Climate change REsilience framework for health SYStems and hospiTALs

DA2.1 - Description of downscaling models & methods	
Contractual Delivery Date: 31/01/2022	**Actual Delivery Date**: 28/02/2022
Type: Report	**Version**: v0.2
Dissemination Level: Public Deliverable	

Preparation Slip			
	Name	**Partner**	**Date**
From	A. Sfetsos / S. Karozis	NCSRD	25/2/2022
Reviewer			
Reviewer			
For delivery			

Document Log			
Issue	**Date**	**Comment**	**Author / Organization**
V0.1	25/02/2022	Draft version	A. Sfestos - S. Karozis / NCSRD
V0.2	28/2/2002	Final version	A. Sfestos - S. Karozis / NCSRD

Executive summary

This Deliverable describes the existing climate sources that will be utilized within the framework of the LIFE-RESYSTAL project. The present DA2.1 lays the first foundation of the project's climate risk assessment process as it describes the climate information that is available and will be employed within each pilot hospital. It will contain future scenarios of priority climate hazards on a very high-resolution scale, and thus describe how this information will be collected, processed and presented.

In this action the project partners will describe both extreme climate events and also long term changes due to climate change and identified through related IPCC Representative Concentration Pathways (RCP), introduce common climate data repositories, such as Copernicus Climate Service, CMIP5 / CMIP or EURO-CORDEX, and estimate if there is a need to apply a downscaling method. Depending on the complex characteristics of the local topography, downscaling could be applied to generate high quality data that are able to capture the local scale characteristics of the hospital location.

Table of content

Table of abbreviations	
Abbreviations	**Meaning**
CMIP5	Coupled Model Intercomparison Project Phase 5
GCM	Global Climate Model
RCM	Regional Climate Model
CORDEX	Coordinated Regional Climate Downscaling Experiment
EM-DAT	Emergency Events Database
GP	Generalized Pareto
CDF	Cumulative Distribution Function
GEV	Generalized Extreme Value
ESGF	Earth System Grid Federation
IPCC	Intergovernmental Panel on Climate Change
AI/ML	Artificial Intelligence / Machine Learning
GRIB	GRIdded Binary or General Regularly-distributed Information in Binary form
NetCDF	Network Common Data Form
FA	Fichier ARPEGE
NWP	Numerical Weather Predictions
NCAR	National Center for Atmospheric Research
NCL	NCAR Command Language
NCO	netCDF Operators

Introduction

Within LIFE-RESYSTAL, as in all climate change adaptation related projects, climate parameters are the driving force behind any conducted assessment. The climate risk assessment framework, developed within action A of the project, will leverage the vast amount of existing knowledge generated up to-date in the climate research domain. It will place primary emphasis on the processing/transformation of existing climate information to be suitable for supporting the project objectives and the local toolbox (Action C). Furthermore, LIFE-RESYSTAL shall provide all necessary arrangements and interfaces so that existing climate information is ingested into the project's platform (Action C1/C2).

The climate information pertinent to the objectives of LIFE-RESYSTAL will feed into the following components of the LIFE-RESYSTAL generic framework (Figure 1)

- Provide an estimation of the likelihood of the climate induced risks to infrastructures,
- Contribute to the identification of the climate health infrastructure thresholds and to the overall assessment of the vulnerability of interconnected health infrastructure assets to climate risks.

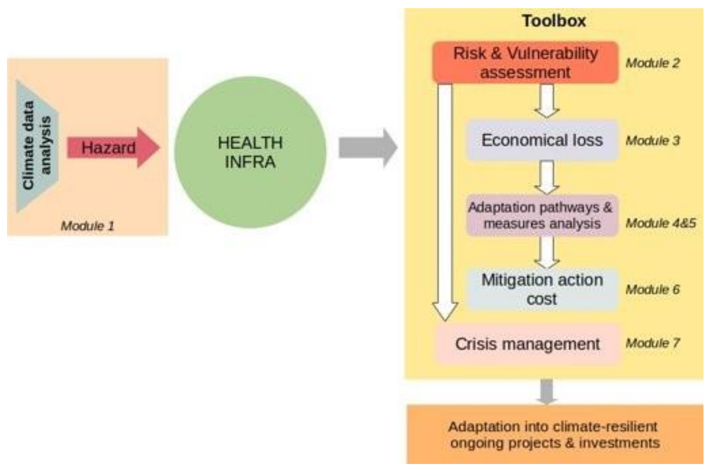

Figure 1: High-level illustration of Local Toolbox design

Climate change related effects have been divided into (Table 1) two different categories:

1. Climate drivers, which are the direct output of simulation models (GCM/RCM, seasonal forecasting model) and observation networks
2. Climate hazards, which are direct consequence of climate drivers, are modelled using post-processing algorithms of the climate drivers and new simulation models from climate simulation models.

Table 1: LIFE-RESYSTAL climate parameters and their interconnections

Climate drivers	Climate hazards
Temperature	Heat waves, cold snaps
Precipitation (rain / snowfall) - humidity	Floods

Winds	Forest Fires
Cloud / fog	Droughts
Solar radiation	Earth movement caused by climate drivers such as rain (landslide, erosion, avalanches)
Sea level rise	Heat Stress (health and safety related)
Ice , frost	
Strom surges, waves	

Climate Data Analyses needed within LIFE-RESYSTAL

Estimation of likelihood/probability of occurrence is a core element of the LIFE-RESYSTAL Risk Assessment Framework (DA2.2), and involves the collection and processing of all related climate information from available sources (historical, simulation, satellite), in order to generate:

1. An estimation of the probability of occurrence of the hazard under the specific scenario.
2. All necessary data that are needed to estimate the impact of the hazard to the infrastructures.

LIFE-RESYSTAL risk assessment could utilise any type of climate information on a temporal and spatial scaling (depending on the requirements of the scenario under examination) using multiple sources of climate data such as:

- historical observations
- climate models (e.g. CORDEX, CMIP5)
- numerical weather prediction models
- satellite data (e.g. Copernicus)
- fused data

The climate data processing is presented in the following sections, and the workflow presented in Figure 2.

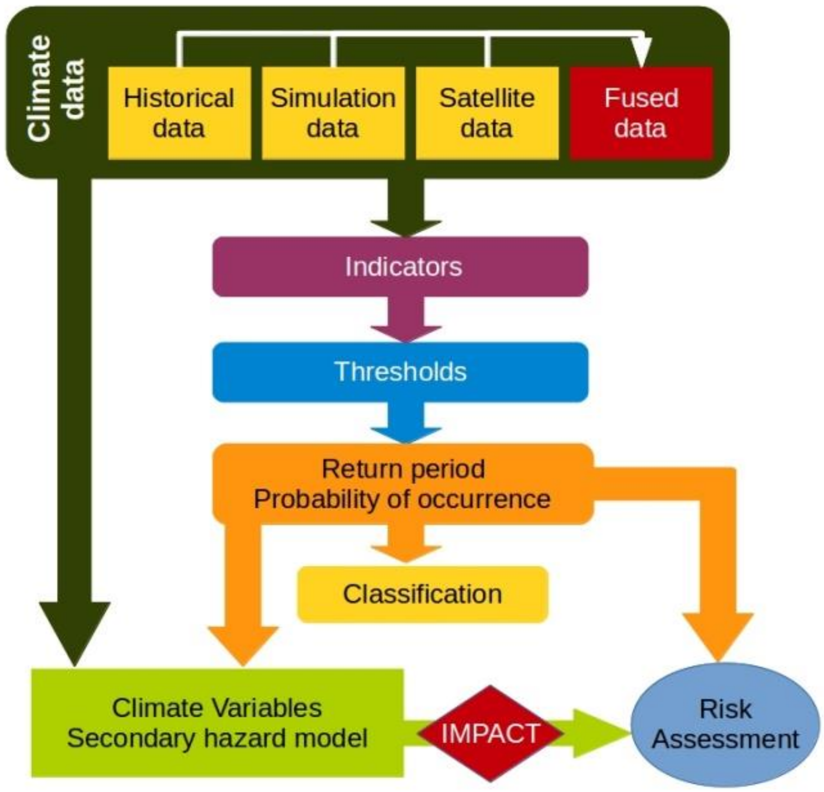

Figure 2: Workflow of the initial procedure of risk assessment

The following example relates to the risk of heat wave, a prolonged period of excessively hot weather, which may be accompanied by high humidity. In Europe, one of the most severe events of extreme temperature/heat wave occurred in 2003 and cost the lives of more than 50,000 people throughout Europe. (EM-DAT).

Table 2: Climate data processing in the framework of LIFE-RESYSTAL Risk model with worked examples

	Example: **Heat Waves**
Collection of climate data from existing databases Collection from available databases climate historical or predicted or processed data, depending on the problem to solve, that are used to calculate the appropriate indices for a certain period and place of interest. **Databases:** ECA&D, CORDEX, CMIP5, etc. **Models and Tools:** GCM, RCM, ESD, etc.	Temperature Humidity
Indicators Indicators measure the actual status of the environment before, during or after an event and serve as a reference status or as a signal for environmental/climate change over time (qualitative or quantitative). Indicators are referring directly to	Temperature Heat Index - Humidex

climate parameters related to the risks or to climate indices that give measure of a risk appearign or not.	
Thresholds Represent quantitative critical values derived from the examined scenario. So it is important to identify where there is a likelihood of unsustainable trends of certain indicators related to environmental issues that show threshold phenomena. These thresholds may be related not only with extreme phenomena (floods, fires, extreme weather events), but to mean climate values, standard deviation of a variable etc., depending on the assessed scenario.	HI > 54 °C
Return period / Probability of occurrence Based on the threshold and the indicators that have been specified, and also the processed data, we calculate the probability of occurrence of the risk scenario or its return period. A Return level with a return period of T = 1/p years is a high threshold x(p) whose probability of exceedance is p (likelihood of rare events).	1:200 yr
Classification The levels of Likelihood are defined by the internationally accepted descriptive terms, classified into a set of five categories.	Very Low –Very rare Low Medium High Very high-Very Likely
Climate variables/ Secondary hazard model Collection of climate variables per case study for further processing or as input data in the secondary hazard model (fire, flood model etc.)	Temperature Humidity
Impact Input of above previous processed data for the impact model	Temperature Humidity

The timescales involved may also be linked to the expected annualized probability of an event (return period), frequently used as an indicator value for the expected probability of occurrence. As a numerical example, the value of 0.5% annual probability is translated as an event occurring 1 in 200 years, which is a higher value of appearance than a 1 in 1000 (0.1%) annual probability event. This concept is also related to the likelihood component of risk resulting in the “cumulative probabilities” of an infrastructure being exposed to a predefined event, and therefore directly linked to the design standard. As a numerical example, for any infrastructure asset with a 100 year design life, there is a 63% cumulative probability of seeing the 100 year snowfall, which is a high enough probability that this event should be considered in the design phase.

Changes in extreme events may be difficult to detect locally, even when powerful methods based on extreme value analysis theory are used. The Generalized Extreme Value (GEV) distribution was introduced into meteorology by (Jenkinson, 1955) and is used extensively to model extremes of natural phenomena such as precipitation (Gellens, 2002), temperature (Nogaj, et al., 2007) and wind speed (Coles and Casson (1999), Walshaw (2000)). It is possible to account for non-stationary conditions (climate change) using extreme value analysisas described in the following paragraphs.

In cases where examined data exhibit homogeneous climate characteristics i.e. stationarity, return periods may be estimated through extreme value theory. Under this examination all parameters of the GEV distribution are homogeneous across the region, or that the scale and shape parameters are homogeneous. Such an assumption would enable records from multiple stations or historical data to be combined to form a larger data sample. The spatial pooling approach has its origin in hydrology where it is known as regional frequency analysis. This approach is most effective for variables, such as precipitation, which have short "decorrelation" distances (that is, where the correlation between observations at different locations falls off quickly as the distance between stations increases).

Spatial pooling can also be done in other ways, such as by averaging parameter estimates from nearby locations. Parameter estimates based on the pooled information across the region are generally less uncertain than those from the data of individual records because the same amount of information is used to estimate a smaller number of parameters. This also leads to a reduction of uncertainty of the estimated extreme quantiles of the distribution.

According to WMO (2009), the best way to determine the return period from a specific climate parameter should comply with the following procedure:

Preparation of data series (in particular observations) for the analysis of extremes

The preparation of data series must include long and quality-controlled daily observational series to evaluate the intensity and frequency of rare events that lie far in the tails of the probability distribution of weather variables (temperature or precipitation) such as the 20-year return value (an event whose intensity would be exceeded once every 20 years, on average, in a stationary climate). T-year return value is the $(1-1/T)^{th}$ quantile of the GEV distribution (Rootzen & Katz, 2013). In some engineering applications, such analysis requires estimation of events that are unprecedented in the available record, say events that occur once in a hundred or thousand years (extreme quantiles of the statistical distribution), while the observation series may be only about 50 years long). Also, continuous data series with at least a daily time resolution are needed to take into account the sub-monthly nature of many extremes.

The concepts of return level and return period are commonly used to carry information about the likelihood of rare events such as floods. A return level with a return period of T = 1/p years is a high threshold x(p) (e.g., annual peak flow of a river) whose probability of exceedance is p. For example, if p = 0.01, then the return period is T = 100 years.

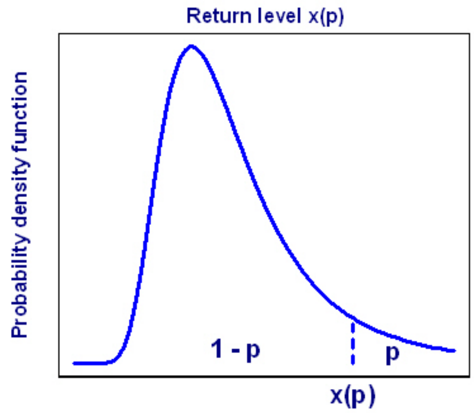

Figure 3: Probability density function.

Two common interpretations of a return level with a return period of T years are:

(i) Waiting time: Average waiting time until next occurrence of event is T years

(ii) Number of events: Average number of events occurring within a T-year time period is one

Utilization of descriptive indices and extreme-value theory to evaluate extremes

Extreme value theory complements the descriptive indices in order to evaluate the intensity and frequency of rare events that lie far in the tails of the probability distribution of weather variables. Two general methods can be used. One method, referred to as the "peaks-over-threshold" or POT method, under suitable conditions, and using a high enough threshold, extremes identified in this way will have a generalized Pareto, or GP, distribution, and the second more generally used method based on an explicit extreme value theory is the so-called "block maximum" method. In this method, one considers the sample of extreme values obtained by selecting the maximum (or in some cases, the minimum) value observed in each block (usually year or season). Statistical theory indicates that the GEV distribution is appropriate for the block maxima when blocks are sufficiently large.

In terms of the tail of a distribution, the corresponding theorem states that the observations exceeding a high threshold, under very general conditions, are approximately distributed as the generalized Pareto (GP) distribution. This distribution has three types: exponential, Pareto, Beta (Figure 3).

The **cumulative distribution function (CDF) of the GEV** distribution is:

Formula 1

$$F(x;\mu,\sigma,\xi) = \begin{cases} \exp\left\{-\left[1+\frac{\xi(x-\mu)}{\sigma}\right]^{-\frac{1}{\xi}}\right\}, & \xi \neq 0 \\ \exp\left\{-\exp\left[-\frac{x-\mu}{\sigma}\right]\right\}, & \xi = 0 \end{cases}$$

Where three parameters, ξ, μ and σ represent a **shape, location, and scale** of the distribution function, respectively. Note that σ and 1 + ξ(x-μ)/σ must be greater than zero. The shape and location parameter can take on any real value. The shape parameter ξ affects the support of the distribution. More specifically, when ξ = 0, the GEV distribution is the Gumbel distribution ((Gumbel, 1958), used extensively in hydrology, meteorology and engineering). with support R; when ξ > 0, it corresponds to the Fréchet distribution and when ξ < 0, it corresponds to the (reversed) Weibull distribution families (Figure 4).

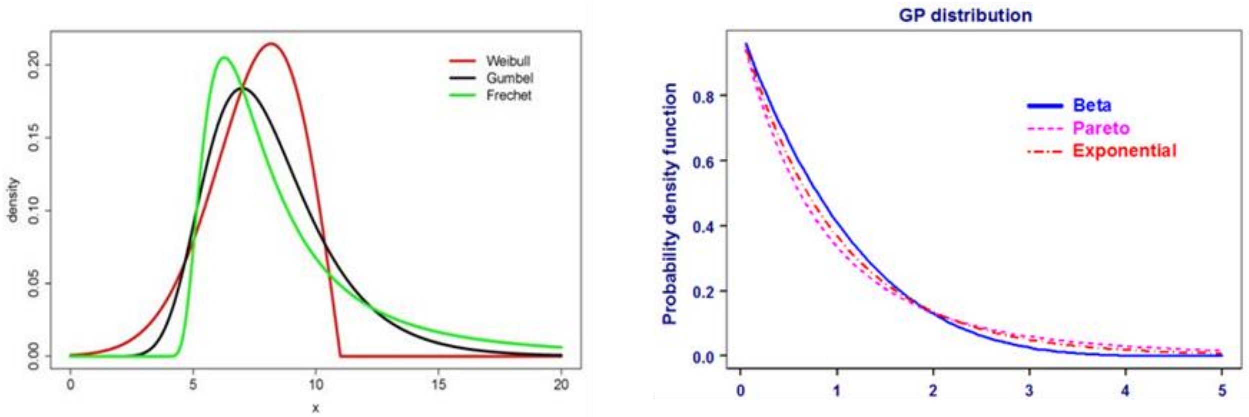

Figure 4: GEV distributions and POT distributions.

When comparing observations and model output, it is important to know what scales are represented by the observational data sets and how they might differ from model output to avoid misinterpretation. Global climate models are not yet able to provide scenarios with sufficient detail at the regional and local scale for many applications.

Their coarse spatial resolution affects in particular the projections for changes in extremes because extremes are often smaller in extent than the effective spatial resolution of the models. For this reason, downscaling (or regionalization) of global climate model projections using regional climate models (nested in the global models) or statistical techniques provides additional useful information (Giorgi, 2008).

The application of this classic theory assumes that time series are stationary. Adjusted techniques are recommended when there are indications for non-stationarity.

Limitations and caveats

The use of the GEV analysis must be performed with following limitations considered:

1. All recorded time-series contain sampling errors which can be reduced by the proper site sitting and continuous insurance of the high data quality.
2. Since some of the extreme parameters are usually defined on an annual time scale (e.g., annual maximum daily precipitation amount, or annual maximum hourly wind speed), the length of the time-series should include several decades.
3. For some applications, data should be separated by the underlying mechanisms. For example, separate return periods of the extreme precipitation amounts can be estimated from the convective or large scale (e.g. fronts) extreme precipitation events.
4. GEV distribution is correct distribution for the largest quantities in an infinite population of an independent variable. For example, all wind speeds at 10-m at some location define the population, while annual maximum values define largest quantities. In practice, observations are finite in length and care must be taken to ensure the independence.

Trend calculation and other statistical approaches for assessing changes in extremes, mean values and other statistical metrics

Trends are the simplest component of climate change and provide information on the first-order changes over the time domain considered. This implies that the physical mechanisms behind the detected trends remain unknown. The calculated trends represent changes that can be due to natural internal processes within the climate system and/or external forcing, which can either be natural, such a solar irradiance and volcanic aerosols, or anthropogenic, such as greenhouse gases.

It is also possible to indicate a definitive trend in the mean value of a climate variable for a specific period that should be taken into consideration because it exceeds a critical value and may cause important environmental consequences.

It is important to test the goodness-of-fit of the fitted distribution (e.g. Kharin & Zwiers (2000)) and to assess the uncertainty of the estimates of the distribution's parameters by calculating standard errors and confidence intervals for these estimates. The latter can be done in a relatively straightforward way when the distribution has been fitted by maximum likelihood because in this case the underlying statistical theory provides expressions that generally give good approximations for these quantities (for example 90%, 95% and 99%).

Although very long period return values can be calculated (for example, once-in-thousand-year levels) from the fitted distribution, the confidence that can be placed in the results may be minimal if the length of the return period is substantially greater than the period covered by the sample of extremes. Estimating return levels for very long return periods is prone to large sampling errors and potentially large biases due to inexact knowledge of the shape of the tails of a distribution. Generally, confidence in a return level decreases rapidly when the period is more than about two times the length of the original data set.

The least square method is one of the approaches for trend estimation; nevertheless this method may be sensitive to individual values, such as a single outlying observation that lies either near the beginning or the end of the available data record. Such observations have "high leverage", meaning that the fitted trend can be strongly affected by their inclusion or exclusion from the data record. In such instances, a non-parametric method may therefore be more statistically robust because the indices generally have non-Gaussian distributions. For instance, it is possible to use Kendall's Tau (Kendall, 1938), which measures the relative ordering of all possible pairs of data points, where the year is used as the independent variable and the extreme index as the dependent variable.

The probability of detecting a trend in any time series depends on the trend magnitude, the record length, and the statistical properties of the variable of interest, in particular the variance. A trend is said to be detected when a test of the null hypothesis that no trend is present is rejected at a high significance level, such as five per cent or one per cent. (Frei & Schär, 2001) show that for precipitation, there is only a one-in-five chance of detecting a 50 per cent increase in the frequency of events with an average return period of 100 days in a 100-year record.

Climate Data Typology

The climate models' products include simulations of regional climate and global climate models (RCMs and GCMs). The relationship between RCMs and GCMs, are discussed in several review papers: Giorgi and Mearns (1991), McGregor (1997), Giorgi and Mearns (1999), Wang et al. (2004), Laprise et al. (2008) and Rummukainen (2010, 2015).
For the purpose of the LIFE-RESYSTAL project, the use of the latest generation of the RCM simulations

over Europe will be employed. These simulations are available on the approximately 12.5 km grid spacing and for the climate projection use the latest IPCC scenarios: RCP2.6, RCP4.5 and RCP8.5. They are available through the ESGF system and WCRP CORDEX project (EURO-CORDEX initiative) or alternative portals and ftp servers run by the specific RCM groups. Also, results and description of the several other projects and initiatives that provide access to corresponding climate model simulations are presented. The dominant data format of all products is NetCDF (both NetCDF3, and NetCDF4 classical and compressed), and should be related with the LIFE-RESYSTAL platform accordingly.

Main Climate Data Repositories

The providers of climate, in situ and satellite environmental data in Europe are the national meteorological services joined in EUMETNET, European Space Agency (ESA) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). There are several other organisations that facilitate free and unrestricted exchange of meteorological and other environmental data and information, products and services in real- or near-real time on a global and European scale, like World Meteorological Organisation (WMO), European Environmental Agency (EEA) and European Centre for Medium Range Weather Forecast (ECMWF).

Table 3: Overview of the potential producers and providers of observations and gridded data.

Organisation / Project	data access
National meteorological services	WMO-synop observations, www.ogimet.com EUMETNET-National meteorological services
EC INSPIRE	INSPIRE geoportal
COPERNICUS	http://www.copernicus.eu/main/data-access http://climate.copernicus.eu/
JRC/European Flood Awareness System (EFAS)	https://www.efas.eu/
JRC/European Drought Observatory (EDO)	http://edo.jrc.ec.europa.eu/edov2/php/index.php?id=1000
JRC/The European Forest Fire Information System (EFFIS)	http://forest.jrc.ec.europa.eu/effis/
Met Office Hadley Centre	http://www.metoffice.gov.uk/hadobs/hadex/index.html http://www.metoffice.gov.uk/hadobs/hadex/data/download.html
WorldClim	http://www.worldclim.org/
EEA	http://www.eea.europa.eu/data-and-maps/indicators/global-and-european-temperature-1/assessment

EURO-CORDEX

EURO-CORDEX is a major European initiative under the scope of the global CORDEX efforts (Giorgi and Gutowski 2015). The results of large number of RCM simulations on the 50-km and 12.5-km horizontal resolutions are available through the Earth System Grid Federation (ESGF) system, which consists of several data nodes. For example, one can explore the content of the entire database by accessing to e.g., https://esg-dn1.nsc.liu.se/projects/esgf-liu/. In addition, large number of new simulations are being performed and in the process of the postprocessing. Based on the specific climate information needed for LIFE-RESYSTAL, different subset of the EURO-CORDEX simulation / derived products can be selected.

In addition, various simulations are available for research or commercial applications (please check ESGF database for more information about specific simulation). A subset of these simulations are already interpolated from the default RCM grids to regular 0.44°x0.44° and 0.11°x0.11° grids (this may be relevant for some post-processing and visualization software).

MED-CORDEX

MED-CORDEX (https://www.medcordex.eu/) simulations follow CORDEX framework over the Mediterranean region, and include two general categories of simulations (we list below types of simulations relevant for the LIFE-RESYSTAL project):

- ✓ CORE simulations: 50 km atmosphere-only RCMs (similar to EURO-CORDEX but over smaller domain). These simulations differ in comparison to TIER-1 simulations (see below) in specific atmospheric or climatological conditions where interaction between processes in atmosphere and ocean is important.
- ✓ TIER-1 simulations: 12.5 km atmosphere-only RCMs (similar to EURO-CORDEX but over small domain) and coupled RCM simulations (i.e., simulations having interaction between atmosphere and oceans) on a highest spatial resolution available to specific modelling group.

All MED-CORDEX experiments (i.e. ocean-only simulations, couple atmosphere-ocean simulations etc) are listed on the following URL: https://www.medcordex.eu/simulations.php. Table 4 presents archived atmosphere-only RCM simulations.

Table 4: MED-CORDEX RCM simulations available through www.medcordex.eu portal.

Institute	RCM	Res. (km)	ERA-Interim	HIST period 1950-2005	RCP8.5 2006-2100	RCP4.5 2006-2100
ENEA	RegCM 3.1	30	1982-2010			
ICTP	RegCM4.3 v1	50	1979-2008	1970-2005: HadGEM2-ES		2006-2100: HadGEM2-ES
ICTP	RegCM4.3 v4	50	1979-2012			
ICTP	RegCM4.3 v7	50	1979-2008	1970-2005: MPI-ESM, HadGEM2-ES	2006-2100: MPI-ESM, HadGEM2-ES	
CNRM	ALADIN 5.2	50	1979-2012	1950-2005: CNRM-CM5	2006-2100 CNRM-CM5	
LMD	LMDZ	30	1979-2009			
U. Belgrade	EBU	50	1989-2008			
IPSL	WRF 3.1.1 v1	50	1989-2008			
IPSL	WRF 3.1.1 v2	50	1989-2008			
UCLM	PROMES	50	1989-2008			

GUF	CCLM 4-8-11 v1	50	1989-2008			
GUF	CCLM 4-8-11 v2	50	1979-2008			
GUF	CCLM 4-8-18 v1	50	1979-2011	1950-2005: MPI-ESM	2006-2100: MPI-ESM	
GUF	CCLM 4-8-18 v2	50	1979-2011			
CMCC	CCLM 4-8-19 v1	50	1979-2012	1950-2005: CMCC-CCLM4-8-19	2006-2100: CMCC-CCLM4-8-19	2006-2100: CMCC-CCLM4-8-19
CMCC	CCLM 4-8-19 v2	50	1979-2012			
ICTP	RegCM4.3 v1	12	1979-2010	1970-2005: HadGEM2-ES	2006-2050: HadGEM2-ES	
CNRM	ALADIN5.2	12	1979-2012	1950-2005: CNRM-CM5	2006-2100: CNRM-CM5	2006-2100: CNRM-CM5
UCLM	PROMES	12	1989-2009			
GUF	CCLM4-8-18 v1	10	1989-2008			
GUF	CCLM4-8-18 v2	10	1989-2008			

The use of MED-CORDEX simulations is suggested only in those case studies cases where atmosphere-ocean effects are important.

Reanalysis products

The use of the reanalysis in modern meteorology and climatology is enabled by the advanced combination of models and observations. Typical weather forecasts are performed using NWP models whose computer code changes over the course of time. For example, new numerical methods are implemented, new model physics are introduced, old model components are improved or developed, etc. In most modelling centres, these changes are done several times per year. The idea between the reanalysis is to use the frozen model code and simulate long periods (several decades or even a century). This enables to explore climatology and perform trend analysis in support to standard evaluation of pure observational records.

Table 5: State of the art global reanalyses.

Acronym	Institution	Spatial resolution	Period	Comments	Data access
ERA-5	ECMWF	~25 km	1979-2016/present	used to force EURO-CORDEX and MED-CORDEX RCMs	http://apps.ecmwf.int/datasets/

ERA-Interim	ECMWF	~80 km	1979-2016/present	used to force EURO-CORDEX and MED-CORDEX RCMs	http://apps.ecmwf.int/datasets/
ERA-20C	ECMWF	~125 km	1900-2010		http://apps.ecmwf.int/datasets/
ERA-40	ECMWF	~125 km	1957-2002	used to force ENSEBMLES RCMs	http://apps.ecmwf.int/datasets/
JRA-55	JMA	~125 km	1958-2016/present	full observing system reanalysis	http://jra.kishou.go.jp/JRA-55/index_en.html#download
MERRA-2	NASA	~ 50 km	1980-2016/present		http://disc.sci.gsfc.nasa.gov/mdisc/
CFSR	NCEP	~ 38 km	1979-2016/present		https://nomads.ncdc.noaa.gov/#cfsr
20CRv2c	NOAA-CIRES	~ 200 km	1851-2014		http://portal.nersc.gov/project/20C_Reanalysis/

Two issues may limit direct applicability of the EURO-CORDEX simulations for the purpose of the LIFE-ESYSTAL project:

a) The existence of the systematic errors when comparing simulated and observed climatology (when available) motivates further model development. This task is beyond the scope of the project, and can be replaced by the application of appropriate bias adjustment (or bias correction) technique, as a measure of the statistical post-processing before further use of the RCMs' results in forcing different impact models.
b) Model output for specific grid cell can be interpreted as an average value of the entire grid area (e.g. for CORDEX = 12.5 km x 12.5 km area). The spatial/horizontal resolution of the cell may be limiting for some high precision applications, especially if focus is typical urban areas, or areas with highly localized climate patterns. This issue can be alleviated to some extend using the climate downscaling methods.

Climate Downscaling Methods

Statistical downscaling makes it possible to estimate the effect of a large-scale climate change on local climatic variables, such as temperature, precipitation or aspects influenced by climate, using multi-variate and AI/ML techniques. It is implemented through R-package functions which also facilitate data processing and more general analysis. The Input Data to statistical downscaling include different climate parameters appropriately defined (units) such as precipitation (mm/day), temperature (deg. C), sea level pressure (hPa), sea surface temperature (deg. C), wind (m/s), sea level (m) etc and can produce results such as number of events exceeding a critical value (impacted area, precipitation, synoptic storms, etc).

The idea behind statistical downscaling is (1) to detect statistical relationships between simulated fields from GCMs or RCMs and real local observations in historical climate, and (2) to apply these relationships on GCMs’ or RCMs’ future climate projections in order to reconstruct possible statistics of the local climate in future periods. This means basically three ingredients are needed: (1) long time-series of high-quality local observations (e.g. time-series of total precipitation, 2m air temperature and/or 10m wind speed), (2) GCMs or RCMs projections of either same quantities as in (1) or of large-scale quantities than can strongly influence near-surface variables, and (3) development of statistical links between (1) and (2). The science of statistical downscaling is in progress and is often associated with the bias correction or bias adjustment methods. Bias adjustment is of particular concern when results of GCMs and RCMs are needed for force various impact models (e.g. forest fire spread model, detailed hydrological models, crop productivity models, etc.). For a quick introduction into this topic consider e.g. climate4impact.eu portal and Benestad et al. (2008).

The second, but highly time consuming downscaling process can be achieved through the so-called **dynamic downscaling** process where weather/climate models are forced by coarser spatial resolution models and generate high resolution, multi-dimensional climate simulations. Within LIFE-RESYSTAL. NCSRD will employ the Weather Research and Forecasting (WRF) Model, a state-of-the-art atmospheric mesoscale modeling system designed for both meteorological research and numerical weather prediction. It offers a host of options for atmospheric processes and can run on a variety of computing platforms. It features two dynamical cores, a data assimilation system, and a software architecture facilitating parallel computation and system extensibility. The model serves a wide range of meteorological applications across scales from tens of meters to thousands of kilometres. The WRF system contains two dynamical solvers, referred to as the ARW (Advanced Research WRF) core and the NMM (Non-hydrostatic Mesoscale Model) core.

Climate Operators

Each weather forecast service stores its real-time and historic climate data in a predefined data format over the area of interest typically a domain of the model integration. These data in Europe may be typically available in the following formats:

(1) GRIB
(2) NetCDF
(3) FA

For point based forecasts, many other formats are available (.csv, xml, ascii). Formats for point based forecasts will not be described here further.

GRIB

GRIB (GRIdded Binary or General Regularly-distributed Information in Binary form) [1] is a concise data format used for storage and transport of gridded meteorological data, such as those from NWP or climate models. It is designed to be self-describing, compact and portable across computer architectures, which is facilitated by the fact that the size of GRIB files is typically 1/2 to 1/3 of the size of normal binary files (floats). The GRIB standard was designed and maintained by the WMO[A]. A GRIB file contains one or more data records, arranged as a sequential bit stream. Each record begins with a header, followed by packed binary data. The header is composed of unsigned 8-bit numbers (octets). It contains information about: (i) the qualitative nature of the data (field, level, date of production, forecast valid time, etc.), (ii) the header itself (meta-information on header length, header byte usage, presence of optional sub-headers), (iii) the method and parameters to be used to decode the packed data, and (iv) the layout and geographical characteristics of the grid the data is to be plotted on.

Over the years, the WMO issued three editions of the GRIB standard: GRIB0, GRIB1 and GRIB2. GRIB0 was used to a limited extent by some projects, but now is outdated, unsupported and very rarely used. GRIB1 is used operationally worldwide by most meteorological centers for NWP outputs. Its future development is suspended, and there is a medium term plan to replace it completely by newer and advanced GRIB2 format. Unlike GRIB1 which needs separate parameter table to unpack the data, GRIB2 improves upon the standard with the compression and the inclusion of mentioned parameter table. Compression is ensured by using the same compression software used for images which gains in a roughly 50% reduction in file size over GRIB1.

NetCDF

NetCDF (Network Common Data Form) [2] is a file format designed to support the creation, access, and sharing of array-oriented scientific data in a form that is self-describing (contains information defining the data within the file) and portable (data are in the form which can be accessed by computers with different ways of storing integers, characters, and floating-point numbers)[B]. It is used extensively in the atmospheric and oceanographic communities to store variables, such as temperature, pressure, wind speed, etc. Array values may be accessed directly, without knowing the details of how the data are stored. Auxiliary information about the data, such as what units are used, may be stored with the data. Generic utilities and application programs can access netCDF datasets and transform, combine, analyse, or display specified fields of the data. The development of such applications has led to improved accessibility of data and improved re-usability of software for array-oriented data management, analysis, and display. NetCDF format was developed and is updated by the Unidata program at the University Corporation for Atmospheric Research (UCAR).

Until version 3.6.0, all versions of netCDF employed only one binary data format, now referred to as netCDF classic format. In version 3.6.0 a new binary format was introduced, 64-bit offset format. Nearly identical to netCDF classic format, it uses 64-bit offsets, and allows users to create far larger datasets. In version 4.0.0 a third binary format was introduced, so-called HDF5 format. Starting with this version, the netCDF library can use HDF5 files as its base format. By default, netCDF uses the classic format. In order to use the 64-bit offset or netCDF-4/HDF5 format, one has to set the appropriate constant when creating the file.

[1] WMO Code Form FM 92-IX Ext. GRIB (http://www.wmo.int/pages/prog/www/WDM/Guides/Guide-binary-2.html)

[2] NetCDF Users Guide (http://www.unidata.ucar.edu/software/netcdf/docs/user_guide.html)

FA

FA (Fichier ARPEGE)[3] files are type of NWP output files used within the ALADIN and HIRLAM consortium. Essentially they are very similar to GRIB files, i.e. may be consisted of either grid-point data or spectral coefficients and can be compressed or uncompressed. Major difference between those two types of files is in fact that FA files cannot be split in multiple files each of whom would contain only one field with its own header.

Scripts for processing climate data

CDO is set of command line instructions used to perform most common data manipulation operations on standard climate and NWP datasets. The NCAR Command Language (NCL) is a free interpreted language designed specifically for scientific data processing and visualization.NCO and is set of command line instructions used to perform most common data manipulation operations on standard climate and NWP datasets. The detailed description of these tools is presented below:

Climate Data Operators (CDO)	
Partner	**DHMZ**
Scope and Use	CDO is set of command line instructions used to perform most common data manipulation operations on standard climate and NWP datasets. One of its advantages is the user-friendly syntax.
Link to LIFE-RESYSTAL framework	CDO can be used to prepare input files for the Toolbox, or as an intermediate step for any software than can read standard meteorological formats (e.g. netCDF3, netcdf4, GRIB1, GRIB2).
Use in case studies	CDO is used e.g. to (1) select specific periods, (2) select specific sub-areas from the regional climate models' output available from the ESGF system.
Coding Language	CDO instructions can be incorporated in standard Unix shell scripts such as Bash scripts.
Input Data	Input data for CDO should be in GRIB1, GRIB2, and netCDF3 or netCDF4 format. Several other formats also supported (check the CDO Documentation https://code.mpimet.mpg.de/projects/cdo/wiki/Cdo#Documentation)
Output Data	Output data from CDO can be in the GRIB1, GRIB2, netCDF3, netCDF4 or ASCII formats.
Uncertainty (if any)	-
References / Citations	Schulzweida, U. (2017), CDO User Guide, https://code.mpimet.mpg.de/projects/cdo/embedded/cdo.pdf

[3] A Guide to the FA file format (http://www.cnrm-game-meteo.fr/gmapdoc/spip.php?article42)

Basic Information of tool	Version 1.9.1 is available as of October 2017. Under Ubuntu system, the installation is trivial "sudo apt-get install cdo"
Photos / screen shoots	CDO is a command line tool without the graphical interface.

NCL	
Partner	**Partner Name**
Scope and Use	The NCAR Command Language (NCL), a product of the Computational & Information Systems Laboratory at the National Center for Atmospheric Research (NCAR) and sponsored by the National Science Foundation, is a free interpreted language designed specifically for scientific data processing and visualization.
Link to LIFE-RESYSTAL framework	The NCAR Command Language (NCL) is used in order to read the netCDF climate global data and either extract the information needed in ASCII format, or print the values of a variable in a specific date and point (lon/lat).
Use in case studies	All
Coding Language	NCL instructions can be incorporated in standard Unix shell scripts such as Bash scripts
Input Data	Format ascii, NetCDF, binary, grib Type Spatial / Temporal
Output Data	Format ascii, netcdf, binary, brib, png, jpg Type Spatial / Temporal
Uncertainty (if any)	NaN
References / Citations	http://dx.doi.org/10.5065/D6WD3XH5
Basic Information of tool	NCL comes with many useful built-in functions and procedures for processing and manipulating data. There are over 600 functions and procedures that include routines for: ● use specifically with climate and model data ● computing empirical orthogonal functions, Fourier coefficients, singular value decomposition, averages, standard deviations, sin, cosine, log, min, max, etc. ● retrieving and converting date information ● drawing primitives (lines, filled areas, and markers), wind

	barbs, weather map symbols, isosurfaces, and other graphical objects • robust file handling • 1-dimensional, 2-dimensional, and 3-dimensional interpolation, approximation, and re-gridding • facilitating computer analysis of scalar and vector global geophysical quantities (most are based on the package known as Spherepack) • retrieving environment variables and executing system commands NCL supports calling C and Fortran external routines, which makes NCL infinitely configurable.
Photos / screen shoots	-

NetCDF Operator (NCO)	
Partner	**DHMZ**
Scope and Use	NCO is set of command line instructions used to perform most common data manipulation operations on standard climate and NWP datasets (similar functionality in some aspects to CDO).
Link to LIFE-RESYSTAL framework	NCO can be used to prepare input files for the Toolbox, or as an intermediate step for any software than can read standard meteorological formats (e.g. netCDF3, netcdf4; similar to CDO).
Use in case studies	NCO is used e.g. to extract or modify metadata from the regional climate models' output available from the ESGF system.
Coding Language	NCO instructions can be incorporated in standard Unix shell scripts such as Bash scripts (similar to CDO).
Input Data	Input data for NCO should be in the netCDF3 or netCDF4 format. Several other formats also supported (check the NCO Documentation http://nco.sourceforge.net/nco.html)
Output Data	Output data from NCO can be in the netCDF3, netCDF4, binary or ASCII formats (similar to CDO).
Uncertainty (if any)	-
References / Citations	Zender, C. S. (2008), Analysis of Self-describing Gridded Geoscience Data with netCDF Operators (NCO), Environ. Modell. Softw., 23(10), 1338-1342, doi:10.1016/j.envsoft.2008.03.004. Zender, C. S. (2017), netCDF Operator (NCO) User Guide, `http://nco.sf.net/nco.pdf`.
Basic Information of tool	Version 4.7.0 is available as of November 2017. Under Ubuntu system, the installation is trivial "sudo apt-get install nco"
Photos / screen shoots	NCO is a command line tool without the graphical interface.

Climate Standards

The aim of this chapter is to present, the structure and content of metadata covering results of climate data. As the global climate community has embraced the CF (Climate and Forecast) metadata conventions and applied this in recent CMIP5 (Coupled Model Intercomparison Project Phase 5[4]) and CORDEX (Coordinated Regional Climate Downscaling Experiment[5], including EURO-CORDEX) programmes.

[4] https://cmip.llnl.gov/cmip5/
[5] http://cordex.org/

Concerning the introduction of file-naming and variable-naming structure into the LIFE-RESYSTAL platform and individual components, we shall use the standardized formats (e.g. NetCDF and XML), and standardized set of file file-naming and metadata entries (such as CF, CMIP5 and CORDEX) for the purpose of more systematic linkage of different steps when exploring the impact of climate events on the health infrastructure.

In principle there are several major initiatives and programmes that consider various types of the metadata at the different level. Their goal is to have community accepted standards that allow for the safe, no-error and transparent use of the climate data, independent of the user and the application platform. This includes Open Geospatial Consortium (OGC), ISO at the highest level of generalization, then moves to CF, CORDEX and WMO-TD No. 1186 at the next level where community relevant (in this case meteorology and climatology) recommendations are given.

All types of data can be produced by the research communities and national and international bodies (often following community driven efforts for the common formats and conventions), and user generated data (often needs additional work to format specific data before sharing). Common format to share these data is NetCDF. Table 6 summarizes main elements in the NetCDF file related to variable dimensions, its local metadata (there may be several variables in the same file; e.g. air-temperature, precipitation amount, wind speed)

Table 6: Main components of the NetCDF header under the CF metadata framework. Description of the components is given by.

Dimensions	These state the dimensions of the structure that the data will reside in and must always have different names. Dimensions can be spatial, temporal or any other variable. CF convention recommends that the dimensions be listed in the order TZYX, where T is time, Z is height or depth, Y is latitude and X is longitude.
Variables	**Data type:** • floating point ('float'), • double precision ('double') • integer ('int') numbers **Units** **Content description:** • standard_name • long_name **Missing data values:** describe the dummy values for missing data within the file.
Global Attributes	**Type Format:** CF-1.4 **title:** a succinct description of the data set **institute:** the organisation where the data were produced **source:** how the data were produced, e.g. model type, run number and circumstances **history:** an audit trail of data set processing **references:** web page or reference to the report and/or scientific publication **comment:** other useful information not covered elsewhere that adds value **author:** the person(s) who generated the data

In Table 7 an overview is given of the typical types of information that describe dataset from the CORDEX initiative that brings together different research groups working on regional climate model development and application. Information are given at the level of (1) filename and (2) metadata content.

Table 7: Exctactable information from the results of the CORDEX regional climate models in NetCDF format.

Climate model data and metadata in NetCDF format	
Source of data	Results of the regional climate models' simulations saved in the NetCDF format using CF and CORDEX metadata description.
References to relevant formats and protocols	NetCDF: https://www.unidata.ucar.edu/software/netcdf/ CF: http://cfconventions.org CORDEX: http://www.cordex.org/experiment-guidelines/experiment-protocol-rcms/
Typical filenames and content	*For example:* tas_EUR-11_CNRM-CERFACS-CNRM-CM5_rcp45_r1i1p1_SMHI-RCA4_v1_day_20060101-20101231.nc Files of this type contain a variable of interest (e.g. tas: near-surface air temperature) and various additional quantities that describe the spatial and temporal characteristics of the data.
Description of the typical filenames	*For example:* **tas**: variable saved: near-surface air temperature **EUR-11**: domain: European domain at the 0.11° (~12.5 km) horizontal resolution. **CNRM-CERFACS**: institution that performed global climate model (GCM) simulation **CNRM-CM5**: specific GCM **rcp45**: greenhouse gases concentration scenario **r1i1p1**: additional details of the GCM simulation related to the model physics and initialization method. **SMHI**: institution that performed regional climate model (RCM) simulation **RCA4_v1**: specific RCM **day**: frequency of data: daily data **20060101**: first simulated time event in this file YYYYMMDD **20101231**: last simulated time event in this file YYYYMMDD
Dimensions	*For example:* dimensions: bnds = 2 ; time = UNLIMITED ; // (1826 currently) rlon = 424 ;

	rlat = 412 ;
Variables and **local attributes** saved in the respective file [NetCDF specific]	*For example:* variables: double **height** ; height:axis = "Z" ; height:long_name = "height" ; height:positive = "up" ; height:standard_name = "height" ; height:units = "m" ; double **time_bnds**(time, bnds) ; double **rlon**(rlon) ; rlon:standard_name = "grid_longitude" ; rlon:long_name = "longitude in rotated pole grid" ; rlon:units = "degrees" ; rlon:axis = "X" ; double **rlat**(rlat) ; rlat:standard_name = "grid_latitude" ; rlat:long_name = "latitude in rotated pole grid" ; rlat:units = "degrees" ; rlat:axis = "Y" ; char **rotated_pole** ; rotated_pole:grid_mapping_name= "rotated_latitude_longitude" ; rotated_pole:grid_north_pole_latitude = 39.25 ; rotated_pole:grid_north_pole_longitude = -162. ; double **time**(time) ; time:standard_name = "time" ; time:units = "days since 1949-12-01 00:00:00" ; time:calendar = "standard" ; time:long_name = "time" ; time:bounds = "time_bnds" ; time:axis = "T" ; float **tas**(time, rlat, rlon) ; tas:grid_mapping = "rotated_pole" ; tas:_FillValue = 1.e+20f ; tas:missing_value = 1.e+20f ; tas:standard_name = "air_temperature" ;

	tas:long_name= "Near-Surface Air Temperature" ; tas:units = "K" ; tas:coordinates = "lonlat height" ; tas:cell_methods = "time: mean" ; double **lon**(rlat, rlon) ; lon:standard_name = "longitude" ; lon:long_name = "longitude" ; lon:units = "degrees_east" ; double **lat**(rlat, rlon) ; lat:standard_name = "latitude" ; lat:long_name = "latitude" ; lat:units = "degrees_north" ;
Global attributes saved in the respective file [NetCDF specific]	*For example:* global attributes: :Conventions = "CF-1.4" ; :contact = "rossby.cordex@smhi.se" ; :creation_date = "2013-07-03-T23:48:30Z" ; :experiment = "RCP4.5" ; :experiment_id = "rcp45" ; :driving_experiment = "CNRM-CERFACS-CNRM-CM5, rcp45, r1i1p1" ; :driving_model_id = "CNRM-CERFACS-CNRM-CM5" ; :driving_model_ensemble_member = "r1i1p1" ; :driving_experiment_name = "rcp45" ; :frequency = "day" ; :institution = "Swedish Meteorological and Hydrological Institute, Rossby Centre" ; :institute_id = "SMHI" ; :model_id = "SMHI-RCA4" ; :rcm_version_id = "v1" ; :project_id = "CORDEX" ; :CORDEX_domain = "EUR-11" ; :product = "output" ; :references="http://www.smhi.se/en/Research/Research-departments/climate-research-rossby-centre" ; :tracking_id = "0a9ffd2d-8fbb-4573-ba8b-8344ebfee6f9" ; :rossby_comment = "201144: CORDEX Europe 0.11 deg | RCA4 v1 | CNRM-CERFACS-CNRM-CM5 | r1i1p1 | rcp45 | L40" ; :rossby_run_id = "201144" ;

	:rossby_grib_path= "/nobackup/rossby16/rossby/joint_exp/cordex/201144/raw/"

When primary climate data are not given in NetCDF format, other data formats are valid too. Table 8 lists LIFE-RESYSTAL recommendation of the metadata information needed with the CSV data format.

Table 8: LIFE-RESYSTAL recommendation for processed data and extreme events in CSV format.

Climate model data output: CSV format	
Headers	Data type: double precision ('double') Header description with units: • TIME , • LAT , LON , • T_00 [C] , • WS_00[m/s] , • T_06[C] , • WS_06[m/s] , • T_12[C] , • R_D[mm] , • WS_12[m/s] , • RH_D[%] , • SM_D[kg/m2] , • T_18[C] , • WS_18[m/s] Missing data values : -9999
External Attributes	Type Format —csv title: scenario[rcpXY]_startyear[YYYY]_endyear[YYYY]_pointID[idXXXXX].csv (for example rcp26_2006_2010_id142771.csv) coordinate system : LAT, LON climatic model : any climatic scenario : any institute : NSCRD source : post processing of NetCDF EURO-CORDEX climate data with NCL scripts and FORTRAN code Comment :

File-naming recommendation for the climate data

Based on the experience in other initiatives, some good-practice approaches may be followed. For the example of the regional model data given in Table 6, tas_EUR-11_CNRM-CERFACS-CNRM-CM5_rcp45_r1i1p1_SMHI-RCA4_v1_day_20060101-20101231.nc

the CORDEX instructions where followed, while the processed regional climate data are summarized in Table 8,

rcp26_2006_2010_id142771.csv

with elements between underscore ("_") separately defined. We recommend similar structure of the file names for climate data in other formats, too. This enables efficient listing, searching and filtering procedures.