IS-ENES3 Milestone M10.4

Climate indices and file metadata specifications and tools

Update of M10.3

Reporting period: 01/01/2022 – 31/03/2023

Author: Lars Bärring (SMHI)

Release date: 23/02/2023

ABSTRACT

This report updates work reported in Milestone 10.3 on 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 the CLIX-META public GitHub repository. Progress have been made along two lines: i) compilation of a detailed picture of inconsistencies in well-established definitions of indices present in software tools and in widely used datasets. And ii) to work with the Climate and Forecasting (CF) Conventions community to expand the Conventions to cater for metadata requirements to correctly describe climate indices. The first part provides the platform for disentangling the inconsistencies and developing the necessary metadata machinery. The second part is focussing on expanding CF metadata elements such as standard names, units, and cell methods. In addition to these technical achievements there is an important social dimension in building a common understanding across communities. There is a growing insight in the user community that metadata are important, and likewise in the metadata community that climate indices are an important source for climate data that needs to be supported by the CF Conventions.

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 overall objective of this work is to establish a standardised metadata description of common climate indicators[1], 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. This Milestone M10.4 report documents the progress since the release of Milestone M10.3 [2] in November 2021, which provides the rationale for the work, and more complete overview and introduction. Specific progress since M10.3 is detailed in paragraphs starting with the keyword "Progress" in boldface.

2. Description of work: Methodology and Results

The methodology 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, and in computer-oriented text format.

Table 1 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.

1 In this report the term "climate index" (pl. "climate indices") are used, which is equivalent to the alternative term "climate indicator". The former is prevalent in the more technical and data-oriented literature, which is the primary target context for this report. The latter is becoming dominant in user and stakeholder communities. Irrespective of which term is used they refer to the same concept.

Table 1. Brief explanation of the different index types.

Index category	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.

Category 1: 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. While this is semantically correct it is counter-intuitive from a [end-]user perspective.

Category 2: 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 on this subject on the CF GitHub site. Relevant cell method is available (sum).

Progress regarding categories 1 and 2: Regarding these two groups of indices progress on metadata specification is complicated by the lack of consistency between different sources for indicator definitions. (Table 2). And this carries over to inconsistencies in the data produced by different software tools. It was essential to clarify these inconsistencies in detail because the metadata is crucially dependent on the unit of the data. While Table 2 only relates to the second category of indicators, *Count percentile occurrences*, because the aim is to as far as possible unify

the metadata of these two categories. And the current *CF standard name* that applies to the first group is associated with a *CF canonical unit* (i.e. "unitless), which is the same as the canonical unit for the percent. That is, by strictly following the CF Conventions regarding standard names and their canonical units it is easy to mix up inconsistent data. For example the data value 40 produced by the software tool Climpact2 (unit: percent of a year) from one data source can be confused with the data value 40 produced the software tool XCLIM (unit "1" for count of days) because the canonical units are the same, even though their actual units are "%" and "1".

Table 2. Different sources provide inconsistent definitions of category 2 indices.

Source	Ref.	Units	Text indicating units
WMO/TD-No. 1500	[4]	days	"Count of days where ..."
ETCCDI legacy webpage	[5]	%	"Percentage of days when..."
ECA & D	[6]	days	"Counted is the number of days..."
EU–INDECIS	[16]	days	"Total numbers of days with..."
Climpact2 (May 2016) Manual	[7]	%	"Percentage of days where..."
XCLIM	[8]	days	"Number of days with..."
ICCLIM	[9]	% or days	Unit is selected by the users

Category 3: Spell length, Temperature sum, Statistics, Running statistics

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

Category4: 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 issue 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.

Progress regarding category 4: There has been an extensive conversation within the CF community regarding how best to introduce robust and specific metadata elements for describing temperature differences without introducing breaking changes to the Conventions. Currently, several alternatives are being explored. This conversation has also spawned a general overhaul of how units are described in the CF Conventions document. A substantial part of this was concluded during the CF2022 workshop 13-15 September in Santander.

Category 5: 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.

Category 6: 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 oranges.

Progress regarding 6: Substantial effort was used trying to get an overview of the inconsistencies in index definitions. A series of virtual conversations between key software developers and data producers were organised in late 2021 and early 2022. Based on this it was possible to get an overview of the differences in index definitions, see Table 3. Possible ways forwards towards more harmonised definitions were discussed. However, a complicating factor is that software tools and datasets have their well-established user communities, which means that the necessary changes are likely to be received as disruptive. But this also underlines the need to have correct and exhaustive metadata that enable users and software tools alike to determine whether superficially compatible datasets actually are so.

Table 3. Overview of the inconsistency in definition of percentile-based precipitation indices.

Source	Ref.	Source type	r75p	r95p	r99p	r75pTOT	r95pTOT	r99pTOT	r75pFRAC	r95pFRAC	r99pFRAC
ETCCDI legacy webpage	[5]	Online document	--	--	--	--	Total precipitation amount from days above percentile (mm) **[Note 1]**		--	--	--
WMO-TD 1500	[4]	Document	--	(Number of days when precipitation is above percentile **[Note 3]**)		--	Total precipitation amount from days above percentile (mm) **[Note 2]**		--	--	--
climdex.pcic	[10]	Software (R)	--	--	--	--	Total precipitation amount from days above percentile (mm)		--	--	--
Climpact2 (May 2016) Manual	[7]	Software (R)	--	Total precipitation amount from days above percentile (mm)		--	Fraction of total precipitation amount from days above percentile (%)		--	--	--
Albert Klein Tank's list" [Note 4]	[11]	Online document	Number of days when precipitation is above percentile			Fraction of total precipitation amount from days above percentile (%)			--	--	--
Alternative table comparing indices used in different projects [Note 4]	[12]	Online document	Number of days when precipitation is above percentile			Fraction of total precipitation amount from days above percentile (%)			--	--	--
HADEX3 (based on CLIMPACT software)	[13]	Dataset	--	--	--	--	Fraction of total precipitation amount from days above percentile (%)		--	--	--
Donat et al. 2013 (Table 1)	[14]	Document	--	Total precipitation amount from days above percentile (mm)		--	Fraction of precipitation amount from days above percentile (%)		--	--	--
ECA & D	[6]	Dataset	Number of days when precipitation is above percentile			Fraction of total precipitation amount from days above percentile (%)			--	--	--
EOBS and NGCD by Copernicus (based on GRIDCLIMIND sof tware)	[15]	Dataset	--	--	--	Total precipitation amount from days above percentile (mm)			Fraction of total precipitation amount from days above percentile (%)		
ICCLIM	[9]	Software (Python)	Number of days when precipitation is above percentile (either expressed as number of days, or as % of the total number of *wet* days in the period)			Fraction of precipitation amount from days above percentile (%)			--	--	--
XCLIM	[8]	Software (Python)	Number of days when precipitation is above percentile			Fraction of precipitation amount from days above percentile (dimensionless).			--	--	--
EU Project INDECIS	[16]	Dataset and software (R)	--	--	--	--	Fraction of precipitation amount from days above percentile (%) **[Note 5]**		--	--	--
ECA & D	[6]	Software (R)	--	--	--	Fraction of precipitation amount from days above percentile (%)			Fraction of total precipitation amount from days above percentile (%)		
GRIDCLIMIND	[17]	Software (R)	--	--	--	Fraction of precipitation amount from days above percentile (%)			Fraction of total precipitation amount from days above percentile (%)		
IPCC/AR6/WGI/Annex VI (IPCC Interact ive Atlas)		Document	--	Total precipitation amount from days above percentile (mm)		--	--	--	--	--	--

General	* Wet days are defined as days having at least 1 mm of accumulated precipitation. * Percentile thresholds are based on wet days only. * For indices showing "Fraction of precipitation amount above percentile (%)" the denominator is total precipitation during wet days (i.e. precipitation during non-wet days are not included).
Note 1	The definition does not explicitly provide the units, which instead is inferred from the text: "Annual total PRCP when RR > 95p." i.e. there is no mentioning of "percentage of" or "fraction of".
Note 2	The definition does not explicitly provide the units, which instead is inferred from the text: "Annual total PRCP when RR > 95p." i.e. there is no mentioning of "percentage of" or "fraction of". Moreover, the text (pp. 27) discusses analysing the ratio R95pTOT / PRCPTOT, which is the same as r95pTOT expressed as a percentage.
Note 3	The report does not include the indices r95p and r99p as such, but states on p.15: "*For instance, the threshold used in the very wet days index R95p (the number of days with rainfall above the 95th percentile of daily accumulations) often refers to larger amounts in wet climates than dry climates.*"
Note 4	The lists includes related indices and alternative names (acronyms) used by various projects of the time: R90T=r90pTOT expressed as %, R90N number of days above 90 percentile,
Note 5	The INDECIS project created new names instead of using well-established ones: PVWD=r95pTOT and PEWD=r99pTOT, both expressed as %

Category 7: Other (complex indices)

This "group" consists of a diverse set of indices. The metadata description for some of these will be limited to finding a standard name that will be accepted into the official CF standard name table. For many it will depend on progress regarding CF issues identified for the other index categories.

Progress regarding 7: For some of the indices it is now clear that further development of metadata specification depends on that open issues pertaining to other index categories are first solved. As there are ongoing conversations regarding several of these issues we expect that the situation will improve.

Results: clix-meta status

Table 4 summarises the metadata maturity of the different index categories detailed in the previous sections. While substantial progress has been made since M10.3, the still open issue in the CF/discuss repository prevent the "yellow" categories to fully covered by CF compliant metadata. In the period after M10.3 the clix-meta repository has seen two main releases, and one bug-fix release. The current version is 0.5.

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 clearly identified issue(s) prevents the metadata to be completed. Nevertheless they can regarded as "ready", although with limited metadata support. And red is for the complex indices for which a detailed metadata specification is in development without clearly identified issues.

Index type	ETCCDI		ET-SCI		ECA&D		CLIPC		TOTAL	
count_occurrences	7		9		7		13		36	
count_percentile_occurrences	4		1		2		6		13	
spell_length	2		x		2		12		16	
temperature_sum	x		3		2		2		7	
statistics	4		5		9		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		4		5	
precipitation statistic using percentile threshold	2		2		2		1		7	
Other (complex)	3		17		17		5		42	
TOTAL (ready / not ready)	21	3	22	17	26	17	47	5	115	42

3. Difficulties overcome

There are mainly two difficulties have been dealt with:

- While the inconsistencies in index definitions have been known since early days (at least a decade) the full implications have not been understood by all involved parties until the virtual conversations were organised. While the problem as such has not been solved, there is now better understanding of the negative implications. Moreover, this overview forms the fundamental basis for developing the necessary metadata standard.
- Some metadata elements that are important for describing climate indices did not exist in the CF Conventions. Because these elements potentially involve central aspects and concepts of the Conventions proposals towards that end has sparked off a vivid discussion with many views, which is necessary to not overlook some aspect. This has contributed to elevate and spread the insight that climate indices is a "data type" that should be included in and catered for by the CF Conventions, at the same time this has also slowed down progress.

4. Next steps

The work will continue, partly within the framework of the Horizon project CLIMATEUROPE2.

5. References

[1] https//cfconventions.org
[2] https://is.enes.org/milestones/
[3] https://github.com/clix-meta/clix-meta
[4] https://library.wmo.int/index.php?lvl=notice_display&id=138#.Y_ZW3kjMLJE
[5] http://etccdi.pacificclimate.org/list_27_indices.shtml
[6] https://eca.knmi.nl//indicesextremes/index.php, https://www.ecad.eu/indicesextremes/indicesdictionary.php
[7] https://climpact-sci.org/about/project/
[8] https://github.com/Ouranosinc/xclim
[9] https://github.com/cerfacs-globc/icclim
[10] https://github.com/pacificclimate/climdex.pcic.ncdf
[11] http://etccdi.pacificclimate.org/docs/ETCCDMIndicesComparison1.pdf
[12] https://www.ecad.eu/documents/ETCCDMIndicesComparison.pdf
[13] https://www.metoffice.gov.uk/hadobs/hadex3/
[14] https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50150
[15] https://surfobs.climate.copernicus.eu/userguidance/indicesdictionary.php
[16] http://www.indecis.eu/docs/Deliverables/INDICIS-list_4.2.pdf
[17] https://github.com/ECA-D/gridclimind
[18] https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Annex_VI.pdf