Climate Data and Visualization

NI4OS-service

Marco Miani • Computational Support Specialist m.miani@cyi.ac.cy

30 October 2020

CARE-C (Climate and Atmosphere Research Centre)

The Content of this webinar talk

1. Climate Data - what are they and what do they describe?

2. High Performance Computing

3. Visualizing Climate Data

Computational Support Specialist – Climate and Atmosphere Research Center (CARE-C)

Research Topics

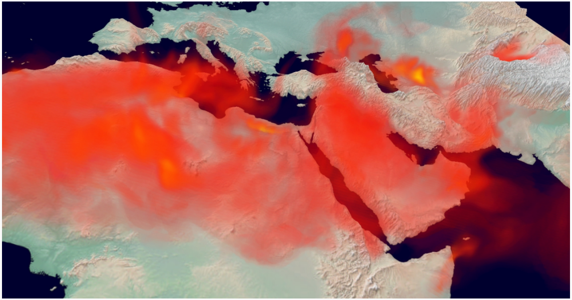
- Atmospheric and climate/earth modelling
- Emission inventories, modelling and analysis
- Dynamical downscaling of climate change and weather extremes
- Modelling and analysis of the urban environment
- Air quality and dust modelling and forecasting

Responsibilities

- Maintenance and management of various geophysical data-sets
- Statistical analysis for the study of temporal and spatial variations of atmospheric and climate data
- Data science and visualisation
- Presentation and interpretation of scientific results

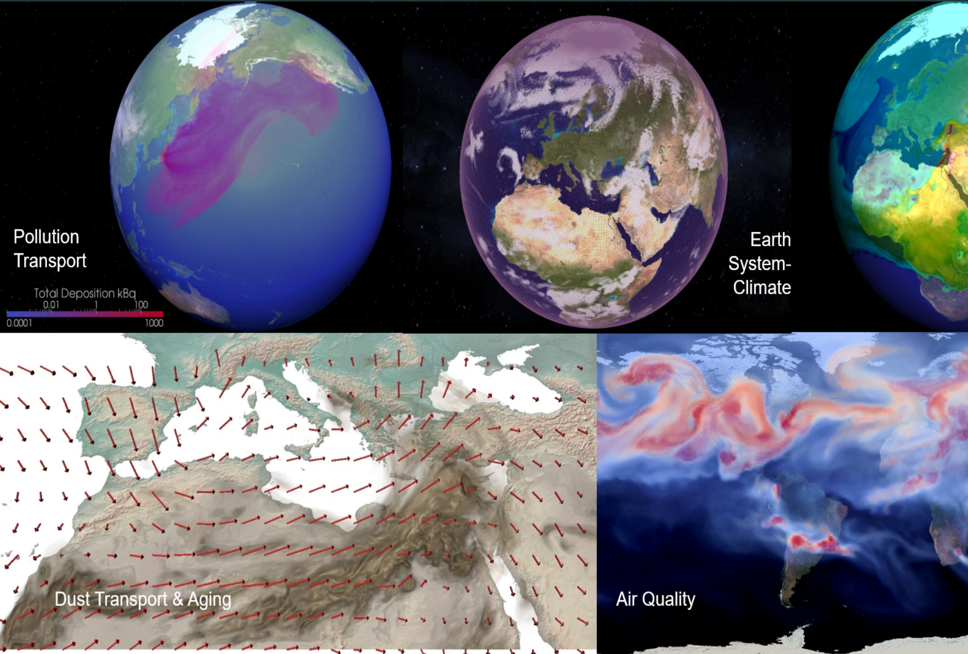
Regional Atmospheric Composition Modelling

- WRF-Chem as Regional Model
- Regional dust-air pollution links
- Aerosol effects on rainfall
- Energy forecasting applications
- Climate-scale effects



- Air quality modelling in the summer over the Eastern Mediterranean using WRF/Chem: Chemistry and aerosol mechanisms intercomparison. ([1])
- Air quality modelling over the Eastern Mediterranean: Seasonal sensitivity to anthropogenic emissions. ([2])
- Evaluation of EU air quality standards through modeling and the FAIRMODE benchmarking methodology ([3])

Environmental Predictions Group



Climate Data - what are they and what do they describe?

What turns data into actual climate data?

The **short-term** state of the atmosphere is named weather (e.g., temperature, precipitation, humidity, cloudiness, wind, et al.), and it can vary from minute to minute and location to location.

Climate is a description of the **long-term** pattern of weather conditions at a location. The expression “long-term” usually means 30 years or more, believed to be a good length of time to establish the usual range of conditions at a given location throughout the year.

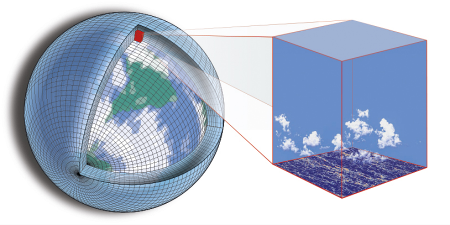
The difference between weather variability and long-term climate trends is like the difference between the path of a dog and the path of the person walking the dog

Adapted from : `https://www.climate.gov/maps-data/primer/comparing-climate-and-weather`
Animation : `https://youtu.be/e0vj-0imOLw`

Climate data: numerical model vs. real-world observations

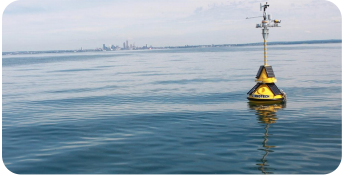
Numerical model

- 180×360 grid points
- 30 vertical levels
- **daily** output, over **30 yrs.**
- multiple quantities (T, p, u, ...)

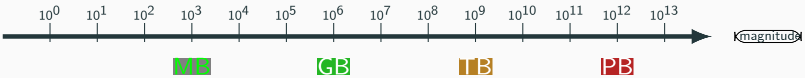


In-situ observations

- hourly data
- increasing coverage
- smart devices are being progressively used

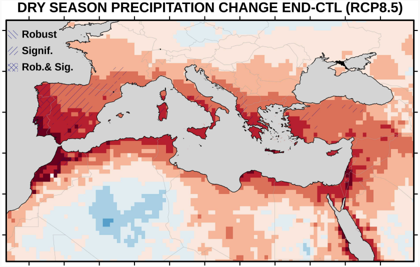


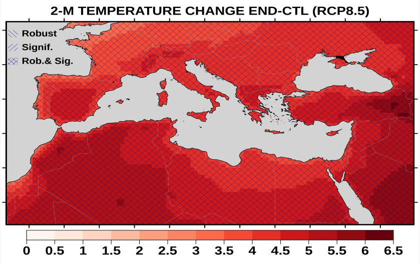
Extremely large data-sets (≈ peta-bytes): special binary **formats** needed.



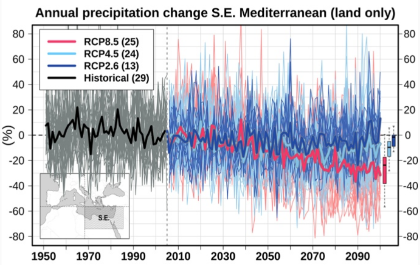
Model generated climate projections

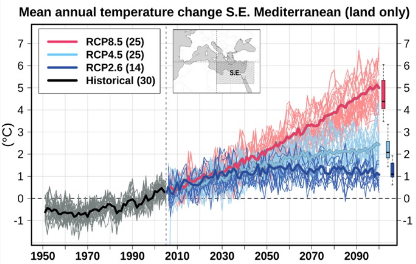
Maps





Time-series





"A multi-model, multi-scenario, and multi-domain analysis of regional climate projections for the Mediterranean", Ziitis et. al. See [4]

Unmanned System Research Laboratory @ The Cyprus Institute

Mission → to provide high-quality observations of atmospheric pollutants and other parameters relevant to air quality and climate change.

Activity → regular unmanned flights (Agia Marina Xyliatou) to monitor long-range transported pollution and dust aerosols from the largest desert regions

Increasing amount of sensors on board • long-term monitoring observations necessary for assessing climate change in our region

Climate Data: Conventional Formats

Selection of common formats

→ `HDF4`, `HDF5`
Hierarchical **D**ata **F**ormat (NASA)

→ `GRIB1`, `GRIB2` – **Gri**dded **B**inary
(World Meteorological Organization)

→ `netCDF3`, `netCDF4`
Network **C**ommon **D**ata **F**orm (NCAR)

binary file:

- non human readable
- sequence of bytes
- memory efficient
- quickly accessible

Focus of this talk: model-generated, **netCDF** data at global scale

Source: `https://climatedataguide.ucar.edu/climate-data-tools-and-analysis/common-climate-data-formats-overview`

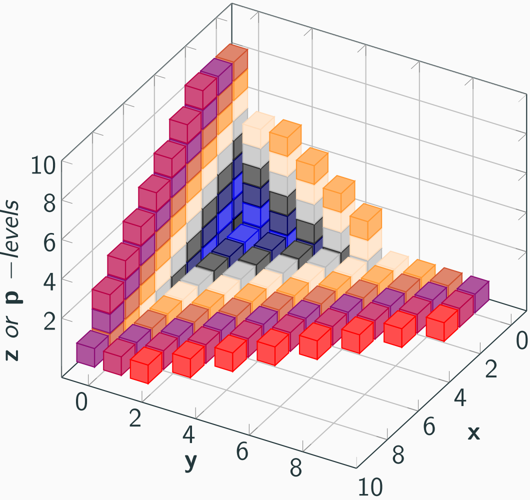
`netCDF` format is array-oriented:

- **Self-Describing**: a `netCDF` file includes information about the data it contains, i.e.: attributes or metadata. (See: [5])
- **Portable**: a `netCDF` file can be accessed by computers with different ways of storing integers, characters, and floating-point numbers.
- **Scalable**: small subsets of large datasets in various formats may be accessed efficiently through `netCDF` interfaces, *even from remote servers*.
- **Appendable**: Data may be appended to a properly structured `netCDF` file without copying the dataset or redefining its structure.

Adapted from: [6] (`www.unidata.ucar.edu/software/netcdf/`)
More sources: [5], [7], [8].

The `netCDF` structure: data and metadata

Structure for **one single timestep**, t=t_0



data: numerical values of physical quantities (`T`, `p`, `v`,...)

metadata:

alphanumeric description;

"data describing data"

(sub)fields such as:

- size
- dimensions
- type
- unit & time ref
- and much more!

collection of self-describing, portable objects (See: [7])

FAIR (climate) data

Findable

Machine-readable metadata are essential for automatic discovery of data-sets and services

Accessible

How can data be accessed, possibly including authentication and authorisation

Interoperable

the data need to interoperate with applications or workflows for analysis, storage, and processing

Reusable

Metadata and data should be well-described so that they can be replicated and/or combined in different settings

Example of Fair Data

National Oceanic and Atmospheric Administration
Atlantic Oceanographic and Meteorological Laboratory
Physical Oceanography Division (PhOD)

Current Missions (29 through 37): 2020 Hurricane Season

Glider ID	Status	Region	Deployment	# of profiles	Date of Recovery	Data Access
SG609	Active	North Atlantic (off Puerto Rico)	07/19/2019	1112	N/A	FTP link
SG610	Active	Caribbean Sea (off Puerto Rico)	07/14/2019	992	N/A	FTP link
SG630	Active	North Atlantic (off Puerto Rico)	07/19/2019	1130	N/A	FTP link
SG635	Active	North Atlantic (off Puerto Rico)	07/19/2019	1150	N/A	FTP link
SG649	Active	Caribbean Sea (off Puerto Rico)	07/15/2019	1106	N/A	FTP link
SG663	Active	Caribbean Sea (off Dominican Republic)	07/15/2019	1098	N/A	FTP link
SG664	Active	Caribbean Sea (off Puerto Rico)	07/14/2019	1088	N/A	FTP link
SG665	Active	North Atlantic (off Puerto Rico)	07/18/2019	1038	N/A	FTP link
SG669	Active	Caribbean Sea (off US Virgin Islands)	07/15/2019	1288	N/A	FTP link

```
(NCL_stable1) Marco@linux:~/Downloads$ ncdump -h SG609_0001_dn_AOML.nc | grep ":history"
                :history = "Created on 2020-07-28T15:54:54Z using SGnc2AOMLnc_py" ;
(NCL_stable1) Marco@linux:~/Downloads$
(NCL_stable1) Marco@linux:~/Downloads$
(NCL_stable1) Marco@linux:~/Downloads$
(NCL_stable1) Marco@linux:~/Downloads$
(NCL_stable1) Marco@linux:~/Downloads$ ncdump -h SG609_0001_dn_AOML.nc | grep variables
variables:
                ctd_time:ancillary_variables = "ctd_time_qc" ;
                longitude:ancillary_variables = "lon_qc" ;
                latitude:ancillary_variables = "lat_qc" ;
                ctd_depth:ancillary_variables = "ctd_depth_qc" ;
                temperature:ancillary_variables = "temperature_qc" ;
                conductivity:ancillary_variables = "conductivity_qc" ;
                salinity:ancillary_variables = "salinity_qc" ;
                density:ancillary_variables = "density_qc" ;
                aanderaa4831_dissolved_oxygen:ancillary_variables = "aanderaa4831_dissolved_oxy
(NCL_stable1) Marco@linux:~/Downloads$
```

Data catalogue with **FTP link**!

Data query of content: variable names.

(netcdf: **1 single command line**!)

Source: *Sustained and Targeted Ocean Observations for Improving Tropical Cyclone Intensity and Hurricane Seasonal Forecasts*
https://www.aoml.noaa.gov/phod/goos/gliders/data.php

High Performance Computing

Cy-Tera: the Cypriot supercomputer @ The Cyprus Institute

What is HPC and how does it work?

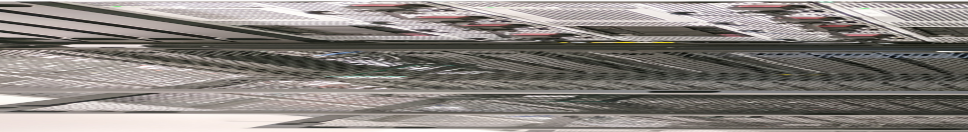
→ *Aggregating* computing power to deliver much higher performance than desktop computer

→ solve large and complex problems in science, engineering, or business

→ Located at the Cyprus Institute, Cy-Tera is the first supercomputer in Cyprus, and the biggest **open access** supercomputer in the Middle East.

Examples?

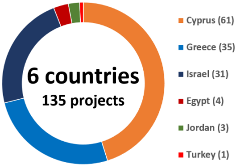
- Computational Chemistry
- Computational Fluid Dynamics
- Computational Particle and Nuclear Physics
- Climate Modelling



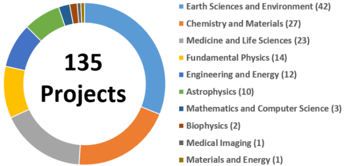
Cy-Tera: the Cypriot supercomputer @ The Cyprus Institute

features	Cy-Tera
Peak Performance	600 TFlop/s
Nr of Nodes	17 forty-core
Processors/node	2 twenty-core
Memory/node	
(compute node)	1992
total memory	5TB
Disk storage	3.2 PB Storage

Source: `http://web.cytera.cyi.ac.cy/resources/`



Distribution of big projects on Cy-Tera per country



Distribution of big projects on Cy-Tera per scientific field

"Numerical Study of Propeller Diameter Effects for a Self-Propelled Conventional Submarine" – Full reference: [9]

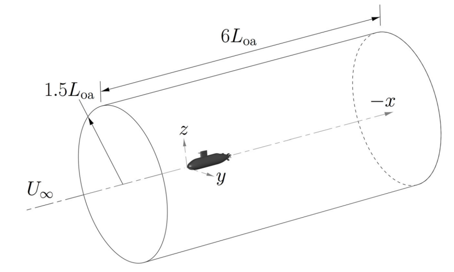
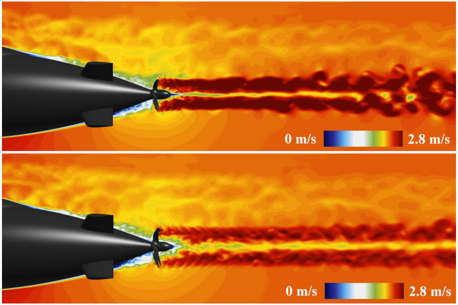
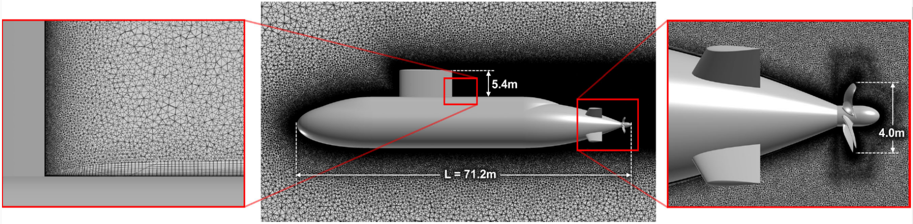
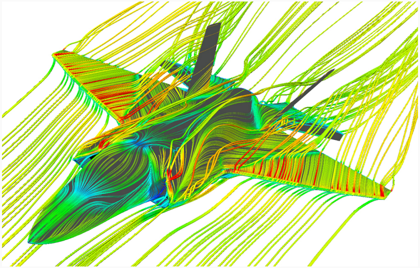


Figure 2: Schematic of the computational domain and axis system.

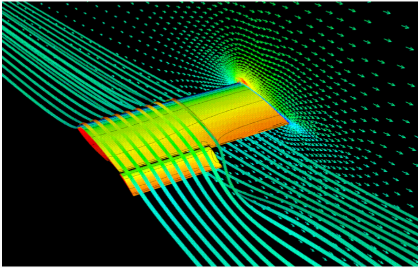




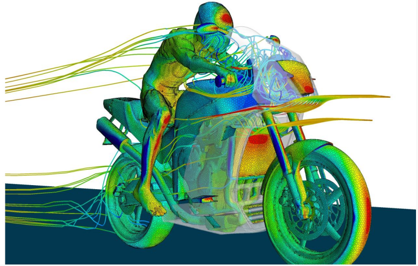
Computational Fluid Dynamics - Application in Engineering



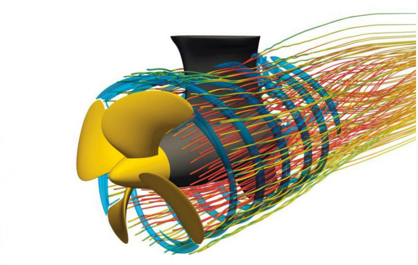
Supersonic-regime airflow (F-35 jet fighter)



Effect of flap on wing's trailing edge @ low speed

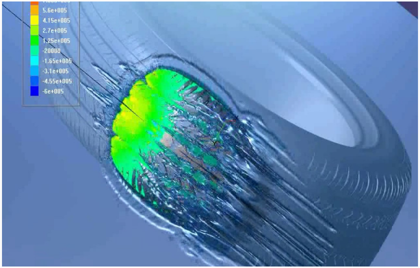


Pressure field around motorbike and pilot

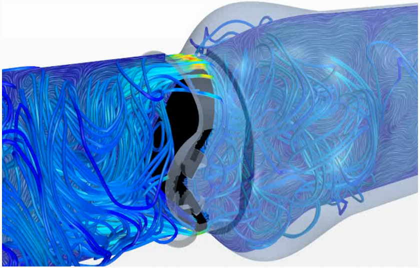


Efficiency of propeller blades

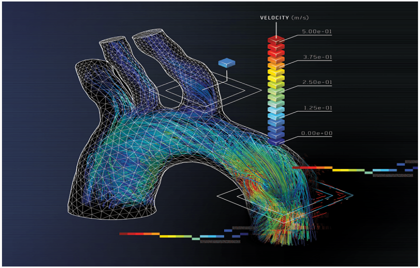
Computational Fluid Dynamics - Innovative Applications



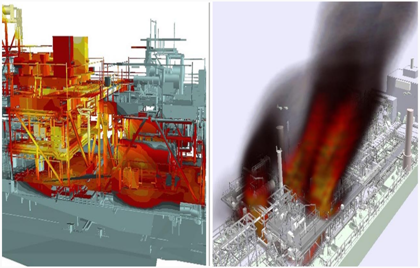
Aquaplaning mitigation



Biomedical Eng.: Vascular flow trough artificial valve



Hemodynamics



Combustion modelling (fire propagation)

Bottlenecks in Climate Modelling (HPC)

→ **Increasing Complexity and Parametrization**
"Simple" model coupling (ocean-atmosphere), up to more complex scenarios: bio-geo-chemistry, criosphere, biosphere, social issues.

→ **Spatial Resolution**
Models can't compute continually in space and time. Instead, the globe is "discretized" forming a "computational grid". Smaller grid cells mean accurate results, but implies higher computing efforts.

→ **Ensembles**
Models are sensible to slight initial disturbances. Various initial configurations are repeated many times to filter out randomness from statistically proven trends.

→ **Temporal Resolution and Duration**
Many scientific investigations require simulations of the earth system over a period of several centuries to preserve significance.

Adapted from: `https://www.dkrz.de/about-en/aufgaben/hpc`

Visualizing Climate Data

High Resolution Climate Modelling

Description:
DYAMOND++ - A High Resolution Climate Model Setup. German Climate Computing Center (DKRZ). Duration approx. 9 minutes (with audio).

Source: `https://youtu.be/5Y_oDaFRLaI?t=17`
Domain: *interactive video – climate data*

Cloud-resolving Simulations

Description:
$HD(CP)^2$ – Cloud-resolving simulation over Germany through ICON high resolution. Vertical stratification can be fully appreciated in this video. German Climate Computing Center (DKRZ).

Source: `https://youtu.be/HhwHuZR2uKo`
Domain: *climate data*

F5-class Tornado

Description:
A simulation of an F5 tornado produced by Cloud Model 1 (`CM1`), including the spectacular "seeding" technique. The Fujita scale (F-Scale), rates tornado intensity, based primarily on the damage inflicted on human-built structures and vegetation. Produced in Vapor (NCAR).

Source: `https://youtu.be/RSfHpzlZAZg`
Domain: *climate data*

Underwater Ocean Glider

Description:
Animating in-situ measured data measured by ocean glider, an autonomous underwater vehicle (AUV) during its underwater journey to the ocean abyss. Sam Jones (SAMS).

Source: `https://youtu.be/w6vONiftikc`
Domain: *applied oceanography*

Large eddy simulation of a Wind Farm

Description:
Interaction between large wind farms with multiple wind turbines, and the flow in the atmospheric boundary layer. Physics of Fluids Group, University of Twente.

Source: `https://youtu.be/qEtcCjln-0Q`
Domain: *Engineering* - **C***omputational* **F***luid* **D***ynamics*

Questions?

References i

[1] G. K. Georgiou, T. Christoudias, Y. Proestos, J. Kushta, P. Hadjinicolaou, and J. Lelieveld, "Air quality modelling in the summer over the eastern mediterranean using WRF-chem: chemistry and aerosol mechanism intercomparison," *Atmospheric Chemistry and Physics*, vol. 18, no. 3, pp. 1555–1571, 2018. [Online]. Available: https://acp.copernicus.org/articles/18/1555/2018/

[2] G. K. Georgiou, J. Kushta, T. Christoudias, Y. Proestos, and J. Lelieveld, "Air quality modelling over the eastern mediterranean: Seasonal sensitivity to anthropogenic emissions," *Atmospheric Environment*, vol. 222, p. 117119, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1352231019307587

References ii

[3] J. Kushta, G. K. Georgiou, Y. Proestos, T. Christoudias, P. Thunis, C. Savvides, C. Papadopoulos, and J. Lelieveld, "Evaluation of eu air quality standards through modeling and the fairmode benchmarking methodology," *Air quality, atmosphere, and health*, vol. 12, no. 1, p. 73—86, 2019. [Online]. Available: https://europepmc.org/articles/PMC6327007

[4] G. Zittis, P. Hadjinicolaou, M. Klangidou, Y. Proestos, and J. Lelieveld, "A multi-model, multi-scenario, and multi-domain analysis of regional climate projections for the Mediterranean," *Regional Environmental Change*, vol. 19, no. 8, pp. 2621–2635, nov 2019. [Online]. Available: http://link.springer.com/10.1007/s10113-019-01565-w

[5] NOAA, "World Ocean Database ragged array netCDF format." [Online]. Available: https://www.nodc.noaa.gov/OC5/WOD/netcdf_descr.html

[6] UniData.ucar.edu, "Network common data form (netcdf)." [Online]. Available: https://www.unidata.ucar.edu/software/netcdf/

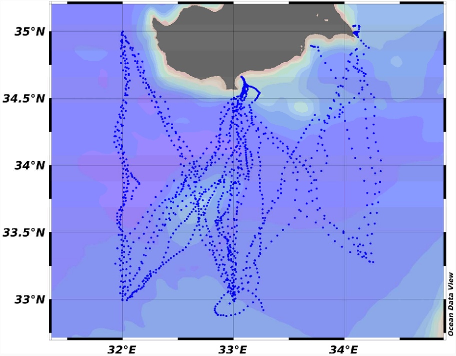
[7] netCDF, "An introduction to netcdf." [Online]. Available: https://www.unidata.ucar.edu/software/netcdf/docs/netcdf_introduction.html

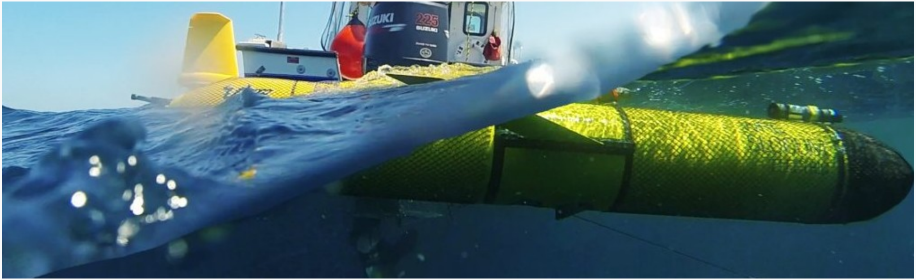
[8] NOAA, "What is netcdf?" [Online]. Available: https://www.esrl.noaa.gov/psd/data/gridded/whatsnetCDF.html

[9] D. Norrison, K. Petterson, and W. Sidebottom, “Numerical study of propeller diameter effects for a self-propelled conventional submarine,” in *Fifth International Symposium on Marine Propulsors*, Espoo, Finland, 2017.

Backup slides

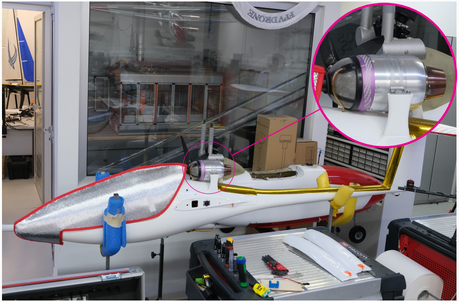
Meanwhile, somewhere beneath the sea surface...





The importance of including vertical layers in climate data

UAV LAB, where the magic happens



cdo – Climate Data Operator https://code.mpimet.mpg.de/projects/cdo/

```
ymonsum        Multi-year monthly sum                           (1|1)
ymonvar        Multi-year monthly variance                      (1|1)
ymonvar1       Multi-year monthly variance (n-1)                (1|1)
yseasadd       Add multi-year seasonal time series              (2|1)
yseasavg       Multi-year seasonal average                      (1|1)
yseasdiv       Divide multi-year seasonal time series           (2|1)
yseasmax       Multi-year seasonal maximum                      (1|1)
yseasmean      Multi-year seasonal mean                         (1|1)
yseasmin       Multi-year seasonal minimum                      (1|1)
yseasmul       Multiply multi-year seasonal time series         (2|1)
yseaspctl      Multi-year seasonal percentile values            (3|1)
yseasrange     Multi-year seasonal range                        (1|1)
yseasstd       Multi-year seasonal standard deviation           (1|1)
yseasstd1      Multi-year seasonal standard deviation (n-1)     (1|1)
```

A collection of many operators for standard processing of climate and forecast model data: from simple statistical and arithmetic functions, data selection and sub-sampling tools, to spatial interpolation. CDO was developed to have the same set of processing functions for `GRIB` and `netCDF` data-sets in one package.

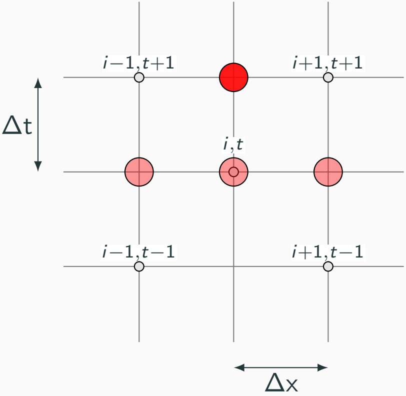
Extremely simple to use and fast in calculating:

`cdo yseasmin TheInputFile.nc TheOutputFile.nc`

operator (yseasmin); existing and created netcdf files (TheInputFile.nc, TheOutputFile.nc)

Finite Difference: a simple 1D case and its "stencil"

State of a system Ψ for each i,t on a **computational grid**, i.e. $\Psi_{i,t}$:



Continuous differentiation:

$$\frac{\partial \Psi}{\partial t} + u_x \cdot \frac{\partial \Psi}{\partial x} = \kappa \frac{\partial^2 \Psi}{\partial x^2}$$

Discretized approximation:

$$\frac{\partial \Psi}{\partial t} \approx \frac{\Psi_{i,t+1} - \Psi_{i,t}}{\Delta t}$$

$$\frac{\partial \Psi}{\partial x} \approx \frac{\Psi_{i+1,t} - \Psi_{i-1,t}}{\Delta x}$$

$$\frac{\partial^2 \Psi}{\partial x^2} \approx \ldots$$

- prediction of Ψ (at time t+1).
- needed to compute prediction: Ψ's initial state over space and time.