SeaDataCloud

ISTITUTO NAZIONALE
DI GEOFISICA E VULCANOLOGIA

SeaDataCloud data products for the Mediterranean Sea and the underlying data gaps

Simona Simoncelli

P. Oliveri, D. Delrosso, G. Mattia, and the SeaDataCloud WP team

Ravenna, 6 Dec 2021

OUTLINE

- SeaDataCloud and SeaDataNet
- the marine data value chain
- SDC data products and objectives
- workflow
- aggregated datasets for the EU marginal seas
- climatologies
- new data products
- dissemination strategy
- demonstration of SeaDataNet data access portal
- Conclusions

SeaDataCloud

What is SeaDataNet?

https://www.seadatanet.org/

- **distributed Marine Data Infrastructure** for management of large and diverse sets of data deriving from in situ
- a Pan-EU **network of data centres** active in data collection **providing on-line integrated databases of standardised quality**
- **AISBL association** with 33 members —> legal entity for sustainability of the Consortium

34 countries

90s	Metadata directories Medar/MedAtlas
2002-2005	Sea-Search (FP5)
2006-2011	SeaDataNet (FP6)
2011-2015	SeaDataNet II (FP7)
2016-2020	SeaDataCloud (H2020)

SeaDataCloud

What is SeaDataNet?

https://www.seadatanet.org/

- provides **on-line access to in-situ data, meta-data and products** through a unique portal interconnecting the interoperable node platforms operated by the data centres
- develops and adopts **common communication standards, adapted technology** to ensure the platforms interoperability
- assures **quality, compatibility and coherence of the data from many sources** by adoption of standardized methodologies and common software tools
- Data, value added products and dictionaries serve wide **uses**: research, operational oceanography, climate, industrial projects, teaching, marine environmental assessment

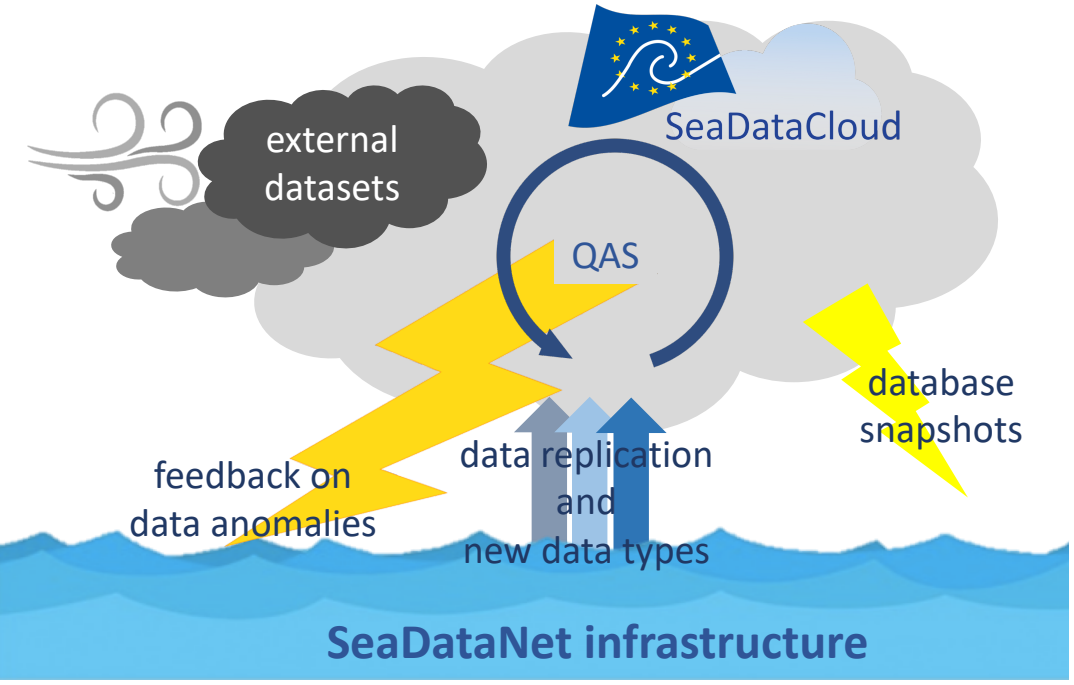
SeaDataCoud

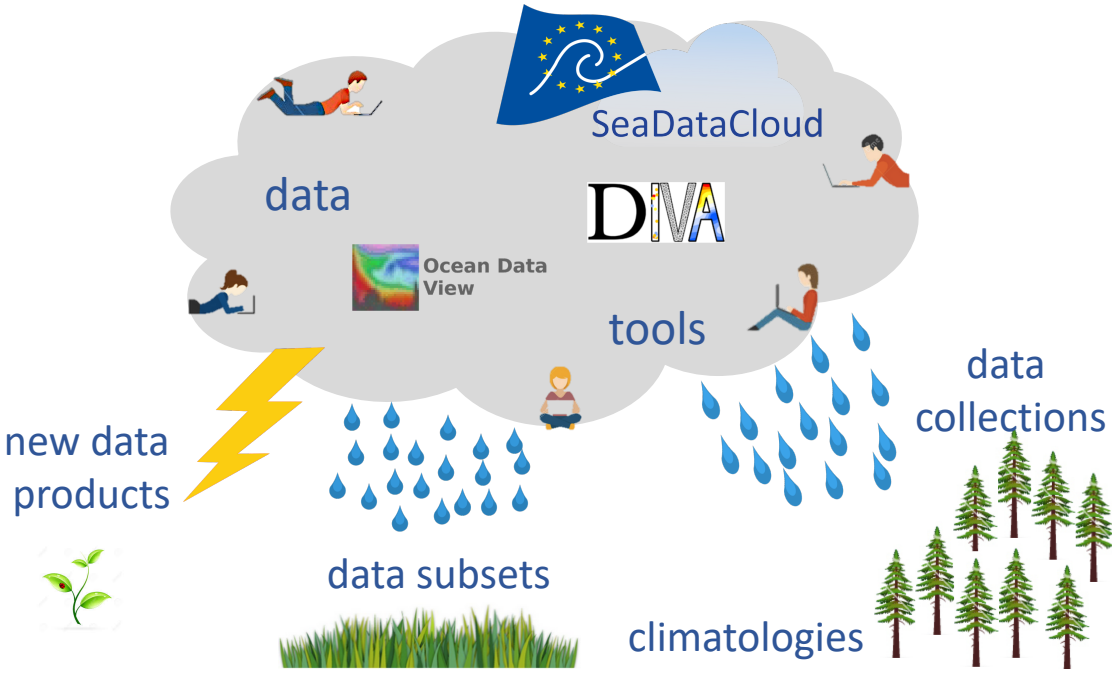
(H2020 program Nov2016-Apr2021)

Innovation: a cloud environment hosts a replication of SDN database content:

- improves data access service
- offers a collaborative environment to perform data-driven research

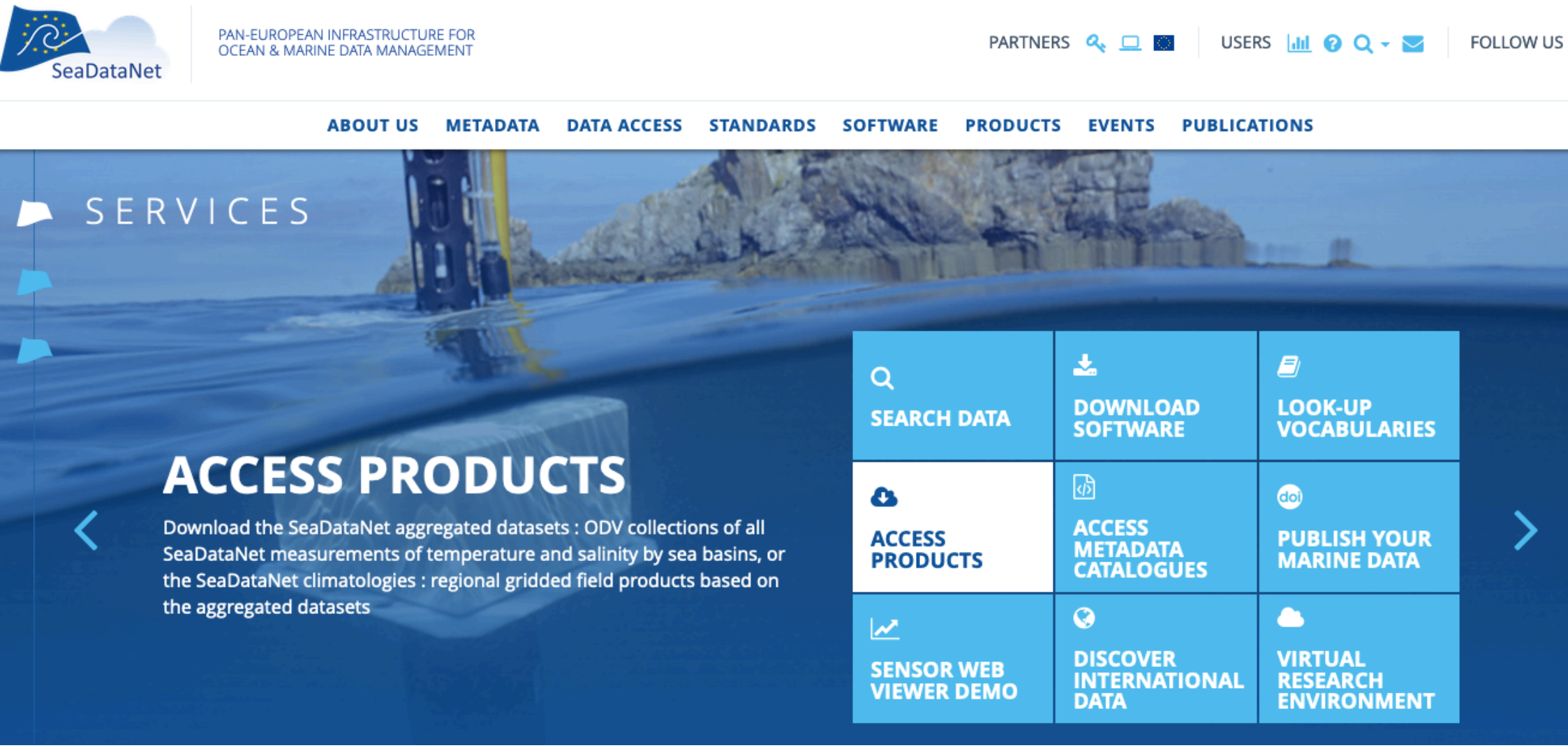
Pilot version co-designed to offer an efficient working experience through expert users' engagement that provided recommendations about the VRE and its auxiliary software functioning



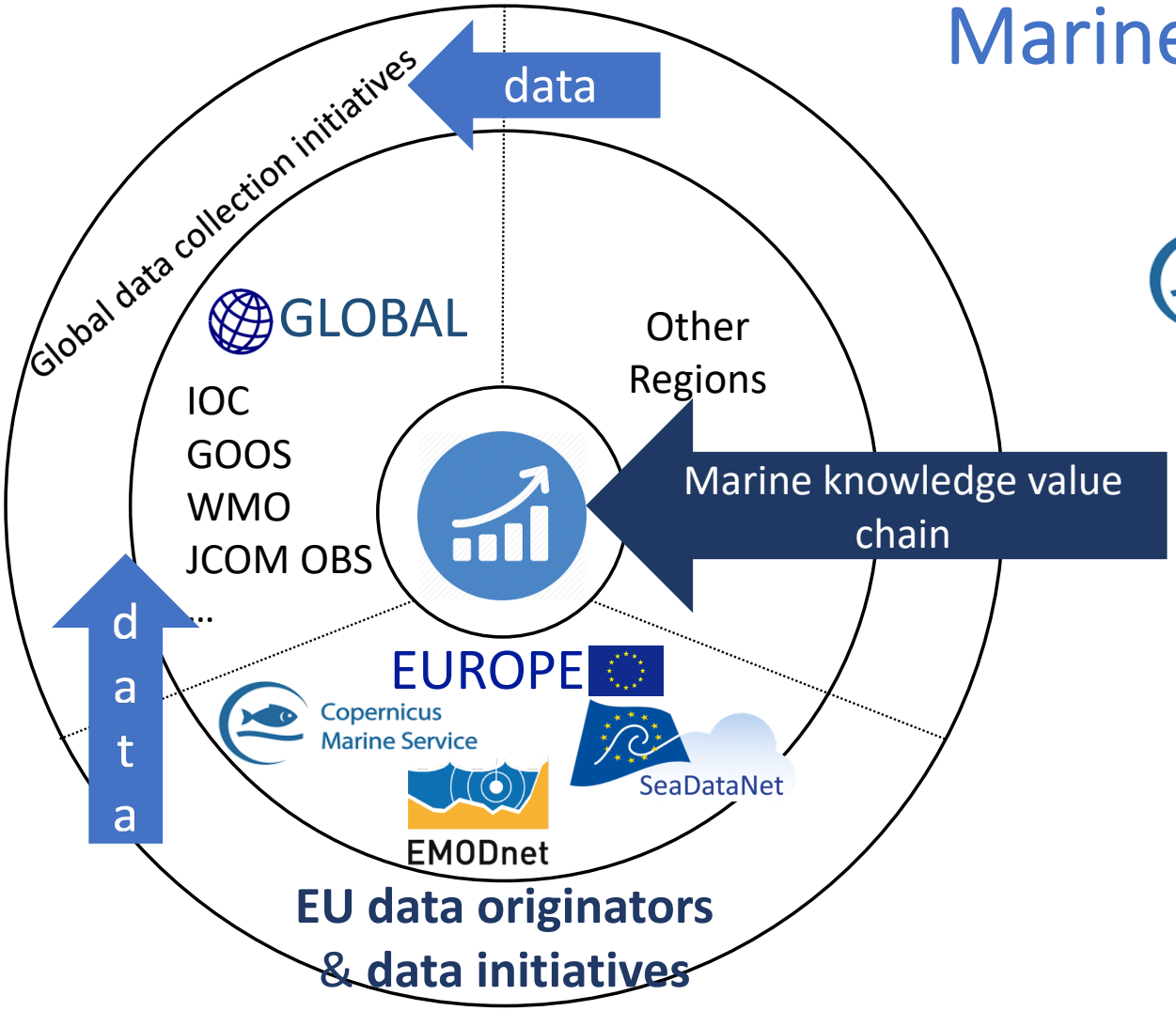


Simoncelli et al. (2022) https://doi.org/10.1016/B978-0-12-823427-3.00001-3

https://www.seadatanet.org/



Marine Data and Data Products Initiatives



adapted from *Martín Míguez et al.(2019)*

Copernicus Marine Service

DG-GROW

provides NRT&DM observations and predictions on the status and dynamics of the ocean and the ecosystem

SeaDataNet

DG-RTD

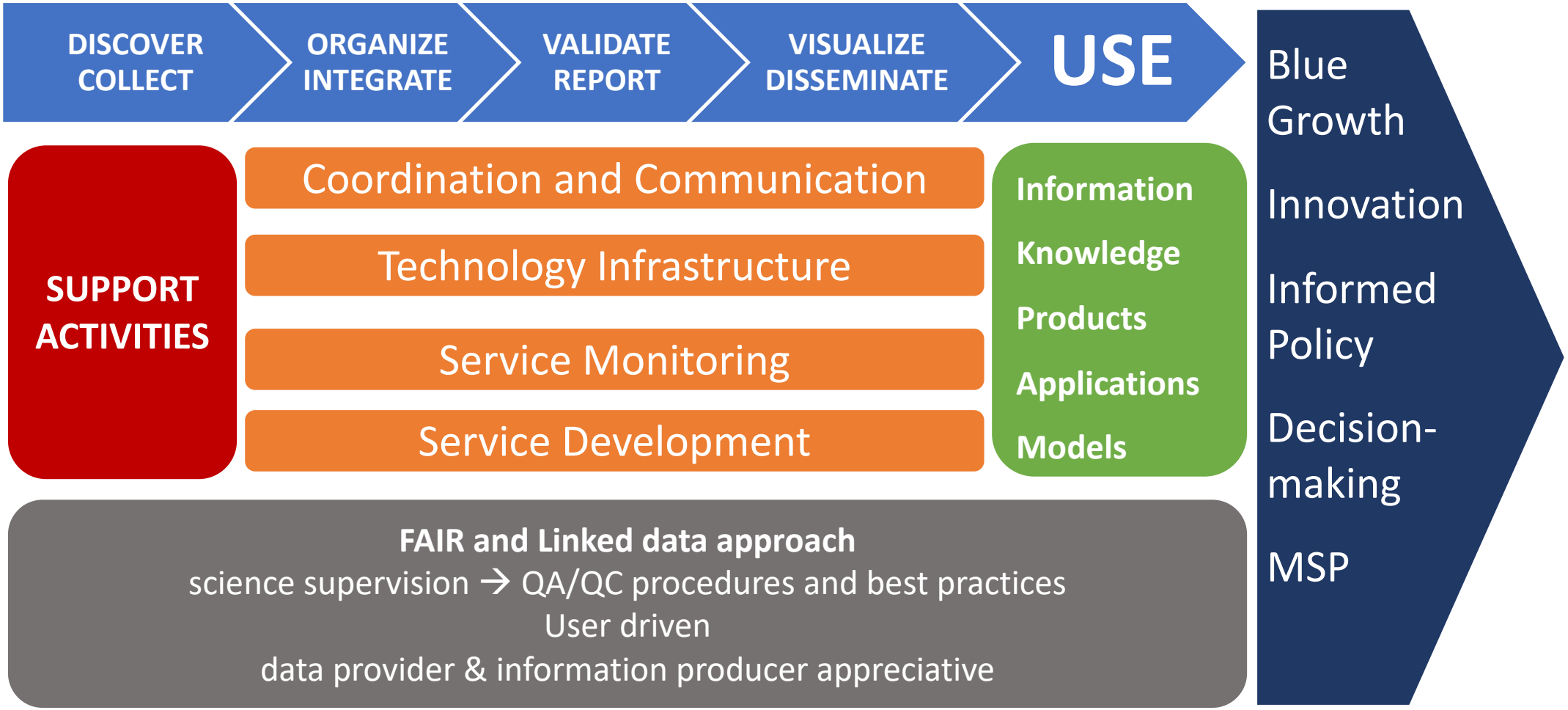
distributed Marine Data Infrastructure for the management of DM in situ data sets; it develops standards and vocabulary; it provides services/tools and data products

EMODnet

DG-MARE

organizes marine data in a unique infrastructure, deriving information from data, creating products targeted to multiple user communities

Marine Data Value Chain



Simoncelli et al. 2022 https://doi.org/10.1016/B978-0-12-823427-3.00001-3

SeaDataCloud data products

GOAL: to provide the best data products, based on SDN in situ T and S observations for EU seas and the global ocean

- data and full metadata description are integrated and harmonized within the database thanks to standardized Quality Assurance and Quality Control (QA/QC) methodologies, conducted at various stages of the data life cycle
- a prior QC is done by data providers, who submit the data with corresponding Quality Flags (QF)
- formats and standards are harmonized and checked during the ingestion process
- a Quality Assurance Strategy (QAS) has been implemented in order to guarantee the consistency of the database content and promote a high quality of the derived products

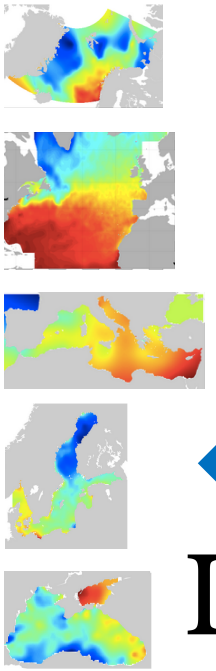
https://www.seadatanet.org/About-us/SeaDataCloud

SeaDataCloud

Workflow

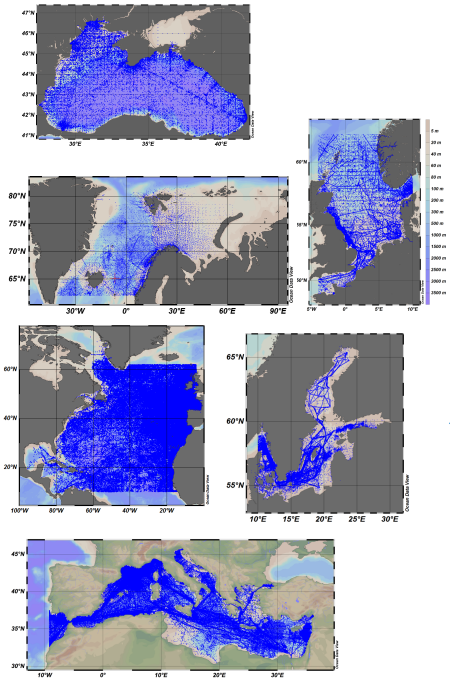
Products

climatologies

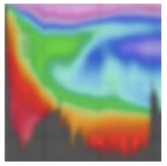


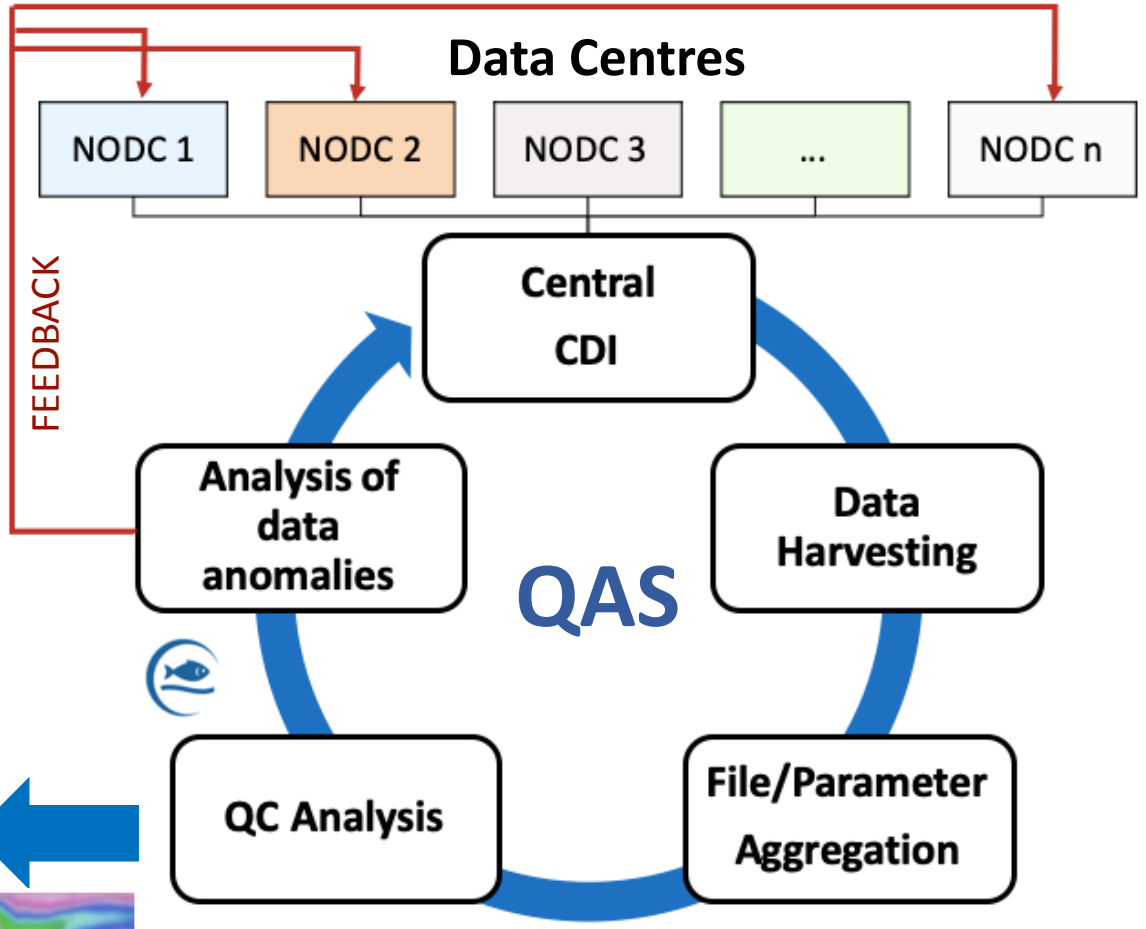
DIVA

aggregated datasets



Ocean Data View





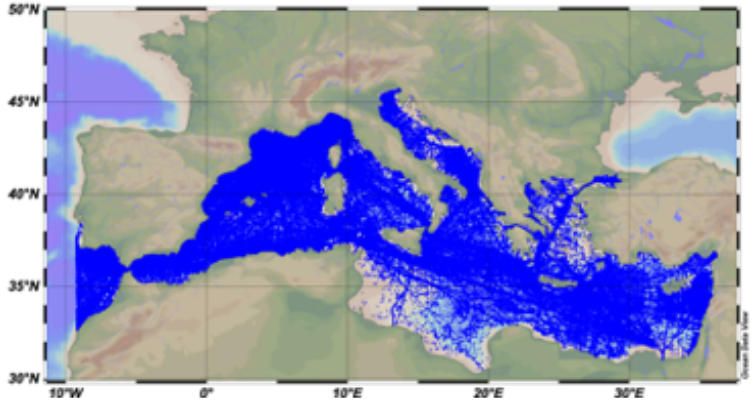
SeaDataCloud

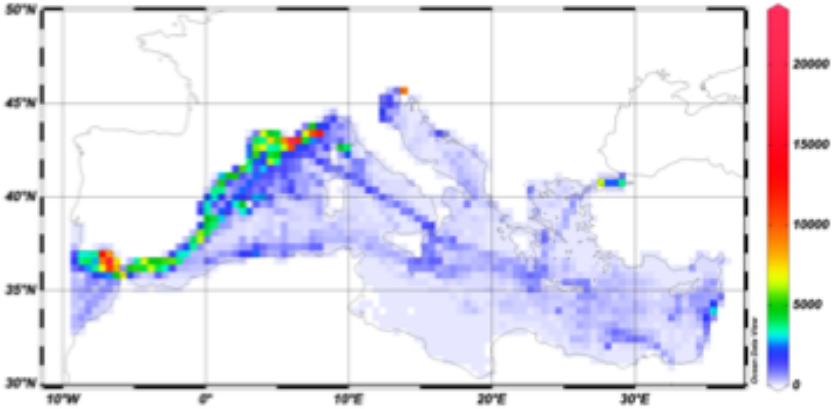
Aggregated Datasets

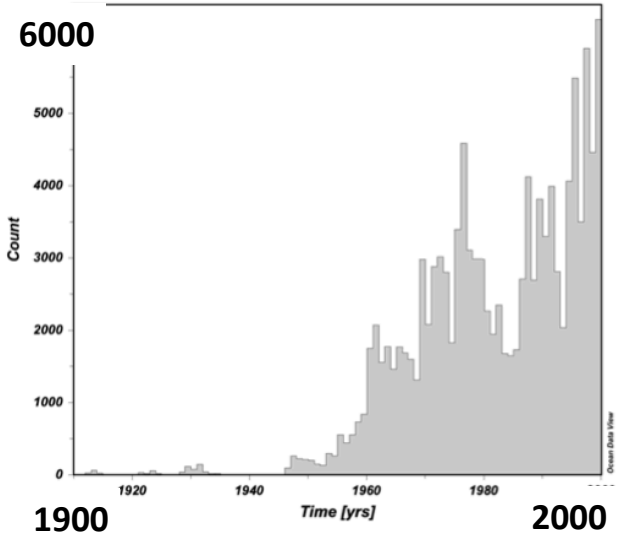
- 2 versions of T and S aggregated datasets have been released (2018, 2020) for the North Atlantic Ocean (NAT), North Sea (NS), Baltic Sea (BAL), Arctic (ARC), Mediterranean Sea (MED) and Black Sea (BLS)
- basic QC steps by visual inspection in ODV: space and time data distribution and coverage, scatter plots (spikes, outliers), stations falling on land, wrong/missing data and statistics about QF
- assigned QFs have been modified by the regional experts where data anomalies (bad data flagged as good) were detected
- Metadata analysis by instrument type or by data providers allowed to identify omission (data existing in literature but not publicly shared) and systematic errors (format, flagging) at the data center level
- validation results included in a Product Information Document (PIDoc) annexed to each dataset, which provides also important usability instructions and acknowledges all data originators

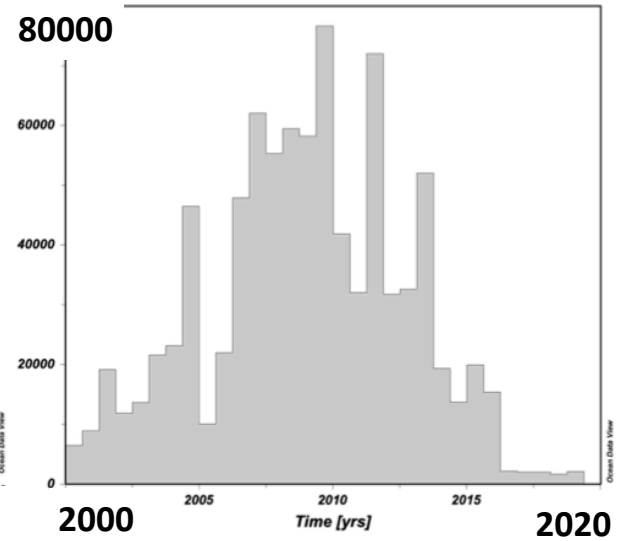
Mediterranean Sea Aggregated Dataset

SDC_MED_DATA_TS_V2 includes open access in situ data of water column temperature and salinity obtained harvesting all measurements contained within SDN infrastructure at the end of Jul2019 —> 1003259 stations belonging to 30 data providers and 124 data originators



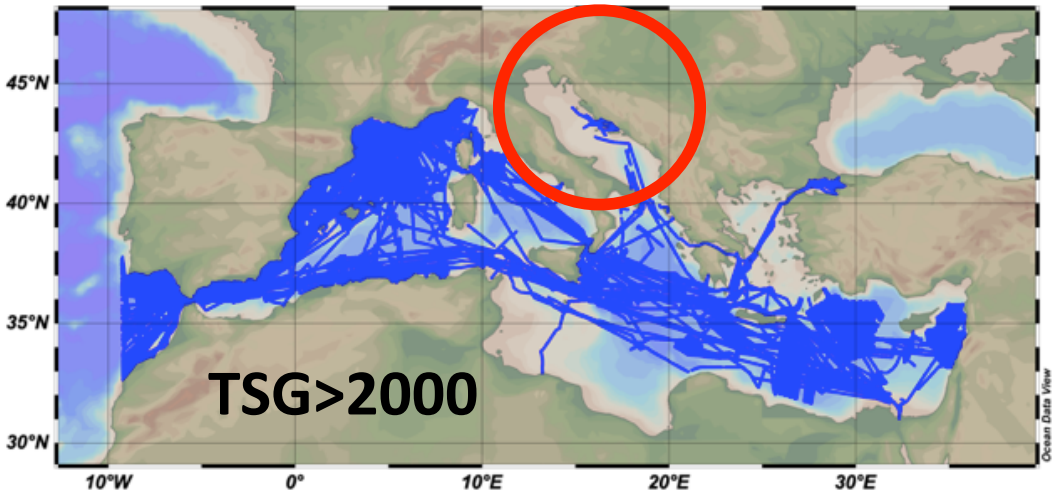


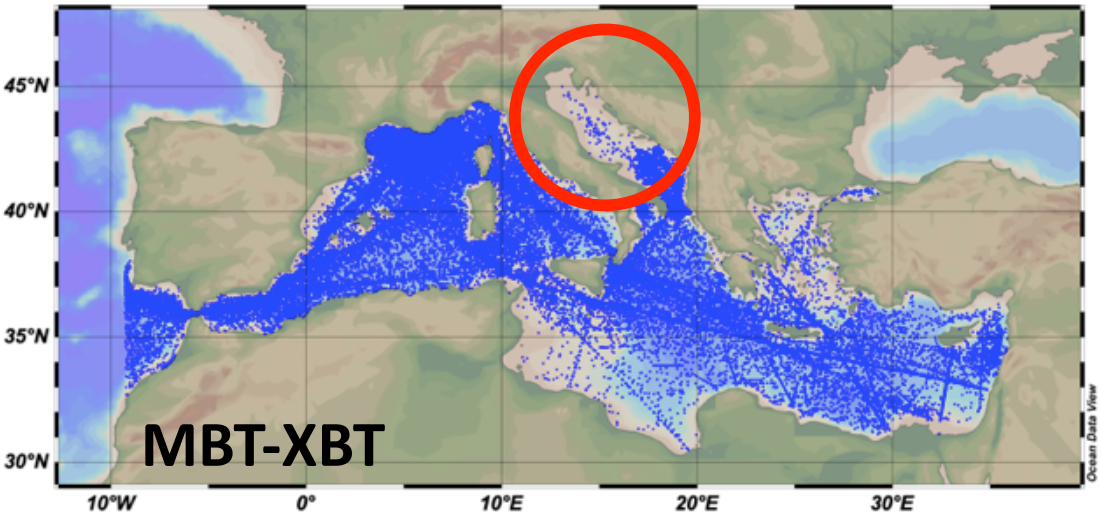


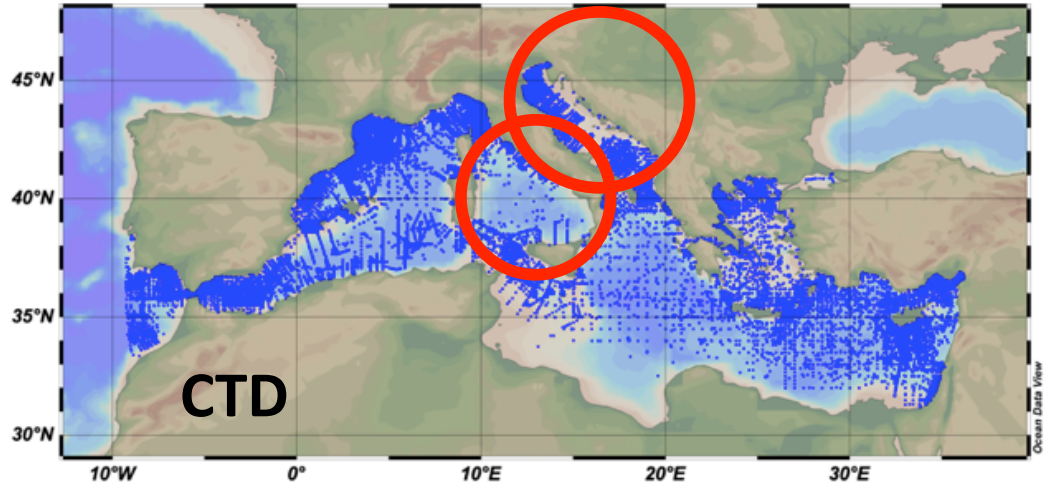


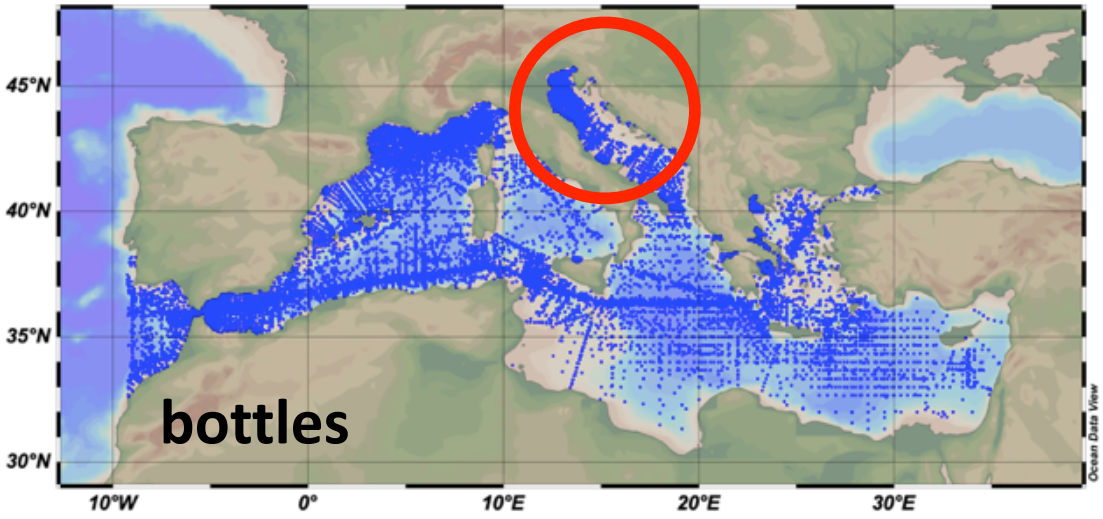
par	# stations	%	# samples
total	1003259		44423115
T	997255	99,4%	43626555
S	931350	92,8%	31458014
TS	926223	92,3%	30880568

Mediterranean Sea Aggregated Dataset



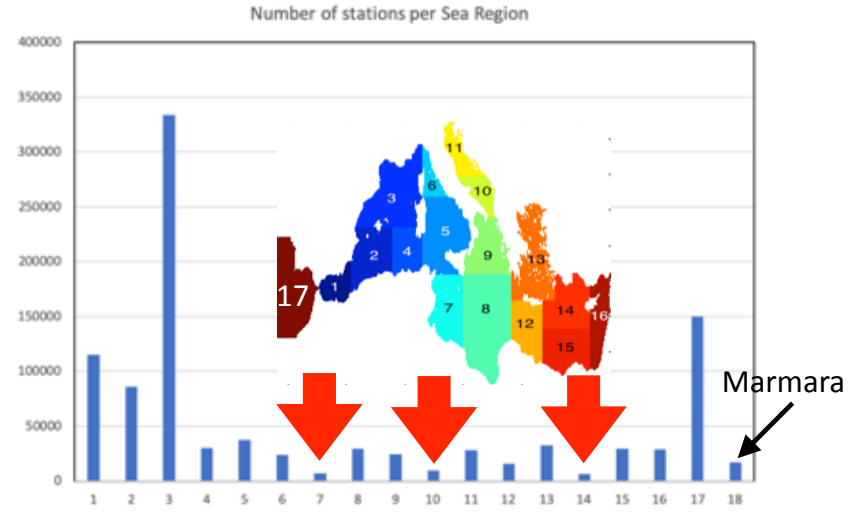






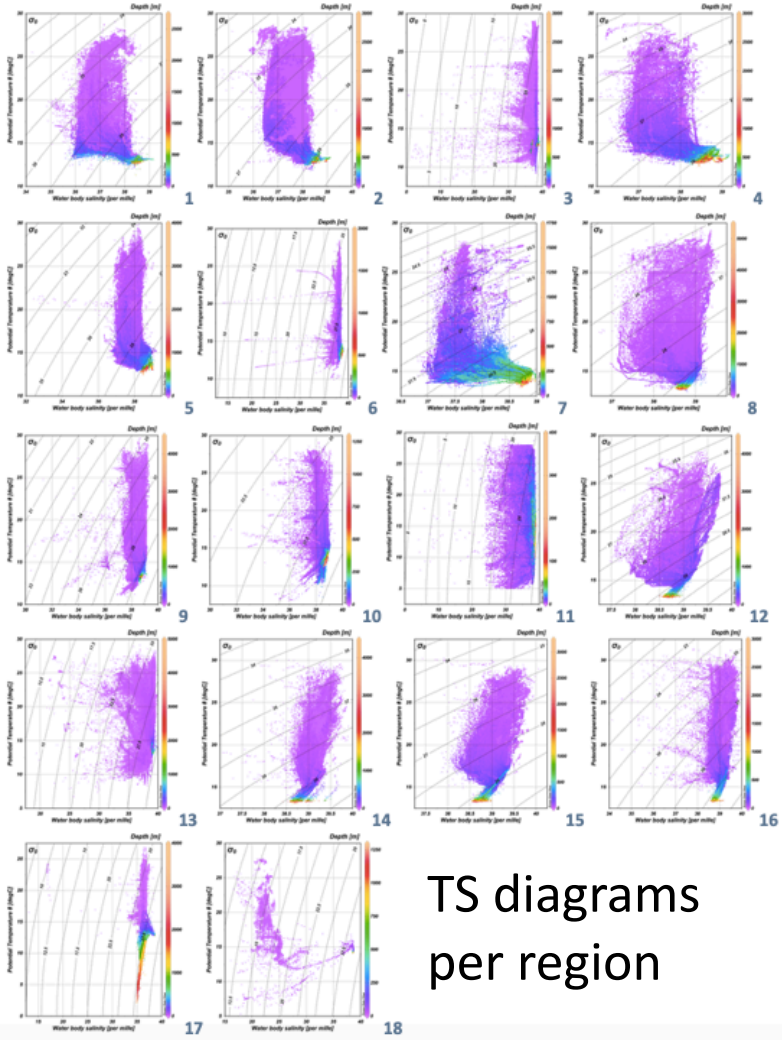
Mediterranean Sea Aggregated Dataset

Quality Control procedure—> per sea region (18), per depth layer, per instrument type



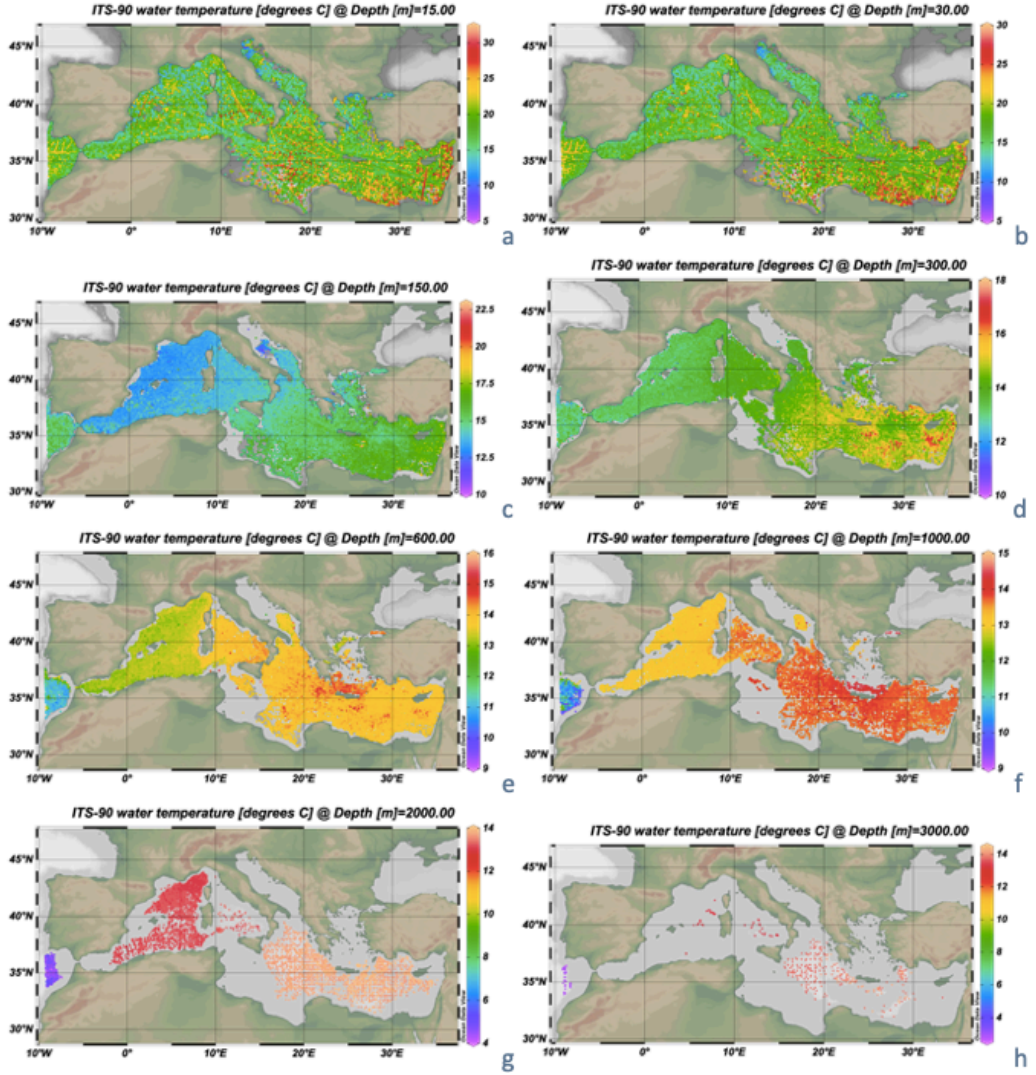
Instrument/Gear Type	V1	%	V2	%
CTD	52031	7	70036	7
bathythermograph	56558	8	59574	6
discrete water sampler	32258	4	37485	4
thermosalinograph	**555269**	**75**	**808364**	**81**
thermistor chains	22	0	22	0
continuous water sampler	1577	0	1654	0
salinity sensor; water temperature sensor	19852	3	25445	3
salinometers	100	0	131	0
none info	**21973**	**3**	**374**	**0**
	739640		1003085	

Mediterranean Sea Aggregated Dataset



TS diagrams per region

Temperature at various depth levels: 15, 30, 150, 300, 600, 1000, 2000, 3000m



Regional Climatologies

- designed with **harmonized approach** to cover the time period after 1955 on World Ocean Atlas (WOA18) standard vertical levels
- monthly fields 1955-2018 and seasonal decadal fields have been produced with associated error fields
- SDC datasets integrated with external sources: World Ocean Database (WOD18) and/or Coriolis Ocean Dataset for ReAnalysis (CORA5.2)
- computed with **DIVAnd** tool
- **validation** through consistency analysis considering WOA18 as reference
- methodology, validation and usability have been reported in **Product Information Documents (PIDoc)** available with the NetCDF files in the SDC Product Catalogue
- **GLO climatology** (*Shahzadi et al. 2021*) computed with data from WOD18 → the spatial coverage of SDN data at global scale is still too sparse

Name	horiz resolution	Time coverage	Seasonal	Monthly	External data sets
GLO_1	1/4°	1900-2017		x	WOD18
GLO _2	1/4°	2003-2017		x	WOD18
ARC _1	1x1/2°	decades	x	x	WOD18
ARC _2	1x1/2°	1955-2014	x	x	WOD18
BAL_1	1/16x1/32°	decades	x	x	CORA5.2
BAL_2	1/16x1/32°	1955-2018	x	x	CORA5.2
NS_1 (V1)	1/8°	1955-2014		x	WOD18
NS_2 (V1)	1/8°	decades	x		WOD18
NAT _1	1/2°	decades	x	x	CORA5.3
NAT_2 (V1)	1/4°	1955-2017	x	x	CORA5.1
NAT_3 (V1)	1/4°	decades	x	x	CORA5.1
MED_1	1/8°	1955-2018	x	x	CORA5.2
MED_2	1/8°	1955-1984	x	x	CORA5.2
MED_3	1/8°	1985-2018	x	x	CORA5.2
MED_4	1/8°	decades	x		CORA5.2
SDC_1	1/8°	1955-1994	x	x	WOD18 and CORA5.2
SDC_2	1/8°	1995-2019	x	x	WOD18 and CORA5.2
SDC_3	1/8°	1955-2019	x	x	WOD18 and CORA5.2
SDC_4	1/8°	decades	x		WOD18 and CORA5.2

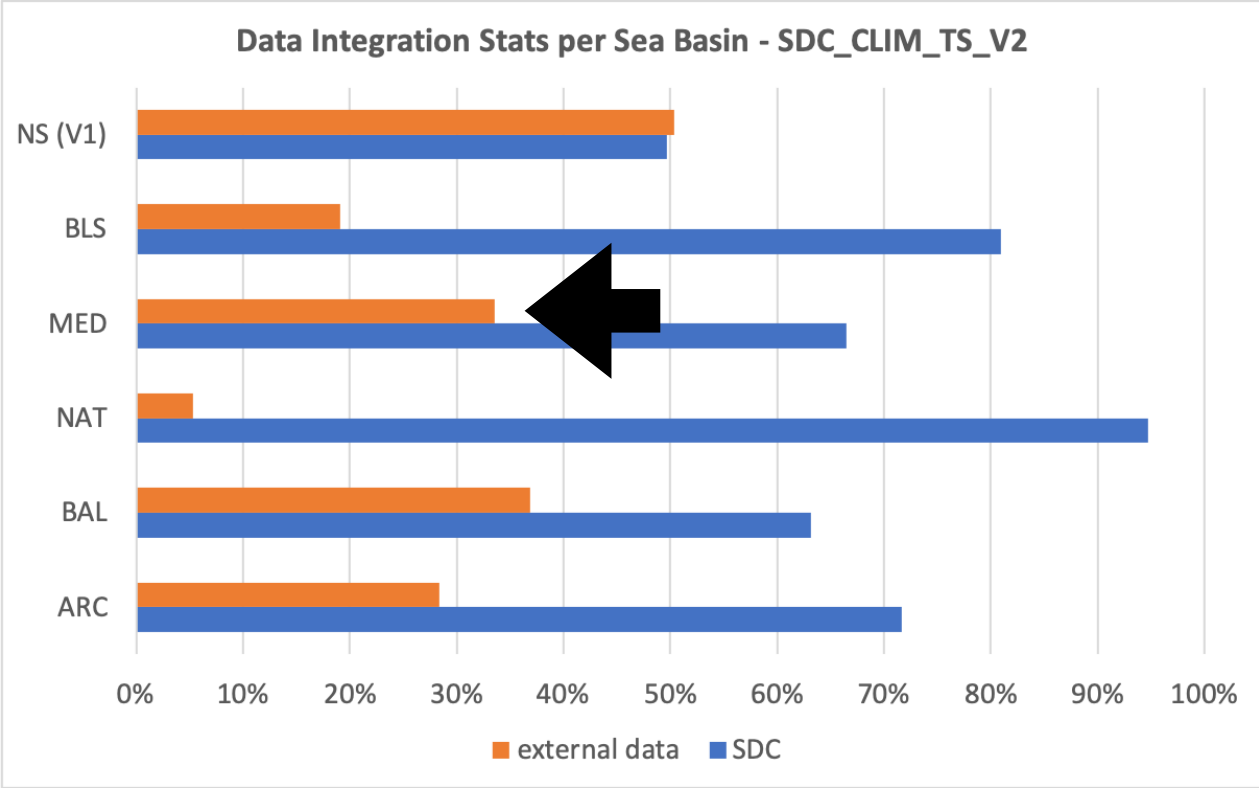
SeaDataCloud

all climatologies on **WOA depth levels**

Integration with external sources

External data imported in ODV and additional QC has been performed through visual inspection before the integration with SDC datasets —> data spatial and temporal consistency

Percentage of SDC data and additional external sources per sea basin:

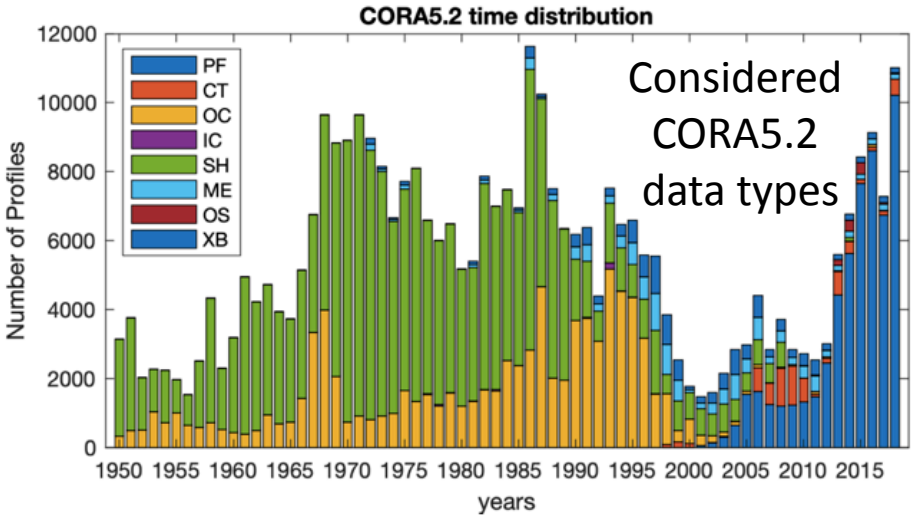


SDC represents the main contributor

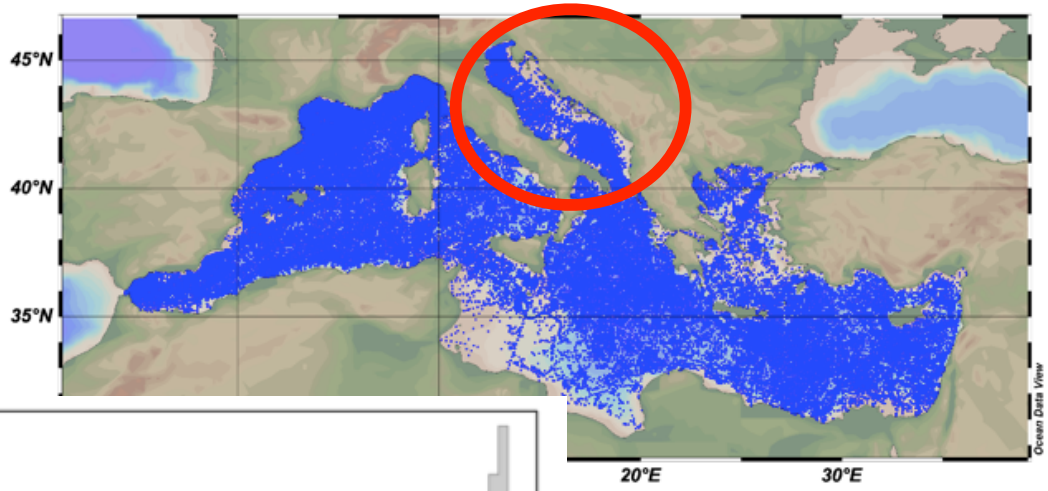
→SDC version of duplicate casts have been retained

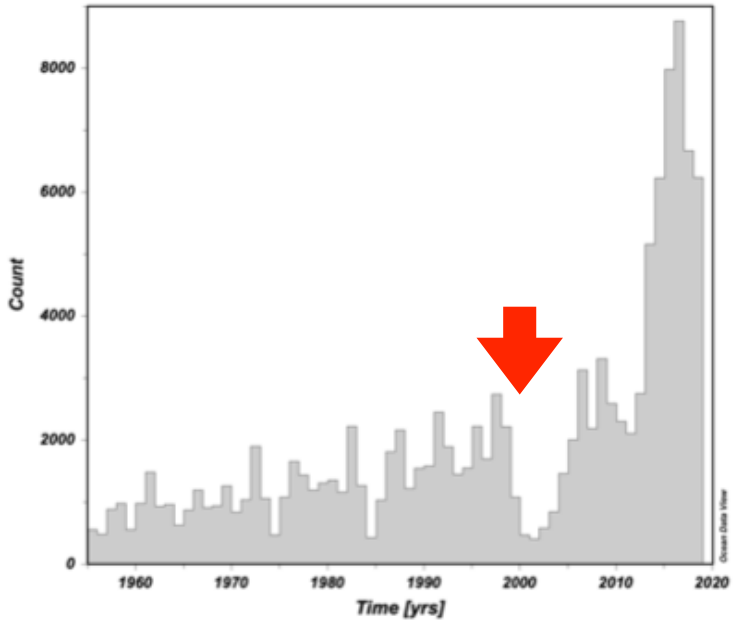
→percentage of complementary external data ranges from 5% in the North Atlantic to 50% in the North Sea

Mediterranean Sea CORA5.2 dataset



The data were reformatted to ODV spreadsheet and imported to ODV collection for additional QC analysis through visual inspection





CORA5.2 subset (1955-2018):

- 355652 stations
- 123827 TS profiles

Mediterranean Sea merged dataset

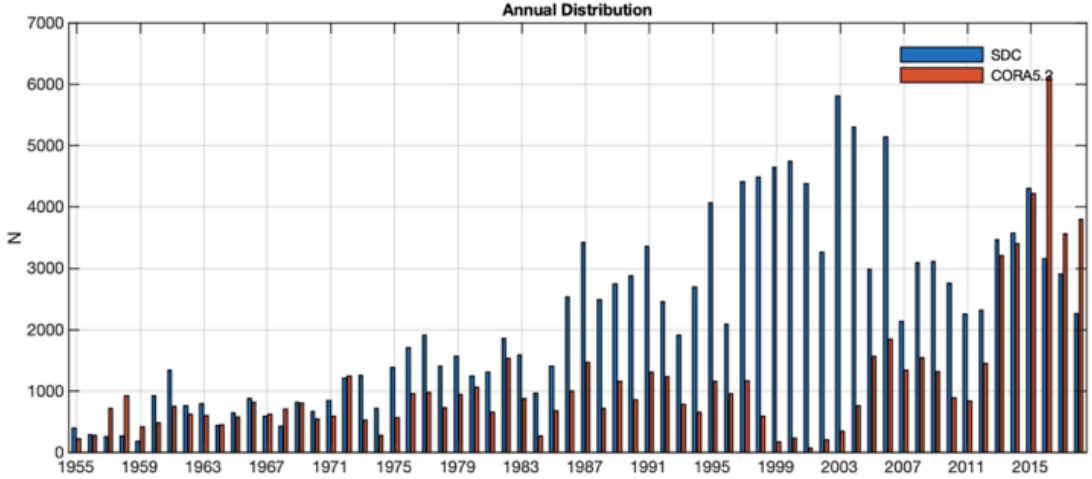
Duplicate check (ODV) looking for "quasi-perfect" duplicates within 0.0001deg (~11m) and 0.0069 days (10 minutes):

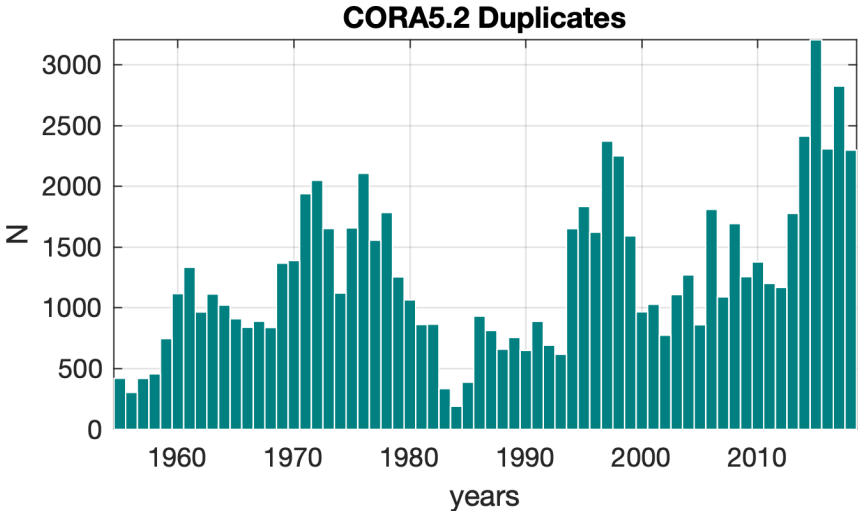
- 88528 (15.5%) duplicates over the 572185 stations, 86971 of which are from CORA5.2
- SDC version of the duplicate data has been retained in the climatology input dataset

Remaining internal duplicates:

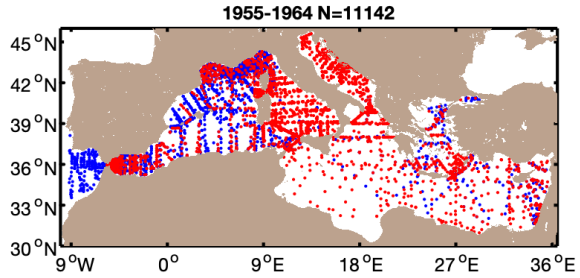
1. repeated samples on the same location with coincident time (62 groups of duplicates—> superobs)
2. profiles sampled with different instruments at the same location (5543 profiles)

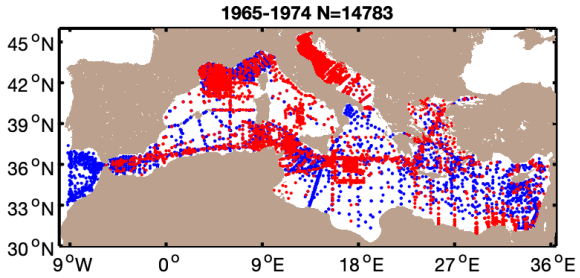
Input dataset: 212734 stations, 66% from SDC and additional 34% from CORA 5.2

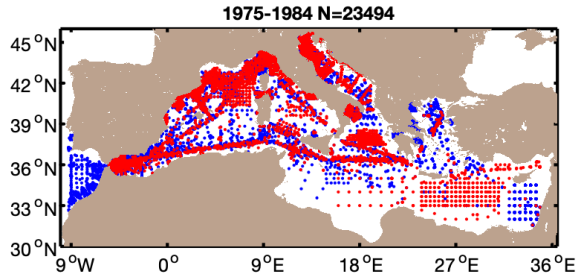


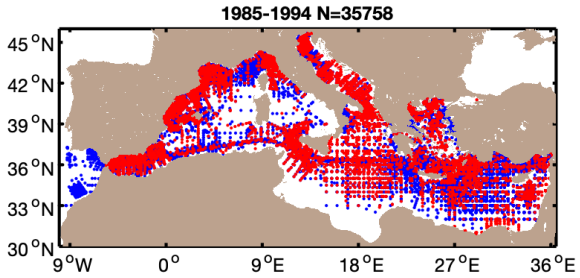


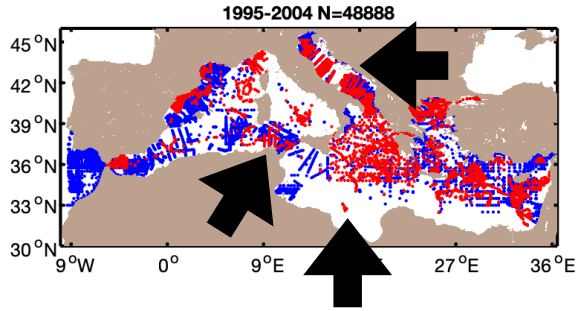
Mediterranean Sea merged dataset

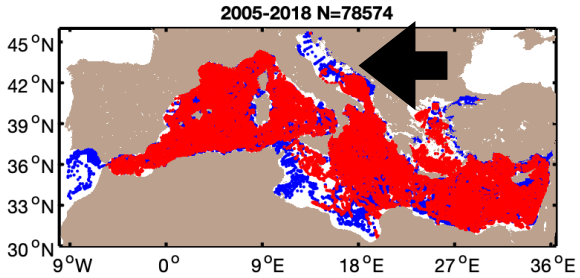


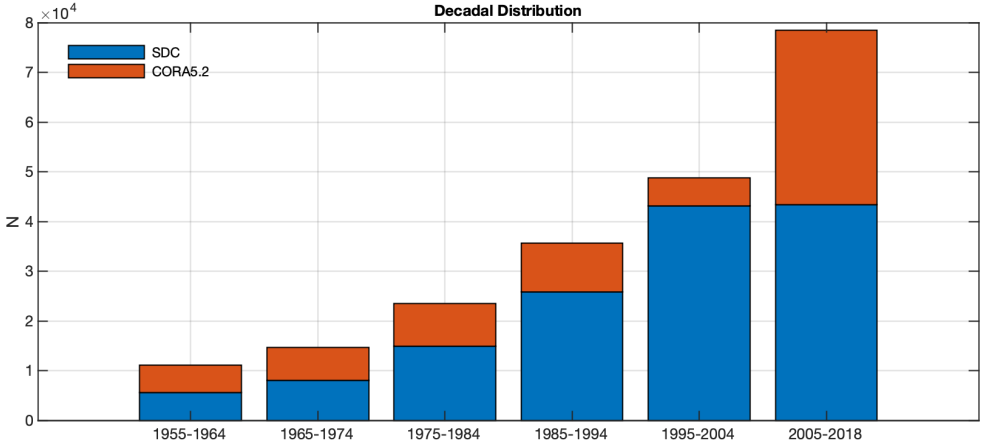








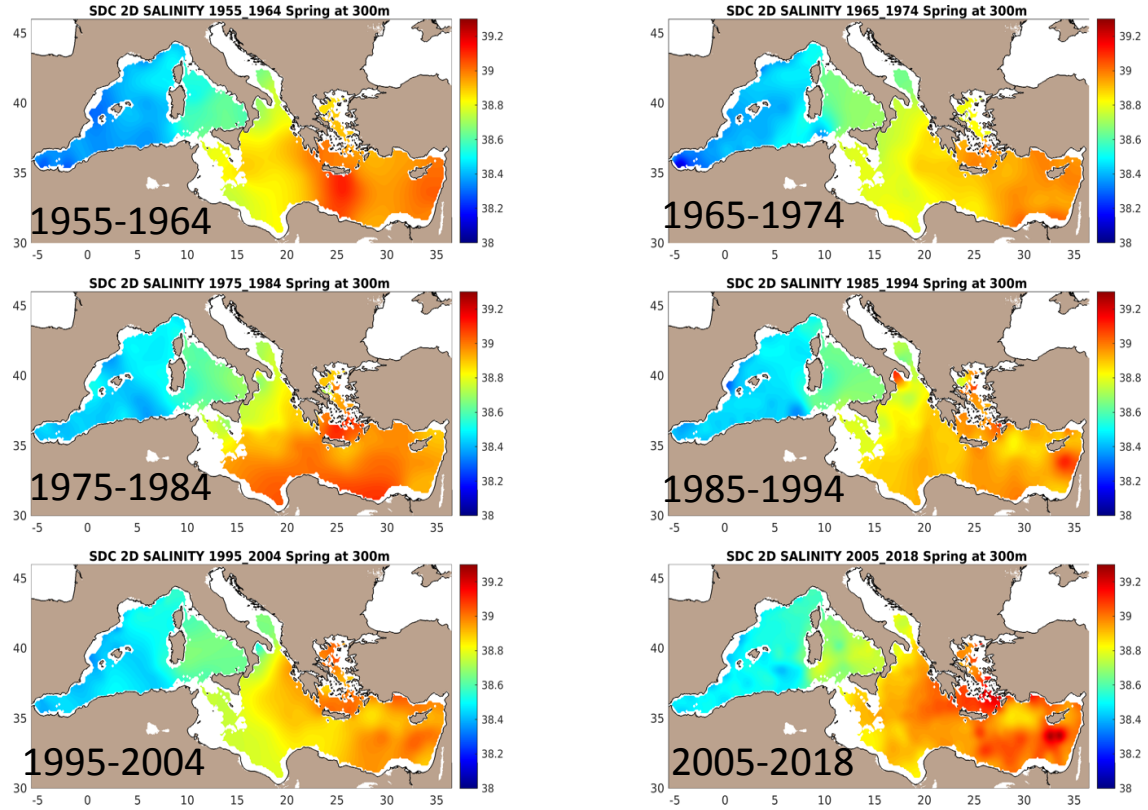




Mediterranean Sea Climatology

SDC_**MED**_CLIM_TS_V2

Spring Salinity at 300m



Salinity background: annual

Temperature background: monthly up to 2000m, computed considering a sliding three-month window; annual below 2000m depth.

3 background fields with different time coverage:

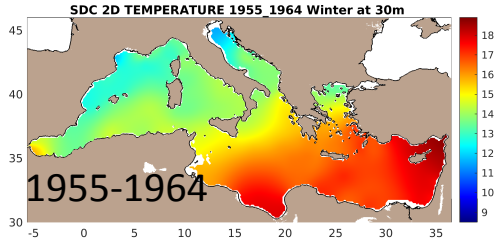
- 1955-2018 bkg —> 1955-2018 analysis
- 1955-1984 bkg —> 1955-1984, 1955-1964, 1965-1974 and 1975-1984 analyses;
- 1985-2018 bkg —> 1985-2018, 1985-1994, 1995-2004 and 2005-2018 analyses

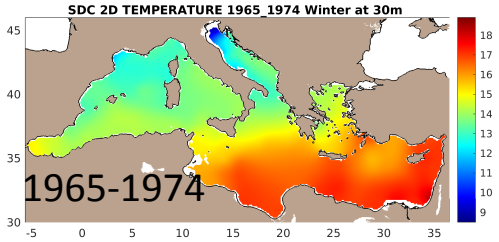
	T Bkg	T An	S Bkg	S An
L (x,y)	5.5°	2°	5.5°	2°
epsilon2	12	1.2 (0-2000m) 0.4 increase up to 6	12	1.2 (0-300m) 0.2 increase up to 6

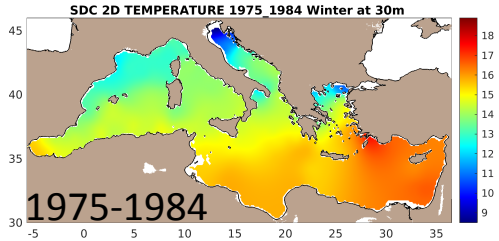
Mediterranean Sea Climatology

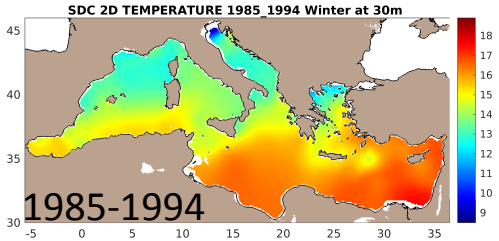
SDC_**MED**_CLIM_TS_V2

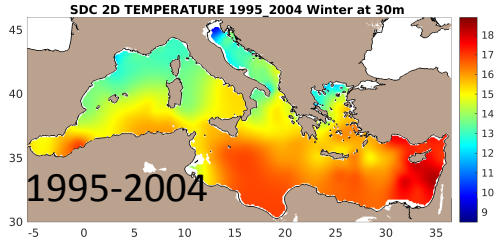
Spring Temperature at 30m

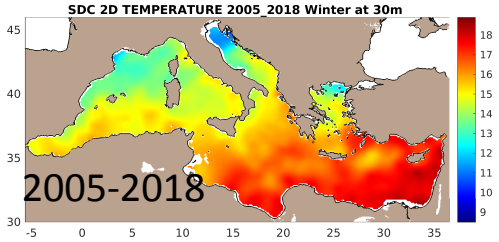


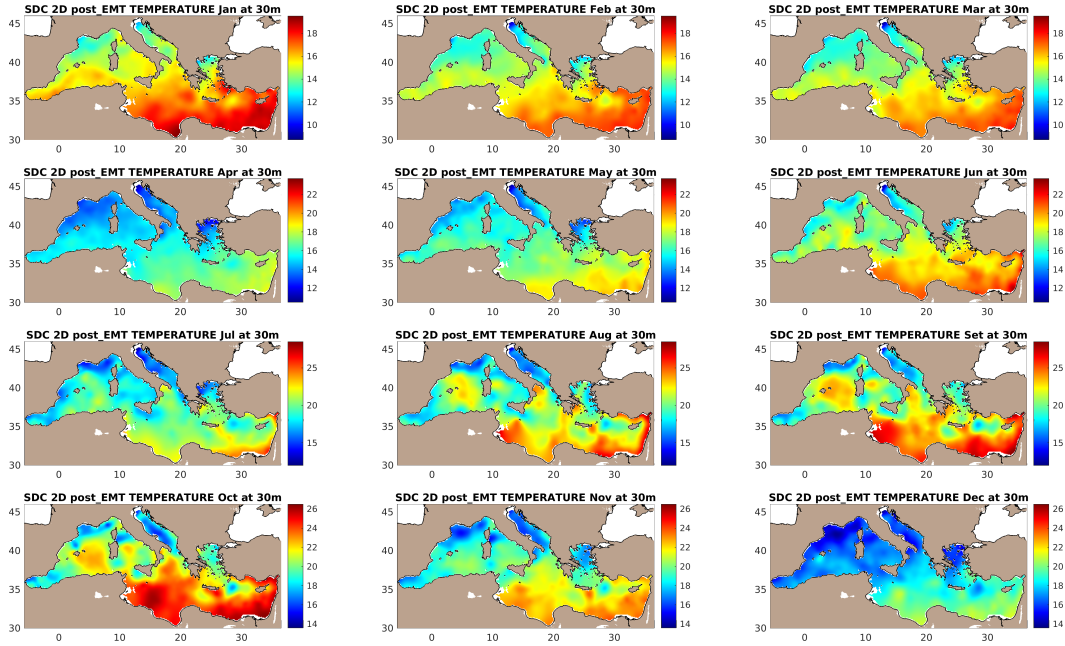












Temperature at 30m
1985-2018

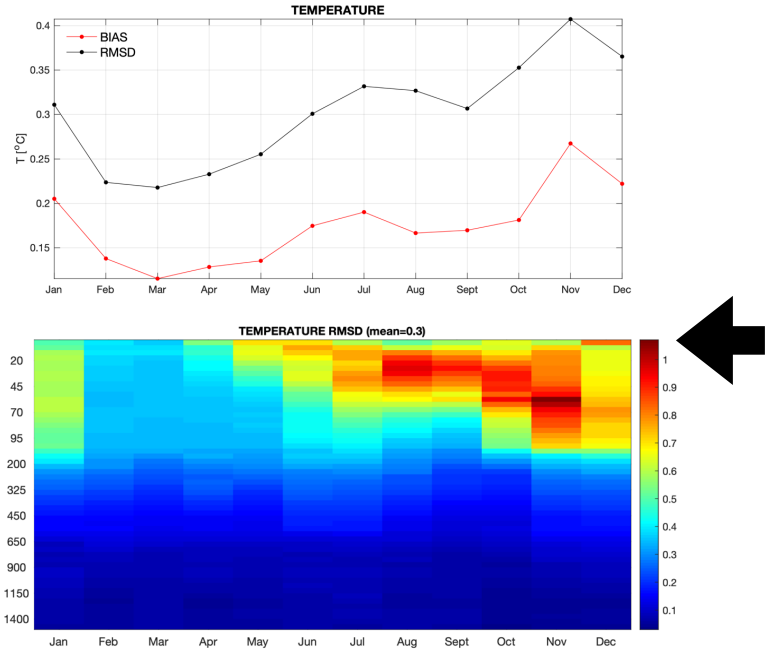
Consistency Analysis

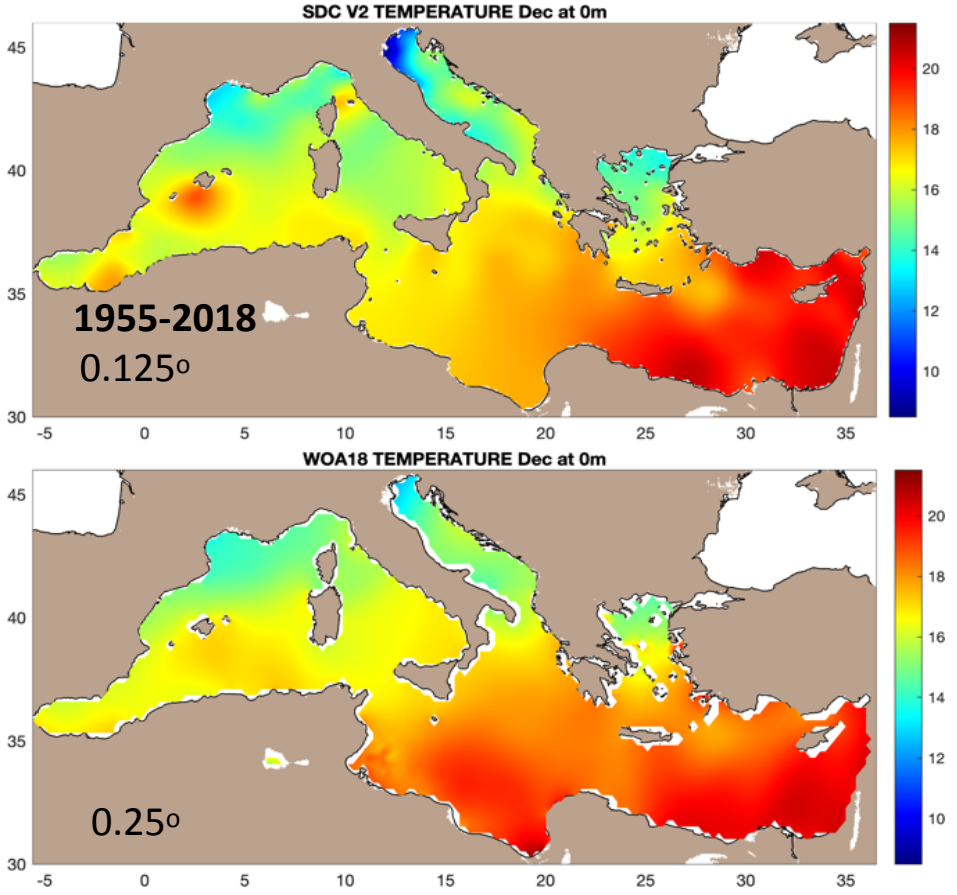
Validation is essential for products reliability → global WOA18 climatology as reference

Twofold scope: (1) to detect possible drawbacks; (2) to highlight strength/weaknesses of SDC climatologies → usability information

Qualitative (visualization): ARC, BAL, NAT, NS

Quantitative (quality indices): GLO, MED, BLS





Northern Adriatic Sea

T and S monthly and seasonal high resolution climatologies for the Northern Adriatic Sea produced with DIVAnd v2.6.1 (Data-Interpolating Variational Analysis in n dimensions, *Barth et al., 2014*)

Product details:

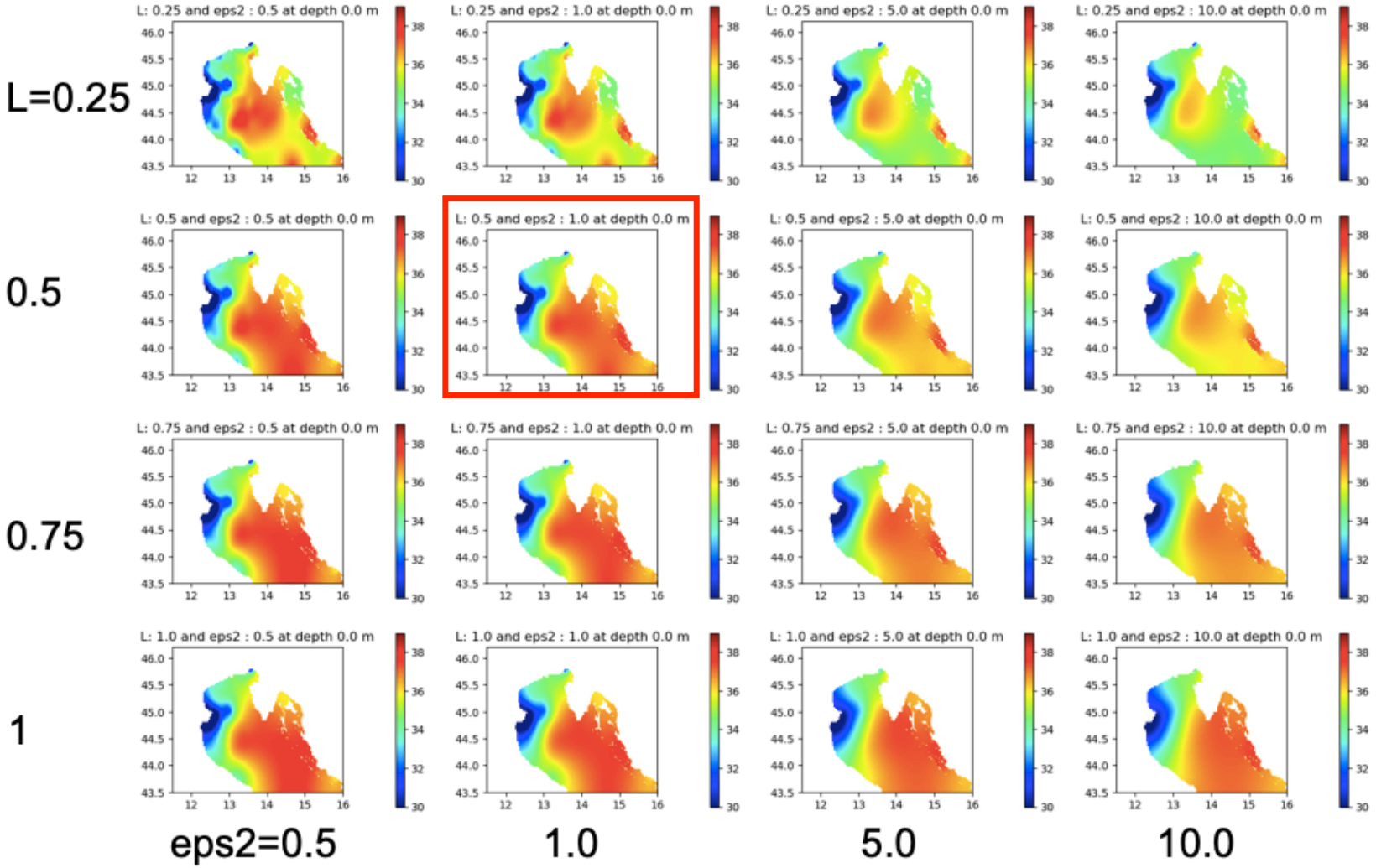
- horizontal resolution: 1/36^{o} (~ 3 km)
- vertical resolution: 5 m (11 equally spaced layers from the surface down to 50 m depth)
- temporal resolution: monthly and seasonal for the periods 1955-2016, 1955-1984, 1985-2016, seasonal for the decades from 1955 to 2016

DIVAnd implementation and settings:

- topography from GEBCO_2019 bathymetry
- **L=2**deg background, **L=0.5**deg analysis
- **eps2=12** background, **eps2=1** analysis
- monthly background for temperature, annual background for salinity
- 2D interpolation for each layer of the grid (vertical correlation length = 0)
- iterative process which computes the residuals first and then re-run the analysis excluding the observations with residuals larger than two standard deviations from the mean of the residuals population

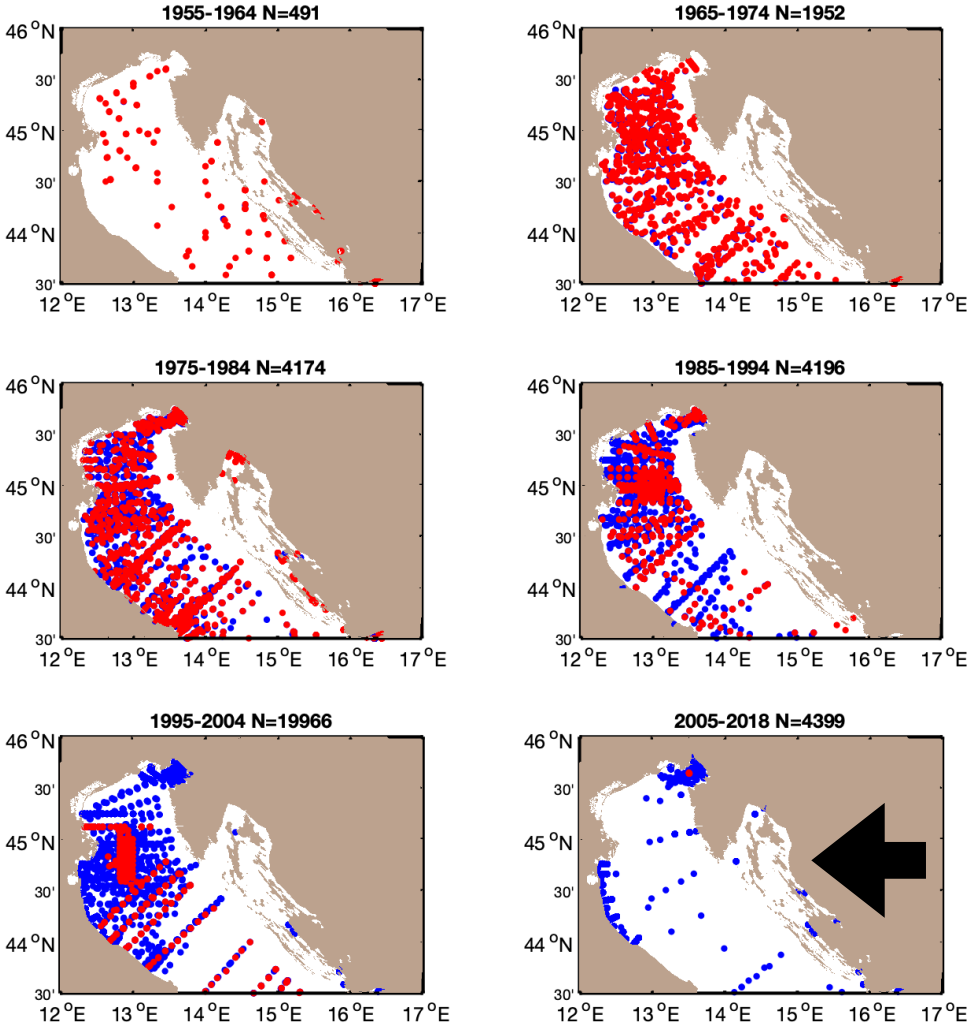
Delrosso et al. (2021) in https://archimer.ifremer.fr/doc/00720/83160/

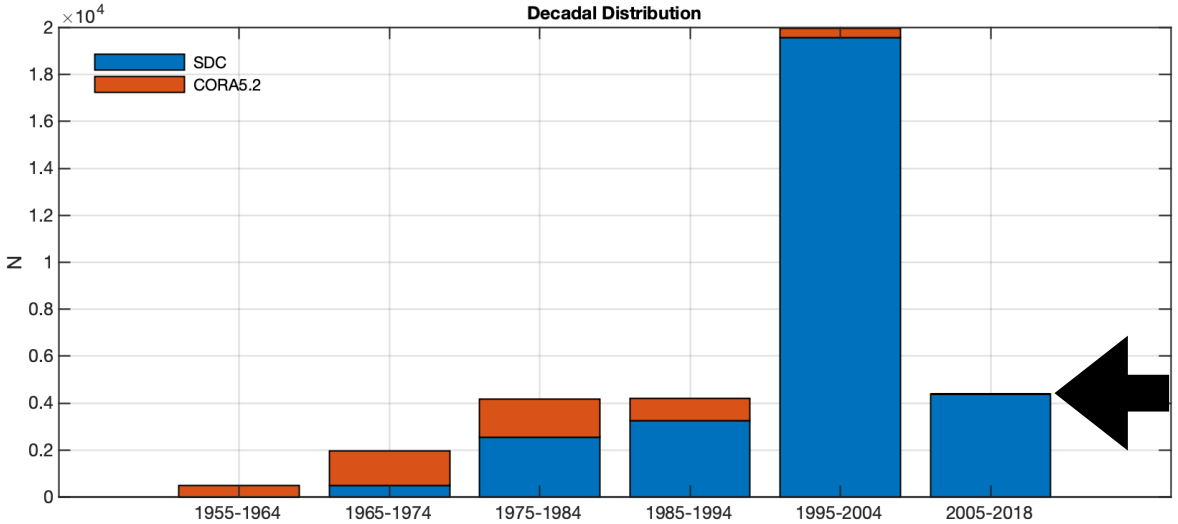
Tuning/optimization of *L* (correlation length) and *epsilon2* (the inverse of the signal-to-noise ratio) parameters has been performed, in order to find a robust balance between the minimization of residuals and the smoothness of the climatological fields produced.

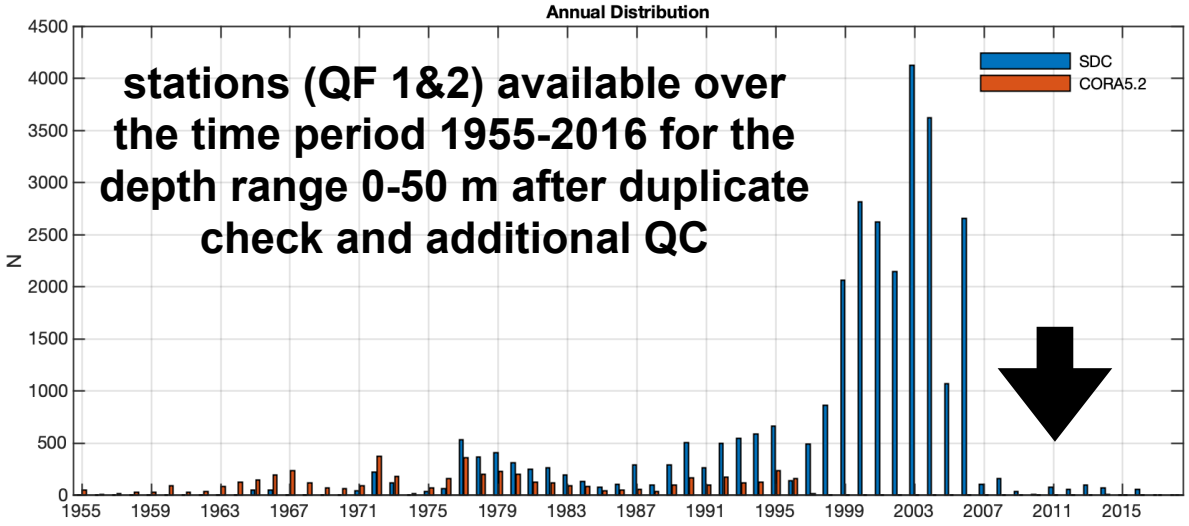


EMODnet PHYSICS
Oceans at your fingertips

Northern Adriatic Sea







stations (QF 1&2) available over the time period 1955-2016 for the depth range 0-50 m after duplicate check and additional QC

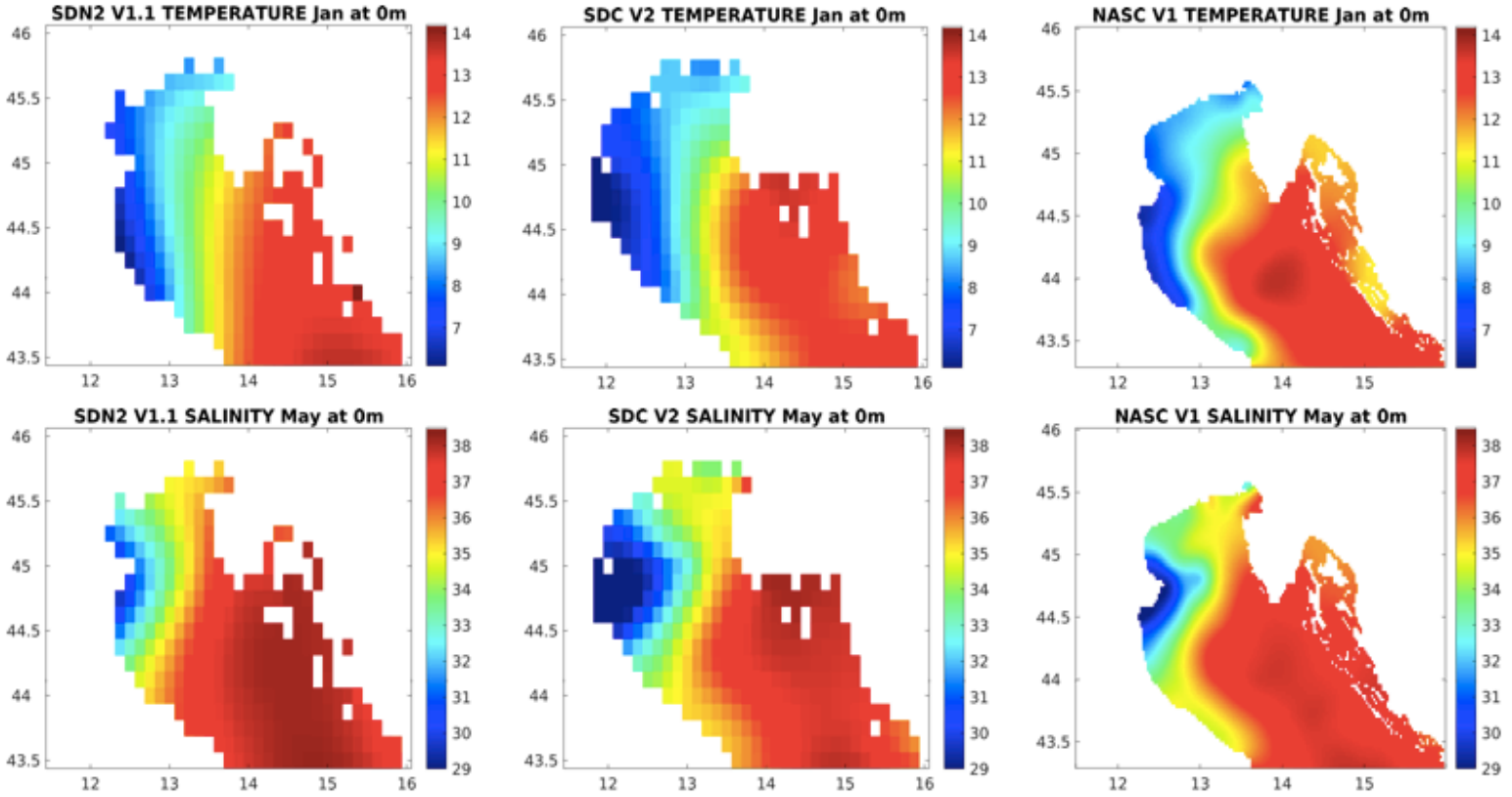
EMODnet PHYSICS
Oceans at your fingertips

Northern Adriatic Sea

SDN2 V1.1
(1900-2013)
Hor res.: 1/8°
33 iode levels
L: 2° (analysis); 10° (bckg)

SDC_MED_CLIM_TS_V2
(1955-2018)
Hor res.: 1/8°
92 WOA levels
L: 2° (analysis); 5.5° (bckg)

NASC V1
(1955-2016)
Hor res.: 1/36°
11 WOA levels
L: 0.5° (analysis); 2° (bckg)



- good agreement with the considered benchmark products
- more realistic topography
- small scale patterns become apparent
- Po river plume is better resolved

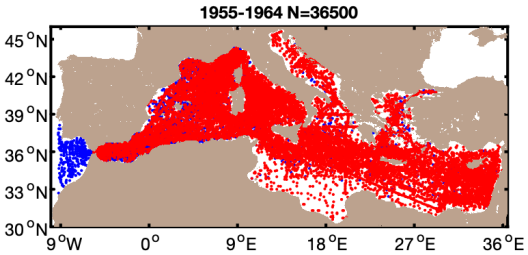
New Data Products

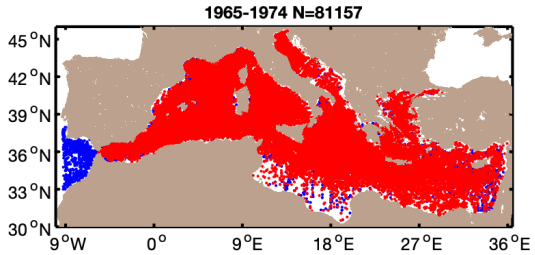
The experts team explored the feasibility of new data products and the capability of SDN infrastructure to release systematically advanced products to monitor the ocean state in view of the UN Ocean Decade and in line with other initiatives (EMODnet, CMEMS)

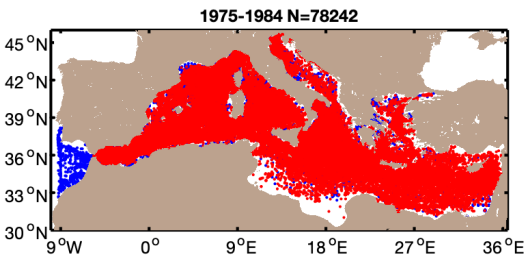
11 new products have been released:

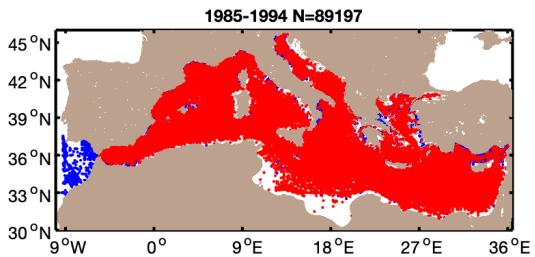
GLO_DP1	Density and Brunt-Väisälä climatology (2003-2017)
GLO_DP2	Apparent Oxygen Utilization (AOU) at 1/4° (2003-2017)
BAL_DP1	regional and sub-regional T and S monthly statistics
NAT_DP1	MLD monthly climatology at 1/4°
MED_DP1	MLD monthly climatology at 1/8°
MED_DP2	OHC anomalies time series and trends (0-700m; 0-2000m)
BLS_DP1	Monthly climatology of Cold Intermediate Layer (CIL) cold content at 1/8°
BLS_DP2	Decadal seasonal CIL cold content at 1/8°
BLS_DP3	Sliding decades CIL cold content at 1/8°
ULG_DP1	Currents map from HF radars data (*Barth et al., 2021*)

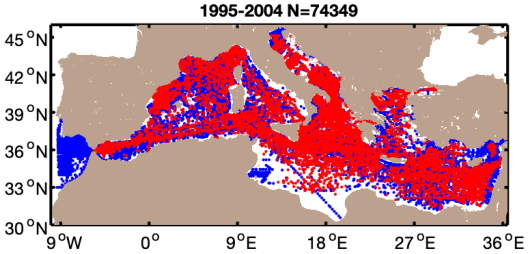
Med Sea Ocean Heat Content (OHC)

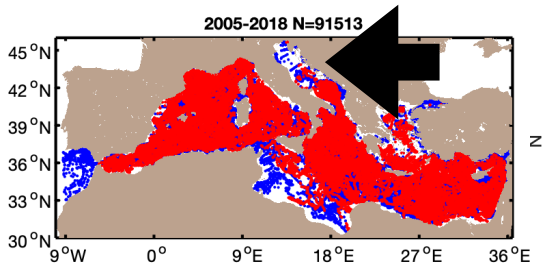


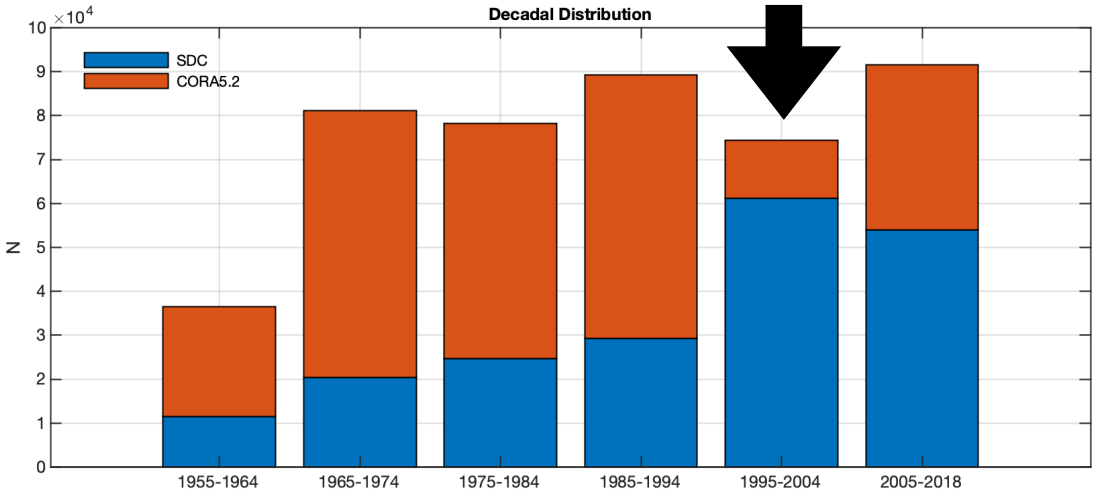


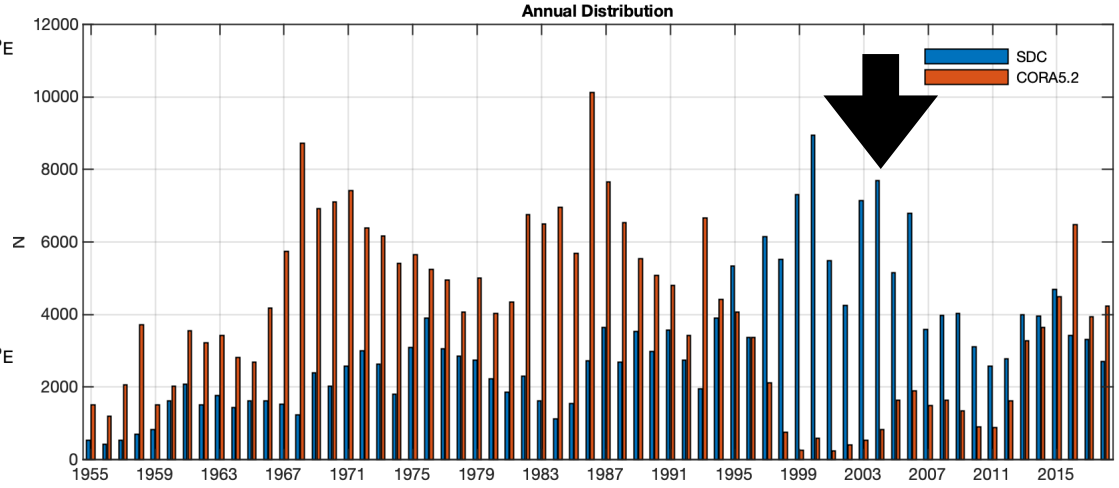






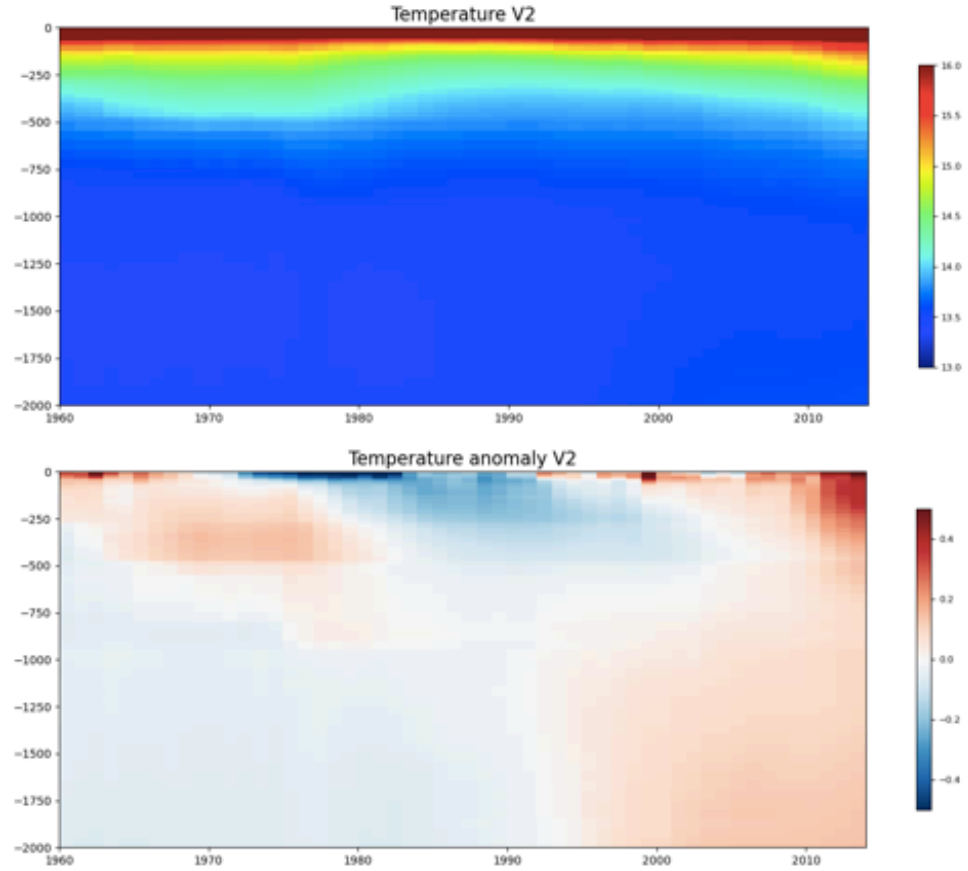






Med Sea Ocean Heat Content (OHC)

SDC OHC product consists of sliding decadal temperature fields and annual OHC anomalies deriving from them in layers: (1) 0-700m; (2) 0-2000m

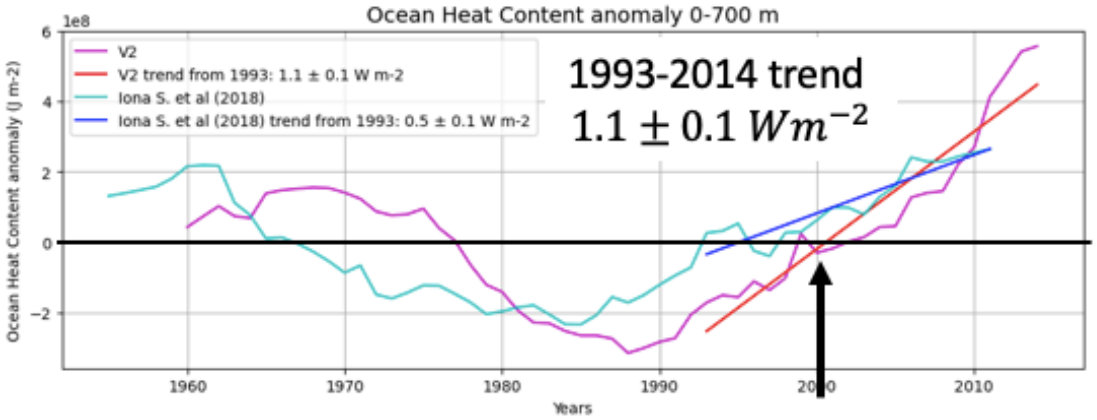


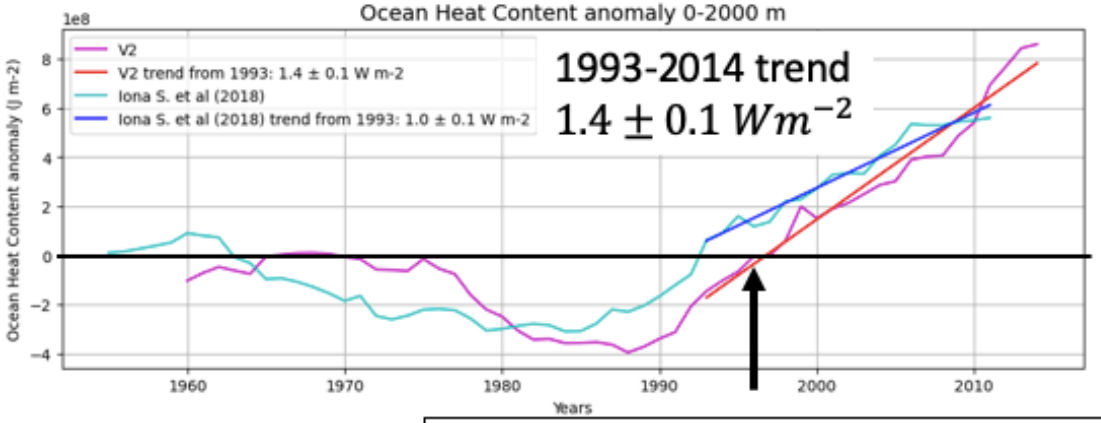
T baseline climatology computed from averaging decadal climatologies: 1955-1964, 1965-1974, 1975-1984, 1985-1994, 1995-2004, 2005-2014

- Positive anomalies characterize the surface layer until early '70s
- Negative anomalies dominate until mid '90s, then positive anomalies prevail again
- Intermediate layer behavior (~200-600m) appears shifted 10 years in time with respect to the surface layer
- Negative anomalies dominate below 700m until mid '90, positive anomalies prevail afterwards

Simoncelli and Oliveri (2021) https://doi.org/10.13155/79146

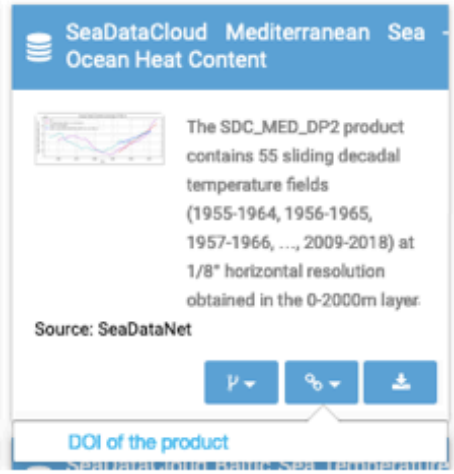
Med Sea Ocean Heat Content (OHC)





in agreement with *Cheng et al. (2021) AAS*

dataset available at SeaDataCloud web catalog
https://www.seadatanet.org/Products#/



Late '90s → positive anomalies and progressive OHC increase

Product Information Document → *Simoncelli and Oliveri (2021) https://doi.org/10.13155/79146*

Conclusion

- SDN data value chain ends with the generation of data products, whose quality reflects the coordination capacity in managing multidisciplinary in situ data but also in developing and adopting software/tools through continuous feedback
- regional data collections showed a progressive increase of the available data and relative quality
- a novel metadata analysis allowed to monitor/acknowledge the EU data sharing landscape, to detect systematic (format, flagging) errors and data/metadata omissions
- Climatologies: decadal gridded fields, integration of external data, DIVAnd software uptake, good consistency with WOA and added value due to the highest resolution and regional QC
- **SDC co-development/production** process is in place and will be further advanced exploiting the new SDC Virtual Research Environment (VRE)
- data products and PIDocs have DOIs and are available at

https://www.seadatanet.org/Products#/search?from=1&to=30

SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR
OCEAN & MARINE DATA MANAGEMENT

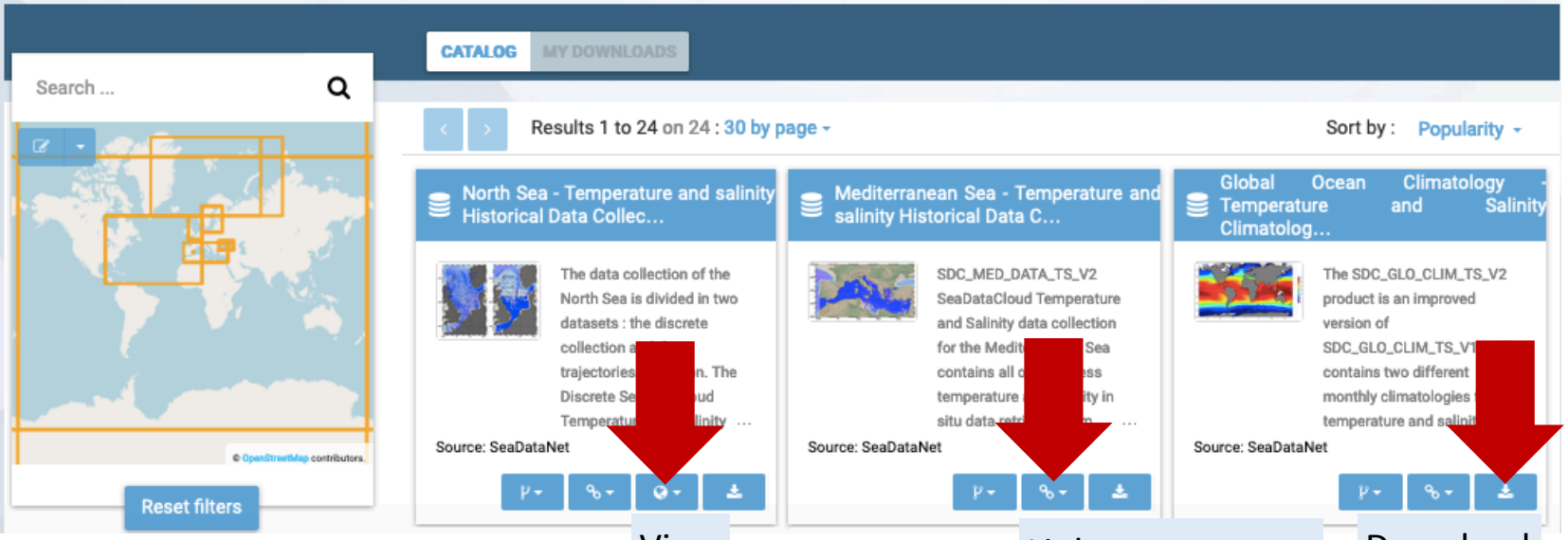
https://www.seadatanet.org/Products#/

ABOUT US METADATA DATA ACCESS STANDARDS SOFTWARE PRODUCTS EVENTS PUBLICATIONS

DATA PRODUCTS

DESCRIPTIONS & DOCUMENTATION | USER MANUAL

SeaDataNet provides aggregated datasets (ODV collections of all unrestricted SeaDataNet measurements of temperature and salinity by sea basins) and climatologies (regional gridded field products) based on the aggregated datasets and data from external data sources such as the COriolis Ocean Dataset for Reanalysis (CORA) and the World Ocean Database (WOD) for all the European sea basins and the Global Ocean. Each SeaDataCloud product is described in a Product Information Document (PIDoc) that can be accessed from the product's landing page. Click on the User Manual tab on the top-right corner of the page for instructions.

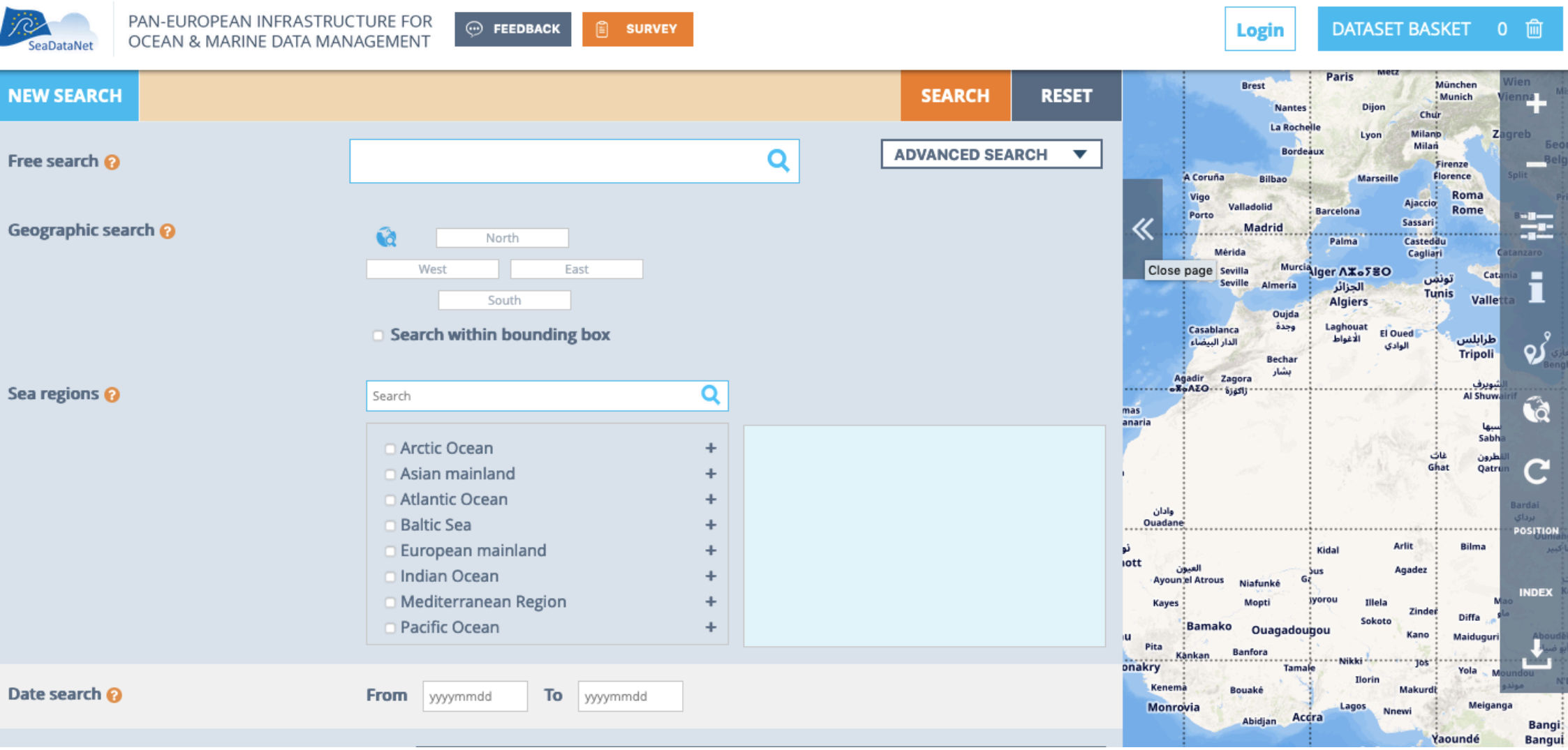


View

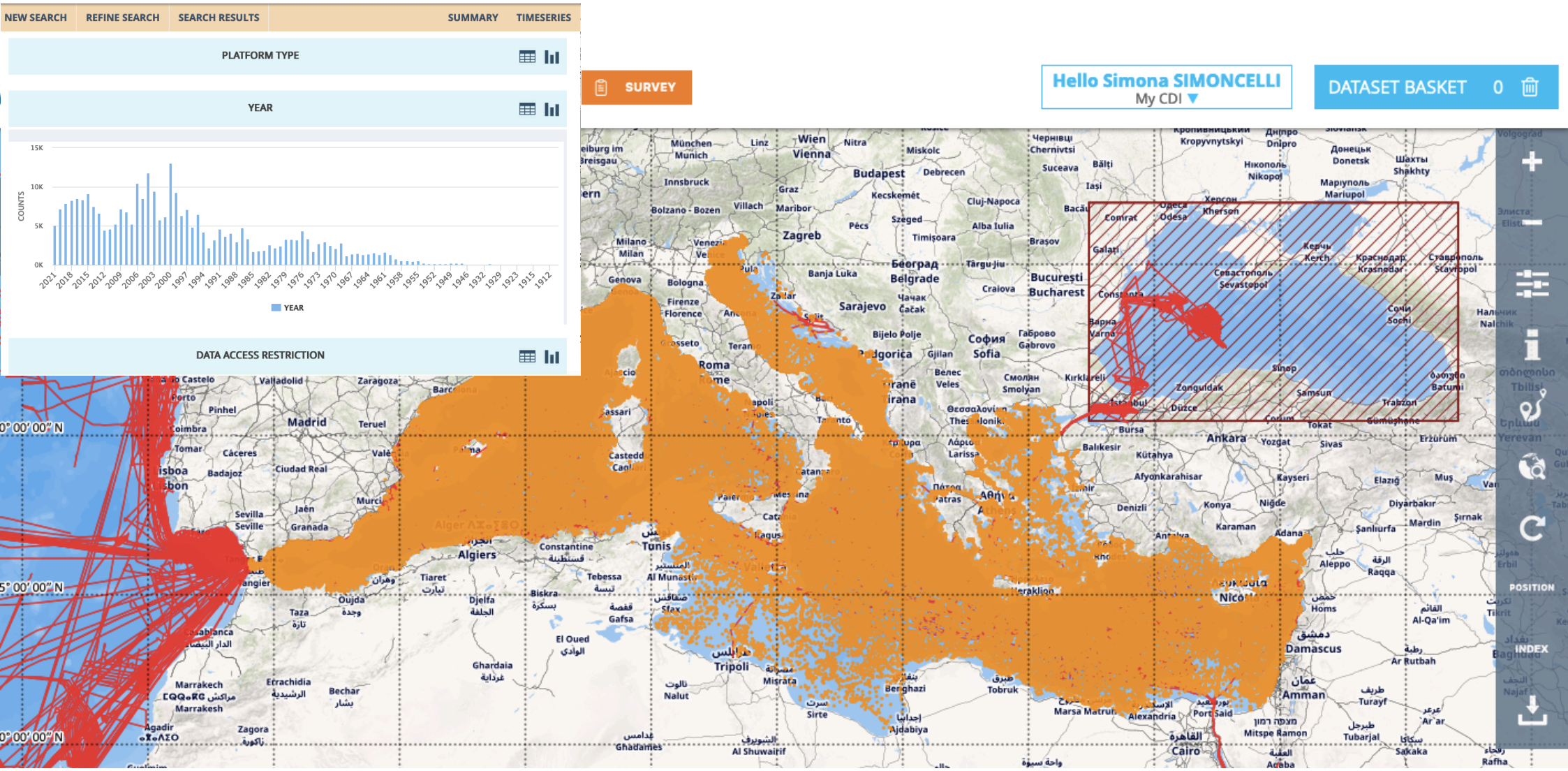
Link
DOI landing page
ERDDAP

Download

https://www.seadatanet.org/



Url for this query = https://cdi.seadatanet.org/search/welcome.php?query=1929&query_code={1AA56DD6-543F-49E0-B2BD-A0A22DA54DA3}



SeaDataCloud

Thanks to all partners

Ifremer

METU-IMS

INSTITUTE OF MARINE RESEARCH

ALMA MATER STUDIORUM A.D. 1088

SMHI

museum
Royal Belgian Institute of Natural Sciences

AWI

Université de Liège

MARIS

INGV

References and Acronyms

WOA	https://www.ncei.noaa.gov/products/world-ocean-atlas
WOD	https://www.ncei.noaa.gov/products/world-ocean-database
CORA	*Szekely et al., 2019* https://archimer.ifremer.fr/doc/00595/70726/
ODV	https://odv.awi.de/
DIVAnd	*Barth et al., 2014* http://dx.doi.org/10.5194/gmd-7-225-2014
QAS	Quality Assurance Strategy
PIDoc	Product Information Documents
VRE	Virtual Research Environment https://www.seadatanet.org/Software/VRE

*D Delrosso, **S Simoncelli**, P Oliveri, A Guarnieri, A Novellino* (2021). EMODNET PRELIMINARY HIGH-RESOLUTION TEMPERATURE AND SALINITY CLIMATOLOGIES FOR THE NORTHERN ADRIATIC SEA In: Fernandez et al. (eds.) (2021). Advances in operational oceanography : expanding Europe's ocean observing and forecasting capacity. Proceedings of the 9th EuroGOOS International Conference. 3 – 5 May 2021, Online Event 2021, EuroGOOS. Brussels, Belgium. 574 pp. https://archimer.ifremer.fr/doc/00720/83160/