Subcategory. Climate Data Sources

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

About the Resource

This learning resource provides practical guidance on using QGIS and ArcGIS to visualise, process, analyse, and map climate data. It is intended for students, PhD candidates, teachers, researchers, and professionals working with meteorological observations, reanalysis data, and climate projections. The material covers the main climate-data formats, spatial resolution, multidimensional NetCDF files, QGIS installation and interface, opening and visualising data, raster clipping, and mathematical operations.

> The examples are presented primarily for QGIS because it is a free and open-source geographic information system. Equivalent tools are available in ArcGIS Pro, although their names and locations within the interface may differ.

Subcategories for subcategory “Climate Data Sources”

1. Key Features of Climate Data
2. Main Climate Data Formats
3. Spatial Resolution
4. Multidimensional NetCDF Data
5. Installing and Using QGIS
6. QGIS Interface
7. Opening Climate Data in QGIS and ArcGIS
8. Visualising Climate Data
9. Clipping Climate Rasters
10. Mathematical Operations on Rasters
11. Additional learning materials on the topic

1. Key Features of Climate Data

Climate data have a number of specific characteristics that distinguish them substantially from most other spatial data used in geographic information systems. First, they are continuous in both space and time, meaning that every point in a territory is characterised by certain climate parameters that are constantly changing. Analysis can be carried out at different spatial levels—from global and continental to regional and local—and at different temporal scales, ranging from hourly and daily observations to monthly, seasonal, multi-year values and climate normals.

Climate information is characterised by a considerable diversity of indicators. Although the number of basic meteorological variables is relatively limited (air temperature, precipitation, wind, humidity, atmospheric pressure, solar radiation, etc.), hundreds of derived climate indicators are calculated from them. These include, for example, the number of hot or frost days, the duration of dry periods, accumulated effective temperatures, fire danger indices, extreme precipitation indicators, drought indices, and many others. Derived indicators are most often used to assess climate risks, territorial vulnerability, and to plan adaptation measures.

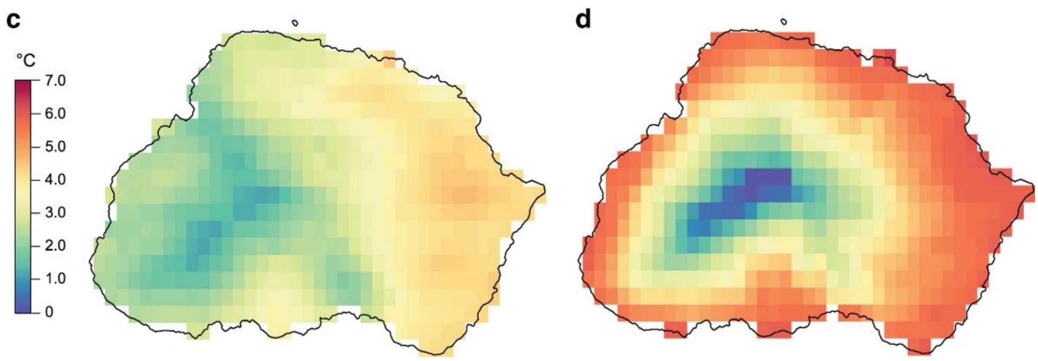
Another important feature is the diversity of climate data sources. GIS applications use meteorological observations, reanalysis data that combine measurements with numerical atmospheric models, and the outputs of climate models projecting future changes. Numerous global climate models and their configurations, represented in different generations of international CMIP projects, are used for climate projections. Simulations are performed under several future greenhouse gas emission scenarios, making it possible to assess a wide range of potential climate changes.

All of this results in extremely large volumes of information. For each climate indicator, there may simultaneously be historical observation series, current estimates, projected data for different time horizons, absolute values, and relative changes compared with a baseline period. The combination of numerous time steps, spatial layers, models, and scenarios produces datasets ranging from tens of gigabytes to many terabytes. Climate data can therefore reasonably be classified as Big Data.

At the same time, working with climate data requires critical assessment of data quality. The accuracy of results may be affected by spatial resolution, incomplete observation series, model uncertainty, differences between climate development scenarios, and the methods used to process the information. Therefore, using climate data in GIS requires not only proficiency in spatial analysis tools but also an understanding of data provenance, limitations, and uncertainty. This is essential for obtaining reliable results in scientific research and decision-making.

2. Main Climate Data Formats

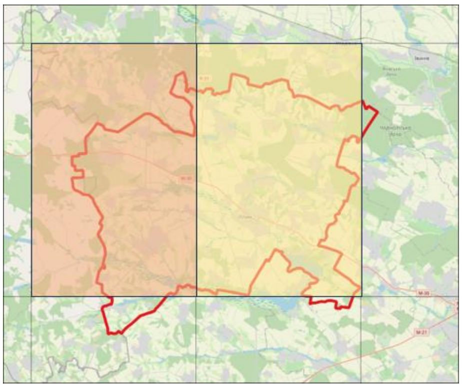
In geographic information systems, climate data are mainly used in raster format (Figure 1), which provides a spatial representation of continuous climate fields. The most common formats are GeoTIFF (.tif, .geotiff), GRIB (.grib), and NetCDF (.nc). GeoTIFF is most often used for individual raster layers, such as maps of mean annual temperature or precipitation. GRIB is widely used for the operational exchange of meteorological data and numerical weather forecasts, whereas NetCDF is an international standard for storing multidimensional climate and oceanographic data.



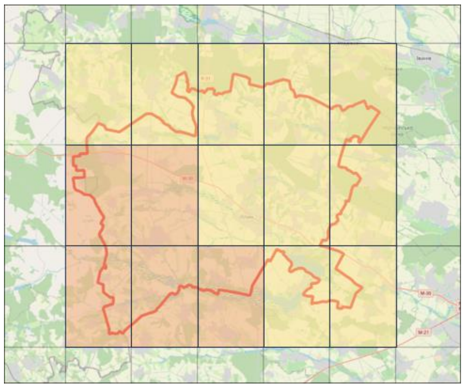
https://www.nature.com/articles/sdata2018177/figures/1*Figure 1 – Example of raster air-temperature data (source: https://www.nature.com/articles/sdata2018177/figures/1)*

3. Spatial Resolution

The spatial resolution of climate data can vary considerably depending on the source and intended purpose of the dataset. For global climate models, it is generally between 2.5° and 1°, corresponding to hundreds of kilometres on the ground. By contrast, modern regional climate models, reanalysis products, and high-resolution climate datasets may have resolutions of 0.25°, 0.11°, or even approximately 0.1°, making it possible to assess climate change at the level of individual administrative areas, river basins, or communities (Figure 2).



0.25 × 0.25 degrees (~28 × 18 km)

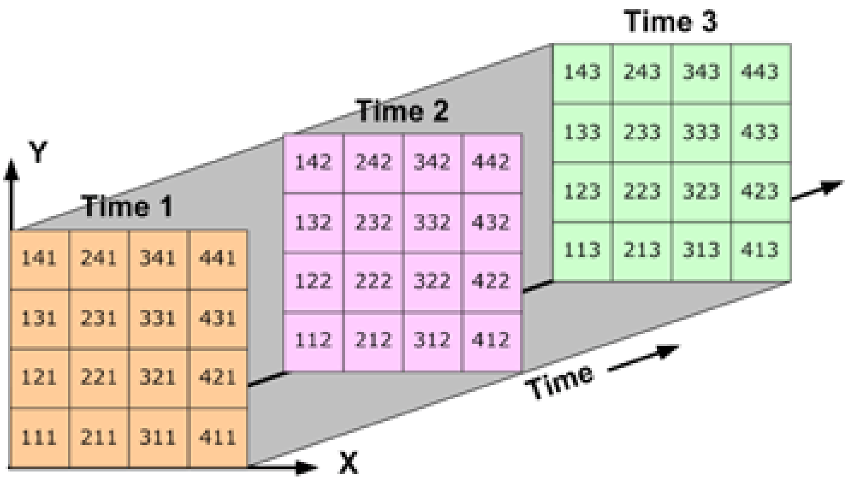


0.1 × 0.1 degrees (~11 × 7 km)

Figure 2 – Comparison of raster cell sizes at different spatial resolutions at the latitude of central Ukraine (49° N)

4. Multidimensional NetCDF Data

For large time series, the NetCDF (Network Common Data Form) format is most commonly used. It was specifically designed to store multidimensional scientific data. Unlike conventional raster formats, a single NetCDF file can contain several climate variables at once, such as minimum, mean, and maximum air temperature, together with additional time and spatial dimensions. For example, one file may contain daily temperature values for every cell of a regular grid over a 30-year period, effectively combining thousands of individual raster layers into a single data structure.



https://desktop.arcgis.com/en/arcmap/latest/manage-data/netcdf/fundamentals-of-netcdf-data-storage.htm*Figure 3 – Structure of spatial data within a multiband NetCDF raster (source: https://desktop.arcgis.com/en/arcmap/latest/manage-data/netcdf/fundamentals-of-netcdf-data-storage.htm)*

In most modern GIS applications, each time slice or individual variable from a NetCDF file can be displayed as a separate raster layer. This makes it possible to perform spatial analysis in the same way as with conventional rasters. In particular, any layers can be used in the Raster Calculator to perform mathematical operations and create new datasets. For example, users can calculate the difference between mean temperature in baseline and future climate periods, determine changes in precipitation, calculate climate anomalies, or create integrated climate-hazard indicators by combining several climate indices. NetCDF is therefore one of the key formats for climate-change analysis in modern geographic information systems.

5. Installing and Using QGIS

QGIS is a free and open-source geographic information system available for Windows, Linux, and macOS. To install it, download the installation package from the official QGIS website (https://qgis.org/download/) and run the standard installation wizard. It is recommended to use the current Long Term Release (LTR) available from the official QGIS website, as it generally provides the best compatibility with commonly used spatial data formats and additional modules.

QGIS supports the full workflow for working with spatial data. Its main capabilities include loading and displaying vector and raster layers, editing spatial features, performing spatial analysis, processing data with geoprocessing tools, creating thematic maps, and preparing them for printing. The software supports a wide range of geospatial formats, including Shapefile, GeoPackage, GeoTIFF, NetCDF, GRIB, CSV, GeoJSON, and many others. QGIS also integrates with the Python programming language, enabling the automation of large-scale geospatial data processing.

For reliable operation and convenient use of reference resources, we recommend setting the QGIS interface to English: Settings – Options – General – User interface translation. By default, the QGIS interface language after installation corresponds to the operating system language.

6. QGIS Interface

The QGIS interface consists of several main components (Figure 4). The central area is occupied by the Map Canvas, where spatial data are displayed. On the left is the Layers Panel, which is used to manage layer order, visibility, and properties.

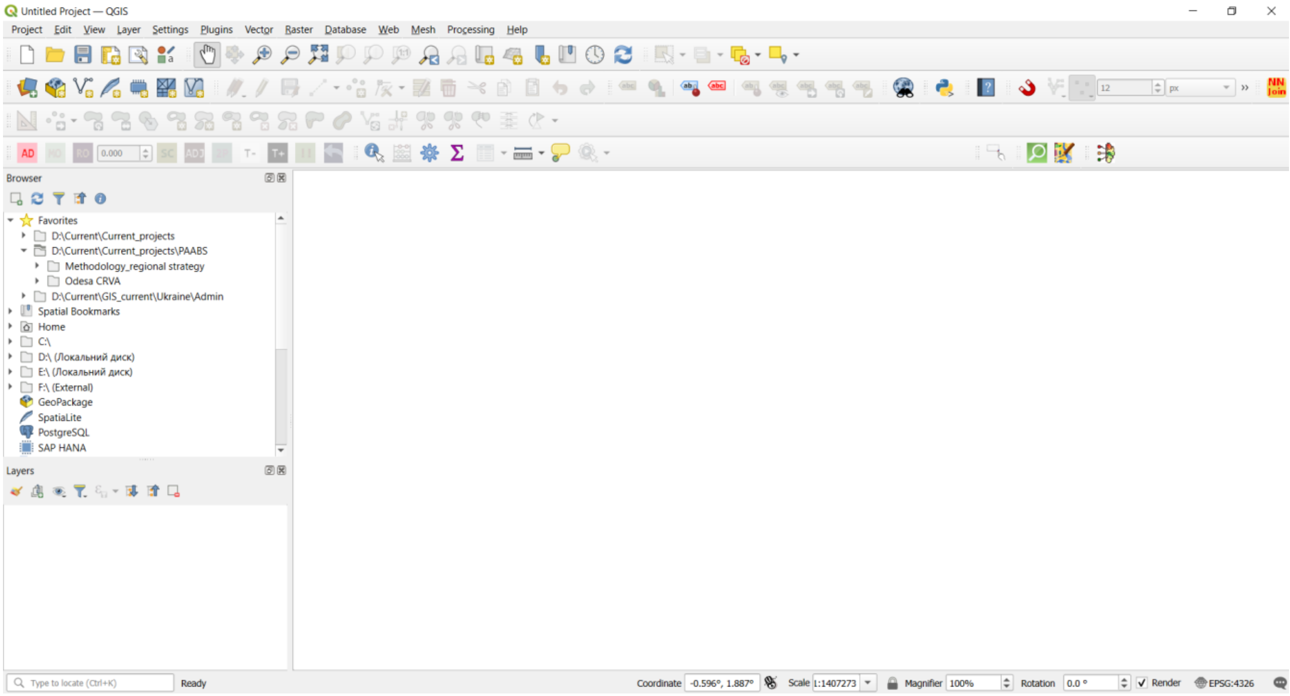


Figure 4 – QGIS desktop interface

The menu bar and toolbars at the top provide access to the main program functions. Additional panels may be opened on the right or at the bottom, including the Processing Toolbox, Browser Panel, and Layer Styling Panel. The status bar at the bottom of the window displays cursor coordinates, map scale, the project coordinate reference system, and other service information. This interface structure supports both simple mapping projects and complex spatial-analysis tasks involving climate data.

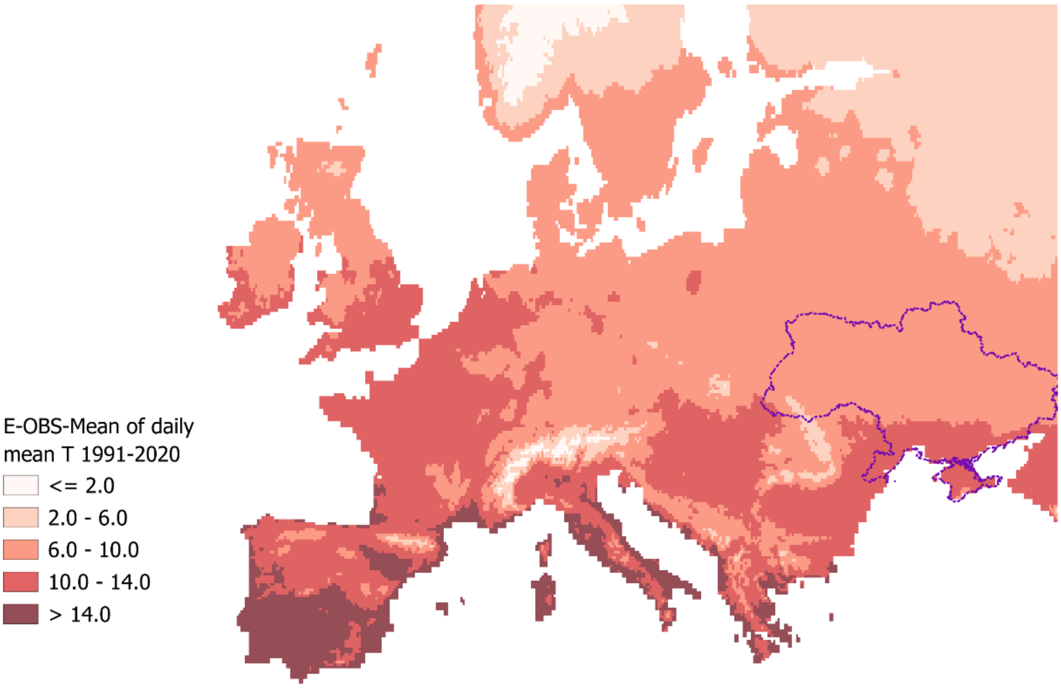
7. Opening Climate Data in QGIS and ArcGIS

Spatial climate data are loaded into GIS in much the same way as other raster datasets, although the procedure depends on the file format. Rasters in GeoTIFF (.tif) and GRIB (.grib) formats can be opened directly as ordinary raster layers using the raster data import tools in the desktop versions of QGIS or ArcGIS. Once loaded, they can be visualised, styled, analysed, and used in spatial-analysis tools like any other raster data.

NetCDF (.nc) files containing multidimensional time series or several climate variables are opened using specialised multidimensional-data tools. After loading the file, the user can select the required climate variable (for example, temperature or precipitation), time interval, or specific time slice, which is displayed as a separate raster layer. Modern versions of QGIS and ArcGIS Pro also support direct work with multidimensional rasters, allowing users to view time series, aggregate them, calculate statistical characteristics, and perform raster calculations without first exporting individual layers. This greatly simplifies the analysis of large climate datasets and improves the efficiency of work with projected climate scenarios.

8. Visualising Climate Data in QGIS and ArcGIS

After loading climate raster data into QGIS or ArcGIS, the visualisation method should first be configured according to the type of indicator. For continuous climate variables such as temperature, precipitation, or wind speed, continuous colour gradients are recommended because they provide a smooth representation of spatial variation. For discrete or classified indicators, such as climate-hazard or vulnerability classes, stepped colour palettes with clearly defined class boundaries are more appropriate (Figure 5).



https://cds.climate.copernicus.eu/*Figure 5 – Example of climate-data visualisation using value categories (data source: https://cds.climate.copernicus.eu/)*

> GOOD PRACTICE. When comparing maps for different periods, models, or scenarios, use the same colour-scale limits. This ensures that the results are compared visually in a consistent and accurate manner.

Particular attention should be paid to selecting the range of values to be displayed. The automatic scaling applied by GIS when a layer is loaded is not always optimal, especially when several maps representing different periods or climate scenarios need to be compared. In such cases, the same manually defined minimum and maximum values should be used for the colour scale on all maps to ensure correct visual interpretation.

For maps showing changes relative to a baseline period, such as changes in temperature or precipitation, divergent colour scales should be used. Their central value, usually zero, represents no change, while increases and decreases are shown with contrasting colours. This approach greatly facilitates the analysis of spatial patterns of climate change.

> DATA INTERPRETATION. For each map, specify the data source, time period, spatial resolution, model, scenario, variable, and units of measurement.

When preparing cartographic materials, it is also recommended to include a legend specifying the units of measurement, information about the climate-data source, spatial resolution, time period, climate model, and scenario used in the calculations. This ensures correct interpretation of the results and improves their reproducibility in research and practical applications.

9. Clipping Climate-Data Rasters

When analysing climate data, it is often necessary to limit them to a study area, such as an administrative region, territorial community, river basin, or protected area. In QGIS and ArcGIS, this is done by clipping or masking the raster using a polygon boundary (Clip Raster by Mask Layer in QGIS or Extract by Mask in ArcGIS). As a result, only the part of the raster located within the specified polygon is retained, while values outside it are replaced with NoData or removed entirely from the new file (Figure 6).

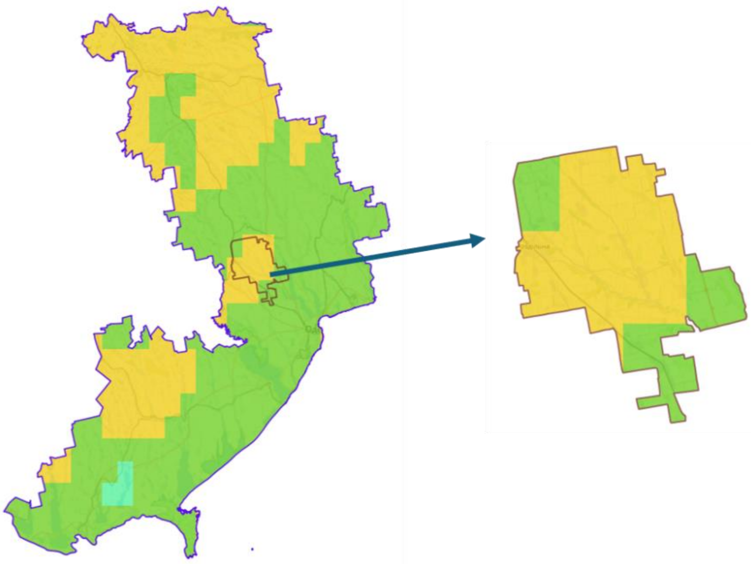


Figure 6 – Example of clipping a climate-data raster to a community boundary

> IMPORTANT. Before clipping, check that the raster and vector layers use the same coordinate reference system. Differences between coordinate reference systems may cause incorrect layer alignment or displacement of the results.

Before clipping, it is necessary to ensure that the raster and vector data use the same coordinate reference system. If the systems differ, the vector boundary or raster should first be reprojected; otherwise, the layers may not align correctly or the results may be displaced. It is also advisable to check polygon geometry, especially if the polygons were created by merging or editing boundaries.

Clipping can substantially reduce data volume, accelerate subsequent calculations, and focus the analysis exclusively on the study area. This is particularly important when working with high-resolution climate rasters or multidimensional NetCDF files that may contain very large amounts of information. For time series, clipping is generally performed once after selecting the required variable or time slice, after which all subsequent operations are carried out on rasters restricted to the study-area boundaries.

When analysing administrative units or river basins, clipping is one of the first stages of data preparation before calculating mean values, spatial statistics, integrated climate indicators, or producing thematic maps. Using the same polygon boundary for all climate indicators ensures spatial consistency and improves the reliability of subsequent comparisons across climate scenarios and time periods.

10. Mathematical Operations on Rasters

One of the most important operations in climate-data analysis in QGIS and ArcGIS is the use of the Raster Calculator to perform mathematical operations between raster layers. This tool is most commonly used to determine the magnitude of climate change by comparing projected values with a baseline climate period. For example, the difference between mean annual temperature for 2021–2040 and the mean value for 1991–2020 is calculated by simply subtracting the corresponding rasters. The result is a new raster layer in which each cell contains the projected temperature change for a specific location (Figure 7).

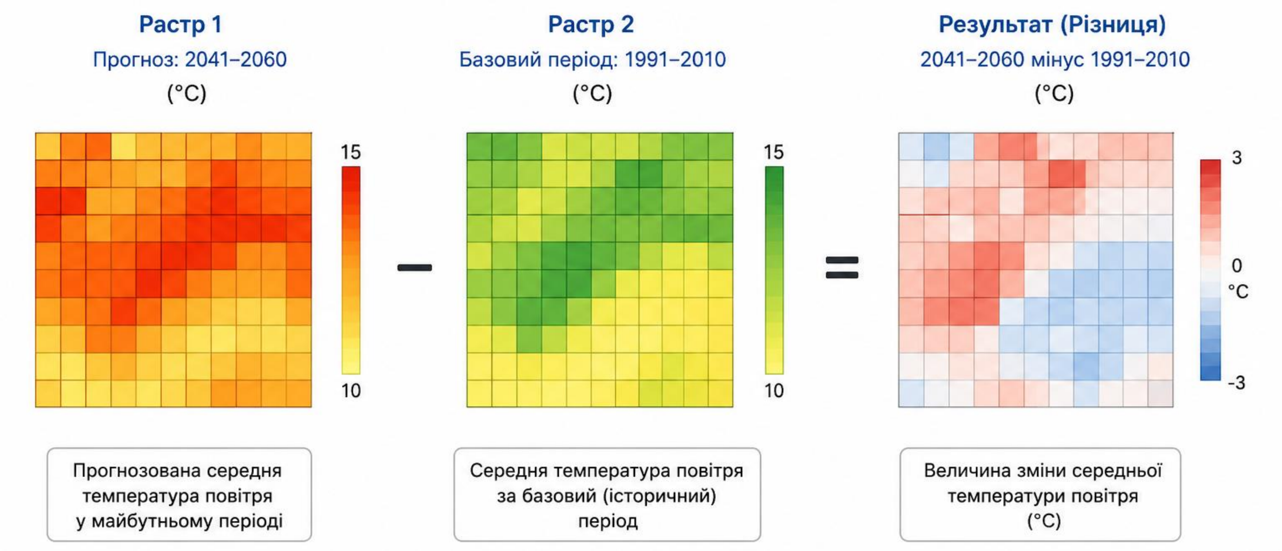


Figure 7 – Principle of Raster Calculator operation

A similar approach is used to analyse other climate indicators, including precipitation, wind speed, the duration of dry periods, and the number of hot or frost days. Depending on the nature of the indicator, both absolute changes (for example, °C, mm, or number of days) and relative percentage changes can be calculated. This enables meaningful comparison of indicators with different units of measurement or substantially different absolute values.

The Raster Calculator can do much more than simple addition or subtraction. It supports complex mathematical expressions, logical operations, conditional functions, and simultaneous analysis of several raster layers. For example, users can calculate the mean of several climate models to obtain an ensemble projection, determine the ratio of actual precipitation to potential evapotranspiration as an indicator of moisture conditions, multiply climate indicators by weighting coefficients when constructing integrated climate-hazard or vulnerability indices, and apply conditional operators to classify values according to specified thresholds. For example, cells can automatically be assigned different scores depending on whether the projected number of hot days exceeds a certain threshold.

The Raster Calculator is also widely used to create integrated maps by combining several climate indicators. Individual indicators may first be normalised or reclassified and then combined using weighting coefficients into a single index of climate hazard, exposure, sensitivity, or vulnerability. This makes the tool one of the basic means of spatial climate-data analysis in modern geographic information systems, widely used both in scientific research and in the preparation of regional climate-change assessments and adaptation strategies.

11. Additional learning materials on the topic

1. A Gentle Introduction to GIS: https://docs.qgis.org/latest/en/docs/gentle_gis_introduction/index.html
2. QGIS Training Manual: https://docs.qgis.org/latest/en/docs/training_manual/index.html
3. QGIS Tutorials and Tips: https://www.qgistutorials.com/en/
4. Video Tutorials: https://jaiindimple.github.io/posts/qgis-playlist/