# IS-ENES3 Milestone M10.3

## Climate indicators/indices(*) and file metadata specifications and tools

*(*) In this report the terms climate index and climate indicator are used interchangeably.*

*Reporting period: 01/07/2020 to 31/12/2021*

Authors:
Lars Bärring (SMHI)

Release date:
19/11/2021

ABSTRACT

This report summarises the work towards an international community standard for climate index metadata and tools for supporting, maintaining and further developing this standard. The work builds on two pillars the Climate and Forecasting (CF) Conventions and a publicly available dedicated github repository holding the information. The information is available in human-friendly format as a spreadsheet file, and in computer-friendly format as a yaml-file.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

# Table of contents



## 1 Objectives

The objective of this work is to establish a standardised metadata description of common climate indicators, and basic tools for developing and managing the metadata description. To the extent possible this should build on the well-established Climate and Forecasting Conventions (CF Conventions) [1], in the following shortened to CF.

## 2 Overview

The term *climate indicator* has no clear definition (the same goes for the term *climate index*). Consequently, it is used in an open-ended fashion depending on the context. We have adopted a tiered approach (Table 1) introduced by the European CLIP-C project [2]. The focus of the work reported here is on tier 1 climate indicators, i. e. those derived from climate data only and designed to inform about the climate drivers relevant for various sectors.

*Table 1. The tiered approach introduced by the European project CLIP-C [2].*

| Tier | Input data | Application area |
|---|---|---|
| 1 | Climate (geophysical) data | Informs about climate drivers |
| 2 | Climate data + data on impacts | Informs about climate impacts |
| 3 | Climate data + impacts data + monetary valuation | Informs about economical ramifications |

The reason for focussing on tier 1 climate indicators is that they are the “classical” indicators that form the backbone of existing collections of climate indicators. Moreover, tier 2 and tier 3 indicators are still in early development and essentially depend on non-climatic data sets and data sources, e.g., a wide range of environmental and national socio-economic statistics. As such tier 2 and tier 3 indicators are unique to the context of the study and/or research context.

This work is based on three components: 1) compilation and review of well-established collections of climate indicators, 2) grouping of climate indicators according to 'anatomy' and computational algorithms, 3) developments towards a consistent description of indicator data using the CF Conventions.

The key tool in this work is the public web repository **clix-meta** [3]. This repository is the central place for organising and handling all the collected information, which in a human-friendly format is stored in a spreadsheet file.

The CF Conventions is a well-established community-lead specification of metadata in climate data files. With respect to climate indicators, it is both permissive and restrictive regarding what and how the metadata have to be organised and formatted. It is *permissive* in the sense that only the bare essential information for understanding what the data represents are mandatory, and that any additional information can be included with few limitations. However, for such free (i.e., non-managed) information there are no rules, which means that it is difficult or intractable to develop common standardised workflows that would depend on this particular metadata information. In practice, it would be difficult for a workflow tool to identify climate indicator data in a file and process it accordingly. Currently, the CF Conventions [1] is *restrictive* in the sense that the managed components are not always well suited to handle climate indicator metadata. In particular the following elements of the CF Conventions machinery is of relevance for detailed description a climate indicator:

- *Standard name* (recommended if available): a controlled vocabulary describing the data variable (in this case indicator) according to certain agreed rules and conventions. A proposal for a standard name is only accepted into the vocabulary after a screening procedure.

- *Long name* (recommended): a short descriptive free text describing the data variable. Typically, this can function as a title for a plot or similar. While this element is not required it is strongly recommended because it serves the dual purpose of providing a succinct description of the data that is understandable to humans, and a semi-standardised title for plots etc.

- *Unit* (required if standard name is used, else recommended): unit of the data variable. In practice information about the unit of the data is compulsory, or the data will be difficult to understand and analyse irrespective of whether it done manually or in an automated workflow.

- *Cell method* (recommended if available): a controlled vocabulary of methods applied to create the climate indicator. Currently, the controlled vocabulary contains a rather limited set of methods, namely: *point* (default), *sum*, *maximum*, *maximum_absolute_value*, *median*, *mid_range*, *minimum*, *minimum_absolute_value*, *mean*, *mean_absolute_value*, *mean_of_upper_decile*, *mode*, *range*, *root_mean_square*, *standard_deviation*, *sum_of_squares*, *variance*.

The current list of cell methods is not (nearly) enough to cover the rich variety of climate indices included in the repository. Nevertheless, in combination with standard names the existing list covers a substantial part of the climate indicators included in the repository. Much of the work is about how best make use of existing CF mechanisms and to interact with the CF community to explore ways to introduce new elements to make the CF Conventions better suited to describe climate indicator metadata.

## 3. Well established collections of climate indicators

The work within IS-ENES-3 is based on earlier work within the previous phase of IS-ENES (IS-ENES-2). From these earlier efforts it was clear that two partly overlapping collections of climate indicator definitions form the backbone of most datasets available. The first one was produced by the CCl/WCRP/JCOMM Expert Team on Climate Change Detection and Indices, *ETCCDI* [4]. This was the taken over and expanded upon by the WMO Expert Team on Sector-specific Climate Indices [5].

New additions to these core collections of index definitions include those used by the European Climate Assessment & Dataset, ECA&D [6], the European projects CLIP-C [2] and INDECIS [7], and Copernicus C3S Sectoral Information System (SIS) products.

The **clix-meta** repository [3] currently (2021-10-29) includes all indices defined by ETCCDI, ET-SCI and CLIP-C, and most defined by ECA&D (Table 2). In addition to these well-established indices there are a few additional indices that are used to explore new types of indices that have been suggested by various stakeholders. It is currently not always possible to provide detailed metadata for all these indices. The reason for this will be discussed in the next two sections.

*Table 2. Summary of the indices currently (2021-10-29) included in the* ***clix-meta*** *repository.*

| Status | ETCCDI | ET-SCI | ECA&D | CLIP-C |
|---|---|---|---|---|
| Ready | 18 | 13 | 23 | 36 |
| Not ready | 9 | 20 | 25 | 16 |
| Total | 27 | 33 | 48 | 52 |

## 4. Development of a consistent description of climate indicator data

Typically, index definitions and index calculations focus on input data having daily resolution. For some indices it is possible to use data having monthly resolution instead of daily resolution. Because the ETCCDI and ET-SCI indices originally were defined for specific limitations of observational data in mind there are two technical considerations to take into account:

- Daily mean temperature is defined to be the arithmetic mean of the daily maximum and daily minimum temperature. The reason for this is that in several countries this how daily mean temperature is – or at least has been – calculated, or that more comprehensive methods for calculating the daily mean temperature vary between countries and/or for different historic periods. For climate model output or reanalysis data the computed daily mean temperature is often used instead.

- For indices based on daily total precipitation only data above 1 mm/day is used. The reason for this is twofold; there is large uncertainty in observations of small precipitation amounts, and in dry climates the number days with zero or only small amounts will totally dominate.

Table 3 gives a summary of the different types of indices according to the 'anatomy' of the index calculation function. In the following the status of the metadata will be summarised for each index type in turn, and then summarised in Table 4.

*Table 3. Brief explanation of the different index types.*

| Index type | Brief explanation |
|---|---|
| Count occurrences | Count of number of days when a constant threshold is exceeded. |
| Count percentile occurrences | Count the number of days when a percentile threshold is exceeded. A constant percentile probability is used to calculate the percentile thresholds for a specified reference period. |
| Spell length | Count the number of days in the longest period of consecutive days exceeding a threshold. |
| Temperature sum | Accumulate over time the temperature exceeding a constant threshold. |
| Statistics | Apply a simple statistic, such as mean, minimum or maximum. |
| Running statistics | Apply a simple statistic to the data under a moving window of N days. |
| Diurnal temperature range | Mean (or other statistic) of the diurnal temperature range. |
| Extreme temperature range | Difference between highest and lowest temperature in a period. |
| Interday diurnal temperature range | Mean of the absolute value of inter-day diurnal temperature difference. |
| Percentile statistic | Similar to *Statistics* but involving percentiles. |
| Precipitation statistics using percentile threshold | Precipitation statistics based on percentile thresholds calculated for a specified reference period. Either count the number of days when the percentile threshold is exceeded, **or** calculate the total precipitation during those days as a proportion of the total precipitation during the period. |
| Other (complex indices) | See text. |

### Count occurrences

These indices generally have well established metadata. Standard names, units and cell methods exists. A quirk however is that while the result is a count of days the actual unit is “1”, i.e., it is unit-less. The reason for this is purely semantic in that a count does not have a unit. On the CF github site there is currently an open ticket regarding refinement of the relevant standard names to more precisely handle some of the ET-SCI indices.

### Count percentile occurrences

Despite that this index type at first sight reminds of “count occurrences” it is a much more complex algorithm. Without going into technical details, the first step is to calculated the thresholds for a constant percentile probability based on a data for a climatological reference period. The next step is to count the number of days exceeding this threshold. The unit is percentage of the total number of days in the analysed period. Currently relevant standard names do not exist, but there is an open issue (ticket) on this subject on the CF github site. Relevant cell method is available (sum).

### Spell length, temperature sum, Statistics, and Running statistics

These indices have well established metadata; standard names, unit, and cell method.

### Diurnal temperature range, Extreme temperature range, Interday diurnal temperature range

These three specific indices do not have standard names, but they have units, and the first two have suitable cell methods. Moreover, there is an open ticket regarding how best to provide relevant metadata for temperature differences and certain other statistically processed temperature variables. As these three indices are very specific and currently do not have variants or alternative that they can be easily mixed up with the lack of standard names is a non-blocking issue.

### Percentile statistic

There is one specific index, “tx95t”, in the ET-SCI list that is quite different from all other indices in that it is the 95th percentile calculated over a number of years (typically 30 years), rather than on an annual or sub-annual basis. This is contrary to three CLIP-C indices that are annual percentiles. Contacts with ET-SCI experts suggest that this index is not frequently requested. Thus, the discrepancy between the two types of definitions remains to be clarified, but this is not a priority.

### Precipitation statistics using percentile threshold

In principle these indices are calculated in the same way as for the “count percentile occurrences” with a major complication that this work has put the limelight on: there is an inconsistency between the different index collections. Basically, there are two categories of index names: r95p and r99p on the one hand, and r75ptot, r95ptot and r99ptot. In both cases the numeric part indicates the percentile probability (cumulative probability) to be used. The resulting index can be expressed in terms of number of days when the percentile threshold is exceeded, or it can be expressed as percentage of the total precipitation that is falls during days when the percentile threshold is exceeded. And there is partial inconsistency between the different index collections, and their reference software implementations, as to which index name is associated with which unit. This should preferably be sorted out to avoid mistakes and inadvertent "comparison of "apples and pears".

### Other (complex indices)

This "group" consists of a diverse set of indices, that can broadly be divided into the following groups

- Well-defined indices that are based on an algorithm that is too complex to be described by the metadata machinery available in CF. Here the strategy is to simply define standard names. Typically. these indices are focusing on a specific sector or climatic impact. This group comprises 11 indices.
- Four closely linked indices that are used to describe the growing season length according to the ETCCDI definition. There is one index each for the start, end, length and degree-days of the growing season.
- Indices designed to quantify cold/warm spells and heat waves based on percentile thresholds. This group, which can be further divided into three subgroups, comprises 18 indices in total.
- Assorted other indices.

All in all, the metadata description for some of these will be limited to find a standard name that will be accepted into the official CF standard name table. For others it may depend on already opened tickets in the CF github site.

*Table 4. Overview of index type per collection and whether metadata is completed. The colour coding is green for indices where metadata is ready, yellow for indices where a clearly identified issue prevents the metadata to be completed, and red is for the complex indices for which a detailed metadata specification is yet to be considered.*

| Index type | ETCCDI | ET-SCI | ECA&D | CLIPC | TOTAL |
|---|---|---|---|---|---|
| count_occurrences | 7 | 9 | 9 | 13 | 38 |
| count_percentile_occurrences | 4 | 1 | 2 | 6 | 13 |
| spell_length | 2 | x | x | 12 | 14 |
| temperature_sum | x | 3 | 2 | 2 | 7 |
| statistics | 7 | x | 10 | 9 | 26 |
| running statistic | 1 | 1 | | | 2 |
| diurnal_temperature_range | 1 | x | x | x | 1 |
| extreme_temperature_range | x | x | 1 | x | 1 |
| interday_diurnal_temperature_range | x | x | 1 | x | 1 |
| percentile statistic | x | 1 | x | 3 | 4 |
| precipitation statistic using percentile threshold | 2 | 2 | 2 | 1 | 7 |
| Other (complex) | 3 | 16 | 21 | 6 | 46 |
| TOTAL (ready / not ready) | 18 / 9 | 13 / 20 | 23 / 25 | 36 / 16 | 90 / 70 |

## 5. Overview of the *clix-meta* repository

The **clix-meta** [3] github repository contains two main components: a spreadsheet file (in a widely accepted format) holding the index definition information, and a set of python functions that are set up to convert the human-friendly spreadsheet information to computer-friendly yaml-files. This is done automatically every time the spreadsheet file is updated.

The spreadsheet file contains the following five tables:

- *README*: brief overview of the different tables.
- *index_definitions*: The main table holding metadata for the individual indices.
- *variables*: Metadata relevant for the input variables, including known aliases.
- *index_functions:* Specifications relevant for the external calculation code snippets
- *ECA&D:* Table of ECA&D indices.

The dual purpose of the *index_definitions* table is to specify the relevant CF compliant metadata describing each index (i.e., standard name, long name, cell methods, and unit), and provide additional information that is akin to CMIP data requests and CORDEX Archive Specifications, which provide information for defining the Data Reference Syntax (DRS) for climate index information.

In particular, a DRS for climate indices may depending on the input data use the same DRS as defined for CMIP, CORDEX or any other input data set. For CMIP and CORDEX the following DRS elements are proposed:

- **variable name** will be the climate index acronym, i.e., the VarName column of the *index_definition* table.
- **product** will be set to “climate-index”.
- **frequency** will reflect the temporal resolution of the data, i.e. “mon”=monthly, “sem”=seasonal, and “yr”=yearly. The *index_definition* table specifies a default for each index (it is anticipated that this can be changed by users).

Other DRS elements depend on the input data source.

The **clix-meta** repository has an active issue tracking function for interaction with users, e.g., for asking questions, requesting new index definitions or other features, or for reporting bugs.

## 6. Difficulties overcome

There have been three main difficulties to handle to be able to reach to the point reported in this document:

i Users and producers of climate index data are typically not very interested or knowledgeable in the highly technical and formalistic aspects of climate indices, neither from a detailed algorithmic perspective or from a metadata standards (CF Conventions) perspective. This is however fundamental for producing a metadata description of an individual index and even more when developing a standard encompassing a range of widely different indices. Due to this the interaction with users and stakeholders is slow and often not providing the concrete input needed for a time-efficient development of the metadata standard. Instead, the required understanding of stakeholder needs has to be patched together little-by-little from several sources and informal interactions. However, this is the state of affairs that has to be accepted and worked with. The situation is however expected to improve for two reasons: i) the c**lix-meta** repository is gaining momentum and linked to two popular index calculations tools (*icclim*[9] and *xclim*[10]), ii) identification of the inconsistency among precipitation indices is of interest to users, which was evident from the recently concluded IS-ENES Workshop on climate indices - Eastern Europe perspective [8].

ii During the IS-ENES3 project time the interest and engagement for climate index metadata from the CF Community has steadily increased. Moving the whole CF web site and community interaction tools to the CF github site has substantially helped to initiate and maintain focussed discussions in github issues.

iii Direct and indirect effects of the COVID-19 pandemic on general working conditions and efficiency in interaction between a wide spectrum of people and groups.

## 7. Next steps

The following concrete actions are planned:

- Engage with the CF community to move the relevant open issues to agreement and acceptance. This will lead to that some about 15 currently “Not ready” indices be become “Ready”.
- Discuss in a smaller group possible standard names for a number of the complex indices. This is the first step towards getting such standard names into the CF standard name table. This will lead to that about 10 currently “Not ready” indices to become “Ready”.
- Engage with software developers (in particular *icclim*[9] and *xclim*[10]) to facilitate smooth integration of the **clix-meta** repository into software tools.
- Together with relevant parties explore the possibility to resolve the current inconsistency among indices based on precipitation statistic using percentile threshold. This might however prove difficult because of legacy and backward compatibility.
- Continue interaction with representatives (i.e., Instituto de Física de Cantabria, Spain, and also representatives of the IPCC Task Group on Data) of the IPCC Interactive Atlas [11] regarding how **clix-meta** may facilitating publication of the climate index datafiles underlying this online map tool.

## References


[1] https//cfconventions.org
[2] http://www.clipc.eu/content/content.php?htm=46
[3] https://github.com/clix-meta/clix-meta
[4] https://www.wcrp-climate.org/etccdi-publications-and-docs
[5] https://climpact-sci.org/about/project/
[6] https://eca.knmi.nl//indicesextremes/index.php
[7] http://indecis.eu/
[8] https://is.enes.org/documents/milestones/m3-5-workshop-on-climate-indices-eastern-europe-perspective/view
[9] https://github.com/cerfacs-globc/icclim
[10] https://github.com/Ouranosinc/xclim
[11] https://interactive-atlas.ipcc.ch/