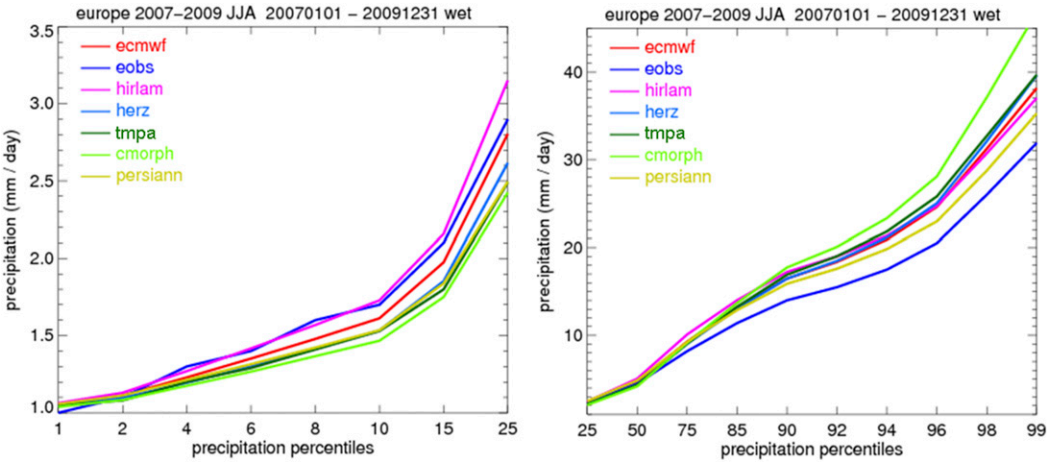

FIG. 4. Frequency distribution of wet-day precipitation intensity for (left) low percentiles and (right) high percentiles for the summer seasons (JJA) of the years 2007–09.

tend to overestimate the frequency of short WS durations (1 and 2 days) and underestimate those for longer durations (3 and 4 days). All satellite-based datasets fail, however, to accurately replicate the WS duration distributions in the humid regime (BRT). During both seasons they largely overestimate the frequency of short WS and underestimate those for long durations. In winter the differences can be larger than in summer except for PERSIANN, whose distribution is close to those of the reanalysis datasets. This is also true for the GER region, where E-OBS, the reanalysis datasets, and PERSIANN are close to the reference.

The differences between the mean WS durations (not shown here) are similar to those for NWET: if NWET is overestimated also the mean WS duration is overestimated. Bias ratios for WS durations are generally lower for all candidate datasets by up to 20%. In regions where WS durations are shorter, the DS durations tend to be longer, for example, in summer over ESP (max ~6) and ROM (max ~5), and where WS durations tend to be longer the DS durations are shorter, for example, over the United Kingdom (max ~1). The DS durations from E-OBS and the reanalysis datasets agree best with the reference dataset, while the satellite PMW-based datasets tend to overestimate DS durations. Similar to the WS analysis, DS durations from the PMW-based datasets perform best in GER and ROM in summer, fair in ESP, and fail in BRT.

b. Detection of (extreme) precipitation events

To assess the skill of detecting individual extreme events, we analyzed collocation of individual precipitation extremes in space and time in different datasets. For this purpose, we used the FSS criterion (see section 3b) derived from the (temporally and spatially) matched daily estimates at 0.25° resolution. Figure 6 displays domain-averaged FSS score values for Q90 for summer as a function of spatial (y axis) and temporal (x axis) scales for the candidate datasets. For the 1-day temporal scale and 0.25° spatial scale, all datasets except E-OBS have very low mean FSS values ($\leq$0.3). Hence, none of the non-station-based datasets show skill in detecting extreme precipitation events at the highest temporal and spatial scale. The mean FFS for E-OBS is 0.52 with around 60% of the grid boxes showing useful skill (i.e., FSS $>$ $\text{FSS}_{\text{useful}}$). For all datasets FSS increases with increasing temporal and spatial scales. In general, the skills are more sensitive to the spatial scales rather than to temporal scales. Thus, the skills increase to a lesser extent with increasing the temporal scale than for increasing the spatial scale. At the maximum temporal scale (7 days) the domain average FSS values are between 20% and 35% below the skills for the maximum spatial scale (2.25°). Among the nonstation datasets HERZ performs best. In summer at the 3-day scale and 1.25° spatial scale, its domain average FSS is about 0.68, with about 90% of the grid boxes showing useful skill (i.e., FSS $>$ $\text{FSS}_{\text{useful}}$). For all other nonstation datasets, similar FSS values are only observed at longer temporal and larger spatial scales, for example, for HIRLAM at 3 days and 1.75° (domain average FSS ~0.66), and for CMORPH, TMPA, and PERSIANN at 7 days and 2.25° (domain average FSS ~0.67, ~0.66, and ~0.63, respectively). All datasets exhibit a seasonal cycle in the skill of detecting precipitation extremes. E-OBS and the regional reanalyses perform better in winter (see Fig. 7) than in summer while the satellite-based products perform better in winter. Among the non-station-based datasets, HERZ shows the largest skill at all scales and for both seasons. CMORPH and TMPA perform similar

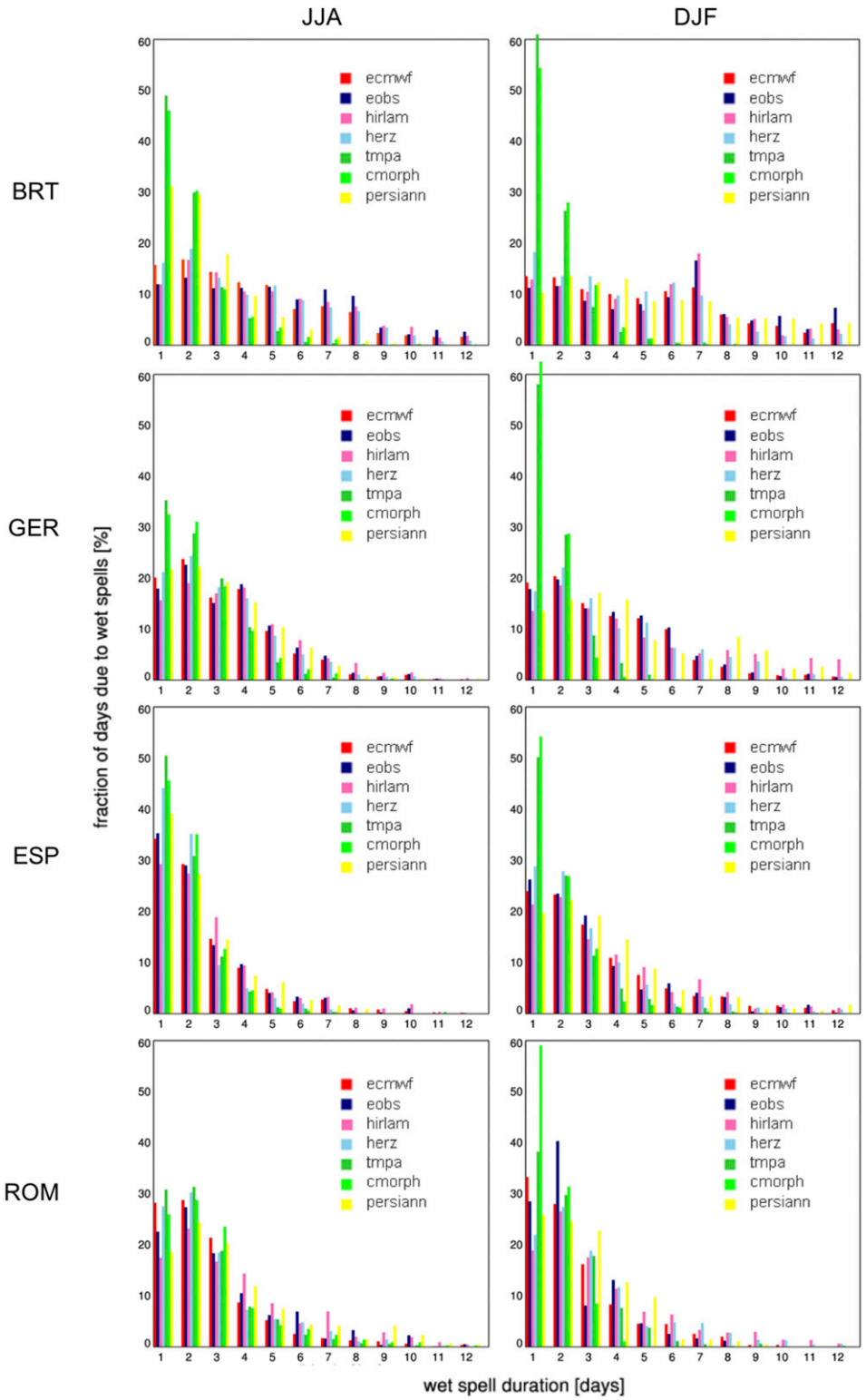

FIG. 5. Distributions of WS durations normalized to the total number of wet days for the four selected climate regions BRT, GER, ESP, and ROM for the reference and candidate datasets for (left) JJA and (right) DJF.

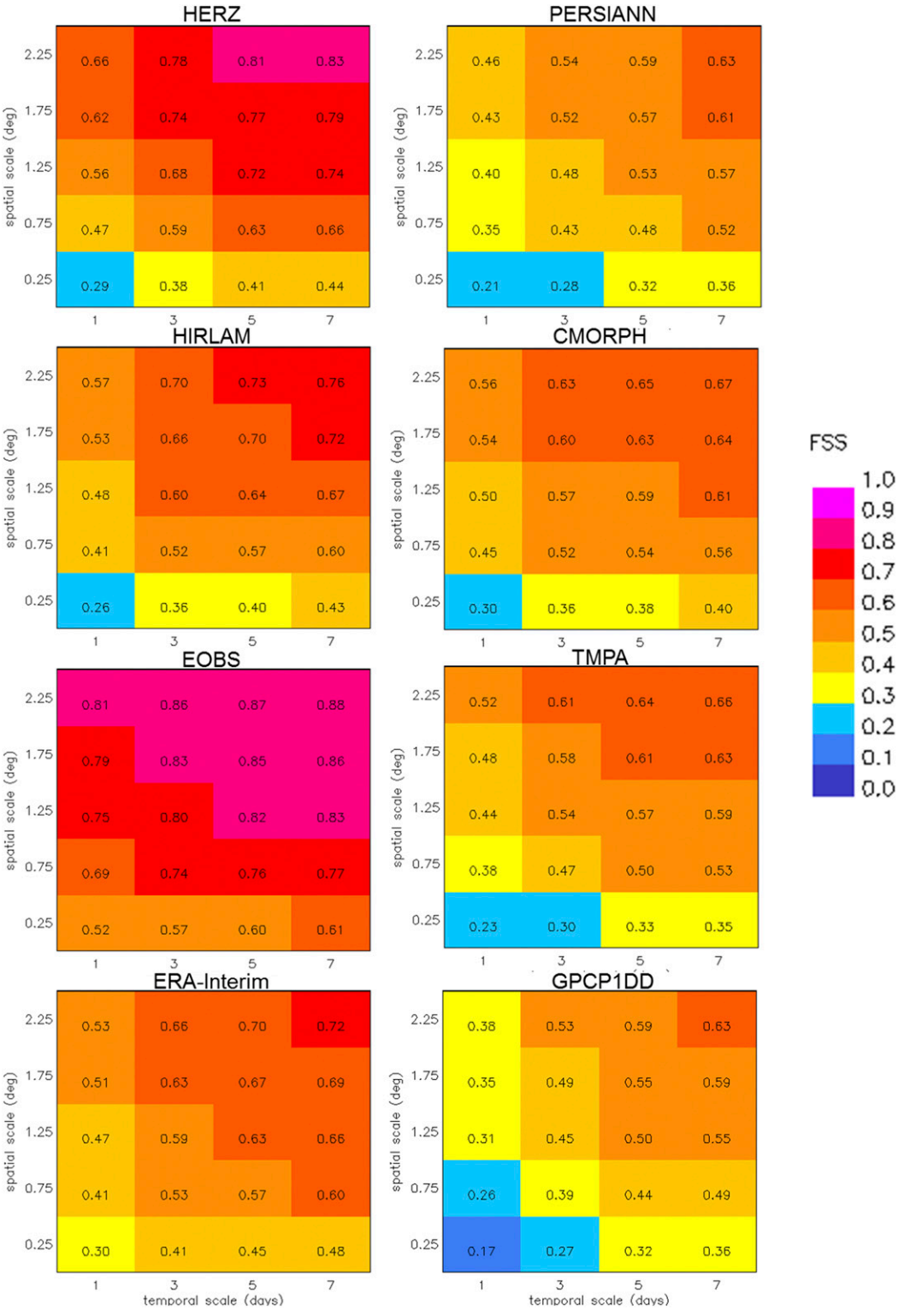

FIG. 6. FSSs as a function of increasing temporal (x axis) and spatial (y axis) scales based on the 90th percentile threshold for summer (JJA) of the years 2007–09 for the candidate and the benchmark datasets. The temporal scale increases from left (original 1-day resolution) to right (7-day resolution), and the spatial scale from bottom (original 0.25° resolution) to top (2.25° resolution).

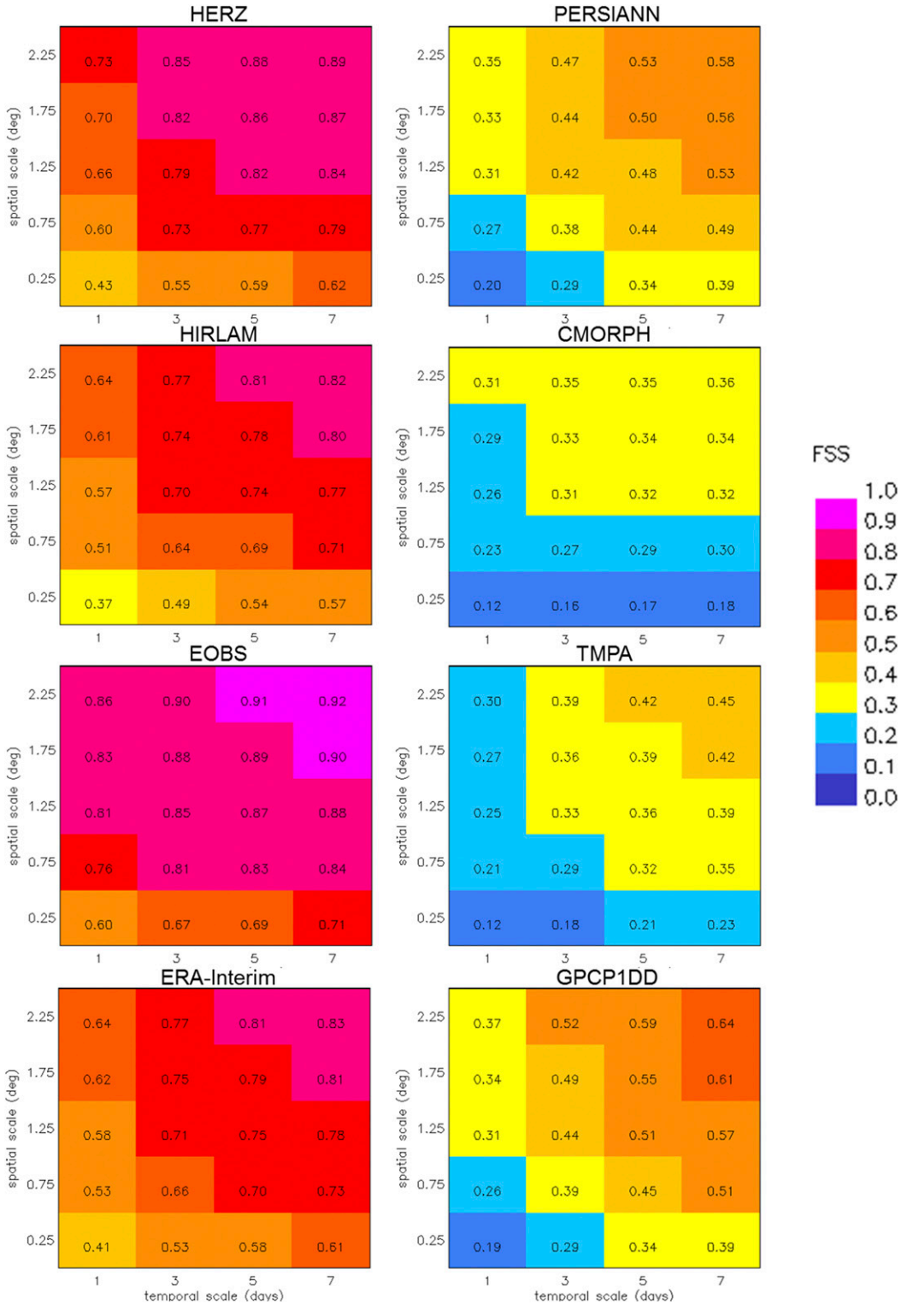

FIG. 7. As in Fig. 6, but for winter (DJF).

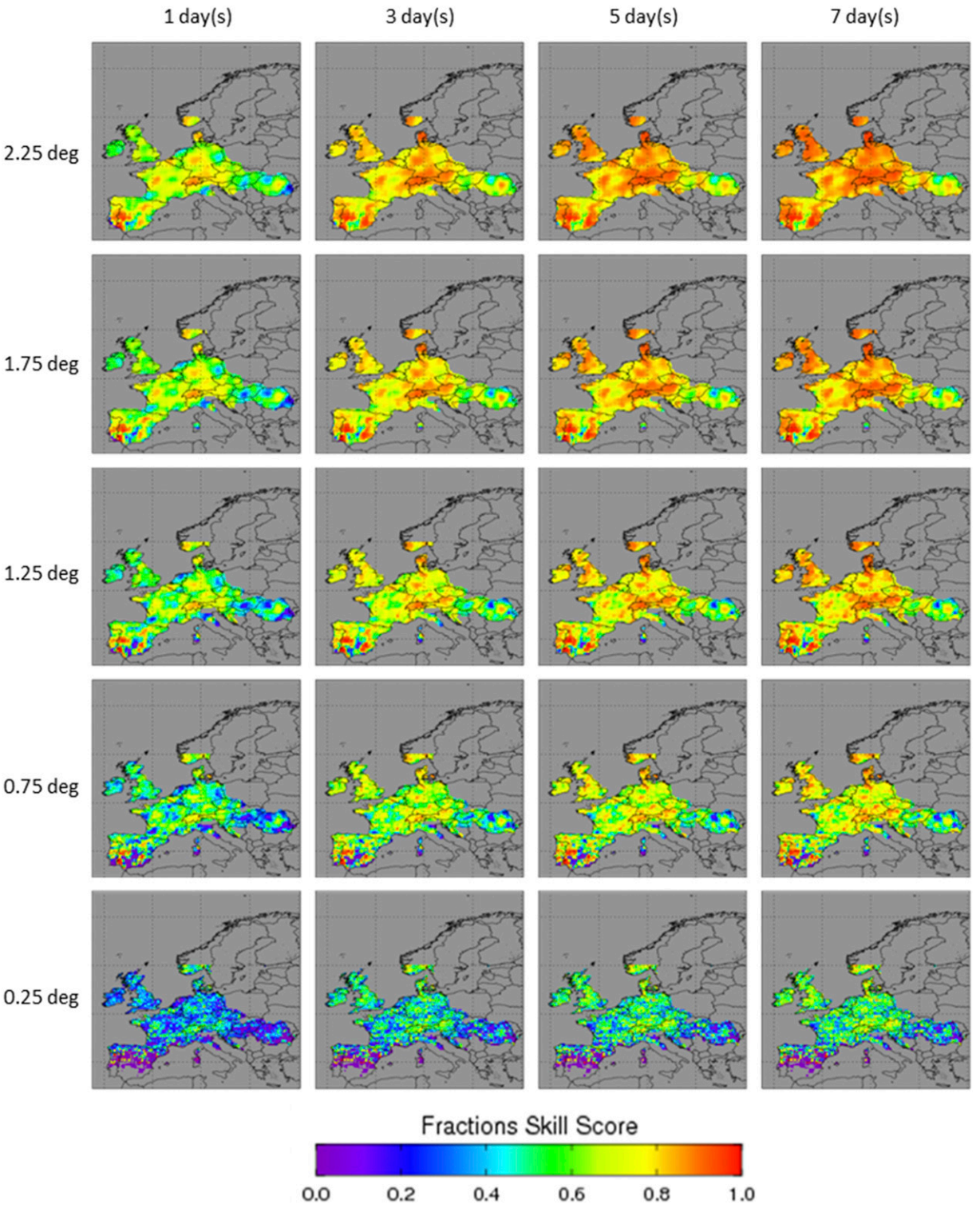

FIG. 8. Spatial distribution of the FSS for HERZ for increasing temporal (1, 3, 5, and 7 days) and spatial (0.25°, 1.25°, 2.25°, and 3.25°) scales based on the 90th percentile threshold for summer (JJA) of the years 2007–09.

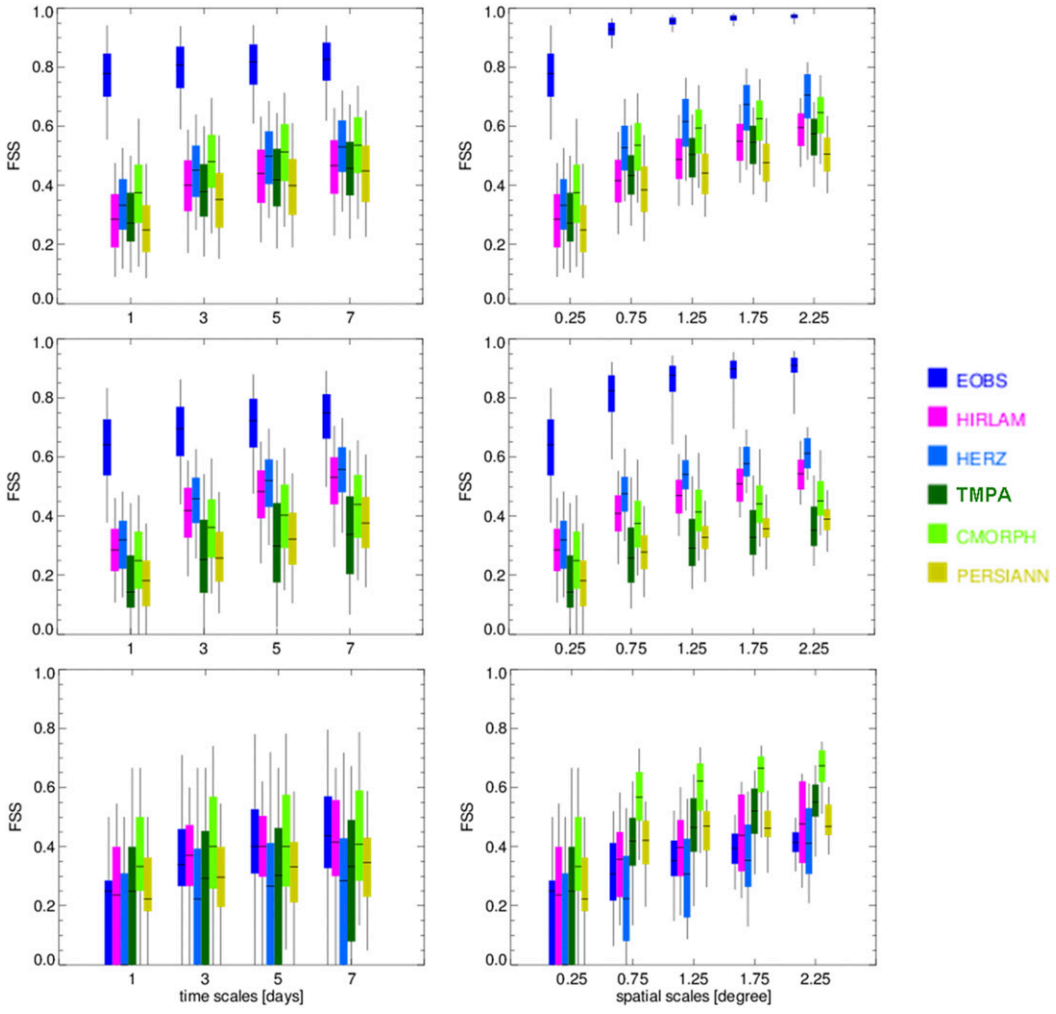

FIG. 9. (left) Temporal and (right) spatial scaling dependencies for the selected climate regions (top) GER, (middle) BRT, and (bottom) ROM by the candidate datasets for summer (JJA) of the years 2007–09.

to HIRLAM in the detection of summer Q90 events from the smallest scales up to 3 days and 1.25°, but slightly worse above. PERSIANN has the lowest skill in summer but outperforms the PMW-based satellite products in winter (see Fig. 7).

The FSS of the candidate datasets for the different scales can vary considerably in space, as shown exemplarily for HERZ in Fig. 8 for summer. At the starting scale (i.e., 1 day and 0.25°) only 10% of the grid boxes have FSS values larger than FSS_{useful}. While the FSS then, for example over Germany, rapidly increases with increasing space–time scales, the values over Romania change much more slowly. Figure 9 shows the temporal (left) and spatial (right) scaling dependencies of the FSS for summer for all candidate datasets for the selected climate regions defined earlier (that is GER, BRT, and ROM). The results for the ESP region are not included as for many grid boxes the FSS could not be calculated due to too few extreme events (minimum is four).

E-OBS performs best over GER (top panels) where it has the largest station density (FSS $>$ 0.9 at 3 days and 1.25°). Over ROM (bottom panels), where E-OBS has much lower station densities, the FSS for 3 days and 1.25° ranges from 0.6 to 0.8 and from 0.4 to 0.7, respectively. Over GER temporal scaling hardly effects the FSS (top-left and middle-left panels), whereas spatial scaling still leads to further improvement. For the other datasets temporal scaling is always beneficial as they all differ from the reference dataset with regard to the definition of a day (in contrast to E-OBS). For BRT and GER increasing the temporal scale beyond 3 days further improves in FSS, but not for ROM. Finally, Fig. 9 also illustrates that over GER and ROM the skills of

individual satellite-based datasets are comparable or better than those from the reanalyses.

To analyze the added value of the higher-resolution datasets compared to the lower-resolution global benchmarks datasets, we also show in Fig. 6 (summer) and Fig. 7 (winter) results of the global ERA-Interim reanalysis and GPCP1DD datasets, which were preprocessed as described in section 2e. For summer (see Fig. 6), both regional reanalysis datasets are more skillful than ERA-Interim at all scales, except for the smallest 0.25° spatial scale. At this scale, ERA-Interim has slightly larger FSS values than both regional reanalysis datasets for all temporal scales considered (i.e., 1–7 days). This feature hints at a double penalty–related behavior due to small positioning errors in the higher resolution datasets. The latter produce often more realistic small-scale precipitation patterns (e.g., convective precipitation dominating in summer), which may be somewhat misplaced though. Once relaxing the requirement for exact (spatial) matching by allowing slight displacements (that is increasing spatial neighborhoods) the feature vanishes. Above the 0.75° spatial scale, HIRLAM performs similar or slightly better than ERA-Interim whereas HERZ performs significantly better than ERA-Interim. For satellite-based datasets, summer season FSS values for GPCP1DD are similar to those from PERSIANN, while both CMORPH and TMPA outperform GPCP1DD at all scales. The differences are larger for CMORPH than for TMPA and more pronounced at smaller than at larger (time and space) scales. The higher skill of the high-resolution satellite products in detecting precipitation extreme events is limited to the summer season though. In winter (see Fig. 7), the skill of all satellite-based datasets is lower than in summer. GPCP1DD reaches similar or even slightly higher FSS values than PERSIANN, TMPA, and CMORPH at all scales. For the reanalysis datasets, the skill is higher in winter with HERZ performing best at all scales. The skills from HIRLAM and ERA-Interim are comparable but lower than for HERZ. For example, at the 3-day temporal and 1.25° spatial scale the FSS for HERZ is 0.79, for HIRLAM 0.70, and for ERA-Interim 0.71. The additional benefit of the regional reanalyses compared to the global ERA-Interim reanalysis in detecting precipitation extremes is hence larger in summer than in winter.

5. Discussion

Whenever the quality of datasets is assessed the question arises of how these differences can be explained. This question goes along with the question about the uncertainties associated with the revealed estimates and their comparability and also relates to the independence of the reference and the candidate datasets and the independence of the datasets among each other.

a. Uncertainties in the reference

It is important to bear in mind that the ECWMF data, which represents the reference within this study, is also prone to errors and uncertainties. In the case of gauge-based datasets, two important sources of uncertainty include (Prein and Gobiet 2017) 1) the stochastic sampling error due to a sparse network density and 2) the systematic measuring error due to wind-induced undercatch.

To bridge the scale gap between point measurements made at stations and the areal averages from model predictions or satellite measurements, station observations are gridded to the resolution of the model or satellite grid. The representativeness of the station-based dataset is largely determined by the spatial sampling, that is, density of the network, especially when the decorrelation length of precipitation is shorter than the mean station distance. For precipitation decorrelation lengths depend on the spatial structure of rainfall patterns, which vary depending on season and region. Thus, convective precipitation has typically shorter decorrelation lengths than stratiform precipitation. Spatial decorrelation lengths generally increase with accumulation intervals (Hatfield et al. 1999) and decrease with increasing intensity (Booij 2002). Osborn and Hulme (1997) found decorrelation lengths for European daily precipitation of about 200 km in summer and 300 km in winter. Ideally, the average station distance should be smaller than the decorrelation length; thus, a minimum of one station per 0.25° grid box for the reference dataset (see also requirement by Haylock et al. 2008) should suffice in general. But at certain regions representativeness may be challenged, for example, in mountainous regions like the Alps, where decorrelation lengths tend to be much shorter, and in less highly sampled regions (e.g., Romania).

The uncertainty due to wind-induced undercatch depends on instrument characteristics (size, shape, exposition of the gauge, etc.), the type of precipitation, and meteorological conditions (wind, precipitation type, air temperature, humidity, radiation) during the precipitation event. Correction methods exit, but differ between countries. As the required information is often not available, undercatch correction is especially challenging for continental-scale and global datasets (Schneider et al. 2014). Therefore, bulk correction factors for monthly climatological conditions are often applied to global datasets such as GPCP, PERSIANN-CDR, and GPCC (Legates and Willmott 1990). The error is large in

snow regions and cold seasons. Hence, under these conditions, the difference found for the various candidate datasets can partly be attributed to uncertainty within the reference dataset.

b. *Uncertainties in the candidate datasets and their comparability*

When comparing different types of datasets (e.g., in situ, model-based, and satellite-based), their structural differences caused by the origin of the estimates must be considered. The quantity actually measured, its uncertainty, its sampling, and its relation to precipitation all have an influence on the resulting precipitation estimate. While the gauge-based datasets (E-OBS and ECMWF) are based on point-based direct measurements of accumulated daily precipitation at the surface, the other datasets used here contain precipitation estimates either based on observations only somehow related to surface precipitation as in case of satellites or on models, where precipitation is the outcome of a chain of simulated processes.

Satellite-based products build upon the assumption that precipitation is associated with cold cloud tops (IR-based techniques, such as PERSIANN) or with microphysical properties of frozen hydrometeors in the upper portion of the cloud (PMW scattering-based retrievals over land used within TMPA and CMORPH; Petty 1995). Due to these assumptions both retrievals work best for convective rainfall as found in the tropics and subtropics and during the warm season over midlatitude continental areas. Consequently, precipitation during midlatitude winter and in maritime-influenced regimes, which is mainly of stratiform character, is often missed or estimated with large uncertainties, which might explain the observed underestimation of precipitation in winter and over northwestern Europe. For the PMW-based products, the large biases in winter are due to retrieval deficiencies for low intensities and over cold/frozen surfaces (Petty 1995). For the IR-based products the false interpretation of high-level but nonprecipitation clouds leads to an overestimation of precipitation frequency as depicted by PERSIANN. As both precipitation ''proxies''—cloud-top height and upper-air ice particles—are only rather loosely coupled to surface rainfall intensity, the relationships may depend on the precipitation regime and thus vary from season to season and from region to region. The resulting precipitation estimates are therefore also affected by the specifics of the regions used for calibrating the retrieval algorithms (Petty 1995; Smith et al. 1992).

The reanalysis products evaluated here also provide only indirect estimates of precipitation, as none of them assimilates observed precipitation. The resulting precipitation fields are therefore based on the representation of its generating processes in the model and on the quality of the initial state achieved by the assimilation of, for example, winds, moisture, and temperature. Increasing horizontal and vertical resolution proved to provide better results due to a more realistic representation of the topography as well as the physical and dynamical processes (Kendon et al. 2012; Knist et al. 2019; Rauscher et al. 2010; Zeng et al. 2016).

On top of uncertainties related to the estimation method, sampling affects the quality of the datasets. As both satellite and model products come as areal average (in place of a point estimate), here only temporal sampling determines the uncertainty. For reanalyses, estimates are already provided as accumulations over time (e.g., over 6 h) and are thus comparable to accumulated precipitation measured by the gauges. Satellite-based estimates only provide snapshots at certain times dictated by satellite orbit and instrument characteristics. Depending on the orbit type the number of snapshots varies between 15 min (GEO) and twice daily (LEO) for individual sensors. Multisensor products (such as TMPA and CMORPH) often come with a 3-hourly resolution, thus daily values are always based on a limited number of snapshots.

c. *Independence of the datasets*

Dependencies between datasets have to be taken into account for any comparison or evaluation exercise. Although the number of stations used in the reference and in the E-OBS datasets differs considerably they cannot be considered independent because of the overlap of stations. Thus, it does not come as a surprise that E-OBS closely agrees with the reference, has only small biases (within 10%), and exhibits the largest skill with the respect to the detection of extreme events.

The other datasets, both the regional reanalysis datasets and the satellite-based datasets do not use any in situ precipitation measurements, except for PERSIANN which is calibrated to the GPCP monthly 2.5° gridded dataset, which merges satellite and in situ observations. Thus, overland gauges in particular have a considerable impact on its precipitation estimate. The impact is, however, limited to monthly or coarser time scales; accordingly, precipitation characteristics based on daily estimates (such as INT, NWET) are only indirectly influenced by the gauge information.

The three satellite-based products used in our evaluation are also not completely independent since they share common observations through the use of PMW and IR data, albeit in very different ways. For example, CMORPH uses the IR data only for deriving motion

vectors and not for precipitation estimation as in TMPA, and PERSIANN only uses IR data. Consequently, although the two PMW-based datasets (CMORPH and TMPA) show generally very similar regional bias patterns for NWET, INT patterns differ considerable (except for the summer season) with the largest differences in winter when CMORPH largely underestimates INT, whereas TMPA tends to overestimate.

While both regional reanalyses evaluated here use the ERA-Interim reanalysis as lateral boundary conditions, they differ with respect to their spatial resolution, assimilation scheme (3D-Var versus nudging), dynamical model core type (hydrostatic versus nonhydrostatic), and parameterization schemes. Accordingly, their precipitation fields may differ considerably. While in summer the spatial bias patterns for INT and the frequency distributions at high intensities are fairly similar, the bias pattern for NWET and the low intensities of the frequency distribution were found to differ considerably. HIRLAM exhibits the commonly known feature of weather and climate models of showing too frequent precipitation (see, e.g., Trenberth et al. 2003; Sun et al. 2006) with respective biases of 40% and 27% for summer and winter. The overestimation of low intensities, another typical model feature, is only visible for HIRLAM in summer, not for HERZ.

6. Summary and conclusions

We have evaluated several widely used delayed-mode precipitation datasets with high temporal and spatial resolution with respect to their ability to represent precipitation characteristics and extremes over western and central Europe. Three years (2007–09) of gridded station-based precipitation fields provided by ECMWF were used as reference because of its uniquely dense gauge network. The candidate datasets encompassed one in situ dataset (E-OBS), two model-based regional reanalyses (HERZ, HIRLAM), and three satellite-based datasets (PERSIANN, TMPA, CMORPH). First, we assessed the ability of the candidate datasets to replicate spatial precipitation patterns and frequency distributions of intensities and WS durations. Then, we determined their skill of precipitation event detection with a focus on extremes, using temporally and spatially matched pairs of precipitation estimates. The results illustrate the general utility of the evaluated datasets, especially during summer months over the region. Their utility may, however, be limited with respect to the representation of certain precipitation quantiles (e.g., E-OBS for high quantiles), during certain seasons (winter for TMPA, CMORPH), over dry regions (especially PERSIANN, HIRLAM), over coastal areas (TMPA, CMORPH), and for certain precipitation regimes (ocean-influenced weather types for TMPA, CMORPH). None of the non-station-based datasets showed skill for the detection of extreme precipitation events at the highest temporal and spatial scales (1 day, 0.25°). But for periods of and longer than 3 days and areas of and larger than 1.25° grid size, the products exhibit considerable skill over large parts of the domain. Both the reanalysis and satellite-based products also revealed added value compared to coarser-resolution global benchmark datasets.

The results illustrate that there is not one single best data product. Each of them is prone to specific time and space varying uncertainties. E-OBS as another ground-based dataset compared best with the reference for obvious reasons and is therefore, as expected, often in closest agreement to the reference. But the results for E-OBS also highlight the importance of station density as a source for uncertainty. The satellite-based precipitation datasets have the greatest difficulties in representing precipitation frequency caused by uncertainties in the rain/no rain detection and due to approximations and assumptions in the retrieval. The former leads to false alarm overestimation of nonprecipitation high clouds, most prominently seen for PERSIANN, to warm rain being missed, and to difficulties over cold/frozen/snow surfaces. The regional reanalysis datasets—although not constrained by precipitation observations—are well able to represent the precipitation characteristics. Remarkably, their performances were still found to differ considerably due to differences, for example, in the spatial resolution, which influences the detail of the topography used and impacts the processes resolved by the model. The regional reanalysis HERZ, which uses a nonhydrostatic model with comparatively high spatial resolution (6 km), is able to overcome known models' problems of showing too frequent too low precipitation as observed for HIRLAM.

To provide guidance on where products are more or less accurate, estimates of the associated uncertainties should be incorporated in the product generation and provided with the datasets. In most cases such uncertainty estimates are not available yet. This information will help users to decide about the fit for purpose for their individual application. The availability of such uncertainty estimates would not only ease the use for specific applications (in for example hydrology or the detection of trends and patterns of change) but would also help to better compare different data products.

Ensembles represent a promising tool to quantify uncertainty. They are increasingly used in the modeling community to quantify uncertainties for weather and climate predictions. In the meantime, also probabilistic

reanalyses become available both on global (e.g., ERA-5) and on regional scales, for example, the COSMO-REA12 (Bach et al. 2016), a 3D-Var ensemble reanalysis from the U.K. Met Office, and the MESCAN-SURFEX surface reanalysis, all of them produced within the EU-funded Uncertainties in Ensembles of Regional ReAnalyses (UERRA) project. Ensemble approaches are also already tested for quantification of in situ data products. For example, a new E-OBS ensemble dataset is under development providing a useful uncertainty estimate by producing an ensemble of grids using conditional stochastic simulations. Ensembles as means of uncertainty estimate may also be an option for satellite-based products as suggested by Povey and Grainger (2015).

As the fuzzy methodology has proven to be better suited (compared to deterministic methods) for comparisons involving such a highly variable parameter as precipitation, especially when it comes to extremes, it would be of particular interest to extend its application both for the space (toward global) and for the time scale (toward higher resolution). This would allow one to get a more complete picture on the useful scales of the global products (in different regions) and to judge the reliability of products (especially models) in representing fluctuations at subdaily time scales, which form the dominant component for the variance of precipitation around the world (Covey et al. 2018).

Acknowledgments. We acknowledge the use of the E-OBS data set 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 and TMPA3B42RT data were provided by the NASA/Goddard Space Flight Center's (GSFC) Mesoscale Atmospheric Processes Laboratory and assessed at ftp://meso.gsfc.nasa.gov/pub/1dd-v1.2/ and ftp://trmmopen.gsfc.nasa.gov/pub/merged/3B42RT/, respectively. We also acknowledge the use of the PERSIANN CDR developed within NOAA's CDR Program, which was acquired from NOAA's National Centers for Environmental Information at https://www.ncei.noaa.gov/data/precipitation-persiann/access/. CMORPH Version 1 was provided by NOAA's Climate Prediction Center (CPC) and downloaded via ftp from ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/. ERA-Interim total precipitation fields were acquired from the ERA-Interim Data Server at http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/. We are grateful to Tomas Landelius and Per Undén (Swedish Meteorological and Hydrological Service, HIRLAM) and Christoph Bollmeyer (formerly at University of Bonn, HERZ) for providing the datasets and to Anna Ghelli (ECMWF) for helping with assessing the high density station data (denoted ECMWF within this study). Olga Zolina benefited from the support by CNRS IGE and UGA and by Helmholtz-RSF Grant 18-47-06202. Software used within this study partly builds upon pieces of code provided by Elisabeth Ebert (https://www.cawcr.gov.au/projects/verification/#Tools_packages).

REFERENCES

Adler, R. F., C. Kidd, G. Petty, M. Morissey, and H. M. Goodman, 2001: Intercomparison of global precipitation products: The third Precipitation Intercomparison Project (PIP-3). *Bull. Amer. Meteor. Soc.*, **82**, 1377–1396, https://doi.org/10.1175/1520-0477(2001)082<1377:IOGPPT>2.3.CO;2.

——, and Coauthors, 2003: The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.*, **4**, 1147–1167, https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

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, https://doi.org/10.1029/2010JD014741.

Allan, R. P., and B. J. Soden, 2008: Atmospheric warming and the amplification of precipitation extremes. *Science*, **321**, 1481–1484, https://doi.org/10.1126/science.1160787.

Ashouri, H., and Coauthors, 2015: PERSIANN-CDR daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Amer. Meteor. Soc.*, **96**, 69–84, https://doi.org/10.1175/BAMS-D-13-00068.1.

Bach, L., C. Schraff, J. D. Keller, and A. Hense, 2016: Towards a probabilistic regional reanalysis system for Europe: evaluation of precipitation from experiments. *Tellus*, **68A**, 32209, https://doi.org/10.3402/tellusa.v68.32209.

Bollmeyer, C., and Coauthors, 2015: Towards a high-resolution regional reanalysis for the European CORDEX domain. *Quart. J. Roy. Meteor. Soc.*, **141**, 1–15, https://doi.org/10.1002/qj.2486.

Booij, M. J., 2002: Modelling the effects of spatial and temporal resolution of rainfall and basin model on extreme river discharge. *Hydrol. Sci. J.*, **47**, 307–320, https://doi.org/10.1080/02626660209492932.

Brdar, S., M. Baldauf, A. Dedner, and R. Klöfkorn, 2013: Comparison of dynamical cores for NWP models: Comparison of COSMO and Dune. *Theor. Comput. Fluid Dyn.*, **27**, 453–472, https://doi.org/10.1007/s00162-012-0264-z.

Brienen, S., A. Kapala, H. Machel, and C. Simmer, 2013: Regional centennial precipitation variability over Germany from extended observation records. *Int. J. Climatol.*, **33**, 2167–2184, https://doi.org/10.1002/joc.3581.

Collins, M., and Coauthors, 2013: Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1029–1136.

——, and Coauthors, 2018: Challenges and opportunities for improved understanding of regional climate dynamics. *Nat. Climate Change*, **8**, 101–108, https://doi.org/10.1038/s41558-017-0059-8.

Covey, C., C. Doutriaux, P. J. Gleckler, K. E. Taylor, K. E. Trenberth, and Y. Zhang, 2018: High-frequency intermittency in observed and model-simulated precipitation. *Geophys. Res. Lett.*, **45**, 12 514–12 522, https://doi.org/10.1029/2018GL078926.

Dahlgren, P., and N. Gustafsson, 2012: Assimilating host model information into a limited area model. *Tellus*, **64A**, 15 836, https://doi.org/10.3402/tellusa.v64i0.15836.

——, T. Landelius, P. Kållberg, and S. Gollvik, 2016: A high-resolution regional reanalysis for Europe. Part 1: Three-dimensional reanalysis with the regional High-Resolution Limited-Area Model (HIRLAM). *Quart. J. Roy. Meteor. Soc.*, **142**, 2119–2131, https://doi.org/10.1002/qj.2807.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteor. Appl.*, **15**, 51–64, https://doi.org/10.1002/met.25.

——, 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, https://doi.org/10.1175/BAMS-88-1-47.

Gehne, M., T. M. Hamill, G. N. Kiladis, and K. E. Trenberth, 2016: Comparison of global precipitation estimates across a range of temporal and spatial scales. *J. Climate*, **29**, 7773–7795, https://doi.org/10.1175/JCLI-D-15-0618.1.

Ghelli, A., and F. Lalaurette, 2000: Verifying precipitation forecasts using upscaled observations. *ECMWF Newsletter*, No. 87, ECMWF, Reading, United Kingdom, 9–17.

Giorgi, F., C. Jones, and G. R. Asrar, 2009: Addressing climate information needs at the regional level: The CORDEX framework. *WMO Bull.*, **58**, 175–183.

Groisman, P. Y., R. W. Knight, D. R. Easterling, T. R. Karl, G. C. Hegerl, and V. A. N. Razuvaev, 2005: Trends in intense precipitation in the climate record. *J. Climate*, **18**, 1326–1350, https://doi.org/10.1175/JCLI3339.1.

Haiden, T., M. J. Rodwell, D. S. Richardson, A. Okagaki, T. Robinson, and T. Hewson, 2012: Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. *Mon. Wea. Rev.*, **140**, 2720–2733, https://doi.org/10.1175/MWR-D-11-00301.1.

Hatfield, J. L., J. H. Prueger, and D. W. Meek, 1999: Spatial variation of rainfall over a large watershed in central Iowa. *Theor. Appl. Climatol.*, **64**, 49–60, https://doi.org/10.1007/s007040050110.

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**, D20119, https://doi.org/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, https://doi.org/10.1007/s00382-009-0698-1.

Huffman, G. J., and Coauthors, 2001: Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeor.*, **2**, 36–50, https://doi.org/10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2.

——, and Coauthors, 2007: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeor.*, **8**, 38–55, https://doi.org/10.1175/JHM560.1.

Isotta, F. A., R. Vogel, and C. Frei, 2015: Evaluation of European regional reanalysis and downscalings for precipitation in the Alpine region. *Meteor. Z.*, **24**, 15–37, https://doi.org/10.1127/metz/2014/0584.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor.*, **5**, 487–503, https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643, https://doi.org/10.1175/BAMS-83-11-1631.

Kendon, E. J., N. M. Roberts, C. A. Senior, and M. J. Roberts, 2012: Realism of rainfall in a very high-resolution regional climate model. *J. Climate*, **25**, 5791–5806, https://doi.org/10.1175/JCLI-D-11-00562.1.

Kidd, C., 2001: Satellite rainfall climatology: A review. *Int. J. Climatol.*, **21**, 1041–1066, https://doi.org/10.1002/joc.635.

——, and V. Levizzani, 2011: Status of satellite precipitation retrievals. *Hydrol. Earth Syst. Sci.*, **15**, 1109–1116, https://doi.org/10.5194/hess-15-1109-2011.

——, P. Bauer, J. Turk, G. J. Huffman, R. Joyce, K. L. Hsu, and D. Braithwaite, 2012: Intercomparison of high-resolution precipitation products over northwest Europe. *J. Hydrometeor.*, **13**, 67–83, https://doi.org/10.1175/JHM-D-11-042.1.

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, https://doi.org/10.1175/1520-0442(2003)016<3665:TIIODT>2.0.CO;2.

Knist, S., K. Goergen, and C. Simmer, 2019: Evaluation and projected changes of precipitation statistics in convection-permitting WRF climate simulations over Central Europe. *Climate Dyn.*, https://doi.org/10.1007/s00382-018-4147-x, in press.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel, 2006: World map of the Koppen-Geiger climate classification updated. *Meteor. Z.*, **15**, 259–263, https://doi.org/10.1127/0941-2948/2006/0130.

Legates, D. R and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatol.*, **10**, 111–127, https://doi.org/10.1002/joc.3370100202.

Lockhoff, M., O. Zolina, C. Simmer, and J. Schulz, 2014: Evaluation of Satellite-retrieved extreme precipitation over Europe using gauge observations. *J. Climate*, **27**, 607–623, https://doi.org/10.1175/JCLI-D-13-00194.1.

Osborn, T. J., and M. Hulme, 1997: Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. *J. Climate*, **10**, 1885–1908, https://doi.org/10.1175/1520-0442(1997)010<1885:DOARBS>2.0.CO;2.

Petty, G. W., 1995: The status of satellite-based rainfall estimation over land. *Remote Sens. Environ.*, **51**, 125–137, https://doi.org/10.1016/0034-4257(94)00070-4.

Povey, A. C., and R. G. Grainger, 2015: Known and unknown unknowns: uncertainty estimation in satellite remote sensing. *Atmos. Meas. Tech.*, **8**, 4699–4718, https://doi.org/10.5194/amt-8-4699-2015.

Prein, A. F., and A. Gobiet, 2017: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *Int. J. Climatol.*, **37**, 305–327, https://doi.org/10.1002/joc.4706.

Rauscher, S. A., E. Coppola, C. Piani, and F. Giorgi, 2010: Resolution effects on regional climate model simulations of seasonal precipitation over Europe. *Climate Dyn.*, **35**, 685–711, https://doi.org/10.1007/s00382-009-0607-7.

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, https://doi.org/10.1175/2007MWR2123.1.

Sapiano, M. R. P., and P. A. Arkin, 2009: An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *J. Hydrometeor.*, **10**, 149–166, https://doi.org/10.1175/2008JHM1052.1.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.*, **115**, 15–40, https://doi.org/10.1007/s00704-013-0860-x.

Seneviratne, S. I., and Coauthors, 2012: Changes in climate extremes and their impacts on the natural physical environment. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 109–230.

Simmer, C., and Coauthors, 2016: HErZ: The German Hans-Ertel Centre for Weather Research. *Bull. Amer. Meteor. Soc.*, **97**, 1057–1068, https://doi.org/10.1175/BAMS-D-13-00227.1.

Smith, E. A., A. Mugnai, H. J. Cooper, G. J. Tripoli, and X. Xiang, 1992: Foundations for statistical-physical precipitation retrieval from passive microwave satellite measurements. Part I: Brightness-temperature properties of a time-dependent cloud-radiation model. *J. Appl. Meteor.*, **31**, 506–531, https://doi.org/10.1175/1520-0450(1992)031<0506:FFSPPR>2.0.CO;2.

Stampoulis, D., and E. N. Anagnostou, 2012: Evaluation of global satellite rainfall products over continental Europe. *J. Hydrometeor.*, **13**, 588–603, https://doi.org/10.1175/JHM-D-11-086.1.

Steppeler, J., G. Doms, U. Schättler, H. Bitzer, A. Gassmann, U. Damrath, and G. Gregoric, 2003: Mesogamma scale forecasts using the nonhydrostatic model LM. *Meteor. Atmos. Phys.*, **82**, 75–96, https://doi.org/10.1007/s00703-001-0592-9.

Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *J. Climate*, **19**, 916–934, https://doi.org/10.1175/JCLI3672.1.

Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, **84**, 1205–1217, https://doi.org/10.1175/BAMS-84-9-1205.

——, Y. Zhang, and M. Gehne, 2017: Intermittency in precipitation: Duration, frequency, intensity, and amounts using hourly data. *J. Hydrometeor.*, **18**, 1393–1412, https://doi.org/10.1175/JHM-D-16-0263.1.

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, https://doi.org/10.1175/2008BAMS2652.1.

Wahl, S., C. Bollmeyer, S. Crewell, C. Figura, P. Friederichs, A. Hense, J. Keller, and C. Ohlwein, 2017: A novel convective-scale regional reanalysis COSMO-REA2: Improving the representation of precipitation. *Meteor. Z.*, **26**, https://doi.org/10.1127/metz/2017/0824.

Wilks, D., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.

Zeng, X.-M., M. Wang, Y. Zhang, Y. Wang, and Y. Zheng, 2016: Assessing the effects of spatial resolution on regional climate model simulated summer temperature and precipitation in China: A case study. *Adv. Meteor.*, **2016**, https://doi.org/10.1155/2016/7639567.

Zolina, O., A. Kapala, C. Simmer, and S. K. Gulev, 2004: Analysis of extreme precipitation over Europe from different reanalysis: A comparative assessment. *Global Planet. Change*, **44**, 129–161, https://doi.org/10.1016/j.gloplacha.2004.06.009.

——, C. Simmer, A. Kapala, and S. K. Gulev, 2005: On the robustness of the estimates of centennial-scale variability in heavy precipitation from station data over Europe. *Geophys. Res. Lett.*, **32**, L14707, https://doi.org/10.1029/2005GL023231.

——, ——, K. Belyaev, A. Kapala, and S. Gulev, 2009: Improving estimates of heavy and extreme precipitation using daily records from European rain gauges. *J. Hydrometeor.*, **10**, 701–716, https://doi.org/10.1175/2008JHM1055.1.

——, ——, S. K. Gulev, and S. Kollet, 2010: Changing structure of European precipitation: Longer wet periods leading to more abundant rainfalls. *Geophys. Res. Lett.*, **37**, L06704, https://doi.org/10.1029/2010GL042468.

——, ——, K. P. Belyaev, S. K. Gulev, and K. P. Koltermann, 2013: Changes in European wet and dry spells over the last decades. *J. Climate*, **26**, 2022–2047, https://doi.org/10.1175/JCLI-D-11-00498.1.

——, ——, A. Kapala, P. Shabanov, P. Becker, H. Mächel, S. Gulev, and P. Groisman, 2014: Precipitation variability and extremes in central Europe: New view from STAMMEX results. *Bull. Amer. Meteor. Soc.*, **95**, 995–1002, https://doi.org/10.1175/BAMS-D-12-00134.1.