SeaDataCloud

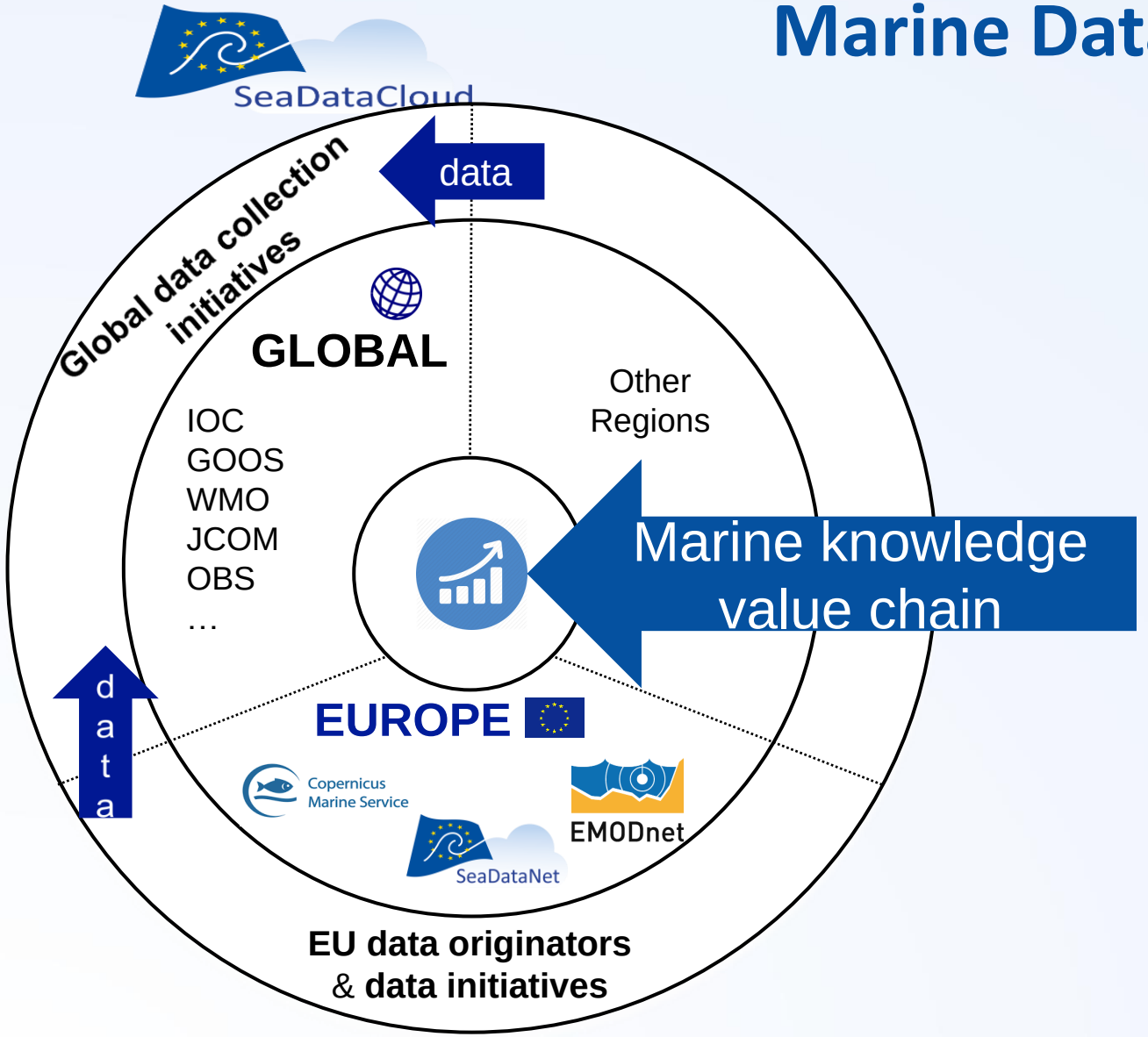
SeaDataCloud Products and links with IQuoD

S. Simoncelli and WP11 team

sdn-userdesk@seadatanet.org – www.seadatanet.org

6th Workshop IQuOD, 29-31 October 2019, Brest

Marine Data and Data Products Initiatives



Copernicus Marine Service
DG-GROW

→CMEMS provides generic info (obs and forecast/rea) on the PHY state and dynamics of the ocean and marine ecosystems

SeaDataCloud
DG-RTD

→ Pan-EU data infrastructure for ocean and marine data management, underpinning several EMODnet thematic lots. It assembles and distributes marine data, metadata and data products (T and S)

EMODnet
DG-MARE

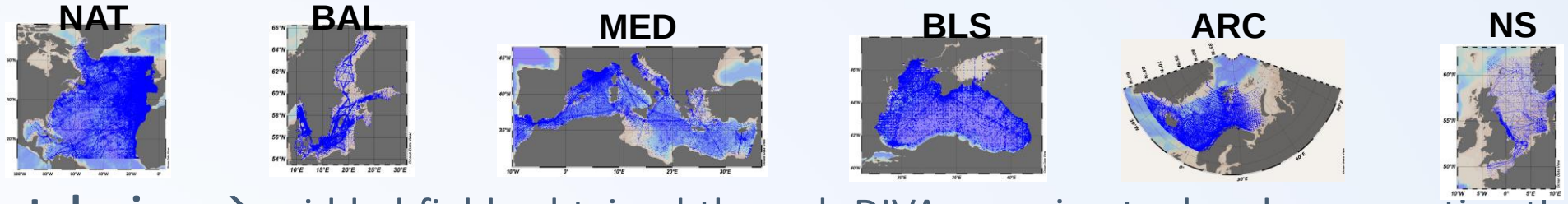
→ assembles and distributes marine data, metadata and data products spanning on 7 thematic areas

adapted from *Martín Míguez et al.(2019)* - The European Marine Observation and Data Network (EMODnet): Visions and Roles of the Gateway to Marine Data in Europe. Front. Mar. Sci.

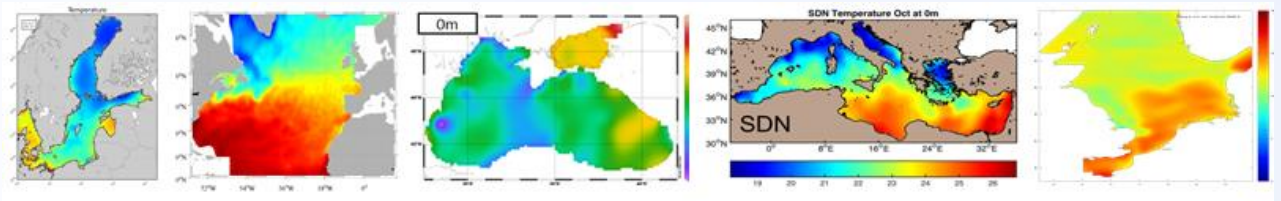
SeaDataCloud Products

GOAL: to provide the best **data products** from SeaDataNet at **regional and global scale** and serve diverse user communities (op. oceanography, climate, marine environment, institutional, academia)

1. **Aggregated data sets EU marginal seas**→ historical temperature and salinity data harvested from the central CDI and validated by regional leaders



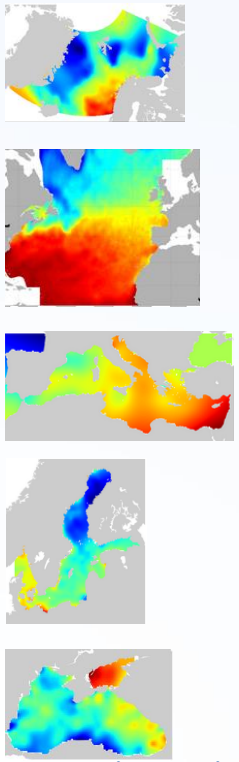
2. **Climatologies** → gridded fields obtained through DIVA mapping tool and representing the climate of the ocean at **regional and global scale**



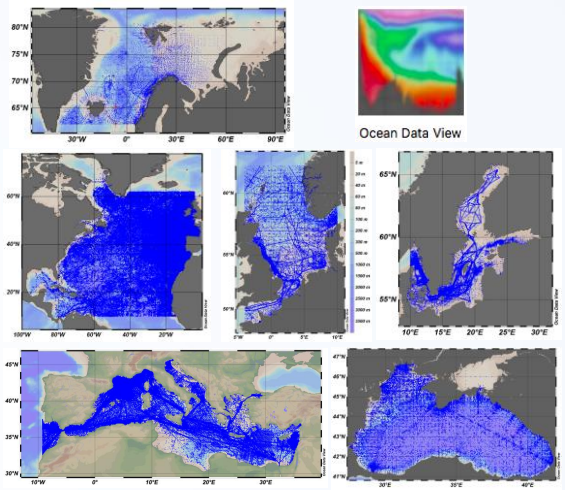
3. **New data products** → multi-platform and multi-disciplinary approach combining in situ (e.g. Argo, ships, drifters, fixed platforms) and remote sensed observations, Ocean Monitoring Indicators for tracking ocean mechanisms and/or climate modes and trends

Quality Assurance Strategy (QAS)

SDC_CLIM_TS_V1 (Jul19)
SDC_CLIM_TS_V2→ Sept20



SDC_DATA_TS_V1 (Jun18)
SDC_DATA_TS_V2 → Dec19



DIVA

sdn-userdesk@seadatanet.org

Quality Control Analysis

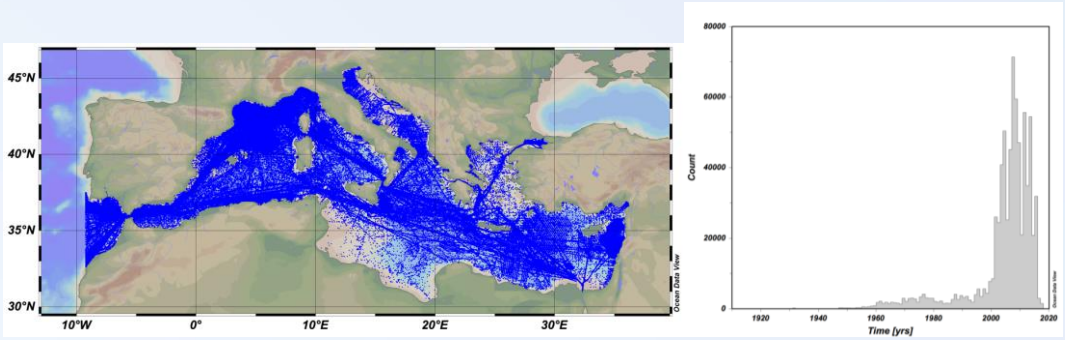
SeaDataCloud

ODV5.0

Ocean Data View

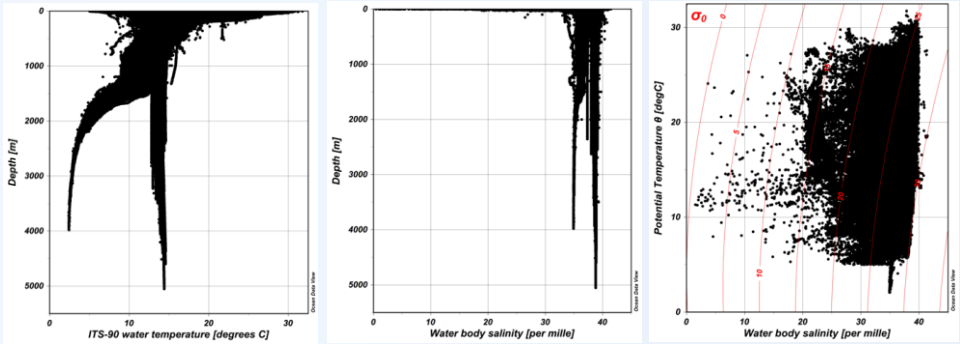
Common guidelines

1. Spatial data distribution
2. Temporal data distribution (annual, seasonal and monthly)
3. Quality Flags statistics
4. Scatter plots of good/probably good (QF1/QF2) observations
5. Gross range check
6. Analysis of QF0 data (not checked) to disclose good data
7. Visual check to identify wrong profiles (spikes, outliers)
8. Identification of stations on land
9. Identification of wrong/missing data (time, measurements)
10. Stability check



par	# stations	%	# samples
total	739784		
T	737102	99,6	41223938
S	667232	90,2	28518744
TS	665388	89,9	28119926

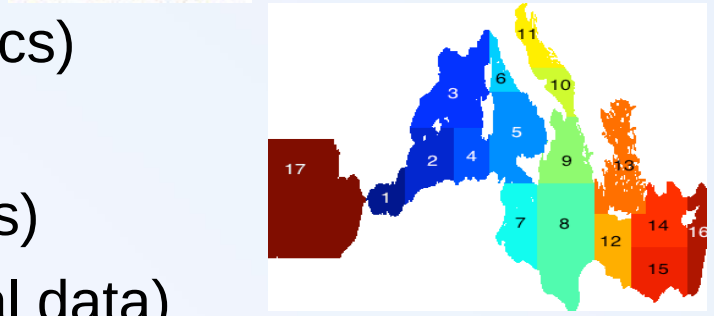
%	QF0	QF1-2	Q3-4
dpt	3.0→0	96.9→99.8	0.1
T	2.7→0	97.0→99.8	0.3
S	4.5→0	94.6→99.2	0.9
dpt&T&S	3.0	94.4	0.3

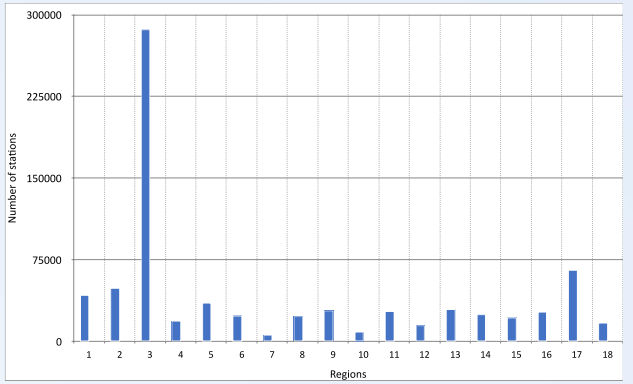


Quality Control Analysis

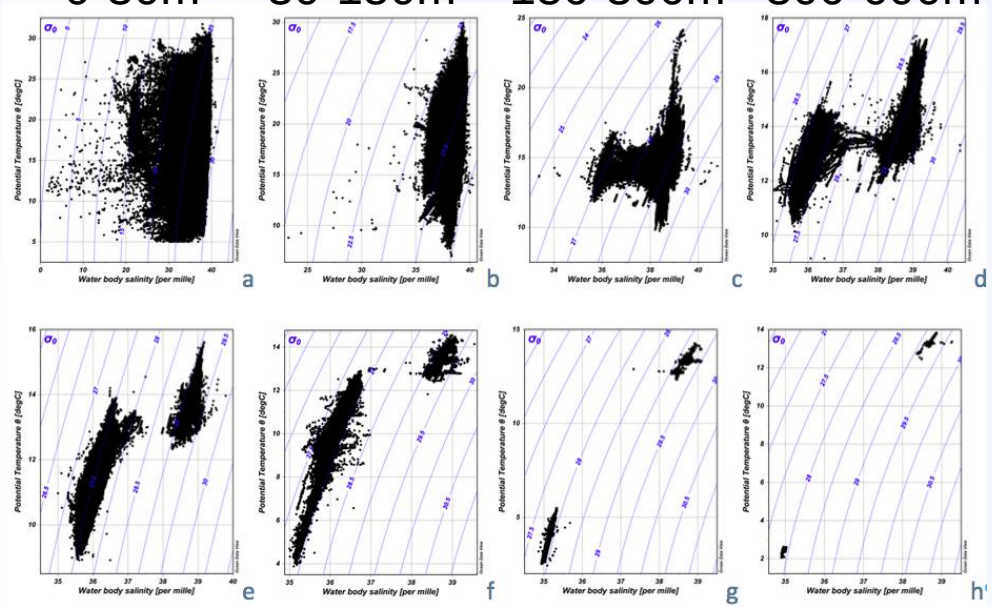
Specific checks per

- **areas** (similar hydrodynamic characteristics)
- **layers** (surface, intermediate, bottom)
- **time periods** (decades or specific periods)
- **Instrument type** (consistency of historical data)

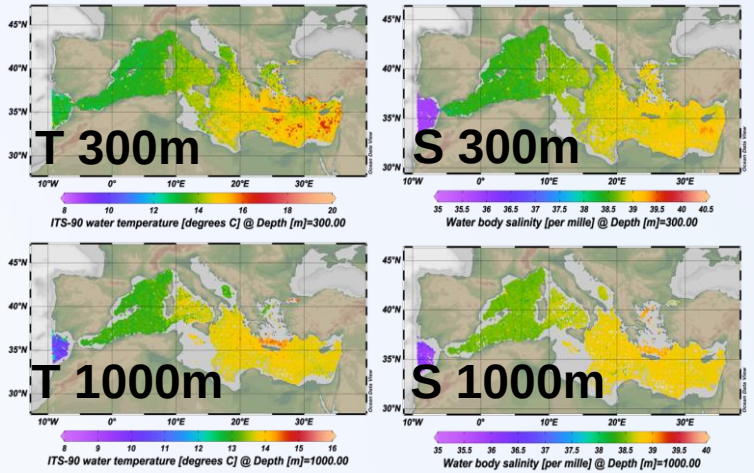


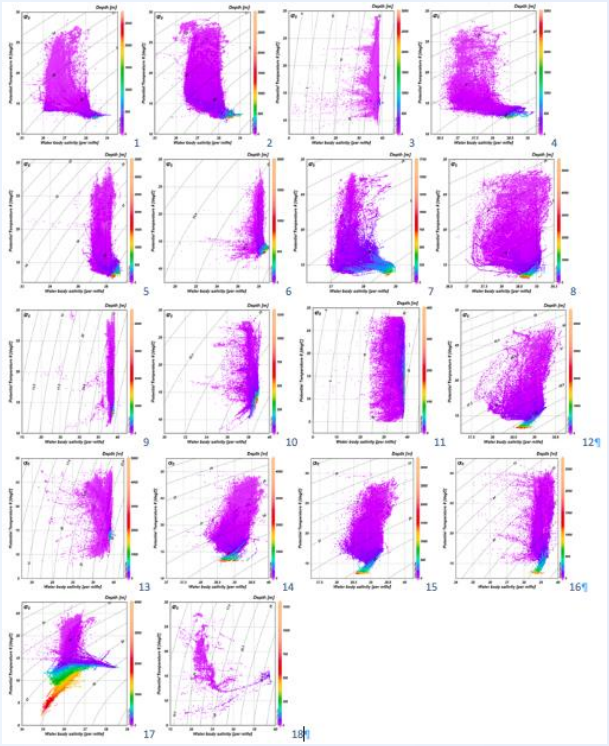


0-30m 30-150m 150-300m 300-600m



600-1000m 1000-2000m 2000-3000m 3000-5500m



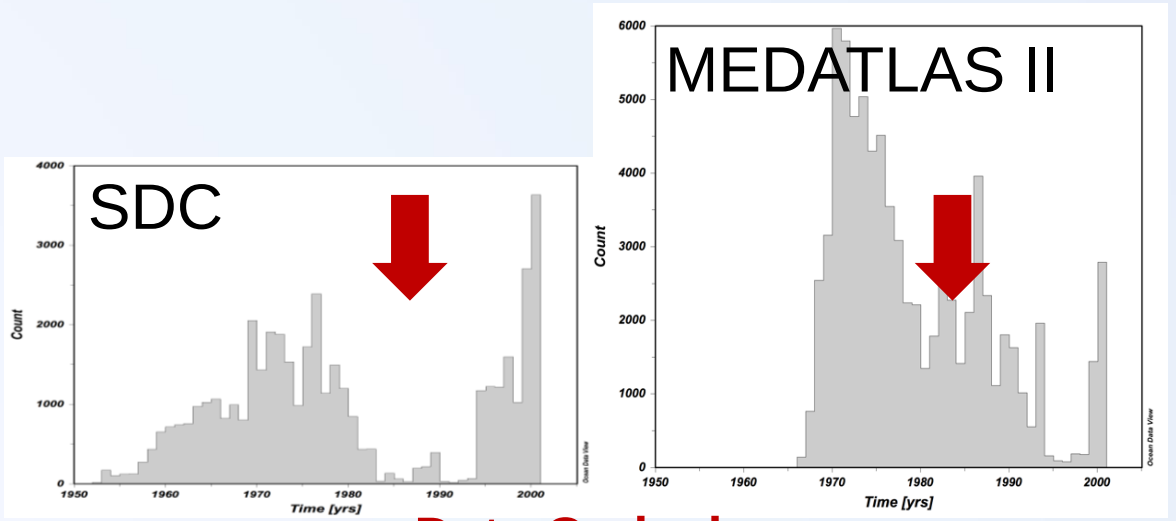


Metadata Analysis

- New **data distributors/originators statistics** → QC filtering by EDMO code → detection of systematic (format, flagging) errors
- New **instrument type statistics** → analysis of monitoring space-time coverage → detection of data omissions

Instrument/Gear Type	# stations	%
CTD	52031	7
bathythermograph	56558	8
discrete water sampler	32258	4
thermosalinograph	555269	75
thermistor chains	22	0
continuous water sampler	1577	0
salinity sensor; water temperature sensor	19852	3
salinometers	100	0
salinity sensor	143	0
water temperature sensor	1	0
none info	21973	3

MED XBTs 1950-2000



Data Omission

!!!Actions to ingest missing XBTs!!!

!!!Actions to complete crucial metadata information!!!

SeaDataCloud

SDC_V1 Climatologies

sdn-userdesk@seadatanet.org – www.seadatanet.org

6th Workshop IQuOD, 29-31 October 2019, Brest

Objectives

1. To produce T and S climatologies with increased horizontal and **vertical resolution** with respect to SDN2 ones
2. To compute **decadal** or pentadal climatologies depending on the data distribution in time and space and the regional long-term variability
3. To perform **consistency analysis with available WOA climatologies** and/or other products like CMEMS climatologies computed by satellite reprocessed data sets and reanalysis products, EN4 gridded products
4. **To include all available SDC restricted observations and integrate external data sets in order to provide decadal climatologies**

Decision: to conform to WOA schema and adopt the same vertical standard levels →

- to test the new DIVAnd software
- to analyze CLIM consistency with WOA fields and guarantee good results

Product	horiz resolution	time coverage	Annual	Seasonal	Monthly	External Source
ARC_1	1x1/2°	1955-2014		x	x	WOD18
ARC_2	1x1/2°	6 decades		x	x	WOD18
NS_1	1/8°	1955-2014	x		x	WOD18
NS_2	1/8°	6 decades		x		WOD18
BAL_1	1/16°x1/32°	1955-2014		x	x	CORA5.2
BAL_2	1/16°x1/32°	6 decades		x	x	CORA5.2
NAT_1	1/2°	1955-2015		x	x	CORA5.1
NAT_2	1/2°	6 decades		x	x	CORA5.1
NAT_3	1/4°	1955-2015		x	x	CORA5.1
NAT_4	1/4°	6 decades		x	x	CORA5.1
MED_1	1/8°	1955-2017		x	x	CORA5.2
MED_2	1/8°	1955-1984		x	x	CORA5.2
MED_3	1/8°	1985-2017		x	x	CORA5.2
MED_4	1/8°	6 decades		x		CORA5.2
BLS_1	1/8°	1955-2017			x	WOD18, CORA5.1
BLS_2	1/8°	1955-1994			x	WOD18, CORA5.1
BLS_3	1/8°	1995-2017			x	WOD18, CORA5.1
BLS_4	1/8°	6 decades		x		WOD18, CORA5.1

Regional Climatologies

Harmonized approach

- cover the time period after 1955
- adopted WOA standard vertical levels
- all have monthly fields over the whole time span 1955-2014 (2017) and seasonal decadal fields
- **all have been created integrating SDC data with external sources, WOD and/or CORA**

Global Climatologies

Name	horizontal resolution	time coverage	monthly	Data Source
GLO_1	1/4°	1900-2017	x	WOD13
GLO_2	1/4°	2003-2017	x	WOD13

A global SDC product has been created for the first time

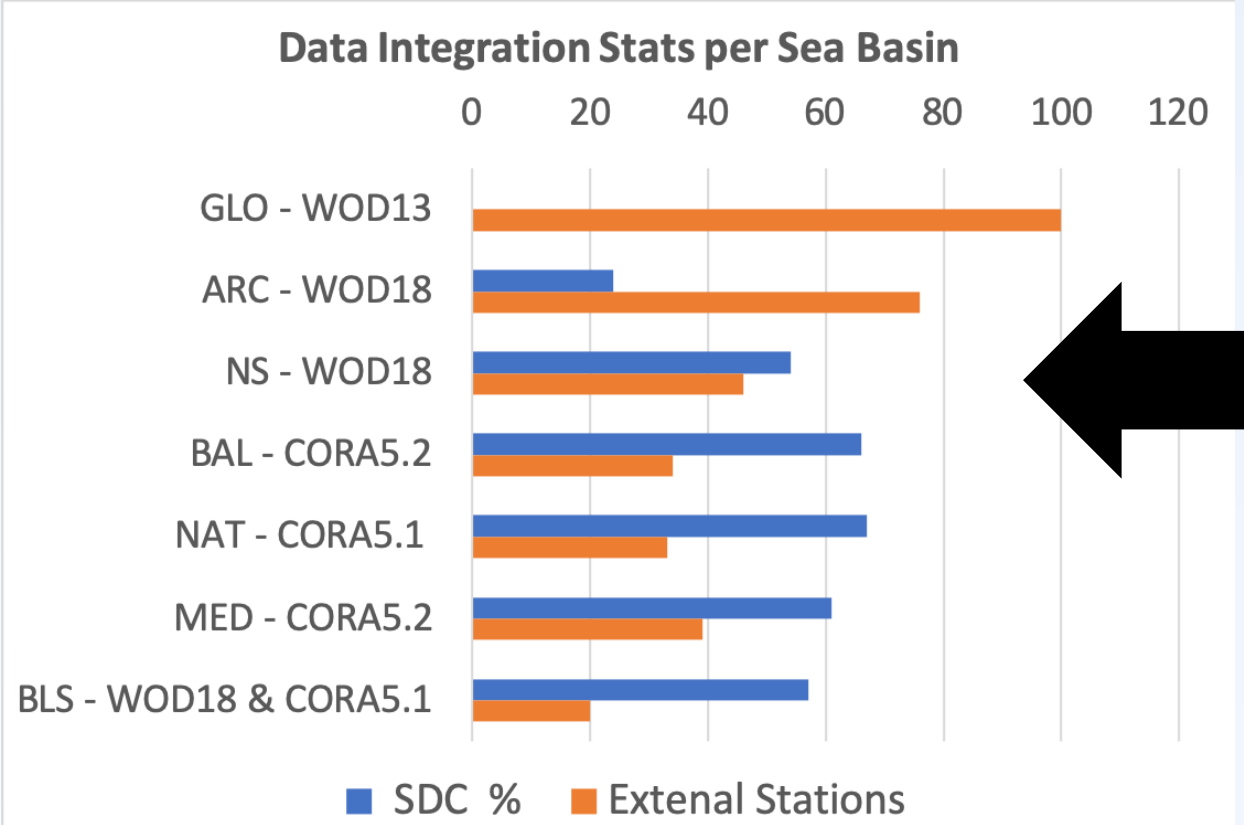
→ two different monthly climatological fields for T and S with a different time coverage, computed from WOD data since spatial coverage of SDN data at global scale is still too sparse

Next releases the idea is:

→ to integrate SDN and WOD18 data

→ to improve the background estimate and maximize product's quality

Input Dataset

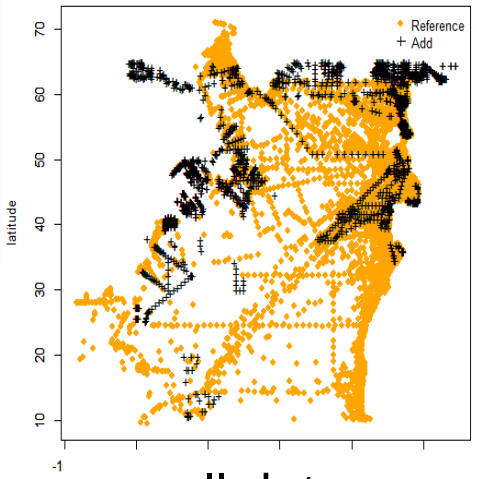


% of data from SDC and the external source in the climatology data set per sea basin

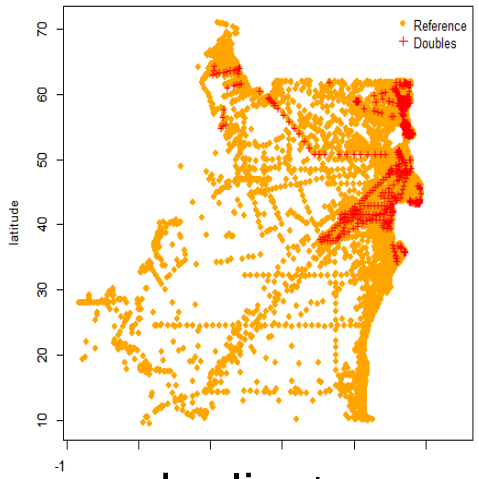
Duplicates

	WOD	CORA
ARC		
NS	2.4% (DIVAnd) 2.6% (ODV)	
BAL		
NAT		
MED		19.9% (DIVAnd)
BLS	16.7%	3,3% 9.2% with WOD

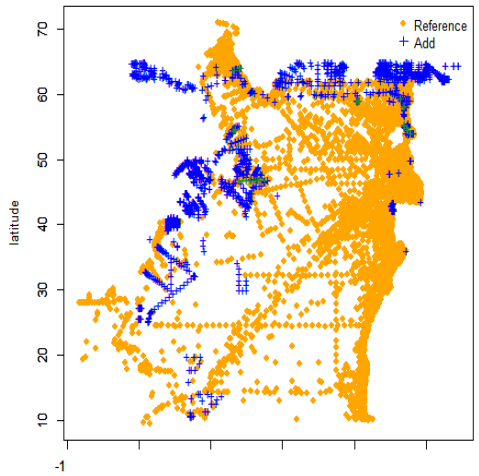
North Atlantic



all data

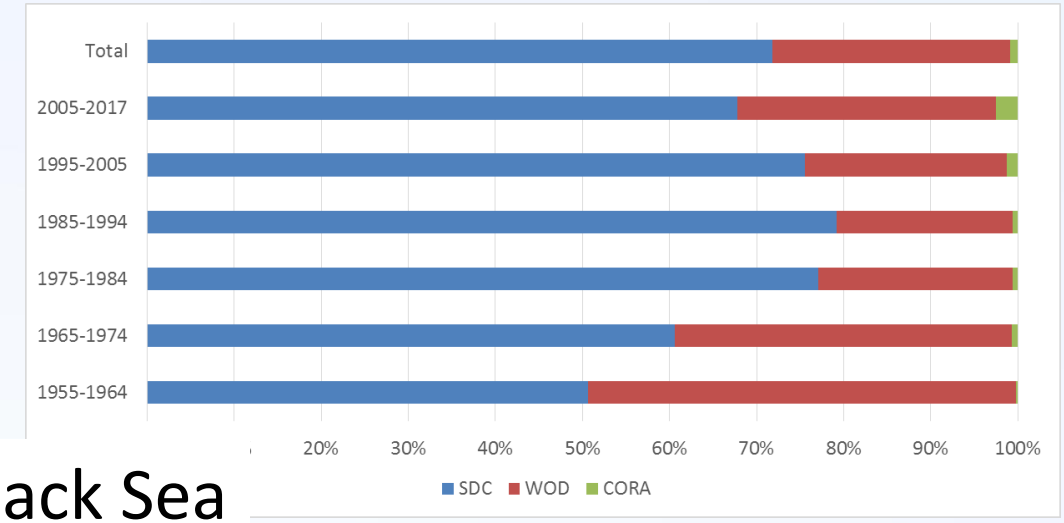


duplicates

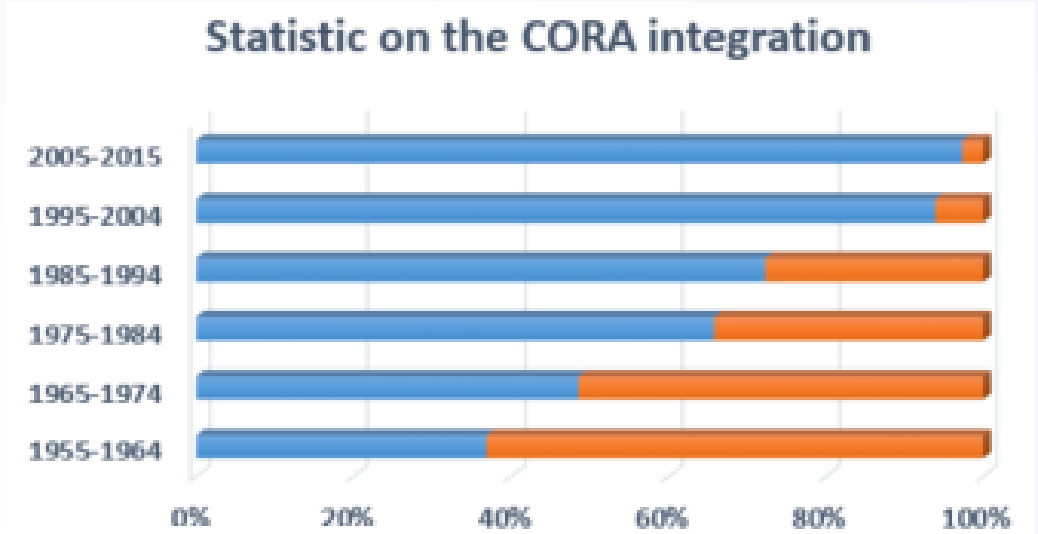


new

Climatology data sets

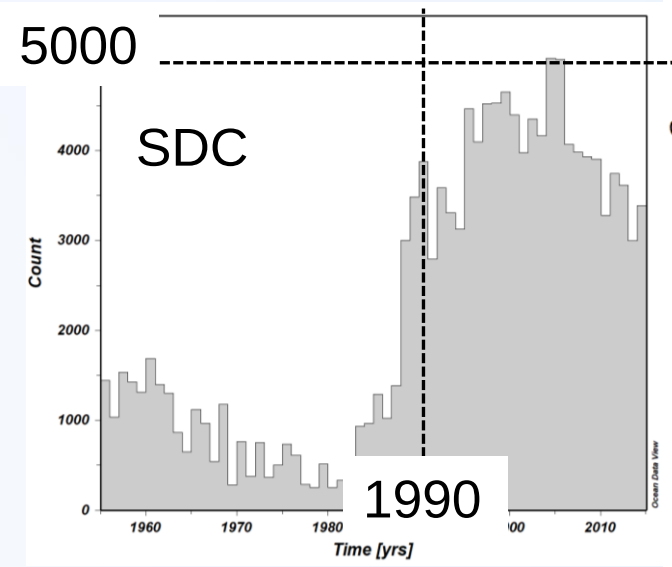


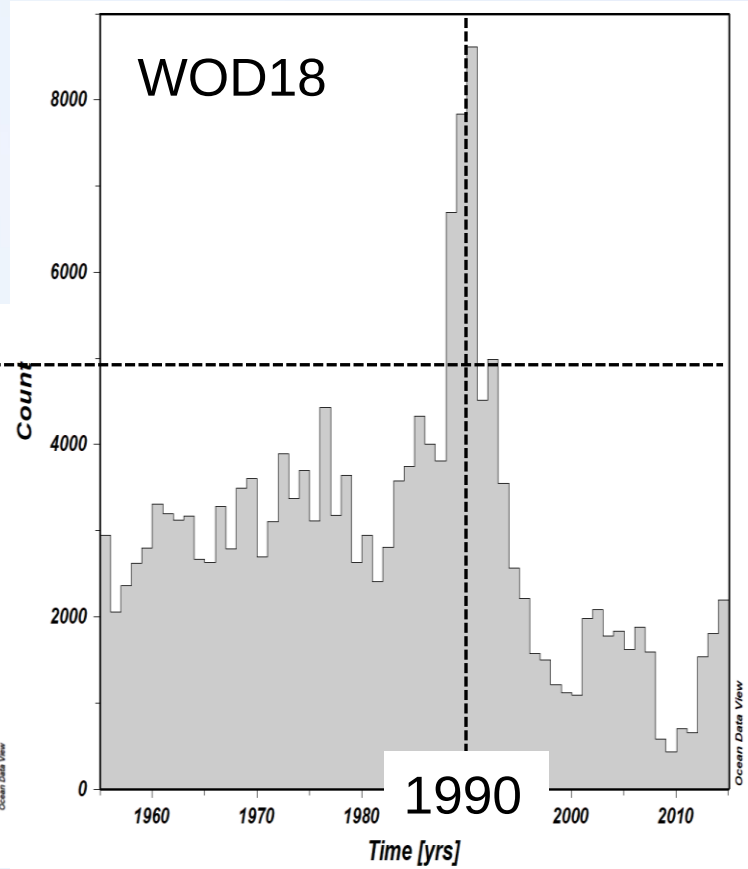
Black Sea



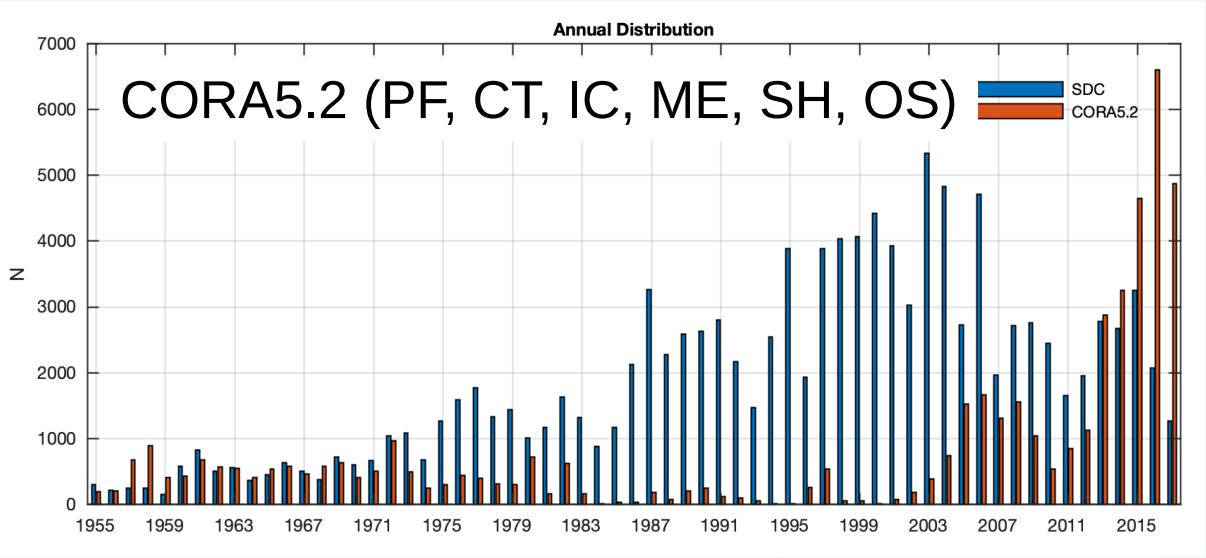
North Atlantic

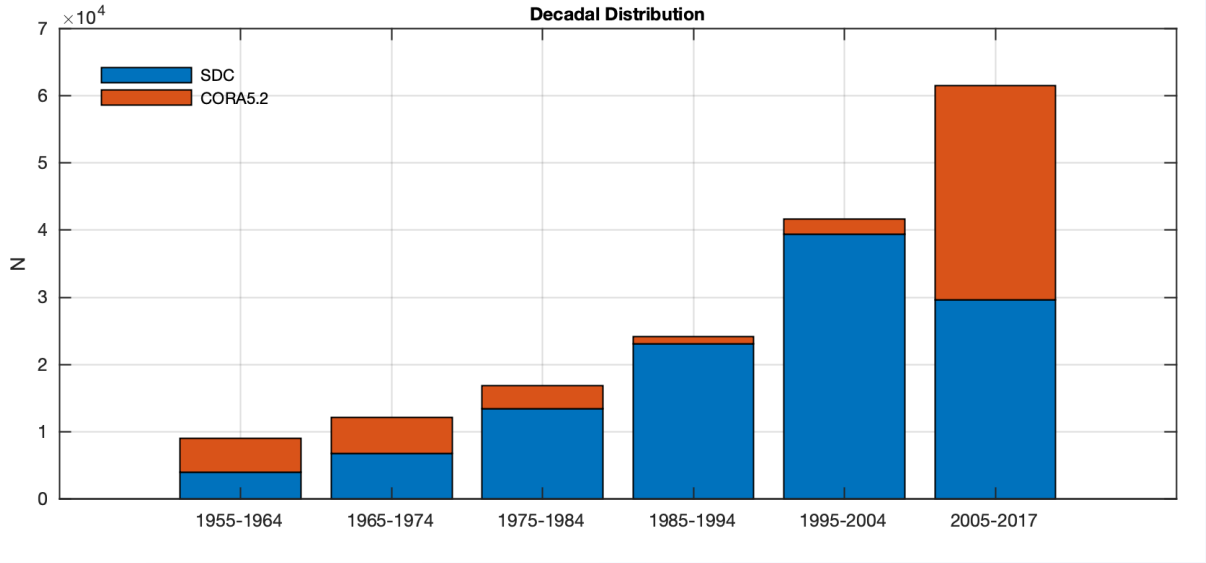
North Sea



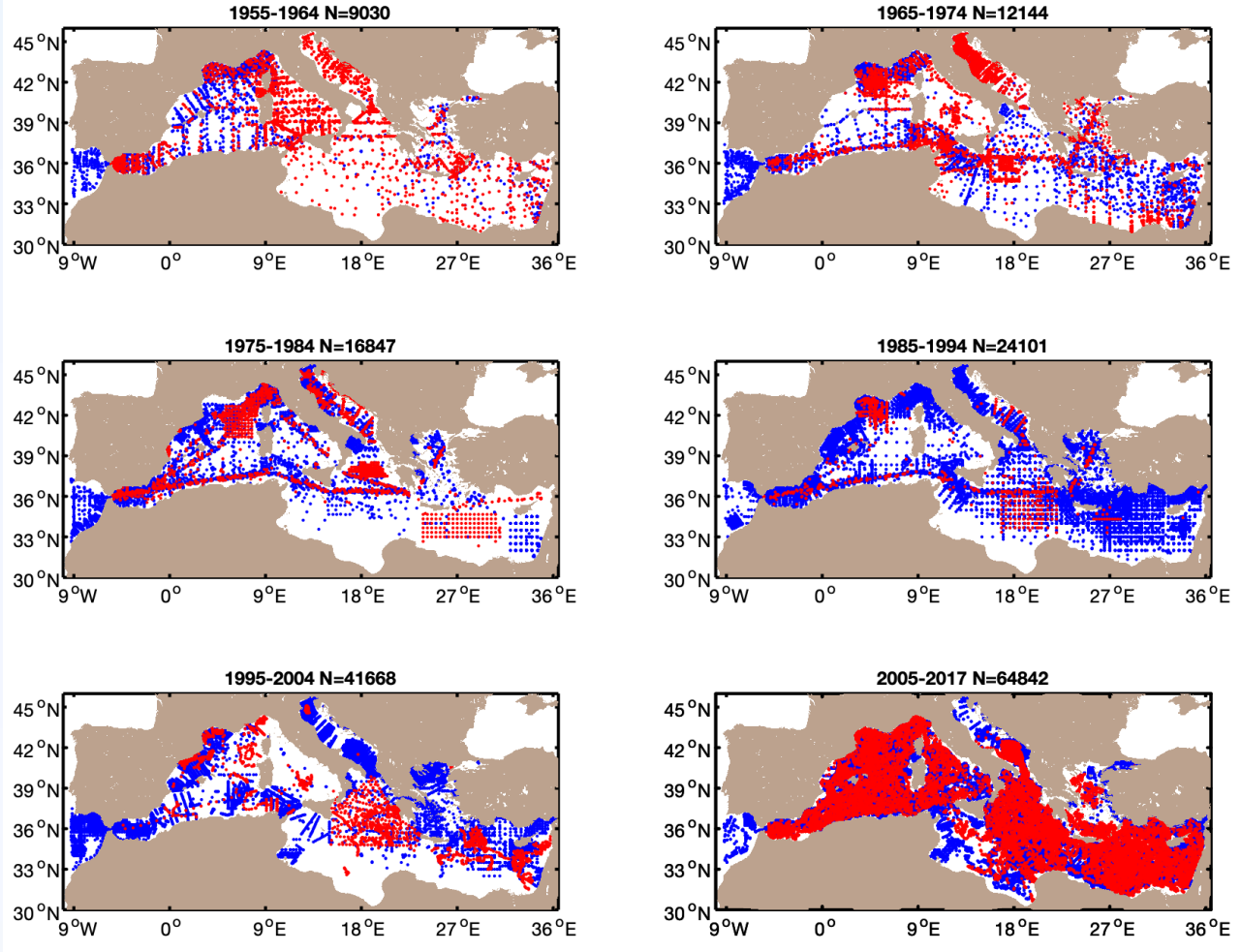


Integrated Dataset





Mediterranean Sea



Integration and Quality Control

Both integrated data sets (WOD and CORA) presented data anomalies (outliers, bad cruises' data, stations on land) thus a dedicated additional QC has been performed:

- before data merging on the external data set
- after data merging on the integrated data set to further assure data consistency

This activity will need particular attention in the next version in order to keep track of data anomalies through the station identifier and give a feed back to NOAA and Coriolis

divand: generalised, n-dimensional interpolation

https://www.geosci-model-dev.net/7/225/2014/gmd-7-225-2014.pdf

https://github.com/gher-ulg/divand.jl

divand-1.0: *n*-dimensional variational data analysis for ocean observations

A. Barth[1,*], J.-M. Beckers[1], C. Troupin[2], A. Alvera-Azcárate[1], and L. Vandenbulcke[3,4]

[1]GHER, University of Liège, Liège, Belgium
[2]IMEDEA, Esporles, Illes Balears, Spain
[3]seamod.ro/Jailoo srl, Sat Valeni, Com. Salatrucu, Jud. Arges, Romania
[4]CIIMAR, University of Porto, Porto, Portugal

[*] *Invited contribution by A. Barth, recipient of the EGU Arne Richter Award for Outstanding Young Scientists 2010.*

Correspondence to: A. Barth (a.barth@ulg.ac.be)

Received: 7 June 2013 – Published in Geosci. Model Dev. Discuss.: 23 July 2013
Revised: 18 October 2013 – Accepted: 12 December 2013 – Published: 29 January 2014

2013: Octave/MATLAB

2016: Julia

faster, better, stronger

DIVAnd User Interface

Jupiter notebooks → interactive computational environments which combine:

1 code fragments that can be executed,

2 text for the description of the application and

3 figures illustrating the data or the results

”Interactive notebooks: Sharing the code”, Nature (2014)
http://www.nature.com/news/
interactive-notebooks-sharing-the-code-1.16261

DIVA	DIVA version	Lc	Lcz	Epsilon2	surfex tend	Background
GLO	**DIVAnd 2.3.1**	200km		0.9	false	spatial mean
ARC	DIVA 4.7.2	1deg		2.0		seasonal 1955-2014
NS	**DIVAnd 2.3.1**	100km		0.1	true	spatial mean
BAL	**DIVAnd 2.3.1**	120km	20m	1.0	true	seasonal 1955-2014
NAT	DIVA 4.7.2	2deg		1.0		spatial mean
MED	**DIVAnd 2.4.0**	2deg	0	0.6 an 6.0 bkg	false	T monthly S annual 1955-2017 1955-1984 1985-2017
BLS	**DIVAnd 2.3.1**	150km	varying with depth	T= 0.3 an 0.5 bkg S=0.1	true	seasonal 1955-2014

A posteriori data Quality Control

During the production phase and additional data rejection has been applied to data that generated artificial bull-eyes features in the gridded fields. Artificial features due to:

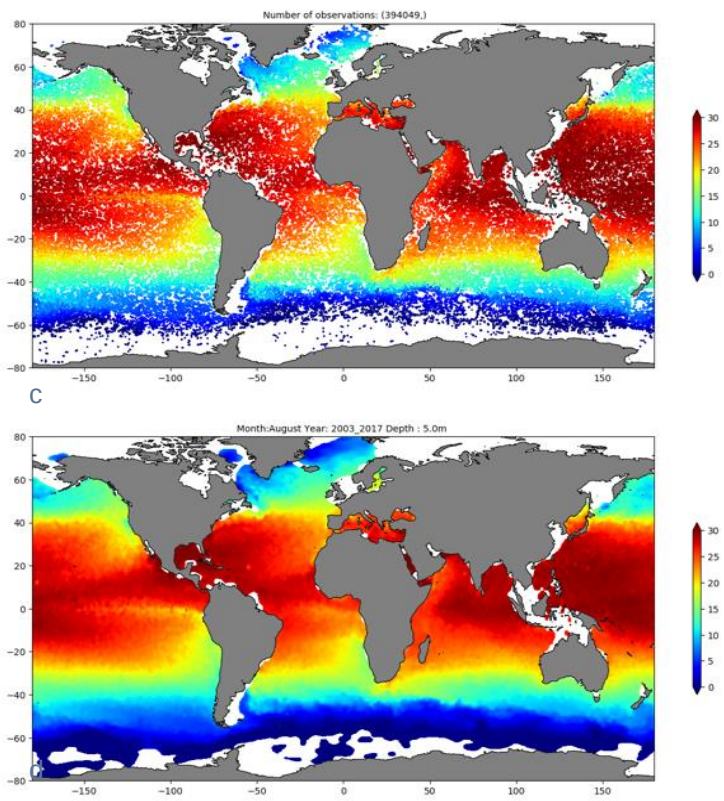
→ non-detected outliers

→ representativeness errors: obs that do not represent the climatological status of the region under analysis, but rather smaller spatial and temporal scales' events

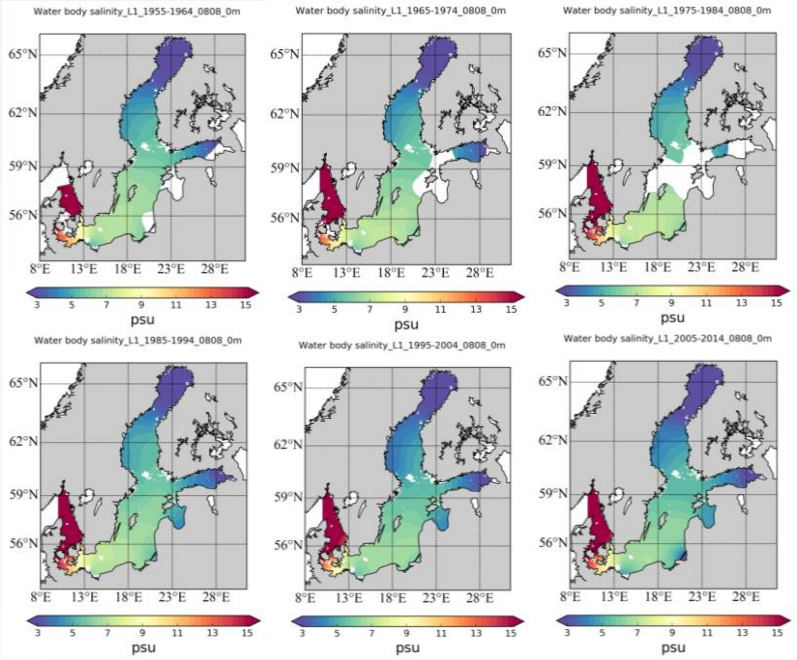
Manual removal or through DIVAnd tool which computes the **residuals** after the climatological fields generation. Iterative approach → residuals > pre-defined thresholds have been removed (BLS and MED)

This phase need to be improved and standardized in the next V2 production

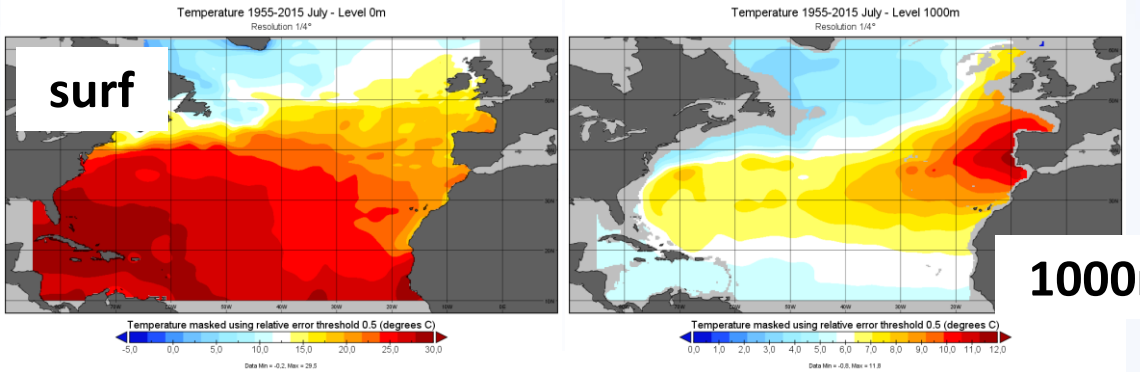
Results



Surface Temperature (Jan)

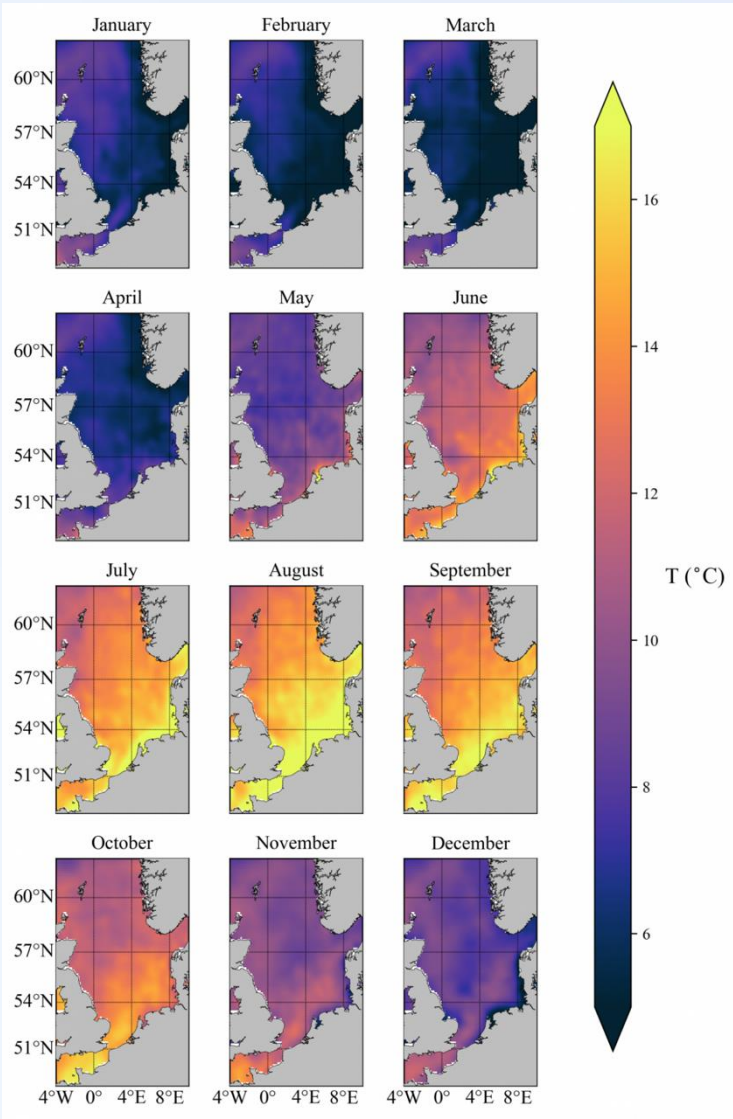


Surface salinity (Aug)



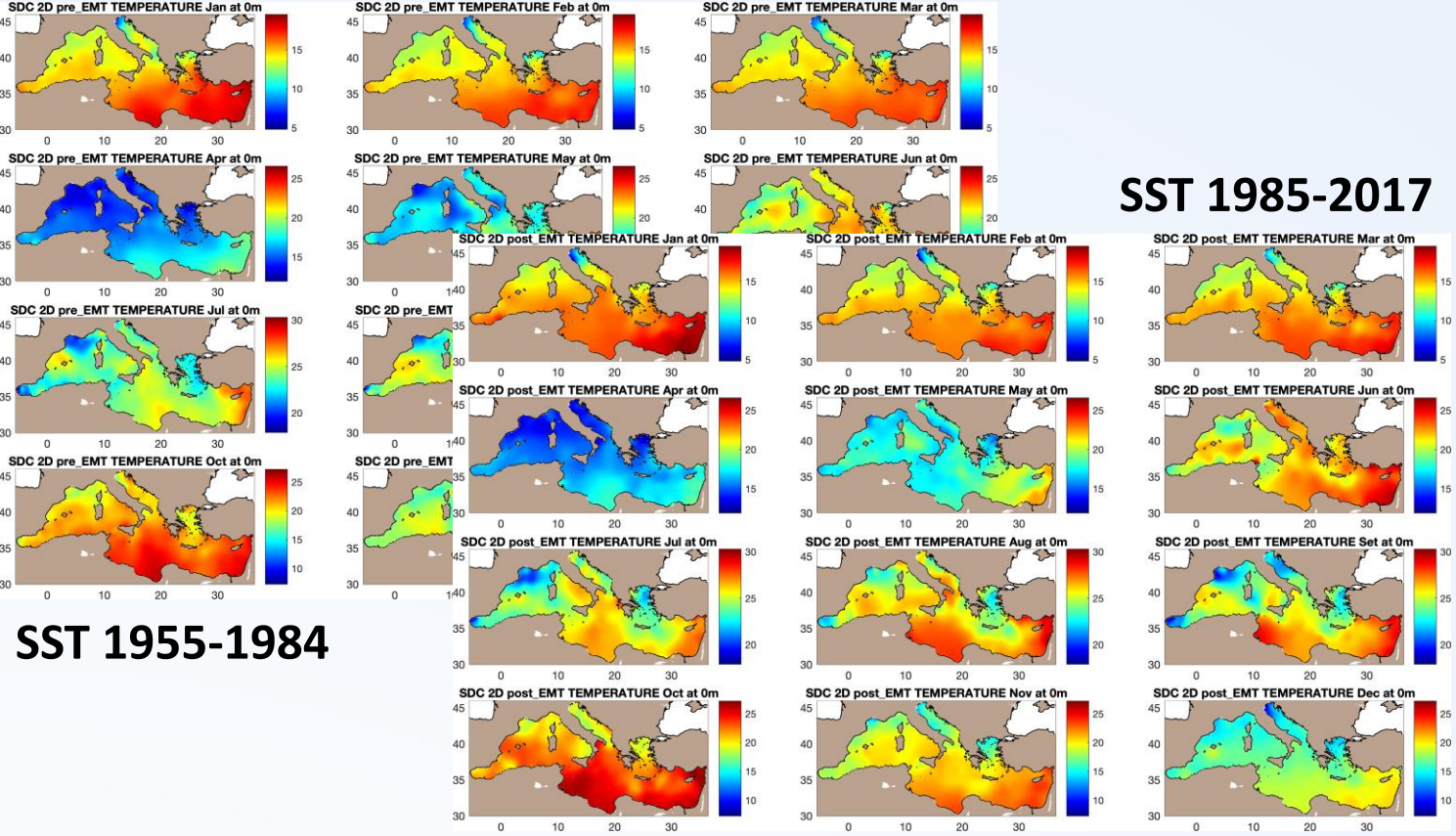
Temperature Jul (1955-2014)

Surface Temperature



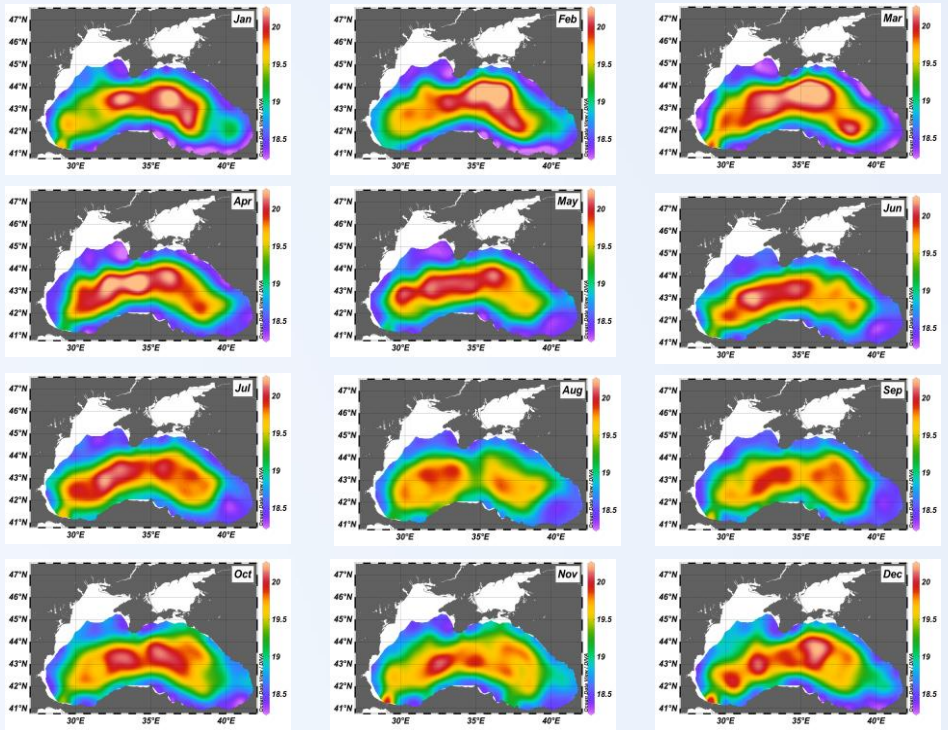
Results

SST 1985-2017

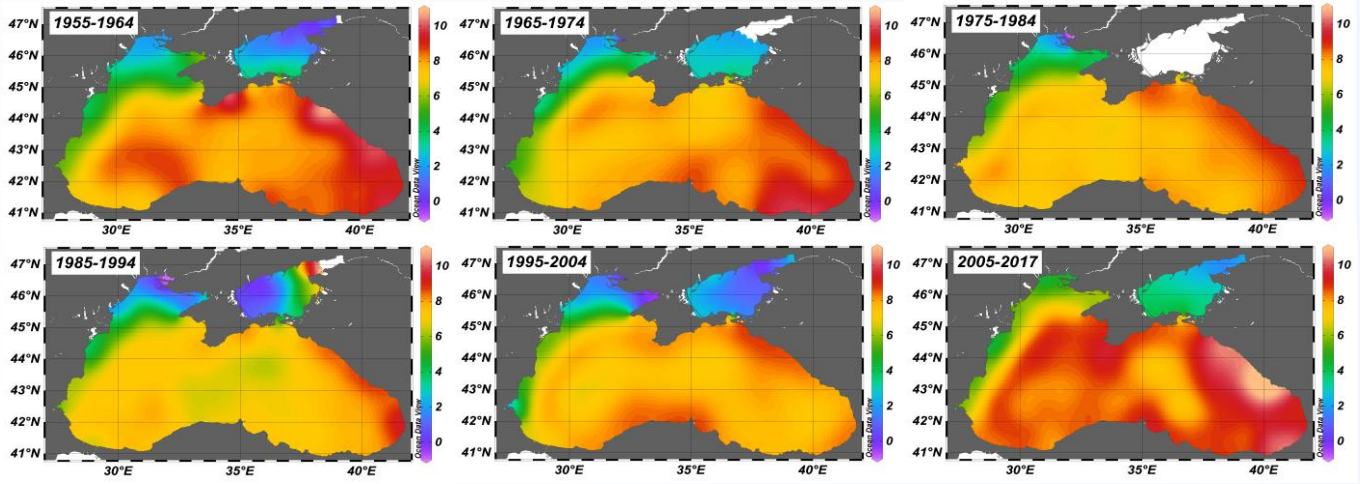


SST 1955-1984

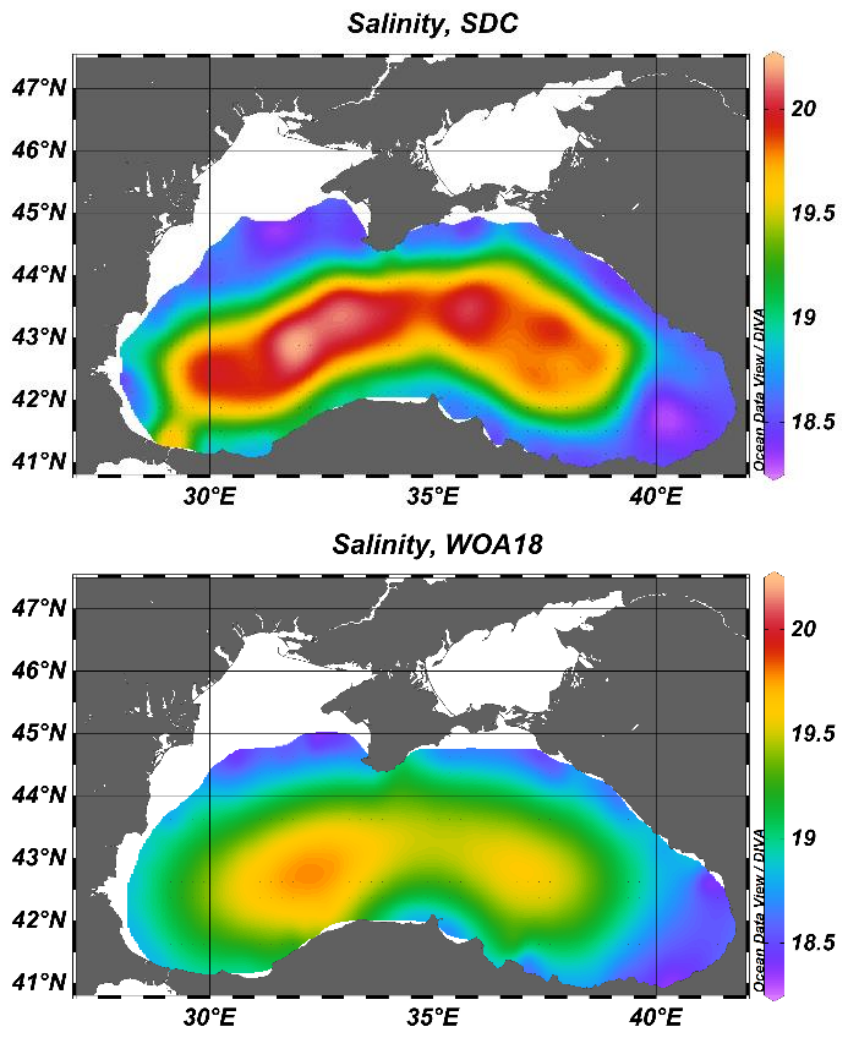
Salinity at 70m (1955-2017)

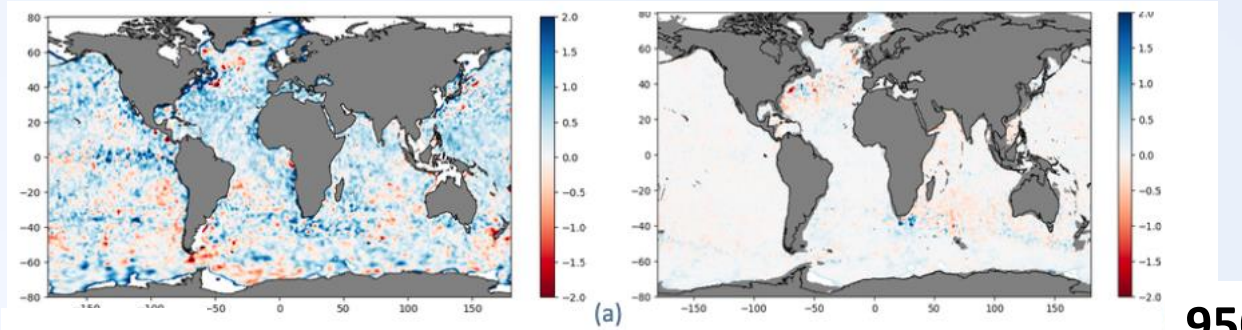


winter SST



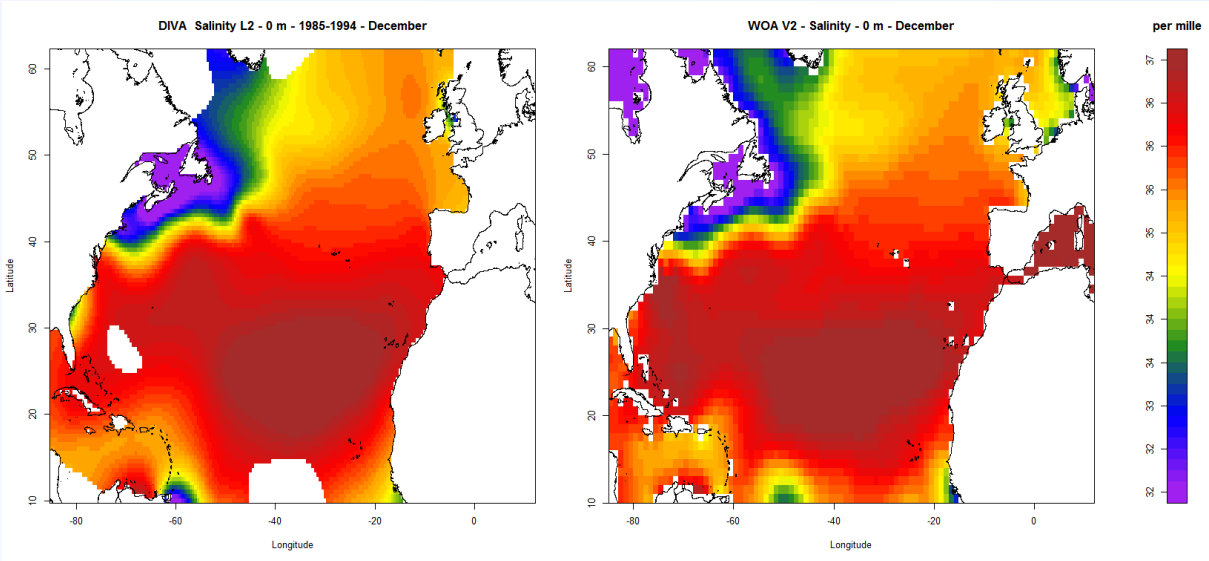
Consistency Analysis





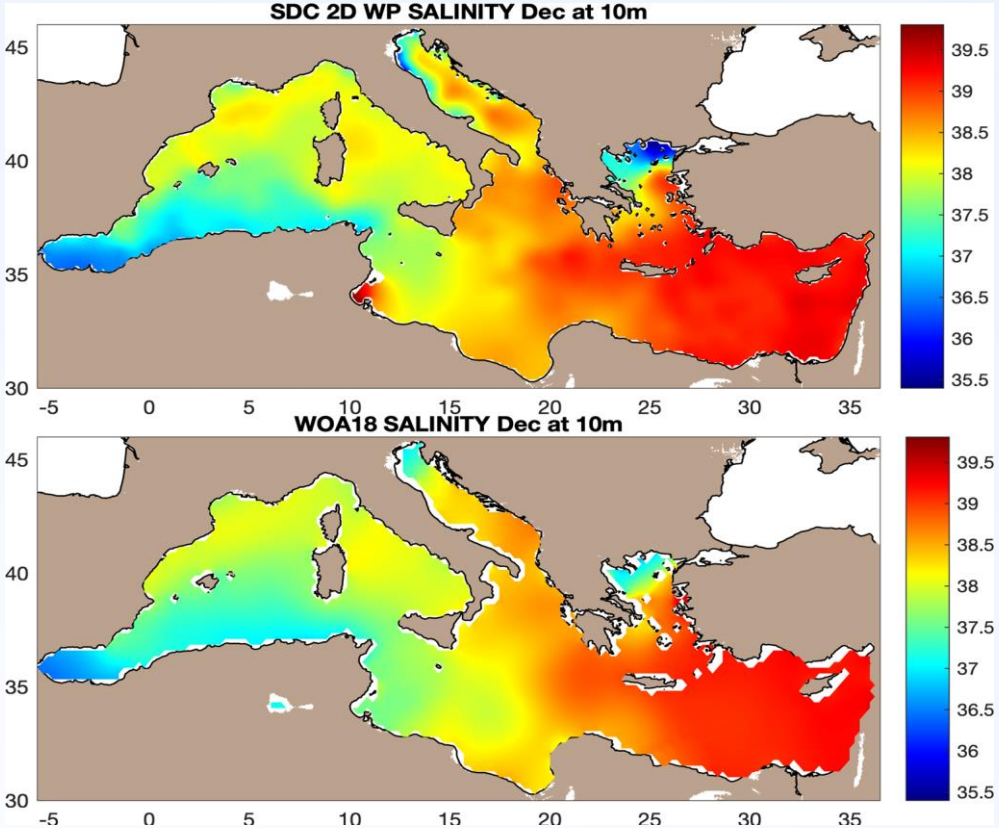
Surface

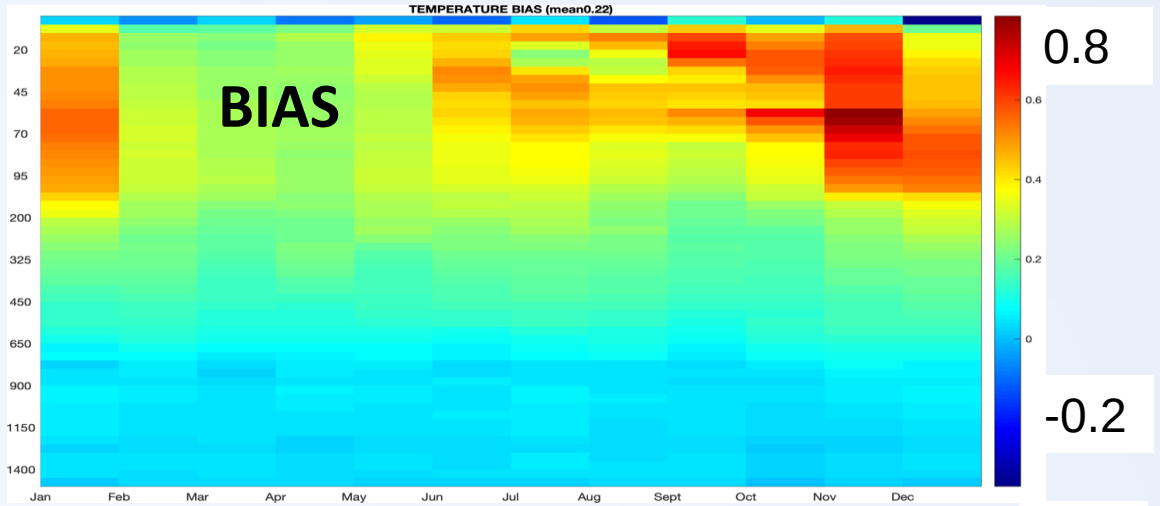
950m

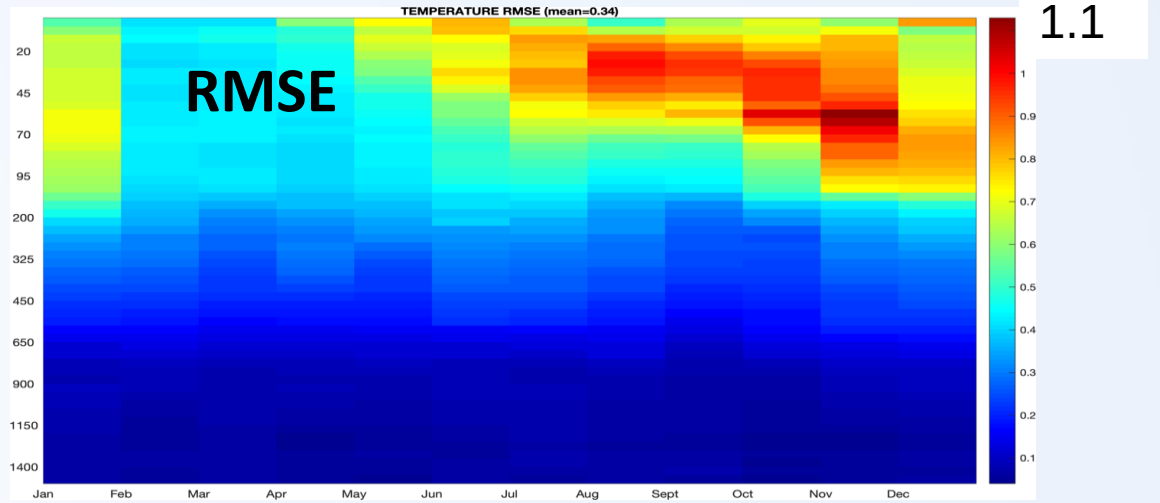


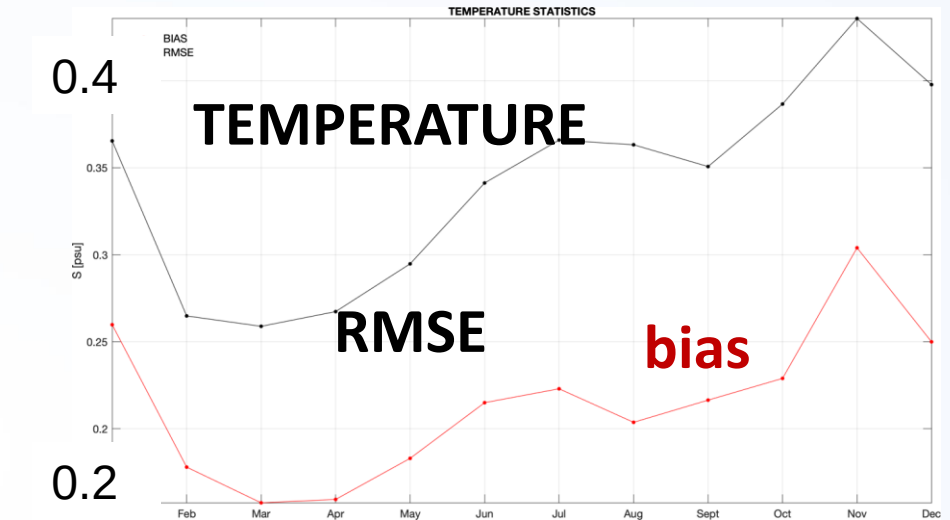
Surface Salinity (Jan)

Consistency Analysis









Conclusions and Future Work on climatologies

- Each climatology, its methodology and validation has been described in the Product Information Document (**PIDoc**)
- each PIDoc passed through a **double stage revision** (WP leader, ULG team, METU) in order to assure a good quality product and documentation → increase user confidence and uptake
- good DIVAnd uptake (5/7) and massive DIVAnd testing which made the code more robust even if some issues were identified and will be handled in the next year
- QC during the data integration process will be improved, together with duplicate detection (improve metadata track of external data)
- A posteriori QC needs to be optimized
- Consistency analysis will consider each field and it will be optimized and standardized

SeaDataCloud

New data harvest: preliminary analysis

sdn-userdesk@seadatanet.org – www.seadatanet.org

Harvesting and Aggregation

STEP 1 HARVESTING

Jul 31st → harvested T&S from the latest SeaDataNet CDI catalogue:

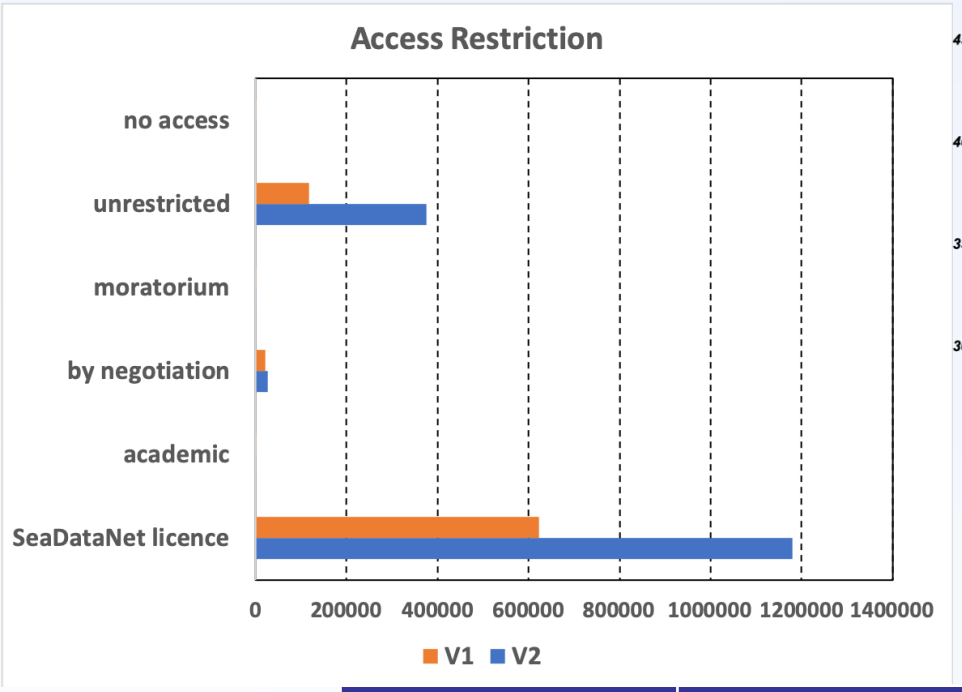
- subset of data sets that are **New** or **Updated** compared to the last harvest at Oct 30th 2017, divided over unrestricted and restricted

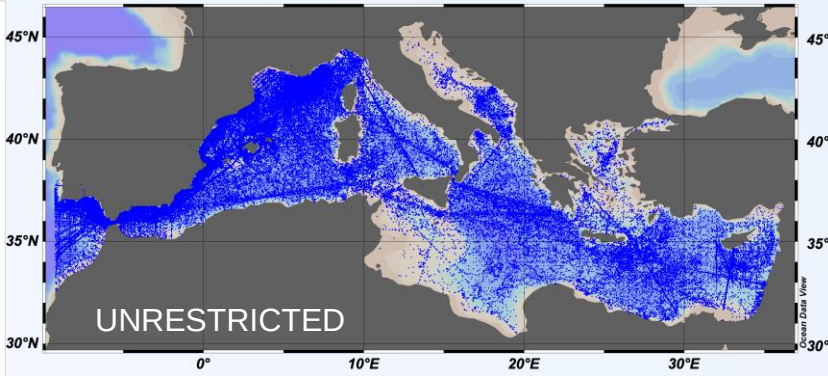
STEP 2 AGGREGATION

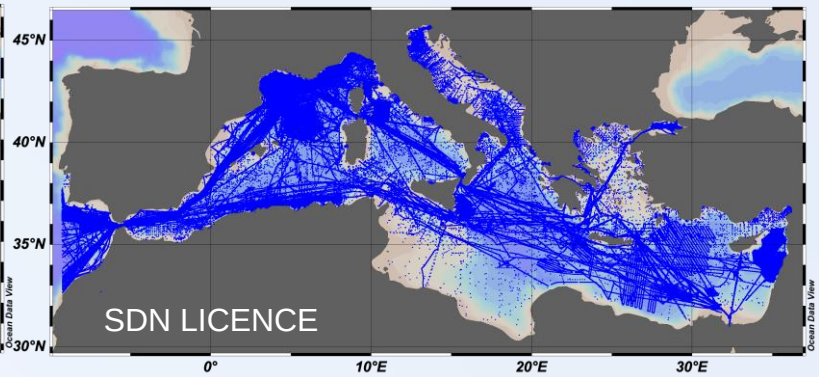
File and parameter aggregation of the "new_and_update_after_20171030" subset

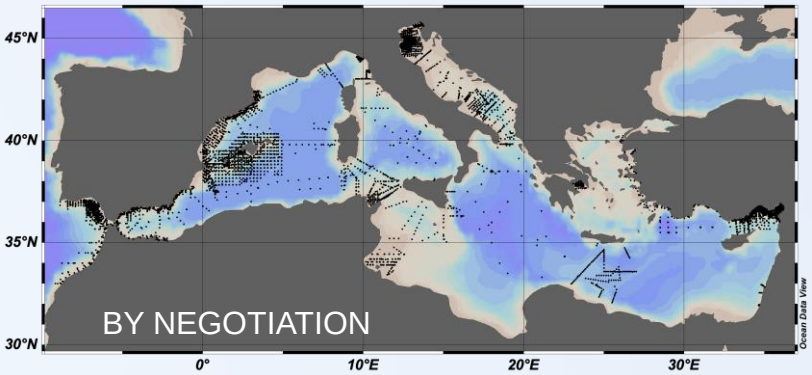
- split the results into regional collections and merged the harvested regional data with the SDC V1 regional collections
- analysis log files generated during import → error and warning messages → **there are still files with serious issues waiting to be corrected by the data centers**

SDC_MED_DATA_TS_V2





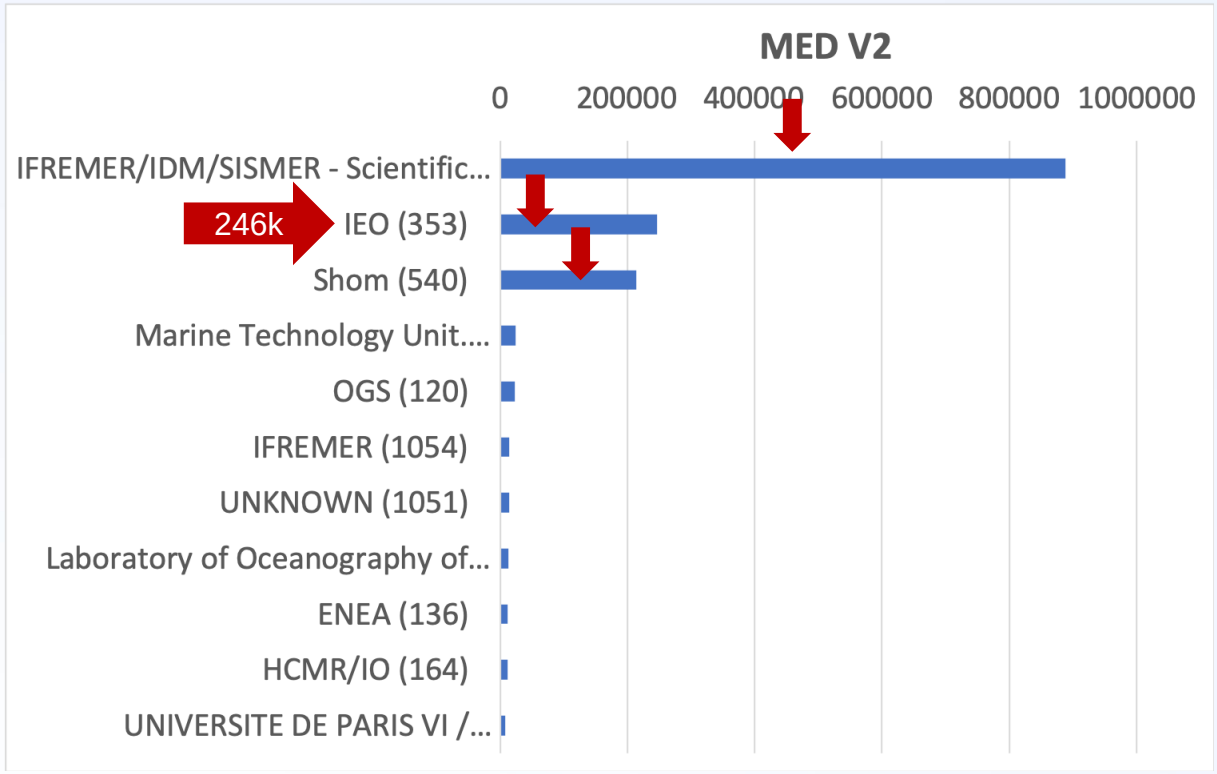


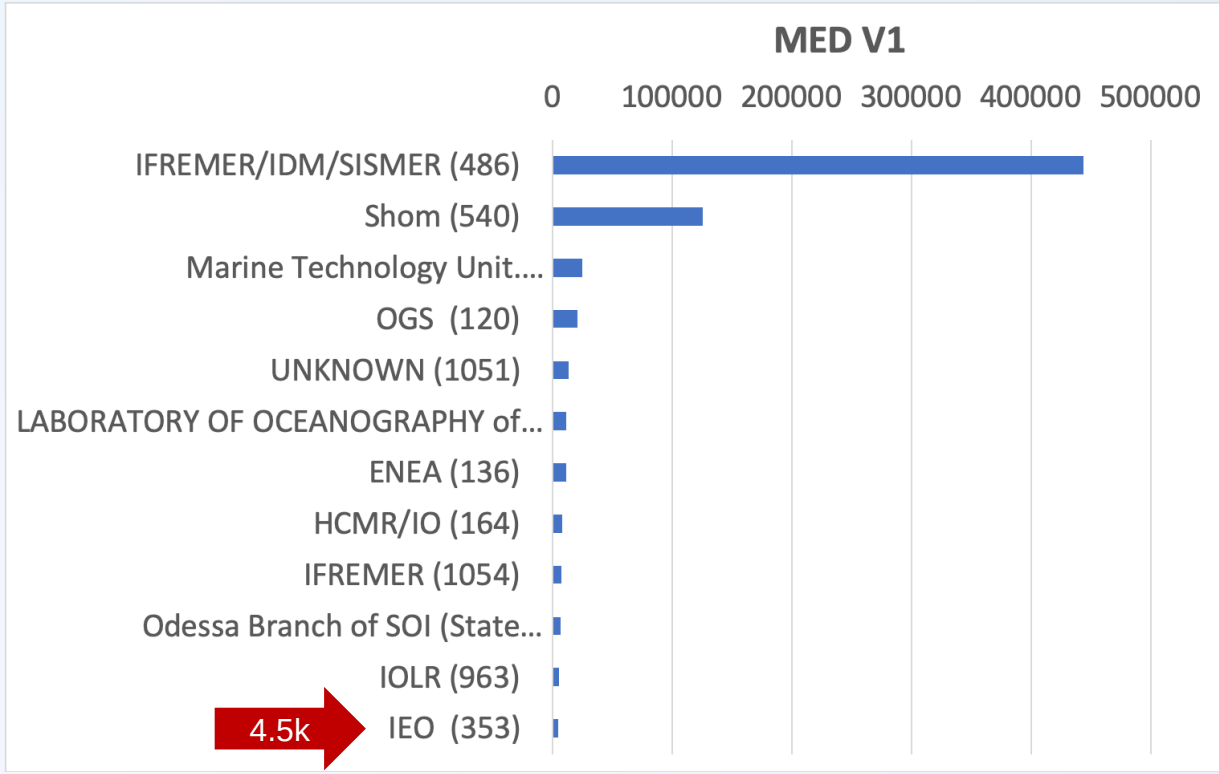


	V1	V2	% increase
free access	739784	1554490	+110
limited access	22221	27673	+25
TOT	762005	1582163	+108

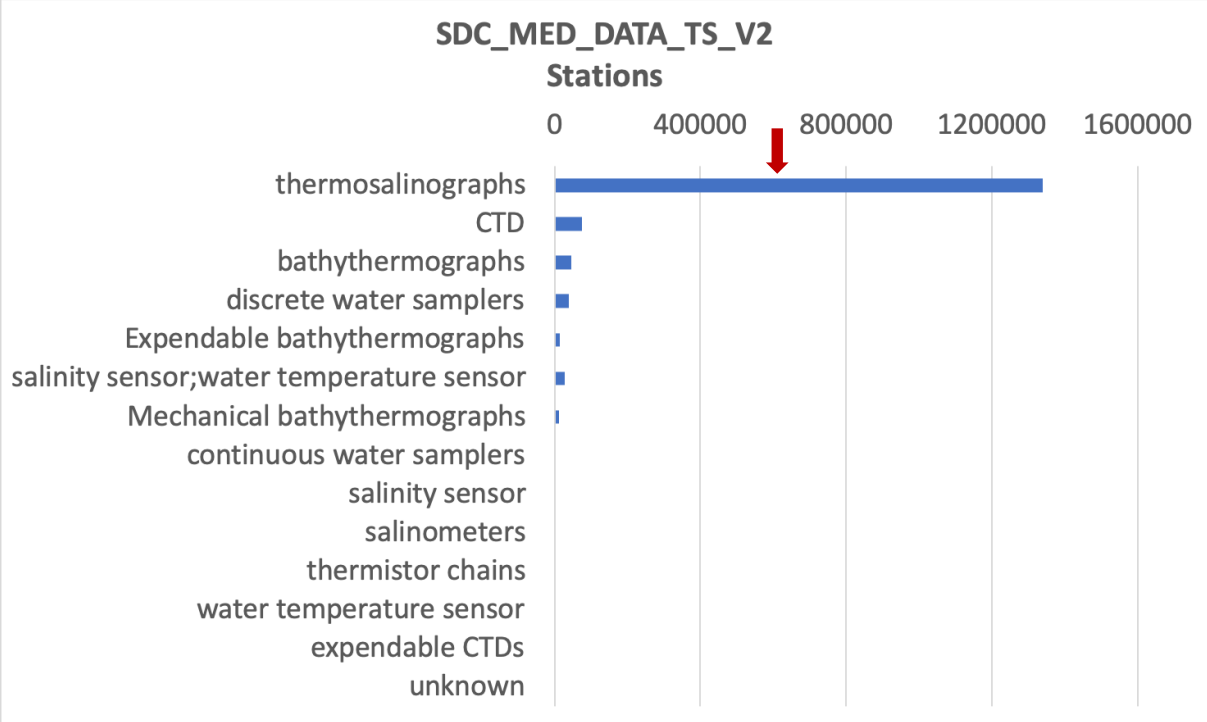
2% of data are restricted

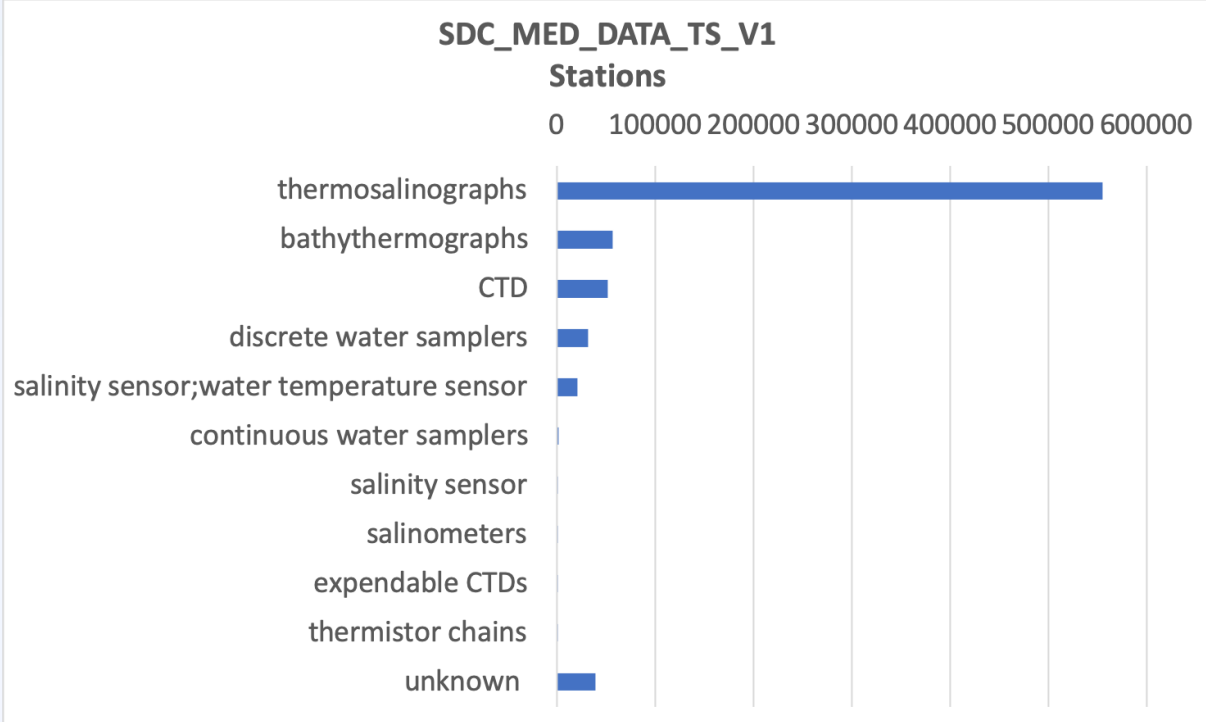
SDC_MED_DATA_TS_V2





Instrument/Gear Type Stats MED

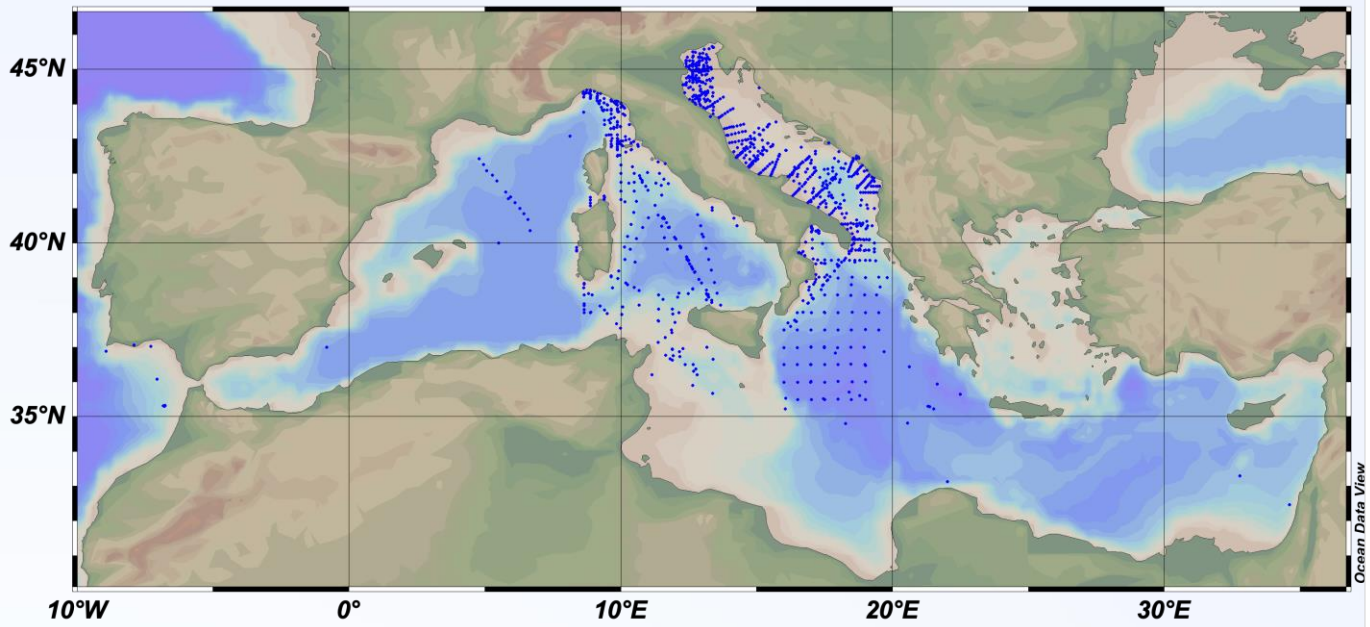


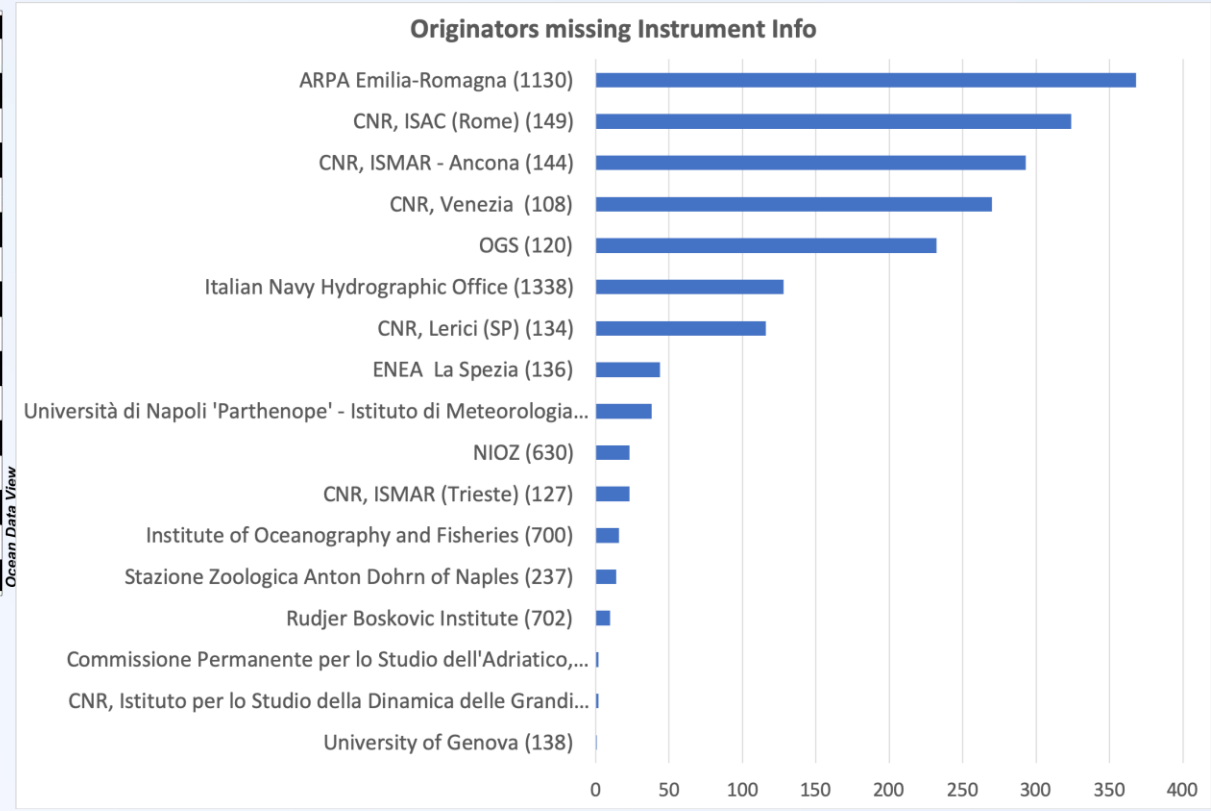


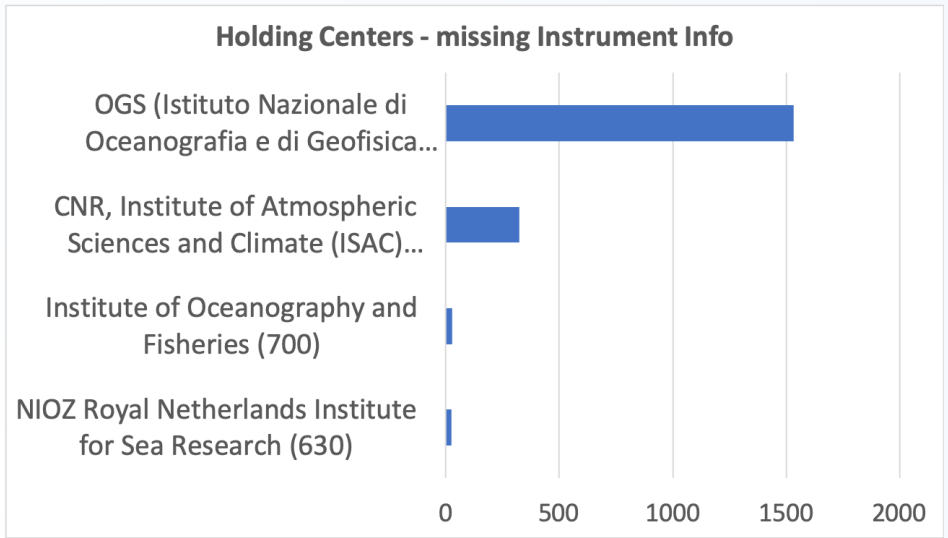
Instrument / gear type: 43 unique string values	Stations	%
thermosalinographs	1339074	86,1
CTD	75849	4,9
bathythermographs	44935	2,9
discrete water samplers	38937	2,5
Expendable bathythermographs	13069	0,8
salinity sensor;water temperature sensor	27047	1,7
Mechanical bathythermographs	11329	0,7
continuous water samplers	1887	0,1
unknown	1890	0,1

Instrument / gear type: 36 unique string values	Stations	%
thermosalinographs	555269	75,1
bathythermographs	56558	7,6
CTD	52031	7,0
discrete water samplers	31559	4,3
salinity sensor;water temperature sensor	20529	2,8
continuous water samplers	1577	0,2
unknown	39494	5,3

Missing Instrument/Gear Type MED

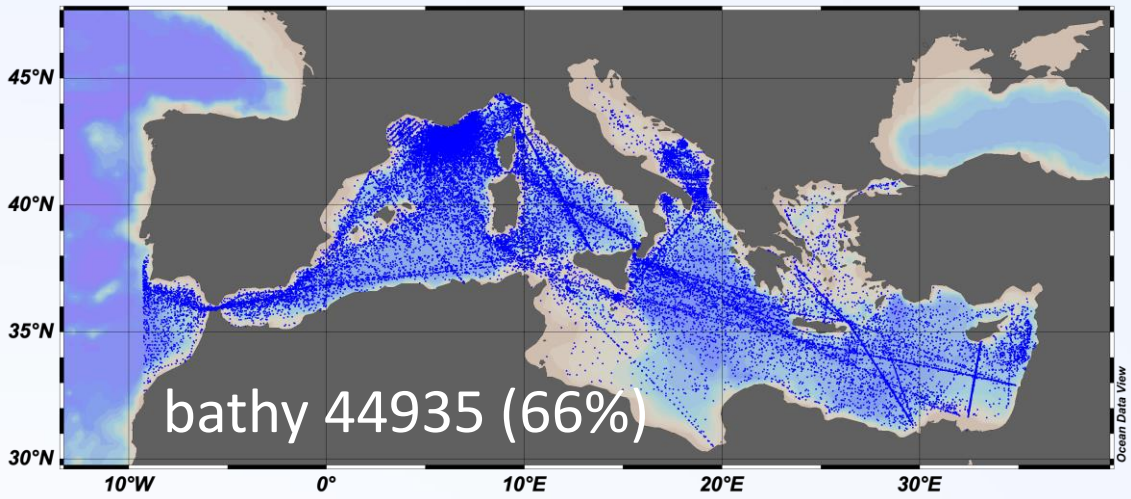


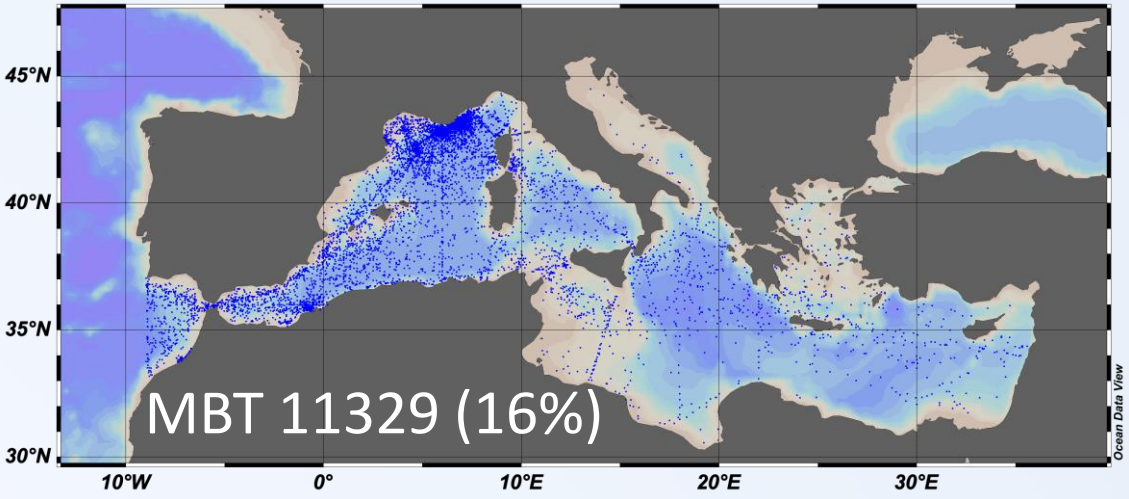


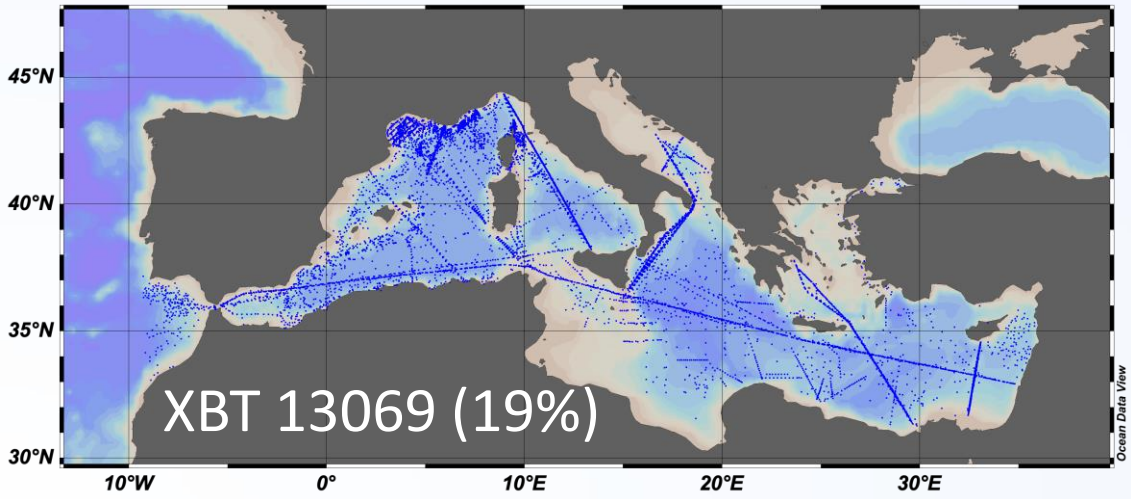


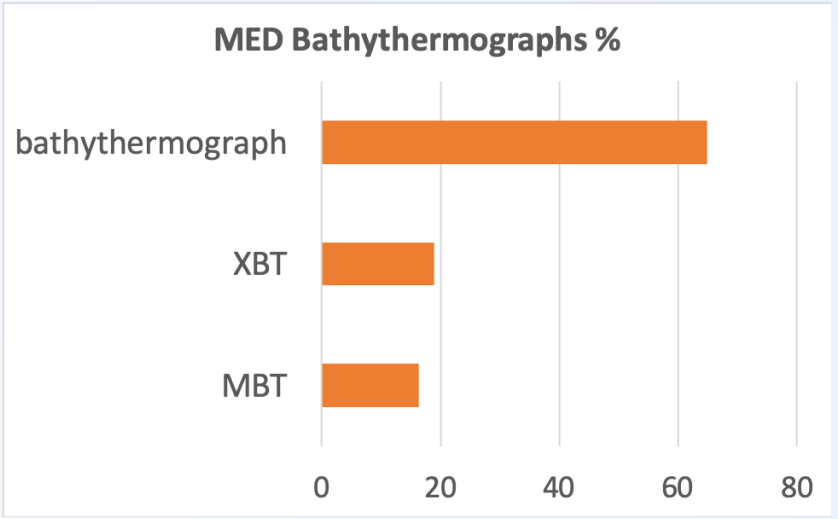
MED XBT

V1	V2	
56558	69333	+23%



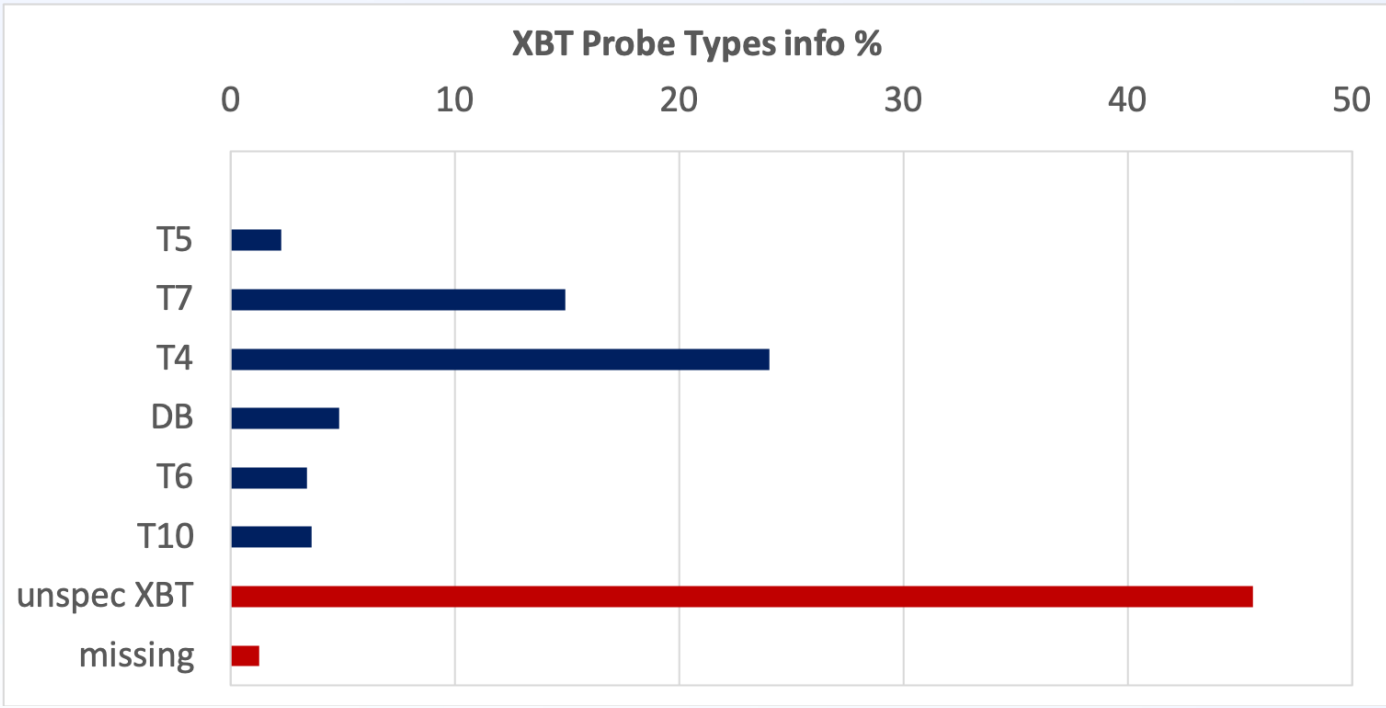




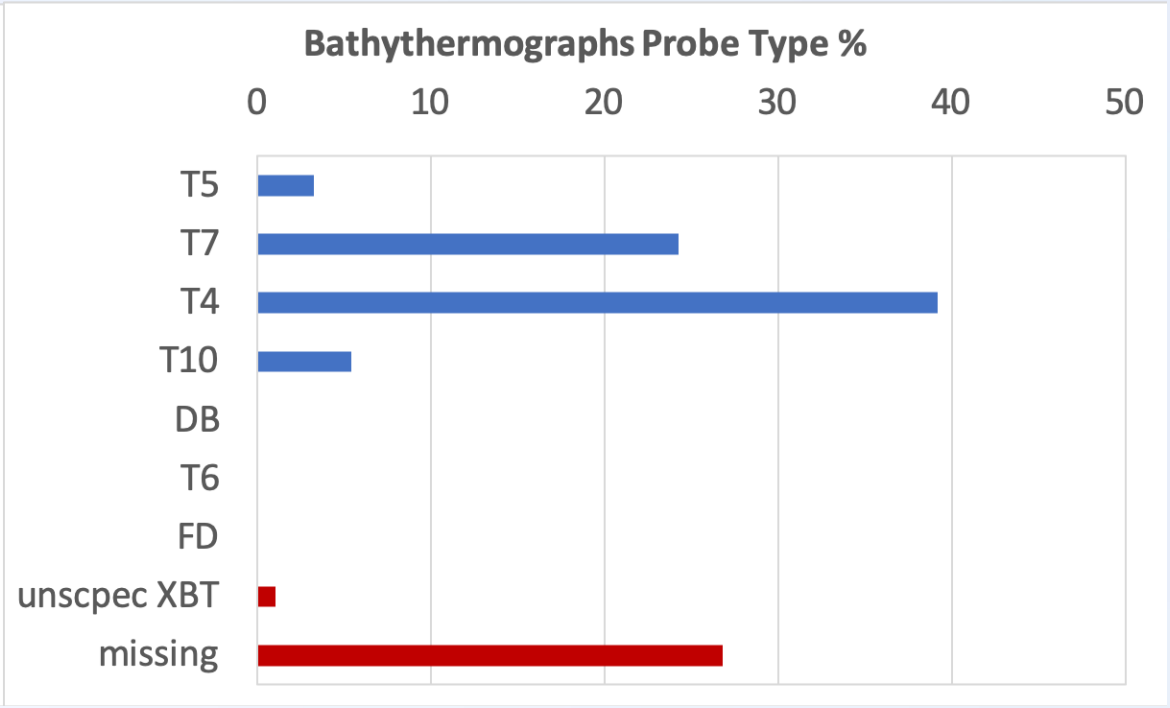


MED XBT

16% of the total BT data



66% of the total BT data



Unspecified or Missing Info

MBT	11329	16%
XBT	6126	9%
BT	12426	18%

27%

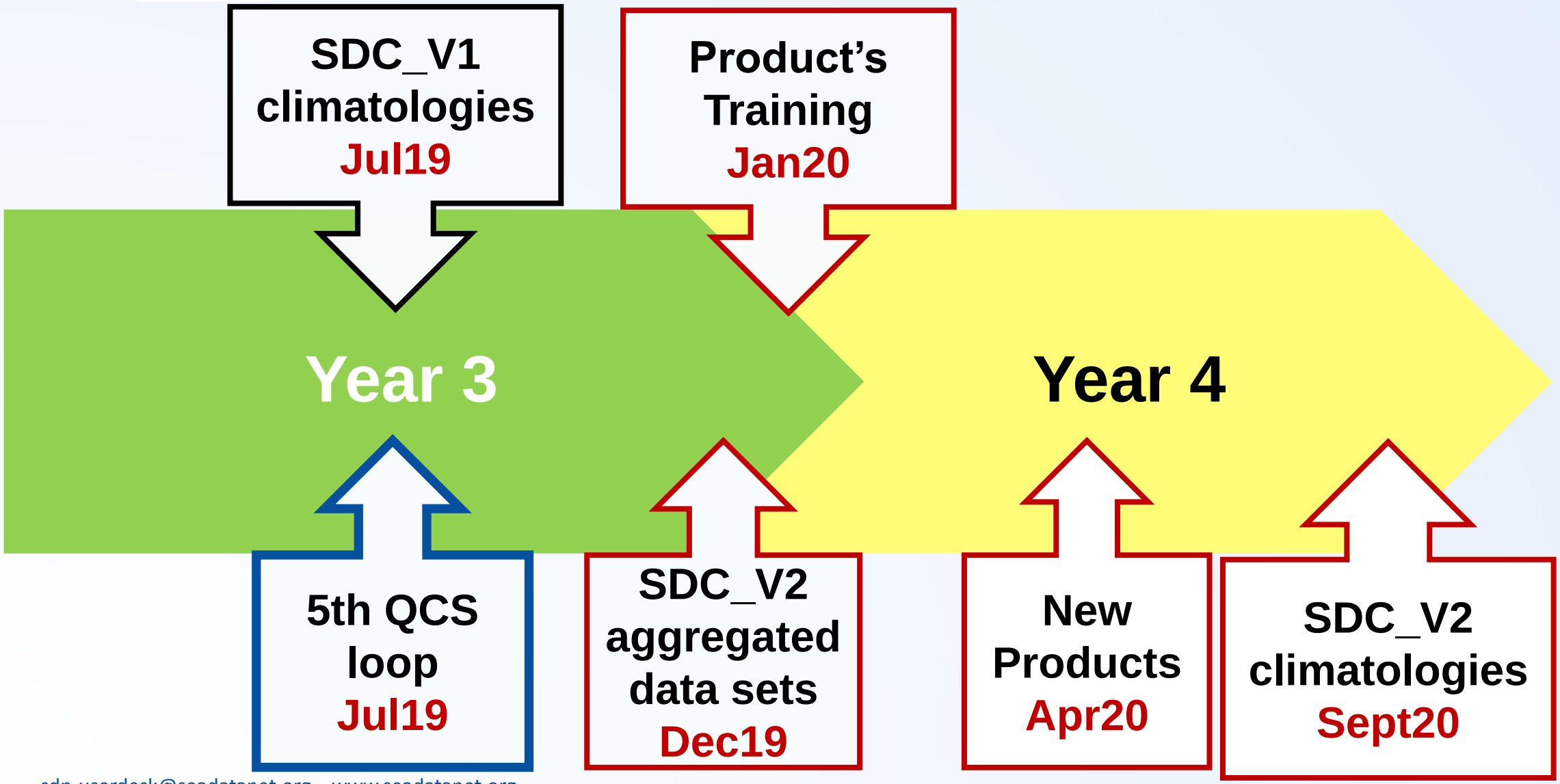
43%

V1 was missing 46% of probe type info

Conclusions and Future Work

- Data aggregation was performed only on new and updated CDIs for the fist time and data combined with V1 → saving time
- V1 PIDocs will be updated with V2 assessment results → saving time for further checks
- ODV developments are progressively addressing User needs providing new functionalities increasing QA/QC efficiency
- metadata need to be populated in order to improve products' quality and the relating scientific results (i.e. Instrument Type, …)
- final V2 assessment due by Dec 15 → release on the catalog Jan20

Work Plan and Timeline



SDN and IQuOD

- results from the next data integration process will keep track of unique stations identifier in order to provide feedback on the detected anomalies and duplicates
- support on the integration of SDN data into IQuOD but we need a feedback too
- data rescue activities will aim to ingest data and metadata omissions
- Work on Med XBTs → to define the best data set with the most complete metadata description and get a regional correction → this could be done similarly for other data types and regions

sdn-userdesk@seadatanet.org – www.seadatanet.org

THANK YOU

EGU General Assembly 2020

Vienna | Austria | 3–8 May 2020

ABOUT AND SUPPORT ABSTRACTS & PROGRAMME REGISTER & VENUE EXHIBITION GUIDELINES LOGINS

ESSI1.1

Informatics in Oceanography and Ocean Science

Co-organized by OS4

Convener: Antonio Novellino | Co-conveners: Luca Bonofiglio[ECS], Cristian Munoz[ECS], Simona Simoncelli

Abstract submission