An Extreme Precipitation Categorization Scheme and its Observational Uncertainty over the Continental United States

EMILY A. SLINSKEY AND PAUL C. LOIKITH

Department of Geography, Portland State University, Portland, Oregon

DUANE E. WALISER AND ALEXANDER GOODMAN

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

(Manuscript received 26 July 2018, in final form 21 March 2019)

ABSTRACT

An extreme precipitation categorization scheme, used to temporally and spatially visualize and track the multiscale variability of extreme precipitation climatology, is applied over the continental United States. The scheme groups 3-day precipitation totals exceeding 100 mm into one of five precipitation categories, or “P-Cats.” To demonstrate the categorization scheme and assess its observational uncertainty across a range of precipitation measurement approaches, we compare the climatology of P-Cats defined using in situ station data from the Global Historical Climatology Network-Daily (GHCN-D); satellite-derived data from the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA); gridded station data from the Parameter-Elevation Regression on Independent Slopes Model (PRISM); global reanalysis from the Modern-Era Retrospective Analysis for Research and Applications, version 2; and regional reanalysis from the North American Regional Reanalysis. While all datasets capture the principal spatial patterns of P-Cat climatology, results show considerable variability across the suite in frequency, spatial extent, and magnitude. Higher-resolution datasets, PRISM and TMPA, most closely resemble GHCN-D and capture a greater frequency of high-end P-Cats relative to the lower-resolution products. When all datasets are rescaled to a common coarser grid, differences persist with datasets originally constructed at a high resolution maintaining a higher frequency and magnitude of P-Cats. Results imply that dataset choice matters when applying the P-Cat scheme to track extreme precipitation over space and time. Potential future applications of the P-Cat scheme include providing a target for climate model evaluation and a basis for characterizing future change in extreme precipitation as projected by climate model simulations.

1. Introduction

Extreme precipitation is associated with a multitude of societal and environmental impacts across the United States. Often accompanying severe weather events, including hurricanes, snowstorms, and atmospheric rivers, these meteorological phenomena pose a threat to property, agriculture, infrastructure, and human life while also playing a key role in the water budget (Kunkel et al. 2013). According to the 2017 National Climate Assessment (NCA) Climate Science Special Report, climate change is

projected to alter the frequency, severity, and seasonality of extreme precipitation across the United States (Easterling et al. 2017). Climate change mitigation policies and adaption initiatives are greatly influenced by societal vulnerabilities to climate impacts like those associated with extreme precipitation. Therefore, a comprehensive understanding and intuitive way to track and project change across space and time at impacts-relevant scales is critical.

Climate model projections of future change in global precipitation generally follow the Clausius–Clapeyron relationship projecting the atmosphere’s water holding capacity to increase exponentially with temperature at roughly 7% °C^{−1} warming (Allen and Ingram 2002; Trenberth et al. 2003; Pall et al. 2007). Consistent with these expectations, a number of studies have suggested that anthropogenic climate warming may be attributable to an increase in the probability and severity of recent

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-18-0148.s1>.

Corresponding author: Emily A. Slinskey, slinskey@pdx.edu

notable heavy precipitation events over the United States such as September 2013 in Colorado (Pall et al. 2017), the rainfall from Hurricane Harvey (Risser and Wehner 2017), and the August 2016 Louisiana event (Wang et al. 2016). However, the sign and magnitude of observed changes in extreme precipitation are not always immediately apparent from observational analysis at local through regional scales. This is due in part to the character of extreme precipitation varying considerably over space and time, making it difficult to detect an anthropogenic signal above natural variability (Easterling et al. 2000; O’Gorman and Schneider 2009). Furthermore, understanding observed and projected changes in the frequency and intensity of key mechanisms associated with extreme precipitation, such as tropical cyclones and atmospheric rivers, is still an area of active research (e.g., Knight and Davis 2009; Prat and Nelson 2013; Gao et al. 2015; Behrangi et al. 2016; Mahoney et al. 2016; Lamjiri et al. 2017).

Several extreme precipitation indices have been developed and applied to a diverse set of datasets using a range of methods to examine changes in extreme precipitation over space and time (Zhang et al. 2011, and references therein). One example is a set of extreme indices developed by the Expert Team on Climate Change Detection and Indices (ETCCDI) as part of the World Climate Research Programme Project on Climate Variability and Predictability (Frich et al. 2002; Alexander et al. 2006). These indices are designed to address a broad range of global climate information needs ranging from the frequency of precipitation threshold exceedances to the maximum length of wet spells. Specific to the United States, precipitation extremes have been monitored using the U.S. Climate Extremes Index (Gleason et al. 2008) in addition to the U.S. Environmental Protection Agency’s climate indicator for annual heavy precipitation aggregated over the continental United States (CONUS; U.S. EPA 2016). While concise and useful, these monitoring approaches provide a great deal of climate information at broad global and national scales, but less information at local to regional scales. The regional variability in extreme precipitation can be large across a single climate region (e.g., the Northwest or Southeast), therefore it is important that monitoring addresses the need for regional relevance while also providing a similarly high level of intuitive interpretability.

The ability to detect, analyze, and track changes in extreme precipitation is also heavily dependent on the reliability of observations and a number of precipitation climatology and dataset intercomparison studies have been conducted at global and regional scales highlighting these differences (e.g., Adler et al.

2001; Guirguis and Avissar 2008). In situ station data are commonly accepted as a primary source and often used as a reference relative to other products. However, station observations are spatially heterogeneous and may be temporally inconsistent, creating observational gaps (Kidd et al. 2017). Satellite-based precipitation measurements, on the other hand, are spatially seamless regardless of in situ gauge density or quality, however these datasets exhibit bias resulting from instrumental and algorithmic error (Sapiano and Arkin 2009; Chen et al. 2013; Behrangi et al. 2014a; Tan et al. 2016). Similarly, bias can be introduced to analysis products through data assimilation and model errors (Bukovsky and Karoly 2007; Bosilovich et al. 2008; Reichle et al. 2017), to gridded in situ products through spatial interpolation (Daly 2006), and simply from spatial resolution (Herold et al. 2017). Additionally, the high spatial and temporal variability characterizing precipitation extremes has been shown to result in exceedingly low agreement among a range of global precipitation measurement products (Donat et al. 2013). Because the dataset one uses has been shown to matter, it is critical to understand and, where possible, constrain observational uncertainty when monitoring and tracking precipitation extremes.

Here we present a climatology of an extreme precipitation categorization scheme as an intuitive way to interpret extreme precipitation climatology, variability, and change over space and time and evaluate its observational uncertainty across a range of datasets. The application of this scheme is motivated by the need for an intuitive, pointwise climate indicator for extreme precipitation that can be provided clearly at scales relevant to societal and environmental impacts. The broad and diverse range of extreme precipitation impacts makes the regional information provided by the indicator suitable for a wide range of interests, including scientists and practitioners, concerned with heavy precipitation climatology, variability, and change at local through CONUS levels. The approach, which is analogous to the familiar Saffir–Simpson hurricane intensity scale, assigns categories from one to five to extreme 3-day precipitation totals at each data point (grid cell or rain gauge). However, unlike the Saffir–Simpson scale, this approach is not designed to rank an individual storm event, but rather provide information at climate scales for pointwise magnitudes of heavy 3-day precipitation totals while being extensible across datasets, time, and space. This approach, adapted from the “R-Cat” categorization scheme first presented in Ralph and Dettinger (2012), can then be stratified by season, geographic

TABLE 1. Datasets used in the intercomparison and their associated specifications.

Agency Source	Dataset		Spatial resolution	Temporal resolution	Data source	Reference
NASA	TMPA	TRMM Multi-Satellite Precipitation Analysis 3B42V7	$0.25^{\circ} \times 0.25^{\circ}$	3 hourly	Satellite	Huffman et al. (2007)
NASA	IMERG	Integrated Multi-Satellite Retrievals for GPM	$0.1^{\circ} \times 0.1^{\circ}$	30 min	Satellite	Huffman et al. (2017)
OSU	PRISM	Parameter-Elevation Regressions on Independent Slopes Model	$0.04^{\circ} \times 0.04^{\circ}$	Daily	Gridded in situ station data	Daly et al. (1994)
NASA	MERRA-2	Modern Era Retrospective-Analysis version 2	$0.625^{\circ} \times 0.5^{\circ}$	3 hourly	Global reanalysis	Molod et al. (2015)
NCEP	NARR	North American Regional Reanalysis	$32 \text{ km} \times 32 \text{ km}$	3 hourly	Regional reanalysis with gauge data assimilation	Mesinger et al. (2006)
NOAA	GHCN-D	Global Historical Climatology Network		Daily	In situ station data	Menne et al. (2012)

subregion, or time period, while change in extreme event categories can be monitored across multiple spatial and temporal scales. By examining the observational uncertainty of this scheme, this study highlights both the utility of the approach as a means to depict the climatology of extreme precipitation, as well as what considerations should be made when choosing a reference dataset.

2. Data

We apply the extreme precipitation categorization scheme to five datasets, each constructed using a different approach and all provided at a relatively high spatial resolution. All datasets used for the intercomparison are summarized in Table 1 and described in more detail below.

a. TRMM 3B42V7

Satellite-derived precipitation data are from NASA's Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) 3B42V7 product (Huffman et al. 2007; Huffman and Bolvin 2015). Prior to its decommissioning in 2015, TMPA was NASA's flagship precipitation measurement product (Liu et al. 2012). TMPA is provided with a 3-hourly temporal and 0.25° latitude–longitude spatial resolution, globally from 50°N to 50°S latitude from 1998 to 2015. TMPA measurements are produced using microwave-calibrated infrared (IR) estimates from multiple geostationary Earth-orbiting and low-Earth-orbiting satellites (Huffman et al. 2007). The final precipitation estimates contain microwave-derived measurements and calibrated thermal IR-derived estimates. The spatial domain accounts for the tendency of microwave and IR estimates to lose skill at higher latitudes (Huffman et al. 2010). The 3B42V7 product incorporates monthly in situ gauge

observations from the Global Precipitation Climatology Center and the Climate Assessment and Monitoring System for bias adjustment.

b. IMERG

As a part of the Global Precipitation Measurement (GPM) mission, the Integrated Multisatellite Retrievals for GPM (IMERG) product was developed as an extension of TMPA after its decommission. IMERG data are provided at 0.1° latitude–longitude resolution every half hour between 60°N and 60°S latitude (Hou et al. 2014; Liu 2016). The GPM core observatory presents an increased orbiting inclination over TRMM, from 35° to 65° respectively, rendering more extensive latitudinal coverage (Huffman et al. 2017). Additionally, more advanced instrumentation, capable of capturing multiple phases of precipitation, is possible through the addition of a higher-frequency radar offering an improved sensitivity to light precipitation as well as to snow and ice. IMERG integrates algorithms from TMPA, the Climate Prediction Center morphing technique, and Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. As of the writing of the paper, IMERG extends from April 2014 to the present, but will be retro-processed to overlap the TRMM era. IMERG and TMPA are freely available via the GES DISC.

c. PRISM

The Parameter-Elevation Regressions on Independent Slopes Model (PRISM) uses point data and a digital elevation model (DEM) to generate gridded precipitation data (Daly et al. 1994). We utilize the daily PRISM product, offered on a 0.04° latitude–longitude grid over the CONUS. The PRISM technique attempts to account for physiographic effects such as coastal

proximity and orography using the linear regression between gauge measurements and the elevation of the gauge taken from a DEM (Daly et al. 1994, 2002, 2008). The gauge measurements used for interpolation were supplied by various sources including the U.S. National Weather Service Cooperative Observer Network and the Natural Resources Conservation Service daily snow-pack telemetry gauges. Station network density relates to population density (Daly et al. 2007). The PRISM product is freely available from Oregon State University's PRISM Climate Group portal.

d. MERRA-2

The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2), atmospheric reanalysis product provides 3-hourly precipitation estimates generated on a $0.5^\circ \times 0.625^\circ$ latitude–longitude grid. MERRA-2 is the latest multiyear reanalysis product produced by NASA's Global Modeling and Assimilation Office using the Goddard Earth Observing System version 5 (Molod et al. 2015; Gelaro et al. 2017; Reichle et al. 2017). This product corrects model generated precipitation estimates with observations, showing marked improvements upon its predecessor MERRA (Rienecker et al. 2011; Reichle et al. 2017). The method for merging observed precipitation into MERRA-2 assimilates aerosols and integrates MERRA-Land reanalysis for correction (Reichle et al. 2017). Estimates are further merged with precipitation generated by the MERRA-2 atmospheric general circulation model weighted according to latitude. MERRA-2 is freely available via the GES DISC.

e. NARR

The North American Regional Reanalysis (NARR) is based on the regional Eta model and its 3D variation data assimilation system initialized from lateral boundary conditions provided by the National Centers for Environmental Information (NCEI; Mesinger et al. 2006) and is freely available through the National Oceanic and Atmospheric Administration's Earth System Research Laboratory. This product is provided at a 3-hourly temporal resolution and a 32-km spatial resolution (Lin et al. 1999). Precipitation gauge observations are used to adjust atmospheric moisture and energy field estimates to improve model-derived precipitation fields.

f. GHCN

In situ daily observations are from the NCEI Global Historical Climatology Network–Daily (GHCN-D) product (Menne et al. 2012). This dataset contains comprehensive in situ climatic data that have undergone extensive

FIG. 1. P-Cat thresholds and associated colors used in subsequent figures.

quality control procedures to limit internal, spatial, and temporal inconsistencies (Durre et al. 2010). For this study, only gauges reporting at least 90% of days over the period of 1998–2015 are included. The data are frequently updated and can be obtained freely via the web from NCEI.

3. Methodology

a. Extreme precipitation categorization scheme

Extreme precipitation totals are grouped into five categories, or “P-Cats,” according to their overall accumulated 3-day total. P-Cats are defined as follows using even 100-mm thresholds as intuitive bounds on each category. A 3-day total between 100 and 199 mm is assigned to P-Cat 1, between 200 and 299 mm to P-Cat 2, between 300 and 399 mm to P-Cat 3, between 400 and 499 mm to P-Cat 4, and greater than 500 mm to P-Cat 5 (Fig. 1). Three-day totals are defined as the sum of the accumulated precipitation for that day and the two preceding days such that if a P-Cat 4 is recorded on 4 January at a given location, the precipitation accumulated over 2–4 January totaled between 400 and 499 mm. This window is then moved forward by one day each time step so that the 3-day total for each day includes the sum of that day and the previous two. The P-Cat approach is a slightly modified version of the rainfall category or “R-Cat” approach introduced by Ralph and Dettinger (2012). The R-Cat scheme is used operationally by the Scripps Institute of Oceanography Center for Western Weather and Water Extremes (<http://cw3e.ucsd.edu/>) to categorize discrete rainfall events associated with AR landfalls over California. Here we use the term “P-Cat” to clarify that this scheme is not only geared toward rainfall, hence the more general “precipitation.” While similar to the R-Cat scale, the P-Cat approach offers an intuitive way to interpret and visualize extreme precipitation climatology across the CONUS applied as an indicator of climate change and variability. Our P-Cats 2–5 are the same as R-Cats 1–4, however we introduce a lower category to capture a wider

geography of extreme precipitation and a greater diversity of the associated meteorological mechanisms. Multiday totals have been suggested as highly relevant to regional hydrologic impacts including flooding and landslides (Ralph and Dettinger 2012). Furthermore, Ralph and Dettinger (2012) indicate that the 3-day window provides the best representation of major storms, with 2-day totals missing storms and 4-day periods revealing negligible differences to 3-day periods.

We note that in using a fixed threshold we are capturing the most extreme 3-day totals defined relative to the CONUS, rather than relative to the climatology of the grid point or station where the P-Cat occurs. As such, some dry portions of the CONUS do not observe P-Cat events during the time period of our analysis while other wetter places experience relatively frequent P-Cats. While this can be viewed as a caveat, the set of fixed thresholds provides an intuitive way to view extreme precipitation climatology and track change in the magnitude of extreme precipitation over space and time. Furthermore, while not applied in this study, variants on the P-Cat approach could be developed that are regionally specific or customized for different datasets. In that sense, the threshold approach can also carry potential for novel climate model evaluation of extreme precipitation and assessment of projections of future changes.

b. Dataset comparison

To assess the effect of observational uncertainty on using the P-Cat approach we compare the magnitude and frequency of P-Cats across a five-dataset suite. Magnitude is assessed by comparing the maximum observed P-Cat at each data point while frequency is examined both through total P-Cat occurrence as well as the average number of P-Cats observed per year or season. Dataset comparisons are performed and summarized over the CONUS as well as over the seven multistate defined NCA regions (Fig. 2; Easterling et al. 2017). All comparison analyses are performed at the annual and seasonal scales with winter defined as December–February (DJF), spring as March–May (MAM), summer as June–August (JJA), and fall as September–November (SON). Comparisons are performed over the period of maximum overlap across all datasets, 1998–2015. Additionally, IMERG is compared with TMPA for the years of overlap (2014–15). In all analyses involving GHCN-D, the station data are used only for qualitative comparison to what can be considered ground truth.

Results for all the datasets are presented both on their native grid and a common grid for comparison to assess the effect of spatial scale on P-Cat frequency and magnitude. Gridded datasets were rescaled, prior to assigning P-Cats, to a common $0.5^\circ \times 0.625^\circ$ grid over the CONUS.

FIG. 2. The seven NCA subregions and the associated abbreviations.

This resolution matches that of the coarsest resolution product included in the study, MERRA-2. To rescale each gridded product, the first-order conservative remapping technique introduced in Jones (1999) was used. Conservative remapping acts to maintain the areal average (Chen and Knutson 2008), unlike alternate methods such as bilinear, bicubic, or distance weighted, and has been used in a number of studies (e.g., Nikulin et al. 2012; Kalognomou et al. 2013; Diaconescu et al. 2015). The spatial correspondence between the patterns of the regridded results are quantitatively summarized using Taylor diagrams, in terms of the centered root-mean-square difference (CRMSD), standard deviation, and correlation coefficient (Taylor 2001). To construct a Taylor diagram, one dataset must be chosen as the reference to measure dataset similarities and differences against. In all Taylor diagrams here, PRISM is used as the reference dataset, chosen because it is the only gridded dataset based primarily on gauge data; however, this is not to say that PRISM is without bias.

4. Results

a. Annual precipitation climatology

As a first-order comparison of dataset precipitation climatology, annual mean precipitation is shown for each dataset on its native grid in Fig. 3. All datasets show similar general climatology patterns, however, using GHCN-D as a reference (Fig. 3a), considerable differences across the datasets emerge. First-order differences relate to the representation of the effect of topography on precipitation, with the high-resolution PRISM (Fig. 3b) best resembling GHCN-D over the mountainous West and the lowest resolution MERRA-2 (Fig. 3e) showing the least detail. TMPA also has a notable dry bias relative to GHCN-D across the mountains of the Northwest despite its relatively

FIG. 3. Average annual precipitation over the period of 1998–2015. Results are for (a) GHCN-D, (b) PRISM, (c) TMPA, (d) NARR, and (e) MERRA-2.

high spatial resolution (Fig. 3c), likely due to limitations in the ability of TMPA to measure snowfall (Bharti and Singh 2015). NARR (Fig. 3d) has a broad dry bias over much of the Southeast compared with GHCN-D and the other three datasets. MERRA-2 is too coarse to resolve most details of individual mountain ranges; however, it does show some qualitative similarities with GHCN-D over the coastal Northwest and northern Rocky Mountains.

b. Maximum P-Cats

The maximum recorded P-Cats are presented for the full year (Fig. 4), for DJF (Fig. 5), and for SON (Fig. 6). Fall and winter are chosen for seasonal analysis because they are concurrent with the most widespread occurrence of heavy precipitation, spanning two primary meteorological mechanisms consistent with the findings in Kunkel et al. (2012): atmospheric rivers often associated with extratropical

cyclones in the West in both seasons (Neiman et al. 2008a,b; Ralph and Dettinger 2011, 2012) and tropical systems in the Southeast in the fall (Knight and Davis 2009; Knutson et al. 2010; Kunkel et al. 2010). Results are summarized across seasons and subregions using Taylor diagrams in Fig. 7.

The spatial distribution of the maximum observed P-Cats in GHCN-D (Fig. 4a, analogous to Fig. 3 from Ralph and Dettinger 2012) generally resembles the precipitation climatology in Fig. 3, with the highest P-Cats coinciding with the highest annual rainfall. This is supported in the West by the prevalence of high-end P-Cats across the coastal mountain ranges, the Sierra Nevada and Cascade ranges, and the Transverse Ranges of Southern California. High-end P-Cats are also more prevalent in the Southeast stretching from Texas eastward to the Carolinas. The maximum P-Cats recorded

FIG. 4. Maximum observed P-Cat at each grid point over the 1998–2015 period. (a)–(d) Maximum P-Cats on native grid and (e)–(h) maximum P-Cats on common MERRA-2 grid. Spatially interpolated datasets are indicated with an asterisk. Color scale is as in Fig. 1. Unshaded grid cells indicate that no P-Cat has occurred during the data record.

during this period are generally lower across the Great Plains, the desert Southwest, and the interior western rain shadows.

All datasets capture the general pattern of relatively high P-Cats in the western mountains and Southeast,

and low P-Cats over the Great Plains and Southwest. However, considerable differences are apparent in P-Cat extent and magnitude. For example, PRISM shows the most widespread P-Cats 4 and 5, likely due at least in part from it having the finest grid resolution

FIG. 5. As in Fig. 4, but for SON.

and being constructed using gauge data. PRISM also shows a multitude of high-end P-Cats over the Southeast, which the other datasets do not capture, possibly indicative of localized convective precipitation that can be captured by the relatively dense gauge network used to construct PRISM here. TMPA (Fig. 4c) also captures a greater occurrence of high-end

extremes compared to NARR and MERRA-2 (Figs. 4d,e).

While regridding reduces some of the P-Cat magnitudes through spatial smoothing, some differences persist (right column of Fig. 4; i.e., regridded to MERRA-2 resolution). In the case that high resolution is necessary for capturing processes leading to extreme

FIG. 6. As in Fig. 4, but for DJF.

precipitation (Herold et al. 2017), such as localized convection, then it is possible that a high-resolution dataset will maintain some high-end totals compared with the coarser products. Potentially illustrative of this effect, PRISM maintains a relatively high number of P-Cats 2–4 after regridding (Fig. 4f). The same effect is apparent for TMPA over the Southeast and Northwest. In

addition to spatial resolution, other factors may also be important in determining the level of agreement after interpolation, including differences in the ability of the analysis products to accurately capture land–atmosphere interaction or potential bias and overestimation in PRISM (Mesinger et al. 2006; Bharti and Singh 2015; Molod et al. 2015).

FIG. 7. Taylor diagrams quantifying the spatial correspondence of the maximum observed P-Cats for TMPA, MERRA-2, and NARR relative to PRISM. Results are for (a) the CONUS annually and seasonally, (b) relevant NCA subregions annually, (c) SON for the NCA subregions, and (d) DJF for the NCA subregions. Each dataset is labeled by a symbol with each season and subregion assigned a color as defined in the legends in the top two panels. The x and y axes correspond to the standard deviation ratio between the indicated dataset and PRISM. The radial axis is the pattern correlation, and the distance between the symbol and the PRISM location is proportional to the CRMSD between the spatial field of the maximum P-Cats of the indicated dataset and PRISM, normalized by the spatial standard deviation of the PRISM field.

The Taylor diagrams in Figs. 7a and 7b summarize the dataset correspondence for the CONUS annually and seasonally and NCA subregions annually, respectively. At the seasonal scale (Fig. 7a), NARR and MERRA-2 show a lower spatial standard deviation across all seasons with TMPA generally exceeding PRISM. TMPA also has a greater spread in pattern correlation resulting in larger CRMSD values compared with NARR and MERRA-2, especially for DJF and MAM. Both NARR and MERRA-2 cluster closely at the CONUS scale across the seasonal cycle. Less spread is apparent at the subregion scale (Fig. 7b) with all datasets revealing similar spatial variance and correlation relative to PRISM.

In SON, the highest observed P-Cats captured by GHCN-D (Fig. 5a) are over the Pacific Northwest, central Texas, and the Gulf and Atlantic Coasts of the Southeast. P-Cats 1 and 2 are common throughout the higher elevations of the West and across the Midwest through the Northeast. Several examples of southwest to northeast oriented bands of P-Cat 2s are apparent in the central United States. For example, one band extends from northern Illinois to southeastern Michigan providing a useful baseline for comparing the details of the other datasets. In many cases, very high-end P-Cats can readily be traced to the contributing storm. For example, the high values over eastern North Carolina are the result of Hurricane Floyd that made landfall in September of

1999, resulting in catastrophic societal impacts (Easterling et al. 2000). The similarities between Figs. 4 and 5 over the Southeast indicate that most of the highest recorded P-Cats occur during SON.

Consistent with our previous findings, PRISM captures the greatest magnitude and spatial extent of high-end totals (Fig. 5b), sharing the most qualitative similarities with the GHCN-D results, including the collocation of the southwest to northeast oriented bands of P-Cat 2s across the Midwest and topographic enhancement in the West. These features are generally captured in the other datasets, however with lower magnitudes. In some cases, regional scale details are not similar across the suite especially in the case of the high-end P-Cats over the Southeast where MERRA-2 and NARR show varying degrees of dissimilarities with the other datasets. As in Figs. 3 and 4, there is a close relationship between spatial resolution and P-Cat magnitude, however even considering a systematic resolution related bias, some fundamental differences persist.

After spatial interpolation, PRISM and TMPA maintain high-end totals over Washington and North Carolina (Figs. 5f,g). MERRA-2 and NARR generally show systematically lower P-Cat magnitudes relative to the regridded PRISM and TMPA, providing further evidence of factors other than resolution being influential on dataset agreement (Figs. 5e,h). In Fig. 7c dataset spread is small between MERRA-2 and NARR, especially across the variance ratio, while TMPA tends to exceed PRISM's spatial variance in most subregions. Note that we omit results for Great Plains North because of its very low number of grid cells with P-Cats.

In DJF (Fig. 6), the overall spatial coverage of stations recording P-Cats is lower than SON, especially across the central United States. GHCN-D shows the most extreme precipitation occurring along the western mountains stretching from northern Washington to southern California and across the southern Midwest and Southeast (Fig. 6a). This is evidence that the intense precipitation from North Pacific extratropical cyclones is maximized by the orographic enhancement of land-falling atmospheric rivers (e.g., Neiman et al. 2008a,b; Guan et al. 2010, 2013; Ralph and Dettinger 2012). Across the eastern half of the CONUS, high-end P-Cats are the result of strong midlatitude cyclones that strengthen along the strong temperature gradients formed by southward excursions of Arctic air masses.

In agreement with GHCN-D, PRISM shows many of the high-end totals that occur across the West (Fig. 6b). TMPA's limitations at capturing snowfall are apparent with considerable underestimation of the magnitude of P-Cats along the Sierra Nevada and Cascades (Fig. 6c). These results are consistent with Behrangi et al. (2014a), emphasizing the inherent challenges associated with

measuring precipitation in remote regions, where station data are sparse, orography and finescale processes are key, and precipitation type limits the utility of TMPA retrievals. Substantial differences in the magnitude of P-Cats captured by NARR and MERRA-2 (Figs. 6d,e) suggest that grid resolution may inhibit the ability of a dataset to capture the impact of localized phenomena, although both datasets capture the broad patterns of P-Cats across the West and Southeast.

While regridding reduces the overall magnitude of P-Cat intensity in PRISM and TMPA, both datasets continue to show more P-Cats 2 and 3. Over the Southeast, resolution does not appear as important at capturing high-end P-Cats, which is consistent with the typical synoptic-scale storms that result in extreme precipitation here in winter. Additionally, this provides evidence that differences across datasets are also driven by dataset construction, and not solely a result of grid resolution. The Taylor diagram in Fig. 7d shows that TMPA exhibits a higher variance relative to PRISM over the Southeast and roughly the same in the Northwest, with all other datasets and subregions showing a slightly lower spatial variance than PRISM and pattern correlation coefficients between 0.9 and 0.99.

c. P-Cat frequency

As for the comparison of P-Cat magnitude in the above section, P-Cat frequency, computed as P-Cats per year or season, is compared across the entire year (Fig. 8), for SON (Fig. 9), and for DJF (Fig. 10). Differences across the data suite are also presented as biases, with reference to PRISM. Results are further summarized using Taylor diagrams in Fig. S1 in the online supplemental material.

The highest annual frequency of P-Cats in GHCN-D (Fig. 8a) generally corresponds spatially to the highest magnitude P-Cats in Fig. 4a. These areas include the Southeast and the mountains of the Pacific Northwest and California where annual P-Cat frequency exceeds 20. For reference, if a station shows an average frequency of 20 P-Cats per year, this would mean that on average 20 days of every year are part of a 3-day precipitation total that exceeds 100 mm. In such cases, heavy precipitation is relatively common and the simple occurrence of a P-Cat may not necessarily be considered highly extreme in a local climatological context. In contrast, a large swath of the eastern half of the domain experiences between 2 and 8 P-Cats annually, while P-Cats are infrequent across the High Plains and the inland West. P-Cats 1 and 2 make up the vast majority of P-Cats per year CONUS-wide with some areas of the West and Southeast recording as many as two high-end P-Cats per year (not shown).

All datasets capture similar principal spatial patterns of annual P-Cat frequency. Qualitatively, PRISM (Fig. 8e) most closely resembles GHCN-D, even capturing many

FIG. 8. (a)–(e) Mean annual frequency of P-Cat occurrence over the 1998–2015 record and (f)–(h) the difference between the mean annual frequency of P-Cat occurrence in the indicated dataset and PRISM. Frequencies are recorded as the number of P-Cats per year. Spatially interpolated datasets are indicated with an asterisk.

FIG. 9. As in Fig. 8, but for SON.

of the small-scale features in areas of complex terrain and regional variations in the Southeast. TMPA (Fig. 8f) shows notable positive frequency bias across the eastern half of the CONUS and over the valleys of the coastal Northwest with lower frequencies across the western

mountains, compared with PRISM. NARR and MERRA-2 both share similarities, with systematically lower P-Cat occurrence compared with PRISM after regridding. NARR shows a greater frequency of P-Cats across the Sierra Nevada compared with MERRA-2, however both

FIG. 10. As in Fig. 8, but for DJF.

datasets show considerable negative frequency biases across most of the West.

During SON (Fig. 9), GHCN-D shows the highest frequency of P-Cat occurrence in the Northwest and

Southeast with values exceeding 10 P-Cats per season along the coasts of Washington and Oregon and between 2 and 4 in southeast Texas and southwest Louisiana (Fig. 9a). This indicates that at least 4 days per fall are part

of a 100-mm or greater 3-day precipitation total on average in these places. There are many commonalities between the frequency map in Fig. 9a and the maximum P-Cat map in Fig. 5a, with many of the regions that experience high values of one also experiencing high values of the other. However, this is not always the case in some parts of the South and along the Atlantic Coast of Florida where P-Cats are common, but rarely exceed P-Cat 2.

Consistent with expectations based on the above results, the observation-based TMPA and PRISM (Figs. 9b,e) share the most similarities with GHCN-D. PRISM captures the overall spatial patterns and frequency magnitudes, but it is also capable of resolving small scale features such as higher frequencies over southeastern Texas. Over the Northwest, as in other analyses, TMPA's limitation at capturing frozen precipitation likely contributes to its negative biases over the mountains (Behrangi et al. 2014a), however it shows a weak positive frequency bias across the lower elevations of the coastal Northwest. NARR and MERRA-2 resemble each other with systematic low frequency bias across the CONUS (Figs. 8f–h).

During winter (Fig. 10a), the P-Cat frequencies are highest across the mountains of Washington, Oregon, and California with elevated P-Cat frequencies also occurring in the higher elevations of Idaho, Utah, and Arizona. In contrast, the other area of high P-Cat occurrence is a broad swath of the South and southern Midwest where Gulf of Mexico moisture fuels heavy precipitation associated with midlatitude cyclones. PRISM (Fig. 10b) captures the mountain ranges across the West and the general pattern in the East (Fig. 10e). However, it underestimates the isolated high frequency P-Cats that GHCN-D captures over the higher terrain of Idaho and Utah. TMPA (Fig. 10b) resembles both PRISM and GHCN-D, but with substantial high frequency biases over the lower elevations of the West Coast and throughout the Southeast (Fig. 10f). A physical explanation for this widespread bias in TMPA is unclear as it is not consistent with findings from other seasons or at the annual scale. TMPA also shows negative biases along the immediate Pacific coast, suggesting frozen precipitation is not the only contributor to underestimation in the West. NARR and MERRA-2 are quite similar with overall negative frequency biases across the CONUS with the exception of some western valleys.

d. Annual P-Cat occurrence

Figures 11–13 show spatially aggregated P-Cat frequencies over time. Here we only show results for annual frequency at the CONUS scale, for DJF over the Northwest and for SON over the Southeast to capture the regions and corresponding seasons where high-end P-Cats are most common. In each figure, the left column

shows the number of P-Cats per category on the native grid of each dataset, while the right column represents the datasets interpolated to the MERRA-2 grid. This means that all things equal, prior to regridding the coarser resolution datasets will have a lower frequency of P-Cat occurrence, simply because there are more data points in the high-resolution cases. In this sense, the left column is intended for qualitative comparison while the right column compares datasets with an equal number of data points.

For most years the full range of P-Cats occurs somewhere over the CONUS according to GHCN-D (Fig. 11a). There is no apparent systematic trend in the frequency of any P-Cat occurrence across the CONUS and the datasets. Comparing each dataset to GHCN-D, datasets tend to show a similar evolution of interannual variability. For example, the year 2000 shows a relative minimum in P-Cat 2s in all datasets. Consistent with results from Figs. 4–10, high-end P-Cats are most common in PRISM (Fig. 11b) while they are uncommon in NARR and MERRA-2 (Figs. 11d,e). When compared on a common grid, P-Cat 1 frequencies are more comparable across the suite. PRISM (Fig. 11f) maintains a number of P-Cats 3 and 4 after regridding. The coefficients of variation for each P-Cat time series, computed as the standard deviation of each dataset's annual frequency divided by its mean, are recorded in Table 2. All datasets show a greater year-to-year variability in higher-end P-Cats relative to lower-end P-Cats. For example, GHCN-D has a coefficient of variation for the annual frequency of P-Cat 5 that greatly exceeds that of P-Cat 1.

During SON over the Southeast (Fig. 12), GHCN-D shows a high number of P-Cats 4 and 5 occurring during 1998 and 1999 (Fig. 12a) with considerable interannual variability throughout the record. PRISM (Fig. 12b) continues to show the greatest number of high-end P-Cats compared with the other datasets. TMPA also captures higher-end P-Cats in the early part of the record (Fig. 12c), including 1999. NARR and MERRA-2 (Figs. 12d,e) show primarily P-Cats 1 and 2, with MERRA-2 showing some P-Cat 3s in 1998 and 1999 suggesting that it realistically represents the high-end totals captured in the finer resolution datasets but with diminished magnitude. This reduced magnitude of extremes likely results in part from the coarser reanalysis resolution, but differences may also stem from the dataset generating algorithms. When compared on a common grid, dataset agreement is much stronger, although NARR stands out as having the lowest P-Cat occurrence, and datasets capture similar interannual variability. The coefficient of variation results continue to show greater variability among the most extreme P-Cats across the five-dataset suite (Table 3).

FIG. 11. Annual frequency of observed P-Cats over the 1998–2015 period. (a)–(d) Annual frequency of P-Cats on native grid and (e)–(h) annual frequency of P-Cats on common grid. Spatially interpolated datasets are indicated with an asterisk. Gray bars represent P-Cat 1, green P-Cat 2, yellow P-Cat 3, orange P-Cat 4, and red P-Cat 5 as in the legend in Fig. 1. Results are plotted on a log scale.

FIG. 12. As in Fig. 11, but for SON and only over the Southeast subregion.

FIG. 13. As in Fig. 11, but for DJF and only over the Northwest subregion.

TABLE 2. Dataset's coefficient of variation values for each P-Cat's annual frequency across the CONUS.

	P-Cat 1	P-Cat 2	P-Cat 3	P-Cat 4	P-Cat 5
GHCN-D	0.1506	0.3602	0.4901	0.9164	1.3614
PRISM	0.2029	0.4466	0.7060	0.9800	1.2972
TPMA	0.2834	0.7035	1.4352	2.2687	4.2426
NARR	0.256	0.6584	1.3625		
MERRA-2	0.3701	0.9433	2.7955	4.2426	

Resolving topography is important for capturing P-Cats in DJF in the Northwest subregion (Fig. 13). GHCN-D and PRISM (Figs. 13a,b) show the most qualitative agreement, including with interannual variability, with NARR also sharing commonalities in year-to-year fluctuations (Fig. 13d). When compared on common grids, overall magnitudes of P-Cat 1 are in reasonable agreement across the suite, however interannual variability is still somewhat different in TPMA (Fig. 13g) compared with PRISM and NARR (Figs. 13f,h). These results further suggest using caution when measuring and monitoring extreme precipitation across areas of complex terrain where orographic effects on precipitation are key and extremes are often associated with frozen precipitation. The datasets annual P-Cat frequency results for DJF in the Northwest continue to show greater variability as the P-Cats increase (Table 4).

e. Comparison of individual storms

While the primary aim of implementing the P-Cat scheme in this study is to track and describe extreme precipitation climatology at the grid point scale, as another way to intercompare the five datasets and as a way to further demonstrate the reliability of the P-Cat approach in capturing extreme precipitation, we show the P-Cat values associated with individual historically impactful storms. Figure 14 shows four examples while additional examples are provided in Figs. S2–S4. Examples were chosen to capture a wide range of storm types occurring across a diverse range of geographic areas. The top row shows the P-Cat values associated with the landfall of Hurricane Floyd in September of 1999. Note that the P-Cat values are based on the 3-day rainfall totals ending on the date specified to the right of each row. All datasets capture high-end P-Cats (3–5) over a similar region, while NARR shows relatively modest P-Cats. This indicates that all datasets with the exception of NARR are capable of capturing the magnitude of rainfall associated with this intense tropical system.

In the second row from the top, P-Cats from a notable atmospheric river event from November 2006 are shown with generally good qualitative agreement across the datasets despite the coarser datasets showing overall lower

TABLE 3. As in Table 2, but for SON and only over the Southeast subregion.

	P-Cat 1	P-Cat 2	P-Cat 3	P-Cat 4	P-Cat 5
GHCN-D	0.4325	0.7217	1.1422	1.856	2.7255
PRISM	0.4839	0.7796	1.4984	1.8646	2.2129
TPMA	0.5626	1.0129	2.2725	3.2571	4.2426
NARR	0.5970	1.0043	3.0870		
MERRA-2	0.6522	1.2568	2.9218	4.2426	

P-Cat magnitudes. There is some indication, however, that PRISM overestimates the magnitude of rainfall over the northern Oregon Coast Range and the Olympic Mountains. The third row from the top compares P-Cats for an intense winter storm that occurred during December 2015. All datasets capture the swath of P-Cats 1 and 2 extending from northeast Texas into central Illinois indicating reasonable qualitative agreement in the magnitude and extent of heavy precipitation from this powerful winter storm. Last, the bottom row shows a comparison for a strong mesoscale convective system that occurred in September 2004. While all datasets can capture the convective precipitation to the extent that it surpasses the P-Cat 1 threshold, TPMA shows a relatively larger area of P-Cats compared with the other datasets suggesting some difference in how it captures convective precipitation in this event. We also note that the P-Cat approach is capable of visualizing the propagation of weather events that produce heavy precipitation totals. Figure S5 shows 5-day evolutions of P-Cats for two hurricanes and one winter storm. The P-Cats show the temporal, spatial, and intensity evolution of the storms highlighting the efficacy of the 3-day total approach to capture observed heavy precipitation events.

f. IMERG intercomparison

Considering the potential benefits of using remote sensing to continuously monitor and track extreme precipitation over time, we compare IMERG data to its predecessor, TPMA in Fig. 15. IMERG has only been online for a short time, prohibiting a comprehensive climatology intercomparison. We therefore leverage the existing overlap period (April 2014–December 2015), as of the writing of paper, using the maximum observed P-Cats as well as total observed P-Cat frequency for comparison. Over this 2-yr period, there is some indication that IMERG captures more small-scale features and better represents extremes over the mountainous West (Fig. 15c). These results are likely attributable, at least in part, to IMERG's higher spatial resolution, but may also be due to improvements in GPM sensors to measure snow (Hou et al. 2014). This qualitatively brings IMERG closer to GHCN-D with

TABLE 4. As in Table 2, but for DJF and only over the Northwest subregion.

	P-Cat 1	P-Cat 2	P-Cat 3	P-Cat 4	P-Cat 5
GHCN-D	0.4082	0.6106	1.4798	2.4589	4.2426
PRISM	0.4393	0.6483	0.9957	1.5989	1.9965
TMPA	0.8906	1.7957	4.2426		
NARR	0.5802	1.3096	2.7069		
MERRA-2	0.6997	2.5205			

exceptions. For example, IMERG does a poorer job at capturing the band of P-Cat 2s stretching from north-east Texas through Missouri compared with TMPA and overestimates P-Cat magnitude over eastern Tennessee and northern Alabama. P-Cat frequencies reveal similarities between TMPA and IMERG (Fig. 15d).

5. Summary and conclusions

Here we present a climatology of a categorization scheme for monitoring and tracking change in extreme precipitation over space and time and assess its observational uncertainty. The approach assigns a category

between one and five to 3-day storm totals (Fig. 1). Intended as a way to track extreme precipitation as a climate indicator, this scheme provides a platform for monitoring change in extreme precipitation across scales, datasets, time, and geography. However, precipitation observation products are all associated with some degree of bias, making it important to understand and attempt to constrain observational uncertainty when analyzing extremes. To demonstrate the utility of the P-Cat scheme as a way to track extreme precipitation events in time and space and to highlight the importance of understanding observational uncertainty, we apply the 3-day total categorization as a basis for a dataset intercomparison across four gridded products, spanning a range of construction methodologies, and in situ station data.

All gridded datasets capture the principal spatial patterns of mean annual precipitation climatology, with higher resolution datasets capturing more orographic features than the lower resolution datasets (Fig. 3). Focusing on extremes, the magnitude (Figs. 4–7) and frequency of P-Cats (Figs. 7–11) are assessed using the P-Cat scheme as a metric for intercomparison.

FIG. 14. P-Cat values for four individual storm events. The color scheme follows previous figures. P-Cat values are computed as the 3-day precipitation total ending on the day indicated at the right of each row. Storms include a tropical cyclone (Floyd), an atmospheric river, a synoptic scale midlatitude winter cyclone, and a mesoscale convective system (MCS).

FIG. 15. Maximum observed P-Cats during April 2014–December 2015, the TMPA/IMERG overlap period. Results are for (a) GHCN-D, (b) IMERG, (c) TMPA, (d) annual frequency per grid cell/station observed over time record.

In general, the higher resolution datasets more closely resemble gauge data across the CONUS and seasons. Specifically, PRISM shares many detailed commonalities with station data while the next highest resolution dataset, TMPA, is also similar overall. NARR and MERRA-2 reanalysis show systematically lower magnitude and frequency of P-Cats across the CONUS and seasonal cycle. TMPA shows systematically lower P-Cat magnitudes and frequencies across the mountains of the West during fall and winter when a large portion of precipitation falls as snow, consistent with known limitations of TMPA at capturing frozen precipitation.

When all datasets are interpolated to a common coarser grid, differences persist but are reduced. In particular, the datasets that were originally constructed at the highest spatial resolution often maintain the highest magnitude of P-Cats, even after coarsening of the gridded data. This feature could result from a number of factors; however, one likely contributor is the fact that a dataset constructed originally at fine resolution is able to capture extreme events that simply could not be resolved at coarser grids (Herold et al. 2017). This may be particularly acute in areas of complex topography where, for example, PRISM is able to resolve local high magnitude events that the other datasets are simply not capable of capturing. Other factors could include other underlying biases in the dataset stemming

from factors such as spatial and temporal heterogeneity in gauges (Kidd et al. 2017), sensor sensitivity to precipitation type (Behrangi et al. 2012, 2014b) or methods of retrieving precipitation from individual sensors in satellites (Kummerow et al. 2011), interpolation methods or misrepresentation due to the sparseness of the observing network (Min et al. 2011), or general deficiencies and model limitations simulating precipitation amounts in reanalysis (Kharin et al. 2013). The annual occurrence of P-Cats shows similar differences across the suite, with a general positive relationship between grid resolution and the number of P-Cats (Figs. 11–13). Preliminary assessment of IMERG, the follow-on satellite product to TMPA, suggests some potential improvements over TMPA in capturing frozen precipitation and finescale extremes (Fig. 15). Ultimately, results suggest satellite data show promise in capturing the overall patterns of heavy precipitation climatology, which could lead to improved monitoring in regions of sparse ground observations.

We acknowledge some assumptions and limitations in use of the P-Cat scheme as a climate indicator for extreme precipitation. First, the use of fixed thresholds for the entire CONUS is intended to highlight the heaviest precipitation across the domain in an intuitive way. As such, some drier regions do not record P-Cats as defined in this study, even though smaller totals may be considered

impactful relative to the local climatology. The synoptic scale of measurement also captures totals at a temporal scale often associated with impacts such as flooding and landslides (Ralph and Dettinger 2012) but does not distinguish between shorter and longer duration totals. This may be relevant for lower-end P-Cats that could result from short duration extreme convective events. It is also possible that a single storm may be counted more than once due to the moving 3-day window used to construct the P-Cat. Finally, while we include the five datasets here in an effort to capture a range of measurement methods while focusing on high resolution products, this analysis could be extensible to other observations.

Overall, the P-Cat characterization scheme applied here offers several opportunities for future research and applications. By comparing P-Cat climatology in climate model simulations of the historical climate to observations, this scheme could provide a novel target for climate model evaluation. As further extension of the P-Cat approach for dataset intercomparison, P-Cat thresholds could be customized to a dataset's grid resolution to account for the inherently lower magnitude of extremes captured at coarser versus finer resolutions, although this could come with somewhat reduced levels of intuitiveness since a P-Cat 1 would be different for different datasets. The P-Cat scheme could also be used for assessing future projections of changes in extreme precipitation in climate models. Last, the P-Cat approach is easily extensible to other regions, facilitating temporal and spatial tracking and monitoring of extremes, dataset intercomparison, model evaluation, and future change assessment.

Acknowledgments. This work was carried out, in part, at the Jet Propulsion Laboratory, California Institute of Technology and at Portland State University, under a contract from National Aeronautics and Space Administration (NASA). Support for this project was provided by the NASA Indicators for the National Climate Assessment (NCA) Program under NASA award NNX16AG60G. Partial support for Emily Slinkey was provided by a JPL student summer internship. We thank Judah Detzer, Heejun Chang, and Andrew Martin for their help with this project. Additionally, we thank Ali Behrangi for his contributions with satellite observation information.

REFERENCES

- Adler, R. F., C. Kidd, G. Petty, M. Morissey, and H. M. Goodman, 2001: Intercomparison of global precipitation products: The Third Precipitation Intercomparison Project (PIP-3). *Bull. Amer. Meteor. Soc.*, **82**, 1377–1396, [https://doi.org/10.1175/1520-0477\(2001\)082<1377:IOGPPT>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<1377:IOGPPT>2.3.CO;2).
- Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, **111**, D05109, <https://doi.org/10.1029/2005JD006290>.
- Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrological cycle. *Nature*, **419**, 224–232, <https://doi.org/10.1038/nature01092>.
- Behrangi, A., M. Lebsock, S. Wong, and B. Lambrigtsen, 2012: On the quantification of oceanic rainfall using spaceborne sensors. *J. Geophys. Res.*, **117**, D20105, <https://doi.org/10.1029/2012JD017979>.
- , K. Andreadis, J. B. Fisher, F. J. Turk, S. Granger, T. Painter, and N. Das, 2014a: Satellite-based precipitation estimation and its application for streamflow prediction over mountainous western U.S. basins. *J. Appl. Meteor. Climatol.*, **53**, 2823–2842, <https://doi.org/10.1175/JAMC-D-14-0056.1>.
- , Y. Tian, B. H. Lambrigtsen, and G. L. Stephens, 2014b: What does CloudSat reveal about global land precipitation detection by other spaceborne sensors? *Water Resour. Res.*, **50**, 4893–4905, <https://doi.org/10.1002/2013WR014566>.
- , B. Guan, P. J. Neiman, M. Schreier, and B. Lambrigtsen, 2016: On the quantification of atmospheric rivers precipitation from space: Composite assessments and case studies over the eastern North Pacific Ocean and the western United States. *J. Hydrometeorol.*, **17**, 369–382, <https://doi.org/10.1175/JHM-D-15-0061.1>.
- Bharti, V., and C. Singh, 2015: Evaluation of error in TRMM 3B42V7 precipitation estimates over the Himalayan region. *J. Geophys. Res. Atmos.*, **120**, 12 458–12 473, <https://doi.org/10.1002/2015JD023779>.
- Bosilovich, M. G., J. Chen, F. R. Robertson, and R. F. Adler, 2008: Evaluation of global precipitation in reanalyses. *J. Appl. Meteor. Climatol.*, **47**, 2279–2299, <https://doi.org/10.1175/2008JAMC1921.1>.
- Bukovsky, M. S., and D. J. Karoly, 2007: A brief evaluation of precipitation from the North American Regional Reanalysis. *J. Hydrometeorol.*, **8**, 837–846, <https://doi.org/10.1175/JHM595.1>.
- Chen, C.-T., and T. Knutson, 2008: On the verification and comparison of extreme rainfall indices from climate models. *J. Climate*, **21**, 1605–1621, <https://doi.org/10.1175/2007JCLI1494.1>.
- Chen, S., and Coauthors, 2013: Evaluation of the successive V6 and V7 TRMM Multisatellite Precipitation Analysis over the continental United States. *Water Resour. Res.*, **49**, 8174–8186, <https://doi.org/10.1002/2012WR012795>.
- Daly, C., 2006: Guidelines for assessing the suitability of spatial climate data sets. *Int. J. Climatol.*, **26**, 707–721, <https://doi.org/10.1002/joc.1322>.
- , R. P. Neilson, and D. L. Phillips, 1994: A statistical–topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158, [https://doi.org/10.1175/1520-0450\(1994\)033<0140:ASTMFM>2.0.CO;2](https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2).
- , W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. *Climate Res.*, **22**, 99–113, <https://doi.org/10.3354/cr022099>.
- , —, —, M. K. Doggett, and J. I. Smith, 2007: Observer bias in daily precipitation measurements at United States cooperative network stations. *Bull. Amer. Meteor. Soc.*, **88**, 899–912, <https://doi.org/10.1175/BAMS-88-6-899>.
- , M. Hablbeib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031–2064, <https://doi.org/10.1002/joc.1688>.
- Diaconescu, E. P., P. Gachon, and R. Laprise, 2015: On the remapping procedure of daily precipitation statistics and indices

- used in regional climate model evaluation. *J. Hydrometeorol.*, **16**, 2301–2310, <https://doi.org/10.1175/JHM-D-15-0025.1>.
- Donat, M. G., L. V. Alexander, H. Yang, I. Durre, R. Vose, and J. Caesar, 2013: Global land-based datasets for monitoring climate extremes. *Bull. Amer. Meteor. Soc.*, **94**, 997–1006, <https://doi.org/10.1175/BAMS-D-12-00109.1>.
- Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Robust automated quality control of daily surface observations. *J. Appl. Meteor. Climatol.*, **49**, 1615–1633, <https://doi.org/10.1175/2010JAMC2375.1>.
- Easterling, D. R., J. L. Evans, P. Ya Groisman, T. R. Karl, K. E. Kunkel, and P. Ambenje, 2000: Observed variability and trends in extreme climate events: A brief review. *Bull. Amer. Meteor. Soc.*, **81**, 417–425, [https://doi.org/10.1175/1520-0477\(2000\)081<0417:OVATIE>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<0417:OVATIE>2.3.CO;2).
- , and Coauthors, 2017: Precipitation change in the United States. *Climate Science Special Report: Fourth National Climate Assessment*, Vol. 1, D. J. Wuebbles et al., Eds., U.S. Global Change Research Program, 207–230, <https://doi.org/10.7930/JOH993CC>.
- Frich, P., L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. M. G. Klein Tank, and T. Peterson, 2002: Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Res.*, **19**, 193–212, <https://doi.org/10.3354/cr019193>.
- Gao, Y., J. Lu, L. R. Leung, Q. Yang, S. Hagos, and Y. Qian, 2015: Dynamical and thermodynamical modulations on future changes of landfalling atmospheric rivers over western North America. *Geophys. Res. Lett.*, **42**, 7179–7186, <https://doi.org/10.1002/2015GL065435>.
- Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- Gleason, K. L., J. H. Lawrimore, D. H. Levinson, and T. R. Karl, 2008: A revised U.S. climate extremes index. *J. Climate*, **21**, 2124–2137, <https://doi.org/10.1175/2007JCLI1883.1>.
- Guan, B., N. P. Molotch, D. E. Waliser, E. J. Fetzer, and P. J. Neiman, 2010: Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements. *Geophys. Res. Lett.*, **37**, L20401, <https://doi.org/10.1029/2010GL044696>.
- , —, —, and —, 2013: The 2010/2011 snow season in California's Sierra Nevada: Role of atmospheric rivers and modes of large-scale variability. *Water Resour. Res.*, **49**, 6731–6743, <https://doi.org/10.1002/wrcr.20537>.
- Guirguis, K. J., and R. Avisar, 2008: A precipitation climatology and dataset intercomparison for the western United States. *J. Hydrometeorol.*, **9**, 825–841, <https://doi.org/10.1175/2008JHM832.1>.
- Herold, N., A. Behrangi, and L. V. Alexander, 2017: Large uncertainties in observed daily precipitation extremes over land. *J. Geophys. Res. Atmos.*, **122**, 668–681, <https://doi.org/10.1002/2016JD025842>.
- 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 D. T. Bolvin, 2015: TRMM and other data precipitation data set documentation. NASA TRMM Doc., 44 pp., http://pmm.nasa.gov/sites/default/files/imce/3B42_3B43_doc_V7.pdf.
- , and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55, <https://doi.org/10.1175/JHM560.1>.
- , R. F. Adler, D. T. Bolvin, and G. Gu, 2010: The TRMM Multisatellite Precipitation Analysis (TMPA). *Satellite Applications for Surface Hydrology*, F. Hossain and M. Gebremichael, Eds., Springer, 3–22.
- , and Coauthors, 2017: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 4.6, 28 pp., https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.6.pdf.
- Jones, P. W., 1999: First- and second-order conservative remapping schemes for grids in spherical coordinates. *Mon. Wea. Rev.*, **127**, 2204–2210, [https://doi.org/10.1175/1520-0493\(1999\)127<2204:FASOCR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2).
- Kalognomou, E., and Coauthors, 2013: A diagnostic evaluation of precipitation in CORDEX models over southern Africa. *J. Climate*, **26**, 9477–9506, <https://doi.org/10.1175/JCLI-D-12-00703.1>.
- Kharin, V. V., F. W. Zwiers, X. Zhang, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, **119**, 345–357, <https://doi.org/10.1007/s10584-013-0705-8>.
- Kidd, C., A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum, 2017: So, how much of the Earth's surface is covered by rain gauges? *Bull. Amer. Meteor. Soc.*, **98**, 69–78, <https://doi.org/10.1175/BAMS-D-14-00283.1>.
- Knight, D. B., and R. E. Davis, 2009: Contribution of tropical cyclones to extreme rainfall events in the southeastern United States. *J. Geophys. Res.*, **114**, D23102, <https://doi.org/10.1029/2009JD012511>.
- Knutson, T. R., and Coauthors, 2010: Tropical cyclones and climate change. *Nat. Geosci.*, **3**, 157–163, <https://doi.org/10.1038/ngeo779>.
- Kummerow, C. D., S. Ringerud, J. Crook, D. Randel, and W. Berg, 2011: An observationally generated a priori database for microwave rainfall retrievals. *J. Atmos. Oceanic Technol.*, **28**, 113–130, <https://doi.org/10.1175/2010JTECHA1468.1>.
- Kunkel, K. E., D. R. Easterling, D. A. R. Kristovich, B. Gleason, L. Stoecker, and R. Smith, 2010: Recent increases in U.S. heavy precipitation associated with tropical cyclones. *Geophys. Res. Lett.*, **37**, L24706, <https://doi.org/10.1029/2010GL045164>.
- , —, D. A. Kristovich, B. Gleason, L. Stoecker, and R. Smith, 2012: Meteorological causes of the secular variations in observed extreme precipitation events for the conterminous United States. *J. Hydrometeorol.*, **13**, 1131–1141, <https://doi.org/10.1175/JHM-D-11-0108.1>.
- , and Coauthors, 2013: Monitoring and understanding trends in extreme storms: State of knowledge. *Bull. Amer. Meteor. Soc.*, **94**, 499–514, <https://doi.org/10.1175/BAMS-D-11-00262.1>.
- Lamjiri, M. A., M. D. Dettinger, F. M. Ralph, and B. Guan, 2017: Hourly storm characteristics along the US West Coast: Role of atmospheric rivers in extreme precipitation. *Geophys. Res. Lett.*, **44**, 7020–7028, <https://doi.org/10.1002/2017GL074193>.
- Lin, Y., K. E. Mitchell, E. Rogers, M. E. Baldwin, and G. I. DiMego, 1999: Test assimilations of the real-time, multi-sensor hourly precipitation analysis into the NCEP Eta model. Preprints, *Eighth Conf. on Mesoscale Meteorology*, Boulder, CO, Amer. Meteor. Soc., 341–344.
- Liu, Z., 2016: Comparison of Integrated Multisatellite Retrievals for GPM (IMERG) and TRMM Multisatellite Precipitation

- Analysis (TMPA) monthly precipitation products: Initial results. *J. Hydrometeorol.*, **17**, 777–790, <https://doi.org/10.1175/JHM-D-15-0068.1>.
- , D. Ostrenga, W. Teng, and S. Kempler, 2012: Tropical Rainfall Measuring Mission (TRMM) precipitation data and services for research and applications. *Bull. Amer. Meteor. Soc.*, **93**, 1317–1325, <https://doi.org/10.1175/BAMS-D-11-00152.1>.
- Mahoney, K., and Coauthors, 2016: Understanding the role of atmospheric rivers in heavy precipitation in the southeast United States. *Mon. Wea. Rev.*, **144**, 1617–1632, <https://doi.org/10.1175/MWR-D-15-0279.1>.
- Menne, M. J., I. Durre, B. G. Gleason, T. G. Houston, and R. S. Vose, 2012: An overview of the Global Historical Climatology Network-Daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, <https://doi.org/10.1175/JTECH-D-11-00103.1>.
- Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, <https://doi.org/10.1175/BAMS-87-3-343>.
- Min, S.-K., X. Zhang, F. W. Zwiers, and G. C. Hegerl, 2011: Human contribution to more-intense precipitation extremes. *Nature*, **470**, 378–381, <https://doi.org/10.1038/nature09763>.
- Molod, A., L. Takacs, M. Suárez, and J. Bacmeister, 2015: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geosci. Model Dev.*, **8**, 1339–1356, <https://doi.org/10.5194/gmd-8-1339-2015>.
- Neiman, P., F. M. Ralph, G. A. Wick, J. D. Lundquist, and M. D. Dettinger, 2008a: Meteorological characteristics and overland precipitation impacts of atmospheric rivers affecting the west coast of North America based on eight years of SSM/I satellite observations. *J. Hydrometeorol.*, **9**, 22–47, <https://doi.org/10.1175/2007JHM855.1>.
- , —, —, Y. H. Kuo, T. K. Wee, and Z. Ma, 2008b: Diagnosis of an intense atmospheric river impacting the Pacific Northwest: Storm summary and offshore vertical structure observed with COSMIC satellite retrievals. *Mon. Wea. Rev.*, **136**, 4398–4420, <https://doi.org/10.1175/2008MWR2550.1>.
- Nikulin, G., and Coauthors, 2012: Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations. *J. Climate*, **25**, 6057–6078, <https://doi.org/10.1175/JCLI-D-11-00375.1>.
- O’Gorman, P. A., and T. Schneider, 2009: The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proc. Natl. Acad. Sci. USA*, **106**, 14 773–14 777, <https://doi.org/10.1073/pnas.0907610106>.
- Pall, P., M. R. Allen, and D. A. Stone, 2007: Testing the Clausius-Clapeyron constraint on changes in extreme precipitation under CO₂ warming. *Climate Dyn.*, **28**, 351–363, <https://doi.org/10.1007/s00382-006-0180-2>.
- , C. M. Patricola, M. F. Wehner, D. A. Stone, C. J. Paciorek, and W. D. Collins, 2017: Diagnosing conditional anthropogenic contributions to heavy Colorado rainfall in September 2013. *Wea. Climate Extremes*, **17**, 1–6, <https://doi.org/10.1016/j.wace.2017.03.004>.
- Prat, O., and B. Nelson, 2013: Precipitation contribution of tropical cyclones in the southeastern United States from 1998 to 2009 using TRMM satellite data. *J. Climate*, **26**, 1047–1062, <https://doi.org/10.1175/JCLI-D-11-00736.1>.
- Ralph, F. M., and M. D. Dettinger, 2011: Storms, floods, and the science of atmospheric rivers. *Eos, Trans. Amer. Geophys. Union*, **92**, 265–266, <https://doi.org/10.1029/2011EO320001>.
- , and —, 2012: Historical and national perspectives on extreme West Coast precipitation associated with atmospheric rivers during December 2010. *Bull. Amer. Meteor. Soc.*, **93**, 783–790, <https://doi.org/10.1175/BAMS-D-11-00188.1>.
- Reichle, R. H., C. S. Draper, Q. Liu, M. Girotto, S. P. P. Mahanama, R. D. Koster, and G. J. M. De Lannoy, 2017: Assessment of MERRA-2 land surface hydrology estimates. *J. Climate*, **30**, 2937–2960, <https://doi.org/10.1175/JCLI-D-16-0720.1>.
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA’s Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, <https://doi.org/10.1175/JCLI-D-11-00015.1>.
- Risser, M. D., and M. F. Wehner, 2017: Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during Hurricane Harvey. *Geophys. Res. Lett.*, **44**, 12 457–12 464, <https://doi.org/10.1002/2017GL075888>.
- Sapiano, M. R., and P. A. Arkin, 2009: An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *J. Hydrometeorol.*, **10**, 149–166, <https://doi.org/10.1175/2008JHM1052.1>.
- Tan, J., W. A. Petersen, and A. Tokay, 2016: A novel approach to identify sources of errors in IMERG for GPM ground validation. *J. Hydrometeorol.*, **17**, 2477–2491, <https://doi.org/10.1175/JHM-D-16-0079.1>.
- Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, <https://doi.org/10.1029/2000JD900719>.
- Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, **84**, 1205–1217, <https://doi.org/10.1175/BAMS-84-9-1205>.
- U.S. EPA, 2016: Climate change indicators in the United States, 2016. 4th ed. EPA 430-R-16-004, 96 pp., www.epa.gov/climate-indicators.
- Wang, S.-Y. S., L. Zhao, and R. R. Gillies, 2016: Synoptic and quantitative attributions of the extreme precipitation leading to the August 2016 Louisiana flood. *Geophys. Res. Lett.*, **43**, 11 805–11 814, <https://doi.org/10.1002/2016GL071460>.
- Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. Klein Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 851–870, <https://doi.org/10.1002/wcc.147>.