V1 Climatologies and external data integration

S. Simoncelli and WP11 team

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

3rd SDC Plenary Meeting, October 18th 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

Regional Climatologies

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

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

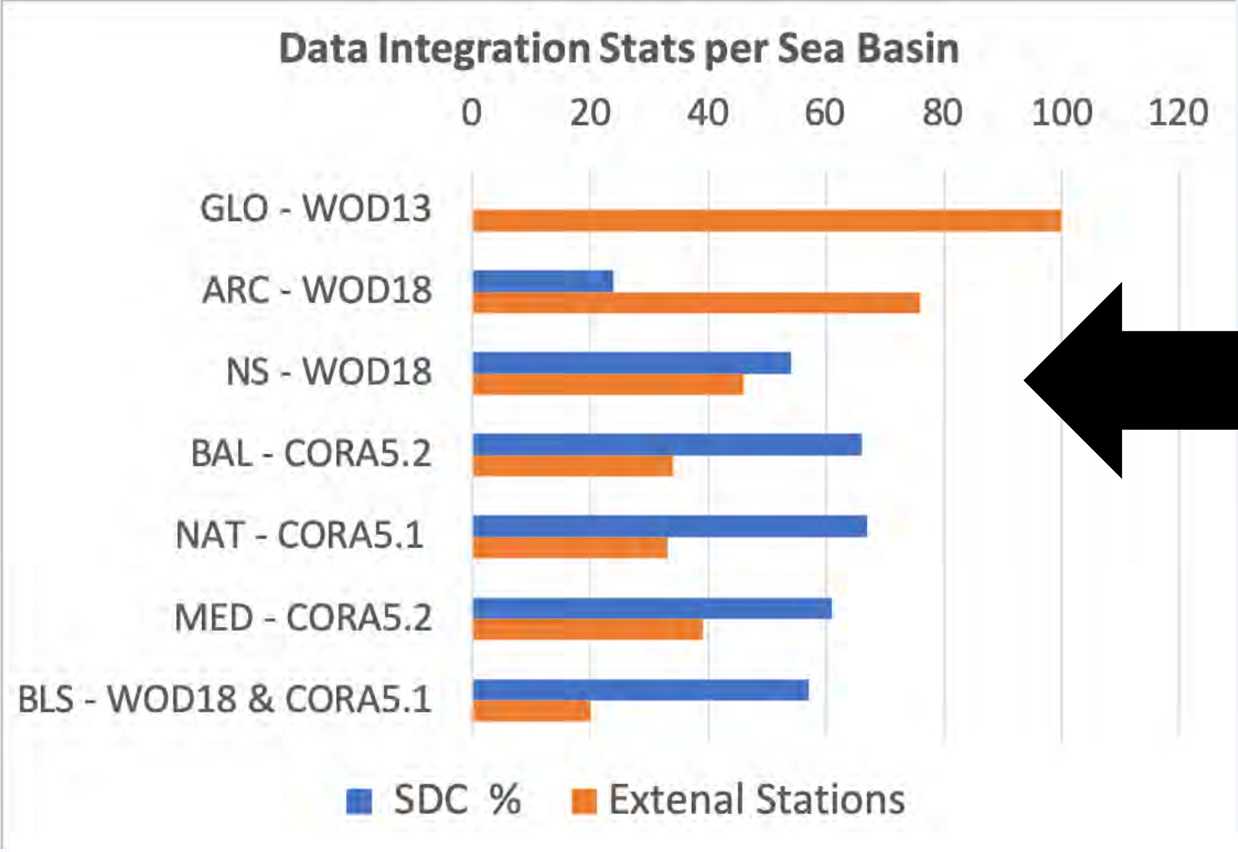
→ two different monthly climatological fields for T and S with a different time coverage, computed from WOD data since spatial coverage of SDN data at global scale is still too sparse

Next releases the idea is:

→ to integrate SDN and WOD18 data

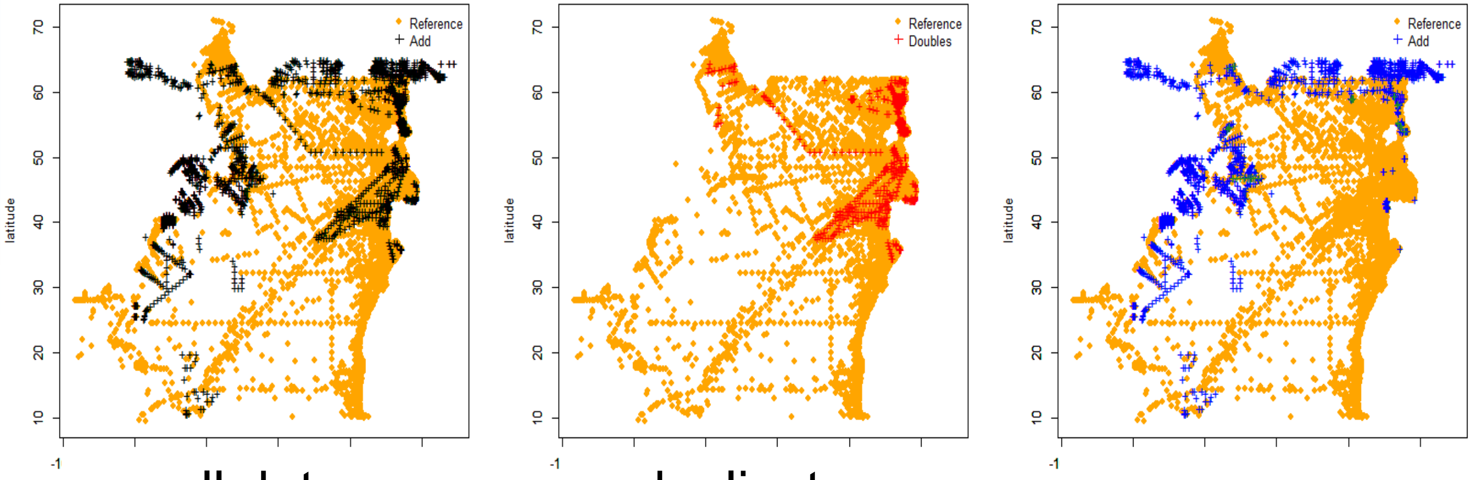
→ to improve the background estimate and maximize product's quality

Input Dataset



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

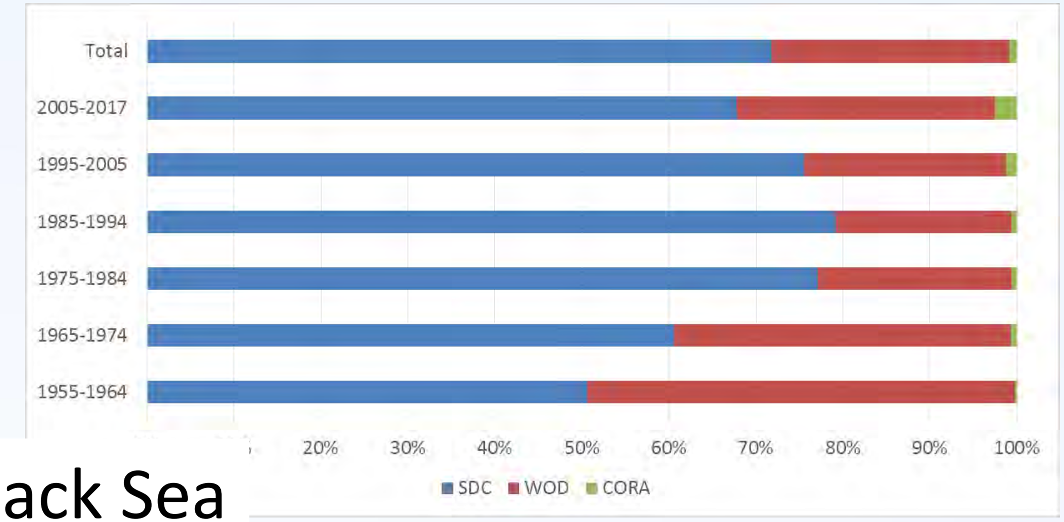
North Atlantic



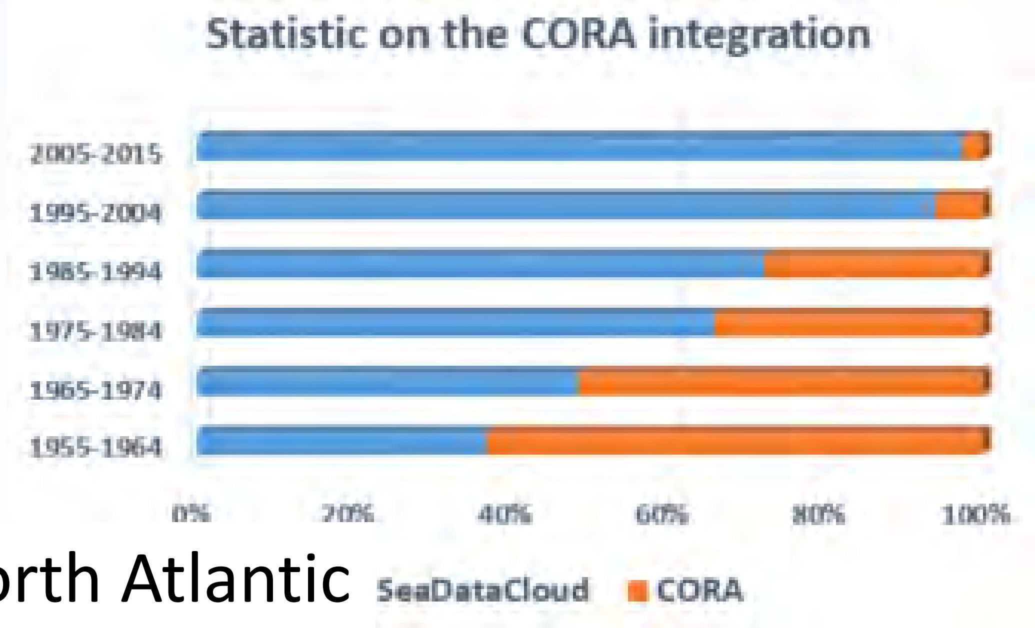
Duplicates

	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

Climatology data sets

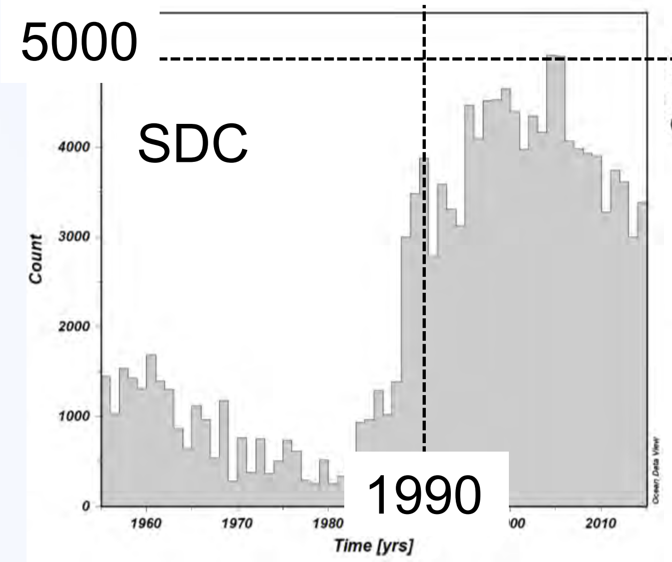


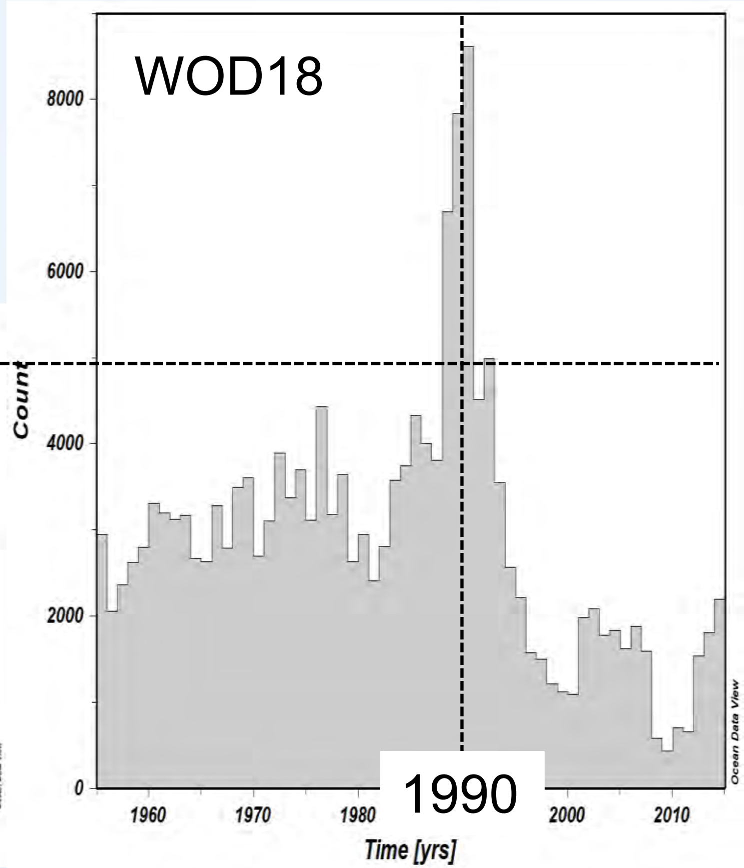
Black Sea



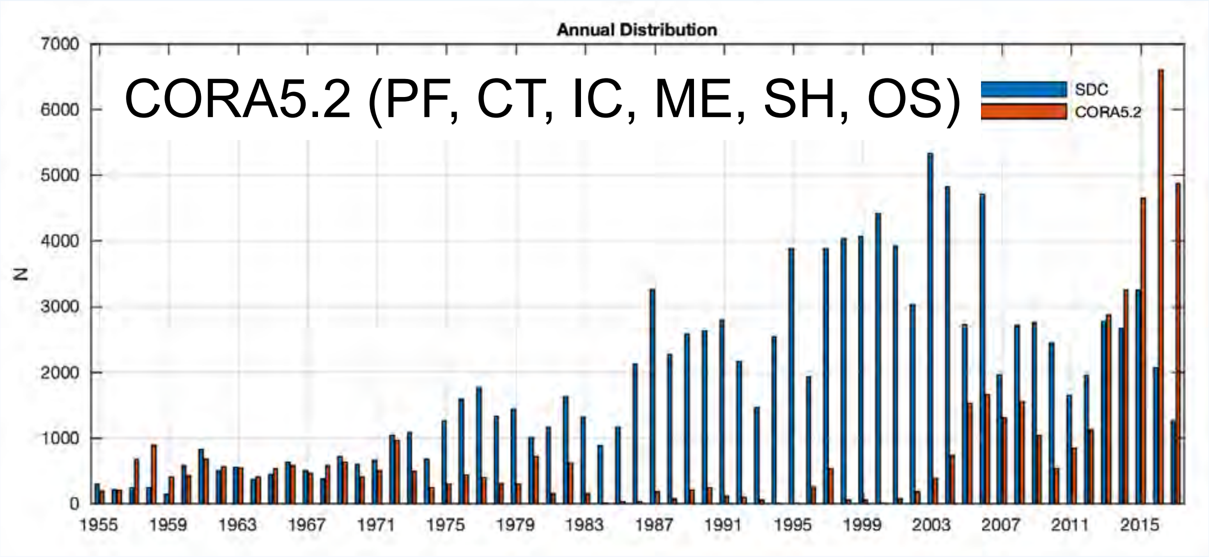
North Atlantic

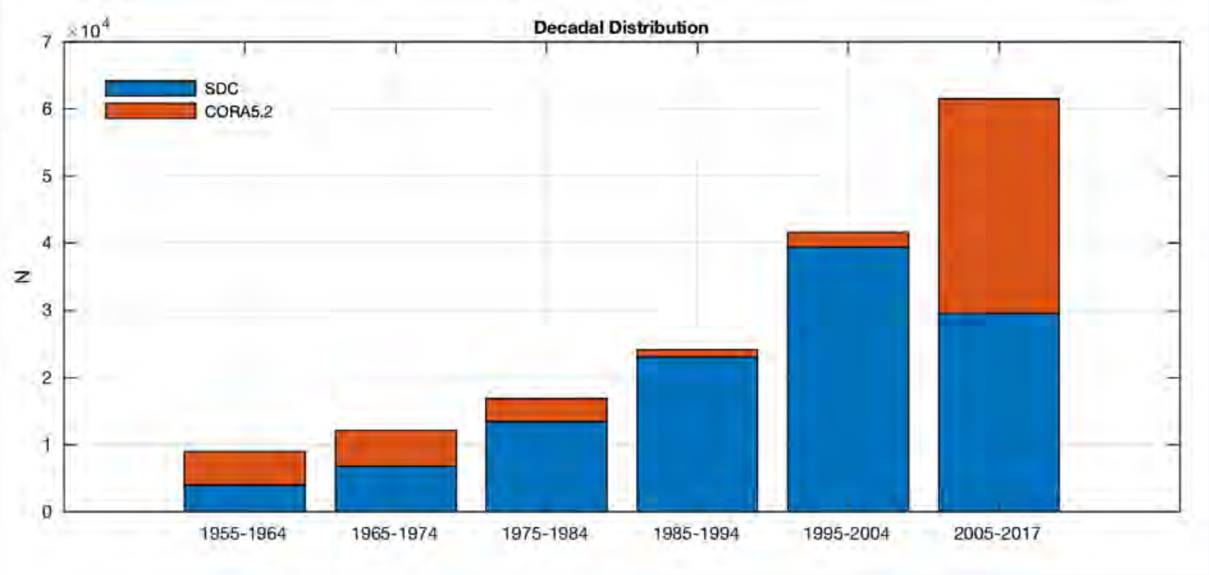
North Sea



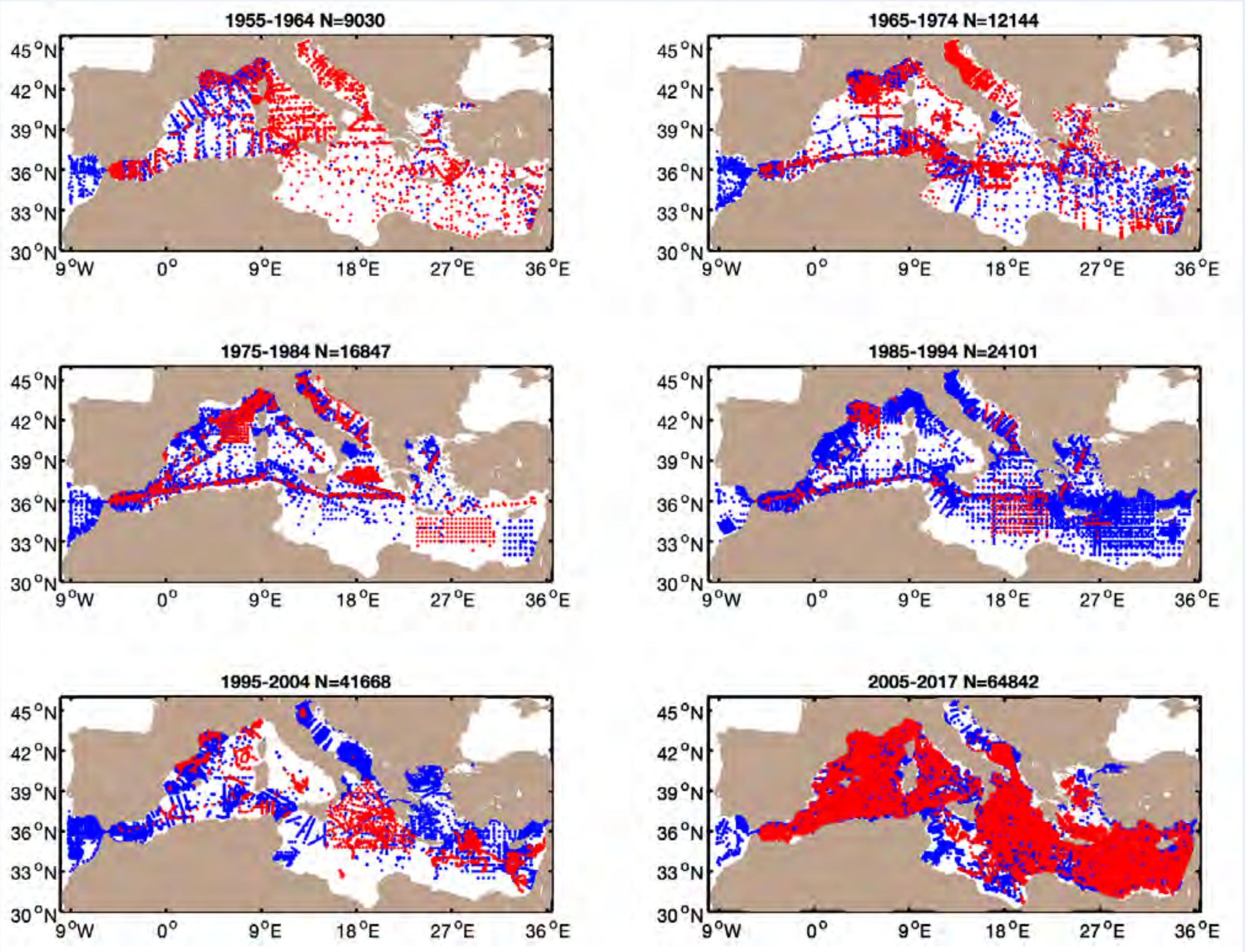


Integrated Dataset





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

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:

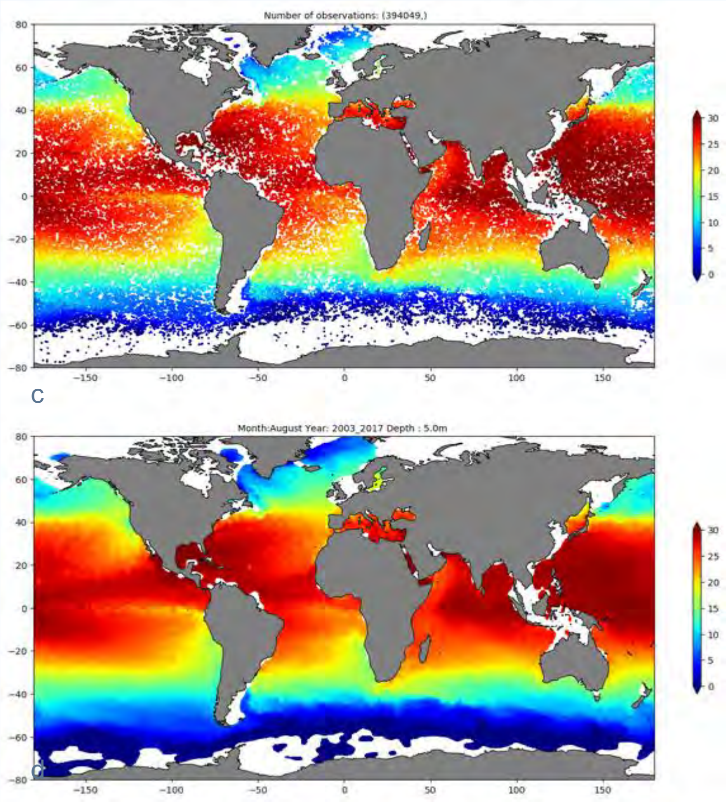
→ non-detected outliers

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

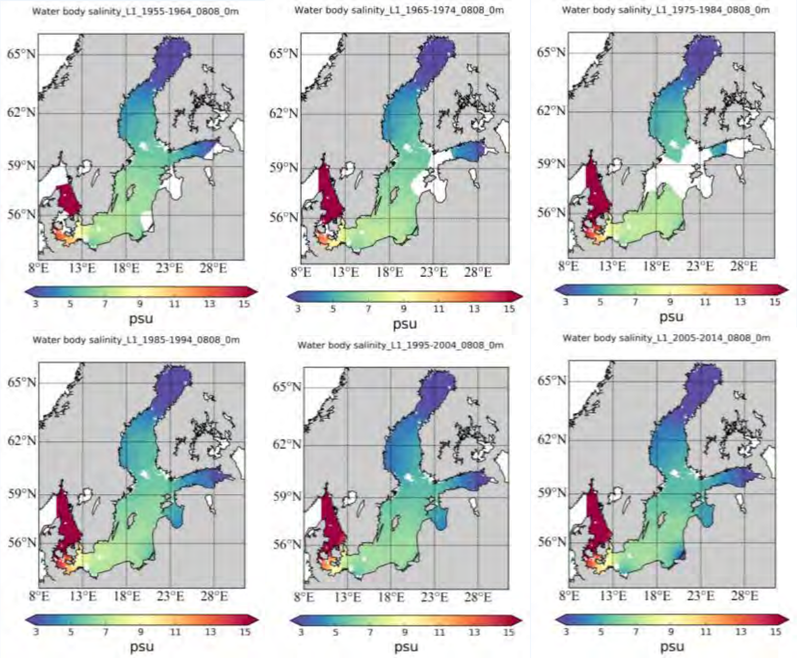
Manual removal or through DIVAnd tool which computes the **residuals** after the climatological fields generation. Iterative approach → residuals > pre-defined thresholds have been removed (BLS and MED).

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

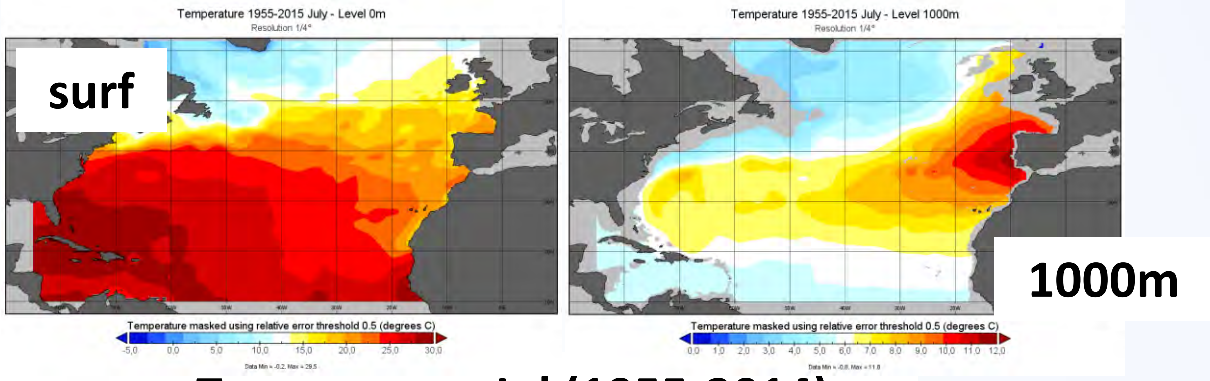
Results



Surface Temperature (Jan)

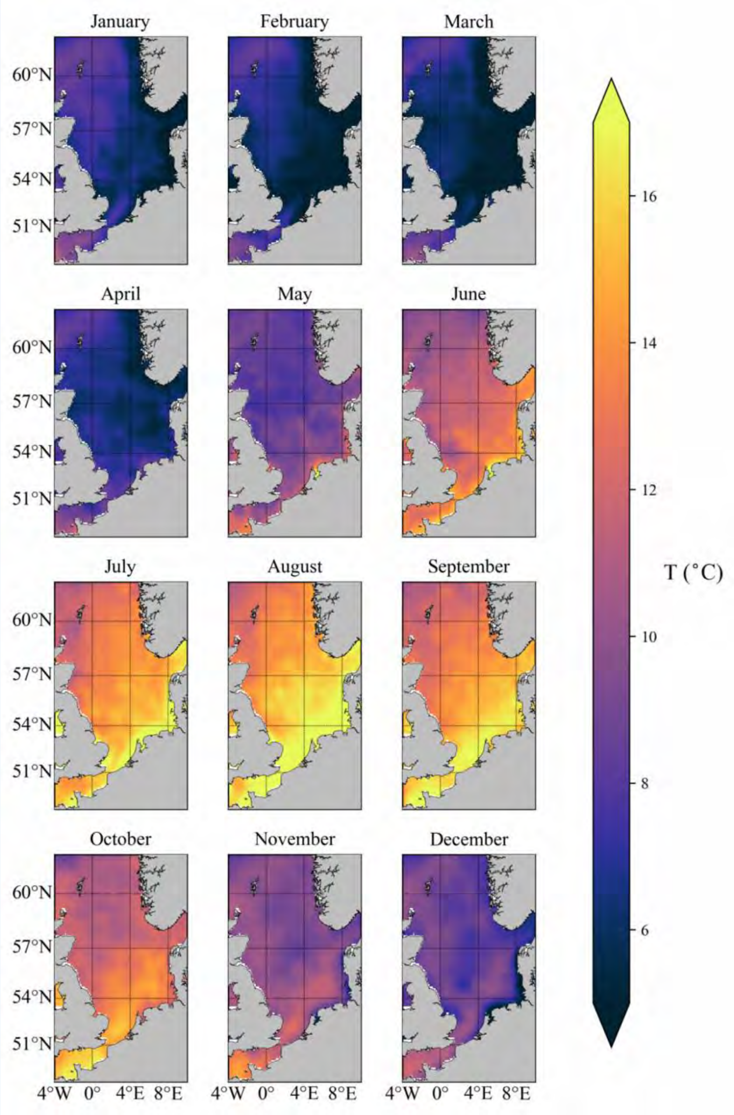


Surface salinity (Aug)

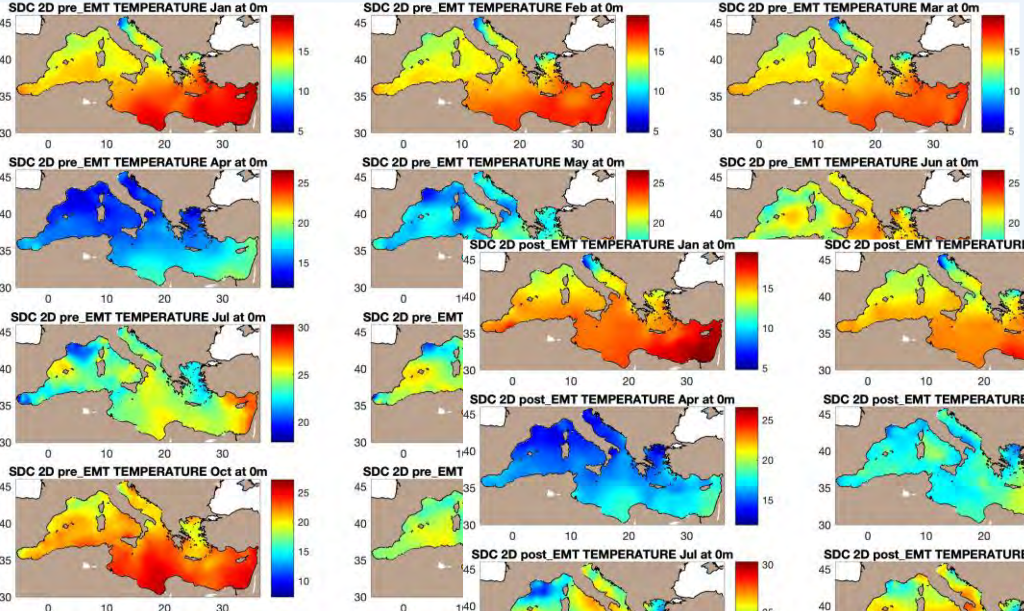


Temperature Jul (1955-2014)

Surface Temperature

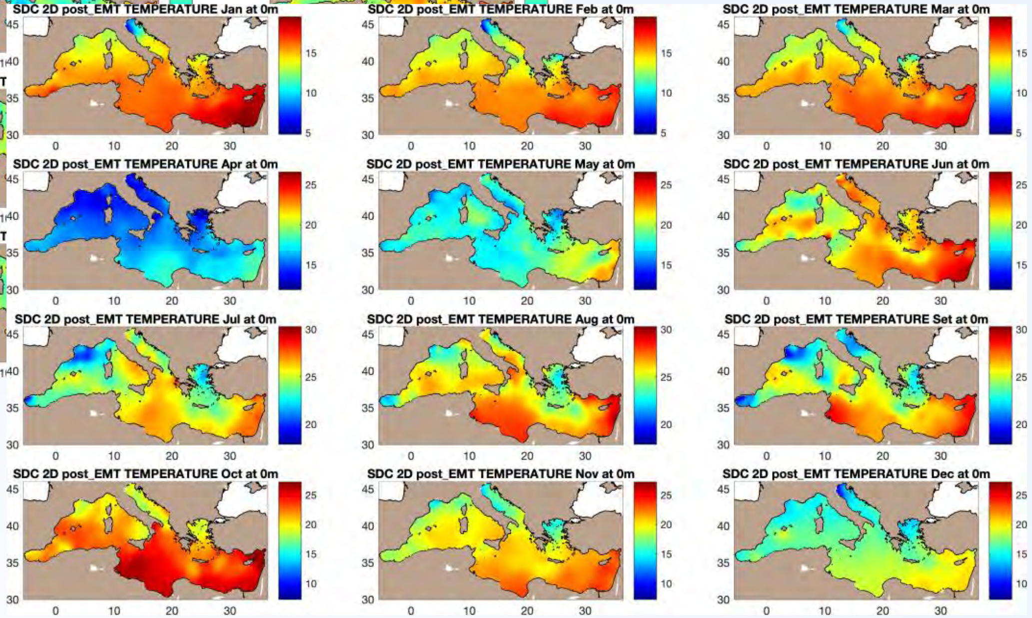


Results

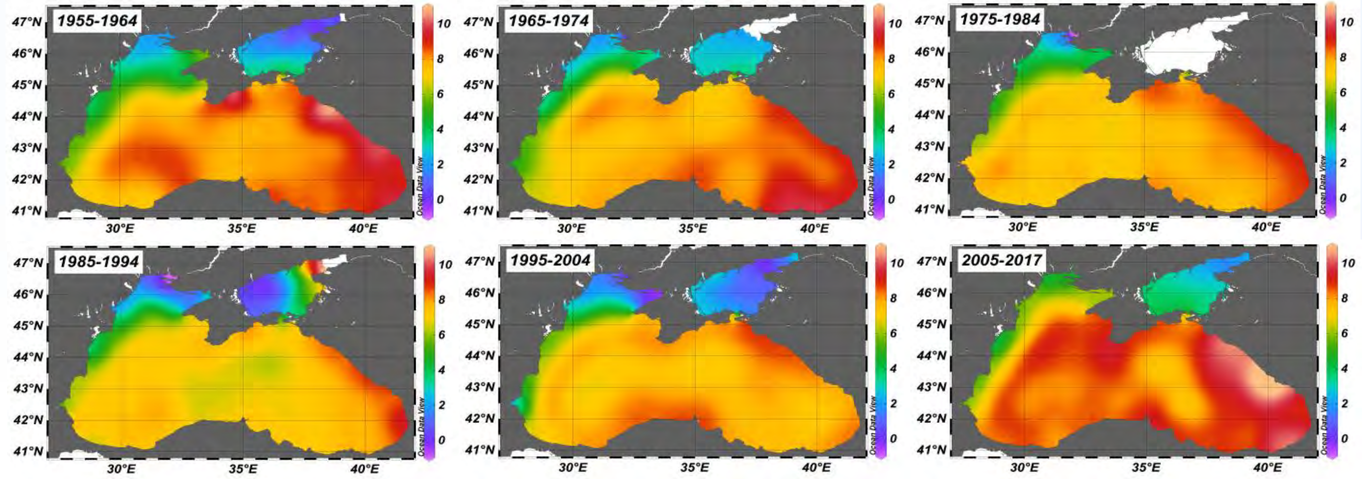


SST 1955-1984

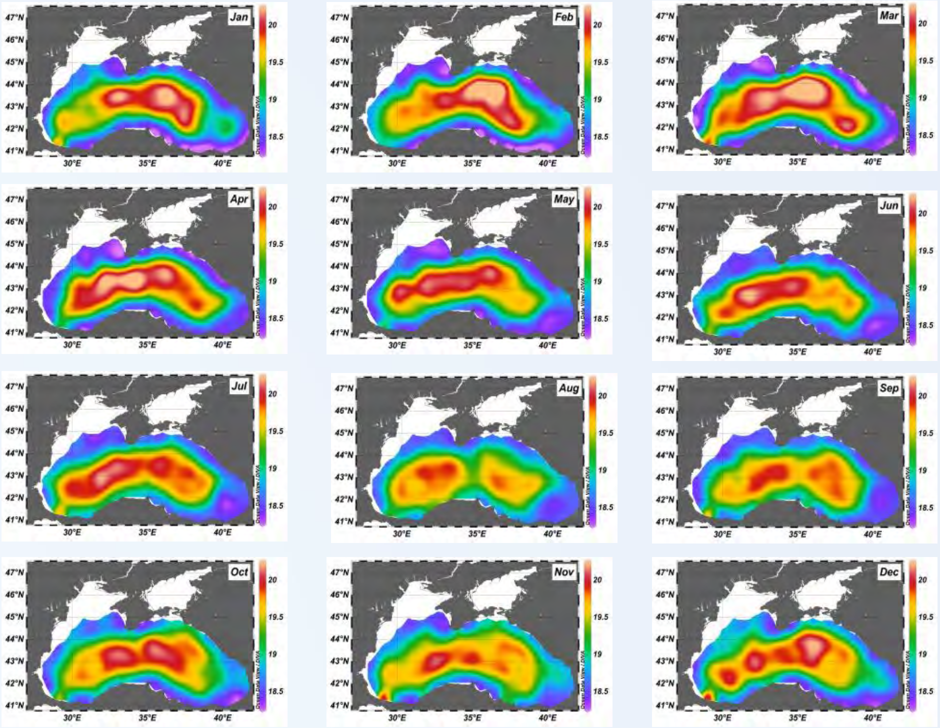
SST 1985-2017



winter SST

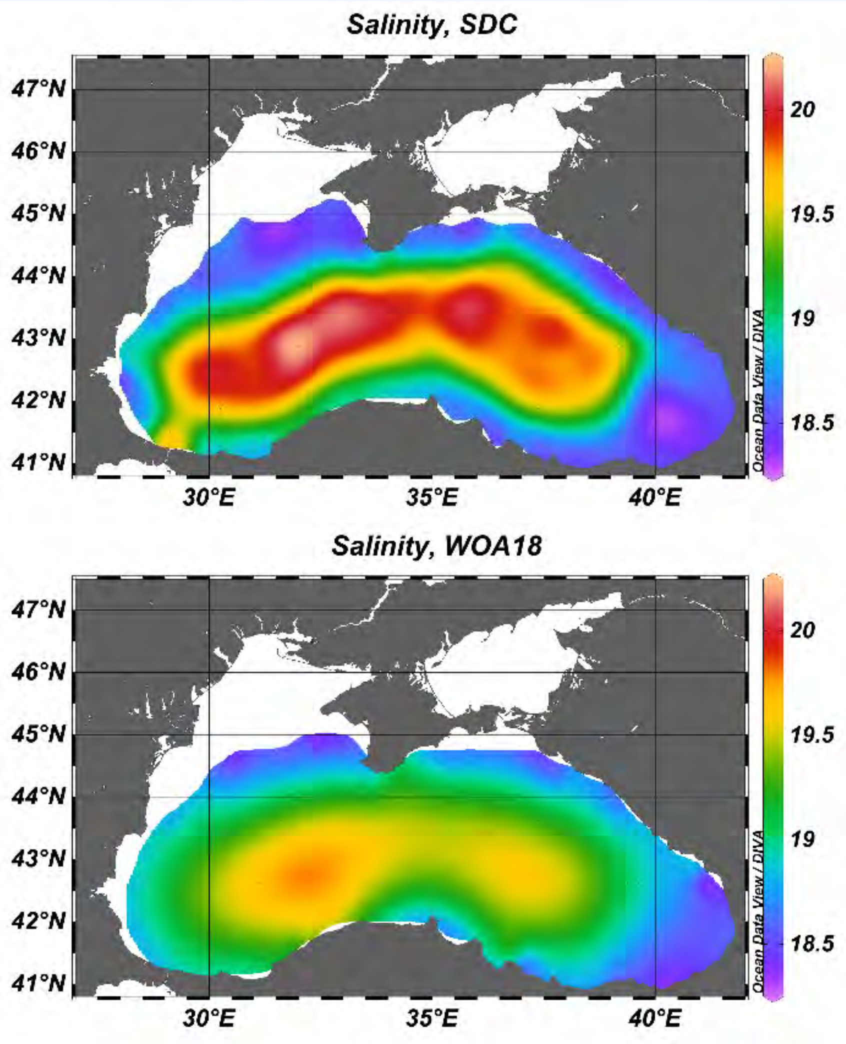


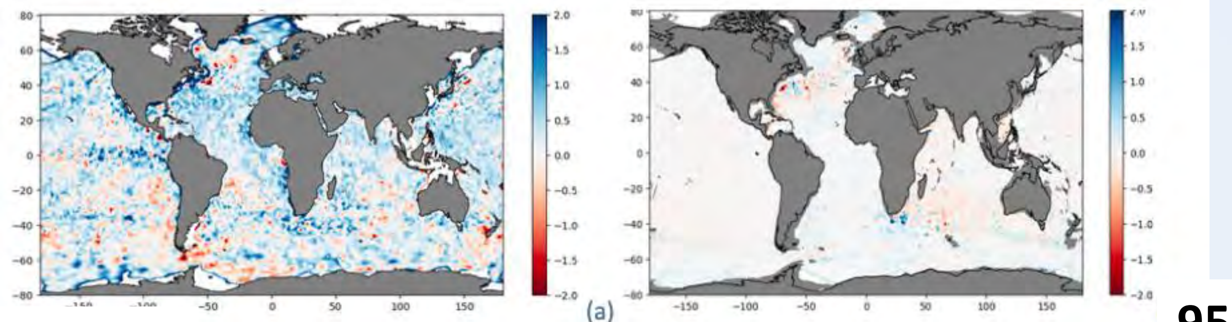
Salinity at 70m (1955-2017)



SeaDataCloud

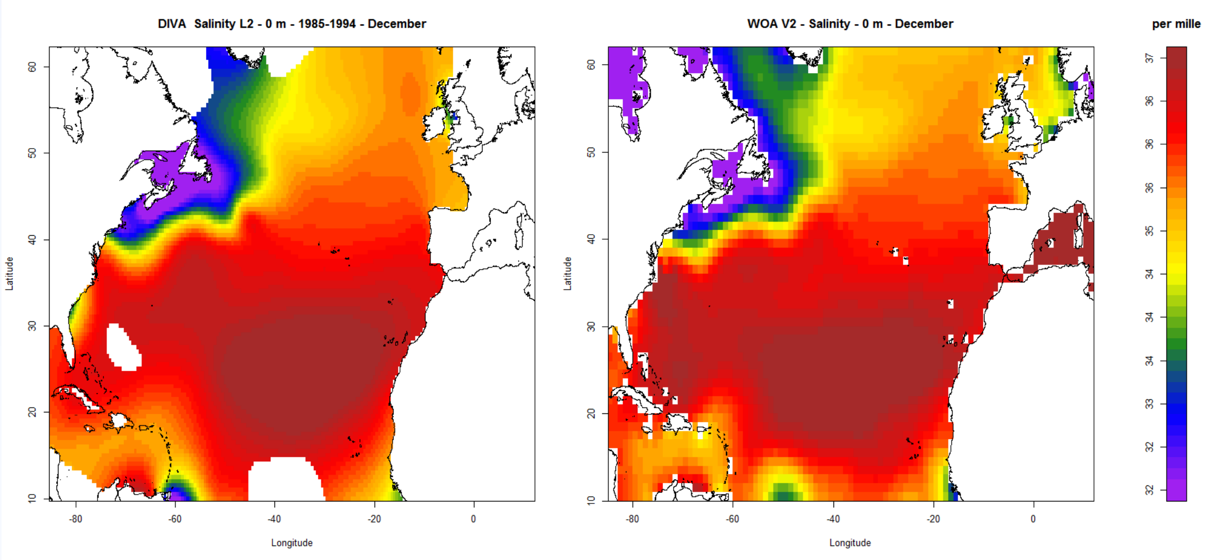
Consistency Analysis





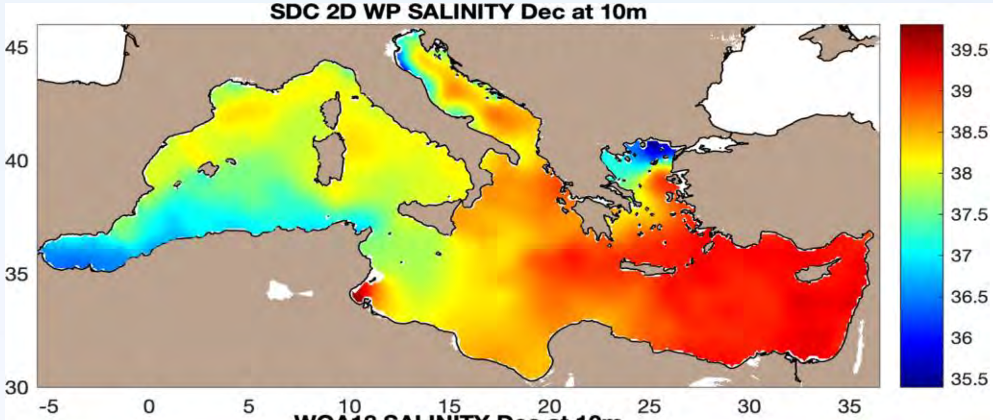
Surface

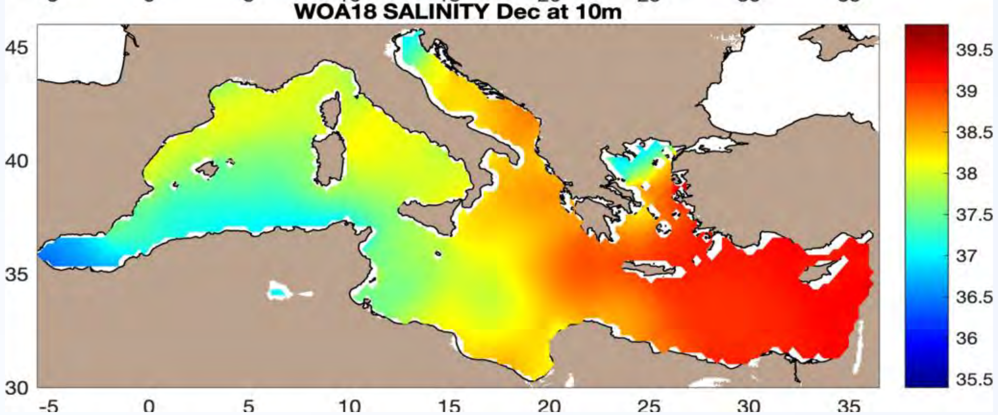
950m

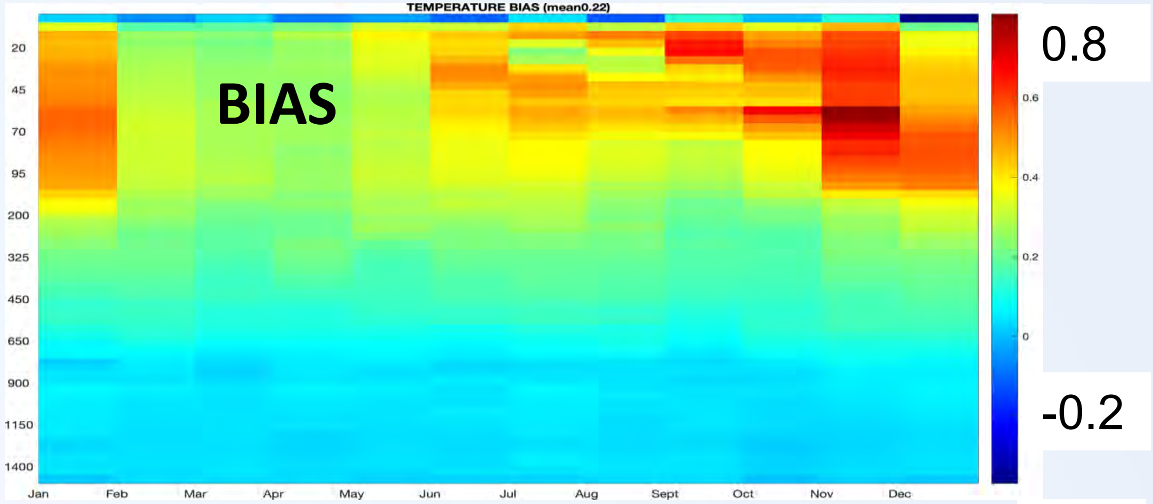


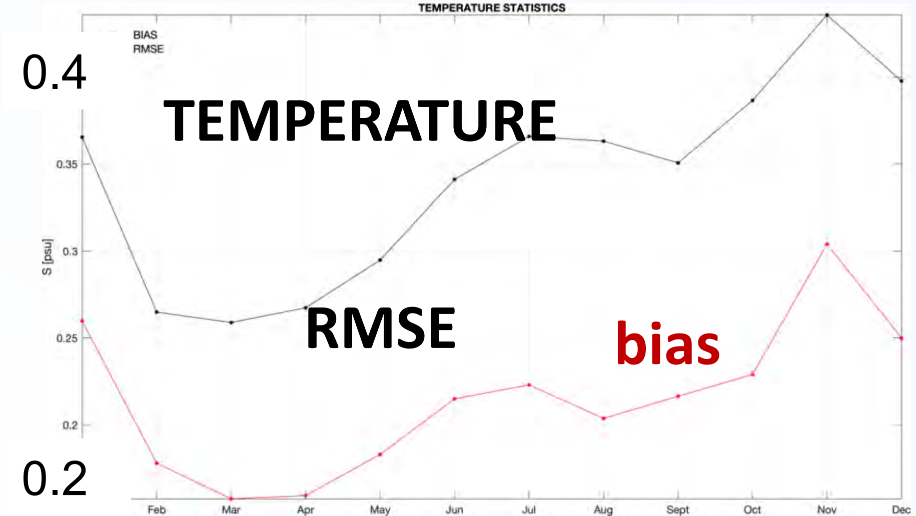
Surface Salinity (Jan)

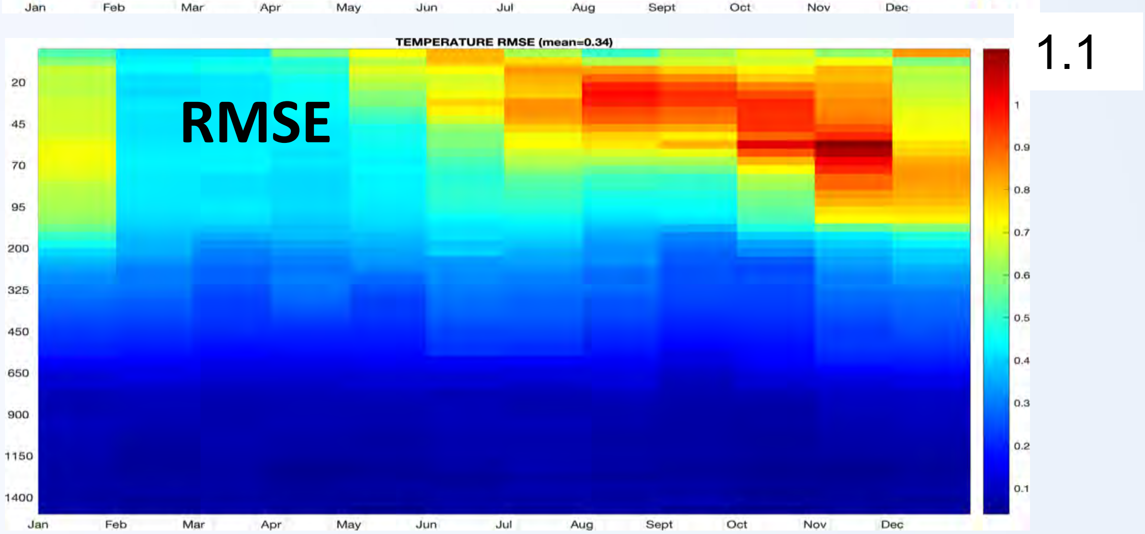
Consistency Analysis











Conclusions and Future Work

- 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 gridded field and will be optimized/standardized

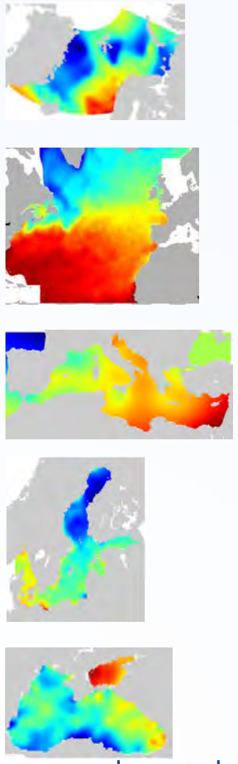
we NEED your FEEDBACK

New data harvest: preliminary analysis

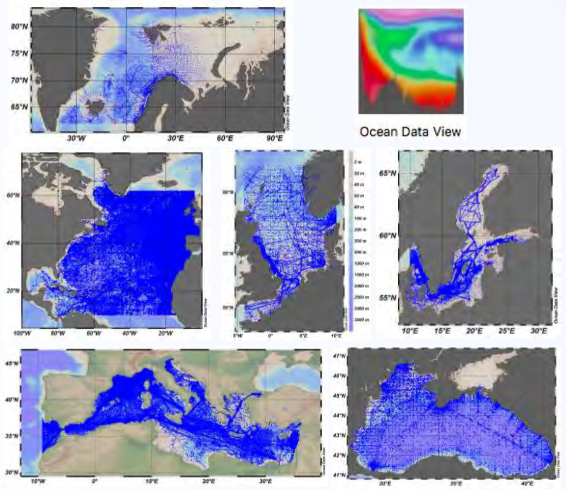
S. Simoncelli and WP11 team

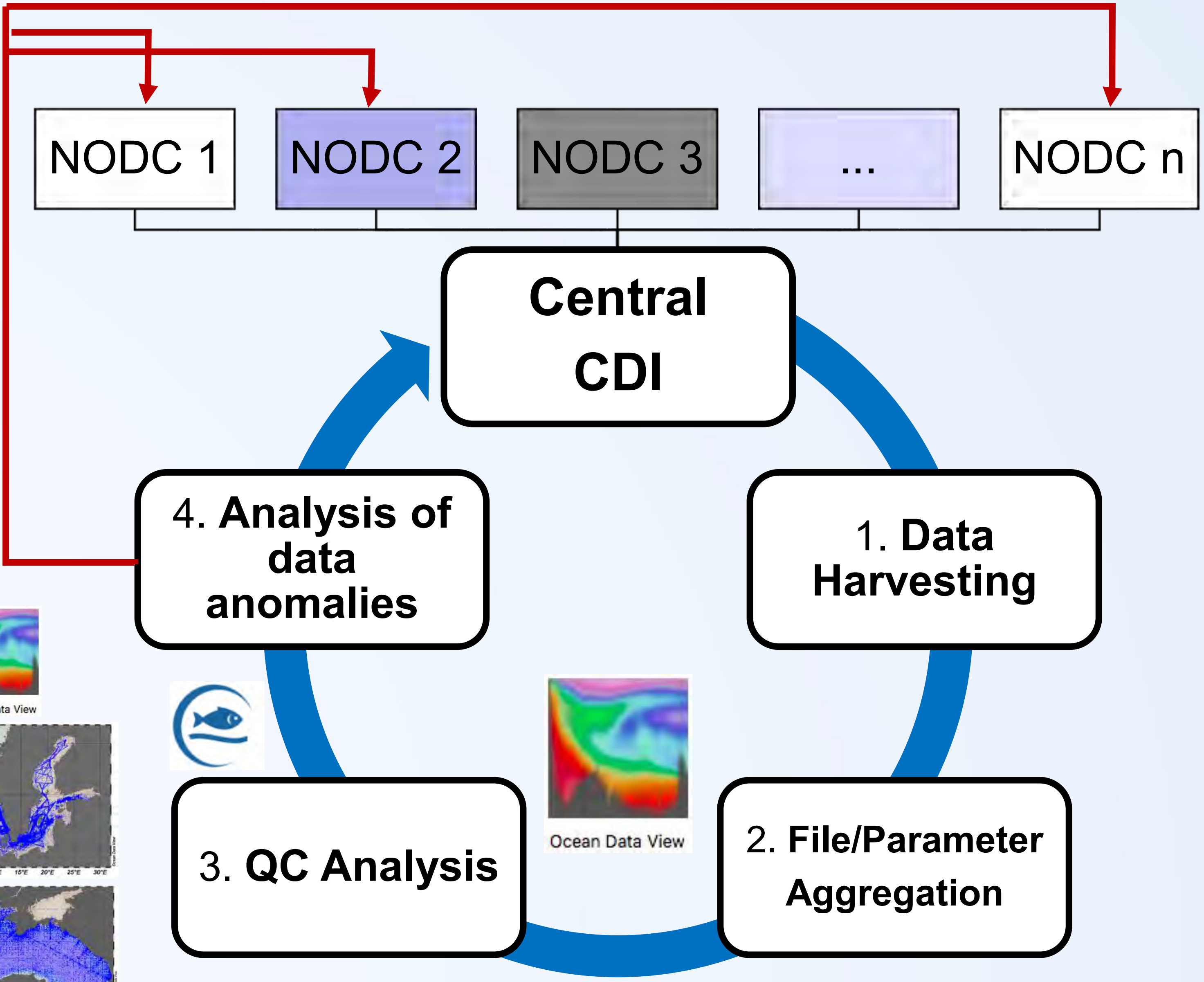
sdn-userdesk@seadatanet.org – www.seadatanet.org

Quality Assurance Strategy (QAS)



DIVA





sdn-userdesk@seadatanet.org

Harvesting and Aggregation

STEP 1 HARVESTING

Jul 31st → harvested 2 data collections for T&S from the latest SeaDataNet CDI catalogue:

- subset of complete data collection ONLY for CDI - data sets that are New or Updated compared to the last harvest at Oct 30th 2017, divided over unrestricted and restricted. (note: the latter is the same as for total collection)

STEP 2 AGGREGATION

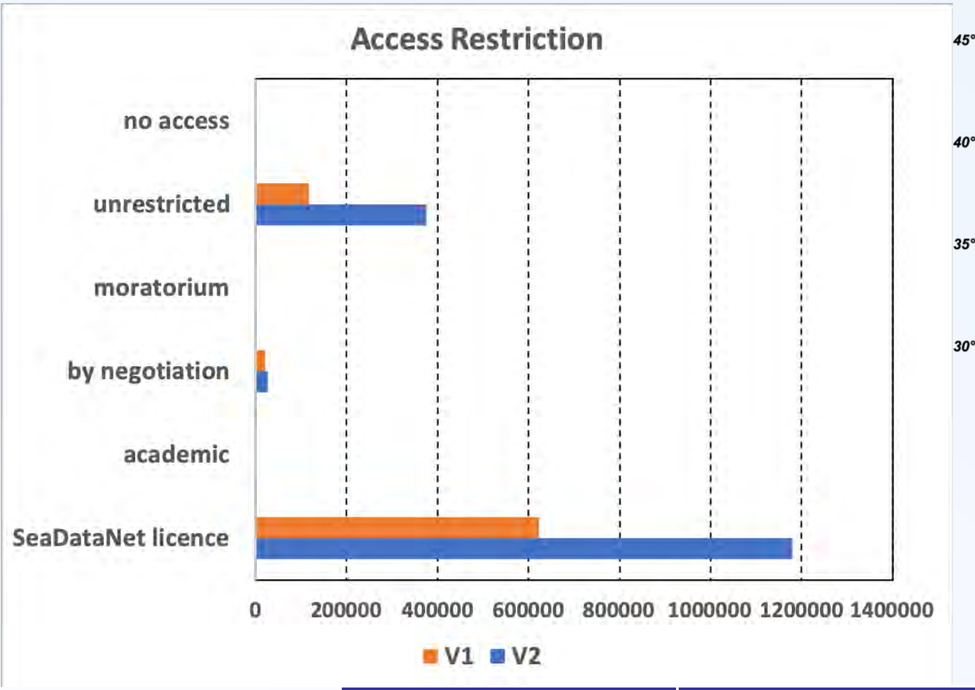
File and parameter aggregation of the "new_and_update_after_20171030" subset

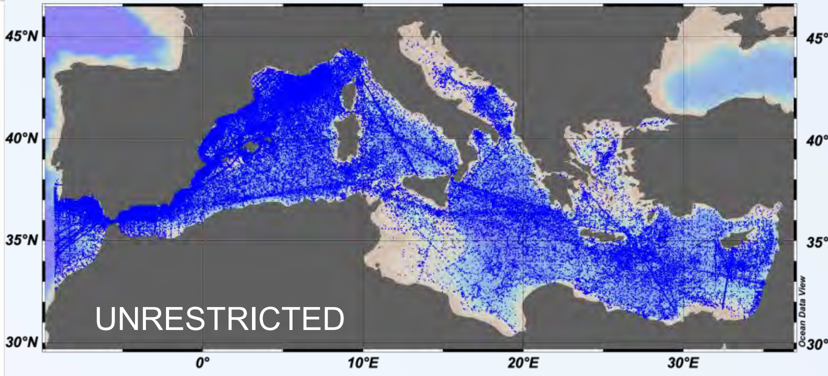
- split the results into regional collections and merged the harvested regional data with the SDC V1 regional collections
- analysis log files generated during import → error and warning messages → **there are still files with serious issues waiting to be corrected by the data centers**

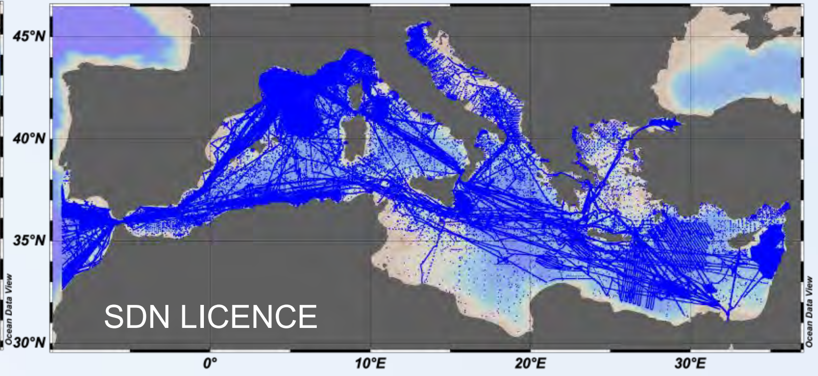
Aggregation

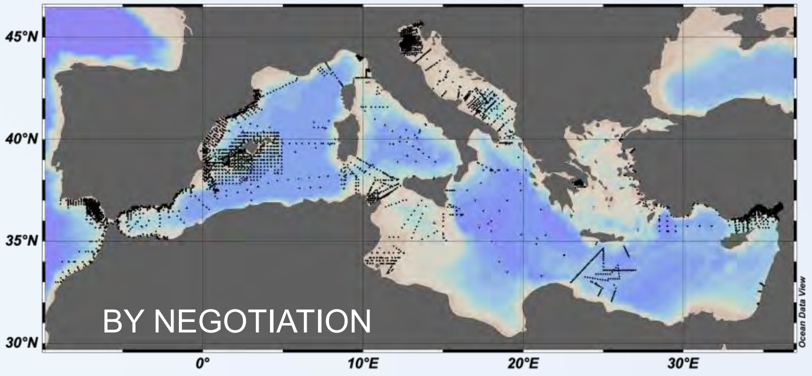
Target collection	# before	# imported	# after	# replaced
ARC	731,286	1,319,511 (nr)	1,437,500	613,297
		1,522 (r)	1,438,924	98
BAL	404,361	146,225 (nr)	516,729	33,857
		10,306 (r)	527,035	0
BLS	137,723	147,255 (nr)	195,264	89,714
		20,116 (r)	215,380	0
MED	739,784	958,730 (nr)	1,554,490	144,024
		27,673 (r)	1,582,163	0
NAT	9,091,769	2,575,876 (nr)	11,453,068	214,577
		38,367 (r)	11,491,431	4
NS PROFILES	162,452	109,106 (nr)	229,955	41,603
		12,007 (r)	241,931	31
NS TRAJECTORIES	580,376	260,516 (nr)	840,892	0
		12,128 (r)	853,020	0

SDC_MED_DATA_TS_V2





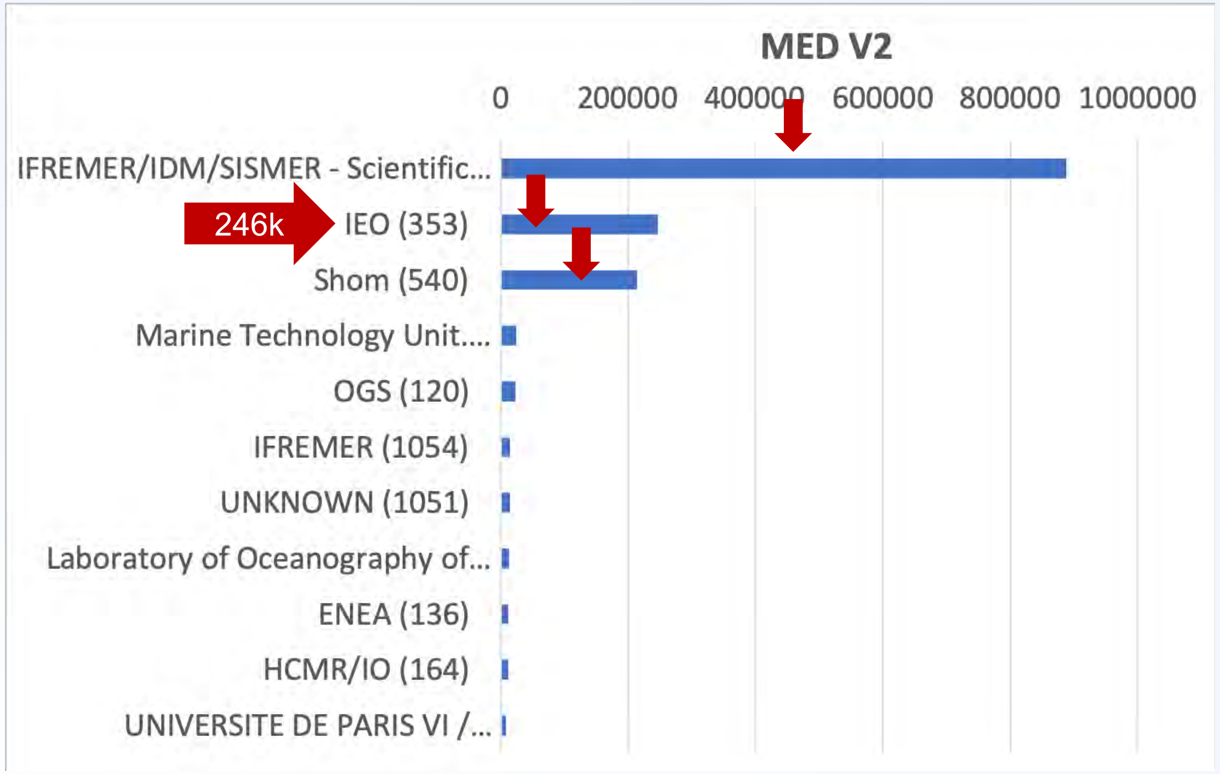


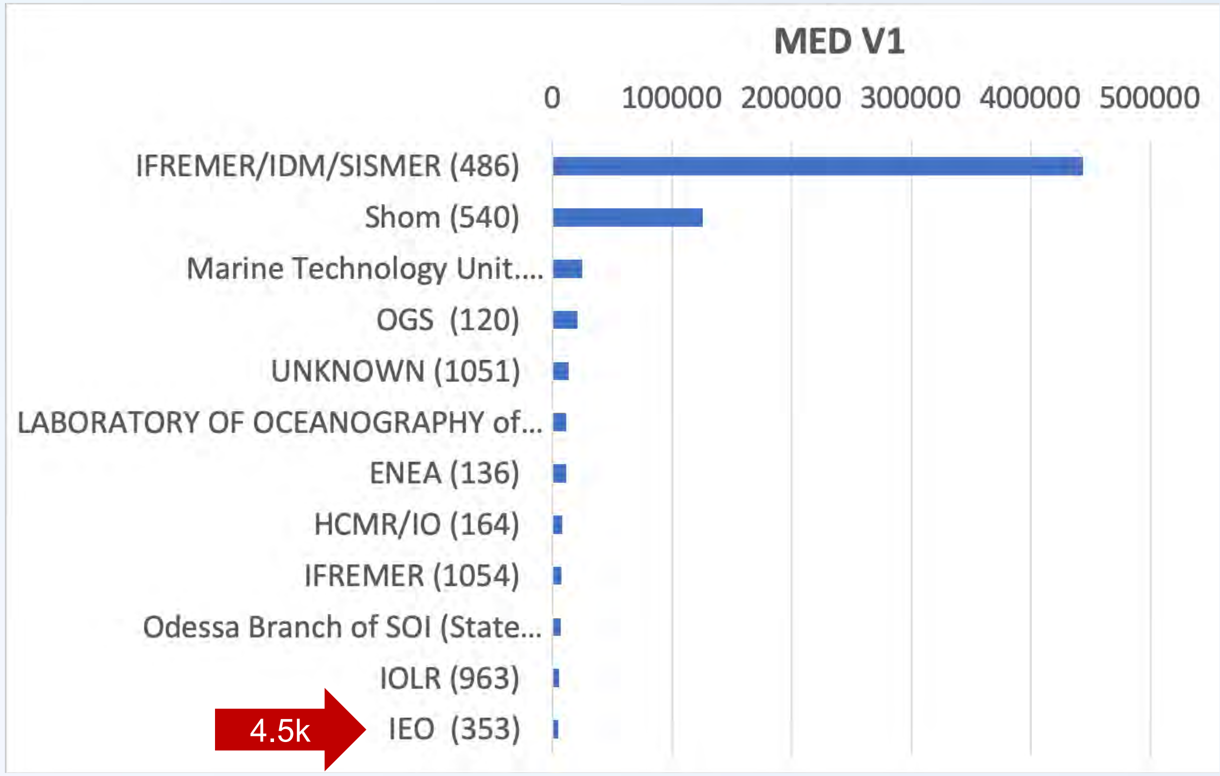


	V1	V2	% increase
free access	739784	1554490	+110
limited access	22221	27673	+25
TOT	762005	1582163	+108

2% of data are restricted

SDC_MED_DATA_TS_V2

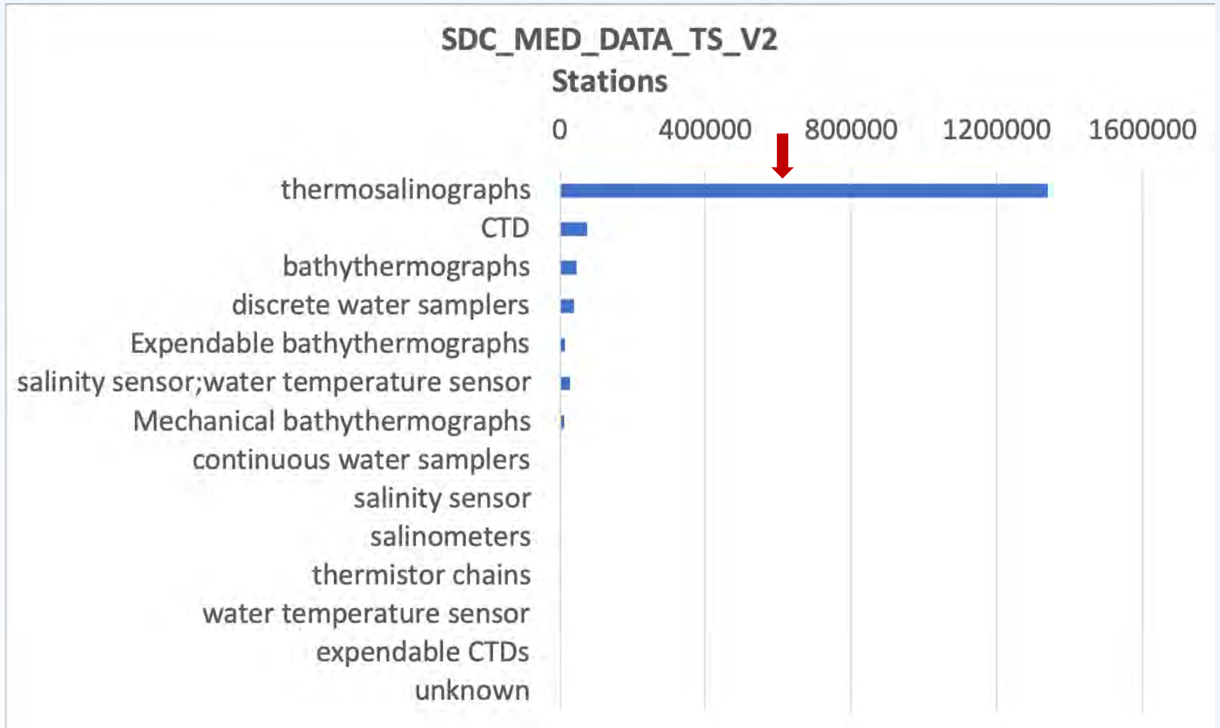


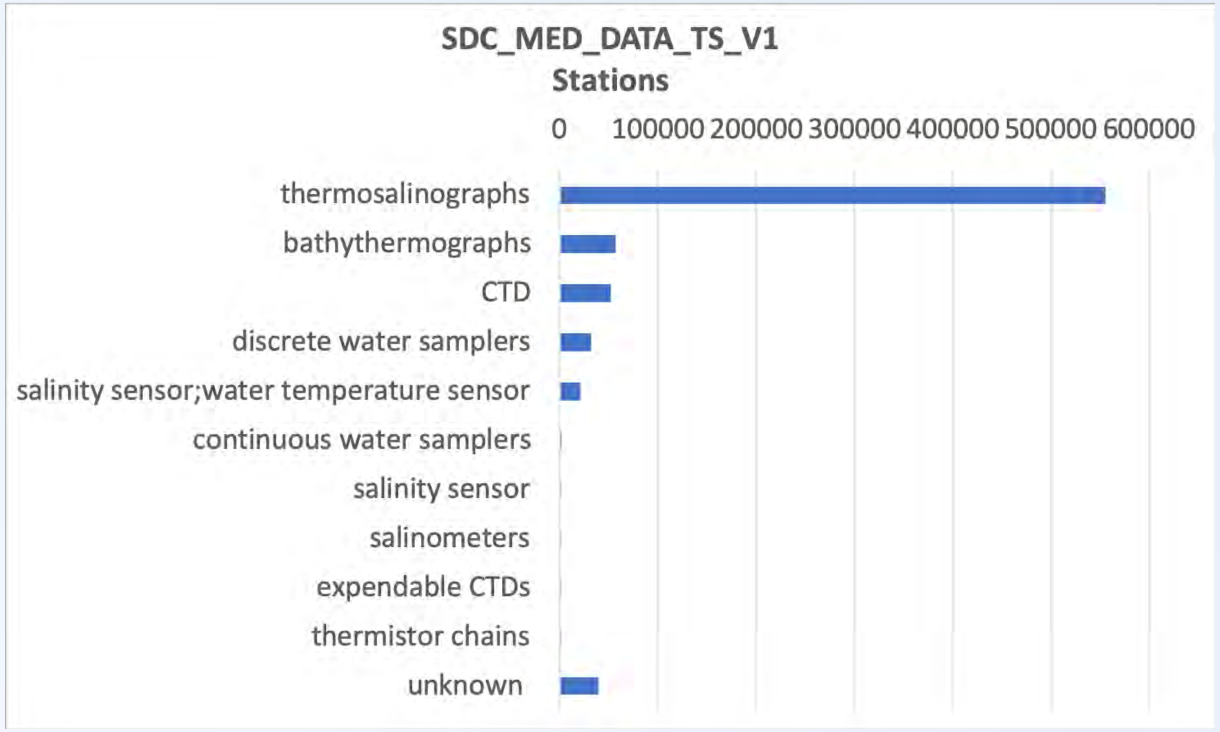


...

...

Instrument/Gear Type Stats MED

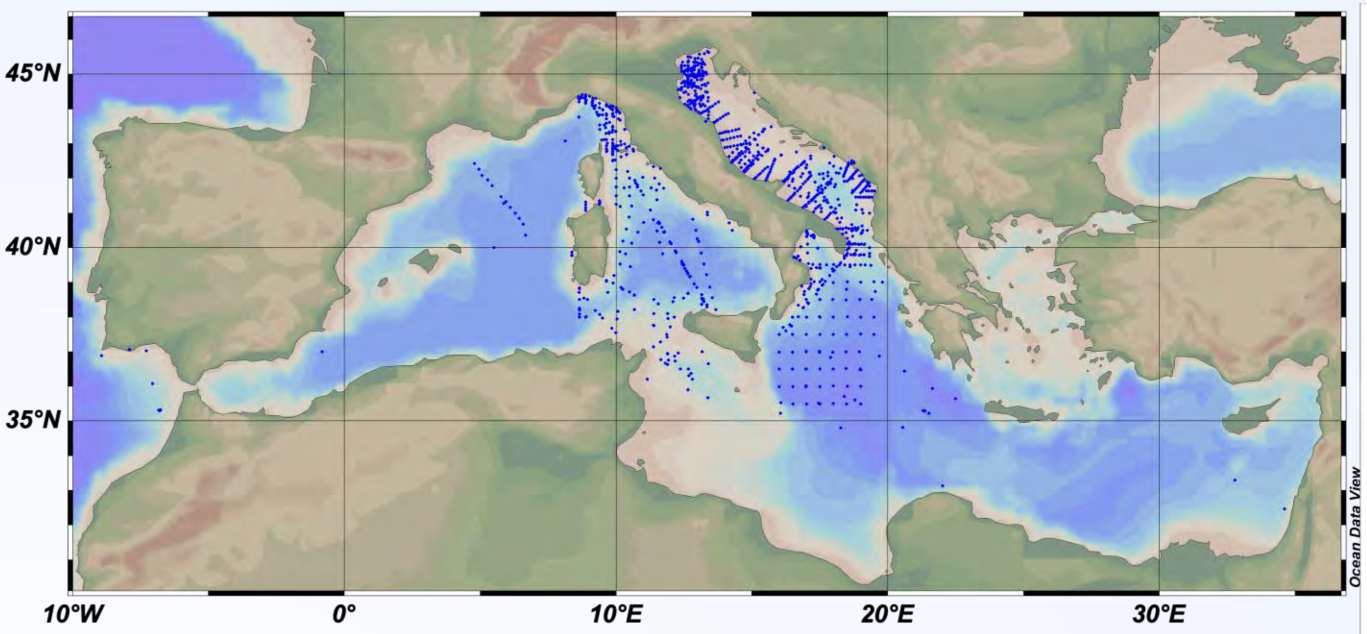


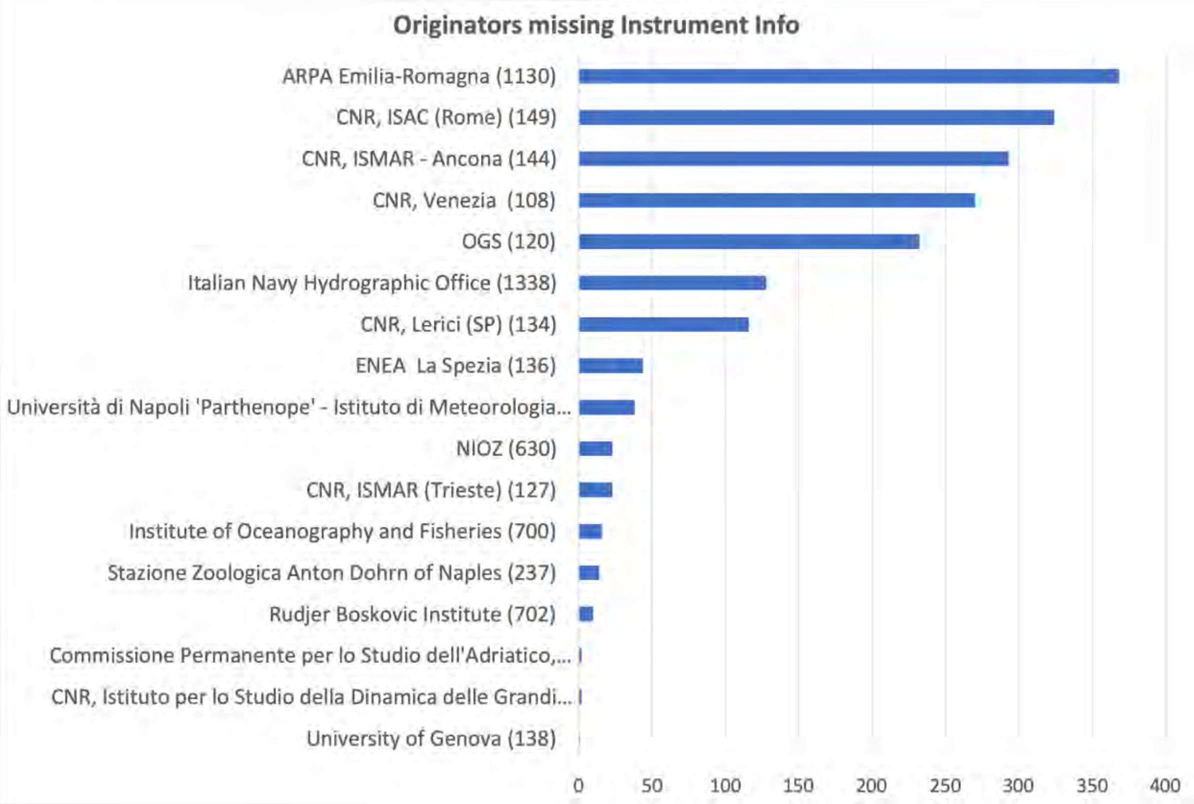


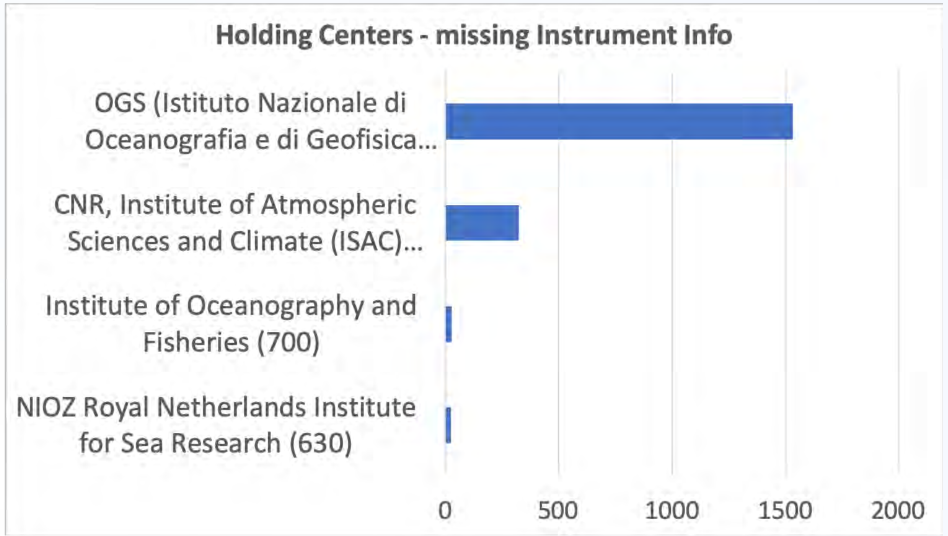
Instrument / gear type: 43 unique string values	Stations	%
thermosalinographs	1339074	86,1
CTD	75849	4,9
bathythermographs	44935	2,9
discrete water samplers	38937	2,5
Expendable bathythermographs	13069	0,8
salinity sensor;water temperature sensor	27047	1,7
Mechanical bathythermographs	11329	0,7
continuous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

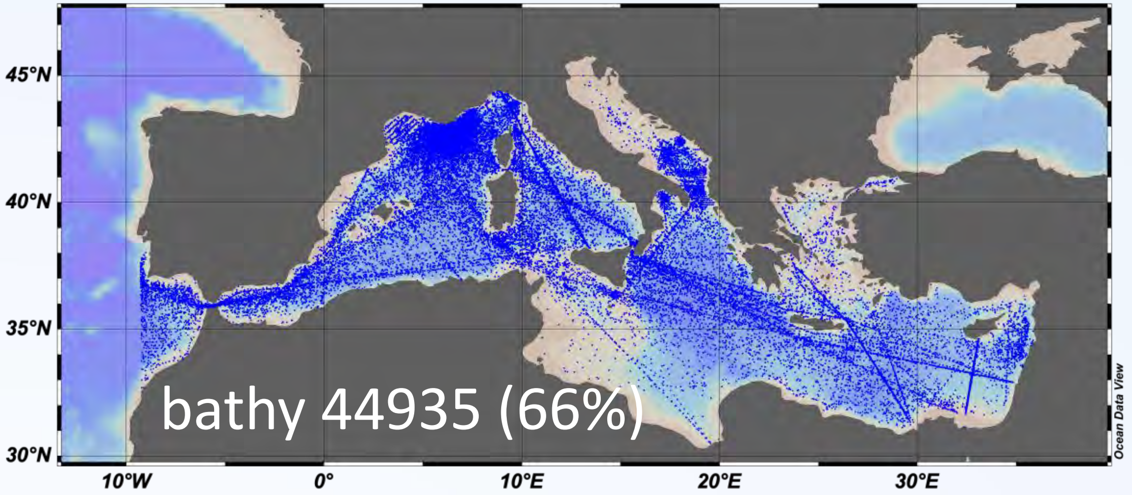


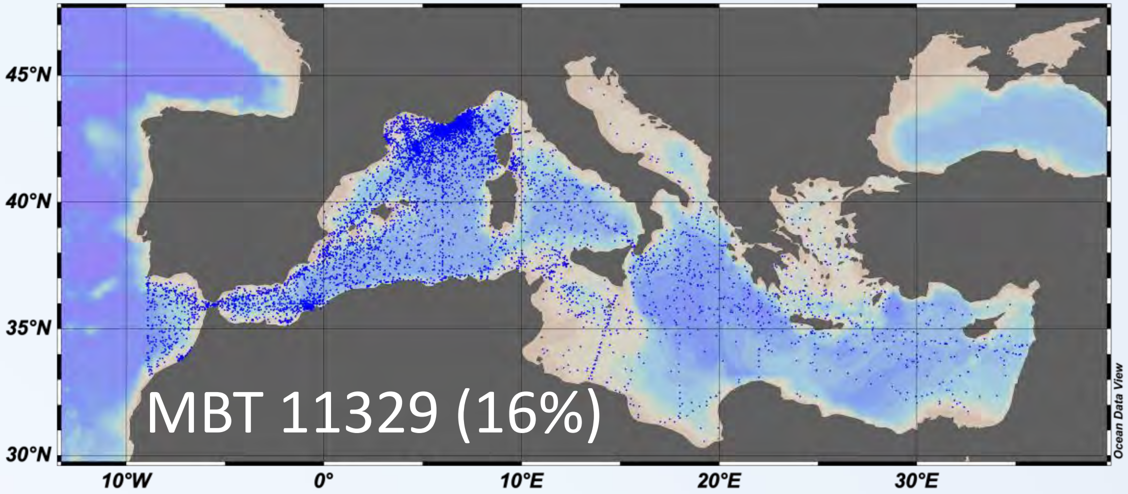


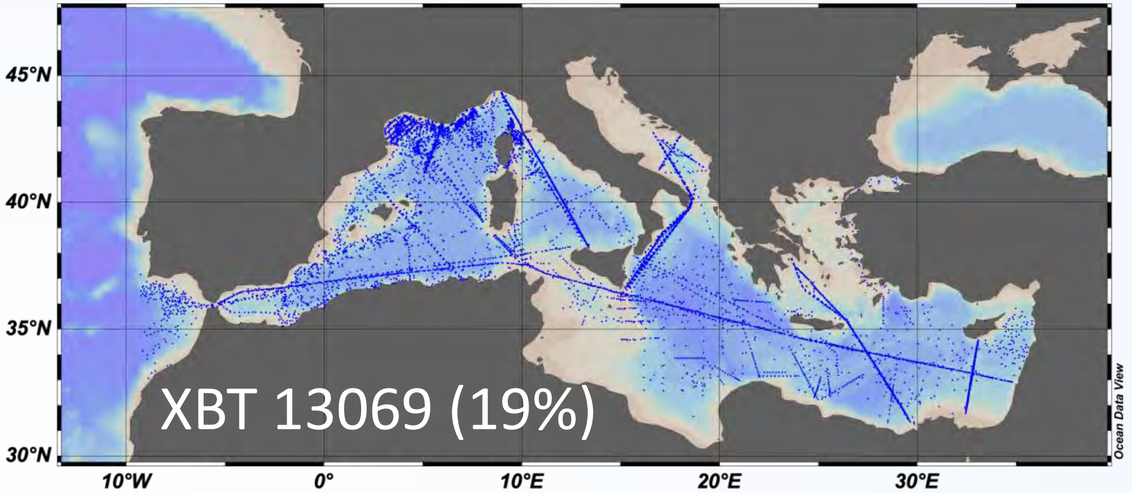


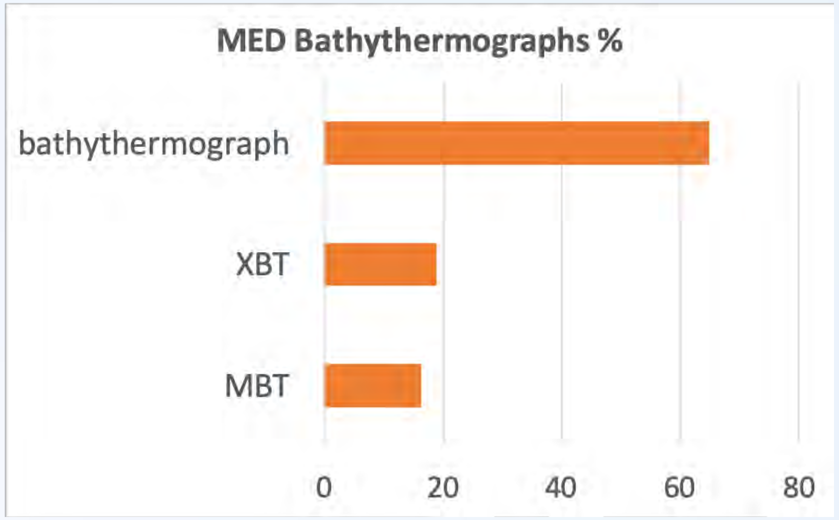
MED XBT

V1	V2	
56558	69333	+23%





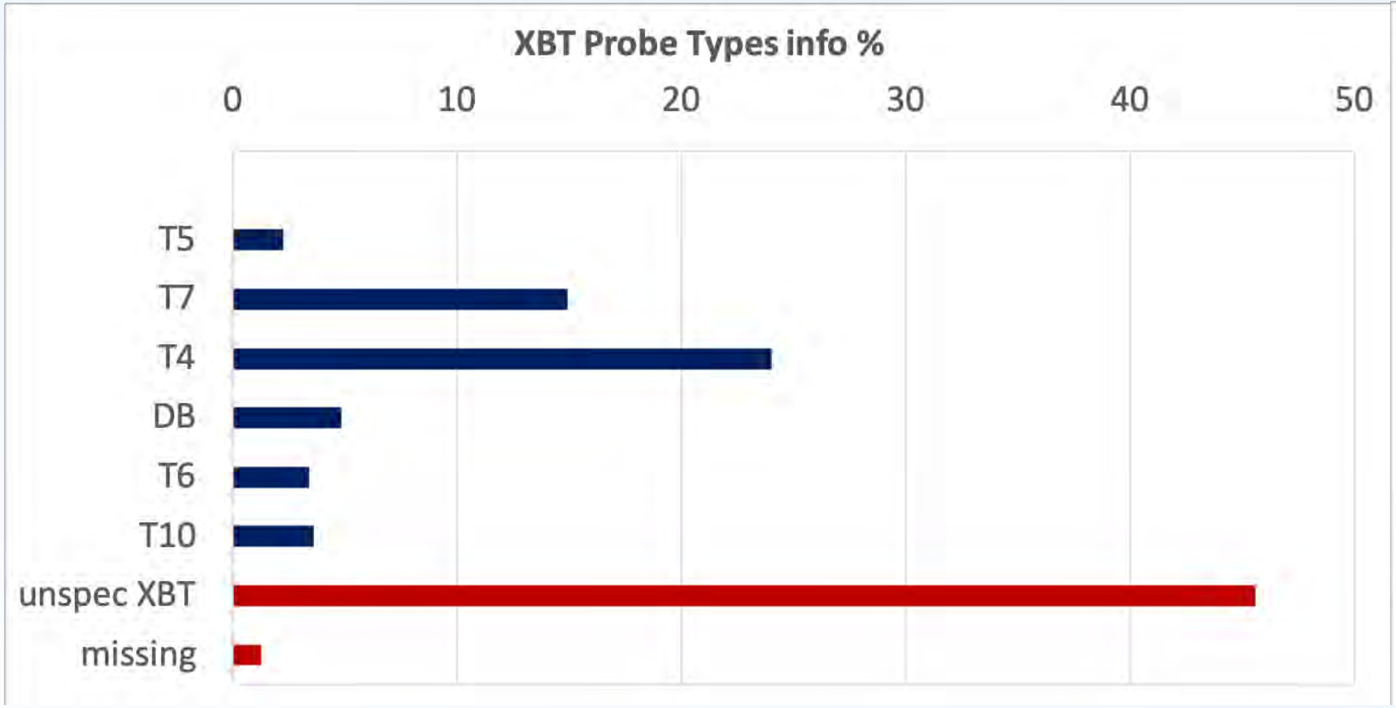


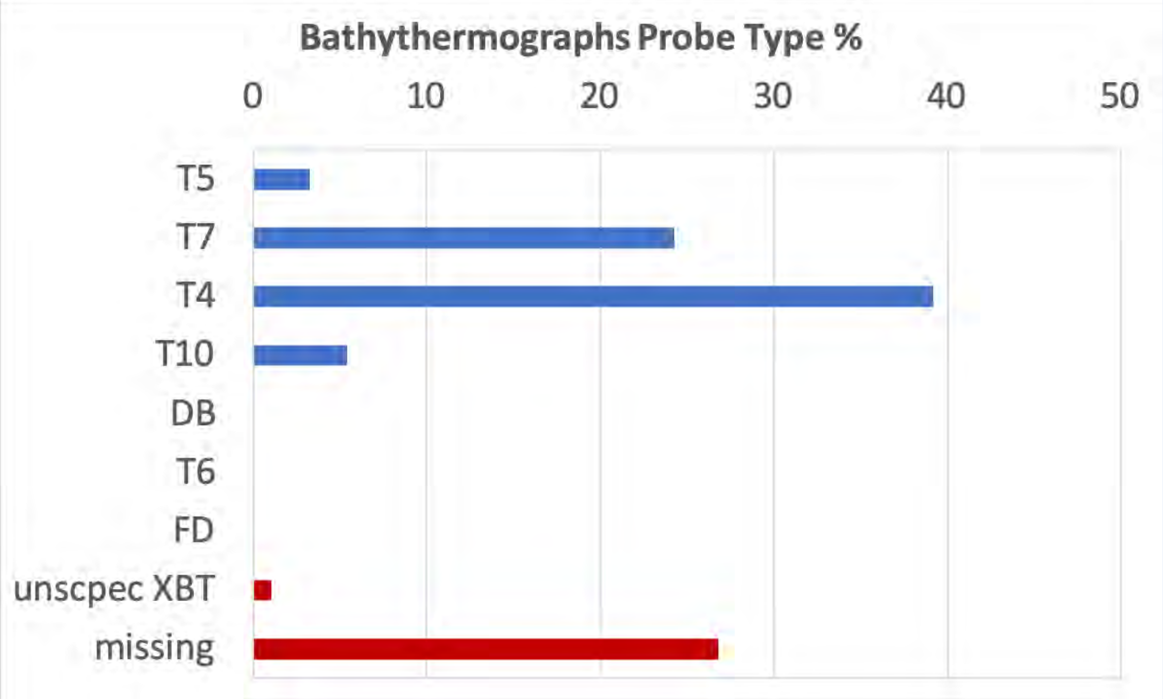


MED XBT

16% of the total BT data

66% of the total BT data





Unspecified or Missing Info

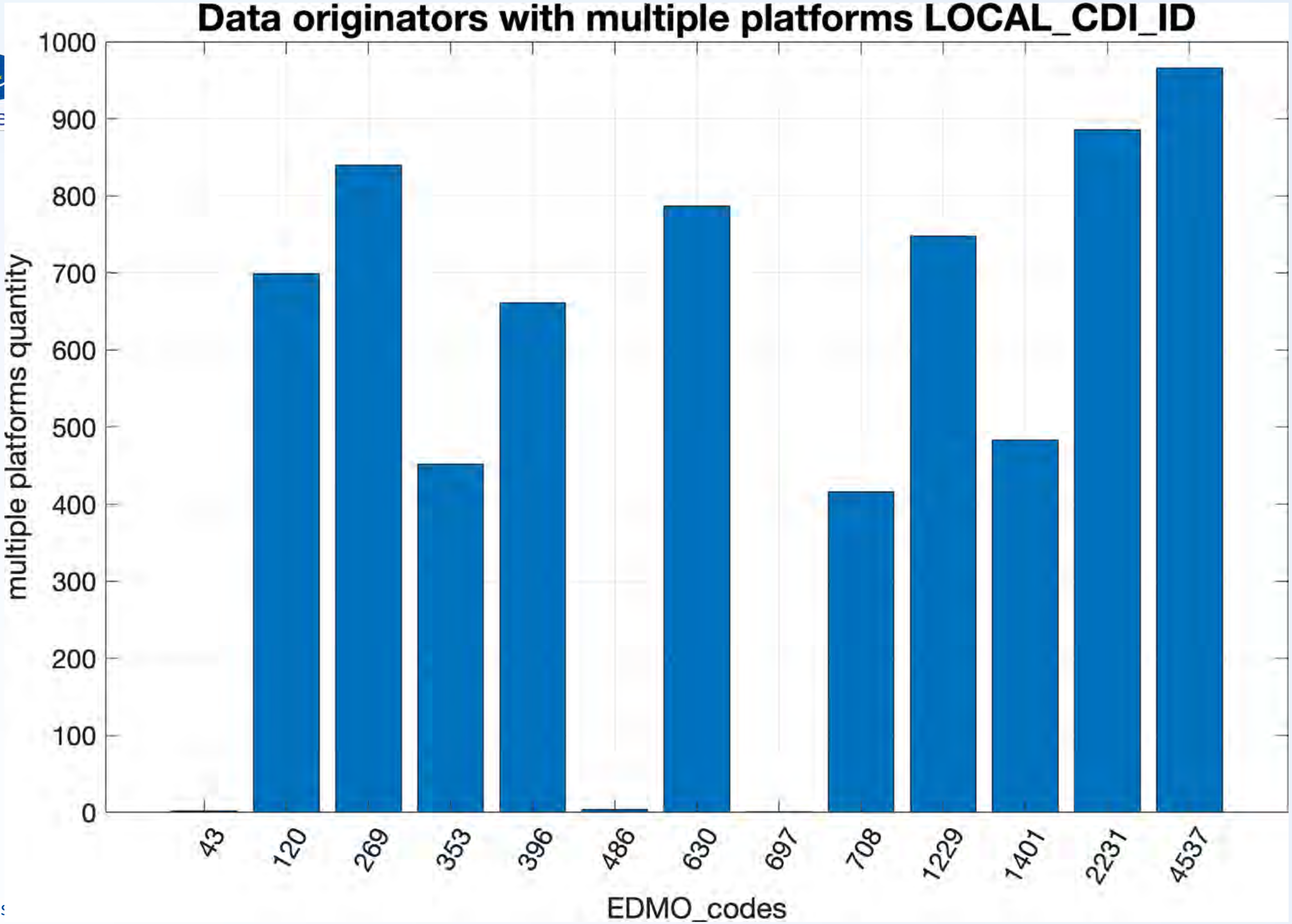
MBT	11329	16%
XBT	6126	9%
BT	12426	18%

Time Series: unrestricted dataset

Time Series: unrestricted dataset

Data Analysis

Total platforms:	99991
REAL AND GOOD platforms:	1709
ALL MISSING DATA:	736
ALL MISSING DEPTH:	26
ALL MISSING TIME:	1
ALL MISSING T DATA:	136
ALL MISSING S DATA:	626
SAMPLING TIME ABOVE 999 days:	11
PLATFORMS WITH MULTIPLE LOCAL_CDI_ID	971


Data originators with multiple platforms LOCAL_CDI_ID
multiple platforms quantity
0
100
200
300
400
500
600
700
800
900
1000
43
120
269
353
396
486
630
697
708
1229
1401
2231
4537
EDMO_codes

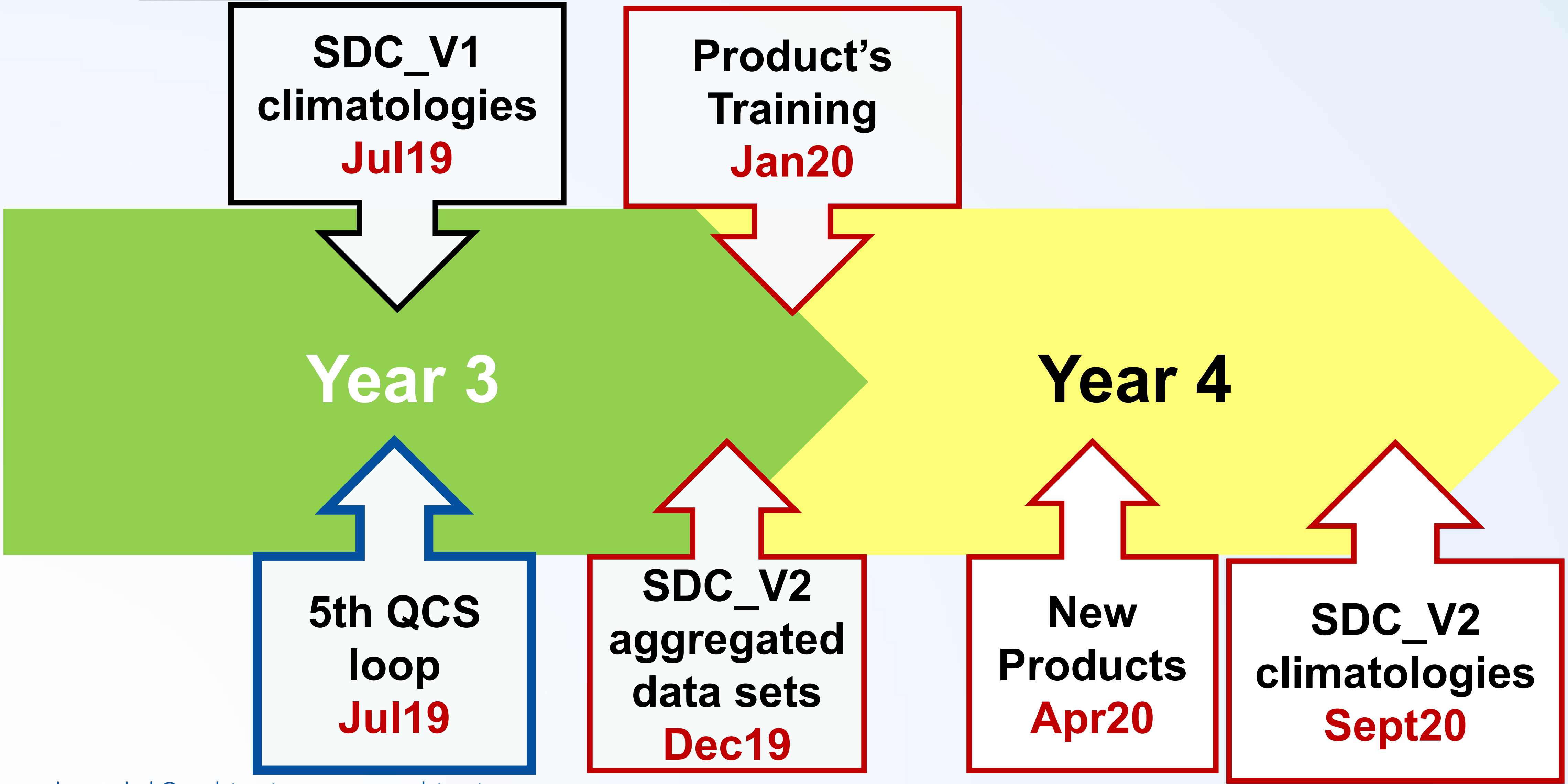
Conclusions and Future Work

- 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 (WP10-WP11 session) 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 assessment due by Dec 15

WP11 last year work plan

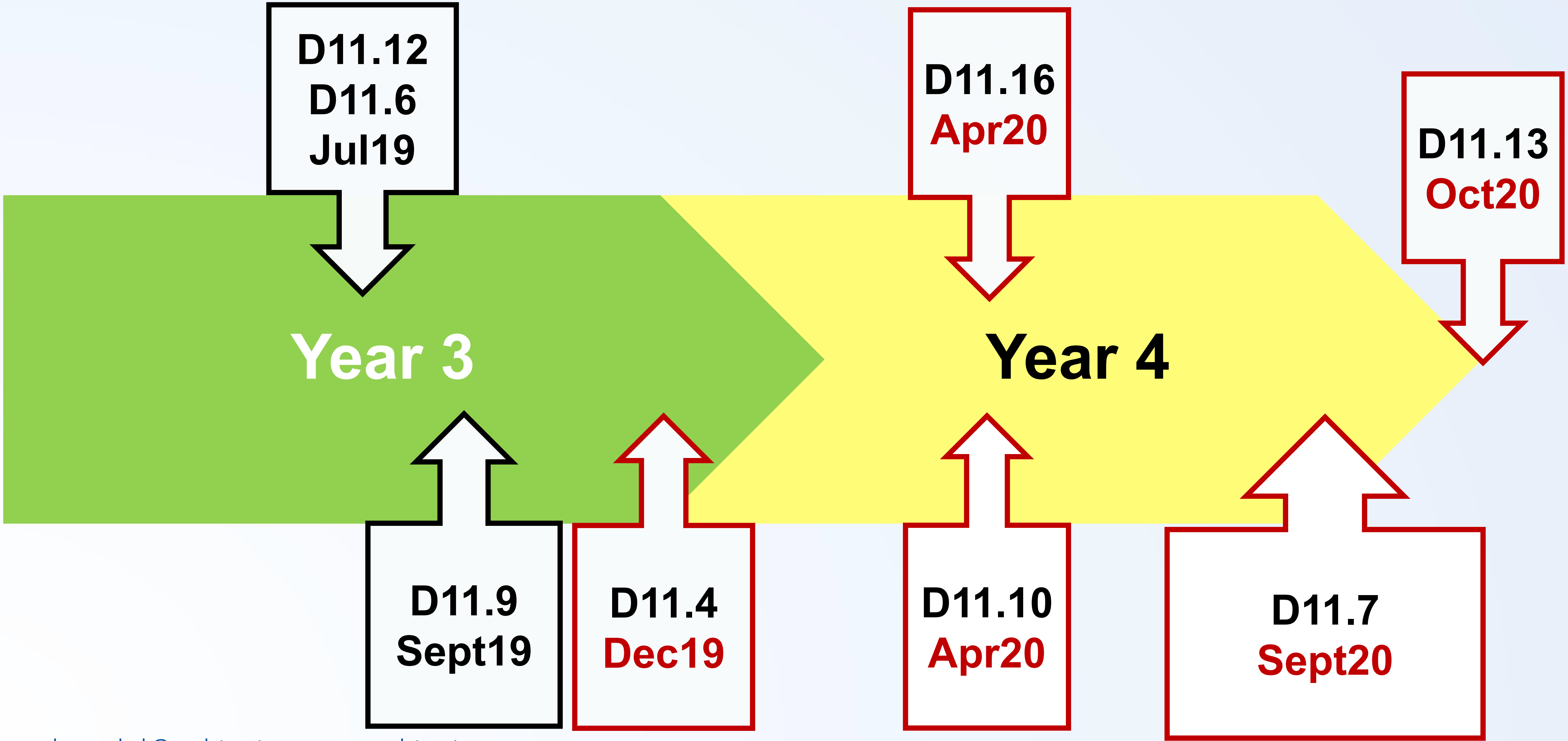
S. Simoncelli and WP11 team

Work Plan and Timeline



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

Deliverables



Deliverables

- D11.4 Release of V2 aggregated data set
- **D11.10 Release of new data product**
- D11.7 Release of V2 climatologies
- **D11.16 final outcome of product’s training activities**
- D11.13 final presentation of SDC product’s catalogue

D11.4 (data sets) and D11.7 (climatologie)

The generation of V2 of products will focus on optimization of procedures

→ V1 has been merged with new and updated CDI

→ DIVAnd uptake was successful

The documentation (PIDocs) has been produced for V1 and now needs to be updated and ameliorated

→more time to dedicate to improve final product’s quality

→2 step revision of PIDocs might involve more people from the partnership to be faster and get additional advises

New Data Products D11.10 (UniBo)

SDC_GLO_DP1	Seasonal mixing Index at 1/4° (2003-2017)
SDC_GLO_DP2	AOU at 1/4° (2003-2017)
SDC_GLO_DP3	Nu concentrations along isopycnals at 1/4° (2003-2017)
SDC_NWS_DP1	regional and sub-regional T and S monthly statistics at different depth layers
SDC_BAL_DP1	regional and sub-regional T and S monthly statistics at different depth layers
SDC_NAT_DP1	Monthly climatology for MLD at 1/4°
SDC_NAT_DP2	sliding decades
SDC_MED_DP1	Monthly climatology for MLD at 1/8°
SDC_MED_DP2	OHC time series and trend (0-700m; 0-2000m)
SDC_BLS_DP1	Monthly climatology for cold intermediate layer (CIL) cold content at 1/8°
SDC_BLS_DP2	Decadal CIL cold content at 1/8°
SDC_ULG_DP1	Currents climatologies from HF radars

Training on products generation (D11.16)

- Training session is planned for the last week of Jan or first week of Feb
- Location will be Bologna (waiting for UniBo confirmation) or Liege

Objectives

- new data products → how to use DIVAnd to get Ocean Monitoring Indicators, BGC indicators, coastal HR maps through DIVAnd
- from Data Integration to DIVAnd → how to harmonize the working flow
- A posteriori QC to assure data representativeness (residuals, stats)
- DIVAnd → products harmonization, full 3D optimization, solving issues, increase production efficiency

WP Performance

- WP team is very collaborative and put a lot of efforts in producing the climatologies and integrating for the first time external datasets
- ODV developments help in optimizing the workflow and QC
- the continuous support of ULG and the feedbacks on the software from producers allowed to solvee many issues, making the code more robust
- the publishing of data products, dissemination activities needed a lot of dedication too
- deliverables were often postponed due to the many dependencies from other WPs activities, but we succeeded!

THANK YOU

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