Received: 1 June 2023 | Revised: 2 January 2024 | Accepted: 3 January 2024

DOI: 10.1002/joc.8369

RESEARCH ARTICLE

International Journal of Climatology RMetS

Uncertainty and outliers in high-resolution gridded precipitation products over eastern North America

Tangui Picart[1,2] | **Alejandro Di Luca**[1,2] | **René Laprise**[1,2]

[1]Department of Earth and Atmospheric Sciences, University of Quebec in Montreal, Montreal, Quebec, Canada

[2]Centre pour l'étude et la simulation du climat à l'échelle régionale (ESCER), University of Quebec in Montreal, Montreal, Quebec, Canada

Correspondence
Tangui Picart, Department of Earth and Atmospheric Sciences, University of Quebec in Montreal, 201, avenue du Président-Kennedy, Local PK-6150 Montreal, QC H2X 3Y7, Canada.
Email: tangui.picart@laposte.net

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RGPIN-2020-05631

Abstract

Several observational precipitation products that provide high temporal (≤3 h) and spatial (≤0.25°) resolution gridded estimates are available, although no single product can be assumed worldwide to be closest to the (unknown) "reality." Here, we propose and apply a methodology to quantify the uncertainty of a set of precipitation products and to identify, at individual grid points, the products that are likely wrong (i.e., outliers). The methodology is applied over eastern North America for the 2015–2019 period for eight high-resolution gridded precipitation products: CMORPH, ERA5, GSMaP, IMERG, MSWEP, PERSIANN, STAGE IV and TMPA. Four difference metrics are used to quantify discrepancies in different aspects of the precipitation time series, such as the total accumulation, two characteristics of the intensity-frequency distribution, and the timing of precipitating events. Large regional and seasonal variations in the observational uncertainty are found across the ensemble. The observational uncertainty is higher in Canada than in the United States, reflecting large differences in the density of precipitation gauge measurements. In northern midlatitudes, the uncertainty is highest in winter, demonstrating the difficulties of satellite retrieval algorithms in identifying precipitation in snow-covered areas. In southern midlatitudes, the uncertainty is highest in summer, probably due to the more discontinuous nature of precipitation. While the best product cannot be identified due to the lack of an absolute reference, our study is able to identify products that are likely wrong and that should be excluded depending on the specific application.

KEYWORDS

North America, observational uncertainty, radar, reanalysis, satellite

1 | INTRODUCTION

Several applications require accurate historical precipitation values: for example, calculating the Earth water cycle budget, estimating trends in hydrometeorological extremes, verifying weather forecasts and evaluating weather and climate simulations (Beck et al., 2017b; Hossain & Huffman, 2008; Nissen & Ulbrich, 2017; Trenberth, 2011). Precipitation gauge measurements usually provide accurate precipitation estimates at the local

scale, except for solid precipitation due to instrumental issues related to wind-induced snowfall undercatch when a shield is not used (Nitu et al., 2018; Rasmussen et al., 2012). However, for specific applications, including model evaluation, station data suffer from spatial representativity issues (Chen & Knutson, 2008; Gervais et al., 2014; Herold et al., 2016; Prein & Gobiet, 2017). For this reason, gridded precipitation datasets have been developed merging in situ and remotely sensed observations with optimal interpolation or kriging, and—in the case of reanalyses—merging observations and background model with data assimilation techniques (e.g., Harris et al., 2020; Meyer-Christoffer et al., 2011). Gridded precipitation products are generally assumed to provide accurate precipitation estimates in regions with high density of stations, which is unfortunately not the case for the whole globe.

In the past few decades, great progress has been made to improve estimates of precipitation at high temporal (subdaily time scales) and spatial (usually finer than 0.25°) resolutions by combining gauge measurements with meteorological radar and satellite remote sensing data (Beck et al., 2019a, 2017b; Derin & Yilmaz, 2014; Sun et al., 2018; Trenberth et al., 2017), or by assimilating observations in reanalysis products (e.g., Hersbach et al., 2019). Their refined temporal and spatial resolution and their nearly worldwide coverage make these products highly valuable for numerous applications, including the evaluation of precipitation simulated by weather and climate models that require high-resolution areal-mean estimations (e.g., Di Luca et al., 2021; Fosser et al., 2015; Huffman et al., 2015; Prein & Gobiet, 2017). However, the quality of these products depends on several factors, including the type of satellite sensor (e.g., infrared or microwave) used to estimate the precipitation (Beck et al., 2017b, 2019a; Derin & Yilmaz, 2014; Liu & Allan, 2012; Sun et al., 2018), the underlying coverage of the surface (e.g., errors can be larger in snow-covered areas; Derin & Yilmaz, 2014; Tian & Peters-Lidard, 2010), the meteorological phenomena producing precipitation (e.g., satellite-based products tend to overestimate the intensity of convection; Beck et al., 2017b; Derin & Yilmaz, 2014; Gehne et al., 2016; Sun et al., 2018), or the phase and type of hydrometeors present in the atmosphere (e.g., larger uncertainty for snowfall than for rainfall; Lockhoff et al., 2019; Prein & Gobiet, 2017; Tian & Peters-Lidard, 2010). Corrections of these products based on gauged measurements tend to reduce their differences (Gehne et al., 2016; Prein & Gobiet, 2017), but these corrections depend strongly on the availability and quality of the gauge data used (Sun et al., 2018). For example, in regions with high station density, gauge-corrected products are usually unbiased regarding monthly or even daily time scales, depending on the type of correction being applied. However, most in situ observations are only available at a daily frequency, limiting the development of corrections at subdaily time scales (Zolina et al., 2014).

Several studies have evaluated differences among precipitation products using time-mean values (e.g., Gehne et al., 2016; Prein & Gobiet, 2017; Sun et al., 2018). Sun et al. (2018) compared the global distribution of time-mean precipitation between 2003 and 2010 across 30 products obtained using satellite, in situ and reanalysis data. Their findings show that most products agree well on the overall spatial patterns, but they show large discrepancies across products over several land regions (including North America) where mean precipitation estimates vary from 600 mm·year^{-1} to almost 900 mm·year^{-1}. A similar global analysis was carried out by Gehne et al. (2016) for monthly-mean precipitation between 2001 and 2012 using 13 precipitation products (eight observation-based products and five reanalyses). They found large differences near the intertropical convergence zone, over land areas in the summer hemisphere (notably over western North America), and in the Atlantic over the western boundary of the Gulf Stream. They also demonstrated that bias-corrected versions of the products are more consistent for annual, monthly and daily precipitation values when compared to their uncorrected counterparts.

Agreement in the time-mean precipitation does not ensure, however, a consistent representation of the daily or subdaily temporal variability, nor of the synchronicity of precipitation events (Catto et al., 2015; Di Luca et al., 2021; Gehne et al., 2016; Sun et al., 2018). Over North America, Gehne et al. (2016) showed that PERSIANN, GPCP and TMPA have similar annual and monthly-mean precipitation values but different intensity distributions of daily precipitation values, showing that precipitation occurs differently across the three datasets. To identify differences in precipitation products arising from the variability or the chronology of events, several metrics have been used. Aghakouchak et al. (2012) assessed the mean error (bias) and the random error (variability and timing error) between the satellite and radar-based precipitation products CMORPH, TMPA and STAGE IV over continental United States. For most of the domain, differences between products for daily and 3-hourly data were largely explained by the random error with small contribution from the bias. Guilloteau et al. (2022) compared IMERG and a product based on gauge and radar precipitation measurements (the Multi-Radar/Multi-Sensor System, MRMS) over the southeast United States between January 2018 and April 2020 using a similar decomposition of the error. They showed that the representation of the fine-scale (spatial and temporal) precipitation is the main source of disparities

between both products. The temporal and spatial consistency of extreme daily precipitation was investigated by Timmermans et al. (2019) over the continental United States for five gauge-based and three satellite-based products (including TMPA and CMORPH). Products based on precipitation gauges shared strong similarities in their representation of extreme precipitation, while the consistency between satellite-based products and gauge-based ones depends strongly on the region: high in the southeast and low in the west and over mountainous regions.

Studies comparing precipitation products generally aim to assess their performance or consistency, two different but often related concepts. Performance quantifies the difference between the estimate and a reference dataset that is considered to provide accurate precipitation estimates and is usually based on precipitation gauges or radar measurements. Consistency quantifies the level of agreement between a set of products and is estimated based on comparisons between them and does not require a reference dataset. High consistency is obtained when products are similar, implying that the observational uncertainty is low. High consistency is often assumed to imply good performance, but this implication can be unsound when the products being evaluated are strongly dependent on each other (e.g., due to similar input data), leading to erroneous conclusions. This paradigm can be illustrated using results from the study of Gehne et al. (2016). They show that TMPA and GPCP monthly-mean precipitation values are similar over North America, suggesting low observational uncertainty. However, GPCP and TMPA are both based on similar gauge observations. Hence, in this case, the low observational uncertainty is associated with a strong dependence among products coming from the use of a similar input dataset and will not necessarily imply good performance. The discussion of the relationship between performance and dependence has been extensively addressed when evaluating large ensembles of climate models (Knutti, 2010; Knutti et al., 2017; Nearing et al., 2016). For satellite-based products, the dependence can come from using the same input data (e.g., the microwave and infrared data assimilated in CMORPH, IMERG, TMPA and GSMaP use common input sources such as the data from the TRMM microwave imager), similar data and methods used to correct the satellite estimates (e.g., both TMPA and IMERG use the GPCC v6 Monitoring gridded product) or common approaches to convert radiances (or reflectivities) to precipitation estimates (e.g., CMORPH and GSMaP both use a similar technique to estimate changes in precipitation rates from infrared satellite data; Bytheway et al., 2020).

Today, the coexistence of multiple precipitation products and the lack of well-established metrics to understand the discrepancies and quantify their uncertainties (i.e., their level of consistency) makes it difficult to determine whether a product is suitable for a study. The aim of this study is to develop and apply a methodology to quantify the uncertainty of a set of eight 3-hourly gridded precipitation products over eastern North America. The uncertainty is quantified using four metrics that evaluate different aspects of the precipitation gridded products in a hierarchical way: a metric quantifying errors in mean precipitation, two metrics incorporating information about 3-hourly precipitation variability, and a fourth metric additionally considers the timing of precipitation events. The quantification of the uncertainty associated with this ensemble of gridded observation datasets includes a method for identifying the least likely products (i.e., outliers) at individual grid points and seasons. The study also includes an analysis of the sensitivity of our results to the temporal (3-hourly, 6-hourly and daily) and spatial (0.25° and 0.75°) scales of the data.

The article is organized as follows. Section 2 describes the datasets used, including the types of corrections used by each product. The methodology is presented in section 3. In section 4, the uncertainty used is introduced (section 4.1) and its metrics (section 4.2), regional (section 4.3) and seasonal (section 4.4) variabilities are investigated. The sensitivity to the choice of the resolution (temporal and spatial) of the data is presented in section 4.5. Some discussion about the uncertainty metric and the relative performance/dependence of the products and concluding remarks are presented in section 5.

2 | DATA

The study focuses on eastern North America, bounded by latitudes 25°N and 60°N, and longitudes 100°W and 60°W (Figure 1). The study separates the analysis according to six regions: northeast Canada (NCA), southeast Canada and the Great Lakes (CAGL), northeast United States (NEUS), southeast United States (SEUS), Gulf of Mexico (GM) and Atlantic Ocean (AO). The comparison between precipitation products is carried over a 5-year period from 2015 to 2019, which is the longest period common to all products (Table 1).

The precipitation products used are based on multiple instruments (satellite based, surface radar and rain gauges) and even modelled data (e.g., reanalysis products). These products were selected because they are widely utilized not only in North America but also globally. Table 1 provides an overview of the eight precipitation products used in the study, detailing their complete names and versions, spatiotemporal resolution, data coverage and input data sources. Additional information about the correction methods employed by each product

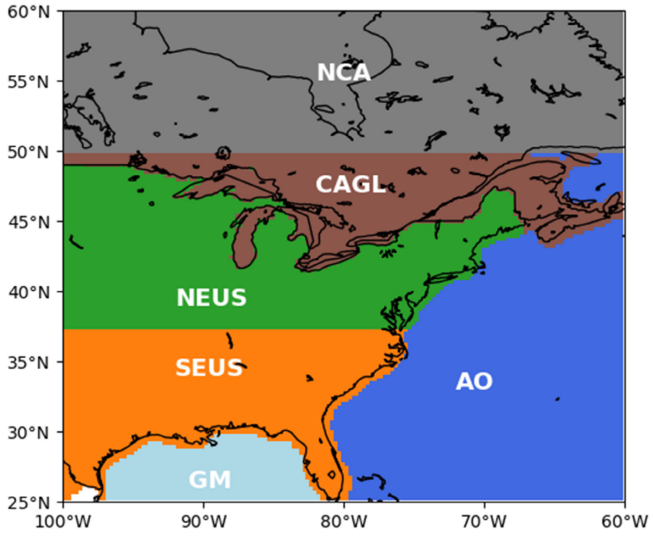

FIGURE 1 Domain of the study including the six regions: northeast Canada (NCA, grey), southeast Canada and the Great Lakes (CAGL, brown), northeast United States (NEUS, green), southeast United States (SEUS, orange), Gulf of Mexico (GM, pale blue) and Atlantic Ocean (AO, blue). [Colour figure can be viewed at wileyonlinelibrary.com]

is given in Table 2. Furthermore, a comparison of the satellite sensors utilized by each product is provided in Table S2, Supporting Information.

The Climate Prediction Center (CPC) morphing technique (CMORPH) precipitation version 1 estimate incorporates microwave (MW) and infrared (IR) observations from multiple instruments (Table S2) and is corrected using gauge observations. The instantaneous precipitation is estimated purely from MW observations on polar-orbiting platforms, and between two satellite overpasses, changes in the precipitation field are estimated by taking into account the cloud motions derived from IR observations (the morphing process; Joyce et al., 2004). Finally, the estimate is debiased using a gauge-based analysis over land and using a product merging satellite and gauge observations over ocean (see Table 2 for details).

The Global Satellite Mapping of Precipitation (GSMaP) version 7 estimate integrates MW and IR observations (Table S2) and is corrected using gauge observations following an approach like CMORPH. The instantaneous precipitation is estimated from MW observations. However, in GSMaP, the interpolation method between two polar-orbiting satellite overpasses considers both cloud motion and changes in cloud top height retrieved from IR observations. Then, the estimate is debiased over land using a gauge-based gridded product, no correction is applied over ocean (Table 2).

The Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)–Cloud Classification System (CCS)–Climate Data Record (CDR) is a corrected version of the PERSIANN–CCS product. First, in the PERSIANN–CCS product, the IR dataset (Table S2) is partitioned using temperature thresholds to identify and separate individual cloud patches. Second, precipitation at each location is estimated based on an empirical relationship between cloud classification and the local brightness temperature. The empirical relationship was obtained from radar observations over the United States (Hong et al., 2004; Hsu et al., 1997). Then, the estimate is debiased over land and ocean using monthly-mean precipitation from a product merging satellite and gauge observations (Table 2).

Tropical Rainfall Measuring Mission (TRMM) multi-satellite precipitation analysis (TMPA) product 3B42 Version 7 incorporates MW and IR observations from multiple sources (Table S2) and is corrected using gauge observations. The instantaneous precipitation is estimated for each MW sensor using the Goddard profiling algorithm (GPROF). For the study period, the version GPROF2010 is used (Huffman & Bolvin, 2018). These MW estimates are climatologically calibrated to match TRMM Combined Radar-Radiometer Algorithm and merged. When MW data is not available, precipitation is exclusively estimated based on IR observations. However, this method exhibits inferior performance compared to the use of MW observations (Huffman et al., 2007). Finally, the estimate is corrected over land using a monthly product based on gauge observations, no correction is applied over ocean (Table S2).

The Integrated Multi-satellitE Retrievals for Global precipitation measurement (IMERG) version 06 integrates IR and MW observations from a higher number of sensors than TMPA (Table S2) and is corrected using gauge observations. MW observations for each sensor are calibrated to match GPM Combined Radar-Radiometer Algorithm and are subsequently converted to precipitation estimates using an updated version of the Goddard profiling algorithm, GPROF2014v2. All MW estimates are calibrated with the Combined Radar-Radiometer product (CORRA) and with the Global Precipitation Climatology Project (GPCP) monthly Satellite-Gauge estimates. Simultaneously, an IR estimate is computed using the PERSIANN-CCS approach and calibrated with MW estimates. MW precipitation estimates are then merged using the CMORPH interpolation method. When the gap between the two MW observations is longer than 90 min, the IR precipitation estimate is considered. The merged estimate is finally corrected over land using the same reference gauge dataset as TMPA (Table 2).

STAGE IV dataset is based on surface radar measurements from National Weather Service River Forecast Centers (RFCs) over continental United States (Lin & Mitchell, 2005; Nelson et al., 2016). RFCs use a network

TABLE 1 Abbreviation, name and version, reference, native spatial (Δx) and temporal (Δt) resolution, available period and spatial coverage, and input data sources of the precipitation datasets used in the study.

Abbreviation	Full name (version)	References	Δx (°)	Δt (h)	Available period	Spatial coverage	Data source
CMORPH	Climate Prediction Center morphing technique (V1.0)	Joyce et al. (2004)	0.25	1	1998/01–present	Quasi-global (60°N–60°S)	Satellite/gauge
GSMaP	Global Satellite Mapping of Precipitation (V7)	Ushio et al. (2009)	0.1	1	2014/03–present	Quasi-global (60°N–60°N)	Satellite/gauge
PERSIANN	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (CCS-CDR)	Hsu et al. (1997, 1999) and Sadeghi et al. (2021)	0.04	3	1983/01–present	Quasi-global (60°N–60°S)	Satellite/gauge
TMPA	Tropical Rainfall Measuring Mission multi-satellite precipitation analysis (3B42)	Huffman et al. (2007)	0.25	3	1998/01–2019/12	Quasi-global (50°N–50°S)	Satellite/gauge
IMERG	Integrated Multi-satellitE Retrievals for Global precipitation measurement (precipitationCal, V6)	Huffman et al. (2018, 2020)	0.1	0.5	2000/06–present	Global	Satellite/gauge
STAGE IV	NCEP/EMC US Gridded Radar-Estimated Precipitation with Bias Removal (stage IV)	Lin and Mitchell (2005)	~0.05	1	2002/01–present	CONUS	Surface radar/gauge
ERA5	European Centre for Medium-range Weather Forecasts Reanalysis 5 High-Resolution	Hersbach et al. (2020)	0.25	1	2008/01–present	Global	Model with data assimilation (satellite/gauge/surface radar)
MSWEP	Multi-Source Weighted-Ensemble Precipitation (V2.8)	Beck et al. (2019a, 2019b) and GloH20 (2021)	0.1	3	1998/01–2019/12	Global	IMERG/ERA5/gauge

of Weather Surveillance Radar-1988 Doppler radars. RFCs in our domain use the MRMS estimator based on radar and gauge observations. Manual quality control of the RFC estimates is also performed (Nelson et al., 2016). Finally, RFC precipitation estimates are combined, and bias-adjusted in near real-time based on automated gauge measurements (Prat & Nelson, 2015).

Precipitation estimates from the European Centre for Medium-Range Weather Forecasts Reanalysis version 5 (ERA5; Hersbach et al., 2020) are based on a state-of-the-art numerical weather prediction model assimilating observational data from multiple sources. ERA5 has a finer temporal and spatial resolution than the previous ERA-Interim reanalysis (Dee et al., 2011). ERA5 has benefited from notable improvements, such as the assimilation of humidity-sensitive satellite channels using the new all-sky assimilation approach and the use of new parameterizations of physical processes (large-scale clouds, microphysics). It is important to note that ERA5 assimilates most of the satellite radiance measurements

TABLE 2 Details of corrections applied to satellite-based products: dataset used as reference, correction reference, spatial (Δx) and temporal (Δt) resolutions of the data used as reference or at which the correction is performed.

	Correction dataset	Correction reference	Δx	Δt	Note
CMORPH (over land)	CPC Unified Gauge-Based Analysis	Xie et al. (2017)	0.5°	Daily	Correction of daily precipitation and debiased of the monthly precipitation
CMORPH (over ocean)	GPCP 5-day precipitation		7.5°	95 days	Debiased of the mean precipitation
GSMaP (over land)	CPC Unified Gauge-Based Analysis	GPM Global Rainfall Map Algorithm Development Team (2014) and Mega et al. (2019)	0.5°	Daily	Correction of hourly precipitation and debiased of the daily precipitation
	Radar rain-gauge network of Japan (2015)		0.1°	Hourly	
PERSIANN	GPCP monthly analysis v2.3	Sadeghi et al. (2021)	2.5°	Monthly	Filtering of nonzero values generated by the Neural Network model. Debiased of the monthly precipitation
TMPA (over land)	GPCC v6 Monitoring analysis	Huffman et al. (2010)	1°	Monthly	Debiased of the monthly precipitation
IMERG (over land)	GPCC v6 Monitoring analysis	Huffman et al. (2018)	1°	Monthly	Debiased of the monthly precipitation
MSWEP	Daily gauge data (Beck et al., 2019b)	Beck et al. (2019b) and GloH20 (2021)	0.1	Daily	Debiased of the monthly precipitation when gauge observations are available

(IR and MW) used by the other products and radar reflectivity measurements over continental United States (used in STAGE-IV). ERA5 does not assimilate gauge-based products.

Multi-Source Weighted-Ensemble Precipitation (MSWEP) version 2.8 combines precipitation estimates from IMERG (0.1°/1 h) and ERA5 (0.25°/1 h) with daily gauge observations from various sources (Beck et al., 2019b, 2017a; GloH20, 2021). Prior to merging, ERA5 estimate is corrected to address its wet day bias. The merging process incorporates local climate conditions, orography, and air temperature to account for regional variations. Additionally, the intensity distribution of the final estimate is corrected to reduce spurious drizzle and artificial peaks resulting from the merging of different products. The estimates are then corrected over land based on daily gauge measurements (Table 2).

3 | METHODS

3.1 | Precipitation data preprocessing

The eight datasets provide precipitation estimates at different temporal and horizontal resolutions. To facilitate the comparison, all datasets were spatially interpolated to the coarser grid mesh (0.25° lat/lon grid) and upscaled at the lowest temporal resolution (3-hourly interval). TMPA has not been interpolated as it shares the same grid as ERA5. Spatial interpolation was performed using the "conservative" algorithm from the xESMF python library (Zhuang et al., 2020), which ensures the precipitation conservation on the coarser ERA5 grid boxes, as has been done in previous studies (e.g., Di Luca et al., 2021; Prein & Gobiet, 2017). The extent of the reduction in resolution (and the associated smoothing effect) varies between the different observations, with the finest resolution being 0.04° and the coarsest (except for 0.25°) being 0.1°.

Specifically, STAGE IV was first interpolated using a nearest-neighbour approach from its native 4-km polar stereographic projection to a 0.01° lat/lon grid before being conservatively interpolated to the ERA5 grid. This product is only considered for the northeast United States (NEUS) and the southeast United States (SEUS) regions.

MSWEP and GSMaP grid (available at 0.1°) were first shifted by +0.025° in latitude and longitude to ensure that the number of input cells (2.5 cells) used to produce an ERA5 output cell (0.25°) remains constant as this substantially affects the frequency of precipitating events (see Figure S1). This operation was not required for the IMERG grid as its grid was already aligned with the ERA5 (two ERA5 grid cells contain exactly 5 IMERG

grid cells). Due to file compression issues, ERA5 precipitation estimates can contain some small negative values ($\geq -10^{-8}$ mm $(3\,\mathrm{h})^{-1}$); these values were set to zero before interpolation.

For all datasets, 3-hourly precipitation values lower than a threshold of 0.125 mm $(3\,\mathrm{h})^{-1}$ are assumed to be zero to assure a similar treatment of low precipitation values. At this threshold value, the most substantial portion of total precipitation that goes unaccounted for is 2.7% from the ERA5 product (refer to Table S1 for specific details). The threshold used here is consistent with previous studies using an intensity-frequency decomposition of the mean precipitation (Catto et al., 2015; Di Luca et al., 2021).

3.2 | Hierarchical precipitation difference metrics

Multiple metrics are used to quantify the precipitation difference between two products and are all calculated at individual grid points.

The first metric, denoted as ΔP_M, is the absolute value of the difference between the time-mean precipitation of two given products (i.e., the absolute bias). At a grid point, the mean precipitation, P_M, is given by

$$P_M = \frac{1}{N_T}\sum_t P_t, \tag{1}$$

where the index t stands for the 3-hourly time interval, N_T the total number of intervals and P_t the 3-hourly precipitation time series. Figure S2 shows the mean precipitation maps of each product. The metric ΔP_M between two given datasets, noted by superscripts (1) and (2) hereinafter, is thus given by

$$\Delta P_M = \frac{1}{N_T}\left|\sum_t \left(P_t^{(1)} - P_t^{(2)}\right)\right| = \left|P_M^{(1)} - P_M^{(2)}\right|. \tag{2}$$

Note that given eight datasets, a total of 28 distinct pairs (and thus comparisons) are possible, $8!/(2!\times 6!)$.

The second and third metrics quantify differences in the precipitation variability. The second metric quantifies differences in the intensity distribution. For a given choice of intensity bins, the total precipitation P_{B_j} within the bin $B_j = (I_j, I_{j+1}]$ is given by

$$P_{B_j} = \sum_{I_j < P_t \leq I_{j+1}} P_t. \tag{3}$$

A novel precipitation metric ΔP_B can be defined as

$$\Delta P_B = \frac{1}{N_T}\sum_{B_j}\left|P_{B_j}^{(1)} - P_{B_j}^{(2)}\right|. \tag{4}$$

The ΔP_B metric is useful to identify differences between products that arise from differences in the intensity distribution. As such, it provides a stricter estimation of differences between two products compared with ΔP_M (that allows for all types of error compensations) verifies that $\Delta P_B \geq \Delta P_M$.

The choice of bins B_j must be made with care. It must span the range of all possible intensities. Also, the bin number must be chosen to ensure a sufficiently low noise (which increases with the bin number) and a precise representation of the intensity distributions. In our study, a logarithmic distribution with 130 bins was used when analysing the entire period, with the lowest intensity being 0.125 mm $(3\,\mathrm{h})^{-1}$ and the highest being 10^3 mm $(3\,\mathrm{h})^{-1}$. When calculations are made at the monthly and seasonal basis (hence with fewer data), we used 20 bins ranging from 0.125 mm $(3\,\mathrm{h})^{-1}$ to 10^3 mm $(3\,\mathrm{h})^{-1}$. Figure S3 shows the dependence of ΔP_B values on the number of bins.

A third metric is based on the commonly used intensity-frequency precipitation decomposition (e.g., Catto et al., 2015; Di Luca et al., 2021). At a grid point, the mean precipitation is given by

$$P_M = \frac{\sum_t P_t}{N_T} = \frac{\sum_t P_t}{N_{P_t>0}} \cdot \frac{N_{P_t>0}}{N_T} = I.F, \tag{5}$$

with $N_{P_t>0}$ the number of 3-hourly periods with precipitation, I the mean intensity, which is the precipitation averaged over all 3-hourly periods with precipitation, and F the precipitation frequency, which is the fraction of 3-hourly periods with precipitation. The mean precipitation used here is the same as in Equation (1). Figure S4 shows the frequency maps of each product.

The intensity-frequency metric ΔP_{IF} is defined as

$$\Delta P_{IF} = i.|\Delta F| + f.|\Delta I| + R, \tag{6}$$

with

$$i = \min\left(I^{(1)}, I^{(2)}\right), f = \min\left(F^{(1)}, F^{(2)}\right), \tag{7}$$

$$\Delta F = F^{(1)} - F^{(2)}, \Delta I = I^{(1)} - I^{(2)}, \tag{8}$$

$$R = \left|\left(I^{(1)} - i\right).\left(F^{(1)} - f\right) - \left(I^{(2)} - i\right).\left(F^{(2)} - f\right)\right|. \tag{9}$$

The terms $i.|\Delta F|$ and $f.|\Delta I|$ measure differences in mean precipitation amounts of two products due to the

difference in the frequency and the intensity of precipitation, respectively. The metric ΔP_{IF} is constructed to highlight compensations between the intensity and frequency terms within ΔP_M. In case of a compensation between the frequency and the intensity, the differences $\left(I^{(1)}-I^{(2)}\right)$ and $\left(F^{(1)}-F^{(2)}\right)$ have opposite signs (i.e., the product with the highest frequency also has the lowest intensity) and the residual term $R=0$ adding to $\Delta P_{IF}=i.|\Delta F|+f.|\Delta I|$. Otherwise, if no compensation occurs, R>0 and $\Delta P_{IF}=\Delta P_M$. By construction $\Delta P_{IF}\geq\Delta P_M$.

The fourth metric, denoted as ΔP_A, is the time-average of absolute differences between 3-hourly precipitation values (i.e., the mean absolute difference),

$$\Delta P_A=\frac{1}{N_T}\sum_t\left|P_t^{(1)}-P_t^{(2)}\right|. \tag{10}$$

This metric is null only if the two precipitation time series are identical, thus preventing any form of error compensation. ΔP_A will be used as a reference in our analysis. In contrast to ΔP_M, ΔP_B and ΔP_{IF}, this metric is most meaningful when the timing of meteorological events coincides between the two datasets (e.g., it cannot be used to evaluate differences between free running climate model simulations). By construction, $\Delta P_A\geq\Delta P_B\geq\Delta P_M$ but ΔP_A is not necessarily greater in values than ΔP_{IF}.

To illustrate differences between the four metrics, Figure 2a shows TMPA and MSWEP precipitation time series for the grid point closest to Montreal, Canada, for 2015. At this grid point and for this period, the comparison between TMPA and MSWEP has a value of 0.01 mm $(3\,\mathrm{h})^{-1}$ for ΔP_M and of 0.47 mm $(3\,\mathrm{h})^{-1}$ for ΔP_A. Figure 2b shows the intensity distributions of TMPA and MSWEP for 2015. In this case, the metric ΔP_B has a value of 0.35 mm $(3\,\mathrm{h})^{-1}$ and the metric ΔP_{IF} has a value of 0.51 mm $(3\,\mathrm{h})^{-1}$. This confirms that the various metrics allow for different levels of compensation among their discrepancies with a large part of discrepancies cancelling out when using ΔP_M and no cancellation at all for the ΔP_A metric.

All metrics have units of precipitation rate, mm $(3\,\mathrm{h})^{-1}$. To better reflect local differences, the relative difference is calculated by normalizing a given metric, ΔP, by the sum of mean precipitation in both products:

$$\Delta P(\%)=100\times\frac{\Delta P\left(\mathrm{mm}\,(3h)^{-1}\right)}{P_M^{(1)}+P_M^{(2)}}. \tag{11}$$

Defined in this way, relative metrics vary between 0% and 100% and show the fraction of the mean precipitation that is not consistent between two products. For example, using the 3-hourly precipitation time series for the year 2015 shown in Figure 2a, the relative differences between MSWEP and TMPA are $\Delta P_M=1.7\%$, $\Delta P_B=50\%$, $\Delta P_{IF}=73\%$ and $\Delta P_A=68\%$.

4 | RESULTS

4.1 | Constructing an uncertainty metric

In this section, we derive a metric to quantify the uncertainty of a set of products associated with the representation of precipitation at an individual grid point. The process of deriving the uncertainty metric is illustrated using the precipitation time series from three locations for specific seasons.

Figure 3a,b shows boxplots of differences between each product with the other seven products for ΔP_M and ΔP_A in winter for a grid point located in central Maine. ΔP_M values vary from 0.3% for the IMERG-STAGE IV

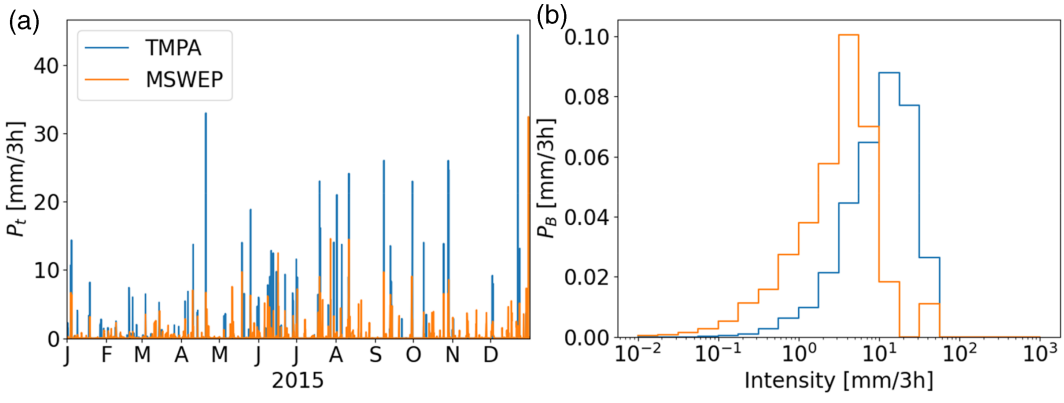

FIGURE 2 TMPA and MSWEP 3-hourly precipitation rates (P_t, panel a) and total precipitation accumulated over different bins (P_B; panel b). Both figures are based on data from the year 2015 and for a grid point located near Montreal, Canada (45.5°N, 73.5°W). The intensity distribution is represented using 20 logarithmic bins. [Colour figure can be viewed at wileyonlinelibrary.com]

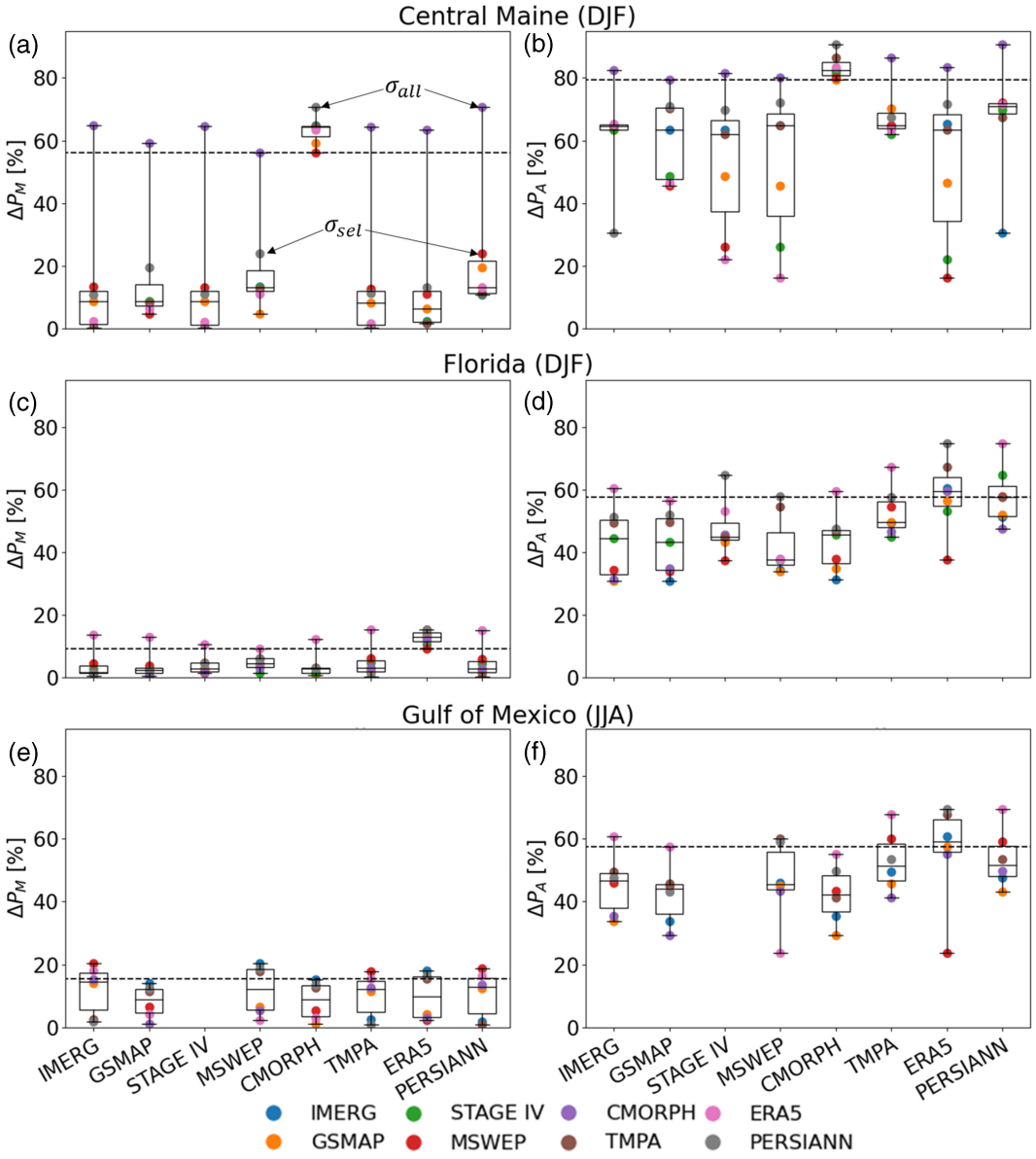

FIGURE 3 Difference values calculated for the absolute bias (ΔP_M, left panels) and the mean absolute difference (ΔP_A, right panels) using 3-hourly precipitation time series from three grid points: central Maine (46.0°N, 68.5°W) in winter (a, b), Florida (30.7°N, 85.5°W) in winter (c, d) and Gulf of Mexico (27.0°N, 90.0°W) in summer (e, f). The dashed line indicates the $(nb_{obs}-1)^{th}$ largest difference value. The uncertainty metrics, σ_{all} and σ_{sel}, are defined in section 3.2. [Colour figure can be viewed at wileyonlinelibrary.com]

pair to 72% for the CMORPH-PERSIANN pair and ΔP_A values vary from 16% for the MSWEP-ERA5 pair to 90% for the CMORPH-TMPA pair. CMORPH shows median values for ΔP_M and ΔP_A of 64% and 82%, respectively, that are much larger than for other products which have median values for ΔP_M between 6.5% (ERA5) and 13.2% (MSWEP) and for ΔP_A between 62% (STAGE IV) and 71% (PERSIANN). The much larger median value for CMORPH results from large errors in CMORPH precipitation associated with a general underestimation of precipitation in winter high latitudes, an issue that has been shown in several studies (e.g., Sun et al., 2018; Xie et al., 2017). The observational uncertainty in the central Maine location could be characterized using the largest difference among all products, in Figure 3a denoted as σ_{all}. However, σ_{all} is determined by the difference between CMORPH and PERSIANN and is thus contingent on the erroneous value of the CMORPH product, leading to an overestimation of the actual uncertainty.

The example above shows that the estimation of the observational uncertainty can be improved if we can remove outliers from the uncertainty calculation. To identify outliers, we assume that the majority of precipitation products are distributed around the true precipitation

and that products with large median differences, the outliers, are therefore wrong. In this study, outliers are identified as those products for which the median value is larger than the $(n_{obs}-1)$ largest individual metric value. That is, in the case where eight products are available, a product will be considered as an outlier only if its median value is larger than or equal to the seventh largest individual value (denoted with the dashed line in Figure 3a). According to this definition, a product will be identified as an outlier if it differs substantially from most products and no more than two products will be identified. Finally, the observational uncertainty is estimated as the largest value among the products that were not identified as outliers and is denoted as σ_{sel} (i.e., the observational uncertainty of selected products).

In winter for the central Maine location (Figure 3a,b), CMORPH median is larger than the seventh largest values for both ΔP_M and ΔP_A metrics, respectively, 65% and 82% and is correctly identified as an outlier. The observational uncertainty decreases from 72% (σ_{all}) to 24% (σ_{sel}) for ΔP_M and from 90% to 72% for ΔP_A.

The identification of outliers, as conducted here, can be affected by the interdependence between products. If multiple products are dependent, the assumption that the majority of products are distributed around the true value becomes problematic. In such instances, the identified outliers could actually be products that are closer to the truth. That is, in instances where there is a strong dependence among most products, our method might fail to properly identify outliers. In our case, a small ΔP value between two datasets suggests a high level of agreement and can be related with products being dependent on each other. For example, the low ΔP_A value between MSWEP, ERA5 and STAGE IV likely reflects that MSWEP was constructed using ERA5 precipitation estimates and that ERA5 assimilated radar measurements on which STAGE IV is based (see section 2). However, the medians for MSWEP, ERA5 and STAGE IV are similar to those of the other products, indicating that the median value is minimally influenced by product dependence. This assumption is only true if co-dependent products account for less than half of all products.

Figure 3c,d shows boxplots of ΔP_M and ΔP_A values during winter in a grid point in Florida. The ΔP_M ERA5 median of 14% is much larger than the median of any other product, all under 5% and seventh largest value, 9%. For ΔP_M, ERA5 is considered as an outlier. For ΔP_A, ERA5 and PERSIANN have relatively high median values, respectively, 60% and 58%, in comparison to other medians (under 50%). ERA5 and PERSIANN medians are greater or equal the seventh largest value, thus, they both products are considered as outliers. The observational uncertainty σ_{sel} decreases compared with σ_{all} from 15% to 7% for ΔP_M and from 75% to 55% for ΔP_A.

Figure 3e,f shows boxplots of ΔP_M and ΔP_A values during summer in a grid point in the Gulf of Mexico. For ΔP_M, products can be divided into two groups of products between which the comparisons have low values: on one side, IMERG, TMPA and PERSIANN and on the other side, MSWEP, GSMaP, ERA5 and CMORPH. Products all have medians under the seventh largest value, 15%, and none is excluded. For ΔP_A, ERA5 has the largest median, 60%, which is higher than the seventh largest value, 57%, and is considered as an outlier. The observational uncertainty remains unchanged for ΔP_M and decreases from 68% to 60% for ΔP_A.

4.2 | The products uncertainty and the difference metrics

Figure 4 shows the monthly median for the eight precipitation products and the four metrics averaged over the NEUS region. Figure 4a shows that, for the metric ΔP_M, the median of the CMORPH product is much higher than any other product from October to May, reaching a value of nearly 60% in February. Products other than CMORPH exhibit a median that varies from about 15% in winter to less than 10% in summer. Interestingly, the σ_{sel} values are also larger in winter than in summer, which shows that the larger observational uncertainty in mean precipitation in winter does not only result from including CMORPH but is a consistent feature among selected products.

Figure 4b,c shows that results for ΔP_B and ΔP_{IF} display strong similarities and values are systematically higher than for ΔP_M. The large disparity between CMORPH and other products in winter remains but, in addition, ΔP_B and ΔP_{IF} also reveal that TMPA estimates are substantially different compared to the other products year-round. While TMPA shows differences in mean precipitation of at most 15%, like other products, it shows differences of about 50% and 70% for ΔP_B and ΔP_{IF}, respectively. This suggests that the intensity and frequency distribution of 3-hourly precipitation of TMPA is at odds with other products. For both metrics, the median of GSMaP also appears to be slightly higher than other products in winter while ERA5 median tends to be slightly higher than others in summer.

Figure 4d shows that the ΔP_A metric captures some additional disparities between products and shows larger values than ΔP_M and ΔP_B. PERSIANN shows low values of ΔP_M, ΔP_B and ΔP_{IF}, but its median is as large as the one from TMPA for the ΔP_A metric showing its timing of precipitation is at odds with other products. ERA5 has

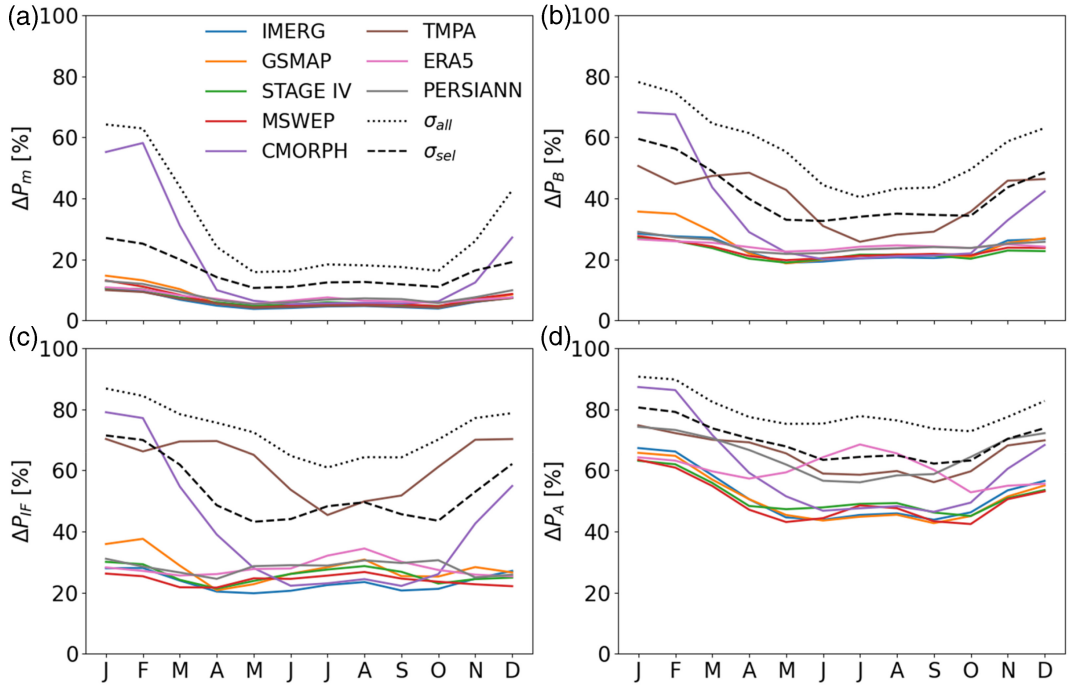

FIGURE 4 Monthly median for each product for each metric: the absolute bias (ΔP_M, a), the difference in the intensity distribution (ΔP_B, b), the difference in the used intensity-frequency precipitation decomposition (ΔP_{IF}, c) and the mean absolute difference (ΔP_A, d) averaged over grid points in the NEUS region. Monthly mean values for the σ_{all} and σ_{sel} uncertainty metrics are also included. [Colour figure can be viewed at wileyonlinelibrary.com]

the largest ΔP_A median values in summer months, with a maximum value of 67% in July. This last result might be related with some known issues of ERA5 at representing the timing of precipitation during the day (Watters et al., 2021). It is also consistent with the literature showing that the more the precipitation regime tends towards deep convection, the more accurate the satellite estimates and the less accurate the model-based estimates (e.g., Ebert et al., 2007). TMPA has a high ΔP_A value throughout the year, with a mean annual value of around 66%.

Figure 5 shows the mean precipitation and precipitation frequency for each product and for each season. As presented in Equation (5), the mean precipitation, P_M, can be expressed as the product of the precipitation frequency, F, and mean intensity, I (grey lines in Figure 5 show values of constant mean intensity). TMPA shows a mean precipitation like most other products and ΔP_M values for comparison with TMPA are small, so the same applies to its median. However, TMPA systematically generates fewer precipitation events with a higher intensity than other products, its mean intensity is near 5 mm $(3\,\text{h})^{-1}$ while other products have intensities below 3 mm $(3\,\text{h})^{-1}$. Comparisons with TMPA for ΔP_B and ΔP_{IF} are thus large in values, as its median. In both winter and spring, CMORPH tends to underestimate the mean precipitation largely due to an underestimation of the frequency of events.

Figure 6 shows the fraction of grid points where each product is considered as an outlier by our method over NEUS, for the four metrics. The sum of all fractions can exceed 100% because two products can be considered as outliers at some points. For the ΔP_M metric, Figure 6a shows that CMORPH is considered as an outlier for around 90% of the grid points in January and February. From April to October, PERSIANN and ERA5 have the largest monthly fractions with values between 20% and 35%.

Figure 6b,c shows that results based on ΔP_B and ΔP_{IF} share some strong similarities. In winter, CMORPH and TMPA are identified as outliers at 75% and 36% of grid points, respectively, for ΔP_B. TMPA is also excluded at most points for the rest of the year. These results are consistent with Figure 4. CMORPH has a higher median in winter, due to its underestimation of average precipitation, and TMPA throughout the year, due to its poor representation of precipitation variability. Interestingly, in summer, ERA5 and PERSIANN are considered outliers at 28% and 15% of grid points for ΔP_B and 15% and 3% for ΔP_{IF}, while their medians (Figure 4) are only slightly higher than the others. The spatial variability of the median within NEUS could explain this result.

Figure 6d shows that for the ΔP_A metric, only CMORPH, TMPA, ERA5 and PERSIANN are considered as outliers. CMORPH is excluded for winter and early spring months with a maximum of 70% in January, while

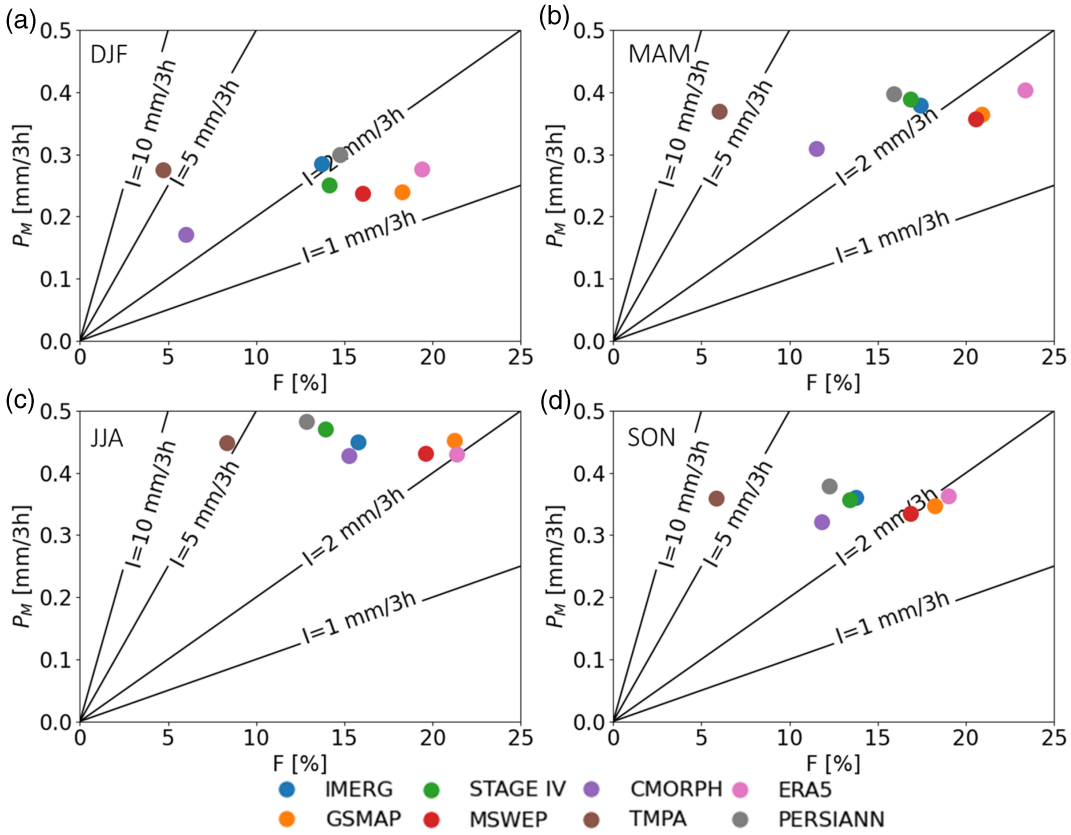

FIGURE 5 Mean precipitation, P_M, as a function of frequency of precipitation events, F, for each product for (a) DJF, (b) MAM, (c) JJA and (d) SON. Results are for an area comprising the two regions NEUS and SEUS. Lines with constant precipitation intensity (I) are shown in black (see Equation (6)). [Colour figure can be viewed at wileyonlinelibrary.com]

ERA5 is excluded mainly for summer months with a maximum of 80% in July. TMPA and PERSIANN are considered as outliers by at least 20% of the grid points all year long, TMPA fraction reaches a maximum of 66% in May and PERSIANN of 86% in October. These results are consistent with Figure 4.

4.3 | The regional variability of products uncertainty

Figure 7a shows the spatial distribution of σ_{sel} values for the ΔP_A metric over eastern North America. A strong spatial variability is visible across the domain, with high latitudes showing the largest uncertainty values. Over NCA, σ_{sel} has a regional mean value of 76%, reaching 95% locally. The uncertainty over US regions is lower than for Canadian regions (CAGL and NCA), with regional mean values of 62% and 66% for SEUS and NEUS, respectively. Further work is needed to quantify uncertainty sources in both regions, but differences are likely related with the lower density of stations (Kidd et al., 2017; Sun et al., 2018) over Canada that will affect precipitation corrections, or errors in satellite retrieval algorithms when representing precipitation in snow-covered areas (Henn et al., 2018; Trenberth et al., 2017). AO and GM have regional mean σ_{sel} values of about 62%, thus showing similar values to those over land.

Figure 7b–g shows the fraction of grid points where each product is identified as an outlier for individual regions on an annual basis. The products identified as outliers over the studied domain are predominantly PERSIANN (58% of domain grid points), ERA5 (34%) and TMPA (16%). Specifically, PERSIANN is excluded at 43%, 33% and 96% of grid points over NEUS, CAGL and NCA, respectively. Figure S5 shows that over SEUS, NEUS and AO, PERSIANN is excluded for along the east coast. ERA5 is identified as an outlier at 43%, 66%, 96% and 45% for NEUS, SEUS, GM and AO, respectively. Figure S5 shows that ERA5 is identified as an outlier at most grid points with latitudes lower than 35°N and points with longitudes between 100°W and 90°W over US land. The result may be explained by the poor representation of convective precipitation in ERA5 (Lavers et al., 2022). Finally, TMPA is excluded at 30% and 35% of grid points over NEUS and CAGL, mostly for grid points surrounding

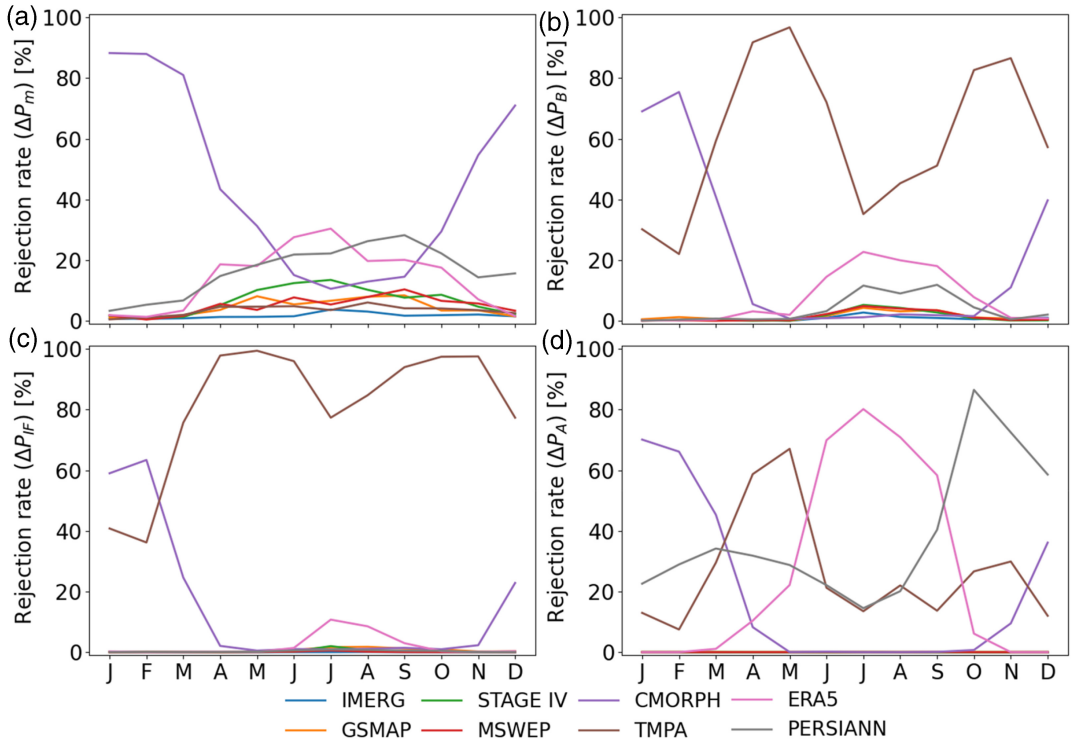

FIGURE 6 Fraction of grid points for which each product is identified as an outlier for each month of the year over the NEUS. Results are shown for the absolute bias (ΔP_M, a), the difference in the intensity distribution (ΔP_B, b), the difference in the used intensity-frequency precipitation decomposition (ΔP_{IF}, c) and the mean absolute difference (ΔP_A, d). [Colour figure can be viewed at wileyonlinelibrary.com]

the Great Lakes (Figure S5). GSMaP and CMORPH are also excluded for a small number of grid points (<13%) over CAGL and NCA.

Figure 8a shows the spatial distribution of $\sigma_{\text{sel}}(\Delta P_B)$. The uncertainty for ΔP_B is lower than for ΔP_A. The lowest values are found over SEUS and NEUS, with an averaged uncertainty of 25% and 28%, and largest ones are found over NCA with an average value of 44%. A discontinuity in $\sigma_{\text{sel}}(\Delta P_B)$ is generally observed between United States and Canada and between land and Great Lakes grid points, likely due to the difference in station density (Kidd et al., 2017; Sun et al., 2018), which strongly affects product's corrections. Northern regions (NCA and CAGL) and ocean regions (GM and AO) show similar regional mean uncertainty values, between 40% and 45%, but their spatial distributions differ widely. Over AO and GM, $\sigma_{\text{sel}}(\Delta P_B)$ values vary between 15% and 57% while over NCA and CAGL grid point values vary between 17% and 83%.

Figure 8b–g shows the fraction of grid points for which each product is excluded within $\sigma_{\text{sel}}(\Delta P_B)$. The fraction of points where PERSIANN is excluded decreases drastically in comparison to results using the ΔP_A metric, which is consistent with previous results (Figure 6). TMPA is identified as an outlier at most grid points over SEUS, NEUS and CAGL and at grid points over the northern part of AO. Over ocean, MSWEP, ERA5 and TMPA are identified as an outlier. MSWEP is excluded at most ocean grid points south of 30°N latitude (not shown). Over oceanic, ERA5 and MSWEP intensity distributions are very similar and both products strongly underestimate precipitation intensities compared with the other products (not shown). MSWEP, combining ERA5 and IMERG, exhibits a distribution more distant from other products than ERA5. This difference could be due to its specific merging technique, or the wet day bias correction applied to ERA5.

On the contrary, CMORPH and GSMaP are identified as outliers only at grid points in the northern part of the domain. GSMaP is excluded for ΔP_B and ΔP_A for areas where GSMaP strongly underestimates the mean precipitation (Figure S2). GSMaP is mainly excluded in regions with strong and unrealistic spatial discontinuities in its precipitation and mean frequency fields (Figures S2 and S4, respectively).

4.4 | The seasonal variability of products uncertainty

Figure 9a shows monthly mean values of $\sigma_{\text{sel}}(\Delta P_A)$ averaged over each individual region. For high-latitude

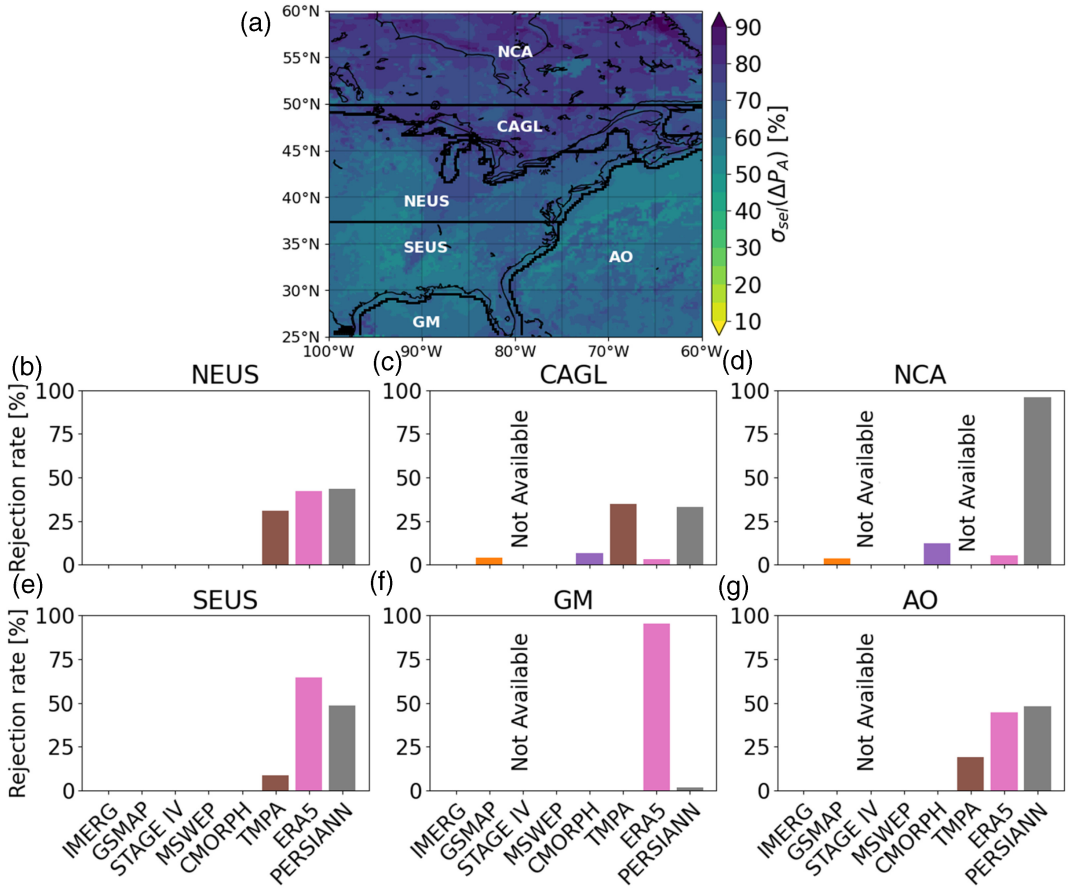

FIGURE 7 Spatial distribution of the σ_{sel} uncertainty metric (a), and the fraction of grid points where each product is identified as an outlier for individual regions (b–g). All metrics are computed using the mean absolute difference, ΔP_A. [Colour figure can be viewed at wileyonlinelibrary.com]

regions (NCA, CAGL and NEUS), the uncertainty is larger in winter than in summer. For example, NCA has a mean uncertainty of 90% in winter and around 70% in summer. This difference may be explained by the presence of snow on the surface and the larger uncertainty in precipitation gauge observations (e.g., wind-induced undercatch of snowfall). For low-latitude regions (SEUS, GM and AO), the uncertainty varies little throughout the year, with a small maximum in summer. The increase in uncertainty in summer is probably related with the predominance of convective precipitation that is associated with high spatial and temporal variability (Beck et al., 2017b; Prein & Gobiet, 2017) and with the increased presence of cirrus clouds that may be mistaken for raining clouds (Tian et al., 2009).

Figure 9b–e shows that the fraction of grid points where products are excluded varies strongly between summer (red bars) and winter (blue bars). For northern regions (Figure 9b–d), CMORPH is excluded for most grid points in winter but only a few in summer. On the contrary, the fraction of grid points where ERA5 is identified as an outlier is higher in summer (e.g., 95% in SEUS) than in winter (e.g., 6% in SEUS) for all regions. PERSIANN is excluded for most grid points when CMORPH and ERA5 are not considered as outliers. As shown by Figures 4 and 6 for NEUS, PERSIANN tends to be consistently distant from the other product year-round and for most grid points.

Figure 10a shows that, for ΔP_B, the uncertainty remains higher in winter than in summer for high latitudes (NEUS, CAGL and NCA). For instance, for NCA, the uncertainty reaches a maximum of 72% in January and a minimum of 41% in July. For ΔP_B, the largest uncertainties in GM occur in spring with a maximum of 57% in May. The large uncertainty is the spring is caused by large disparities in the estimation of the mean precipitation. Figure 10b–g shows the fraction of grid points where each product is identified as an outlier in each region. The main difference between results for ΔP_B and ΔP_A is that PERSIANN is rarely identified as an outlier for ΔP_B, except over NCA in summer where PERSIANN tends to overestimate precipitation intensities (not shown).

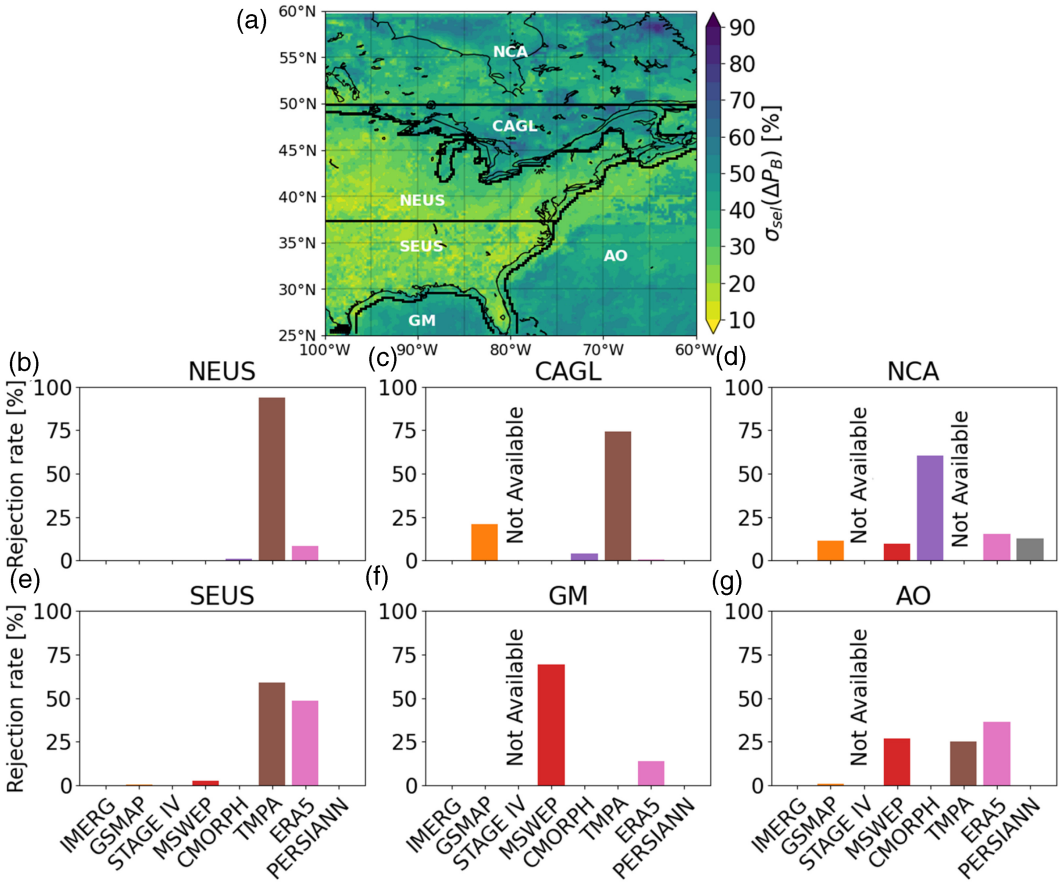

FIGURE 8 As in Figure 7 but using the difference in the intensity distribution, ΔP_B. [Colour figure can be viewed at wileyonlinelibrary.com]

4.5 | The sensitivity of products uncertainty to the choice of temporal and spatial resolutions

So far, the analysis was carried out using precipitation data available every 3 h over a common regular 0.25° latitude–longitude grid. For some applications, we might be interested on knowing the precipitation uncertainty based on data at different temporal and spatial scales. The sensitivity of the uncertainty to the spatiotemporal resolution of the data is assessed here by considering different temporal (3, 6 and 24 h) and spatial (0.25° and 0.75°) resolutions. Decreasing the temporal or spatial resolution implies a decrease in the number of samples and thus the number of bins in the calculation of ΔP_B needs to be adjusted. Here, 20 bins are used for all resolutions to ensure the comparability of the results.

Figure 11 shows the σ_{sel} observational uncertainty for each region and for each of the four metrics. As expected, the observational uncertainty measured with ΔP_M does not depend on the temporal or spatial resolution of the data. For ΔP_B, ΔP_{IF} and ΔP_A, a decrease in the temporal or the spatial resolution of the data leads to substantial decreases in the observational uncertainty in all regions. For instance, Figure 11c shows that the ΔP_A uncertainty decreases from 73% for 3-hourly data to 54% for daily data for a spatial resolution of 0.25° over CAGL. For most regions, the resolutions 0.25°/6-hourly and 0.75°/3-hourly have similar averaged uncertainties. It is however important to note that locally, a decrease of temporal or spatial resolution does not necessary lead to a decrease in the overall uncertainty for ΔP_{IF} and ΔP_B.

5 | DISCUSSION AND CONCLUSIONS

A methodology was presented to quantify the observational uncertainty among a set of eight high-resolution gridded precipitation datasets and to identify, at individual grid points, the products that are most likely to be outliers. At each grid point, the uncertainty is then estimated as the largest precipitation difference among products excluding the outliers, thus avoiding the overestimation of uncertainty arising from erroneous or unrealistic products. Four metrics were used to quantify several aspects of the

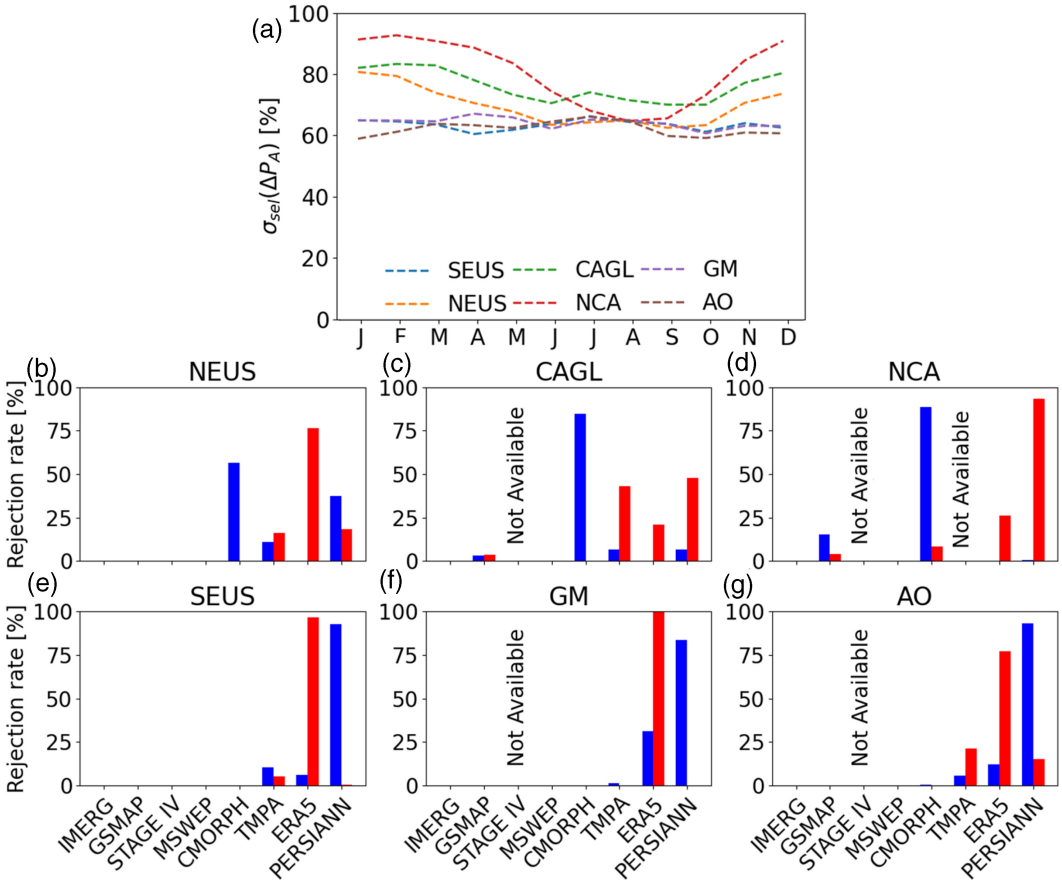

FIGURE 9 Annual cycle of the σ_{sel} uncertainty metric for each region (a) and fraction of grid points where each product is identified as an outlier in summer (red bar) and winter (blue bar) for all six individual regions (b–g). All metrics are computed using the mean absolute difference, ΔP_A. [Colour figure can be viewed at wileyonlinelibrary.com]

precipitation time series in a hierarchical way: from the simple total accumulation (denoted as ΔP_M), to the variability of precipitation by looking at the intensity and frequency (ΔP_{IF}) or the intensity distribution (ΔP_B), to the inclusion of the chronology (i.e., timing) of events (ΔP_A). The different metrics can be viewed as allowing varying degrees of error compensation, with the long-term mean difference (ΔP_M) allowing for complete compensation of errors, the sum of 3-hourly absolute differences (ΔP_A) permitting no compensation, and the two other metrics (ΔP_B and ΔP_{IF}) allowing for partial compensation of errors.

The newly introduced measure of observational uncertainty, which exclude outliers, yields substantially lower values compared to a conventional measure of uncertainty that is based on all products, sometime leading to reductions as large as 50%. For example, the observational uncertainty decreases in winter over NCA from 95% to 48% for mean precipitation ΔP_M, from 97% to 70% for the intensity distribution ΔP_B and from 99% to 90% for the precipitation absolute differences ΔP_A. It should be noted, however, that the identification of outliers is a non-trivial task, particularly in the absence of a high-resolution precipitation product that can be used as a reference. According to our method, outliers are identified as products which are notably different from most others, based on a statistical description of differences between all products. That is, the method assumes that a product that is far from most is an outlier and as such is likely to be wrong. However, there may be such a case that a product identified as outlier is closer to the "truth" than other products. Indeed, as Knutti (2010) and Knutti et al. (2017) have argued regarding the selection of climate models from an ensemble, this assumption can present challenges when the products comprising the ensemble exhibit high dependency. While the precipitation products considered are constructed using a variety of instruments and methodologies, including different corrections from in situ datasets, it is evident that there are dependencies among them. This is clearly observed with MSWEP precipitation estimates which are directly derived from IMERG and ERA5 precipitation estimates.

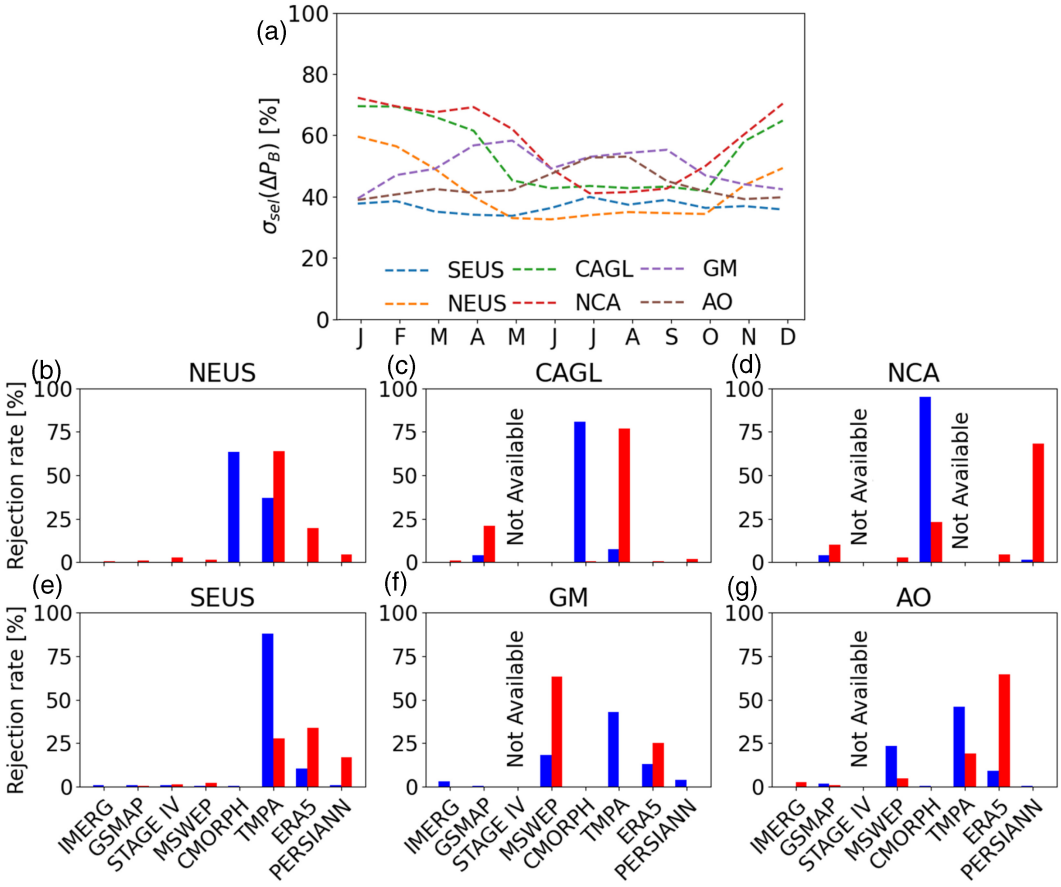

FIGURE 10 As in Figure 9 but using the difference in the intensity distribution, ΔP_B. [Colour figure can be viewed at wileyonlinelibrary.com]

Our findings demonstrate numerous cases where products were identified as outliers, in line with the outcomes and conclusions of earlier studies. This consistency bolsters our confidence in the results:

- CMORPH v1.0 in winter over northern regions for all metrics. The poor performance of CMORPH in cold-season high-latitude has been discussed in detail by Xie et al. (2017).
- ERA5 in summer over southern regions for the precipitation absolute differences ΔP_A metric. This result is likely related with the poor representation of convective precipitation in ERA5 (Gomis-Cebolla et al., 2023; Lavers et al., 2022; Xin et al., 2021), including the representation of the diurnal cycle (Watters et al., 2021).
- TMPA 3B42 in most regions and seasons for all metrics (ΔP_B, ΔP_{IF} and ΔP_A) but the long-term mean difference (ΔP_M). The poor performance of TMPA at representing the frequency and intensity of precipitating events is in agreement with findings in other studies such as Gehne et al. (2016) and Sun et al. (2018).
- GSMaP v7 in winter over Canada for all metrics. At high latitudes, GSMaP misses precipitation events due to the presence of snow, which leads to an underestimation of the mean precipitation (Tian et al., 2010; Wang & Yong, 2020).
- PERSIANN CCS-CDR over Canada for the intensity distribution difference metric (ΔP_B). For northern latitudes, PERSIANN overestimates the intensity of precipitation compared to other products. This issue have been documented for other versions of the products (Nguyen et al., 2018; Sadeghi et al., 2021).

In addition, our approach identifies additional outliers that cannot be directly associated to findings in previous studies:

- MSWEP v2.8 and ERA5 over ocean for the intensity distribution difference metric (ΔP_B). Both MSWEP and ERA5 underestimate precipitation intensities when compared to other products including IMERG which has been shown to underestimate precipitation intensities along the US coast (Derin et al., 2022). We can

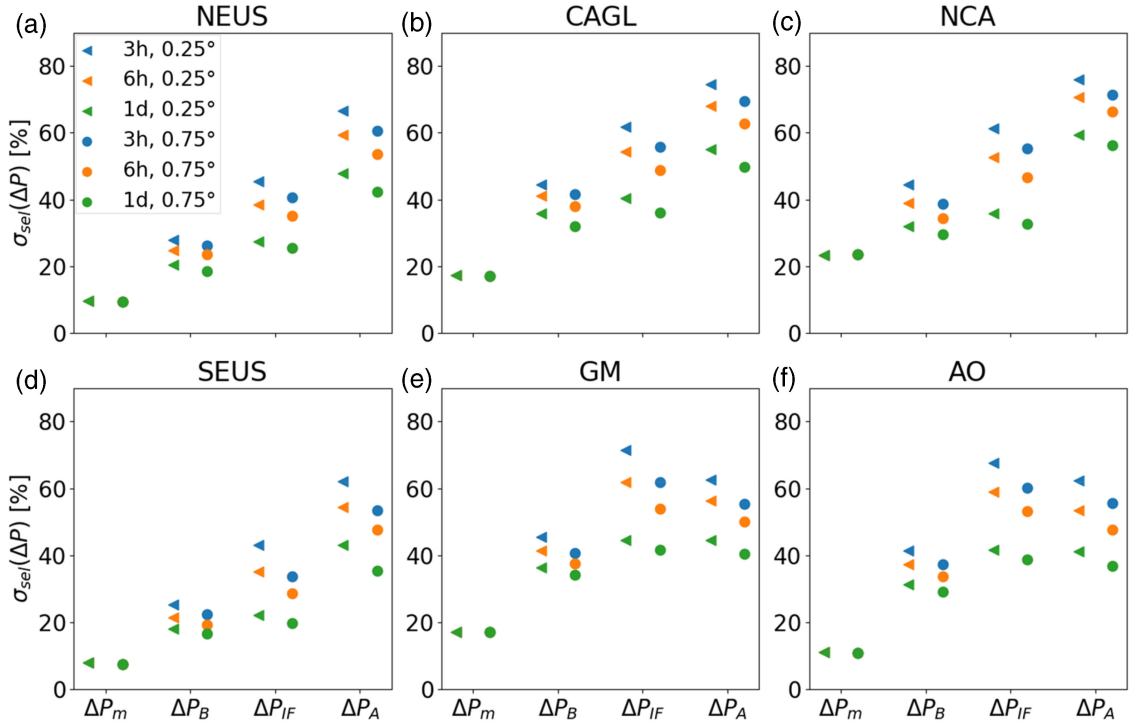

FIGURE 11 Regional mean (a–f) observational uncertainty, σ_{sel}, estimated using data with different spatiotemporal resolutions for each difference metric: the absolute bias (ΔP_M), the difference in the intensity distribution (ΔP_B), the difference in the used intensity-frequency precipitation decomposition (ΔP_{IF}) and the mean absolute difference (ΔP_A). [Colour figure can be viewed at wileyonlinelibrary.com]

thus assume that, at least for this region, they are further from the truth than others.

- PERSIANN CCS-CDR for most regions for the precipitation absolute difference metric (ΔP_A). PERSIANN shows similar mean precipitation and an intensity-frequency distribution to other products with low values of ΔP_M, ΔP_B and ΔP_{IF}. However, the chronology of precipitation events differs substantially between PERSIANN and other products, leading to high values of ΔP_A. We are not aware of other literature showing this issue.

IMERG v6 and STAGE IV are not systematically identified as outliers in any region or season. While this does not necessarily imply that these two datasets are the best, it does show that both datasets are rarely very different from most others.

Results show that the choice of the metric leads to very large differences in the observational uncertainty values. For example, the uncertainty in the absolute bias (ΔP_M) is generally within 20% while it can attain 90% when considering the time-average of absolute differences of 3-hourly precipitation (ΔP_A). This shows that metrics capture disparities in different characteristics of precipitation time series and that the choice of the metric needs to be carefully considered according to the aim of the study. In this study, the absolute difference, ΔP_A, is often used as a reference because it captures all disparities between products but its use as the main metric would be unwise whenever the chronology of precipitation events is not a key aspect to consider, such as for climate model evaluation (Ashouri et al., 2015; Beck et al., 2017b). For the same reason, if a study focuses on monthly mean precipitation, there might be no need to eliminate a product simply because it poorly represents the variability in 3-hourly precipitation. While it is difficult to provide general guidelines about the use of the products considered in our analysis, we believe that our results can inform the selection of specific products depending on the application at hand (i.e., for a given region, season, and quantity of interest).

Our results show that the observational uncertainty and the outlier identification exhibit strong regional and seasonal variations. The uncertainty obtained from the annual analysis does not accurately reflect the seasonal uncertainty, especially for northern regions. For the NCA region, the precipitation absolute difference metric (ΔP_A) annual uncertainty is 76% rising to as much as 90% in winter. The same applies to the identification of outliers and products identified as outliers according to the annual analysis are generally not identified as outliers for the seasonal analysis. For example, for NCA, the annual analysis identifies PERSIANN as an outlier while the winter season analysis shows that CMORPH is the main

outlier. It is also important to note that for most regions, notably regions with few direct observations, such as Canada or the oceans, the uncertainty is generally large, and the use of different product types seems unavoidable. In this case, our method is particularly relevant by identify products that should not be used.

Our results also show that the observational uncertainty decreases substantially for all metrics, except the mean precipitation difference ΔP_M, as the horizontal or the temporal resolution of the data decreases. This is expected and related with the smoothing effect of decreasing the resolution (Nitu et al., 2018; Norris et al., 2019; Trenberth et al., 2017; Trenberth & Zhang, 2018). For instance, a change in the temporal resolution from 3 h to 1 day decreases the uncertainty by a quarter for northern Canada and by almost half for southeast United States for ΔP_A. However, lowering the resolution not only reduces the uncertainty (which can be seen as a way to reach more practical values for a given application), but also the information within datasets.

Considering the growing availability of precipitation products with high spatiotemporal resolution, further research is required to determine how to best estimate precipitation values and their associated observational uncertainty. In particular, future methods need to explicitly address the presence of highly interdependent precipitation products within the ensemble. In addition, it would be valuable to take advantage of the broader availability of subdaily in situ observations (e.g., Collet et al., 2022; Smith et al., 2011) to continue improving precipitation products.

AUTHOR CONTRIBUTIONS

Tangui Picart: Conceptualization; investigation; writing – original draft; writing – review and editing; methodology; software. **Alejandro Di Luca:** Funding acquisition; methodology; writing – review and editing; supervision. **René Laprise:** Funding acquisition; writing – review and editing; supervision.

ACKNOWLEDGEMENTS

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), funding application number RGPIN-2020-05631. The authors would like to thank Katja Winger, François Roberge and Frédérik Toupin for maintaining a user-friendly local computing facility and for downloading and preparing some of the precipitation datasets. Finally, the authors thank the two anonymous reviewers for their helpful comments and suggestions for improving this paper.

DATA AVAILABILITY STATEMENT

Data analysed in this study are openly available: CMORPH v1 (Joyce et al., 2004) at ftp://ftp.cpc.ncep.noaa.gov/precip/global_CMORPH/3-hourly_025deg, GSMaP v7 (Ushio et al., 2009) at https://sharaku.eorc.jaxa.jp/GSMaP/guide.html, TMPA 3B42 (Huffman et al., 2007) at https://doi.org/10.5067/TRMM/TMPA/3H/7, IMERG V6 (Huffman et al., 2018) at https://doi.org/10.5067/GPM/IMERG/3B-HH/06, PERSIANN (Sadeghi et al., 2021) at https://persiann.eng.uci.edu/CHRSdata/PERSIANN-CDR/adj_3hB1/, STAGE IV (Lin & Mitchell, 2005) at https://doi.org/10.26023/PB7G-6Y5W-8Y0B, ERA5 (Hersbach et al., 2020) at https://doi.org/10.24381/cds.adbb2d47, MSWEP V2.8 (Beck et al., 2019b) at http://www.gloh2o.org/mswep/. An example of application of our method is provided can be found at https://github.com/TanguiP/Example_uncertainty_outliers.

ORCID

Tangui Picart https://orcid.org/0009-0000-9185-6741

REFERENCES

Aghakouchak, A., Mehran, A., Norouzi, H. & Behrangi, A. (2012) Systematic and random error components in satellite precipitation data sets. *Geophysical Research Letters*, 39, 3–6. Available from: https://doi.org/10.1029/2012GL051592

Ashouri, H., Hsu, K.L., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D. et al. (2015) PERSIANN-CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, 96, 69–83. Available from: https://doi.org/10.1175/BAMS-D-13-00068.1

Beck, H.E., Van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B. et al. (2017a) MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, 21, 589–615. Available from: https://doi.org/10.5194/hess-21-589-2017

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A.I.J.M., Weedon, G.P. et al. (2017b) Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Advances in Global Change Research*, 69, 625–653. Available from: https://doi.org/10.1007/978-3-030-35798-6_9

Beck, H.E., Pan, M., Roy, T., Weedon, G.P., Pappenberger, F., Van Dijk, A.I.J.M. et al. (2019a) Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, 23, 207–224. Available from: https://doi.org/10.5194/hess-23-207-2019

Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., Van Dijk, A.I.J.M. et al. (2019b) MSWep v2 Global 3-hourly 0.1° precipitation: methodology and quantitative assessment. *Bulletin of the American Meteorological Society*, 100, 473–500. Available from: https://doi.org/10.1175/BAMS-D-17-0138.1

Bytheway, J.L., Hughes, M., Mahoney, K. & Cifelli, A.R. (2020) On the uncertainty of high-resolution hourly quantitative

precipitation estimates in California. *Journal of Hydrometeorology*, 21, 865–879. Available from: https://doi.org/10.1175/JHM-D-19-0160.1

Catto, J.L., Jakob, C. & Nicholls, N. (2015) Can the CMIP5 models represent winter frontal precipitation? *Geophysical Research Letters*, 42, 8596–8604. Available from: https://doi.org/10.1002/2015GL066015

Chen, C.-T. & Knutson, T. (2008) On the verification and comparison of extreme rainfall indices from climate models. *Journal of Climate*, 21, 1605–1621. Available from: https://doi.org/10.1175/2007JCLI1494.1

Collet, F., Di Luca, A. & Chen, T.-C. (2022) North America ISD to ERA5 (NA-ISD2ERA) Catalogue. Borealis, V1.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S. et al. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597. Available from: https://doi.org/10.1002/qj.828

Derin, Y., Kirstetter, P.-E., Brauer, N., Gourley, J.J. & Wang, J. (2022) Evaluation of IMERG satellite precipitation over the Land–Coast–Ocean continuum. Part II: quantification. *Journal of Hydrometeorology*, 23, 1297–1314. Available from: https://doi.org/10.1175/JHM-D-21-0234.1

Derin, Y. & Yilmaz, K.K. (2014) Evaluation of multiple satellite-based precipitation products over complex topography. *Journal of Hydrometeorology*, 15, 1498–1516. Available from: https://doi.org/10.1175/JHM-D-13-0191.1

Di Luca, A., Argüeso, D., Sherwood, S. & Evans, J.P. (2021) Evaluating precipitation errors using the environmentally conditioned intensity-frequency decomposition method. *Journal of Advances in Modeling Earth Systems*, 13, 1–45. Available from: https://doi.org/10.1029/2020MS002447

Ebert, E.E., Janowiak, J.E. & Kidd, C. (2007) Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society*, 88, 47–64. Available from: https://doi.org/10.1175/BAMS-88-1-47

Fosser, G., Khodayar, S. & Berg, P. (2015) Benefit of convection permitting climate model simulations in the representation of convective precipitation. *Climate Dynamics*, 44, 45–60. Available from: https://doi.org/10.1007/s00382-014-2242-1

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

Gervais, M., Gyakum, J.R., Atallah, E., Tremblay, L.B. & Neale, R.B. (2014) How well are the distribution and extreme values of daily precipitation over North America represented in the community climate system model? A comparison to reanalysis, satellite, and gridded station data. *Journal of Climate*, 27, 5219–5239. Available from: https://doi.org/10.1175/JCLI-D-13-00320.1

GloH20. (2021) MSWEP V2.8 technical documentation, GloH20, pp. 1–8.

Gomis-Cebolla, J., Rattayova, V., Salazar-Galán, S. & Francés, F. (2023) Evaluation of ERA5 and ERA5-Land reanalysis precipitation datasets over Spain (1951–2020). *Atmospheric Research*, 284, 106606. Available from: https://doi.org/10.1016/j.atmosres.2023.106606

GPM Global Rainfall Map Algorithm Development Team. (2014) Global Satellite Mapping of Precipitation (GSMaP) for GPM Algorithm Theoretical Basis Document (ATBD): Algorithm Theoretical Basis Document (ATBD): Algorithm Ver. 6 17.

Guilloteau, C., Foufoula-Georgiou, E., Kirstetter, P., Tan, J. & Huffman, G.J. (2022) How well do multisatellite products capture the space–time dynamics of precipitation? Part II: building an error model through spectral system identification. *Journal of Hydrometeorology*, 23, 1383–1399. Available from: https://doi.org/10.1175/JHM-D-22-0041.1

Harris, I., Osborn, T.J., Jones, P. & Lister, D. (2020) Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7, 1–18. Available from: https://doi.org/10.1038/s41597-020-0453-3

Henn, B., Newman, A.J., Livneh, B., Daly, C. & Lundquist, J.D. (2018) An assessment of differences in gridded precipitation datasets in complex terrain. *Journal of Hydrology*, 556, 1205–1219. Available from: https://doi.org/10.1016/j.jhydrol.2017.03.008

Herold, N., Alexander, L.V., Donat, M.G., Contractor, S. & Becker, A. (2016) How much does it rain over land? *Geophysical Research Letters*, 43, 341–348. Available from: https://doi.org/10.1002/2015GL066615

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. Available from: https://doi.org/10.1002/qj.3803

Hersbach, H., Bell, B., Berrisford, P., Horányi, A., Sabater, J.M., Nicolas, J. et al. (2019) Global reanalysis: goodbye ERA-Interim, hello ERA5. *ECMWF Newsletter*, 159, 17–24. Available from: https://doi.org/10.21957/vf291hehd7

Hong, Y., Hsu, K.-L., Sorooshian, S. & Gao, X. (2004) Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology*, 43, 1834–1853. Available from: https://doi.org/10.1175/JAM2173.1

Hossain, F. & Huffman, G.J. (2008) Investigating error metrics for satellite rainfall data at hydrologically relevant scales. *Journal of Hydrometeorology*, 9, 563–575. Available from: https://doi.org/10.1175/2007JHM925.1

Hsu, K.L., Gao, X., Sorooshian, S. & Gupta, H.V. (1997) Precipitation estimation from remotely sensed information using artificial neural networks. *Journal of Applied Meteorology*, 36, 1176–1190. Available from: https://doi.org/10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2

Hsu, K., Gupta, H.V., Gao, X. & Sorooshian, S. (1999) Estimation of physical variables from multichannel remotely sensed imagery using a neural network: Application to rainfall estimation. *Water Resources Research*, 35, 1605–1618. Available from: https://doi.org/10.1029/1999WR900032

Huffman, G., Bolvin, D., Braithwaite, D., Hsu, K. & Joyce, R. (2018) Algorithm theoretical basis document (ATBD) NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG). Nasa 29.

Huffman, G.J., Adler, R.F., Bolvin, D.T., Gu, G., Nelkin, E.J., Bowman, K.P. et al. (2007) The TRMM Multisatellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8, 38–55. Available from: https://doi.org/10.1175/JHM560.1

Huffman, G.J., Adler, R.F., Bolvin, D.T. & Nelkin, E.J. (2010) The TRMM Multi-Satellite Precipitation Analysis (TMPA). In: Gebremichael, M. & Hossain, F. (Eds.) *Satellite Rainfall Applications for Surface Hydrology*. Springer, pp. 3–22. Available from: https://doi.org/10.1007/978-90-481-2915-7_1

Huffman, G.J. & Bolvin, D.T. (2018) TRMM and other data precipitation data set documentation, pp. 1–48.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C. et al. (2015) NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD) Version 4.5 26.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K.-L., Joyce, R.J., Kidd, C. et al. (2020) Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG). In: Levizzani, V., Kidd, C., Kirschbaum, D.B., Kummerow, C.D., Nakamura, K. & Turk, F.J. (Eds.) *Satellite precipitation measurement*, Vol. 1. Advances in Global Change Research. Springer International Publishing, pp. 343–353. Available from: https://doi.org/10.1007/978-3-030-24568-9_19

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

Kidd, C., Becker, A., Huffman, G.J., Muller, C.L., Joe, P., Skofronick-Jackson, G. et al. (2017) So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society*, 98, 69–78. Available from: https://doi.org/10.1175/BAMS-D-14-00283.1

Knutti, R. (2010) The end of model democracy? *Climatic Change*, 102, 395–404. Available from: https://doi.org/10.1007/s10584-010-9800-2

Knutti, R., Sedláček, J., Sanderson, B.M., Lorenz, R., Fischer, E.M. & Eyring, V. (2017) A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, 44, 1909–1918. Available from: https://doi.org/10.1002/2016GL072012

Lavers, D.A., Simmons, A., Vamborg, F. & Rodwell, M.J. (2022) An evaluation of ERA5 precipitation for climate monitoring. *Quarterly Journal of the Royal Meteorological Society*, 148, 3152–3165. Available from: https://doi.org/10.1002/qj.4351

Lin, Y. & Mitchell, K.E. (2005) *The NCEP stage II/IV hourly precipitation analyses: development and applications*. Presented at 85th AMS Annual Meeting, American Meteorological Society - combined preprints 1649–1652.

Liu, C. & Allan, R.P. (2012) Multisatellite observed responses of precipitation and its extremes to interannual climate variability. *Journal of Geophysical Research: Atmospheres*, 117, D03101. Available from: https://doi.org/10.1029/2011JD016568

Lockhoff, M., Zolina, O., Simmer, C. & Schulz, J. (2019) Representation of precipitation characteristics and extremes in regional reanalyses and satellite- and gauge-based estimates over Western and Central Europe. *Journal of Hydrometeorology*, 20, 1123–1145. Available from: https://doi.org/10.1175/JHM-D-18-0200.1

Mega, T., Ushio, T., Matsuda, T., Kubota, T., Kachi, M. & Oki, R. (2019) Gauge-Adjusted Global Satellite Mapping of Precipitation. *IEEE Transactions on Geoscience and Remote Sensing*, 57, 1928–1935. Available from: https://doi.org/10.1109/TGRS.2018.2870199

Meyer-Christoffer, A., Becker, A., Finger, P., Rudolf, B., Schneider, U. & Ziese, M. (2011) GPCC Climatology Version 2011 at 0.5°: monthly land-surface precipitation climatology for every month and the total year from rain-gauges built on GTS-based and historic data. https://doi.org/10.5676/DWD_GPCC/CLIM_M_V2011_025

Nearing, G.S., Tian, Y., Gupta, H.V., Clark, M.P., Harrison, K.W. & Weijs, S.V. (2016) A philosophical basis for hydrological uncertainty. *Hydrological Sciences Journal*, 61, 1666–1678. Available from: https://doi.org/10.1080/02626667.2016.1183009

Nelson, B.R., Prat, O.P., Seo, D.J. & Habib, E. (2016) Assessment and implications of NCEP stage IV quantitative precipitation estimates for product intercomparisons. *Weather and Forecasting*, 31, 371–394. Available from: https://doi.org/10.1175/WAF-D-14-00112.1

Nguyen, P., Ombadi, M., Sorooshian, S., Hsu, K., AghaKouchak, A., Braithwaite, D. et al. (2018) The PERSIANN family of global satellite precipitation data: a review and evaluation of products. *Hydrology and Earth System Sciences*, 22, 5801–5816. Available from: https://doi.org/10.5194/hess-22-5801-2018

Nissen, K.M. & Ulbrich, U. (2017) Increasing frequencies and changing characteristics of heavy precipitation events threatening infrastructure in Europe under climate change. *Natural Hazards and Earth System Sciences*, 17, 1177–1190. Available from: https://doi.org/10.5194/nhess-17-1177-2017

Nitu, R., Roulet, Y.-A., Wolff, M., Earle, M., Reverdin, A., Smith, C. et al. (2018) WMO Solid Precipitation Intercomparison Experiment (SPICE) (2012–2015). WMO-IOM Report No. 131. Available from: https://www.researchgate.net/publication/349114825_WMO_Solid_Precipitation_Intercomparison_Experiment_SPICE

Norris, J., Chen, G. & David Neelin, J. (2019) Changes in frequency of large precipitation accumulations over land in a warming climate from the CESM large ensemble: the roles of moisture, circulation, and duration. *Journal of Climate*, 32, 5397–5416. Available from: https://doi.org/10.1175/JCLI-D-18-0600.1

Prat, O.P. & Nelson, B.R. (2015) Evaluation of precipitation estimates over CONUS derived from satellite, radar, and rain gauge data sets at daily to annual scales (2002–2012). *Hydrology and Earth System Sciences*, 19, 2037–2056. Available from: https://doi.org/10.5194/hess-19-2037-2015

Prein, A.F. & Gobiet, A. (2017) Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *International Journal of Climatology*, 37, 305–327. Available from: https://doi.org/10.1002/joc.4706

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A.P. et al. (2012) How well are we measuring snow? *Bulletin of the American Meteorological Society*, 93, 811–829.

Sadeghi, M., Nguyen, P., Naeini, M.R., Hsu, K., Braithwaite, D. & Sorooshian, S. (2021) PERSIANN-CCS-CDR, a 3-hourly 0.04° global precipitation climate data record for heavy precipitation studies. *Scientific Data*, 8, 1–11. Available from: https://doi.org/10.1038/s41597-021-00940-9

Smith, A., Lott, N. & Vose, R. (2011) The integrated surface database: recent developments and partnerships. *Bulletin of the American Meteorological Society*, 92, 704–708. Available from: https://doi.org/10.1175/2011BAMS3015.1

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S. & Hsu, K.L. (2018) A review of global precipitation data sets: data sources, estimation, and Intercomparisons. *Reviews of Geophysics*, 56, 79–107. Available from: https://doi.org/10.1002/2017RG000574

Tian, Y. & Peters-Lidard, C.D. (2010) A global map of uncertainties in satellite-based precipitation measurements. *Geophysical Research Letters*, 37, 1–6. Available from: https://doi.org/10.1029/2010GL046008

Tian, Y., Peters-Lidard, C.D., Adler, R.F., Kubota, T. & Ushio, T. (2010) Evaluation of GSMaP precipitation estimates over the contiguous United States. *Journal of Hydrometeorology*, 11, 566–574. Available from: https://doi.org/10.1175/2009JHM1190.1

Tian, Y., Peters-Lidard, C.D., Eylander, J.B., Joyce, R.J., Huffman, G.J., Adler, R.F. et al. (2009) Component analysis of errors in satellite-based precipitation estimates. *Journal of Geophysical Research: Atmospheres*, 114, 1–15. Available from: https://doi.org/10.1029/2009JD011949

Timmermans, B., Wehner, M., Cooley, D., O'Brien, T. & Krishnan, H. (2019) An evaluation of the consistency of extremes in gridded precipitation data sets. *Climate Dynamics*, 52, 6651–6670. Available from: https://doi.org/10.1007/s00382-018-4537-0

Trenberth, K.E. (2011) Changes in precipitation with climate change. *Climate Research*, 47, 123–138. Available from: https://doi.org/10.3354/cr00953

Trenberth, K.E. & Zhang, Y. (2018) How often does it really rain? *Bulletin of the American Meteorological Society*, 99, 289–298. Available from: https://doi.org/10.1175/BAMS-D-17-0107.1

Trenberth, K.E., Zhang, Y. & Gehne, M. (2017) Intermittency in precipitation: duration, frequency, intensity, and amounts using hourly data. *Journal of Hydrometeorology*, 18, 1393–1412. Available from: https://doi.org/10.1175/JHM-D-16-0263.1

Ushio, T., Sasashige, K., Kubota, T., Shige, S., Okamoto, K., Aonashi, K. et al. (2009) A kalman filter approach to the global satellite mapping of precipitation (GSMaP) from combined passive microwave and infrared radiometric data. *Journal of the Meteorological Society of Japan*, 87, 137–151. Available from: https://doi.org/10.2151/jmsj.87A.137

Wang, H. & Yong, B. (2020) Quasi-global evaluation of IMERG and GSMaP precipitation products over land using gauge observations. *Water*, 12, 243. Available from: https://doi.org/10.3390/w12010243

Watters, D., Battaglia, A. & Allan, R.P. (2021) The diurnal cycle of precipitation according to multiple decades of global satellite observations, three CMIP6 models, and the ECMWF reanalysis. *Journal of Climate*, 34, 5063–5080. Available from: https://doi.org/10.1175/JCLI-D-20-0966.1

Xie, P., Joyce, R., Wu, S., Yoo, S.H., Yarosh, Y., Sun, F. et al. (2017) Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *Journal of Hydrometeorology*, 18, 1617–1641. Available from: https://doi.org/10.1175/JHM-D-16-0168.1

Xin, Y., Lu, N., Jiang, H., Liu, Y. & Yao, L. (2021) Performance of ERA5 reanalysis precipitation products in the Guangdong-Hong Kong-Macao greater Bay Area, China. *Journal of Hydrology*, 602, 126791. Available from: https://doi.org/10.1016/j.jhydrol.2021.126791

Zhuang, J., Dussin, R., Jüling, A. & Rasp, S. (2020) JiaweiZhuang/-xESMF: v0.3.0 adding ESMF.LocStream capabilities (v0.3.0). Zenodo. https://doi.org/10.5281/zenodo.3700105

Zolina, O., Simmer, C., Kapala, A., Shabanov, P., Becker, P., MäcHel, H. et al. (2014) Precipitation variability and extremes in Central Europe: new view from STAMMEX results. *Bulletin of the American Meteorological Society*, 95, 995–1002. Available from: https://doi.org/10.1175/BAMS-D-12-00134.1

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Picart, T., Di Luca, A., & Laprise, R. (2024). Uncertainty and outliers in high-resolution gridded precipitation products over eastern North America. *International Journal of Climatology*, *44*(4), 1014–1035. https://doi.org/10.1002/joc.8369