



EPA Public Access

Author manuscript

Environ Model Softw. Author manuscript; available in PMC 2021 January 01.

About author manuscripts

Submit a manuscript

Published in final edited form as:

Environ Model Softw. 2020 January 1; 123: 1–104570. doi:10.1016/j.envsoft.2019.104570.

Demonstration of an online web services tool incorporating automatic retrieval and comparison of precipitation data

Jan Sitterson^a, Sumathy Sinnathamby^b, Rajbir Parmar^d, Joshua Koblich^a, Kurt Wolfe^d, Christopher D. Knightes^{c,*}

^aOak Ridge Associated Universities, United States

^bOak Ridge Institute for Science and Education, United States

^cUS EPA Office of Research and Development National Health and Environmental Effects Research Laboratory, Narragansett, RI, 02882, United States

^dUS EPA Office of Research and Development National Exposure Research Laboratory, Athens, GA, 30605, United States

Abstract

Input data acquisition and preprocessing is time-consuming and difficult to handle and can have major implications on environmental modeling results. US EPA's Hydrological Micro Services Precipitation Comparison and Analysis Tool (HMS-PCAT) provides a publicly available tool to accomplish this critical task. We present HMS-PCAT's software design and its use in gathering, preprocessing, and evaluating precipitation data through web services. This tool simplifies catchment and point-based data retrieval by automating temporal and spatial aggregations. In a demonstration of the tool, four gridded precipitation datasets (NLDAS, GLDAS, DAYMET, PRISM) and one set of gauge data (NCEI) were retrieved for 17 regions in the United States and evaluated on 1) how well each dataset captured extreme events and 2) how datasets varied by region. HMS-PCAT facilitates data visualizations, comparisons, and statistics by showing the variability between datasets and allows users to explore the data when selecting precipitation datasets for an environmental modeling application.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Corresponding author: knightes.chris@epa.gov (C.D. Knightes).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Disclaimer

This paper has been reviewed in accordance with US Environmental Protection Agency's peer and administrative review policies and approved for publication. The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the US Environmental Protection Agency.

Software availability

Name: Hydrological Micro Services Precipitation Comparison and Analysis Tool.

Developed by US EPA Office of Research and Development National Exposure Research Laboratory, Athens, GA.

Contact: <https://www.epa.gov/ceam/forms/contact-hms-helpdesk>.

Website: https://qed.epacdx.net/hms/workflow/precip_compare/

A new user can request a user ID and password by clicking the provided link on the login page.

GitHub: <https://github.com/quanted/hms/blob/dev/Web.Services/Models/WSPrecipCompare.cs>.

Keywords

Precipitation; Web services; Comparison; Preprocessing data

1. Introduction

Precipitation data are critical inputs required in environmental modeling, including hydrologic, water quality, climate, atmospheric deposition, erosion, and agricultural models. Precipitation is governed by complicated nonlinear and extremely sensitive atmospheric physical processes (B'ardossy & Plate, 1992) and has significant variability over space and time (Krajewski et al., 2003). Inability to represent spatial rainfall produces uncertainties with streamflow and non-point source pollution modeling (Shen et al., 2012). Studies have reported precipitation input as the main source of uncertainty in calibrating a watershed hydrology model (SWAT) (Cao et al., 2018; Chaplot et al., 2005; Hernandez et al., 2000; Tuo et al., 2016). It is therefore both crucial and challenging to have an accurate representation of precipitation for environmental modeling.

Traditionally, three mechanisms provided precipitation data for environmental modeling: rain gauges, weather radar, and satellite-based sensors (Sikorska and Seibert, 2018); each has its strengths and weaknesses. Rain gauge data are often referred to as the most accurate representation of precipitation at a precise location (Kim et al., 2014; Price et al., 2014). Observational rain gauge data may have missing values due to station maintenance or equipment malfunctions, as well as inaccuracies from sampling errors, calibration uncertainty, or random errors. Due to low spatial uniformity, an assumption is made that rainfall amounts over an area are represented by a single gauge station. Radar data provide spatially distributed precipitation data on a much finer scale (Gao et al., 2017). Bias in radar datasets stems from signal blockage by topographic effects, bright band contaminations, range dependency, and radar calibration errors. Satellites are the only way to retrieve globally homogenous estimations of precipitation (Tapiador et al., 2012). Gauge data tend to underestimate precipitation events while radar and satellites can misinterpret hail as heavy rainfall (Awange et al., 2016; Kidd and Huffman, 2011; Tapiador et al., 2012). Precipitation data types mentioned above vary in degree of temporal and spatial resolution; underlying assumptions and methods (e.g., sampling frame, how data are interpolated); data quality; units; timespan of record; data dissemination methods; data formats; ways missing data are handled; standard versus local time; and temporal aggregations. The inherent tradeoffs of individual datasets are often overlooked (Bishop and Beier, 2013; Daly et al., 2007). Furthermore, deviations of up to 300 mm have been reported in estimated annual precipitation between multiple datasets (Sun et al., 2017). These deviations can impact the output of hydrologic, climate, agricultural, and other types of models used to influence best management practices, regulations, and decision making. Since selection of precipitation data has crucial effects on model performance, the choice of precipitation data needs to come from an intentional and informed decision.

Modelers currently spend significant time retrieving precipitation data directly from source websites and preprocessing the data before model input. Barriers in data retrieval can

include difficulty collecting input data from different sources, incompatibility of data formats, lack of available data products, and limitations of computing resources (C. Zhang et al., 2019). These can cause multiple data requests, time-consuming downloads, and preprocessing procedures for one source, which must be repeated for every dataset used in a modeling project. When multiple precipitation datasets are used in calibration and validation, the performance of hydrological models is improved (Finger et al., 2015). Process-based models requiring large amounts of data are used less often due to data restrictions and data processing limitations (Fatichi et al., 2016). Spatial and temporal resolutions of current datasets limit modeling efforts due to the level of detail in available data and computing resources (Regan et al., 2019). In addition, each dataset has its own format, resolutions, units and timespan, making it difficult to compare them quickly.

To simplify data gathering, web services provide raw water quality data through online script requests and programming packages like the Water Quality Data Portal (<https://www.waterqualitydata.us/portal/>). Unfortunately, there is still a need to process and format the data once downloaded. Most existing models lack built-in data provisioning services and it is difficult for users without expert knowledge to obtain data due to the complexity of some web service interfaces and protocols (Huang et al., 2011). Data provisioning is not usually a seamless part of environmental models and exists as external components or services. Phuong et al. (2019) presented an open source python library to help ease gridded dataset availability and preprocessing. Use of web services, however, make it possible to integrate data provisioning programmatically as part of the modeling process or workflow. Most data acquisition projects involve interaction between data services, model services, and an application that integrates the services (Carlson et al., 2014). Interaction of services may require a strong background in computer science or customized programming language scripts that cannot be widely accessed or reused, hindering the reproducibility of a scientific study (Samourkasidis et al., 2019). A primary constraint on efficient use of models is provisioning data from disparate data sources and services (Carlson et al., 2014).

Discovering, preprocessing, and evaluating precipitation data often constitute a necessary and significant part of environmental modeling. Obtaining multiple precipitation datasets is costly in time and resources, and the problem is amplified when decisions need to be made on the selection of a data source. The goal of this study is to introduce and demonstrate the United States Environmental Protection Agency's (US EPA) Hydrologic Micro Services Precipitation Comparison and Analysis Tool (HMS-PCAT) for precipitation data provisioning and analysis. We present (1) the design, available precipitation data sources, data processing operations, and strengths and limitations of the precipitation tool as a data web service, and (2) how the tool can gather and compare different precipitation sources across the conterminous US. This online tool automatically retrieves, processes, compares, and visualizes precipitation time-series data at point or catchment locations using methods that eliminate the need for interaction between user and computer code. Many studies have compared gridded precipitation datasets to gauge data for use in modeling projects (Behnke et al., 2016; Gao et al., 2017; Muche et al., 2019; Sun et al., 2017). Oftentimes, one chooses a gridded dataset that best matches local station data. HMS-PCAT provides visual and statistical comparisons of precipitation sources to inform modelers about datasets to use in environmental modeling projects. In a demonstration of the tool, we showcase its ease of use

by downloading data and performing additional comparisons. We used the tool to investigate the differences and variability between datasets and regions in the US. Threshold values indicating light, wet, heavy, and very heavy precipitation intensities, as well as the differences in maximum values, were explored to analyze the ability to capture extreme events across multiple datasets. This data retrieval and evaluation tool will aid in gathering and preprocessing precipitation input data for environmental modeling projects.

2. Hydrologic Micro Services description

Historically, legacy models have worked independently to solve specific questions. Most do not have efficient automated input data provisioning services, resulting in modelers having to spend more time on input data gathering and preprocessing rather than analyzing outputs. Environmental modeling is moving forward to meet the needs of multimedia platforms and interoperability between new data and models. To understand why results differ between models, transparency of the methodology and reproducibility of data are needed. We developed a hydro-informatics platform called Hydrologic Micro Services (HMS) to address the importance of interoperability, transparency, reproducibility, and efficiency in environmental modeling (Parmar et al., 2018).

HMS was created to break down barriers of old models and datasets that are constrained by legacy formats and connect them to newer formats, advanced models, and workflows. The motivation was to address problems of the hydrology and water quality modeling community. We developed HMS for users in private, public, and academic sectors at the local, state, and federal levels. The HMS platform is a collection of hydrology and water quality data provisioning web services and modeling components. Data provisioning web services purvey raw data through online script requests for hydrologic parameters including precipitation, air temperature, solar radiation, soil moisture, evapotranspiration, surface and subsurface flow, and runoff. A component is a distinct software module that users can incorporate and link into existing models. For example, components that have been incorporated and compiled into HMS include Normalized Difference Vegetation Index (NDVI), sediment diagenesis, eutrophication, and kinetic transformation of nutrients and chemicals. HMS enables users to rapidly characterize the hydrology of a watershed, reducing their time on data gathering and preprocessing, and easily parameterize model workflows. A web service of the HMS platform is the Precipitation Comparison and Analysis Tool (HMS-PCAT), which is the focus of this paper. HMS-PCAT provides an intuitive online interface for precipitation data source exploration and data download, irrespective of computing platforms or coding languages (Parmar et al., 2018).

HMS-PCAT facilitates precipitation data access, retrieval, preprocessing, and comparison statistics computations for environmental modeling so more time can be spent analyzing model results. The tool is a workflow composed of web services provided by HMS that automates accessing and downloading precipitation time-series data from their original sources and compiles comprehensive statistics and metadata. We developed HMS-PCAT to harmonize precipitation data from multiple sources with different data retrieval protocols, formats, spatial and temporal resolutions, and time references. The following sections introduce the software design, precipitation data sources that were implemented, how the

data is processed automatically, and challenges we faced while creating the tool. HMS-PCAT provides precipitation information to fill knowledge gaps for effective water resource management.

2.1. Software design

We designed and developed HMS-PCAT as an assembly of data provisioning web services that provide a consistent way to access precipitation data across disparate sources with different formats, resolutions, and access protocols. HMS-PCAT is accessible by browser and integrates internal and external web services to provide data processing, geospatial, and statistical computations that create an accessible database (Fig. 1). The underlying web services have been implemented as Representational State Transfer Application Programming Interface (RESTful API) over Hyper Text Transfer Protocol (HTTP). This approach allows components to communicate over standard internet protocols by removing obstacles associated with heterogeneous programming languages and operating systems. The tool's browser-based user interface was built with standard programming languages (HTML/JavaScript/CSS). The user interface implements several cases for pulling and comparing precipitation data. Each case requires three types of inputs: location, temporal extent and aggregation level, and gridded-data source for comparison. Upon receiving this input from the browser-based website, each aspect of the input object is validated for accuracy by the HMS server and a task is created. The HMS server distributes the information to appropriate internal services to pull data from external cloud services (Fig. 1). The internal services perform a series of operations on the data including: unit and time zone conversions, aggregation of data into the requested time interval, management of missing data, calculations of relevant statistics, and compiling of metadata. A database temporarily stores the data until retrieval is finished. Final numerical analysis and comparisons are performed before the time series dataset is returned to the web browser. The interoperable and readable time series output file can be downloaded as JavaScript object notation (.JSON) or a comma separated value (.CSV) file or viewed through the user interface as an interactive line graph. Downloading time series data allows users to transfer data across multiple models and infrastructures, as well as edit and perform additional statistical analyses or interpretations.

2.2. Precipitation Data Sources

Data from federal agencies such as NASA, NOAA, and USDA are frequently cited as sources of precipitation data used in environmental modeling (Behnke et al., 2016; Gao et al., 2017; Golden et al., 2010; Muche et al., 2019). Five precipitation datasets; four interpolated and gridded, and one set of collected data, of the conterminous US are implemented in HMS-PCAT (Table 1). Additional datasets will be added in later development of the tool. Spatial coverage and grid resolutions of the precipitation data sources are shown in Fig. 2. A brief description of each precipitation source implemented in HMS-PCAT is presented below.

2.2.1. NLDAS and GLDAS—The North American Land Data Assimilation System (NLDAS) combines North American radar data, Climate Prediction Center gauge data, and satellite data from the Climate Prediction Center Morphing Technique (CMORPH). NLDAS has an hourly time step with data across North America available from 1981 to the present,

with a maximum time lag of four days (Rodell, 2019). The Global Land Data Assimilation System (GLDAS) combines satellite data and ground-based observational data covering Earth between 90° north and 60° south (Rodell, 2019). GLDAS data are given every 3 h and take at least a month to process. GLDAS data exist in two different time ranges: Version 1.0 contains data from 1948 to 2010, and Version 2.0 containing data from 2010 to the present. If a data request covers both ranges, HMS-PCAT will gather both versions and combine the data. Both NLDAS and GLDAS data are provided in the Greenwich Mean Time (GMT) zone, which is then converted to local time in the HMS-PCAT algorithm. NLDAS is provided in $\text{kg m}^{-2} \text{h}^{-1}$, which HMS-PCAT converts to mm h^{-1} , given the conversion that 1 kg m^{-2} is equivalent to 1 mm of water thickness. Units for GLDAS are provided in $\text{kg m}^{-2} \text{s}^{-1}$, which is aggregated into three hourly data. Both NLDAS and GLDAS data are summed to provide daily output in mm d^{-1} . Both sources are publicly available on the NASA web service via “hydrology data rods”, which are large volumes of organized time series data in an American Standard Code for Information and Interchange (ASCII) text format (<https://disc.gsfc.nasa.gov/information/tools?title=Hydrology%20Data%20Rods>). Data can be obtained from both NLDAS and GLDAS by querying data rod access endpoints with a designated request, using location, time span, and the desired dataset variable as parameters.

2.2.2. PRISM—The Parameter-elevation Regressions on Independent Slopes Model (PRISM) provides climatology information by combining ground gauge stations from multiple sources and radar products based on digital elevation models. Data cover the contiguous US from 1981 to 2017. The database is updated yearly by adding data for the complete year. The data are provided in JSON format with default units of mm d^{-1} in the local time zone (Daly et al., 2008). PRISM data are dispensed as a layer containing all data within user-specified spatial extents, making data download and extraction slow. HMS-PCAT calls a web service hosted by Colorado State University for downloading PRISM precipitation data as a data rod at a specific location for a faster process (<http://csip.engr.colostate.edu:8083/csip-climate/m/prism/1.0>).

2.2.3. DAYMET—The Daily Surface Weather and Climatological Summaries (DAYMET) is a dataset of rain gauge data interpolated and extrapolated by the DAYMET algorithm (Thornton et al., 2017). The interpolation provides data over Canada, Mexico, the United States of America, and Puerto Rico. Daily rainfall is rounded to the nearest whole number provided in mm d^{-1} and is available from 1980 to the latest full year. The extra day in leap years are omitted. Data are obtained as a CSV file from a NASA-hosted web service by querying a web service with the desired dataset, location, and timespan (<https://daymet.ornl.gov/single-pixel/api/data>).

2.2.4. NCEI—This dataset from the National Center for Environmental Information (NCEI) provides precipitation data collected and recorded at land-based rain gauge stations from the Global Historical Climate Network-Daily (GHCN-D) and the Cooperative Observer Network (COOP). NCEI has access to about 53,000 stations worldwide, some with data going back as far as 1901 (NOAA, 2017). Start and end years for NCEI stations depend upon the specific station. The temporal resolution is hourly, with units in mm and data are provided by the station latitude and longitude point. As part of quality control, a flag is

placed where there is a missing measurement or data quality inconsistency. NCEI data are obtained in JSON format, using the Climate Data Online web service by issuing a request along with an access token parameter unique to the user (<https://www.ncdc.noaa.gov/cdo-web/webservices/v2#gettingStarted>). Tokens are required by NCEI to access any datasets, which can be specified in the request, along with the station ID and time span.

2.3. Data processing

Data processing steps in HMS-PCAT are important for ensuring an accurate comparison of datasets. Location, temporal resolution, and time series format for each dataset must be consistent to relate precipitation data. The processing workflow within HMS-PCAT involves minimal user input with automated discovery, retrieval, and evaluation of precipitation data on the software side. The user inputs a location and date range, then chooses a temporal aggregation method and desired gridded data sources (Fig. 3). Additional details on input combinations are discussed later. The tool lets users save time on data retrieval and data processing by automating the procedure as follows:

1. Validates user input and sends requests to source websites
2. Pulls data for location and specified time period
3. Aggregates data into specified temporal aggregation algorithms (daily, monthly, annual, extreme event) in local time zone and flags missing data
4. Merges all individual datasets into one data file
5. Computes statistics and metadata
6. Provides data visualizations on web page
7. Formats data for export or incorporation into other modeling components

Location input can be retrieved by (1) National Hydrography Dataset (NHDPlus V2) catchment identifier (COMID) or (2) NCEI gauge station identifier (Station ID). If Station ID or COMIDs are unknown, a hyperlink is provided to a nationwide map where this information can be obtained (Fig. 3). Different combinations of inputs determine how HMS-PCAT processes data. The combinations for location inputs are: (i) COMID is provided and the nearest (to catchment centroid) NCEI station is used, along with gridded data at catchment centroid location; (ii) COMID is provided, with nearest NCEI station used with spatially aggregated gridded data for the catchment; (iii) COMID and specific NCEI station are provided with gridded data at the catchment centroid; or (iv) COMID and specific NCEI station are provided with spatially aggregated gridded data for the catchment. For spatial aggregation of gridded data, an external service call is made to the EPA Waters cloud service (<https://www.epa.gov/waterdata/waters-web-services>) to obtain a polygon shapefile for the catchment corresponding to the provided COMID. The polygon shapefile is overlaid on the respective grid of the data source and a calculation to provide area averaged data is made. For options (i) and (ii), the search range is expanded until a station with data are found if the closest NCEI station does not have precipitation data for the specified time-period. The system gives an error if no suitable station is found within one latitude/longitude degree.

Temporal resolution varies from one data source to another, as shown in Table 1; thus, a temporal aggregation algorithm is necessary to compare data sources one-to-one. The temporal aggregations automatically format the data to cover complete years within the specified start and end year, and include daily, monthly, annual, or extreme precipitation event (Fig. 3). If the extreme precipitation event is chosen, two user specified threshold values are required: one for rainfall accumulation for the previous five days and one for the sixth day amount. In all temporal aggregations, data covers January 1st of the start year to December 31st of the end year.

To format the data, HMS-PCAT software builds a request to external web services using unique access tokens to identify and pull data using the location and start/end dates. The request is processed, and a time series is produced by shifting time data to the local time zone if required. Time series from other data sources are unified with the NCEI time series. Missing or invalid data are flagged and given a -9999 value and formatted into a data structure. Statistical calculations are automated using the Math.NET library. Flagged missing data points are excluded from all datasets in the comparison statistics. These calculations and missing or invalid data flags are documented in the tool metadata for transparency. To create a time series output, a column was generated for each dataset and a row for each temporal aggregation specified. Metadata is at the end of the time series which includes the summary statistics performed on the time series. The time series is exportable as JSON or CSV.

2.4. Challenges

HMS-PCAT addresses many difficulties modelers face in recovering and handling precipitation data. We were confronted with a series of challenges in the creation of this tool. The US EPA firewall prohibits certain external website domains from being accessed (Fig. 1). This required a firewall exception to be made by our IT services, so that data source websites could be accessed. HMS-PCAT has a limited number of precipitation data sources (Table 1). Each data source has its own limitations and challenges associated with its integration into HMS-PCAT. Some are not publicly available, are not formatted as web services, or do not have the necessary metadata information. Another challenge we encountered was source websites that limit the number of data requests that can be completed within a given time range. Delay functions and unique access tokens were used to meet website conditions. Examples are the large requests for GHCN-D data covering multiple years which are broken up into smaller delayed requests and combined programmatically in HMS-PCAT.

The data processing workflow in HMS-PCAT is affected by inconsistency in how precipitation sources handle leap years and missing data. For leap years, data obtained from NLDAS, PRISM, and NCEI provide all 366 days, while GLDAS requests contain data for February 29th and eliminate December 31st to maintain 365 days. DAYMET simply ignores the leap day and treats the leap year as having only 365 days. To maintain 366 days per leap year for each data source, HMS must adjust the leap year data to include February 29th and append a zero-value day to December 31st. By addressing this, data aggregation becomes more streamlined and datasets can be compared more accurately. Dealing with missing

values and how they are represented in each data source is challenging because users want a seamless time series. Adding a -9999 value to every missing or invalid data point in each dataset addresses discrepancies among precipitation sources.

3. PCAT demonstration

HMS-PCAT can be used to download data for a time series analysis, data visualization, and comparison for initial observations, and as decision support when choosing a precipitation dataset. The tool accesses precipitation data from external websites, performs temporal aggregations; computes statistics; produces tables and graphs for data visualization; and prepares data files for user download or integration with other existing services. One example of using HMS-PCAT for a dataset retrieval and comparison tool across the conterminous US is described in the following sections.

3.1. Study area

We investigated precipitation data over the conterminous US and divided the study area by climatic regions according to Bukovsky (2011) (Fig. 4). These regions are simplified from Ricketts et al. (1999) ecoregions, and capture important features in regional climate and topography (Bukovsky, 2011). They have been used previously to evaluate precipitation and temperature data sources over the conterminous US (Behnke et al., 2016; Kampe et al., 2010). Seventeen NCEI gauge stations closest to the Bukovsky region's centroid are used in our demonstration. Selected stations are Global Historical Climate Network-Daily (GHCN-D) gauges, with more than 100 years of precipitation data. Except for one station (USC00057656 Silverton, CO in the South Rockies which had 11.27% missing days), all had a minimum of 90% daily records of precipitation data for the reference period (Table 2). The reference period chosen is from January 1981 to December 2017 (37 years) given constraints of the gridded datasets.

3.2. Results

Results of using HMS-PCAT to retrieve, compare, and evaluate data from multiple sources are presented, as well as examples of downloading the processed data. HMS-PCAT gathered all five precipitation sources for each location in Table 2. Data from each region was downloaded as a CSV and combined in a database for additional analysis to demonstrate the ease of working with preprocessed data. Comparison statistics calculated by HMS-PCAT were extracted from the metadata and validated with external code and data visualizations for a regional analysis. The large amounts of data retrieved and downloaded were also investigated for variations in regions and datasets on their ability to capture a range of precipitation intensity events.

3.2.1. HMS-PCAT data visualization—Data was retrieved using HMS-PCAT by the NCEI Station ID and the start and end date for our reference period, with a daily aggregation, and choosing all sources for comparison. As a result of the HMS-PCAT web service, Fig. 5 is a screenshot of the output provided, including the metadata, an interactive time series graph, a table of summary statistics, and the Pearson's correlation matrix for data visualization and comparison. Metadata includes the location where the data was pulled and

the number of missing data points. The time series graph shows values recorded by each dataset on a given day. Statistics on the datasets include standard deviation, mean, median, and percentile values (75th, 95th, and 99th). The Pearson's correlation coefficient matrix demonstrates the linear comparison of the precipitation datasets and does not put weight on a 'reference' dataset. The heat map of correlation coefficients shows the degree of correlation between any two datasets. From these results, we can evaluate variation among datasets and start to make decisions on appropriateness of each dataset for a project.

3.2.2. Regional variation—Using the HMS-PCAT data download capability, we compared different regions and distinguished the variability of precipitation datasets in diverse environments. Of the 17 regions, 11 showed PRISM and NCEI having the highest Pearson's correlation coefficient (above 0.8) from 1981 to 2017 (Figure A1 in Appendix). Three regions showed NCEI and DAYMET with the highest correlation. In some regions all datasets showed a similar relationship with a correlation of over 0.5 (Fig. 6A the Pacific NW region); other regions indicated that some datasets were vastly different from others in their ability to record values (e.g., Fig. 6B shows the correlation between NCEI and NLDAS is 0.19 in the Mid-Atlantic region). The Mid-Atlantic and Appalachian regions had a similar relationship with the lowest dataset correlations. Other regions showed higher correlations between two gridded datasets, DAYMET and GLDAS. An example is the South Rockies region shown in Fig. 6C. The Southern Rockies NCEI gauge station had the highest elevation at 2830.1 m and the largest number of missing days (1524 out of 13,514).

3.2.3. Extreme events—To check how well the precipitation datasets capture extreme weather conditions, five climate indices (CLIMDEX) described by X. Zhang et al. (2011) were used: the single day maximum value recorded, and the number of days precipitation (P) is considered light ($P < 1$ mm), wet ($P \geq 1$ mm), heavy ($P \geq 10$ mm), and very heavy ($P \geq 20$ mm). These climate indices are commonly used to analyze climate variability (Alexander et al., 2006; Behnke et al., 2016; Donat et al., 2013; Mucic et al., 2019; Sillmann et al., 2013). The number of days precipitation classified as light, wet, heavy, very heavy, and the maximum value recorded is shown in Fig. 7. By showing each dataset, one can see variability in days recorded across the US. DAYMET recorded no days < 1 mm, while other datasets show more data in this range. PRISM, NLDAS, and GLDAS had a higher percentage of data recorded in this range than NCEI in all regions. NLDAS has slightly more days recorded as wet, while DAYMET recorded more heavy days compared to the other datasets. The Deep South and the Southeast region recorded the most days with very heavy precipitation events.

The maximum value recorded is also an indicator of extreme events. Fig. 7E shows the maximum daily value recorded by each data source for each region. The same shade of red across each row indicates the ability of each dataset to capture the maximum value. The Great Lakes region had the greatest variation in maximum recorded event. NCEI recorded 178 mm as the maximum, DAYMET 126 mm, NLDAS 62.4 mm, GLDAS 102 mm, and PRISM 172 mm; this is a difference of 115.6 mm between greatest and lowest maximum recorded for the region. The Great Basin and Pacific NW regions showed the least difference between the dataset's maximum recorded value.

4. Discussion

HMS-PCAT provides easy access to multiple precipitation datasets, both gauge station and gridded in one location. This tool is both powerful and efficient for users acquiring data. HMS-PCAT software finds the location of gauge stations or catchment identifiers and pinpoints its grid for each gridded dataset. The tool pulls the precipitation data from a web service, then aggregates the datasets into one file. Quick data comparison on the webpage can determine which datasets closely match, using a summary statistics table, Pearson's Correlation Matrix, and an interactive times series plot. Data visualizations and dataset statistics are provided on the output page with an option to download the data for further analysis. One can use this tool to decide the most appropriate dataset to use, to quickly compare datasets, and to gather data for environmental modeling. HMS speeds up data gathering by automating tedious discovery, retrieval, and processing procedures. One major advantage of HMS-PCAT is that users do not have to maintain and update each resource. Using this tool can significantly benefit scientists dealing with data acquisition who are not skilled programmers.

We evaluated HMS-PCAT by showing the amount of data it can handle, as well as an analysis on the data it produces. We took a closer look into variations of 17 different regions within the US and evaluations of precipitation intensity events recorded by each dataset.

4.1. Regional variation

As one would expect, there are regional differences in rainfall amounts. Orographic and topographic effects complicate climatic processes in mountainous regions. High elevations can be quite different from low elevations and the differences cannot be captured in coarse spatial resolutions (Guan et al., 2005). The Mid-Atlantic and Appalachian regions had the lowest correlation coefficients, which could be caused by the Appalachian Mountains affecting weather patterns. Gridded datasets with fine spatial resolution (PRISM and DAYMET) in these regions showed a higher correlation than coarser datasets (NLDAS and GLDAS). Spatial resolution may also play a role in the ability to correlate with other datasets. PRISM and DAYMET were the highest resolution datasets and showed high correlation with NCEI in the Pearson's matrix (refer to appendix Figure A1); similar results were shown in Muche et al. (2019). NLDAS showed low correlations across each region, even though it does not have the coarsest resolution. Results of NLDAS bias were shared by Behnke et al. (2016). The Pacific NW was the only region with all datasets above a correlation coefficient of 0.5 for the reference period.

Interpolation methods of gridded datasets involving gauge data can skew data toward gauge records. DAYMET and PRISM were originated using station data (Daly et al., 2002; Thornton et al., 1997, 2017). Similar observations have been made between station data and PRISM data (Golden et al., 2010; Muche et al., 2019). Having gridded data align with station records can benefit users wanting to fill missing data points. Datasets that are not independent may introduce unwanted bias in the comparison, which is why we also showed general trends in evaluating threshold values and maximum values.

4.2. Extreme events

Variability between datasets in recording rainfall intensity can have major impacts on model results for streamflow and erosion calculations. The maximum daily rainfall amount is useful in engineering applications and can show long-term changes in extreme events when analyzed on an annual basis (X. Zhang et al., 2011). The Great Lakes region had the greatest inconsistency between datasets in maximum recorded amount. How well extreme events are captured differs by dataset estimation techniques and methodology. Days with $P > 0$ and $P < 1$ mm, differ dramatically by dataset. DAYMET is unable to account for these low rainfall events because the internal algorithm either assumes the day was dry or was rounded to 1 mm. NCEI had a low percentage of data in this range which is likely due to estimation of rainfall amounts by observer bias as described by Daly et al. (2007). The forcing generation technique for radar-rainfall estimations used in NLDAS and GLDAS can contribute to the high number of days recorded with $P < 1$ mm (Luo, 2003).

Miscalculations by rain gauges during heavy precipitation events may be due to water loss from wind and erratic behavior of mechanical aspects of the gauge (Lanza and Stagi, 2008; Molini et al., 2005). Rain gauges frequently underestimate rainfall during large storms (Price et al., 2014; Radcliffe and Mukundan, 2017). Gauge stations often include observer bias in recorded values through favoring or avoiding some precipitation quantities (Daly et al., 2007). The South Rockies had the most missing days in the NCEI dataset, which could account for low correlations. Regions with low rainfall have gridded datasets that closely match the observed dataset, while regions with high rainfall do not agree with the observed dataset which can be due to datasets misinterpreting heavy rainfall events.

The Pacific NW is a wet region and has more 'heavy' days than the Southwest region which is more arid. The Pacific NW had more days considered light, wet, and heavy than the Southeast and Deep South regions (Fig. 7), but the overall cumulative sum for those regions is higher than the Pacific NW (Figure A2 in Appendix). This results from the amount and magnitude of very heavy days. The single day maximum for the Pacific NW is expressively lower than the southern regions' magnitude (Fig. 7). Heavy daily rainfall is a common occurrence in the eastern half of the US (Behnke et al., 2016).

5. Conclusion

HMS-PCAT offers a standardized interface for simple access to multiple precipitation datasets to save time and resources and to increase the efficiency and quality of modeling projects. The tool is publicly available at https://qed.epacdx.net/hms/workflow/precip_compare/ and provide an email address to request a user ID. We will continue to update the HMS platform to provide additional services and components as research and development continues. HMS-PCAT can be used as a data evaluation tool for data source selection by showing comparison statistics between datasets on a web interface. Being able to quickly view multiple precipitation datasets and their summary statistics can enable users to make well informed decisions on selecting their data source for research or modeling projects. Further applications of the tool and its use in environmental modeling will be investigated. HMS-PCAT overcomes challenges faced when retrieving and processing precipitation data. It aids in improving environmental modeling efforts related to water

resource research by providing public access to multiple precipitation datasets in an interoperable format and automates data processing and calculations.

Acknowledgements

Funding: This work was funded through the US EPA's Office of Research and Development's Safe and Sustainable Water Resources Research Program. This research did not receive any specific grant from funding agencies in the commercial or not-for-profit sectors.

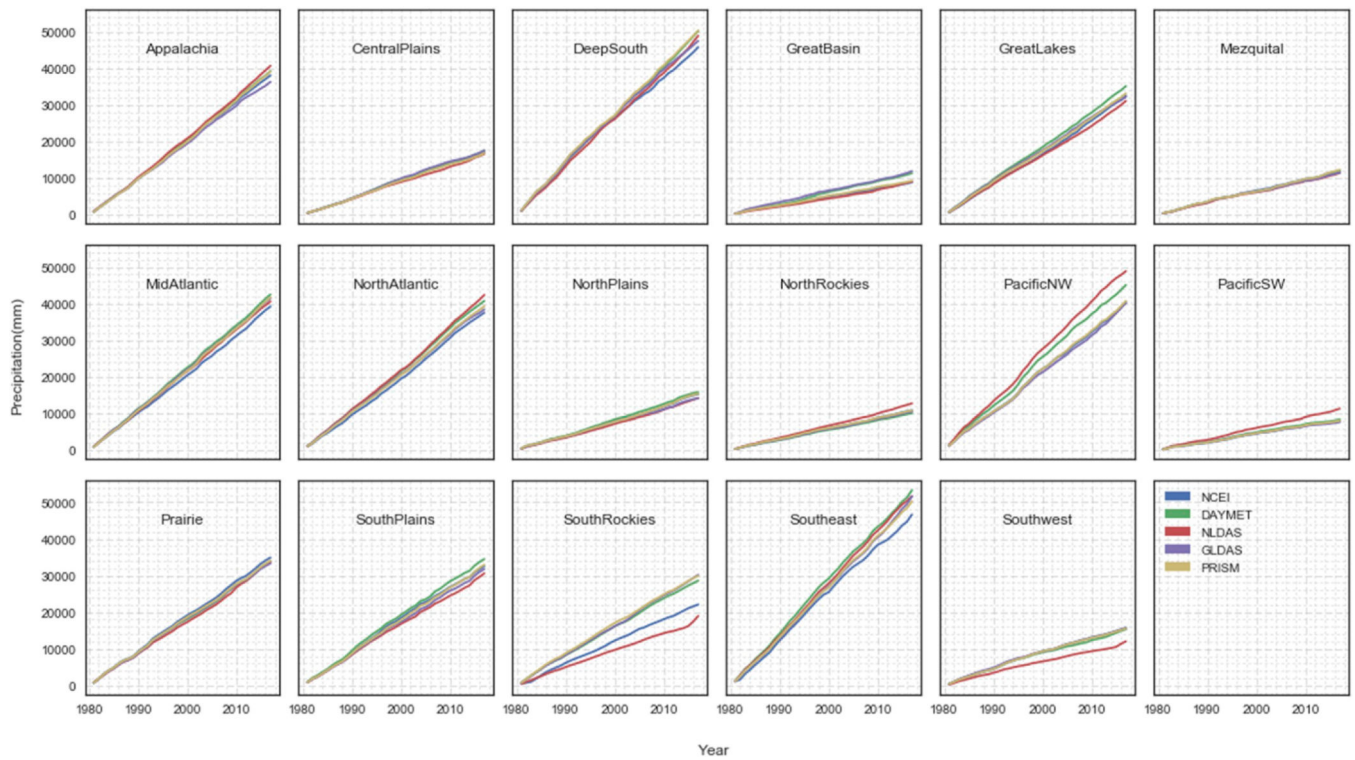
Appendix

Supporting Material for Regional Variation

HMS-PCAT gathered five precipitation sources (NCEI, NLDAS, GLDAS, DAYMET, PRISM) in 17 locations for the 1981 to 2017 reference period. Data were evaluated through external code. Data visualizations on variations between precipitation datasets for regions of the United States were performed. In support of results in section 3.2.2, the variation analyses for all regions are presented.



Fig. A1. A Pearson’s correlation matrix for each of the 17 regions studied. Pearson’s correlation coefficient matrix demonstrates the linear comparison of the precipitation datasets. Darker red indicates a high correlation between two datasets while darker blue shows low correlations.

**Fig. A2.**

Cumulative sum graphs of precipitation in each region, starting at year 1981 and going to the end of 2017. NCEI is shown in blue, DAYMET is green, NLDAS is red, GLDAS is purple, and PRISM is yellow.

References

- B'ardossy A, Plate E, 1992 Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resour. Res* 28, 1247–1259.
- Alexander LV, Zhang X, Peterson TC, Caesar J, Gleason B, 2006 Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.: Atmosphere* 111 10.1029/2005JD006290.
- Awange JL, Ferreira VG, Forootan E, Khandu, Andam-Akorful SA, Agutu NO, He XF, 2016 Uncertainties in remotely sensed precipitation data over Africa. *Int. J. Climatol* 36 (1), 303–323. 10.1002/joc.4346.
- Behnke R, Vavrus S, Allstadt A, Albright T, Thogmartin WE, Radeloff VC, 2016 Evaluation of downscaled, gridded climate data for the conterminous United States. *Ecol. Appl* 26 (5), 1338–1351. 10.1002/15-1061. [PubMed: 27755764]
- Bishop DA, Beier CM, 2013 Assessing uncertainty in high-resolution spatial climate data across the US Northeast. *PLoS One* 8 (8), e70260 10.1371/journal.pone.0070260. [PubMed: 23936401]
- Bukovsky MS, 2011 Masks for the Bukovsky regionalization of north America. Retrieved from: <http://www.narccap.ucar.edu/contrib/bukovsky/>.
- Cao Y, Zhang J, Yang MX, Lei XH, Guo BB, Yang L, Qu JS, 2018 Application of SWAT model with CMADS data to estimate hydrological elements and parameter uncertainty based on SUFI-2 algorithm in the Lijiang river basin, China. *Water* 10 (6), 742 10.3390/w10060742 ARTN.
- Carlson J, David O, Lloyd W, Leavesley G, Rojas K, 2014 Data Provisioning for the Object Modeling System (OMS). Institute of Technology Publications.

- Chaplot V, Saleh A, Jaynes D, 2005 Effect of the accuracy of spatial rainfall information on the modeling of water, sediment, and NO₃-N loads at the watershed level. *J. Hydrol* 312, 223–234.
- Daly C, Gibson WP, Taylor GH, Johnson GL, Pasteris P, 2002 A knowledge-based approach to the statistical mapping of climate. *Clim. Res* 22 (2), 99–113. 10.3354/cr022099.
- Daly C, Gibson WP, Taylor GH, Doggett MK, Smith JI, 2007 Observer bias in daily precipitation measurements at United States cooperative network stations. *Bull. Am. Meteorol. Soc* 88 (6), 899–910. 10.1175/BAMS-88-6-899.
- Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Pasteris PP, 2008 Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol* 28 (15), 2031–2064. 10.1002/joc.1688.
- Donat MG, Alexander LV, Yang H, Durre I, Vose R, Caesar J, 2013 Global land-based data sets for monitoring climatic extremes. *Bull. Am. Meteorol. Soc* 94.
- Faticchi S, Vivoni ER, Ogden FL, Ivanov VY, Mirus B, Gochis D, Tarboton D, 2016 An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *J. Hydrol* 537, 45–60. 10.1016/j.jhydrol.2016.03.026.
- Finger D, Vis M, Hiss M, Seibert J, 2015 The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments. *Water Resour. Res* 10.1002/2014WR015712.
- Gao JG, Sheshukov AY, Yen H, White MJ, 2017 Impacts of alternative climate information on hydrologic processes with SWAT: a comparison of NCDC, PRISM and NEXRAD datasets. *Catena* 156, 353–364. 10.1016/j.catena.2017.04.010.
- Golden HE, Knightes CD, Cooter EJ, Dennis RL, Gilliam RC, Foley KM, 2010 Linking air quality and watershed models for environmental assessments: analysis of the effects of model-specific precipitation estimates on calculated water flux. *Environ. Model. Softw* 25 (12), 1722–1737. 10.1016/j.envsoft.2010.04.015.
- Guan H, Wilson JL, Makhnin O, 2005 Geostatistical mapping of mountain precipitation incorporating autosearched effects of terrain and climatic characteristics. *J. Hydrometeorol* 6 (6), 1018–1031. 10.1175/JHM448.1.
- Hernandez M, Miller SN, Goodrich DC, Goff BF, Kepner WG, Edmonds CM, Jones KB, 2000 Modeling runoff response to land cover and rainfall spatial variability in semi-arid watersheds. *Environ. Monit. Assess* 64 (1), 285–298. 10.1023/A:1006445811859.
- Huang M, Maidment D, Tian Y, 2011 Using SOA and RIA's for water data discovery and retrieval. *Environ. Model. Softw* 36 (11), 1309–1324. 10.1016/j.envsoft.2011.05.008.
- Kampe TU, Johnson BR, Kuester M, Keller M, 2010 NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. *J. Appl. Remote Sens* 4 10.1117/1.3361375.
- Kidd C, Huffman G, 2011 Review global precipitation measurement. *Meteorol. Appl* 18 (3), 334–353. 10.1002/met.284.
- Kim KP, Katie, Gene Whelan, Kurt Wolfe, Paul Duda, Mark Gray, Yakov Pachepsky, 2014 Using Remote Sensing and Radar MET Data to Support Watershed Assessment Comprising IEM.
- Krajewski W, Ciach G, Habib E, 2003 An analysis of small-scale rainfall variability in different climatic regimes. *Hydrol. Sci. J* 48 (2).
- Lanza LG, Stagi L, 2008 Certified accuracy of rainfall data as a standard requirement in scientific investigations. *ADGEO* 16, 43–48.
- Luo L, 2003 Validation of the north American land data assimilation system (NLDAS) retrospective forcing over the southern great plains. *J. Geophys. Res* 108 (D22) 10.1029/2002jd003246.
- Molini A, Lanza LG, La Barbera P, 2005 The impact of tipping-bucket raingauge measurement errors on design rainfall for urban-scale applications. *Hydrol. Process* 19 (5), 1073–1088. 10.1002/hyp.5646.
- Muche ME, Sinnathamby S, Parmar R, Knightes CD, Johnston JM, Wolfe K, Smith D, 2019 Comparison and Evaluation of Gridded Precipitation Sources with Monitored Precipitation in an Agricultural Watershed Using SWAT. Manuscript submitted for publication (under review - Journal of the American Water Resources Association).

- NOAA, 2017 National Oceanic and Atmospheric Administration. National Centers for Environmental Information Data Access. Retrieved 04/20, 2017, from. <https://www.ncdc.noaa.gov/data-access>.
- Parmar R, Knightes CD, Smith D, Wolfe K, Koblich J, Sitterson J, Purucker T, 2018 Hydrological Micro services. In: Proceedings of the 9th International Congress on Environmental Modelling and Software, Fort Collins, CO Paper presented at the.
- Puong J, Bandaragoda C, Istanbuluoglu E, Beveridge C, Strauch R, Setiawan L, Mooney SD, 2019 Automated retrieval, preprocessing, and visualization of gridded hydrometeorology data products for spatial-temporal exploratory analysis and intercomparison. *Environ. Model. Softw* 116, 119–130. 10.1016/j.envsoft.2019.01.007.
- Price K, Purucker ST, Kraemer SR, Babendreier JE, Knightes CD, 2014 Comparison of radar and gauge precipitation data in watershed models across varying spatial and temporal scales. *Hydrol. Process* 28 (9), 3505–3520. 10.1002/hyp.9890.
- Radcliffe DE, Mukundan R, 2017 PRISMvs. CFSR precipitation data effects on calibration and validation of SWAT models. *JAWRA J. Am. Water Resour. Assoc* 53 (1), 89–100. 10.1111/1752-1688.12484.
- Regan R, Juracek KE, Haya LE, Markstrom SL, Viger RJ, Driscoll JM, Norton PA, 2019 The U. S. Geological Survey National Hydrologic Model infrastructure: rationale, description, and application of a watershed-scale model for the conterminous United States. *Environ. Model. Softw* 10.1016/j.envsoft.2018.09.023.
- Ricketts TH, Dinerstein E, Olson D, Loucks C, Echbaum W, DellaSala D, Walters S, 1999 Terrestrial Ecoregions of North America: A Conservation Assessment. Island Press, Washington, DC.
- Rodell M, 2019 5 21 LDAS land data assimilation system LDAS: project goals. Retrieved May 29, 2019, from. <https://ldas.gsfc.nasa.gov/>.
- Samourkasidis A, Papoutsoglou E, Athanasiadis IN, 2019 A template framework for environmental timeseries data acquisition. *Environ. Model. Softw* 10.1016/j.envsoft.2018.10.009.
- Shen Z, Chen L, Liao Q, Liu R, Hong Q, 2012 Impact of spatial rainfall variability on hydrology and nonpoint source pollution modeling. *J. Hydrol* 472–473, 205–215. 10.1016/j.jhydrol.2012.09.019.
- Sikorska A, Seibert J, 2018 Value of different precipitation data for flood prediction in an alpine catchment: a Bayesian approach. *J. Hydrol* 10.1016/j.jhydrol.2016.06.031.
- Sillmann J, Kharin VV, Zwiers FW, Zhang X, Bronaugh D, 2013 Climate extreme indices in the CMIP5 multi- model ensemble. Part 2: future climate projections. *J. Geophys. Res.: Atmosphere* 118.
- Sitterson J, Knightes C, Parmar R, Wolfe K, Avant B, Smith D, 2018 A survey of precipitation data for environmental modeling. In: Proceedings of the 9th International Congress on Environmental Modelling and Software, Fort Collins, CO Paper presented at the.
- Sun Q, Miao C, Duan Q, Ashouri H, Sorooshian S, Hsu K-L, 2017 A review of global precipitation data sets: data sources, estimation, and intercomparisons. *Rev. Geophys* 1–29. 10.1002/2017rg000574.
- Tapiador FJ, Turk FJ, Petersen W, Hou AY, García-Ortega E, Machado LAT, de Castro M, 2012 Global precipitation measurement: methods, datasets and applications. *Atmos. Res* 104–105, 70–97. 10.1016/j.atmosres.2011.10.021.
- Thornton PE, Running SW, White MA, 1997 Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol* 190, 214–251.
- Thornton PE, Thornton MM, Mayer BW, Wei Y, Devarakonda R, Vose RS, Cook RB, 2017 Daymet: daily surface weather and climatological Summaries. Daymet V3. From, 2017–04-19. <https://daymet.ornl.gov/>.
- Tuo Y, Duan Z, Disse M, Chiogna G, 2016 Evaluation of precipitation input for SWAT modeling in Alpine catchment: a case study in the Adige river basin (Italy). *Sci. Total Environ* 573, 66–82. 10.1016/j.scitotenv.2016.08.034. [PubMed: 27552731]
- Zhang X, Alexander L, Hegerl GC, Jones P, Tank AK, Peterson TC, Zwiers FW, 2011 Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Clim. Change* 2 (6), 851–870. 10.1002/wcc.147.

Zhang C, Di L, Sun Z, Lin L, Yu EG, Gaigalas J, 2019 Exploring cloud-based web processing service: a case study on the implementation of CMAQ as a service. *Environ. Model. Softw* 113, 29–41. [10.1016/j.envsoft.2018.11.019](https://doi.org/10.1016/j.envsoft.2018.11.019).

EPA Author Manuscript

EPA Author Manuscript

EPA Author Manuscript

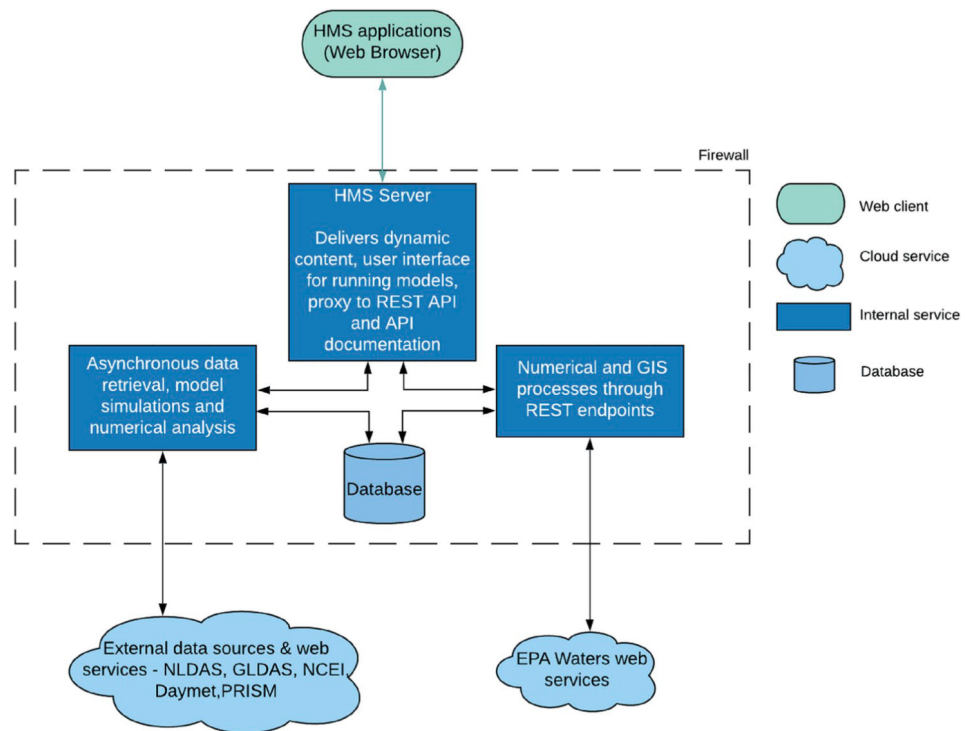


Fig. 1.

Infrastructure diagram of the web services in HMS-PCAT. The web client communicates with internal services to pull data from external cloud services. Pulled precipitation data are compiled and temporarily stored in a database before final analysis is performed on the data and sent back to the web client.

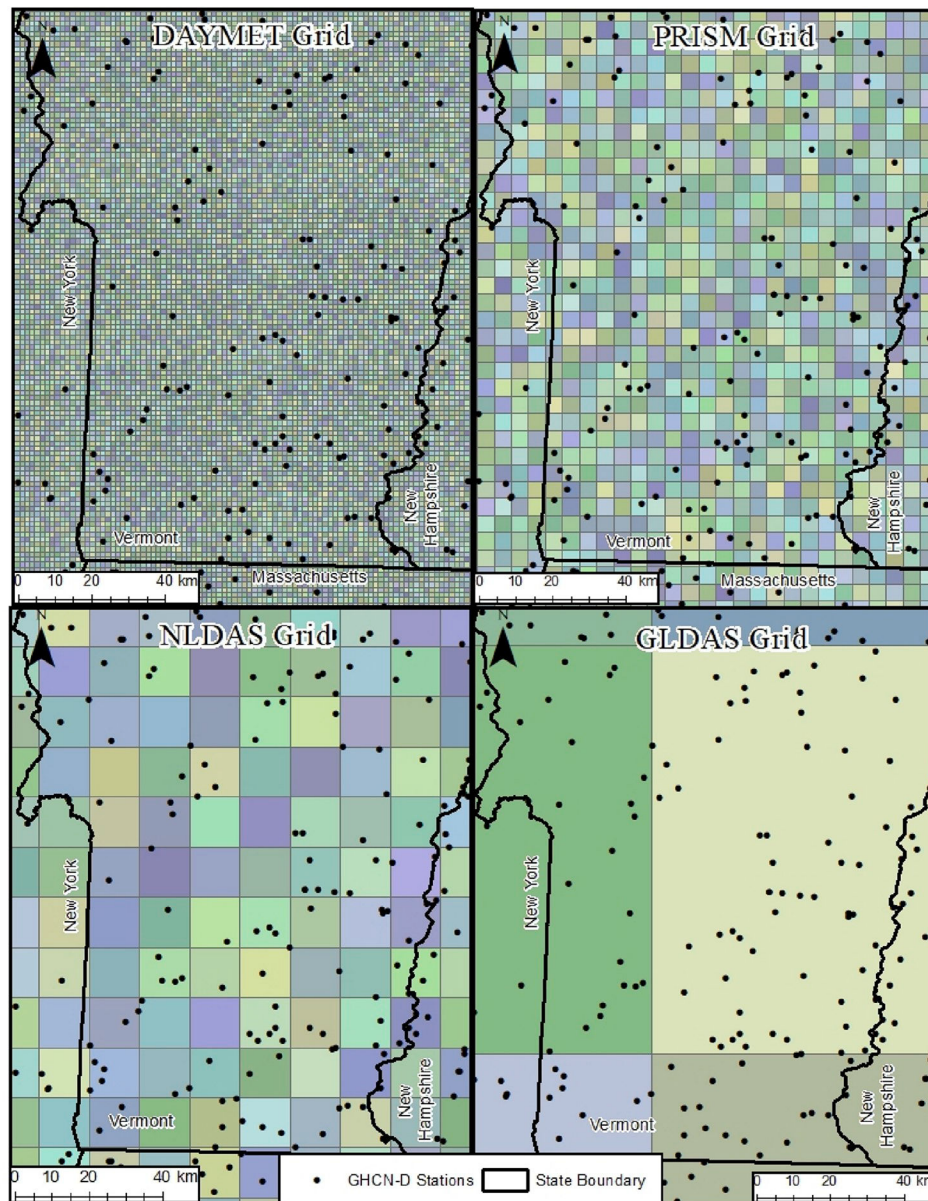


Fig. 2. Spatial resolution of precipitation datasets used in HMS-PCAT showing DAYMET at 1 km, PRISM at 4 km, NLDAS at 0.125-degree, and GLDAS at 0.25-degree grids in the state of Vermont, USA. NCEI GHCN-D gauge station locations are shown as points. Image based on Golden et al. (2010).

HMS: Precipitation Comparison

Overview
Data Request
Output Data
Data Algorithms

Select Precipitation Compare Version Version 2.0 ▾

Location

NHDPlus COMID [Use Map to locate NHDPlus catchment](#)

Use weighted spatial average NCEI Station ID [Use Map to locate NCEI station](#)

Note: Nearest (to the catchment centroid) NCEI Station will be used if NCEI station ID is not provided above.

NCEI Station ID [Use Map to locate NCEI station](#)

Temporal

Start Year End Year

Temporal Aggregation (select one of the following)

Daily

Monthly

Annual

Extreme Precipitation Event

5 Day total precip threshold (mm) >= Daily precip threshold (mm) >=

Data Sources

Select gridded data sources to be compared with observed data from NCEI weather station

NLDAS GLDAS DAYMET PRISM

Submit

Fig. 3. Screenshot of the HMS-PCAT user interface for data request and processing. Data will be pulled and compared to the selected NCEI station ID for the start and end year for each data source selected on a daily time step.

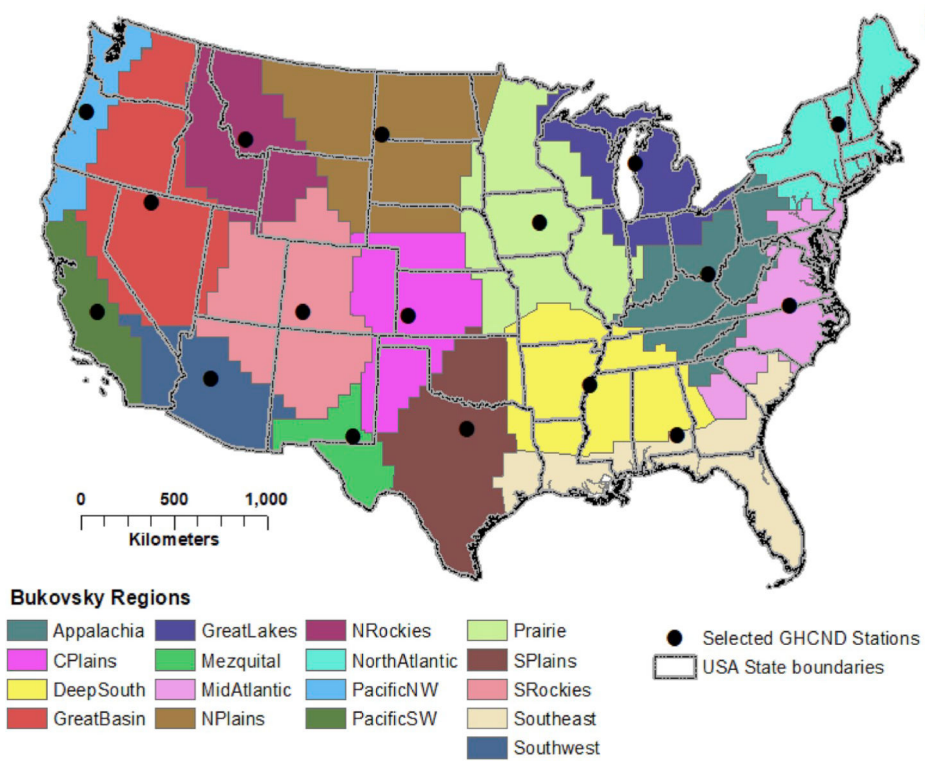


Fig. 4. The 17 Bukovsky regions of the conterminous US and the point locations of GHCN-D stations closest to the region centroid utilized in this demonstration of HMS-PCAT.

HMS: Precipitation Comparison

Overview Data Request **Output Data** Retrieve Data Data Algorithms

Precipitation: NCEI, DAYMET, PRISM, GLDAS, NLDAS Metadata

Metadata	Value
ncei_elevation	66.6
ncei_mindate	1893-01-01
ncei_maxdate	2019-02-28
ncei_latitude	44.6341
ncei_name	CORVALLIS STATE UNIVERSITY, OR US
ncei_datacoverage	1
ncei_id	GHCND:USC00351862
ncei_elevationUnit	METERS
ncei_longitude	-123.19
ncei	daily

Export JSON Export CSV

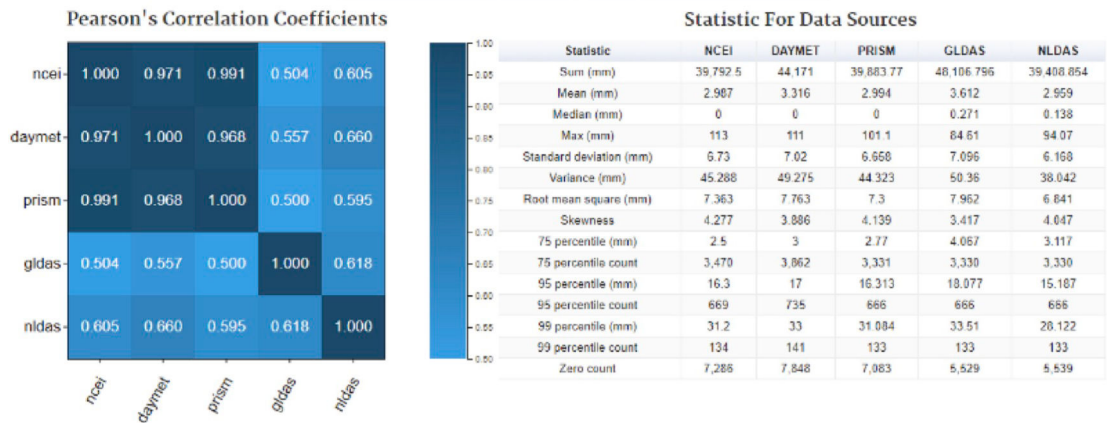
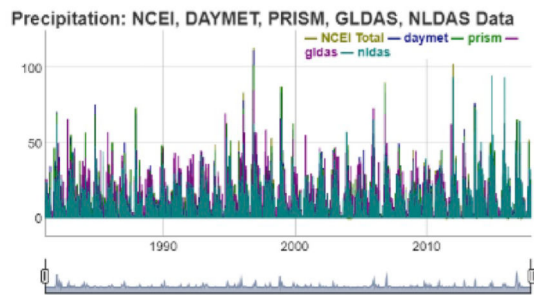
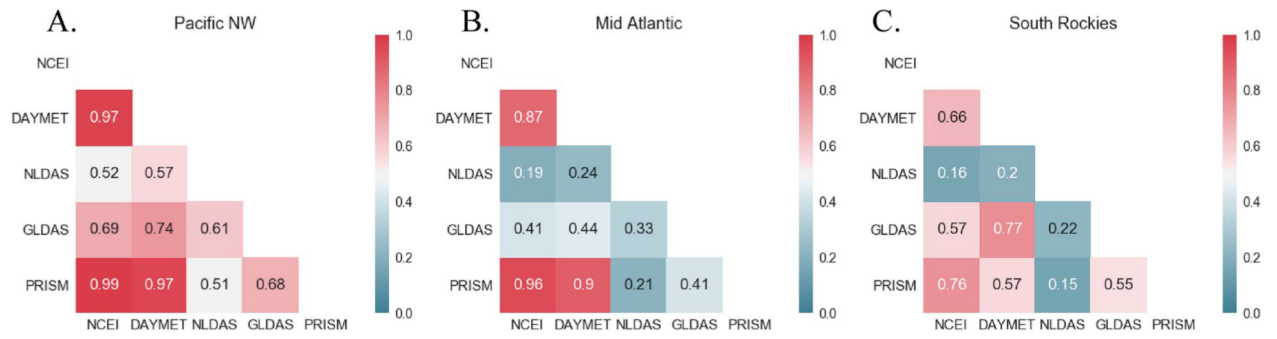


Fig. 5. Output page screen shot of HMS-PCAT showing data and results pulled from all five sources at the Pacific NW region station ID GHCND:USC00351862 from January 1981 to December 2017.

**Fig. 6.**

The Pearson's correlation matrix calculated by HMS-PCAT shows how each dataset compares to others in the Pacific NW (A), Mid-Atlantic (B), and South Rockies (C), respectively, from right to left.

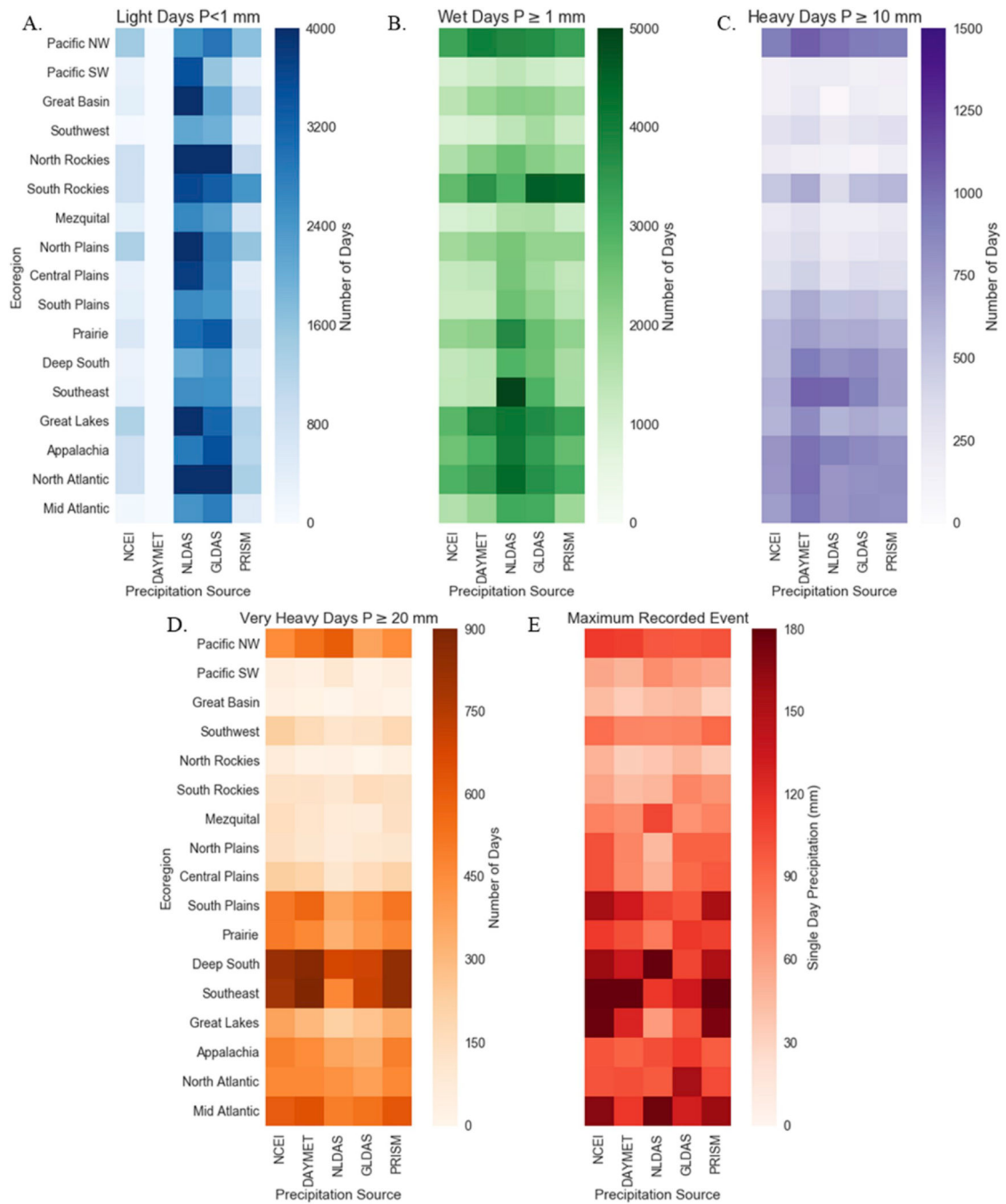


Fig. 7. Data from HMS-PCAT were used to calculate the total number of days out of the 37-year (13,514 day) time frame for threshold values of precipitation (P). CLIMDEX threshold indices are categorized by precipitation as (A) light, $P < 1$ mm; (B) wet, $P \geq 1$ mm, (C) heavy, $P > 10$ mm, and (D) very heavy days, $P > 20$ mm. Single day maximum precipitation value recorded in mm is shown in panel E.

Table 1

Description of different precipitation datasets used in environmental modeling. Includes units, temporal aggregation, spatial resolution, available time period, spatial coverage, maximum time lag for data processing and availability, and data creation method. Based on Sitterson et al. (2018).

Type/Name	Units	Temporal Aggregation	Resolution (Degree Grid)	Time Period	Coverage	Time lag	Method	Source/Reference
Rain Gauge								
GPCC Full Data	mm mo ⁻¹	Monthly	0.5 × 0.5	1901–2013	7000 US, 65,000 Worldwide	n/a	Weighted Method for grid	GPCC
^a NCEI	mm	Hourly, Daily	By Station	Varies	72 N – 15 S, –60 E 130 W	6 months	Gathering of multiple stations GHCN, COOP, QCLCD	NCEI
^a Daymet	mm d ⁻¹	Daily	0.0089 × 0.0089	1980–2017	N. America and Puerto Rico	1 year	Spatial truncation/Gaussian weighting filters of ground station locations	Daymet
Combined								
^a NLDAS	kg m ⁻² h ⁻¹	Hourly	0.125 × 0.125	1981–Present	N. America	4 days	Integration of CMORPH and RADAR	LDAS
^a GLDAS	kg m ⁻² s ⁻¹	3 Hourly	0.25 × 0.25	1949–Present [#]	90 N 60 S	2 months	Incorporation of satellites and ground-based observations	LDAS
^a PRISM	mm d ⁻¹	Daily, Monthly,	0.04 × 0.04	1981–2017	CONUS	1 year	Climatologically Aided Interpolation (CAD) of gauge stations with RADAR	PRISM
CMAP Pentad RT	mm d ⁻¹	Daily	2.5 × 2.5	1979–Dec 2016	88 N 88 S	1 month	Filling in gaps from gauge data with satellite (CMORPH)	CMAP
TRMM	mm h ⁻¹	3 Hourly	0.25 × 0.25	1998–2015	35 N 35 S to 50NS	n/a	Microwave, Infrared	TRMM
GPM	mm h ⁻¹	30 Minute	0.1 × 0.1	2014–Present	60 N 60 S	4–6 h	Microwave, Infrared, Satellite Precip Radar	GPM
CMORPH	mm h ⁻¹	30 Minute, 3 Hourly	0.07277 × 0.07277, 0.25 × 0.25	2002–Present	60 N 60 S	18 h	Morphing of Microwave and Infrared	CMORPH
PERSIANN CCS	mm h ⁻¹	Hourly	0.04 × 0.04	2003–Present	60 N 60 S	1–2 days	Infrared, Cloud segmentation algorithm	PER SIANN- CCS
PERSIANN CDR	mm d ⁻¹	Daily	0.25 × 0.25	1983–2015	60 N 60 S	n/a	Infrared, Artificial Neural Network	PERSIANN CDS
Radar								
NEXRAD	mm h ⁻¹	1,3 Hourly	1.1-nm × 1	1994–Present	160 sites in the US	2–4 days	Radar, Precipitation Processing System	NEXRAD
TDWR	mm h ⁻¹	Hourly	1.1-nm × 1	2001–Present	45 sites in US	4 days	Radar, Precipitation Processing System	RADAR

^aCurrently implemented in HMS-PCAT

[#]Includes version 1 and 2 data.

Table 2

Gauge stations used in demonstration with corresponding Bukovsky (2011) region, elevation of station, and percent missing data (in reference period 1981 to 2017 that were recorded as missing or invalid).

Station ID	Station Name	Bukovsky Region	Elevation (meter)	Percent Missing (%)
GHCND:USC00351862	Corvallis State University, OR	Pacific NW	68.6	0.95%
GHCND:USC00043747	Hanford 1 S, CA	Pacific SW	72.2	1.56%
GHCND:USC00265818	Orovada 3 W, NV	Great Basin	1280.2	4.89%
GHCND:USC00029166	Walnut Grove, AZ	Southwest	1147.3	1.82%
GHCND:USC00242409	Dillon U of Montana Western, MT	North Rockies	1563.3	1.00%
GHCND:USC00057656	Silverton, CO	South Rockies	2830.1	11.27%
GHCND:USC00291469	Carlsbad, NM	Mezquital	951	2.37%
GHCND:USC00320995	Bowman, ND	North Plains	908.3	0.61%
GHCND:USC00144464	Lakin, KS	Central Plains	913.8	0.56%
GHCND:USC00419532	Weatherford, TX	South Plains	291.1	1.29%
GHCND:USC00135198	Marshalltown, IA	Prairie	265.2	3.27%
GHCND:USC00033242	Helena, AR	Deep South	59.4	1.82%
GHCND:USC00018323	Troy, AL	Southeast	165.2	6.14%
GHCND:USC00205065	Manistee 3 SE, MI	Great Lakes	204.2	7.59%
GHCND:USC00336781	Portsmouth Sciotoville, OH	Appalachia	164.6	1.10%
GHCND:USC00436995	Rutland, VT	North Atlantic	189	1.14%
GHCND:USC00441746	Clarksville, VA	Mid-Atlantic	100.6	4.56%