Glossary

Introduction to Climpact: Generating Climate Indices to Support Climate Services

Updated 18 Aug 2023

Term	Definition
Absolute threshold	Threshold expressed as a fixed value. For example, does the temperature exceed 32º Celsius?
Alternative data sources	In Climpact this refers primarily to remote sensing (satellite observations) or reanalysis.
Anthropogenic climate change	Climate change caused by human impacts on the earth.
Assessment	Process in which scientists survey existing scientific studies and peer-reviewed literature.
Attribution	Defined as the process of evaluating the relative contributions of multiple causal factors to a change or event with an assignment of statistical confidence (IPCC, 2007). So, attribution means offering some explanation of what caused a change in the data.
Base period	Also known as the reference period. Generally, a subset of the complete climate record. Ideally, you should have climate observation data from before and after the base period. The times before and after the base period will each be compared to this base period. WMO recommends a standard base period of 1961–1990 for climate change analysis. This facilitates comparison of climate indices between groups of countries. This period is considered the most stable as it was before most global warming effects became evident. According to the WMO, the base period should be at least 30 years, and preferably longer. The base period is usually contained within the overall period of analysis.

Best fit line (or regression line)	The line that best represents the pattern of the points in a scatter plot, using a method called the least squares method. This line represents an equation to model the relationship between this pair of variables. It helps us illustrate the overall pattern of the data, and whether the two variables have a linear relationship.
Box-and-whisker plots	A data plot method that provides a panoramic view of a series. Useful in detecting parts of the series which may be problematic because of how they are outside the range of the other values, i.e. they are "outliers". (For information on box-and-whisker plots, or box-plots, see https://www.mashupmath.com/blog/box-and-whisker-plots-explained.)
Climate change	A change in the state of the climate identified by changes in the mean and/or the range of variability that persists for an extended period, typically decades or longer.
Climate index	A plot of the calculated value in any climate variable over time (e.g. number of warm/cold days or nights, number of frost days in a year, etc.) Climate indices allow a statistical study of variations of the dependent climatological variables.
Climate prediction	Quantitative, reproducible expressions of probability of future climate conditions on timescales ranging from seasons to decades or longer, and on spatial scales ranging from local to regional and global.
Climate scenario	Descriptions of possible futures in the world under given climate conditions which may include considerations of socio-economic sectoral impacts and feedback effects.
Climate variability	Expected fluctuations in the climate system over short or moderate time scales due to natural internal processes or anthropogenic (human-caused) external forcing.
Climatol	Homogenization and plotting tool. Highly regarded tool for data homogenization. Provides additional plotting tools such as "climograms". Most features are not part of Climpact. Does not calculate indices. A stand-alone tool created in R.
Climdex.org	Core climate indices calculation from global data only. Calculates Climpact indices from existing data in specific publicly-available global databases. Does not support use of custom data.

Climpact	Application designed to work with daily observation data of temperature and rainfall from any source. The primary data source should be instrumental observations (from measurements made with instruments in stations on the ground). In their absence—and only when necessary due to lack of available instrumental data—alternate data sources may be used. These might include remote-sensed data (observations from satellites) or reanalysis (a blend of observations with numerical modeling).
Confidence interval	A value that provides an indication of how certain we are about the accuracy of the trend. The confidence interval is calculated as the difference between the lower and upper bound values. The lower and upper bound values represent the maximum and minimum likely slopes of the trend line.
Correlation	A measure of the strength of the relationship between two or more variables. The higher the correlation, the stronger the relationship between the variables. The lower the correlation, the weaker the relationship. Correlations may be positive (also referred to as "direct") or negative (also referred to as "inverse"). The classical correlation measure is a linear correlation, also known as the Pearson correlation using a scale of -1 to +1, where 0 means no relationship, -1 means a perfect inverse relationship (a negative correlation) and +1 means a perfect direct relationship (a positive correlation). This number is also referred to as the "linear correlation coefficient", (or just "correlation coefficient") and is usually represented by the letter 'r'.
Data rescue activities (DARE)	Process of finding data available from resources that are not immediately apparent. This may include such things as digitization of old paper records or rescuing data from old analog instruments.
Decimal Value Histogram	Plot produced in Climpact, useful for identifying rounding biases based on how frequently each decimal value occurs. They show how frequently each of the 10 possible values of the first decimal place are used (e.g, $x.0$ to $x.9$). A "rounding bias" is when data was rounded before recording, for example changing a precipitation value of 10.56 mm to 10.6. This kind of change can distort trends in the data over time.
Detection	The process of demonstrating that an observed change is significantly different (in a statistical sense) than can be explained by natural internal variability of the climate system. In

	other words, detection is just showing that there is a significant change in the data, without explaining why it has occurred.
Detrending	A process in which we remove the overall trend in the data in order to better see the patterns of variation that we are interested in.
Duration indices	Measure the length of time of or between certain events, or the aggregate number of days meeting a certain criteria. For example, the growing season length (GSL) measures the number of days between the first and last instances in the year of daily mean temperatures exceeding temperatures needed for growth.
Expert Team on Climate Change Detection and Indices (ETCCDI)	In 1999, the Expert Team on Climate Change Detection and Indices (ETCCDI) (created by the WMO Commission of Climatology) developed an internationally coordinated set of core climate indices consisting of 27 descriptive indices for moderate extremes, along with a software package called RClimdex. Successor to the Expert Team on Climate Risk and Sector-specific Climate Indices, (ET-CRSCI)
Expert Team on Climate Information for Decision Making (ET-CID)	Successor to the Expert Team on Sector-Specific Climate Indices (ET-SCI)
Expert Team on Sector-Specific Climate Indices (ET-SCI)	Expert Team on Sector-Specific Climate Indices (ET-SCI) was an international team of climate scientists, formed in 2019, by the WMO Commission for Climatology (CCI). This team was dedicated to improving the availability and consistency of sector-specific climate indices through the creation of software, regional workshops, research, and training materials.
Extra-QC	Quality control tool. An extension to RClimDex. Added quality control tools. Also included in Climpact.
Extreme Event	Unexpected, unusual, severe, or unseasonal weather; weather at the extremes of the historical distribution—the range that has been seen in the past. (Wikipedia)
FClimDex	Batch processing tool built on a set of FORTRAN routines using the same methods and algorithms as RClimDex. Added batch processing mode for processing many stations at once. This is included in Climpact.

Forecast	Expected weather conditions in a certain region based on many factors on a time-scale ranging from a few hours to 14-days.
Frequency	Describes the proportion of a dataset that occurs at a certain value or within a range of values. Calculated by dividing the number of occurrences of a particular value by the total number of values. It may then be expressed as a percent by multiplying by 100. For example, in the set (1, 5, 8, 12, 8, 22, 8), the frequency of the value 8 is (3/7)*100 = 42.9%
Frequency indices	Measure the number of days above or below a certain threshold. For example, the number of days each year when precipitation is greater than 10 mm (R10mm), or the percentage of days each year when maximum temperature is above the 90th percentile (TX90p). Thus the threshold used may be absolute or percentile-based.
Historical data	Any available data on weather and climate events that may help in climate analyses.
Homogeneity	A measure of whether data has been collected in a consistent enough way to be able to detect climate trends. In order to have confidence in trends that we may see in our climate indices, we have to be sure that the data has been recorded consistently over time.
Infilled Data	There is a common temptation to use estimations to replace or "infill" missing data. Such estimations are made with statistical or mathematical formulas. For example, they might use the simple average of observations or linear interpolation. However, the use of infilled data will greatly increase the uncertainty of the indices and reduce the ability to identify actual changes in the climate over time.
Intensity indices	Indicate the minimum or maximum of some variable at a set frequency. For example, the maximum daily precipitation each year. Or the minimum value of daily maximum temperatures each month.
IPCC	WMO and the United Nations Environment Programme (UNEP) established the IPCC in 1988 with the role of assessing the scientific, technical, and socioeconomic information relevant for understanding the risk of human-induced climate change. The IPCC was created to provide policymakers with regular scientific assessments on climate change, its implications, and

	potential future risks, as well as to put forward adaptation and mitigation options. IPCC assessments document the state of knowledge on climate change and identify where there is agreement in the scientific community on topics related to climate change and where further research is needed. IPCC assessments and other documents can be accessed on the IPCC website at: www.ipcc.ch.
IPCC Reports	Updated every 5 years. Document the current state of knowledge on climate change and identify areas of scientific agreement.
Linear correlation	Also known as a Pearson correlation, or correlation score, it uses a scale of -1 to +1, where 0 means no relationship, -1 means a perfect inverse relationship (a negative correlation) and +1 means a perfect direct relationship (a positive correlation). This number is also referred to as the "linear correlation coefficient", (or just "correlation coefficient") and is usually represented by the letter 'r'.
Linear trend	Trendline created using the least squares method. Trend estimation method that expresses data as a linear function of time. See best fit line.
Local expertise	Resources at the local level that can provide important insights into the circumstances and history of a region, especially as relevant to climate analyses.
Mean	Calculated by adding up the values and dividing by the number of values. This is the same as the average. Note that the mean can be greatly affected by even one extreme value in the dataset. For example, in the set (1, 5, 8, 12, 22), the mean is the sum of the values (48), divided by 5, which equals 9.6.
Median	The middle value of a dataset when the values are placed in order from lowest to highest. This is not affected by extremes or outliers. It is always just the value in the middle of the dataset. For example, in the set (1, 5, 8, 12, 22), the median is 8.
Monitoring	Long-term observations of those quantities (e.g., temperature, CO_2 concentrations and precipitation) which describe the state of the Earth's surface and atmosphere. For example, the North American Climate Extremes Monitoring tool allows users to examine trends and occurrences of certain types of extreme or threshold events at the station-by-station level: https://www.ncdc.noaa.gov/news/product-highlight-north-america-climate-extremes-monitoring

National Communications	A specific type of report released periodically by governments to fulfill agreements made under the UNFCCC. Include information on expected potential climate change impacts. Are part of the efforts to identify strategies for societies to respond to specific climate change trends.
Nationally Determined Contributions (NDCs)	Main focus is on mitigation, though might mention adaptation. Focuses on rate of emissions, commitment to reduction and include plans for reducing greenhouse gas (GHG) emissions. Actions that countries plan to undertake to address climate change.
Pearson correlation	Also known as linear correlation, or correlation score, it uses a scale of -1 to +1, where 0 means no relationship, -1 means a perfect inverse relationship (a negative correlation) and +1 means a perfect direct relationship (a positive correlation). This number is also referred to as the "linear correlation coefficient", (or just "correlation coefficient") and is usually represented by the letter 'r'.
Percentile	A measure of the value below which a given percentage of observations falls. For example, the 20th percentile is the value below which 20% of the observations fall.
Period of analysis	Time period chosen for use from the available data. It is important to have a clear and well-documented set of criteria by which you define the period of analysis you will use out of the data set available.
Physical reasoning	An approach that relies on a cause-effect analysis for looking at climate relationships. For example, suppose we find a correlation between a change in sea surface temperature and a corresponding change in the amount of precipitation in a nearby area. Because of our understanding of climate processes, and the close proximity, we can theorize a relationship between these changes.
Quality control (QC) process	Process of looking for errors in data that might have been introduced during the digitization or observation process. This might include finding errors such as misplaced decimal points, incorrectly entered values, measurements from uncalibrated instruments, or accidentally duplicated values.
R package	R packages are extensions to the R statistical programming language. R packages contain code, data, and documentation in a standardised collection format that can be installed by users of R, typically via a centralised software repository such as CRAN (the Comprehensive R Archive Network). (Wikipedia)

R^2 value (determination coefficient)	A statistical measure of how close the data points are to the line of best fit on a scatter plot. The higher the R^2 is, the more of a relationship there is. Here we see that the average minimum temperature is related to the wheat production only by 3% ($R^2=0.03$). So, essentially, there is no relationship found here.
Range	The difference between the highest and lowest values. Often used together with other parameters, such as mean, mode, and median to express the total variability of a data set over a time period. For example, it can be used with temperature data to express the annual highest and lowest temperatures of a location.
RClimDex	Tool for core climate indices calculation from custom data. Built in R programming language. Calculates the 27 core indices, and basic quality control features. Supports custom data sets. These features are all contained in Climpact.
Reanalysis data	Estimates based on the use of climate models, combined with available observations at the large scale. Observation data are not available for all locations globally.
Regional time series	Method of integrating representative information from multiple locations or larger regional areas, supporting trend analyses at a regional level.
Relative Threshold	A threshold based on a value compared to other values in a sample. For example, the 90th percentile for temperature may vary with elevation.
Remote sensing	satellite observations
RH-Test	Homogenization tool. Provides data homogenization features that are not in Climpact. A stand-alone tool created in R.
Sen's slope method	A special, more robust, type of linear slope that is less impacted by outliers. It reflects the median slope of all ordered pairs of points in a dataset and so is particularly effective for calculating trends in datasets that might contain extreme values.
Standard Deviation	A measure of variability or dispersion. A large standard deviation indicates that observations/data points are far from the mean and a small standard deviation indicates that they are clustered closely around the mean.

Standardized Precipitation Index (SPI)	Provides a standardized distribution tool used to establish categories for very wet to very dry conditions. It can describe intensity, frequency and duration of extreme precipitation events.
Statistical significance	At the most basic level, statistical significance is a numerical measure of whether a result is due to a relevant factor rather than chance. Uses a binary approach. According to some prominent climate scientists, limiting statistical inferences to this type of binary conclusions is unnecessarily restrictive and can be misleading.
Thresholds	A quantity that must be exceeded for a certain event or result or condition to occur. For example, if precipitation exceeds 200 mm in a day in a given region, flooding is likely to occur.
Time series	A series of data points listed by time. Daily time-series of precipitation, maximum temperature, minimum temperature and diurnal temperature range allow the user to view the data and identify obvious problems by eye, such as missing data (indicated by red circles) or potentially unrealistic values.
Trend	"Trend" typically refers to a line or curve that connects or passes through two or more data points in a plot, showing the general direction of change over time. Climpact calculates trends based on the linear method.