

WP5 – Training session: RGeostats

Ellip Solutions for IMR dataset

INTAROS – General Assembly

Haus der Wissenschaft,
Sandstrasse 4/5
28195 Bremen, Germany

14.00-17.30

January 10th 2019

ORS Fabien, ARMINES (lead)
RENARD Didier, ARMINES (co-lead)



Outline

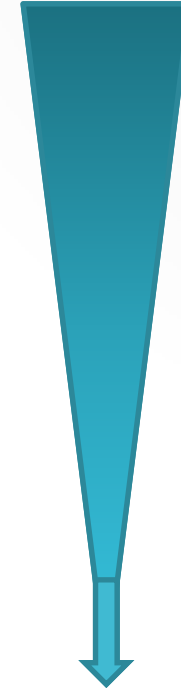
Thursday 10th

1. Creating iAOS Processing Services
2. Geostatistics and RGeostats
3. Ellip Notebooks using RGeostats
4. Ellip Workflow using RGeostats
5. IMR Case Study - RGeostats in Action!

Friday 11th

Geostatistics Course & Exercices

Details level



Part #1

Creating iAOS Processing Services

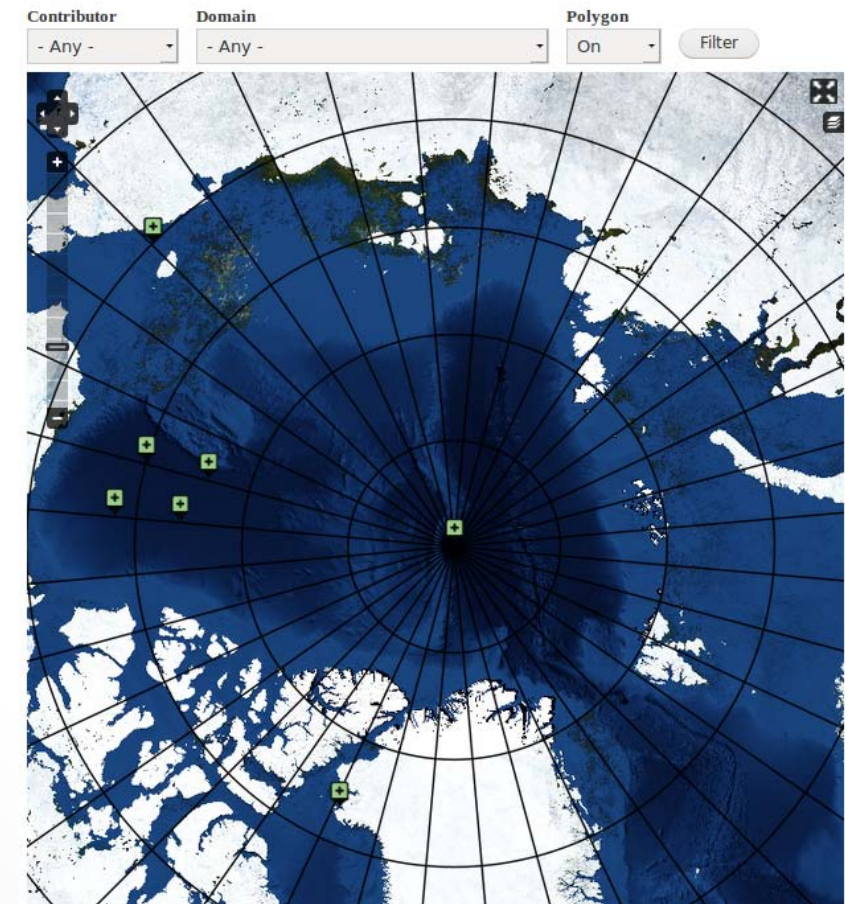


Creating iAOS Processing Services

Why?

- Visualizing and analyzing Pan-Arctic data on the iAOS
- Studying spatial and temporal correlations
- Mapping multiple variables
- Combining different data sources
- Learning more from available data with a deeper analysis

Map of observing sites



Creating iAOS Processing Services

For who?

- Developers:
 - INTAROS partners (data providers, iAOS developers, WP6 projects...)
 - INTAROS arctic scientific community (researchers, scientists...)
- iAOS users (*visitors*):
 - Scientists (climate, meteorology, biology, oceanography,...)
 - Companies (tourism, earth resources, environments...)
 - Non Governmental Organization (climate change, pollution, life preservation...)
 - European commission (politics)
 - World citizens



Creating iAOS Processing Services

For who?

- **Developers:**
 - INTAROS partners (data providers, stakeholders...)
 - Arctic scientific community (researchers, scientists)
- iAOS users (*visitors*):
 - Scientists (climatologists, meteorologists, biologists, oceanologists,...)
 - Companies (tourism, earth resources, environments)
 - Non Governmental Organization (climate change, pollution, life preservation)
 - European commission (politics)
 - Curious world citizens



The rest of this talk is mainly targeted to **developers**

Creating iAOS Processing Services

New Ellip Solution provided by TERRADUE (sept. 2018)



<https://notebooks-qa.terradue.com>

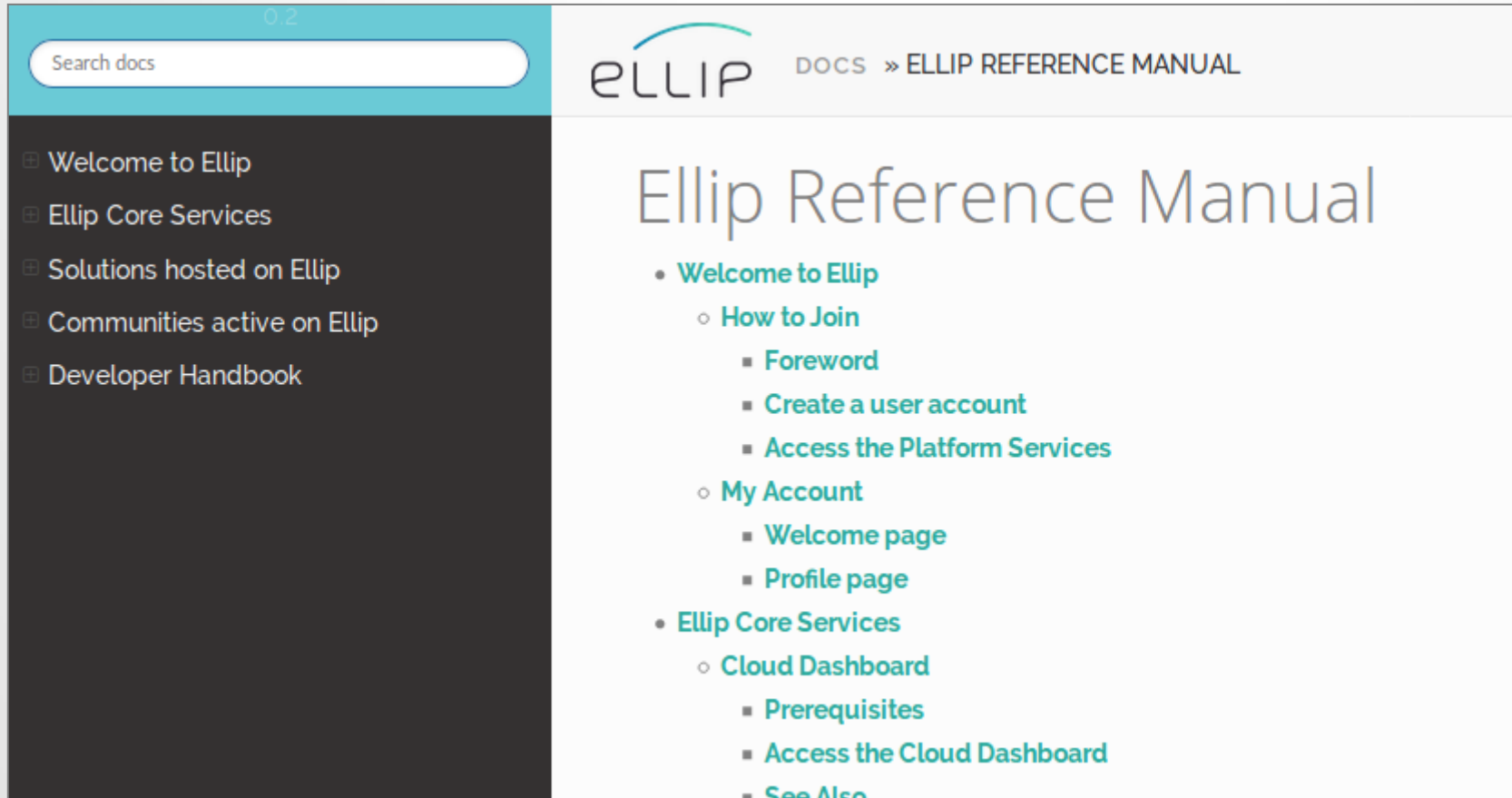
Objectives:

- Make data processing service **design easier**
- **Build, deploy** and **monitor** data processing services



Creating iAOS Processing Services

Full set of Ellip documentation



The screenshot shows the Ellip Reference Manual page. On the left is a dark sidebar with a search bar and a list of navigation items: Welcome to Ellip, Ellip Core Services, Solutions hosted on Ellip, Communities active on Ellip, and Developer Handbook. The main content area has a light blue header with the Ellip logo and the text 'DOCS » ELLIP REFERENCE MANUAL'. Below the header is the title 'Ellip Reference Manual' and a table of contents with the following items:

- Welcome to Ellip
 - How to Join
 - Foreword
 - Create a user account
 - Access the Platform Services
 - My Account
 - Welcome page
 - Profile page
- Ellip Core Services
 - Cloud Dashboard
 - Prerequisites
 - Access the Cloud Dashboard
 - See Also

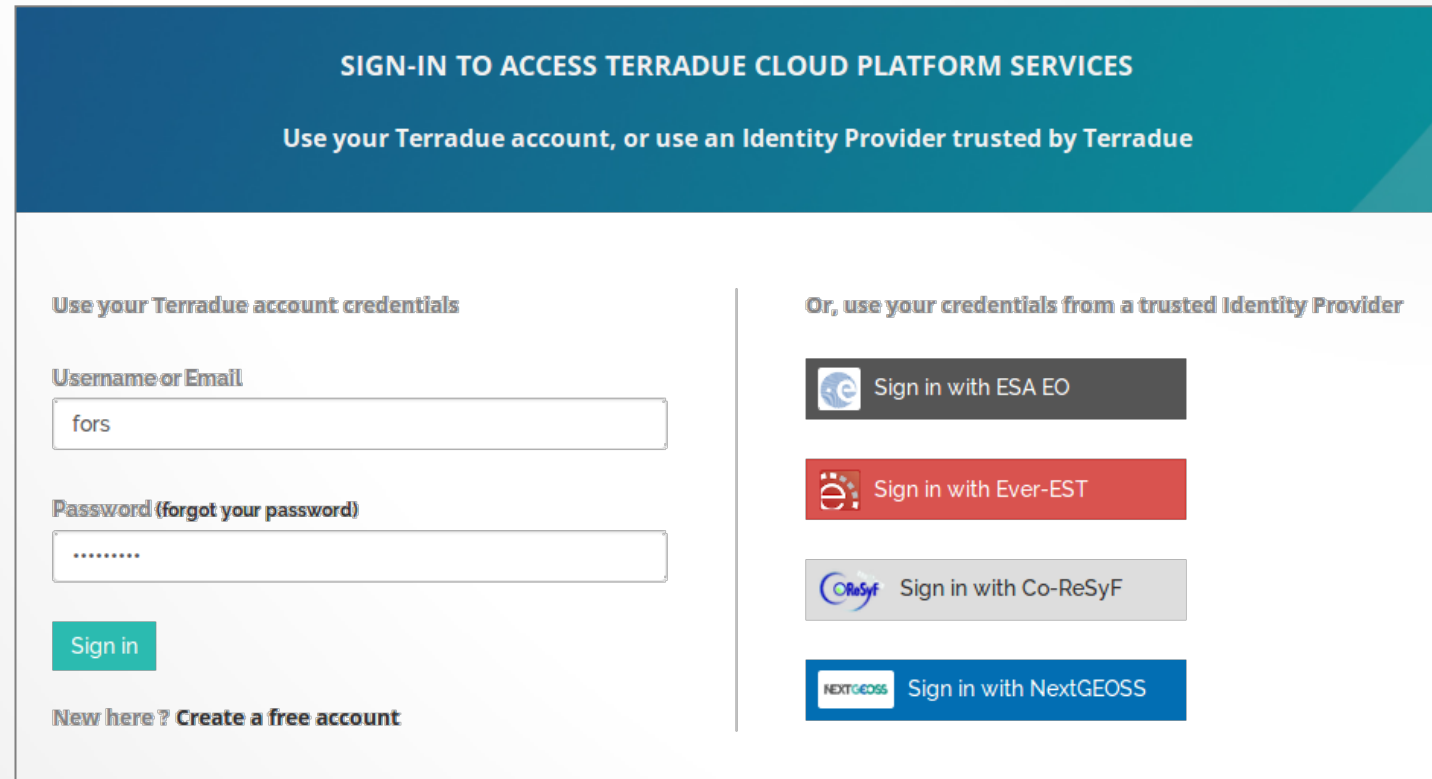
<https://docs.terradue.com/ellip>



Creating iAOS Processing Services

Become an Ellip user

1. Request iAOS access to Terradue's Ellip Notebooks solution
2. Install Google Chrome browser
3. Sign-in to the Terradue Portal



The screenshot shows the Terradue sign-in interface. At the top, a dark teal banner contains the text "SIGN-IN TO ACCESS TERRADUE CLOUD PLATFORM SERVICES" and "Use your Terradue account, or use an Identity Provider trusted by Terradue". Below this, the page is divided into two main sections. The left section, titled "Use your Terradue account credentials", contains a "Username or Email" field with the text "fors", a "Password (forgot your password)" field with masked characters ".....", a teal "Sign in" button, and a link "New here? Create a free account". The right section, titled "Or, use your credentials from a trusted Identity Provider", lists four options: "Sign in with ESA EO" (dark grey button), "Sign in with Ever-EST" (red button), "Sign in with Co-ReSyF" (light grey button), and "Sign in with NextGEOSS" (blue button).

<https://www.terradue.com>



Creating iAOS Processing Services

Configure the workshop

1. Open Google Chrome and sign-in to Terradue:

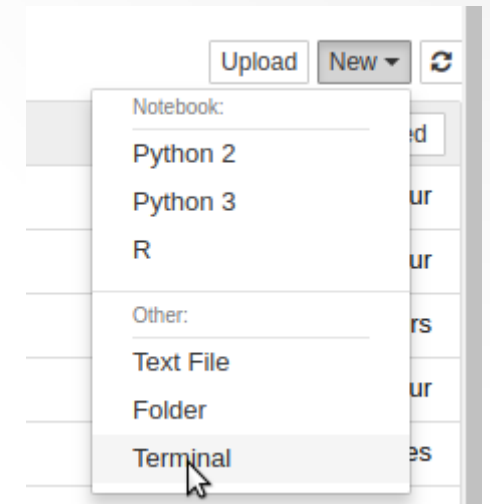
<https://www.terradue.com>

2. Browse to Jupyter Notebook and open a Terminal:

<https://notebooks-qa.terradue.com>

2. Execute the 3 following commands:

```
$ rm -f configure_workshop.sh
$ wget http://rgeostats.free.fr/doc/Files/configure_workshop.sh
$ chmod +x configure_workshop.sh
$ ./configure_workshop.sh
```

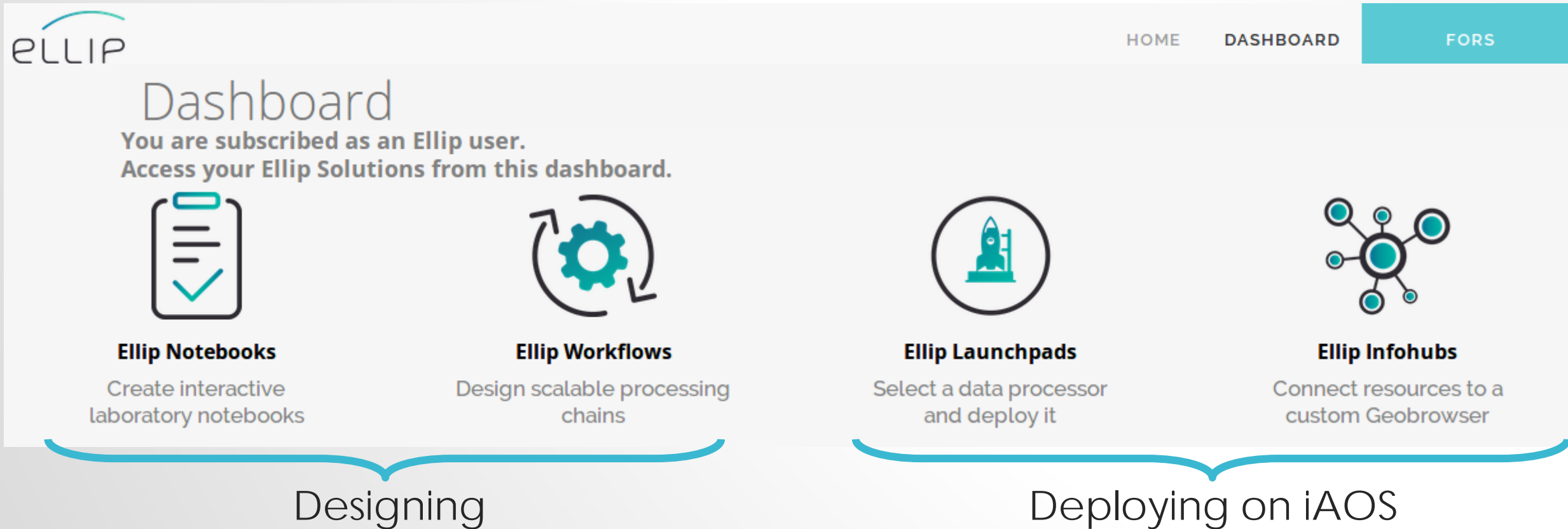


This may take several minutes...



Creating iAOS Processing Services





4 steps to create and use data processors



ELLIP HOME DASHBOARD **FORS**

Dashboard

You are subscribed as an Ellip user.
Access your Ellip Solutions from this dashboard.

- 
Ellip Notebooks
Create interactive laboratory notebooks
- 
Ellip Workflows
Design scalable processing chains
- 
Ellip Launchpads
Select a data processor and deploy it
- 
Ellip Infohubs
Connect resources to a custom Geobrowser

Designing **Deploying on iAOS**



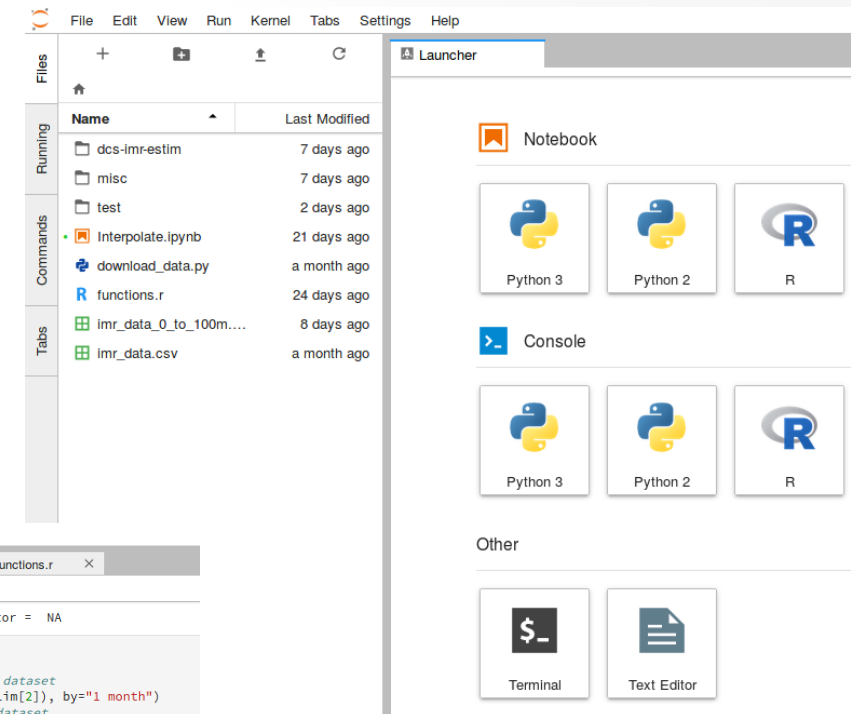
Creating iAOS Processing Services

1. Designing data processors

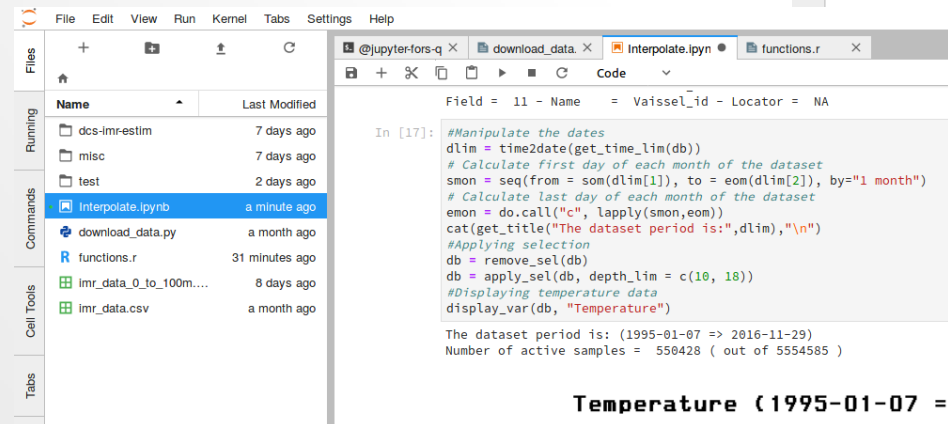


Notebooks

1. Access Jupyter Notebooks and your lab: <https://notebooks-qa.terradue.com>
2. Configure your Jupyter environment
3. Write your Notebooks (Python or R)
4. Test your Notebooks and analyze results



- Python Kernel
- R Kernel
- Terminal



Creating iAOS Processing Services

2. Building your workflow (1/2)

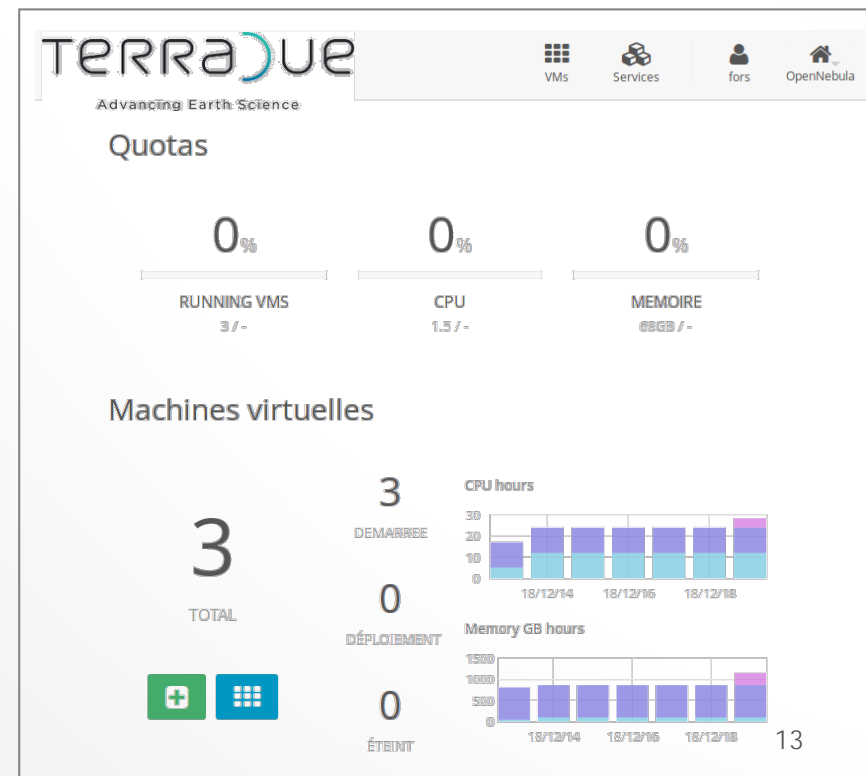
1. Access to your Cloud Dashboard <https://cloud.terradue.com/>
2. Create an **Ellip Workflows** Virtual Machine (VM)

While the VM is deploying...

3. Install SSH key pairs
4. Configure the VPN client (**openvpn**)



Workflows



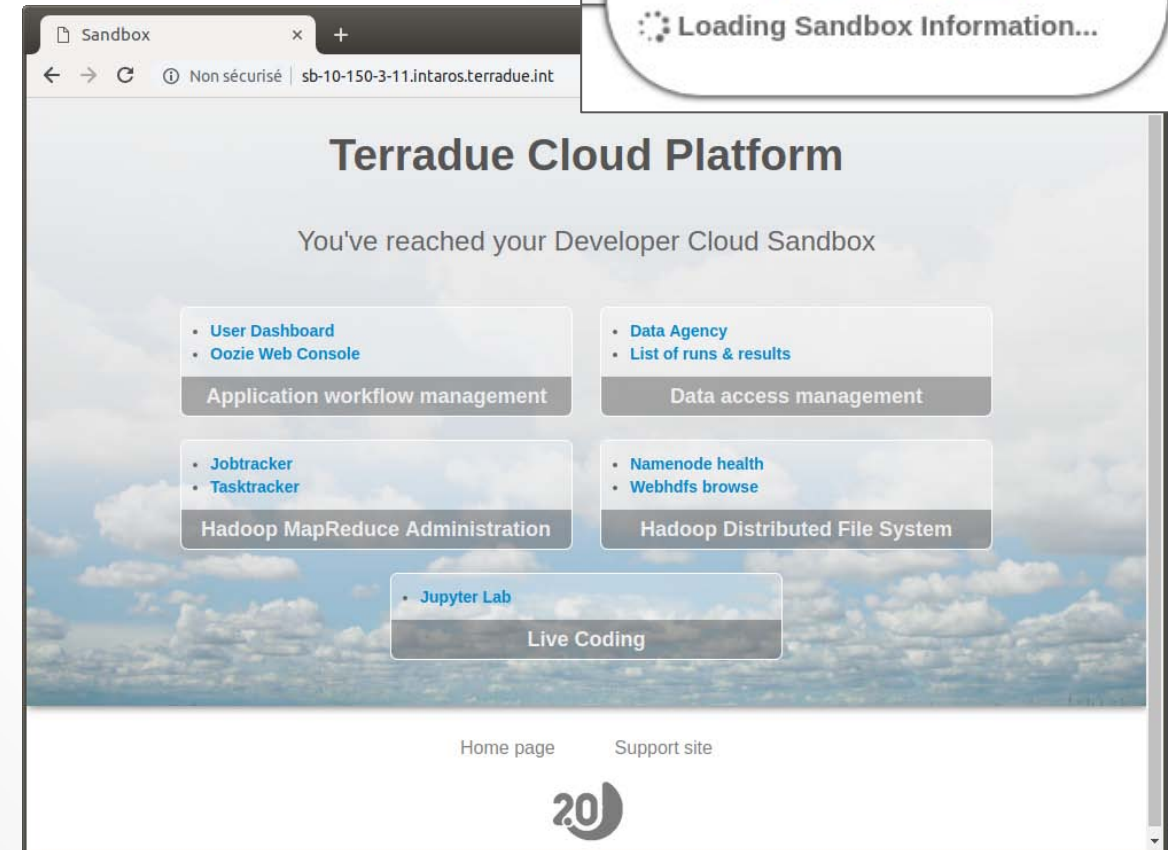
Creating iAOS Processing Services

2. Building your workflow (2/2)



Workflows

1. Connect to Terradue's **VPN**
2. Connect via **ssh** to your VM
3. Configure your VM system environment
4. Create an Ellip workflow application using **ciop**, **git** and **mvn**
5. Test your application with the **Sandbox**



Creating iAOS Processing Services

3. Deploying your workflow

1. Manage successive workflow releases (**ciop-release**)
2. Package it in a Web Processing Service (**WPS**)
3. Make it available to geo-portals



Launchpads

Select a data processor and deploy it

NextGEOSS

Data processors

- gridded-
data.armines.nextgeoss.terradue.com Cluster 2
- maxent.wur.nextgeoss.terradue.com

Available clusters

- Cluster 1 Deploy Undeploy
- Cluster 2 Deploy Undeploy



Creating iAOS Processing Services

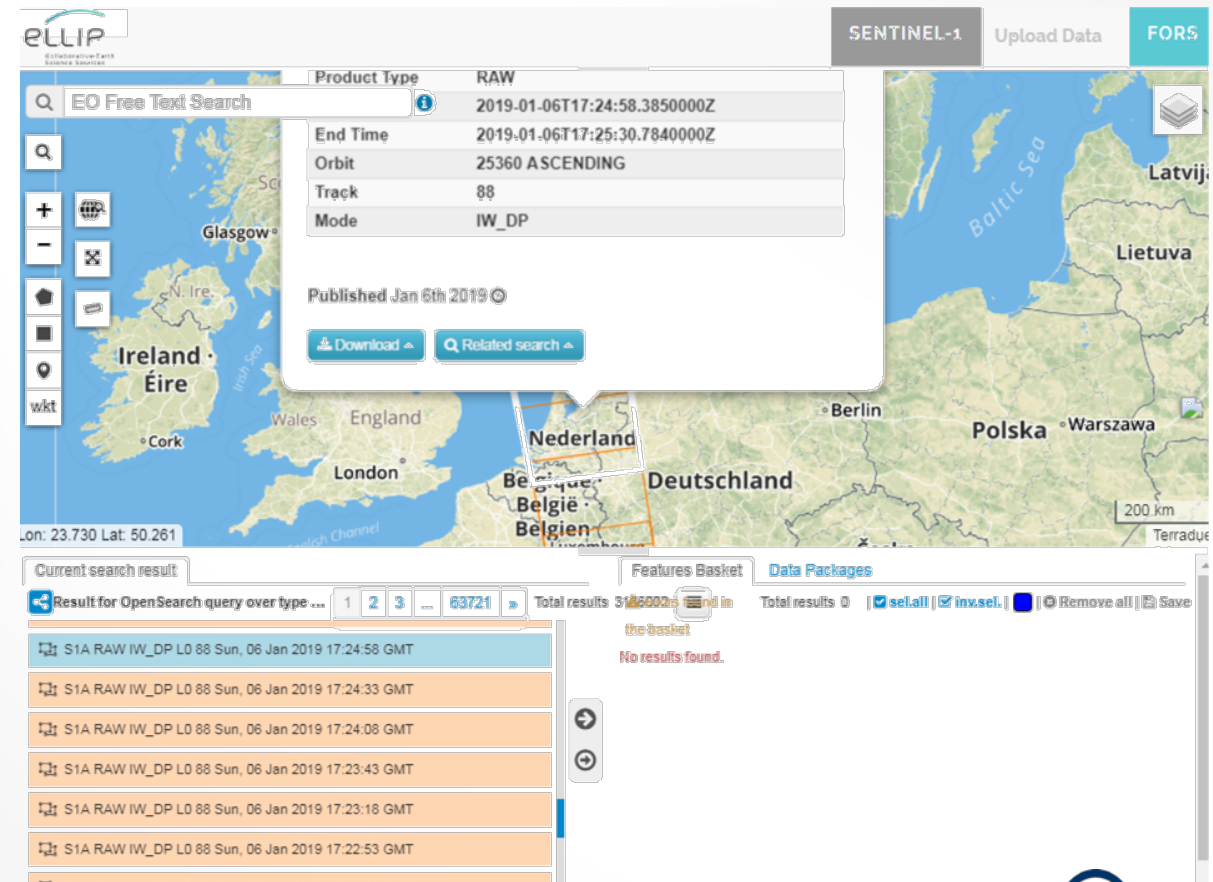
4. Running your workflow

1. Create a geo-browser application
2. Connect to your data
3. Launch your deployed Web Processing Services



Infohubs

This could be an example of interface for iAOS visitors



The screenshot shows the eLLIP web interface. At the top, there are tabs for 'SENTINEL-1', 'Upload Data', and 'FORS'. A search bar is visible with the text 'EO Free Text Search'. Below the search bar, a map of Europe is displayed with a search overlay showing details for a specific product: 'Product Type RAW', '2019-01-06T17:24:58.385000Z', 'End Time 2019-01-06T17:25:30.784000Z', 'Orbit 25360 ASCENDING', 'Track 88', 'Mode IW_DP', and 'Published Jan 6th 2019'. Below the map, there is a list of search results for 'S1A RAW IW_DP L0 88 Sun, 06 Jan 2019' with various timestamps. The interface also includes a 'Features Basket' and 'Data Packages' section.



Part #2

Geostatistics and RGeostats

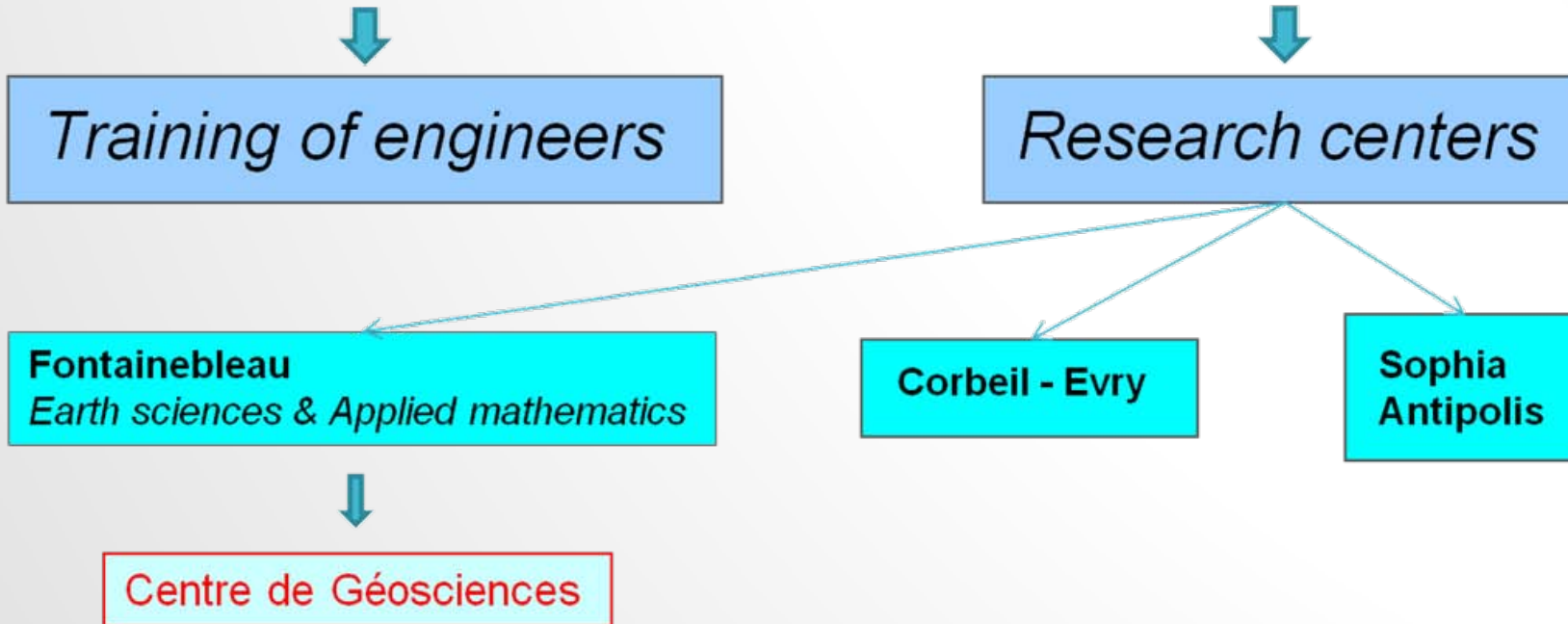


Who are we?



60 Bd St Michel in Paris

MINES ParisTech



Who are we?

Centre de Géosciences in Fontainebleau: Géostatistique



Names in Geostatistics

- D.G. Krige (mining engineer) & H.S. Sichel (statistician):
Gold mining (1950)
- L. Gandin:
Objective Analysis of Meteorological Fields (1965)
- B. Matérn:
Importance of spatial dependence and variation in forestry(1960)
- G. Matheron:
 - Formalization of the theory of regionalized variables, Ecole des Mines de Paris (1962)
 - Inception of Mathematical Morphology discipline
- Many fellows / students:
 - M. David, A. Journel, J. Davis, D. Merriam, R. Webster, N. Cressie, ...



Geostatistics

- Describe the spatial characteristics of the variable (variogram): classification, spatial correlation
- Estimation, interpolation (kriging)
- Simulations: possible alternative scenarios
- Appraisal of uncertainty
- Risk assessment

Applications

- Soil Science, Topography, Geology
- Mining, Petroleum
- Hydrology
- Biology, Epidemiology, Health, Ecology
- Environmental monitoring and assessment
- Forestry, Fishing, Agronomy
- Atmospheric Sciences
- Any discipline with spatial data



RGeostats

- Created in Centre de Géosciences of MINES ParisTech
- Initiated for Fish Industry within a European project: GEFA (2001)
- Named RGeoS and expanded possibilities:
 - Based on a separate commercial library Geoslib (written in C)
 - Mapping its functions in R language (using Rcpp layer)
- Renamed in RGeostats in 2014
- Used for testing new methodologies and/or for teaching purpose
- (Free) Download from: <http://cg.ensmp.fr/rgeostats>
- Provides demonstration scripts, FAQ, forum for posting questions, asking for help and benefiting from users experience



Trainings

More extensive courses in Geostatistics are available:

- *Les méthodes de la Géostatistique* (in French): October 2019 in Paris & Fontainebleau
- C.F.S.G. (in English) : ten months from September to July
- *Mining professional training* (during academic year in English)
- *Introduction to Geostatistics* (in English): 18-22 Feb 2019 in Paris

More information:

<http://www.geosciences.mines-paristech.fr>

Contact:

nathalie.dietrich@mines-paristech.fr



Bibliography

G. Matheron:

- 1962-1963: *Treatise of applied geostatistics* (in French), Technip and BRGM editions, Paris
- 1965: *Regionalized variables and their estimation* (in French), Masson, Paris
- 1967: *Elements for a theory of porous media* (in French), Masson, Paris
- 1968: *Treatise of applied geostatistics* (in Russian), MIR, Moscow
- 1969: *Theory of random sets* (in French), Ecole des Mines de Paris
- 1969: *Geostatistics course* (in French), Mines Paris
- 1969: *Universal kriging* (in French), Mines Paris
- 1970: *Mathematical morphology* (in French), Mines Paris
- 1970: *The theory of regionalized variables and its applications*, Mines Paris
- 1972-1975: *Random sets and integral geometry*, Wiley, New York
- 1978-1989: *Estimating and choosing*, Springer, Berlin



Bibliography

From others:

- Armstrong M., Galli A., Beucher H., Le Loc'h G., Renard D., Doligez B., Eschard R., Geffroy F., *Plurigaussian Simulations in Geosciences*, 2011. 176p.
- Chilès J.P., Delfiner P. *Geostatistics : modeling spatial uncertainty*. N.Y. : Wiley, 1999. 695p. 2nd edition 2012.
- Journel A., Huijbregts C. *Mining geostatistics*. London : Academic Press, 1978. 600p.
- Lantuéjoul C. *Geostatistical simulation : models and algorithms*. Berlin : Springer, 2001. 256p.
- Rivoirard J. *Introduction to disjunctive kriging and non-linear geostatistics*. Oxford : Clarendon Press, 1994. 181p.
- Wackernagel H. *Multivariate geostatistics: an introduction with applications*. 3rd ed. Berlin : Springer, 2003. 387p.



Part #3

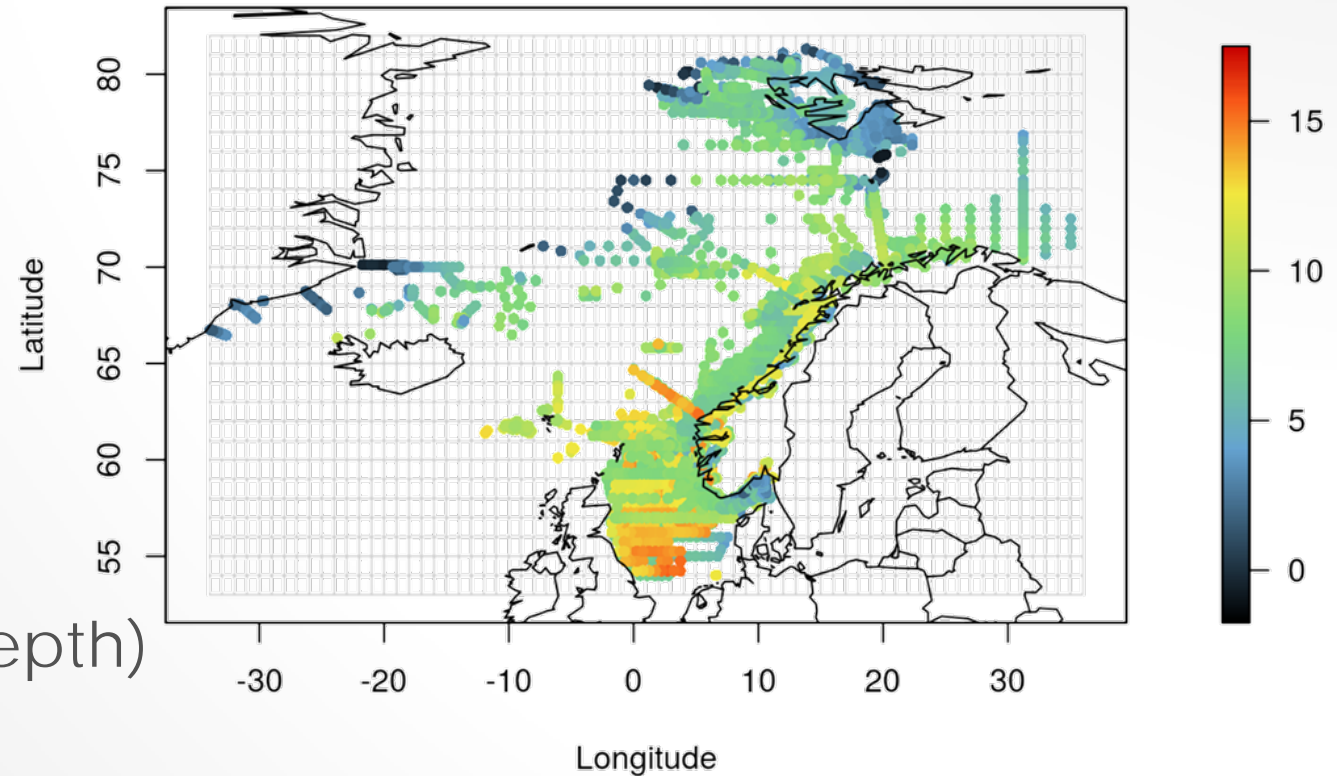
Ellip Notebooks using RGeostats



Ellip Notebooks using RGeostats

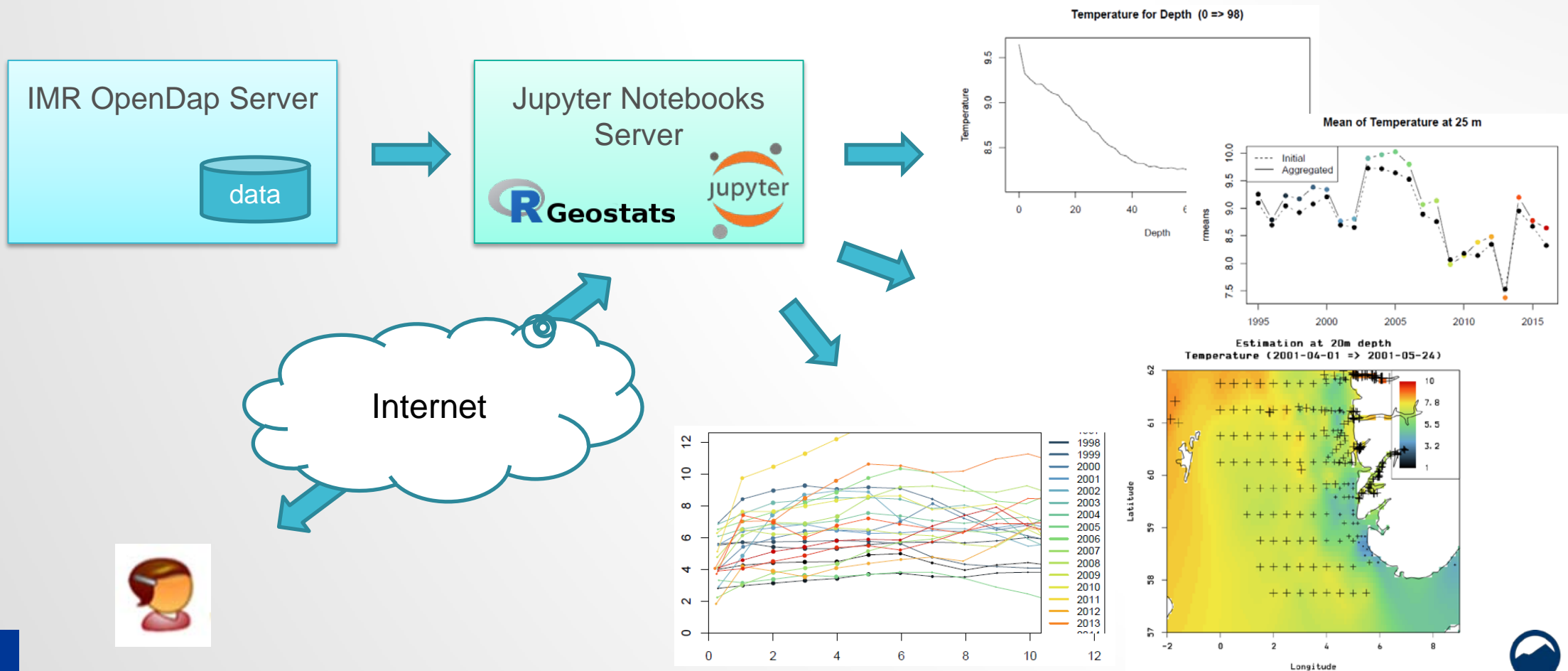
IMR Dataset global overview

- 7 vessels
- from 1995 to 2016
- 3 variables measured:
 - Temperature
 - Salinity
 - Conductivity
- 63 500 positions {long, lat}
- 63 500 vertical profiles (in depth)
- A few million samples
- 84 NetCDF files (~60 Mb each)



Ellip Notebooks using RGeostats

Work in the cloud!



Ellip Notebooks using RGeostats

Available unitary jobs for IMR dataset

One **Python** job for fetching data:

1. Subsetting by location, depth or time period from IMR OpenDAP server

Independent RGeostats **R scripts** jobs:

2. Display basemap of a variable
3. Subsetting / Filtering
4. Statistics by cells
5. Multi-variable correlation
6. 2D aggregation along time or depth
7. Variable plotting through time or depth
8. Interpolation map using kriging
9. ...

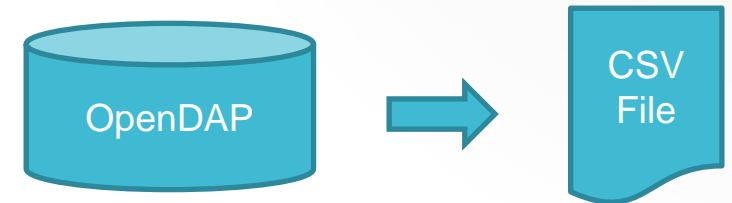


Ellip Notebooks using RGeostats

Python job for downloading data

Python script fetching data into a CSV file

IMR OpenDAP server:
Thredds server



Catalog URL:

http://opendap1-test.nodc.no/thredds/catalogs/physics/physics_point_yearly.xml

File URLs:

http://opendap1-test.nodc.no/thredds/dodsC/physics/point/yearly/58xx_CTD_yyyy.nc

Variables dumped in CSV:

Longitude, Latitude, Depth, Time, Vessel_id, Vessel_name,
Profile_id, **Temperature, Salinity, Conductivity**



Ellip Notebooks using RGeostats

Download known issues

IMR OpenDAP access directly through R:

- **Rdap** package: Not operational and not maintained
- **Pydap** module through **reticulate** package: no translation of Pydap indexing within reticulate

⇒ Current workaround: **netCDF4** python module

⇒ Fetching / subsetting data with python script

⇒ Creating CSV intermediate file

⇒ Loading this CSV file for creating the RGeostats Db

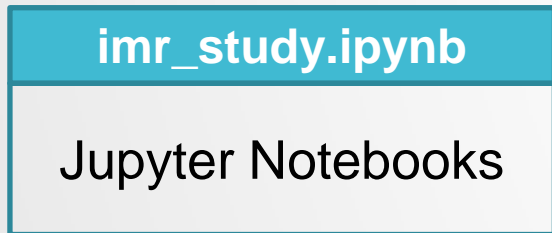
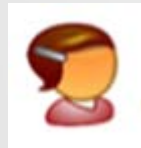
But:

- No numpy advanced indexing support
- No meta dataset covering all boats (yet)
- IMR Remote access to Hyrax OpenDAP server is slow (has to be solved)



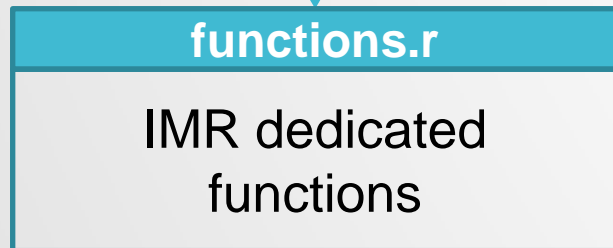
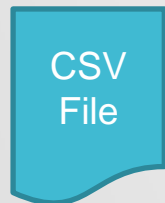
Ellip Notebooks using RGeostats

Wrapping RGeostats functions



Unitary jobs for each notebook cell:

Statistics / Aggregation / **Interpolation**



read_csv / write_csv / load_data
date2time / time2date / aggregate_depth
apply_sel / remove_sel / **interpolate_2D**
display_var / display_stats / **display_result**



db.create / db.locate / db.sel / db.plot
vario.calc / model.auto / kriging
blockstat / db.regularize



Ellip Notebooks using RGeostats

Interpolation service example : **estimate.R**

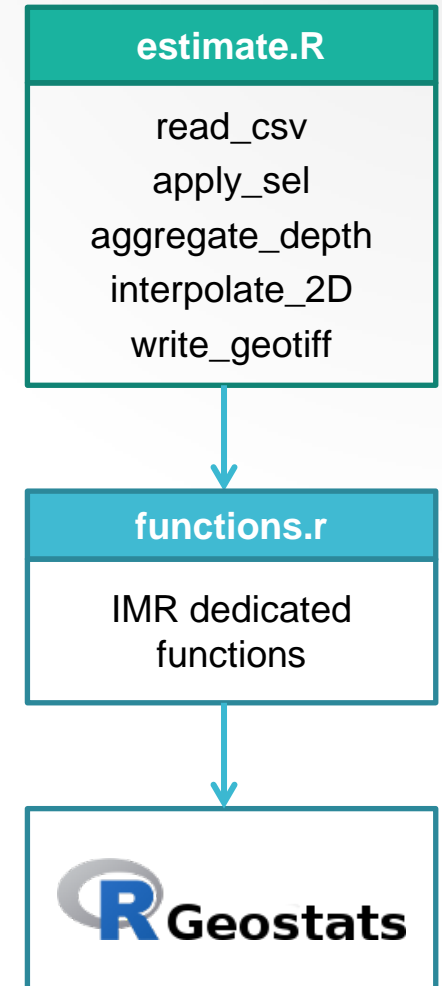
```
# Load the CSV file
dbr = read_csv(csv_file)

# Select measures in the required interval (time and aera)
dbr = apply_sel(dbr, ...)

# Aggregate all measures around 20m depth
dbr = aggregate_depth(dbr, 20, ...)

# Interpolate Temperature values
res = interpolate_2D(dbr, "Temperature", ...)

# Display results
display_result(dbr, res,...)
```



Part #4

Ellip Workflow using RGeostats



Ellip Workflow using RGeostats

Application creation

New workflow:

`mvn archetype:generate`

Type: **3** (R workflow with 2 nodes)

Group Id: **com.terradue**

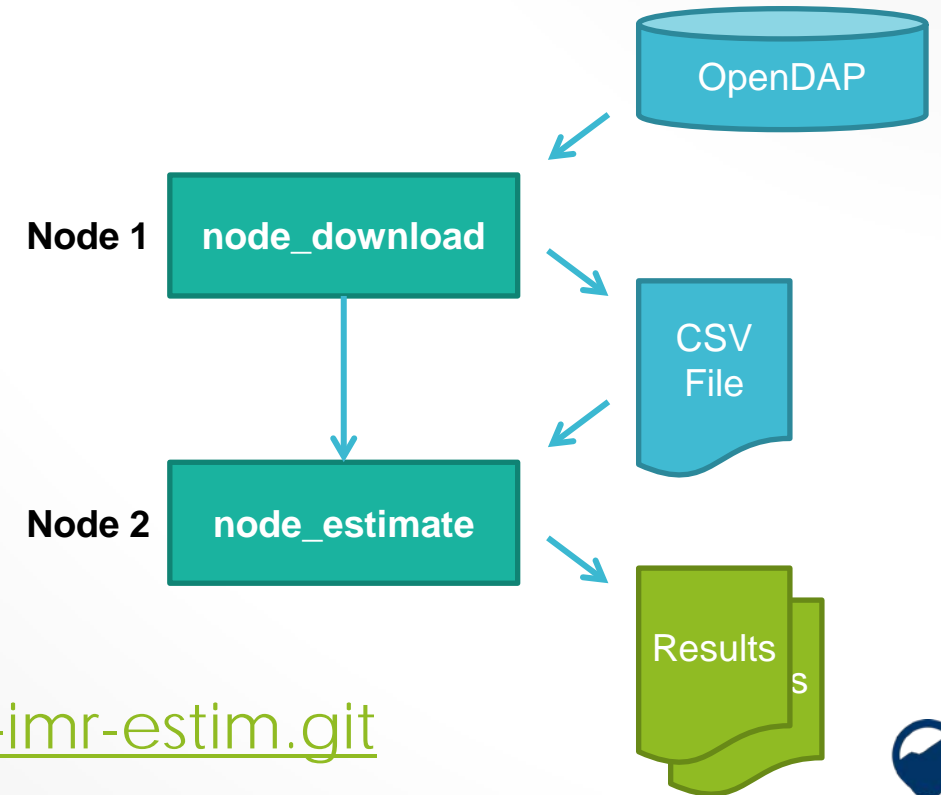
Community: **ec-intaros**

Name: **dcs-imr-estim**

Git repository:



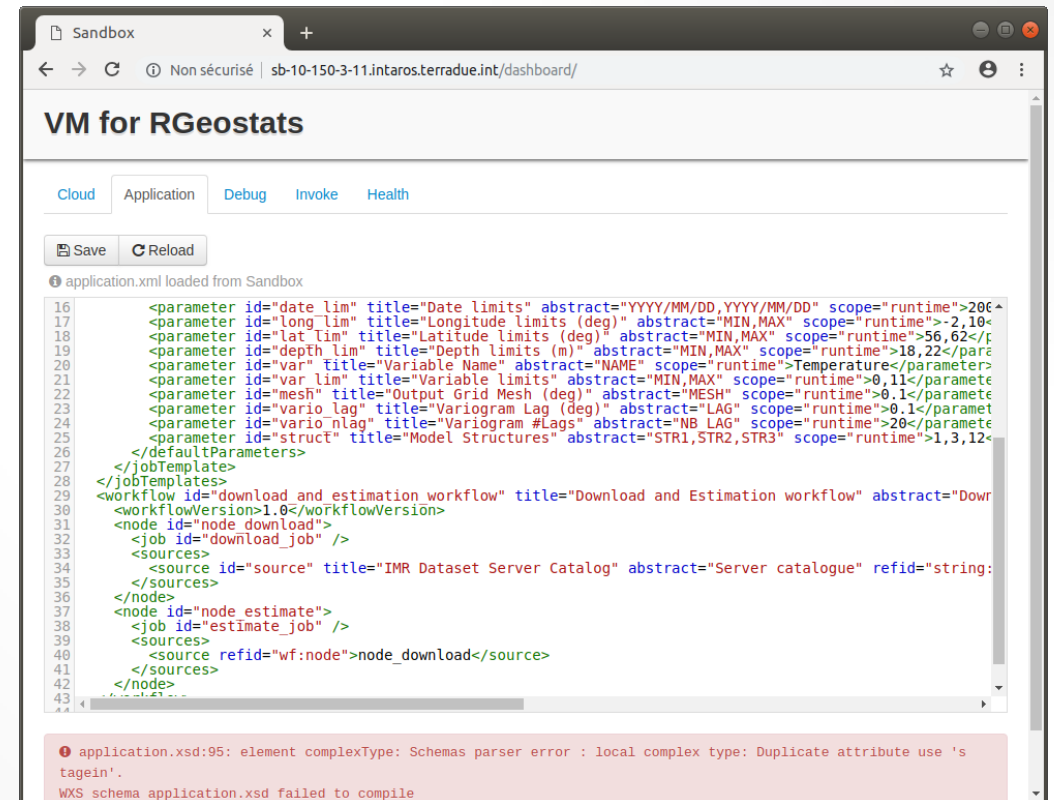
<https://gitlab.com/ec-intaros/dcs-imr-estim.git>



Ellip Workflow using RGeostats

Application.xml file

- Job Templates:
 - Streaming executable = R entry points:
node_download/run.R
node_estimate/run.R
 - Job parameters
 (long_lim: Longitude limits
 vario_nlag: Variogram lags)
- Workflow sequence:
 - Nodes identification
 - Data sources
 (catalog URL or results from previous node)



```

16 <parameter id="date_lim" title="Date limits" abstract="YYYY/MM/DD,YYYY/MM/DD" scope="runtime">20</parameter>
17 <parameter id="long_lim" title="Longitude limits (deg)" abstract="MIN,MAX" scope="runtime">-2,10</parameter>
18 <parameter id="lat_lim" title="Latitude limits (deg)" abstract="MIN,MAX" scope="runtime">56,62</parameter>
19 <parameter id="depth_lim" title="Depth limits (m)" abstract="MIN,MAX" scope="runtime">18,22</parameter>
20 <parameter id="var" title="Variable Name" abstract="NAME" scope="runtime">Temperature</parameter>
21 <parameter id="var_lim" title="Variable limits" abstract="MIN,MAX" scope="runtime">0,11</parameter>
22 <parameter id="mesh" title="Output Grid Mesh (deg)" abstract="MESH" scope="runtime">0.1</parameter>
23 <parameter id="vario_lag" title="Variogram Lag (deg)" abstract="LAG" scope="runtime">0.1</parameter>
24 <parameter id="vario_nlag" title="Variogram #Lags" abstract="NB LAG" scope="runtime">20</parameter>
25 <parameter id="struct" title="Model Structures" abstract="STR1,STR2,STR3" scope="runtime">1,3,12</parameter>
26 </defaultParameters>
27 </jobTemplate>
28 </jobTemplates>
29 <workflow id="download and estimation workflow" title="Download and Estimation workflow" abstract="Down
30 <workflowVersion>1.0</workflowVersion>
31 <node id="node_download">
32 <job id="download_job" />
33 <sources>
34 <source id="source" title="IMR Dataset Server Catalog" abstract="Server catalogue" refid="string:
35 </source>
36 </sources>
37 </node>
38 <node id="node_estimate">
39 <job id="estimate_job" />
40 <sources>
41 <source refid="wf:node">node_download</source>
42 </sources>
43 </node>
44 </workflow>

```

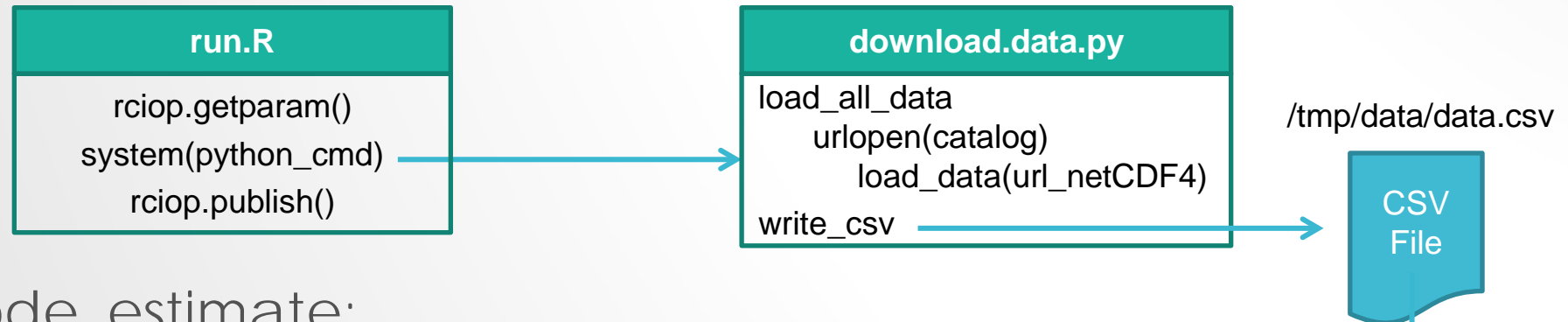
application.xsd:95: element complexType: Schemas parser error : local complex type: Duplicate attribute use 'stagein'. WXS schema application.xsd failed to compile



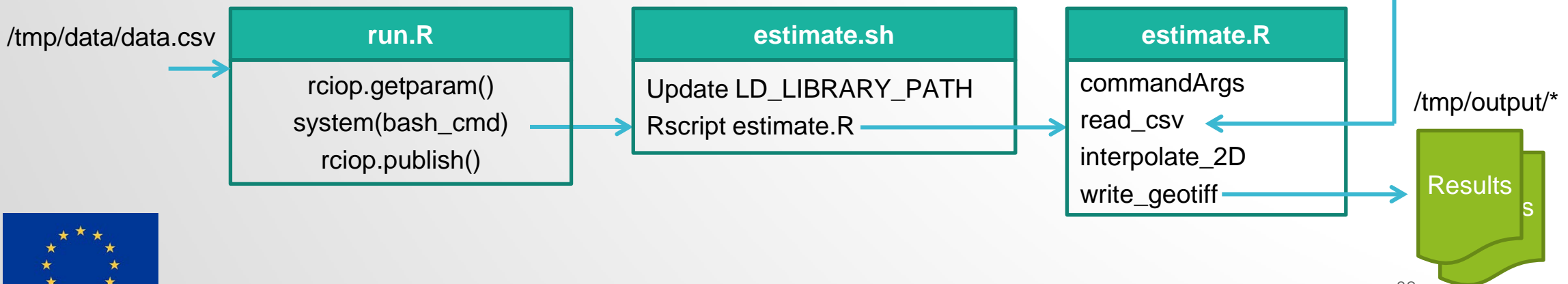
Ellip Workflow using RGeostats

Nodes description

- node_download:



- node_estimate:



Ellip Workflow using RGeostats

Run the workflow

1. Synchronise the application with your git repository (**git pull**)
2. Install the application (**mvn clean install**)
3. Launch the application (**ciop-run**)



Ellip Workflow using RGeostats

Run the workflow

```
$ ci op-run
```

```
2019-01-06 09:13:51 [INFO ] - Workflow submitted
2019-01-06 09:13:51 [INFO ] - Closing this program will not stop the job.
2019-01-06 09:13:51 [INFO ] - To kill this job type:
2019-01-06 09:13:51 [INFO ] - ci op-stop 0000065-190101000008245-oozi e-oozi -W
2019-01-06 09:13:51 [INFO ] - Tracking URL: 2019-01-06 09:13:51 [INFO ] - http://sb-10-150-3-11.intaros.terradue.int:11000/oozi e/?job=0000065-190101000008245-oozi e-oozi -W
```

```
Node Name      : node_download
Status         : OK
```

```
Node Name      : node_estimate
Status         : OK
```

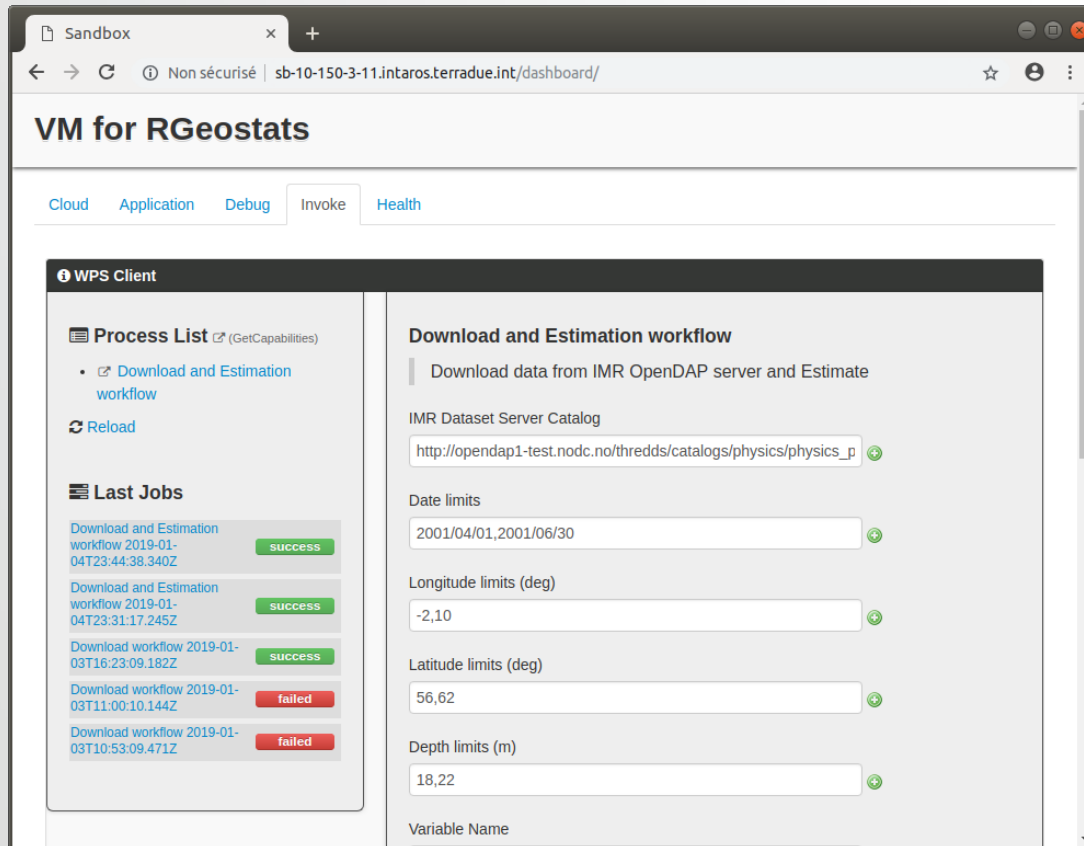
```
Publishing results...
```

```
2019-01-06 09:19:52 [INFO ] - Workflow completed.
2019-01-06 09:19:52 [INFO ] - Output Metalink: http://sb-10-150-3-11.intaros.terradue.int:50070/webhdfs/v1/ciop/run/download\_and\_estimation\_workflow/0000065-190101000008245-oozi e-oozi -W/results.metalink?op=OPEN
```



Ellip Workflow using RGeostats

Invoke the workflow (1/2)



VM for RGeostats

Cloud Application Debug **Invoke** Health

WPS Client

Process List (GetCapabilities)

- Download and Estimation workflow

Reload

Last Jobs

Download and Estimation workflow 2019-01-04T23:44:38.340Z	success
Download and Estimation workflow 2019-01-04T23:31:17.245Z	success
Download workflow 2019-01-03T16:23:09.182Z	success
Download workflow 2019-01-03T11:00:10.144Z	failed
Download workflow 2019-01-03T10:53:09.471Z	failed

Download and Estimation workflow

Download data from IMR OpenDAP server and Estimate

IMR Dataset Server Catalog

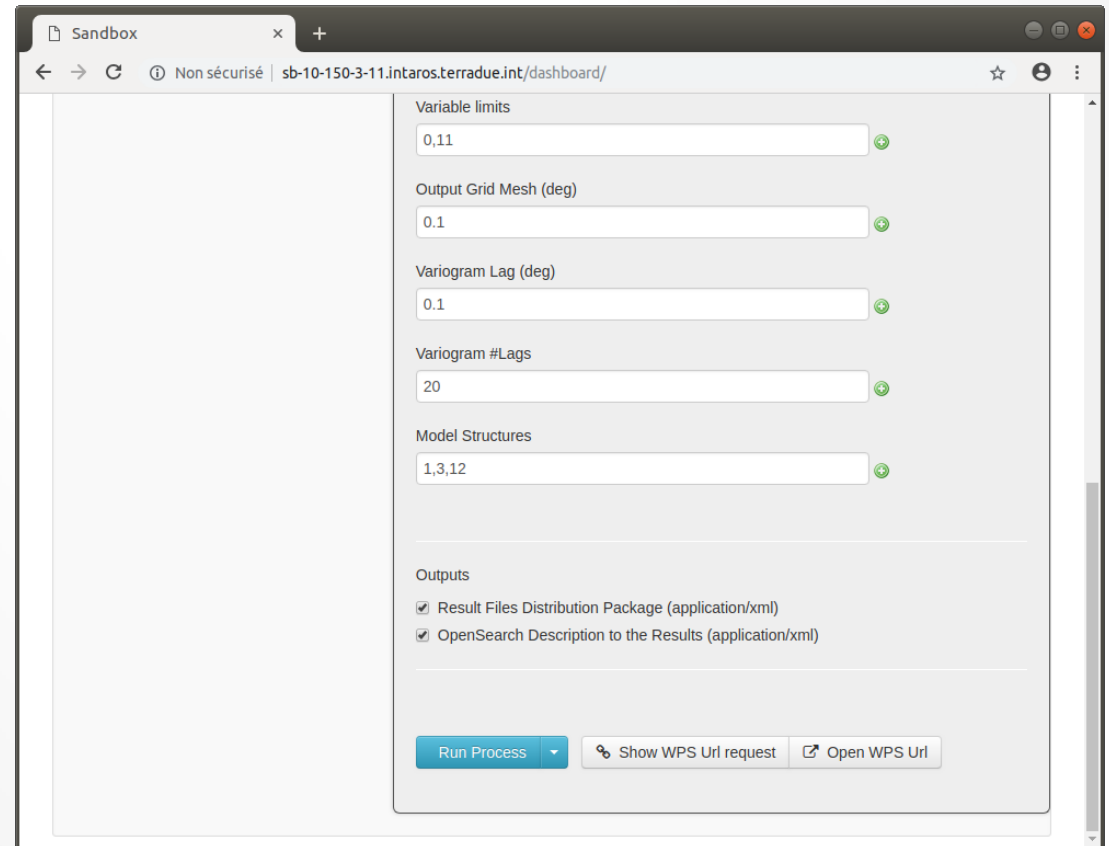
Date limits

Longitude limits (deg)

Latitude limits (deg)

Depth limits (m)

Variable Name



Variable limits

Output Grid Mesh (deg)

Variogram Lag (deg)

Variogram #Lags

Model Structures

Outputs

Result Files Distribution Package (application/xml)

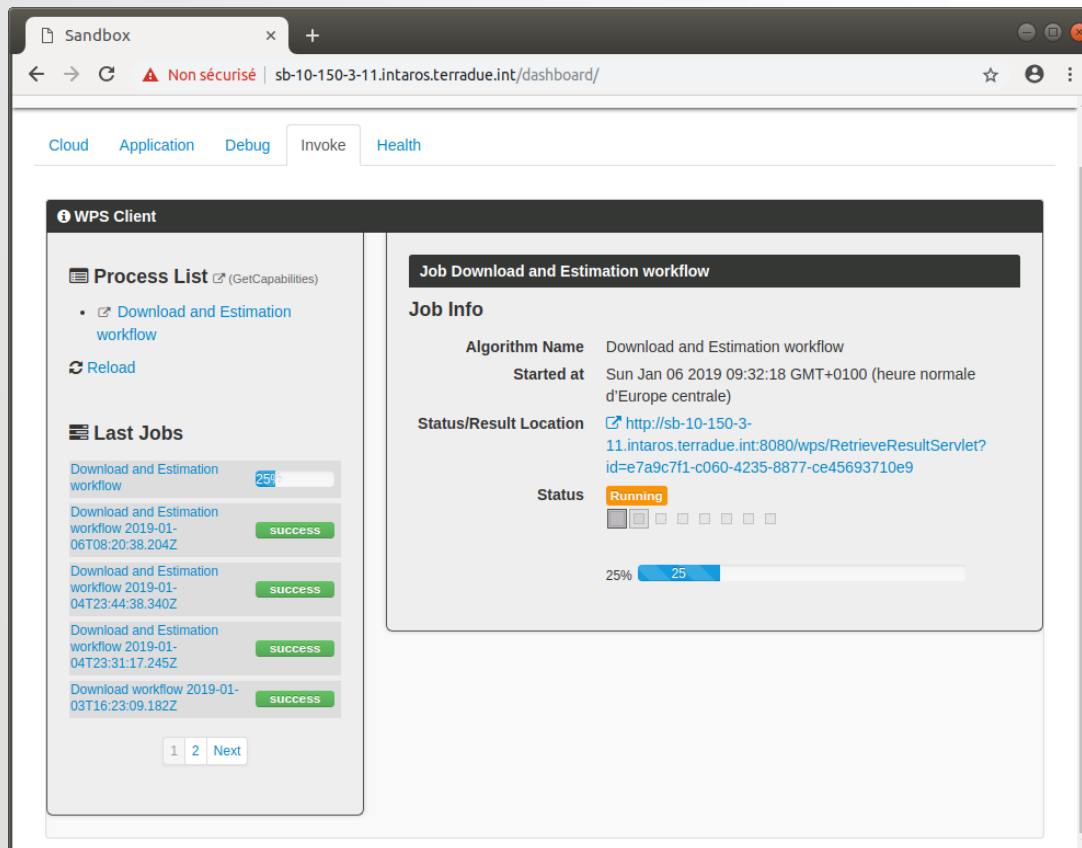
OpenSearch Description to the Results (application/xml)

Run Process Show WPS Url request Open WPS Url



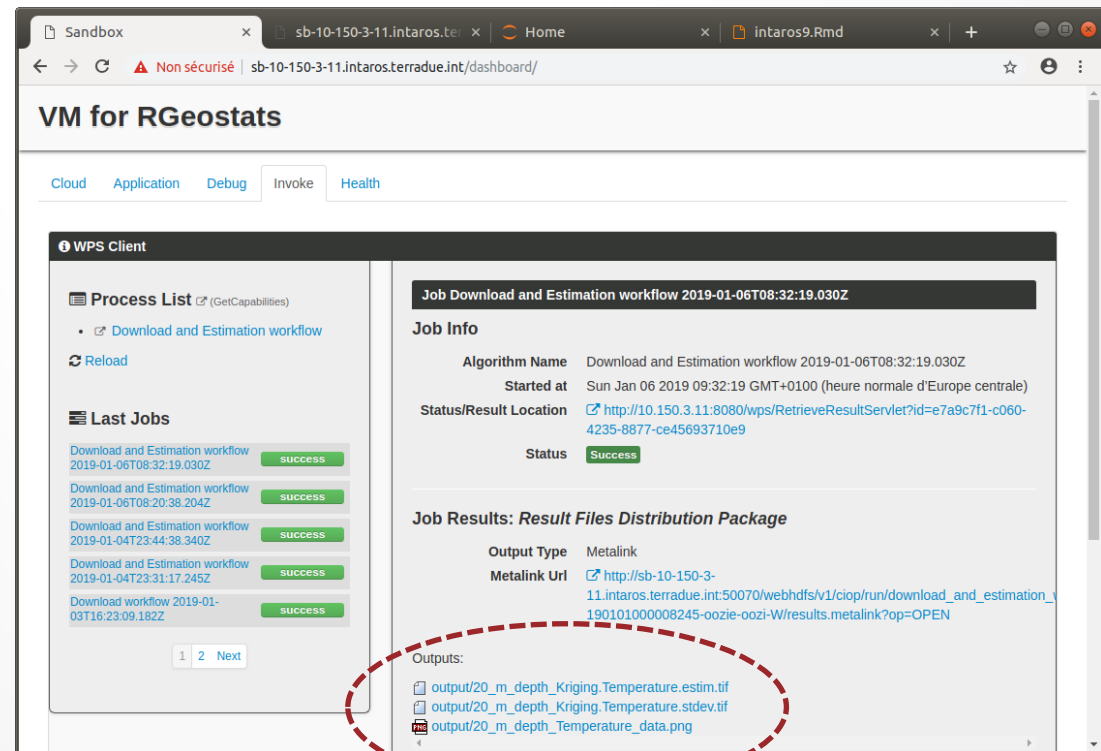
Ellip Workflow using RGeostats

Invoke the workflow (2/2)



The screenshot shows the WPS Client interface with the following details:

- Process List:** Download and Estimation workflow
- Last Jobs:** A list of five completed jobs, each with a 'success' status.
- Job Info:**
 - Algorithm Name:** Download and Estimation workflow
 - Started at:** Sun Jan 06 2019 09:32:18 GMT+0100 (heure normale d'Europe centrale)
 - Status/Result Location:** <http://sb-10-150-3-11.intaros.terradue.int:8080/wps/RetrieveResultServlet?id=e7a9c7f1-c060-4235-8877-ce45693710e9>
 - Status:** Running (indicated by an orange 'Running' label and a progress bar at 25%)



The screenshot shows the WPS Client interface with the following details:

- Process List:** Download and Estimation workflow
- Last Jobs:** A list of six completed jobs, each with a 'success' status.
- Job Info:**
 - Algorithm Name:** Download and Estimation workflow 2019-01-06T08:32:19.030Z
 - Started at:** Sun Jan 06 2019 09:32:19 GMT+0100 (heure normale d'Europe centrale)
 - Status/Result Location:** <http://10.150.3.11:8080/wps/RetrieveResultServlet?id=e7a9c7f1-c060-4235-8877-ce45693710e9>
 - Status:** Success (indicated by a green 'Success' label)
- Job Results: Result Files Distribution Package:**
 - Output Type:** Metalink
 - Metalink Url:** http://sb-10-150-3-11.intaros.terradue.int:50070/webhdfs/v1/ciop/run/download_and_estimation_19010100008245-oozie-oozi-W/results.metalink?op=OPEN
- Outputs:** A list of three output files:
 - output/20_m_depth_Kriging.Temperature.estim.tif
 - output/20_m_depth_Kriging.Temperature.stdev.tif
 - output/20_m_depth_Temperature_data.png



Ellip Workflow using RGeostats

Debug the workflow

Oozie Run

- download_and_estimation_workflow_dc48e17c-7310-42f9-a4b7-223b0a80d0aa
0000067-190101000008245-oozie-oozi-W
succeeded 32 minutes ago
- download_and_estimation_workflow_ede6d839-ec0b-4ae4-a54d-0b0d207f7a38
0000066-190101000008245-oozie-oozi-W
succeeded 44 minutes ago
- download_and_estimation_workflow_13c0c6b9-8e7f-44b0-bb54-101e12846781
0000065-190101000008245-oozie-oozi-W
succeeded an hour ago
- download_and_estimation_workflow_b0e5b12d-242d-410e-80d8-2749fec69c8e
0000064-190101000008245-oozie-oozi-W
succeeded a day ago
- download_and_estimation_workflow_0c5a08cb-7477-4cf0-a7f2-ec8c932b08a
0000063-190101000008245-oozie-oozi-W
succeeded a day ago
- download_and_estimation_workflow_1feb42f-b2f3-493a-90ca-894ae3065d90

Run Information

```

graph TD
    A[fs hdfscommands] -- ok --> B[map-reduce n-node_download]
    B -- ok --> C[map-reduce n-node_estimate]
    C -- ok --> D[map-reduce publish-results]
  
```

Hadoop job_201812122359_0264 on sb-10-150-3-11

User: tonicat
Job Name: oozie:action:map-reduce:W-download_and_estimation_workflow_412159b2-6d1e-4f7d-bc66-11e4660c523-An-node_estimate-0000068-190101000008245-oozie-oozi-W
Job File: hdfs://sb-10-150-3-11.intaros.terradue.int:8020/vault/hadoop-0.20/cachemmapredmapredlogstomcatr-staggsppb_201812122359_0264/job.xml
Submit Host: sb-10-150-3-11.intaros.terradue.int
Submit Host Address: 10.150.3.11
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Succeeded
Started At: Sun Jan 06 12:23:06 CET 2019
Finished At: Sun Jan 06 12:23:35 CET 2019
Finished In: 29sec
Job Cleanup: Successful

Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	100.00%	2	0	0	2	0	0/0
reduce	100.00%	1	0	0	1	0	0/0

	Counter	Map	Reduce	Total
	SLOTS_MILLIS_MAPS	0	0	7,067
	Launched reduce tasks	0	0	1
	Total time spent by all reduces waiting after reserving slots (ms)	0	0	0

stderr logs

```

Loading required package: XML
Loading required package: methods
Warning message:
package 'rciop' was built under R version 3.4.1
2019-01-06T12:23:25 [INFO][user process] Parameters values:
2019-01-06T12:23:25 [INFO][user process] - date_param = 2001/04/01,2001/06/30
2019-01-06T12:23:25 [INFO][user process] - long_param = -2,10
2019-01-06T12:23:25 [INFO][user process] - lat_param = 56,62
2019-01-06T12:23:25 [INFO][user process] - depth_param = 18,22
2019-01-06T12:23:25 [INFO][user process] - var_param = Temperature
2019-01-06T12:23:25 [INFO][user process] - varlim_param = 0,11
2019-01-06T12:23:25 [INFO][user process] - mesh_param = 0,1
2019-01-06T12:23:25 [INFO][user process] - vario_lag_param = 0,1
2019-01-06T12:23:25 [INFO][user process] - vario_nlag_param = 20
2019-01-06T12:23:25 [INFO][user process] - struct_param = 1,3,12
2019-01-06T12:23:25 [INFO][user process] Processing input: /tmp/data/data.csv
2019-01-06T12:23:25 [INFO][user process] Bash command: /application/util/R/estimate.sh /tmp/data/data.csv 2001/04/01,2001/06/30 -2,10 56,62 18,22 Temperature 0,11 0,1 0,1 20 1,3,12
During startup - warning messages:
1: Setting LC_CTYPE failed, using "C"
2: Setting LC_COLLATE failed, using "C"
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
5: Setting LC_MONETARY failed, using "C"
6: Setting LC_PAPER failed, using "C"
7: Setting LC_MEASUREMENT failed, using "C"
NOTE: rgdal::checkCRSArgs: no proj_defs.dat in PROJ.4 shared files
NOTE: rgdal::checkCRSArgs: no proj_defs.dat in PROJ.4 shared files
NOTE: rgdal::checkCRSArgs: no proj_defs.dat in PROJ.4 shared files
NOTE: rgdal::checkCRSArgs: no proj_defs.dat in PROJ.4 shared files
NOTE: rgdal::checkCRSArgs: no proj_defs.dat in PROJ.4 shared files
NOTE: rgdal::checkCRSArgs: no proj_defs.dat in PROJ.4 shared files
NOTE: rgdal::checkCRSArgs: no proj_defs.dat in PROJ.4 shared files
2019-01-06T12:23:29 [INFO][user process] Bash command finished
CONTAINER ID unset
2019-01-06T12:23:29 [INFO] [ciop-publish] Starting publish of results /tmp/output
2019-01-06T12:23:31 [INFO] [ciop-publish] DFS copyFromLocal /tmp/output /ciop/run/download_and_estimation_workflow/0000068-190101000008245-oozie-oozi-W/_results
mkdir: cannot create directory -p: file exists
mkdir: cannot create directory /ciop/run/download_and_estimation_workflow/0000068-190101000008245-oozie-oozi-W/_results: File exists
2019-01-06T12:23:32 [INFO] [ciop-publish] Publish of results done. List below:
  
```



Ellip Workflow using RGeostats

Deploy the workflow

```
$ ci op-release

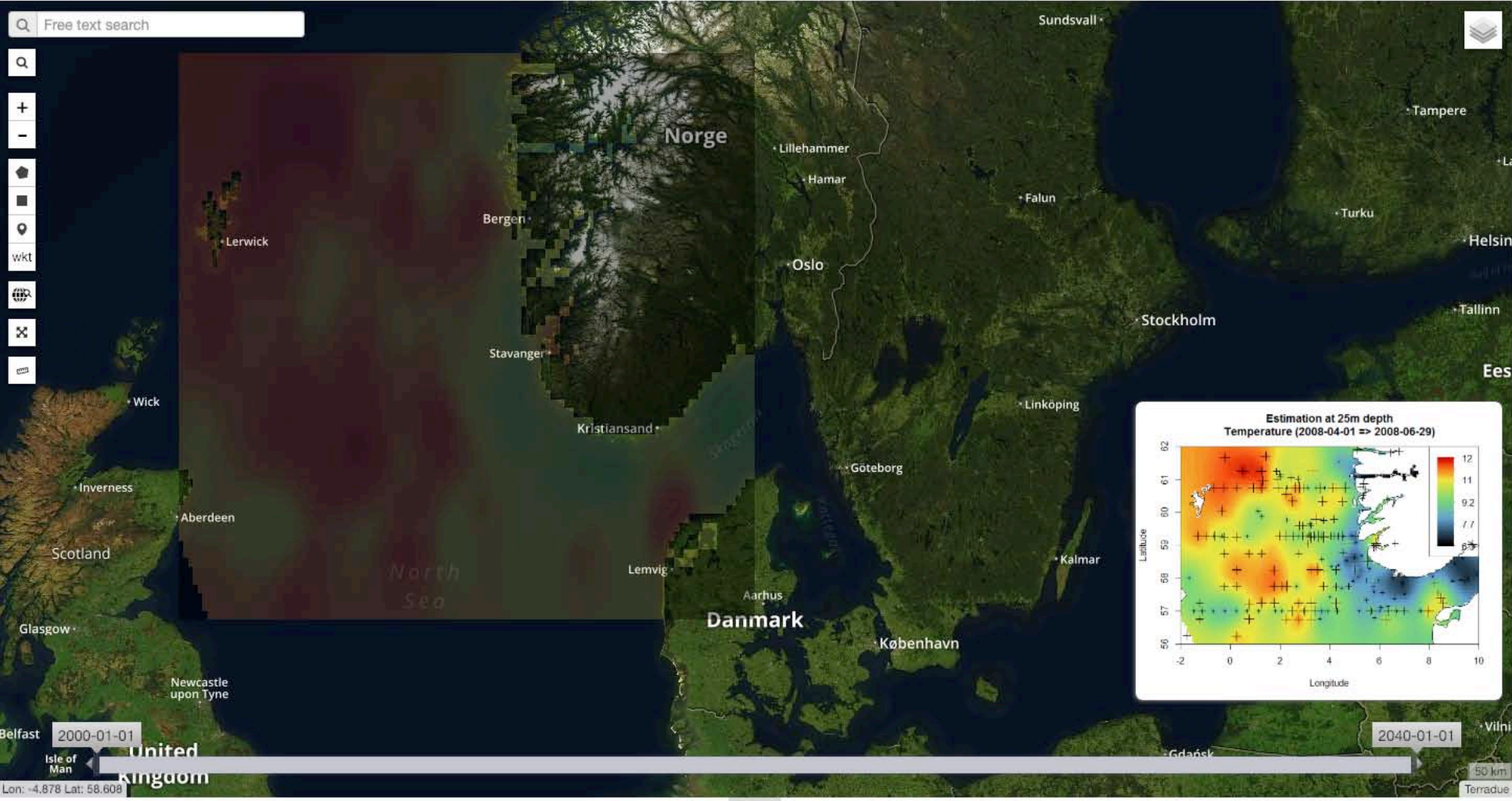
***
*** Application release for 'dcs-imr-estim'
***

...

Writing objects: 100% (1/1), 159 bytes, done.
Total 1 (delta 0), reused 0 (delta 0)
To https://gitlab.com/ec-intaros/dcs-imr-estim.git
 * [new tag]          1.0 -> 1.0

***
-> Stage 1. Checking your repository 'dcs-imr-estim' DONE
-> Stage 2. Setting the release and development versions DONE
-> Stage 3. Releasing 'dcs-imr-estim' DONE
-> Stage 4. Synchronising the remote repository DONE
```





Processing service Download and Estimation workflow
Started at Jan 9th 2019 10:17
Finished at Jan 9th 2019 10:24
Created by crossi
Status/Result Location [↗](#)
Status Success
Visibility restricted
Share [🔗](#)

Parameters

Name	Value
source	http://opendap1-test.nodc.no/thredds/catalogs/physics/physics_point_yearly.xml
date_lim	2001/04/01,2001/06/30
long_lim	-2,10
lat_lim	56,62
depth_lim	18,22
var	Temperature
var_lim	0,11
mesh	0.1
vario_lag	0.1
vario_nlag	20
struct	1,3,12

[🔄 Resubmit Job](#)

Success
 The job was completed successfully.

Current search result

[Discovery feed for local data](#) Total results 2

- .20_m_depth_Kriging_Temperature_estim.tif
- .20_m_depth_Kriging_Temperature_stddev.tif

Features Basket Data Packages

Total results 0 [sel.all](#) [inv.sel](#) [Remove all](#) [Save](#)

No results found.

Results

Found layers in the result. [📍 Show results](#)

[XML Result](#)

[Support](#)

Ellip Workflow using RGeostats

Useful commands (1/3)

Connect to your VM

```
sudo openvpn --config client.ovpn  
ssh 10.150.3.11
```

- => Connect to VPN
- => Connect to your VM with no password (ssh key pairs)

Creating and installing application

```
mvn archetype:generate  
mvn clean install
```

- => Create a new workflow
- => Install the application



Ellip Workflow using RGeostats

Useful commands (2/3)

Local git repository

```
git init
git config credential.helper store
git add -A .
git commit -m <msg>
git branch <branch>
git checkout <branch>
```

- => Initialize a local repository
- => Store credentials (no more password)
- => Add content to a local repository
- => Commit changes to the local repo
- => Create a branch named <branch>
- => Work on the branch named <branch>

Remote git repository interaction

```
git remote add origin <url>
git push
git pull
git clone
```

- => Synchronize with remote repository
- => Push local changes to server
- => Update local repo from the server
- => Create a local repo from the server



Ellip Workflow using RGeostats

Useful commands (3/3)

Ciop tool

ciop-getparam
ciop-log
ciop-copy
ciop-publish
ciop-run
ciop-release

- => Retrieve job's parameter value (user input)
- => Log something to job log files
- => Copy online source file to local file system
- => Publish job productions as output results
- => Execute a job or a complete workflow
- => Create and deploy a new application release



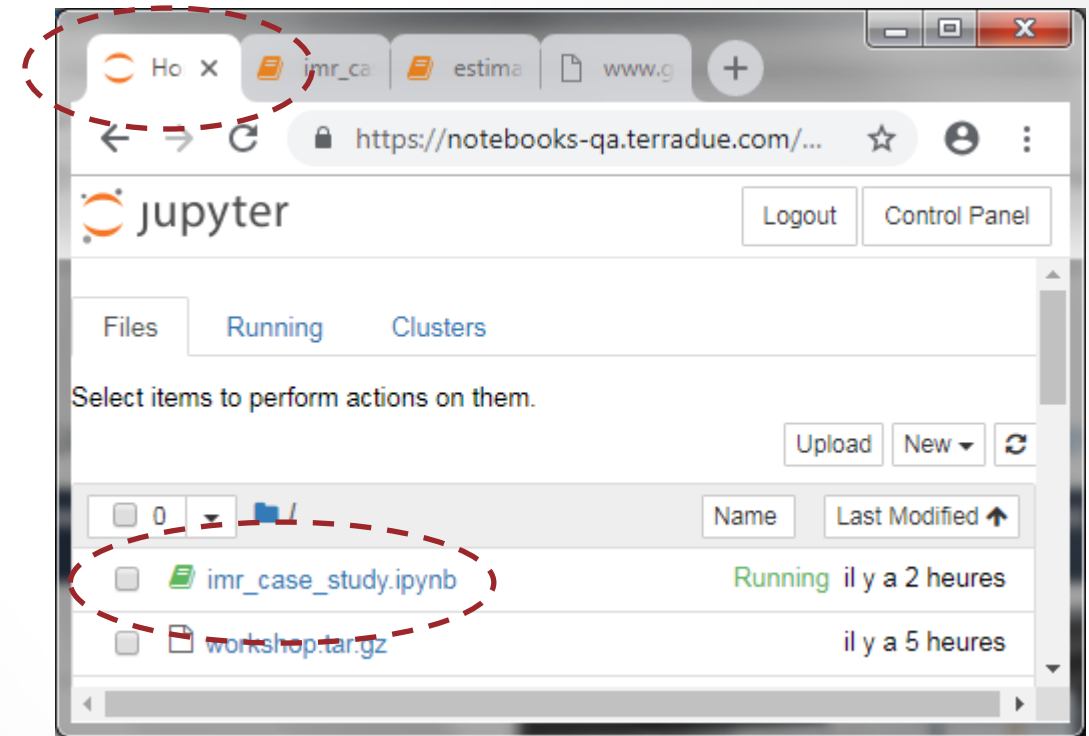
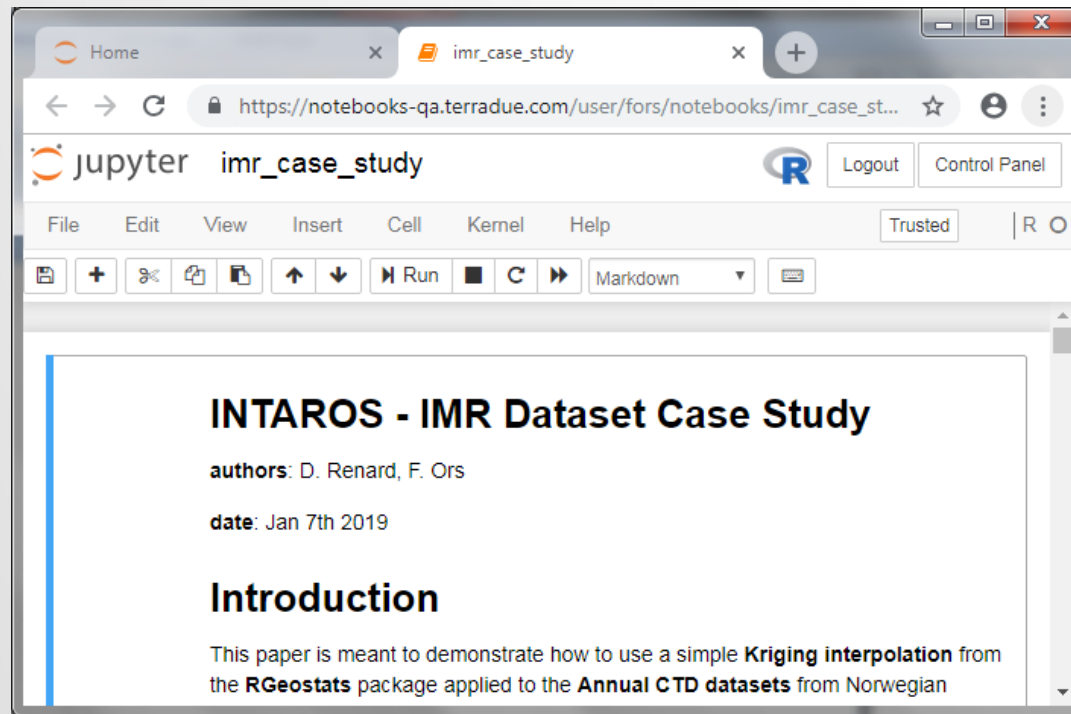
Part #5

IMR Case Study – RGeostats in Action!



Getting Ready for Action!

- Click on the **Home** tab of your Jupyter Notebooks (in Chrome)
- Click on the **imr_case_study.ipynb**



Note: You may prefer the 'lab view': replace **tree** by **lab** in URL



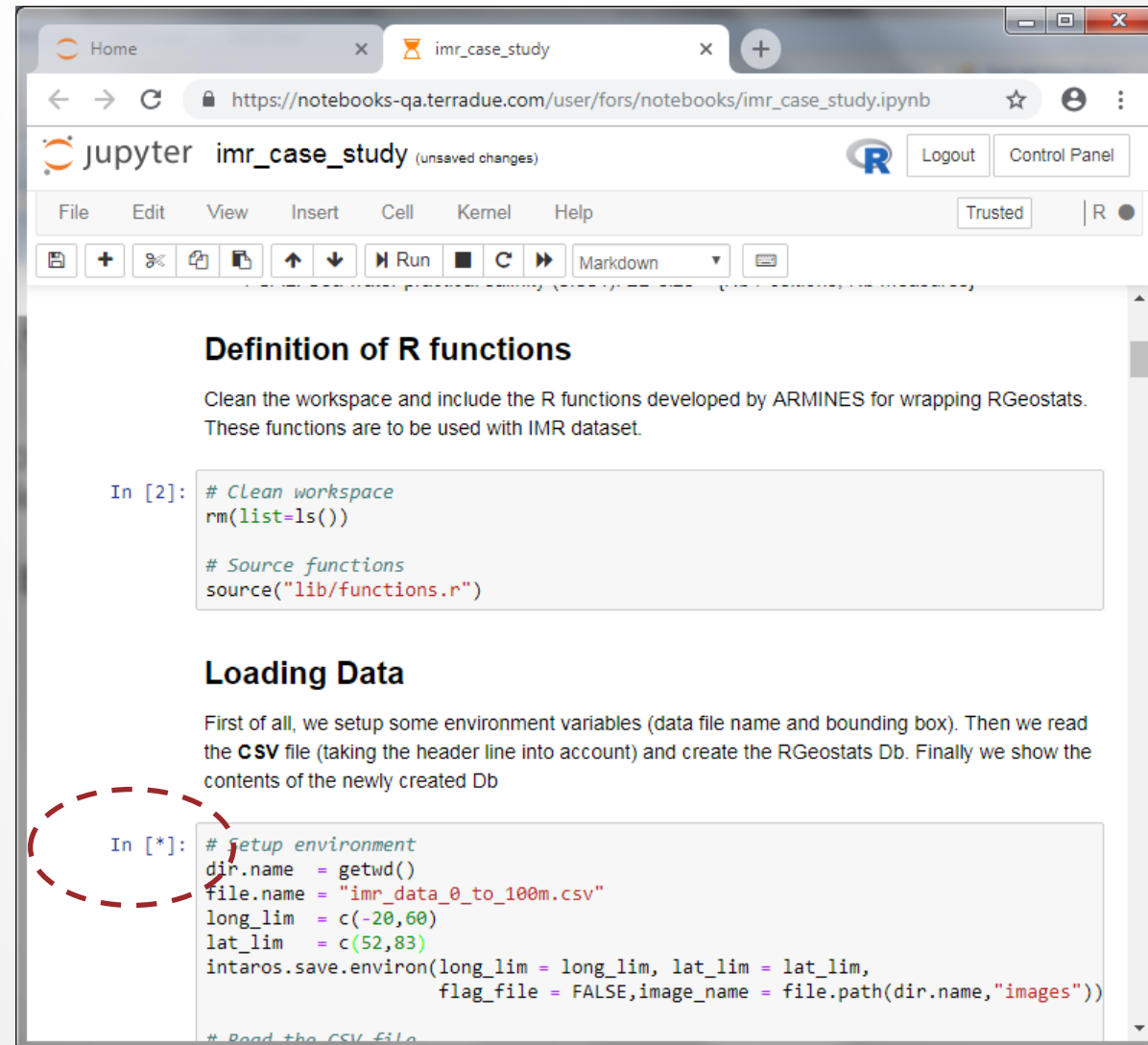
Getting Ready for Action!

- Execute all code cells in the **Introduction** section:

=> Hitting **Shift + Enter**

- Definition of R functions
- Loading Data (slow)
- Dataset global statistics (2 cells)

Note: The symbol **In [*]** indicates that a cell is running



```

In [2]: # Clean workspace
rm(list=ls())

# Source functions
source("lib/functions.r")

Definition of R functions

Clean the workspace and include the R functions developed by ARMINES for wrapping RGeostats.
These functions are to be used with IMR dataset.

Loading Data

First of all, we setup some environment variables (data file name and bounding box). Then we read
the CSV file (taking the header line into account) and create the RGeostats Db. Finally we show the
contents of the newly created Db

In [*]: # Setup environment
dir.name = getwd()
file.name = "imr_data_0_to_100m.csv"
long_lim = c(-20,60)
lat_lim = c(52,83)
intaros.save.environ(long_lim = long_lim, lat_lim = lat_lim,
                    flag_file = FALSE,image_name = file.path(dir.name,"images"))

# Read the CSV file

```

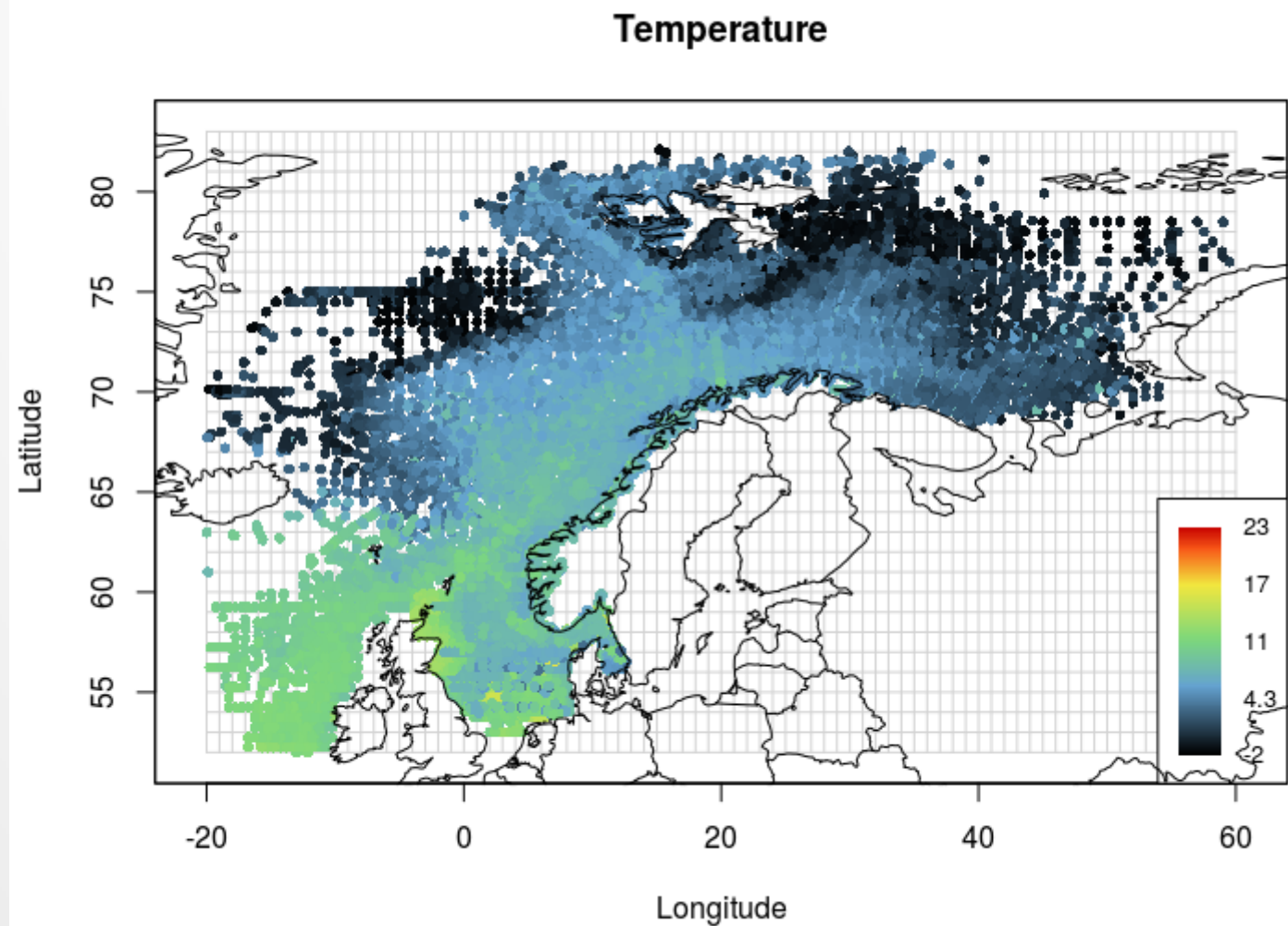


Studying Temperature variable



All Database

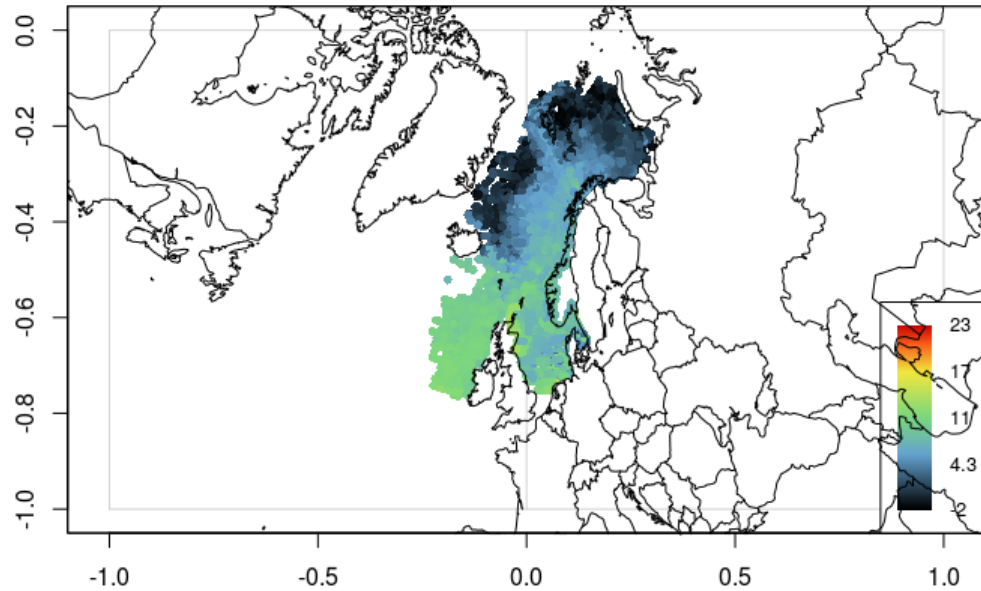
- Display Temperature Variable
- 5 million samples displayed



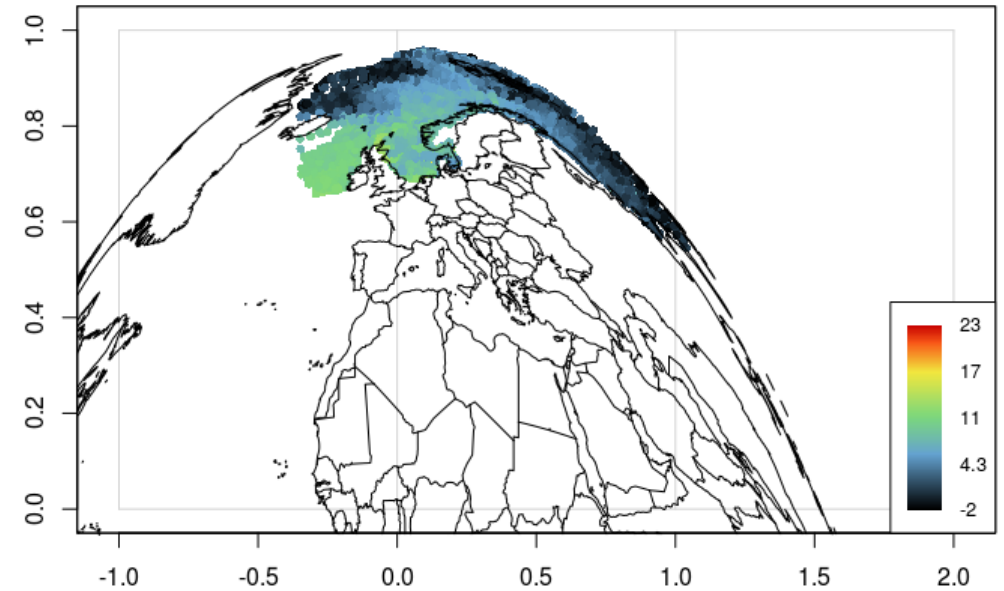
Different Projections

- Define New Projections (Gnomonic and Mecca)

Temperature (Projection Gnomonic)

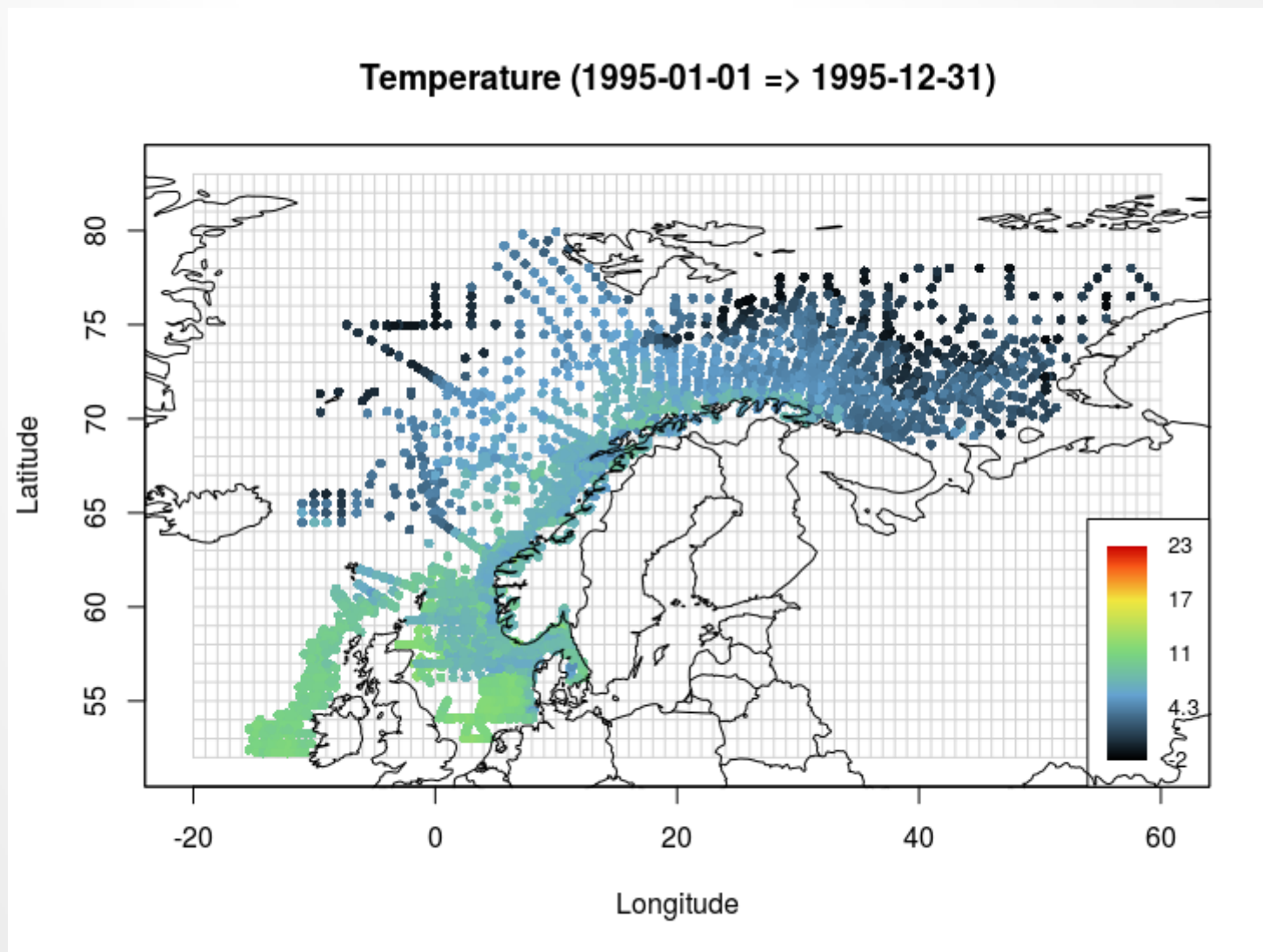


Temperature (Projection Mecca(10))



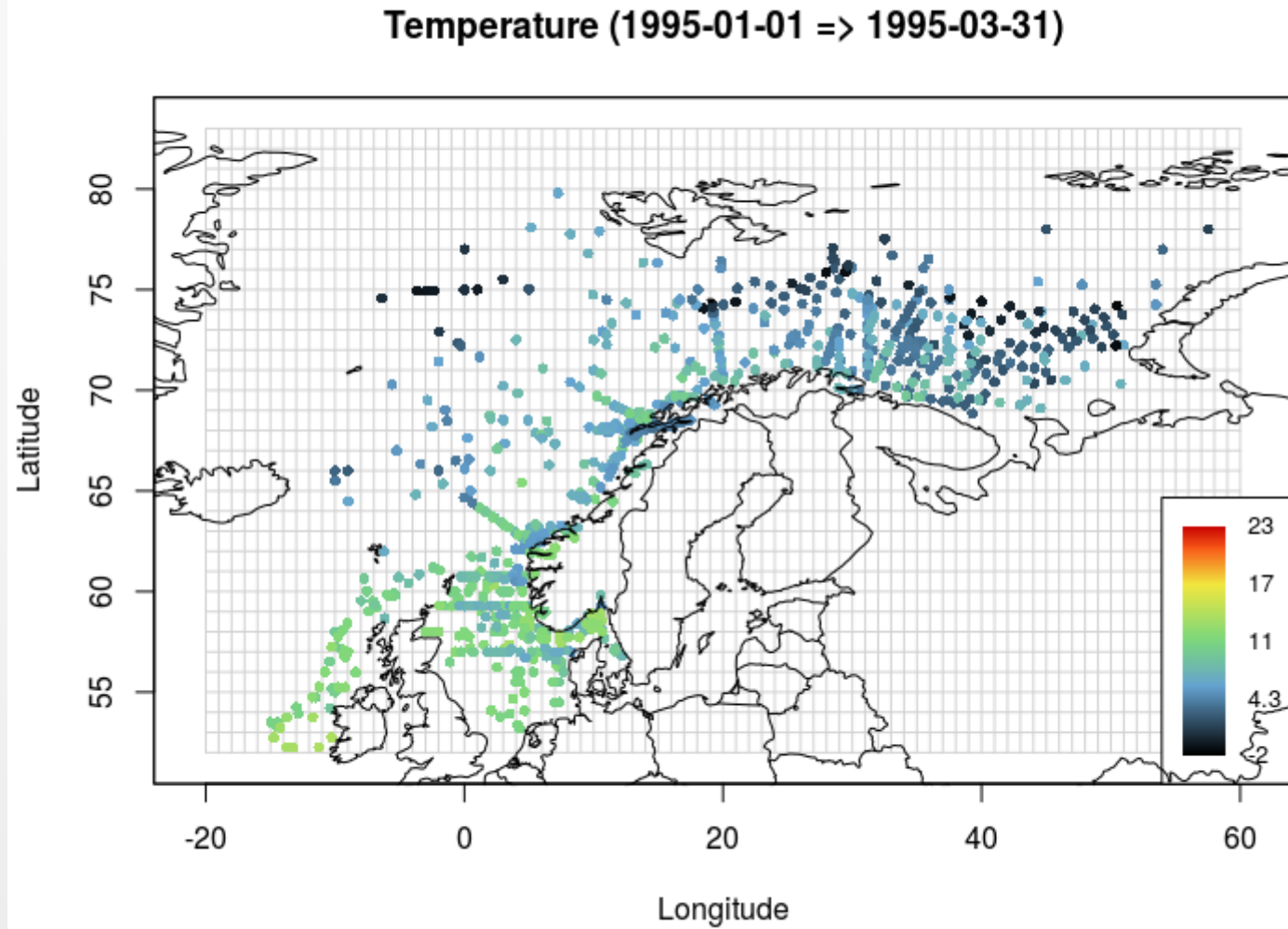
Year Campaign

- Define Target Year (1995)
- Display Variable



Year/Trimester at 20m Depth

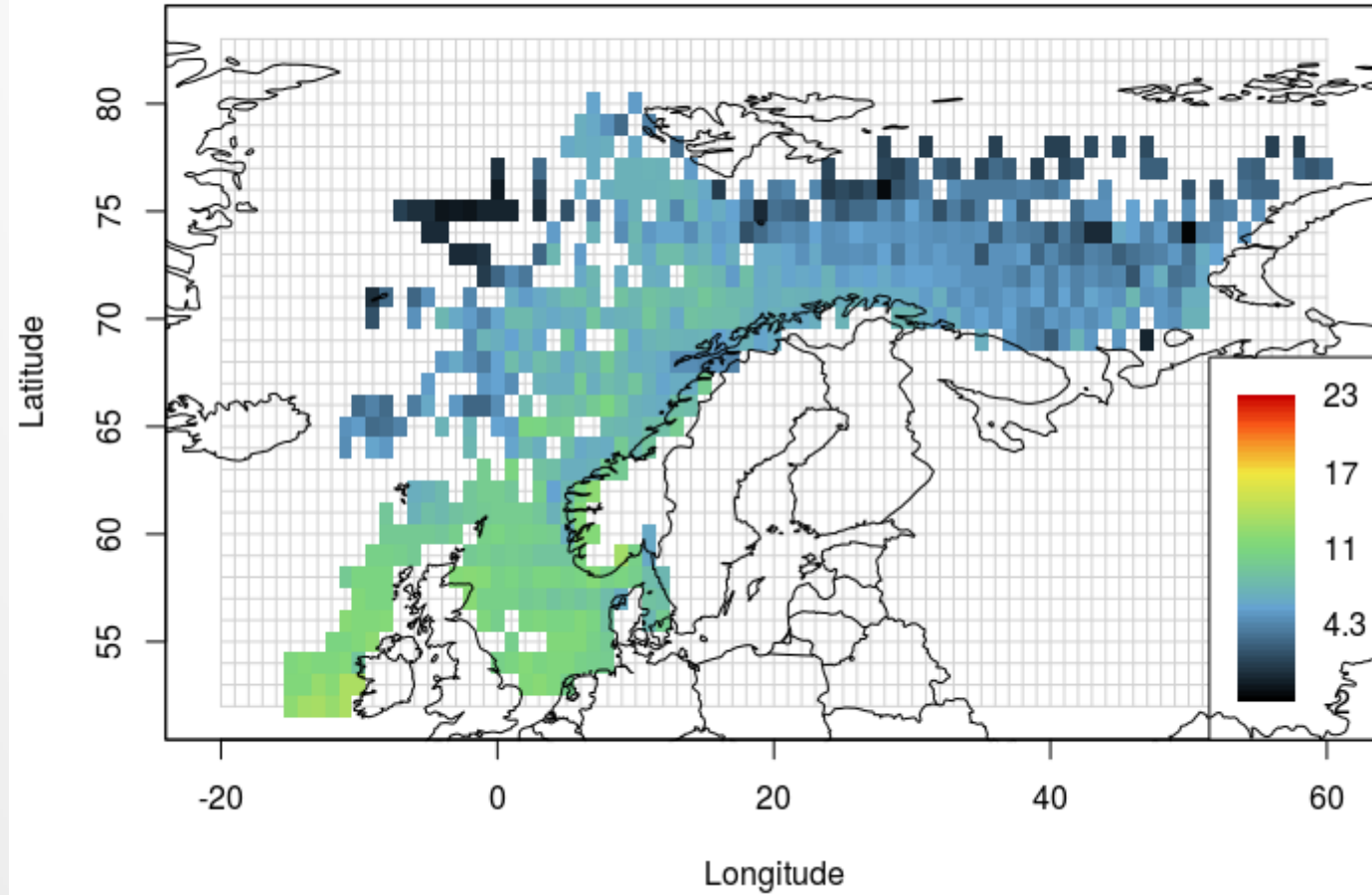
- Select Depth Interval (20m)
- Define Target Year (1995)
- Define Target Trimester (1st)
- Display Variable



Block Average at 20m Depth

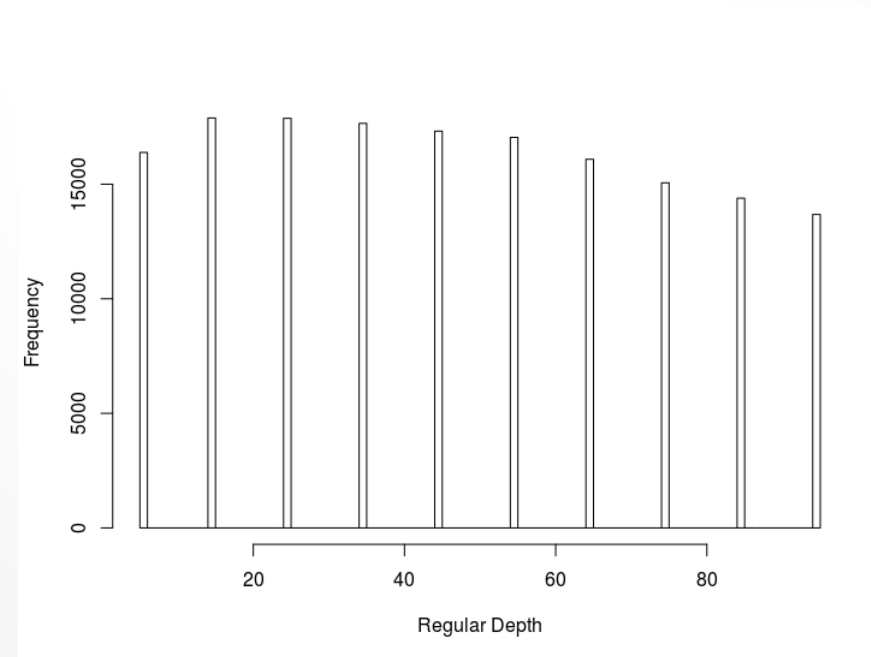
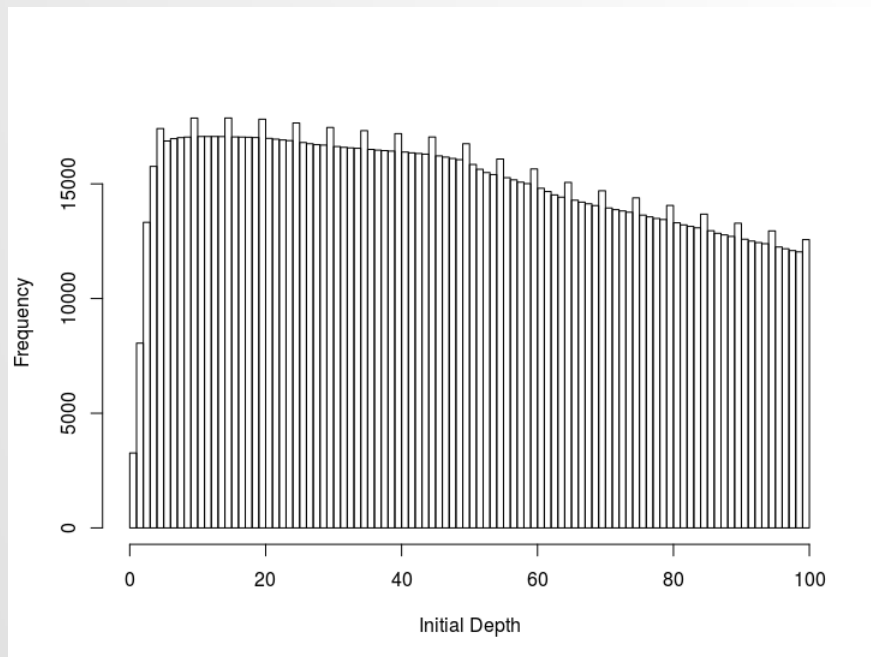
- Select Depth Interval (20m)
- Define Target Year (1995)
- Average per Cell (1 degree)
- Display Variable

Block Average for Temperature (1995-01-01 => 1995-12-31)



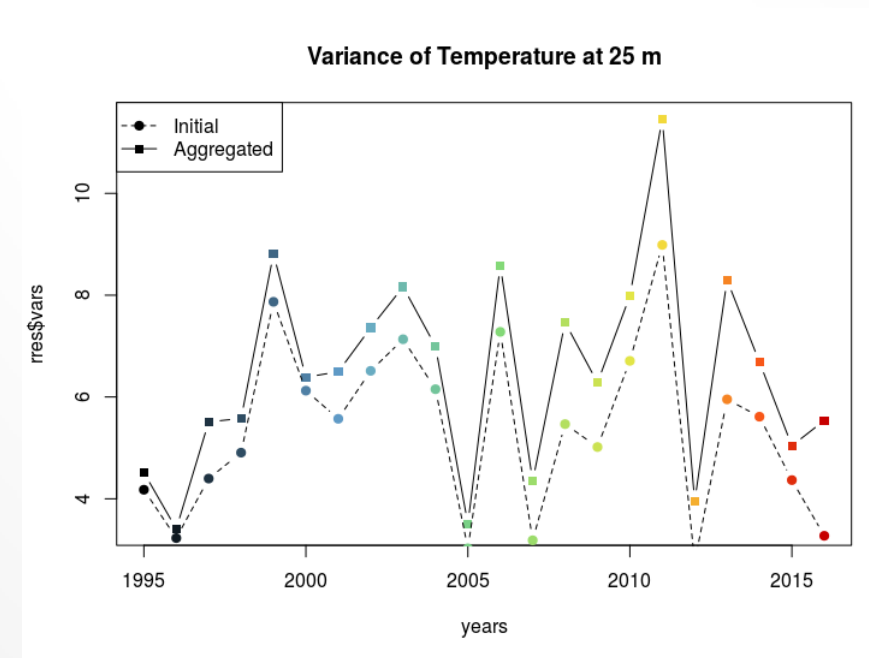
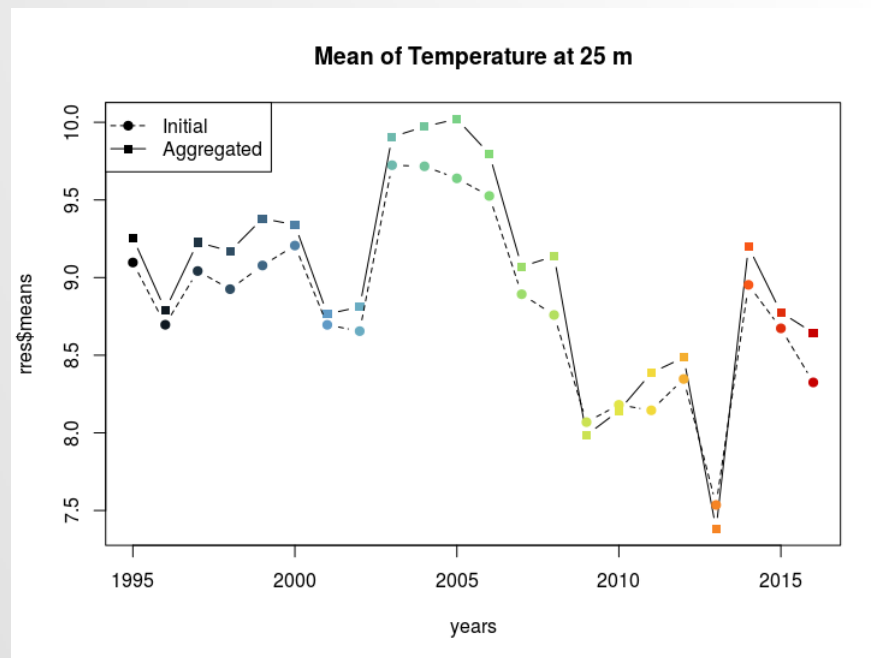
Histogram of Measurement Depths

- Focus on a Smaller Area (South West of Norway)
- Aggregate per Depth Intervals (10m)
- Histogram of Depths (before and after Aggregation)



Statistics per Year

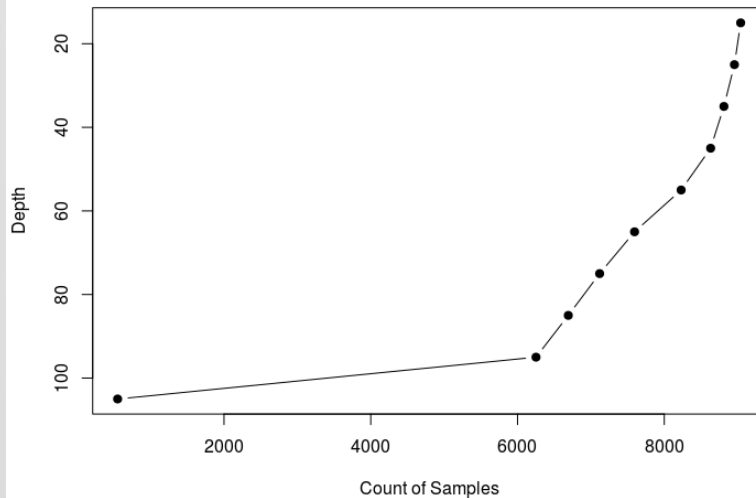
- Select Depth Interval (25m)
- Aggregate per Time Intervals (1 year)
- Mean and Variance of Temperature (before and after Aggregation)



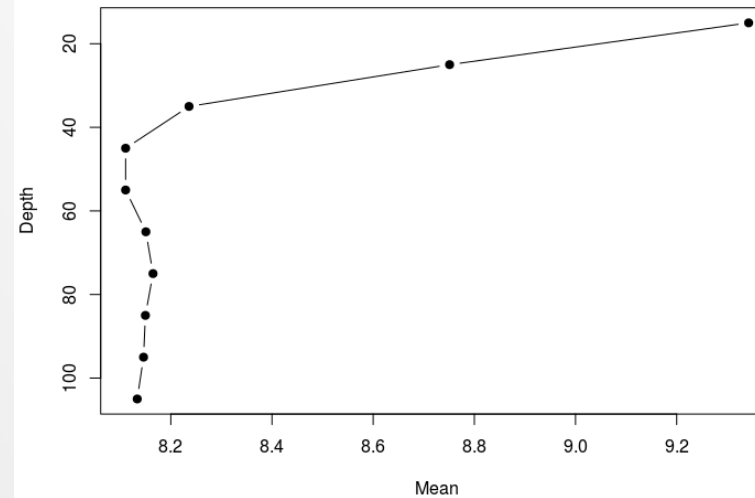
Statistics per Depth

- Select Year (2008)
- Average per Depth Interval (10m)
- Count, Mean and Variance of Temperature per Depth

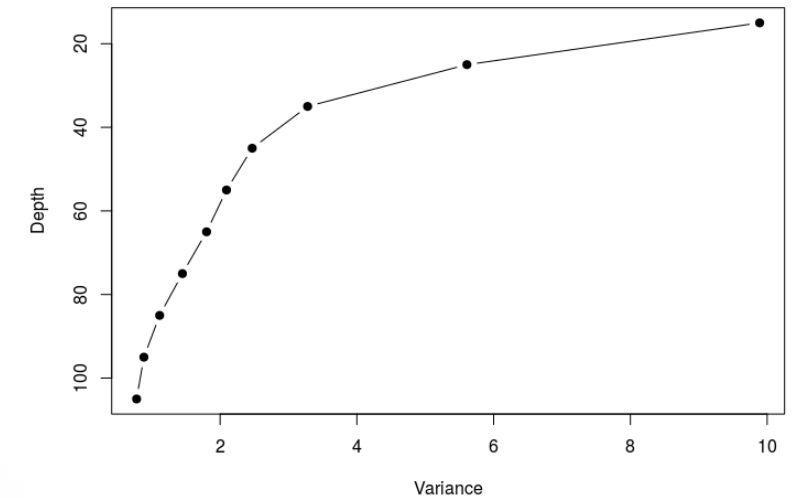
Number of samples for Temperature



Mean for Temperature

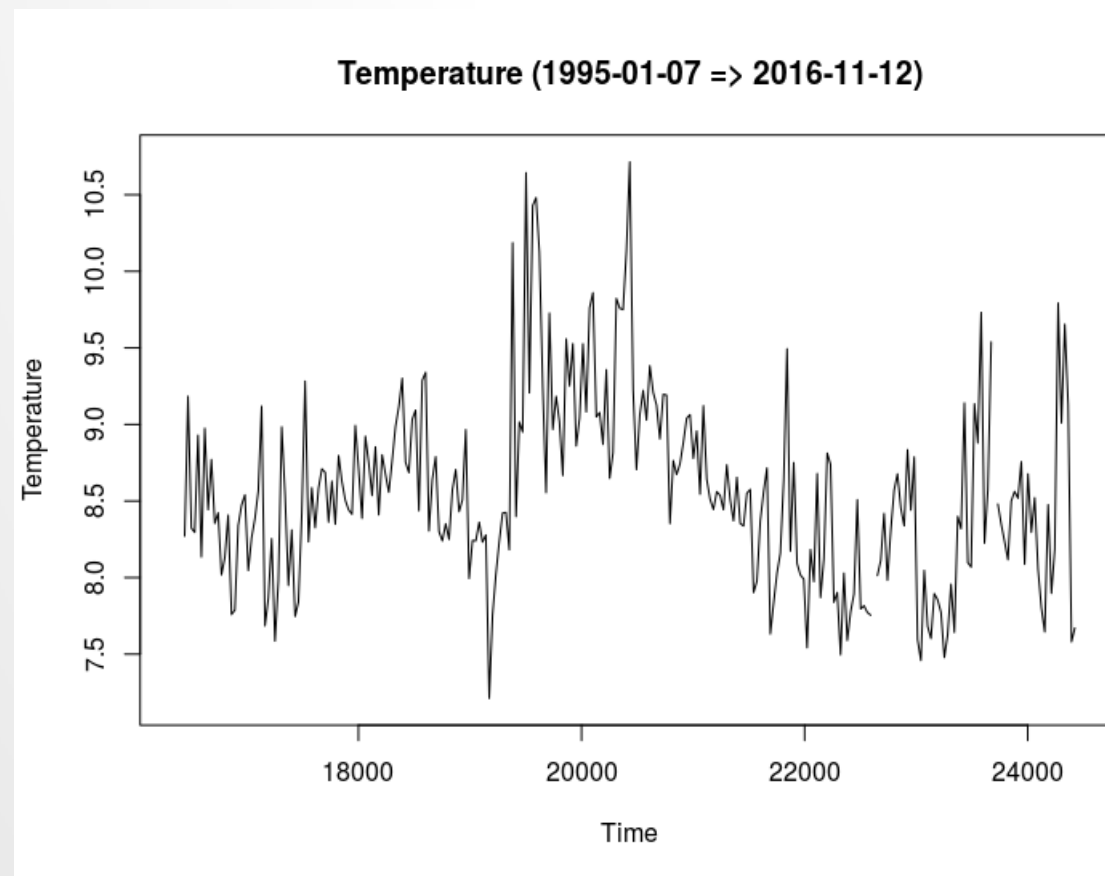


Variance for Temperature



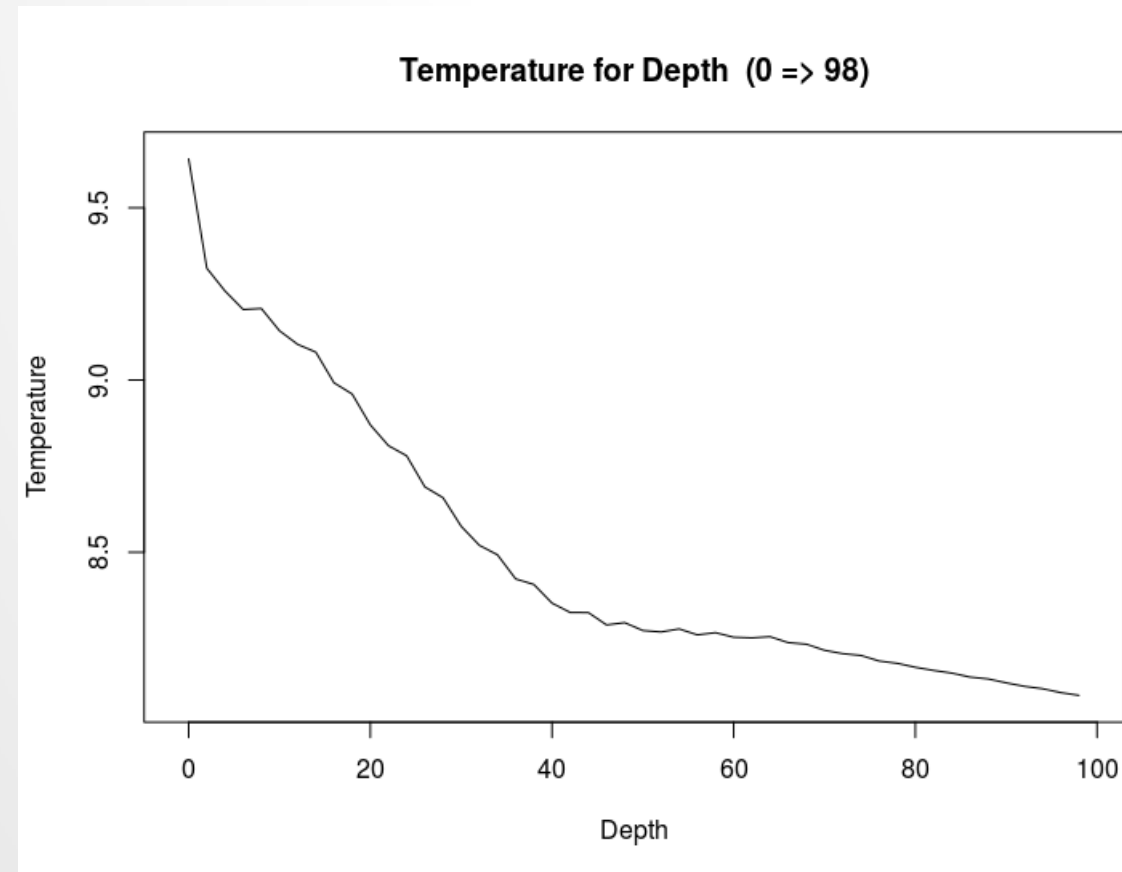
Variation along Time

- Aggregate by Time Intervals (30 days)



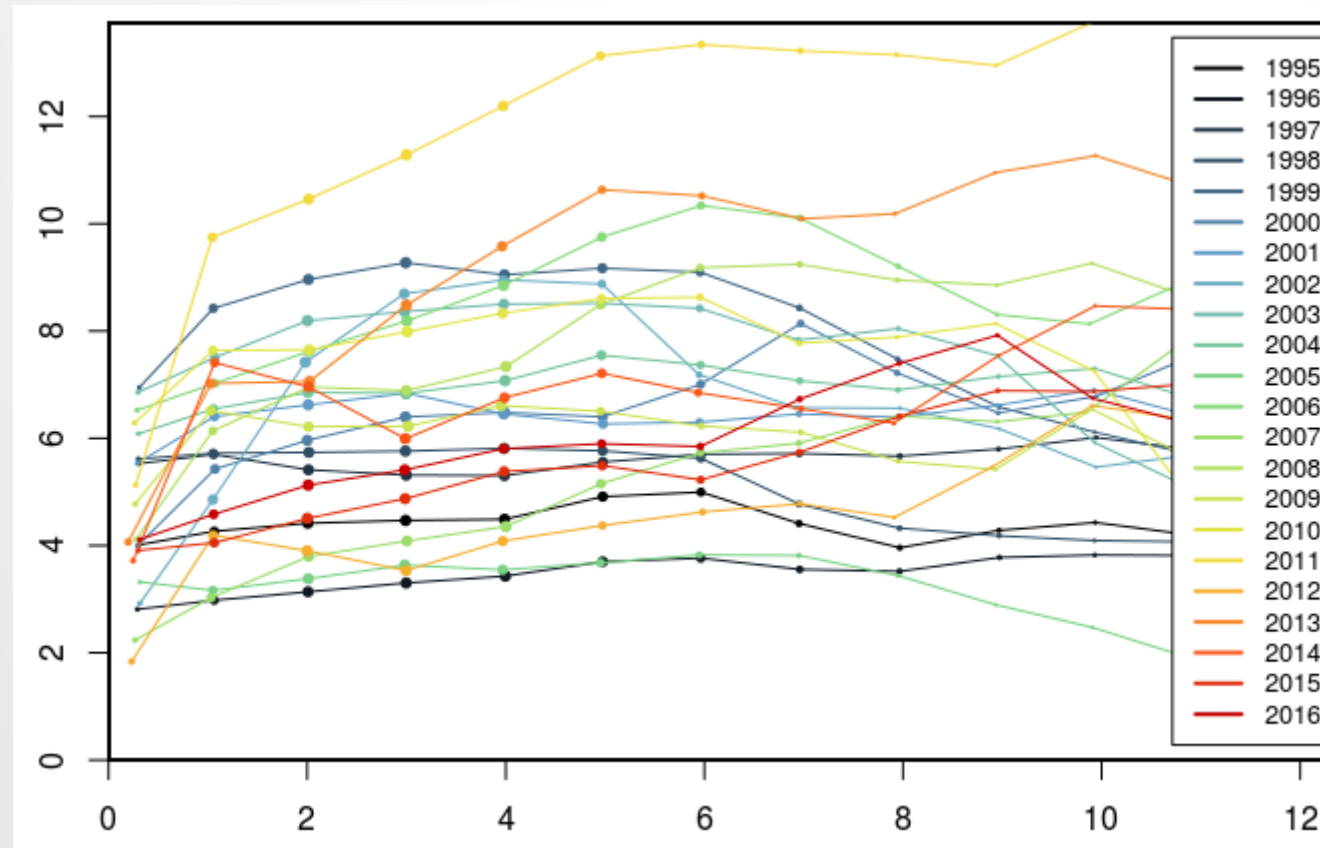
Variation along Depth

- Aggregate by Depth Intervals (2m)



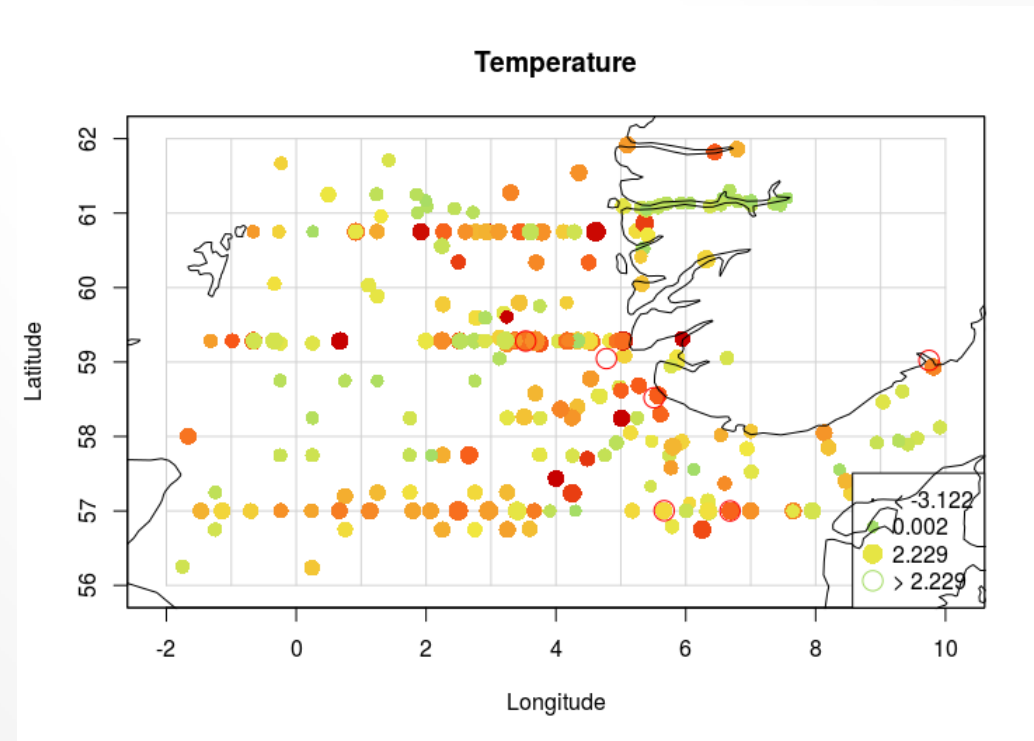
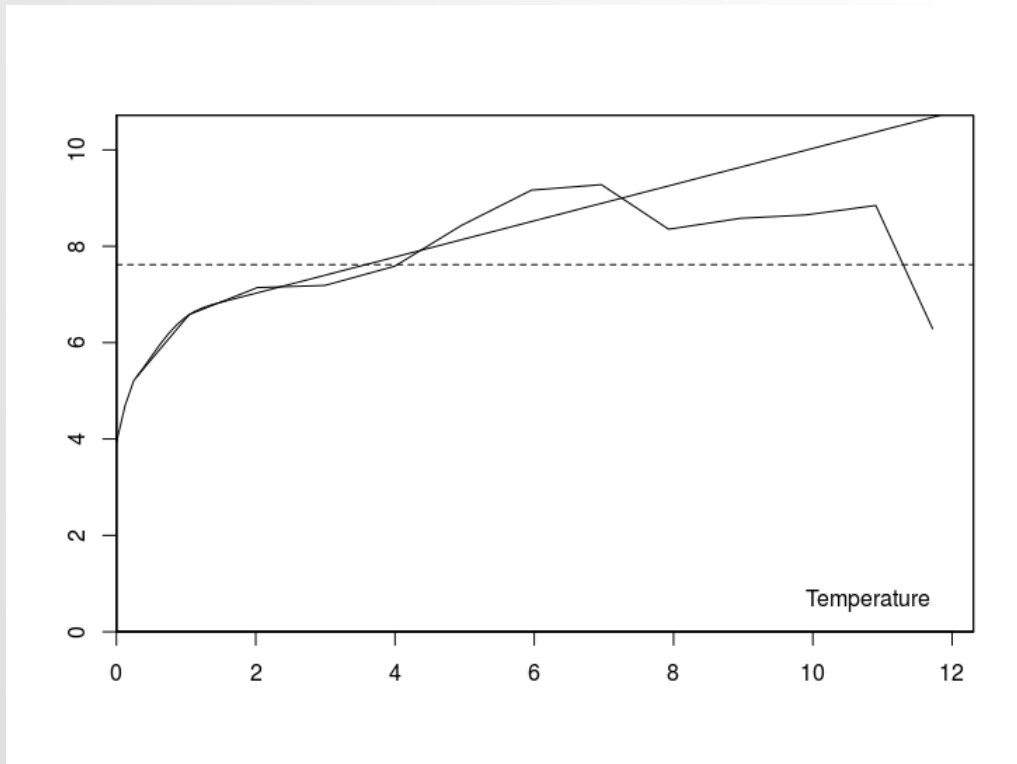
Horizontal Variogram per Year

- Select Depth Interval (25m)
- Omni-directional Variogram per Year for Temperature



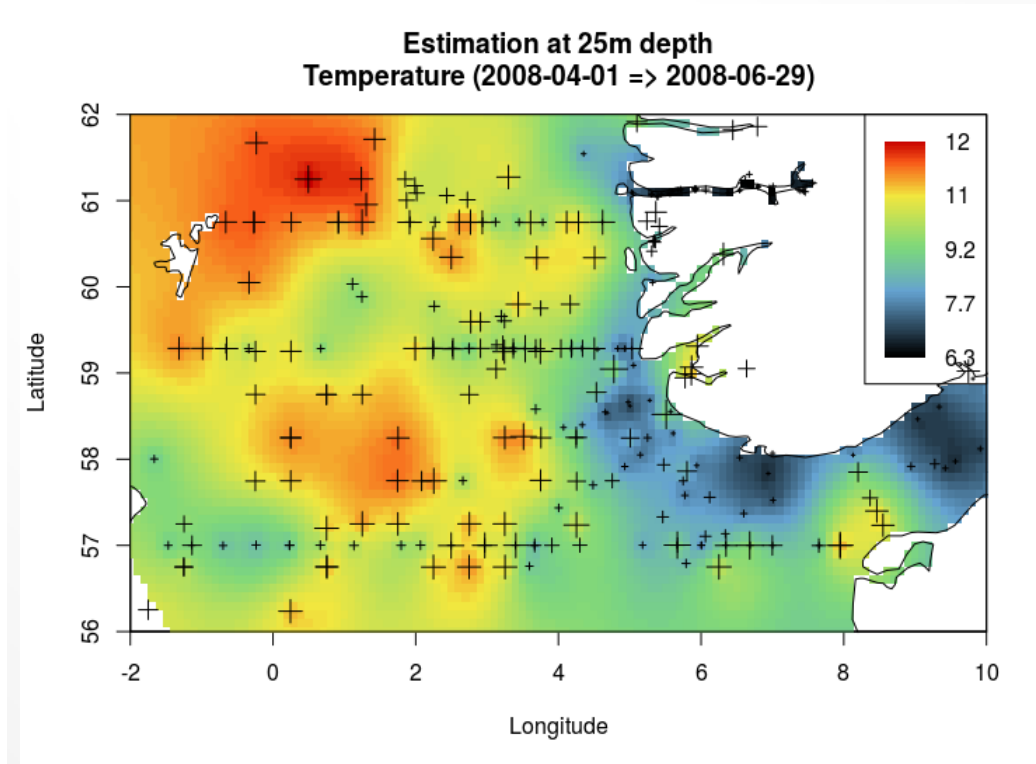
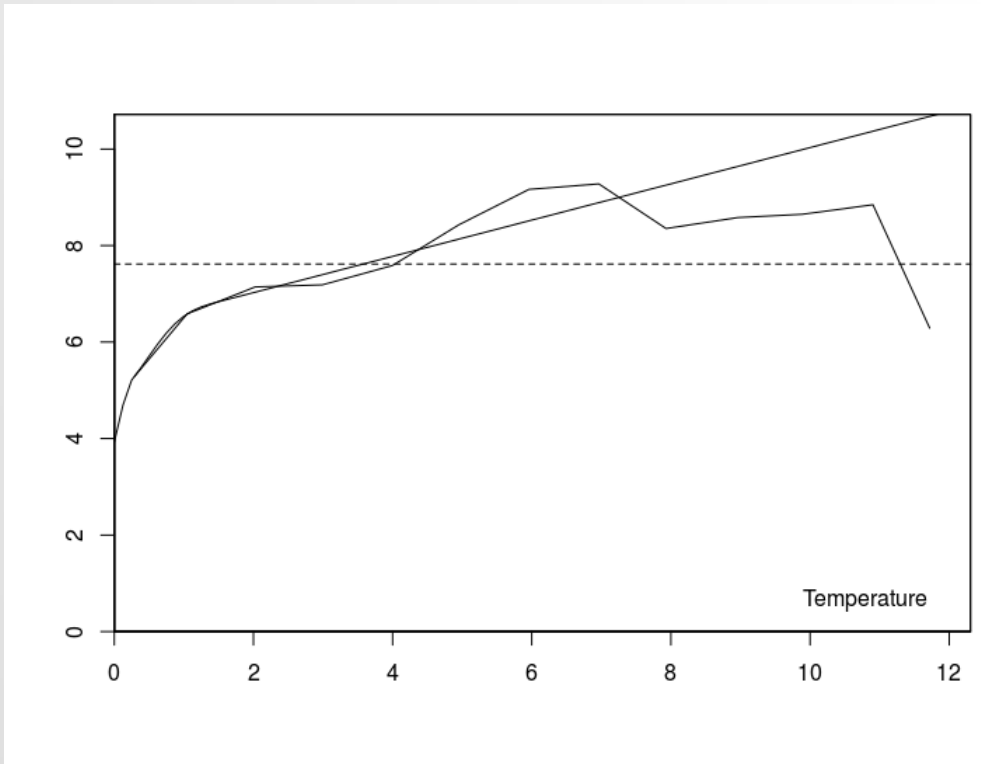
Cross-Validation

- From now: use Aggregated database per Depth Intervals (10m)
- Calculate Average Horizontal Variogram and Fit the Model of Temperature
- Perform the Cross-Validation at Data Samples



2-D Estimation of Temperature

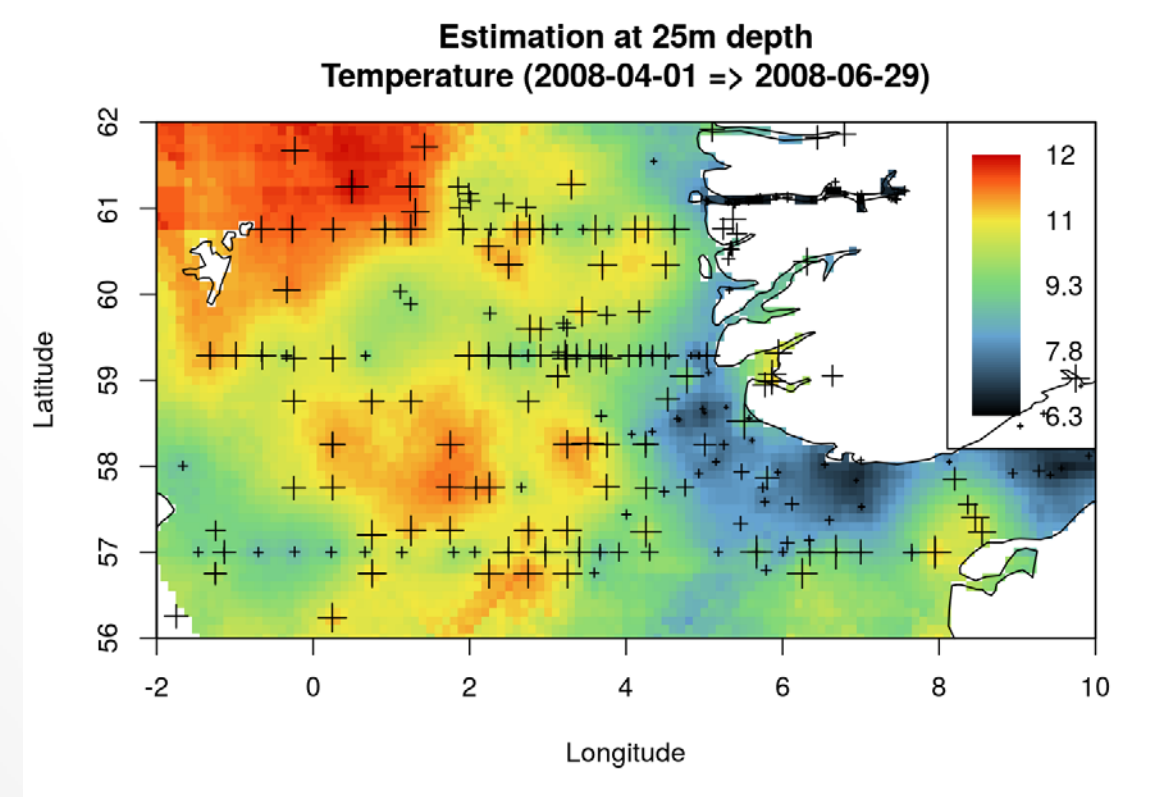
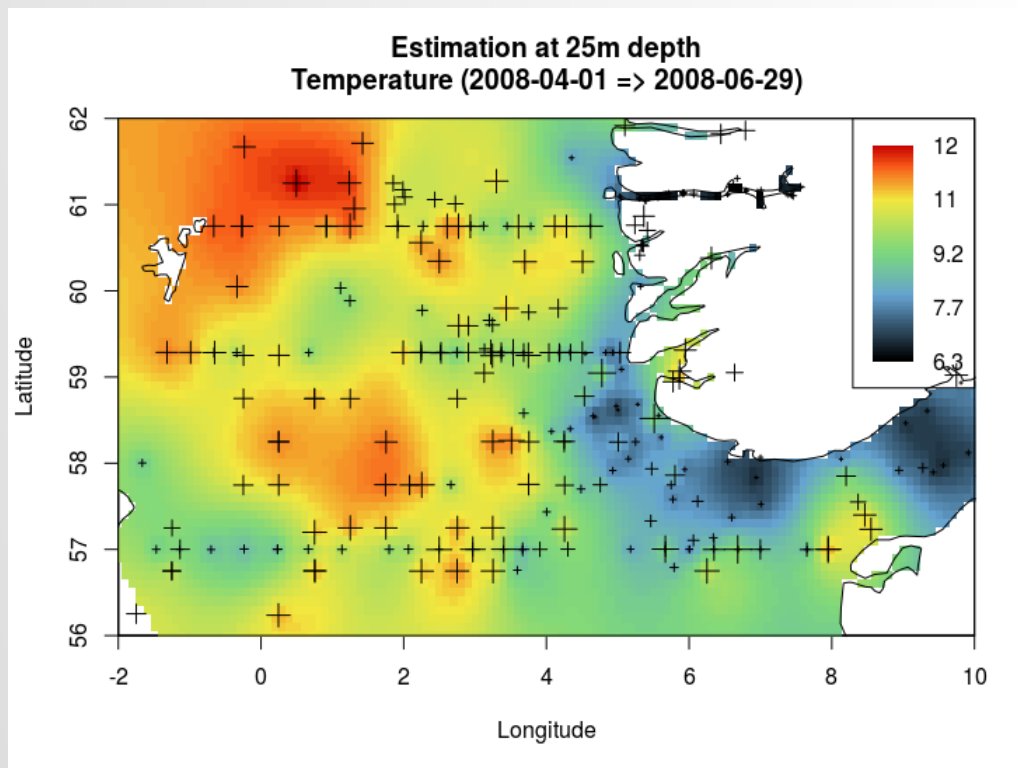
- Select Depth Interval (25m) and Focus on 2nd Trimester of Year 2008
- Calculate Average Horizontal Variogram and Fit the Model for Temperature
- Estimate the Temperature in 2-D in South-West of Norway



Moving Neighborhood

Comparing Unique (left) and Moving (right) Neighborhoods

- 2-D Moving Neighborhood: Maximum Number of Samples = 40

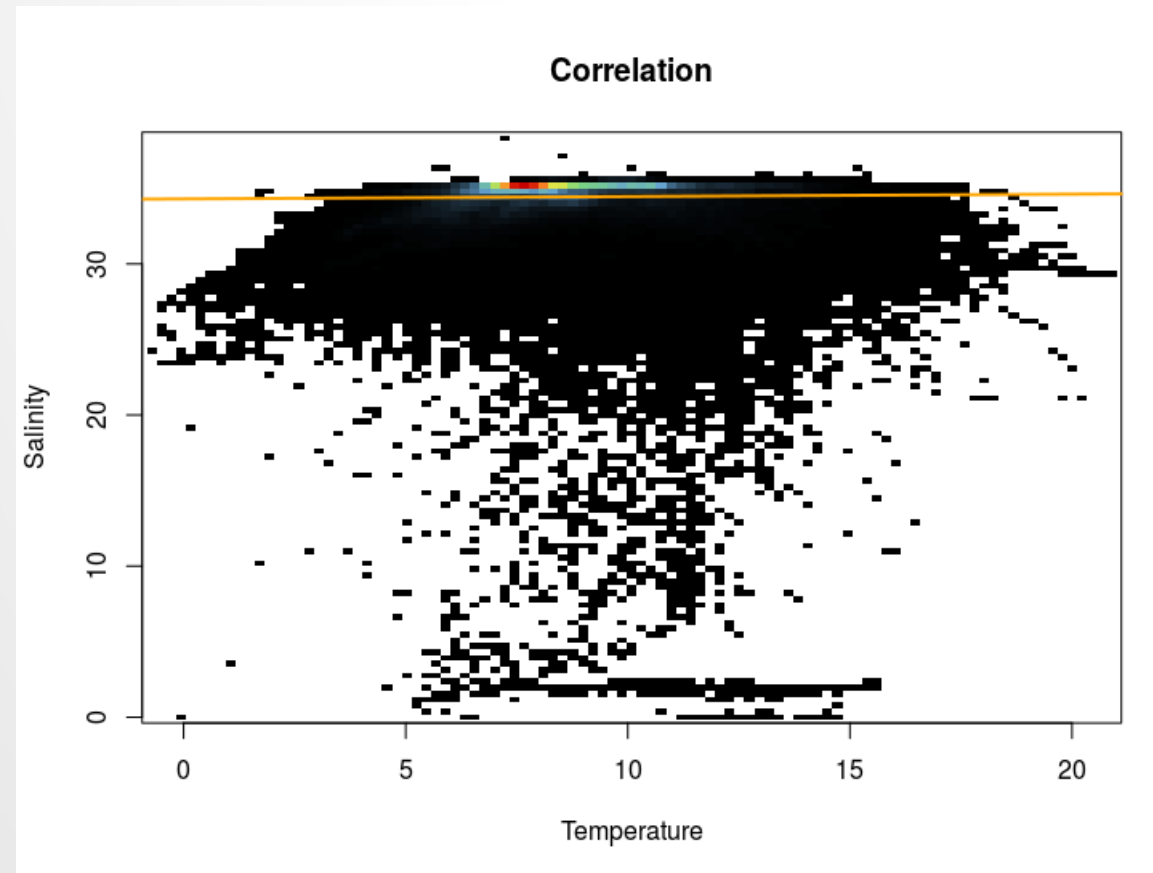


Bivariate Approach: Temperature & Salinity



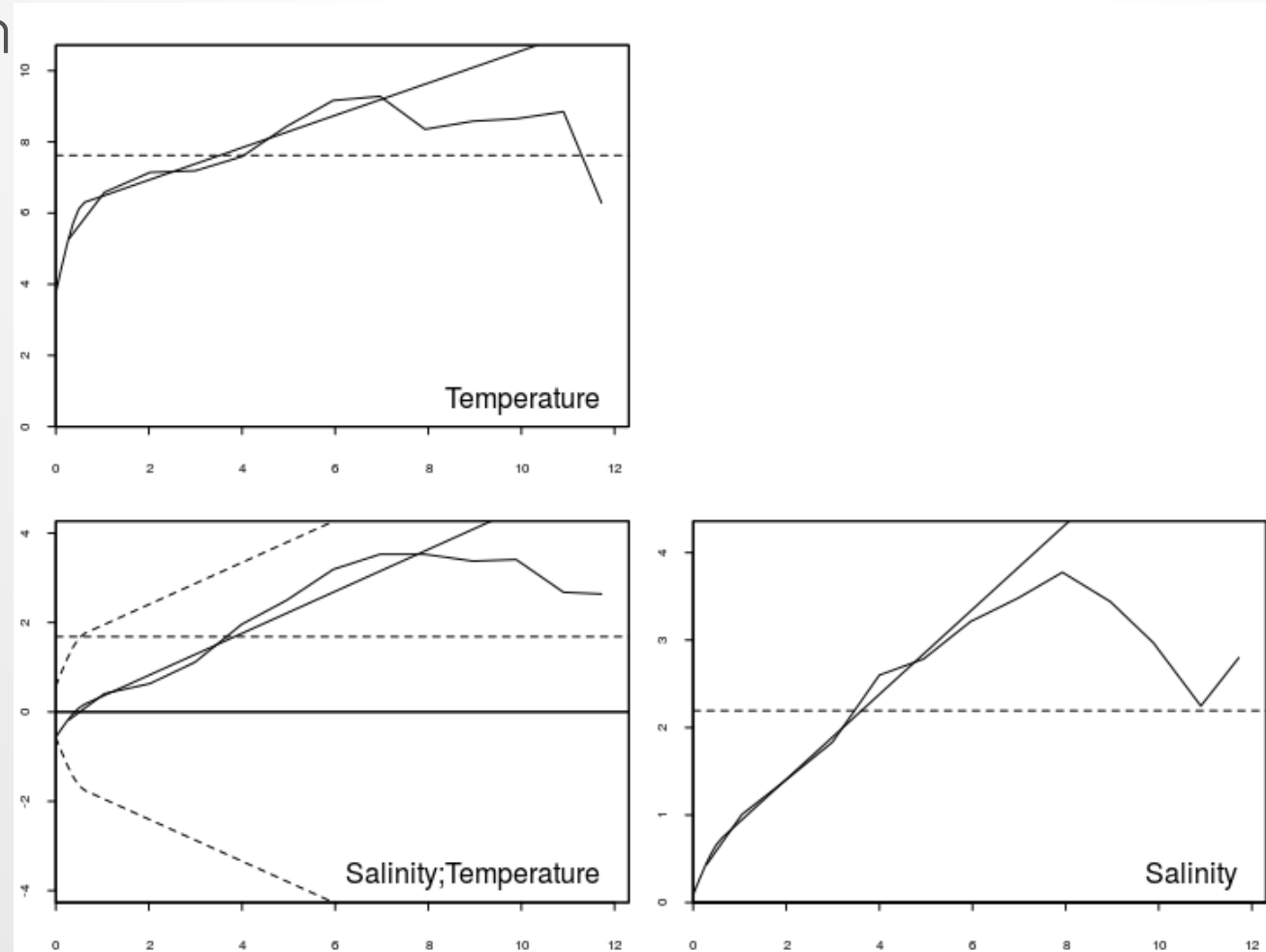
Temperature vs. Salinity

- Define New Variables of Interest
- Correlation and Linear Regression



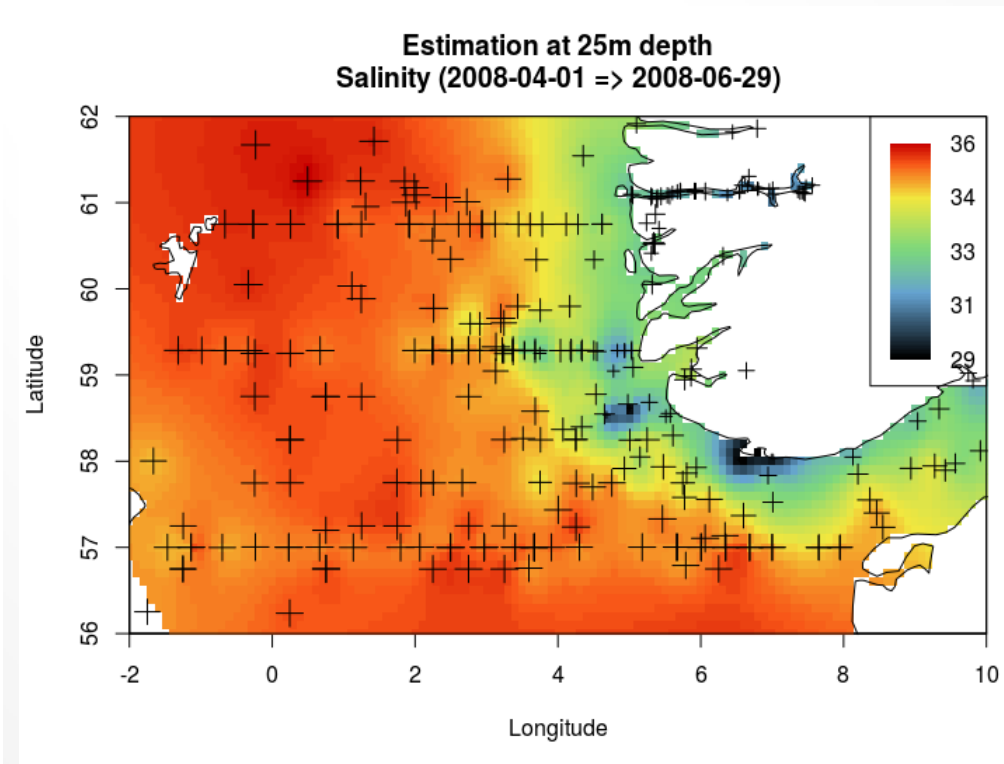
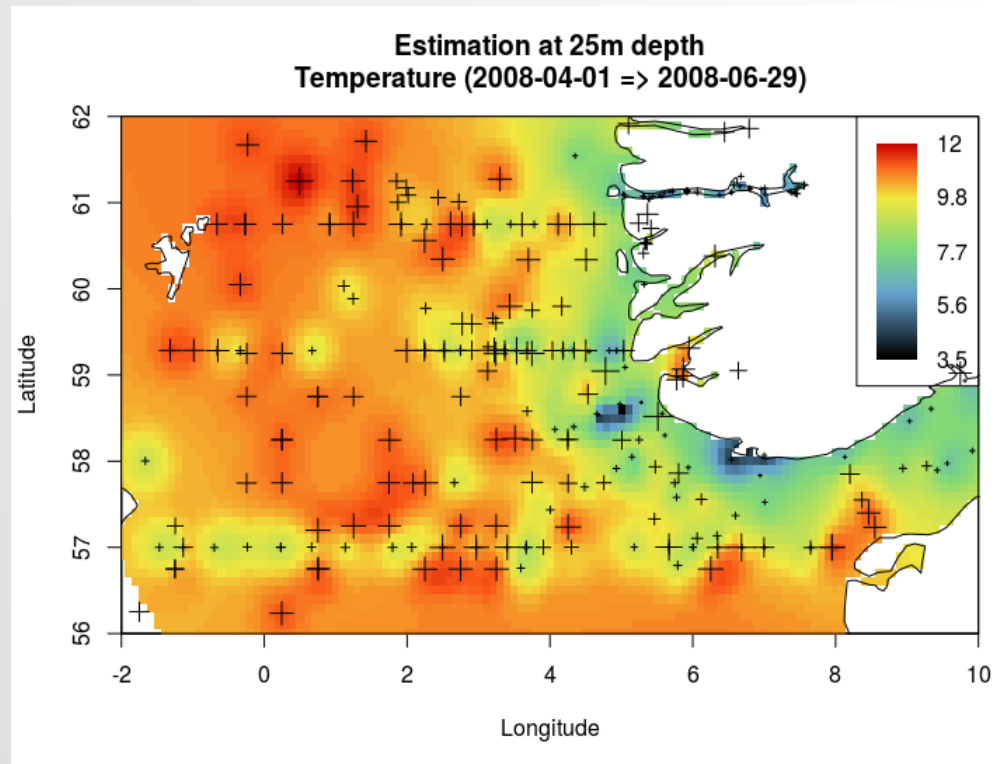
2-D Bivariate Variogram

- Select Depth Interval (25m) and Focus on 2nd Trimester of Year 2008
- Bivariate Variogram



2-D Estimation (Bivariate)

- Select Depth Interval (25m) and Focus on 2nd Trimester of Year 2008
- Calculate Average Horizontal Bivariate Variogram and Fit the Model
- Estimate the Temperature and the Salinity at 25m Depth by Cokriging



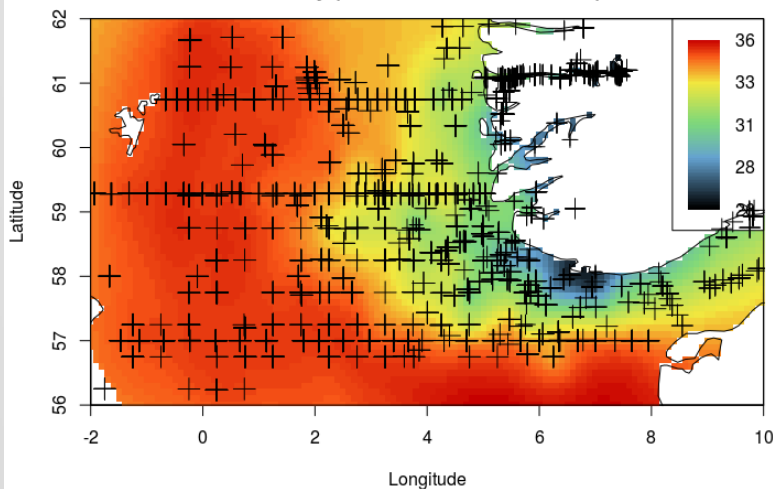
3-D Estimation of Salinity



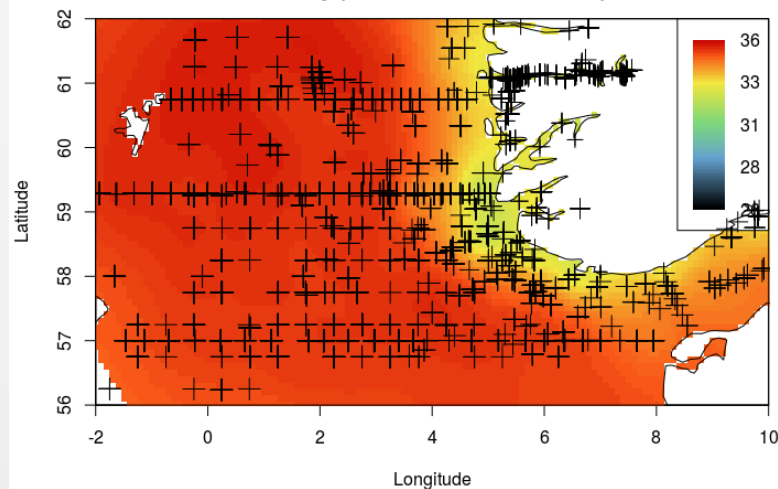
3-D Estimation of Salinity

- Define New Variable of Interest
- Focus on 2nd Trimester of Year 2008
- Calculate 3-D Variogram (by Bench) and Fit the Model for Salinity
- Estimate Salinity by Benches of 10m Thick from 5m to 95m Depth

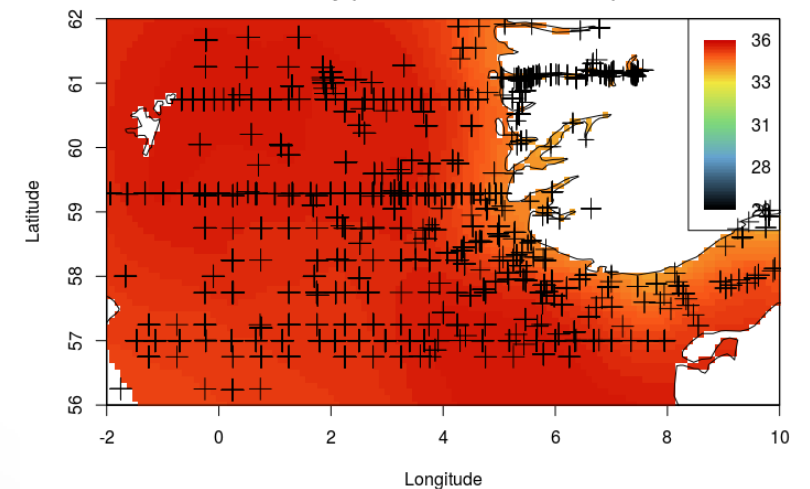
Estimation at 5m depth
Salinity (2008-04-01 => 2008-06-29)



Estimation at 35m depth
Salinity (2008-04-01 => 2008-06-29)

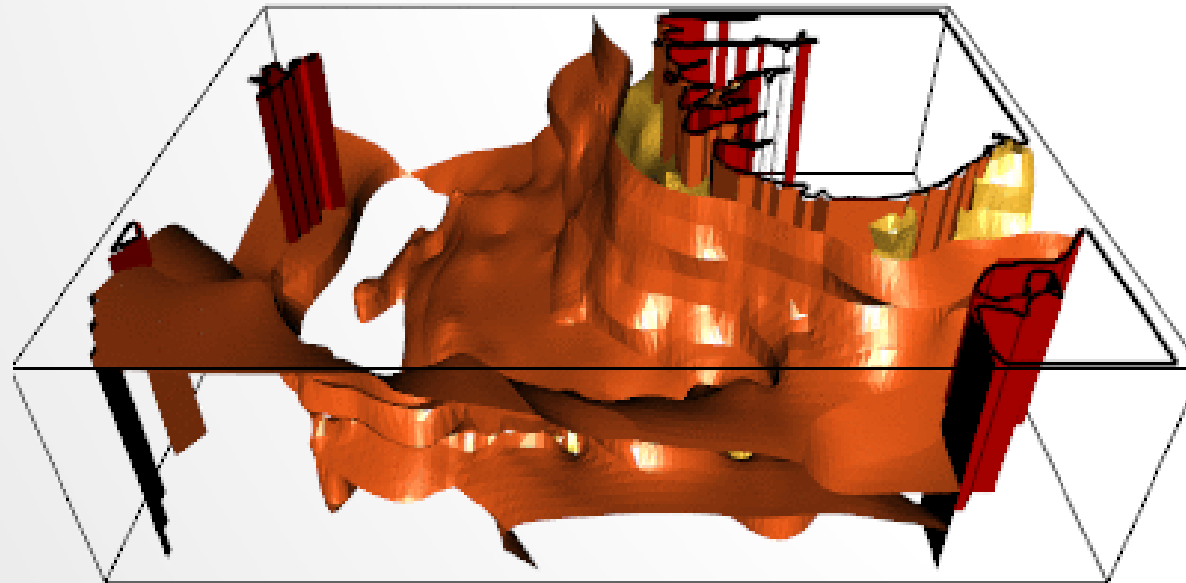


Estimation at 85m depth
Salinity (2008-04-01 => 2008-06-29)



3-D Estimation by Iso-Surfaces

- Show 3-D Estimation of Salinity with Iso-Surface



X

End of presentation

Thank you for attention!

