Toward merging MOPEX and CAMELS hydrometeorological datasets: compatibility and statistical comparison

Katharine Sink, Tom Brikowski

Department of Sustainable Earth Systems Sciences, University of Texas at Dallas, Richardson, 75048, United States

5   *Correspondence to*: Katharine Sink (katharine.sink@utdallas.edu)

Abstract. This study compares two large hydrometeorological datasets, the Model Parameter Estimation Experiment (MOPEX), and the Catchment Attributes and Meteorology for Large-sample Studies (CAMELS), focusing on 47 shared watersheds within the continental United States. The evaluation spans daily, monthly, seasonal, and annual scales for the overlapping water years of 1981 to 2000. Spatial aggregations are conducted based on Köppen-Geiger climate regions along with annual Budyko evaporative

10   and aridity indices. Results indicate significant differences between the datasets at daily timesteps, highlighting the challenge of high temporal resolution data reconciliation; however, compatibility markedly improves with temporal aggregation at monthly, seasonal, and annual scales. While MOPEX shows a warm bias for temperature and CAMELS shows a wet bias for precipitation, statistical analyses demonstrate that both datasets are representative of climatic conditions and extreme events. Our findings validate the results of previous research employing either dataset. Furthermore, this study serves as a foundation for the merging

15   and extension of MOPEX and CAMELS datasets.

1 Introduction

Comprehensive historical datasets are crucial for investigating and projecting surface water availability given the complex response of watersheds to natural and anthropogenic forcings. In particular, comparative and large sample hydrology (LSH) rely on large datasets comprised of numerous catchments to derive relationships, develop new models and uncertainty estimates, and classify

20   locations that span different climatic and physiographic regions (Addor et al., 2020; Gupta et al., 2014), yet significant discrepancies make combining and comparing such datasets difficult. Indeed Addor et al. (2020) state the "lack of common standards impedes the comparison of basins from different datasets".

This paper explores and attempts to resolve the principal issues confronting the merger of two of the most commonly used LSH

25   datasets for the continental United States, (CONUS), the Model Parameter Estimation Experiment, MOPEX (Duan et al., 2006; Schaake et al., 2006), and the Catchments Attributes and Meteorology for Large-sample Studies, CAMELS (Addor et al., 2017; Newman et al., 2015). In general, there is an abundance of data available for climate variables, streamflow, and catchment characteristics, including ground and remote-sensed parameters; however, varying spatial and temporal resolutions among variables such as precipitation and temperature often hinder intercomparison and merging of datasets (Guo, 2017). A wide range

30   of data sources with varying analysis and derivation methods can introduce uncertainty, especially when metadata (Kelleher and Braswell, 2021) or uncharacterized anthropogenic influences are excluded (Addor et al., 2020).

MOPEX and CAMELS are two prominent datasets that encompass a combination of daily temperature, precipitation, potential evapotranspiration, and streamflow values for selected catchments. Additionally, these datasets provide essential catchment

35   characteristics such as area, elevation, vegetation, and soil texture, employing the United States Geological Survey (USGS) hydrologic unit code (HUC) subbasin classification (Seaber et al., 1987). While the consolidation of attributes and hydroclimatic data simplifies the acquisition process, challenges arise due to differences in spatial coverage and data sources, which currently limit the opportunity to effectively utilize both the MOPEX and CAMELS datasets simultaneously or confirm findings and expand on studies employing either dataset.

40   Researchers often face the necessity of choosing one data set over the other, leading to a situation where the unique strengths and limitations of each data set influence the selection process. Numerous studies have engaged in the generalization and categorization of watersheds within the CONUS using either the MOPEX or CAMELS dataset, which underscores the widespread impact and influence of these two large-sample datasets, making them arguably the most prolific resources within hydrological studies focused on the CONUS. Their prevalence in hydrologic studies is reflected in the citation counts data

45   derived from *Clarivate Web of Science* (Clarivate, 2024); with MOPEX (Duan et al., 2006) currently cited in 489 scientific papers and CAMELS (Addor et al., 2017) cited in 352. Here we undertake a unique comparative study between the MOPEX and CAMELS datasets, using exploratory data analysis to evaluate their comparability, accuracy, and implications for past, present, and future research. The results aim to bolster confidence in analytical and modeling outcomes derived from either dataset, thereby fostering robust hydrological research, and supporting effective water resource management in the CONUS. The

50   evaluation is conducted over water years common to both datasets (1981–2000), emphasizing 47 common subbasins. Many previous dataset comparison studies have addressed global climate datasets (Essou et al., 2016; Newman et al., 2019), precipitation (Buban et al., 2020; Levy et al., 2017; Muche et al., 2020; Prat and Nelson, 2015; Sitterson et al., 2020; Sun et al., 2018), temperature (Oubeidillah et al., 2014), and evapotranspiration products (Carter et al., 2018; Chao et al., 2021; E. Han et al., 2015). These studies contribute to the ongoing efforts to advance the understanding of hydrological processes and improve

55   the reliability of hydrologic models (Gupta et al., 2014); however, there has yet to be a study comparing these two large sample watershed-based datasets. The unique combination of MOPEX and CAMELS datasets could potentially encompass 1,050 watersheds within the CONUS, providing detailed attributes and daily values for the period 1948 to present (in progress, Sink, et al. 2025). This study focuses on 47 of the 52 watersheds shared by both datasets and demonstrates that both can be extended by leveraging external gridded datasets, while accounting for the associated margin of error and biases in the process.

60   **2 Hydrometeorological datasets**

2.1 MOPEX

The MOPEX intercomparison project was conceived by several organizations including the World Meteorological Organization (WMO), the International Association of Hydrogeologists (IAH) Predictions in Ungauged Basins (PUB) initiative, and the Global Energy and Water Cycle Experiment (GEWEX) in 1996 (Duan et al., 2006). Its aim was to establish guidelines for parameter

65   estimation techniques while simultaneously decreasing uncertainty (Schaake et al., 2006). MOPEX contains precipitation, minimum and maximum temperature, and streamflow data for 431 CONUS basins on a daily time step for 1949–2003. MOPEX variables are based on weather station observations from the National Climatic Data Center (NCDC), which are then averaged by catchment area using an inverse distance weighting method. For more details regarding data selection and processing, refer to Duan et al. (2006) and Schaake et al. (2006).

2.2 CAMELS

CAMELS, sponsored by the US Bureau of Reclamation and the US Army Corps of Engineers, is a dataset of daily forcing data along with benchmark model performance results using the coupled Snow-17 snow model and the Sacramento Soil Moisture Accounting Model (SAC-SMA) for 671 basins within the CONUS covering the years 1980–2014 (Newman et al., 2015). CAMELS contains precipitation, temperature, and streamflow data on daily time steps in addition to detailed soil characterizations and
75   geology. The Daymet Version 2 forcing dataset interpolates observations to a 1 km x 1 km grid using a Gaussian weighting process (Thornton et al., 2012), which are simply averaged over the catchment area in CAMELS. For an in-depth discussion regarding data selection and processing for CAMELS, refer to Addor et al. (2017) and Newman et al. (2015).

2.3 Dataset comparison

Both datasets selected basins with apparently minimal anthropogenic impacts, highlight processing methods, and provide access
80   to basin characteristics including boundary files, fractional spatial coverage of soil type, vegetation type, land cover, area, and elevation (Table 1). The documentation of catchment attributes, along with daily data for streamflow, temperature, precipitation, and potential evapotranspiration, significantly streamlines the initial phases of data investigation, consolidation, and processing, making the datasets exceptionally valuable for research and analysis.

85   **Table 1. Comparisons between MOPEX and CAMELS. Acronyms are Hydro-Climatic Data Network (HCDN), National Climatic Data Center (NCDC), Cooperative Observer Program (COOP), Snow Telemetry Network (SNOTEL), Parameter-elevation Regressions on Independent Slopes Model (PRISM), North American Land Data Assimilation System (NLDAS), National Oceanic and Atmospheric Administration (NOAA), Sacramento Soil Moisture Accounting Model (SAC-SMA), State Soil Geographic (STATSGO) database, Global Lithological Map (GLiM), Global Hydrogeology Maps (GLHYMPS) of permeability and porosity, Moderate Resolution Imaging**
90   **Spectroradiometer (MODIS), International Geosphere-Biosphere Programme (IGBP), University of Maryland (UMD), Normalized Difference Vegetation Index (NDVI).**

Characteristic	MOPEX	CAMELS
Basins	431	671
Temporal coverage	1948–2003	1980–2015
Streamflow (daily)	USGS HCDN (Slack and Landwehr, 1992)	USGS HCDN-2009 (Lins, 2012)
Temperature (daily)	NCDC COOP, SNOTEL	Daymet, Maurer, NLDAS
Potential Evapotranspiration	NOAA (Farnsworth et al., 1982)	Priestly-Taylor
Actual Evapotranspiration		SAC-SMA model
Soil Properties	STATSGO (Miller and White, 1998)	STATSGO, Pelletier et al. (2016)
Geology		GLiM, GLHYMPS
Greenness Fraction (NDVI)	NLDAS	MODIS
Vegetation Type	IGBP, UMD	MODIS

For this study, temperature and precipitation values from the datasets were evaluated on daily, monthly, seasonal, and annual temporal scales between 1981 and 2000, based on water years spanning 1 October 1980 to 30 September 2000. The readily
95   available CAMELS dataset is derived from Daymet (Thornton et al., 2012), but also provides Maurer (Maurer et al., 2002) and NLDAS (Xia et al., 2012) forcing values. Derived variables were omitted for most analyses in this study because evapotranspiration, when calculated using the water balance, will only differ based on the precipitation since both MOPEX and CAMELS obtain the other balance component, streamflow, from the USGS National Water Information System (NWIS). Potential evapotranspiration values are highly dependent on the estimation method used and require additional information such as wind
100   speed, solar radiation, and temperature (Andréassian et al., 2004; Lemaitre-Basset et al., 2022; Pimentel et al., 2023). Potential evapotranspiration values can also be estimated during modeling.

This study provides researchers with detailed analyses regarding the uncertainties within the datasets and between them for a 20-year period. To obtain a longer record, the datasets can be extended to the present using Daymet (Thornton et al., 2012) and the

105 results from this study can be applied to additional basins by climate region. The merged MOPEX and CAMELS datasets will incorporate up to 1,050 watersheds, temporally extended from 1948 to 2023. (in progress, Sink et al., 2025).

2.4 Study Area

MOPEX contains 431 catchments and CAMELS contains 671 (red and blue points, respectively, Fig. 1) within the CONUS. The spatial coverage differs between the two with CAMELS deliberately incorporating more basins within the Great Plains and

110 southwestern US. Each catchment is identified based on the USGS NWIS stream gauge identification number (Table 2), representing its downstream outlet.

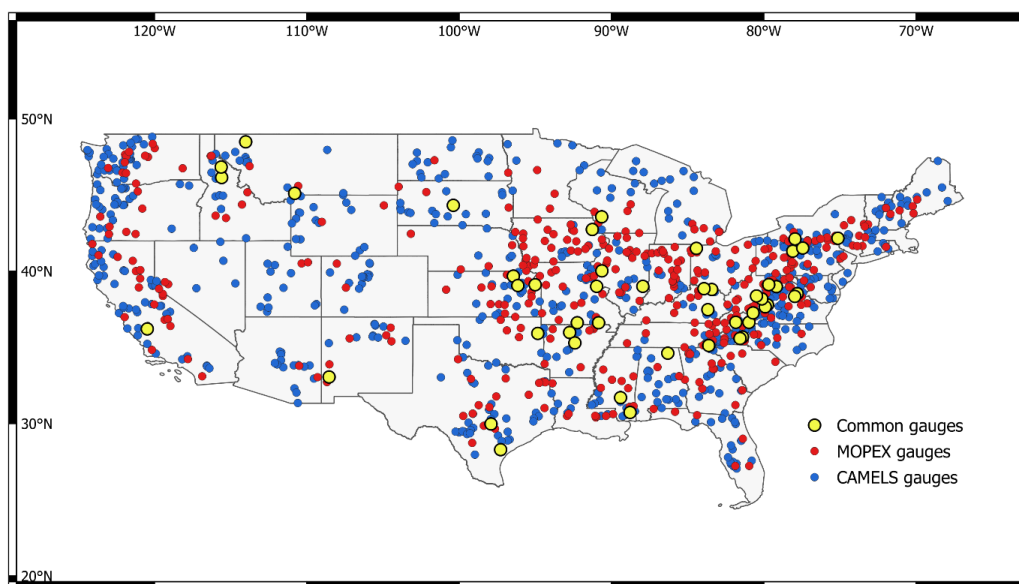



Figure 1. Locations of 431 USGS NWIS stream gauges in MOPEX (red points), 671 gauges in CAMELS datasets (blue points), and common gauges (yellow points) within the CONUS that appear in both the MOPEX and CAMELS datasets.

115 The datasets have 52 basins in common, 47 of which were used in this study (yellow points, Fig. 1). Five watersheds were omitted from this study because of incomplete streamflow records, or the gauge catchment was only a portion of the watershed. Catchment climate variables precipitation (PRCP) and temperature (TAIR) data were area weighted (average of observations values over the area of the basin) using the Hydro-Climatic Data Network (HCDN) basin delineations (Slack and Landwehr, 1994).

120 **Table 2. Common watersheds between MOPEX and CAMELS. Basins are described by GaugeID (NWIS identification) along with the station name, location (city, state), latitude (decimal degrees), longitude (decimal degrees), elevation (meters), area (square kilometers) and climate. Basins are grouped by climate type and then sorted by increasing gauge identification number.**

GaugeID	Station Name	Location	Latitude	Longitude	Elevation (m)	Area (km²)	Climate
06441500	Bad River	Fort Pierre, SD	44.33	-100.38	683.42	8,152.55	Arid
08171300	Blanco River	Kyle, TX	29.98	-97.91	379.23	1,067.47	Arid
08189500	Mission River	Refugio, TX	28.29	-97.28	67.31	1,808.29	Arid
09430500	Gila River	Gila, NM	33.06	-108.54	2,227.36	4,804.93	Arid
11224500	Los Gatos Creek	Coalinga, CA	36.21	-120.47	658.03	247.44	Arid
01664000	Rappahannock River	Remington, VA	38.53	-77.81	216.10	1,605.10	Continental

GaugeID	Station Name	Location	Latitude	Longitude	Elevation (m)	Area (km^2)	Climate
01667500	Rapidan River	Culpepper, VA	38.35	-77.98	193.47	1,209.75	Continental
02016000	Cowpasture River	Clifton Forge, VA	37.79	-79.76	645.04	1,194.55	Continental
02018000	Craig Creek	Parr, VA	37.67	-79.91	648.68	852.34	Continental
03173000	Walker Creek	Bane, VA	37.27	-80.71	750.95	773.32	Continental
03237500	Ohio Brush Creek	West Union, OH	38.80	-83.42	272.36	1,003.21	Continental
03238500	White Oak Creek	Georgetown, OH	38.86	-83.93	285.37	568.50	Continental
03346000	North Fork Embarras River	Oblong, IL	39.01	-87.95	173.50	814.69	Continental
04185000	Tiffin River	Stryker, OH	41.50	-84.43	250.64	1,064.00	Continental
05408000	Kickapoo River	La Farge, WI	43.57	-90.64	348.35	689.33	Continental
05412500	Turkey River	Garber, IA	42.74	-91.26	327.65	3,858.21	Continental
05514500	Cuivre River	Troy, MO	39.01	-90.98	226.22	2,407.41	Continental
05585000	La Moine River	Ripley, IL	40.02	-90.63	197.96	3,354.61	Continental
06191500	Yellowstone River	Corwin Springs, MT	45.11	-110.79	2,547.95	6,783.59	Continental
06885500	Black Vermillion River	Frankfort, KS	39.68	-96.44	394.81	1,062.87	Continental
06888500	Mill Creek	Paxico, KS	39.06	-96.15	412.34	842.35	Continental
06892000	Stranger Creek	Tonganoxie, KS	39.12	-95.01	304.55	1,092.72	Continental
07057500	North Fork River	Tecumseh, MO	36.62	-92.25	324.68	1,456.44	Continental
01423000	West Branch Delaware River	Walton, NY	42.17	-75.14	593.67	859.68	Temperate
01543500	Sinnemahoning Creek	Sinnemahoning, PA	41.32	-78.10	547.44	1,778.26	Temperate
01548500	Pine Creek	Cedar Run, PA	41.52	-77.45	546.71	1,557.05	Temperate
01606500	South Branch Potomac River	Petersburg, WV	38.99	-79.18	836.38	1,684.55	Temperate
02143000	Henry Fork	Henry River, NC	35.68	-81.40	399.35	216.67	Temperate
02143040	Jacob Fork	Ramsey, NC	35.59	-81.57	411.33	66.48	Temperate
02472000	Leaf River	Collins, MS	31.71	-89.41	123.24	1,927.13	Temperate
02479300	Red Creek	Vestry, MS	30.74	-88.78	65.19	1,144.20	Temperate
03069500	Cheat River	Parsons, WV	39.12	-79.68	961.46	1,856.85	Temperate
03164000	New River	Galax, VA	36.65	-80.98	766.60	2,952.74	Temperate
03182500	Greenbrier River	Buckeye, WV	38.19	-80.13	934.51	1,364.97	Temperate
03186500	Williams River	Dyer, WV	38.38	-80.48	1,057.61	329.68	Temperate
03281500	South Fork Kentucky River	Booneville, KY	37.48	-83.68	376.49	1,838.22	Temperate
03473000	South Fork Holston River	Damascus, VA	36.65	-81.84	916.29	784.81	Temperate
03504000	Nantahala River	Rainbow Springs, NC	35.13	-83.62	1,039.71	134.52	Temperate
03574500	Paint Rock River	Woodville, AL	34.62	-86.31	337.61	813.80	Temperate
04221000	Genesee River	Wellsville, NY	42.12	-77.96	658.41	750.88	Temperate
07056000	Buffalo River	St. Joe, AR	35.98	-92.75	459.08	2,149.36	Temperate
07068000	Current River	Doniphan, MO	36.62	-90.85	293.50	5,318.59	Temperate
07197000	Baron Fork	Eldon, OK	35.92	-94.84	348.86	808.45	Temperate
07261000	Cadron Creek	Guy, AR	35.30	-92.40	197.55	445.81	Temperate
12358500	Middle Fork Flathead River	West Glacier, MT	48.50	-114.01	1,559.24	2,939.19	Temperate
13337000	Lochsa River	Lowell, ID	46.15	-115.59	1,548.18	3,053.42	Temperate
13340600	North Fork Clearwater River	Canyon Ranger Station, ID	46.84	-115.62	1,417.79	3,354.62	Temperate

3 Methodology

3.1 Climate characterization of the watersheds

125   Understanding how catchments partition annual precipitation into runoff and evapotranspiration under varying climatic conditions is crucial for hydrological modeling and water resource management. The Budyko function describes the long-term water and energy balance using annual evaporative (evapotranspiration/precipitation) and aridity (potential evapotranspiration/precipitation) indices (Budyko, 1974). The annual indices were determined for both datasets and subsequently combined during KMeans clustering to obtain the overall climate representation for each basin. KMeans clustering, an unsupervised machine learning

130   algorithm, that seeks to minimize the within cluster sum of squares (Hartigan and Wong, 1979), was utilized to divide the 47 selected MOPEX-CAMELS shared basins into three climate groups based on their annual evaporative and aridity indices, with a classification accuracy of 84 %. The arid (aridity index > 1.5), continental (aridity index 1.5 to 0.82), and temperate (aridity index < 0.82) zones represent the three KMeans groups. For this study, the basin climate region classifications (arid, continental, temperate) are based on the KMeans clustering results, which agree closely (but not perfectly) with the Köppen-Geiger climate

135   classification (Fig. 2).

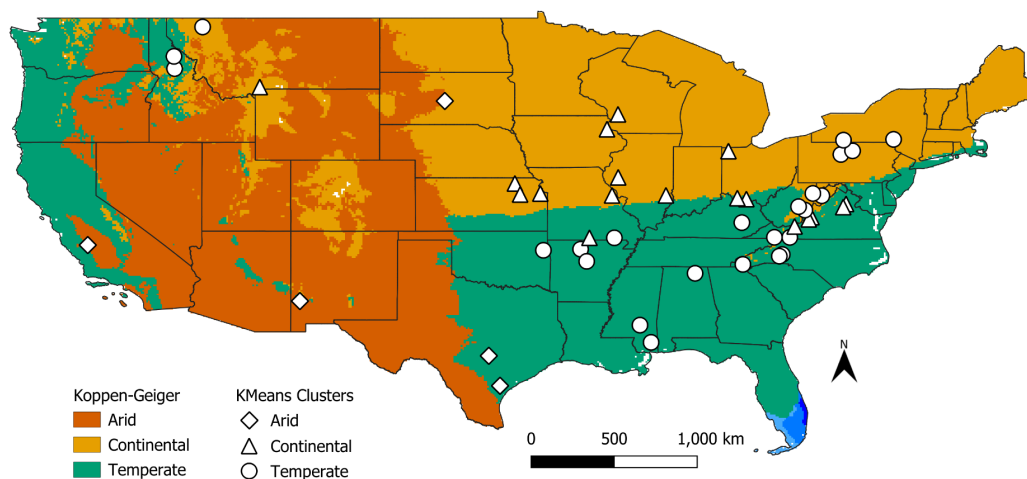



Figure 2. Regions of the CONUS divided into Köppen Geiger climate classification (arid, continental, temperate) are represented by vermillion, orange, and bluish green respectively. The results of the KMeans clustering are based on the annual aridity and evaporative indices for MOPEX-CAMELS shared basins shown by point symbols (diamond, triangle, circle). Climate groupings for analyses are
140   **represented by the KMeans clusters which agree closely but imperfectly with the Köppen-Geiger classification.**

   A Budyko diagram, plotting evaporative versus aridity indices, clarifies the predominant hydrologic processes versus climate type (Fig. 3) for the common basins. Unfortunately, CAMELS evapotranspiration (ET, Newman et al., 2015) values are computed differently than MOPEX values (which uses catchment water balance, the difference between precipitation and streamflow, to estimate ET); a water balance-based estimate for CAMELS basins (CAMELS-WB, open symbols, Fig. 3) was derived in this study

145   to facilitate comparison of the two datasets. ET provided in the CAMELS dataset is generated from SAC-SMA modeling results (CAMELS SAC-SMA, solid symbols, Fig. 3), and exhibits large discrepancies from CAMELS-WB and MOPEX ET values for most catchments (arrows, Fig. 3). Furthermore, several CAMELS SAC-SMA gauges plotted above the water limit (i.e. to absurd values in the Budyko context) and were 10 to 12 % larger than the water-balance-calculated evapotranspiration indices. The higher model-derived ET for CAMELS could reflect additional non-precipitation sources of water to the catchment, but that was not

150   evaluated in this study. The largest discrepancies between model-derived ET/P and water balance derived ET/P for CAMELS for

each climate region are 46.41 %, 12.25 %, and 59.82 % for arid, continental, and temperate regions respectively. Average discrepancies for both CAMELS evapotranspiration values are largest in arid regions, 12.22 %, followed by average discrepancies of 11.34 % in temperate regions, and 4.88 % in continental regions. Studies that have made use of MOPEX data obtain evapotranspiration via the water balance approach (Berghuijs et al., 2014; Coopersmith et al., 2012; P. F. Han et al., 2023; Sawicz et al., 2014) and further research conducted using the CAMELS dataset should also apply the water balance approach instead of the SAC-SMA derived evapotranspiration, to avoid vertical displacement in Budyko space that is an artifact of model derived evapotranspiration values. The water balance evapotranspiration values were calculated using precipitation that does not include snowmelt; however, most of the larger discrepancies are present in arid regions (vermillion, Fig. 3) where snowmelt is negligible.

Differences between water-balance calculated ET for MOPEX versus CAMELS vary by climate type and may partly result from variations in sample distribution. Most of the shared watersheds fall into temperate and continental climates, but the western US is not as heavily represented based on the distribution of the catchments and the restriction to shared basins. Only eight shared catchments lie west of the hundredth meridian (Fig. 1). The arid region basins lie close to the water limit (ET/P = 1, Fig. 3), while the temperate region basins are close to the energy limit (PET/P = 1). The continental region catchments can be seen as a transitional climate which can be either energy or water limited. Annual MOPEX and CAMELS evaporative and aridity indices are plotted separately to highlight the improvements when utilizing the water balance evapotranspiration values for CAMELS. The largest difference between MOPEX and CAMELS evaporative indices using evapotranspiration water balance values is 16.21 % in temperate regions with an overall average difference of 4.6 % for all 47 basins. The mean difference between MOPEX and CAMELS evaporative indices, with water balance calculated evapotranspiration, is 1.89 %, 2.53 %, and 7.14 % for arid, continental, and temperate regions respectively.

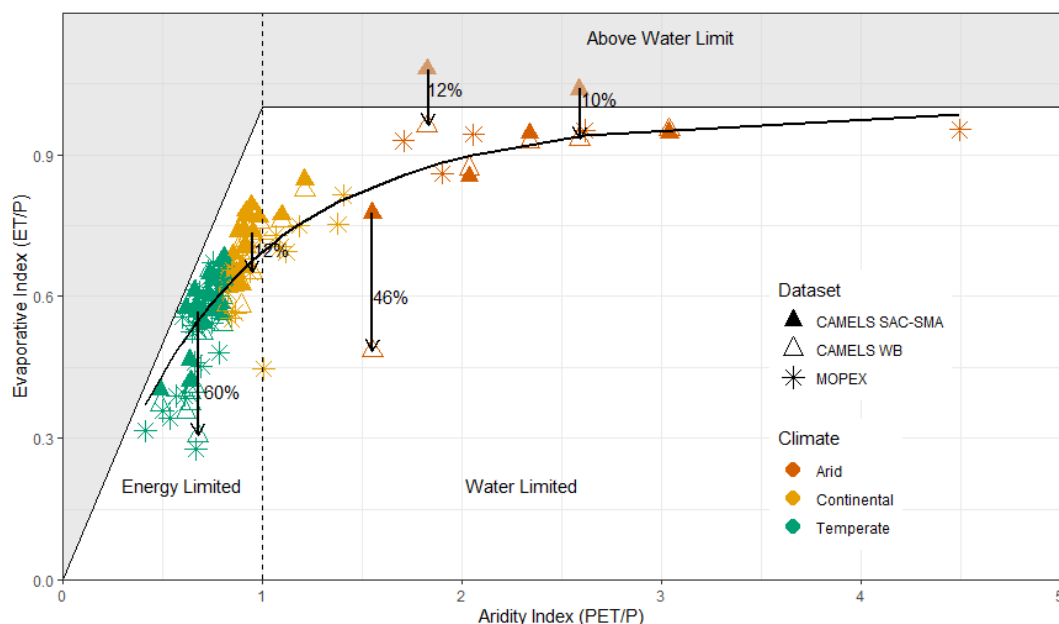



Figure 3. Budyko diagram with the aridity and evaporative indices plotted for each of the 47 catchments (1981-2000). The overall aridity index and evaporative index is plotted for each catchment for the three ET values, resulting in 141 points. The three ET values include MOPEX (asterisk), CAMELS (solid triangle) with SAC-SMA derived ET, and CAMELS WB (open triangle) with water balance calculated evapotranspiration. Evaporative index values > 1 are non-physical.

3.2 Exploratory data analysis

All statistical analyses were conducted using R Statistical Software (v. 4.3.3; R Core Team, 2024). When basins are consolidated by climate region, the number of values used in calculations are dependent on the number of gauges unless otherwise specified (Table 3). Each gauge has 7,305 daily observations beginning on 1 October 1980 and ending on 30 September 2000. Monthly values are based on water years which begin in October of the previous calendar year and end in September of the current calendar year. Seasons are winter (December, January, February), spring (March, April, May), summer (June, July, August), and fall (September, October, November) and the months are grouped by water year, resulting in all four seasons within each water year.

Table 3. Number of observations used for various statistical analyses on temporal scales per dataset.

Time	Range	Per Gauge	Arid (5 gauges)	Continental (18 gauges)	Temperate (24 gauges)
Days	1 October 1980–30 September 2000	7,305	36,525	131,490	175,320
Months	October–September	240	1,200	4,320	5,760
Seasons	Fall 1981–Fall 2000	80	400	1,440	1,920
Water Years	1981–2000	20	100	360	480

3.2.1 Uncertainty and variability within datasets

The central tendency (mean, median), variability (variance, standard deviation, coefficient of variation), and distribution (skewness) of precipitation and temperature were independently evaluated for MOPEX and CAMELS. Uncertainty for the mean value was determined using two-sided confidence intervals computed via the bootstrap method. Bootstrapping is a statistical technique that estimates the sampling distribution of a statistic by iteratively resampling, with replacement, from the observed data when the population or sample distribution is unknown (Efron and Tibshirani, 1993). This nonparametric method utilizes the observed data to derive the robust estimates and sampling distributions (Helsel et al., 2020). In this study, bootstrapping was implemented using the *Hmisc* R package (Harrell Jr, 2024) to calculate the mean value for daily, monthly, seasonal, and annual precipitation and temperature for MOPEX and CAMELS, separately. Analyses involved 10,000 resamples, and the two-sided 95-percent confidence intervals were determined by the 0.025 and 0.975 quantiles. This approach provides a robust method for estimating the uncertainty and variability associated with the mean values on different temporal scales.

3.2.2 Uncertainty and variability between datasets

Several hypothesis tests were conducted to compare observations between MOPEX and CAMELS. The nonparametric (binomial) sign test is used to compare two groups and assess whether one group is consistently higher than the other (Helsel et al., 2020). For a two-sided test, the null hypothesis posits that about half of the differences will be positive and half will be negative, resulting in a median difference of zero between paired observations. For context, paired observations compare the same date (day), month, season, or water year from each dataset. To conduct this test, MOPEX values were subtracted from CAMELS values, where a positive (negative) difference indicates that the CAMELS value is greater than (less than) the MOPEX value, with no consideration for the magnitude. The differences were computed for daily (7,305 pairs), monthly (240 pairs), seasonal (80 pairs), and annual (20 pairs) precipitation and temperature values for each basin ("Per Gauge" column, Table 3). Given that temperature may include negative instances, strict inequalities were applied. Subsequently, the outcomes were assigned a positive (n_+), negative (n_-), or zero value, and the values were tallied. A binomial distribution was used to calculate the probability of observing a value of n, which is 0.5. A 95% confidence interval results in a significance level of $p<0.05$. Hypothesis testing and significance make use of the *rstatix* (Kassambara, 2023) and *stats* R packages.

210

Independent difference hypothesis tests included the Fligner-Killeen test and the t-test. The non-parametric Fligner-Killeen test was conducted to check whether MOPEX and CAMELS have equal variances, with the null hypothesis assuming variances are equal across all samples. It is less sensitive to departures from normality compared to the Bartlett and Levene tests. The absolute value of the residuals (AVR) is calculated from each group median for j=1 to k groups and i=1 to n_j observations where

215
$$AVR_{ij} = |x_{ij} - median_j| \qquad (1)$$

The AVR is ranked and weighted, resulting in a set of scores. A linear-rank test is then computed on the set of scores (Helsel et al., 2020).

Welch's t-test is a modification of the Student's t-test that does not assume equal variance. The null hypothesis posits that the two group means are identical. The test statistic, t, is calculated as shown in Eq. 2

220
$$t = \frac{m_A - m_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \qquad (2)$$

Where S_A and S_B are the standard deviation of the two groups A and B, along with the means m_A and m_B. And the degrees of freedom, df, is calculated as shown in Eq. 3

$$df = \frac{\left(\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}\right)^2}{\left(\frac{S_A^4}{n_A^2(n_A-1)} + \frac{S_B^4}{n_B^2(n_B-1)}\right)} \qquad (3)$$

Statistical significance for the Fligner-Killeen test and t-test are based on a p-value less than 0.05.

225 Bias, the mean absolute error (MAE), and standard error (SE) were also used to assess the variability within each group. The standard error provides an estimate of the standard deviation of the sampling distribution of the difference between means. The margin of error (MOE) was also determined based on a 95-percent confidence interval with a critical value (α) of 1.96. The critical value is multiplied by the standard error of the difference of the means, which provides the confidence interval for the true difference between the means. The nonparametric Spearman rank correlation coefficient (ρ) was also employed to assess the strength of association between variables. This method is robust to the distribution of data and is less influenced by outliers.

3.3 Validation

A support vector machine (SVM) is a supervised machine learning algorithm used for linear or non-linear classification and regression. SVMs finds the ideal hyperplane, separating the data points in different classes (Shmilovici, 2023). This modeling is robust, efficient, and versatile. Predictive machine learning models can aid in accuracy assessments (Patton et al., 2022) and a

235 linear kernel, C-classification SVM was used to compare MOPEX and CAMELS datasets as a simple binary classification problem using the *e1071* (Meyer et al., 2023) R package. The two datasets were merged into a composite dataset for each climate region, and each was identified by either a zero (CAMELS) or one (MOPEX), representing the target variable. The composite dataset was then split into training and test sets, with 75 % of the data allotted to training and 25 % to testing. Data were randomly selected to avoid any potential bias due to formatting, etc. The classification was conducted on all three climate regions separately for daily,

240 monthly, seasonal, and water year aggregations. Precipitation, temperature, and evapotranspiration values were used as the predictor variables. A double mass curve was also used to check the consistency of the data by plotting the cumulative annual precipitation of CAMELS versus MOPEX. If the data are proportional, then the points will plot as a straight line (Searcy et al., 1960).

4 Results

245   Evaluation and comparison of the internal uncertainty and variability of individual dataset parameters are key to understanding the consistency between the MOPEX and CAMELS datasets, and the potential for merging and extending these datasets. For each dataset, climate parameter variability primarily depends on level of aggregation (daily, monthly, annual) and secondarily on climate type. Between datasets, potentially important biases in climate variables are evident, varying by climate type and aggregation level. This paper presents a thorough exploratory data analysis and supports the main finding that the two datasets exhibit similar

250   uncertainty and variability, both within and between them. Though there are consistent biases, they are minimal for aggregations beyond a daily time step, making them suitable for combined application in climate studies and hydrologic modeling at monthly or annual aggregations.

4.1 Uncertainty and variability within datasets

Analysis of internal variability of the MOPEX and CAMELS datasets indicates relatively similar characteristics. Statistics for

255   precipitation included in the datasets are shown in Table 4 and were calculated over all shared basins within a climate region. Temperature statistics are shown in Table 5. The mean values and corresponding confidence intervals are based on the averages derived from bootstrapping results, shown in Figs. 4-6 for monthly, seasonal, and annual precipitation values and Figs. 8-10 for monthly, seasonal, and annual temperature values. The tables also highlight the commensurate central tendencies, variabilities, and dispersion values within the datasets and provide insight into the biases and comparisons between the datasets that will be discussed

260   in Sect. 4.2.

Table 4. Statistics for MOPEX (M) and CAMELS (C) precipitation totals by climate region. Bootstrapping mean values for each climate region and the lower and upper confidence limit are based on two-sided 95 % confidence interval and 10,000 replicates with replacement. Variance, standard deviation, and skew are based on the average of values calculated for each basin within a region.

265

	PRCP	Median	Median	Mean ±CI	Mean ±CI	Variance	Variance	St Dev	St Dev	Skew	Skew
	(mm)	C	M	C	M	C	M	C	M	C	M
ARID	Day	0.00	0.06	1.76 ±0.06	1.78 ±0.06	33.12	32.78	5.50	5.51	5.18	6.82
	Month	36.92	40.38	53.65 ±10.59	54.13 ±10.22	3088.97	2846.59	54.63	52.47	1.55	1.50
	Season	143.97	148.50	159.08 ±22.75	160.56 ±22.03	13249.96	12495.40	114.91	111.23	0.91	0.81
	Year	607.69	638.06	643.74 ±182.57	649.56 ±165.36	66151.24	55442.78	236.10	213.50	0.21	0.01
CONT	Day	0.03	0.18	2.85 ± 0.04	2.70 ±0.04	46.45	42.53	6.74	6.45	3.99	4.52
	Month	78.21	73.76	86.86 ±5.20	82.17 ±4.86	2645.29	2253.29	50.90	47.00	1.14	1.14
	Season	248.88	235.19	257.56 ±9.79	243.68 ±9.61	9504.04	8246.90	97.17	90.54	0.60	0.68
	Year	1040.57	982.88	1,042.31 ±79.39	986.09 ±76.57	35233.38	31811.23	178.89	170.55	0.01	0.04
TEMP	Day	0.10	0.39	3.73 ±0.04	3.50 ±0.04	61.98	60.14	7.69	7.55	3.56	4.20
	Month	103.36	95.94	116.09 ±6.03	111.42 ±6.46	3655.81	3278.64	60.25	57.04	1.00	1.12
	Season	325.39	300.32	344.33 ±12.30	330.48 ±14.36	14851.10	13598.74	121.20	115.60	0.92	0.96
	Year	1315.96	1243.95	1393.13 ±122.75	1337.09 ±158.06	75120.98	76096.40	267.91	271.02	1.12	0.89

Table 5. Statistics for MOPEX (M) and CAMELS (C) temperature means by climate region. Bootstrapping mean values for each climate region and the lower and upper confidence limit are based on two-sided 95 % confidence interval and 10,000 replicates with replacement. Variance, standard deviation, and skew are based on the average of values calculated for each basin within a region.

	TAIR	Median	Median	Mean ±CI	Mean ±CI	Variance	Variance	St Dev	St Dev	Skew	Skew
	(˚C)	C	M	C	M	C	M	C	M	C	M
ARID	Day	14.77	15.33	14.39 ±0.10	15.01 ±0.10	70.17	65.68	8.17	7.93	-0.32	-0.25
	Month	13.83	15.34	14.35 ±1.07	14.97 ±1.04	32.28	29.08	5.59	5.30	-0.06	-0.16

	Season	13.73	15.49	14.39 ±1.05	15.01 ±1.04	30.26	27.31	5.43	5.16	0.01	-0.12
	Year	13.47	15.35	14.39 ±4.44	15.01 ±4.29	33.28	30.27	5.76	5.49	0.18	0.01
CONT	Day	11.32	11.37	10.61 ± 0.06	10.77 ±0.06	103.63	98.58	10.13	9.89	-0.34	-0.26
	Month	11.25	11.29	10.57 ±0.37	10.72 ±0.34	13.34	11.09	3.63	3.32	-1.57	-1.42
	Season	11.30	11.32	10.60 ±0.36	10.75 ±0.32	10.93	8.99	3.30	3.00	-1.91	-1.81
	Year	11.54	11.66	10.62 ±1.33	10.77 ±1.19	9.35	7.64	3.05	2.76	-2.30	-2.33
TEMP	Day	11.14	11.35	10.57 ±0.05	10.92 ±0.04	80.52	76.14	8.94	8.70	-0.36	-0.28
	Month	10.07	10.63	10.47 ±0.40	10.89 ±0.38	21.71	19.96	4.64	4.45	-0.02	-0.05
	Season	9.96	10.64	10.49 ±0.38	10.92 ±0.36	19.63	18.15	4.42	4.26	0.05	-0.01
	Year	10.09	10.65	10.51 ±1.64	10.94 ±1.57	19.04	17.67	4.36	4.20	0.11	0.03

270

The coefficient of variation (CV) was calculated for each catchment on all temporal scales for precipitation (Fig. 4). Analyses indicate that the largest dispersion is observed among daily values and decreases with greater temporal aggregation. Daily precipitation shows considerably high variation, with CAMELS mean CV values of 3.28, 2.39, and 2.12 and MOPEX mean CV values of 3.23, 2.42, and 2.22 in arid, continental, and temperate regions respectively (Fig. 4a). Considerably high variation is still observed on monthly scales (Fig. 4b) but decreases to moderate variability for seasonal temporal aggregations for all regions, and low variability, less than one, on an annual scale. Temperature demonstrates the same decrease in variability as precipitation from daily to annual temporal aggregation for all regions and is not shown.

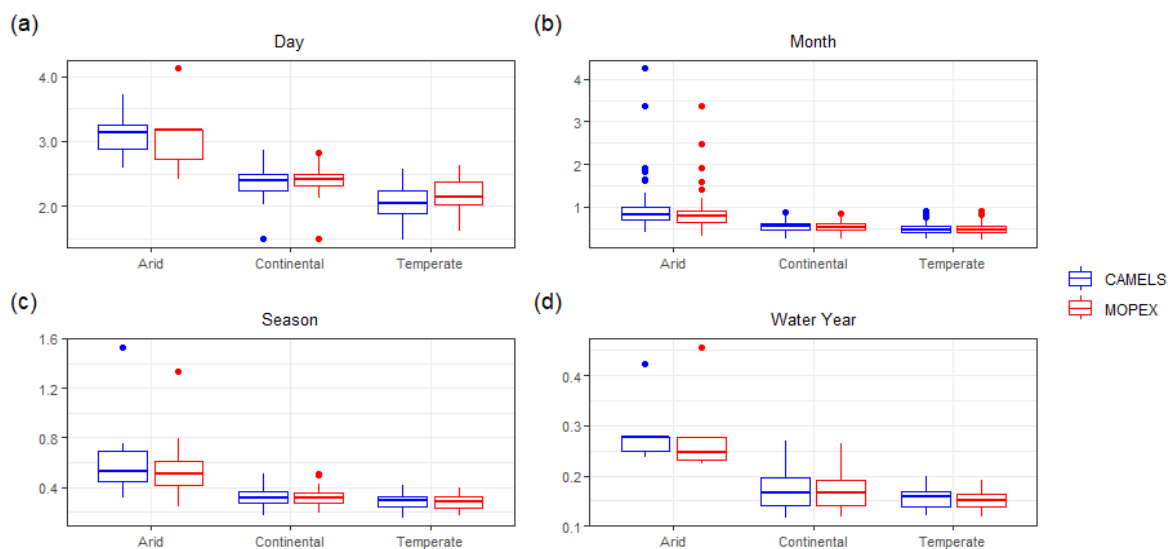



Figure 4. Coefficient of variation of precipitation for CAMELS (blue) and MOPEX (red) for each climate region, shown by temporal aggregation a) day, b) month, c) season, and d) water year. Each boxplot represents the value of all basins within the climate region based on total precipitation (mm). Note progressively declining y-axis range from (a) to (d).

The daily mean precipitation and temperature were calculated for each basin and the corresponding 95-percent confidence intervals were established using 10,000 samples with replacement. The results in Table 4 illustrate that overall daily precipitation means are larger for CAMELS than for MOPEX (except for arid regions); however, it is noteworthy that the confidence intervals exhibit overlap for most regions, suggesting similar degrees of uncertainty. The most pronounced divergence in precipitation means, a

difference of 7 percent, is observed in temperate catchments where the overall CAMELS daily mean is 3.73 mm day^{-1} and MOPEX is 3.50 mm day^{-1}.

290

When examining total mean monthly precipitation, both datasets exhibit comparable monthly fluctuations (Fig. 5), but CAMELS exhibits a small positive bias in non-arid climate regions. Arid regions display the most variability, with the largest confidence intervals (± 13.74 mm month^{-1} for CAMELS and ± 12.93 mm month^{-1} for MOPEX) observed in June, and the smallest (± 7.00 mm month^{-1} for CAMELS and ± 7.73 mm month^{-1} for MOPEX) observed in November. Despite this variability, these regions show

295   the greatest temporal consistency between MOPEX and CAMELS values, with total precipitation highest in May and June and lowest in April (Fig. 5a). Additionally, arid regions demonstrate the most notable overlap of the mean values and confidence intervals of the two datasets. Continental regions show an increase in total monthly precipitation in May, June, and July (Fig. 5b). There is the least amount of variation in February (± 4.12 mm month^{-1} for CAMELS and ± 3.79 mm month^{-1} for MOPEX), contrasting with the largest in July (± 6.31 mm month^{-1} for CAMELS and ± 6.08 mm month^{-1} for MOPEX). Temperate regions

300   show decreased precipitation in August, September, and October with less overlap between dataset confidence intervals (Fig. 5c). The smallest confidence intervals differ between datasets with April (± 4.92 mm month^{-1}) for CAMELS and July (± 4.54 mm month^{-1}) for MOPEX, but both share the largest amount of variability in December (± 8.19mm month^{-1} for CAMELS and ± 10.21 mm month^{-1} for MOPEX).

Figure 5. Average total monthly precipitation for CAMELS (blue) and MOPEX (red) by a) arid, b) continental, and c) temperate climate region. The mean value is determined using all basins within the climate region and each corresponding month for 1981-2000. Error bars represent two-sided 95 % confidence interval, derived from bootstrapping with replacement for 10,000 replicates.

Seasonal precipitation confidence intervals exhibit the most variability yet also the greatest consistency in arid regions (Fig. 6),

310   which coincides with monthly precipitation analyses (Fig. 5). The range of potential values decreases in continental and temperate regions. MOPEX values are larger than CAMELS in arid regions in the summer and winter seasons (which corresponds to larger monthly values in December, January, June, July, and August). For arid regions (Fig. 6a), the greatest variance is in the winter season (± 25.24 mm season^{-1} for CAMELS and ± 25.37 mm season^{-1} for MOPEX). Continental regions (Fig. 6b) show the greatest uncertainty in summer for CAMELS (± 10.58 mm season^{-1}) and fall for MOPEX (± 10.24 mm season^{-1}). Temperate regions (Fig.

6c) have the largest differences in variance between datasets with little to no overlap of confidence intervals, notably in the spring. Winter has the greatest confidence intervals for CAMELS (± 17.69 mm season^{-1}) and MOPEX (± 22.89 mm season^{-1}).

Figure 6. Average total seasonal precipitation for CAMELS (blue) and MOPEX (red) by a) arid, b) continental, and c) temperate climate region. The mean value is determined using all basins within the climate region and each corresponding season for 1981-2000. Error bars represent two-sided 95 % confidence interval, derived from bootstrapping with replacement for 10,000 replicates.

For average total annual precipitation, arid regions continue to show the largest variability but remain the most consistent between the two datasets (Fig. 7a); other climate regions exhibit a small positive precipitation bias for CAMELS. Arid region confidence intervals are greater for CAMELS (between ± 70.91 mm year^{-1} in 1996 and ± 326.94 mm year^{-1} in 1987) than MOPEX (between ± 33.96 mm year^{-1} in 1996 and ± 298.17 mm year^{-1} in 1985). Annual means in continental (Fig. 7b) and temperate regions (Fig. 7c) are consistently higher in CAMELS but confidence intervals do overlap with MOPEX. The smallest uncertainty is in continental regions with intervals slightly larger for CAMELS (± 48.03 mm year^{-1} in 1986 to ± 136.19 mm year^{-1} in 1996) compared to MOPEX (± 43.57 mm year^{-1} in 1992 to ± 126.91 mm year^{-1} in 1996). Temperate regions have greater uncertainty associated with MOPEX values (± 92.22 mm year^{-1} in 1992 to ± 221.13 mm year^{-1} in 1982) rather than CAMELS (± 72.39 mm year^{-1} in 1981 to ± 152.43 mm year^{-1} in 1995).

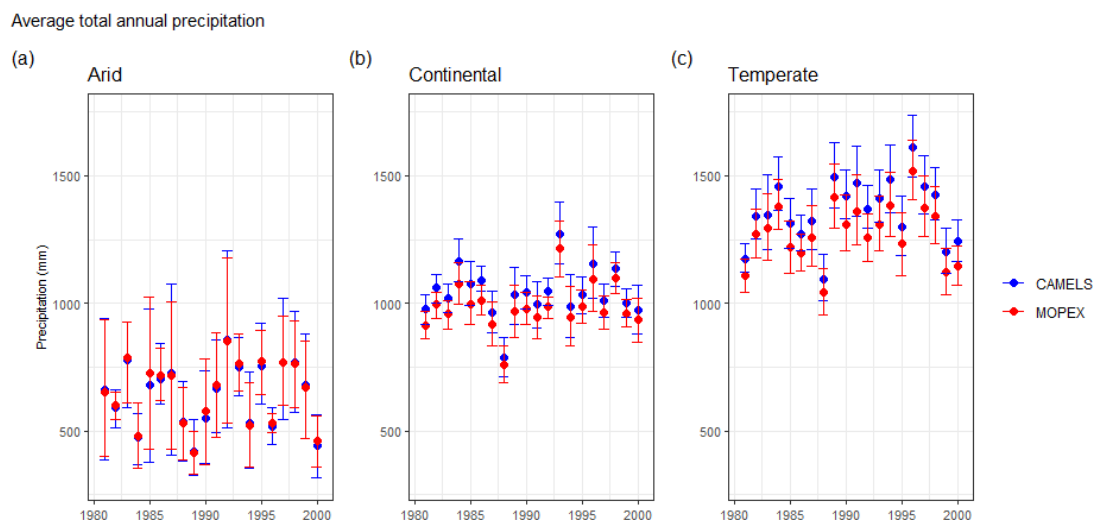



Figure 7. Average total annual precipitation for CAMELS (blue) and MOPEX (red) by a) arid, b) continental, and c) temperate climate region. The mean value is determined using all basins within the climate region and each corresponding water year for 1981-2000. Error bars represent two-sided 95 % confidence interval, derived from bootstrapping with replacement for 10,000 replicates.

Evaluation of daily temperature indicates a consistent pattern discerned in Table 5. The means for daily temperature are consistently larger for MOPEX with the largest differences (mean difference of 0.62 ˚C day^{-1}) observed in arid regions. Monthly temperatures show consistent trends in both datasets, with higher temperatures in July and August and lower temperatures in January and December in all regions. MOPEX and CAMELS are quite similar in their mean values and monthly variability (Fig. 8). Akin to precipitation, arid regions contain the most variability, followed by temperate regions. The largest uncertainty is in December for CAMELS (± 1.37 ˚C month^{-1}) and MOPEX (± 1.32 ˚C month^{-1}) in arid and continental regions (CAMELS ± 0.46 ˚C month^{-1}, MOPEX ± 0.40 ˚C month^{-1}) and in February for temperate regions (CAMELS ± 0.45 ˚C month^{-1}, MOPEX ± 0.43 ˚C month^{-1}). The smallest uncertainty is in July (CAMELS ± 0.80 ˚C month^{-1}, MOPEX ± 0.71 ˚C month^{-1}) for arid regions, October (CAMELS ± 0.32 ˚C month^{-1}, MOPEX ± 0.31 ˚C month^{-1}) for continental regions, and August (CAMELS ± 0.36 ˚C month^{-1}, MOPEX ± 0.33 ˚C month^{-1}) for temperate regions.

Average monthly temperature

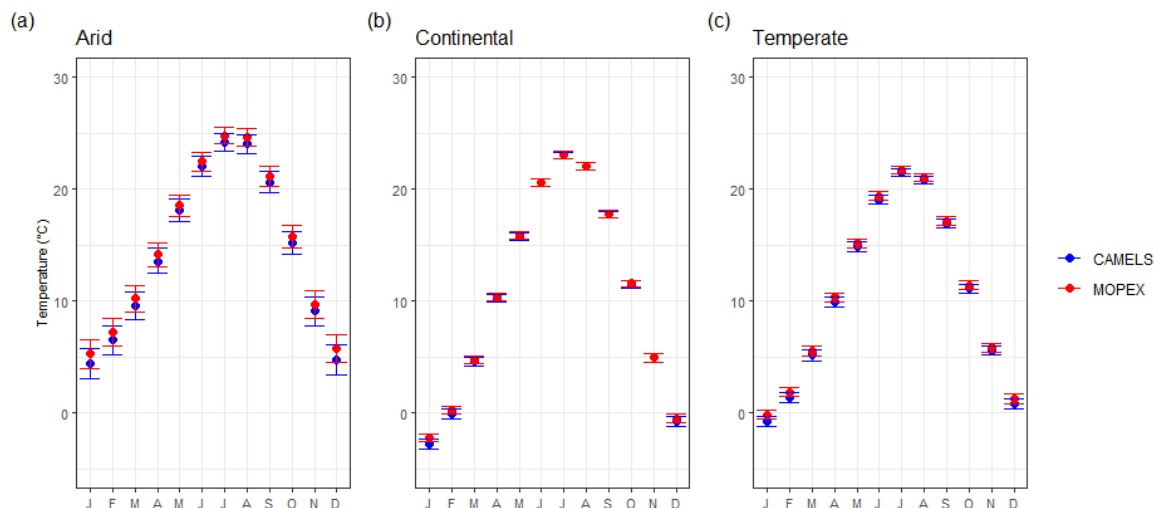



Figure 8. Average monthly mean temperature for CAMELS (blue) and MOPEX (red) by a) arid, b) continental, and c) temperate climate region. The mean value is determined using all basins within the climate region and each corresponding month for 1981-2000. Error bars represent two-sided 95 % confidence interval, derived from bootstrapping with replacement for 10,000 replicates.

Seasonal temperature is also the most variable in arid regions with confidence intervals ranging from $\pm\,0.77$ ˚C to $\pm\,1.28$˚C season^{-1} compared to intervals ranging from $\pm\,0.30$ ˚C to $\pm\,0.40$ ˚C season^{-1} for continental and temperate regions (Fig. 9). Winter is consistently the most variable season among all regions, resulting in the largest confidence intervals.

Average seasonal temperature

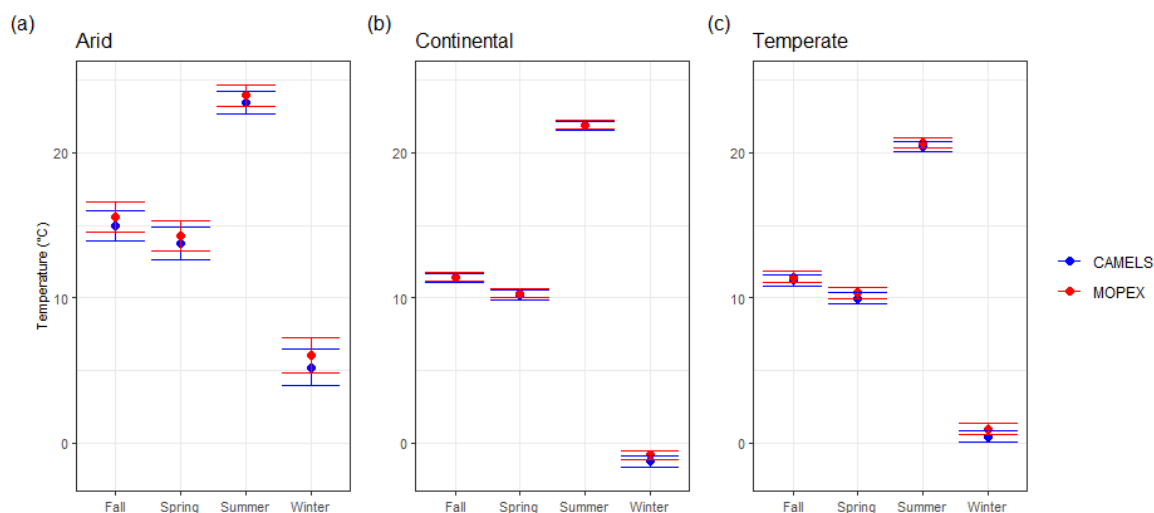



Figure 9. Average mean seasonal temperature for CAMELS (blue) and MOPEX (red) by a) arid, b) continental, and c) temperate climate region. The mean value is determined using all basins within the climate region and each corresponding season for 1981-2000. Error bars represent two-sided 95 % confidence interval, derived from bootstrapping with replacement for 10,000 replicates.

Annually, for temperature, arid confidence intervals are more than double the range of those found in continental and temperate regions (Fig. 10), strongly influenced by the small number of arid sites (Table 3). MOPEX means are consistently larger than

CAMELS indicating a warm bias with the largest bias in arid regions (Fig. 10a). Continental regions have the most similarity between mean values and the smallest amount of uncertainty, with confidence intervals ranging from ± 1.24 ˚C year^{-1} to ± 1.47 ˚C

365   year^{-1} for CAMELS and from ± 1.10 ˚C year^{-1} to ± 1.31 ˚C year^{-1} for MOPEX (Fig. 10b). For temperate regions, MOPEX has a slightly smaller variance compared to CAMELS (Fig. 10c) with confidence intervals ranging from ± 1.44 ˚C year^{-1} to ± 1.73 ˚C year^{-1} (MOPEX) versus ± 1.47 ˚C year^{-1} to ± 1.82 ˚C year^{-1} (CAMELS).

370

Figure 10. Average annual temperature for CAMELS (blue) and MOPEX (red) by a) arid, b) continental, and c) temperate climate region. The mean value is determined using all basins within the climate region and each corresponding water year for 1981-2000. Error bars represent two-sided 95 % confidence interval, derived from bootstrapping with replacement for 10,000 replicates.

4.2 Uncertainty and variability between datasets

375   Important differences between the datasets are detailed below, but in general in time-aggregated values MOPEX exhibits higher temperature, while CAMELS exhibits higher precipitation. Statistical and bootstrapping results from Sect. 4.1 supports these findings. The comparison of paired observations indicates that individual daily MOPEX values for precipitation and temperature are generally larger than CAMELS; in contrast, when CAMELS precipitation values are aggregated monthly, seasonally, or annually, they are typically larger than MOPEX in continental and temperate regions (Fig. 11). This analysis is based on the counts

380   of negative (MOPEX > CAMELS), positive (CAMELS > MOPEX), and zero values (CAMELS = MOPEX). Out of the 7,305 days recorded for each basin, precipitation values for MOPEX surpass CAMELS 48 % (62,638 days out of 131,490 total days) and 49 % (86,496 days out of 175,320 total days) of the time in continental and temperate climates respectively, and 40 % of the time in arid regions (Fig. 11a). In arid climates, MOPEX and CAMELS precipitation values are equal 46 % of total days, while in continental and temperate climates, they are equal 28 % and 22 % of total days. The same binomial test was conducted to analyse

385   total monthly precipitation for each catchment. Direct comparisons were made for each month across all water years (i.e. January 1981, January 1982), tallying negative and positive differences, resulting in 240 months per catchment. When aggregated on a monthly scale, CAMELS typically exhibits greater total monthly precipitation, particularly in continental (66 %) and temperate (69 %) regions. Identical ("SAME", Fig. 11b) total values are negligible. In contrast, arid regions indicate larger MOPEX values in 55 % of all months and only 3.25 % of all months have the same total value (Fig. 11b). Seasonal comparisons, based on 80

seasons per catchment, indicate the same pattern with total precipitation greater in MOPEX for 52 % of all seasons in arid regions, and CAMELS greater in continental and temperate regions for 76 % and 78 % of all seasons respectively (Fig. 11c). On an annual scale, 20 years per watershed, the comparison reveals that total precipitation for arid regions is evenly split, with CAMELS and MOPEX dominating 51 % and 49 % of all years respectively. In contrast, continental and temperate regions are largely dominated by CAMELS, constituting 88 % of all years (Fig. 11d). All comparisons, except for arid seasonal and annual, failed to reject the null hypothesis, which expects a median difference of zero between paired observations.

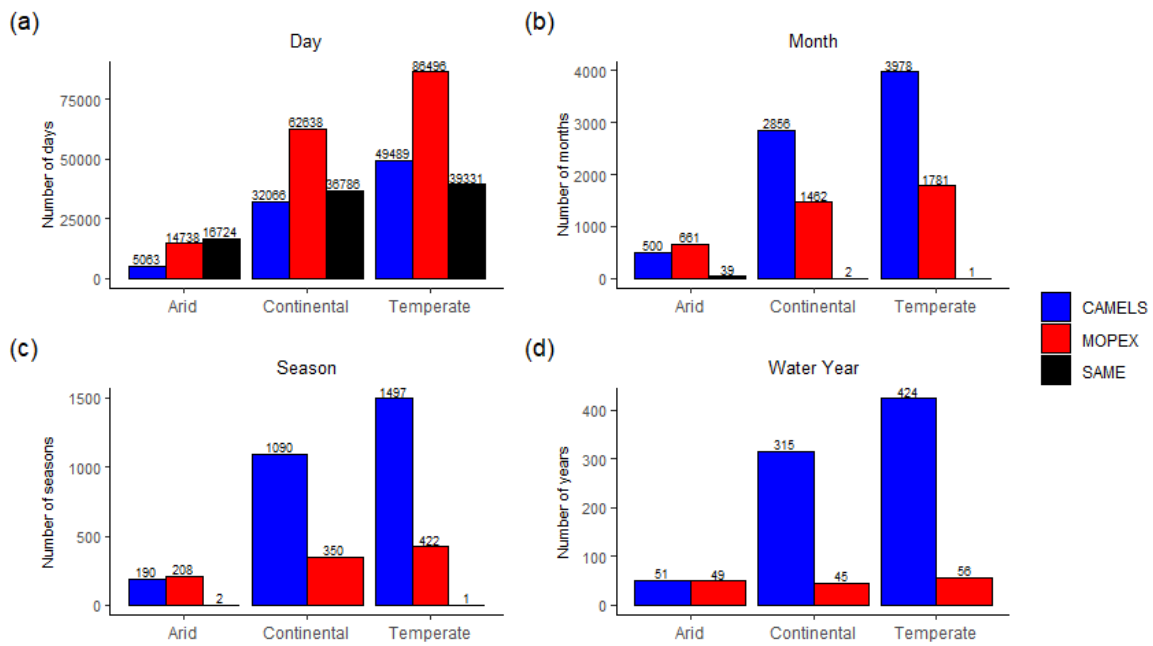



Figure 11. Sign tally results from non-parametric binomial sign test for a) daily, b) monthly, c) seasonal, and d) annual precipitation values. Counts on the y-axis reflect the number of basins (gauges) within each climate region times the number of temporal periods. All results are based on CAMELS minus MOPEX values. Positive values (CAMELS) indicate that CAMELS > MOPEX (blue bars), negative (MOPEX) values indicate MOPEX > CAMELS (red bars), and zero (SAME) indicate that CAMELS = MOPEX (black bars).

Regarding temperature, MOPEX exceeds CAMELS 72 % of total days in arid, 58 % in continental, and 65 % in temperate regions respectively (Fig. 12a). These regions all exhibit the same mean daily temperature values (CAMELS = MOPEX) only 0.03 % of total days. On a monthly scale, MOPEX mean temperature values are larger for all regions, with arid at 81 %, continental at 58 %, and temperate at 74 % of total months with no equal values (Fig. 12b). Seasonal temperature is greater for MOPEX values 85 %, 58 %, and 77 % of all seasons for arid, continental, and temperate regions (Fig. 12c). As for annual mean temperatures, MOPEX values are greater for arid regions in 91 % of all years, while continental and temperate regions show MOPEX dominance in 65 % and 79 % of all years respectively (Fig. 12d). All temperature differences were statistically significant.

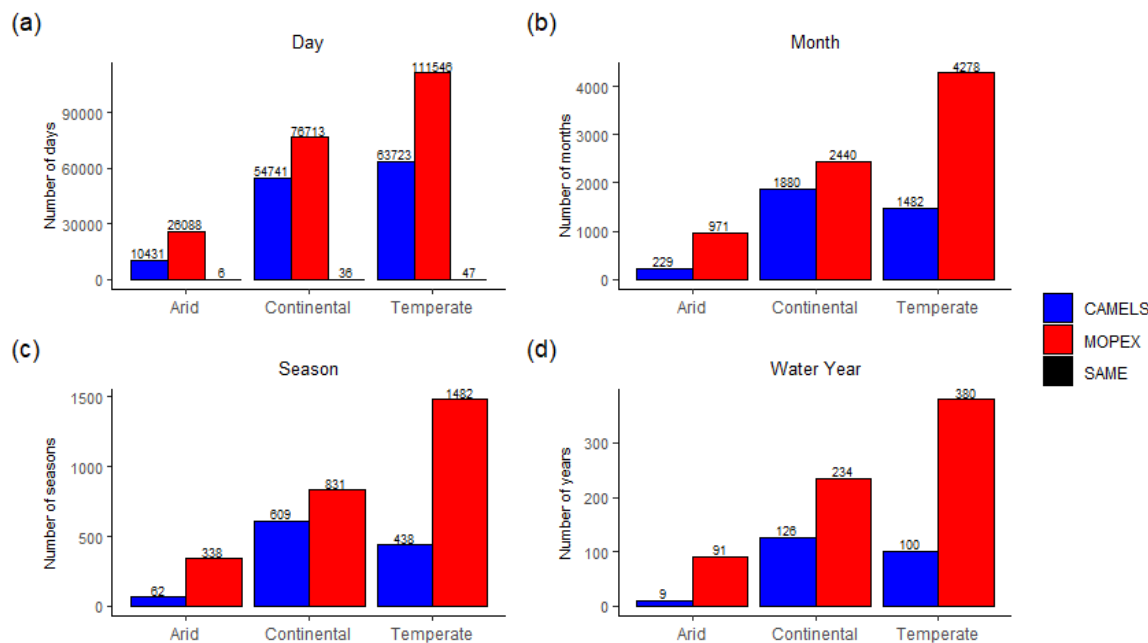



Figure 12. Sign tally results from non-parametric binomial sign for a) daily, b) monthly, c) seasonal, and d) annual temperature values. Counts on the y-axis reflect the number of basins (gauges) within each climate region times the number of temporal periods. All results are based on CAMELS minus MOPEX values. Positive values (CAMELS) indicate that CAMELS > MOPEX (blue bars), negative (MOPEX) values indicate MOPEX > CAMELS (red bars), and zero (SAME) indicate that CAMELS = MOPEX (black bars).

The numerical differences between each pair of same day precipitation values, CAMELS minus MOPEX, reveal substantial differences for extreme events. Specifically, there are 20 instances of daily precipitation values differing by more than 100 mm in separate comparisons across all catchments. This indicates notable variations in daily precipitation values between the two datasets. Daily values do not consistently coincide, as exemplified by the comparison of the same maximum precipitation events for each climate region between 1981 and 2000 (Table 6). In the temperate region, for instance, CAMELS reports the maximum precipitation (181.04 mm day^{-1}) occurring on 7 April 1983 for gauge 02479300 while MOPEX, for the same date, reports a precipitation total of 64.07 mm. MOPEX reports the maximum precipitation (183.25 mm day^{-1}) as occurring on 20 January 1993 at the same gauge (while CAMELS shows a precipitation value of 59.73 mm day^{-1}). Consequently, this study does not recommend direct daily comparisons between MOPEX and CAMELS due to discrepancies in single precipitation events.

Table 6. Largest precipitation event on record for each climate region. Max indicates the maximum daily measurement on record for that dataset between 1981–2000 along with the corresponding value in the other dataset on that date for comparison.

Climate	GaugeID	Date	CAMELS (max)	MOPEX	GaugeID	Date	MOPEX (max)	CAMELS
Arid	08171300	17 Oct 1998	152.12 mm	216.23 mm	08171300	17 Oct 1998	216.23 mm	152.12 mm
Continental	02016000	5 Nov 1985	126.82 mm	11.46 mm	03237500	1 Mar 1997	140.08 mm	68.68 mm
Temperate	02479300	7 Apr 1983	181.04 mm	64.07 mm	02479300	20 Jan 1993	183.25 mm	59.73 mm

A positive precipitation bias for CAMELS is visible for all watersheds within a climate region for all temporal aggregations (positive values, Fig. 13). Monthly, precipitation biases for arid regions range from -52.57 mm month^{-1} to 99.28 mm month^{-1},

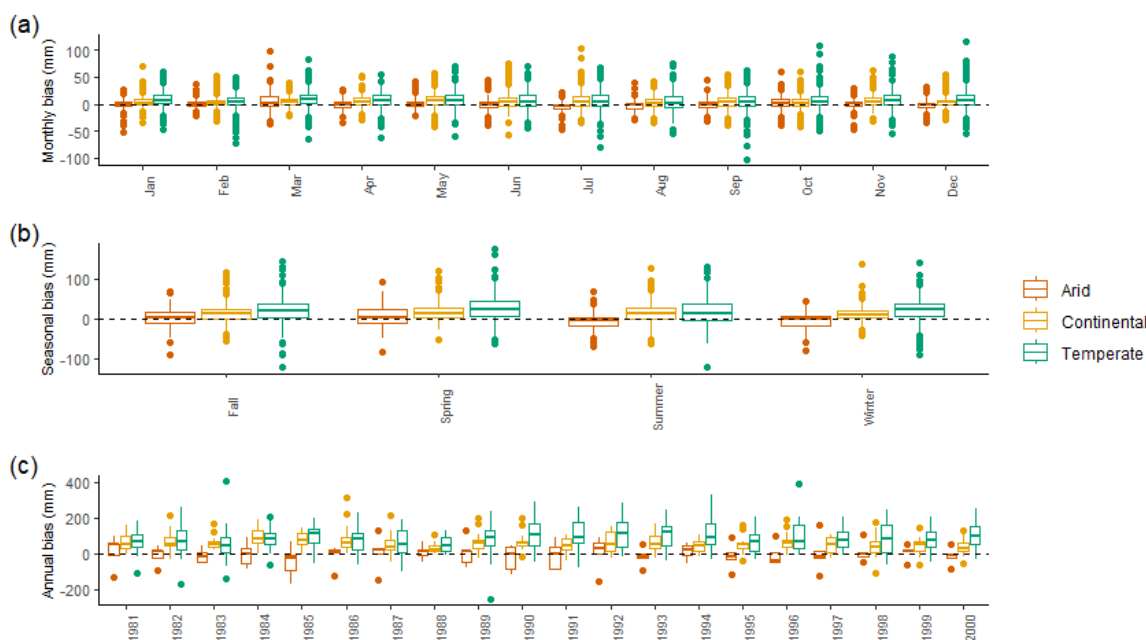
430 continental regions range from -57.60 mm month^{-1} to 103.22 mm month^{-1}, and temperate regions are between -102.68 mm month^{-1} to 117.29 mm month^{-1}, comparing 240 months per catchments (Fig. 13a). Seasonal precipitation biases for arid, continental, and temperate regions are -92.21 mm season^{-1} to 94.65 mm season^{-1}, -64.04 mm season^{-1} to 137.32 mm season^{-1}, and -123.50 mm season^{-1} to 174.88 mm season^{-1} (Fig. 13b). Total annual precipitation bias ranges are -174.43 mm year^{-1} to 160.03 mm year^{-1}, -111.77 mm year^{-1} to 315.40 mm year^{-1}, and -256.88 mm year^{-1} to 405.25 mm year^{-1} based on 20 years per catchments (Fig. 13c).

435

Figure 13. Monthly (a), seasonal (b), and annual (c) precipitation biases. All basins are combined by climate region (arid, continental, and temperate) and box plots represent the number of observations indicated in Table 3. Precipitation biases are based on total values of CAMELS minus MOPEX. Positive values indicate CAMELS > MOPEX and negative values indicate MOPEX > CAMELS.

A negative temperature bias for CAMELS vs MOPES is visible for all watersheds in a climate region for all temporal aggregations

440 (negative values, Fig. 14). Daily temperature values differ between the datasets by as much as ± 28 ˚C day^{-1} with MOPEX demonstrating a greater positive bias (Fig. 12a). Monthly temperature biases for arid regions range from -5.29 ˚C to 2.00 ˚C month^{-1}, continental regions range from -6.43 ˚C to 0.70 ˚C month^{-1}, and temperate regions range from -5.51 ˚C to 2.26 ˚C month^{-1} (Fig. 14a). Seasonal temperature biases range from -2.71 ˚C to 1.70 ˚C season^{-1}, -4.24 ˚C to 0.53 ˚C season^{-1}, and -2.84 ˚C to 0.85 ˚C season^{-1} (Fig. 14b) and mean annual temperature biases decrease to -2.17 ˚C to 0.09 ˚C year^{-1}, -2.44 ˚C to 0.39 ˚C year^{-1}, and -1.89

445 ˚C to 0.36 ˚C year^{-1} (Fig. 14c) for arid, continental, and temperate regions respectively.

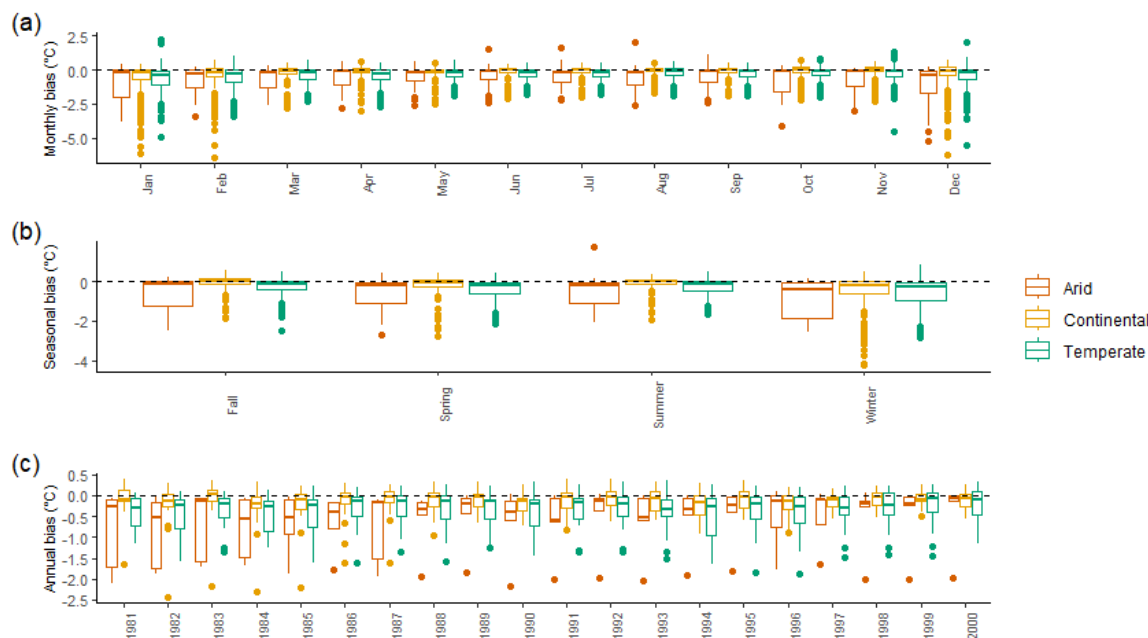



Figure 14. Monthly (a), seasonal (b), and annual (c) temperature biases. All basins are combined by climate region (arid, continental, and temperate) and box plots represent the number of observations indicated in Table 3. Temperature biases are based on mean values of CAMELS minus MOPEX. Positive values indicate CAMELS > MOPEX and negative values indicate MOPEX > CAMELS.

While the ranges of biases for precipitation and temperature are shown in Fig. 13 and Fig. 14 respectively, the magnitude of differences between MOPEX and CAMELS precipitation and temperature values is clarified by averaging biases over all basins in a climate region for daily, monthly, seasonal, and annual time aggregations for 1981–2000 (Fig. 15). Given that the differences are either negative (MOPEX > CAMELS) or positive (CAMELS > MOPEX), the mean reflects the overall bias since equal differences will negate each other. In direct pairwise comparisons, MOPEX daily precipitation values tend to be larger than CAMELS, however, when CAMELS values exceed MOPEX, the numerical difference is greater. Daily averages (not shown) for precipitation bias are -0.02 mm day^{-1} (MOPEX > CAMELS) for arid regions, 0.15 mm day^{-1} (CAMELS > MOPEX) for continental, and 0.23 mm day^{-1} (CAMELS > MOPEX) for temperate regions, indicating a wet bias in arid regions for MOPEX and a wet bias for CAMELS in continental and temperate regions. When precipitation values are aggregated on a monthly scale (Fig. 15a), CAMELS values exceed MOPEX values by 2.94 mm month^{-1} (Feb) to 6.79 mm month^{-1} (May) in continental regions and by 4.41 mm month^{-1} (Aug) to 9.31 mm month^{-1} (Mar) in temperate regions. In arid climates, CAMELS exceeds MOPEX by 0.18 mm month^{-1} (Feb), 5.07 mm month^{-1} (Mar), 1.11 mm month^{-1} (May), and 2.56 mm month^{-1} (Oct) while MOPEX exceeds CAMELS (negative values) by 2.58 mm month^{-1} (Jan), 0.84 mm month^{-1} (Apr), 0.14 mm month^{-1} (Jun), 4.38 mm month^{-1} (Jul), 3.16 mm month^{-1} (Aug), 0.42 mm month^{-1} (Sep), 0.36 mm month^{-1} (Nov), and 2.88 mm month^{-1} (Dec). Average seasonal precipitation differences are larger for CAMELS in continental regions (Fig. 15b), ranging between 12.13 mm season^{-1} (DJF) to 15.66 mm season^{-1} (MAM), and temperate regions, ranging between 16.78 mm season^{-1} (JJA) to 24.88 mm season^{-1} (MAM). Average arid precipitation differences are larger in JJA and DJF by 7.68 mm season^{-1} and 5.27 mm season^{-1} for MOPEX values and larger in SON and MAM by 1.79 mm season^{-1} and 5.34 mm season^{-1} for CAMELS. Mean annual differences mirror the biases observed in monthly aggregations (Fig. 15c). Annual CAMELS precipitation values are 1.62 mm year^{-1} to 12.05 mm year^{-1} larger (1981, 1987, 1988, 1989, 1992, 1994, 1997, 1998, 1999) than MOPEX, while MOPEX values are 5.63 mm year^{-1} to 46.91 mm year^{-1} larger than

CAMELS for annual totals (1982, 1983, 1984, 1985, 1986, 1990, 1991, 1993, 1995, 1996, 2000) in arid regions. CAMELS values in continental regions are 27.84 mm year^{-1} to 89.23 mm year^{-1} larger, and temperate regions are 51.46 mm year^{-1} to 112.03 mm year^{-1} larger than MOPEX values.

Overall average precipitation bias CAMELS - MOPEX

475

Figure 15. Magnitude of precipitation bias averaged over all watersheds in a climate region (arid, continental, and temperate) based on a) monthly, b) seasonal, and c) annual totals. All differences are CAMELS minus MOPEX values. Positive bias indicates CAMELS > MOPEX while negative bias indicates MOPEX > CAMELS.

The average of daily temperature differences indicated MOPEX values were greater than CAMELS by 0.62 ˚C day^{-1} for arid basins,

480 0.15 ˚C day^{-1} for continental basins, and 0.35 ˚C day^{-1} for temperate basins, suggesting a warmer bias in all MOPEX values. For monthly aggregations, temperature exhibits larger values for MOPEX by 0.41 ˚C month^{-1} to 0.95 ˚C month^{-1}, 0.01 ˚C month^{-1} to 0.54 ˚C month^{-1}, and 0.24 ˚C month^{-1} to 0.64 ˚C month^{-1} in arid, continental, and temperate regions respectively (Fig. 16a). Seasonally, temperature differences indicate a warm MOPEX bias with average differences of 0.51 ˚C season^{-1} to 0.85 ˚C season^{-1} in arid, 0.03 ˚C season^{-1} to 0.41 ˚C season^{-1} in continental, and 0.25 ˚C season^{-1} to 0.52 ˚C season^{-1} in temperate regions (Fig.

485 16b). Mean annual temperature differences indicate MOPEX is greater than CAMELS by 0.43 ˚C year^{-1} to 0.86 ˚C year^{-1}, 0.07 ˚C year^{-1} to 0.29 ˚C year^{-1}, and 0.23 ˚C year^{-1} to 0.47 ˚C year^{-1} for arid, continental, and temperate regions respectively (Fig. 16c).

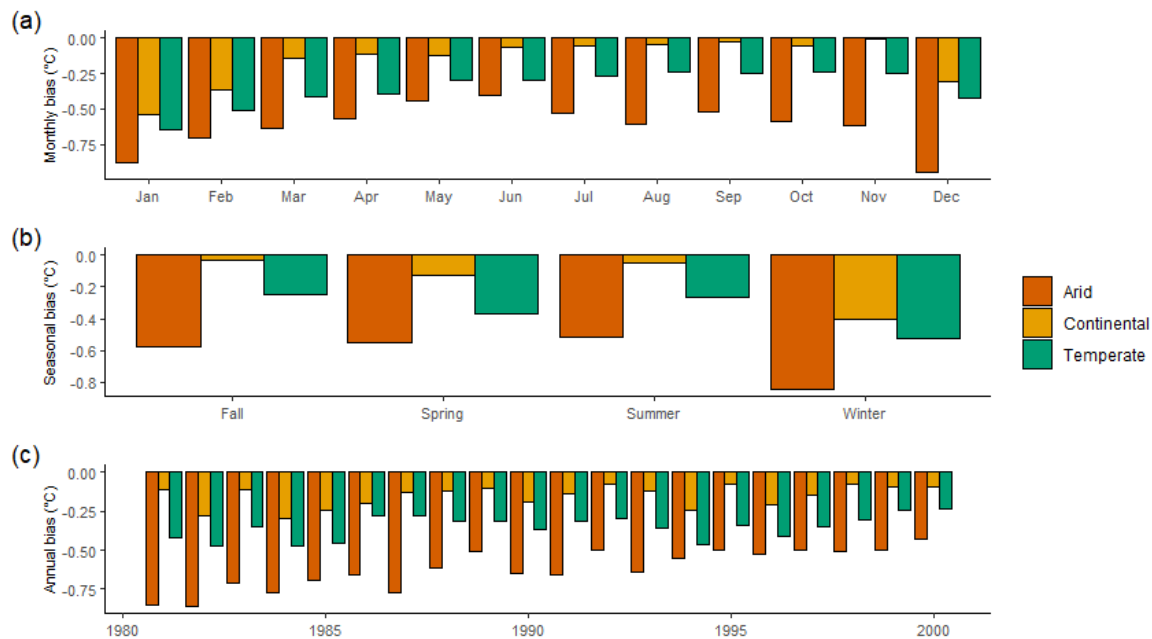



Figure 16. Magnitude of temperature bias averaged over all watersheds in a climate region (arid, continental, and temperate) based on a) monthly, b) seasonal, and c) annual totals. All differences are CAMELS minus MOPEX values. Positive bias indicates CAMELS > MOPEX while negative bias indicates MOPEX > CAMELS.

Spatial distribution of precipitation and temperature mean biases between the two datasets shows some geographic concentration, especially of positive (CAMELS) bias for precipitation in the Eastern U.S. (Fig. 17a). Arid regions show an overall wet bias for MOPEX (two watersheds have a slight wet bias for CAMELS) while continental and temperate regions have a wet bias for CAMELS for all temporal aggregations. Temperature biases in Fig. 17b show an overall warm bias for MOPEX for all regions with the exception of four continental watersheds and two temperate watersheds.

500

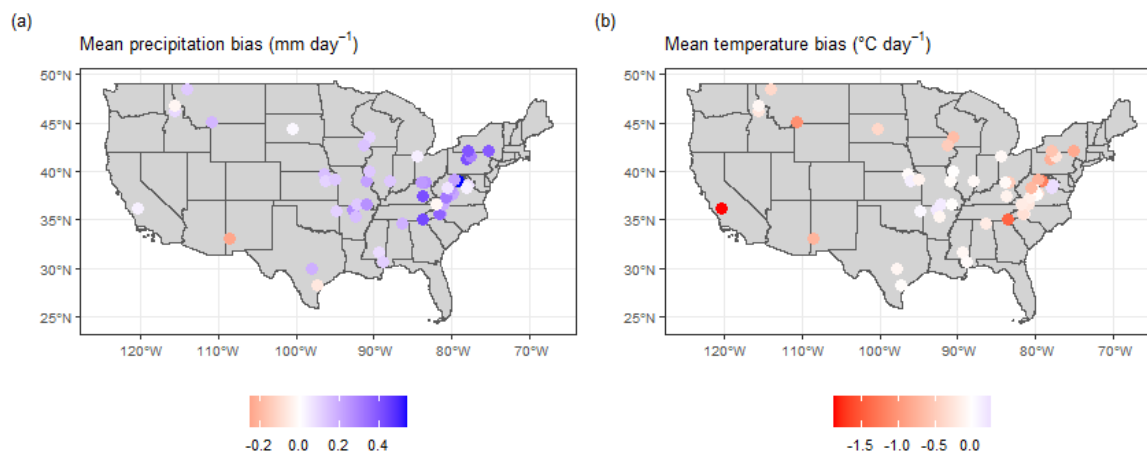



Figure 17. Overall bias for a) mean precipitation at each gauge location based on CAMELS minus MOPEX. Color bar represents bias in mm per day where negative values (red) indicate a MOPEX bias and positive values (blue) indicate a CAMELS wet bias. Overall bias for b) mean temperature at each gauge location based on CAMELS minus MOPEX. Color bar represents bias in degrees Celcius per day where negative values (red) indicate a MOPEX bias and positive values (blue) indicate a CAMELS warm bias.

To assess the magnitude of disparities between CAMELS and MOPEX, we also examined the mean absolute error (MAE), with temperate regions exhibiting greatest MAE in precipitation (Fig. 18), and arid regions greatest in temperature (Fig. 19). The overall error considers both positive and negative differences equally, with the magnitude escalating from 1.30 mm day^{-1} in arid regions to 2.70 mm day^{-1} in continental and 3.19 mm day^{-1} in temperate regions for daily precipitation. Monthly precipitation MAE ranges from 2.37 mm month^{-1} to 5.72 mm month^{-1} in arid regions, 3.74 mm month^{-1} to 6.80 mm month^{-1} for continental regions, and 5.48 mm month^{-1} to 9.62 mm month^{-1} in temperate regions (Fig. 18a). Seasonal MAE ranges from 1.79 mm season^{-1} to 7.68 mm season^{-1}, 12.13 mm season^{-1} to 15.66 mm season^{-1}, and 16.78 mm season^{-1} to 24.88 mm season^{-1} for arid, continental, and temperate regions (Fig. 18b). Annual MAE ranges from 1.62 mm year^{-1} to 46.91 mm year^{-1}, 27.84 mm year^{-1} to 89.23 mm year^{-1}, and 51.46 mm year^{-1} to 112.03 mm year^{-1} for arid, continental, and temperate regions (Fig. 18c).

515

Figure 18. Mean absolute error for a) monthly, b) seasonal, and c) annual precipitation. MAE is based on mean totals between all basins within a climate region (arid, continental, and temperate).

Daily temperature MAE averages 1.01 ˚C day^{-1} in arid regions, 0.87 ˚C day^{-1} in continental, and 0.82 ˚C day^{-1} in temperate regions. Monthly temperature MAE ranges from 0.41 ˚C month^{-1} to 0.95 ˚C month^{-1}, 0.04 ˚C month^{-1} to 0.54 ˚C month^{-1}, and

520   0.24 ˚C month^{-1} to 0.64 ˚C month^{-1} for arid, continental, and temperate regions respectively (Fig. 19a). Seasonal MAE for temperature ranges from 0.51 ˚C season^{-1} to 0.85 ˚C season^{-1}, 0.03 ˚C season^{-1} to 0.41 ˚C season^{-1}, and 0.25 ˚C season^{-1} to 0.52 ˚C season^{-1} for arid, continental, and temperate regions (Fig. 19b). Annual temperature MAE ranges from 0.43 ˚C year^{-1} to 0.86 ˚C year^{-1}, 0.07 ˚C year^{-1} to 0.29 ˚C year^{-1}, and 0.23 ˚C year^{-1} to 0.47 ˚C year^{-1} for arid, continental, and temperate regions (Fig. 19c).

525

Figure 19. Mean absolute error for a) monthly, b) seasonal, and c) annual temperature. MAE is based on means between all basins within a climate region (arid, continental, and temperate).

The statistical results for all regions are summarized in Table 7 and are calculated over all days, months, seasons, and water years (refer to Table 3). Overall statistics remove the observed fluctuations in monthly, seasonal, and annual data but provide a

530 generalized value by climate region. The margin of error (MOE) was derived from the standard error (SE) of the difference of the means and coincides with the bootstrapping results. Arid regions have the largest MOE for precipitation and temperature.

Table 7. Statistical results for comparisons between CAMELS and MOPEX values for all basins within a climate region. Analyses were conducted over the total number of values. Bias represents mean CAMELS minus mean MOPEX. Standard error (SE), margin of error (MOE), mean absolute error (MAE), and Spearman rank (R^2) are also based on mean values.

		Precipitation (mm)					Temperature (°C)				
		Bias	SE	MOE	MAE	R^2	Bias	SE	MOE	MAE	R^2
ARID	Day	-0.02	0.04	± 0.08	1.30	0.665	-0.62	0.07	± 0.14	1.01	0.992
	Month	-0.48	2.24	± 4.38	3.99	0.975	-0.62	0.37	± 0.72	0.62	0.996
	Season	-1.45	8.02	± 15.71	5.02	0.977	-0.62	0.59	± 1.16	0.62	0.993
	Year	-5.82	35.90	± 70.37	12.09	0.964	-0.62	0.72	± 1.41	0.62	0.986
CONT	Day	0.15	0.03	± 0.05	2.70	0.571	-0.16	0.04	± 0.08	0.87	0.994
	Month	4.69	1.13	± 2.22	5.16	0.973	-0.15	0.20	±0.40	0.16	0.999
	Season	13.88	3.92	± 7.69	14.05	0.978	-0.15	0.33	±0.64	0.15	0.998
	Year	56.21	14.95	± 29.31	56.22	0.956	-0.15	0.22	±0.42	0.15	0.989
TEMP	Day	0.23	0.03	± 0.05	3.19	0.609	-0.35	0.03	± 0.06	0.82	0.995
	Month	4.67	1.11	± 2.18	7.52	0.962	-0.36	0.17	± 0.33	0.35	0.999
	Season	13.85	3.88	± 7.60	21.07	0.964	-0.35	0.27	± 0.52	0.35	0.998
	Year	56.04	18.99	± 37.22	84.28	0.949	-0.35	0.27	± 0.53	0.35	0.996

535

Fligner-Killeen's test for equality of variances indicated that arid regions are the most similar in precipitation variance for all temporal aggregations except daily (Table 8). Statistically significant differences between variances were found for continental regions (daily, monthly, seasonally) and temperate regions (daily and monthly) but not on an annual basis. Temperature values are more consistent with statistically significant differences between variances indicated for daily values only. These results are corroborated by observations previously presented and outlined in Sect. 4.1, as shown in Tables 4 and 5, and Figs. 4 through 10.

540

Table 8. Fligner-Killeen's test for homogeneity of variance results. Df1 is the number of groups minus 1, statistic is Chi-squared. P-values are reported as * p < 0.05, ** p < 0.01, *** p < 0.001.

		Precipitation			Temperature		
		df1	Chi-squared	p	df1	Chi-squared	p
ARID	Day	1	9661.2	**<0.001*****	1	38.22	**<0.001*****
	Month	1	0.15	0.70	1	1.16	0.28
	Season	1	0.10	0.75	1	1.00	0.32
	Year	1	0.09	0.76	1	1.35	0.25
CONT	Day	1	14432	**<0.001*****	1	60.52	**<0.001*****
	Month	1	23.24	**<0.001*****	1	2.86	0.09
	Season	1	4.96	**0.03***	1	1.98	0.16
	Year	1	0.69	0.41	1	2.62	0.11
TEMP	Day	1	11096	**<0.001*****	1	89.83	**<0.001*****
	Month	1	21.54	**<0.001*****	1	3.32	0.07
	Season	1	2.68	0.10	1	1.86	0.17
	Year	1	0.67	0.41	1	2.69	0.10

545

Differences in the mean values, evaluated by Welch's t-test, indicated that there were no statistically significant differences in the mean for daily, monthly, seasonal, and annual arid precipitation (Table 9). Despite the largest variance, arid regions are the most similar with the smallest amount of bias between means. Differences in mean precipitation values in continental and temperate regions are statistically significant on all temporal scales, however, the largest difference between mean values is only 6.36 % (daily temperate). Temperature mean differences are only statistically significant at daily aggregations for all climate regions, with the exception of monthly temperature.

550

555

560

Table 9. Welch's t-test for comparison of means with unequal variance results. The n represents the number of values, df is calculated degrees of freedom, statistic is calculated t. P-values are reported as * p < 0.05, ** p < 0.01, *** p < 0.001.

		Precipitation				Temperature			
		n	df	statistic	p	n	df	statistic	p
ARID	Day	36525	73045.58	-0.37	0.71	36525	72950.21	-8.70	**<0.001***
	Month	1200	2393.65	-0.22	0.83	1200	2395.78	-1.70	0.09
	Season	400	797.34	-0.18	0.85	400	797.19	-1.05	0.29
	Year	100	197.15	-0.16	0.87	100	197.54	-0.86	0.39
CONT	Day	131490	262467.77	5.91	**<0.001***	131490	262734.92	-3.76	**<0.001***
	Month	4320	8589.14	4.14	**<0.001***	4320	8632.98	-0.75	0.45
	Season	1440	2867.28	3.58	**<0.001***	1440	2875.99	-0.47	0.64
	Year	360	715.67	3.76	**<0.001***	360	711.07	-0.71	0.48
TEMP	Day	175320	350554.08	8.72	**<0.001***	175320	350322.75	-10.64	**<0.001***
	Month	5760	11481.67	6.32	**<0.001***	5760	11513.38	-2.11	**0.03***
	Season	1920	3829.89	5.43	**<0.001***	1920	3836.39	-1.32	0.19
	Year	480	957.97	4.44	**<0.001***	480	956.60	-1.29	0.20

565   The non-parametric Wilcoxon signed-rank test (Helsel et al., 2020) was also conducted to evaluate the median differences and the results indicated statistically significant differences for daily precipitation in all regions, and for monthly, seasonal, and annual precipitation in continental and temperate regions. Temperature median differences were only statistically significant for daily values.

570   As previously noted, there are no discrepancies in runoff between MOPEX and CAMELS datasets because both contain identical daily streamflow values sourced from the USGS. However, evapotranspiration estimates derived from the water balance approach will differ due to variations in precipitation, since runoff remains consistent across the datasets. Runoff and water balance derived evapotranspiration were included in correlation analyses to evaluate the relationships among all variables for both datasets and to determine consistency in the strength and direction of their associations.

575

Daily precipitation Spearman Rank correlation values between CAMELS (_C) and MOPEX (_M) ranged from 0.58 to 0.74, 0.48 to 0.86, and 0.46 to 0.88 for arid, continental, and temperate regions respectively (Fig. 20). The highest precipitation correlation values were observed for monthly and seasonal aggregations (red cells, Fig. 20), with annual values following closely. Monthly precipitation correlations are the lowest in July and August for all regions (0.84 to 0.91). Monthly and seasonal aggregations are

580   the most consistent between MOPEX and CAMELS, followed by annual and then daily for precipitation and temperature. Temperature shows a high similarity between MOPEX and CAMELS for all temporal aggregations and regions, ranging from 0.99 to 1.0. Correlations between runoff and precipitation are positive for all regions and temporal aggregations in both datasets with the largest difference of 0.08 in daily continental (0.14 for MOPEX, 0.22 for CAMELS). Water balance evapotranspiration values show improved agreement, greater than 0.90, for monthly, seasonal, and annual aggregations. Daily evapotranspiration coefficients

585   are between 0.59 to 0.64, indicating less consistency between MOPEX and CAMELS.

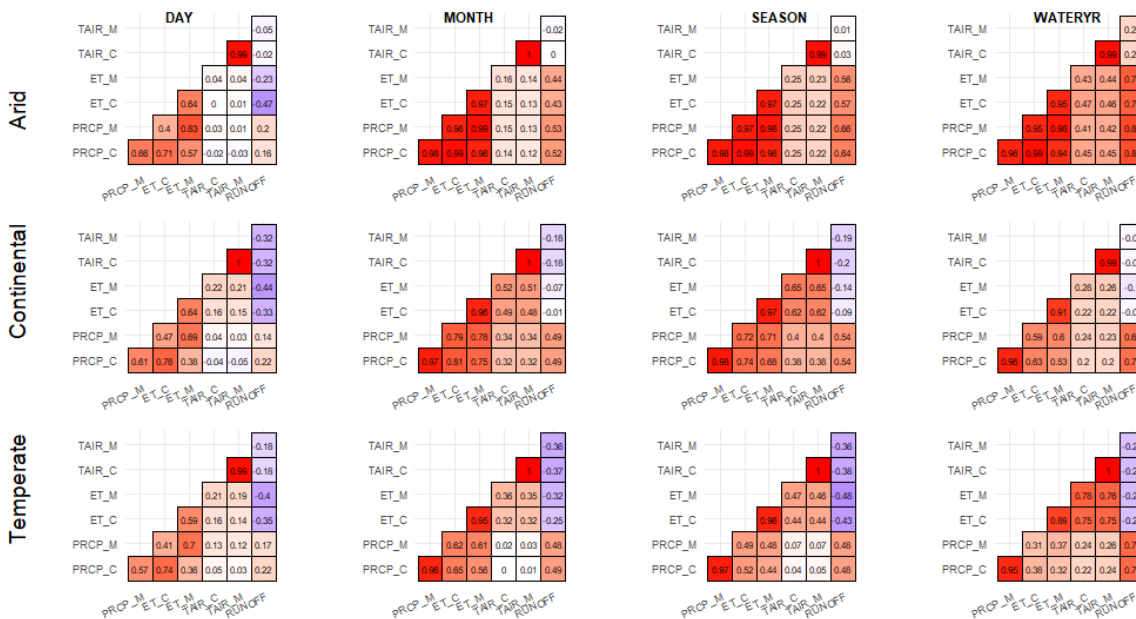



Figure 20. Spearman rank correlation values between CAMELS (_C) and MOPEX (_M) datasets for precipitation, temperature, and water balance calculated evapotranspiration. Runoff represents both datasets.

590

Runoff efficiency is the amount of precipitation that becomes runoff and can be used to evaluate trends and climate impact. This coefficient provides an additional metric of dataset compatibility. The annual efficiency for each basin was determined for CAMELS and MOPEX using total precipitation and total runoff and then plotted, resulting in an R^2 value of 0.988 for all climate regions combined (Fig. 21). This correlation was conducted to illustrate the annual compatibility of the datasets and the ability of both to convey consistent attributes among watersheds for derived parameters, such as runoff efficiency.

595

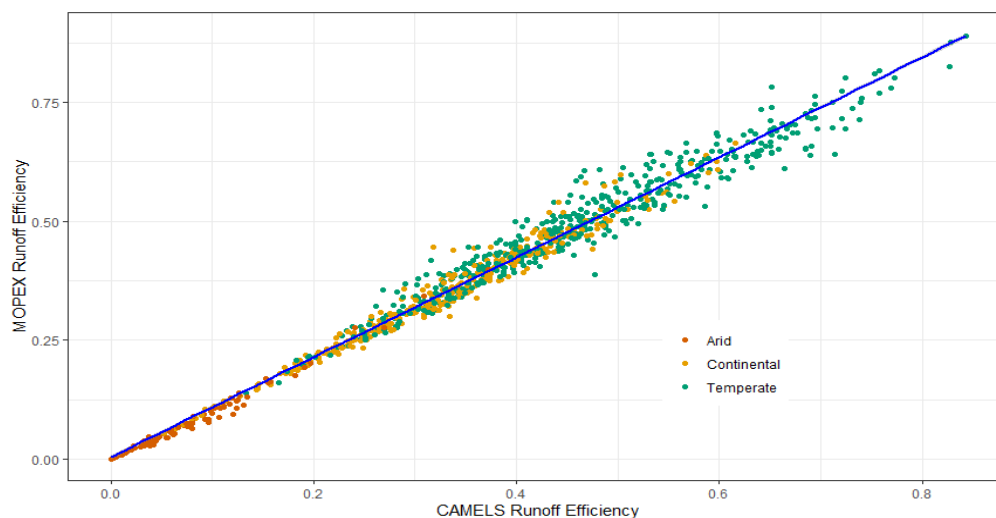



Figure 21. Annual runoff coefficient (runoff/precipitation) for each basin. Colored points represent climate region (arid, continental, and temperate). The blue line indicates the best linear fit.

4.3 Evaluation of precipitation and temperature extremes between datasets

600   While data extremes were not the focus of this study, a few precipitation and temperature extreme indices were evaluated for CAMELS and MOPEX values. The number of heavy precipitation days, where daily precipitation is greater than or equal to 10 mm, are more prolific in CAMELS, consistent with the wet bias (Fig. 22a). Despite the difference in the number of days, the two datasets show the same trends over time from 1981 to 2000. The number of dry days (precipitation < 1mm) per year are greater in CAMELS for all climate regions, with the largest discrepancies in arid regions (Fig. 22b). This study has shown that

605   CAMELS has a wet bias for continental and temperate regions and MOPEX has a wet bias for arid regions (Fig. 15). The differences in the number of dry days show that CAMELS daily precipitation values are overall, larger than MOPEX values (Fig. 22b).

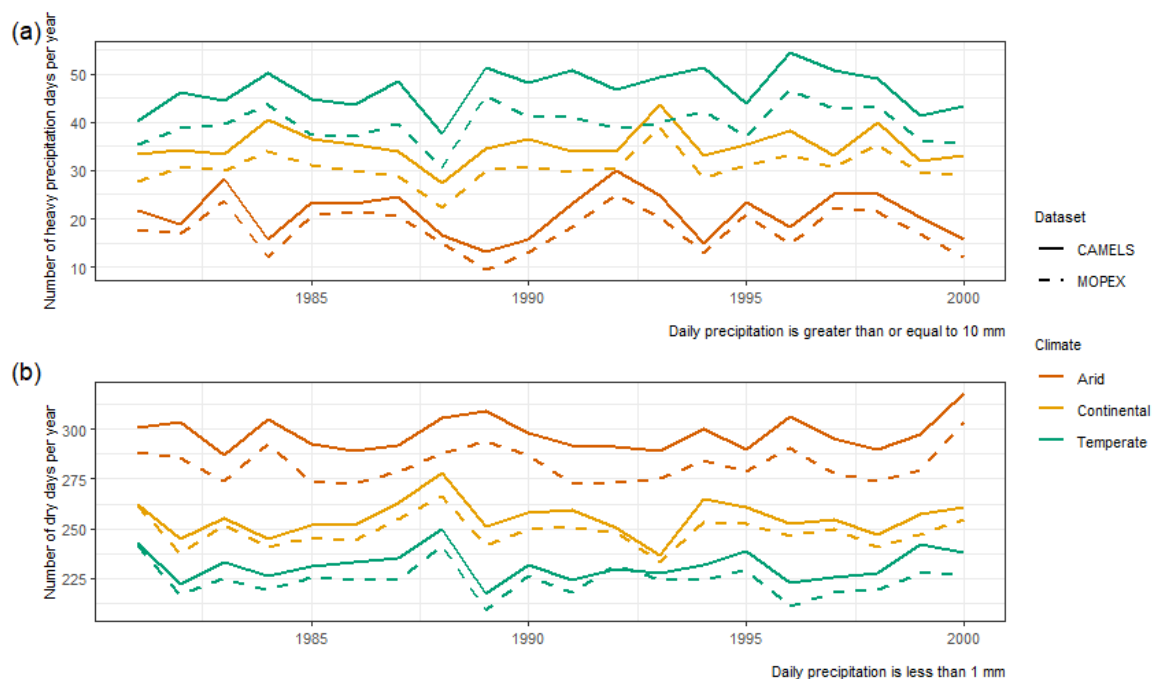



Figure 22. Precipitation indices which show the a) annual count of heavy precipitation days where precipitation is ≥ 10 mm and b) annual
610   **count of dry days where precipitation amount is < 1mm. Colors represent the climate region (arid, continental, and temperate), dashed lines represent MOPEX, and solid lines represent CAMELS.**

     The extremely wet day rainfall, R99p, represents the annual total precipitation when daily rainfall is greater than the 99th percentile and when plotted for both datasets by climate region, very similar trends are observed (Fig. 23). In a broad temporal context, analysis consistently shows that precipitation values tend to be larger in CAMELS, regardless of the temporal scale

615   considered beyond paired daily values. This pattern is observed in monthly, seasonal, and annual aggregations as well as summarized daily mean for continental and temperate regions.

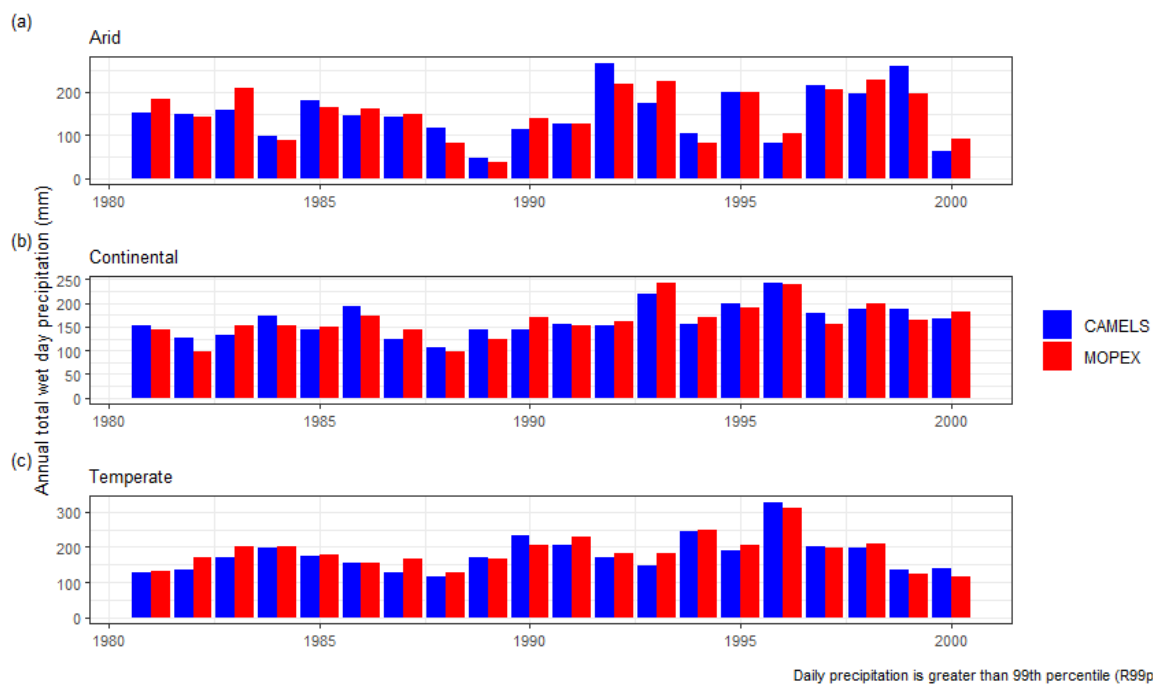



Figure 23. Annual total precipitation when daily precipitation amount on a wet day (≥ 1mm) is greater than the 99th percentile for a) arid, b) continental, and c) temperate regions. Colors (blue and red) represent the dataset.

620   In contrast, temperature values exhibit a different trend, with MOPEX consistently showing larger values irrespective of the temporal aggregation or climate region (Fig. 16). The number of frost days (Fig. 24a) indicate the annual count per year where temperature falls below 0 ˚C. CAMELS has a greater number of cold days which corresponds to warmer MOPEX bias. The warm MOPEX bias is most prevalent in arid regions when evaluating the number of summer days per year, the annual count of days with temperatures above 25 ˚C (Fig. 24b).

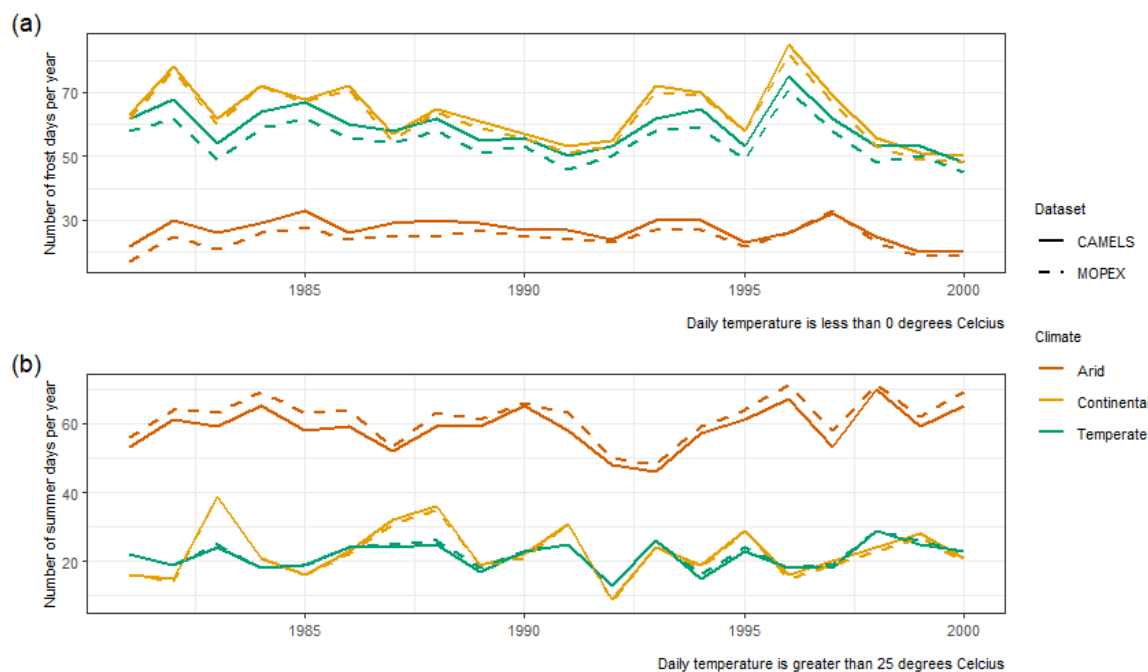



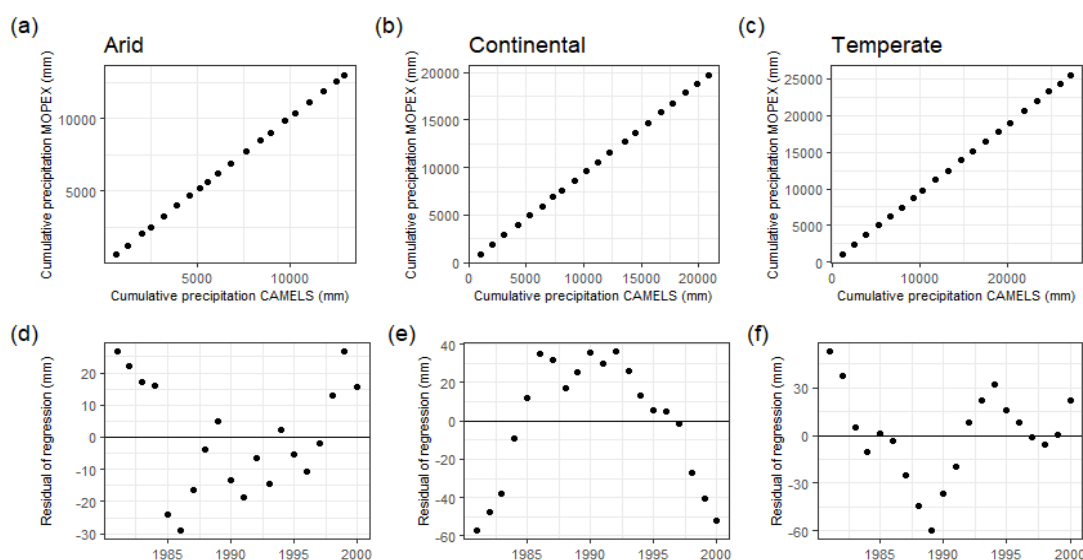
Figure 24. Temperature indices which show the a) annual count of frost days where temperature is < 0 ˚C and b) annual count of summer days where temperature is > 25 ˚C. Colors represent the climate region (arid, continental, and temperate), dashed lines represent MOPEX, and solid lines represent CAMELS.

4.4 Validation

Machine learning has been used to successfully validate data and models (Patton et al., 2022). For this study, data were separated by climate region and then precipitation, temperature, and water balance derived evapotranspiration were used for daily, monthly, seasonal, and annual classifications. The support vector machine performed binary classification on standardized values. The classification accuracy values shown in Table 10 represent the model's ability to identify which dataset the values belong to. If both datasets are considered equal, then the probability of choosing the correct dataset based on a selection of precipitation, temperature, and evapotranspiration values would be 0.5. Accuracy ranges from 49 to 53 % with most classifications close to 50 %. The model's difficulty to successfully classify the data demonstrates the relative similarity of MOPEX and CAMELS for all climate regions and temporal aggregations.

Table 10. Classification accuracy results for support vector models.

	Day	Month	Season	Year
Arid	0.517	0.490	0.495	0.480
Continental	0.505	0.521	0.529	0.523
Temperate	0.508	0.524	0.530	0.510

Annual precipitation similarity was evaluated using the double mass curve (Searcy et al., 1960), a comparative analysis that can identify changes in values over time, examine data consistency, and provide validation (Gao et al., 2017). Cumulative values of two variables plotted against each other should display a linear relationship if the ratio between them is constant. Breaks in the
645   slope can indicate changes in the data and the time it occurred. When cumulative precipitation values are plotted for CAMELS and MOPEX, the slopes are 0.99, 1.06, and 1.07 for arid (Fig. 25a), continental (Fig. 25b), and temperate regions (Fig. 25c) respectively. There are apparent trends in the residuals for continental (Fig. 25e) and temperate (Fig. 25f) regions which could be due to bias, however, the residuals are small, within ± 60 mm.

650   **Figure 25. Cumulative annual precipitation of CAMELS versus MOPEX for each climate region a) arid, b) continental, c) temperate for all water years 1981 – 2000. Residuals of regression are shown for each double mass curve d) arid, e) continental, f) temperate.**

5 Discussion

The comprehensive results above indicate important biases between the CAMELS and MOPEX datasets which vary in potential importance by climate region, geographic location and degree of temporal aggregation. The findings underscore the need for
655   careful consideration of dataset disparities, acknowledging the impact of temporal scale and methodology. As one would expect, comparability between CAMELS and MOPEX increases with greater temporal aggregation. The uncertainty and variability within each dataset were evaluated independently by calculating variance, standard deviation, and mean confidence intervals with bootstrapping (Tables 4-5). The uncertainty and variability between the datasets were evaluated by paired tests, bias, MAE, MOE, R^2, and hypothesis tests. By these measures monthly, seasonal, and annual precipitation and temperature values are
660   comparable for MOPEX and CAMELS (Table 7). "Magnitudes of difference or trends within data products may be comparable to the magnitude of difference between data products (Levy et al., 2017)." Essentially, the datasets share similar uncertainties and variances.

Statistically significant differences between variance and mean were more prevalent for precipitation than temperature, however,
665 bootstrapping results indicate that both datasets have similar uncertainties for their mean values, with frequent overlap of the
confidence intervals (Tables 4-5). The most substantial differences in mean values are observed in daily aggregations, with
discrepancies of 1.13 %, 5.41 %, and 6.36 % for arid, continental, and temperate regions, respectively. At the annual scale, these
differences decrease to 0.87 % for arid regions and 4.11 % for temperate regions. Continental regions unexpectedly show an
increased difference of 5.54 % at the annual scale, which could be attributed to higher interannual variability or spatial
670 heterogeneity. Median differences in precipitation show significant improvement with temporal aggregation: arid regions
decrease from 200 % to 4.88 %, continental regions from 142.86 % to 5.70 %, and temperate regions from 118.37 % to 5.63 %.

Arid regions have the largest margin of error for precipitation, but the smallest percent difference between their daily, monthly,
seasonal, and annual mean values. Conversely, for temperature, arid regions have the largest percent difference between mean
values in addition to the largest margins of error compared to continental and temperate regions. Correlations improve from 0.57
675 to 0.67 for daily precipitation values to greater than 0.90 for monthly, seasonal, and annual totals. Temperature correlation values
are 0.99 for all regions. Positive correlation between temperature/evapotranspiration and runoff in arid regions is possibly due to
increased rates of soil drying and decreased soil moisture retention, which then leads to greater runoff. Intense precipitation
events with higher temperatures can also lead to less infiltration and more runoff. Climate patterns, such as rainy or dry seasons,
are captured similarly in both datasets and the more consistent pattern may contribute to higher monthly and seasonal
680 correlations compared to annual.

MOPEX has an overall warm bias for all climate regions and a wet bias for arid regions. CAMELS has an overall wet bias for
temperate and continental regions. MOPEX is approximately 0.62 ˚C, 0.16 ˚C, and 0.35 ˚C warmer than CAMELS for arid,
continental, and temperate regions respectively on all temporal scales. For precipitation, MOPEX is larger than CAMELS by 0.02
mm day^{-1}, 0.48 mm month^{-1}, 1.45 mm season^{-1}, and 5.82 mm year^{-1} for arid regions. CAMELS shows a wet bias for continental,
685 ranging from 0.15 mm day^{-1} to 56.21 mm year^{-1}, and temperate, with biases of 0.23 mm day^{-1} to 84.27 mm year^{-1}. Other comparison
studies have also shown a warm MOPEX bias when compared to Daymet and a wet Daymet bias compared MOPEX (Essou et al.,
2016) and PRISM (Muche et al., 2020). Arid regions have the largest variance and uncertainty which could be due to high
evaporation rates, diverse landscapes, or the limited number of stations could influence bootstrapping results, incorporating fewer
values compared to temperate or continental landscapes.
690

Precipitation exhibits a more pronounced positive bias on an annual scale (Fig. 13c), primarily due to the accumulation of small
positive biases observed on monthly and seasonal scales. On finer temporal scales, these individual biases may partially offset each
other due to seasonality, leading to less noticeable discrepancies yet when data is aggregated annually, consistent overestimations
are amplified, resulting in a more evident positive bias. This pattern highlights the presence of a systematic wet bias, where
695 precipitation is consistently overestimated across temporal scales. The effect becomes more apparent at larger aggregation
intervals, particularly due to increasing precipitation totals.

For temperature, although larger biases are observed in the 1980s, this pattern is limited to arid regions. Even in later years, the
outliers for arid regions remain close to -2 °C. The medians for 1981, 1983, and 1987 are near zero, indicating minimal central
700 tendency bias for those years. When averaged over time, as shown in Fig. 16, there is a slight improvement in biases for arid
regions, with a reduction of approximately 0.25 °C from 1981 to 2000. An increase in station density could also contribute to the
observed improvement in biases. The GHCNd database, accessed via the National Centers for Environmental Information (NCEI),

indicates a noticeable rise in the number of precipitation and temperature stations during the late 1990s and early 2000s. This increased station coverage likely enhanced the spatial representation of observations, reducing biases and improving the accuracy
705   of aggregated data.

Biases have potential consequences for dependent variables like evapotranspiration and runoff. A wet bias in daily precipitation could propagate and inflate the estimates of evapotranspiration and runoff, with hydrological models possibly predicting higher water availability across the system than actually present. Precipitation and temperature data sources differ between MOPEX and
710   CAMELS, resulting in the observed biases, however, statistical analyses and validation in this study demonstrate that the datasets closely resemble one another. Precipitation biases are between -0.25 to 0.54 mm day^{-1} and temperature biases are between -1.88 to 0.27 ˚C day^{-1}. These biases exhibit little influence on general hydrologic indices such as runoff efficiency (Fig. 21). Paired same-day comparisons of CAMELS versus MOPEX indicate that the largest discrepancies are between daily values for precipitation and temperature. Combining these datasets is not recommended for evaluating daily events such as maximum 1 day
715   precipitation, however, both datasets exhibit the same general trends for seasonality and climatic indices (Sect. 4.3).

Climate data inherently carries a degree of uncertainty, and the biases described above derive from varying data sources and interpolation of those data to the watershed footprint. Factors such as heterogeneity, gaps in coverage, and interpolation methods contribute to deviations from precision. Data sources can include gauges, satellites, or a combination of both. No two datasets will be identical, as discrepancies occur across various scales. The primary objective is to choose the most appropriate, quality-
720   controlled, accurate, and representative dataset for the research at hand. Previous comparison studies have highlighted the inconsistencies and biases between data products (Levy et al., 2017; Sun et al., 2018; Mallakpour et al., 2022; Prat et al., 2015).

Differences are expected between the datasets based on their sources and processing. Essou et al. (2016) compared three gridded datasets that interpolated daily temperature and precipitation values from the same observation network and noted that "the differences between gridded products may largely be attributed to the interpolation schemes which differ substantially from one
725   dataset to another." MOPEX estimated daily mean areal precipitation and temperature by interpolating NCDC COOP and SNOTEL station data using an inverse distance algorithm. CAMELS used gridded daily precipitation and temperature data from Daymet version 2 which used a Gaussian convolution kernel interpolation (Thornton et al., 2012). While there are some notable differences in daily events, the results from this study show that MOPEX and CAMELS consistently demonstrate the same trends in space and time.

730   **6 Conclusions**

In this study, we evaluated two large sample datasets, MOPEX and CAMELS, comparing precipitation and temperature. The current MOPEX dataset contains data for 431 watersheds within the CONUS while CAMELS includes data for 671 watersheds. The datasets were combined for this study and 47 common basins were compared for water years 1981 through 2000 on daily, monthly, seasonal, and annual scales. Precipitation, temperature, and streamflow data are areally weighted by delineated
735   boundaries available as shapefiles. The main conclusions from the statistical comparison between CAMELS and MOPEX at daily, monthly, seasonal, and annual scales are summarized as follows:

1) Daily pairwise comparisons are not recommended due to the variability in extreme precipitation event measurements. However, both datasets capture similar patterns and basin behavior, e.g. when evaluating the number of rainy days or dry days per year.

2) Comparison improves significantly with monthly, seasonal, and annual aggregations.

3) Compatibility is constrained by basin water balance and requires basin averaged values, i.e. ET values from CAMELS must be used with caution, and often cannot be reconciled with MOPEX or other water-balance based estimates

The comparative analysis of the MOPEX and CAMELS datasets reveals distinct biases and variability patterns across climate regions and temporal scales. CAMELS generally exhibits a positive precipitation bias at monthly, seasonal, and annual aggregations in continental and temperate regions, while MOPEX shows higher daily precipitation values, particularly in arid regions. Temperature analysis highlights a consistent warm bias in MOPEX across all regions and time scales, with notable disparities in daily values. Despite these differences, the datasets show overlapping confidence intervals for many metrics, suggesting similar levels of uncertainty.

The observed variations, particularly for extreme precipitation events, underscore the necessity for cautious interpretation of dataset-specific results. For applications requiring precision, such as hydrological modeling or climate analysis, direct substitution of daily values between MOPEX and CAMELS is not advisable without considering these biases. Instead, leveraging insights from both datasets can provide a more comprehensive understanding of regional and temporal climate characteristics.

Ongoing research aims to extend the MOPEX dataset from 2003 to 2023 and the CAMELS dataset from 2014 to 2023, leveraging the Daymet dataset (Thornton et al., 2021). MOPEX and CAMELS will be integrated into a cohesive resource that combines catchment attributes and human impact classifications based on the GAGESII framework (Falcone et al., 2010). The enhanced dataset will support model calibration and freshwater balance studies at the watershed scale (in progress, Sink et al.). Basin-scale analyses and forecasts are expected to benefit from more precise water balance constraints, improving their accuracy and predictive power.

Code and data availability. A repository with R Code used for analyses and resulting data is available from https://github.com/k-sink/Toward_merging. Colors are based on https://jfly.uni-koeln.de/color/#pallet by Okabe-Ito (2008). Although the MOPEX dataset (Schaake et al., 2006) was publicly available at the time of analysis, recent attempts to access the data at http://hydrology.nws.noaa.gov/pub/gcip/-mopex/US_Data were unsuccessful. The CAMELS dataset (Addor et al., 2017) is available from https://gdex.ucar.edu/dataset/camels.html. Streamflow data and attributes are available from https://waterdata.usgs.gov/nwis/inventory.

Author contributions. KS and TB conceptualized the research project. KS developed and performed the formal analysis. KS prepared the paper with contributions from the co-author.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Thank you to my advisor, Tom Brikowski, for his helpful discussions and comments during this research project. Data processing and visualization were performed with R (R Core Team, 2024). UTD SESS contribution #1728.

References

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., and Mendoza, P. A.: Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges. *Hydrological Sciences Journal*, *65*(5), 712–725. https://doi.org/10.1080/02626667.2019.1683182, 2020.

780 Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, *21*(10), 5293–5313. https://doi.org/10.5194/hess-21-5293-2017, 2017.

Andréassian, V., Perrin, C., and Michel, C.: Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. *Journal of Hydrology*, *286*(1–4), 19–35. https://doi.org/10.1016/j.jhydrol.2003.09.030, 2004.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25°
785 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrol. Earth Syst. Sci., 21, 589–615, https://doi.org/10.5194/hess-21-589-2017, 2017.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*, 1–12. https://doi.org/10.1038/sdata.2018.214, 2018.

Berghuijs, W. R., Sivapalan, M., Woods, R. A., and Savenije, H. H. G.: Patterns of similarity of seasonal water balances: A window into
790 streamflow variability over a range of time scales. *Water Resources Research*, *50*, 5638–5661. https://doi.org/10.1002/2014WR015692, 2014.

Buban, M. S., Lee, T. R., and Baker, C. B.: A comparison of the U.S. climate reference network precipitation data to the parameter-elevation regressions on independent slopes model (PRISM). *Journal of Hydrometeorology*, *21*(10), 2391–2400. https://doi.org/10.1175/JHM-D-19-0232.1, 2020.

795 Budyko, M.I. Climate and life, Academic Press, New York, 508pp., ISBN 0121394506, 1974.

Carter, E., Hain, C., Anderson, M., and Steinschneider, S.: A water balance-based, spatiotemporal evaluation of terrestrial evapotranspiration products across the contiguous United States. *Journal of Hydrometeorology*, *19*(5), 891–905. https://doi.org/10.1175/JHM-D-17-0186.1, 2018.

Chao, L., Zhang, K., Wang, J., Feng, J., and Zhang, M.: A comprehensive evaluation of five evapotranspiration datasets based on ground
800 and grace satellite observations: Implications for improvement of evapotranspiration retrieval algorithm. *Remote Sensing*, *13*(12). https://doi.org/10.3390/rs13122414, 2021.

Coopersmith, E., Yaeger, M. A., Ye, S., Cheng, L., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves - Part 3: A catchment classification system based on regime curve indicators. *Hydrology and Earth System Sciences*, *16*(11), 4467–4482. https://doi.org/10.5194/hess-16-4467-2012, 2012.

805 Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y. M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O. N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E. F.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, *320*(1–2), 3–17. https://doi.org/10.1016/j.jhydrol.2005.07.031, 2006.

Essou, G. R. C., Arsenault, R., and Brissette, F. P.: Comparison of climate datasets for lumped hydrological modeling over the continental
810 United States. *Journal of Hydrology*, *537*, 334–345. https://doi.org/10.1016/j.jhydrol.2016.03.063, 2016.

Falcone, J. A., Carlisle, D. M., Wolock, D. M., and Meador, M. R.: GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States. In *Data Papers Ecology* (Vol. 91, Issue 2). http://esapubs.org/archivei, 2010.

Gao, P., Li, P., Zhao, B., Xu, R., Zhao, G., Sun, W., and Mu, X.: Use of double mass curves in hydrologic benefit evaluations. *Hydrological Processes*, *31*(26), 4639–4646. https://doi.org/10.1002/hyp.11377, 2017.

816 Guo, H.: Big Earth data: A new frontier in Earth and information sciences. *Big Earth Data*, *1*(1–2), 4–20. https://doi.org/10.1080/20964471.2017.1403062, 2017.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., and Andréassian, V.: Large-sample hydrology: A need to balance depth with breadth. *Hydrology and Earth System Sciences*, *18*(2), 463–477. https://doi.org/10.5194/hess-18-463-2014, 2014.

820 Han, E., Crow, W. T., Hain, C. R., and Anderson, M. C.: On the use of a water balance to evaluate interannual terrestrial ET variability. *Journal of Hydrometeorology*, *16*(3), 1102–1108. https://doi.org/10.1175/JHM-D-14-0175.1, 2015.

Han, P. F., Sankarasubramanian, A., Wang, X. S., Wan, L., and Yao, L.: One-Parameter Analytical Derivation in Modified Budyko Framework for Unsteady-State Streamflow Elasticity in Humid Catchments. *Water Resources Research*, *59*(9). https://doi.org/10.1029/2023WR034725, 2023.

825 Hartigan, J. A., and Wong, M. A.: Algorithm AS 136: A K-Means Clustering Algorithm. In *Source: Journal of the Royal Statistical Society. Series C (Applied Statistics)* (Vol. 28, Issue 1), 1979.

Helsel, D. R., Hirsch, R. M., Ryberg, K. R., Archfield, S. A., and Gilroy, E. J.: Statistical Methods in Water Resources Techniques and Methods 4 – A3. *USGS Techniques and Methods*, 2020.

Kelleher, C., and Braswell, A.: Introductory overview: Recommendations for approaching scientific visualization with large
830 environmental datasets. *Environmental Modelling and Software*, *143*(June), 105113. https://doi.org/10.1016/j.envsoft.2021.105113, 2021.

Lemaitre-Basset, T., Oudin, L., Thirel, G., and Collet, L.: Unraveling the contribution of potential evaporation formulation to uncertainty under climate change. *Hydrology and Earth System Sciences*, *26*(8), 2147–2159. https://doi.org/10.5194/hess-26-2147-2022, 2022.

Levy, M. C., Cohn, A., Lopes, A. V., and Thompson, S. E.: Addressing rainfall data selection uncertainty using connections between
835 rainfall and streamflow. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-00128-5, 2017.

Lins, H. F.: *USGS Hydro-Climatic Data Network 2009 (HCDN-2009)*. http://wdr.water.usgs.gov/, 2012.

Mallakapour, I., Sadeghi, M., Mosaffa, H., Asanjan, A.A., Sadegh, M., Nguyen, P., Sorooshian, S., Aghakouchak, A.: Discrepancies in changes in precipitation characteristics over the contiguous United States based on six daily gridded precipitation datasets. *Weather and Climate Extremes, 36(100433),* 1-9. https://doi.org/10.1016/j.wace.2022.100433, 2022.

840 Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., and Nijssen, B.: A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States*. *Journal of Climate*, *15*(22), 3237–3251. https://doi.org/10.1175/1520-0442(2002)015, 2002.

Miller, D. A., and White, R. A.: A Conterminous United States Multilayer Soil Characteristics Dataset for Regional Climate and Hydrology Modeling. *Earth Interactions*, *2*(1), 2–2. https://doi.org/10.1175/1087-3562(1998)002, 1998.

845 Muche, M. E., Sinnathamby, S., Parmar, R., Knightes, C. D., Johnston, J. M., Wolfe, K., Purucker, S. T., Cyterski, M. J., and Smith, D.: Comparison and Evaluation of Gridded Precipitation Datasets in a Kansas Agricultural Watershed Using SWAT. *Journal of the American Water Resources Association*, *56*(3), 486–506. https://doi.org/10.1111/1752-1688.12819, 2020.

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., Mizukami, N., Brekke, L., and Arnold, J. R.: Gridded ensemble precipitation and temperature estimates for the contiguous United States. *Journal of Hydrometeorology*, *16*(6), 2481–
850 2500. https://doi.org/10.1175/JHM-D-15-0026.1, 2015.

Newman, A. J., Clark, M. P., Longman, R. J., and Giambelluca, T. W.: Methodological intercomparisons of station-based gridded meteorological products: Utility, limitations, and paths forward. *Journal of Hydrometeorology*, *20*(3), 531–547. https://doi.org/10.1175/JHM-D-18-0114.1, 2019.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson,
855   T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set
        characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*,
        19(1), 209–223. https://doi.org/10.5194/hess-19-209-2015, 2015.

Oubeidillah, A. A., Kao, S. C., Ashfaq, M., Naz, B. S., and Tootle, G.: A large-scale, high-resolution hydrological model parameter data
        set for climate change impact assessment for the conterminous US. *Hydrology and Earth System Sciences*, *18*(1), 67–84.
860   https://doi.org/10.5194/hess-18-67-2014, 2014.

Patton, D., Smith, D., Muche, M. E., Wolfe, K., Parmar, R., and Johnston, J. M.: Catchment scale runoff time-series generation and
        validation using statistical models for the Continental United States. *Environmental Modelling and Software*, *149*.
        https://doi.org/10.1016/j.envsoft.2022.105321, 2022.

Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G., Williams, Z., Brunke, M. A., and Gochis, D.: A gridded
865   global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *Journal
        of Advances in Modeling Earth Systems*, *8*(1), 41–65. https://doi.org/10.1002/2015MS000526, 2016.

Pimentel, R., Arheimer, B., Crochemore, L., Andersson, J. C. M., Pechlivanidis, I. G., and Gustafsson, D.: Which Potential
        Evapotranspiration Formula to Use in Hydrological Modeling World-Wide? *Water Resources Research*, *59*(5).
        https://doi.org/10.1029/2022WR033447, 2023.

870 Prat, O. P., and Nelson, B. R.: 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*(4), 2037–2056. https://doi.org/10.5194/hess-19-
        2037-2015, 2015.

R Core Team (2024). R: A language and environment for statistical computing, available from http://www.r-project.org/, 2024.

Reis, A.A., Weerts, A., Ramos, M., Wetterhall, F. Fernandes, W.: Hydrological data and modeling to combine and validate precipitation
875   datasets   relevant   to   hydrological   applications.   *Journal   of   Hydrology: Regional   Studies,   44,*   1-19.
        https://doi.org/10.1016/j.ejrh.2022.101200, 2022.

Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., and Carrillo, G.: Characterizing hydrologic change through catchment
        classification. *Hydrology and Earth System Sciences*, *18*(1), 273–285. https://doi.org/10.5194/hess-18-273-2014, 2014.

Schaake, J., Cong, S., and Duan, Q.: The US mopex data set. *IAHS-AISH Publication*, *307*, 9–28, 2006.

880 Seaber, P. R., Kapinos, F. P., and Knapp, G. L.: USGS Watershed Delineation. *United States Geological Survey Water Supply Paper*,
        2294, 1987.

Searcy, J. K., Hardison, C. H., and Langbein, W. B.: *Double-Mass Curves With a section Fitting Curves to Cyclic Data Manual of
        Hydrology: Part 1. General Surface-Water Techniques*, 1960.

Shmilovici, A.: Machine Learning for Data Science Handbook. In L. Rokach, O. Maimon, and E. Shmueli (Eds.), *Machine Learning for
885   Data Science Handbook*. Springer International Publishing. https://doi.org/10.1007/978-3-031-24628-9, 2023.

Sitterson, J., Sinnathamby, S., Parmar, R., Koblich, J., Wolfe, K., and Knightes, C. D.: Demonstration of an online web services tool
        incorporating automatic retrieval and comparison of precipitation data. *Environmental Modelling and Software*, *123*.
        https://doi.org/10.1016/j.envsoft.2019.104570, 2020.

Slack, J. R., and Landwehr, J. M.: Hydro-Climatic Data Network (HCDN): A U.S. Geological Survey Streamflow Data Set for the United
890   States for the Study of Climate Variations, 1874-1988. *Open-File Report 92-129,* 1994.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K. L. (2018). A Review of Global Precipitation Data Sets: Data
        Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, *56*(1), 79–107. https://doi.org/10.1002/2017RG000574

Thornton, P. E., Thornton, M., Mayer, B., Wilhelmi, N., Wei, Y., Cook, R.: Daymet: Daily surface weather on a 1 km grid for North America, 1980-2008.Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, TN.

895   http://dx.doi.org/10.3334/ORNLDAAC/1219, 2012.

Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Duan, Q., and Lohmann, D.: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research Atmospheres*, *117*(3), 1–23. https://doi.org/10.1029/2011JD016051, 2012.

900