

Climate Data Indices Tool (CDITool)

Software Documentation

A hands-on guide for the calculation of climate indices from CORDEX datasets

Author: Sven Kralisch (sven.kralisch@uni-jena.de)

Software version: 1.0_00

Date: 19/06/2022

1 Background

Future Climate Change (CC) is one of the major trends to sustainable development in the Hindu Kush Himalaya region and worldwide. Analyzing the outputs of related climate projection models and therefore is a crucial requirement for understanding sectoral CC impacts and for developing effective adaptation strategies. To ease the extraction of meaningful information for different user groups and improve interpretability, the calculation of indices that target specific thematic areas of climate impact is a common approach. However, the processing of often large climate projection datasets, index calculation, and visualization through maps and charts often rely on cumbersome and error-prone workflows. Therefore, a software tool is needed that supports non-expert users in analyzing climate projection datasets through a transparent and intuitive data processing workflow.

The CDITool software addresses these challenges by providing a flexible and user-oriented data analysis workflow. The tool is built on top of the JAMS¹ modelling platform, a software system for creating environmental models from individual data analysis and process simulation components. Therefore, CDITool can be easily adapted and utilizes all the existing JAMS data handling and visualization functions. In summary, the software covers the following functions:

1. Extraction of precipitation and temperature data of NetCDF model outputs from the *Coordinated Regional Climate Downscaling Experiment* (CORDEX)²
2. Calculation of climate indices defined by the *Expert Team on Sector-Specific Climate Indices* (ET-SCI) based on the extracted data
3. Result analysis, visualization and output

The next sections will provide an overview of how to accomplish these steps using CDITool.

2 Step by step for Climate Data Indices Tool

2.1 Preparation and start

CDITool comes as a ZIP archive that includes all the required files for running the software. The most current version of the software can be downloaded from <http://jams.uni-jena.de/cditool>. Since it is a Java application, CDITool requires a Java Runtime Environment (JRE) that is packaged in the archive. To run the software, follow these steps:

1. Extract the ZIP archive into a new folder
2. Start the **cditool.exe** file in the root of the folder

An empty CDITool window will open in result (Figure 1). If required, the version of the software can be accessed by selecting **Help** → **About** from the main menu.

¹ <http://jams.uni-jena.de>

² <https://cordex.org>

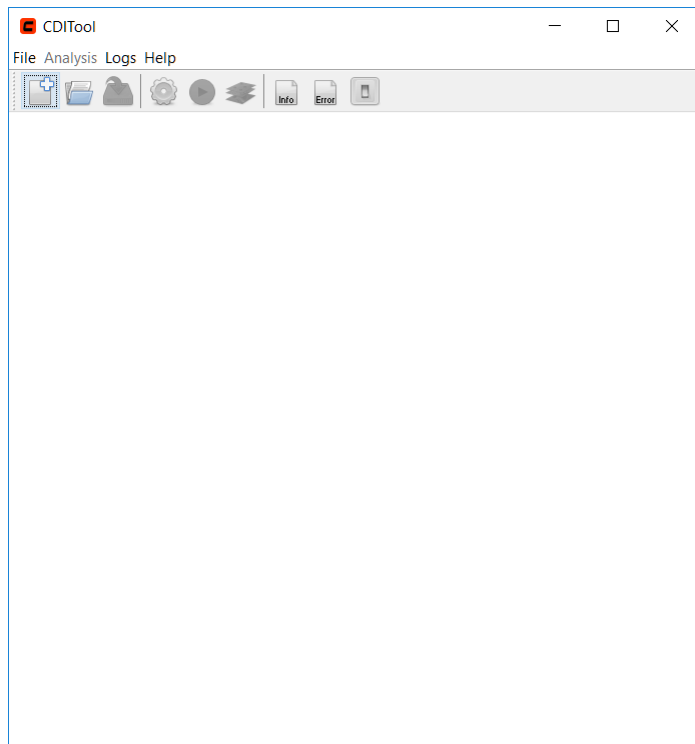


Figure 1: CDITool start view

2.2 Create/open workspace

To start a new data analysis, a workspace has to be created. This is a directory folder in the file system that contains all the data and files of a single index calculation. To create a new workspace, follow these steps:

1. Select **File** → **New Analysis** from the menu or press  in the toolbar
2. In the **Open** dialog, select an existing, empty folder or press the **New Folder** button () to create one
3. Press the **Open** button

In result, the CDITool window will create the new workspace in the desired location and will open the new analysis view (Figure 2). Each analysis view is separated into panels (shown as tabs) that cover the following functions and settings:

- **Data Extraction:** functions for selecting the climate model, defining the grid cells to be analyzed and for extracting the related data
- **Generic:** generic settings of the climate data analysis, covering the analysis period and the core input data to be analyzed
- **Aggregation:** time intervals for which the input parameters and indices will be aggregated
- **Precipitation Indices:** selection of / settings for indices based on precipitation
- **Temperature Indices:** selection of / settings for indices based on temperature
- **Trends:** selection of / settings for trend calculations

The following sections will explain the available options in more detail.

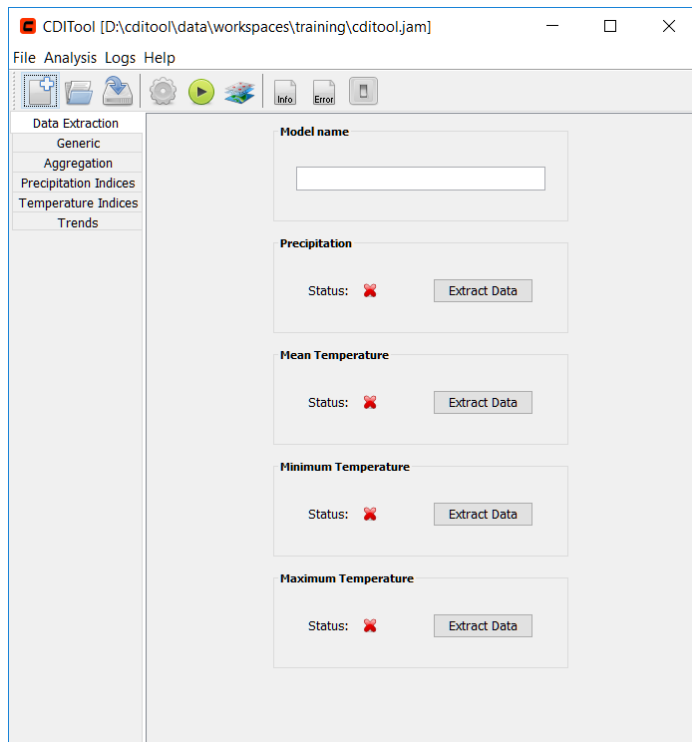


Figure 2: New analysis view

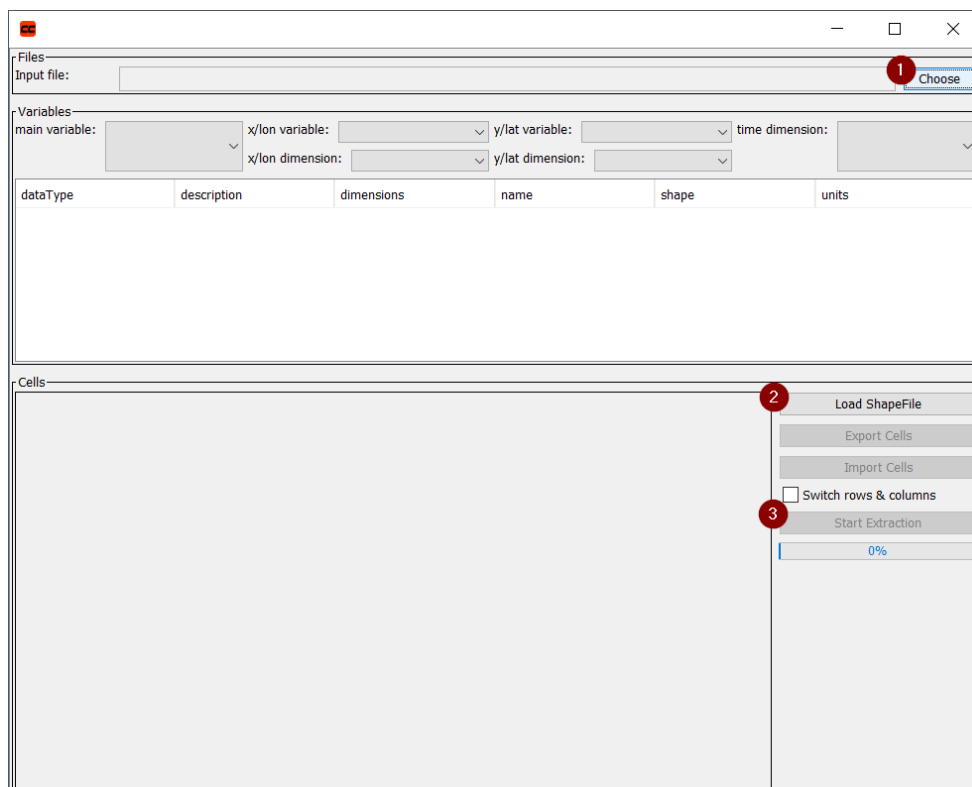


Figure 3: Data extraction dialog window

2.3 Data extraction from climate models

The first step of calculating climate indices from climate models is the definition of the models' grid cells to be analyzed and the extraction of the related data. Four datasets have to be processed, including precipitation, mean temperature, minimum temperature, and maximum temperature

data. Each of these datasets is processed at a time and has to be provided by a NetCDF dataset that covers the full time period to be analyzed.

The **Data Extraction** panel shows a list of the four datasets together with an indicator icon showing whether the dataset was extracted already (green check mark) or not (red cross mark) (Figure 2).

To start the extraction process for one of the datasets, e.g. precipitation, the related **Extract Data** button has to be pressed. In result, a new window will open (Figure 3). The data extraction procedure consists of three steps with the related buttons marked in Figure 3:

1. Selection of a NetCDF input file
2. Selection of model grid cells to be analyzed by means of a polygon Shapefile containing the larger region of interest
3. Extraction of data

After button 1 was pressed, a file input dialog will ask for the NetCDF input file. After the file was chosen, meta-information is extracted and shown in the upper part of the window (Figure 4). The following NetCDF variables, which are marked in Figure 4, need to be defined:

1. the main variable (e.g. precipitation)
2. the longitude variable
3. the latitude variable
4. the longitude dimension variable (differs from #2 for rotated datasets)
5. the latitude dimension variable (differs from #3 for rotated datasets)
6. the time dimension variable

By default, the software will try to detect this information automatically. However, the identified values should be double-checked to avoid erroneous data. A table below the variable selection will further list meta-information extracted from the NetCDF file.

dataType	description	dimensions	name	shape	units
double	time	[time = UNLIMITED; /...	time	[0:45259]	days since 1949-12-01 0...
double		[time = UNLIMITED; /...	time_bnds	[0:45259, 0:1]	
double	longitude	[lon = 13;]	lon	[0:12]	degrees_east
double	latitude	[lat = 13;]	lat	[0:12]	degrees_north
float	Precipitation	[time = UNLIMITED; /...	pr	[0:45259, 0:12, 0:12]	kg m-2 s-1

Figure 4: Definition of NetCDF variables

In the next step, the desired model grid cells must be selected. For this purpose, button 2 (Figure 3) is pressed and a polygon Shapefile with the larger region of interest has to be selected. This Shapefile has to intersect the model domain to obtain meaningful results. After the selection of the Shapefile, the contained polygons are displayed in a map view together with the boundaries of all model grid cells (Figure 5).

The map view features basic functions for panning (hold right mouse button), zooming (mouse wheel) and selection of grid cells (left mouse button). After panning/zooming to the area of interest, model grid cells can be selected. This can be done via the following options:

- A rectangular extent can be selected by left-clicking/holding at the upper/left cell, moving the mouse to the lower right cell and releasing the button (Figure 6, left)
- Individual cells can be selected by left-clicking to a first cell, pressing the CTRL-key and subsequently selecting further cells by left-clicking them (Figure 6, right)

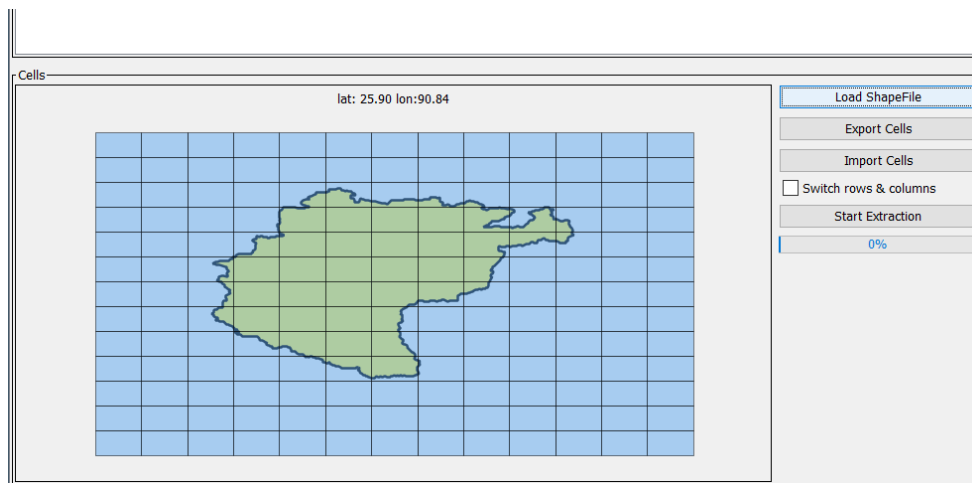


Figure 5: Model grid cells and region of interest

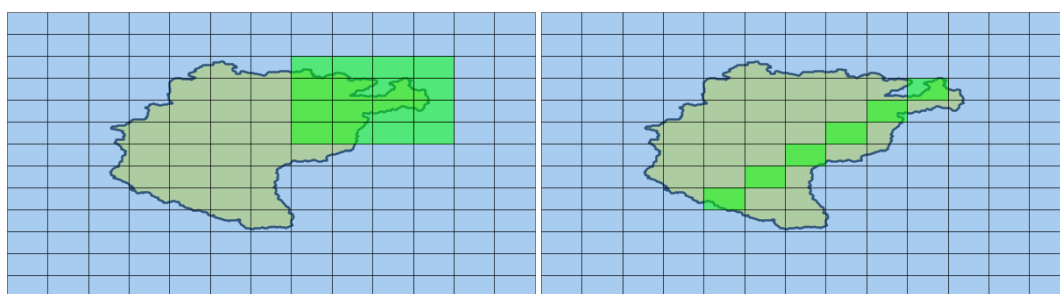


Figure 6: Grid cell selection using rectangular extent (left) or individual selection (right)

After the grid cells have been selected, further, optional functions can be applied (Figure 5, right). A function for exporting and importing of a grid cells selection allows to store the current selection to a text file and to load this selection at a later time. The buttons labelled **Export Cells** and **Import Cells** will open related file save/open dialogs that will ask for user defined file name.

Another function allows to switch the orientation of rows and columns (i.e. latitude and longitude) in the NetCDF data file. While this should not be required for standard NetCDF data files, this function may be handy in situations where both dimensions have been changed due to a reprojection of the original data without adapting the NetCDF metadata. In these cases, the checkbox labelled **Switch rows & columns** can be enabled to ensure correct data extraction.

After all settings have been made, the data extraction process can finally be started by pressing the **Start Extraction** button (Figure 3, button 3). This will extract time series of the selected parameter (e.g. precipitation) for the selected grid cells into a pre-defined text file inside the current workspace. In addition, a Shapefile containing the outline polygons of all selected grid cells will be generated and stored inside the workspace. If the extraction was successful, the status label next to the corresponding dataset in the Data Extraction panel of the CDITool main window will show a green check mark (Figure 7).

This data extraction procedure has to be applied for all required datasets, i.e. precipitation, mean temperature, minimum temperature, and maximum temperature. While the NetCDF input file has to be changed according to the required data, the selected grid cells must stay untouched in order to ensure that all four datasets have been

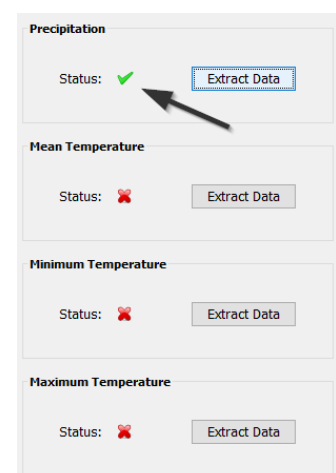


Figure 7: Data extraction panel in main JAMS/CCIndex window after extraction of precipitation data

extracted for identical grid cells. At the same time, it must be ensured that NetCDF input files cover the same time period and contain identical spatial and temporal resolutions, i.e. result from the same climate model.

In the case that only the mean temperature of a climate model is available, identical temperature data can be used to provide all three temperature input datasets as a workaround. However, some of the calculated indices will not gain meaningful results. For example, the index *DTR* (*Daily Temperature Range*) will only contain zero values since minimum and maximum temperature show identical values. In general, index calculation will not be possible before all four datasets have been provided, i.e. showing four green checkmark icons.

In a final step of the data extraction process, a model name can be provided in the **Model name** text field of the **Data Extraction** panel (Figure 2). This model name will be used during the following output of analysis results and will allow to easily compare the indices of different climate models.

2.4 Index calculation settings

Once all input parameters have been extracted, the software is ready to start the index calculation using recommended settings. However, depending on individual requirements these settings that can be changed. As outlined in section 2.2, the settings are subdivided into five groups, i.e. Generic, Aggregation, Precipitation Indices, Temperature Indices, and Trends which are organized in individual panels/tabs.

The **Generic** panel (Figure 8) provides options to enable/disable the aggregation and output of the input variables. Further, the group provides an option to calculate indices only for a user-defined time interval. The latter will be done for the **Individual Time Period** setting once the check box **Default Time Period** was disabled. The **Percentile Time Period** setting defines the time period which will be used for calculating percentile-based indices.

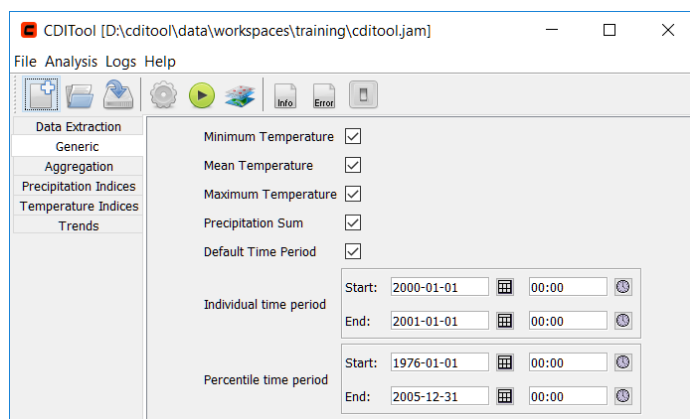


Figure 8: "Generic" settings

The **Aggregation** panel (Figure 9) offers options to enable or disable individual aggregation intervals for climate parameters and indices. Disabling any of these options will result in not aggregating indices for the related time periods, which can save processing time and storage space especially for large datasets.

The **Precipitation Indices** and **Temperature Indices** panel list options related to the calculation of precipitation and temperature indices (Figure 10). Both panels show three options that allow choosing the extend of indices calculated. These are arranged into the following groups (see appendix):

1. Core ET-SCI indices
2. Non-Core ET-SCI indices

3. Miscellaneous, non-ET-SCI indices

As with aggregation options, disabling any of these options will save processing time and storage space.

In addition, these panels contain index-specific options, such as the rain threshold required for calculating the *Rnnmm* precipitation index and other user-defined thresholds.

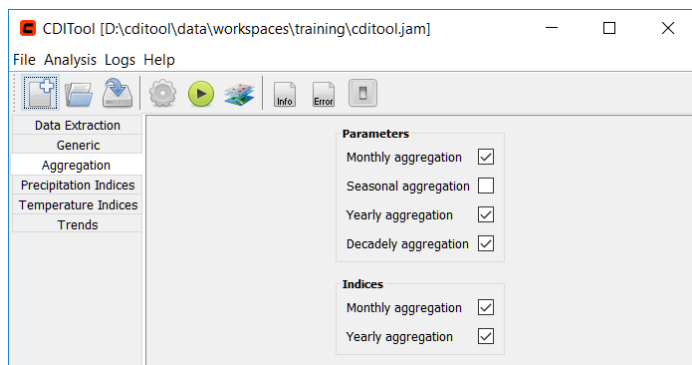


Figure 9: "Aggregation" settings panel

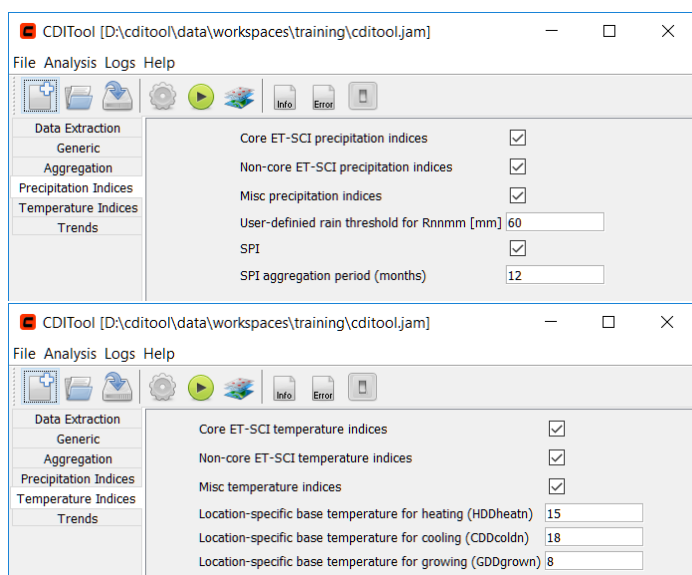


Figure 10: Options related to the calculation of precipitation (upper panel) and temperature (lower panel) indices

The last panel is related to **Trend** calculation (Figure 11) and offers options to enable/disable the calculation of trends and moving averages as well as their related window sizes in number of years.

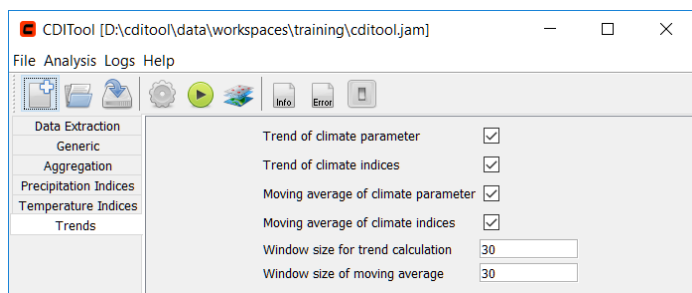


Figure 11: "Trend" settings panel

2.5 Index calculation

After all required input data have been extracted and settings been made, the actual trend calculation can be started. This is done by pressing the green start button in the tool bar (▶) or by choosing **Analysis** → **Start Analysis** from the main menu. In result, a window will open that shows the progress of the index calculation and various status outputs (Figure 12). Using the buttons in the tool bar of this window, the index calculation can be stopped (■). The calculation process is finished once the close button (✖) is enabled and the window can be closed.

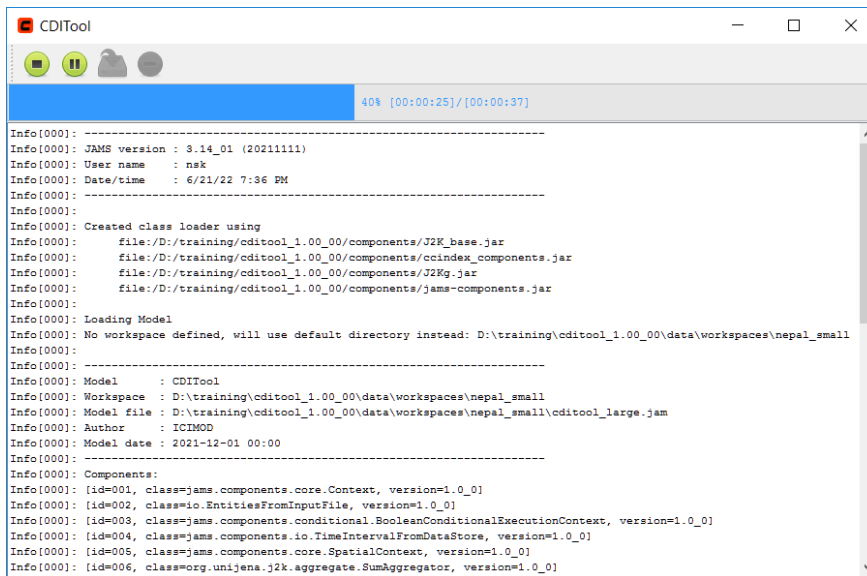


Figure 12: Index calculation progress and status window

2.6 Result analysis and visualization

Once the index calculation is finished, its results can be explored by means of the CDITool data explorer (JADE). This separate tool can be opened by pressing the respective button in the toolbar (🗂️) or by selecting **Analysis** → **Data Explorer** from the main menu. Figure 13 shows the start view of JADE.

JADE is separated into two main panels. The panel on the left-hand side shows a tree view that reflects the organization of input and output data files of the workspace. While the input data consist of a single dataset each, all output data come in two flavors:

1. As an ASCII file that contains the spatio-temporal result datasets, indicated by a file icon (📄).
2. As a Shapefile vector dataset that contains the boundaries of the selected grid cells as geometries and the calculated indices as attribute data of these geometries, indicated by a directory icon (📁).

The panel on the right-hand side shows the content of one or more data files that have been previously selected. Selection of a dataset can be done by double-clicking the respective entry or by right-clicking it and choosing **Show data**. Shapefiles cannot be analyzed with JADE and have to be further processed in standard GIS software such as QGIS³.

Once a dataset is opened in JADE, a new tabulator panel will open in the right-hand side of the window. While the number of open datasets is not limited, they can be closed by clicking the close button in the tool bar (🗑️). The dataset view (Figure 14) is again separated into two areas. The upper panel (1) shows the available data, time steps and grid cells of a dataset. The lower panel

³ <https://www.qgis.org>

(2) shows a table view of these data. The way of how the data are transferred into the table view is defined by the user. Here, JADE offers various functions for aggregating the data in time or space and for outputting selected values as a cross-product. Typical use-cases are:

1. Show time series of the mean value of all grid cells (spatial aggregation)
2. Show the temporal mean across all/selected time steps for all grid cells (temporal aggregation)
3. Show time series of a single/multiple grid cells (cross-products)

The following section will give an overview of these options.

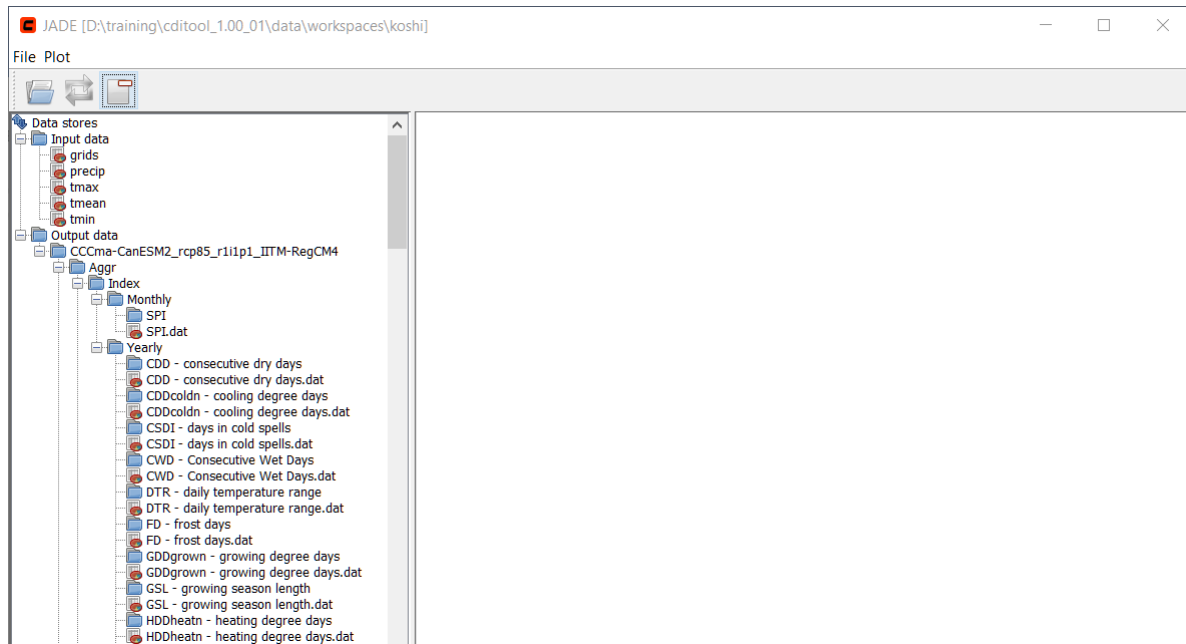


Figure 13: CDITool data explorer (JADE)

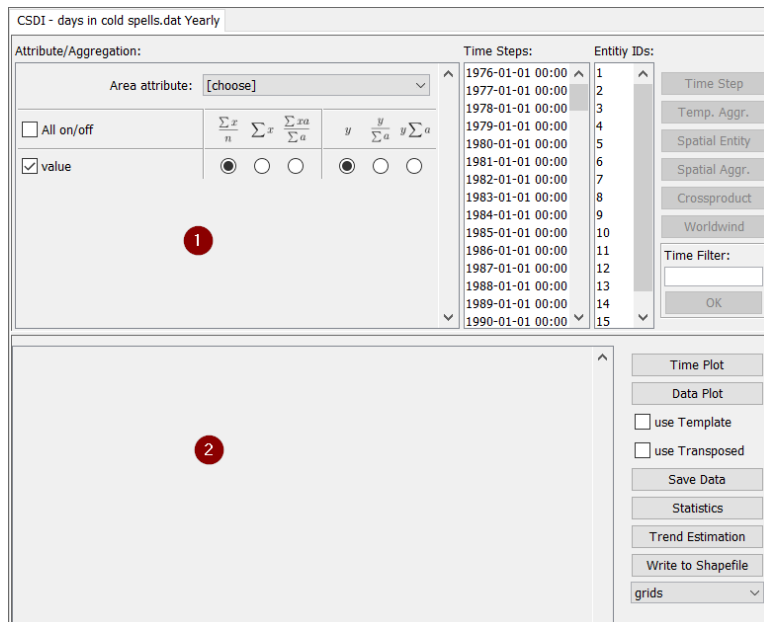


Figure 14: Dataset view panel

2.6.1 Data aggregation

Spatial aggregation

The spatial aggregation requires the selection of the grid cells to be aggregated. This happens by selecting entries from the **Entity IDs** list. A double click into the list will select all grid cells. This can also be achieved by pressing the CTRL-A key combination. Holding the CTRL-key while left-clicking entries from the list will select arbitrary subsets. Left-clicking a first entry and subsequently left-clicking a second entry while holding the SHIFT-key will select all entries in-between. Once the grid cells have been selected, the type of aggregation must be chosen. Figure 15 shows the three available options: (i) mean, (ii) sum, and (iii) area-weighted mean. Since all grid cells cover the same area, the last option is not needed. In order to choose mean or sum aggregation, select the radio button under your preferred option. To finally calculate the aggregates, press the **Spatial Aggr.** button. In result, the table in the lower part of the panel will be filled with a time series of the aggregated values. Figure 16 shows the result of a spatial mean of the CSDI over all grid cells.

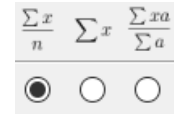


Figure 15: Data aggregation options

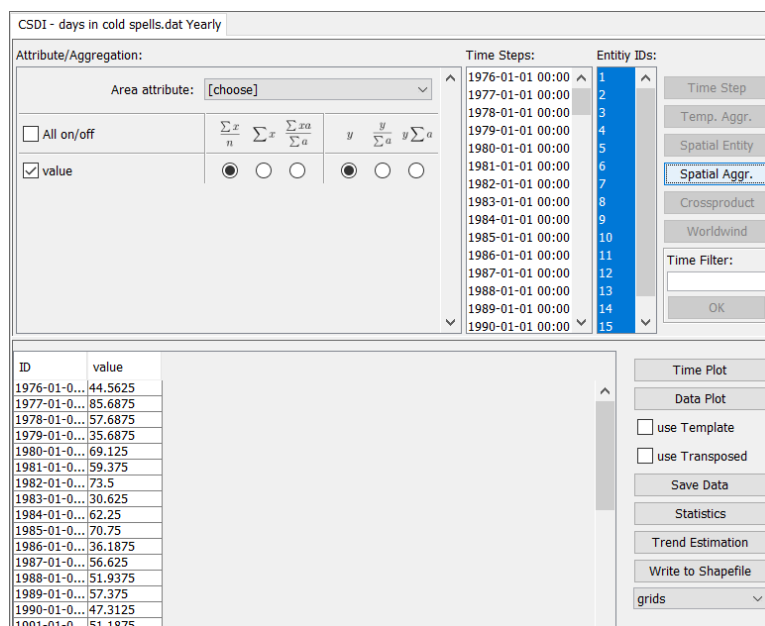


Figure 16: Result of a spatial aggregation of the CSDI index

Temporal aggregation

Temporal aggregation is very similar to spatial aggregation with the difference that date entries from the **Time Steps** list are chosen this time. The selection of date entries follows the same procedure as with the selection of grid cells. The same applies for the configuration of the aggregation type (sum vs. mean). The calculation the aggregates is triggered by pressing the **Temp. Aggr.** button this time. In result, the table in the lower part of the panel will be filled with a list of aggregated values for each grid cell.

Cross-products

The cross-product output option simply extracts the data for all selected dates and grid cells as a table. Alternatively, the data can directly be transferred into the integrated **Worldwind** viewer for further exploration (see section 2.6.2 *Worldwind*). Figure 17 shows an example result of the cross-product output option.

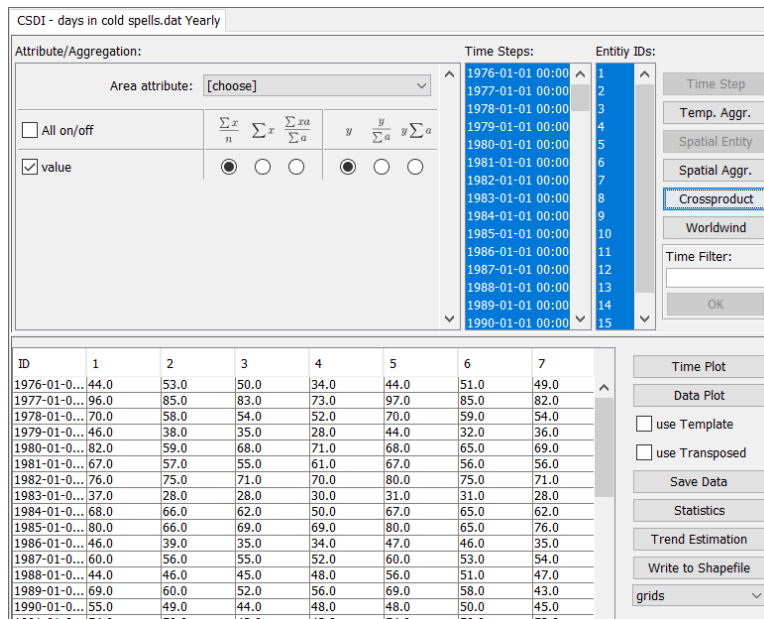


Figure 17: Cross-product output result

2.6.2 Data visualization

Charts

The easiest way of visualizing tabular data is by using charts. Once data are listed in the table view of the lower panel, they can be visualized by means of time plots (first column contains dates) or data plots. This is done by selecting the data that should be visualized (key combo CTRL-A will select all data) and subsequently pressing the **Time Plot** or **Data Plot** buttons. Figure 18 shows an example of a time series visualization using five grid cells. The charts offer zoom functionality (drag & drop inside the plot area), image output functions (right click inside the plot area). A variety of options within the plot viewer allows to further tune the layout of the plot.

Worldwind

Visualization of spatio-temporal data is supported by means of the build-in Worldwind⁴ viewer. For this to work, dates and grid cells of interest have to be selected first, such as for the Cross-product output. Afterwards, the **Worldwind** button must be pressed. In result, the Worldwind viewer will open and zoom to the location of the selected grid cells (Figure 19). In a next step, the data must first be classified in order to visualize them. This is initiated by pressing the **Classify** button. Another window will open, offering options for data classification. These include the number of classes and the classification method. Pressing the **Calculate** button will initiate the classification of the data and will show a histogram of the data distribution. The color setting will further allow to select a start color (linked to the first class) and an end color (linked to the last class). Once the colors have been selected, the classification settings can be applied to the visualization by pressing the **Apply** button.

Once the classification was applied, the view port of the Worldwind viewer can be changed by right-clicking inside the map while moving the mouse. At the same time, the mouse wheel can be used to zoom in or out. To change the display date, the **Time Steps** slider can be used to move forward/backward in time. The play button () can further be pressed to automatically change to the next date after a defined number of milliseconds. The latter can be defined by means of a text field next to the play button. Detail information about the data of a single grid cell can be

⁴ <https://worldwind.arc.nasa.gov>

accessed by left-clicking into the grid cell. Figure 20 shows an example output of the Worldwind viewer.

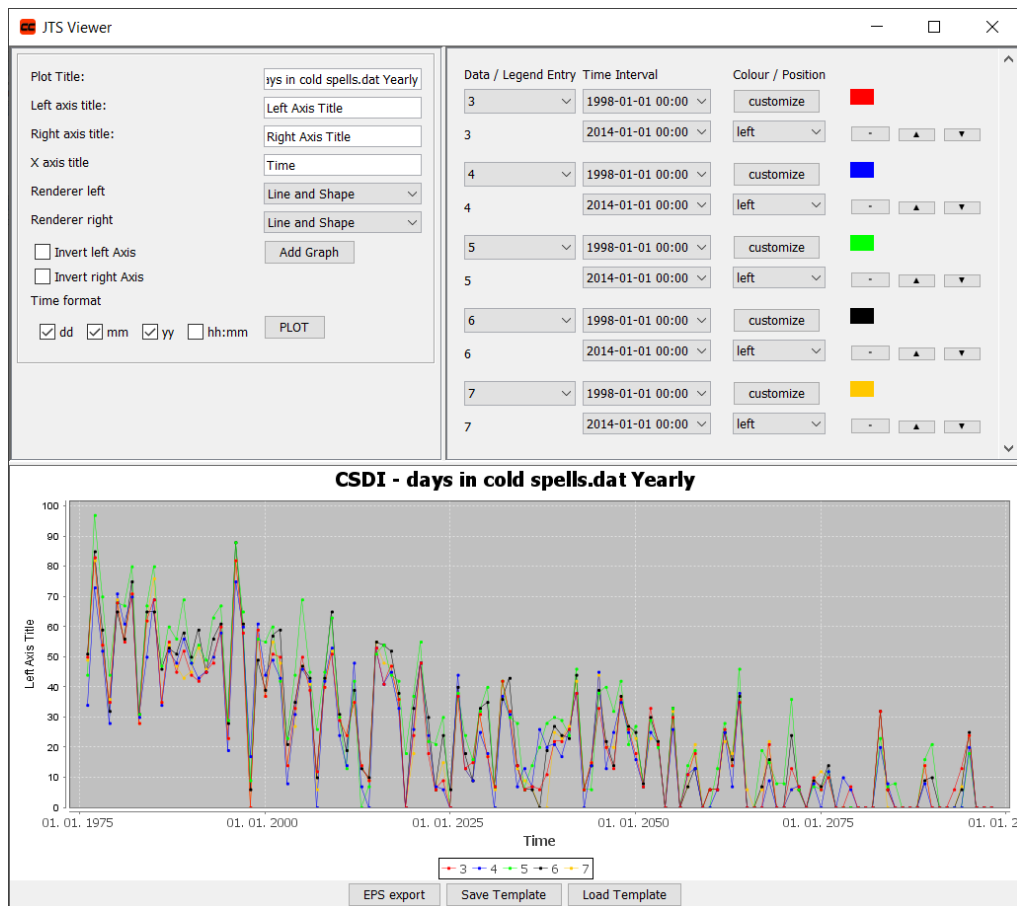


Figure 18: Time series visualization

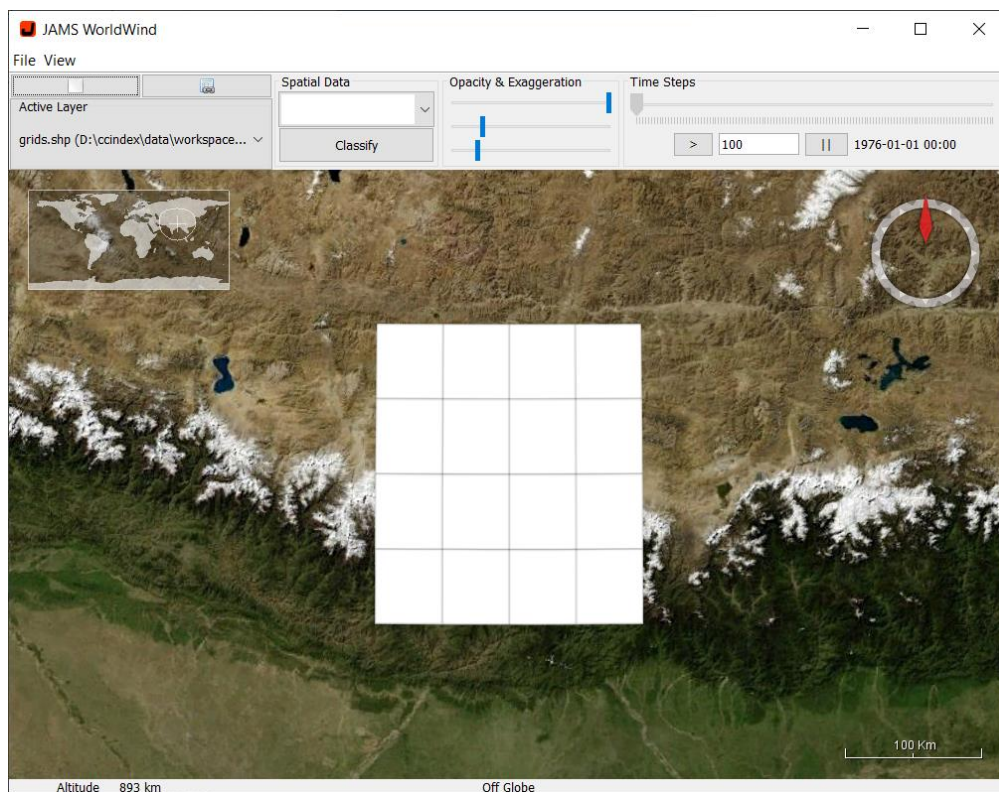


Figure 19: Worldwind viewer

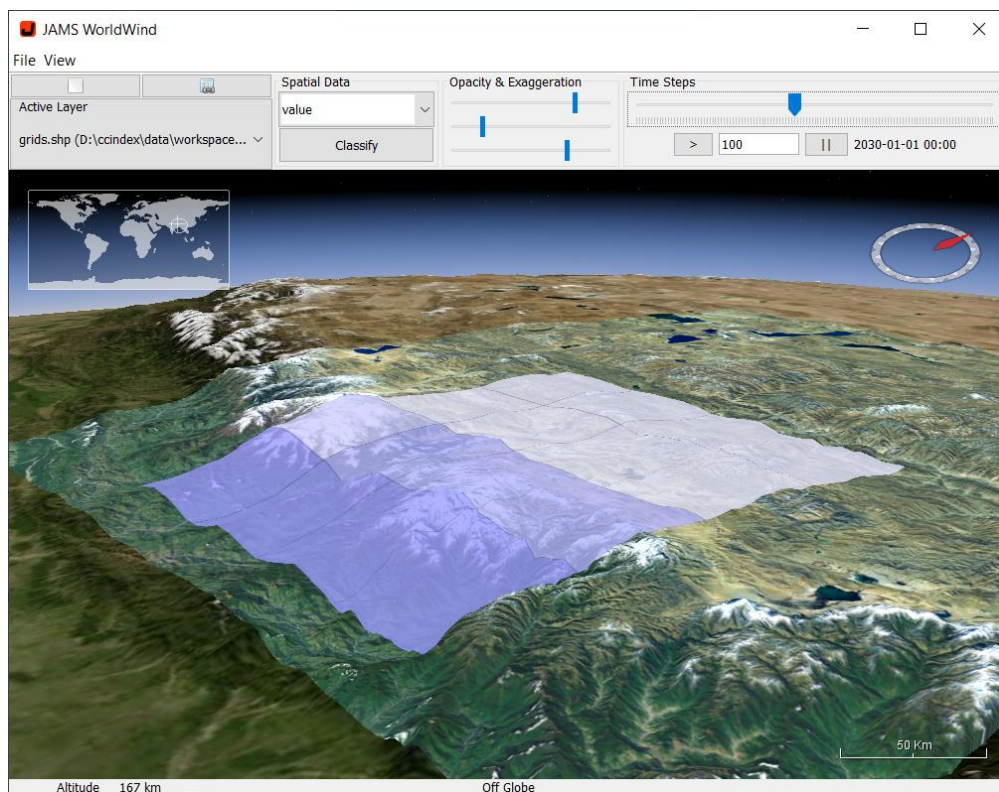


Figure 20: Spatio-temporal data visualization with the Worldwind viewer

3 Appendix: ET-SCI climate indices to be considered in CDITool

Core ET-SCI indices (Mistry, 2019)

Short name	Long name	Definition	Plain language description	Units	Sector(s)
FD	Frost Days	Number of days when $TN < 0^{\circ}C$	Days when minimum temperature is below $0^{\circ}C$	days	H, AFS
TNlt2	TN below $2^{\circ}C$	Number of days when $TN < 2^{\circ}C$	Days when minimum temperature is below $2^{\circ}C$	days	AFS
TNltm2	TN below - $2^{\circ}C$	Number of days when $TN < -2^{\circ}C$	Days when minimum temperature is below $-2^{\circ}C$	days	AFS
TNltm20	TN below - $20^{\circ}C$	Number of days when $TN < -20^{\circ}C$	Days when minimum temperature is below $-20^{\circ}C$	days	H, AFS
ID	Ice Days	Number of days when $TX < 0^{\circ}C$	Days when maximum temperature is below $0^{\circ}C$	days	H, AFS
SU	Summer days	Number of days when $TX > 25^{\circ}C$	Days when maximum temperature exceeds $25^{\circ}C$	days	H
TR	Tropical nights	Number of days when $TN > 20^{\circ}C$	Days when minimum temperature exceeds $20^{\circ}C$	days	H, AFS
GSL	Growing Season Length	Annual number of days between the first occurrence of 6 consecutive days with $TM > 5^{\circ}C$ and the first occurrence of 6 consecutive days with $TM < 5^{\circ}C$	Length of time in which plants can grow	days	AFS
TXx	Max TX	Warmest daily TX	Hottest day	$^{\circ}C$	AFS
TNn	Min TN	Coldest daily TN	Coldest night	$^{\circ}C$	AFS
WSDI	Warm spell duration indicator	Annual number of days contributing to events where 6 or more consecutive days experience $TX > 90$ th percentile	Number of days contributing to a warm period (where the period has to be at least 6 days long)	days	H, AFS, WRH
CSDI	Cold spell duration Indicator	Annual number of days contributing to events where 6 or more consecutive days experience $TN < 10$ th percentile	Number of days contributing to a cold period (where the period has to be at least 6 days long)	days	H, AFS
TXgt50p	Fraction of days with above average temperature	Percentage of days where $TX > 50$ th percentile	Fraction of days with above average temperature	%	H, AFS, WRH
TX95t	Very warm day threshold	Value of 95th percentile of TX	A threshold where days above this temperature would be classified as very warm	$^{\circ}C$	H, AFS
TMge5	TM of at least $5^{\circ}C$	Number of days when $TM \geq 5^{\circ}C$	Days when average temperature is at least $5^{\circ}C$	days	AFS

TMlt5	TM below 5°C	Number of days when TM < 5 °C	Days when average temperature is below 5°C	days	AFS
TMge10	TM of at least 10°C	Number of days when TM ≥ 10 °C	Days when average temperature is at least 10°C	days	AFS
TMlt10	TM below 10°C	Number of days when TM < 10 °C	Days when average temperature is below 10°C	days	AFS
TXge30	TX of at least 30°C	Number of days when TX ≥ 30 °C	Days when maximum temperature is at least 30°C	days	H, AFS
TXge35	TX of at least 35°C	Number of days when TX ≥ 35 °C	Days when maximum temperature is at least 35°C	days	H, AFS
HDDheatn	Heating Degree Days	Annual sum of n - TM (where n is a user-defined location-specific base temperature and TM < n)	A measure of the energy demand needed to heat a building	degree-days	H
CDDcoldn	Cooling Degree Days	Annual sum of TM - n (where n is a user-defined location-specific base temperature and TM > n)	A measure of the energy demand needed to cool a building	degree-days	H
GDDgrown	Growing Degree Days	Annual sum of TM - n (where n is a user-defined location-specific base temperature and TM > n)	A measure of heat accumulation to predict plant and animal developmental rates	degree-days	H, AFS
CDD	Consecutive Dry Days	Maximum number of consecutive dry days (when PR < 1.0 mm)	Longest dry spell	days	H, AFS, WRH
R20mm	Number of very heavy rain days	Number of days when PR ≥ 20 mm	Days when rainfall is at least 20mm	days	AFS, WRH
PRCPTOT	Annual total wet-day PR	Sum of daily PR ≥ 1.0 mm	Total wet-day rainfall	mm	AFS, WRH
R95pTOT	Contribution from very wet days	100*r95p / PRCPTOT	Fraction of total wet-day rainfall that comes from very wet days	%	AFS, WRH
R99pTOT	Contribution from extremely wet days	100*r99p / PRCPTOT	Fraction of total wet-day rainfall that comes from extremely wet days	%	AFS, WRH
SPI	Standardised Precipitation Index	Measure of drought using the Standardised Precipitation Index on time scales of 3, 6 and 12 months. See McKee et al. (1993) and the WMO SPI User guide (World Meteorological Organization, 2012) for details. Calculated using the SPEI R package.	A drought measure specified as a precipitation deficit	unitless	H, AFS, WRH

Non-core ET-SCI indices (Mistry, 2019)

Short name	Long name	Definition	Plain language description	Units	Sector(s)
DTR	Daily Temperature Range	Mean difference between daily TX and daily TN	Average range of maximum and minimum temperature	°C	
TNx	Max TN	Warmest daily TN	Hottest night	°C	
TXn	Min TX	Coldest daily TX	Coldest day	°C	
TMm	Mean TM	Mean daily mean temperature	Average daily temperature	°C	
TXm	Mean TX	Mean daily maximum temperature	Average daily maximum temperature	°C	
TNm	Mean TN	Mean daily minimum temperature	Average daily minimum temperature	°C	
TX10p	Amount of cool days	Percentage of days when TX < 10th percentile	Fraction of days with cool day time temperatures	%	
TX90p	Amount of hot days	Percentage of days when TX > 90th percentile	Fraction of days with hot day time temperatures	%	
TN10p	Amount of cold nights	Percentage of days when TN < 10th percentile	Fraction of days with cold night time temperatures	%	
TN90p	Amount of warm nights	Percentage of days when TN > 90th percentile	Fraction of days with warm night time temperatures	%	
CWD	Consecutive Wet Days	Maximum annual number of consecutive wet days (when PR ≥ 1.0 mm)	The longest wet spell	days	
R10mm	Number of heavy rain days	Number of days when PR ≥ 10 mm	Days when rainfall is at least 10mm	days	
Rnnmm	Number of customised rain days	Number of days when PR ≥ <i>nn</i>	Days when rainfall is at least a user-specified number of mm	days	
SDII	Daily PR intensity	Annual total PR divided by the number of wet days (when total PR ≥ 1.0 mm)	Average daily wet-day rainfall intensity	mm/day	
R95p	Total annual PR from heavy rain days	Annual sum of daily PR > 95th percentile	Amount of rainfall from very wet days	mm	
R99p	Total annual PR from very heavy rain days	Annual sum of daily PR > 99th percentile	Amount of rainfall from extremely wet days	mm	
Rx1day	Max 1-day PR	Maximum 1-day PR total	Maximum amount of rain that falls in one day	mm	
Rx5day	Max 5-day PR	Maximum 5-day PR total	Maximum amount of rain that falls in five consecutive days	mm	

Sectors: H=Health, AFS=Agriculture and Food Security and WRH=Water Resources and Hydrology.

Mistry, M. N. (2019, February 4). A High-Resolution (0.25 degree) Historical Global Gridded Dataset of Climate Extreme Indices (1970-2016) using GLDAS data. *Supplement to: Mistry, MN (2019): A high resolution global gridded historical dataset of climate extreme indices. Data, 4(1), <https://doi.org/10.3390/data4010041>*. PANGAEA. doi:[10.1594/PANGAEA.898014](https://doi.org/10.1594/PANGAEA.898014)