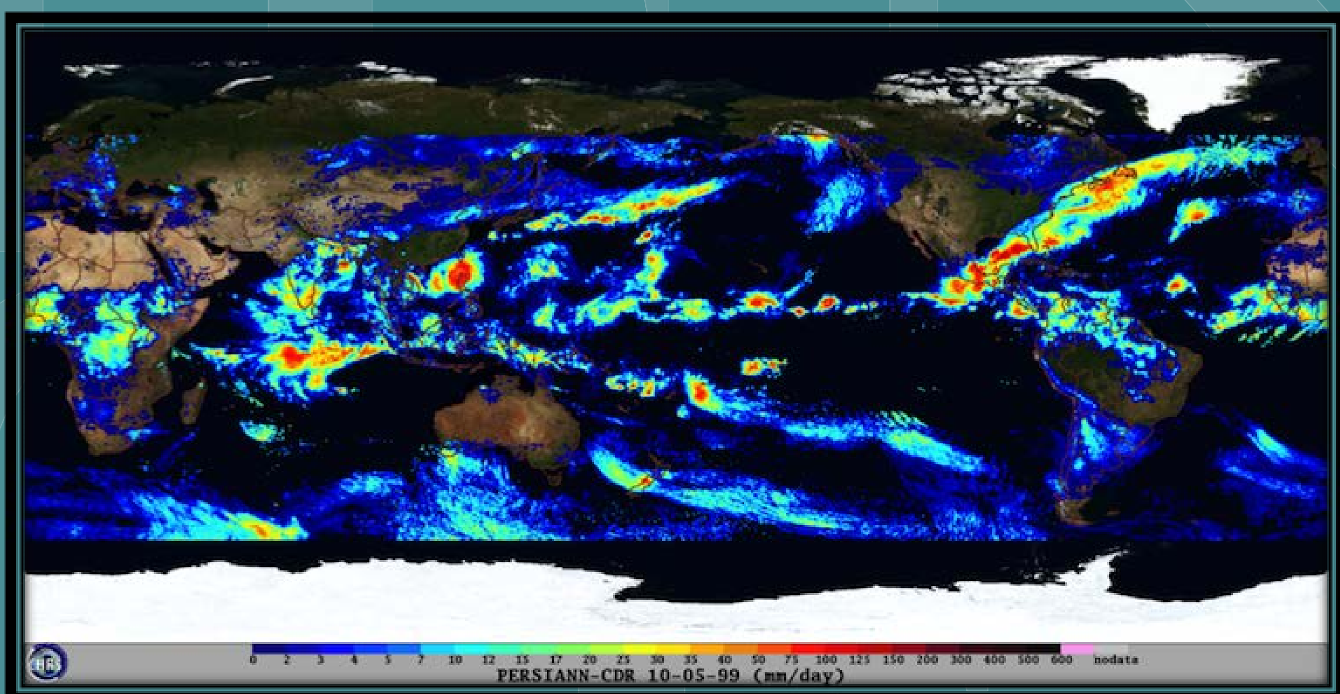
EPA/600/R-17/441 |September 2017| www.epa.gov/research

United States
Environmental Protection
Agency

A Survey of Precipitation Data for Environmental Modeling



EPA/600/R-14/152
September 2017

A Survey of Precipitation Data for Environmental Modeling

Jan Sitterson[1], Chris Knightes[2], Rajbir Parmar[2], Kurt Wolfe[2], Brian Avant[3], Amber Ignatius[4], Deron Smith[1]

[1]Oak Ridge Associated Universities
[2]U.S. Environmental Protection Agency, Office of Research and Development
[3]Oak Ridge Institute for Science and Education
[4]Oak Ridge Institute for Science and Education now at University of North Georgia

Chris Knightes - Project Officer

Office of Research and Development
National Exposure Research Laboratory
Athens, Georgia, 30605

Notice

The U.S. Environmental Protection Agency (EPA) through its Office of Research and Development funded and managed the research described here. The research described herein was conducted at the Computational Exposure Division of the U.S. Environmental Protection Agency National Exposure Research Laboratory in Athens, GA. Any mention of trade names, products, or services does not imply an endorsement by the U.S. Government or the U.S. Environmental Protection Agency. The EPA does not endorse any commercial products, services, or enterprises. This document has been reviewed by the U.S. Environmental Protection Agency, Office of Research and Development, and approved for publication.

Abstract

This report explores the types of precipitation data available for environmental modeling. Precipitation is the main driver in the hydrological cycle and modelers use this information to understand water quality and water availability. Models use observed precipitation information for modeling past or current conditions, while simulated data are used to predict future conditions as well as re-create historic conditions. Rain gauge-, radar-, and satellite-based measurements are categorized in the observed precipitation dataset. Calculated precipitation data from numerical weather predictors, stochastic models, and nonparametric models are part of the simulated data available for modeling. Temporal resolution, data availability, spatial resolution, and method of measuring precipitation are described for each dataset; global datasets and datasets of the contiguous United States are explained in this report. Our goal is to inform modelers of the various types, resolutions, and sources of precipitation data available for environmental modeling. We discuss only a few frequently cited datasets in detail due to the vast amounts of precipitation data available for modeling purposes.

Foreword

The U.S. Environmental Protection Agency (EPA) is charged by Congress with protecting the Nation's land, air, and water resources. Under a mandate of national environmental laws, the Agency strives to formulate and implement actions leading to a compatible balance between human activities and the ability of natural systems to support and nurture life. To meet this mandate, EPA's research program is providing data and technical support for solving environmental problems today and building a science knowledge base necessary to manage our ecological resources wisely, understand how pollutants affect our health, and prevent or reduce environmental risks in the future.

The National Exposure Research Laboratory (NERL) Computational Exposure Division (CED) develops and evaluates data, decision-support tools, and models to be applied to media-specific or receptor-specific problem areas. CED uses modeling-based approaches to characterize exposures, evaluate fate and transport, and support environmental diagnostics/forensics with input from multiple data sources. It also develops media- and receptor-specific models, process models, and decision support tools for use both within and outside of EPA.

The goal of the Hydrologic Micro Services (HMS) project is to develop a collection of inter-operable water quantity and quality modeling components. Components can be integrated to rapidly compose work flows to address water quantity and quality related questions. Each component may have multiple implementations ranging from macro (coarse) to micro (detailed) levels of modeling the physical processes. The components leverage existing internet-based data sources and sensors. They can be integrated into a work flow in two ways: calling a web service or downloading component libraries. It is generally more efficient to call a web service for less computational intensive components, yet, local copies of components are needed if the component requires large amounts of input/output data.

Elaine Hubal, Acting Division Director for CED

Table of Contents

List of Tables

List of Figures

Acronyms and Abbreviations

CAM	Community Atmosphere Model
CESM	Community Earth Systems Model
CHIRPS	Climate Hazards Group Infrared Precipitation with Station
CMAP	Climate Prediction Center Merged Analysis of Precipitation
CMIP	Coupled Model Intercomparison Project
CMORPH	Climate Prediction Center Morphing
ECHAM	European Centre Hamburg Model
GCM	Global Circulation Model
GEOS-16	Geostationary Operational Environmental Satellite System
GLDAS	Global Land Data Assimilation System
GPCC	Global Precipitation Climatology Centre
GPM	Global Precipitation Mission
NASA	National Aeronautics and Space Administration
NCDC/NCEI	National Climatic Data Center / National Center for Environmental Information
NEXRAD	Next Generation Weather Radar
NLDAS	North American Land Data Assimilation System
NWP	Numerical Weather Prediction
ORD	Office of Research and Development
PERSIANN	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
PRISM	Parameter-elevation Relationship on Independent Slopes Model
RADAR	Radio Detection and Ranging
RCP	Representative Concentration Pathways
SWAT	Soil and Water Assessment Tool
TDWR	Terminal Doppler Weather Radar
TRMM	Tropical Rainfall Measuring Mission
US EPA	United States Environmental Protection Agency
WGEN	Weather Generator
WRF	Weather Research and Forecasting

1. Introduction

This report provides a survey of precipitation data sources and generation methods for environmental modeling. As a main component of the hydrological cycle and contaminant fate and transport, precipitation data are needed for hydrological modeling, erosion modeling, and water quality research. We compile descriptions of several precipitation datasets and data generation methods available globally and for the contiguous United States, specifically. The datasets and data generation methods included are publicly available online and often cited by the modeling community. Some of the datasets are purely observed and derived from rain gauges, radar, and satellites, while others are simulated datasets generated by mathematical equations to predict future weather conditions and recreate past events. Additionally, some precipitation data products are derived from a combination of observed data and model equations to generate weather estimates.

Precipitation has great importance because it influences drinking water availability, supports agriculture, and maintains freshwater resources. It is a vital component in the global hydrological cycle due to its direct effect on the circulation of Earth's latent heat (Ebert, 2007). Chahine (1992) states that "the hydrological cycle is the largest movement of any substance on Earth's surface". Most water movement occurs through precipitation and evaporation. Controlled by the sun's radiation, water evaporates from the ocean and the land's surface where it moves with winds in the atmosphere. It then condenses into clouds to fall back to Earth's surface as precipitation flowing toward the oceans to complete the global hydrological cycle (Chahine, 1992). With the exception of arid climates, precipitation often exceeds evaporation over land and the excess drains to a reservoir or recharges groundwater (Fig. 1). Precipitation is highly variable and influences vegetation, droughts, floods, and the movement of minerals and chemicals. In agriculture and urban areas, precipitation drives contaminant and nutrient transport in water systems through runoff.

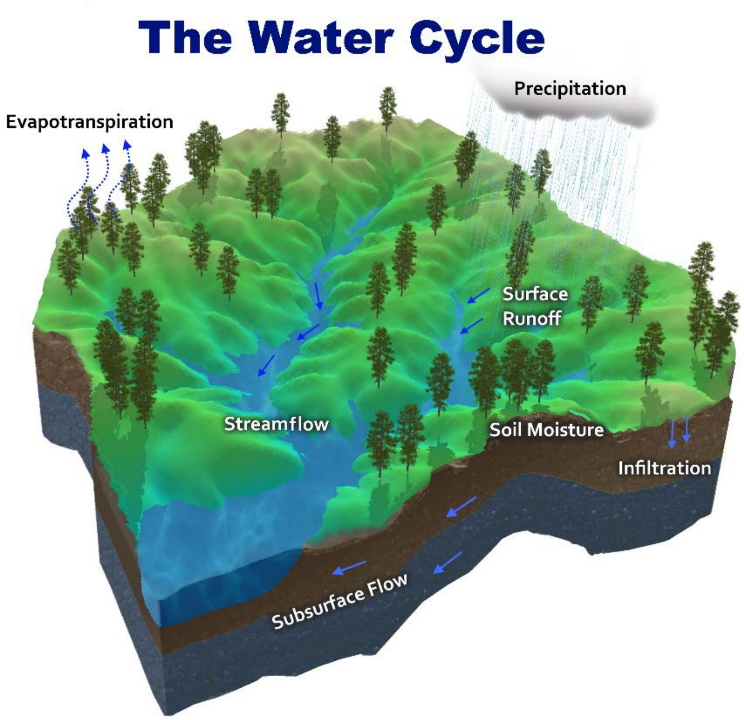


Figure 1. A simplified diagram of the water cycle within a watershed. (Brewster, 2017; ESRI, 2015)

Precipitation data are integral inputs for many watershed, air, erosion and agricultural models as well as climate-predicting projects. It determines flood/drought conditions, hydrologic transportation of contaminants, best management practices, and regulations. Precipitation data are generated through direct observation and model simulation. Observed data are captured directly from rain gauge stations, or technologically observed from radar and satellites. Simulated precipitation data are mathematically generated through parameterizations, statistical probability, or historical trends. There are many types of precipitation datasets with different spatial and temporal resolutions available for project needs (Table 1). Each has strengths and weaknesses depending on its intended use. This report presents different types of available precipitation datasets with details including temporal and spatial resolution, potential errors in the dataset, and optimal performance scenarios. It also discusses the benefits and deficiencies of certain precipitation datasets. Focusing on a project's purpose and understanding its questions, goals, and needs are vital for selecting input data since exploratory, planning, and regulatory purposes have different input criteria and uncertainty thresholds (Harmel et al., 2014).

2. History of Precipitation Data

Rain gauge records have been available for hundreds of years. The first scientific report on differences in measured precipitation using height from rain gauges was by William Heberden in 1770 (Tapiador et al., 2012). As technology has advanced, rain gauge data has become more accurate in determining the amount of rainfall at a particular location. During World War II, radar operators searching for enemy ships and aircraft found that precipitation caused 'false' echoes on their screens (NOAA, 2017); thus began the development of precipitation detection radar. Radar-based precipitation information overcame the lack of spatial resolution in rain gauge data (Hu et al., 2014). As research in meteorology continued, there was a need to study macro-scale rainfall which then led to the use of satellites to monitor cloud cover and precipitation events around the world. Scientific advancements in spacecraft satellites and high-resolution sensors provide information to calculate precipitation amounts. Observed data gives a realistic view of precipitation in the past or in real time, but cannot predict future conditions. Using mathematical equations, simulated precipitation was therefore created to fill data gaps and predict future scenarios.

With so many precipitation datasets being available, rain gauge data are universally considered the best source of reference data for precipitation observations (Tapiador et al., 2012). Some weaknesses of rain gauges include being able to observe precipitation at only one site in space and often underestimating rainfall amounts. Despite these limitations, many researchers use rain gauge data because it is assumed to represent the most accurate source of information at the exact location, and installing a rain gauge is easy and fairly inexpensive (Kim, 2014; Price et al., 2014). Rain gauge networks have been around for at least a century and, therefore, provide the longest precipitation record. Because rain gauge data are widely used and free of assumptions, research methodology involving rain gauge networks is well accepted. Research on climatology requires long-term datasets of precipitation, and only rain gauge data have enough historical information for this type of research. Other methods of observing precipitation do not have the necessary decadal time series although they do have better spatial resolution.

3. Differences in Observed and Simulated Precipitation Data

Both observational and simulated precipitation datasets have strengths and weaknesses. Observed datasets give information about past or current rainfall events, but often have gaps in the time series due to lack of measurement. Observed data are also more localized spatially due to providing information at specific sites or over an area. Rain gauge data are observed at a single location in space and interpolation methods must be used to estimate precipitation across a broader spatial extent, or assumed to represent constant precipitation over a region. Precipitation simulations can provide past or future precipitation quantities in a seamless time series over a global extent, at different spatial discretizations. Generally, simulated data are best used for non-extreme weather patterns, mountainous regions, and colder weather; observed data performs best in warm weather and documents extreme events very well (Table 2) (Harmel et al., 2002). Satellite-derived data perform better than numerical models in warm seasons and over the tropics (Ebert, 2007; Hu et al., 2014). Studies have shown that input precipitation data from observed and simulated datasets impact watershed model outputs (Golden et al., 2010; Tuo et al., 2016). Modeling projects that use more than one type of dataset may be more accurate in reproducing precipitation patterns than a single dataset, but the spatial and temporal resolutions of different datasets must be considered in calibration (Tuo et al., 2016). Observed and simulated data can generate outputs in different spatial scales and can be provided as accumulated precipitation per hour, day, or month. Simple algorithms fix time or measurement differences and find a common spatial resolution.

Table 1. Comparison of the major types of precipitation datasets.

Type	Method	Spatial Extent	Time Series	Time Step	Best Performance
Observed	Direct measurement or technologically observed	Specific or range of area	Includes gaps of past or current data	Half-hourly, hourly, daily, monthly, yearly	Warm weather, extreme events
Simulated	Numerical calculations based on historical events	Global, downscaled to regional	Seamless, future predictions, past data	Daily, monthly, yearly	Cold weather, mountainous regions, non-extreme events

Table 2. Description of Precipitation Datasets

	Type/Name	Precip Output	Temporal Resolution	Resolution (Degree Grid)	Time Period	Coverage	Time lag	Method	Source
Satellite	TRMM	mm/hr	3 hourly	0.25x0.25	1998-2015	35N 35S to 50NS	n/a	Microwave, Infrared	TRMM
	GPM	mm/hr	0.5 hourly	0.1x0.1	2014-	60N 60S	4-6 hours	Microwave, Infrared, Satellite Precip Radar	GPM
	CMORPH	mm/hr	0.5, 3 hourly	0.07277x0.07277, 0.25x0.25	2002-	60N 60S	18 hours	Morphing of Microwave and Infrared	CMORPH
	PERSIANN CCS	mm/hr	Hourly	0.04x0.04	2003-	60N 60S	1-2 days	Infrared, Cloud segmentation algorithm	PERSIANN-CCS
	PERSIANN CDR	mm/day	Daily	0.25x0.25	1983-2015	60N 60S	n/a	Infrared, Artificial Neural Network	PERSIANN CDS
Radar	NEXRAD	mm/hr	1, 3 hour	1x1 N. America	1994-	160 sites in the US	2-4 days	Radar, Precipitation Processing System	NEXRAD
	TDWR	mm/hr	Hourly	1x1 N. America	2001-	45 sites in US	4days	Radar, Precipitation Processing System	RADAR
Rain Gauge	GPCC Full Data	mm/mo	Monthly	0.5x0.5	1901-2013	7000 US, 65000 Worldwide	n/a	Weighted Method for grid	GPCC
	NCDC	inch	Hourly	By Station	1951-	72N -15S, -60E 130W	6 months	Gathering of multiple stations GHCN, COOP, QCLCD	NCDC
	Daymet	mm/day	Daily	0.0089x0.0089	1980-2015	N. America	1 year	Spatial truncation of Gaussian weighting filters of ground station locations	Daymet
Combined	NLDAS	kg/m2/hr	hourly	0.125x0.125	1979-	N. America	4 days	Integration of CMORPH and RADAR	LDAS
	GLDAS	kg/m2/hr	3 hourly	0.25x0.25	2000-Dec 2016	90N 60S	2 months	Incorporation of satellites and ground-based observations	LDAS
	PRISM	mm/mo	Monthly, Yearly	0.04x0.04	1981-	CONUS	1 month	Climatologically Aided Interpolation (CAI) of gauge stations with RADAR	PRISM
	CMAP Pentad RT	mm/day	Daily	2.5x2.5	1979-Dec 2016	88N 88S	1 month	Filling in gaps from gauge data with satellite (CMORPH)	CMAP
Simulated	WRF	mm/hr	Daily	0.03x0.03	User Specified	Global	n/a	NWP Microphysics/Cumulus Schemes	WRF
	ECHAM	mm/day	Daily	0.703125x0.703125	User Specified	Global	n/a	Numeric Weather Prediction and Parameterization	ECHAM
	CESM-CAM	mm/day	Daily	0.35x0.35	User Specified	Global	n/a	NWP and Non-parametric, CMIP5	CAM
	WGEN	mm/day	Daily	HRU	1960-2100	Site Specific	n/a	Stochastic	WGEN

3.1 Spatial Resolution

The spatial resolution in purely observed datasets are not uniform due to random station locations or radar blocking. Gridded precipitation data, i.e., those that provide precipitation information at each point across an entire domain at a specified grid resolution, are useful in environmental modeling. To produce a gridded precipitation dataset, values at locations without observations are manipulated using many methods including, nearest neighbor, weighted average, geostatistics, mechanistic methods, etc. Such methods assume that points close to each other are better correlated (Tuo et al., 2016). Since rainfall is not distributed evenly, rainfall estimates are often misleading; interpolation of observed data is a significant limitation in accurately modeling responses to rainfall because data cannot be validated at every position. Globally-simulated datasets must be downscaled to reflect a study area. This can cause error because global generalizations may not represent local processes. The statistical relationships between large climatic parameters and local variables affecting precipitation (e.g., temperature and its effects on evaporation) are needed to downscale global outputs accurately (Wilks & Wilby, 1999). A major problem in scaling to a grid for observed and simulated data is that precipitation is not evenly distributed and values may differ within a grid. With all precipitation datasets, coarser spatial resolutions lead to more approximations about rainfall distribution, and interpolation introduces known biases to the results (Tapiador et al., 2012).

3.2 Precipitation Data Limitations

There is no way to determine the exact weather condition at every point in space, which means that all datasets have limitations. Observational data often have missing values due to station maintenance or equipment malfunction; error sources can be due to sampling errors, calibration uncertainty, or random errors. Instrument and calibration uncertainty also pose potential sources of bias. Due to the inability to accurately measure frozen precipitation in all observational techniques, observed datasets are most accurate during warm weather conditions. The length of recorded data also differs between datasets. Radar and satellite data do not yet have records old enough for climatology research, which requires at least 30 years of historical data. Simulated outputs from mathematical equations do not depict precipitation events with as much detail as observed datasets. Correctly simulating patterns, seasonal variations, and characteristics of precipitation with mathematical models is an area of active research (Eyring et al., 2016). Harmel et al. (2002) revealed that weather variability, especially in extreme events, is difficult to predict since the event does not fit common mathematical distributions. Model drift is a common problem among simulated precipitation datasets due to the probability distribution being skewed away from observed data; modelers often can compensate for this using a correction factor after simulation runs. The algorithm or parameterization scheme selected in a numerical weather prediction model influences model uncertainty since some schemes work better in certain locations. Combining different output datasets improves regional and global precipitation data results (Ebert, 2007; Huffman et al., 1995; NOAA, 2017).

4. Observed Data

Observational data provide a historical record of past precipitation events using direct rainwater catchment in rain gauges or with technical instruments from a distance (e.g., sensors). A historical record of precipitation is helpful in looking at changes and trends in day-to-day climate. Observed precipitation data are specific to the sampling location and range of detection (Table 3), which often leaves gaps in the spatial and temporal resolution of the dataset. Observational systems give an accurate depiction of the amount of rain produced by an extreme event like hurricanes or monsoons. Rain gauge data can provide estimates of rain accumulation at exact locations; when using rain gauge data, an observed value represents uniform precipitation in the area around the gauge. If two or more rain gauges are used for a study area, a method for interpolating data across the region, such as the Thiessen Polygon Method, can be applied. Radar and satellite data have larger ranges of detection and can serve as a warning devices for current and near future precipitation events. Many studies have compared observational datasets and their ability to estimate precipitation amounts and their effect on model output. Gao et al. (2017) studied the impacts of three different precipitation sources (rain gauge, radar, and a combined reanalysis dataset) in SWAT streamflow simulations.

Table 3. Summary of observed precipitation dataset characteristics.

Observed	Method	Spatial Extent	Spatial Resolution	Temporal Resolution	Years of data	Precipitation Output	Error
Rain Gauge	Physically collected on the ground	Specific locations	Lat.-Lon. of station, 0.009x0.009 or 0.5x0.5-degree grid interpolation	Hourly, Daily, Monthly, Yearly	100 years	Underestimates heavy rainfall events	Random error, mechanical issues, location
Radar	Technologically collected on the ground	Radial area around station (radius 230km)	1x1 degree grid, lat. and long of station	Hourly, 3-hourly,	30-40 years	Overestimates heavy rainfall events, underestimates light rainfall	Signal blockage, hail misreading
Satellite	Technologically collected from space	Latitude range (60°N, 60°S)	0.04x0.04-degree grid 0.1x0.1-degree grid 0.25x0.25-degree grid	Half-Hourly, Hourly, Daily	20 years or less	Underestimates rain from warm top clouds	Frozen precipitation, multilayer clouds

4.1 Rain Gauge

Rain gauge precipitation data represent the direct capture and measurement of rainwater at a specific location. Inexpensive and easy to install rain gauges are found all over the world. There are many different methods for capturing direct rainfall varying in complexity from measured cylinders to sophisticated weighing gauges. One of the most common methods for determining the amount of precipitation is the tipping bucket method, which records the time and a count of the number of times a premeasured bucket tips from overfilling with rainwater (Tapiador et al., 2012). The tipping bucket method becomes more inaccurate with increasing rainfall intensities due to catching and counting errors (Shedekar et al., 2016). Gauge data are the most accurate representation of precipitation at a precise location (Kim, 2014; Price et al., 2014) but limitations stem from mechanical issues or operational errors. Random errors in datasets can occur due to damage to the gauge from wildlife or humans, and gauge data often underestimates precipitation amount due to wind effects, frozen precipitation, and rain particles that evaporate before contact with the gauge (Kidd & Huffman, 2011; Tapiador et al., 2012). Despite these limitations, many studies use gauge stations in modeling total maximum daily loads for management purposes. Liu et al. (2008) used precipitation data from five stations to model nitrogen transportation in three models (WASP[1], EFDC[2], and HSPF[3]). Below is a short description of a few often cited rain gauge data sources and datasets that primarily use rain gauge data.

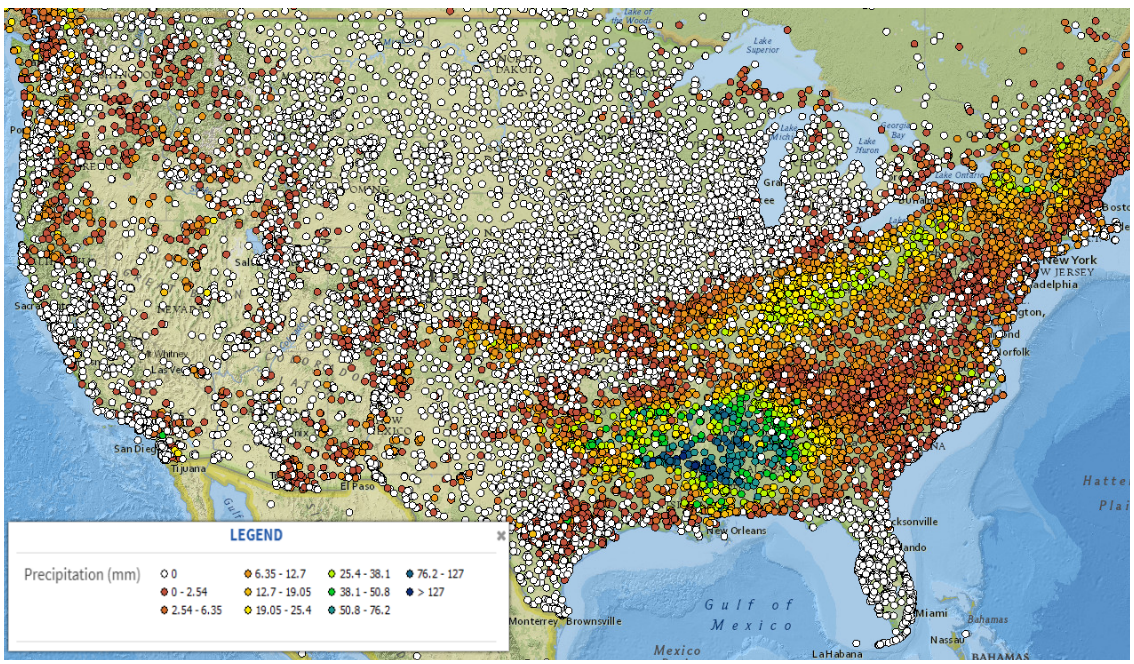


Figure 2. Map of rain gauge stations in the United States showing precipitation detected on April, 12, 2016. Each dot represents one rain gauge at a specific location, colored by the amount of precipitation measured. Image from NOAA's website.

4.1.1 NCDC / NCEI

The National Climatic Data Center (NCDC), now named the National Center for Environmental Information (NCEI), provides precipitation data recorded at rain gauge

[1] Water Quality Analysis Simulation Program epawasp.twool.com
[2] Environmental Fluid Dynamics Code https://www.epa.gov/exposure-assessment-models/efdc
[3] Hydrological Simulation Program Fortran https://www.epa.gov/exposure-assessment-models/hspf

stations around the world. NCEI uses a network of volunteer centers to collect daily weather observations from local rain gauges. NCEI has access to about 53,000 stations worldwide some with data going as far back as 1901 (NOAA, 2017), these data can be accessed online and queried by location (Latitude, Longitude) of a gauge station. The spatial coverage of NCEI gauge stations in the United States is shown in Figure 2; the network of gauge stations and precipitation data can be accessed at NOAA's website.

4.1.2 GPCC

The Global Precipitation Climatology Centre (GPCC), another source for global precipitation data from rain gauge stations on the ground, is comprised of about 67,000 rain gauge stations worldwide. This dataset is available from 1901 to 2013 on a monthly time step and provides interpolated data on three grid sizes, with the finest resolution on a 0.5-degree grid (NOAA, 2017). A spherical adaptation of Shepard's empirical weighting scheme is used to transpose gauge stations to a grid point (Becker, 2013). GPCC data are accessible at UCAR's website.

4.1.3 Daymet

Daymet is a daily dataset of rain gauge data interpolated and extrapolated by the Daymet algorithm. It uses ground station data from NCEI with its model algorithm to produce gridded estimates of daily weather parameters. Interpolation for the gridded resolution uses the spatial convolution of a truncated Gaussian filter from the local station density (Thornton, 2017). Daily rainfall is rounded to the nearest whole number. The interpolated spatial resolution is about a 0.009-degree grid, or approximately 1km resolution, over North America. Data are accessible since 1980, to the latest full year, due to interpolation at Daymet's website.

4.2 RADAR

Radio Detection and Ranging (RADAR), detects precipitation in the troposphere. Radar weather systems are mostly found on the ground, with a few on satellites. They send radio waves into the atmosphere in pulses and radio waves are sent back when the wave makes contact with a raindrop. The system calculates the distance and direction of the rain and uses the Doppler Effect to provide precipitation characteristics like reflectivity and droplet size (NOAA, 2017). A reflectivity-to-rainfall equation – the Precipitation Processing Subsystem (PPS) -- estimates rainfall amounts (Nelson et al., 2010). The relationship between reflectivity and rainfall amount varies for different forms of precipitation, resulting in uncertainty. An example of modified return values for different precipitation types is that interpretation of hail from radar can send a signal that resembles heavy rainfall. Radar data can make short-term forecasts and show intensity of a storm event as a warning to the public (NOAA, 2017). Because radar stations are very expensive, some countries cannot afford radar equipment. With its higher range of detection, radar data covers more area than rain gauge data. Radar technology can locate precipitation within a range of 230 km from the station and reports data close to real time (NOAA, 2017). Although radar detects rainfall on a larger scale than rain gauges, ground-based radar cannot reach high altitude clouds and signals can be blocked by topographic effects. Bias in the radar dataset comes from signal blockage, bright band contaminations, range dependency, and radar calibration errors.

4.2.1 NEXRAD

Next Generation Weather Radar (NEXRAD) is the largest collection of forecasting radars accessible for research in North America. It is composed of 160 WSR-88D ground radars across the United States (see Figure 2). NEXRAD has two levels of output data, Level-II and Level-III. Level-II is raw meteorological data including reflectivity, radial velocity, and spectrum width. Level-III is a set of computer-processed products that include hourly precipitation and bias-corrected precipitation from rain gauges. NEXRAD observations are taken every hour and can be gridded on a one-degree grid two to four days after observations using a spatial weighting scheme. NEXRAD's station network has been operational since 1994, which is not long enough for climatology research (Nelson et al., 2010). Price et al. (2013) investigated whether NEXRAD data, corrected with rain gauge data by the Multisensor Precipitation Estimation (MPE) algorithm, would improve simulations in watershed models; they found that adjusted radar precipitation estimates using gauge data consistently performed better than non-adjusted radar data. NEXRAD data archive can be accessed at [NOAA's website](#).

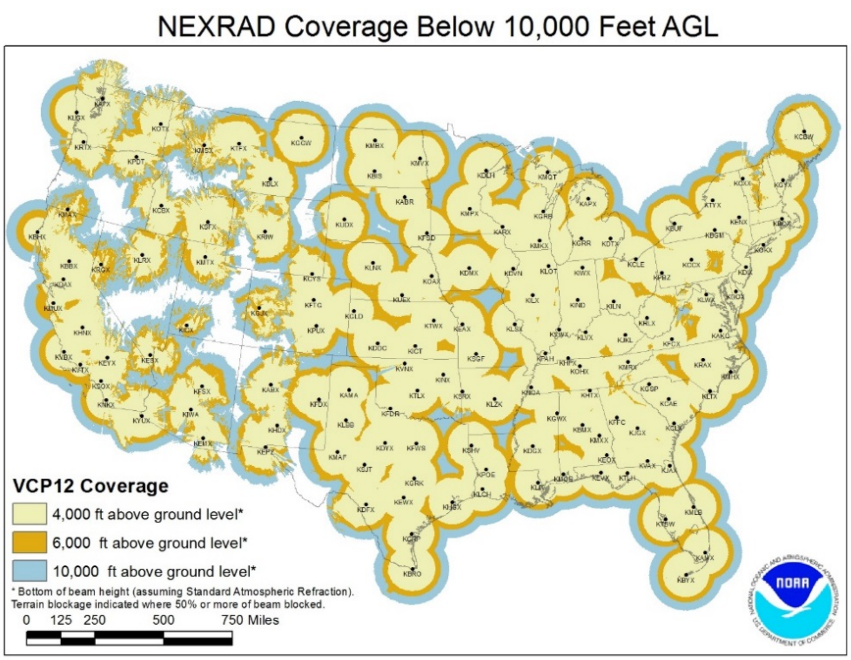


Figure 3. A map showing the NEXRAD site locations in the contiguous United States. Image from (NOAA, 2017)

4.3 Satellite

Satellite-based precipitation data are derived from infrared and microwave measurements taken from satellites in space. Infrared information gives cloud top temperature which can be used in an algorithm to produce rainfall amounts. The microwave measurement gives information about cloud depth and layer characteristics which is physically related to the formation of precipitation (Duan et al., 2016). Satellites are the only way to retrieve global homogeneous estimates of precipitation (Tapiador et al., 2012). They provide an estimate of precipitation in millimeters per hour at high resolutions over their span of orbits. Some datasets use a single satellite for specific coverage (TRMM), while others use an array of satellites like GPM for global coverage (Table 4). Combining diverse satellite observations lowers the potential to miss precipitation events. Missing data can be found in satellite-derived data due to instrument error or lack of spatial coverage. Satellite observations cannot detect frozen precipitation or snowfall accumulation very well due to the complexity of the radiative properties of snowflakes and ice crystals (Kidd & Huffman, 2011). Satellite-derived precipitation data tends to underestimate rain from warm top clouds due to the infrared sensory tools used (Awange et al., 2016). Multilayer cloud systems also pose a threat to miscalculations because cloud layers can block the sensor's ability to detect the precipitating layer (Tapiador et al., 2012). This technology has been around for at least 30 years but, since each satellite-derived dataset has different temporal and spatial resolution, period of activation and methods of calculating precipitation, each dataset has information for only the time the satellite was operational. Although there are many satellite precipitation datasets and satellite-derived products, only the four most recommended and most recent satellite datasets are described below. The newest precipitation dataset, the GOES-16 satellite which was launched in November 2016, will provide atmospheric measurements of Earth with more spectral bands than its predecessors. For further information about the use of satellite data in modeling see Bitew and Gebremichael (2011) who used PERSIANN and CMORPH datasets in the hydrological model MIKE-SHE and Tramblay et al. (2016) who compared CMORPH, RFE, TRMM, and PERSIANN in the a hydrological model for water resource management in ungauged areas.

Table 4. Description of precipitation datasets that use satellite technology.

Name	Spatial Coverage	Time step	Spatial Resolution	Time Frame	Category	Method
TRMM	35N, 35S	3hr	0.25x0.25	1998-2015	Single Satellite	Microwave, Infrared
GPM	60N 60S	0.5 hour	0.1x0.1	2014-present	Multi-Satellite-radar	Microwave, Infrared, Satellite Precip Radar
CMORPH	60N 60S	0.5, 3-hour	0.07277x0.07277, 0.25x0.25	2002-present	Multi-Satellite	Morphing of Microwave and Infrared
PERSIANN CCS	60N 60S	Hourly	0.04x0.04	2003-present	Multi-Satellite	Infrared & cloud segmentation algorithm

Name	Spatial Coverage	Time step	Spatial Resolution	Time Frame	Category	Method
PERSIANN CDS	60N 60S	Daily	0.25x0.25	1983-2015	Multi-Satellite	Infrared, Artificial Neural Network
CHIRPS	50N, 50S	Daily	0.05x0.05	1981-present	Satellite-gauge	Infrared with station data
CMAP	Global	Monthly	2.5x2.5	1979-2006	Satellite-gauge	Infrared with station data
NLDAS	N. America	Hourly	0.125x0.125	1979-present	Multi-Satellite-radar	Integration of CMORPH and RADAR
GLDAS	Global	3hr	0.25x0.25	2000-Dec 2016	Satellite-gauge	Incorporation of satellites and ground-based observations

4.3.1 TRMM

The Tropical Rainfall Measuring Mission (TRMM) was a single satellite used to detect precipitation and tropical storms near the equator to better understand climate and weather patterns. TRMM used microwave and infrared information to calculate precipitation every three hours at a resolution of 0.25-degrees (see Figure 4). Although TRMM was deactivated in 2015, its data from 1998 to 2015 has been used in numerous publications. TRMM took tropical measurements of precipitation covering 35 degrees north to 35 degrees south and had multiple reanalysis products. TRMM data can be accessed at NASA's website (Skofronick-Jackson, 2017).

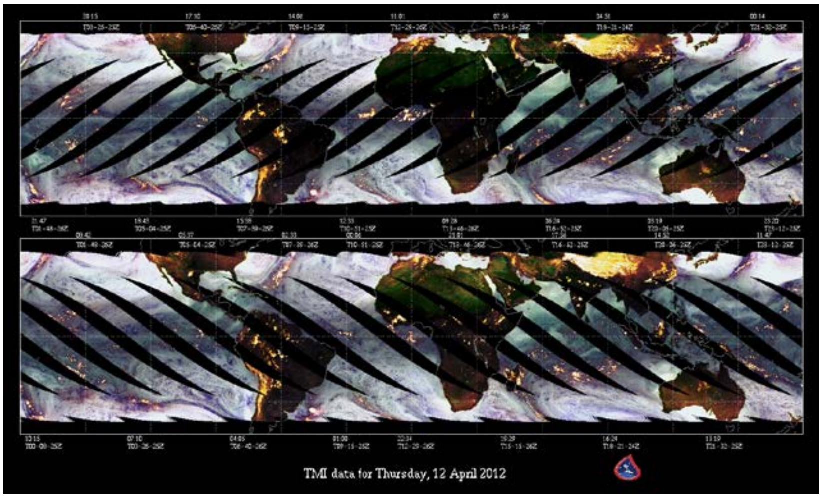


Figure 4. TRMM's orbit path captures the spatial resolution over the tropics on April 12th, 2012. Yellow areas depict probable rainfall. Image from NASA.

4.3.2 GPM

The Global Precipitation Mission (GPM) of 2014 is a continuation of TRMM's mission. Partners from the United States, Japan, France, India and the European Union allow GPM to track precipitation across the globe using 10 satellites. With a fine global resolution of 0.1-degree taken every half hour, GPM provides precipitation measurements using microwave, infrared, and radar technology. It has greater accuracy than its predecessor and covers the earth between 60 degrees north and 60 degrees south. Due to calculation time, the dataset can be accessed four to six hours after observation on NASA's website (Skofronick-Jackson, 2017).

4.3.3 CMORPH

The Climate Prediction Center Morphing (CMORPH) product is another highly recommended satellite dataset for precipitation. It gets its name from the method of calculating precipitation by morphing microwave and infrared information from multiple satellites at a 0.0730-degree resolution. Observations are taken every half hour and accumulated every three hours with up to an 18-hour lag in data accessibility. CMORPH has provided precipitation data since 2002 over the span of 60 degrees north to 60 degrees south. CMORPH data can be accessed at NOAA's website.

4.3.4 PERSIANN

The Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) has satellite derived datasets calculated from infrared imagery and artificial neural network algorithms (CHRS, 2004). This dataset's coverage is between 60 degrees north and 60 degrees south. PERSIANN has two products that provide precipitation data, one on an hourly time step (CCS) and one on a daily time step (CDR). The PERSIANN CCS (Cloud Classification System) has a resolution of 0.04-degrees with data from 2003 to the near present. Due to the complexity of algorithms, the CCS data can be accessed one or two days after observation. PERSIANN CDR (Climate Data Record) has a resolution of 0.25-degrees with data from 1983 to June 2016, with an even greater lag in accessibility. All PERSIANN data can be accessed at CHRS's data portal.

4.4 Combining Datasets

Combining observed precipitation measurements often involves two or more types of observations. An example of combined precipitation data are reanalysis products such as North American Regional Reanalysis (NARR) and Climate Forecast System Reanalysis (CFSR). Reanalysis products are generated when multiple precipitation datasets are merged onto a regularly-spaced grid to produce a consistent spatiotemporal output (NOAA, 2017). Blending and merging observed datasets can significantly improve precipitation estimates (Ebert, 2007; NOAA, 2017). Since rain gauge data are often sparse in some areas and may contain missing values, satellite and radar data has been combined with gauge data to fill the gaps, as in the Climate Prediction Center Merged Analysis of Precipitation (CMAP) and Climate Hazards Group Infrared Precipitation with Station data (CHIRPS). In addition to the previously stated datasets, there are many more datasets that combine multiple observational methods including the popular LDAS and PRISM datasets that are described in detail below. An example of the use of combined datasets is Radcliffe and Mukundan (2017) which compared the effect of PRISM and CFSR, in a SWAT model of streamflow.

4.4.1 LDAS

The North American Land Data Assimilation System (NLDAS) combines North American radar data and satellite data from CMORPH which allows higher resolutions and better accuracy for detecting precipitation. NLDAS has an hourly time step on a 0.125 -degree grid of North America and a maximum time lag of four days for data retrieval. The Global Land Data Assimilation System (GLDAS) combines satellite data and ground-based observational data to provide precipitation and other variables on a spatial resolution of 0.25-degrees, covering the Earth between 90 degrees north and 60 degrees south. GLDAS data are given every three hours and takes a least a month to process. EPA's BASINS system combines NLDAS data with NCEI data for plugging in missing values in order to have a near-seamless time series of precipitation data. Lee et al. (2010) compared NLDAS to NCDC station data in the HSPF tool to improve streamflow predictions for water quality assessments. NLDAS and GLDAS data can be accessed at *NASA's website*.

4.4.2 PRISM

The Parameter-elevation Relationship on Independent Slopes Model (PRISM) provides climatology information by combining ground gauge stations from multiple sources and radar products. The data is provided on a four by four kilometer spatial resolution covering the contiguous United States from 1981 to present on a monthly or annul temporal resolution. The method used to produce the gridded dataset is a combination of the Climatologically Aided Interpolation (CAI) method, Digital Elevation Model (DEM), and radar interpolation. With this methodology, the longer time-step is able to capture orographic precipitation patterns in mountainous areas better than a daily interpolation (Daly et al., 2008). PRISM data can be retrieved from the PRISM Climate Group website.

5. Simulated Data

Even after combining different types of observed data, there may still be missing values in the datasets. Simulated data, based on computer models, can fill these gaps to produce a continuous time series for model input. Weather prediction models are mathematically-driven models that simulate precipitation from the past as well as the future. Simulated precipitation measurements make computation easier for modelers because there are no missing values nor time spent on data retrieval. Three main types of models simulate precipitation data: Numerical Weather Predictors (NWP), stochastic models, and nonparametric models, as described in Table 5. Simulating weather characteristics is difficult because climatic processes can occur below the grid size of the model which leads to generalizations that introduce bias. Despite these limitations, some models closely mimic true weather patterns and can be used to study and manage water quality and water supply (Harmel et al., 2002). An example of the application of simulated precipitation data is the paper by Golden et al. (2013). Golden et al. used simulated rainfall amounts from Global Circulation Models in three watershed models (VELMA, GBMM, TOPLOAD) to determine how future changes in climate may impact watershed mercury transport.

Table 5. Summary of simulated precipitation dataset characteristics.

Simulated	Method	Inputs	Spatial Resolution	Spatial Extent	Error
Numerical Weather Prediction	Cumulus and microphysics schemes	Atmospheric conditions and thresholds	0.03x0.03-degree grid 0.703x0.703-degree grid	Global or limited area model (LAM)	Scheme selection
Stochastic	Probability	20-year history of precipitation	Site specific	Global or delineated area	Skewed distributions
Non-Parametric	Historical trends	Long historical records and emission scenarios	0.35x0.35-degree grid 1.4x1.4-degree grid	Global or regional downscaled	Model drift from observed data

5.1 Numerical Weather Prediction (NWP)

Numerical Weather Prediction models integrate differential equations that describe fluid flows to predict rainfall and other atmospheric conditions. Two major types of equations used to estimate precipitation describe microphysics and cumulus clouds. Microphysics parameterization schemes resolve the process of rain production, and cumulus parameterization schemes describe effects of cumulus clouds in rain events. Combining them determines rainfall occurrences and amount (Yang et al., 2015). Many schemes are needed to produce a set of rainfall predictions: for example, the Weather Research and Forecasting (WRF) model uses seven microphysics parameterization schemes and three cumulus parameterization schemes. More details on specific schemes can be found in Yang et al. (2015). Scheme selection is a main source of error within NWP models because certain schemes work better in certain circumstances. Some physical processes in rain production occur at small scales or in specific climates and cannot be properly modeled over a larger resolution. Two frequently used numerical weather prediction models are described below.

5.1.1 WRF Model

The Weather Research and Forecasting (WRF) model, known previously as the fifth generation Mesoscale Model (MM5), is a numerical weather predictor used in climate forecasting models such as CMAQ[4], NRCM[5], and NCEP Eta[6]. The Kain-Fritsch Scheme is a cumulus scheme in WRF that models precipitation based on condensation exceeding a threshold value (Yang et al., 2015). WRF uses past observation data or idealized atmospheric conditions and thresholds in schemes to generate rainfall. The daily precipitation output has a fine resolution of 0.03-degree grid that can be scaled to fit the modeler's area of interest. The source code for WRF can be found on UCAR's website.

5.1.2 ECHAM

The European Centre Hamburg Model (ECHAM) is a numerical weather prediction method used as the atmospheric model in climate models such as MPI-ESM[7] and ECHO-G[8]. ECHAM6 is the latest version using parameterization schemes include mass flux in cumulus convection and cloud microphysics to determine daily precipitation. A detailed description of the model can be found in Stevens et al. (2013) and Giorgetta et al. (2013). ECHAM has multiple sets of spatial resolutions with the finest resolution of 0.7031x0.7031 degrees for experimental use. ECHAM source code is freely available to the public at the Max Planck Institute for Meteorology website.

5.2 Stochastic

Stochastic models, among the simplest prediction models, use statistics and probabilities associated with weather data to predict atmospheric parameters (Harmel et al., 2002). Model output generates data that is statistically consistent with the observed data input. These models generate daily weather at a single point location or through a more complicated process of multi-site generation (Mehrotra et al., 2006). To generate precipitation, a Markov Chain Model determines the probability of having a wet day or a dry day, then finds the probability of a wet day following a dry or wet day (Wilks & Wilby, 1999). Historical precipitation measurements of 20 years or more is recommended to initiate the Markov chain; then, an equation using mean daily rainfall, standard deviation of daily rainfall, and a skew coefficient gives the amount of rainfall on a particular wet day. Stochastic models often fail to accurately describe the length of dry or wet periods and model output can be skewed based on historical input data, thus requiring statistical verification.

[4] Community Multiscale Air Quality Modeling System https://www.epa.gov/cmaq
[5] Nested Regional Climate Model https://rda.ucar.edu/datasets/ds601.0/
[6] National Center for Environmental Prediction Eta model https://rda.ucar.edu/datasets/ds609.2/
[7] Max-Planck-Institute Earth Systems Model (Stevens et al., 2013) https://www.mpimet.mpg.de/en/science/models/mpi-esm/
[8] ECHAM4 Atmospheric model coupled with HOPE-G oceanic model ("Lawrence Livermore National Laboratory Program for Climate Model Diagnosis and Intercomparison," 2005)

5.2.1 Weather Generators (WGEN)

Weather Generators are used for statistically simulating atmospheric conditions in many models. The Water Erosion Precipitation Project (WEPP) and the Soil and Water Assessment Tool (SWAT) both use stochastic weather generators. SWAT's default is Cooperative Observer Network (COOP) gauge data from 1960 to 2010, for historical reference, to begin the Markov Chain Model (Tuo et al., 2016). Weather generators create precipitation accumulation for the specified day, month, and year. SWAT can predict precipitation accumulation over a delineated area of interest, out to year 2100. The WGEN algorithm and Fortran code can be found in Richardson (1984); variations of this stochastic model have been created with the same internal structure as WGEN-- for instance, WXGEN[9], CLIGEN[10], and GEM[11]. Comparing results of different weather generators produces different predictions for each weather simulator due to their stochasticity (Migliaccio & Srivastava, 2007).

5.3 Nonparametric Models

Nonparametric models resample historic data to find trends and weather characteristics for future data. They can be thought of as smoothed, conditional bootstrapping or kernel density estimates (Rajagopalan et al., 1997). Nonparametric simulations use large numbers of observational data to create a probability density function that best describes the data (Sharma, 2000). A Gaussian kernel function is commonly used to describe weather patterns (Rajagopalan et al., 1997; Sharma, 2000). Nonparametric models produce only values that occurred from the historical dataset, but data may be regenerated that violates a boundary condition, such as the rain versus snow threshold temperature (Rajagopalan et al., 1997). This can cause error in the model's product, so careful consideration of input data is important. It is assumed (but cannot be guaranteed) that models which accurately predict historic weather patterns are more likely to accurately predict future weather patterns (Rupp et al., 2013).

5.3.1 GCM

Global Circulation Models (GCM) use a combination of nonparametric trends and numeric predictions to generate precipitation on a global scale. GCMs are good representations of temporal trends on a large scale, but often vary when downscaled to a regional level. The Coupled Model Intercomparison Project Phase 5 (CMIP5) was designed to evaluate how realistic 20+ GCMs are at recreating past climate data, projecting future climate change, and understanding differences among models (Taylor, 2009). Each model is given inputs of historical climate data from 1800- 2005 and future emission scenarios to

[9] Erosion/Productivity Impact Calculator Weather Generator (Wallis & Griffiths, 1995)
[10] USDA's Climate Generator (Meyer, 2004)
[11] Generation of weather Elements for Multiple applications (Harmel et al., 2002)

simulate near-term (to year 2035) or long-term (to year 2100) weather conditions. The Intergovernmental Panel on Climate Change (IPCC) provided CMIP5 with four future atmospheric scenarios represented by radiative forcing values in year 2100 called Representative Concentration Pathways (RCP) (Taylor, 2009). Each GCM used these scenarios to predict weather conditions:

RCP2.6 assumes that greenhouse gas emissions will peak around 2030, then decline;
RCP4.5 assumes the peak will be around 2050, then level off at 4.5W/m^2;
RCP6.0 assumes the peak will be around 2090, then level off at 6W/m^2;
RCP8.5 estimates emission will continuously rise throughout the 21st century;

This set of scenarios is the newest decision from IPCC on future climate scenarios. RCP 4.5, 6, and 8.5 are comparable to Special Report on Emission Scenarios (SRES) B1, B2, A1F1, respectively. Simulated precipitation data can be retrieved from a specific model or as a multi-model mean. An assessment of CMIP5 models based on observational data for reliability can be found in Rupp et al. (2013). The CNRM-CM5 model within CMIP5 performed with the least error in reproducing global precipitation, followed by CESM1-CAM5 (Rupp et al., 2013). Please note the experimental design for CMIP6 has only been recently published, and data is available but is still being evaluated (Eyring et al., 2016). CMIP5 data can be downloaded at the WorldClim website.

6. Discussion

Precipitation is a difficult variable to measure precisely. In calibrating the SWAT model for river basin modeling, Tuo et al. (2016) found that precipitation is the main source of uncertainty. Observed and simulated precipitation datasets have strengths and weaknesses in providing an accurate representation of rainfall amounts. Simulated datasets from numerical weather prediction, stochastic models, and nonparametric models provide a seamless time series and perform well in cold weather, mountainous regions, and non-extreme conditions. Simulated future data are applicable for managing and planning purposes since there is a need for information about changes in future precipitation. Observed datasets often include gaps in the time series due to lack of measurement, but they perform best in warm conditions and reflect extreme weather events well. Direct rainfall measurement from rain gauges is preferred by researchers since assumptions are not made and they have long measurement records. Many studies comparing differences in precipitation datasets for regional analysis have been performed, (e.g., Costa & Foley, 1998; Fekete et al., 2004; Tapiador et al., 2012).

Precipitation plays a large role in the availability of drinking water, erosion, and transportation of contaminants. Selection of precipitation data has crucial effects on hydrological model performance; thus, choosing precipitation datasets based on method, time step, and resolution needs to be carefully thought out (Tuo et al., 2016). Regulatory, planning, and exploratory purposes require different levels of uncertainty. Regulatory projects must have very little error and uncertainty while exploratory projects encourage uncertainty. A need for advancements in precipitation accuracy, length of record, and free availability is still a recurring problem in the modeling community. There is no “best” precipitation dataset, only the most appropriate for a given purpose. As Harmel et al. (2002) said, *"Historical data provide only one realization or 'picture' of a previous weather pattern that may not represent future climate scenarios"*.

References

Awange, J. L., Ferreira, V. G., Forootan, E., Khandu, Andam-Akorful, S. A., Agutu, N. O., & He, X. F. (2016). Uncertainties in remotely sensed precipitation data over Africa. *International Journal of Climatology, 36*(1), 303-323. doi: 10.1002/joc.4346

Becker. (2013). A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901-present.

Bitew, M. M., & Gebremichael, M. (2011). Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model. *Water Resources Research, 47*(6). doi: 10.1029/2010wr009917

Brewster, R. (2017). Paint.net 4.0.17 (Version 4.0.17). Retrieved from https://www.getpaint.net/

Chahine, M. T. (1992). The Hydrological Cycle and Its Influence on Climate. *Nature, 359*(6394), 373-380. doi: Doi 10.1038/359373a0

CHRS. (2004). Center for Hydrometeorology & Remote Sensing. *Satellite Precipitation-Activities* Retrieved June 4, 2017, from http://chrs.web.uci.edu/research/satellite_precipitation/activities00.html

Costa, M. H., & Foley, J. A. (1998). A comparison of precipitation datasets for the Amazon basin. *Geophysical Research Letters, 25*(2), 155-158. doi: Doi 10.1029/97gl03502

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., . . . Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology, 28*(15), 2031-2064. doi: 10.1002/joc.1688

Duan, Z., Liu, J., Tuo, Y., Chiogna, G., & Disse, M. (2016). Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales. *Science of the Total Environment, 573*, 1536-1553. doi: 10.1016/j.scitotenv.2016.08.213

Ebert, E. J., John; Kidd, Chris. (2007). Comparison of Near-Real-Time Precipitation Estimates from Satellite Observations and Numerical Models. *American Meteorological Society*. doi: 10.1175/BAMS-88-1-47

ESRI. (2015). ArcGIS Release 10.3.1 Redlands, CA: Environmental Systems Research Institute.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development, 9*(5), 1937-1958. doi: 10.5194/gmd-9-1937-2016

Fekete, B. M., Vorosmarty, C. J., Roads, J. O., & Willmott, C. J. (2004). Uncertainties in precipitation and their impacts on runoff estimates. *Journal of Climate, 17*(2), 294-304. doi: Doi 10.1175/1520-0442(2004)017<0294:Uipati>2.0.Co;2

Gao, J. G., Sheshukov, A. Y., Yen, H., & White, M. J. (2017). Impacts of alternative climate information on hydrologic processes with SWAT: A comparison of NCDC, PRISM and NEXRAD datasets. *CATENA, 156*, 353-364. doi: 10.1016/j.catena.2017.04.010

Giorgetta, M., Roeckner, E., Mauritsen, T., Bader, J., Crueger, T., Esch, M., . . . Stevens, B. (2013). The atmospheric general circulation model ECHAM6: Model Description. *Reports on Earth System Science*(135).

Golden, H. E., Knightes, C. D., Conrads, P. A., Feaster, T. D., Davis, G. M., Benedict, S. T., & Bradley, P. M. (2013). Climate change and watershed mercury export: a multiple projection and model analysis. *Environ Toxicol Chem, 32*(9), 2165-2174. doi: 10.1002/etc.2284

Golden, H. E., Knightes, C. D., Cooter, E. J., Dennis, R. L., Gilliam, R. C., & Foley, K. M. (2010). Linking air quality and watershed models for environmental assessments: Analysis of the effects of model-specific precipitation estimates on calculated water flux. *Environmental Modelling & Software, 25*(12), 1722-1737. doi: 10.1016/j.envsoft.2010.04.015

Harmel, R. D., Johnson, G., & Richardson, C. W. (2002). The GEM experience: Weather generator technology development in the USDA. *Bulletin of the American Meteorological Society, 83*(7), 954-957. doi: Doi 10.1175/1520-0477(2002)083<0954:Tgewgt>2.3.Co;2

Harmel, R. D., Smith, P. K., Migliaccio, K. W., Chaubey, I., Douglas-Mankin, K. R., Benham, B., . . . Robson, B. J. (2014). Evaluating, interpreting, and communicating performance of hydrologic/water quality models considering intended use: A review and recommendations. *Environmental Modelling & Software, 57*, 40-51. doi: 10.1016/j.envsoft.2014.02.013

Hu, Q., Yang, D., Li, Z., Mishra, A. K., Wang, Y., & Yang, H. (2014). Multi-scale evaluation of six high-resolution satellite monthly rainfall estimates over a humid region in China with dense rain gauges. *International Journal of Remote Sensing, 35*(4), 1272-1294. doi: 10.1080/01431161.2013.876118

Huffman, G. J., Adler, R. F., Rudolf, B., Schneider, U., & Keehn, P. R. (1995). Global Precipitation Estimates Based on a Technique for Combining Satellite-Based Estimates, Rain-Gauge Analysis, and Nwp Model Precipitation Information. *Journal of Climate, 8*(5), 1284-1295. doi: Doi 10.1175/1520-0442(1995)008<1284:Gpeboa>2.0.Co;2

Kidd, C., & Huffman, G. (2011). Review Global Precipitation Measurement. *Meteorological Applications, 18*(3), 334-353. doi: 10.1002/met.284

Kim, K. P., Katie; Whelan, Gene; Wolfe, Kurt; Duda, Paul; Gray, Mark; Pachepsky, Yakov. (2014). Using Remote Sensing and Radar MET Data to Support Watershed Assessment Comprising IEM.

Lawrence Livermore National Laboratory Program for Climate Model Diagnosis and Intercomparison. (2005). *Model Information of Potential Use to the IPCC Lead Authors and AR4: ECHO-G*. Retrieved June 6, 2017, from http://www-pcmdi.llnl.gov/ipcc/model_documentation/ECHO-G.htm

Lee, S., Ni-Mesister, W., Toll, D., Nigro, J., Gutierrez-Magness, A. L., & Engman, T. (2010). Assessing the hydrologic performance of the EPA's nonpoint source water quality assessment decision support tool using North American Land Data Assimilation System (NLDAS) products. *Journal of Hydrology, 387*(3), 212-220. doi: https://doi.org/10.1016/j.jhydrol.2010.04.009

Liu, Z., Kingery, W. L., Huddleston, D. H., Hossain, F., Chen, W., Hashim, N. B., & Kieffer, J. M. (2008). Modeling nutrient dynamics under critical flow conditions in three tributaries of St. Louis Bay. *J Environ Sci Health A Tox Hazard Subst Environ Eng, 43*(6), 633-645. doi: 10.1080/10934520801893725

Mehrotra, R., Srikanthan, R., & Sharma, A. (2006). A comparison of three stochastic multi-site precipitation occurrence generators. *Journal of Hydrology, 331*(1-2), 280-292. doi: 10.1016/j.jhydrol.2006.05.016
Meyer, C. (2004). *General Description of the CLIGEN Model and its History*. West Lafayette, IN: Retrieved from https://www.ars.usda.gov/ARSUserFiles/50201000/WEPP/cligen/CLIGENDescription.pdf.
Migliaccio, K. W., & Srivastava, P. (2007). Hydrologic components of watershed-scale models. *Transactions of the Asabe, 50*(5), 1695-1703.
Nelson, B. R., Seo, D. J., & Kim, D. (2010). Multisensor Precipitation Reanalysis. *Journal of Hydrometeorology, 11*(3), 666-682. doi: 10.1175/2010JHM1210.1
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
Price, K., Purucker, S. T., Kraemer, S. R., Babendreier, J. E., & Knightes, C. D. (2014). Comparison of radar and gauge precipitation data in watershed models across varying spatial and temporal scales. *Hydrological Processes, 28*(9), 3505-3520. doi: 10.1002/hyp.9890
Radcliffe, D. E., & Mukundan, R. (2017). PRISMvs. CFSR Precipitation Data Effects on Calibration and Validation of SWAT Models. *JAWRA Journal of the American Water Resources Association, 53*(1), 89-100. doi: 10.1111/1752-1688.12484
Rajagopalan, B., Lall, U., Tarboton, D. G., & Bowles, D. S. (1997). Multivariate nonparametric resampling scheme for generation of daily weather variables. *Stochastic Hydrology and Hydraulics, 11*(1), 65-93. doi: Doi 10.1007/Bf02428426
Richardson, C. W. W., D. A. (1984). WGEN: A Model for Generating Daily Weather Variables.
Rupp, D. E., Abatzoglou, J. T., Hegewisch, K. C., & Mote, P. W. (2013). Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. *Journal of Geophysical Research: Atmospheres, 118*(19), 10,884-810,906. doi: 10.1002/jgrd.50843
Sharma, A. (2000). Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3 - A nonparametric probabilistic forecast model. *Journal of Hydrology, 239*(1-4), 249-258. doi: Doi 10.1016/S0022-1694(00)00348-6
Shedekar, V. S., King, K. W., Fausey, N. R., Soboyejo, A. B. O., Harmel, R. D., & Brown, L. C. (2016). Assessment of measurement errors and dynamic calibration methods for three different tipping bucket rain gauges. *Atmospheric Research, 178*, 445-458. doi: 10.1016/j.atmosres.2016.04.016
Skofronick-Jackson, G. (2017). Precipitation Measurement Missions. Retrieved 6/30/2017, from https://pmm.nasa.gov/precipitation-measurement-missions
Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., . . . Roeckner, E. (2013). Atmospheric component of the MPI-M Earth System Model: ECHAM6. *Journal of Advances in Modeling Earth Systems, 5*(2), 146-172. doi: 10.1002/jame.20015
Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A. T., . . . de Castro, M. (2012). Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research, 104-105*, 70-97. doi: 10.1016/j.atmosres.2011.10.021
Taylor, K. S., Ronald; Meehl, Gerald. (2009). A Summary of the CMIP5 Experiment Design. 1-33.

Thornton, P. E., M.M. Thornton, B.W. Mayer, Y. Wei, R. Devarakonda, R.S. Vose, and R.B. Cook. (2017, 2017-04-19). Daymet: Daily Surface Weather and Climatological Summaries. *Daymet V3.* from https://daymet.ornl.gov/

Tramblay, Y., Thiemig, V., Dezetter, A., & Hanich, L. (2016). Evaluation of satellite-based rainfall products for hydrological modelling in Morocco. *Hydrological Sciences Journal, 61*(14), 2509-2519. doi: 10.1080/02626667.2016.1154149

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). *Science of The Total Environment, 573*, 66-82. doi: 10.1016/j.scitotenv.2016.08.034

U.S. Geological Survey (USGS). 2011. National Elevation Dataset (NED). Sioux Falls, SD. In: U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS). 2012. National Hydrography Dataset Plus – NHDPlus, V2.10. Available at: https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus#v2datamap

Wallis, T. W. R., & Griffiths, J. F. (1995). An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator (EPIC). *Agricultural and Forest Meteorology, 73*(1), 115-133. doi: http://dx.doi.org/10.1016/0168-1923(94)02172-G

Wilks, D. S., & Wilby, R. L. (1999). The weather generation game: a review of stochastic weather models. *Progress in Physical Geography, 23*(3), 329-357. doi: Doi 10.1177/030913339902300302

Yang, M.-x., Jiang, Y.-z., Lu, X., Zhao, H.-l., Ye, Y.-t., & Tian, Y. (2015). A weather research and forecasting model evaluation for simulating heavy precipitation over the downstream area of the Yalong River Basin. *Journal of Zhejiang University SCIENCE A, 16*(1), 18-37. doi: 10.1631/jzus.A1400347

EPA
United States
Environmental Protection
Agency

Office of Research
and Development
(8101R)
Washington, DC
20460