

Syllabus

Course Title

Management of climate data databases

General information

General description of the required training/preparation, outlining the main objectives and providing an explanation of the need for training/preparation at the organization/country/region level.

Climate change is a fundamental and complex challenge for both natural systems and human society. As Earth system models become more sophisticated and observation networks expand, the volume, velocity, and variety of climate data are growing exponentially. Contemporary scientific research indicates that the cascading effects of climate variability — from the attribution of extreme events to long-term ecosystem changes — may be more severe and interconnected than previously anticipated. Addressing these challenges requires not only specialized domain knowledge but also advanced skills in managing and curating large and complex datasets.

The primary objective of this course for postgraduate students is to develop advanced competencies in PhD candidates in Earth sciences that are necessary for designing, querying, and managing climate databases to support high-level scientific research and the provision of climate services. Course participants will go beyond simple data acquisition and will master the architecture of relational and non-relational databases, learning how to ensure data integrity and accessibility.

This course bridges the gap between raw meteorological observations and practical scientific conclusions. Postgraduate students will acquire skills in obtaining and processing climate information from a wide range of sources, including remote sensing data, reanalysis outputs, and general circulation model (GCM) outputs. Focusing on the needs of the end user — whether in climate policy, risk assessment, or geophysical research — participants will learn to develop data structures that support rigorous analyses of climate change vulnerability and resilience within defined geographical regions. The course program emphasizes best practices in data management, metadata standards, and the use of distributed data services to support reproducible and scalable research in Earth sciences.

Audience

The primary target audience of the course and any secondary audience, if it may influence decisions regarding the course structure or content.

The expected level of knowledge and skills of the primary audience (current or minimally required), as well as other factors (for example, cultural characteristics, level of technical proficiency, internet access) that should be considered when planning the course, since they may affect the choice of teaching methods, materials, and approaches to audience interaction.

The primary target audience consists of PhD students in specialty E4 “Earth Sciences” and related natural science specialties (E2 “Ecology,” G18 “Geodesy and Land Management,” 104 “Physics and Astronomy” with specialization in geophysics) who are conducting dissertation research in climatology, meteorology, hydrology, oceanography, or physical geography.

Secondary audience (influential in decision-making)

1. Academic supervisors of PhD students – experienced researchers (Doctors and Candidates of Sciences) who define dissertation topics and are interested in ensuring that their students master modern, internationally recognized data management methods (FAIR principles, research reproducibility). Their requirements directly influence the course content.

2. Heads of departments and institutes / dean’s offices – individuals responsible for decisions regarding the inclusion of the course in the doctoral curriculum, allocation of hours, and resources. They are interested in the course’s compliance with educational standards, accreditation requirements, and the overall relevance of the discipline for enhancing the institution’s ranking.

3. Representatives of the National Hydrometeorological Service and sectoral institutes (potential employers) – practitioner specialists who may act as stakeholders of the educational program. Their need for professionals capable of working with modern databases (for example, for preparing climate reports or forecasts) influences the practical component of the course and the prioritization of certain technologies.

4. Educational program administrators and guarantors of educational-scientific programs (ESPs) – individuals responsible for the integrity of doctoral training programs. They ensure that the course organically complements other disciplines (for example, “Statistical Methods in Climatology,” “Modeling of Geophysical Processes”) and does not duplicate them.

Expected level of knowledge and skills (entry requirements):

Fundamental knowledge: Understanding of the basic principles of atmospheric physics, oceanography, hydrology, or related disciplines. Familiarity with the concept of “climate change” at the higher education level (Master’s degree).

Data handling skills: Basic experience working with tabular data (Excel, Google Sheets). General understanding of what databases are (for example, where observational data are stored). Initial skills in scientific computing environments such as Python (with Pandas and Xarray libraries) or R (with tidyverse and stars packages) are desirable but not mandatory.

English language proficiency: Ability to read technical documentation and scientific articles in English (as most database and tool documentation is available in English).

Needs and motivation:

✓ The primary need is to learn how to effectively work with large volumes of climate data (models, reanalysis, satellites) for dissertation research.

✓ Aspiration to master tools that will enhance competitiveness in the international academic job market (publications in leading journals, participation in international projects).

✓ The need to systematize knowledge on how to properly store, process, and document their own field or experimental data.

Competencies

Training needs at the individual level or at the organization/country/region level, as well as a description of how these needs were identified and recognized as relevant.

The competencies that the training will target.

C1. Assessment of the impacts of climate change, climate variability, and climate extremes on various sectors (society, environment, economy, etc.) at different scales, taking into account the full range of interconnections between components of the climate system and transdisciplinary interactions with society.

Learning outcomes and performance criteria

Learning outcomes and performance criteria formulated based on the knowledge and skills to be acquired during the course.

LO1. Explain the process of creating and managing climate datasets, including the sequential application of data rescue, quality control, homogenization, and integration methods within a climate data management system.

LO2. Describe the geographical characteristics and historical events that influence climate observation networks, including political events and changes in observation methods and methodologies.

LO3. Demonstrate skills in using various software tools required for climate data management, including office suites, statistical software, GIS, and specialized packages for data quality control and homogenization.

LO4. Design a climate and metadata database using a climate data management system, including raw, quality-controlled, and homogenized data.

LO5. Create, document, and analyze climate data for specific purposes, including metadata, and explain their applications and associated uncertainties.

Course Content

Provide a content plan that corresponds to the course aims and learning outcomes. This may be the course outline as presented to students, but not necessarily a full curriculum.

Include a general list of all topics considered necessary for coverage. If you think it helps clarify the situation, indicate what will NOT be covered.

Module 1. Workflow and quality control in climate data management.

Explain the process of creating and managing climate datasets, including the sequential application of data rescue methods, quality control, homogenization, and integration into a climate data management system.

Module 2. Geographical and historical context of climate data.

Describe geographical characteristics and historical events that influence climate observation networks, including political events and changes in observation methods and methodologies.

Module 3. Climate data processing and software tools.

Demonstrate skills in using various software tools required for climate data management, including office suites, statistical software, GIS, and specialized packages for data quality control and homogenization.

Module 4. Organization of climate data and metadata management.

Design a climate data and metadata database using a climate data management system, including raw, quality-controlled, and homogenized data.

Module 5. Creation and analysis of climate datasets.

Create and document climate datasets for specific purposes, including metadata, and explain their potential applications and associated uncertainties.

Learning approaches and implementation methods

List the teaching approaches (learning methods) that will be used and explain why they were chosen. For example: classroom learning, online learning, blended learning, workplace learning, online self-learning resources, coaching or mentoring, etc.

Classroom-based learning for the main lecture topics and practical sessions. Some practical exercises may be completed at home or in groups in classrooms, allowing students to support each other. Learning materials will be distributed online, and the lecturer will be available online at designated times for questions and consultations.

Learning tools: online learning, blended learning, workplace learning, online self-study resources.

Strategies of learning

Consider which learning strategies you will use. Provide justification for why you want to apply them, including reasons why they will help participants achieve the planned learning outcomes.

Combine different learning strategies to create a diverse learning environment that will meet different learning styles of participants. This will increase learning effectiveness and help achieve the planned learning outcomes. In this section, it is not necessary to describe specific activities in detail.

- 1. Establishing clear learning objectives.** It is important that participants understand what results are expected from them. The objectives should be challenging but achievable, encouraging participants to reach high standards.
- 2. Flipped classroom strategy.** The flipped classroom is a learning principle in which the main acquisition of new material by students takes place at home, while classroom time is dedicated to practical work — completing tasks, exercises, laboratory and practical studies, individual consultations with the instructor, etc.
- 3. Competency-based approach.** This approach is focused on developing students' practical skills and abilities that they can apply in professional activities. It includes the integration of theoretical knowledge with practical tasks that reflect real working conditions.
- 4. Project-based learning.** Involving students in real projects allows them to apply acquired knowledge in practice, develop analytical and research skills, as well as teamwork abilities.
- 5. Individualization of learning.** Adapting the learning process to the needs and interests of each student, providing the opportunity to choose research or practical topics, promotes deeper engagement and motivation in learning.

Learning activities

Describe the main learning activities that will be included, such as lectures, reading, case studies, discussions, exercises, practical tasks, simulations, role-playing games, etc.

Also describe the roles of teachers and students during these activities.

Main learning activities

1. Interactive lectures with demonstration elements

Description: Short theoretical blocks (20–30 minutes) with visualization of key concepts. Each lecture is accompanied by real-time demonstration — the instructor shows on screen how to connect to a database, write basic queries, or work with a specific tool.

Role of the instructor:

- Explains theoretical concepts (e.g. relational vs non-relational databases, FAIR principles, NetCDF data structure).
- Demonstrates practical implementation in code (Python/R) or interface.
- Answers questions in real time.
- Emphasizes the connection between theory and real research tasks.

Role of the student:

- Listens, takes notes, asks clarifying questions.
- May replicate the instructor's actions on their own computer (optional).
- Records unclear points for discussion during practical sessions.

2. Practical exercises (coding and data work)

Description: The main form of work. Each practical session is built around a specific task: “Download 30 years of observational data”, “Merge satellite data with ground measurements”, “Calculate climate normals and visualize them”. Real datasets (Copernicus, NOAA, local weather stations) are used.

Role of the instructor:

- Defines tasks and explains the expected outcome.
- Acts as a mentor — helps with errors and suggests solutions.
- Shows alternative ways to solve the same problem.
- Conducts code review of selected or typical solutions.

Role of the student:

- Actively works on a computer and writes code.
- Experiments with parameters and performs debugging.
- Consults the instructor or peers when needed.
- Stores solutions in a personal repository (GitHub) for future use.

3. Case study analysis

Description: Analysis of real research or applied problems solved using climate data. For example: “How to aggregate meteorological station data to assess droughts in Ukraine?” or “Why reanalysis data may differ from ground observations and what to do about it?”

Role of the instructor:

- Presents the problem and shows raw data and final research results.
- Guides discussion through prompting questions.
- Demonstrates typical pitfalls in climate data work.
- Explains how proper data management influenced research success.

Role of the student:

- Analyzes materials and proposes solutions.
- Engages in discussion and justifies opinions.
- Draws conclusions on how to avoid errors in own work.
- Suggests additional data sources for deeper analysis.

4. Independent reading assignments

Description: Study of technical documentation (e.g. Copernicus CDS API documentation), scientific papers on data standards (NetCDF, Zarr), or textbook chapters on database architecture.

Role of the instructor:

- Provides a structured reading list with annotations (what to focus on).
- Prepares control questions to check understanding.
- Integrates reading content into lectures and discussions.

Role of the student:

- Independently studies materials outside class time.
- Takes notes and records unclear terms.
- Prepares 1–2 questions for discussion in the next session.

Assessment of Learning

Describe the plan for assessing participants before, during, and/or after the course, including tests, exercises, activities, and projects to be evaluated.

Indicate whether self-assessment or peer assessment will be used. Explain how the assessment is aligned with the learning outcomes.

The total number of points each student can earn is 100. The maximum score for each module is 20 points. Modules 1 and 2 consist of a test with 20 questions (20 points). Practical modules 3, 4, and 5 include 15 points for full completion of the task and 5 points for the defense of the practical work. Thus, the total maximum score a student can obtain for successful completion of the course is 100 points, and the minimum passing score is 60 points.

Learning storyboard (instructional storyboard)

Use this to create a visual scenario of your blended learning activity.





Learning Resources and Tools

List the available resources that you will use for different types of learning activities and recommend to students.

Describe the technologies that you will use to implement the learning solutions, including educational technologies and operational equipment (technical equipment, software, collaboration tools).

A classroom for 5–10 students equipped with a projector for blended learning.

1. Laptops or tablets with Internet access.
2. Moodle website (<https://re.climed.network/course/>), where course information is uploaded: objectives, calendar, participants, test information, forum, presentations, exercises, simulations, grades, and any additional materials.
3. Presentations and exercises prepared by trainers.
4. Access to registration and climate product portals; meteorological data portals, where students will have access to real-time data, NWP products from different centres, observation data, etc.
5. Work with cloud services (e.g. Google Earth Engine, Amazon Web Services, Azure Open Data), where data are already stored and computations are performed on powerful servers. This eliminates the problem of data downloading.

1. Data sources and access platforms

- Copernicus Climate Data Store (CDS) URL: <https://climate.copernicus.eu>

Description: The main European open-access portal for climate data (satellite observations, reanalysis such as ERA5, and climate forecasts). It includes an API for programmatic access, which is critical for practical course tasks. The C3S service also offers free online courses on understanding and using these data.

- ENES Data Space URL: <https://enesdataspace.vm.fedcloud.eu>

Description: A cloud environment for climate data analysis developed for the European climate community. It provides access to prepared CMIP (Coupled Model Intercomparison Project) datasets through a Jupyter interface, enabling computation without downloading large data volumes locally. It is useful for teaching cloud-based data processing.

- Climate Data Guide URL: <https://climatedataguide.ucar.edu/>

Description: A resource from the National Center for Atmospheric Research (NCAR), USA. It provides detailed reviews of various climate datasets (surface stations, satellites, reanalysis), including their strengths, weaknesses, limitations, and recommendations for use. It is essential for

the module on data quality assessment.

2. Learning materials and tutorials

- Project Pythia URL: <https://projectpythia.org/>

Description: A resource with open educational materials for working with geoscience data in Python. It includes ready-made “Cookbooks” with examples of data processing in NetCDF formats, visualization, and cloud data workflows.

- Climate DataLab URL: <http://www.climate-datalab.org/> and <https://github.com/climate-datalab>

Description: A public resource offering end-to-end training in working with climate model data. It explains fundamental concepts such as file formats (NetCDF), coordinate systems, model calendars, climate change scenarios, and model ensembles. The GitHub repositories include Python and R exercises.

- Climpact URL: <https://etrp.wmo.int/course/view.php?id=221>

Description: A course from the World Meteorological Organization (WMO) on using the open-source Climpact software to generate sector-specific climate indices. It aligns with WMO competencies for climate services, which is important for secondary audiences from hydrometeorological services.

Books:

1. Степаненко С.М. (2013). Динаміка та моделювання клімату. – Одеса, Видавництво: „Екологія”, 204 с. (електронний підручник) <http://eprints.onu.edu.ua/id/eprint/6173/>
2. Степаненко С.М., Польовий А.М., Лобода Н.С. (2015) Кліматичні зміни та їх вплив на сферу економіки України: монографія. ОДЕКУ, Одеса. 520 с. <http://eprints.onu.edu.ua/id/eprint/2269>
3. Wilks, D. S. (2019). Statistical Methods in the Atmospheric Sciences (4th ed.). Elsevier. ISBN: 9780128158234
4. Shen, S. S. P., & North, G. R. (2023). Statistics and Data Visualization in Climate Science with R and Python. Cambridge University Press. ISBN: 9781108842570
5. Pal, S. C., Chatterjee, U., Saha, A., & Ruidas, D. (Eds.). (2025). Climate Change: Conflict and Resilience in the Age of Anthropocene. Springer. ISBN: 9783031853593

WMO guiding documents

1. Climate Data Management Systems (CDMSs)

<https://community.wmo.int/en/climate-data-management-systems-cdmss#:~:text=Long%2Dterm%2C%20high%2Dquality,global%20climate%20variability%20and%20change>

2. Climate Data Management System Specifications WMO-No. 1131, 2025, 166 p.

<https://library.wmo.int/ru/records/item/51447-climate-data-management-system-specifications>

3. WMO-No. 100 (Guide to Climatological Practices), 2018, 153 с.

https://biotech.law.lsu.edu/blog/100-2018_en.pdf