



Integrated Arctic Observation System

Research and Innovation Action under EC Horizon2020
Grant Agreement no. 727890

Project coordinator:
Nansen Environmental and Remote Sensing Center, Norway


Deliverable 5.11

Integration of new processing services V2

Start date of project:	01 December 2016	Duration:	60 months
Due date of deliverable:	31 May 2021	Actual submission date:	15 November 2021
		Resubmission after review:	26 August 2022
Lead beneficiary for preparing the deliverable:	TERRADUE		
Other contributing partners	NERSC, ARMINES		
Person-months used to produce deliverable:	1.1		

Authors: Hervé Caumont (Terradue), Simone Vaccari (Terradue), Espen Storheim (NERSC), Torill Hamre (NERSC).

Version	DATE	CHANGE RECORDS	LEAD AUTHOR
0.1	24/05/2021	Initial draft - Table of content First version for internal project review (an update being scheduled to September 2021, in agreement with the project coordinator, in order to accommodate with late started WP6 activities on the showcase applications)	Hervé Caumont
0.2	30/07/2021	Second draft	Hervé Caumont
0.3	06/08/2021	Third draft	Simone Vaccari
0.4	13/10/2021	Update of C-PAD and TS-SIC service descriptions (incurred delays from coordination with WP6)	Torill Hamre
0.5	20/10/2021	Internal Review (incurring delays to implement all the remarks from internal reviewers)	Pedro Goncalves Fabien Ors
1.0	15/11/2021	Final version for EC delivery	Hervé Caumont
1.1	15/11/2021	Final review and submission	Stein Sandven/Hanne Sagen
1.2	25/08/2022	Updated according to external review	Hervé Caumont

Approval	Date:	Sign.
X	26 August 2022	 Stein Sandven

USED PERSON-MONTHS FOR WRITING THIS DELIVERABLE					
<i>No</i>	<i>Beneficiary</i>	<i>PM</i>	<i>No</i>	<i>Beneficiary</i>	<i>PM</i>
1	NERSC	0.1	24	TDUE	1
2	UiB		25	GINR	
3	IMR	-	26	UNEXE	
4	MISU		27	NIVA	
5	AWI	-	28	CNRS	
6	IOPAN		29	U Helsinki	
7	DTU	-	30	GFZ	
8	AU		31	ARMINES	-
9	GEUS	-	32	IGPAN	
10	FMI	-	33	U SLASKI	
11	UNIS		34	BSC	
12	NORDECO		35	DNV GL	
13	SMHI		36	RIHMI-WDC	
14	USFD		37	NIERSC	
15	NUIM	-	38	WHOI	
16	IFREMER		39	SIO	
17	MPG		40	UAF	
18	EUROGOOS		41	U Laval	
19	EUROCEAN		42	ONC	
20	UPM		43	NMEFC	
21	UB		44	RADI	
22	UHAM		45	KOPRI	
23	NORUT		46	NIPR	
			47	PRIC	

DISSEMINATION LEVEL		
PU	Public, fully open	X
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

EXECUTIVE SUMMARY

Within INTAROS, WP5 (Data integration and management) is tasked with designing and implementing evolutions of the cloud platform and tools for the iAOS (integrated Arctic Observation System).

This document in its volume V2 defines additional resources for iAOS data exploitation and processing, in particular in the context of the WP5 support to the WP6 iAOS Showcase applications.

These new resources are delivered as a collection of R and Python-based Jupyter Notebooks, a state of the art approach for collaborative science, interactive data analysis and management of software code experiments.

This V2 report embeds the results previously delivered as part of the V1 release, these can be found in the sections 3. "C-PAD Service Specification" (Characterization of Passive Acoustic Data) and 4. "TS-SIC Service Specification" (Time Series of Sea Ice Concentration). For these two services, the update between D5.7 (V1) and D5.11 (V2) was done only on the structure of the chapters "Software modules required", "Cloud resources allocation" and "Roadmap for future development". No new developments were performed here.

The new results are focused on the exploitation of interoperability protocols (OPeNDAP) as promoted by INTAROS for the iAOS, and on the use of the RGeostats Geostatistical software library. They are presented in section 5. "TS-CTD Service Specification" (Time Series of Conductivity-Temperature-Depth) and in section 6. "iAOS Showcase Service Specification".

Table of Contents

Acronyms and abbreviations	7
Definitions	8
1. Introduction	9
1.1. Purpose and scope of the document	9
1.2. Intended audience of this document	9
1.3. Document Structure	10
2. Overview	11
2.1. Jupyter Notebooks for iAOS processing services	11
2.1.1. Characterization of Passive Acoustic Data (C-PAD)	11
2.1.2. Time Series of Sea Ice Concentration (TS-SIC)	11
2.1.3. Time Series of Conductivity-Temperature-Depth (TS-CTD)	11
2.2. Software libraries for iAOS processing services	12
2.2.1. RGeostats	12
2.2.2. Arctic HYCOS utils	12
2.3. OGC WPS endpoints for iAOS processing services	12
2.2.1. Ellip Workflows applications	12
3. C-PAD Service Specification	13
3.1. Actors interacting with the service	13
3.2. User Stories	13
3.3. Use Cases	13
3.4. Service implementation and examples of use	14
3.4.1. Main features of the C-PAD Service	14
3.4.2. Examples of passive acoustics data analysis and visualisation using C-PAD	15
3.4.3. How to access the C-PAD Service	17
3.4.4. IPR and licensing summary table for C-PAD Service	17
3.5. Software modules required by the C-PAD service	17
3.5.1. Required software	18
3.5.2. IPR and licensing summary table for required software modules	18
3.5.3. Online documentation resources	19
3.6. Cloud resources allocation	19
3.6.1. Data resources	19
3.6.2. ICT resources in the iAOS cloud platform	19
3.7. Roadmap for future development	19
4. TS-SIC Service Specification	20
4.1. Actors interacting with the service	20
4.2. User Stories	20

4.3. Use Cases	21
4.4. Service implementation and examples of use	21
4.4.1. Main features of the TS-SIC Service	22
4.4.2. Examples of sea ice statistics products generated using TS-SIC	22
4.4.3. How to access the TS-SIC Service	23
4.4.4. IPR and licensing summary table for TS-SIC Service	23
4.5. Software modules required by the TS-SIC service	23
4.5.1. Required software	24
4.5.2. IPR and licensing summary table for required software modules	24
4.5.3. Online documentation resources	24
4.6. Cloud resources allocation	24
4.6.1. Data resources	24
4.6.2. ICT resources in the iAOS cloud platform	24
4.7. Roadmap for future development	25
5. TS-CTD Service Specification	26
5.1. Actors interacting with the service	26
5.2. User Stories	26
5.3. Use Cases	29
5.4. Service implementation and examples of use	30
5.4.1 REF Notebook exploiting the CTD 2003 subset	30
5.4.2. DAP Notebook exploiting the CTD 2003 subset	36
5.4.3. How to access the TS-CTD Service	41
5.4.4. IPR and licensing summary table for TS-CTD Service	42
5.5. Software modules required by the TS-CTD service	42
5.5.1. Required software	42
5.5.2. IPR and licensing summary table for required software modules	42
5.6. Cloud resources allocation	42
5.6.1. Data resources	43
5.6.2. ICT resources in the iAOS cloud platform	43
5.7. Roadmap for future development	43
6. iAOS Showcase Applications - Integration of processing services	44
6.1. Pan-Arctic Hydrological Modelling (with Task 6.1 SMHI)	44
6.2. Barents Sea Multi-depth Temperature & Salinity Maps (with Task 6.2 IMR)	45
6.3. Maps for Svalbard Avalanche Forecast Modelling (with Task 6.4 FMI)	46
6.4. Baffin Bay Bottom Temperature Maps (with Task 6.8 Aarhus)	47
7. Conclusion	50
8. References	51
9. Annex - WP6 references to iAOS Showcase applications	52

List of Figures

Figure 1. Examples of power spectrum spectrogram (left) and noise statistics plot (right) generated by the C-PAD Service when analysing WIFAR data	15
Figure 2. Examples of power spectrum spectrogram (left) and noise statistics plot (right) generated by the C-PAD Service when analysing acoustic data collected in Kongsfjorden, Svalbard	16
Figure 3. Examples of power spectrum spectrogram (left) and noise statistics plot (right) generated by the C-PAD Service when analysing passive acoustic data from the PALAOA observatory in Antarctica	17
Figure 4. Data flow in PAMGuide (from supplement to Merchant et al. (2015)	18
Figure 5. Examples of sea ice concentration statistics, monthly means for June - September 2018 in the Fram Strait of area north of Svalbard, generated from daily CMEMS products.	22
Figure 6. Examples of sea ice concentration statistics, monthly means for June - September 2019 in the Fram Strait of area north of Svalbard, generated from daily CMEMS products.	23
Figure 7. TS-CTD service and Cloud Platform support to implement, test and validate remote data access requirements for CTD data products	28
Figure 8. Positions generated from the OpenDAP's Dataset Descriptor Structure (DDS)	31
Figure 9. REF Notebook - Sea Water Temperature and Pressure, between 0-20 meters for the platform GT	33
Figure 10. REF Notebook - Sea Water Temperature and Pressure, between 0-20 meters for the aggregated platforms	34
Figure 11. REF Notebook - Sea Water Temperature and Pressure, between 50-2956 meters for the platform GT	35
Figure 12. REF Notebook - Sea Water Temperature and Pressure, between 50-2956 meters for the aggregated platforms	36
Figure 13. DAP Notebook - Sea Water Temperature and Pressure, between 0-20 meters for the platform GT	38
Figure 14. DAP Notebook - Sea Water Temperature and Pressure, between 0-20 meters for the aggregated platforms	39
Figure 15. DAP Notebook - Sea Water Temperature and Pressure, between 50-2956 meters for the platform GT	40
Figure 16. DAP Notebook - Sea Water Temperature and Pressure, between 50-2956 meters for the aggregated platforms	41

List of Tables

Table 1. Online repository for the C-PAD Service.	17
Table 2. Online repository for software modules required by the C-PAD Service.	18
Table 3. Online repository for the TS-SIC Service.	23
Table 4. Online repository for software modules required by the TS-SIC Service.	24
Table 5. Online repository for the TS-CTD Service.	42
Table 6. Online repository for software modules required by the TS-SIC Service.	42

Acronyms and abbreviations

API	Application Programming Interface
CPU	Central Processing Unit
DevOps	Development to Operations
DIAS	Data and Information Access Service
DDS	OpenDAP's Dataset Descriptor Structure
EGI	European Grid Infrastructure
EO	Earth Observation
FRAM	FRontiers in Arctic marine Monitoring project
HTTP	HyperText Transfer Protocol
IaaS	Infrastructure-as-a-Service
iAOS	integrated Arctic Observation System
ICT	Information and Communication Technologies
IDE	Integrated Development Environment
NERSC	Nansen Environmental and Remote Sensing Center
OGC	Open Geospatial Consortium
OpenDAP	Open Data Access Protocol
OSDD	OpenSearch Description Document
OWS	OGC Web Services
PaaS	Platform-as-a-Service
SDK	Software Development Kit
THREDDS	Thematic Real-time Environmental Distributed Data Services
VM	Virtual Machine
VPN	Virtual Private Network
WMS	Web Map Service
WPS	Web Processing Service

Definitions

Application

A program or group of programs designed for end users. An application, application program or application software is a computer program designed to help people perform an activity.

Cloud resource

Cloud computing resources are allocated to a software system owned by an organisation in order to run operationally within an infrastructure. The customer organisation has access to computer power, storage, bandwidth, and security layers configured for operational needs.

iAOS Showcase Application

A set of tailored 'real-world' applications, showing the benefits of integrating Arctic data from local and scientific observation systems towards better-informed decisions and better-documented processes, demonstrating the benefits of the iAOS functionality to selected stakeholders.

Service specification

A service specification clarifies all the objectives and requirements to be addressed by a specific system component or module. The service specification typically expresses the problem domain, the design principles and the expected outcomes, along with the constraints to be taken in account and the performances to be delivered.

Service implementation

The software installation and configuration outcomes necessary to communicate on the service access conditions and to deliver the service to users.

Use Case

An actor represents a human or another software interacting with the function to be designed and implemented in a computer system. Along the interaction cycle described by the Use Case specification, each system actor has a specific role in the sequence. These roles are also modelled by the Use Case, in order to provide the context and clarify the functional goals to be addressed. A complete use case will include one main interaction flow and various alternate flows (various ways to perform the task, error management, ...).

User Story

A user story (a scenario) expresses one very specific need that a type of system user has. It's usually written out as a couple of sentences, centered on the result that such a user typically expects out of a task. Writing system requirements in the form of User Stories is a way to put system users at the center of the conversation between multi-disciplinary teams (system architects, software engineers, data managers, ...).

1. Introduction

1.1. Purpose and scope of the document

This document covers the activities of the INTAROS Task 5.5 Integration of new processing services. It provides a second volume to the initial INTAROS deliverable D5.7 (Nov. 2019, updated June 2020). The descriptions of the previously developed processing services (C-PAD and TS-SIC, two iAOS services developed by NERSC) have been kept in this second volume, in order to deliver a comprehensive overview of the processing service integration work done over the full INTAROS project duration. For the sections that are inherited from the D5.7 report, an introduction sentence explains the motivation. This is complemented by the integration of a processing service dedicated to CTD datasets (TS-CTD, an iAOS service developed by Terradue) to support data access query optimization using the OPeNDAP interface, a data access standard promoted throughout the INTAROS project and implemented by several iAOS data providers for the online delivery of their data products. This new TS-CTD service was created out of the notice of the lack of OPeNDAP tutorials to write working DAP query parameters for CTD data products. It is made available in order to support the project partners for the development of iAOS data exploitation applications (e.g. the iAOS showcase applications processing CTD datasets), as well as to support any future iAOS user willing to proceed with a similar OPeNDAP query optimisation approach for CTD data products.

Overall, Cloud integration activities for the delivery of iAOS data processing services as part of custom user applications are presented in this report for three categories of use cases:

1. **Jupyter Notebooks:** for interactive use and community sharing
2. **Software libraries:** for software application development
3. **Web Services:** for distributed systems integration

Based on one or several of these use cases, a set of four selected iAOS Showcase applications are then specified. They address the key objective of the INTAROS project to demonstrate advances in ecological and environmental understanding, by merging and synthesizing iAOS data through ecosystem modelling at the regional (Barents Sea) and local (Greenland) scales.

1.2. Intended audience of this document

The main target audience of this document are the INTAROS Consortium Partners (especially the WP5 partners integrating processing services and the WP6 partners developing showcase applications) and European Commission services.

The processing services described in this report are powered by the iAOS Cloud Platform, deployed and operated by Terradue (cf. INTAROS deliverables D5.5 - iAOS requirements and architectural design V2 and D5.8 - iAOS platform and tools V2).

In particular, the Cloud Platform supports partners to perform the integration and analysis of multidisciplinary arctic data from distributed repositories. The iAOS Cloud Platform design is user driven; consequently a set of custom tools for data discovery, aggregation, analysis and visualization was made available. Part of it, the Ellip Solutions provide an application integration environment accessible by the iAOS software developers who need to build scalable, interoperable data processing applications, and make these applications available to user communities “as-a-Service”.

Such Cloud integration capacity for data processing services is designed to cover the needs of Pan-Arctic applications from stakeholders in industry, governance, local communities or research. It has delivered a suite of selected iAOS applications (the iAOS “Showcase” applications), designed and validated during the INTAROS project through a collaboration process between project partners in the INTAROS Work Package 5 and Work Package 6.

1.3. Document Structure

This section describes the document structure:

Section 1 is this introduction section.

Section 2 provides an overview of the iAOS processing services developed by WP5 for their usage by iAOS users, and discusses any specific context related to their usage conditions.

Section 3 summarizes the C-PAD Service Specification (introduced in D5.7).

Section 4 summarizes the TS-SIC Service Specification (also introduced in D5.7).

Section 5 presents the newly developed TS-CTD Service Specification.

Section 6 provides the Service Specification for the iAOS processing services supporting four selected iAOS showcase applications (cf. WP5 collaboration with WP6).

Section 7 concludes on the work done during the INTAROS project for development and deployment (i.e. provide the user access) of iAOS processing services.

2. Overview

2.1. Jupyter Notebooks for iAOS processing services

2.1.1. Characterization of Passive Acoustic Data (C-PAD)

Purpose	The C-PAD service enables users to analyze and present passive acoustic data, in order to support the detection and classification of different sources of noise pollution in the ocean.
Provider	NERSC
Online access	Jupyter Notebook (License CC BY 4.0) https://github.com/ec-intaros/PAMGuide-R-Tutorial
Products delivered	Noise level statistics and spectrograms.

2.1.2. Time Series of Sea Ice Concentration (TS-SIC)

Purpose	The TS-SIC service enables users to extract a time series of sea ice concentration maps for a selected time period in the Fram Strait and North of Svalbard areas.
Provider	NERSC
Online access	Jupyter Notebook (License CC BY 4.0) https://github.com/ec-intaros/enb-sea-ice-concentration
Products delivered	Maps of monthly sea ice statistics (e.g. 10% ice, 50% ice, 70% ice, 90% ice).

2.1.3. Time Series of Conductivity-Temperature-Depth (TS-CTD)

Purpose	The TS-CTD service enables users to access, filter, analyse and visualise CTD datasets queried from OPeNDAP servers, with methods to support developers with inter-comparisons of data subsets while optimizing their OPeNDAP query filters.
Provider	TERRADUE
Online access	Jupyter Notebook (License CC BY 4.0) https://github.com/ec-intaros/opendap-xarray-use-cases
Products delivered	Custom dataframes and visual plots of selected contents from CTD datasets.

2.2. Software libraries for iAOS processing services

2.2.1. RGeostats

Purpose	The RGeostats library enables users to implement and test geostatistical procedures using R scripts. It is adapted to tackle problems with several variables defined in a space of any dimensions.
Provider	ARMINES
Online access	Freeware, COPYRIGHT © by MINES ParisTech / ARMINES http://cg.ensmp.fr/rgeostats https://anaconda.org/Terradue/r-rgeostats https://anaconda.org/Terradue/r-rintaros
Products delivered	R functions

2.2.2. Arctic HYCOS utils

Purpose	The Arctic HYCOS Utils service enables users to download, explore, and extract hydrological data provided by the Arctic-HYCOS project.
Provider	SMHI
Online access	https://github.com/dgustafsson/arctic-hycos-iaos-utils
Products delivered	R functions

2.3. OGC WPS endpoints for iAOS processing services

2.2.1. Ellip Workflows applications

Purpose	The OGC WPS Interface Standard provides standardized inputs and outputs (requests and responses) for geospatial processing services. It also defines how a client software can request the execution of a process, and how the output from the process is handled. It defines an interface that facilitates the publishing and discovery of geospatial processes, and the binding rules for a client application.
Provider	TERRADUE & service integrator partners
Online access	https://ellip.terradue.com
Products delivered	OGC Web Processing Services 1.0 endpoint The RGeostats Geostatistical Library v1 has been deployed as a service on the iAOS Cloud Platform, implementing the processing service for the iAOS Showcase Application “Barents Sea Multi-depth Temperature & Salinity Maps” (cf. INTAROS D5.6). The Geostatistical Library v2 has not been yet deployed as a service.

3.C-PAD Service Specification

Note: this section is based on the previous INTAROS D5.7 report (Integration of new processing services V1) but updated to reflect the current status and future plans for the service.

The iAOS Service “Characterization of Passive Acoustic Data (C-PAD)” enables users to analyse and present passive acoustic data, in order to support the detection and classification of different sources of noise pollution in the ocean. The generated data products address Noise level statistics and spectrograms.

3.1. Actors interacting with the service

The actors interacting with the service are defined as follows:

Scientist - ocean acoustics scientist working in the INTAROS project.

Decision-maker (offshore operations) - user of service, to be invoked through WP6 (Task 6.3).

Stakeholder - member of the ocean acoustics and marine environmental communities.

3.2. User Stories

User Stories are centered on expected results and benefits from the perspective of the system actors and their roles in operating the system. From there, Use Cases (cf. next section) can be more granular, and describe how the system will act to match the expectations.

Story: As a scientist, Bob would like to process passive acoustic data to study the ocean soundscape.

1. Bob selects the service for processing acoustic data
2. Bob defines an area and time range of interest
3. iAOS searches for passive acoustic datasets fulfilling these criteria
4. iAOS presents a list of these datasets
5. Bob selects one of the datasets for processing
6. iAOS runs the passive acoustic data processing service for the selected dataset
7. iAOS presents the estimated ocean soundscape parameters
8. Bob selects to download the results (optional)

3.3. Use Cases

Use Case specifications are meant to capture a system actor point of view (a user, a sub-system) while describing functional requirements. They describe the step by step process expected, for an actor interacting with the system being specified, to complete a goal in using the software system.

The approach for this service is based on a simple interaction sequence where the user can read input data from different sources as follows:

- as a WAV file previously uploaded on the JupyterLab workspace hosted on the Ellip Notebooks solution,

- as a NetCDF data served from an OPeNDAP server.

The user is supported by the application to calibrate the data, and plot the output from the different analysis types available in PAMGuide. The service is implemented as a Jupyter Notebook, and is made available from an online git repository.

Use Case ID: CPAD-1

Use Case Name: Analyse Passive Acoustic Data

Primary Actor: Scientist

Stakeholders and Interests: Decision-maker, Stakeholder

Preconditions: Scientist is identified and authenticated.

Postconditions: Passive acoustic data is calibrated correctly. Power spectrum is computed.

Main success scenario:

1. Scientist selects a file with passive acoustics data
2. Service opens the file and reads the metadata
3. Service displays the metadata
4. Scientist identifies and extracts the calibration metadata
5. Scientist enters calibration metadata as input to the Service
6. Service performs data calibration
7. Scientist selects DFT analysis mode
8. Service performs discrete fourier transform analysis
9. Scientist selects power spectrum density (PSD) spectrogram as plot type
10. Service generates and displays a PSD spectrogram

3.4. Service implementation and examples of use

NERSC has implemented the C-PAD service in R, utilizing Jupyter Notebook as a rapid development and interactive testing tool. C-PAD draws upon several open source tools and libraries for acoustic data processing, analysis and visualisation. This enables other scientists and service developers to quickly put the service into use, without committing resources to commercial tools.

The C-PAD service is based on the PAMGuide software (Merchant et al., 2015), a tool for analysis of passive acoustic data. This tool was initially designed to process audio files (WAV files), but has now been augmented to support vector input. This allows for processing of data stored in e.g. NetCDF format without converting to WAV before processing, which potentially could lead to loss of data, as well as adding computational time.

3.4.1. Main features of the C-PAD Service

The C-PAD service can read passive acoustic data from either a WAV file or a NetCDF file, perform calibration and processing of these data, and present the results in a spectrogram and a noise statistics plot. The calibration factors must be extracted from the input files or external sources (e.g. cruise report or scientific papers), and entered into the service. If the calibration factors are not known, the service will still run, but users must take into account that the lack of calibration increases the uncertainty of the results.

A spectrogram visualises temporal variation of the spectrum of frequencies in the acoustic signal. This enables the user to visually analyse the changes in different frequency bands over time and possibly detect events such as earthquakes, ship traffic and seismic airgun shots. The analysis of acoustic data can also lead to identification of marine mammals through matching the signal variations of the acoustic recordings to known vocalisations of a given species. The noise statistics generated by the C-PAD service visualises the power spectrum (intensity) of the signal for the different frequencies which the signal is composed of. A series of percentiles (numbers of recordings below a certain intensity) is used to indicate how the signal intensity varies with frequency. A “spike” in this diagram indicates that signal strength is close to the same level throughout the period for which the statistics is computed. Whereas a larger difference between high and low percentiles indicates that the signal varies in strength for a particular frequency for the time period of the statistics.

3.4.2. Examples of passive acoustics data analysis and visualisation using C-PAD

This section shows the C-PAD service applied to passive acoustic data from three sources:

- WIFAR: Waves-in-Ice Forecasting for Arctic Operators; data collected by NERSC; <https://www.nersc.no/project/wifar>
- INTAROS: Integrated Arctic Observation System; data collected by CNRS; <https://catalog-intaros.nersc.no/dataset/passive-acoustic-data-from-kongsfjorden-svalbard-september-2018>
- Perennial Acoustic Observatory in the Antarctic Ocean (PALAOA); data collected by PANGAEA / AWI; <https://doi.pangaea.de/10.1594/PANGAEA.773610>

The WIFAR project conducted several field experiments in the Fram Strait and Barents Sea Marginal Ice Zone (MIZ) using ice-going vessels. As part of WIFAR, an integrated ice station was deployed in the Fram Strait during August 2012 and September 2013, for both periods collecting acoustic data from a hydrophone mounted under the ice. The *Figure 1* shows the spectrogram and noise statistics for an acoustic recording collected during the WIFAR project in August 2012 (Geyer et al, 2017; <https://doi.org/doi:10.11582/2017.00012>). The spectrogram shows a strong, constant signal around 20 Hz throughout the 10 minute recording. This frequency is typical for fin whale vocalization. A peak is also seen near 300 Hz which could be ship noise. This dataset is publicly available through the Norwegian Marine Data Centre (NMDC).

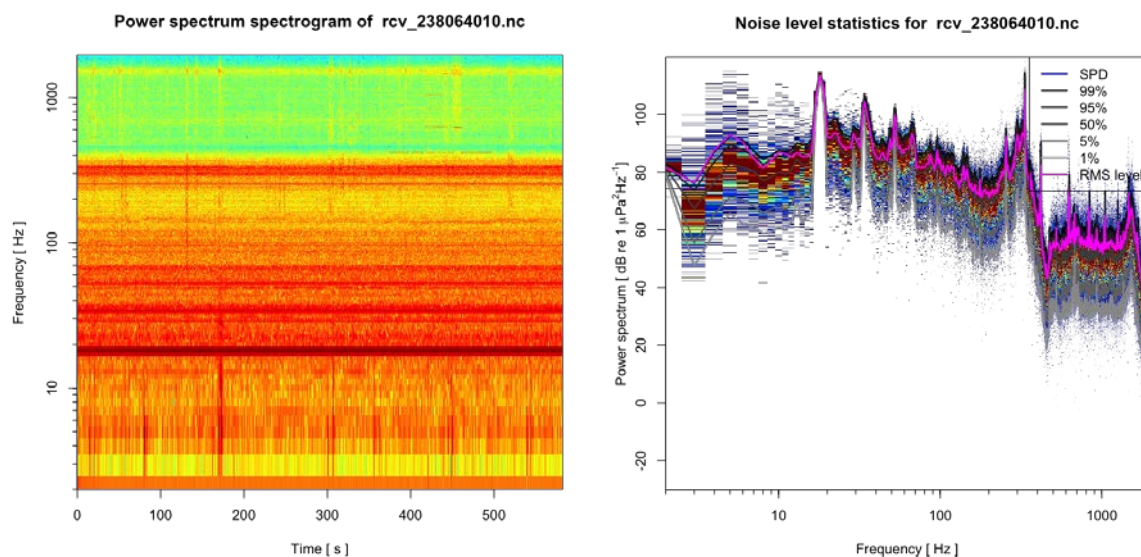


Figure 1. Examples of power spectrum spectrogram (left) and noise statistics plot (right) generated by the C-PAD Service when analysing WIFAR data (Geyer *et al.*, 2016)

In the INTAROS project, the National Center for Scientific Research (CNRS), France has collected passive acoustic data from a mooring in Kongsfjorden, Svalbard. The *Figure 2* shows the spectrogram and noise statistics for an acoustic recording collected during the INTAROS project in September 2018 (Chauvaud *et al.*, 2020). The spectrum is dominated by low-frequency noise below 10 Hz. Local peaks around 10 Hz and 80 Hz are also seen, which could be mammal vocalization. Intermittent broad-band signals are also seen in the spectrogram. This dataset is publicly available through the INTAROS Data Catalog.

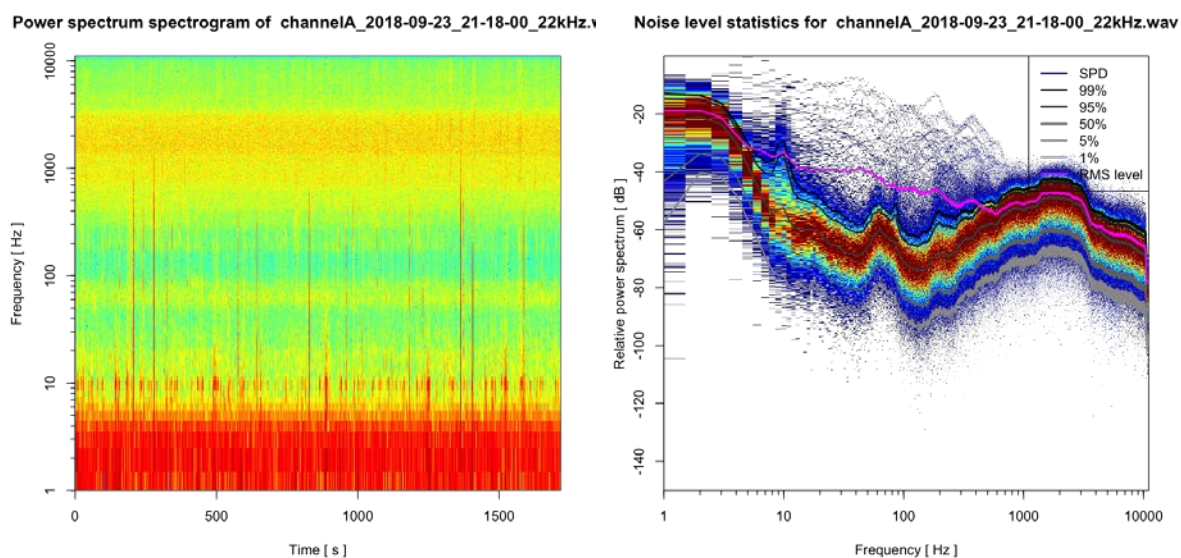


Figure 2. Examples of power spectrum spectrogram (left) and noise statistics plot (right) generated by the C-PAD Service when analysing acoustic data collected in Kongsfjorden, Svalbard (<https://catalog-intaros.nersc.no/dataset/passive-acoustic-data-from-kongsfjorden-svalbard-september-2018>)

The Perennial Acoustic Observatory in the Antarctic Ocean (PALAOA) has collected acoustic data off the Ekström ice shelf, Antarctica, for many years since its establishment in December 2005. Recordings from PALAOA are publicly available in PANGAEA, in WAV format. The

Figure 3 shows the spectrogram and noise statistics for an acoustic recording collected at PALAOA (Kindermann, 2013). This recording is very stable throughout the 10 minute recording period. Peaks are seen from 40 Hz to 400 Hz, but are difficult to identify without further analysis.

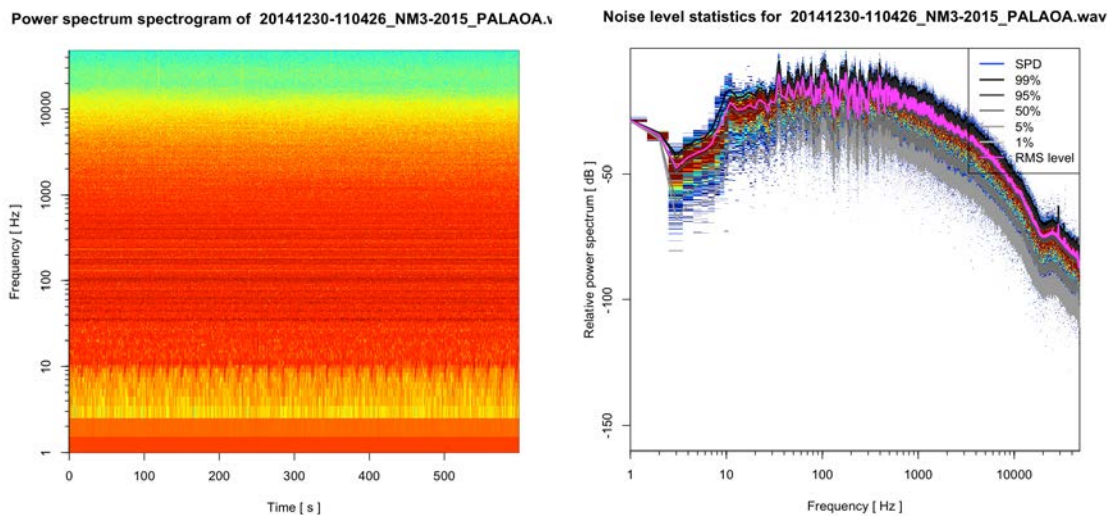


Figure 3. Examples of power spectrum spectrogram (left) and noise statistics plot (right) generated by the C-PAD Service when analysing passive acoustic data from the PALAOA observatory in Antarctica (<https://doi.pangaea.de/10.1594/PANGAEA.773610>)

3.4.3. How to access the C-PAD Service

The C-PAD Service can be used as a basis for further development by scientists and service developers in the INTAROS WP6 and by external parties. Source code for the service (Table 1) and the software it depends on is publicly available (Table 2).

3.4.4. IPR and licensing summary table for C-PAD Service

The software package for the C-PAD Service is available from a public git repository (Table 1). Besides source code, this repository provides a description of dependencies and installation instructions, as well as links to some examples dataset that can be used to test the service.

Table 1. Online repository for the C-PAD Service.

Package	Source URL	License
C-PAD	https://github.com/ec-intaros/PAMGuide-R-Tutorial	CC BY 4.0

3.5. Software modules required by the C-PAD service

The C-PAD Service is developed using PAMGuide, an open source tool for processing and visualisation of passive acoustic data. PAMGuide is written by Nathan Merchant, Centre for Environment, Fisheries & Aquaculture Science (Cefas), United Kingdom. The software package is available for both MATLAB and R development of passive acoustic data services. A detailed

description of how to use PAMGuide is given in the supplementary material of Merchant et al. (2015).

3.5.1. Required software

This service has been developed using the R version of PAMGuide. This software package can read WAV files with passive acoustic data, analyse and visualise these data. Supported plots include spectrograms and noise statistics. The *Figure 4* illustrates a typical use of PAMGuide. After reading acoustic data into the system, data can be calibrated using information in the metadata or provided in accompanying documentation such as cruise reports. If calibration information is not available, the acoustic data can still be processed and analysed, but the user must take into account that the underlying data is uncalibrated and thus have a larger uncertainty. PAMGuide supports several types of plots, of which spectrograms are the most commonly used to visualise temporal variation of the signal. For statistics analysis, PAMGuide can present the full specter percentiles or the $\frac{1}{3}$ octave band level percentiles to enable users to analyse statistics of the acoustic recordings.

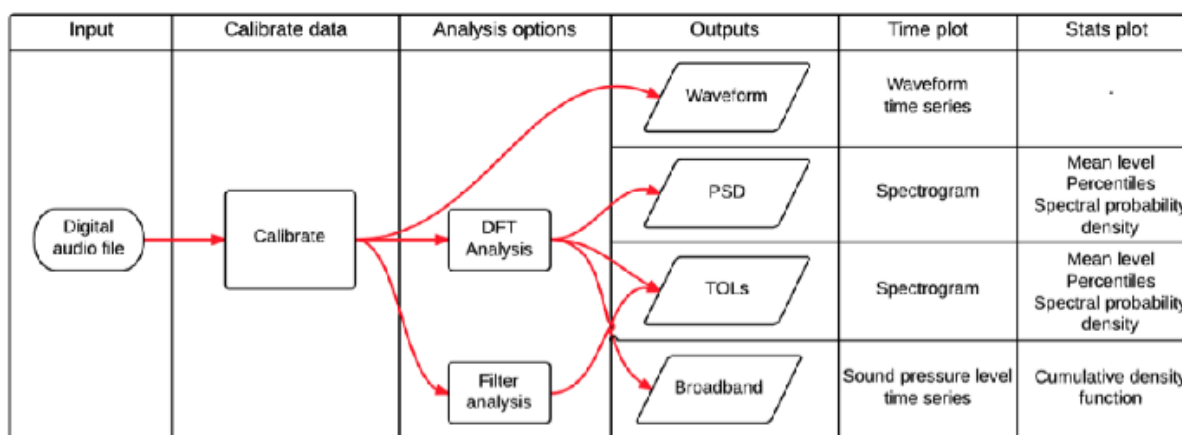


Figure 4. Data flow in PAMGuide (from supplement to Merchant *et al.* (2015))

DFT = discrete Fourier transform; PSD = power spectral density; TOL = 1/3-octave band level

The R version of PAMGuide uses the *tuneR package* to read .WAV files and analyse the data.

In addition, the Characterization of Passive Acoustic Data (C-PAD) Service uses the *ncdf4 package* (cf. *Table 2*) to read passive acoustic data from NetCDF files. This extends PAMGuide to support this widely used data format

3.5.2. IPR and licensing summary table for required software modules

The following software packages were used to develop the C-PAD iAOS Service.

Table 2. Online repository for software modules required by the C-PAD Service.

Package	Source URL	License
PAMGuide	https://sourceforge.net/projects/pamguide/	Copyright (c) 2014 The Authors.
tuneR	https://cran.r-project.org/package=tuneR	GPL-2 GPL-3
ncdf4	https://cran.r-project.org/package=ncdf4	GPL (≥ 3)

3.5.3. Online documentation resources

The following resources are recommended for scientists and service developers that want to run the C-PAD service or customise it for their needs:

- PAMGuide Tutorial:
<http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12330/supinfo>
- tuneR Reference Manual:
<https://cran.r-project.org/web/packages/tuneR/tuneR.pdf>
- ncdf4 Reference Manual:
<https://cran.r-project.org/web/packages/ncdf4/ncdf4.pdf>

3.6. Cloud resources allocation

3.6.1. Data resources

The C-PAD service does not use any satellite EO data. This service analyses and visualises passive acoustic data.

Sample passive acoustic data from the WIFAR (Waves-in-Ice Forecasting for Arctic Operators) project are provided by the Nansen Center (NERSC), and sample acoustic data from INTAROS is provided by CNRS. PerenniAL Acoustic Observatory in the Antarctic Ocean (PALAOA) data are collected by PANGAEA (AWI).

- WIFAR: Waves-in-Ice Forecasting for Arctic Operators; data collected by NERSC; <https://www.nersc.no/project/wifar>
- INTAROS: Integrated Arctic Observation System; data collected by CNRS; <https://catalog-intaros.nersc.no/dataset/passive-acoustic-data-from-kongsfjorden-svalbard-september-2018>
- PerenniAL Acoustic Observatory in the Antarctic Ocean (PALAOA); data collected by PANGAEA / AWI; <https://doi.pangaea.de/10.1594/PANGAEA.773610>

3.6.2. ICT resources in the iAOS cloud platform

The C-PAD service has been developed using Jupyter Notebook, and a first beta version has been successfully tested in the Sandbox of the iAOS cloud platform. A required step is to install PAMGuide and the needed libraries in the developer environment on the iAOS cloud platform. The “Ellip Notebooks” and the “Ellip Workflows” solutions available on the iAOS Cloud Platform are accessible by the INTAROS partners for this purpose.

3.7. Roadmap for future development

The C-PAD Service is available for use and further development in WP6, in particular with partners in Task 6.3 (“Ice-ocean statistics for decisions support and risk assessment”). The service can be used to analyse additional acoustic datasets depending on the needs emerging from this task. A possible extension to the service is to store the data underlying the spectrogram and noise statistics plots in NetCDF files. This will enable applications under development in WP6 to integrate these datasets and combine with other data from INTAROS as well as other projects.

4. TS-SIC Service Specification

Note: this section is based on the previous INTAROS D5.7 report (Integration of new processing services V1), but updated to reflect the current status and future plans for the service.

The iAOS Service “Time Series of Sea Ice Concentration (TS-SIC)” enables iAOS users to extract a time series of sea ice classification maps for a selected time period in the Fram Strait and North of Svalbard. The data in the time series can be used to generate different statistics for sea ice concentration in the region. The generated data products address maps of monthly sea ice concentration for the selected areas.

4.1. Actors interacting with the service

The actors interacting with the service are defined as follows:

Scientist - sea ice scientist working in the INTAROS project.

Decision-maker (offshore operations) - user of service, to be invoked through WP6 (Task 6.3).

Stakeholder - member of the wider sea ice science and offshore operations communities.

4.2. User Stories

User Stories are centered on expected results and benefits from the perspective of the system actors and their roles in operating the system. From there, Use Cases (cf. next section) can be more granular, and describe how the system will act to match the expectations.

User Story 1: As a scientist, John would like to generate sea ice statistics for certain months of the year, to support risk assessment for offshore operators in Arctic waters.

1. John selects the service for sea ice statistics
2. John defines the area of interest, and the desired start and end date
3. iAOS searches for **Sentinel-1 SAR images** within the given area and time range
4. iAOS presents list of the images that satisfy the search criteria
5. John selects which of the available images to classify
6. John gives input for the statistics (point-radius or transect, time range, ...)
7. iAOS runs the sea ice statistics service for the selected SAR images
8. iAOS presents the resulting sea ice statistics
9. John selects to store the results through iAOS, and gets a DOI for the new dataset (optional)
10. John selects to download the results (optional)

User Story 2: As a scientist, John would like to generate sea ice statistics for certain months of the year, to support risk assessment for offshore operators in Arctic waters.

1. John selects the service for sea ice statistics
2. John defines the area of interest, and the desired start and end date
3. iAOS searches for CMEMS sea ice concentration products within the given area and time range

4. iAOS presents list of the **CMEMS SIC products** that satisfy the search criteria
6. John gives input for the statistics (point-radius or transect, time range, ...)
7. iAOS runs the sea ice statistics service for the selected CMEMS SIC maps
8. iAOS presents the resulting sea ice statistics
9. John selects to store the results through iAOS, and gets a DOI for the new dataset (optional)
10. John selects to download the results (optional)

4.3. Use Cases

Use Case specifications are meant to capture a system actor point of view (a user, a sub-system) while describing functional requirements. They describe the step by step process expected, for an actor interacting with the system being specified, to complete a goal in using the software system.

The approach for this service is based on a simple interaction sequence where the user can read input data (sea ice concentration) and generate and publish statistics for a selected area.

The user is supported by the application to select a geographic area and a time range (month) of interest. The service is implemented as a Jupyter Notebook, and is made available from an online git repository.

Use Case ID: TSSIC-1

Use Case Name: **Generate Monthly Sea Ice Statistics**

Primary Actor: Scientist

Stakeholders and Interests: Decision-maker, Stakeholder

Preconditions: Scientist is identified and authenticated. Scientist is a registered user of Zenodo (<https://zenodo.org/>).

Postconditions: Average sea ice concentration is computed for a user defined area and time period.

Main success scenario:

1. Scientist selects a geographic area and a time range of 1 month
2. Service downloads sea ice concentration data for given area and time range
3. Service calculates average sea ice concentration for this area and period
4. Service displays the computed sea ice statistics
5. Scientist selects to store the sea ice statistics through iAOS
6. Scientist selects Zenodo as data centre and provides login information
7. Service generates metadata to accompany the sea ice statistics
8. Service forwards the sea ice statistics and metadata Zenodo, using the provided login information of Scientist
9. Service displays the DOI for the newly registered dataset of sea ice statistics

4.4. Service implementation and examples of use

NERSC has implemented the TS-SIC service in Python. TS-SIC draws upon open source libraries for reading and manipulating scientific data in NetCDF files. The choice of open source software enables other scientists and service developers to quickly put the service into use, without committing resources to commercial tools.

4.4.1. Main features of the TS-SIC Service

The TS-SIC service generates sea ice concentration statistics for a user defined geographic area and time period. The main features of the current version of the service include:

- TS-SIC utilises the CMEMS product “Svalbard ice chart” provided by the Meteorological Institute of Norway, which is a 1km daily product with sea ice concentration (%) for the Svalbard and Barents Sea region. This CMEMS product is available from 4 January 2010 to present.
- The area supported by TS-SIC is limited to 80°W-80°E ; 60°N-85°N.
- The sea ice concentration statistics are available on a monthly basis.
- The sea ice concentration statistics result is provided as a NetCDF file.
- A simple plot of the monthly mean sea ice concentration is provided in a PNG file.

This realizes a subset of User Story 2 described in section 4.2, demonstrating the use of CMEMS products to generate a monthly statistics for sea ice concentration in the area given above. Based on this product, the TS-SIC service can be used to generate monthly statistics for a period of more than 10 years.

4.4.2. Examples of sea ice statistics products generated using TS-SIC

The TS-SIC service was used to generate monthly mean sea ice concentration (SIC) fields using the CMEMS daily SIC product for the Svalbard region. The *Figure 5* and *Figure 6* show the mean SIC fields for June - September 2018 and June - September 2019, respectively. The plots illustrate how the sea ice conditions can vary significantly from one year to the next. In summer of 2018, large areas north of Svalbard and towards Franz Josef Land were ice free already in June. While in 2019 the ice conditions were much more severe in this region, with heavy ice throughout the summer.

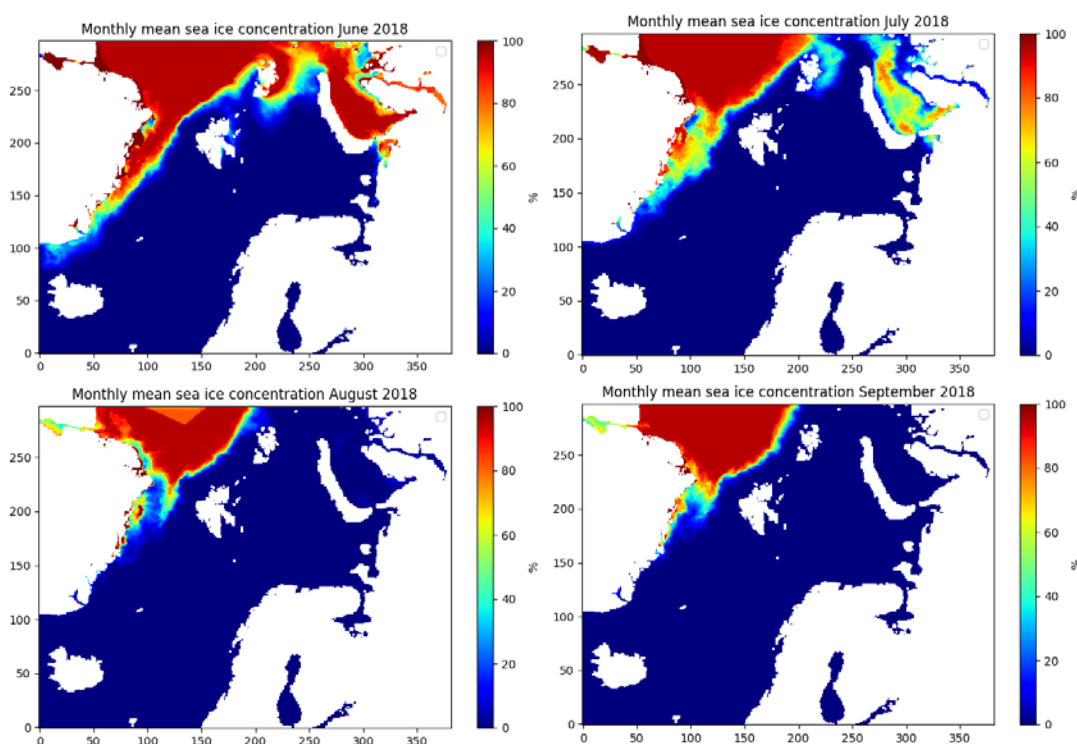


Figure 5. Examples of sea ice concentration statistics, monthly means for June - September 2018 in the Fram Strait of area north of Svalbard, generated from daily CMEMS products.

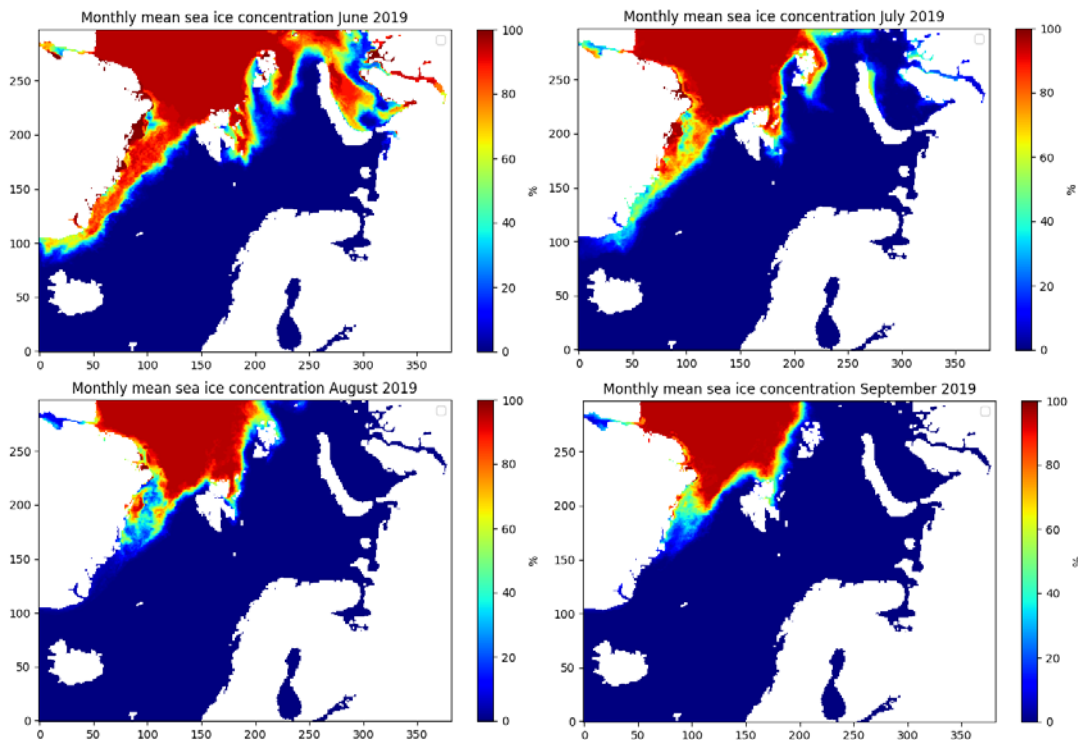


Figure 6. Examples of sea ice concentration statistics, monthly means for June - September 2019 in the Fram Strait of area north of Svalbard, generated from daily CMEMS products.

4.4.3. How to access the TS-SIC Service

The TS-SIC Service can be used by the INTAROS partners in WP6 to generate mean sea ice concentration, for user-defined periods between January 2010 and present.

Partners can modify the service to generate statistics for longer periods than one month or for seasonal means for the last decade. Source code for the service (*Table 3*) and the software it depends on (*Table 4*) is publicly available.

4.4.4. IPR and licensing summary table for TS-SIC Service

The software package for the TS-SIC Service is available from a public git repository (*Table 3*). Besides source code, this repository provides a description of dependencies and installation instructions.

Table 3. Online repository for the TS-SIC Service.

Package	Source URL	License
TS-SIC	https://github.com/ec-intaros/enb-sea-ice-concentration	CC BY 4.0

4.5. Software modules required by the TS-SIC service

The TS-SIC Service is developed using widely used open source tools for manipulating scientific data in NetCDF files (netcdf4) and for plotting data held in NetCDF (matplotlib). Both of these tools have extensive online documentation and tutorials, as well as an active community helping programmers in efficiently using these tools. Both support Python 3.

4.5.1. Required software

This service has been developed using Python 3 and Jupyter Notebook, using open source tools netcdf4 and matplotlib. netcdf4 is used for reading NetCDF files provided by an OPeDAP server and writing NetCDF files for the generated monthly mean sea ice concentration fields. matplotlib is used for generating and storing plots of the monthly fields in PNG files.

4.5.2. IPR and licensing summary table for required software modules

The following software packages were used to develop the TS-SIC iAOS Service.

Table 4. Online repository for software modules required by the TS-SIC Service.

Package	Source URL	License
netcdf4	https://github.com/Unidata/netcdf4-python	Copyright 2008 by Jeffrey Whitaker
matplotlib	https://matplotlib.org/	https://matplotlib.org/users/license.html (based on PSF - Python Software Foundation License)

4.5.3. Online documentation resources

The following resources are recommended for scientists and service developers that want to run the TS-SIC service or customise it for their needs:

- netcdf4 module documentation: <https://unidata.github.io/netcdf4-python/netCDF4/index.html>
- Open-source Project for a Network Data Access Protocol (OPeNDAP) <http://www.opendap.org/>
- matplotlib tutorials: <https://matplotlib.org/tutorials/index.html>
- CMEMS PRODUCT USER MANUAL (CMEMS-SI-PUM-011-002) <https://marine.copernicus.eu/documents/PUM/CMEMS-SI-PUM-011-002.pdf>
- CMEMS QUALITY INFORMATION DOCUMENT (CMEMS-SI-QUID-011-001TO007-009TO014) <https://marine.copernicus.eu/documents/QUID/CMEMS-SI-QUID-011-001to007-009to014.pdf>

4.6. Cloud resources allocation

4.6.1. Data resources

This service uses Sentinel-1 SAR based sea ice concentration products from CMEMS.

4.6.2. ICT resources in the iAOS cloud platform

The TS-SIC Service has been developed using Python and Jupyter Notebook. It is easily installable and usable in the iAOS Cloud Platform. The “Ellip Notebooks” and the “Ellip Workflows” solutions available on the iAOS Cloud Platform are accessible by the INTAROS partners for this purpose. This enables other scientists to reuse the Notebook for generation of sea ice concentration statistics for areas and time periods of their interest.

4.7. Roadmap for future development

The TS-SIC Service is available for use and further development in WP6, in particular with partners in Task 6.3 (“Ice-ocean statistics for decisions support and risk assessment”). The service as implemented in D5.7 can be used to generate monthly means of sea ice concentration in the Svalbard region for inclusion in data analysis in WP6. A possible extension to the service is to integrate SIC products from other data providers to be able to extend the area or time range of the generated statistics.

5. TS-CTD Service Specification

The iAOS Service “Time Series of Conductivity Temperature Depth (TS-CTD)” enables iAOS users to extract a time series of CTD measurements for a selected time period in the Arctic. The data in the time series can be used to generate different statistics for the water column or sea bottom, in particular when ingested as input to a RGeostats module.

The Conductivity Temperature Depth (CTD) datasets in scope of this iAOS service are provided by the INTAROS partner IMR. The datasets encoding follows the Copernicus standards: they are encoded server side following the netCDF-4 classic model, with information content structured according to the following conventions:

- CF-1.6 Copernicus-InSitu TAC-Format Manual-1.41
- Copernicus-InSitu TAC-SRD-1.5
- Copernicus-InSitu TAC-Parameters List-3.2.0
- OceanSITES vertical profile

The dataset collection involved acquisitions from CTD Seabird instruments onboard of several research vessels:

- platform_code: 58AA (Håkon Mosby, Institute of Marine Research, Norway)
- platform_code: 58GS (G.O. Sars, Institute of Marine Research, Norway)
- platform_code: 58GT (Sarsen, Institute of Marine Research, Norway)
- platform_code: 58HJ (Helmer Hanssen, University of Tromsø, Norway)
- platform_code: 58JH (Johan Hjort, Institute of Marine Research, Norway)

The dataset collection is public and free of charge. The user assumes all risks in the use of the data.

5.1. Actors interacting with the service

The actors interacting with the service are defined as follows:

Application developer - user of the service, to be involved through WP6 for the development of any iAOS Showcase Application leveraging CTD data products.

Scientist - marine ecosystems, fish resources management or environment management scientist working in the INTAROS project.

Stakeholder - member of the wider arctic science and offshore operations communities.

5.2. User Stories

User Stories are centered on expected results and benefits from the perspective of the system actors and their roles in operating the system. From there, Use Cases (cf. next section) can be more granular, and describe how the system will act to match the expectations.

The service is designed for accessing, filtering, analysing and visualising CTD data from a remote OPeNDAP server. Two Jupyter Notebooks are provided, a reference (REF) one, and a DAP queries optimisation (DAP) one. They can be executed both during a same work session, in order to support comparisons while looking for optimizing OPeNDAP Queries.

User Story 1: As an application developer, Mary would like her application to extract user-defined temperature profiles, over selected geographic areas, from CTD data hosted on a remote OPeNDAP server.

1. Mary connects with her Ellip user account to the iAOS Cloud Platform to access the Ellip Notebooks solution, and loads the two Jupyter Notebooks REF and DAP from their reference GitHub repositories.
2. Mary interacts with the Operations Support team to request the configuration of sample CTD datasets (e.g. representative months and platforms) from the operational OPeNDAP server she wants to use from her target software application.
3. Mary performs the setup operations on both the REF and DAP Notebooks
4. Mary runs the load and Plot of CTD Positions only (from DDS information) on both the REF and DAP Notebooks.
 - a. This part of the Notebooks leverages the NetCDF data structure as exposed by an OPeNDAP server, and allows to perform an initial data discovery query, that will deliver a lightweight data structure (the OpenDAP's DDS Dataset Descriptor Structure information) useful for the application developer to visually analyse and understand the dataset at hand (geographic and temporal extends covered, eventual data gaps, ...).
5. Mary configures and runs on the REF Notebook the load and plot functions for needed CTD Data Variables and Attributes, downloading initially the full sample CTD dataset, and then filtering locally to the Notebook the data structures of interest to generate reference Plots of these variables.
6. Mary finally configures and runs on the DAP Notebook the load and plot functions for needed CTD Data Variables and Attributes, downloading only the optimised data subsets from the full sample CTD dataset, and directly creating the related optimised data structures of interest to generate control Plots of these variables, to be compared with the reference plots previously created on the REF Notebook.
7. Mary optimises the DAP Notebook query filters and validates the functions to be integrated in her target software application.
8. Mary integrates the validated DAP Notebook query filter functions in her target software application, to deliver optimised download times and data structures to the data processing module of her application.

In the *Figure 7* we summarise the approach for supporting users of the iAOS Cloud Platform in optimizing their OPeNDAP queries to remote repositories exposing that interface.

Data has been prepared and uploaded on the iAOS OPeNDAP reference servers (hosted on the iAOS Cloud Platform domain on Terradue Cloud Platform), from where it can be accessed directly online via the OPeNDAP protocol.

A subset of NetCDF data has been chosen from available CTD data acquisition campaigns collected in the year 2003.

Note: other datasets can be copied on demand from the [NODC source](#), in order to support INTAROS partners willing to perform tests while integrating their application as a Jupyter Notebook or RGeostats program.

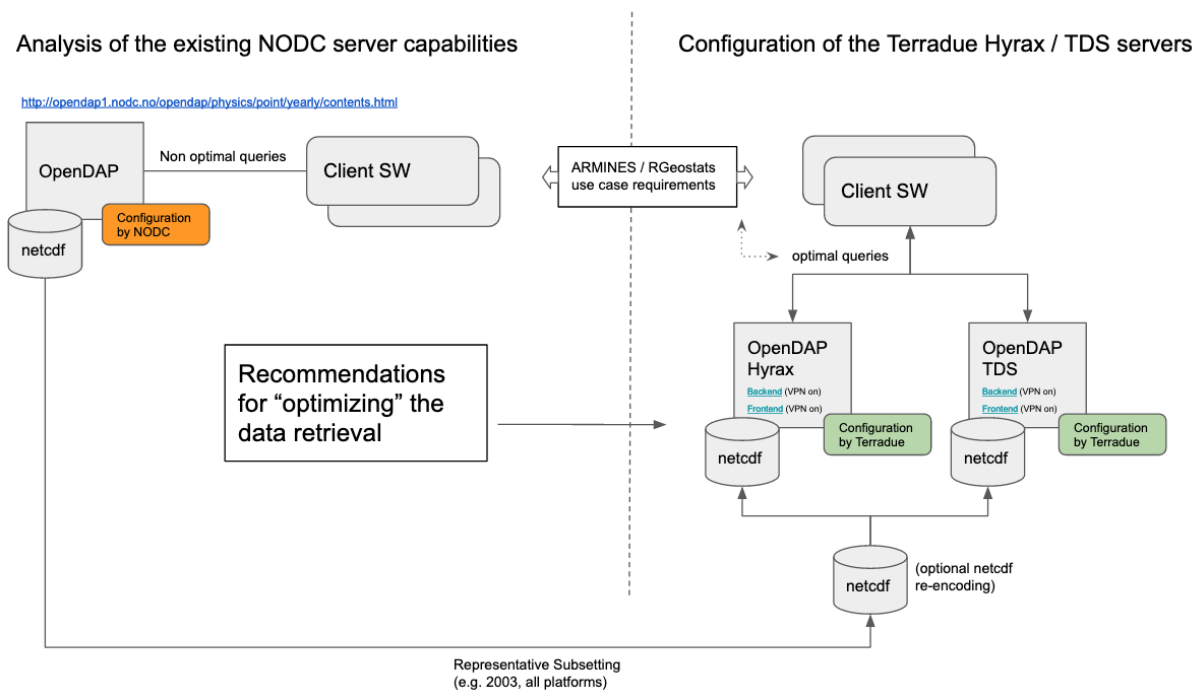


Figure 7. TS-CTD service and Cloud Platform support to implement, test and validate remote data access requirements for CTD data products

The datasets in use for the initial definition of the TS-CTD service are Arctic Ocean - In Situ Observation prepared by the INTAROS partner IMR (Institute of Marine Research, Norway) for the CMEMS Copernicus service.

Two options are provided within the provided Notebooks to access CTD data products via remote calls to an OPeNDAP interface, one for the HYRAX Server technology and one for THREDDS Data Server (TDS) technology, as they are the two leading implementations on the OPeNDAP standard and there are some not well documented differences in their query capabilities:

iAOS Hyrax Server instance

- https://opendap.terradue.com/hyrax/data/subset_2003/
 - Access point for Ellip Solutions registered users
 - VPN connexion to the iAOS Cloud Platform required

iAOS Thredds Data Server instance

- https://opendap.terradue.com/thredds/dodsC/subset_2003/
 - Access point for Ellip Solutions registered users
 - VPN connexion to the iAOS Cloud Platform required

For the datasets, the year and the platform codes are defined.

This information is needed, and is known a priori, as it allows completing the service access URL to access each specific NetCDF file.

The naming convention of the NetCDF files is:

58<platform_code>_CTD_<year>.nc.nc4

For example, for the platform 'GT' and year 2003, the name of the NetCDF file is:

58GT_CTD_2003.nc.nc4

See Section 9. Annex “WP6 references to iAOS Showcase applications” for the selected usages of the iAOS Cloud Platform that can benefit from the TS-CTD service to address scientific questions.

5.3. Use Cases

The two Notebooks are labelled as follows:

1. **Test1 REF - OPeNDAP Xarray use cases:** a reference notebook for the download of full data files (data bulks as they are encoded server side), and storing their content in memory for processing: the bulk data structure can be further filtered to extract data chunks related to an analysis goal, using Python functions, and with plotting for data visualisation;
2. **Test1 DAP - OPeNDAP Xarray use cases:** a Data Access Protocol (DAP) Notebook for the validation of OPeNDAP query filters, in order to download optimized data chunks related to an analysis goal. These data chunks are then plotted for data visualisation.

Both the REF and DAP Notebooks follow the same use case:

Use Case ID: TSCTD-1

Use Case Name: **Compare plots of filtered CTD data for optimized queries**

Primary Actor: Application Developer

Stakeholders and Interests: Scientist, Stakeholder

Preconditions: Application Developer is identified and authenticated on the iAOS Cloud Platform, as a registered user of the Ellip Solutions.

Postconditions: validated DAP Notebook for a selected, representative sample dataset and OPeNDAP Server technology.

Main success scenario:

1. Set-up of the Server URL, Year and Platform Codes
2. Load and Plot of CTD Positions only (from DDS information)
 - a. This part of the Notebooks leverages the NetCDF data structure as exposed by an OPeNDAP server, and allows to perform an initial data discovery query, that will deliver a lightweight data structure (the OpenDAP’s Dataset Descriptor Structure information)
3. Load and Plot all needed CTD Data (from Variables and Attributes)

In each Notebook, two examples of data filtering queries are pre-defined, describing how to visualise different data ranges:

- extraction of data across the first elements of DEPTH measurements, e.g. 0-20;
- extraction of data across the last elements of DEPTH measurements, e.g. 50-last.

Application developers are supported with these two examples to compare the two ways to querying data (Bulk and Optimized) in order to learn how to exploit efficiently (and without mistake) all the OPeNDAP server-side capabilities for optimized data filtering and retrieval across a network connection.

Application developers are also supported with these 2 notebooks to create and validate new use cases of CTD data structure extraction, matching their specific application software needs and validated for the specific OPeNDAP remote server technology that serves their CTD data

of interest, and ultimately incorporate their validated DAP Notebook as the data access module of their software application.

5.4. Service implementation and examples of use

The service is available as a set of two Jupyter Notebook files (.ipynb) downloadable from the INTAROS GitHub organisation, within the repository “opendap-xarray-use-cases”: <https://github.com/ec-intaros/opendap-xarray-use-cases>

The two data processing Notebooks are executable within a JupyterLab environment, such as the Ellip Notebooks one, provided to the INTAROS partners as part of the iAOS Cloud Platform.

5.4.1 REF Notebook exploiting the CTD 2003 subset

Jupyter Notebook file: **Subset 2003 Analysis with Xarray**
<https://github.com/ec-intaros/opendap-xarray-use-cases>

- **Test1 REF - Subset2003 Analysis.ipynb**

This Notebook provides an overview, as well as practical examples, for online access and analysis of netCDF data files holding CTD data acquisitions.

It loads full data structures in memory into potentially large Python xarray structures and performs reference plotting of the retrieved variables after local filtering.

The recommended approach to generate analysis content with such a REF Notebook is to proceed from selected data subsets copied on the Terradue OpenDAP servers.

The first steps are to specify the data source (hyrax or thredds) available on the iAOS Cloud Platform’s Data Agency, and perform the retrieval of the DDS information (OpenDAP’s Dataset Descriptor Structure).

After the set-up, the data dimensions can be accessed through the DDS, to understand the size of the datasets in an efficient way, without downloading all data into local memory. The dimensions under consideration for these NetCDF files are: 'TIME', 'LATITUDE', 'LONGITUDE', 'DEPTH', 'POSITION'.

Exploratory Data Analysis: create Position Dictionary and Dataframe

The position_dict is a dictionary that is generated by iteratively reading the URL of each platform, selecting only with 'TIME', 'LATITUDE', 'LONGITUDE' dimensions. In the dictionary are saved: the actual data, loaded into an xarray for data handling, analysis and visualisation;

and the campaign's main attributes: platform code & name, data type, title, instrument, longitude & latitude, and vertical min & max.

A position dataframe is then generated per each platform, to match and combine the three 'TIME', 'LATITUDE', 'LONGITUDE' dimensions for each measurement. Subsequently, the platform-specific dataframes are combined in a position_df dataframe.

The positions of all measurements of the given platforms are visualised in different colours on an interactive plot (Figure 8). By hovering on the Jupyter Notebook, on the coloured dots, information such as index, x&y, and platform can be easily retrieved and visualised, as shown as an example in the figure.

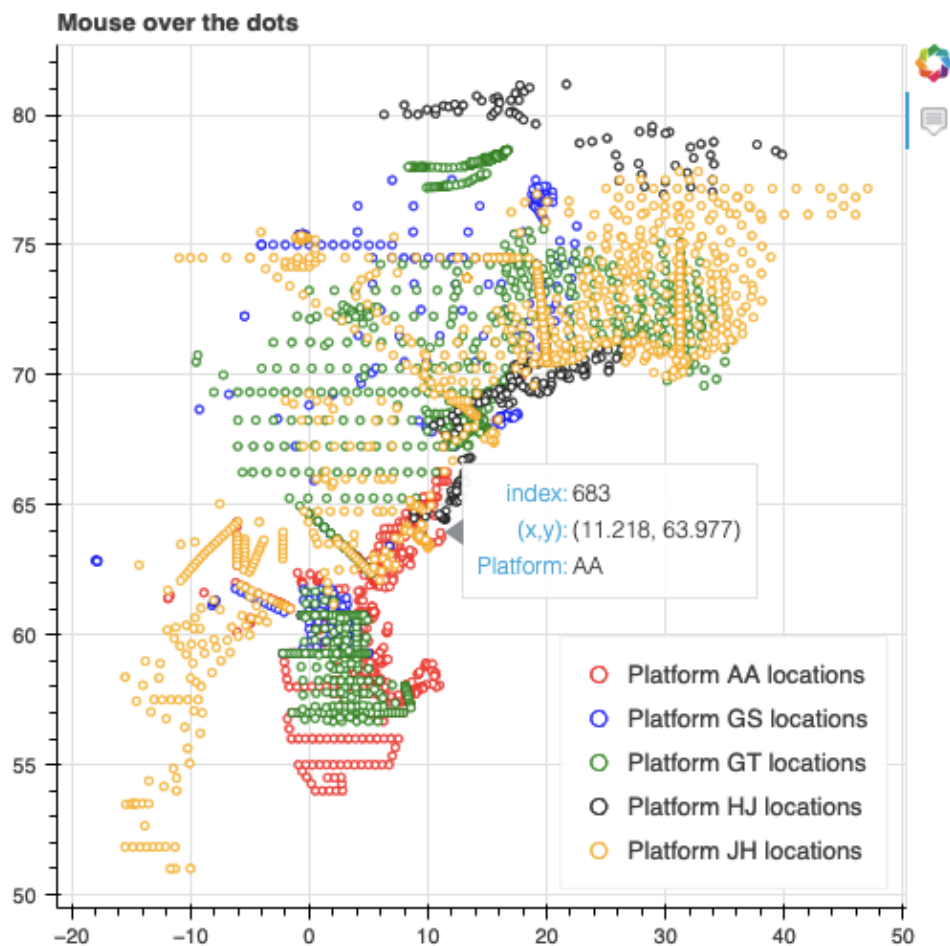


Figure 8. Positions generated from the OpenDAP's Dataset Descriptor Structure (DDS)

Filter positions

The next step is to filter positions. The following filters are included in the provided Notebook, and can be modified or extended by an application developer:

- By bounding box (BBOX)
- By BBOX and Month of collection
- By BBOX and Time (eg hour) of collection
- By customised time range

Going deeper into the data analysis, the next step is accessing all data (variables and their attributes). This is therefore more computationally-demanding, especially if multiple datasets of many measurements are analysed at once.

Fetch Data & Create Data Dictionary (*data_dict*)

All data (variables and their attributes) for each platform are read iteratively, and saved into a dictionary *data_dict* which contains:

- the actual data, loaded into an xarray for data handling, analysis and visualisation
- the campaign's main attributes: *platform code & name, data type, title, instrument, longitude & latitude, and vertical min & max*)

An overview dataframe *overview_df* is then generated to show the detailed information about each campaign at sea: *platform code & name, data type, title, instrument, longitude & latitude, and vertical min & max*).

The four variables that are available in this sample dataset are:

- PRES: Sea Water Pressure
- TEMP: Sea Water Temperature
- PSAL: Sea Water Practical Salinity
- CNDC: Sea Water Electrical Conductivity

Then, the filtered dataframe (eg *position_df_bbox, position_df_bbox_mm, position_df_bbox_hh*) to use for the analysis is assigned to the *df_toPlot* variable. The parameters to be defined are:

- Platform Code
- Selected Variables
- DEPTH range (e.g. 0-20)

The filtering is done for all variables of the defined platform and DEPTH range, to the previously-filtered data (eg by "BBOX", by "BBOX and Month", or by "BBOX and Hour"). The following are the two types of filters that are given as examples:

- Filtered data by BBOX and One Variable
- Filtered data by BBOX (All Variables)

The output of both filters is a *filtered_xarr* xarray dataset, containing one or all the variables within the specified DEPTH range. The reference plots are generated for the available variables of the filtered xarrays. On the y-axis is shown the TIME¹ of the measurement (in float format, which needs to be converted to datetime format), and on the x-axis is the DEPTH of the measurement.

This notebook shows how to generate two types of plots:

- Plotting individual Variables per individual Platform
- Plotting individual Variables across aggregated Platforms

¹The time axis follows the netcdf convention applied to the dataset by the data provider. These are the values accessible from the data records.

In the parameters set associated to each of these figures (top line), the corresponding timeframe is indicated (e.g. Year=2003)

The first plot is more straightforward, as it automatically generates the plot(s) of the variable(s) that has(have) been generated in the Data Filtering section. The example plots below (*Figure 9*) represent the Sea Water Temperature and Pressure between 0-20 meters for the selected locations of the platform GT.

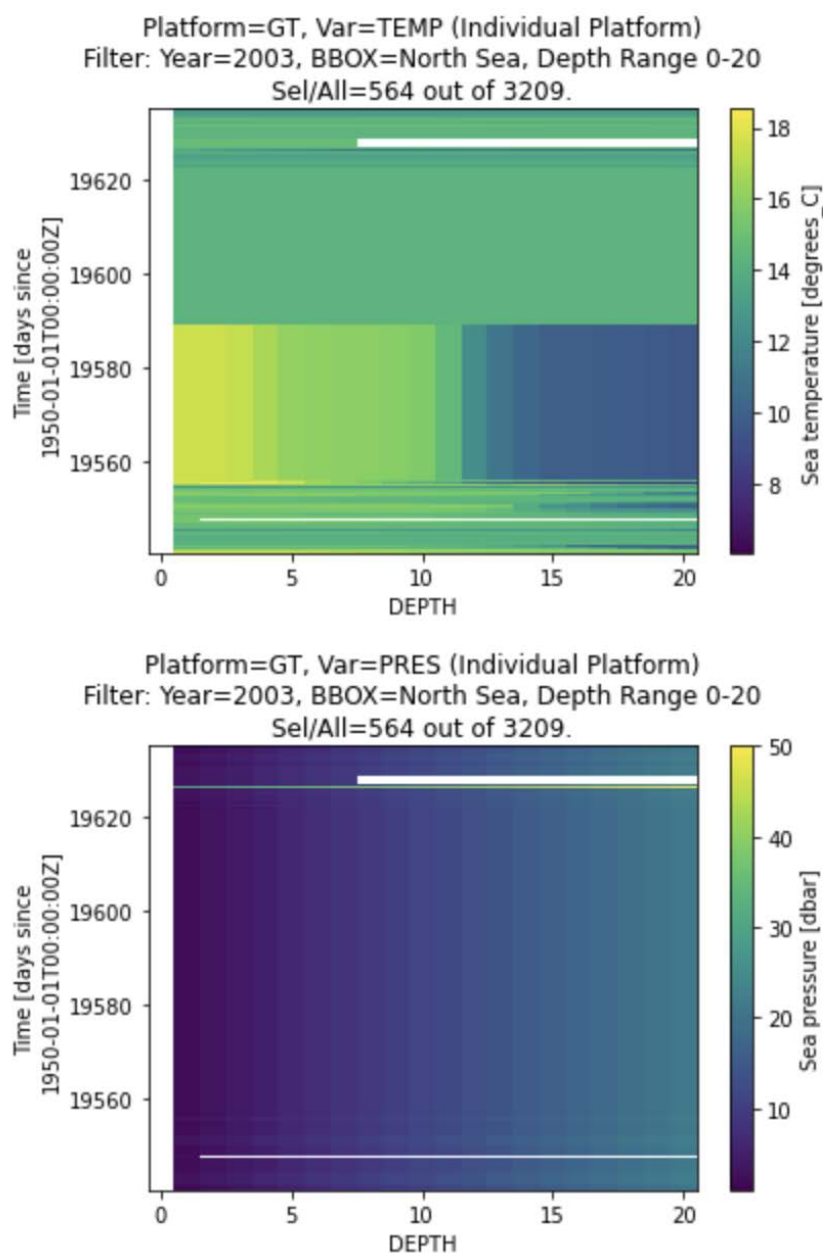


Figure 9. REF Notebook - Sea Water Temperature and Pressure, between 0-20 meters for the platform GT

The second plot is more complex, as it needs an additional operation before executing. This consists in generating and then aggregating all data across all platforms, for a specific variable.

To do so, the dimensions of the DEPTH of the variables of all platforms must be equal, otherwise it is not possible to combine them into a new, aggregated, xarray. Two options are provided to accomplish this:

- Aggregate with minimum DEPTH elements, i.e. use minimum common DEPTH across all platforms' DEPTHS
- Aggregate with maximum DEPTH elements, i.e. use maximum DEPTH across all platforms' DEPTHS, and fill empty values with nans

The example (*Figure 10*) plots below represent the four variables aggregated among the available platforms, between 0 and 20 meters below sea level.

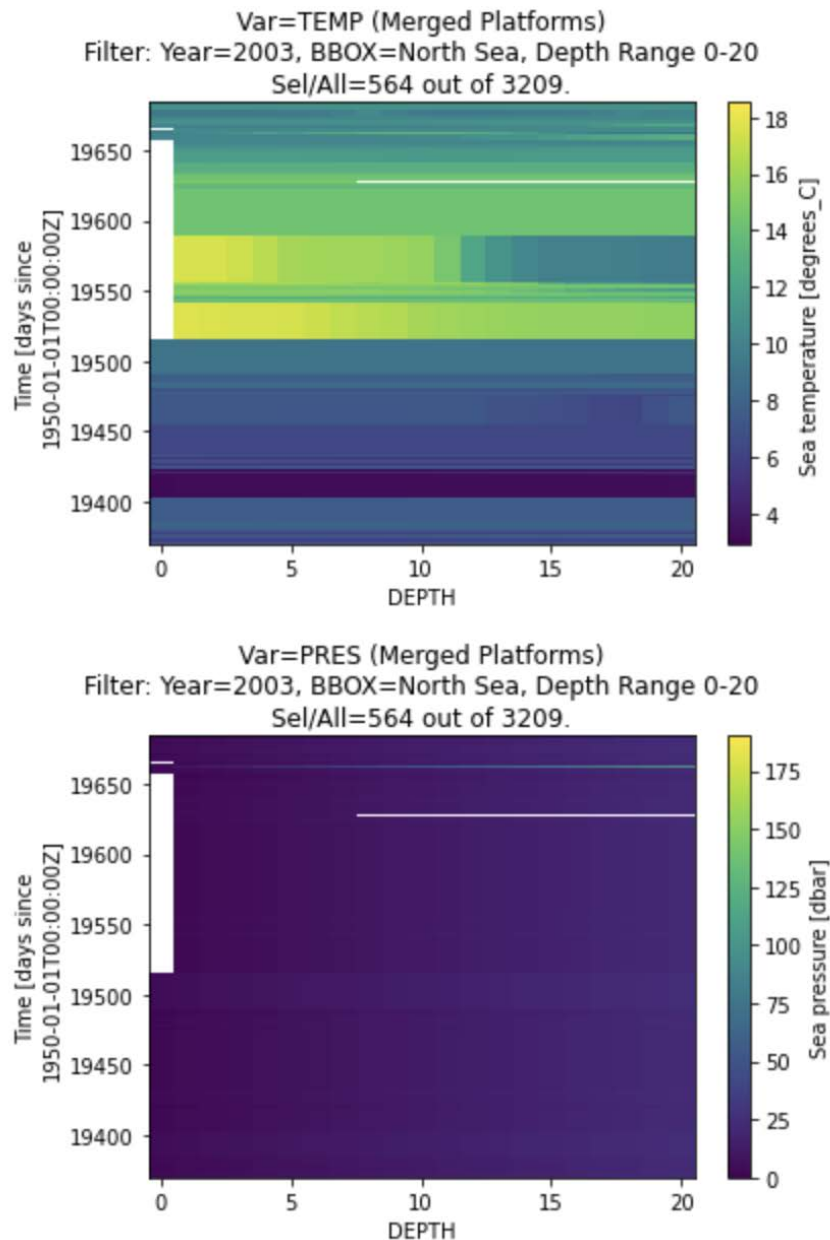


Figure 10. REF Notebook - Sea Water Temperature and Pressure, between 0-20 meters for the aggregated platforms

Another example reuses the same structure as in the 0-20 range section, but focuses on data extracted from the 50 meters onwards of DEPTH (*Figure 11*)

It plots for Sea Water Temperature and Pressure respectively, between 50-2956 meters for the selected locations of the platform GT.

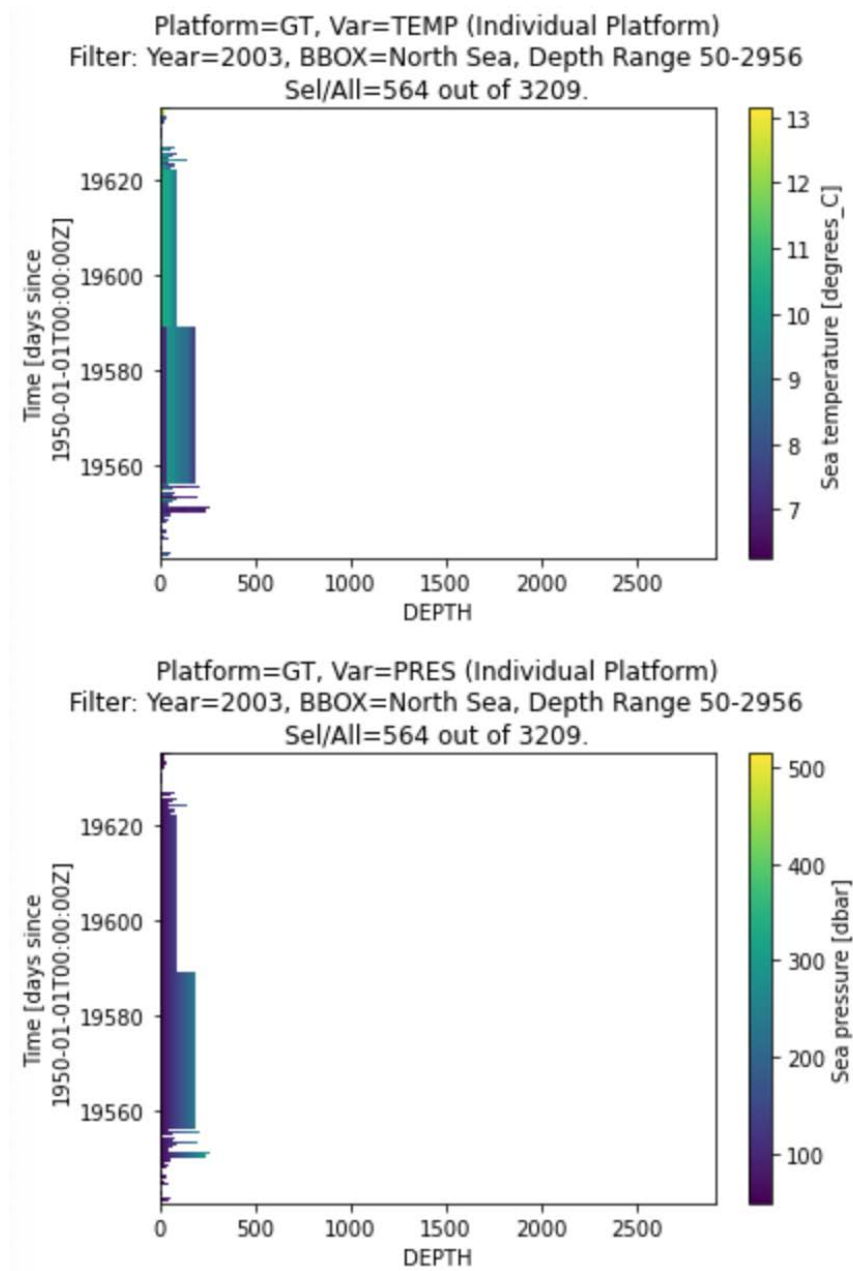


Figure 11. REF Notebook - Sea Water Temperature and Pressure, between 50-2956 meters for the platform GT

Another example (*Figure 12*) is provided with plots for Sea Water Temperature and Pressure respectively, between 50-2956 meters for the aggregated platforms.

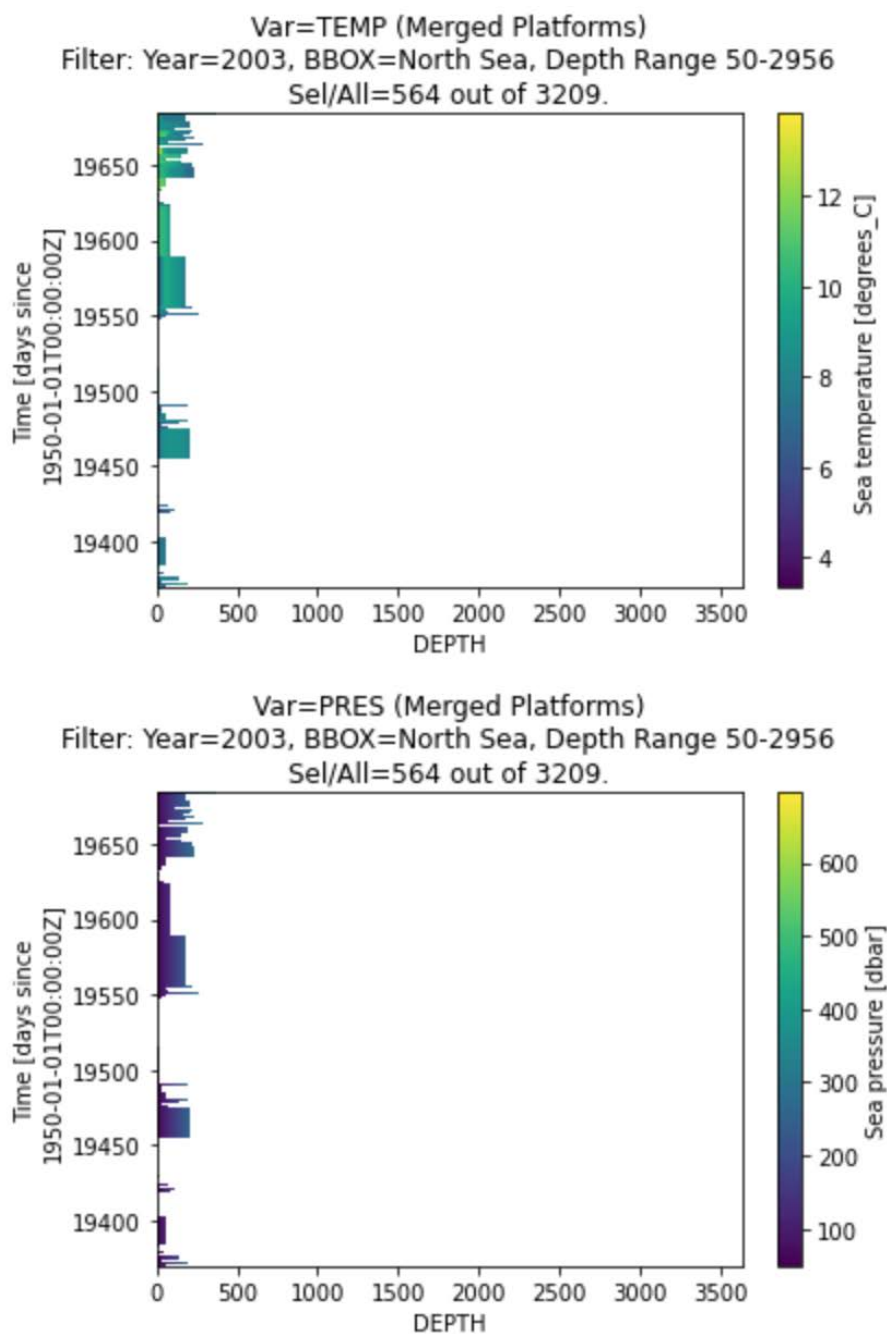


Figure 12. REF Notebook - Sea Water Temperature and Pressure, between 50-2956 meters for the aggregated platforms

5.4.2. DAP Notebook exploiting the CTD 2003 subset

Jupyter Notebook file: **Subset 2003 Analysis with Xarray**

<https://github.com/ec-intaros/.opendap-xarray-use-cases>

- **Test1 DAP - Subset2003 Analysis.ipynb**

This Notebook provides an overview, as well as practical examples, for online access and analysis of NetCDF data files holding CTD data acquisitions.

It loads server-side filtered data structures in memory into optimized Python xarray structures and performs the plotting of the retrieved variables.

The recommended approach to generate analysis content with such a DAP Notebook is to validate the Notebook queries from the selected data subsets copied on the Terradue OpenDAP servers, and then adapt these queries to the operational OpenDAP servers serving the full data collections (typically directly from the NODC source).

The first steps are to specify the data source (hyrax or thredds) available on the iAOS Cloud Platform's Data Agency, and perform the retrieval of the DDS information (OpenDAP's Dataset Descriptor Structure).

After the set-up, the data dimensions can be accessed through the DDS, to understand the size of the datasets in an efficient way, without downloading all data into local memory. The dimensions under consideration for these NetCDF files are: 'TIME', 'LATITUDE', 'LONGITUDE', 'DEPTH', 'POSITION'.

Exploratory Data Analysis: create Position Dictionary and Dataframe

This step is the same as the one presented above in the Exploratory Data Analysis section of the REF Notebook.

Load and Plot selected Data (Variables within DEPTH range)

This step is accessing data of only selected variables and within a specified DEPTH range, to avoid fetching unnecessary data and in order to minimise the data volume over the network (and therefore the duration of transfer) and the footprint of the data in memory.

The selected variables of interest need to be specified in the *var_list*. The four variables that are available in this dataset are:

- PRES: Sea Water Pressure
- TEMP: Sea Water Temperature
- PSAL: Sea Water Practical Salinity
- CNDC: Sea Water Electrical Conductivity

The DEPTH range must also be defined in this case, as data is fetched with the specific DEPTH range directly from URL. For some limitations to the DAP syntax, at least one range boundary needs to correspond to one of the two extremes. For example, in a data array of 100 elements from 0 to 99, the following scenarios are possible:

- select the first 20 elements, corresponding to the values range 0 - 19 --> [first:1:intermediate] (eg [0:1:19]) work
- select the last 20 elements, corresponding to the values range 80 - 99 --> [intermediate:1:last] (eg [80:1:99]) work
- select the intermediate 60-80 elements, corresponding to the values range 60 - 79 --> [intermediate_1:1:intermediate_2] (eg [60:1:79]) does NOT work

The next step is to create the Data Dictionary (*data_dict*). Define the Selected Variables and then the DEPTH range, noting that:

- `metadata[pc]['depth_m_v1']`: either this is equal to the lower bound (ie index=0)
- `metadata[pc]['depth_m_v2']`: or is equal to the upper bound (ie index=-1)

Once variables and depth range are defined, the data and their attributes are read iteratively for each platform, and saved into a dictionary `data_dict`, following the same approach described in the REF notebook.

Subsequently, the filtered dataframe (eg `position_df_bbox`, `position_df_bbox_mm`, `position_df_bbox_hh`) to use for the analysis is defined to the `df_toPlot` variable.

- The *Figure 13* shows the plots for Sea Water Temperature and Pressure respectively, between 0-20 meters for the selected locations of the platform GT.

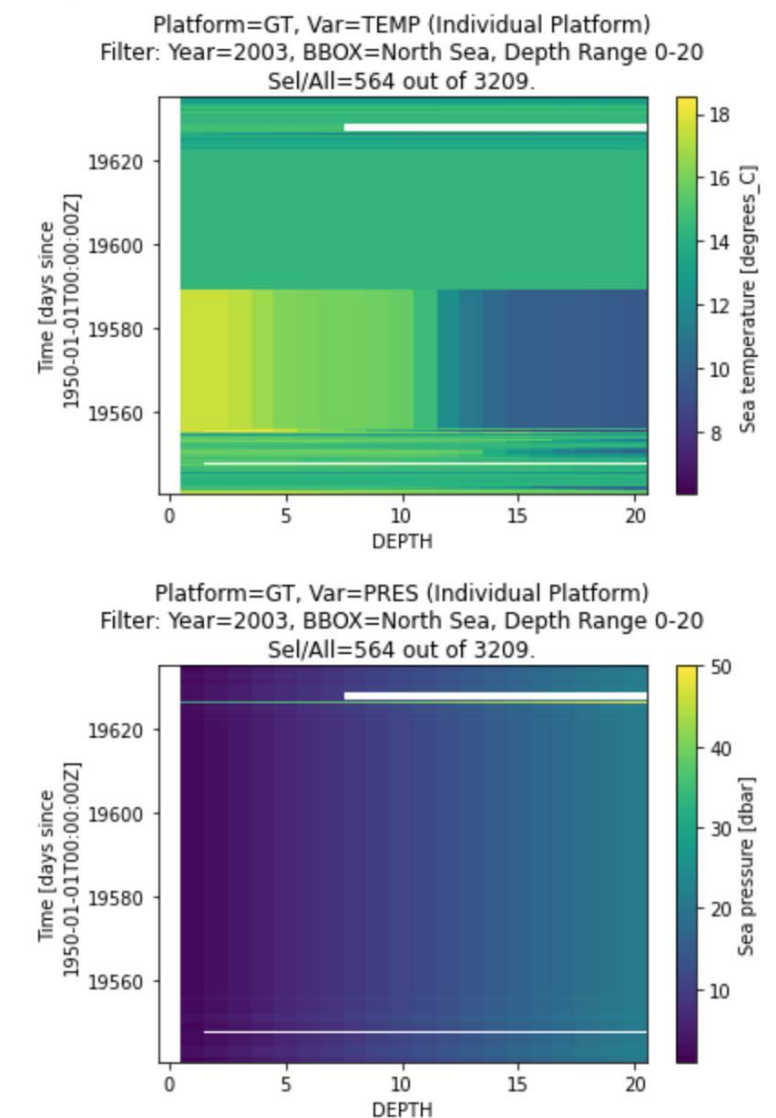


Figure 13. DAP Notebook - Sea Water Temperature and Pressure, between 0-20 meters for the platform GT

- The Figure 14 shows the plots for Sea Water Temperature and Pressure respectively, between 0-20 meters, for the selected locations of the aggregated platforms.

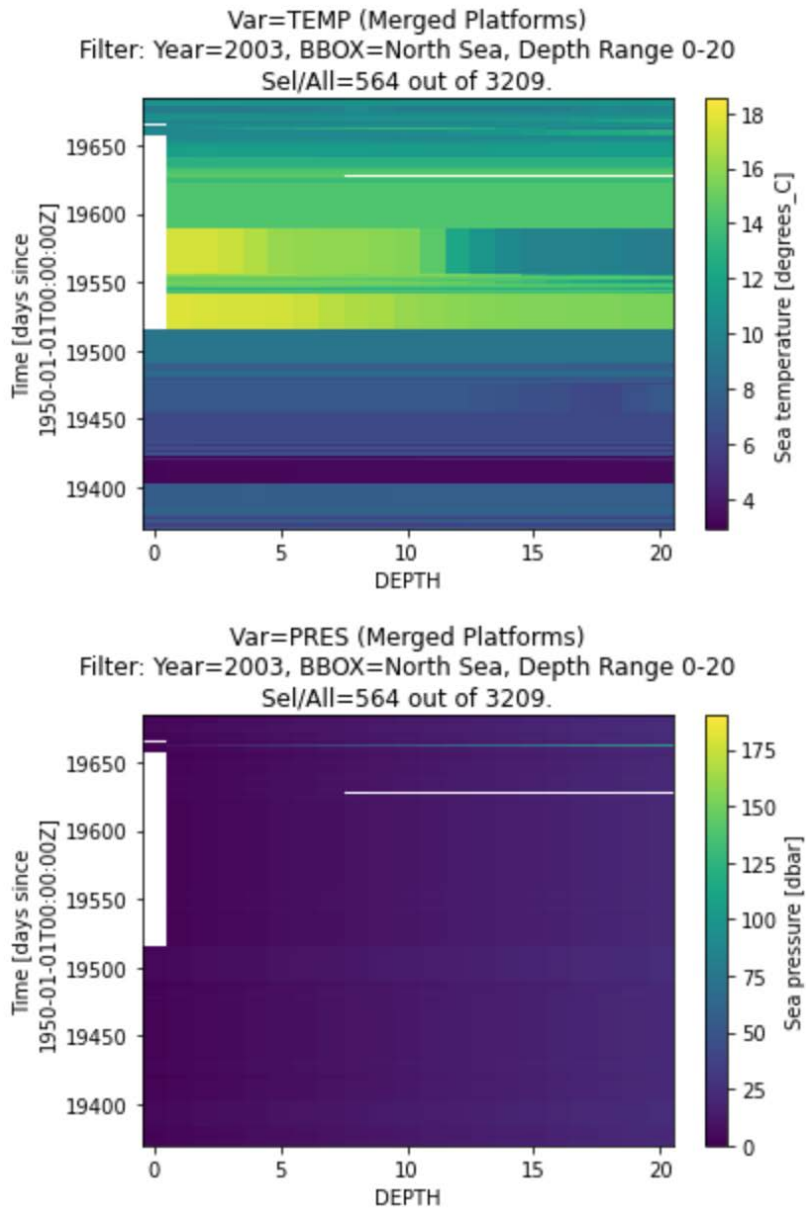


Figure 14. DAP Notebook - Sea Water Temperature and Pressure, between 0-20 meters for the aggregated platforms

The next step follows the same structure and analysis as the DAP 0-20 range section, but focuses on data extracted from the 50 meters onwards of DEPTH.

- The Figure 15 shows the plots for Sea Water Temperature and Pressure respectively, between 50-2956 meters for the selected locations of the platform GT.

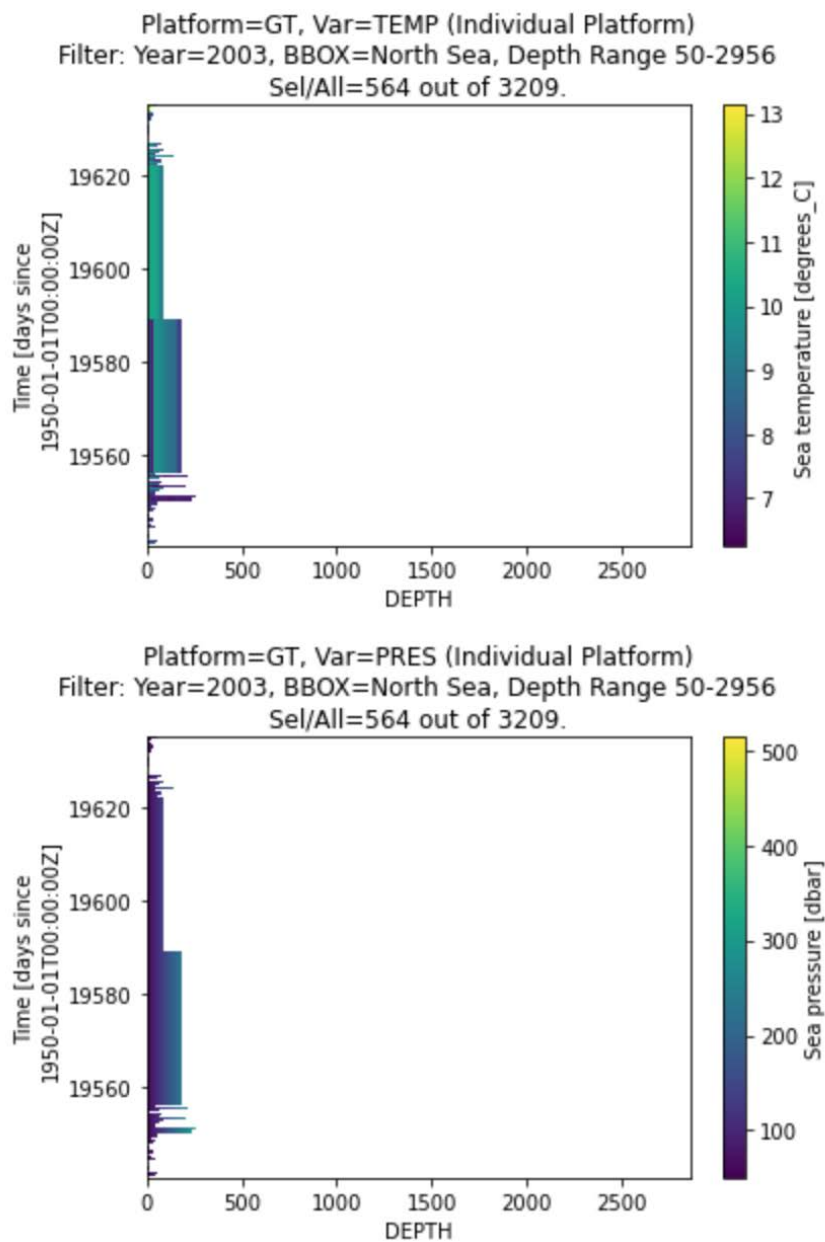


Figure 15. DAP Notebook - Sea Water Temperature and Pressure, between 50-2956 meters for the platform GT

- The *Figure 16* shows the plots for Sea Water Temperature and Pressure respectively, between 50-2956 meters for the aggregated platforms.

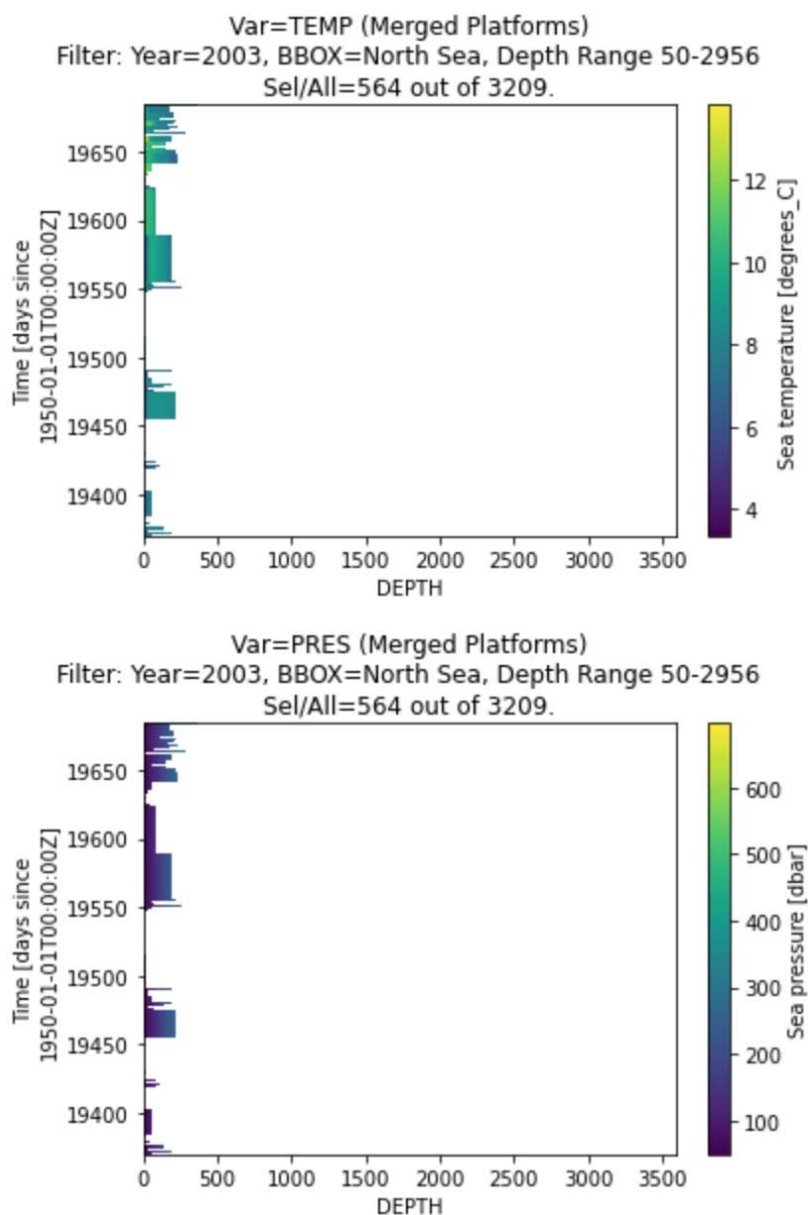


Figure 16. DAP Notebook - Sea Water Temperature and Pressure, between 50-2956 meters for the aggregated platforms

5.4.3. How to access the TS-CTD Service

The TS-CTD Service can be used by the INTAROS partners in WP6 to analyse CTD datasets and optimise their remote API calls to download CTD profiles strictly matching their application software needs. The source code for the service (*Table 5*) and the software it depends on (*Table 6*) is publicly available.

5.4.4. IPR and licensing summary table for TS-CTD Service

The software package for the TS-CTD Service is available from a public git repository (*Table 5*). Besides source code, this repository provides a description of dependencies and installation instructions.

Table 5. Online repository for the TS-CTD Service.

Package	Source URL	License
TS-CTD	https://github.com/ec-intaros/opendap-xarray-use-cases	CC BY 4.0

5.5. Software modules required by the TS-CTD service

The TS-CTD Service is developed using widely used open source tools for querying OPeNDAP server interfaces (DAP2 / DAP4), manipulating scientific data within Python data structures (panda dataframes, numpy, xarray) and for plotting data (matplotlib).

All of these tools have extensive online documentation and tutorials, as well as an active community helping programmers in efficiently using these tools. Both support Python 3.

5.5.1. Required software

This service has been developed using Python 3 and Jupyter Notebook, using open source tools pandas, numpy, xarray and matplotlib.

pandas dataframe is used for lightweight visual analysis of positions (match and merge LAT, LONG and TIME) for each sensor platform served from the OPeNDAP DDS data model. numpy is used to manipulate data structures. xarray is used to combine data arrays across platforms, for each variable, before plotting. matplotlib is used for generating and storing plots of the filtered data structures.

5.5.2. IPR and licensing summary table for required software modules

The following software packages were used to develop the TS-SIC iAOS Service.

Table 6. Online repository for software modules required by the TS-SIC Service.

Package	Source URL	License
pandas	https://pandas.pydata.org/	BSD 3
numpy	https://numpy.org/	BSD 3
xarray	https://xarray.pydata.org/	Apache License 2.0
matplotlib	https://matplotlib.org/	https://matplotlib.org/users/license.html (based on PSF - Python Software Foundation License)

5.6. Cloud resources allocation

5.6.1. Data resources

This service is tested for CTD products from IMR / NODC.

5.6.2. ICT resources in the iAOS cloud platform

The TS-CTD Service has been developed using Python and Jupyter Notebook. It is easily installable and usable on the iAOS Cloud Platform. The “Ellip Notebooks” and the “Ellip Workflows” solutions available on the iAOS Cloud Platform are accessible by the INTAROS partners for this purpose. This enables other application developers to reuse the Notebook for assessment and optimisation of remote data queries for ingestion and post-processing in their specific applicative workflows.

5.7. Roadmap for future development

The TS-CTD Service is available as described in this report for use and further development in WP6, in particular with partners in Task 6.2 “Improved ecosystem understanding and management” working on ocean and ecosystem modelling of the Barents Sea. A dedicated iAOS Application Showcase titled “Barents Sea Multi-depth Temperature & Salinity Maps” has been released to this end by WP5, including a demo video recording of the software capacities, defined as part of a joint WP5-WP6 seminar to share return of experience and new ideas in using the developed application.

In relation to the WP6 deliverable D6.13 “Extension of ecosystem management systems: Use the existing environmental and fisheries reporting and management systems of the Barents Sea and off Greenland”, the service can be used to demonstrate how data from an iAOS federated provider may allow for implementing similar procedures in other parts of the Arctic.

A possible extension to the service is to strengthen the Geostatistical model in order to improve the data explanatory analysis of CTDs. For example: take in account the vertical variations that can be used with a functional kriging, introduce the polynomial drifts to handle temperatures at high latitudes, or use the Salinity as a co-regionalized variable. Ultimately, building a global unique 4-D model (spatio-temporal) will allow scientists to take into account anisotropies and non stationarities. This will enable ocean and ecosystem models such as the ones under development in WP6 to integrate CTD-based datasets and combine with other data from INTAROS as well as other projects.

6. iAOS Showcase Applications - Integration of processing services

This section presents new results (not addressed in D5.7) focused on the exploitation of interoperability protocols (OPeNDAP) as promoted by INTAROS for the iAOS, and on the use of the RGeostats Geostatistical software library.

We provide hereafter a summary of the four iAOS Showcase applications that have been investigated in collaboration by WP5 and WP6 partners during the project.

6.1. Pan-Arctic Hydrological Modelling (with Task 6.1 SMHI)

Objectives

- Have the “observational” data available for search & 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 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

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

Delivered results

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

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
---	--	--

Foreseen improvements

- The approach was successful with the new model running for the Yakutian case study: <https://hype-eras.org/forecasts/>
- Software tools improvements in support of the exploitation of flow estimates in ungauged basins and of Pan-Arctic forecasts of river discharge in near real time, thus helping with predictions of spring floods, river ice breakup and freshwater flow to Arctic Ocean:
 - Improved API integrations
 - e.g. software updates for the HYCOS datasets pre-processing
 - Improved guidance to users, in order to
 - retrieve resources (API endpoints, software scripts)
 - generate data export files for local applications

6.2. Barents Sea Multi-depth Temperature & Salinity Maps (with Task 6.2 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 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 (operational and test instances):
<http://opendap1.nodc.no/opendap/physics/point/yearly/contents.html>
<http://opendap1-test.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.

Data exploitation

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

Delivered 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

Foreseen improvements

Approach:

- Improve the initial geostatistical spatial analysis performed in 2018
 - Build a unique spatio-temporal variogram model for Temperature
 - Take into account co-variables like (salinity, bathymetry...)
 - Handle currents nonstationarities 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

6.3. Maps for Svalbard Avalanche Forecast Modelling (with Task 6.4 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 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>

Data exploitation

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

Delivered 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 information sources (with relevant co-variables) over a small scale grid, at regular 6 hours intervals

Foreseen improvements

Approach:

- 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

6.4. Baffin Bay Bottom Temperature Maps (with Task 6.8 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 sources

CTDs (ICES) <ul style="list-style-type: none"> ● 1977 to 2017 ● 3700 vertical profiles ● 1.34M samples ● 1 CSV file (95Mo) 	Trawls (GINR) <ul style="list-style-type: none"> ● 1988 to 2016 ● Catches near sea floor only ● 51K samples ● 1 CSV file (11Mo)
Bottles (ICES) <ul style="list-style-type: none"> ● 1960 to 2017 ● 7800 vertical profiles ● 167K samples ● 1 CSV file (11Mo) 	Bathymetry (Gebco) <ul style="list-style-type: none"> ● Grid lag = 1/250 degree ● Grid size = 10320x4560 nodes ● 47M samples ● 1 NetCDF file (94Mo)

Additional data sources provided by Aarhus / Mikael 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/GTSP>
- NASA project in Greenland (new CTD)
<https://omg.jpl.nasa.gov/portal/browse/OMGEV-AXCTD/>

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

Delivered 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)

Foreseen improvements

Approach:

- 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

7. Conclusion

This D5.11 report concludes the INTAROS contribution to new iAOS processing services. It constitutes a second volume to the initial INTAROS deliverable D5.7 (Nov. 2019, updated June 2020).

The descriptions of the previously developed processing services (C-PAD and TS-SIC, two iAOS services developed by NERSC) have been kept in this second volume, in order to deliver a comprehensive overview of the processing service integration work, done over the full INTAROS project duration.

In the last year of the INTAROS project, an additional processing service has been integrated (TS-CTD, an iAOS service developed by Terradue) to support data access query optimization using the OPeNDAP interface. This new iAOS service was created in order to support the project partners for the development of iAOS data exploitation applications (e.g. the iAOS showcase applications), as well as to support any future iAOS user willing to proceed with a similar OPeNDAP query optimisation approach.

The processing services described in this report are powered by the iAOS Cloud Platform, deployed and operated by Terradue (cf. INTAROS deliverables D5.5 - iAOS requirements and architectural design V2 and D5.8 - iAOS platform and tools V2). The iAOS Cloud Platform allows partners to perform the integration and analysis of multidisciplinary arctic data from distributed repositories. The design of the iAOS Cloud Platform is user driven; consequently a set of custom tools for data discovery, aggregation, analysis and visualization has been made available to the Cloud Platform users. Part of it, the Ellip Solutions provide an application integration environment accessible by the iAOS software developers who need to build scalable, interoperable data processing applications, and make these applications available to user communities “as-a-Service”.

Such Cloud integration capacity for data processing services is defined to cover the needs of Pan-Arctic applications from stakeholders in industry, governance, local communities or research. It has delivered a suite of selected iAOS applications (the iAOS “Showcase” applications), designed and validated during the INTAROS project through a collaboration process between project partners in the INTAROS Work Package 5 and Work Package 6.

Such iAOS applications address the key objective of the INTAROS project to showcase advances in ecological and environmental understanding, by merging and synthesizing iAOS data through ecosystem modelling at the regional (Barents Sea) and local (Greenland) scales.

8. References

Chauvaud, L., Mathias, D., Jolivet, A., Amice, E., Le Bec, T., Marec, C., Merny, P., and collaborators (2020). Underwater soundscape in Kongsfjorden. Presentation at online INTAROS Ocean Workshop – April 2020.

Geyer, F., Sagen, H., and P. F. Worcester (2017). “WIFAR/UNDER-ICE acoustic recording from an integrated ice station in the Fram Strait marginal ice zone in 2012”. NIRD Research Data Archive, <https://doi.org/doi:10.11582/2017.00012>

Geyer, F., Sagen, H., Hope, G., Babiker, M., and P. F. Worcester, “ Identification and quantification of soundscape components in the Marginal Ice Zone,” J. Acoust. Soc. Am. 139, 1873–1885 (2016). <https://doi.org/10.1121/1.4945989>

Kindermann, Lars (2013): Acoustic records of the underwater soundscape at PALAOA with links to audio stream files, 2005-2011. Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA, <https://doi.org/10.1594/PANGAEA.773610>

Merchant, N. D., Fristrup, K. M., Johnson, M. P., Tyack, P. L., Witt, M. J., Blondel, P. and Parks, S. E. (2015), Measuring acoustic habitats. *Methods Ecol Evol*, 6: 257-265. doi:10.1111/2041-210X.12330

9. Annex - WP6 references to iAOS Showcase applications

Task 6.1 deliverables

D6.1 (v1.4) Climate model initialization: first report on the added value of using data from INTAROS (Resp. BSC, with SMHI and NERSC; **M42**).

Section 6 presents the added value of the integrated Arctic Observation System (iAOS) for enhancing and making available hydrological model predictions for the major Arctic rivers. In particular it describes the approach to combine the river discharge data from the Arctic Hydrological Cycle Observing system (Arctic-HYCOS, that was assessed and enhanced in INTAROS WP2) with the pan-arctic hydrological model Arctic-HYPE provided by SMHI (<http://hypeweb.smhi.se>), to predict and monitor fresh water inflow to the Arctic Ocean and changes in Arctic hydrological regimes. Cf. processing service related to the iAOS Showcase Application “Pan-Arctic Hydrological Modelling”.

D6.11 Climate model initialization: first: final report on the added value of using data from INTAROS (Resp. SMHI, contribution from BSC and NERSC; **M58**).

Task 6.2 deliverables

D6.2 Impact of climate change on Greenland ecosystems and fish resources, first report on decadal scale changes in fish community structure along the West Greenland coast (Resp. AU, GINR; **M42**).

Section 3 presents the Trawl surveys data from the summer bottom trawl surveys conducted by the Greenland Institute of Natural Resources. It described the distribution of the bottom trawls along the West Greenland shelf and the near bottom temperature measured for each trawl. In collaboration with Aarhus University (Task 6.8), this provided input to the definition of the processing service for the iAOS Showcase Application “Baffin Bay Bottom Temperature Maps”.

D6.3 Extension of ecosystem management systems: Use the existing environmental and fisheries reporting and management systems of the Barents Sea and off Greenland to demonstrate how data from an iAOS may allow for implementing similar procedures in other parts of the Arctic, first report (Resp. IMR, AU; **M42**).

Section 1.1 introduces selected cases (Barents Sea and Disko Bay, Greenland) to analyse how data, including data to be incorporated in the INTAROS iAOS, may contribute to advances in ecological and environmental understanding and allow for expanding existing environmental and fisheries reporting and management systems into new geographic areas, and the collaboration with task 6.8 (Fisheries and environment management). Section 4.1 describes the evaluation of existing and future iAOS to support models and Section 4.2 introduces the role of ecosystem models in existing and future iAOS. Cf. processing service related to the iAOS Showcase Application “Barents Sea Multi-depth Temperature & Salinity Maps”.

D6.12 Impact of climate change on Greenland ecosystems and fish resources, final report on decadal scale changes in fish community structure along the West Greenland coast (Resp. AU, GINR; **M58**)

D6.13 Extension of ecosystem management systems: Use the existing environmental and fisheries reporting and management systems of the Barents Sea and off Greenland to demonstrate how data from an iAOS may allow for implementing similar procedures in other parts of the Arctic, final report (Resp. IMR, AU; **M58**).

Task 6.4 deliverables

D6.16 Natural hazard assessment in the Arctic and evaluation of long-term impact of climate change on natural hazards.

(Resp. GEUS; DTU, FMI, UPM; **M58**)

D6.17 Ice discharge from Greenland icesheet: Model-based demonstration of calculations of ice discharge from glaciers to the ocean, aimed to predict the contribution of glaciers to sea level rise.

(Resp. UPM; **M58**)

Task 6.8 deliverables

D6.10 Report on ecosystem management for managers: Ecological model results from the Barents Sea and Disko Bay will be presented and discussed with fisheries and environmental managers at workshops respectively in Norway and on Greenland

(Resp. IMR, AU; **M54**)

----- END of DOCUMENT-----



INTAROS

This report is made under the project
Integrated Arctic Observation System (INTAROS)
 funded by the European Commission Horizon 2020 program
 Grant Agreement no. 727890.



Project partners:

