SEBASTIEN

Deliverable Lead	CMCC
Deliverable due date	2022/9/30
Status	FINAL
Version	V1.0
Project	SEBASTIEN

Co-financed by the Connecting Europe Facility of the European Union

Agreement: INEA/CEF/ICT/A2020/2373580**Action**: 2020-IT-IA-0234

DOCUMENT INFORMATION

Title	2.2 List of suitable data sources and of newly acquired data
Agreement	INEA/CEF/ICT/A2020/2373580
Action	2020-IT-IA-0234
Creator	Alfredo Reder (CMCC), Francesco Renzi (Nature4.0)
Deliverable Description	List of suitable data sources and of newly acquired data
Contributors	Alfredo Reder (CMCC), Francesco Renzi (Nature4.0), Paola Mercogliano (CMCC)
Requested deadline	M9
Reviewer	Sergio Noce (CMCC), Marco Milanesi (UNITUS)

Outline

1. Executive Summary

This Deliverable aims to identify the existing data sources and ways for their access and extraction in developing SEBASTIEN applications and services. It moved from the Deliverable *D2.1 List of indicators/indices to be proposed to stakeholders* who identified a trial list of indicators for bioclimate, territorial and animal segments. The goal of D2.2 is to identify High-Value Datasets coming from multi-sources and multi-thematic portals (Italian Open Data Portal, SIAN, SINANET, SCIA, ISTAT, EAA, Copernicus Land Monitoring Service, Copernicus Climate Change Service, Copernicus-linked observatories, Copernicus Open Access Hub, LEO open data portal and OIE-WAHIS portal) suitable to derive these indicators and support services development (WP6). In addition, many other input data used in SEBASTIEN come from geospatial and non-geospatial datasets leveraging previous experiences of project partners (e.g., HIGHLANDER and LEO projects). This Deliverable follows the subdivision proposed in D2.1. The next Sections will introduce datasets containing weather and climate information (§2), datasets containing territorial information (§3) and datasets containing animal welfare information (§4). The Deliverable will be accompanied by an Annex in which specific features of the datasets are reported.

2. Datasets containing weather and climate information

2.1 Foreword

Historically, there has been a clear separation between weather and climate predictions, albeit both exploit similar numerical tools. Weather prediction refers to the prediction of daily weather patterns from a few days up to about two weeks in advance. Climate prediction refers to predicting climate fluctuations averaged over a season and beyond. However, convergence is occurring, stimulated by the growing realisation that weather and climate occur on a temporal and spatial scales continuum (see Figure 1). Consistent phenomena on a range of scales along this continuum lead to predictability on sub-daily scales, weeks, months, years, decades and beyond (Hoskins, 2012).

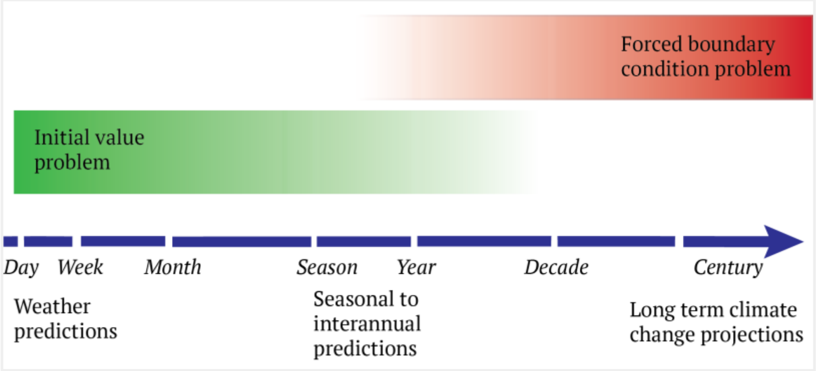

Forced boundary condition problem
Initial value problem
Day
Week
Month
Season
Year
Decade
Century
Weather predictions
Seasonal to interannual predictions
Long term climate change projections

Figure 1: Differences between weather and climate predictions

2.2 Weather forecasting

Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given location. Weather forecasts are made by collecting quantitative data on the current state of the atmosphere in a given area and using a scientific understanding of atmospheric processes to predict how the atmosphere will evolve there.

There are different types of weather forecasts, which are: nowcasting, defined by the WMO (World Meteorological Organization) as a detailed analysis and description of the current weather and then forecasting for a period from 0 to 6 h, short-range forecasting (will last 1-2 days), medium range forecasting[1] which generally covers a period ranging from about three days to seven days in advance and long-range forecasting[2] which typically provides information about expected future atmospheric and oceanic conditions, averaged over periods of one to three months. However, the chaotic nature of the atmosphere, the massive computational power required to solve the equations that describe the atmosphere, the error involved in measuring the initial conditions and an incomplete understanding of atmospheric processes mean that forecasting accuracy decreases with forecast length.

Nowcasting and short-term forecasts are the most reliable; medium-range forecasts are fairly accurate but with a certain margin of error: 75-85% in the first two days and less than 50% after eight days. On the other hand, long-range forecasts indicate weather patterns, so they are not very accurate.

2.3 Open Access state-of-the-art weather forecasting data

COSMO 2I

The Consortium for Small-scale Modeling (COSMO) model is a non-hydrostatic limited-area atmospheric model. It is based on the primitive thermo-hydrodynamical equations describing compressible flow in a moist atmosphere, with various physical processes taken into account by parameterisation schemes (Schaettler et al., 2019). The forecast model is initialised twice a day starting from Cosmo-5M lateral boundary conditions, with a forecast range up to 48h, covering one week backwards.

NCEP-GFS 0.25

The National Centers for Environmental Prediction (NCEP) operational Global Forecast System auxiliary analysis and forecast grids are on a 0.25° by 0.25° global latitude-longitude grid. Grids include analysis and forecast time steps at a 3-hourly interval from 0 to 240 and a 12-hourly interval from 240 to 384. Model forecast runs occur at 00, 06, 12, and 18 UTC daily.

[1] https://glossary.ametsoc.org/wiki/Medium-range_forecast

[2] https://www.ecmwf.int/en/forecasts/documentation-and-support/long-range

ARW-WRF

The Weather Research and Forecasting (WRF) model is a next-generation mesoscale numerical weather prediction system for atmospheric research and operational forecasting applications. In this specific case, the ARW-WRF is powered by the University of Naples “Parthenope” using three horizontal resolutions working in operational mode with four daily runs and a forecast range of up to one week.

2.4 Climate information

Depending on the time scale considered (see Figure 2), it is possible to categorise the available climate information into:

Observations - They are crucial to understanding the past and current features of a climate system; they include data from a variety of instrumental data records, ranging from historical weather observations to the latest measurements from space

Climate reanalysis - They combine past observations with models to generate consistent time series for a large set of climate variables for recent and current climate; they are among the most-used datasets in the geophysical sciences

Seasonal forecasts - They combine outputs from several state-of-the-art seasonal prediction systems from providers in Europe and elsewhere to estimate weather statistics on monthly and seasonal time scales, which places it somewhere between conventional weather forecasting and climate forecasting

Climate projections - They give projections of future climate for different scenarios for concentrations of greenhouse gases, aerosols and other atmospheric constituents based on outputs from multiple global and regional climate models

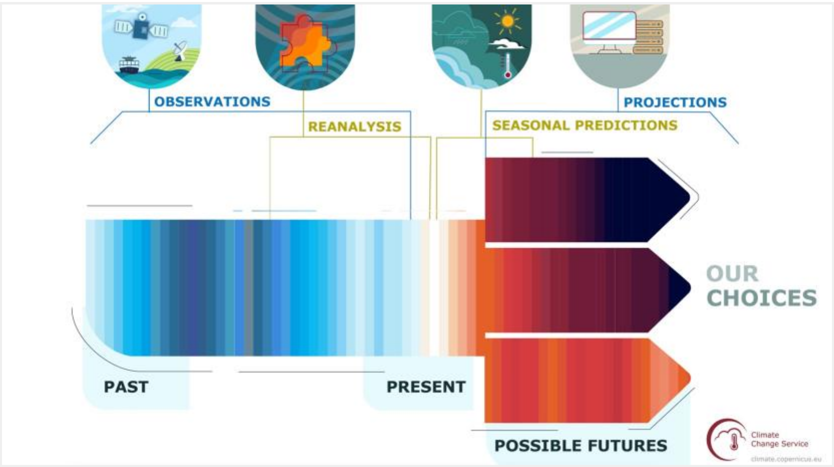


Figure 2: Climate information available for different time scales

2.4.1 Observations

The availability of homogeneous and continuous data (in terms of time and space) represents an essential requirement for characterising past and current climates and extreme past weather events and the associated physical and socio-economic impacts (Street et al., 2019). In this regard, the most obvious support is represented by in-situ weather station measurements. However, they are rarely available over long periods and often outline a scarce homogeneity and density of observation points assumed as a reference for large areas. Therefore, several gridded observational datasets have been developed for Europe and European countries to cope with such limitations in recent years. These products feature different temporal (e.g., from hourly to daily) and spatial (e.g., from ≃ 1 km to ≃ 10–20 km) resolutions, covering various periods; their reliability is strictly related to the density of station networks from which they derive. Moreover, they are often merged with radar and satellite information to cope with some limitations, e.g. for precipitation, rain gauges and weather radars constitute important devices for operational precipitation monitoring. However, rain gauges provide accurate yet spotty precipitation estimates, while radars offer high temporal and spatial resolution yet at a limited absolute accuracy (Wüest et al., 2010).

Concerning precipitation, observational gridded datasets could feature some limitations (Isotta et al., 2014): (i) precipitation underestimation at high elevation due to the not adequately accounted precipitation lapse rate and that induced by stations sparseness and mask-effect issues for radar data; (ii) systematic wind-induced rain gauge under-catch, and (iii) wetting and evaporation losses; (iv) interpolation methods, which systematically induce underestimation of high intensities (smoothing effect) and overestimation of low intensities (moist extension into dry areas); (v) density of station network, whose various coverage over Europe makes such a dataset more reliable in some regions than in others (Cornes et al., 2018). However, the effect of these limitations tends to reduce with spatial resolution. Moreover, additional sources of uncertainty are the methods adopted to merge rain gauge measurements and radar data and the statistical disaggregation approaches.

2.4.2 Climate Reanalysis

A climate reanalysis gives a numerical description of the recent climate by combining models with observations. It contains estimates of atmospheric parameters such as air temperature, pressure and wind at different altitudes, and surface parameters such as rainfall, soil moisture content, ocean-wave height and sea-surface temperature. The estimates are produced for all locations on Earth, and they span a long period that can extend back several decades or more. Climate reanalysis generates large datasets that can take up several petabytes of space and are best processed with cloud-based tools, to avoid large download volumes.

Climate reanalysis generally provides a multivariate, spatially complete, and coherent record of the atmospheric state – far more complete than any observational dataset can achieve. Its main advantages are to provide (i) regularly gridded data, even in places where there are no or few observations; (ii) a coherent, complete set of variables describing the atmospheric state; (iii) reconstruction of the record of past weather since observations constrain it. Reanalysis systems differ in the set of observations assimilated, the model used, and the way the error statistics are estimated and corrections are applied. Various reanalysis methods exist, e.g., 4D variational analysis (4D-VAR), 3D-VAR, nudging, and optimal interpolation.

The European Centre for Medium-Range Weather Forecast (ECMWF) periodically uses its forecast models and data assimilation systems to 'reanalyse' archived observations, creating global datasets describing the recent history of the atmosphere, land surface, and oceans.

Reanalysis requires notable developments to ensure its products' best possible temporal consistency, which can be adversely affected by biases in models and observations and the ever-changing observing system. Current research at ECMWF focuses on developing consistent reanalyses of the coupled climate system, including the atmosphere, land surface, ocean, sea ice, and the carbon cycle, extending back as far as a century or more. An additional research topic is directed towards improved handling of systematic model bias.

2.4.3 Seasonal Forecasts

Seasonal forecasts provide estimates of weather statistics on monthly and seasonal time scales, which places it somewhere between conventional weather forecasting and climate forecasting. In this sense, although seasonal forecasting shares some methods and tools with weather forecasting, it is part of a different paradigm that requires further treatment. Seasonal forecasts tell us how likely the next season will be wetter, drier, warmer or colder than 'usual' for that time of year. This type of long-term forecasting is possible because of the behaviour of specific components of the Earth system that evolve more slowly than the atmosphere (e.g. the ocean, the cryosphere) and in a predictable manner. Seasonal forecasts differ from a classic weather forecast, which gives a lot more precise detail - both in time and space - of the evolution of the state of the atmosphere over a few days into the future. Beyond a few days, the chaotic nature of the atmosphere limits the possibility of predicting precise changes at local scales. This is one of the reasons long-range forecasts of atmospheric conditions have significant uncertainties. To quantify these uncertainties, long-range forecasts use ensembles, and meaningful forecast products reflect a distribution of outcomes. Given the complex, non-linear interactions between the individual components of the Earth system, the best tools for long-range forecasting are climate models, which include as many of the system's key components as possible. Typically, such models include the atmosphere, ocean and land surface representations. They are initialised with data describing the system's state at the starting point of the forecast and used to predict the evolution of this state in time. While uncertainties coming from imperfect knowledge of the initial conditions of the components of the Earth system can be described with the use of ensembles, uncertainties arising from approximations made in the models are very much dependent on the choice of model. A convenient way to quantify the effect of these approximations is to combine outputs from several models, independently developed, initialised and operated.

2.4.4 Climate projections

Climate projections are simulations of Earth's climate in future decades (typically until 2100) based on assumed 'scenarios' for the concentrations of greenhouse gases, aerosols, and other atmospheric constituents that affect the planet's radiative balance. They are obtained by running numerical models of Earth's climate, which may cover either the entire globe or a specific region, e.g. Europe. These models are referred to as Global Climate Models (GCMs) – also known as General Circulation Models – or Regional Climate Models (RCMs), respectively.

Although featuring a coarser spatial resolution (e.g., ≃ 1000 km), GCM simulations are pivotal sources for quantitatively understanding how the climate of the earth may change over the 21 century. They are disseminated through different phases of coupled model intercomparison project (CMIP). The two last generations of GCMs are released in the frame of the CMIP5 and CMIP6. Their data underpin the Intergovernmental Panel on Climate Change 5th Assessment Report (IPCC AR5) and Intergovernmental Panel on Climate Change 6th Assessment Report (IPCC AR6). In addition to finer spatial resolutions, enhanced parameters of the cloud microphysical process, and other Earth system processes and components such as biogeochemical cycles and ice sheets (Eyring et al., 2019), CMIP6 varies from CMIP5 for the definition of future scenarios. CMIP5 projections are available based on 2100 radiative forcing values for four GHG concentration pathways emitted in the years to come (van Vuuren et al., 2011), i.e. Representative Concentration Pathways (RCP) 2.6, 4.5, 6.0 and 8.5 where the suffixes represent the radiative forcing values in the year 2100 (2.6, 4.5, 6, and 8.5 W/m^2, respectively). Conversely, CMIP6 projections rely on the combined pathways of Shared Socioeconomic Pathway (SSP) and RCP, being then more robust future scenarios.

Despite the widespread use of GCM projections by the international community to make decisions on climate change mitigation, the impacts of a changing climate and the adaptation strategies needed to address them occur on more regional and national scales. High-resolution RCMs can provide climate change information on regional and local scales in relatively fine detail, which cannot be obtained from coarse-scale GCMs. This is manifested in a better description of small-scale regional climate characteristics and a more accurate representation of extreme events. Consequently, the outputs of such RCMs are indispensable in supporting regional and local climate impact studies and adaptation decisions. Furthermore, RCMs are not independent of the GCMs, since the GCMs provide lateral and lower boundary conditions to the regional models. In that sense, RCMs can be viewed as magnifying glasses of the GCMs.

2.5 Open Access state-of-the-art climate datasets from Copernicus Climate Datastore

The C3S Climate Data Store (CDS) is a one-stop shop for past, present and future information about the climate. It provides a single access point to a wide range of quality-assured climate datasets distributed in the cloud. CDS datasets include observations, historical climate data records, estimates of Essential Climate Variables (ECVs) derived from Earth observations, global and regional climate reanalyses of past observations, seasonal forecasts and climate projections. Access to data is open, free and unrestricted. Along with the data, the CDS includes a set of tools for analysing and predicting the impacts of climate change. Users of the CDS can access these tools to develop their applications online. The CDS has been designed to support a wide range of users with different needs by facilitating the processing of large data volumes and creating simple visualisations based on multiple data sources. CDS data and tools form the backbone of the C3S Sectoral Information

System (SIS), which provides tools and applications for dealing with climate impact in different industrial sectors, including energy, water management and agriculture.

2.5.1 Observations

E-OBS Gridded Observations

E-OBS (Cornes et al., 2018; Haylock et al., 2008) is a daily gridded land-only observational dataset over Europe at a horizontal resolution of 0.1° (≃11 km) for 1950–2021. It contains data for precipitation amount, mean/maximum/minimum temperature, relative humidity, sea level pressure, wind speed, and surface shortwave downwelling radiation. E-OBS relies on the blended time series from the station network of the European Climate Assessment & Dataset (ECA&D) project. It is calculated following a two-stage process to derive the daily field and the uncertainty in these daily estimates. The latest version (i.e., v25.0e) was released in April 2022, covering the period 01/01/1950-31/12/2021. The dataset is daily, meaning the observations cover 24 hours per time step. The exact 24-hour period can be different per region. The reason for this is that some data providers measure between midnight to midnight while others might measure from morning to morning. Since E-OBS is an observational dataset, no attempts have been made to adjust the time series for this 24-hour offset.

2.5.2 Climate Reanalysis

ERA5 Reanalysis

ERA5 (Hersbach et al., 2020) is the fifth generation ECMWF reanalysis representing the most plausible description of the current climate nowadays. It has global coverage with a spatial resolution of 0.25◦ (≃31 km) and provides outputs at an hourly scale from 1950 to now (with a latency of 5 days). Such features make ERA5 suitable for a wide range of applications, such as monitoring climate change, research, education, policy-making and business in sectors such as renewable energy and agriculture (Buontempo et al., 2020). It forms the basis for monthly C3S climate bulletins. It is used in the World Meteorological Organization's annual assessment of the State of the Climate presented at the Conference of the Parties of the United Nations Framework Convention on Climate Change (UNFCCC). ERA5 data are available as hourly and monthly products on pressure levels (upper air fields) and single levels (atmospheric, ocean-wave and land surface quantities).

ERA5-Land Reanalysis

ERA5-Land (Muñoz-Sabater et al., 2021) is a reanalysis dataset released by the ECMWF. It is a replay of the land component of the ERA5 climate reanalysis, at a finer horizontal resolution and with a series of improvements making it more accurate for all types of land applications. It has hourly temporal resolution and a spatial resolution of 0.1° x 0.1° (~9km) from 1950 to the present. It provides several different types of climatic data, including precipitation and temperature fields. ERA5-Land does not assimilate additional observation data directly. Instead, the evolution of the model is driven by the atmospheric fields obtained from the lowest ERA5 model level, which is 10 m above the surface, interpolated from the ERA5 resolution (~31 km) to ERA5-Land resolution (~9 km) via a linear interpolation method and with additional lapse-rate correction derived from ERA5 (Dutra et al., 2020).

UERRA (Regional Reanalysis for Europe from 1961 to 2019)

UERRA (Ridal et al., 2017; Bazile et al., 2017) in the MESCAN-SURFEX option is a reanalysis at ~5.5 km, providing estimations of the climate in Europe from 1961 to 2019 at 00, 06, 12, and 18 UTC (N.B., only a 06 UTC for precipitation). It descends from the UERRA-HARMONIE, a reanalysis (~11 km) based on a 3-D data assimilation system assuming along the lateral borders data from ERA40 for the years before 1979 and ERA-Interim for the years until 2019. It combines the UERRA-HARMONIE with the MESCAN system and the land surface platform SURFEX to derive daily accumulated precipitation. To this aim, additional surface observations are considered.

CERRA (Sub-daily Regional Reanalysis data for Europe from 1984 to present)

The Copernicus European Regional ReAnalysis (CERRA) dataset represents a step forward from UERRA, providing spatially and temporally consistent sub-daily historical reconstructions of atmospheric and surface meteorological variables for Europe from 1984 to the present day, using ERA5 as lateral boundary conditions. It was produced using the HARMONIE-ALADIN limited-area numerical weather prediction, and data assimilation system referred to as the CERRA system. The CERRA system employs a 3-D variational data assimilation scheme of the atmospheric state at every assimilation time. CERRA inputs are observational data, lateral boundary conditions from ERA5 global reanalysis (as prior estimates of the atmospheric state), and physiographic datasets (describing the surface characteristics of the model). The observing system has developed over time. Although the data assimilation system can resolve data holes, the much sparser observational networks in the past periods can impact the quality of analyses leading to less accurate estimates. The uncertainty estimates for reanalysis variables are provided by the CERRA-EDA, a 10-member ensemble of data assimilation systems.

2.5.3 Seasonal Forecasts

C3S Seasonal forecast data

The C3S provides a multi-system seasonal forecast service, where data produced by state-of-the-art seasonal forecast systems developed, implemented and operated at forecast centres in several European countries is collected, processed and combined to enable user-relevant applications. The centres currently providing forecasts to C3S are ECMWF, The Met Office and Météo-France, Deutscher Wetterdienst (DWD), Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), National Centers for Environmental Prediction (NCEP) and Japan Meteorological Agency (JMA). Each model simulates the Earth system processes that influence weather patterns in slightly different ways and makes slightly different approximations, leading to other kinds of model error.

Each model simulates the Earth system processes that influence weather patterns slightly differently and makes slightly different approximations, leading to different types of model error.

The C3S seasonal multi-system composition and the full content of the database underpinning the service are described in its documentation. The data is grouped in several catalogue entries (CDS datasets), currently defined by the type of variable (single-level or multi-level, on pressure surfaces) and the level of post-processing applied (data at original time resolution, processing on temporal aggregation and post-processing related to bias adjustment). The variables available in this dataset include forecasts created in real-time (since 2017) and retrospective forecasts (hindcasts) initialised at equivalent intervals during 1993-2016.

2.5.4 Climate projections

EURO-CORDEX data

The EURO-CORDEX initiative represents the European branch of the Coordinated Downscaling Experiment (CORDEX) (Jacob et al., 2014; Giorgi and Gutowski, 2015) of the World Climate Research Programme (WCRP). It includes data from many experiments, models, domains, resolutions, ensemble members, time frequencies and periods computed following the CORDEX experiments protocol. In general, these experiments consist of RCM simulations representing different future socio-economic scenarios (forcings), different combinations of GCMs and RCMs and different ensemble members of the same GCM-RCM combinations. This experiment design allows studies to address questions related to the key uncertainties in future climate change. These uncertainties come from differences in the scenarios of future socio-economic development, imperfections of regional and global models used and internal (natural) variability of the climate system. The term "experiment" refers to three main categories:

1. *Evaluation*: Experiments driven by ECMWF ERA-Interim reanalysis for a past period (typically 1980-2010); they can be used to evaluate the quality of the RCMs using perfect boundary conditions as provided by a reanalysis system
2. *Historical*: Experiments covering a period for which modern climate observations exist (typically 1950-2005); GCMs provide their boundary conditions. These experiments follow the observed changes in climate forcing and show how the RCMs perform for the past climate when forced by GCMs; they can be used as a reference period for comparison with scenario runs for the future
3. *Scenario*: Experiments covering a future period (typically 2006-2100) relying on RCP forcing scenarios (i.e. RCP 2.6, 4.5 and 8.5 scenarios); their boundary conditions are provided by GCMs

Specifically, the same experiments in CORDEX were done using different RCMs. In addition, for each RCM, there is a variety of GCMs, which can be used as lateral boundary conditions. The GCMs used are coming from the CMIP5 archive. Moreover, the uncertainty related to the internal variability of the climate system is sampled by running several simulations with the same RCM-GCM combination. On the forms, these are indexed as separate ensemble members. Finally, for each

GCM, the same experiment was repeatedly done using slightly different conditions (like initial conditions or different physical parameterisations, for instance), producing an ensemble of closely related experiments.

2.6 Additional climate dataset for Italy

In addition to the various datasets for climate information described in the previous paragraphs, several products for Italy are helpful for the scope of the SEBASTIEN project. In particular, SEBASTIEN will leverage past and ongoing experiences involving project partners, such as the Highlander project. Current advancements in supercomputer power have made it possible to run very high-resolution climate models with kilometre-scale grids (< 4 km). Such models are acknowledged as "convection-permitting" as, at such a high resolution, they can explicitly solve convection on the model grid without needing a convective parameterisation scheme. Their development is returning a step-change in the ability to understand past and future climate change at local scales, supporting the characterisation of extreme weather events that most impact society. The expected improvements of convection-permitting regional climate models (CP-RCMs) are for the representation of hourly precipitation characteristics (i.e., daily cycle, the spatial structure of precipitation, intensity distribution and extremes) towards dynamics matching reality and the representation in detail of surface heterogeneities (e.g., mountains, coastal regions, and urban areas). The capabilities of the CP-RCMs have been exploited within the Highlander project to deliver two new products of paramount significance for Italy, i.e., VHR-REA_IT (Very High-Resolution REAnalysis for ITaly; Raffa et al., 2021) and VHR-PRO_IT (Very High-Resolution PROjections for ITaly; Raffa et al., submitted).

2.6.1 Observations

SCIA-ISPRA

SCIA represents the national system for collecting, processing and disseminating climate data created by the Italian Environmental Protection Agency (ISPRA). It responds to the need to harmonise and standardise processing methods and make available data, indices and indicators helpful in representing and evaluating the state, variations and trends of the climate in Italy. Based on time series of observations from different monitoring networks, decadal, monthly and annual statistics are calculated and represented. Furthermore, the climate data series are subjected to validity checks with homogeneous methodologies, according to the World Meteorological Organisation (WMO) guidelines.

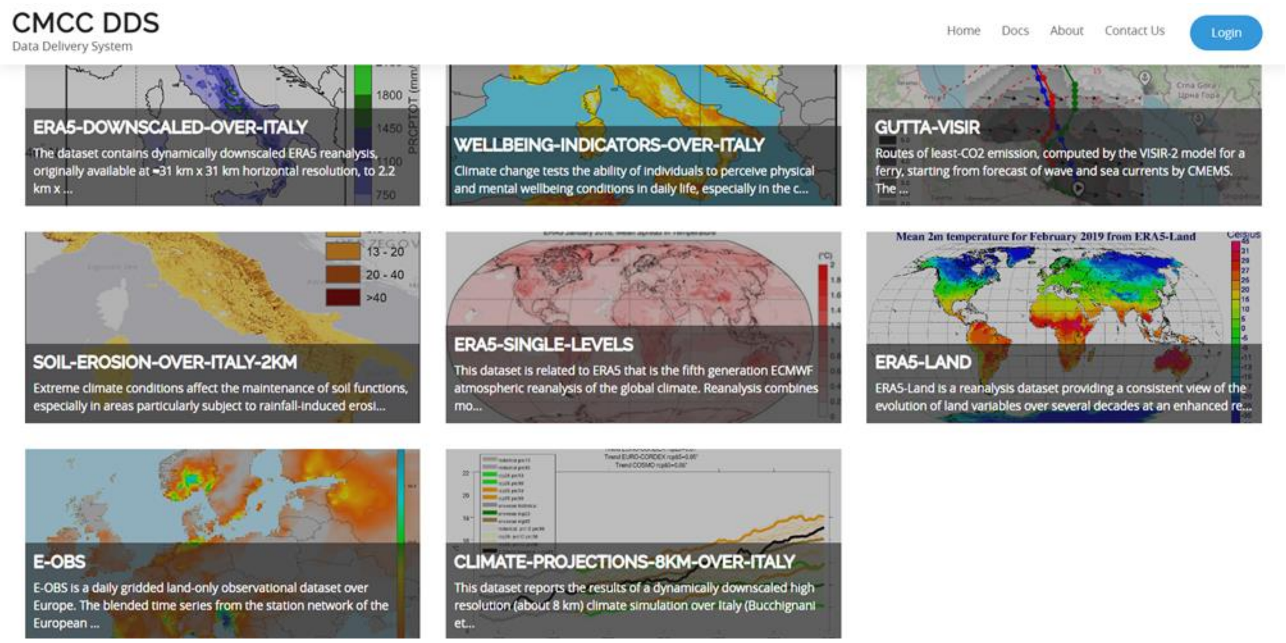
In particular, SCIA-ISPRA (Desiato et al., 2011) is a daily gridded dataset available at a horizontal resolution of approximately 5 km for maximum and minimum temperatures and 10 km for precipitation. This dataset is derived by interpolating data from local weather stations. It covers the 20th century to the present day (i.e., 1981-2021 for temperature; 1951-2021 for precipitation), with varying reliability due to the number of interpolated values available each year. Precipitation data from the manual stations of the hydrographic networks refer to the 24-hour interval from 9 a.m. on the previous day to 9 a.m. on the reference date.

2.6.2 Climate Reanalysis-based

VHR-REA_IT Dataset: Very High-Resolution Dynamical Downscaling of ERA5 Reanalysis over Italy by COSMO-CLM

VHR-REA_IT (Very High-Resolution REAnalysis for ITaly) is a new dataset for recent climate developed by dynamically downscaling ERA5 Reanalysis, initially available at ≃ 31 km horizontal resolution to ≃ 2.2 km resolution. The downscaling activity is performed with the COSMO model in CLimate Mode (COSMO-CLM), switching on the module TERRA-URB to account for the urban parameterisations. The temporal resolution of outputs is hourly (like for ERA5). Runs cover the whole Italian territory (and neighbouring areas according to the necessary computation boundary; Lon = 5°W-20°E; Lat = 36°N-48°N). It provides detailed (in terms of space-time resolution) and comprehensive (in terms of variables) climatological data for the last 30 years (January 1989 - December 2020).

The dataset contains hourly data on a rotated grid with temporal coverage from 01/01/1989 00:00 to 31/12/2020 at 23:00. These data are provided in NetCDF format (dimensions = time, longitude, latitude, single vertical level), generally on single levels (i.e., 2 or 10 metres from surface depending on the selected variables) except for soil moisture available at 7 soil levels (i.e., depth = 1, 3, 9, 27, 81, 243, 729 cm from the surface). The reference coordinate system is WGS84 (EPSG 4326). The outputs have been stored in the CMCC Data Delivery System (DDS). Through the DDS Web User Interface (see Figure 3), users can quickly build queries related to the VHR-REA_IT dataset by choosing variables, geographical area of interest and period. Then, according to these criteria, users can retrieve data using the DDS Python client. Data will also be available in the Highlander platform and could be accessed similarly to the DDS service.



ERA5-DOWNSCALED-OVER-ITALY

The dataset contains dynamically downscaled ERA5 reanalysis, originally available at ≃31 km x 31 km horizontal resolution, to 2.2 km x ...

WELLBEING-INDICATORS-OVER-ITALY

Climate change tests the ability of individuals to perceive physical and mental wellbeing conditions in daily life, especially in the c...

GUTTA-VISIR

Routes of least-CO2 emission, computed by the VISIR-2 model for a ferry, starting from forecast of wave and sea currents by CMEMS. The ...

SOIL-EROSION-OVER-ITALY-2KM

Extreme climate conditions affect the maintenance of soil functions, especially in areas particularly subject to rainfall-induced erosi...

ERA5-SINGLE-LEVELS

This dataset is related to ERA5 that is the fifth generation ECMWF atmospheric reanalysis of the global climate. Reanalysis combines mo...

ERA5-LAND

ERA5-Land is a reanalysis dataset providing a consistent view of the evolution of land variables over several decades at an enhanced re...

E-OBS

E-OBS is a daily gridded land-only observational dataset over Europe. The blended time series from the station network of the European ...

CLIMATE-PROJECTIONS-8KM-OVER-ITALY

This dataset reports the results of a dynamically downscaled high resolution (about 8 km) climate simulation over Italy (Bucchignani et...

Figure 3: Data Delivery System (DDS) User Interface

The VHR-REA_IT dataset features some limitations that users must consider appropriately for correct use. Albeit it is obtained by dynamically downscaling a reanalysis (i.e., ERA5), some biases may be noticed due to the absence of a data assimilation procedure that is hard to be included over the whole domain with the same characteristics due to the resolution of this new dataset. In this sense, it is also important to stress how the use of an urban parameterisation correctly leads to an increase in temperature in urban centres. Such an increase is hard to detect using ERA5 reanalysis or E-OBS observations due to their resolution (at least 5 times lower for E-OBS and about 15 times for ERA5). It could be evaluated against observations provided by urban meteorological stations. However, these measurements are hard to retrieve as synoptic stations are often used.

To sum up, some biases may be encountered; they should be evaluated with punctual observations and appropriately removed through bias correction procedures to feed impact models correctly.

2.6.3 Climate projections

ITALY8km-CM data

Italy8km-CM (Bucchignani et al., 2016; Zollo et al., 2016) is a climate simulation released by the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) over the period 1971–2100 at a spatial resolution of 0.0715° (≃ 8 km) and a 6h output frequency (available at daily resolution), covering the Italian peninsula and some neighbouring areas. It is produced using the COSMO model in CLimate Mode (COSMO-CLM). Italy8km-CM is driven by the GCM CMCC-CM (Scoccimarro et al., 2011), employing the IPCC RCP4.5 and RCP8.5 scenarios.

VHR-PRO_IT Dataset: Very High-Resolution Projections over Italy under IPCC RCP8.5 scenario

VHR-PRO_IT (Very High-Resolution PROjections for ITaly) is a new climate projection dataset developed by dynamically downscaling the GCM CMCC-CM over the period 1989-2050, adopting the IPCC RCP8.5 scenario to the same spatial (≃ 2.2 km) and temporal (hourly) resolution of VHR-REA_IT. An intermediate dynamical downscaling has been conducted by configuring the regional climate model COSMO-CLM at ≃ 8 km over Italy (i.e., Italy8km-CM climate projection). Firstly, ITALY8km-CM has been interpolated to the rotated latitude-longitude grid of the COSMO-Model through the INT2LM program. This tool provides the initial and boundary data necessary to run the COSMO-Model. Generally, data from the global models GME (icosahedral grid point model of DWD), IFS (spectral model of ECMWF) and the regional COSMO-Model itself, as in this case of ITALY8km-CM, can be processed directly, avoiding the pre-processing phase. Finally, a long-term climate simulation has been performed by setting an automatic restart procedure to prevent interruptions of simulation due to the maximum walltime of the SLURM (Simple Linux Utility for Resource Management) partition.

Runs cover the Italian territory (and neighbouring areas according to the necessary computation boundary). It provides a detailed (in terms of space-time resolution) and comprehensive (in terms of meteorological fields) dataset of projected climatological data for at least 60 years (01/1989-12/2050).

The dataset contains hourly data on a rotated grid with temporal coverage from 01/01/1989 at 00:00 to 31/12/2050 at 23:00 (i.e., 1989-2005 for the historical period; 2006-2050 for the future period). These data are delivered in NetCDF format (dimensions = time, longitude, latitude, single vertical level), generally on single levels (i.e., 2 or 10 metres from surface depending on the selected variables), except for soil moisture available at seven soil levels (i.e., depth = 1, 3, 9, 27, 81, 243, 729 cm from the surface). The reference coordinate system is WGS84 (EPSG 4326). The outputs have been stored in the CMCC Data Delivery System (DDS). Through the DDS Web User Interface (see Figure 3), users can quickly build queries related to the VHR-PRO_IT dataset by choosing variables, geographical area of interest and period. Then, according to these criteria, users can retrieve data using the DDS Python client. Data will also be available in the Highlander platform and could be accessed similarly to the DDS service.

It should be pointed out that biases of varying magnitude depending on the area and meteorological variables of interest can hinder using raw climate data as input for impact analysis. To overcome this problem, several approaches, known as bias correction methods, have been proposed in recent years (Maraun, 2016; Casanueva et al., 2020). They can be defined as "all methods that calibrate an empirical transfer function between simulated and observed distributional parameters and apply this transfer function to the simulated output from the considered model" (Maraun et al., 2017). Therefore, the use of bias correction techniques to post-process the data before their use in impact models is strongly recommended.

3. Datasets containing territorial information

The spatial indicators include elevation, slope, aspect, soil type and land cover. The first three are derived directly or indirectly from information from digital elevation models, while specific datasets must be considered for soil type and land cover. Based on this initial screening, the spatial information required as input is the digital elevation model, the soil type dataset and the land cover classification. Such information can be retrieved from the Copernicus Land Monitoring Service and specific Italian databases. This Section describes the primary sources to obtain these datasets.

3.1 Digital Elevation Model

A Digital Elevation Model (DEM) represents the bare ground (bare earth) topographic surface of the Earth , excluding trees, buildings, and other surface objects. It means the altitude (in metres) and is the base to derive slope and aspect. More specifically, a Digital Surface Model (DSM) also captures the natural and built/artificial features of the environment.

Digital Elevation - Global 30 Arc-Second Elevation (GTOPO30)

GTOPO30 is a global digital elevation model (DEM) resulting from a collaborative effort led by the staff at the U.S. Geological Survey's EROS Data Center in Sioux Falls, South Dakota. It features a horizontal grid spacing of 30 arc seconds (approximately 1 kilometre), resulting in a DEM having dimensions of 21,600 rows and 43,200 columns. GTOPO30 has been derived from several raster and vector sources of topographic information, and it has been divided into 33 tiles to ease its distribution. The horizontal coordinate system is decimal degrees of latitude and longitude referenced to WGS84. The vertical units represent elevation in metres above mean sea level. The elevation values range from -407 to 8752 m. In the DEM, ocean areas have been masked as "no data" and assigned a value of 9999. Lowland coastal areas have an elevation of at least 1 m, so if a user reassigns the ocean value from -9999 to 0, the land boundary is maintained. Due to the nature of the DEM raster structure, small islands in the ocean of less than 1 square kilometre are not represented. Further information on the characteristics of GTOPO30, including the data distribution format, the data sources, production methods, accuracy, and hints for users, is found in the GTOPO30 README file.

Copernicus Land Monitoring Service - EU Digital Elevation Model (EU-DEM)

EU-DEM (updated to 1.1 ver.) is a digital surface model (DSM) of EEA members and cooperating countries representing the first surface illuminated by the sensors. It provides Pan-European elevation data at 1 arc-second (+/-30 metres at the equator) postings. The EU-DEM offers complete coverage of the EEA countries (i.e. the so-called EEA39) consisting of 33 member states and 6 cooperating ones. In terms of surface area, the EU-DEM covers 5.84M km². The EU-DEM is a hybrid product based mainly on SRTM and ASTER GDEM (globally relevant and noteworthy datasets) but also publicly available Russian topographic maps for regions north of 60°N latitude. The data is fused by a weighted averaging approach, and it has been generated as a contiguous dataset divided into 5-degree by 5-degree tiles. The spatial reference system is geographic, lat/lon with horizontal datum ETRS89, ellipsoid GRS80 and vertical datum EVRS2000 with geoid EGG08.

TINITALY/01

TINITALY/01 is a DEM in triangular irregular network format (TIN) created for the whole Italian territory under the umbrella of the National Institute of Geophysics and Volcanology (in Italian: Istituto Nazionale di Geofisica e Vulcanologia, INGV) (Tarquini et al., 2007). The DEM was obtained from heterogeneous vector datasets, mainly consisting of elevation contour lines and elevation points from several sources. First, the input vector database was carefully cleaned to get a seamless TIN derived using the DEST algorithm (Favalli and Pareschi, 2004). Then, the TINITALY/01 DEM was converted in grid format (10-m cell size) according to a tiled structure composed of 193, 50-km side square elements. The grid database is freely available as a 10 m-cell grid (in GeoTIFF format) in the UTM WGS 84 zone 32 projection system. It comprises over 3 billion cells and occupies 12 GB of disk memory.

3.2 Soil Information

Soil is a natural resource that can be classified into different soil types, each with distinct characteristics that provide growing benefits and limitations. Therefore, identifying the type of soil is paramount to support the healthy growth of plant life, as well as to derive specific features from selected soil parameters such as organic carbon, pH, water storage capacity, soil depth, cation exchange capacity and clay fraction, total exchangeable nutrients, lime and gypsum contents, sodium exchange percentage, salinity, textural class and grain size.

Harmonized World Soil Database v 1.2 (HWSD)

The FAO - Harmonized World Soil Database (HWSD) is a 30 arc-second (~1 km) raster database with over 15000 different soil mapping units. It combines existing regional and national updates of soil information worldwide (SOTER, ESD, Soil Map of China, WISE) with the information contained within the 1:5000000 scale FAO-UNESCO Soil Map of the World (FAO, 1971-1981). The resulting raster database consists of 21600 rows and 43200 columns linked to harmonised soil property data. The use of a standardised structure allows for the linkage of the attribute data with the raster map to display or query the composition in terms of soil units and the characterisation of selected soil parameters (organic Carbon, pH, water storage capacity, soil depth, cation exchange capacity of the soil and the clay fraction, total exchangeable nutrients, lime and gypsum contents, sodium exchange percentage, salinity, textural class and granulometry). The reliability of the information contained in the database varies with geographical context. More detailed information for this dataset is retrievable from its technical report.

European Soil Database v2.0

The European Soil Database (ESDB) provides Pan-European data for 73 primary or derived soil attributes. Its core data includes information with a spatial coverage on (i) soil classification, (ii) texture, (iii) parent material, (iv) impermeable layer within the soil profile, (v) soil water regime, (vi) most important limitation to agricultural use. This information is also used to estimate further attributes through pedotransfer rules, including hydrological properties like available water capacity and chemical properties like base saturation. The spatial information of soil attributes presented on a nominal scale of 1:1.000.000 is available through the European Soil Data Centre in vector and raster formats with complete coverage of Europe. The Soil Profile Analytical Database (SPADE/M Version 2.0) is a complementary database within the European Soil Data Centre containing profile data for modelling at the European level. Measured and estimated soil profile data is held at the SPADE/M, including information on soil physical (structure, particle size distribution, bulk density etc.) and chemical (organic carbon content, pH, gypsum content etc.) properties that can influence hydrogeological behaviour.

SoilGrids250m

SoilGrids is a global digital soil mapping system that uses state-of-the-art machine learning methods to map the spatial distribution of soil properties across the globe. SoilGrids prediction models are fitted using over 230000 soil profile observations from the WoSIS database and a series of environmental covariates. Covariates were selected from a set of over 400 environmental layers from Earth observation-derived products and other environmental information, including climate, land cover and terrain morphology. The outputs of SoilGrids are global soil property maps at six standard depth intervals (according to the GlobalSoilMap IUSS working group and its specifications) at a spatial resolution of 250 metres. Prediction uncertainty is quantified by the lower and upper limits of a 90% prediction interval. The additional uncertainty layer displayed at soilgrids.org is the interquartile range and median ratio. In addition, maps of the following soil properties are available: pH, soil organic carbon content, bulk density, coarse fragments content, sand content, silt content, clay content, cation exchange capacity, total nitrogen, as well as soil organic carbon density and soil organic carbon stock.

3.3 Land Cover

The land cover represents a key source as it provides spatial information on different types (classes) of physical coverage of the Earth's surface, e.g. forests, grasslands, croplands, lakes, and wetlands. This information is paramount for investigating morphological and functional changes occurring in terrestrial ecosystems and the environment (Sleeter et al., 2018). Land cover change may affect different ecosystem services, resulting in loss of biodiversity, disruption of the hydrological cycle, increase in soil erosion, microclimatic discomfort and runoff (IPCC, 2022).

CORINE Land Cover (CLC) 2018

CLC2018 is one of the Corine Land Cover (CLC) datasets produced within the frame of the Copernicus Land Monitoring Service, referring to the land cover status of the year 2018. CLC service has a long-time heritage (formerly known as the "CORINE Land Cover Programme"), coordinated by the European Environment Agency (EEA). It provides consistent and detailed information on land cover and its changes across Europe.

CLC datasets are based on the classification of satellite images produced by the national teams of the participating countries - the EEA members and cooperating countries (EEA39). National CLC inventories are then further integrated into a seamless land cover map of Europe. The resulting European database relies on standard methodology and nomenclature with the following base parameters: 44 classes in the hierarchical 3-level CLC nomenclature; minimum mapping unit (MMU) for status layers is 25 hectares; minimum width of linear elements is 100 metres. Change layers have higher resolution, i.e. minimum mapping unit (MMU) is 5 hectares for Land Cover Changes (LCC), and the minimum width of linear elements is 100 metres. For Italy, ISPRA provides some thematic insights at Level IV.

The CLC service delivers essential data to support the implementation in key priority areas of the Environment Action Programmes of the European Union (e.g. protecting ecosystems, halting the loss of biological diversity, tracking the impacts of climate change, monitoring urban land take, assessing developments in agriculture or dealing with water resources directives). CLC belongs to the Pan-European component of the Copernicus Land Monitoring Service, part of the European Copernicus Programme coordinated by the European Environment Agency, providing environmental information from a combination of air- and space-based observation systems and in-situ monitoring. Additional details about the CLC product description, including mapping guides, can be found in the technical guideline and the CLC class description guide.

ESA CCI Land Cover map

This dataset provides global maps describing the land surface into 22 classes, defined using the United Nations Food and Agriculture Organization’s (UN FAO) Land Cover Classification System (LCCS). In addition to the land cover (LC) maps, four quality flags are produced to document the reliability of the classification and change detection. To ensure a continuity, these land cover maps are consistent with the series of global annual LC maps from the 1990s to 2015 produced by the European Space Agency (ESA) Climate Change Initiative (CCI), which are also available on the ESA CCI LC viewer.

To produce this dataset, the entire Medium Resolution Imaging Spectrometer (MERIS) Full and Reduced Resolution archive from 2003 to 2012 was first classified into a unique 10-year baseline LC map. This is then back- and updated using changes detected from (i) Advanced Very-High-Resolution Radiometer (AVHRR) time series from 1992 to 1999, (ii) SPOT-Vegetation (SPOT-VGT) time series from 1998 to 2012 and (iii) PROBA-Vegetation (PROBA-V) and Sentinel-3 OLCI (S3 OLCI) time series from 2013. Beyond the climate-modelling communities, this dataset's long-term consistency, yearly updates, and high thematic detail on a global scale have made it attractive for many applications such as land accounting, forest monitoring and desertification, and scientific research.

Land Use/Cover Area frame statistical Survey (LUCAS)

Land Use/Cover Area frame Survey (LUCAS) by EUROSTAT provides harmonised statistics on land use (for instance, agriculture, forestry, recreation or residential use) and land cover (for example, crops, grass, broad-leaved forest, or built-up area) across the European Union. Since 2006, Eurostat has carried out this survey every three years. The latest LUCAS survey, covering all the 27 European Union (EU) countries and the UK, refers to 2018. A new release will be available in 2022.

LUCAS is carried out by direct observations of surveyors in a small area centred on the selected point. Since 2012, all EU countries have been covered, and over more than 330 000 points have been analysed on different land cover types (cropland, grassland, forest, built-up areas, transport network). On these points, the surveyors have examined the land cover and land use, irrigation management and structural elements in the landscape. In addition, a 500 gr topsoil sample is taken in one out of 10 points. These samples are analysed in a laboratory and used for assessing environmental factors, such as updating European soil maps, validating soil models, and measuring the quantity of organic carbon in the soil, which is an essential factor influencing climate change.

In addition to the core LUCAS, specific information is collected in each survey (e.g., topsoil sample and transects). The LUCAS surveys generate three types of information: (i) micro-data containing the statistical information collected in every sample point, (ii) point and landscape photos, and (iii) statistical tables with aggregated results by land cover and land use at the geographical level. The LUCAS surveys are used to monitor the social and economical use of land, ecosystems and biodiversity. Sustainable Development Indicators and Agro-Environmental indicators on soil are examples of LUCAS data use. At the same time, the micro-data collected in it also serve to produce, verify and validate the Corine Land Cover (CLC).

4. Datasets for animal welfare indicators

This chapter describes datasets suitable to calculate animal welfare indicators. In particular, only datasets that present data on the Italian situation at the subnational scale and constantly updated have been reported. The last feature required to be reported below is to be open access.

4.1 LEO (Livestock Environment Opendata)

The LEO project aims to collect in a single database all the information related to livestock in Italy to support and improve the quality of this sector reducing, at the same time, the impact on animal wellness and the environment. The project was born in 2017 in response to the request of the MiPAAF (Ministry of Agriculture, Food and Forestry Policies) to link all useful information for the Italian livestock sector in a single and open-access database.

The leadership of the project is entrusted to the AIA (*Associazione Italiana* Allevatori, one of the SEBASTIEN partner). Other partners are IZSAM – BDN, Istituto Spallanzani – Banca Dati fertilità, IZSUM, Bluarancio, some research institutions (i.e., Università Cattolica di Piacenza, Università della Tuscia, Università di Palermo, ConSDABI), and the main Italian livestock databases (i.e., Banca dati Nazionale – BDN; Banca dati del sistema allevatori – SIAll e Banca dati della Fertilità Maschile).

The 86 datasets are divided into 6 areas.

Climate data (ClimData)

Climatic data datasets provide information about the climate conditions in the areas covered by LEO project, including THI (Temperature Humidity Index). THI is a parameter widely used for the evaluation of the animal heat stress. Another relevant parameter is the THI-LOAD that is an index developed inside LEO project in order to consider the duration of the stress and the subsequent risk assessment, in particular for indoor rearing conditions.

The datasets included in this category are:

- Daily climatic data (precipitation, temperature, humidity, THI);
- Hourly climatic data (temperature, humidity, THI, THI Load)

Precision Livestock Farming (PLFData)

This category contains data collected in real time by smart devices that make them available on the internet. In particular, a device is applied during the milking to collect data such as milk conductivity, milking time and milk amount. This data are particularly useful for both check the animal during a delicate and important phase and for analyse the chemical characteristics of the milk, in particular the presence of high values of CL^- ions that are related to mastitis in cattles.

The datasets included in this category are:

- Average milk electrical conductivity
- Daily milking

Laboratory data

Laboratory data is the broader category of LEO database (50 datasets over 86) and some of its datasets are in common with wellness category. Laboratory data contains results of analysis performed on milk took from animals. When the aggregated data are available, they refer to samples taken from a tank. The concentration of various substances in milk are strictly related to animal conditions such as ketosis or mastitis but they are also indicators of the quality of the product obtained.

The datasets included in this category are:

- Urea - single animal and aggregated data
- Milk cheese-making properties: A30, K20, R, IAC - single animal and aggregated data
- β-hydroxybutyrate (BHBA) - single animal and aggregated data
- Acetone - single animal
- Total saturated fatty acids and total unsaturated fatty acids - single animal and aggregated data
- Milk electrical conductivity - single animal and aggregated data
- De novo fatty acids - single animal and aggregated data
- Mixed fatty acid - single animal and aggregated data
- Preformed fatty acids - single animal and aggregated data
- Milk ph - single animal and aggregated data
- Fatty acids C18:0 and C18:1 - single animal and aggregated data
- Cryoscopic index (milk freezing-point index) - single animal and aggregated data
- Casein - single animal and aggregated data
- Lactose - single animal and aggregated data
- Bacterial load - aggregated data
- IBR (Infectious bovine rhinotracheitis) analysis on milk - single animal and aggregated data
- Differential somatic cells count (DSCC) - single animal
- Progesterone presence in milk - single animal
- Citric acid - single animal
- Fat - single animal
- Proteins - single animal
- Somatic cells (CSS) - single animal

Wellness

This category collects historical data on the health check performed on the animals for the project goals and their results, what illnesses have been analysed and the method used to perform the aforementioned analysis. In addition, two datasets regarding analysis performed on the animal sperm (spermatic membrane integrity and DNA fragmentation index) belong to this category. These two parameter are related to the fertility of the animals.

The datasets included in this category are:

- Health checks
- Animal sperm analysis
- DNA fragmentation index
- Screening

Genetic data

This category groups the datasets that are related to the genotype of the animal. In particular, a set of characteristics that affect animal productivity are reported in these datasets in form of indexes. Being genetic traits, the data reported can be used to select the best breeding animals.

The datasets included in this category are:

- Somatic cell index
- Conception-calving index
- Longevity index
- Food conversion index
- Fitness index
- Lactation continuation index
- First calving age index

Collected data

Collected data contains all datasets related to animals' history with information such as birth, abortions, elimination or farm management and structure.

The datasets included in this category are:

- Dry events
- Abortion
- Elimination
- Unique identifiers
- Pregnancy diagnosis
- Delivery
- Insemination

- Embryonic reabsorption
- Implantation
- Information on farms
- Information on slaughterhouse
- Horses’ characteristics
- Farms
- Slaughter
- Birth
- Cattle entering the farm
- Cattle exiting the farm
- Information on structures
- Herds of horses’ characteristics
- Animal registry

Each dataset, except climatic ones, is structured using the following fields (where applicable):

- IdAnimale - Animal unique code
- IdMisuraPrimaria - measure unique identifier
- codiceIstat – ISTAT code of the district where the event took place
- siglaProvincia – District abbreviation where the event took place
- codiceSpecieAIA – Species AIA code
- codiceRazzaAIA – Breed AIA code
- nomeMisura – measurement name
- valoreMisura – measurement value
- unitaDiMisura – unit of measurement
- giorno – day
- mese – month
- anno – year

In case of aggregated data, the animal ID is substitute by location ID.

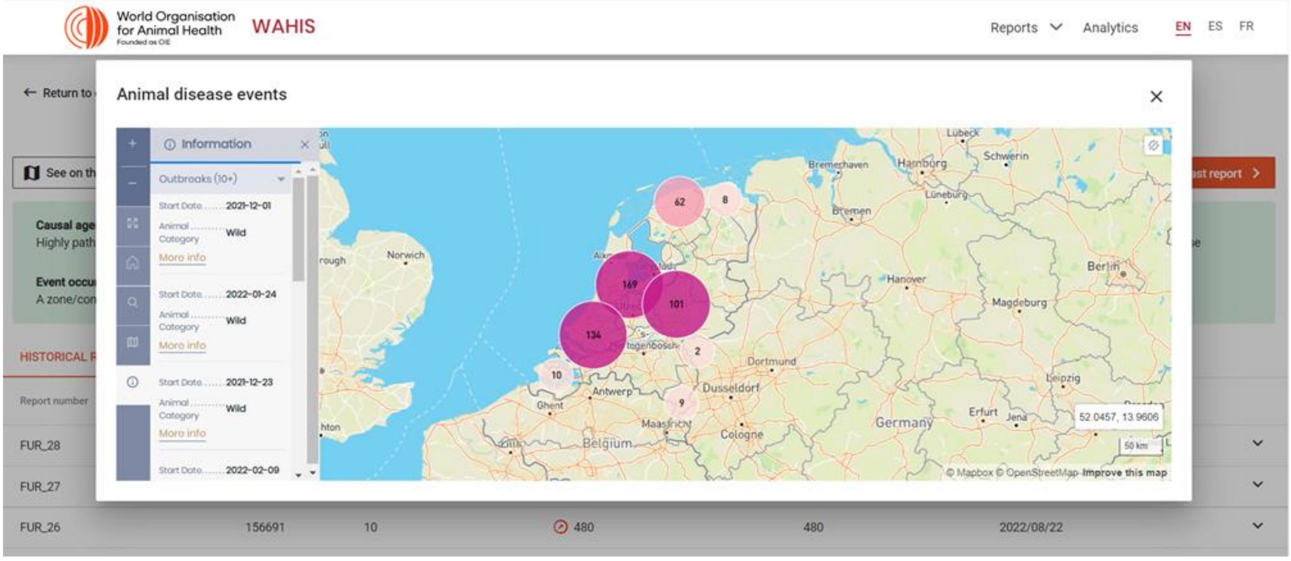
Most datasets contain data since 2017 with monthly updates. Each dataset has its specifications. It is released on “CC-BY-4.0 licence” and can be accessed from www.leo-italy.eu in CSV and JSONL format or SPARQL Protocol. Regarding SPARQL Protocol, it can be used to query the databases and the result can be exported in 10 different format. Examples of queries performed on Collected data, PLFData and Laboratory data databases are reported on the website.

4.2 OIE-WAHIS (OIE World Animal Health Information System)

OIE-WAHIS (OIE World Animal Health Information System) is a comprehensive database through which information on the animal health situation is reported and disseminated worldwide. OIE-WAHIS data reflects the information gathered by the Veterinary Services from OIE Members (182 countries) and non-Members Countries and Territories on OIE-listed diseases in domestic animals and wildlife, as well as on emerging diseases and zoonoses for a total of 205 disease of which 16 bovine specific, 15 regarding sheep and goats and other 27 affecting terrestrial multiple species. All this information can be publicly accessed and visualised. OIE-WAHIS replaces and significantly extends the former web interface named WAHIS, providing access to all reported data since 2005. OIE-WAHIS is divided into three main parts: the early warning system (animal disease events), the monitoring system (six-monthly reports) and additional information collected annually (annual reports).

- The early warning system is designed to inform the international community through alert messages (immediate notifications and follow-up reports) on relevant animal disease events for OIE-listed diseases and emerging diseases.
- Through the monitoring system, countries and territories provide information for all OIE-listed diseases (disease status, quantitative data and control measures) through six-monthly reports.
- Additional information is collected in the annual reports concerning animal population, zoonoses in humans, and capacity and capability of national Veterinary Services, including human resources, vaccine production and diagnostic capabilities.

The data provided are spatially located at the province level, updated every two hours, and released under an open licence (OIE-WAHIS product licence). It covers almost all livestock types. Of particular interest for a service is the early warning system.
Information is divided into reports containing all the information about an outbreak (see Figure 4). The OIE-WAHIS maps can be exported in geojson, shapefile, kml, csv, xlsx, gpx. Although there are methods to be kept informed about new information, it does not seem to be an API for data retrieval. However, reports can be downloaded in excel format, and it is possible to save a report filter.



Animal disease events

Start Date 2021-12-01
Animal Category ...... Wild

Start Date 2022-01-24
Animal Category ...... Wild

Start Date 2021-12-23
Animal Category ...... Wild

Start Date 2022-02-09

FUR_28					
FUR_27					
FUR_26	156691	10	480	480	2022/08/22

Figure 4: OIE-WAHIS user interface

4.3 Sentinel datasets

This paragraph introduces two open datasets of satellite images that could be used to analyse the status of grazing fields. In particular, two missions, among the ones called Sentinel, will be described. The European Space Agency (ESA) launched these missions to support the Copernicus project. Each mission relies on a constellation of satellites to meet revisit and coverage requirements, providing robust data sets for Copernicus services.

Sentinel-1

The Sentinel-1 mission is the European Radar Observatory for the Copernicus joint initiative of the European Commission (EC) and the European Space Agency (ESA). The mission is composed of a constellation of two satellites, Sentinel-1A and Sentinel-1B, sharing the same orbital plane.

Sentinel-1 is designed to work in a pre-programmed, conflict-free operation mode, imaging all global landmasses, coastal zones and shipping routes at high resolution and covering the global ocean with vignettes. This design ensures the reliability of service required by operational services and a consistent long-term data archive built for applications based on long-time series.

The Sentinel-1 makes use of a Synthetic Aperture Radar (SAR) that has the advantage of operating at wavelengths not impeded by cloud cover or a lack of illumination and can acquire data over a site during day or nighttime under all weather conditions. The satellite provides both HH and HV polarised (HH+HV, VV+VH, VV, HH) images with an approximate resolution of 20 m, depending on the chosen mode. Sentinel-1 features four modes and three processing levels. The resolution relies on a combination between the selected mode and level (further details can be found here).
The data provided by this satellite can be used to estimate water content and the presence of vegetation for a specific location.

Sentinel 1 has a 12-day repeat cycle with 175 orbits per cycle for a global orbital period of 98.6 minutes. All SENTINEL-1 SAR data are systematically processed to create pre-defined product types and are available globally, regionally and locally within a defined timescale.

Global products will be systematically generated for all acquired data. They include Level-0, detected Level-1 and Level-2 ocean products. These products are available within 1 hour of observation over NRT areas with a subscription and, in every case, within 24 hours of observation. Regional products are systematically generated over well-defined regions or areas for a subset of the total acquired data. For example, level-1 SLC products are made available within 1 hour of observation over specific NRT areas and systematically over specified areas within 24 hours of observation. The systematic processing approach allows the systematic generation of a pre-defined set of Level-1 products after acquisition (either in NRT or within 24 hours) without ordering required for each product to be generated.

For critical GMES and national services requiring data in quasi-real-time, notably maritime surveillance, data are transmitted by the satellite in real-time for reception by local collaborative ground stations supporting these services. This condition requires that SENTINEL-1 is inside the coverage of these collaborative ground stations. In addition to systematic and routine production, rush processing with high priority at the ground receiving stations will support emergency/security-related observation requests. These observations will be minimised within the strict necessary duration to avoid overriding the pre-defined observation scenario.

Data are distributed on open and free bases by ESA in SAFE (Standard Archive Format for Europe) format.

Sentinel-2

SENTINEL-2 is a wide-swath, high-resolution, multi-spectral imaging mission supporting Copernicus Land Monitoring studies, including monitoring vegetation, soil and water cover, and observation of inland waterways and coastal areas.

The Sentinel-2 mission will provide systematic coverage over the following areas:

- all continental land surfaces (including inland waters) between latitudes 56° South and 82.8° North
- all coastal waters up to 20 km from the shore
- all islands greater than 100 km^2
- all EU islands
- the Mediterranean Sea
- all closed seas (e.g. Caspian Sea).

In addition, the Sentinel-2 observation scenario includes observations following member states or Copernicus Services requests (e.g. Antarctica, Baffin Bay).

The data are provided in two packets reported in Table 1.

Table 1: Sentinel-2 available data packets

Level-1C	Top-of-atmosphere reflectances in cartographic geometry	Systematic generation and online distribution	600 MB (each 100x100 km2)
Level-2A	Bottom-of-atmosphere reflectance in cartographic geometry	Systematic generation and online distribution and generation on the user side (using Sentinel-2 Toolbox)	800 MB (each 100x100 km2)

The resolutions of the SENTINEL-2 Mission and its payload MSI instrument are threefold:

The **temporal resolution** of a satellite in orbit is the revisit frequency of the satellite to a particular location. The revisit frequency of every single SENTINEL-2 satellite is 10 days, and the combined constellation revisit is 5 days without clouds.

The **spatial resolution** of SENTINEL-2 depends on the particular spectral band, and the three groups are reported in Figure 5, Figure 6 and Figure 7.

The **radiometric resolution** is the capacity of the instrument to distinguish differences in light intensity or reflectance. The greater the radiometric resolution, the more accurate the sensed image will be. Radiometric resolution is routinely expressed as a bit number, typically between 8 and 16 bits. The radiometric resolution of the MSI instrument is 12-bit, enabling the image to be acquired over a range of 0 to 4095 potential light intensity values.

Data are distributed on open and free bases by ESA in SAFE (Standard Archive Format for Europe) format.

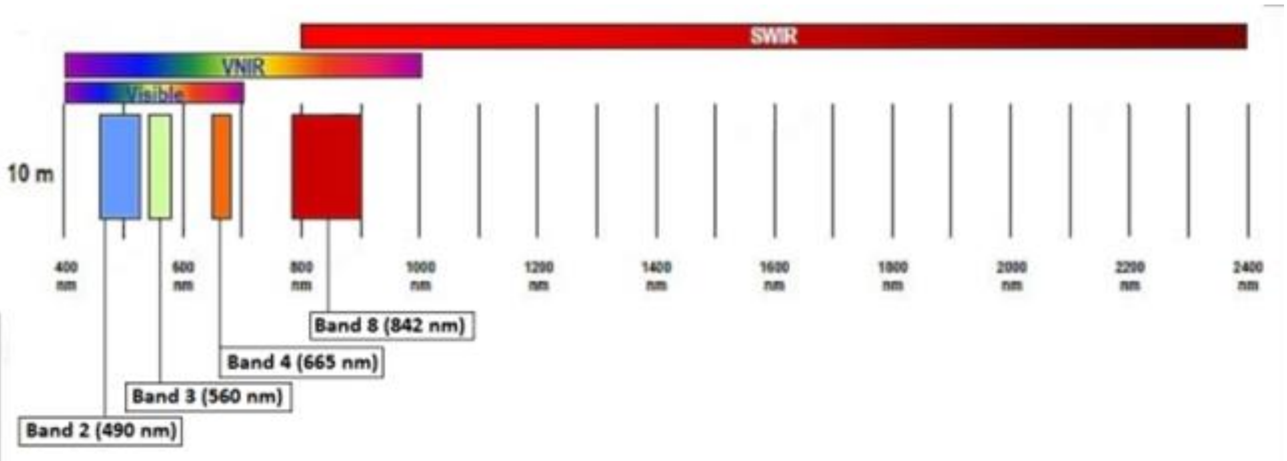


Figure 5: SENTINEL-2 10 m spatial resolution bands: B2 (490 nm), B3 (560 nm), B4 (665 nm) and B8 (842 nm)

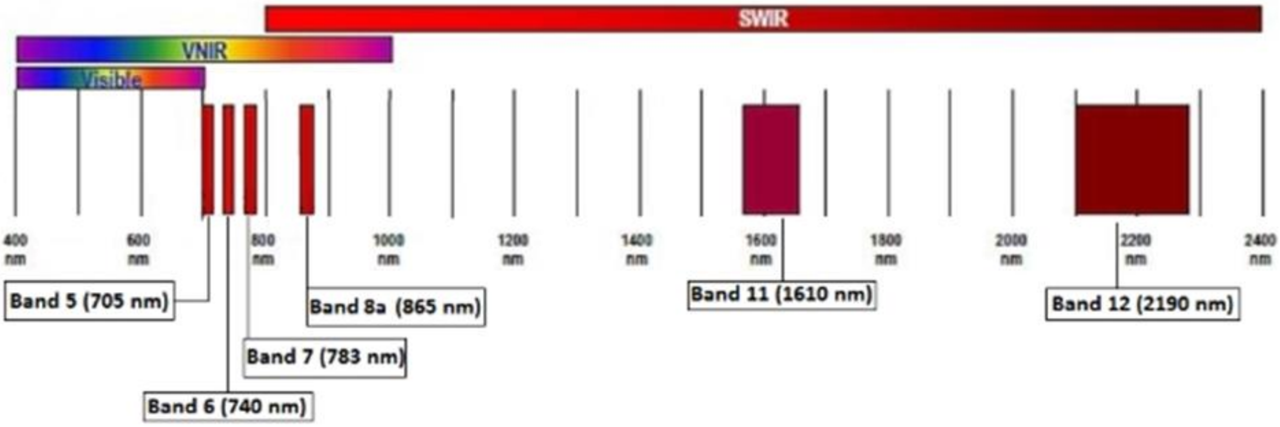


Figure 6: SENTINEL-2 20 m spatial resolution bands: B5 (705 nm), B6 (740 nm), B7 (783 nm), B8a (865 nm), B11 (1610 nm) and B12 (2190 nm)

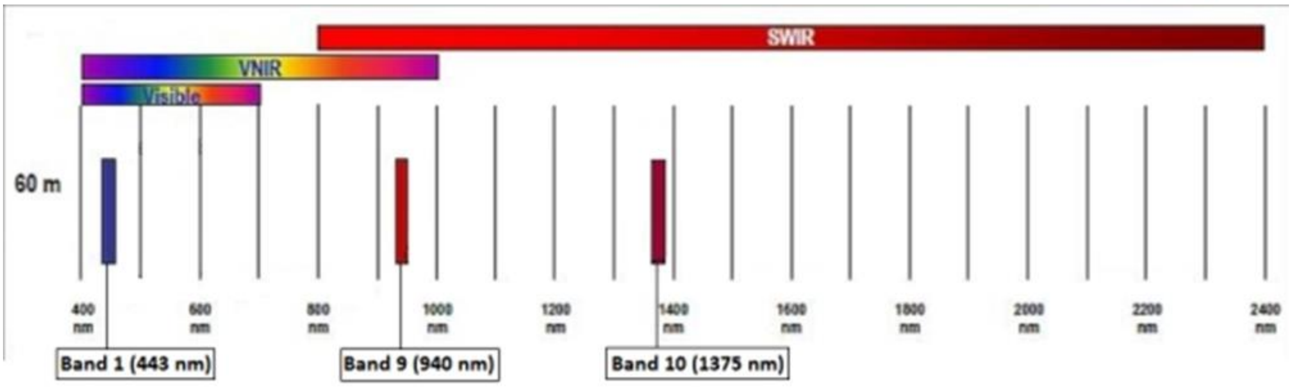


Figure 7: SENTINEL-2 60 m spatial resolution bands: B1 (443 nm), B9 (940 nm) and B10 (1375 nm)

4.4 IoT datasets

In this paragraph, datasets from IoT sensors for real time monitoring are reported and described. Highlander project aims to obtain a smarter management of the lands and to reduce the risks associated with climate change. In particular, the use of high performance computing and the development and use of new technologies in the boundaries of the project itself are the means to obtain its goal. In this environment, a sensor network for animal wellbeing has been developed and tested for the project itself. During Sebastien project, an expansion of this network will take place and the data collected from it will be used.

Highlander datasets

In the framework of Highlander project, some low cost IoT sensors for animal health monitoring have been developed and tested on some cows. In particular, an accelerometer has been placed on the leg of some animals to monitor their movements. Furthermore, a sensor capable of measuring cutaneous temperature has also been developed. Finally, a collar for collecting data about air humidity and temperature as well as acceleration data has been tested. Reported devices are able to expose the collected data over the internet. All the information retrieved are provided through a link where a csv file can be downloaded, the link be related to the identification number of each collar, and will be accessible through Highlander portal. Each file contains all the data collected from one animal. The data from three axes accelerometers are provided as average and standard deviation of the values collected during one minute every two minutes for a total of 20 points per hour. The air temperature and humidity and the animal skin temperature are collected every 10 minutes and sent every hour instead.

Sebatien IoT sensors

The sensors network described in the previous paragraph will be supported and expanded by Sebastien project with the implementation of new features. In particular, sensors for the measurement of animal heartbeat and air quality are being added to the existing system. The heartbeat sensor is expected to be placed on the animal’s ear and will send data with the same frequency of the temperature one. The sensor will be integrated in the existing measurement system, thus the format of the data and the access routine will be the same.

The air quality monitoring sensor will be used in sheds. It is able to collect data regarding CH4, CO2, NO2 and NH3 concentration in addition to air temperature and humidity. Each device is able to send data over the internet. The data will be available in csv format through download while the upload frequency will be once every 30 minutes. The access point for the data described in this paragraph will be Sebastien portal.

5. Conclusions

The Deliverable identifies the existing data sources and ways for their access and extraction in developing SEBASTIEN applications and services. In general, these data sources are under open access licences and usable via API. The document represents a living database, which can be updated if additional datasets will be developed and released in the future.

6. References

Bazile, E.; Abida, R.; Verelle, A.; Le Moigne, P.; Szczypta, C. *MESCAN-SURFEX Surface Analysis*. Deliverable D2.8 of the UERRA Project **2017**. Available online: http://www.uerra.eu/publications/deliverable-reports.html

Bucchignani, E., Montesarchio, M., Zollo, A.L., & Mercogliano, P. High-resolution climate simulations with COSMO-CLM over Italy: performance evaluation and climate projections for the 21st century. *Int. J. Climatol.* **2016**, 36(2), 735–756

Casanueva, A. et al. Testing bias adjustment methods for regional climate change applications under observational uncertainty and resolution mismatch. *Atmospheric Science Letters* **2020**, 21(7), e978

Cornes, R.; van der Schrier, G.; van den Besselaar, E.J.M.; Jones, P.D. An ensemble version of the E-OBS temperature and precipitation datasets. *J. Geophys. Res. Atmos.* **2018**, 123, 9391–9409, doi:10.1029/2017JD028200

Desiato, F., Fioravanti, G., Fraschetti, P., Perconti, W., Toreti, A. Climate indicators for Italy: calculation and dissemination. *Advances in Science and Research* **2011**, 6, 147-150, doi:10.5194/asr-6-147-201

Dutra, E., Muñoz-Sabater, J., Boussetta, S., Komori, T., Hirahara, S., and Balsamo, G. Environmental Lapse Rate for High-Resolution Land Surface Downscaling: An Application to ERA5. *Earth Space Sci.* **2020**, 7, e2019EA000984, https://doi.org/10.1029/2019EA000984.

Eyring et al. Taking climate model evaluation to the next level. *Nat. Clim. Chang.* **2019**, 9, 102–110. https://doi.org/10.1038/s41558-018-0355-y

Favalli, M., Pareschi, M.T.. Digital elevation model construction from structured topographic data: the DEST algorithm, *J. Geophys. Res.* **2004**, 109, F04004, doi: 10.1029/2004JF000150

Giorgi, F.; Gutowski, W.J. Regional dynamical downscaling and the CORDEX initiative. *Annu. Rev. Environ. Resour.* **2015**, 40, 467–490, doi:10.1146/annurev-environ-102014-021217

Haylock, M.R.; Hofstra, N.; Klein Tank, A.M.G.; Klok, E.J.; Jones, P.D.; New, M. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res. Atm.* **2008**, 113, doi: 10.1029/2008jd010201

Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, 146, 1999–2049, doi:10.1002/qj.3803

Hoskins, B. The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Q. J. R. Meteorol. Soc.* **2012**, 139, 573–584. https://doi.org/10.1002/qj.1991

IPCC. *Sixth Assessment Report*. Intergovernmental Panel on Climate Change **2022**. Available at: https://bit.ly/3zB8WUT. Accessed 03/06/2022 [Accepted version subject to final edits]

Isotta, F.A.; Frei, C.; Weilguni, V.; Perčec Tadić, M.; Lassegues, P.; Rudolf, B.; Pavan, V.; Cacciamani, C.; Antolini, G.; Ratto, S.M.; et al. The climate of daily precipitation in the Alps: Development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *Int. J. Clim.* **2014**, 34, 1657–1675, doi:10.1002/joc.3794

Jacob, D.; Petersen, J.; Eggert, B.; Alias, A.; Christensen, O.B.; Bouwer, L.M.; Braun, A.; Colette, A.; Deque, M.; Georgievski, G.; et al. EURO-CORDEX: new high-resolution climate change projections for European impact research. *Reg. Environ. Change.* **2014**, 14, 563–578, doi:10.1007/s10113-013-0499-2

Maraun, D. Bias Correcting Climate Change Simulations-a Critical Review. *Current Climate Change Reports* **2016**, 2, 211–220

Maraun, D. et al. Towards process-informed bias correction of climate change simulations. *Nature Climate Change* **2017**, 7, 764–773

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, *Earth Syst. Sci. Data* **2021**, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-202

Raffa, M., Adinolfi, M., Reder, A., Marras, G.F., Mancini, M., Scipione, G., Santini, M., & Mercogliano, P. VHR-REA_IT Dataset: Very High Resolution Dynamical Downscaling of ERA5 Reanalysis over Italy by COSMO-CLM. *Data* **2021**, 6, 88.

Raffa, M., Reder, A., Marras, G.F., Mancini, M., Scipione, G., Santini, M., & Mercogliano, P. VHR-PRO_IT Dataset: Very High Resolution Projections over Italy under IPCC RCP8.5 scenario. *Scientific Data* (**under review**)

Ridal, M.; Olsson, E.; Unden, P.; Zimmermann, K.; Ohlsson, A. *Uncertainties in Ensembles of Regional Re-Analyses*. Deliverable D2.7 HARMONIE Reanalysis Report of Results and Dataset **2017**. Available online: http://www.uerra.eu/component/dpattachments/?task=attachment.download&id=296.

Schaettler U., Doms G., Schraff C. *A description of the nonhydrostatic regional COSMO-model part VII: user's guide*. Consortium for Small-Scale Modelling **2019** https://www.cosmo-model.org/content/model/documentation/core/cosmo_userguide_6.00.pdf

Sleeter, B.M., Loveland, T., Domke, G., Herold, N., Wickham, J., Wood N. *Land Cover and Land-Use Change*. In Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume II [Reidmiller, D.R., C.W. Avery, D.R. Easterling, K.E. Kunkel, K.L.M. Lewis, T.K. Maycock, and B.C. Stewart (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, **2018** pp. 202–231. doi: 10.7930/NCA4.2018.CH5

Street, R., Buontempo, C., Mysiak, J., Karali, E., Pulquerio, M., Murray, V., et al. How could climate services support disaster risk reduction in the 21st century. *Int. J. Disaster Risk Reduc.* **2019** 34, 28–33. https://doi.org/10.1016/j.ijdrr.2018.12.001

Tarquini S, Isola I, Favalli M, Mazzarini F, Bisson M, Pareschi MT, Boschi E. TINITALY/01: a new Triangular Irregular Network of Italy. *Ann. Geophys.* **2007**, 50(3), 407-25

van Vuuren DP, Edmonds J, Kainuma M, et al. The representative concentration pathways: an overview. *Clim Change* **2011**, 109, 5

Wüest, M., Frei, C., Altenhoff, A., Hagen, M., Litschi, M., and Schär, C.: A gridded hourly precipitation dataset for Switzerland using rain-gauge analysis and radar-based disaggregation. *Int. J. Climatol.* **2010**, 30, 1764–1775, https://doi.org/10.1002/joc.2025

Zollo, A.L., Rillo, V., Bucchignani, E., Montesarchio, M., Mercogliano, P. Extreme temperature and precipitation events over Italy: assessment of high-resolution simulations with COSMO-CLM and future scenarios. *Int. J. Climatol.* **2016**, 36(2), 987–1004