Evaluation of IMERG and MSG-CPP Precipitation Estimates over Europe Using EURADCLIM: A Gauge-Adjusted European Composite Radar Dataset

Journal of Hydrometeorology

van der Plas, Emiel; Overeem, Aart; Meirink, Jan Fokke; Leijnse, Hidde; Bogerd, Linda

https://doi.org/10.1175/JHM-D-23-0184.1

Evaluation of IMERG and MSG-CPP Precipitation Estimates over Europe Using EURADCLIM: A Gauge-Adjusted European Composite Radar Dataset

EMIEL VAN DER PLAS,[a] AART OVEREEM,[a] JAN FOKKE MEIRINK,[a] HIDDE LEIJNSE,[a] AND LINDA BOGERD[b]

[a]*KNMI, Royal Netherlands Meteorological Institute, De Bilt, Netherlands*
[b]*WUR, Wageningen University and Research, Wageningen, Netherlands*

(Manuscript received 17 October 2023, in final form 7 May 2024, accepted 6 June 2024)

ABSTRACT: A new pan-European climatological dataset was recently released that has a much higher spatiotemporal resolution than existing pan-European interpolated rain gauge datasets. This radar dataset of hourly precipitation accumulations, European Radar Climatology (EURADCLIM) (Overeem et al.), covers most of continental Europe with a resolution of 2 km × 2 km and is adjusted employing data from potentially thousands of government rain gauges. This study aims to use this dataset to evaluate two important satellite-derived precipitation products over the period 2013–19 at a much higher spatiotemporal resolution than was previously possible at the European scale: the IMERG late run and the Meteosat Second Generation (MSG) cloud physical property product from the SEVIRI instrument. The latter is only available during daytime, so the analyses are restricted to daytime conditions. A direct gridcell comparison of hourly precipitation reveals an apparently low coefficient of correlation. However, looking into slightly more detail at statistics pertaining to longer time scales or specific areas, the datasets show good correspondence. All datasets are shown to have their specific biases, which can be transient or more systematic, depending on the timing or location. The MSG precipitation seems to have an overall positive bias, and the IMERG dataset suffers from some transient overestimation of certain events.

KEYWORDS: Climatology; Hydrometeorology; Mixed precipitation; Radars/Radar observations; Satellite observations

1. Introduction

Precipitation can be measured in different ways, which all have their strengths and weaknesses. Rain gauges yield the most direct measurement, providing confident information on whether and how much it is raining at a specific location. Ground-based weather radars make indirect observations with several uncertainties but cover larger domains at relatively high spatiotemporal resolution. In addition, biases can be reduced considerably by calibration with rain gauge networks. From space, precipitation can be estimated via microwave radiometers on a constellation of low-Earth-orbiting satellites, or even more indirectly through visible and infrared radiance measurements from geostationary satellites such as Meteosat Second Generation (MSG). Satellite quantitative precipitation estimation (QPE) has a clear advantage over ground-based radar with its quasi-global coverage, and retrievals can be made over sea and mountains or otherwise inaccessible terrain, although the surface parameters (e.g., frozen surfaces and mountains) can influence the quality of the retrieval considerably. In addition, low-Earth-orbiting satellites do not have continuous coverage and rely on interpolation between overpasses. Nevertheless, satellites are the only means to obtain precipitation information over vast areas of Earth's land surface and oceans.

Western Europe has a large variety of different climate zones, and many countries have dense observational networks. This makes Europe an interesting testbed for the comparison and evaluation of precipitation datasets. Eventually, this could lead to improved satellite precipitation retrieval algorithms or better calibration. Hence, ground validation in Europe is eventually also relevant for satellite QPE in those regions across the globe with similar precipitation events, but that lack proper reference data.

The release of the climatological pan-European gauge-adjusted radar dataset European Radar Climatology (EURADCLIM) has opened up a range of possibilities for validation at the pan-European scale (Overeem et al. 2023). Until recently, only gridded precipitation datasets based on interpolated rain gauge accumulations were available over Europe, in the best case daily accumulations at 0.1° and 0.25° grids (Cornes et al. 2018). Apart from obvious applications such as the evaluation of numerical weather prediction (NWP) suites, the relatively large area over which the EURADCLIM dataset is available makes it possible to validate satellite precipitation retrievals with respect to a reliable gridded ground-based dataset over the very different precipitation climate zones in Europe.

The EURADCLIM dataset has attempted to overcome some of the shortcomings of the Operational Programme for the Exchange of Weather Radar Information (OPERA) radar composite product (instantaneous surface rain rate) from which it was derived. Ground-based weather radar precipitation estimates suffer from several challenges. Nonmeteorological echoes, called clutter, cause precipitation overestimation. These echoes may be due to the refraction of radar beams toward Earth's surface, obstacles around radar sites (buildings and orography), or interference. This clutter can be structural, but it is also very short lived. Also, errors in the retrieval can result in precipitation underestimation, with precipitation estimates on average decreasing with increasing distance from

Corresponding author: Emiel van der Plas, emiel.van.der.plas@knmi.nl

DOI: 10.1175/JHM-D-23-0184.1

the radar. For instance, attenuation of radar signals due to extreme rainfall or due to changes in the vertical profile of reflectivity yields lower rain rates. Mountainous areas pose their own set of issues with regard to retrieving reliable precipitation information due to (partial) beam blockage. See Fabry (2015) and Rauber and Nesbitt (2018) for an overview of sources of error in radar QPE. In EURADCLIM, additional algorithms have been applied to mitigate nonmeteorological echoes and the QPE has been much improved by merging with rain gauge data from the European Climate Assessment and Dataset (ECA&D; https://www.ecad.eu). Despite its remaining shortcomings, EURADCLIM is a unique dataset, with high-resolution gridded precipitation data for the period 2013–20.

In this paper, the EURADCLIM dataset is used to validate two satellite-based precipitation products: the Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (GPM) late-run (IMERG-L) (half-hourly sum) dataset (Hou et al. 2014; Skofronick-Jackson et al. 2018) and the precipitation from the cloud physical properties (CPPs) as derived from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) instrument aboard the geostationary MSG satellite (Roebeling and Holleman 2009). This is done to get a perspective on the quality of their quantitative precipitation estimates, which can be relevant for nowcasting purposes (CPPs), hydrological applications, etc. Those datasets have clear advantages, such as large or even worldwide coverage (60°N–60°S, IMERG), but also their own caveats, such as uncertainty due to spatiotemporal interpolation (IMERG), and only being available during daylight (MSG). Both MSG-CPP rain (Roebeling and Holleman 2009) and IMERG have been validated extensively, e.g., in Bogerd et al. (2020) and Gaona et al. (2016), using a Dutch rain gauge–adjusted radar product, and in Navarro et al. (2019) against E-OBS. This is (to the author's knowledge) one of the first studies that compares IMERG and MSG precipitation climatologies for such an extensive area.

The datasets and the methodology will be described in sections 2 and 3. Given the size and the detail of the datasets, a choice has been made of a number of ways to look at differences and similarities between them. These will be presented in section 4. Finally, a short conclusion section is dedicated to the discussion of a number of these results.

2. Data

a. *Gauge-adjusted radar data: EURADCLIM*

The European climatological high-resolution gauge-adjusted radar precipitation dataset EURADCLIM data are described in Overeem et al. (2023). This is a gridded precipitation dataset of 1 and 24 h accumulates every clock hour that covers the major part of Europe from 2013 to 2020 on an approximately 2 km × 2 km grid. It is publicly available (Overeem et al. 2022). In Fig. 1, we give an impression of the area.

The OPERA dataset is a composite of many (on average 138) radars across western and central Europe. Each radar has its own processing, after which they are combined to give

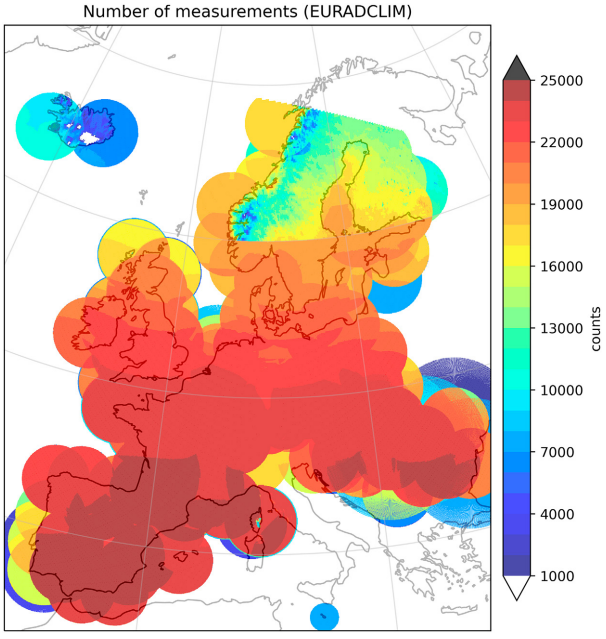

FIG. 1. The number of measurements that were taken into account under the condition of simultaneous availability (daylight hours) of the three datasets over the period 2013–19.

a precipitation estimate over the whole area every 15 min (Saltikoff et al. 2019). In this process, already large amounts of nonmeteorological echoes and artifacts have been removed, both by the National Meteorological and Hydrological Services (NMHS) and by the OPERA program (Saltikoff et al. 2019). For EURADCLIM, additional removal of nonmeteorological echoes was performed by employing two statistical methods and the Cloud Property Dataset using SEVIRI, Edition 2 (CLAAS-2), satellite cloud-type mask. Moreover, disaggregated 1-h rain gauge accumulations from the ECA&D (https://www.ecad.eu) project (Klein Tank et al. 2002; Klok and Klein Tank 2009) are combined with the OPERA radar data, using a local mean-field bias adjustment succeeded by a spatial adjustment, resulting in the EURADCLIM dataset (Overeem et al. 2023). This does, however, not prevent all spurious radar echoes and other artifacts to pollute the data, resulting in some remaining features that are very easy to spot in accumulations over time, but are not easy to remove safely in an automated procedure on gridded data. In addition, the adjustment with rain gauge data is not expected to resolve all inaccuracies in the OPERA radar accumulations.

EURADCLIM is a climatological dataset in the sense that it contains a historical archive over multiple years. Given changes in the number of rain gauges, the radar sites, and the algorithms applied to radar data, as well as changes in environmental conditions, the quality of OPERA radar data, and hence EURADCLIM, can change in the course of time. This is demonstrated by Park et al. (2019) (e.g., Fig. 5) for the unadjusted (OPERA) radar data for the period May–October for the years 2015–17.

The MSG and IMERG data are distributed on different grids, both coarser than the EURADCLIM grid. We chose to

resample those onto the EURADCLIM grid to still have the highest resolution data available for analysis. To prevent interpolation artifacts, nearest-neighbor resampling was used: within a radius of 5 km from the center of the target grid cell, a value is sought on the original grid. This value is assigned to the new grid cell.

b. *Satellite-derived precipitation product: MSG-CPP*

MSG is a series of four geostationary satellites carrying the SEVIRI aboard. SEVIRI is an imager with 12 channels in the visible to infrared spectral range. Full-disk SEVIRI measurements are made every 15 min at a subsatellite resolution of 3×3 km^2. The MSG-CPP product relies on the retrieval of a range of cloud physical properties from a combination of channels (Benas et al. 2017). Subsequently, precipitation rate estimates are derived from cloud-condensed water path, particle effective radius, cloud thermodynamic phase, and cloud height retrievals.[1] More details on the algorithm and its validation can be found in Roebeling and Holleman (2009) and Roebeling et al. (2012). Measurements from the two Dutch weather radars were used to tune the MSG-CPP precipitation algorithm. Note that a correction for parallax effects is not included in this study.

Because shortwave channels are required to derive the MSG-CPP products, they are only available during daytime (solar zenith angle smaller than 84°). This has substantial consequences for the comparison presented in this paper. All the data presented here are limited by the availability of the precipitation estimates based on the SEVIRI instrument. This means that all the radar data and IMERG satellite precipitation data are clipped in a way that if, for one grid cell within a 1-h accumulation, the MSG value is unavailable, the point is discarded from the comparison. The availability is dependent on the diurnal cycle and the season, with large parts of Scandinavia not getting enough daylight to have a proper MSG precipitation estimate.

c. *Satellite network-derived precipitation product: IMERG*

A polar-orbiting satellite with a radiometer and radar constitutes the GPM *Core Observatory*, which is part of the international GPM constellation (Huffman et al. 2007), consisting of various satellites. The GPM *Core Observatory*, which was launched in 2014, extended the measurement capabilities of the Tropical Rainfall Measuring Mission (TRMM) with a dual-frequency precipitation radar (DPR) and a multichannel GPM Microwave Imager (GMI), which made it possible to measure light precipitation intensities and falling snow. Because it consists of polar-orbiting satellites, the sampling is different from, e.g., the geostationary MSG satellite data. Most of the data (60%) have sampling less than 1 h apart and 80% less than 3 h apart. Above the European area that is considered in this paper, the revisit times are between approximately 1 and 3 h, depending on the observation type (Hou et al. 2014).

Apart from the data of the individual contributing instruments, the data are also offered as a gridded integrated product or so-called IMERG (Hou et al. 2014). All the data are combined to form a half-hourly precipitation product on a $0.1° \times 0.1°$ spatial resolution, where interpolation is used to fill data gaps, based on geostationary satellite precipitation products. This precipitation product comes in three iterations.

The early run (IMERG-E) is a product released with a 4-h delay with respect to the measurement time. The product that we used in this study is the late run (IMERG-L v06), which is released with a 14-h latency.[2] The main difference between the early run and the late run is that the former only uses forward extrapolation in time, while the latter uses forward and backward propagation to interpolate between the measurements and uses measurements that were not processed in time to be used in the early run. Finally, there is also a final run (IMERG-F), which uses GPCC ground stations (Global Precipitation Climatology Centre; Becker et al. 2013; O et al. 2017) to adjust the measurements at different time scales and for different geographical areas. This is by far the best IMERG product over regions with good coverage of ground measurements, but it is released 4 months after the measurement time. This is the main reason not to use it in this validation.

Monthly gauge adjustments are only applied to the IMERG final run, not to the late run. The late run does include Combined Ku Radar-Radiometer Algorithm (CORRA) product data from the GPM core satellite, which incorporates climatological adjustments from GPCC gauges. However, the gauge influence is expected to be limited due to the limited coverage of GPCC gauges, and the adjustments are only climatological and are only applied on overpasses from the core satellite.

3. Method

Evaluation approach

As mentioned, the data of the three independent precipitation estimates have been resampled to the grid of the EURADCLIM data. Subsequently, the data were accumulated to clock-hour intervals, so two 30-min IMERG measurements were added, as were four MSG 15-min measurements. MSG grid cells flagged to be outside daylight are dismissed, and a gridcell hour is only counted if all four MSG 15-min values are available. If any of the files had an issue, the hour would be removed from the analysis.

Comparisons between gridded datasets may suffer from representativity issues. Even when we sample them all onto the same grid, they all come from different measurements, leading to precipitation in a grid cell for one dataset but not in the other, while looking at the same scene. Though this is more of an issue for NWP verification applications, for the current comparisons it is clear that comparing values of individual grid cells or pixels may not be very informative,

[1] Near-real-time MSG-CPP data are publicly available via the KNMI data portal at https://msgcpp.knmi.nl/.

[2] IMERG-L data are freely available at the NASA portal: https://disc.gsfc.nasa.gov/.

because of the highly localized nature of precipitation. For the satellite data, this would be a subgrid scale. Furthermore, the MSG data may have some parallax issues, especially at higher latitudes. According to Roebeling and Holleman (2009), the shift due to collocation issues and to parallax in midlatitudes may be as much as 20 km, or approximately 10 EURADCLIM pixels. The nature of the retrieval method, being based on the estimation of cloud liquid water path using light of different wavelengths, results in less contiguous precipitation patterns.

This indicates that it makes sense to look at downsampled data, in the sense that for every pixel one can look at the average amount of precipitation in an area of, e.g., 3 × 3 pixels around the pixel of interest, similar to how fraction skill scores are determined, as introduced by Roberts and Lean (2008). This results in smoother maps, retaining the information of the data source with the highest resolution, without resorting to rigorous downsampling. A number of neighborhoods were used to create smoothed data in this sense: 0, 1, 3, 5, 11, 15, and 25 grid cells, equivalent to square areas with sides of 1, 3, 7, 11, 23, 31, and 51 pixels. The average for neighborhood size n is then taken over a square area defined by a central point (i_0, j_0) with lower-left corner $(i_0 - n, j_0 - n)$ and upper-right corner $(i_0 + n, j_0 + n)$. In this paper, the kernel with 31 × 31 pixels (nb = 15, approximately 60 km × 60 km) is often used as an illustration, as this is a large enough area to resolve the potential parallax issue with respect to the MSG data, as mentioned above.

Also, for a single pixel there is a relatively large chance that the different sources registered different amounts of rain, but for a larger area this mismatch becomes smeared out. On the other hand, looking at a too large area would lose relevance for a local analysis. We attempted to alleviate the representativity issue in this paper by looking at different neighborhoods around the pixel of interest, both for scatter density plots as contingency table statistics and, e.g., bias maps.

We will first discuss yearly accumulations and then look more at monthly accumulation to discuss any seasonal effects. Zooming in further, we will look at some specific cases where heavy precipitation caused flooding in central Europe.

4. Results

a. Scatter density plots

For comparison, the most straightforward thing to do is to make a log–log scatterplot of the two datasets that are considered.

In Fig. 2, scatter density plots are given of hourly precipitation over the whole period from 2013 to 2019, with EURADCLIM data on the horizontal axis and MSG (left panels) and IMERG data (right panels) on the vertical axis. This is data that have been drawn from a reduced grid (for data management purposes), by taking a point every 50 EURADCLIM grid cells. The top row shows the scatter density plot for pixel-by-pixel comparisons ("neighborhood" nb = 0), and the bottom row shows a scatter density plot when we look at the comparison of an average over a 31 × 31 pixel area (neighborhood nb = 15), in EURADCLIM pixels, so approximately 60 × 60 km^2. The R^2 is calculated as

$$R^2 = 1 - \frac{\mathrm{SS}_{\mathrm{res}}}{\mathrm{SS}_{\mathrm{tot}}}.$$

In this equation, $\mathrm{SS}_{\mathrm{res}} = \sum_{i,j} n_{\mathrm{hist},i,j}(b_{\mathrm{sat},i} - b_{\mathrm{rad},j})^2$ is the residual sum of squares and $\mathrm{SS}_{\mathrm{tot}} = \sum_{i,j} n_{\mathrm{hist},i,j}(b_{\mathrm{sat},i} - \overline{\mathrm{satobs}})$ is the total sum of squares, where $b_{\mathrm{sat},i}$ and $b_{\mathrm{rad},j}$ are the central precipitation rate values of the satellite and radar bins, respectively, $\overline{\mathrm{satobs}}$ is the average of the satellite observations, and $n_{\mathrm{hist},i,j}$ are the counts in the specific bin of the histogram. The R^2 becomes slightly negative for the top row in Fig. 2 ($R^2 = -0.07$ and -0.10 for MSG and IMERG, respectively) and slightly higher for the nb = 15 case: $R^2 = 0.07$ and 0.15.

The most important message from these plots is that a straightforward pixelwise comparison of hourly precipitation data is not a good indicator of how well any observation system captures precipitation. Looking at averages in a spatial or temporal sense helps dealing with representativity issues, but it is interesting to see what patterns emerge in different statistics. In this paper, we shall look at a number of ways to compare the satellite precipitation products to the calibrated European ground-based radar data.

b. Contingency table statistics

A well-established way to evaluate a data source with respect to dataset that is assumed to be reliable (in this case EURADCLIM radar product) is contingency table statistics [CTS, well described in, e.g., Ebert and McBride (2000)]. With respect to some thresholds, a quantity is reduced to a number of binary true/false values (e.g., precipitation exceeds 1 mm h^{-1} or not), where true values (i.e., exceedances) are also called events. When both the satellite and the radar measure an event at a certain point in space and time, this is called a hit; when the radar measures an event and the satellite does not, it is a miss; and when a satellite measures an event and the radar does not, it is called a false alarm. From these counts, scores can be constructed between 0 (sometimes −1) and 1, such as the probability of detection (POD), false alarm ratio [or probability of false detection (POFD)], and the critical success index (CSI). The POD (number of hits divided by the number of hits and misses) measures the ratio of corresponding events with respect to the total number of events as seen by the radar. The POFD (number of false alarms divided by the total number of nonevents seen by the radar) is an indication of how specific the satellite is in what it registers as an event. In this paper, we have chosen to only highlight the CSI to quantify the agreement between the datasets. The CSI is defined as the total number of corresponding events (hits) divided by the total number of satellite events plus the number of misses (hits + false alarms + misses). This number is not affected by the number of nonevents (dry hours), which may vary considerably over the European mainland. The CSI can be seen as a measure of the general quality or correspondence of precipitation events between the two datasets.

In Fig. 3, a map is shown of the CSI for the same reduced grid of the EURADCLIM domain as was used for Fig. 2, for the whole 2013–19 period. The top row compares the CSI

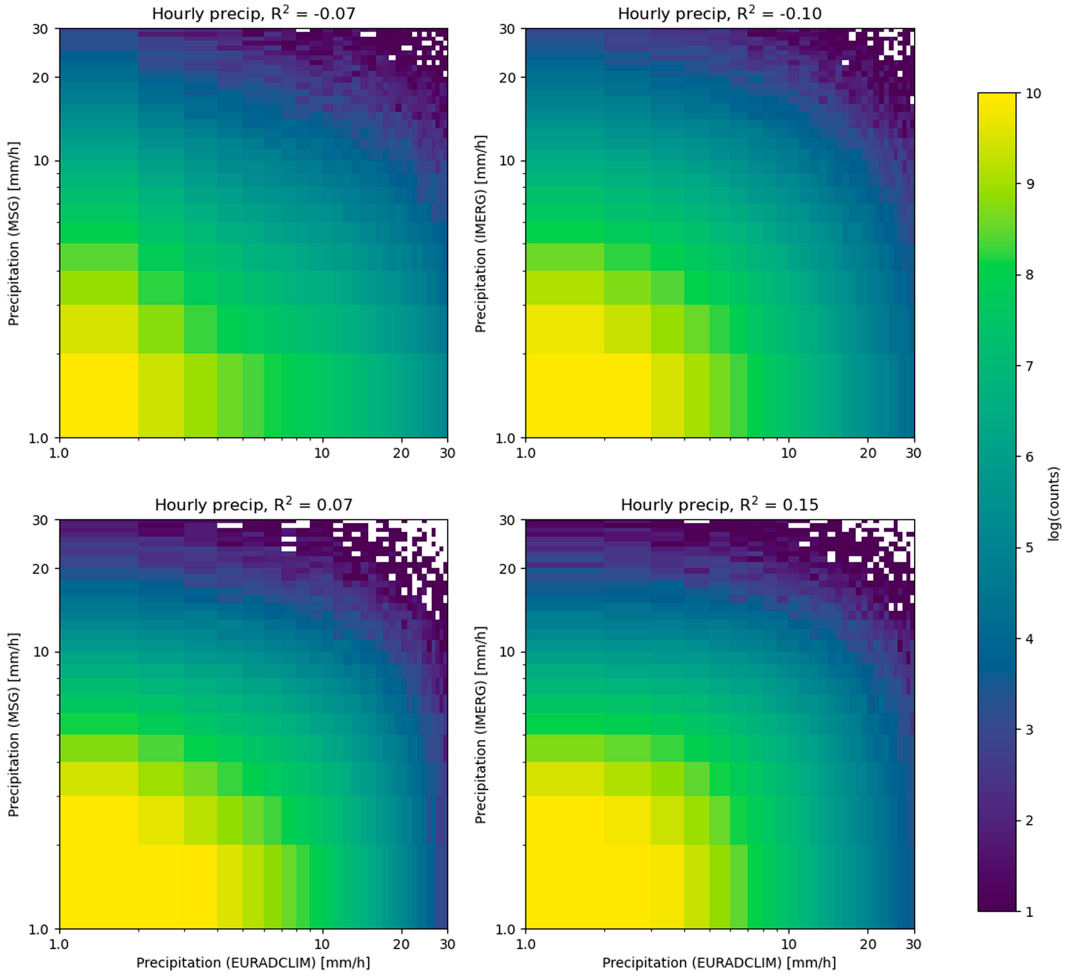

FIG. 2. Scatter density plots of (left) MSG and (right) IMERG vs EURADCLIM for hourly precipitation for the whole period on a reduced grid (see text) on a log–log scale. (top) Cell-by-cell comparison (neighborhood size 0); (bottom) an average over 31 × 31 EURADCLIM grid cells around the grid cells of the reduced grid (≈60 km × 60 km).

with respect to EURADCLIM for MSG (on the left panel) and for IMERG (on the right panel), when only the pixels of the reduced grid themselves are being regarded. The bottom row shows the same comparison, only when exceedance is taken as a mean over an area of 31 × 31 pixels around these same pixels. If we compare left to right, we see significantly higher scores for IMERG than for MSG. For MSG, large parts of Europe are in the 0.1–0.15 range, whereas IMERG is consistently in the 0.2–0.3 range. If we compare top to bottom, we see that both datasets benefit from considering them on a somewhat larger scale. One feature to note here already is how, especially in the MSG nb = 15 image in the lower-left panel, the Alps show up as having a particularly poor critical success index. This can be attributed to imperfections in the cloud masking algorithm, causing snow cover on the ground to be interpreted as a thick cloud that has a high probability of producing high-intensity rain. This aspect will be revisited later. Furthermore, the MSG image seems to display a gradient in the CSI from north to south. In principle, the north experiences more rain, but because of the zenith angle of the geostationary satellite with respect to these regions this may impact the overall quality of the retrievals, making them worse at higher latitudes. This north-to-south gradient in skill is not really visible in the IMERG maps. The poor scores above sea are due to the EURADCLIM dataset not being calibrated to have good values at these locations, by lack of ground stations at sea. The same holds for locations at the edge of the covered area, such as in eastern Europe.

c. *Yearly accumulation*

A first glance at Fig. 4, showing comparisons of yearly accumulations of precipitation, suggests that the three products are in good agreement. Most of the precipitation is well captured by all three observation systems. The west coasts of countries like Spain, Ireland, and Norway stand out, indicating yearly sums of more than 500 mm.

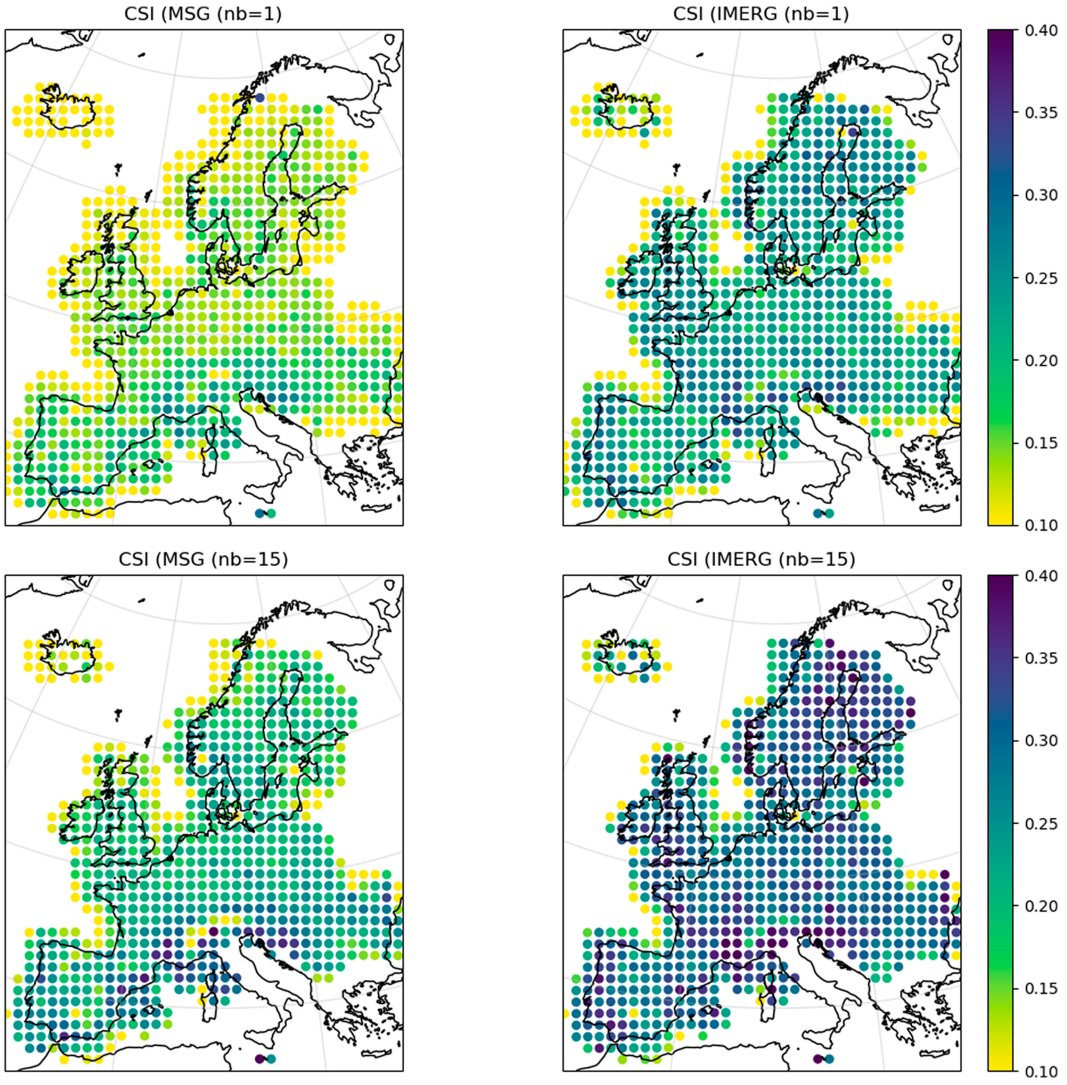

FIG. 3. Maps of the CSI for precipitation exceeding 1 mm h^{-1} for (left) MSG and (right) IMERG vs EURADCLIM for hourly precipitation for the whole period on a reduced grid (see text). (top) Pixel-by-pixel comparison (neighborhood size 0); (bottom) an average over 31 × 31 EURADCLIM pixels around the pixels of the reduced grid (≈60 km × 60 km).

Differences are particularly clear over mountainous areas that are covered by snow during parts of the year. All three systems have their own weaknesses here: Radars may have spurious echoes or miss precipitation because the mountains block the radar's signal. However, especially the MSG algorithm fails to make the distinction between snow-covered land and (intense) precipitation. For certain months, the IMERG product does not yield data above Scandinavia and Iceland for probably the same reason. The IMERG product interpolates in the temporal dimension, which leads to relatively smooth precipitation estimates, but may not capture the dynamics of a storm in the same level of detail. This aspect will be revisited in section 4f.

In some locations, a clear anisotropy in the bias with respect to EURADCLIM can be seen, such as along the coastlines and, e.g., in northern Ireland. This is due to the shape of the radar observation cone, where close to the instrument higher intensities are registered than further away. Also, permanent artifacts due to, for example, blocking can be seen as a local drop in the bias of the satellite products.

These issues are immediately apparent from figures showing the bias of MSG and IMERG with respect to EURADCLIM: The edges of the dataset (e.g., over sea) turn red, indicating that EURADCLIM is clearly underestimating the precipitation over sea, due to the fact that over sea the typically declining sensitivity of the radar with distance is not adjusted by the weighing of precipitation measurements by ground stations. For MSG, the mountain ranges, like the Pyrenees, the Alps, and the Norwegian inland, turn deep red, indicating an overestimation by the MSG retrieval algorithm, interpreting the

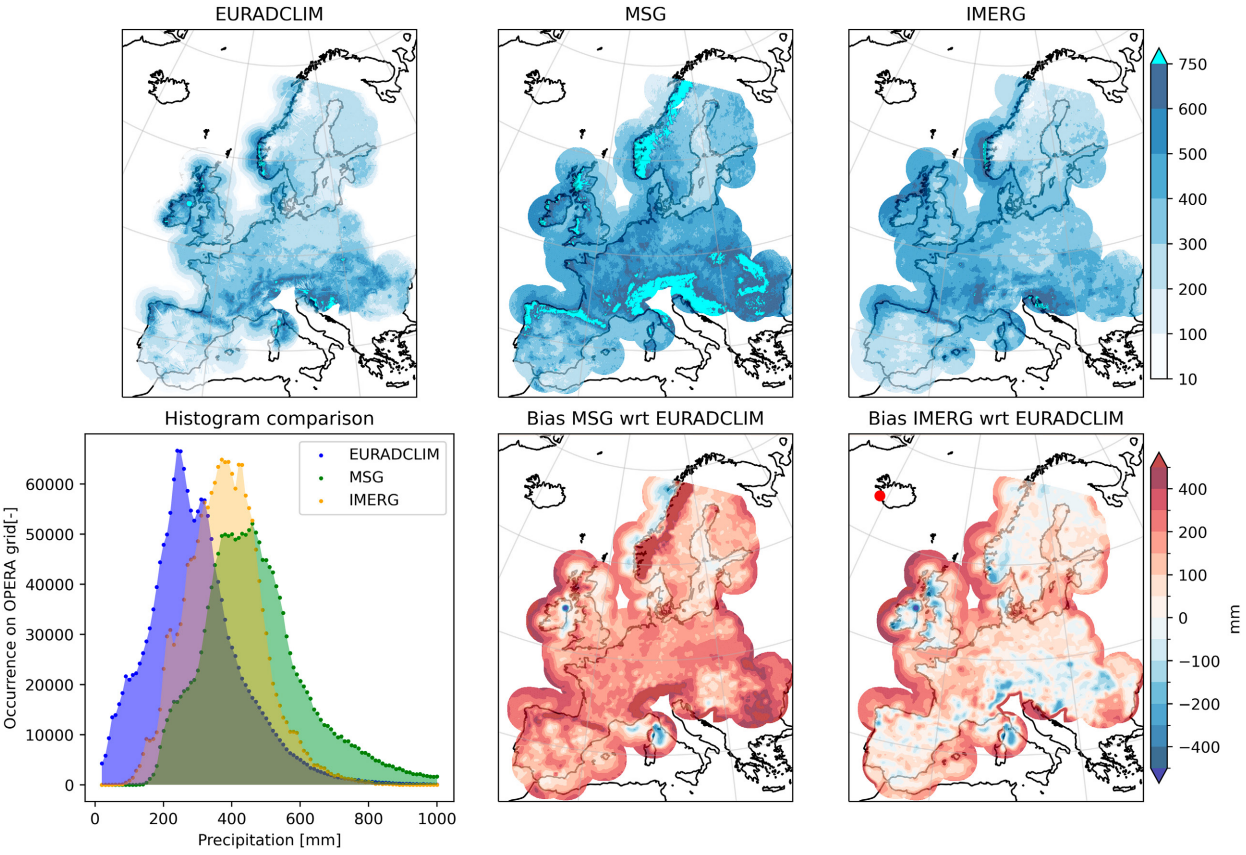

FIG. 4. (top) Average yearly accumulation over the period 2013–19, according to EURADCLIM, MSG, and IMERG for the simultaneous availability (a.o. daylight hours). (bottom left) The (total) histogram of the three sources. The bias of (bottom middle) MSG and (bottom right) IMERG with respect to EURADCLIM, using an averaging kernel of 31 × 31 pixels (nb = 15).

high reflectance of snow-capped mountains as persistent precipitation.

For the largest part of continental Europe, IMERG and MSG report higher yearly precipitation sums than the EURADCLIM radar climatology dataset. In Roebeling and Holleman (2009), a negative bias with respect to the not-adjusted Dutch weather radar of approximately 0.1 mm h^{-1} was reported. Here, the yearly accumulation over Europe suggests a positive bias of approximately 200–300 mm, which would translate to a bias of 0.02–0.03 mm h^{-1}.

The histogram of the counts in the yearly total sum is shown. This shows the overrepresentation of low values for EURADCLIM data (in blue). When we look at a similar histogram but only over continental France, this peak of lower values seems to be gone, and we assume that for a large part the overrepresentation of low values can be attributed to the fact that the radar values over sea are typically too low, as mentioned above.

Furthermore, the MSG distribution shows a slightly fatter tail for the high values than the IMERG data. The yearly sums in the Alps locally add up to 5000 mm yr^{-1}, so well beyond this histogram, but most of the values between 500 and 1000 mm yr^{-1} are also in the mountainous regions (and some in central France). The highest values in the IMERG yearly sums are more likely related to actual precipitation events that are overestimated, and not spurious reflections or unphysical outliers (see also section 4f).

Apart from looking at the total precipitation sum per area per interval, it is interesting to see if the distribution of the perceived precipitation can be compared. If we look at the exceedance of a threshold on a map, we would see that radar images are often contaminated with anomalous echoes, while the MSG images suffer from the snow cover during winter. In Fig. 5, we show several high percentile values for the hourly rainfall depths for each month in 2016. We see that the three sources agree quite well for median rainfall intensity, but they start to diverge for the extremes. In the distributions of the total monthly sum (left panels), we see that in Spain and France there are some locations in January–March with very high monthly accumulations, up to 400 mm, according to MSG, but not according to EURADCLIM and IMERG, so these are probably the pixels with snow in the Picos de Europa, Pyrenees, and Alps, that look like intense precipitation events to MSG. The distribution of hourly sums, in the right panels, shows events with high intensity, but not extremely high, which suggests that the high monthly sum is the consequence of seeing the moderately high values of snowed caps every day, every hour, thus amounting to such a large accumulation.

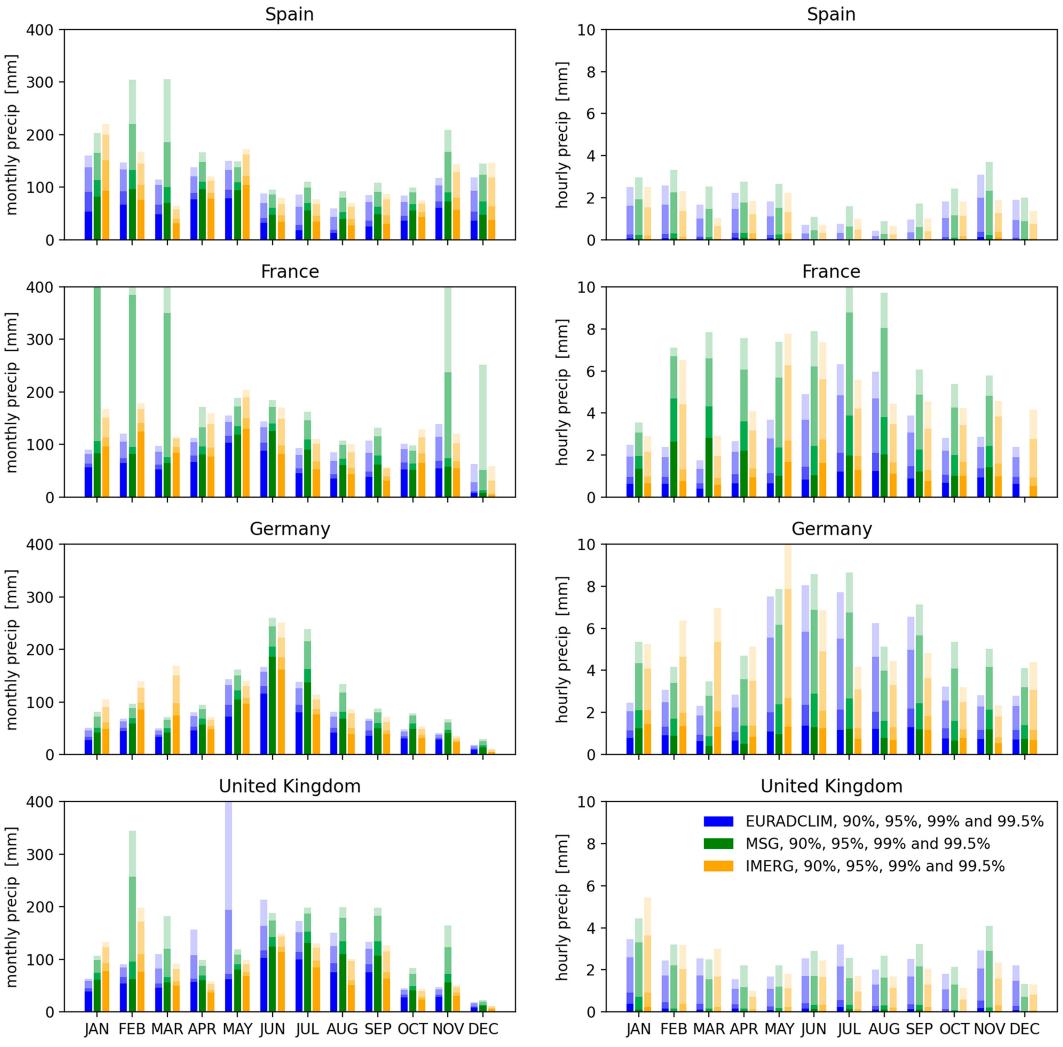

FIG. 5. Distribution of the (left) monthly and (right) hourly rainfall accumulations over (from top to bottom) Spain, France, Germany, and the United Kingdom per month of the year 2016. The opaquest bar indicates the 90th percentile, and the 95th, 99th, and 99.5th percentiles are gradually more transparent.

In the United Kingdom (lower panels), we see something similar in February, with a high sum for MSG, but also a high sum for EURADCLIM in May due to uncalibrated radar artifacts accumulating. Both outliers are not visible in the hourly sum distribution on the right, suggesting that they are also the result of a steady buildup of values that are not extreme by themselves.

In Roebeling and Holleman (2009), a notion was made of the fact that in the MSG retrieval algorithm the uncertainty in rain rate for very high-intensity precipitating systems is very large. So, the rain rate of opaque clouds may be overestimated, leading to higher intensities in the convective season (right panels, June–September).

Furthermore, IMERG detects significantly higher values over Germany in March 2016, but significantly lower extremes in the second half of the year. This might be the result of unjustified extrapolation of intense precipitation events in central Europe. This is in line with the findings of, e.g., Bogerd et al. (2020).

d. *Seasonal variation*

When we look at monthly accumulations of the three datasets, in Fig. 6 for the months January, April, July, and October 2016, we can see how some of the differences mentioned in section 4c are already visible. In January, the snow cover on the Alps is visible in the MSG product, as well as the snow in Norway in April. Furthermore, we see some radar location artifacts, e.g., in April 2016 as a spotty precipitation pattern over continental Spain.

More interestingly, while for most of these figures a qualitative agreement is observed, both MSG and IMERG reveal a slightly higher amount of precipitation over north–west Europe than the EURADCLIM dataset. We can zoom into precipitation within the borders of, e.g., continental France, as in Fig. 7.

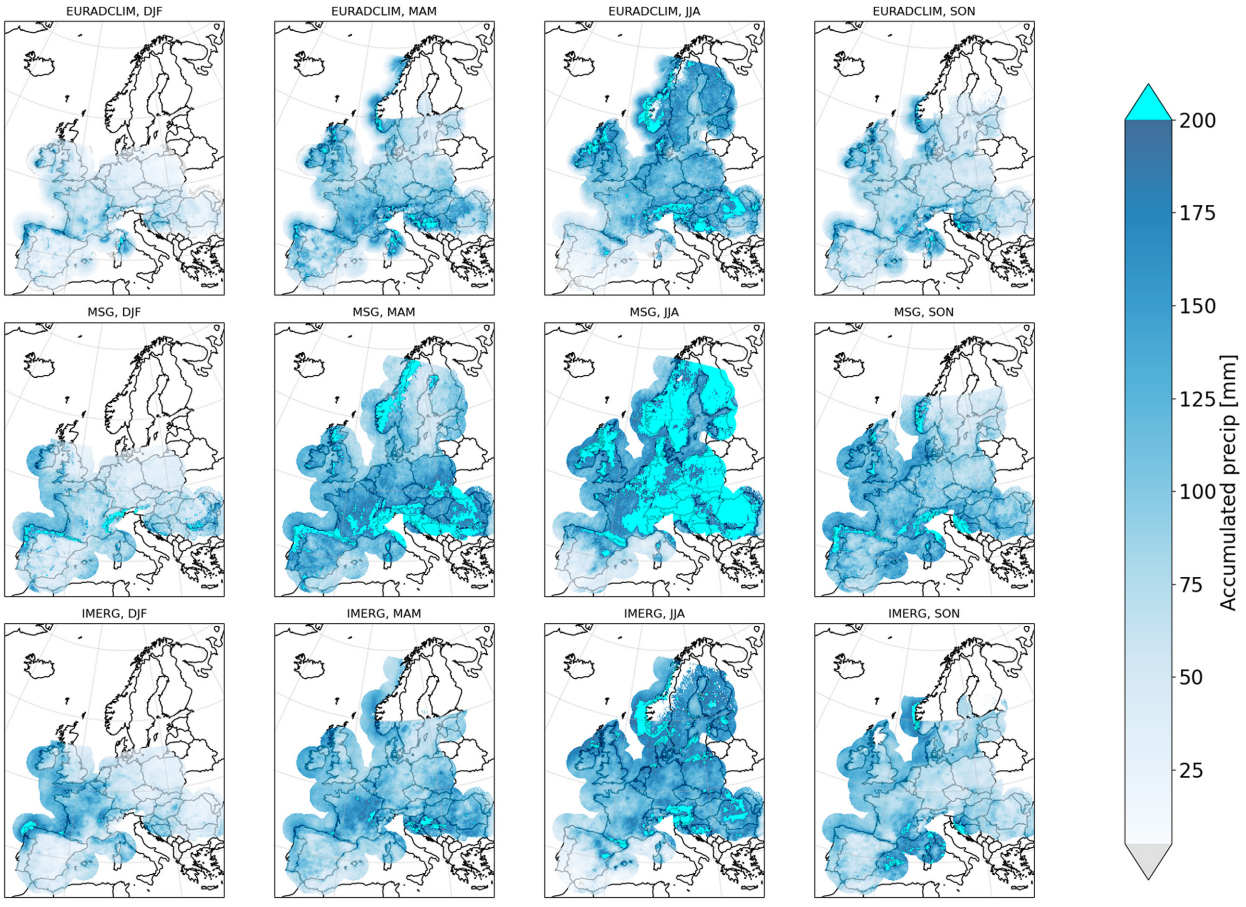

FIG. 6. (top) Seasonal EURADCLIM, (middle) MSG, and (bottom) IMERG precipitation accumulation for (from left to right) December–February (DJF), March–May (MAM), June–August (JJA), and September–November (SON), averaged over the years 2013–19.

We can see that in France, in April 2016, the satellite-based datasets MSG and IMERG agree on fairly heavy precipitation in the east and south of France, but the IMERG dataset also reports large amounts of rain in the Bordeaux coastal area. As we will revisit later, this may be due to the erroneous interpolation of a heavy storm. The histogram shows that if we cut the precipitation off at the coast the surplus incidence of low precipitation events of EURADCLIM disappears and that IMERG displays both lower and higher precipitation events: underestimates (in the north) and overestimates (in the Alps and Bordeaux areas).

e. Variation of precipitation volume over the day

One of the main differences between the MSG precipitation retrieval and the ground radar and IMERG constellation is that the MSG-based retrieval infers precipitation from cloud properties, in particular the cloud liquid/ice water path. It is well known that the retrieval of cloud water path from passive visible/near-infrared imagery tends to have a positive bias at high solar zenith angles, i.e., at the beginning and end of the day (Greenwald 2009). This will likely translate into an overestimation of precipitation intensity.

If we look at the time series around a specific location, we can plot the total precipitation sum as a function of the hour of day (UTC), as shown in Fig. 8 for a particular grid cell in central Germany. The right panel shows the cumulative precipitation for various spatial averages around this specific pixel, so we can see if biases are due to a lack of representativity or locally consistent. To avoid some of the representativity issues, in which a single pixel could be an outlier because of an anomalous echo or some other retrieval artifact, the cumulative precipitation of the average over neighborhood sizes increasing from 1, 3, 5, 11, and 15 to 25 pixels is also given. The average for neighborhood size n is then taken over a square area defined by a central point (i_0, j_0) with lower-left corner $(i_0 - n, j_0 - n)$ and upper-right corner $(i_0 + n, j_0 + n)$.

In this case, IMERG and MSG with neighborhood radius 0 have a bias of almost a factor 2 on a yearly basis. The IMERG bias becomes smaller when we look at the average over a large area around the location (largest 51×51 pixels, approximately 100 km $\times$ 100 km squared). We can see that the IMERG bias can largely be attributed to a small number of events, whereas MSG shows a steadier increase with respect to the EURADCLIM line.

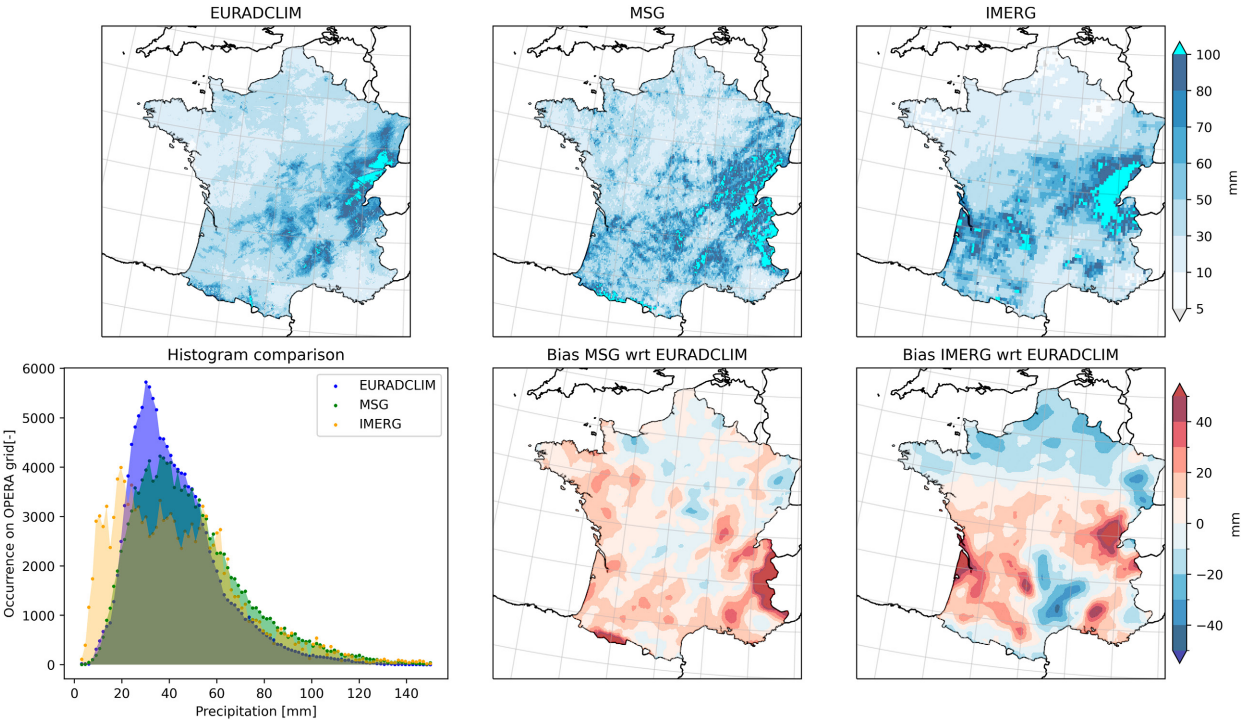

FIG. 7. Monthly accumulated data for EURADCLIM, MSG, and IMERG clipped to France for April 2016, as in Fig. 4.

We can see that for this particular location in central Germany that, if we look at the average over a small area around the grid cell (middle panel, 7 × 7 pixels), there seems to be a tendency for higher precipitation amounts at the end of the afternoon. However, this effect largely dissipates if we look at larger spatial averages (e.g., as in the right panel with an average over approximately 100 km × 100 km). Here, we do see relatively high MSG precipitation amounts at the beginning and end of the daylight interval, although no firm conclusions can be drawn from this limited analysis. This phenomenon deserves more careful analysis, because clouds will appear to have higher opacity when illuminated from a low angle. This low-angle illumination does not only occur at dusk and dawn but also occur in winter more than in summer and at different times for different longitudes. If we would perform statistics over a larger area, mapping along the day of these conditions, i.e., solar zenith angle per grid cell per hour, should filter all precipitation events to compare similar events. Here, it is indicated that an effect of illumination angle may already be present in the climatological analysis. Note furthermore that here we chose a relatively small and relatively large kernel to stress that information of sufficient resolution is in the data, if not averaged away when using a large kernel to make a comparison for climatological purposes only.

f. *Case study: Flooding event in May 2016*

In this section, we will show how the different precipitation measurement systems capture a series of storms over central

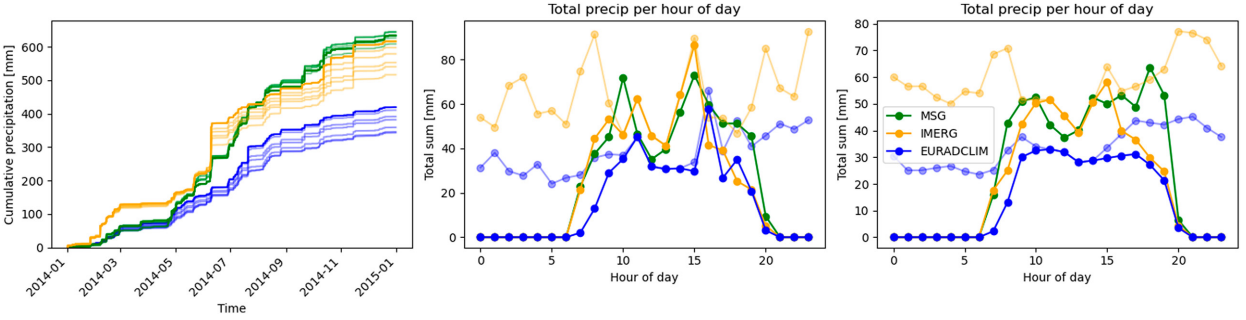

FIG. 8. (left) We show the cumulative precipitation sum in 2014 for one particular grid cell in central Germany for the three observation systems, with the thick line for the grid cell itself and semitransparent lines for neighborhoods of 1, 3, 5, 11, 15, and 25 pixels. (middle), (right) The total sum for an average around this pixel over 7 times 7 pixels and for an average over 51 times 51 pixels around this pixel (neighborhood 25), as a function of time of day (UTC). Here, the transparent lines show the values for EURADCLIM and IMERG if we do not account for the MSG (daylight) availability.

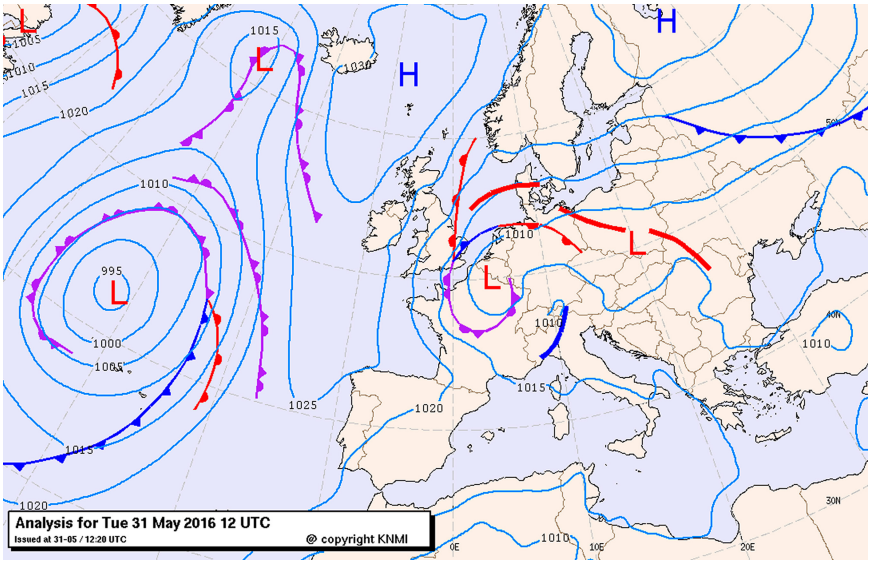

FIG. 9. Analysis of the synoptic meteorological situation over Europe for 1200 UTC 31 May 2016.

Europe in May 2016. A large active low pressure system over the Alps produced a number of storms over southern Germany, the Benelux countries, and France (see Fig. 9). High pressure above Scandinavia kept the system more or less in place for an extended period of time. In this period, rainfall of locally more than 100 mm fell in the catchment areas of the Seine and the Inn River.

In Fig. 10, an accumulation of the total precipitation between 29 May and 3 June 2016 is shown. For this small interval in time, already considerable differences have built up between the different precipitation measurement systems. Especially, MSG and EURADCLIM report large sums over this interval, with IMERG here not showing the same amounts, especially in France.

To study how the biases are built up over time, we can look at the cumulative sum of the precipitation for one pixel. In Fig. 10, two red dots indicate places where significant flooding was reported. The cumulative sum during the whole year 2016 and the period between 10 May and 5 June 2016 for these two pixels are shown as the opaque lines in Figs. 11 and 12. For the pixel in France, we see that IMERG indicates a significantly higher yearly sum, and EURADCLIM and MSG are closer together. Zooming in on the specific period, we can see that IMERG shows two precipitation events that shift the cumulative sum significantly, before the period depicted in Fig. 10. For completeness also, the measurement by the nearest rain gauge station is shown, as well as the interpolation of EURADCLIM at the specific location of the station. This "bursty" behavior that leads to excursions in the cumulative distribution suggests that high measured precipitation intensities may be erroneously interpolated, leading to periods of time with very high cumulative precipitation. The duration of a high-intensity event may thus be overestimated. This would not show up as a

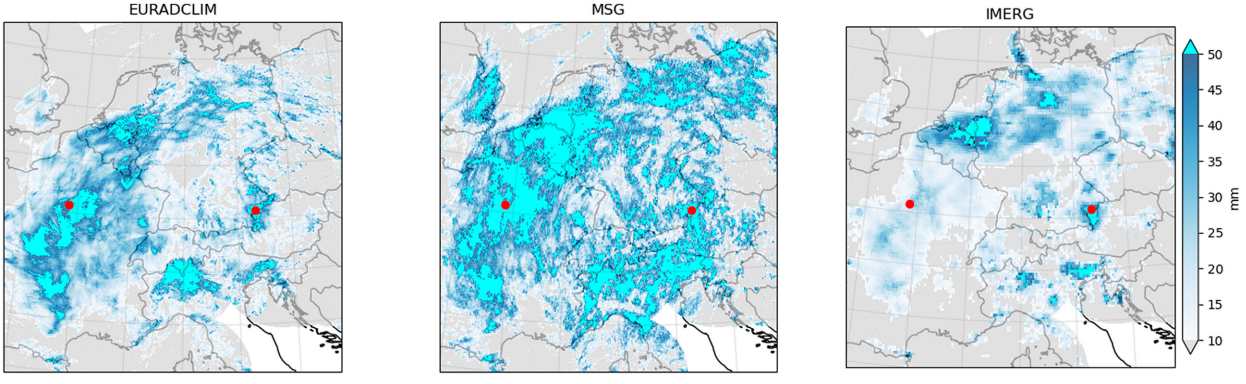

FIG. 10. Accumulation of precipitation for EURADCLIM, MSG, and IMERG, respectively, over the period between 29 May and 4 Jun 2016. Here, the figure was zoomed to the area containing southern Germany and northern France, the areas that were affected most by the flooding. The red dots indicate places of which the cumulative precipitation is shown in Figs. 11 and 12.

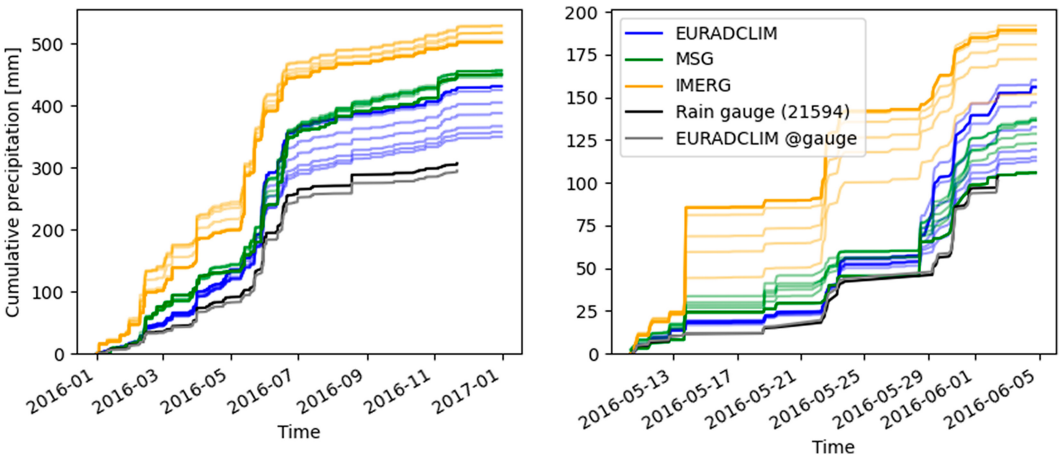

FIG. 11. Cumulative precipitation data for EURADCLIM, MSG, and IMERG for a pixel in central France for (left) the whole year 2016 and (right) 10 May–5 Jun 2016. The cumulative precipitation for averages of an area around 1, 3, 5, 11, 15, and 25 pixels is also shown as slightly transparent lines. The precipitation measured by the nearest rain gauge (black) and the interpolated EURADCLIM precipitation over that exact location (gray).

consistent bias, but as a different distribution, with outliers not as much in the intensity of the precipitation, but in the product of intensity and duration.

In the same way as in Fig. 8, we see in the figure that for this particular grid cell in France the EURADCLIM and IMERG data are high with respect to its neighbors, and the MSG data are relatively low. The averages are closer together than the single-pixel lines would suggest. For the German grid cell, the lines are closer to each other, indicating more agreement between the data sources. That being said, both IMERG and to a lesser extent MSG show excursions from the other data due to events with large precipitation sums in short times. The EURADCLIM radar product is expected to be most consistent in its retrieval, if the nonmeteorological echoes are well dealt with. IMERG may suffer from interpolation issues causing long perceived duration (or absence) of high-intensity events.

5. Conclusions

Precipitation is a quantity that is highly localized in both the spatial and temporal dimensions, which make climatological data hard to properly analyze. Looking at yearly accumulation sums may not tell the story of how well precipitation is captured on a daily or hourly basis. Taking these numbers at face value leads to figures such as Fig. 2, suggesting hardly any correlation at all.

The new EURADCLIM dataset (Overeem et al. 2022) makes it possible to do various kinds of analysis in both the spatial and temporal domains. Though it is not perfect, it can be regarded as a gridded hourly dataset that combines the reliability of the rain gauges with the high-resolution spatiotemporal information provided by the radar. The comparison to the satellite products did confirm that especially far away

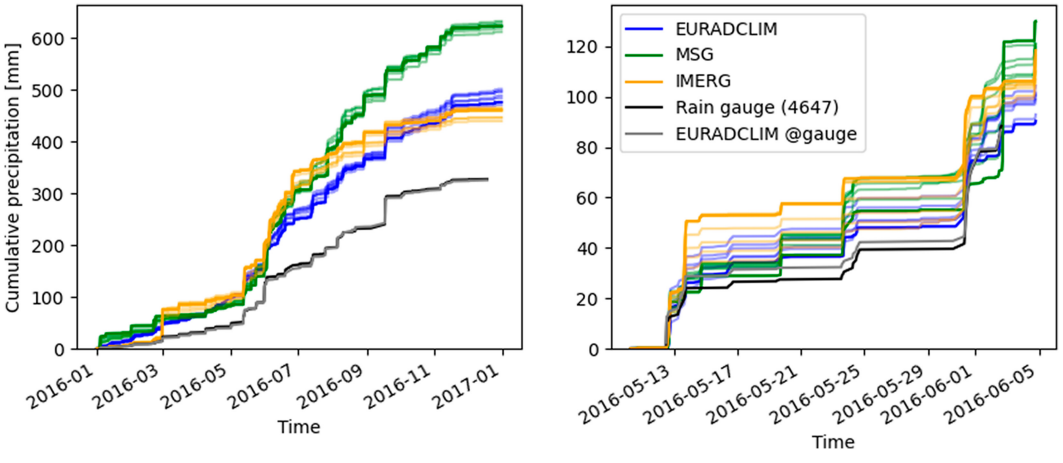

FIG. 12. Cumulative precipitation data for EURADCLIM, MSG, and IMERG for a pixel in southern Germany for (left) the whole year 2016 and (right) 10 May–5 Jun 2016. The cumulative precipitation for averages of an area around 1, 3, 5, 11, 15, and 25 pixels is also shown as slightly transparent lines.

from the rain gauges, like over sea, EURADCLIM precipitation shows a persistent underestimation (Fig. 4).

Both precipitation retrievals from the cloud physical property data from the SEVIRI instrument on board of the MSG as the IMERG system retrievals display the same structures of precipitation, especially on longer time scales. The retrieval of precipitation amounts by looking at the reflectance of clouds, as done by the MSG-CPP product, leads to artifacts caused by snow, especially in mountainous areas. The IMERG dataset looks better in some respects, on the longer time scales, but on shorter time scales the overestimation of precipitation events may occur due to (temporal) interpolation between overpasses. It is noted here that for the EURADCLIM data the MSG CLAAS-2 cloud-type mask was used to remove nonmeteorological echoes. This could introduce a dependence between the two datasets.

In Bogerd et al. (2020), an extensive analysis was presented in how IMERG overestimates light rain events yet underestimates heavy precipitation over the Netherlands. In this paper, we have focused more on climatology and how heavy precipitation may be responsible for local biases. On maps of accumulated precipitation over longer time scales, we see a general agreement, IMERG and MSG both with a positive bias.

We have seen that biases of satellite products that build up over time can be very localized. Apart from the reflections of the snow leading to a steady buildup of apparent precipitation in MSG, we have seen a relatively large bias during summer over central Europe (less so over the Iberian Peninsula). We have seen an indication of another source of potential biases in the overestimation of precipitation in low-sun circumstances, though the analysis in this article is at its best incomplete. In time series or cumulative diagrams, IMERG especially shows isolated events with very high precipitation that neither MSG nor EURADCLIM see.

The fact that in the case study of a period with some very heavy precipitation that caused local flooding in 2016 two locations on roughly the same latitude can behave quite differently with respect to the perceived precipitation intensity is interesting in itself: Apparently MSG generally has a steady bias that is visible in maps of accumulated rainfall and collocation issues that make it necessary to perform a smoothing over 20 km $\times$ 20 km regions (or maybe a smaller radius for the longitudinal component), but IMERG seems to overestimate the duration of high-intensity events that are more localized according to EURADCLIM, in both space and time, leading to biases resulting from those extended "bursts."

It is noted that both the MSG-CPP and particularly the IMERG precipitation products are under constant development: The new IMERG version 7 will most likely be better balanced than version 6 used in this evaluation. Also, new products based on data from a new generation of geostationary satellites will give new insights and improvements.

Further work is needed to explore differences among these datasets. In particular, further exploration is needed of extreme precipitation, MSG parallax and collocation, and short-lived convective precipitation.

Acknowledgments. All figures containing maps were made with Python package Cartopy (Met Office 2010).

Data availability statement. The EURADCLIM data are publicly available at https://dataplatform.knmi.nl/dataset/rad-opera-hourly-rainfall-accumulation-euradclim-1-0. MSG-CPP data are available via msg.cpp.knmi.nl, and IMERG data can be freely downloaded through the NASA portal at https://disc.gsfc.nasa.gov/.

REFERENCES

Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M. Ziese, 2013: A description of the global land-surface precipitation data products of the global precipitation climatology centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst. Sci. Data*, **5**, 71–99, https://doi.org/10.5194/essd-5-71-2013.

Benas, N., S. Finkensieper, M. Stengel, G.-J. van Zadelhoff, T. Hanschmann, R. Hollmann, and J. F. Meirink, 2017: The MSG-SEVIRI-based cloud property data record CLAAS-2. *Earth Syst. Sci. Data*, **9**, 415–434, https://doi.org/10.5194/essd-9-415-2017.

Bogerd, L., A. Overeem, H. Leijnse, and R. Uijlenhoet, 2020: Evaluation of the GPM IMERG-late precipitation product and its constituents over the Netherlands. *2020 Fall Meeting*, Online, Amer. Geophys. Union, Abstract H217-06.

Cornes, R. C., G. van der Schrier, E. J. M. van den Besselaar, and P. D. Jones, 2018: An ensemble version of the E-OBS temperature and precipitation data sets. *J. Geophys. Res. Atmos.*, **123**, 9391–9409, https://doi.org/10.1029/2017JD028200.

Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.*, **239**, 179–202, https://doi.org/10.1016/S0022-1694(00)00343-7.

Fabry, F., 2015: *Radar Meteorology: Principles and Practice*. Cambridge University Press, 276 pp.

Gaona, M. F. R., A. Overeem, H. Leijnse, and R. Uijlenhoet, 2016: First-year evaluation of GPM rainfall over the Netherlands: IMERG day 1 final run (V03D). *J. Hydrometeor.*, **17**, 2799–2814, https://doi.org/10.1175/JHM-D-16-0087.1.

Greenwald, T. J., 2009: A 2 year comparison of AMSR-E and MODIS cloud liquid water path observations. *Geophys. Res. Lett.*, **36**, L20805, https://doi.org/10.1029/2009GL040394.

Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurement mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722, https://doi.org/10.1175/BAMS-D-13-00164.1.

Huffman, G. J., 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.

Klein Tank, A. M. G., and Coauthors, 2002: Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. *Int. J. Climatol.*, **22**, 1441–1453, https://doi.org/10.1002/joc.773.

Klok, E. J., and A. M. G. Klein Tank, 2009: Updated and extended European dataset of daily climate observations. *Int. J. Climatol.*, **29**, 1182–1191, https://doi.org/10.1002/joc.1779.

Met Office, 2010: Cartopy: A Cartographic Python Library with a Matplotlib interface. https://scitools.org.uk/cartopy.

Navarro, A., E. García-Ortega, A. Merino, J. L. Sánchez, C. Kummerow, and F. J. Tapiador, 2019: Assessment of IMERG precipitation estimates over Europe. *Remote Sens.*, **11**, 2470, https://doi.org/10.3390/rs11212470.

O, S., U. Foelsche, G. Kirchengast, J. Fuchsberger, J. Tan, and W. A. Petersen, 2017: Evaluation of GPM IMERG early, late, and final rainfall estimates using WegenerNet gauge data in southeastern Austria. *Hydrol. Earth Syst. Sci.*, **21**, 6559–6572, https://doi.org/10.5194/hess-21-6559-2017.

Overeem, A., E. van den Besselaar, G. van der Schrier, J. Meirink, E. van der Plas, and H. Leijnse, 2022: EURADCLIM: The European climatological gauge-adjusted radar precipitation dataset (1-h accumulations). KNMI DataPlatform, accessed 11 June 2022, https://doi.org/10.21944/7ypj-wn68.

——, ——, ——, ——, ——, and ——, 2023: EURADCLIM: The European climatological high-resolution gauge-adjusted radar precipitation dataset. *Earth Syst. Sci. Data*, **15**, 1441–1464, https://doi.org/10.5194/essd-15-1441-2023.

Park, S., M. Berenguer, and D. Sempere-Torres, 2019: Long-term analysis of gauge-adjusted radar rainfall accumulations at European scale. *J. Hydrol.*, **573**, 768–777, https://doi.org/10.1016/j.jhydrol.2019.03.093.

Rauber, R. M., and S. W. Nesbitt, 2018: *Radar Meteorology: A First Course*. John Wiley and Sons, 496 pp., https://doi.org/10.1002/9781118432662.

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.

Roebeling, R. A., and I. Holleman, 2009: SEVIRI rainfall retrieval and validation using weather radar observations. *J. Geophys. Res.*, **114**, D21202, https://doi.org/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. Hydrometeor.*, **13**, 1552–1566, https://doi.org/10.1175/JHM-D-11-089.1.

Saltikoff, E., and Coauthors, 2019: OPERA the radar project. *Atmosphere*, **10**, 320, https://doi.org/10.3390/atmos10060320.

Skofronick-Jackson, G., D. Kirschbaum, W. Petersen, G. Huffman, C. Kidd, E. Stocker, and R. Kakar, 2018: The Global Precipitation Measurement (GPM) mission's scientific achievements and societal contributions: Reviewing four years of advanced rain and snow observations. *Quart. J. Roy. Meteor. Soc.*, **144**, 27–48, https://doi.org/10.1002/qj.3313.