WP5 - DATA INTEGRATION AND MANAGEMENT

Tasks 5.4-5.5: Implementing the WP6 showcase applications in collaboration with WP6 tasks

INTAROS 4th Year General Assembly

Remote meeting
15.50-16.00
January 12th 2021

Pedro Gonçalves, Terradue (lead)
Torill Hamre, NERSC (co-lead)
Fabien Ors (ARMINES), Didier Renard (ARMINES)
Herve Caumont (Terradue), Ingo Schewe (AWI), Arnfinn Morvik (IMR)

INTAROS

Agenda

- **Implementing iAOS Showcase Applications**
  - For T6.1 with SMHI: Pan-Arctic Hydrological Modelling
  - For T6.2 with IMR: Barents Sea Multi-depth Temperature & Salinity Maps
  - For T6.4 with FMI: Maps for Svalbard Avalanche Forecast Modelling
  - For T6.8 with Aarhus: Baffin Bay Bottom Temperature Maps
- **Promoting iAOS Showcase Applications**
  - Communications
  - Reusable tools

INTAROS

WP5 / WP6 coordination for the definition of iAOS Showcases

"WP6 will integrate **remote sensing data** and **in situ observations delivered through WP5**, from a variety of platforms and geographical scales and locations. *Incorporation of these data into analysis and modelling systems, including physical and* ***ecological process models****,* ***climate models*** *and* ***forecast methods****, will provide support for better products to key societal areas*."

-- from WP6 Description of work

Helsinki, 2018 - Joint WP5-6 Workshop (but WP6 starting later)

Bremen, 2019 - RGeostats Workshop

Sopot, 2020 - Interviews with WP6 task leaders for their work plan analysis, and identification of best 'showcase' opportunities to be supported by the iAOS (WP5)

- As return of experience (INTAROS internal) on how the WPs interact in order to illustrate the iAOS added value
- As a set of results-oriented data collections and services, which can support the INTAROS outreach activities in 2020-2021

Remote, 2020 - Intermediate results reviews and definition of final objectives for each Showcase

INTAROS

Implementing
iAOS Showcase Applications

INTAROS

Pan-Arctic Hydrological Modelling (1/3)
T6.1 SMHI

Objectives

- Have the "observational" data available for search and download from the iAOS Portal
- Have the Arctic-HYPE produced at SMHI and provide the data as open data from SMHI repositories:
  - Daily analyses of last 60 days
  - Medium range forecast of coming 10 days
- Improve predictions of spring floods, river ice breakup and freshwater flow to Arctic Ocean, cf.INTAROS D6.1 Climate model initialization v1.4

Data exploitation

- Implement HYCOS pre-processing (both archive of quality controlled data with 4months/2years lag, or provisional datasets)
  - In-house server at SMHI
  - Cloud-based, using Ellip, to compare
- Schedule HYCOS pre-processing operations to be made daily at a certain time
- Setup OpenDAP server for publishing the Arctic-HYPE model results

Data sources

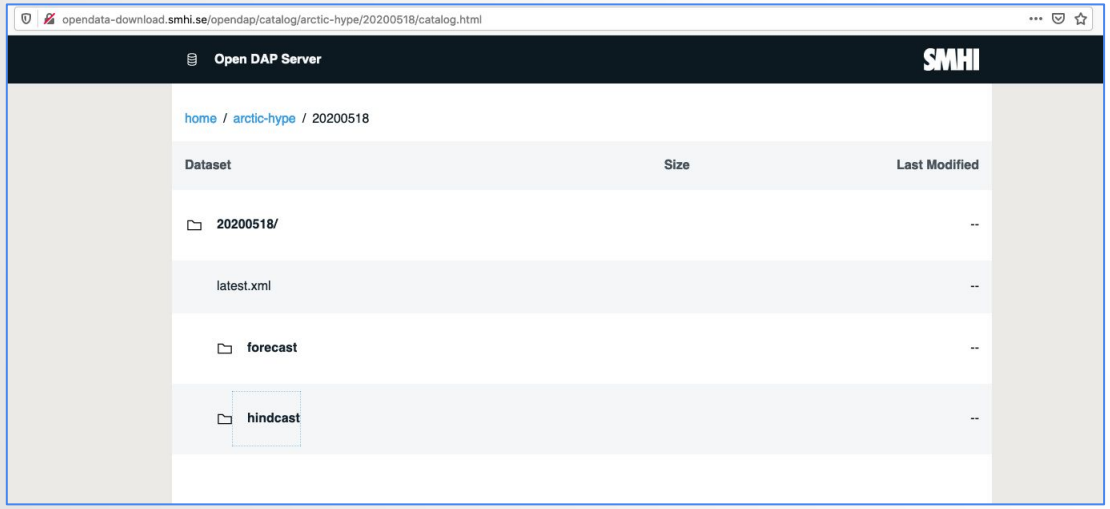
- River discharge data from the Arctic Hydrological Cycle Observing system (Arctic-HYCOS) - assessed and enhanced in INTAROS WP2 https://catalog-intaros.nersc.no/dataset/arctic-hycos-hydrological-data/
- HydroGFD v3 temperature and precipitation data
- ECMWF deterministic medium range weather forecasts

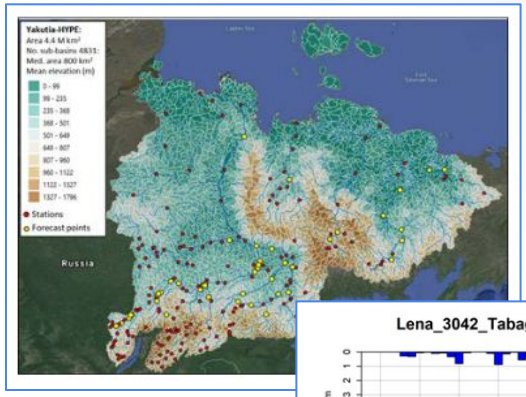
INTAROS

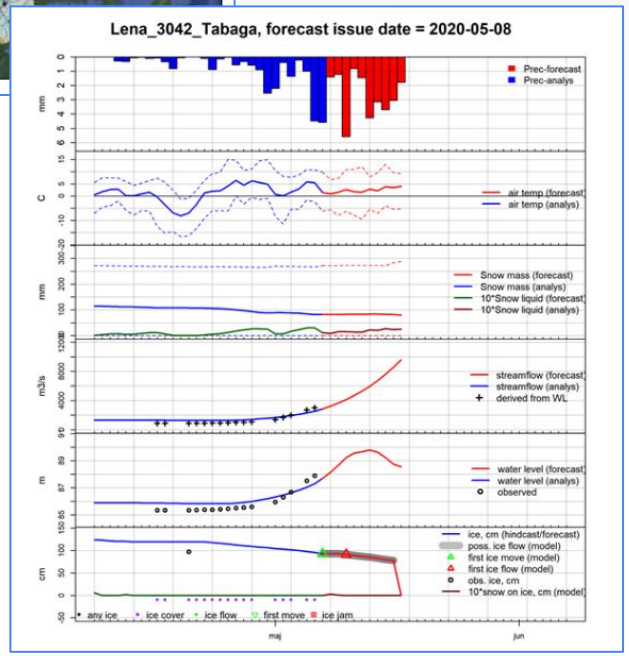
Pan-Arctic Hydrological Modelling (2/3)
T6.1 SMHI

Current results

- OpenDAP server publishing Pan-arctic hydrological model Arctic-HYPE results provided by SMHI http://opendata-download.smhi.se/opendap/catalog/catalog.html







Pan-Arctic Hydrological Modelling (3/3)
T6.1 SMHI

Perspectives

iAOS Showcase (demo)	**Use of Arctic-HYCOS sources (multi-provider)**	**Use of Arctic-HYPE (SMHI) outputs (per station ID, with 1 dimension time)**
Flood forecasting use cases Select a point, or some AOI (in Russia MPI in Yakutsk - L. Lebedeva)	For checking the initial conditions	For checking the model analysis of current condition and the model forecast
Entire freshwater inflow to the ocean (in collaboration with IMR - R.Hordoir)	To integrate discharge into the ocean from observations (only represent 60% draining land, see WP2 deliverables).	To integrate model analysis/predictions on river discharge into the ocean on user defined resolution

Barents Sea Multi-depth Temperature & Salinity Maps (1/3)
T6.2 IMR

Objectives

- Use the **Geostatistical Library** (RIntaros / RGeostats) and build the R software for interpolating maps from **CTD datasets**
- Generate temperature and salinity fields for:
  - modelling of Arctic Ocean biogeochemistry
  - validation of climate model projections (NorCMP)
- Build a **Web Processing Service** for iAOS

Data exploitation

- Use of the iAOS OpenDAP server at NODC
- Explanatory data analysis and variography
- Modelisation of **spatial behavior** for Temperature and Salinity

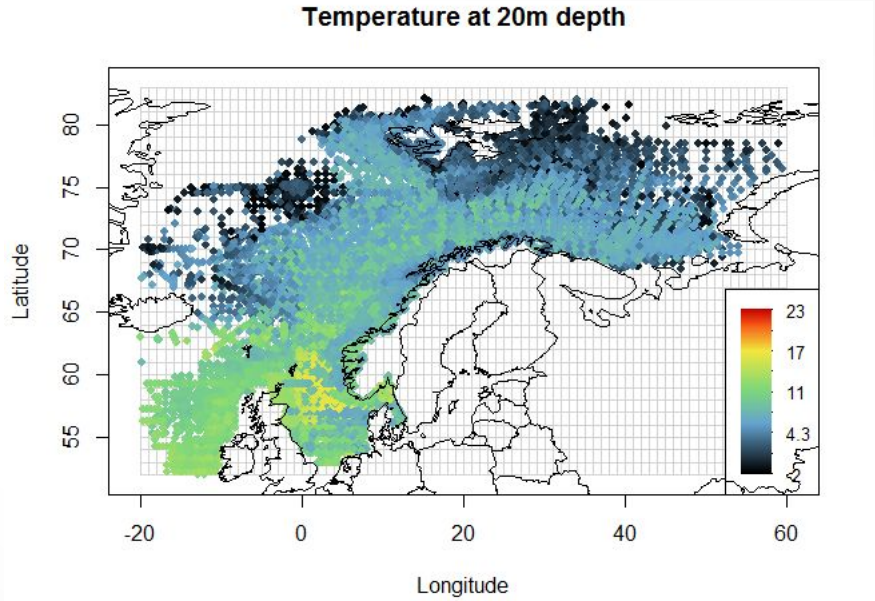
Data sources

Institute of Marine Research (**IMR**) center of Norway, 7 research vessels collecting conductivity, temperature, depth (CTD) data in the North Sea.

- Acquisitions between 7th January 1995 and 29th November 2016.
- 5.5 billion of samples measured over 63 500 positions (vertical profiles).

All files are freely available on an OpenDAP server:
http://opendap1.nodc.no/opendap/physics/point/yearly/contents.html

NetCDF files (one file by year and per vessel). Coordinates are in degrees (Long/Lat) and the timestamp is the number of minutes since the 1st January 1950. The whole dataset volume is 880 GB.

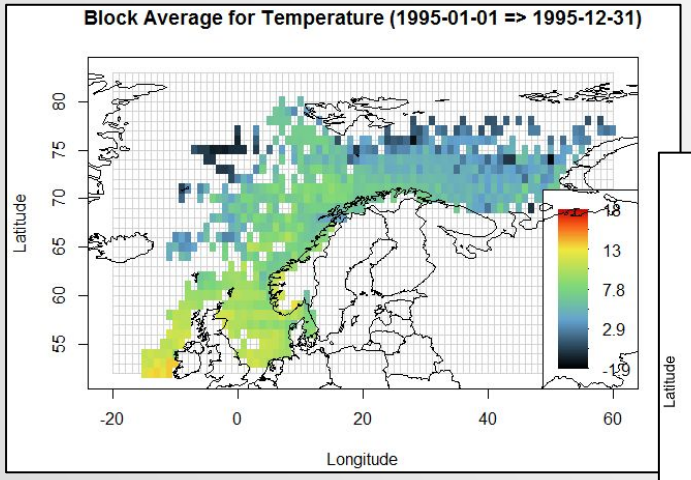


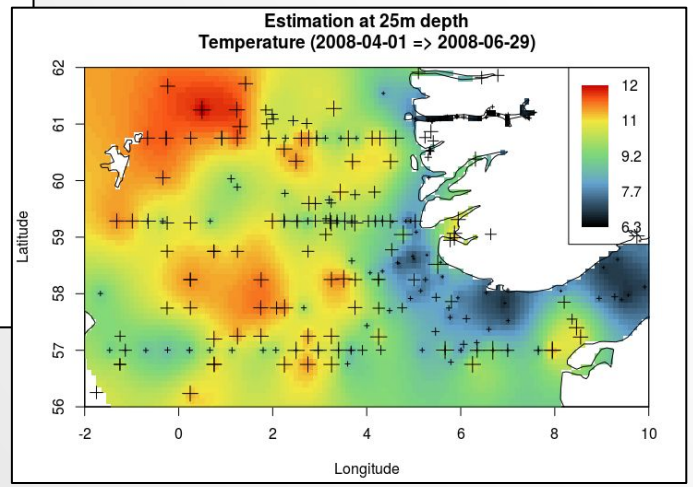
Barents Sea Multi-depth Temperature & Salinity Maps (2/3)
T6.2 IMR

Current results

Standalone solution (R software)

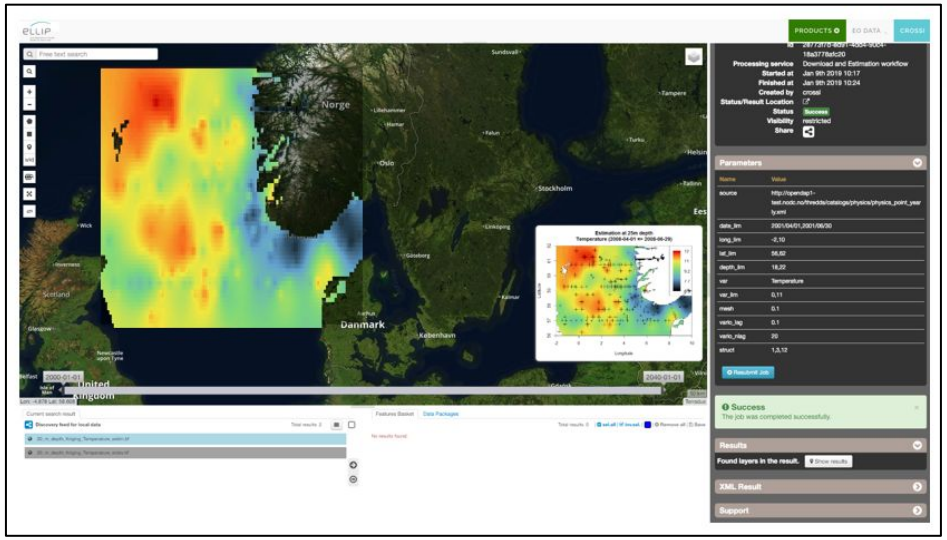
- Map productions per run:
  - Base maps
  - Average per cell
  - Cross-validation (blind test) maps
  - **Estimation** (Temperature / Salinity) and corresponding **uncertainty maps**





Solution as-a-Service (Cloud software)

- Map productions per WPS run:
  - On-demand, self performed by each user from the **Portal**
  - Split tiles for large areas

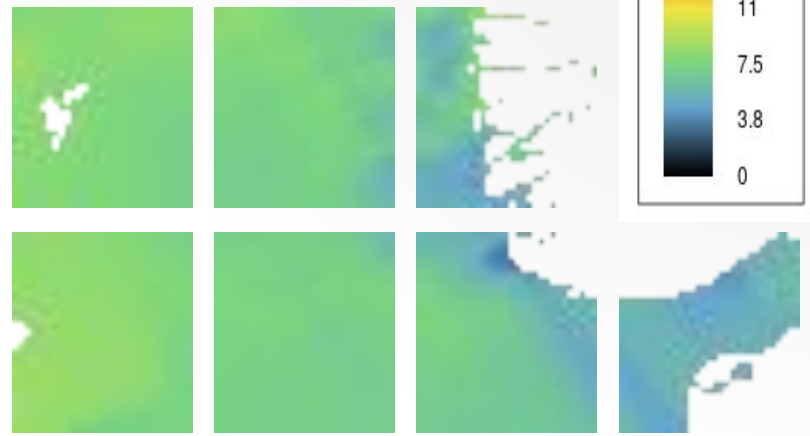


Barents Sea Multi-depth Temperature & Salinity Maps (3/3)
T6.2 IMR

Perspectives

Methodology

- **Improve** the initial geostatistical spatial analysis performed in 2018 by:
  - building a unique spatio-temporal variogram model for Temperature
  - taking into account co-variables like (salinity, bathymetry...)
  - handling currents non stationarities and distance to coastline



Standalone solution (R software)

- **Promote** the Jupyter Notebook and Rmarkdown scripts on iAOS and their documentation on GitHub
- Build a flyer for addressing the reusability of the solution

Solution as-a-Service (Cloud software)

- **Promote** the use of the Ellip Solutions for parallel processing
  https://gitlab.com/ec-intaros/dcs-imr-estim
- Consider the use of a single global variogram model for each computing node in charge of each kriging tile

RGeostats

ELLIP TERRADUE

INTAROS

Maps for Svalbard Avalanche Forecast Modelling (1/3)
T6.4 FMI

Objectives

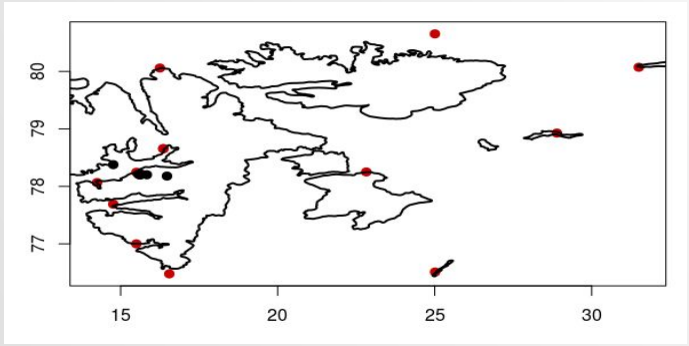
- Use the **Geostatistical Library** (RIntaros / RGeostats) and build the R software for interpolating maps from snow stations, arome model output and terrain model
- Generate **snow depth maps** at regular time intervals as input for avalanche forecast model

Data exploitation

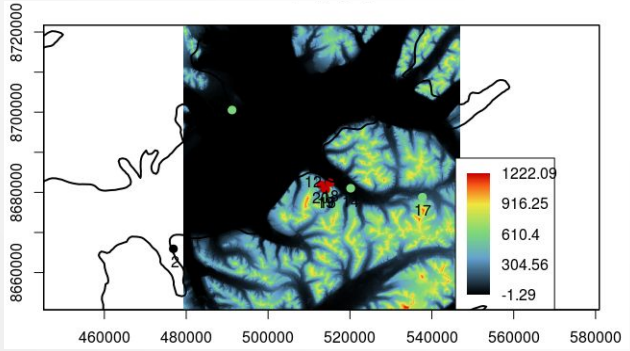
- Explanatory data analysis and variography
- Handle the different spatial distributions and resolutions of the data ("support")
- Modelisation of **spatial and temporal behavior** for **snow depth** through co-variables (temperature, wind speed by class of wind direction)

Data sources

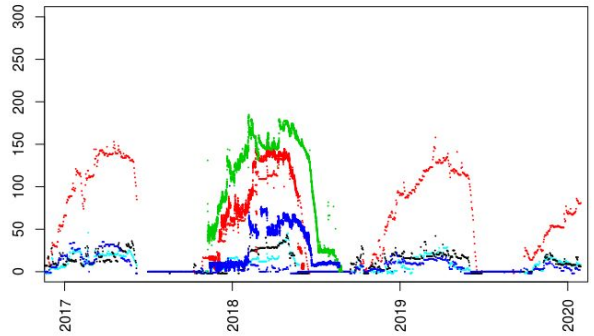
- NMI Frost API historical weather and climate data stations (selected files on shared Drive)
- NMI arome model: https://thredds.met.no/thredds/catalog/aromearcticarchive/catalog.html
- Norwegian Polar Institute Svalbard Terrain Model: https://doi.org/10.21334/npolar.2014.dce53a47



Location of Stations



Terrain Model



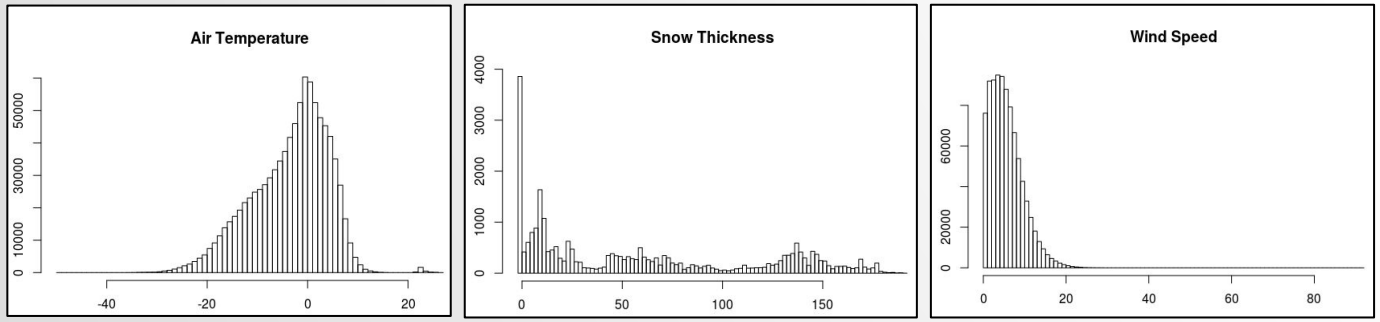
Snow Time Series at Stations

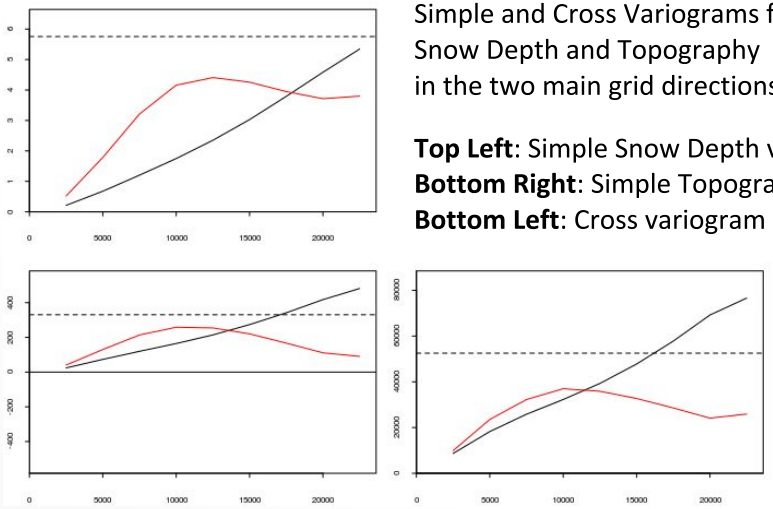
Maps for Svalbard Avalanche Forecast Modelling (2/3)
T6.4 FMI

Current results

- **Pre-analysis** of the data :
  - Few stations: temporal series of snow thickness measured at short time steps
  - Arome models: various maps covering the whole area, every 6 hours, on large scale grid (incl. snow thickness derived from model)
  - Several co-variables
- Processing :
  - Regularization of the station time series by averaging over 6 hours
  - **Correlation** (space-time) of snow depth variable with arome model output
  - Estimation using both informations (with relevant co-variables) over a small scale grid, at regular 6 hours intervals.

Stations histograms





Simple and Cross Variograms for Snow Depth and Topography in the two main grid directions.

Top Left: Simple Snow Depth variogram
Bottom Right: Simple Topography variogram
Bottom Left: Cross variogram

Maps for Svalbard Avalanche Forecast Modelling (3/3)
T6.4 FMI

Perspectives

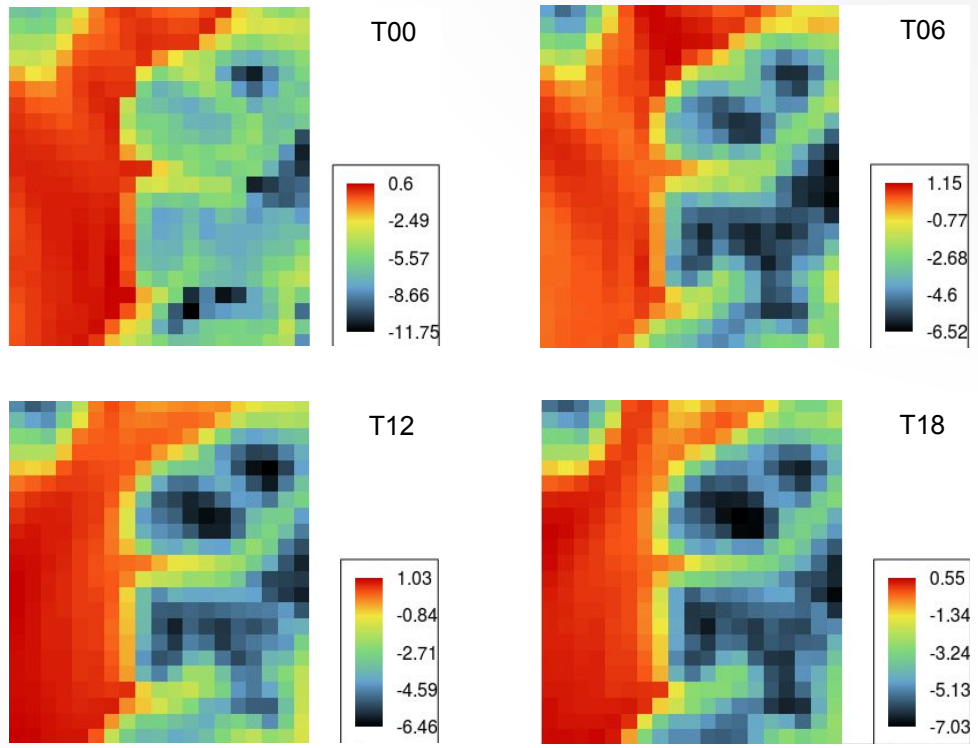
Methodology

- **Finalize** the showcase with a proof-of-concept:
  - Create snow depth variogram models (with relevant co-variables) for each class of wind orientation
  - Generate snow depth map for:
    - a given date and time interval (6 hours)
    - accounting for a global wind orientation

Standalone solution (R software)

- **Promote** the Jupyter Notebook and Rmarkdown scripts on iAOS and their documentation on GitHub
- Build a flyer for addressing the reusability of the solution

Coarse Grid – Time slices (T°C)



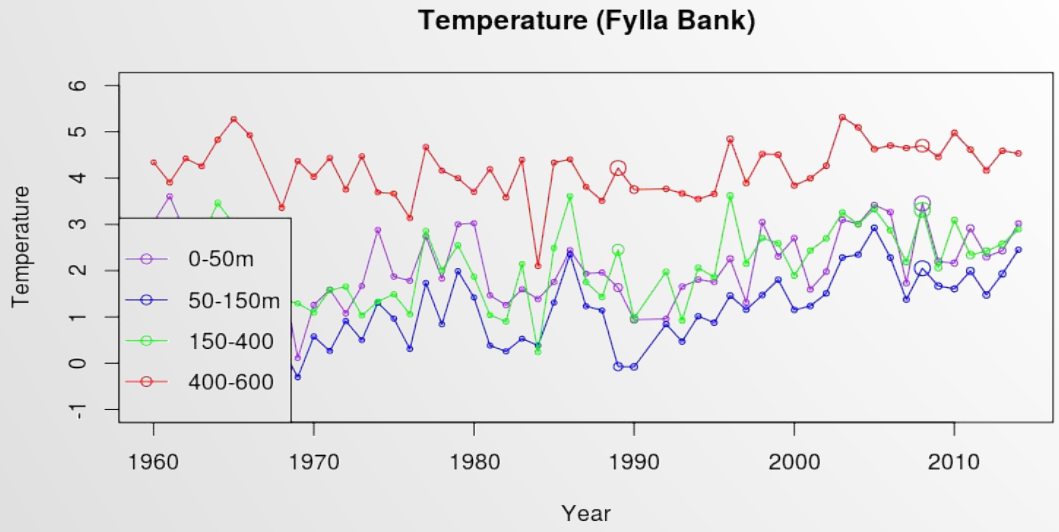
RGeostats

INTAROS

Baffin Bay Bottom Temperature Maps (1/4)
T6.8 Aarhus

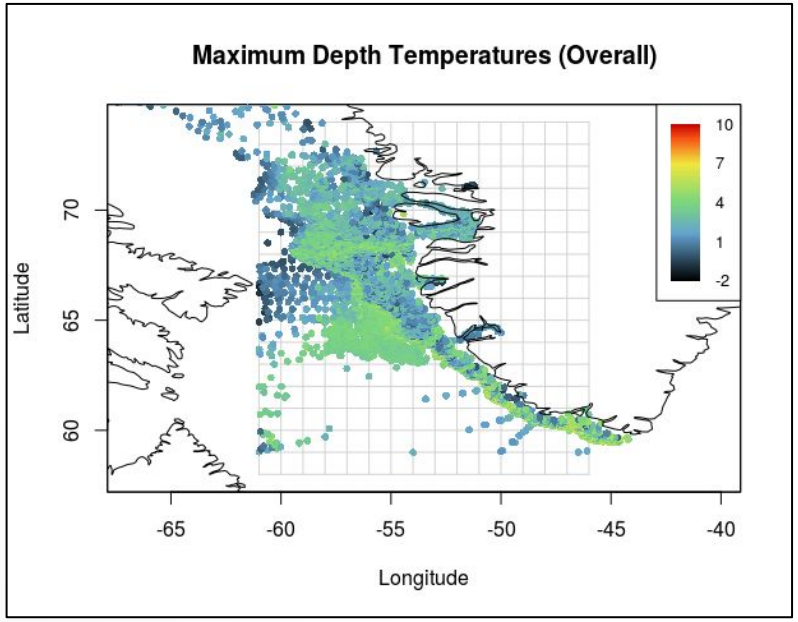
Objectives

- Use the **Geostatistical Library** (RIntaros / RGeostats) and build the R software for interpolating ocean floor temperature maps from CTD and Trawl datasets
- Generate temperature fields at bottom of the ocean in support of:
  - Analysis of **long term global warming** influence
  - Analysis of the **fish stock correlation** to bottom temperature



Data Exploitation

- Explanatory data analysis and variography
- Local and global temperature evolution analysis
- Modelisation of **spatial and temporal behavior** for **ocean floor temperature** through Bathymetry co-variable



Baffin Bay Bottom Temperature Maps (2/4)
T6.8 Aarhus

CTDs (ICES)
- 1977 to 2017
- 3700 vertical profiles
- 1.34M samples
- 1 CSV file (95Mo)

Bottles (ICES)
- 1960 to 2017
- 7800 vertical profiles
- 167K samples
- 1 CSV file (11Mo)

Trawls (GINR)
- 1988 to 2016
- Catches near sea floor only
- 51K samples
- 1 CSV file (11Mo)

Greenland Institute of Natural Resources

Bathymetry (Gebco)
- Grid lag = 1/250 degree
- Grid size = 10320x4560 nodes
- 47M samples
- 1 NetCDF file (94Mo)

New data (M. Sejr, Nov. 2020):

Better quality of arctic bathymetry: https://www.nature.com/articles/s41597-020-0520-9

World ocean database (Temp fields): https://www.nodc.noaa.gov/OC5/WOD/pr_wod.html

Global temperature and salinity profile program (data?): https://www.nodc.noaa.gov/GTSPP

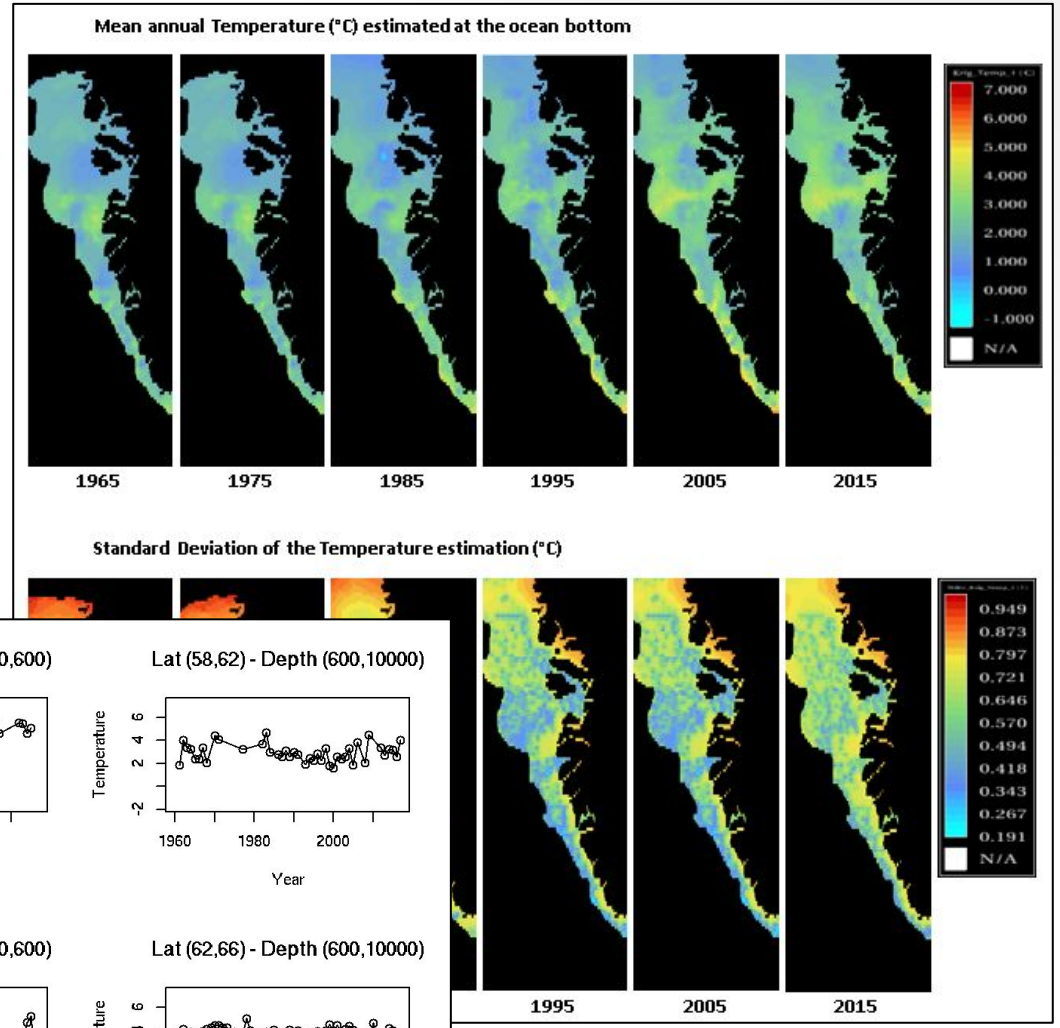
NASA project in Greenland (new CTD): https://omg.jpl.nasa.gov/portal/browse/OMGEV-AXCTD/

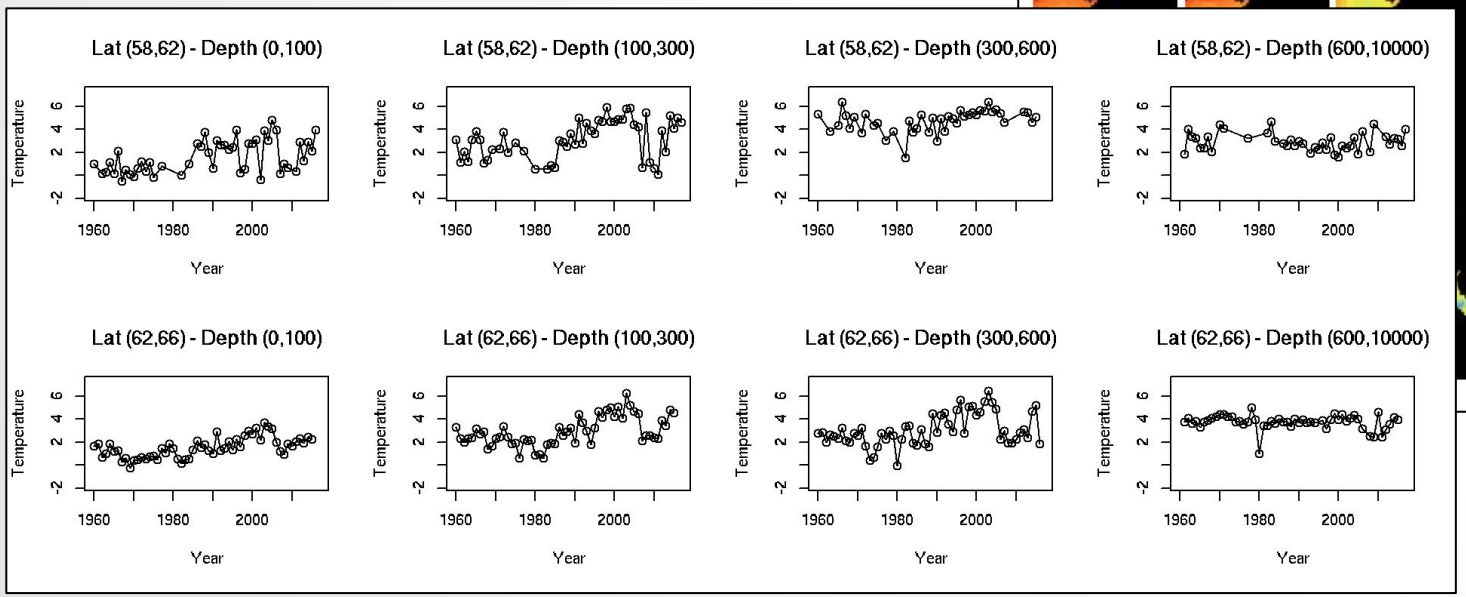
INTAROS

Baffin Bay Bottom Temperature Maps (3/4)
T6.8 Aarhus

Current results

- Temporal evolution (global and local):
  - Between 1960 and 2015, T°C has gained 1°C (around 1995)
  - Lower T°C values around 2008 have been recorded
- Map productions:
  - Basemap of data
  - **Bottom temperature estimation** by year and its standard deviation
- Time series of average temperature by region (Lat/Depth)





Baffin Bay Bottom Temperature Maps (4/4)
T6.8 Aarhus

Perspectives

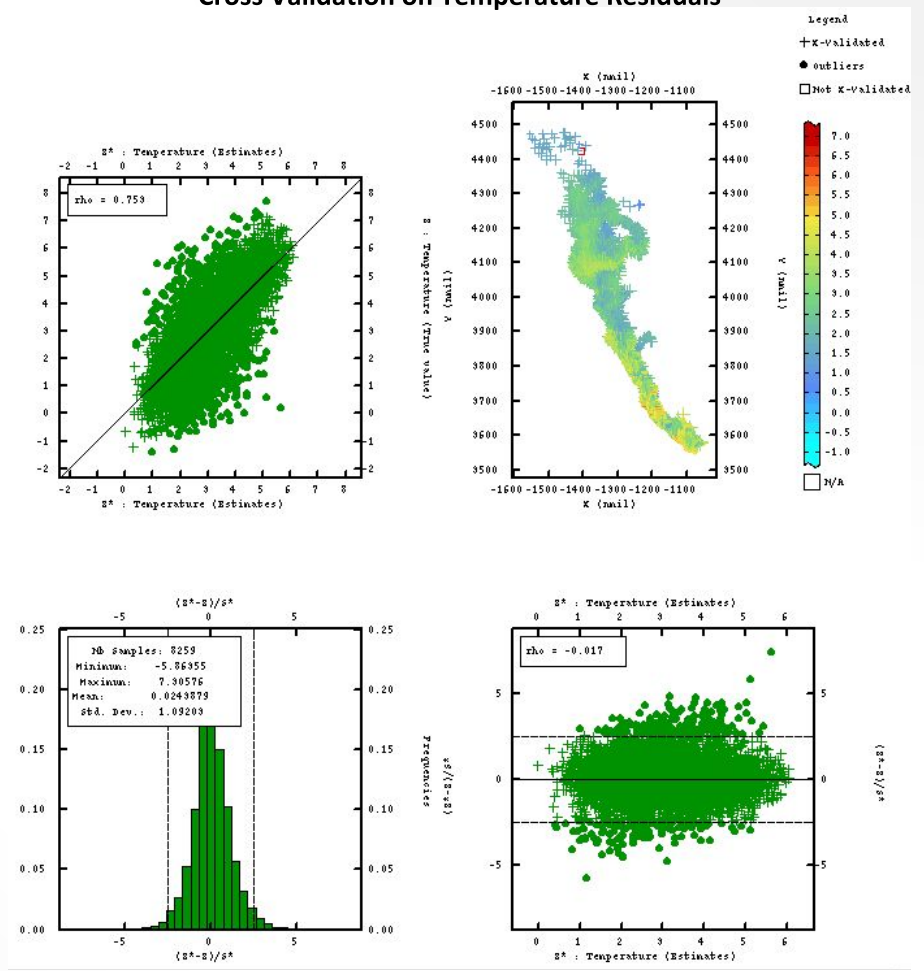
Methodology

- **Improve** first kriging estimation:
  - Reduce the estimation error (currently around 0.7°C)
  - Reduce estimation smoothing
- Next improvements:
  - Multi-directional variography and zonal anisotropies
  - Local cross-validation with additional abyssal (or new) data
  - **Bathymetry** as external drift using non linear regression
  - **Salinity** as co-variable to be studied

Standalone solution (R software)

- **Promote** the Jupyter Notebook and Rmarkdown scripts on iAOS and their documentation on GitHub
- Build a flyer for addressing the reusability of the solution

Cross Validation on Temperature Residuals



Promoting
iAOS Showcase Applications

INTAROS

Communications

For each of the four WP6 Showcased applications, the communications will emphasize the driving domain aspects such as listed hereafter, as well as the reusable tools:

Pan-Arctic Hydrological Modelling (T6.1 SMHI)

- Understanding of climate impact on fine-scale hydrology in the Arctic
- Predictions of river discharge into the ocean for present and future climate

Barents Sea Multi-depth Temperature & Salinity Maps (T6.2 IMR)

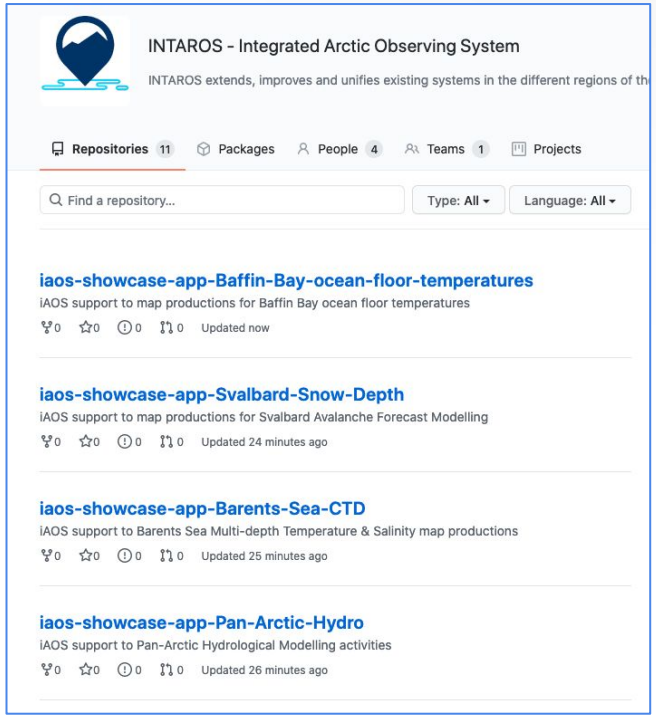
- Inputs for Ecosystem models (e.g. NORWECOM, ATLANTIS, FLEXEM)
- Arctic Ocean biogeochemistry / Climate model projections

Maps for Svalbard Avalanche Forecast Modelling (T6.4 FMI)

- Snow depth interpolation at regular time intervals

Baffin Bay Bottom Temperature Maps (T6.8 Aarhus)

- Analysis of long term global warming influence
- Analysis of the fish stock correlation to bottom temperature



https://github.com/ec-intaros

INTAROS

Reusable tools

Out of the four WP6 Showcased applications, the tools made available on the iAOS to support the implementation of user applications are listed hereafter:

Data Servers

- Arctic-HYPE pan-Arctic hydro model
  http://opendata-download.smhi.se/opendap/catalog/catalog.html
- NODC's CTD 1995-2016
  http://opendap1.nodc.no/opendap/physics/point/yearly/contents.html

R code and Jupyter Notebooks

- RGeostats Bremen Workshop (2019)
  https://github.com/ec-intaros/RGeostats-workshop
- RGeostats resources per showcase
  https://github.com/ec-intaros?q=iaos-showcase

Software libraries

- RGeostats Conda package
  https://anaconda.org/Terradue/r-rgeostats
- RIntaros Conda package
  https://anaconda.org/Terradue/r-rintaros

Cloud services

- Ellip Notebooks solution
  https://ellip.terradue.com/#!notebooks
- Ellip Workflows solution
  https://ellip.terradue.com/#!workflows

End of presentation

Back-up slides

TASK 5.4 - DEVELOPMENT OF GEO-STATISTICAL METHODS FOR DATA INTEGRATION

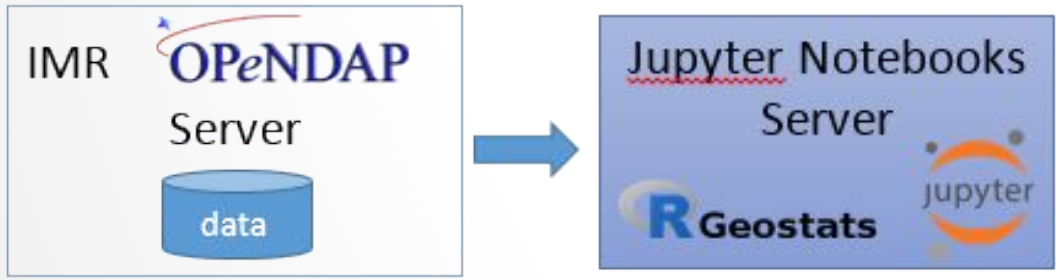
Partners : ARMINES, NERSC, DTU

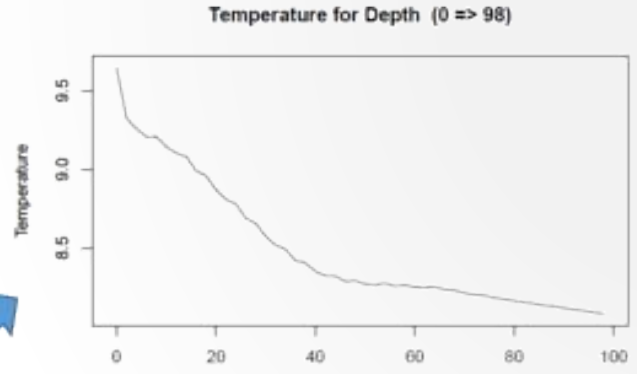
- Deployment of RGeostats package with anaconda for iAOS developers: https://anaconda.org/Terradue/r-rgeostats
- Creation and deployment of RIntaros Geostatistical package for iAOS developers: https://anaconda.org/Terradue/r-rintaros
- Dissemination material produced to outreach the iAOS users community (Bremen Workshop, Terradue Seminar at Fontainebleau, XIVème Journées de Géostatistique)
- Overall, the task activities are progressing according to the schedule.

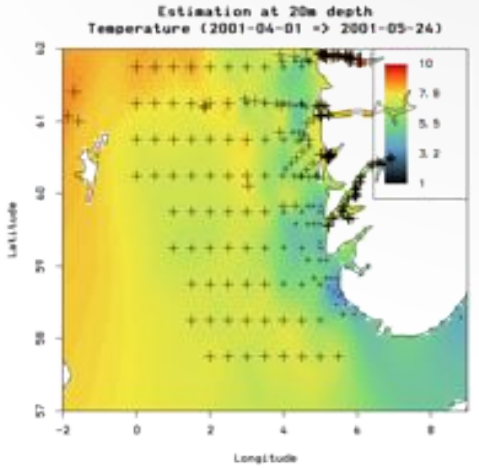
INTAROS

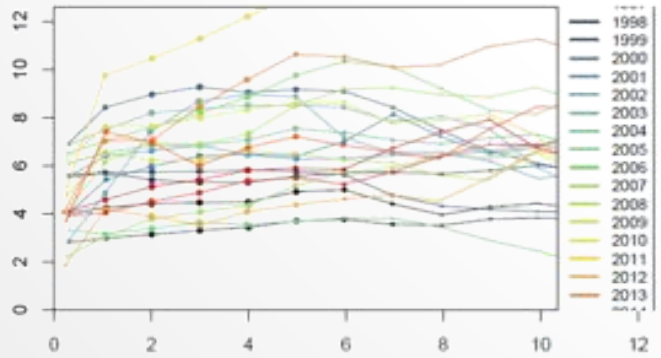
iAOS Platform
IMR Dataset Analysis

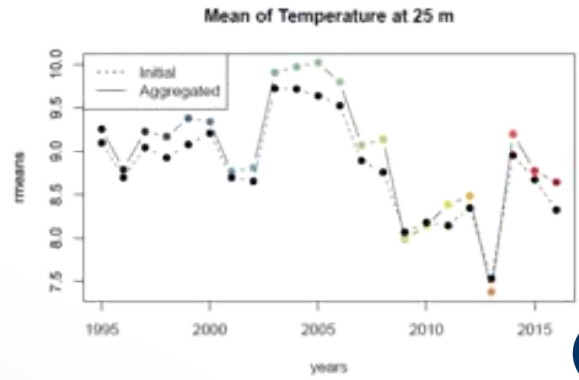
- Use of Jupyter Notebooks Terradue solution
- Access to IMR dataset (OPeNDAP)
- Geostatistical analysis of IMR dataset using RGeostats





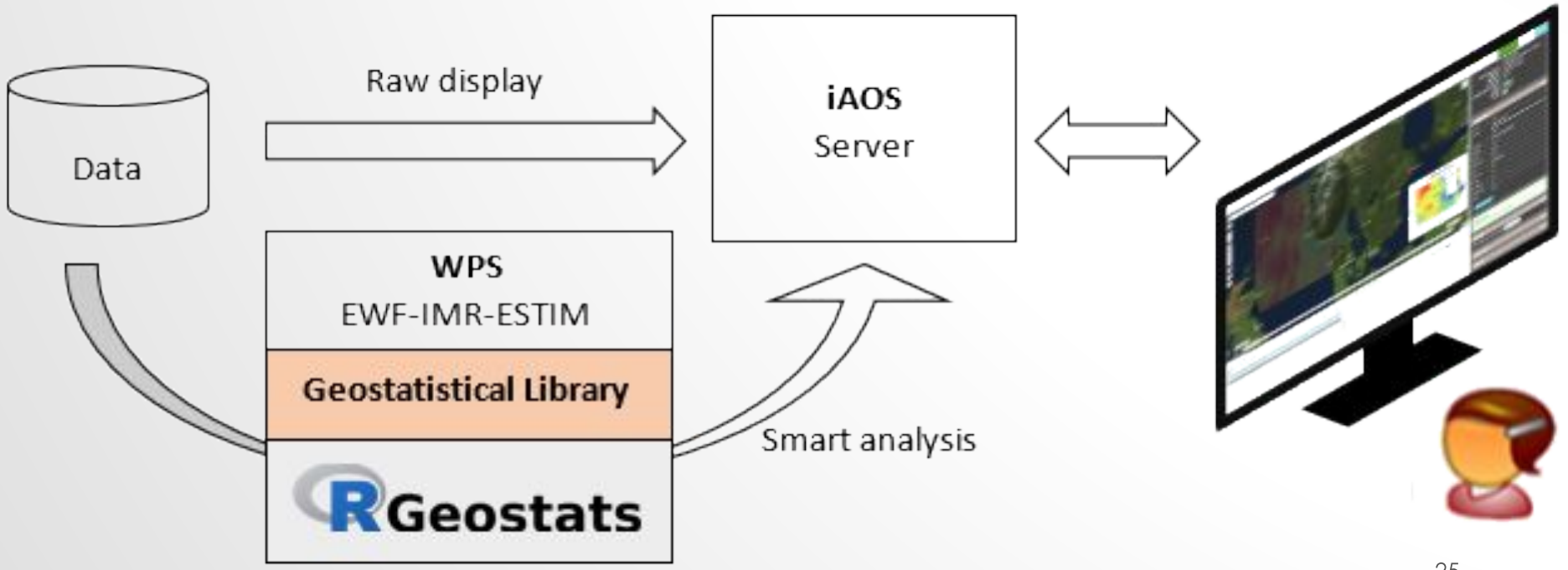






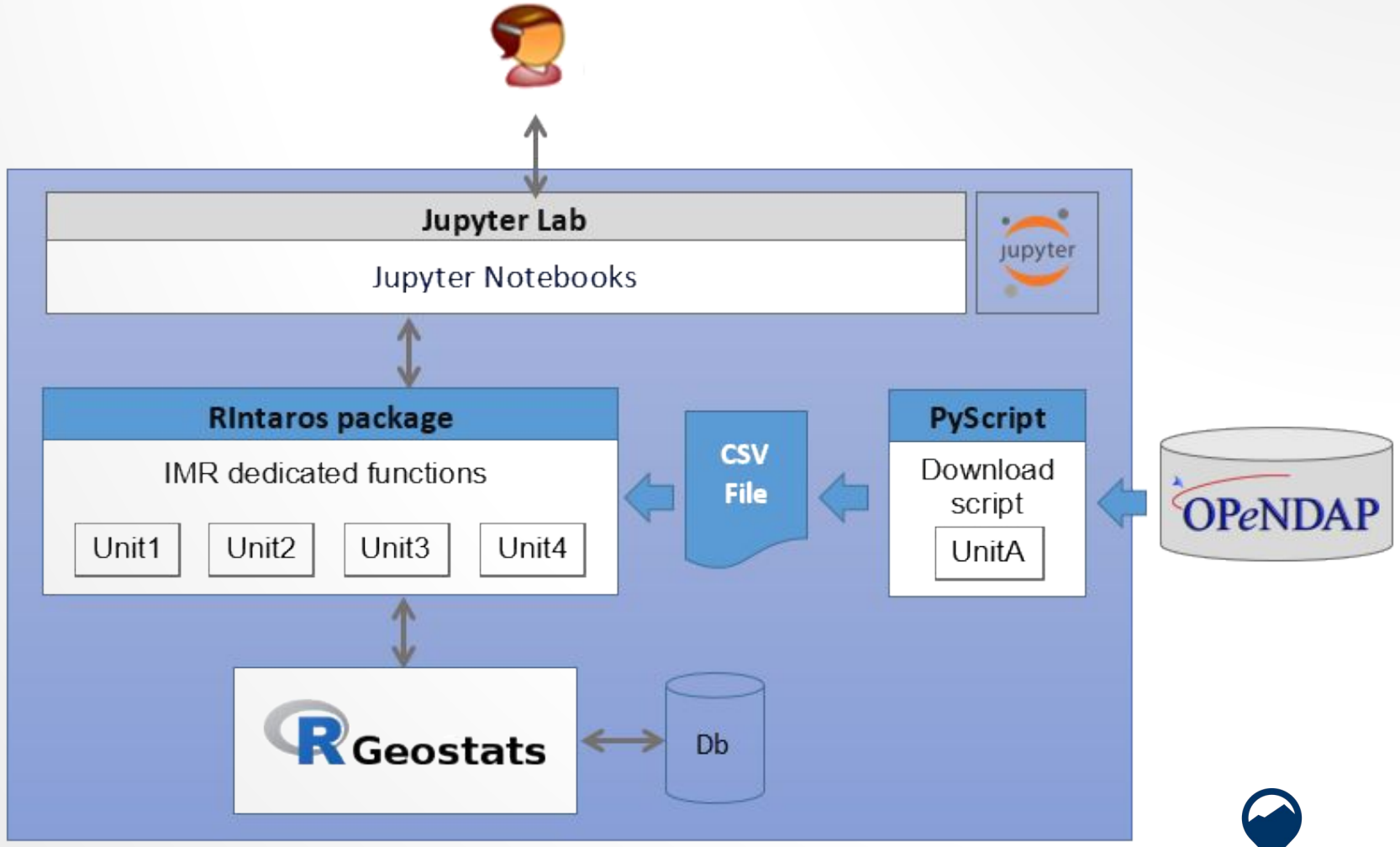
iAOS Platform
Geostatistical Library Creation

- Creation of the Geostatistical Library (**RIntaros**) as a wrapper of RGeostats package for smarter analysis of IMR dataset (to be extended later to another INTAROS datasets)



iAOS Platform
Unitary Scripts using RIntaros

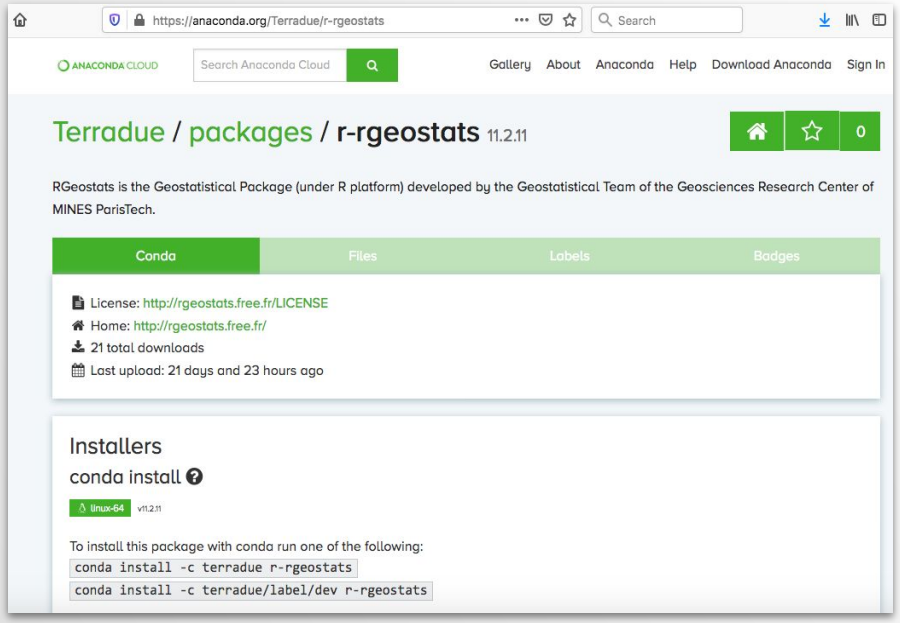
- New Python script for downloading IMR data
- Design of unitary R scripts using RIntaros package for each Geostatistical operation

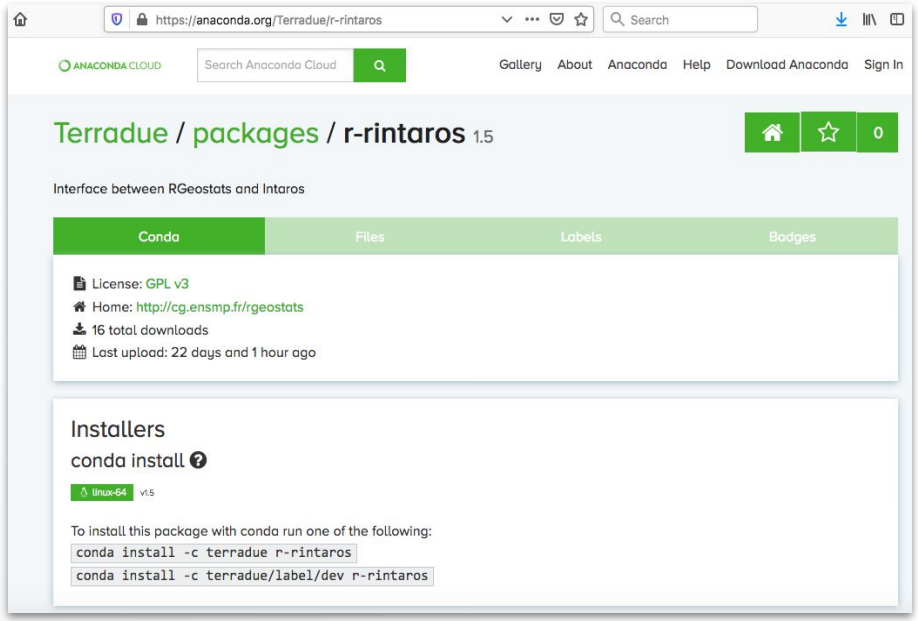


INTAROS

iAOS Platform
RGeostats and RIntaros Deployment

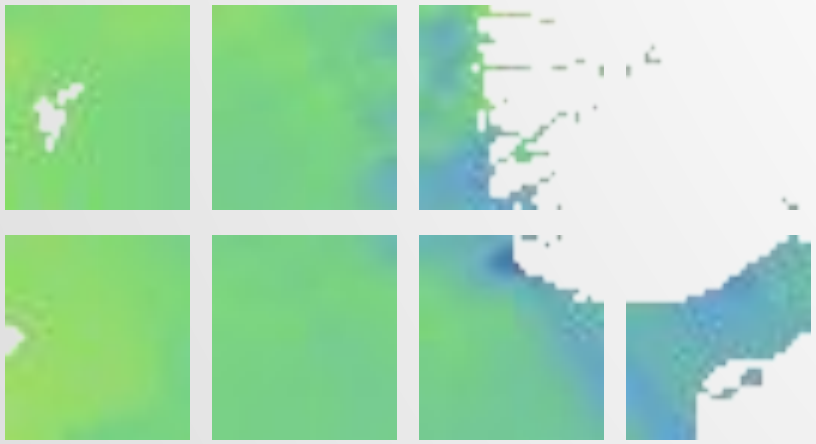
- Deployment of RGeostats and RIntaros packages as anaconda package for iAOS developers

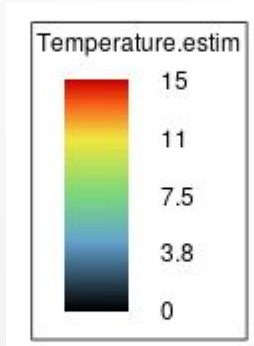


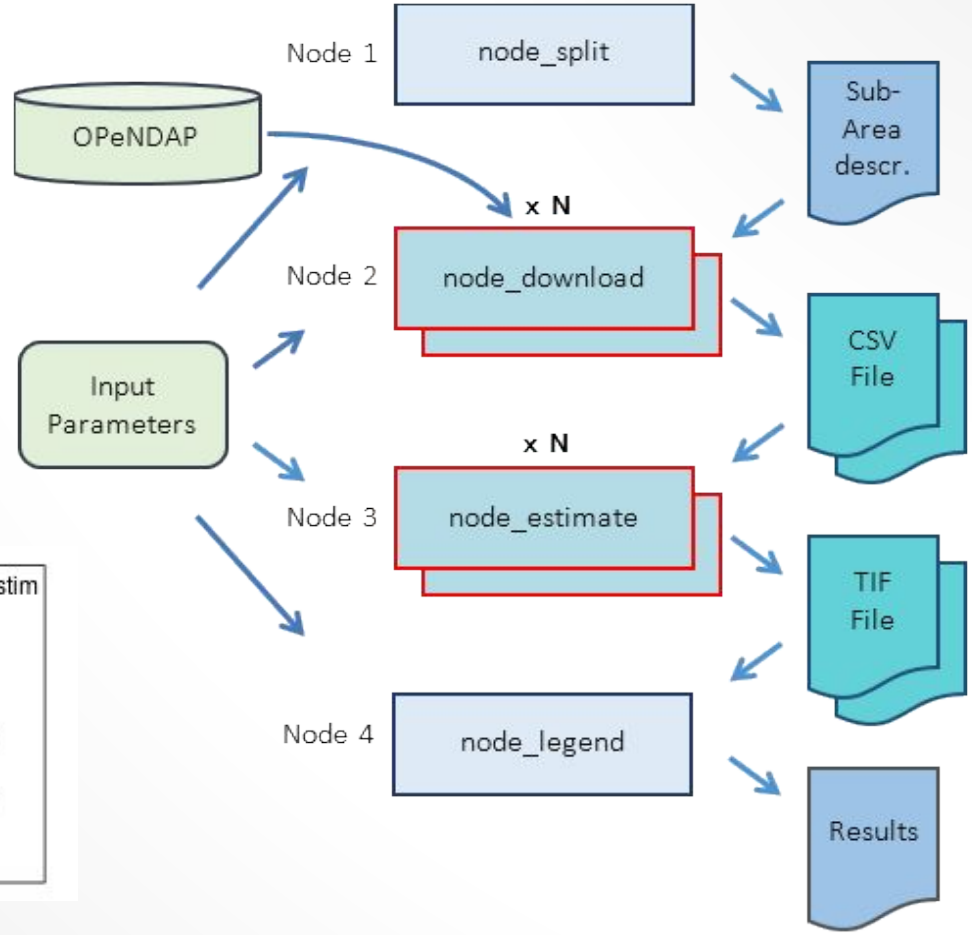


iAOS Platform
Workflow Parallelization

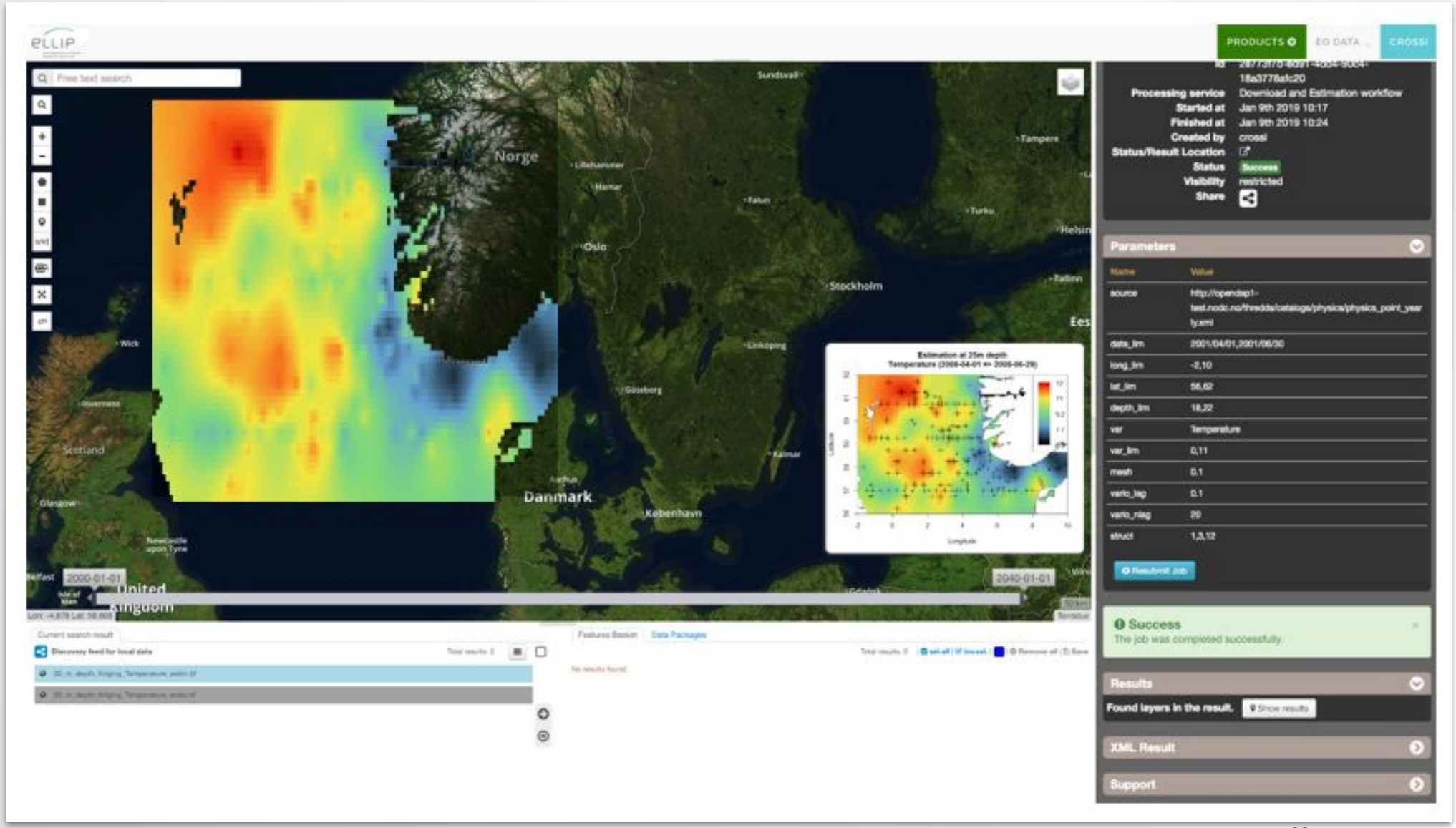
- Improving efficiency:
  - Duplication of downloading node
  - Estimation by tile







iAOS Platform with new **ewf-imr-estim**



TASK 5.5 - INTEGRATION OF NEW PROCESSING SERVICES

Partners : Terradue, NERSC, ARMINES

- User support on the use of Terradue's Ellip solutions to the partners NERSC and ARMINES for their service integration activities.
- New version of dcs-imr-estim Web Processing Service dedicated to IMR dataset analysis for iAOS users: https://gitlab.com/ec-intaros/dcs-imr-estim (Ellip Workflows app)
- Description of the applications integration work into the deliverable D5.7 Processing services for iAOS v1.
- Overall, the task activities are progressing according to the schedule.

INTAROS