

Glossary

This glossary is a non-exhaustive source for definitions of key terms and concepts related to the Analytics Engine and its associated tools and applications. This page includes detailed explanations of data products, use cases, analytical tools, and methodologies used in climate data analysis to support a user's understanding and application of the information in the Analytics Engine. This glossary is designed to complement existing extensive meteorological glossaries, such as the American Meteorological Society's [Glossary of Meteorology](#), with specific terminologies that are focused on the Analytics Engine.

8760

A representative hourly time series for a variable for one year (24 hours x 365 days in a year), used to characterize typical environmental conditions in many energy system modeling applications. 8760s are also referred to as hourly profiles.

Annual maximum series (AMS)

A time series constructed by extracting the maximum (or minimum) value for each year from a longer time series of observations or model outputs. The AMS is a foundational input to extreme value analysis and can be derived from data at any sub-annual time resolution, including hourly, daily, or monthly. The AMS is a specific case of the more general block maximum/minimum series, where the block size is set to one year. When working with climate model simulations, the AMS should be constructed on a simulation-by-simulation basis rather than aggregating across simulations prior to analysis.

Bias correction / adjustment

Various techniques applied to global climate model output to adjust for systematic errors (or biases). For example, a bias adjustment can be applied to model output which underestimates extreme precipitation events, or temperatures that are consistently too high. There are several sources of model biases that bias adjustment can adjust for, including spatial resolutions that are not fine enough to resolve all types of weather phenomena, or because model formulations rely on simplifications of physical processes.

Block maximum series

A block maximum series is a time series where extreme values are calculated from a time series of data by segmenting the time series into "blocks" of equal duration. The block maximum series is a more generalized approach to the Annual Maximum Series, where the block size used is one year.

Cascading event

A cascading event is a type of compound event that can occur when two or more climate or environmental hazards are temporally sequenced. Their combination can lead to exacerbated



impacts, as each hazard builds on the previous one.

CF conventions

[Metadata conventions](#) for formatting and documenting netCDF format data, to ensure standardized reporting and variable naming across different data sources. CF stands for “climate and forecast”. Data supported on the Analytics Engine adhere to CF conventions.

Change signal

The difference between two time periods, typically projected future climate conditions and a historical reference period. Also known as a “delta signal”. A change signal is used to quantify how much a climate variable (such as temperature or precipitation) is expected to change relative to past conditions.

ClimaKitAE

An open-source Python library that contains functionality to work with cloud-optimized gridded climate data. It includes basic tools for creating data objects and visualizations, as well as more advanced tools that support various applications of climate data including analyzing extreme events, understanding regional responses at different global warming levels, and describing uncertainty across different simulations. [ClimaKitAE](#) stands for “climate toolkit Analytics Engine”.

Climate sensitivity

Climate sensitivity is a measurement that describes the equilibrium change in global mean surface temperature resulting from a doubling of atmospheric carbon dioxide (CO₂) concentrations compared to pre-industrial levels (1850-1900).

Climatology

The long-term average or baseline conditions for a given climate variable (such as temperature or precipitation) over a specified historical period. These values are typically calculated from multi-year datasets and are used as a reference point to assess changes, anomalies, or trends in future climate projections. Best scientific practice is to use a 30-year period to ensure the reference period captures a range of variability.

CMIP6

The [Coupled Model Intercomparison Project, Phase 6](#) is a coordinated international effort to produce climate model output with consistent standards in areas such as variable naming and experimental design. This consistency allows different models to be compared to each other more easily. CMIP6 is the latest generation of global climate models (ca. 2020), used in the Intergovernmental Panel on Climate Change [Sixth Assessment Report](#) (IPCC AR6), and in California’s Fifth Climate Change Assessment.



Compound event

A type of climate event that arises due to the combination of two or more climate drivers or hazards occurring simultaneously or successively with negative implications for environmental and human systems. The weather phenomena contributing to the compound hazard may not be extreme events on their own, but combine to create negative impacts.

Cooling degree days (CDDs)

Cooling degree days (CDDs) measure how much and for how long outdoor temperatures exceed a specific threshold, indicating the need for indoor cooling. For each day, CDDs are calculated as the number of degrees the day's temperature is above the chosen base temperature; days that do not exceed the threshold are assigned zero. The threshold (commonly 65°F) represents the temperature above which cooling is typically required.

Cumulative distribution function (CDF)

A cumulative distribution function (CDF) is a statistical tool that quantifies the probability that a variable has a value less than or equal to a given threshold. In the context of climate data, representing the long-term historical record of a variable as a CDF shows how often certain values have occurred, very similar to a historical climatology.

Derived variable

Derived variables are variables that are computed from one or more other variables. An example is "heating degree days" which uses the air temperature variable as an input. Derived variables expand the available variable list beyond the native variables available within the Analytics Engine data sources.

Dynamical downscaling

A method to generate fine scale spatial resolution data from coarse scale global climate models using a dynamical regional weather model. The regional model uses the global climate model output as the inputs to generate additional projections. The Analytics Engine hosts dynamically downscaled model output created using the [Weather Research and Forecasting \(WRF\)](#) model.

Effective sample size (ESS)

An independent estimate of sample size that accounts for the reduction in statistical information caused by temporal autocorrelation in a dataset. When observations within a block are strongly correlated with one another (as is common in meteorological data), the effective number of independent data points is lower than the actual number of observations.

Effective temperature

A heat metric that represents the cumulative influence of temperatures over multiple consecutive days, rather than relying on a single day's measurement. It is calculated as a weighted sum of the



current day's temperature and temperatures from preceding days, with weights that vary depending on the application.

Ensemble member

A single simulation or GCM run within a larger set of simulations (referred to as an ensemble) that are generated to assess the uncertainty and variability in climate model projections. Each ensemble member is typically produced by slightly varying the initial conditions, model parameters, or by using different models altogether.

Extreme value analysis

An area of statistical analysis focused on extremes relative to the median of a probability distribution. Concepts from extreme value analysis can be used to analyze extreme weather events (such as a heat wave or extreme precipitation event), including event return levels and return periods.

Gamma distribution

A flexible probability distribution commonly used in climate and hydrological applications to model variables that are strictly positive and right-skewed, such as precipitation totals or soil moisture.

Generalized extreme value (GEV) distribution

A family of continuous probability distributions developed within extreme value theory to model the largest or smallest value from a large collection of random observations. Some distributions within this family include the Gumbel, Fréchet, and Weibull distributions.

Generalized Pareto distribution

A probability distribution used to model the tail behavior of a variety of data types; it is particularly useful for assessing the probabilities of extreme events, such as rare climate events that exceed a certain threshold.

Global climate model (GCM)

A mathematical model that represents the physical processes in the atmosphere, ocean, cryosphere, and land surface. GCMs are used to simulate the climate system and predict future climate conditions based on different scenarios of greenhouse gas emissions and other factors.

Global warming level

A global warming level is defined as the difference in the global mean air temperature from the historical period (defined on the Analytics Engine as the pre-industrial period 1850-1900). Global warming levels are frequently used in international policy discussions (for example, goals to constrain global warming to 1.5 or 2 degrees Celsius). The standard global warming levels are 1.5°C, 2°C, 3°C, and 4°C.



Goodness of fit

A statistical measure that describes how well a model's outputs align with observed data. It is used to assess the accuracy of models or the accuracy of statistical methods in replicating real-world phenomena.

Gumbel distribution

A probability distribution used to model the distribution of the maximum (or the minimum) value from a number of samples of various distributions. It is commonly applied in extreme value theory, particularly for modeling the distribution of extreme events such as floods, heatwaves, or other climate extremes.

Heat index

A metric that combines air temperature and relative humidity to estimate how hot conditions feel to the human body, often referred to as "apparent temperature". The heat index is widely used to assess heat-related health risk and predict cooling energy demand.

Heating degree days (HDDs)

Heating degree days (HDDs) quantify how much and for how long outdoor temperatures fall below a specific threshold, reflecting demand for indoor heating. The base temperature (commonly 65°F) approximates the point at which heating becomes necessary.

Internal variability

Internal variability represents the natural (stochastic) variations in the Earth's climate, due to interactions between the atmosphere, ocean, land surface, and sea ice. In global climate models with multiple simulations, this uncertainty can result in slightly different outputs based on small differences in the model's physical processes design.

KS test

The Kolmogorov-Smirnov test for nonparametric data evaluates the equality of two different one-dimensional distributions. The KS test can be used to determine when statistical significance is achieved when two samples are sufficiently different from one another.

Localization

Statistical method to produce time-series of future weather at a point-location where a weather station is presently in operation. Localization involves bias adjusting gridded climate model output, based on observations from the weather station.

Maximum likelihood estimation (MLE)

A statistical method used to estimate the parameters of a probability distribution by finding the parameter values that make the observed data most probable under that distribution.



Model run

Model run refers to data from different initial-condition ensemble runs of a GCM. Climate models are often run multiple times with slightly different initial conditions (ensemble members), and each of these runs is termed as a 'model run'.

Model uncertainty

Uncertainty in global climate model output that arises from design differences between models. Global climate models are developed by different research institutions and differ in how they represent the global climate system.

Multivariate event

When several hazards affect the same region at the same time; can also result from the intersection of climate hazards and other environmental hazards. Multivariate events can be temporally compounding, spatially compounding, pre-conditioned, or complex.

Notebook

The Analytics Engine uses Jupyter Notebooks to be a user-friendly and interactive example workflow from start to finish of an application of climate data.

Pearson Type III distribution

A probability distribution that is commonly used in hydrology and climatology to model skewed data, particularly in the context of flood frequency analysis.

Percentage point function (PPF)

The inverse of the cumulative distribution function (CDF); also known as the quantile function. Given a specified probability, the PPF returns the corresponding value of the variable.

Projection

A potential future evolution of a quantity or set of quantities, often associated with climate variables such as temperature or sea level. Unlike predictions, projections do not imply certainty but rather illustrate a range of possible outcomes based on different assumptions and conditions.

Return period

The reciprocal of the return probability. For example, if the maximum temperature at a location has a 10% annual return probability of exceeding 105°F, the return period of that event is 10 years.

Return probability

The probability that a climate or weather event will exceed a specified magnitude in any given year. Return probability is the reciprocal of the return period.



Return value

The magnitude of a climate or weather event associated with a given return probability. Also referred to as a return level.

Scenario uncertainty

Uncertainty that arises from not knowing how people, policies, the economy, and technology will evolve in the future to address the issue of climate change.

Shared socioeconomic pathways (SSPs)

[Shared socioeconomic pathways](#) are inputs into the latest generation of IPCC reports which describe potential pathways the world could take in terms of features such as political, economic, and other societal dynamics and choices which impact greenhouse gas emissions. The Analytics Engine has multiple data sources with several SSPs, including SSP2-4.5, SSP3-7.0, and SSP5-8.5.

Simulation

A computational process used to model and analyze the behavior of complex systems by representing them through mathematical models. In the context of the Analytics Engine, a simulation is a specific ensemble member from a GCM.

Statistical downscaling

A method to generate fine scale spatial resolution data outputs from coarser scale global climate models using statistical relationships between the coarse global climate model output and observed climatological conditions. The Analytics Engine hosts statistically downscaled model output from [LOCA2-Hybrid](#).

Tool

A Python function to work with climate data, analyze data, or produce visualizations. Tools are located within ClimaKitAE.

Toolkit

A Python package consisting of a set of functions [tools] for climate data operations, analyses, and visualizations.

Typical meteorological year

A typical meteorological year ([TMY](#)) is a complete set of meteorological variables at a given location for every hour in a year. TMYs are used in some building and energy system modeling applications to describe typical annual weather conditions at a specific location.



Use-case

Use-cases are written in the following format: [data product] + [application of data product]. Data products consist of the climate data itself being developed by the Analytics Engine development team. Applications of the data product are specific applications as identified by energy industry partners.

Weibull distribution

A probability distribution used to model the distribution of extreme values, particularly for analyzing the reliability of systems and the occurrence of extreme weather events. This distribution falls within the GEV distribution family.

Wet bulb globe temperature (WBGT)

A heat stress metric that estimates the combined effects of air temperature, humidity, wind speed, and solar radiation on the human body, particularly in direct sunlight. WBGT is [widely used to assess heat hazards for outdoor workers](#).

Working group

In the context of the Analytics Engine, a working group refers to a focused discussion involving a diverse group of climate scientists, social scientists, energy sector partners, platform developers, data users, and other partners.

Zarr

[Filetype](#) for storing large, multidimensional arrays (such as gridded climate data with spatial and temporal dimensions), optimized for cloud storage.

References

- American Meteorological Society. *Glossary of Meteorology*. <https://glossary.ametsoc.org/>
- Cal-Adapt Analytics Engine. *About Climate Projections and Models*. https://analytics.cal-adapt.org/guidance/about_climate_projections_and_models
- CF Conventions. *NetCDF Climate and Forecast Metadata Conventions*. <https://cfconventions.org/>
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- Eyring, V., et al. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6). *Geoscientific Model Development*, 9, 1937–1958.
- Gutiérrez, J.M., et al. (2021). *IPCC AR6 Interactive Atlas*. In: Climate Change 2021: The Physical Science Basis. <https://interactive-atlas.ipcc.ch/>
- Intergovernmental Panel on Climate Change (IPCC). *Sixth Assessment Report*. <https://www.ipcc.ch/report/sixth-assessment-report-cycle/>
- Lafferty, D.C., et al. (2023). *LOCA2-Hybrid Downscaled Climate Projections for California*. <https://loca.ucsd.edu/>
- National Renewable Energy Laboratory. *National Solar Radiation Database — Typical Meteorological Year*. <https://nsrdb.nrl.gov/data-sets/tmy>



- O'Neill, B.C., et al. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9, 3461–3482.
- Zarr Development Team. *Zarr: An implementation of chunked, compressed, N-dimensional arrays*. <https://zarr.dev/>

