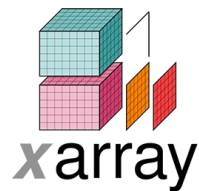


Learning Objectives

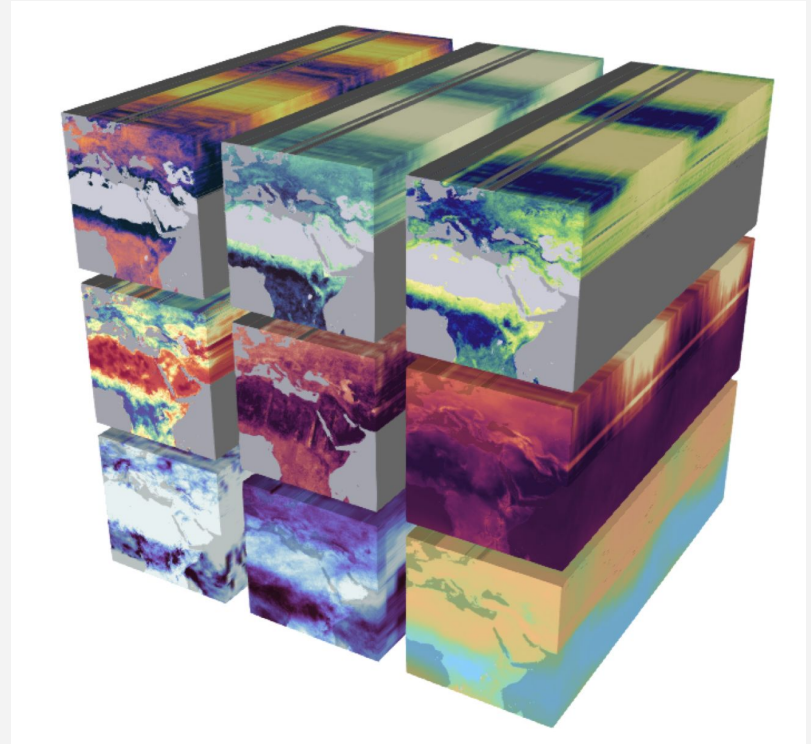
- Students will learn the main characteristics and components of the NetCDF data format.
- Students will become familiar with the core data structures in the xarray Python package and how these mirror the NetCDF structure.



Example

Earth System Data Cube from [Mahecha et al. 2020](#)

- nine environmental variables at high resolution across space and time
- **large:** high-resolution
- **complex:** multiple variables
- **multi-dimensional:** each variable is measured along three dimensions
- metadata must be included to make it functional



Mahecha et al. 2020 . Visualization of the implemented Earth system data cube. The figure shows nice climate variables, measured across space and time.

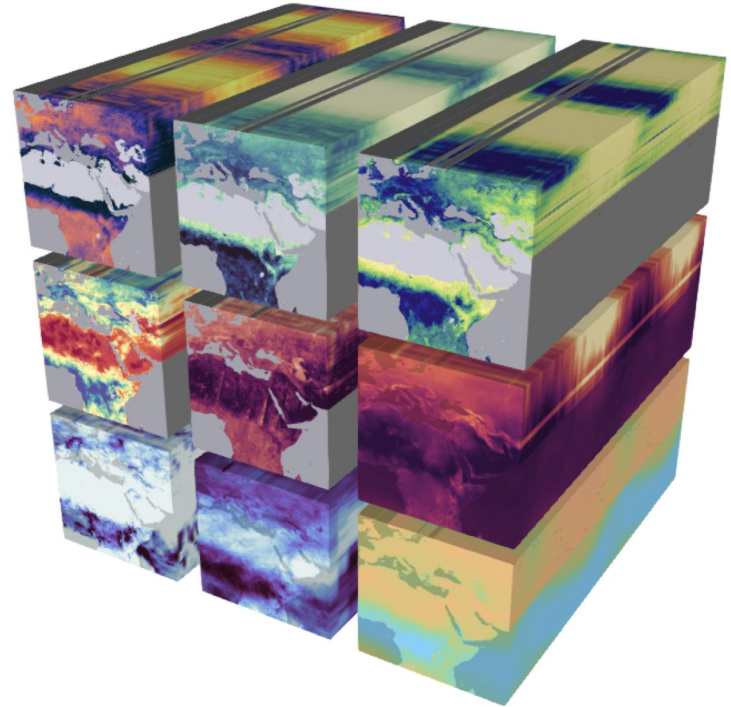
Example

Earth System Data Cube from [Mahecha et al. 2020](#)

- nine environmental variables at high resolution across space and time
- **large:** high-resolution
- **complex:** multiple variables
- **multi-dimensional:** each variable is measured along three dimensions
- metadata must be included to make it functional

Central idea

Exploring large, complex is becoming key to asking BIG questions about humans and nature. We need the right tools to access, process, share, and archive this information!



Mahecha et al. 2020 . Visualization of the implemented Earth system data cube. The figure shows nice climate variables, measured across space and time.

What is NetCDF?

- network Common Data Form
- **file format:** extension .nc

What is NetCDF?

- network Common Data Form
- **file format**: extension .nc
- developed by the University Corporation for Atmospheric Research
- designed for the creation, access, and sharing of **array-oriented scientific data**

What is NetCDF?

- network Common Data Form
- **file format:** extension .nc
- developed by the University Corporation for Atmospheric Research
- designed for the creation, access, and sharing of **array-oriented scientific data**
- **self-describing:**
 - Information describing the data contents of the file is embedded within the data file itself
- **supported on almost all platforms** (machine independent / interoperable)
- **archivable:**
 - Access to all earlier forms of NetCDF data will be supported by current and future versions of the software.

What is [NetCDF](#)?

- network Common Data Form
- **file format:** extension .nc
- developed by the University Corporation for Atmospheric Research
- designed for the creation, access, and sharing of **array-oriented scientific data**
- **self-describing:**
 - Information describing the data contents of the file is embedded within the data file itself
- **supported on almost all platforms** (machine independent / interoperable)
- **archivable:**
 - Access to all earlier forms of NetCDF data will be supported by current and future versions of the software.



[View of Earth and satellite - NASA](#)

From the [NASA Earth Data Program](#):

*Volumes of data distributed in this format run into the **tens of terabytes** coming from both past and existing NASA/NOAA missions.*

*The NetCDF Classic is an approved standard **recommended** for use in NASA Earth Science Data Systems.*

What is xarray?

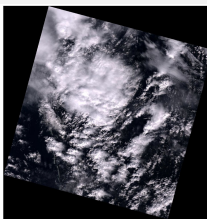
- an **open source Python package**
- **based on the netCDF data model**
- augments NumPy arrays by **adding labeled dimensions, coordinates, and attributes**

What is xarray?

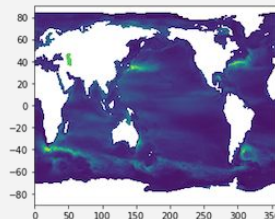
- an **open source Python package**
- **based on the netCDF data model**
- augments NumPy arrays by **adding labeled dimensions, coordinates, and attributes**
- **interoperable** with tools for parallel processing (Dask), visualization (Matplotlib), and analysis (Pandas)
- central part of the Pangeo **open software ecosystem for Big Data geoscience**



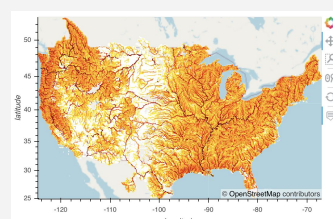
xarray



satellite image
analysis



oceanography



river discharge model

How does NetCDF organize data?

NetCDF file

Variables

Dimensions

Attributes

How does NetCDF organize data?

Variables

- N-dimensional arrays of numbers

What is being measured?

Varying/measured/dependent quantities.

ex: we measure snowfall

Dimensions

Attributes

NetCDF file

Variable: snowfall



Photo: [LA Times 2023](#)

How does NetCDF organize data?

Variables

- N-dimensional arrays of numbers

What is being measured?

Varying/measured/dependent quantities.

ex: we measure snowfall

Dimensions

- describe the axes of the arrays

With respect to what are we measuring the variable?

constant/fixed/independent quantities at which we measure the variables.

ex: we measure snowfall with respect to time or we can measure it with respect to location

Attributes

NetCDF file

Variable: snowfall



Photo: [LA Times 2023](#)

Dimensions: time & location



How does NetCDF organize data?

Variables

- N-dimensional arrays of numbers

What is being measured?

Varying/measured/dependent quantities.

ex: we measure snowfall

Dimensions

- describe the axes of the arrays

With respect to what are we measuring the variable?

constant/fixed/independent quantities at which we measure the variables.

ex: we measure snowfall with respect to time or we can measure it with respect to location

Attributes

- annotations about a variable, dimension or the whole file

How are measuring the variables and dimensions?

ex: units for snowfall, time, and location. author & agency

info.

NetCDF file

Variable: snowfall



Photo: [LA Times 2023](#)

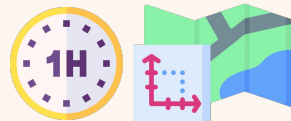
Dimensions: time & location



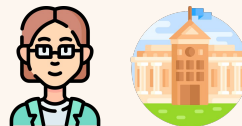
Variable attributes



Dimension attributes



File attributes



Exercise

Part 1 - Imagine the following (very simplified) scenario: we have a network of 25 weather stations.

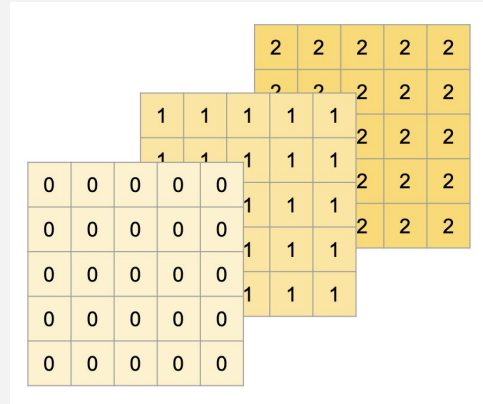
They are located in a square grid: starting at 30°0'N 60°0'E, there is a station every 10° North and every 10° East.

Each station measures the air temperature at a set time for three days, starting on September 1st, 2022.

On the first day, all stations record a temperature of 0°C. On the second day, all temperatures are 1°C, and on the third day, all temperatures are 2°C.

What are the *variables*, *dimensions*, and *attributes* for this data? Make a diagram.

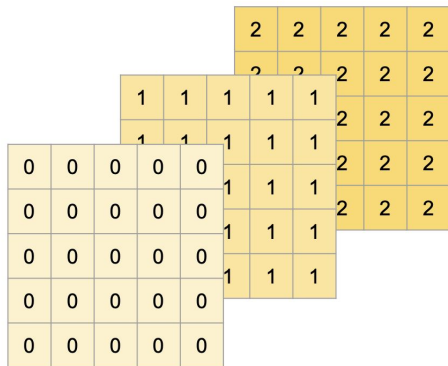
Exercise



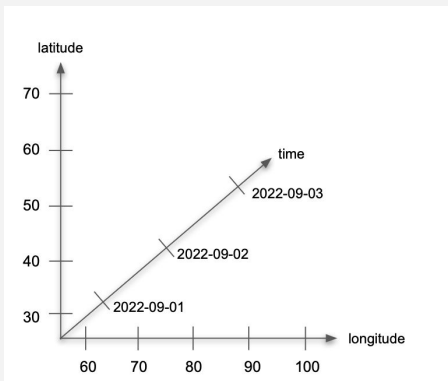
Temperature variable

Variables: Temperature. Can be represented as a 5x5x3 array, with constant values for each day.

Exercise



Temperature variable



Dataset dimensions

Variables: Temperature. Can be represented as a 5x5x3 array, with constant values for each day.

Dimensions:

Time, latitude, and longitude.

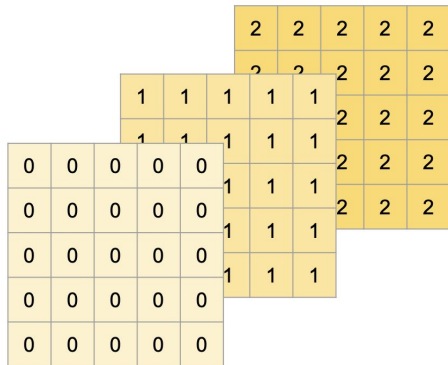
Time indicates when the measurement happened, we can encode it as the dates 2022-09-01, 2022-09-02, and 2022-09-03.

The pairs of latitude and longitude values indicate the positions of the weather stations.

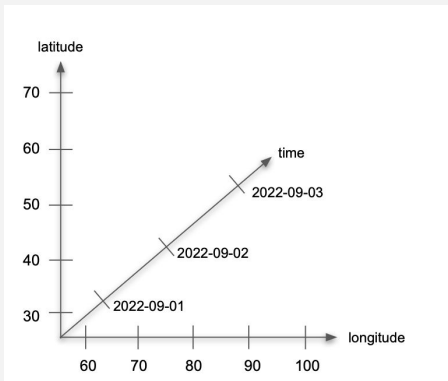
Latitude has values 30, 40, 50, 60, and 70, measured in degrees North.

Longitude has values 60, 70, 80, 90, and 100, measured in degrees East.

Exercise



Temperature variable



Dataset dimensions

Variables: Temperature. Can be represented as a 5x5x3 array, with constant values for each day.

Dimensions:

Time, latitude, and longitude.

Time indicates when the measurement happened, we can encode it as the dates 2022-09-01, 2022-09-02, and 2022-09-03.

The pairs of latitude and longitude values indicate the positions of the weather stations.

Latitude has values 30, 40, 50, 60, and 70, measured in degrees North.

Longitude has values 60, 70, 80, 90, and 100, measured in degrees East.

Attributes:

Variable attributes:

Temperature attributes:

- standard_name: air_temperature
- units: degree_C

Dimension attributes:

Time attributes:

- description: date of measurement

Latitude attributes:

- standard_name: grid_latitude
- units: degrees_N

Longitude attributes:

- standard_name: grid_longitude
- units: degree_E

Dataset attributes:

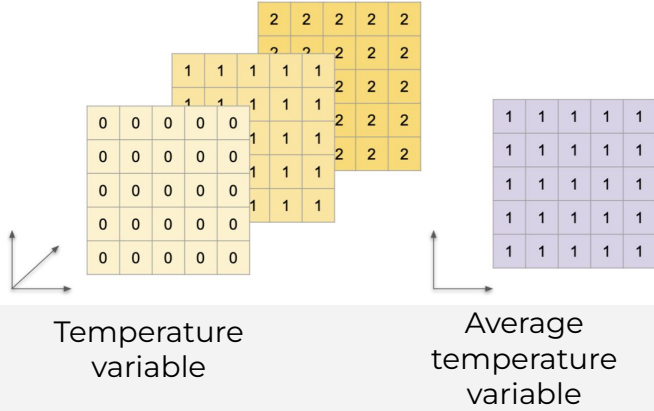
- title: Temperature Measurements at Weather Stations
- summary: an example of NetCDF data format

Exercise

Part 2 - Now imagine we calculate the average temperature over time at each weather station, and we wish to incorporate this data into the same dataset.

How will adding the average temperature data change the dataset variables, attributes, and dimensions?

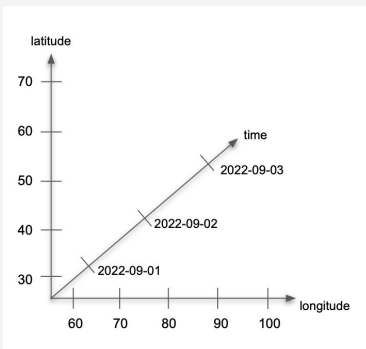
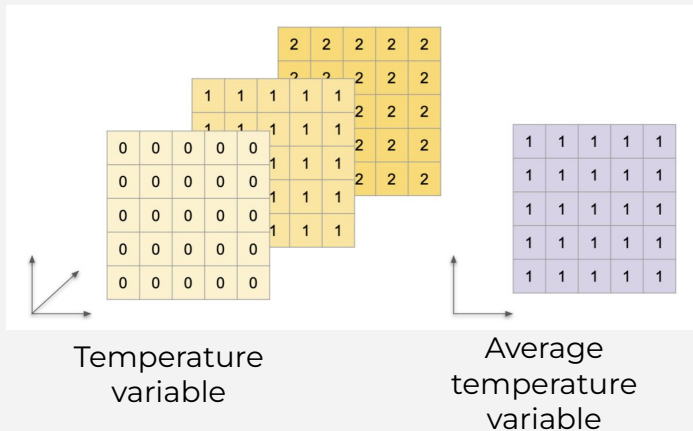
Exercise



Variables: temperature and average temperature.

The temperature data stays the same. We can represent the average temperature as a single 5x5 array with value 1 at each cell.

Exercise



Dataset dimensions

Variables: temperature and average temperature.

The temperature data stays the same. We can represent the average temperature as a single 5x5 array with value 1 at each cell.

Dimensions:

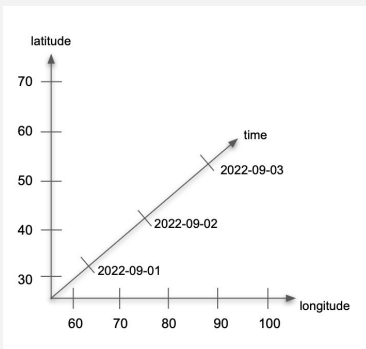
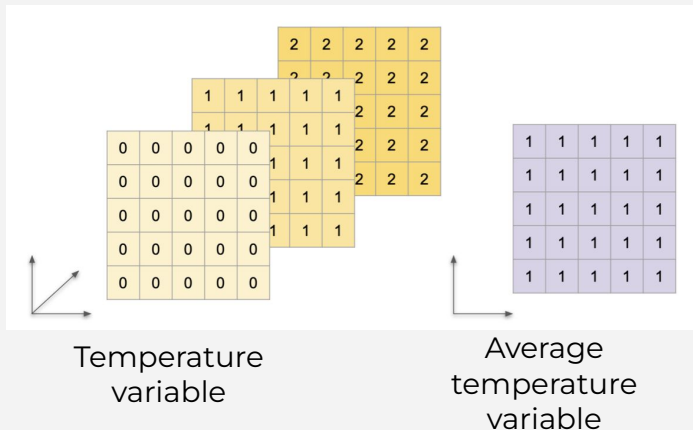
The **complete dataset** still has three dimensions: time, latitude, and longitude.

The **temperature variable uses all three dimensions, and the average temperature variable only uses two** (latitude and longitude).

This is ok! The dataset's dimensions are the union of the dimensions of all the variables in the dataset.

Variables in the same dataset may have all, some, or no dimensions in common.

Exercise



Dataset dimensions

Variables: temperature and average temperature.

The temperature data stays the same. We can represent the average temperature as a single 5x5 array with value 1 at each cell.

Dimensions:

The **complete dataset** still has three dimensions: time, latitude, and longitude.

The **temperature variable uses all three dimensions, and the average temperature variable only uses two** (latitude and longitude).

This is ok! The dataset's dimensions are the union of the dimensions of all the variables in the dataset.

Variables in the same dataset may have all, some, or no dimensions in common.

Attributes:

Keep all the previous attributes and add the attributes for our new average temperature variable:

Average temperature attributes:

- standard_name: average_air_temperature
- description: average temperature over three days

What are coordinates?

Variables

- N-dimensional arrays of numbers

What is being measured?

Varying/measured/dependent quantities.

ex: we measure temperature

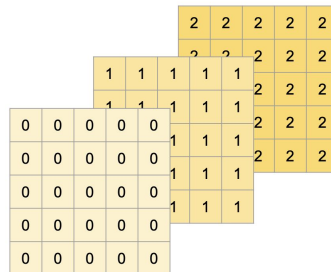
Dimensions

- describe the axes of the arrays

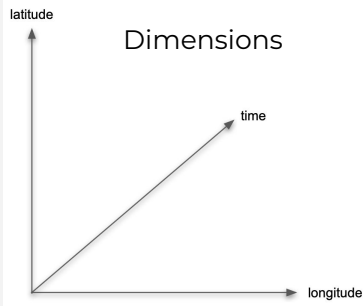
*With respect to what are we measuring the variable?
constant/fixed/independent quantities at which we
measure the variables.*

ex: we temperature with respect to time, lat and long

Variable



0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0



What are coordinates?

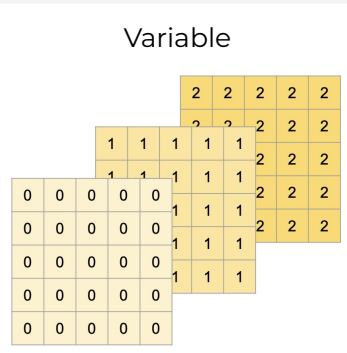
Variables

- N-dimensional arrays of numbers

What is being measured?

Varying/measured/dependent quantities.

ex: we measure temperature

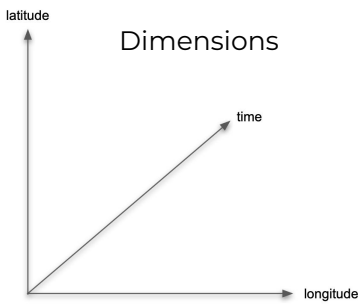


Dimensions

- describe the axes of the arrays

With respect to what are we measuring the variable?
constant/fixed/independent quantities at which we
measure the variables.

ex: we measure temperature with respect to time, lat and long



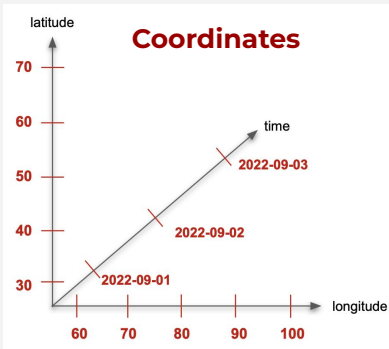
Coordinates

- each dimension has a set of coordinates
- coordinates indicate the dimension's values

At which points in this dimension are we measuring the variable?

tick labels along a dimension axis

ex: we measure the temperature variable at 60, 70, ..., 100 degrees along the longitude dimension



Let's code!

bit.ly/cgg_demo