Subcategory. Climate Data Sets and Sources

Dmytro Diadin, O. M. Beketov National University of Urban Economy in Kharkiv

About the Resource

This resource is part of the **Climate Data and Tools** section of the ClimEd Scientific and Educational Platform. It is intended for students, PhD candidates, teachers, researchers, and professionals working in climate-sensitive sectors who use climate data, reanalysis products, climate scenarios and projections, spatial data, and time series. The material can be used in ClimEd courses, practical and laboratory assignments, qualification and research projects, climate risk assessments, and the development of applied climate services.

The resource covers both services for the rapid viewing and downloading of individual climate indicators and specialised platforms for working with large volumes of climate and geospatial data. The description of each service includes its purpose, main functions, available data formats, advantages, limitations, and recommendations for practical use.

Learning Outcomes

- distinguish among ground-based meteorological observations, reanalysis data, and climate model outputs;
- explain the origin and the spatial and temporal resolution of major climate datasets;
- select among E-OBS, ERA5, ERA5-Land, and EURO-CORDEX according to the purpose of the analysis;
- account for uncertainty, limitations in spatial detail, and model errors;
- identify appropriate sources for downloading data and further processing in GIS, R, or Python.

Subcategories for subcategory “Climate Data Sources”

1. Main Categories of Climate Data
2. Ground-Based Meteorological Observations
3. The E-OBS Climate Dataset
4. ERA5 and ERA5-Land Reanalysis Datasets
5. EURO-CORDEX EUR-11 Regional Climate Projections
6. Dataset Comparison and Source Selection
7. Use of the Resource in Education and Research
8. Related ClimEd Resources

1. Main Categories of Climate Data

Climate data used in scientific research and practical analysis in geographic information systems can broadly be divided into three main categories: ground-based meteorological observations, reanalysis data, and climate model outputs. Each type has its own advantages, limitations, and fields of application; therefore, the choice of dataset depends on the task being addressed.

Key principle. A dataset should be selected not because of its popularity, but because it matches the research objective, spatial scale, time period, required variables, and acceptable level of uncertainty.

Category	Origin	Strengths	Main limitations
Ground observations	Direct measurements at meteorological stations	High accuracy at the observation point; long time series	Uneven spatial coverage; interpolation required
Reanalysis	Combination of observations and a numerical model through data assimilation	Continuous spatial and temporal fields; many variables	Model component; limited local detail
Climate modelling	Outputs from global or regional climate models	Assessment of future changes under different scenarios	Model and scenario uncertainty; not a precise point forecast

2. Ground-Based Meteorological Observations

Ground-based meteorological observations are the primary source of climate information. They are generated from long-term measurements conducted at networks of meteorological stations using standardised methods. Such data include air temperature, precipitation, wind speed and direction, atmospheric pressure, relative humidity, and other meteorological variables. The main advantage of ground observations is their high accuracy at station locations. At the same time, they have limited spatial coverage because the density of observation networks varies considerably among countries and regions. To produce continuous climate maps, point measurements must be interpolated, which may reduce accuracy in areas with insufficient station coverage.

Ground-based meteorological observation data for Ukraine (daily minimum temperature, daily maximum temperature, and daily precipitation totals) for 1946–2020 are available in the repository of the Ukrainian Hydrometeorological Institute:

- https://uhmi.org.ua/data_repo/ClimUAd_Ukrainian_gridded_daily
- https://uhmi.org.ua/data_repo/month

When to use. For analysing actual conditions at a station location, studying local trends and extremes, validating model and reanalysis data, and addressing engineering and applied tasks where point-specific accuracy is important.

Please note. Interpolated gridded products derived from station data are not identical to the original station observations. Analyses should distinguish between point measurements and spatially interpolated fields.

3. The E-OBS Climate Dataset

E-OBS (European Observational Gridded Dataset) is one of the most authoritative European datasets of historical climate information derived from actual ground-based meteorological observations. Unlike reanalysis or climate model outputs, E-OBS does not contain model projections; it consists of spatially interpolated values of meteorological variables obtained from a network of thousands of meteorological stations across Europe. It is therefore widely used as a reference dataset for studying the present climate, assessing climate trends, and validating climate models.

The dataset is developed by the European Climate Assessment & Dataset project (https://www.ecad.eu/), coordinated by the Royal Netherlands Meteorological Institute (https://www.knmi.nl/over-het-knmi/about) in cooperation with the national hydrometeorological services of most European countries. The data are distributed through the Copernicus Climate Change Service (https://cds.climate.copernicus.eu/) and are regularly updated as additional observation stations become available. E-OBS was initially created within the European ENSEMBLES project to validate regional climate models, but it is now one of the most widely used climate datasets for Europe. The most convenient source for downloading E-OBS data is the Copernicus Climate Data Store: https://cds.climate.copernicus.eu/datasets/insitu-gridded-observations-europe?tab=overview.

E-OBS covers almost the entire European land area and provides daily values of the main meteorological variables. The principal variables are:

- mean air temperature (TG);
- minimum temperature (TN);
- maximum temperature (TX);
- precipitation total (RR);
- mean sea-level pressure (PP);
- relative humidity (HU);
- mean wind speed (FG);
- total solar (global) radiation (QQ).

The data are available in NetCDF-4 format at spatial resolutions of 0.1° (approximately 11 × 7 km for Ukraine) and 0.25° (approximately 28 × 18 km). For most variables, the time series begins in 1950, while wind-speed data are available from 1980. The spatial domain extends approximately from 25° N to 71.5° N and from 25° W to 45° E, covering almost all of Europe.

An important feature of recent E-OBS versions is the use of an ensemble approach. In addition to the main “best guess” dataset, users can obtain an ensemble of 20 realisations generated through statistical modelling of the interpolation process. This makes it possible to assess spatial interpolation uncertainty and determine the reliability of values in individual regions. Measures of spatial uncertainty, or ensemble spread, can also be obtained for each grid cell.

The main advantage of E-OBS is its observational basis. The data are derived exclusively from actual measurements rather than numerical atmospheric modelling. The dataset also provides a long time series, relatively high spatial detail at the European scale, a common processing methodology across countries, and extensive use in international research. E-OBS is therefore frequently used as a historical baseline for comparison with EURO-CORDEX or other climate model projections and for calculating climate indices in accordance with the recommendations of the Expert Team on Climate Change Detection and Indices.

The dataset also has several limitations. The most important is that accuracy depends on the density of the meteorological station network. In regions with insufficient station coverage, such as mountainous areas, the edges of the domain, or some countries of South-Eastern Europe, interpolation uncertainty increases substantially. Furthermore, a spatial resolution of 0.1° is insufficient to represent local climate features such as urban heat islands, narrow river valleys, or individual mountain slopes. Data quality also varies over time because the number of available stations changes from year to year, while interpolation tends to smooth local temperature and precipitation extremes. E-OBS should therefore not be regarded as a complete substitute for individual station data, particularly when analysing local phenomena or addressing engineering tasks.

Overall, E-OBS is considered one of the best open sources of historical climate data for Europe. Its combination of long observational records, relatively high spatial detail, open access, and a standardised production methodology makes it a core dataset for climate-change research, climate-risk assessment, and validation of regional climate model outputs.

When to use E-OBS. For studying Europe’s historical climate, calculating climate indices, assessing trends, establishing a baseline period, and validating regional climate models.

Main limitation. The reliability of E-OBS depends on station-network density. It should not be treated as a complete substitute for primary meteorological station data when analysing local phenomena, urban microclimates, or engineering applications.

Sources for Downloading E-OBS

1. Copernicus Interactive Climate Atlas: https://atlas.climate.copernicus.eu/atlas
2. IPCC WGI Interactive Atlas: https://interactive-atlas.ipcc.ch/
3. Copernicus Climate Data Store: https://cds.climate.copernicus.eu/

4. ERA5 and ERA5-Land Reanalysis Datasets

ERA5

ERA5 is the latest global reanalysis dataset developed by the European Centre for Medium-Range Weather Forecasts (https://www.ecmwf.int/) within the Copernicus Climate Change Service (https://climate.copernicus.eu/). Unlike datasets based solely on meteorological observations, ERA5 combines a vast array of data from ground stations, satellites, radiosondes, aircraft, marine buoys, and other sources with a modern numerical atmospheric model. This integration is achieved through data assimilation, allowing the most complete possible reconstruction of the atmospheric state at each point in time, including areas where regular observations are unavailable.

ERA5 covers the entire globe at a spatial resolution of approximately 0.25° × 0.25° (about 28 × 18 km in central Ukraine) and provides hourly, daily, and monthly data from 1940 to the present, with continuous operational updates. The dataset includes hundreds of atmospheric, oceanic, and land-surface parameters, including air temperature, precipitation, wind speed and direction, atmospheric pressure, humidity, solar radiation, snow cover, evaporation, soil moisture, and many others. The data are distributed mainly in NetCDF and GRIB formats, enabling efficient use in GIS and scientific analysis systems.

The principal advantage of ERA5 is the spatial and temporal continuity of the data. Unlike conventional meteorological observations, reanalysis provides values for every cell of a regular grid and for every hour. This allows users to analyse both long-term climate trends and individual extreme weather events. ERA5 is therefore widely used for hydrological modelling, studies of extreme heat, droughts and floods, climate-risk assessment, and the calculation of numerous derived climate indicators.

At the same time, ERA5 is not a dataset of direct observations. Its values are partly generated by a numerical atmospheric model, and some variables may therefore differ from actual measurements at meteorological stations. A spatial resolution of 0.25° also cannot represent local terrain effects, urban heat islands, or the microclimate of individual sites. Despite these limitations, ERA5 is regarded as a global standard among reanalysis datasets and is one of the most widely used sources of climate information for scientific research, modelling of natural processes, and regional climate-impact assessments.

ERA5-Land

ERA5-Land is a specialised version of ERA5 reanalysis designed to provide a more detailed representation of processes occurring over land. Unlike ERA5, which models the broader climate system, including the atmosphere, ocean, and land surface, ERA5-Land uses a land-surface model driven by atmospheric boundary conditions from ERA5. Consequently, the dataset has a higher spatial resolution of 0.1° × 0.1° (approximately 11 × 7 km in central Ukraine), allowing the effects of terrain, vegetation cover, and other land-surface characteristics to be represented in greater detail. ERA5-Land provides hourly data from 1950 onward and includes more than 50 variables, among them land-surface and air temperature, precipitation, soil moisture, evaporation, surface and subsurface runoff, snow cover, heat fluxes, and other characteristics of land water and energy balances.

When to use ERA5. For analysing continuous spatial and temporal fields, hourly variability, extreme events, hydrological modelling, derived indicators, and areas with sparse observation networks.

When to use ERA5-Land. For land-surface applications, including soil moisture, evaporation, runoff, snow cover, and heat fluxes, when greater spatial detail than ERA5 is required.

Interpretation. ERA5 and ERA5-Land are reanalysis datasets, not direct observations. For local studies, they should be compared with meteorological station data, and systematic biases in individual variables should be considered.

Sources for Downloading ERA5 and ERA5-Land

1. Copernicus Interactive Climate Atlas: https://atlas.climate.copernicus.eu/atlas
2. Copernicus Climate Data Store: https://cds.climate.copernicus.eu/
3. Google Earth Engine: https://code.earthengine.google.com/

5. The EURO-CORDEX EUR-11 Climate Dataset

EURO-CORDEX (European Coordinated Regional Climate Downscaling Experiment) is the largest international regional climate-modelling initiative for Europe (https://www.euro-cordex.net/060374/index.php.en). It was established within the global CORDEX initiative (Coordinated Regional Climate Downscaling Experiment) (https://cordex.org/) under the World Climate Research Programme (https://www.wcrp-climate.org/about-wcrp/about-history). Its main objective is to produce high-resolution projections of future climate suitable for assessing climate-change impacts on natural and socio-economic systems. Unlike global climate models, which have relatively coarse spatial resolution, EURO-CORDEX uses regional climate models (RCMs) to downscale global model outputs for Europe.

The most widely used configuration is EURO-CORDEX EUR-11, where “11” denotes a spatial resolution of 0.11° × 0.11°, equivalent to approximately 12 × 12 km. In central Ukraine, the grid-cell size is about 12 × 8 km. This level of detail is almost twice as fine as that of CMIP5 global models and allows the effects of terrain, coastlines, large water bodies, and other regional features on climate formation to be represented more effectively. EURO-CORDEX also includes the EUR-44 dataset at 0.44° resolution (about 50 km), although EUR-11 is generally preferred for contemporary applied research.

EURO-CORDEX covers almost all of Europe and adjacent territories, including the whole of Ukraine, approximately from 27° to 72° N and from 22° W to 45° E, and contains simulations for 1950–2100. Model reconstructions are available for the historical period, while future periods are represented by projections under different greenhouse-gas concentration scenarios. The most commonly used scenarios are RCP4.5, which assumes stabilisation of greenhouse-gas concentrations in the second half of the twenty-first century, and RCP8.5, which represents a high-emissions pathway. Newer simulations also include SSP scenarios consistent with the IPCC Sixth Assessment Report, although many current studies in Ukraine continue to use RCP scenarios.

The dataset contains a wide range of climate variables at daily and monthly temporal resolutions. Key variables include mean, minimum, and maximum air temperature, precipitation, wind speed and direction, relative humidity, atmospheric pressure, solar radiation, evaporation, soil moisture, snow cover, and other characteristics. Hundreds of derived climate indicators can be calculated from these variables, including the number of hot days, duration of dry spells, fire-weather indices, growing degree-day totals, extreme precipitation intensity, drought indices, and other indicators required for climate-risk assessment. Data are distributed mainly in NetCDF format in accordance with the international Climate and Forecast (CF) conventions.

One of the most important features of EURO-CORDEX is its ensemble modelling approach. Projections are produced not by a single model but by combinations of several dozen regional climate models driven by different global climate models. Users therefore obtain an ensemble of possible climate futures rather than a single projection. Contemporary studies consequently recommend analysing ensemble means or several models together instead of relying on a single model. This reduces the influence of individual model characteristics and improves the robustness of assessments.

The main advantage of EURO-CORDEX is the high spatial detail of its projections, which enables climate changes to be assessed at the level of regions, river basins, territorial communities, and individual natural systems. A common modelling framework provides internally consistent results across Europe, which is particularly important for transboundary studies and regional climate-risk analysis. EURO-CORDEX data are widely used in climate-change adaptation planning, vulnerability assessments of economic sectors, water-resources modelling, and analyses of drought, flood, wildfire, and other climate-related risks.

Several limitations must nevertheless be considered. Despite their relatively high spatial resolution, the results remain model projections rather than actual observations and contain uncertainty associated both with the climate model and the selected greenhouse-gas emissions scenario. Even a resolution of approximately 12 km cannot fully represent local microclimatic conditions, individual urban heat islands, or small-scale terrain effects. EURO-CORDEX results should therefore be interpreted as plausible projections of changes in climate conditions rather than as precise weather or climate forecasts for a specific point. To improve confidence, they are usually compared with historical E-OBS or ERA5 data and assessed across ensembles of several models rather than using a single simulation.

The main sources for downloading EURO-CORDEX data are:

Copernicus Interactive Climate Atlas: https://atlas.climate.copernicus.eu/atlas

When to use EURO-CORDEX. For assessing possible future climate changes, climate risks and vulnerability, developing adaptation plans, and investigating water resources, droughts, floods, fire danger, and other climate-related impacts.

Good practice. Do not rely on a single model. Analyse an ensemble of model simulations, compare the historical period with E-OBS or ERA5, and clearly state the scenario, model, period, and aggregation method.

Sources for Downloading EURO-CORDEX

1. IPCC WGI Interactive Atlas: https://interactive-atlas.ipcc.ch/
2. Copernicus Climate Data Store: https://cds.climate.copernicus.eu/

6. Dataset Comparison and Source Selection

The table below provides a quick guide to selecting an appropriate dataset according to the nature of the task. The final choice should consider the specific variable, period, territory, availability of primary documentation, and accuracy requirements.

Dataset	Type	Coverage	Resolution	Main application	Key caution
Station data	Observations	Individual locations	Point-based	Local trends, extremes, validation	Uneven coverage
E-OBS	Interpolated observations	Europe	0.1° / 0.25°	Historical climate, indices, validation	Local extremes may be smoothed
ERA5	Reanalysis	Global	approximately 0.25°	Hourly fields, extreme events, modelling	Contains a model component
ERA5-Land	Land reanalysis	Global land areas	approximately 0.1°	Soil, evaporation, runoff, snow	Does not represent point microclimates
EURO-CORDEX EUR-11	Regional climate models	Europe	0.11°	Future changes and adaptation	Model and scenario uncertainty

For historical analysis. Begin with station data, E-OBS, ERA5, or ERA5-Land, depending on the spatial scale and required temporal resolution.

For assessing future climate. Use EURO-CORDEX ensembles or other climate projections. Reanalysis and observations are not projections of future change.

For local decision-making. Combine several sources: primary observations, gridded historical data, and ensemble projections. Where possible, apply statistical bias correction and local validation.