A Practical Guide for Working with Weather Datasets, Topic #1:

The Main Types of Weather Datasets

Catastrophe and Climate

Patrick Wiese, ASA

October 2022

Overview

This is the first of a series of papers intended to serve as a practical guide for actuaries and researchers who wish to analyze weather datasets. Each paper will describe a particular weather dataset and will be accompanied by an open-source computer program for analyzing the dataset. The computer programs – developed by SOA staff and by volunteers – will reduce the upfront time and effort required to begin working with weather datasets. Because the programs' source code will be open and accessible, researchers will be able to modify and expand it to suit their own purposes.

Each of the weather datasets to be covered in this series of papers is available online and can be downloaded at no cost. For each dataset, the associated paper will provide the following information:

1. The contents of the weather dataset
2. Geographic area covered
3. Time period covered
4. Time increments of the data (hourly, daily, monthly, etc.)
5. Types of weather observations included
6. How to download the data
7. How to tabulate the data using the free computer program provided by the Society of Actuaries

This introductory paper provides an overview of the main types of weather datasets and lists examples of each type.

The second paper in this series will outline strategies for analyzing large weather datasets using an ordinary personal computer. Many weather datasets exceed 100 gigabytes, and some are much larger. A researcher may lack the hard-drive space to store a dataset of this size. Even if sufficient storage space is available, the data is likely to exceed the memory limits of the researcher's computer. However, with a clever approach, much can be accomplished using a personal computer, as will be illustrated in the second paper.

Each subsequent paper will describe a particular weather dataset and will be accompanied by a free, open-source computer program for analyzing the dataset. A wide range of datasets will be covered in this series of papers, including (1) data collected by weather stations; (2) data estimated using Doppler radar and/or sensors on satellites; (3) "reanalysis" datasets generated by weather models that assimilate historical data from many sources (land-based stations, ships, planes, weather balloons, buoys, satellites, and radar) and produce, as an output, spatially and temporally complete historical records; (4) short-term, sub-seasonal and seasonal forecasts and (5) long-range climate projections.

Station Data versus Gridded Data

Most weather datasets are comprised either of station data or gridded data. Therefore, it is helpful to begin our discussion with a high-level overview of these two broad types of weather data.

STATION DATA

Weather stations are facilities with instruments that measure temperature, precipitation, wind speed, air pressure and/or humidity. Most weather stations are located on land, but some are located on buoys or ships. A dataset comprised of weather observations collected by stations will also contain metadata for each station, such as the station's latitude, longitude, and elevation above sea-level.

The Global Historical Climatology Network daily[1] (GHCNd) dataset is a good example of station data. GHCNd consists of daily temperature, precipitation and wind speed observations collected from over 100,000 land-based weather stations. GHCNd provides geographic coverage of much of the world, but the availability of data varies considerably from one country to another and may also vary significantly within a country, as illustrated in the map below. In general, the density of weather stations is greatest in urban areas and lowest in rural areas.

Figure 1
GHCN WEATHER STATIONS IN THE USA AND SOUTHERN CANADA
(Each station is represented by a small red dot; only those stations that have provided data in 2022 are included)

---

[1] https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily

GRIDDED DATA

In contrast to station data, gridded data is spaced uniformly across land, oceans, and lakes[2]. The density of the data does not decline in rural or unpopulated areas. Typically, grid spacing is fixed in degrees, where 360 degrees represents the distance around the earth. However, some datasets use spacing that is fixed in miles or kilometers.

While station data is strictly historical, looking backwards in time, gridded data may be either historical or projected. Historical gridded datasets are generated either from algorithms that interpolate across station data, or from weather models that reconstruct the past in a manner consistent with available historical data, including data from stations, ships, planes, weather balloons, buoys, satellites, and radar. A historical dataset produced via a weather model is referred to as "reanalysis" data. Relative to station datasets which have uneven coverage across space and time, reanalysis data offers the advantage of spatial and temporal completeness.

The spacing between grid points varies from one gridded dataset to another. Finely gridded datasets use a spacing of between 0.10 and 0.25 degrees, while coarsely gridded datasets use spacing of two degrees or greater. Note that at the equator, 0.25 degrees is about 30 kilometers, while two degrees is about 200 kilometers.

Suppose a weather dataset spanning the entire earth uses a one-degree grid. For each line of latitude encircling the earth in an east-west direction, the dataset will have 360 data points, corresponding to each of the 360 degrees of a circle. In the north-south direction, each line of longitude will have 180 points, corresponding to 180 degrees between the two poles. In total, the dataset will have 64,800 grid points, equal to the product of 360 and 180. Each grid point is specified by a latitude and longitude – for example, 41 degrees north (of the equator) and 74 degrees west (of the prime meridian[3]), which is a location about 30 kilometers north of New York City.

A data grid in fixed degree intervals is not uniformly spaced when distance is measured in miles or kilometers. Because the equatorial circumference is about 40,000 kilometers, the spacing between equatorial grid points would be about 110 kilometers (40,000 divided by 360) for a dataset with a one-degree grid. As one moves away from the equator towards either the north or south pole, the earth's circumference declines when measured in an east-west direction. For example, at 60 degrees north of the equator, the east-west circumference of the earth is 50% less than its equatorial circumference; consequently, the spacing between grid points is 50% lower and the data density (measured in grid points per square mile or square kilometer) is correspondingly greater. This is an important fact to keep in mind when working with data gridded in degrees[4].

In addition to latitude and longitude, some gridded weather datasets have "pressure level" as a third dimension. For example, ERA5[5] includes data for 37 pressure levels ranging from 1 to 1000 hPA, where "hPA" stands for "hectopascal pressure unit". One can envision these 37 pressure levels as locations in a vertical column above each latitude/longitude grid point. Thus, a gridded dataset with latitude, longitude and pressure level provides weather data across a three-dimensional model of space. Temperature data, for example, is provided not only for locations just above the surface of the earth, but also extending upwards at various heights (or pressure levels) in the atmosphere. Roughly speaking, 1000 hPA corresponds to areas near the earth's surface. Pressure falls as one moves

---

[2] Some gridded datasets provide coverage over both land and oceans; some provide coverage solely over land, and some provide coverage solely over the oceans.

[3] The "prime meridian" is the line of longitude defined to be zero degrees. This line runs in a north/south direction, and passes through Greenwich, England, which is a borough of London. East-west position on the earth's surface is measured relative to the prime meridian.

[4] If one were to compute average temperature or rainfall across a large geographic region, without accounting for differences in the density of points, the resulting average would be over-weighted for those areas farthest from the equator. A simple adjustment is to calculate a weighted (rather than unweighted) average using cosine of latitude for the weight of each grid point.

[5] "ERA5" is short for "Fifth generation European Center for Medium-Range Weather Forecasts Atmospheric Reanalysis of the Global Climate". A detailed description of this dataset is available here: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5

upwards through the atmosphere. At 5000 meters above the earth's surface, pressure is about 500 hPA, and at 10000 meters, pressure is about 250 hPA.

Geographically gridded datasets also have a time dimension (but this is not considered to be part of the grid). For example, ERA5 data is available on an hourly basis such that each grid point has 24 observations per day. The finer the geographic grid and the shorter the time steps, the larger will be the dataset. Consider an hourly dataset that covers the full surface of the earth using a one-degree grid. As explained earlier, such a dataset will have 64,800 geographic points. Each of these points, in turn, will have 24 daily observations for each weather variable (temperature, precipitation, etc.). Thus, each day of data will include over 1.5 million observations per weather variable, and a year's worth of data will include over half a billion observations per weather variable.

The Main Types of Weather Datasets

Classifying weather datasets into station data versus gridded data is useful, but a more detailed classification scheme is needed to provide an adequate description of the key types of weather data. One possible classification approach is as follows:

MAIN CATEGORIES OF WEATHER DATASETS

1. Historical data
    a. Direct observations collected by weather stations, buoys, weather balloons, ships, and planes
    b. Gridded data interpolated from station data
    c. Gridded estimates produced from data collected by satellites and radar
    d. Gridded reanalysis data generated via weather models that "assimilate" historical data
    e. Severe weather event data (tropical cyclones, severe storms, tornadoes, hail, etc.)

2. Forecasts (all of which are gridded)
    a. Short to medium range (up to 15 days forward in time)
    b. Sub-seasonal to seasonal (15 to 200 days forward in time)
    c. Climate projections (typically, about 100 years forward in time)

Using this approach, datasets are categorized into eight types. The first five of these types are historical data and the remaining three are forecasts. Below, we describe these eight types and offer an example of each.

1A. DIRECT HISTORICAL OBSERVATIONS

"Direct" weather observations are gathered by instruments that are in contact with the medium that is being measured. A standard thermometer, for example, is surrounded by the air or water that it is measuring. A weather observation obtained via direct measurement is referred to as an "in-situ" observation. Land-based weather stations are the most common source of in-situ observations. Additional sources include ships, planes, buoys, and balloons outfitted with weather measurement devices. The GHCNd dataset (described earlier in this paper) is a good example of in-situ data gathered via land-based weather stations.

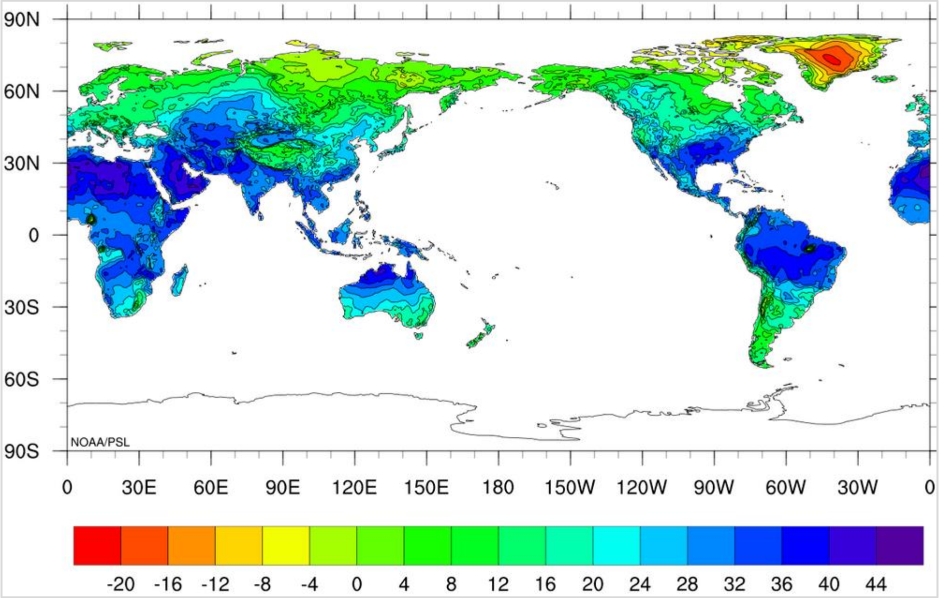
In-situ data is not gridded. Rather, the spatial distribution of the data is determined by the location of the associated weather stations. Because weather stations are not uniformly spaced across land (or sea), the data captured by these stations will reflect the same uneven geographic distribution. In addition, in-situ data is unevenly distributed across time because there were fewer weather stations operating in prior decades compared to recent years.

1B. GRIDDED DATA INTERPOLATED FROM STATION DATA

To address the irregular spacing of ground-based weather stations, researchers have developed algorithms to translate station data into gridded data. For example, the Climate Prediction Center – a federal agency that is part of the National Oceanic and Atmospheric Administration – uses station data to produce the CPC Global Unified Temperature Dataset[6]. This dataset is gridded at 0.5 degrees and covers the period from 1979 to the present. For each day and for each grid point, the data provides daily minimum and maximum temperature. Areas over oceans are not included; rather, the dataset only covers land areas.

Figure 2

HIGH TEMPERATURE (°C) ON 21 SEPTEMBER 2022, FROM THE CPC GLOBAL TEMPERATURE DATASET



Source: CPC Global Unified Temperature data provided by the NOAA PSL, Boulder, Colorado, USA

Gridded data interpolated from station data has several applications: it is helpful for the analysis of weather and climate for areas that are not located close to a weather station; it is useful for monitoring regional climate change; and it can be used to validate climate models which produce gridded data as an output.

1C. ESTIMATES PRODUCED FROM DATA COLLECTED BY SATELLITES AND RADAR

In-situ data is of critical importance for analyzing weather, but weather stations cannot be placed everywhere. For urban areas of wealthy countries, in-situ data is usually plentiful. But in rural or remote areas, as well as in low-income countries, the density of weather stations is typically quite low.

Fortunately, satellites and Doppler radar provide a means to collect weather data across most of the earth. The observations are not direct measurements, but rather indirect measurements that are inferred via remote sensing. In general, indirect estimates of surface temperature and precipitation are not as reliable as direct estimates. However, by comparing indirect estimates with direct observations, scientists have been able to gradually improve the algorithms that translate satellite and radar data into weather estimates.

Weather satellites measure the amount of radiation emitted from the earth's surface, clouds, and atmosphere. This information is digitized and transmitted to a ground-based receiving station where it is translated into estimates of temperature, precipitation, wind, sea-ice coverage, sea-surface height, and the height of ocean waves. Although

---

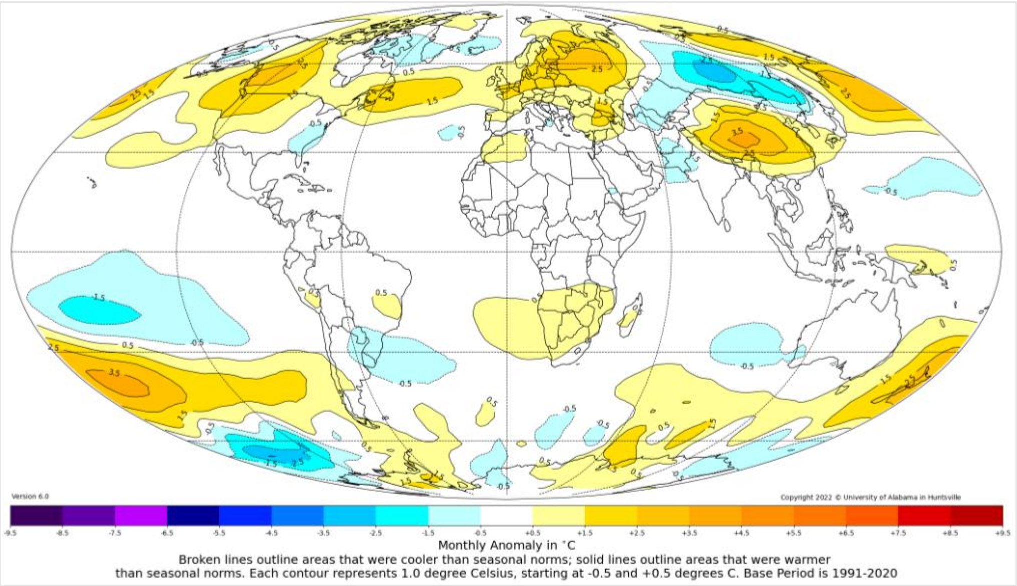
[6] https://psl.noaa.gov/data/gridded/data.cpc.globaltemp.html

satellites have been used to estimate weather data since 1960, significant technical improvements[7] occurred in the late 1970s. For this reason, most satellite-based weather datasets have a starting year of 1978 or later.

The University of Alabama in Huntsville (UAH) temperature dataset[8] is an example of data inferred from satellite-based measurements. The UAH data is gridded at 2.5 degrees, covers over 97% of the earth's surface, and extends from 1978 to the present in monthly time steps. The primary use of the dataset is to analyze long-term trends in global warming.

Figure 3

TEMPERATURE ANOMALIES FOR AUGUST 2022, COMPUTED USING THE UAH TEMPERATURE DATASET



Broken lines outline areas that were cooler than seasonal norms; solid lines outline areas that were warmer than seasonal norms. Each contour represents 1.0 degree Celsius, starting at -0.5 and +0.5 degrees C. Base Period is 1991-2020

Source: Global Temperature Report for August 2022, produced by the University of Alabama in Huntsville

GPM IMERG[9] (Integrated Multi-Satellite Retrievals for the Global Precipitation Mission) is an example of precipitation data generated from satellite-based measurements. GPM IMERG data runs from 1997 to the present and has a temporal resolution of 30 minutes and a spatial resolution of 0.1 x 0.1 degrees (roughly, 10km x 10km). As illustrated in Figure 4, the data provides worldwide coverage.

While satellites peer down into the atmosphere from high above the earth, ground-based Doppler-radar weather stations perform their scans laterally (parallel to the earth's surface) and upwards, using pulses of radio waves to collect the data needed to generate precipitation estimates, including both the type of precipitation (snow, sleet, rain, or hail), the rate of precipitation, and the direction a storm front is moving. Compared to satellites, ground-based radar has several advantages such as superior ability to detect what is happening beneath clouds, and the ability to perform 3D scans storms. However, a ground-based radar station has a maximum scanning radius of only 240 kilometers (150 miles), while a satellite can scan vast areas.

To scan large regions using ground-based radar, a network of radar stations is required. In the United States, the National Weather Service operates a network known as "NEXRAD" consisting of 160 high-resolution Doppler radar stations that provide coverage for most of the US. NEXRAD tracks precipitation in real-time and facilitates the

---

[7] A new generation of polar-orbiting satellites was launched in 1978 (the Advanced TIROS-N series), and this greatly improved the available data.

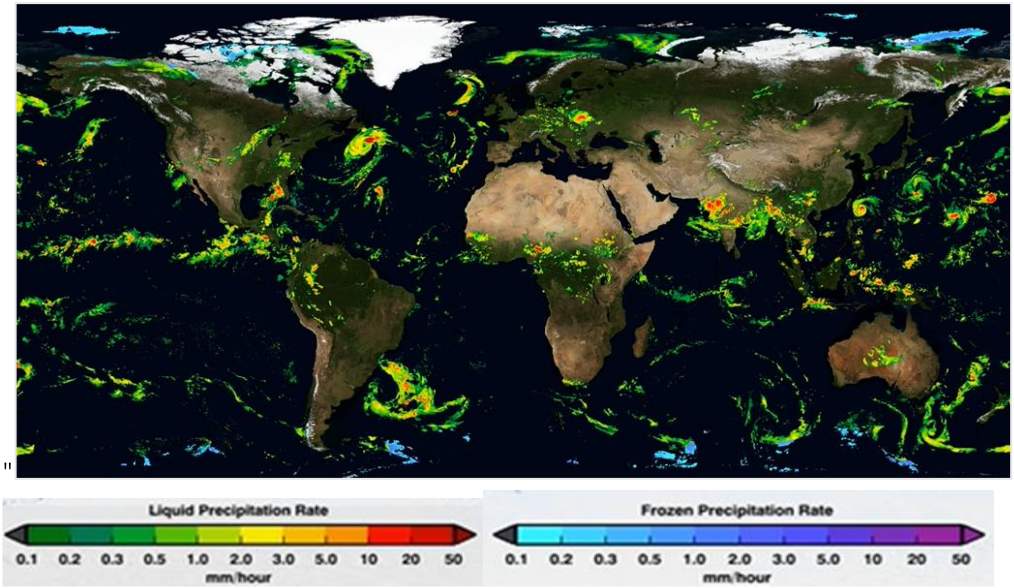
[8] The data is available at https://www.ncei.noaa.gov/products/climate-data-records/mean-layer-temperature-uah, and a report summarizing the data is available at https://www.nsstc.uah.edu/climate/index.html.f

[9] https://gpm.nasa.gov/data/imerg

development of forecasts. Information collected by NEXRAD is used to produce three "levels" of data: "Level 1" consists of the raw data collected by the radar stations, not yet transformed into weather estimates; "Level 2" includes raw data with a small amount of processing; and "Level 3" is data that has been transformed into estimates of the type of precipitation, precipitation rates and totals, and radial wind velocity[10].

Figure 4

PRECIPITATION RATES AT 1 PM UTC[11] ON SEPTEMBER 10, 2022, EXTRACTED FROM THE GPM IMERG DATASET



Source: half-hour precipitation image from GPM IMERG (https://gpm.nasa.gov/data/imerg#latesthalf-hourlyimage)

1D. GRIDDED REANALYSIS DATA

Like datasets interpolated from station data (refer to section 1B), reanalysis datasets are gridded on a geographic lattice. Unlike interpolated datasets, reanalysis datasets use weather models to reconstruct the past in a manner consistent with available historical data, and consistent with the physical laws governing the earth's atmosphere and oceans. The historical data fed into the model(s) can come from many sources, including weather stations, ships, planes, weather balloons, satellites, and radar.

In contrast to station data which has an uneven distribution across space and time, reanalysis data is spatially and temporally complete. Reanalysis data is particularly helpful when analyzing historical trends in a region that has a low density of weather stations. Keep in mind, however, that reanalysis involves weather models (in addition to direct weather observations), and no model is perfectly reliable.

A good example of a reanalysis dataset is ERA5[12], produced by the European Centre for Medium-Range Weather Forecasts. The dataset extends from 1950 to the present in hourly time steps and provides coverage of the entire earth using a 0.25-degree grid. The data is updated daily, with a latency of only five days, and provides a comprehensive set of weather metrics including air temperature, air pressure, sea surface temperature, wind speed, total precipitation, and various measures of soil moisture.

---

[10] Radial velocity is the component of wind velocity parallel to the direction of a radar beam.

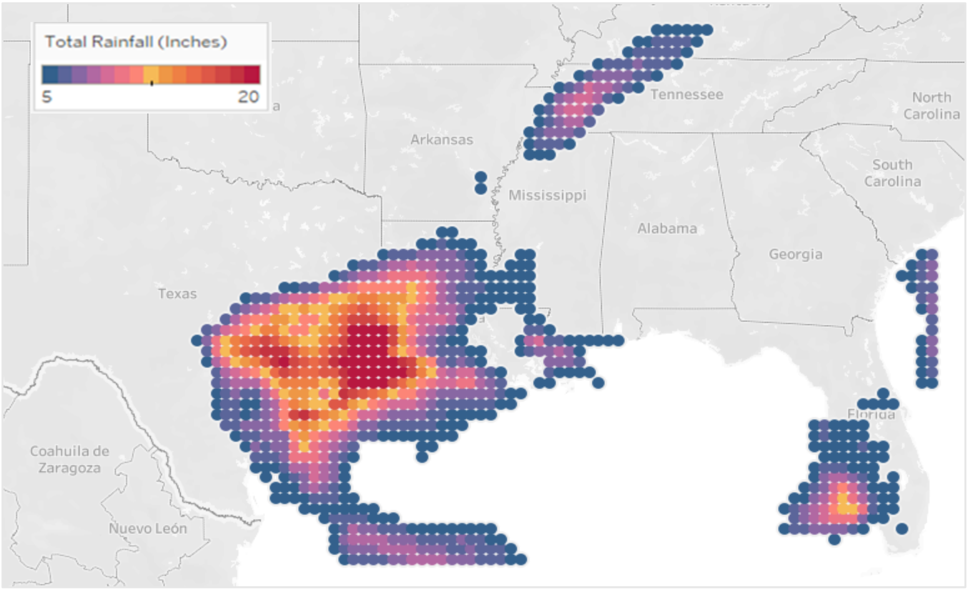
[11] "UTC" stands for "Universal Time Coordinated". In years past, UTC was referred to as "Greenwich Mean Time".

[12] https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5

Figure 5

TOTAL RAINFALL DURING HURRICANE HARVEY, TABULATED FROM THE ERA5 DATASET

(Only those ERA5 grid points with greater than 5 inches of rainfall are shown on this map)



Data source: ERA5 single-level hourly precipitation data, summed across the period from 8-23-2017 to 9-1-2017

1E. SEVERE WEATHER EVENT DATA

The datasets described in 1A, 1B, 1C, and 1D provide weather observations that run continuously across time, capturing both normal weather and severe weather when it occurs. In contrast, some datasets focus exclusively on severe weather events such as hurricanes, tornadoes, winter storms, and droughts. A few examples of databases that focus on severe weather are as follows:

- International Best Track Archive for Climate Stewardship[13] (IBTrACS), containing worldwide data on the paths and intensities of tropical cyclones that occurred between 1842 and the present.
- The Storm Events Database[14] compiled by the National Oceanic and Atmospheric Administration (NOAA), containing data for severe weather events in the United States having sufficient intensity to cause loss of life, injuries, significant property damage, and/or disruption to commerce. This database covers the period from 1950 to the present, and provides loss estimates in addition to data describing each severe weather event.
- The U.S. Drought Monitor[15], covering the period from 2000 to the present, with data summarizing drought conditions across the United States.
- NOAA's Snowstorm Database[16], containing data for the United States for the period from 1900 to the present. Only storms with large areas of heavy snowfall (10+ inches) are included.

---

[13] https://www.ncei.noaa.gov/products/international-best-track-archive

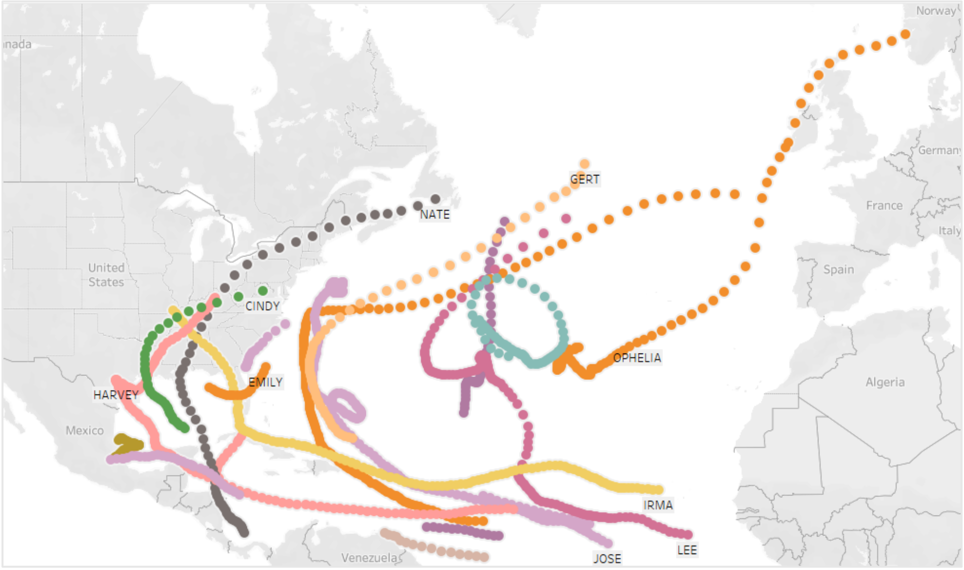
[14] https://www.ncdc.noaa.gov/stormevents/

[15] https://droughtmonitor.unl.edu/

[16] https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00464

Figure 6

NORTH ATLANTIC BASIN HURRICANE TRACKS IN 2017, EXTRACTED FROM IBTrACS DATABASE



2. WEATHER FORECASTS

The datasets described in 1A through 1E focus on weather that occurred in the past. Now, we shift gears and discuss datasets that describe forecasts of future weather. Roughly speaking, weather forecasts can be categorized into three types:

Table 1

MAIN TYPES OF WEATHER FORECASTS

Forecast Type	Projection Horizon	Spatial and Temporal Resolution	Objective
Short to Medium Range	Up to 15 days forward in time	High	To predict hourly and daily weather
Sub-seasonal to Seasonal	15 to 200 days forward in time	Medium	To predict if weekly or monthly weather metrics will be above/below normal levels
Climate Projections	Up to 300 years forward in time	Low	To predict changes in climate

Modern weather models have strong predictive power out to about 7 days forward in time. Beyond the 10th day, the ability to accurately predict daily weather drops significantly, but models have demonstrated some capacity to predict general sub-seasonal and seasonal weather patterns across large geographic regions – for example, to predict if the Northeastern United States will experience above or below average temperatures during the upcoming winter.

A climate projection, in contrast to a weather forecast, has a forecasting range measured in decades or centuries as opposed to days or months. A climate projection provides a sense of how the *distribution* of weather events may gradually change in the decades ahead, as opposed to reliably forecasting individual near-term weather events. In general, climate projections have a lower spatial resolution than do weather forecasts. Short to medium-range weather forecasts are typically gridded at between 0.1 and 0.5 degrees, while climate projections are typically on grids of 1.0 degree or greater.

2A. SHORT TO MEDIUM-RANGE WEATHER FORECASTS

A short-range forecast is typically defined as a projection up to 3 days forward in time, while a medium-range forecast runs up to 15 days forward in time. The period from day 11 through 15 is sometimes referred to as an "extended" forecast.

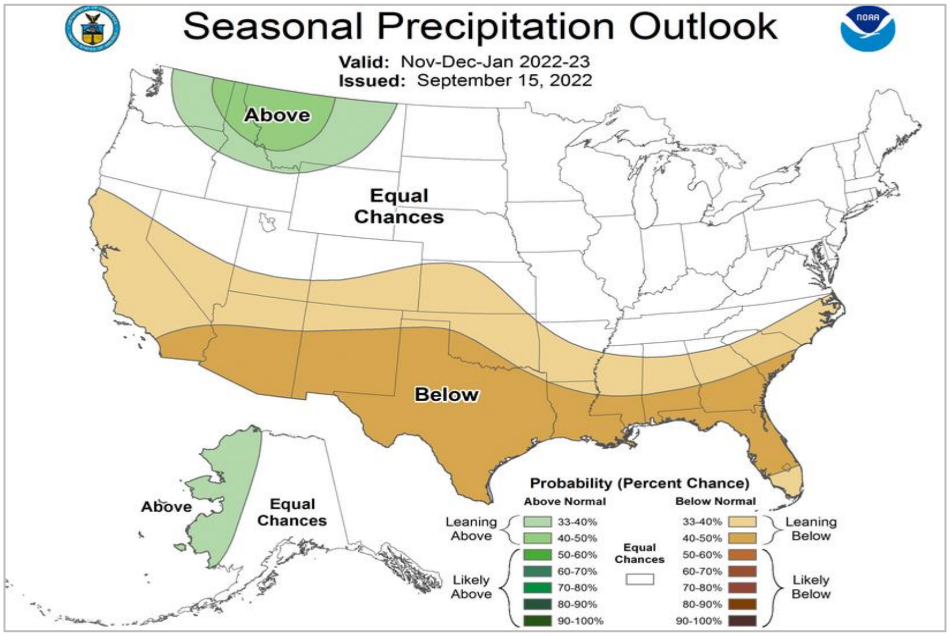
We are accustomed to seeing near-term forecasts on TV or the internet, but it is also possible to download a forecast dataset that covers either the entire planet or a particular user-specified region. For example, the European Center for Medium-Range Weather Forecasts (ECMWF) makes a set of 10-day forecasts[17] available for free. These forecasts are updated four times each day and provide worldwide coverage gridded at 0.4 degrees[18]. Time steps of 3 hours are used for the first 6 days of each forecast, and 6-hour time steps are used for days 7 through 10. A broad range of weather metrics is provided, including air temperature, soil temperature, wind speed, total precipitation, air pressure, and relative humidity.

2B. SUBSEASONAL AND SEASONAL WEATHER FORECASTS

Forecasts that extend from 15 to 45 days into the future are frequently referred to as "sub-seasonal", while forecasts that extend from 45 to 200 days into the future are often referred to as "seasonal". However, these forecasting periods are not strictly defined. For example, NOAA defines sub-seasonal as the period between two weeks through three months[19], while ECMWF defines sub-seasonal as the period from two weeks through one month[20].

Figure 7

SEASONAL PRECIPITATION FORECAST PRODUCED BY NOAA



Source: https://www.cpc.ncep.noaa.gov/products/predictions/long_range/seasonal.php?lead=2

Semantics aside, as models move beyond the 10th forecasting day, their ability to predict weather at a high spatial and temporal resolution declines significantly. However, sub-seasonal and seasonal forecasts offer predictive value

---

[17] ECMWF forecasts can be downloaded via this web page: https://confluence.ecmwf.int/display/DAC/ECMWF+open+data:+real-time+forecasts

[18] At the equator, 0.4 degrees is equivalent to about 44 kilometers; at a latitude of 60N or 60S, 0.4 degrees is equivalent to about 22 kilometers.

[19] https://www.weather.gov/media/sti/S2S/Annotated%20Outline%20for%20Public%20Release_7_12_18.pdf

[20] https://www.ecmwf.int/en/about/media-centre/science-blog/2022/progress-sub-seasonal-forecasting-weather-extremes

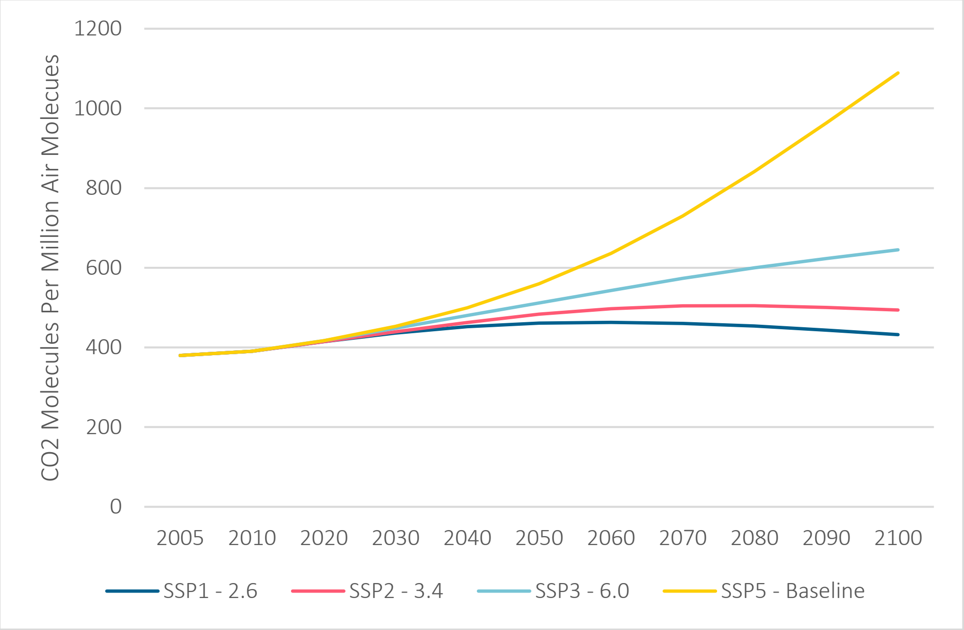
when results are aggregated across large time steps and large geographic regions, and when one's goal is simply to understand if weather conditions are likely to be above normal, normal, or below normal. For example, the three-month forecast[21] in Figure 7 indicates that total precipitation across November, December, and January of 2022 is likely to be below normal in the southern half of the U.S., above normal in a small portion of northwestern U.S., and an equal chance of above or below normal for the remainder of the country.

2C. CLIMATE PROJECTIONS

Climate change is driven by increases in greenhouse gas (GHG) concentrations in the atmosphere. It is unclear if atmospheric GHG concentrations will continue to escalate rapidly for many decades to come, or if society will successfully transition to a low-emission economy that will have a favorable impact on the trajectory of GHG concentrations. Given this uncertainty, climate projection models are outfitted with the ability to run a range of GHG scenarios. A GHG scenario or "pathway" is fed into a model as a set of exogenous inputs, and the model then projects the impact of this scenario on the climate.

Perhaps the most well-known set of GHG scenarios is produced by the Intergovernmental Panel on Climate Change (IPCC). Roughly every 5 years, the IPCC releases a report that presents updated climate projections. Each report includes updated GHG scenarios that are the basis for the climate projections. The latest IPCC report[22] contains 26 scenarios, 4 of which are presented in Figure 8:

Figure 8
FOUR SCENARIOS FOR FUTURE ATMOSPHERIC CO2 CONCENTRATIONS



Data source: the sixth IPCC assessment report

The IPCC's scenarios are referred to as "shared socioeconomic pathways" (SSPs). The SSPs[23] reflect assumptions about population growth, economic growth, the use of sustainable energy sources versus fossil fuels, as well as other factors which, collectively, lead to a prediction for the trajectory of GHG emissions.

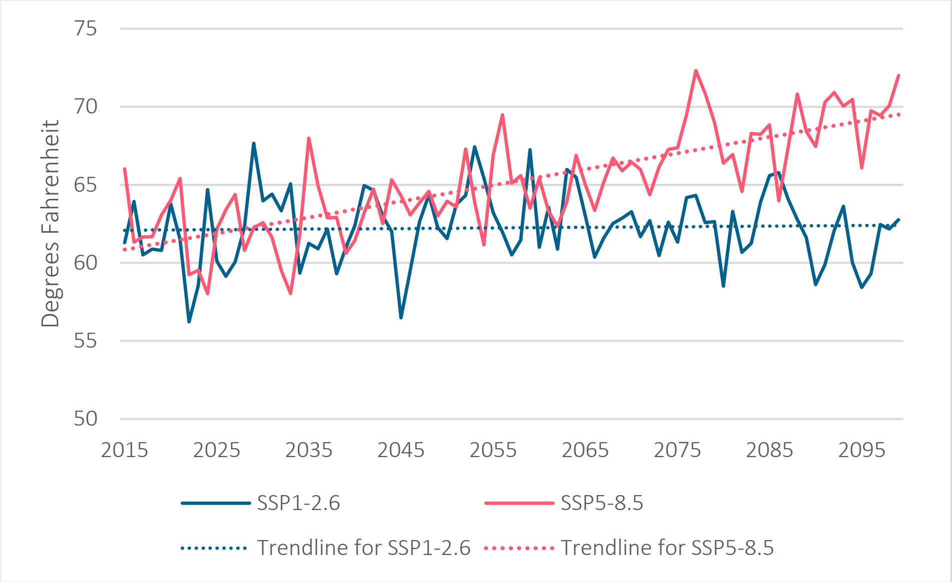
---

[21] https://www.cpc.ncep.noaa.gov/products/predictions/long_range/seasonal.php?lead=2
[22] https://www.ipcc.ch/assessment-report/ar6/
[23] This link provides a good description of the SSPs and provides a mechanism to download the underlying data: https://ourworldindata.org/explorers/ipcc-scenarios?facet=none

Because the modeling of climate change is complex, and because different models produce different results, the IPCC uses an "ensemble" of models to produce projections. For the latest IPCC report, the ensemble incudes over 100 models from more than 50 modeling centers. By running each SSP through each model in the ensemble, a range of forecasts is produced. The average results across all models can be viewed as a best estimate, while the distribution of results provides some sense of the level of uncertainty.

Figure 9 shows a temperature forecast produced by one member of the IPCC's ensemble of models. For this model, results for two GHG scenarios are presented. In practice, a complete analysis would require an examination of results for each of the 26 SSPs, and for each model in the ensemble.

Figure 9
PROJECTED AVERAGE ANNUAL TEMPERATURE (°F) FOR KANSAS CITY
UNDER TWO DIFFERENT GREENHOUSE GAS SCENARIOS



Source: extracted from r1i1p1 CMIP6 simulation results produced by the Alfred Wegener Institute Climate Model

A Preview of the Next Paper in this Series

Many weather datasets exceed 100 gigabytes, and some are much larger. While most climate scientists have access to servers that can store and process massive weather datasets, other types of researchers may wish to perform weather analyses on a standard personal computer. A personal computer rarely offers more than 1000 gigabytes of storage space and 16 to 32 gigabytes of RAM (active memory for running applications and programs). Given these constraints, a researcher will need a clever approach for working with large weather datasets. The next paper in this series will tackle this subject, offering illustrative examples of how to run analyses of a large weather dataset despite the storage and memory limitations imposed by a personal computer.

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org