Модуль 2 · Реанализ и многомерные климатические данные

NetCDF и xarray: многомерные климатические данные

Ph.D., старший преподаватель Мадина Турсумбаева

Кафедра метеорологии и гидрологии

Курс: Современные цифровые технологии в метеорологии

Цель лекции

Познакомить с форматом NetCDF и библиотекой xarray и научить открывать, выбирать и агрегировать многомерные данные реанализа ERA5, а также загружать их через CDS API.

Коллоквиум (рубежный контроль 1) — в конце лекции, по материалам лекций 1–7.

Результаты обучения

После лекции студент будет:

Знать

Структура NetCDF: размерности, координаты, переменные, атрибуты

Разницу между Dataset и DataArray

Уметь

Открыть ERA5-файл и посмотреть его структуру

Выбрать точку и период, перевести K в °C, агрегировать по суткам и месяцам

Загрузить ERA5 через CDS API

Понимать

Почему файл — не таблица

Почему ближайший узел сетки — не станция

Что такое NetCDF

Размерности

time, latitude, longitude — оси, по которым лежат данные.

Координаты

Значения осей: даты, широты, долготы.

Переменные

t2m — температура на 2 м, в кельвинах.

Атрибуты

Единицы (units), название (long_name), источник.

- NetCDF самоописываемый: файл содержит и данные, и их описание.
- Соглашения CF (Climate and Forecast) задают единообразные имена и единицы.

Открываем файл

```
import xarray as xr
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

ds = xr.open_dataset("ERA5.nc")    # путь к вашему файлу
ds                                 # краткое описание структуры
```

Dataset и DataArray

Dataset — набор переменных с общими координатами.

DataArray — одна переменная, например ds.t2m.

Путь к файлу укажите свой. В лабораторной работе файл лежит на диске F:, но в коде путь заменяется на свой.

Переменные, атрибуты и размерности

```
list(ds.data_vars)                  # список переменных
ds.t2m                              # метаданные переменной t2m
ds.t2m.attrs                        # атрибуты: units, long_name
ds.dims, ds.coords                  # размеры и координаты
```

Что проверить

Имя времени: valid_time в новых выгрузках, time — в старых.

Порядок широт: по возрастанию или убыванию.

Единицы

ERA5 хранит температуру в кельвинах. Проверяйте units перед расчётом.

Перевод единиц: K в °C

```
# создаём новый DataArray в градусах Цельсия
t2m_c = ds.t2m - 273.15
t2m_c.name = "t2m_c"
t2m_c.attrs["units"] = "degC"    # атрибуты после арифметики задаём явно
t2m_c
```

Ловушка

Арифметика может не сохранить атрибуты. Если единица в атрибутах важна, задавайте её явно, как в примере.

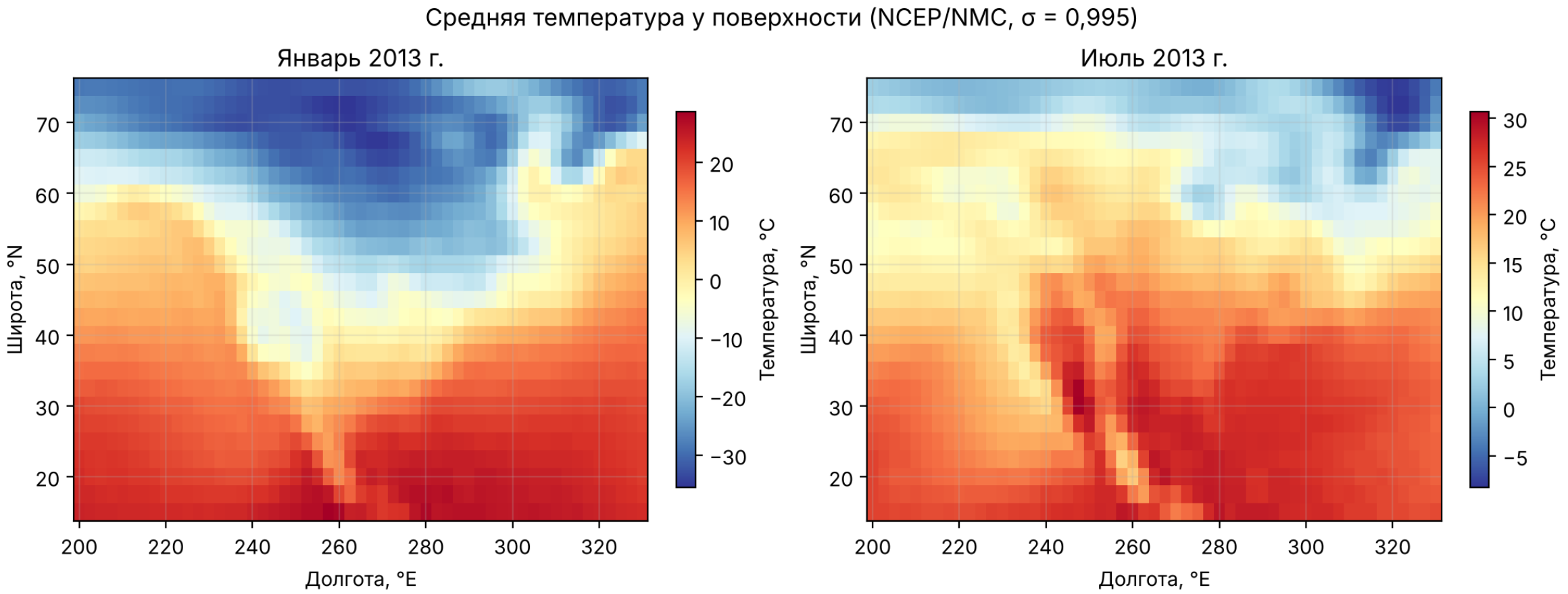
0 K = −273,15 °C: при переводе используйте точное значение, а не 273.

Точка на сетке: ближайший узел

```
lat0, lon0 = 51.17, 71.43          # Астана (пример из лабораторной работы)
lat = t2m_c["latitude"].to_numpy()
lon = t2m_c["longitude"].to_numpy()
# индексы ближайших узлов; долгота — с учётом формата 0–360°
ilat = np.abs(lat - lat0).argmin()
ilon = np.abs(((lon - lon0 + 180) % 360) - 180).argmin()
pt = t2m_c.isel(latitude=ilat, longitude=ilon)     # выбор ближайшей точки
```

- Короткий вариант: t2m_c.sel(latitude=43.25, longitude=76.95, method="nearest") — для Алматы.
- Шаг сетки 0,25°: ближайший узел находится не дальше примерно 17 км от точки.

Поле температуры: пример сетки (NCEP, не ERA5)



Пример из открытого файла NCEP/NMC (уровень σ = 0,995, 2013 г., сетка 2,5°). Это не ERA5: слайд показывает устройство поля, а не значения ERA5.

Выбор периода и простая статистика

```
# выбор за 2024 год
pt_2024 = pt.sel(valid_time=slice("2024-01-01", "2024-12-31"))

float(pt_2024.mean()), float(pt_2024.min()), float(pt_2024.max())
```

sel и slice

sel выбирает по метке (дате, координате); slice задаёт интервал. Границы включаются.

Результат

Среднее, минимум и максимум за год в °C — это одна точка сетки, а не станция.

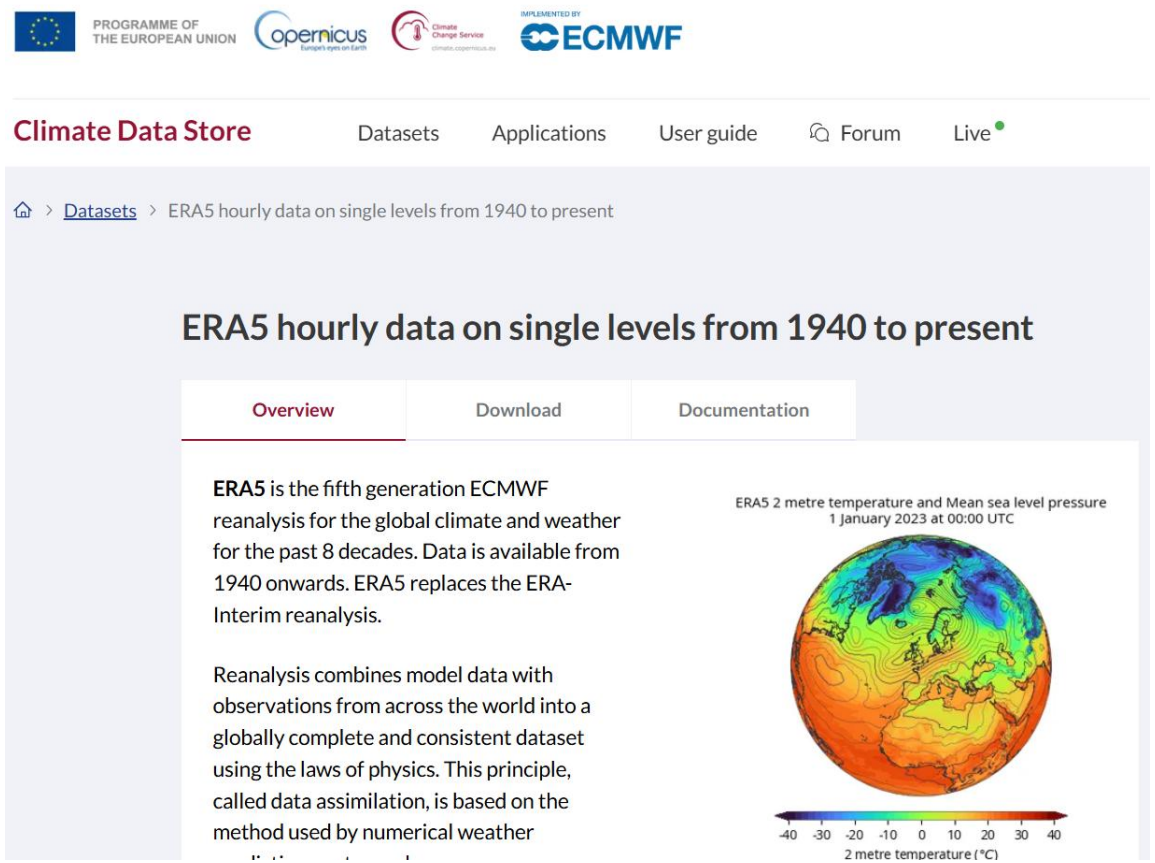
Ресемплинг и сохранение результатов

```
import matplotlib.pyplot as plt
s = pt.to_series()                    # pandas.Series с DatetimeIndex
pt_daily = s.resample("1D").mean()    # суточные средние
pt_monthly = s.resample("1MS").mean() # месячные средние
pt_monthly.plot(marker="o"); plt.ylabel("T, °C"); plt.grid(True); plt.show()

pt_daily.to_csv("t2m_point_daily.csv", header=["t2m_c"])
pt_monthly.to_csv("t2m_point_monthly.csv", header=["t2m_c"])
```

- "1D" — сутки; "1MS" — месяц с первого дня. В pandas 2.2+ конец месяца обозначают "1ME".
- Суточное среднее по 24 срокам — ERA5 почасовой, поэтому полнота здесь практически полная.

Откуда берутся данные: каталог CDS



Что показывает скриншот

Интерфейс CDS: датасет ERA5, переменная 2m temperature, карта за 1 января 2023 г.

Параметры выбираются в форме. Тот же запрос можно записать кодом для cdsapi.

Поле, которое вы видите на карте, в Python открывается через xarray.

Источник: скриншот интерфейса CDS из материалов курса (презентация по ERA5). Здесь не загружался из сети.

Загрузка ERA5 через CDS API

```
import cdsapi
c = cdsapi.Client()        # нужна учётная запись CDS и файл .cdsapirc с ключом
c.retrieve("reanalysis-era5-single-levels", {
    "product_type": "reanalysis",
    "variable": "2m_temperature",
    "year": "2024",
    "month": [f"{m:02d}" for m in range(1, 13)],
    "day": [f"{d:02d}" for d in range(1, 32)],
    "time": [f"{h:02d}:00" for h in range(24)],
    "area": [56, 46, 40, 88],        # север, запад, юг, восток — Казахстан
    "data_format": "netcdf",
}, "era5_t2m_2024.nc")
```

Параметры запроса меняются. Сверяйтесь с формой запроса на странице датасета в CDS перед загрузкой.

Типичные ошибки и их исправление

Ошибка	Что происходит	Как исправить
Время называется time, а код ищет valid_time	Ошибка при выборе периода	Проверить ds.dims
Градусы Цельсия приняты за кельвины	Ошибка на 273 градуса	Проверить units перед расчётом
Долгота 0–360° не учтена	Ближайший узел — на противоположной стороне	Учесть формат долготы
Ближайший узел принят за станцию	Высота и местность различаются	Сравнивать с поправкой (лекция 10)
Атрибут units потерян	Единица не видна в выводе	Задавать атрибуты явно

Заключение

1

Открыть

open_dataset: размерности, координаты, атрибуты

2

Выбрать

sel, isel: ближайший узел сетки и период

3

Преобразовать

K → °C, resample: сутки и месяцы

4

Сохранить

to_csv; новые данные — через CDS API

Главный вывод

NetCDF хранит многомерные данные вместе с координатами и атрибутами, а xarray позволяет открывать, выбирать и агрегировать их. Перед расчётом проверяют единицы и имя оси времени; ближайший узел сетки ERA5 нельзя считать станцией.

Вопросы для самопроверки

1. Чем Dataset отличается от DataArray и что хранится в атрибутах переменной?

2. Почему перед расчётом нужно проверить units и имя оси времени (time или valid_time)?

3. Почему ближайший узел сетки ERA5 нельзя считать станцией?

4. Какие параметры задают в запросе CDS API, чтобы загрузить ERA5 для своего района?

Коллоквиум: рубежный контроль 1

Вопросы для подготовки (лекции 1–7)

1. Чем отличаются float, int, str и bool? Что даст 25.4 + "1"?
2. Что делают read_csv, isna и dropna и в каком порядке их применять?
3. Как посчитать норму за 1991–2020 гг. и аномалию по месяцам?
4. Почему fillna средним опасен для ряда с трендом?
5. Почему станционный ряд читают в одном часовом поясе?
6. Почему r = 0 не означает отсутствия связи?
7. Чем профиль радиозонда отличается от суточного ряда станции?
8. Что такое потенциальная температура и в каких единицах её считать?
9. Как в xarray выбрать ближайший узел сетки и перевести K в °C?
10. Какие атрибуты хранятся в NetCDF и зачем нужны размерности?

Семинар 8 и следующая лекция

Семинар 8 · NetCDF-файлы ERA5 в Python

Открыть файл, выбрать точку, перевести K в °C, сохранить суточные и месячные ряды.

Сравнить свою точку с примером из лабораторной работы.

Лекция 9

Анализ ERA5: тренды, сезонность, аномалии и ограничения реанализа.

Литература

- Hersbach, H., et al. (2020). The ERA5 global reanalysis. QJRMS, 146, 1999–2049. https://doi.org/10.1002/qj.3803
- Hoyer, S., & Hamman, J. (2017). xarray: N-D labeled arrays and datasets in Python. Journal of Open Research Software, 5(1), 10.
- Rew, R., & Davis, G. (1990). NetCDF: an interface for scientific data access. IEEE Computer Graphics and Applications, 10(4), 76–82.
- CF Conventions. Climate and Forecast metadata conventions. https://cfconventions.org
- Copernicus Climate Data Store (CDS). ERA5 hourly data on single levels. https://cds.climate.copernicus.eu
- Raoult, B., Hersbach, H., et al. ERA5: The ECMWF Reanalysis Data. ECMWF, 2018.