Product Information Document (PIDoc)

SeaDataCloud Temperature and Salinity Historical Data Collection for the Baltic Sea (Version 2)

SDC_BAL_DATA_TS_V2

HORIZON 2020

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

SeaDataCloud - Further developing the pan-European infrastructure for marine and ocean data management

Grant Agreement Number: 730960

Product Name

SDC_BAL_DATA_TS_V2

Extended name

SeaDataCloud Temperature and Salinity Historical Data Collection for the Baltic Sea (Version 2)

Product DOI

https://doi.org/10.12770/6a881c9e-1d38-4edd-84d3-c6b212c27eb7

Short description

The SeaDataCloud Temperature and Salinity Historical Data Collection for the Baltic Sea includes open access in situ data on temperature and salinity for the period 1900 – 2019. The data were retrieved from the SeaDataNet infrastructure in August 2019. The dataset format is ODV binary collections [4]. The quality control of the data has been performed with the help of ODV software. Data Quality Flags have been revised following common recommended procedures defined under SeaDataNet 2 project [3] in conjunction with visual expert check.

Authors

Örjan Bäck, Christine Coatanoan, Volodymyr Myroshnychenko

Dissemination	Copyright terms
Public	

How to Cite

Örjan Bäck, Christine Coatanoan, Volodymyr Myroshnychenko (2020). SeaDataCloud Temperature and Salinity Historical Data Collection for the Baltic Sea (Version 2). Product Information Document (PIDoc).

History

Version	Authors	Date	Comments
V0	Ö. Bäck	20/01/2020	First draft
V0.1	V. Myroshnychenko	24/01/2020	Revision
V1	Ö. Bäck	07/02/2020	Final version

Table of contents

Abstract

The SeaDataCloud [2] Temperature and Salinity Historical Data Collection for the Baltic Sea include open access in situ data on temperature and salinity for the period 1900 – 2019. The data were retrieved from the SeaDataNet infrastructure at the end of August 2019. The dataset format is ODV binary collections [4]. The quality control of the data has been performed with the help of ODV software. Data Quality Flags have been revised following common recommended procedures defined under SeaDataNet 2 project [3] in conjunction with visual expert check.

Whenever SDC_BAL_DATA_TS_V2 product is used, this PIDoc should be cited in any publication. We also ask users to remember that hard-working scientists made these measurements, often under severe conditions. Further, the data providers normally possess insight on the quality and context of the data not always shared with the SeaDataCloud team. Hence, inviting data providers and product leaders to collaborate in scientific investigations that depend on their data and data products is considered good and fair practice. Importantly, this will promote further sharing of data and will be beneficial to science.

1. General description of the data collection

The Baltic Sea historical data set contains about 14.4 million values for both salinity and temperature and covers the years 1900-2019. The collection includes open access data retrieved from the SeaDataNet infrastructure at the end of July 2019. The majority of all measurements contain both temperature and salinity data, Table 1. Most of the data are from profiles, dots in Figure 1(a), but there are also some high resolution data that are from trajectories (ferrybox system), looks like solid lines in Figure 1(a). Figure 1(b) shows the data density map with data availability. The overall geographical coverage is good with some exceptions, especially along some of the coasts.

The time distribution in Figure 2(a) presents low data availability for the first 60 years, it increases somewhat after 1960 until around 1990 or late 1980s where it increases again and is stable at high levels. In the latest years there is a decrease in data which is caused by a natural time lag between sampling and data publication in the SeaDataNet system. Seasonal distribution shows a good spread of data during the whole year, Figure 2(b).

Ferrybox data with high resolution in space and time makes some statistical figures skewed. In the collection the total number of stations is 484,624 of which the ferrybox data stands for 213,094, which is 44% of all stations. However when comparing number of actual values you get a better picture. The total number of values is 14,753,042 of which 213,094 are from ferrybox, 1.4%. If looking at the upper 5 metres only (0-5 metres); the ferrybox data stands for almost 9% of all the data.

To get a better understanding of the data distribution, data have been split in two parts; profiles and ferrybox data. Distribution plots have been created separately for each part. Figure 3 shows the geographical distribution of ferrybox data and Figure 4 the data distribution in time. The geographical and seasonal distribution of the ferrybox data is rather good, but the data originates from scattered years so the time coverage is not complete. The data distribution of the profile data, both geographical and temporal, can be seen in Figure 5 and Figure 6. The data coverage is best in the Skagerrak and the Kattegat, also decent in the central parts of the Baltic Proper. The worst coverage is found in some coastal areas in the Baltic Proper and the Gulf of Bothnia, also lacking data in the inner parts of the Gulf of Finland.

Comparing to the historical data set constructed previous in SeaDataCloud (SDC_V1), which in the Baltic covered the years 1900 – 2017, there is an increase by 5% for both temperature and salinity (from around 14 million values to 14.7 million values).

	Number of values	% of total
Depths	14,753,042	
Temperature	14,384,851	97.5
Salinity	14,425,429	97.8
T and S	14,174,006	96.1

Table 1. Number of values for temperature and salinity in the entire data set for the Baltic Sea.

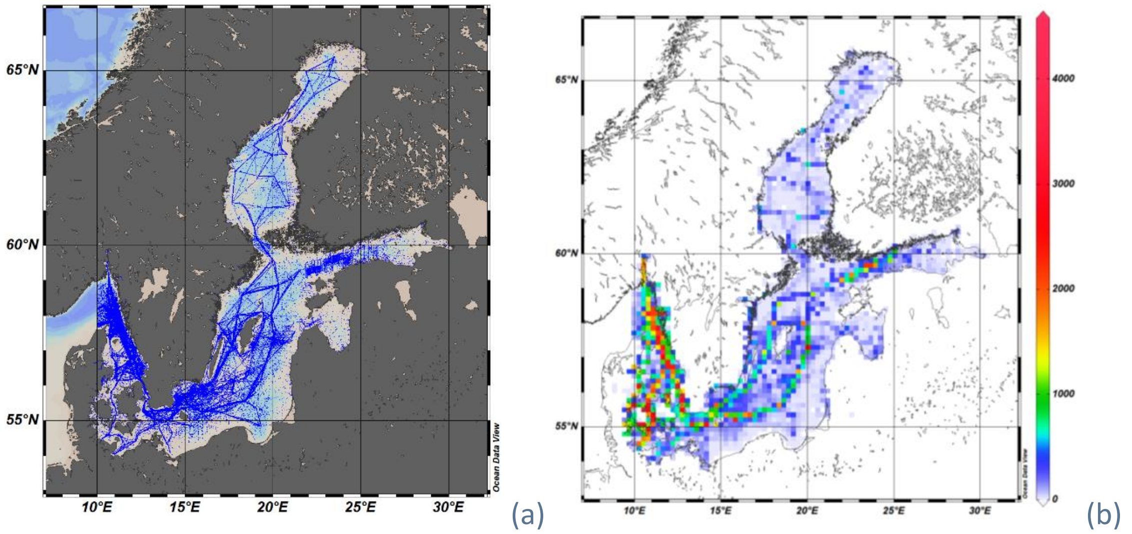


Figure 1. Map with all data, dots are profiles and what appears to be lines are high resolution ferrybox data (a). To the right a data density plot showing where most values have been sampled (b).

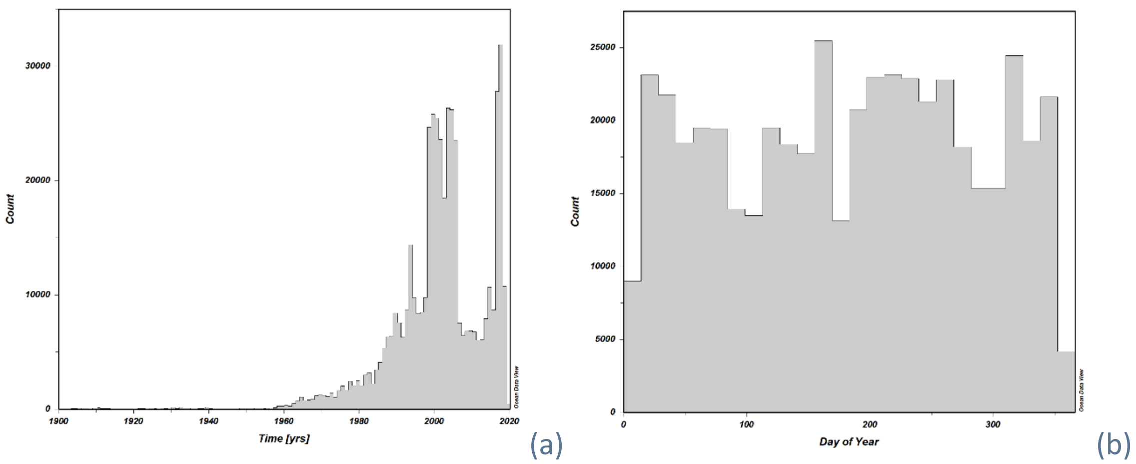
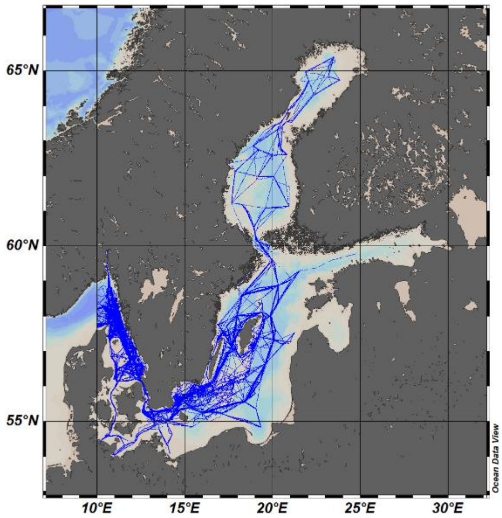
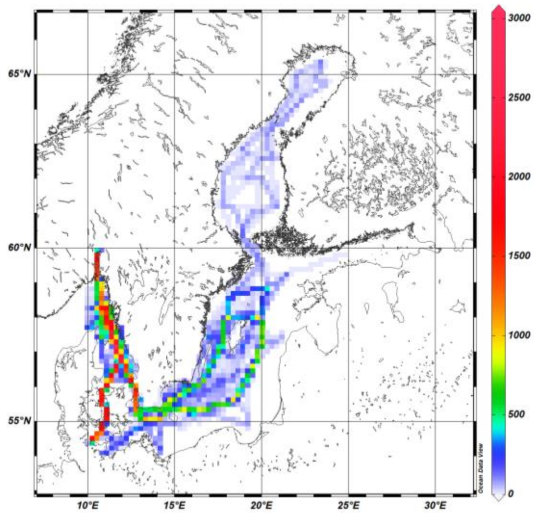


Figure 2. Temporal distribution (a) and seasonal distribution over the year (b) for the entire data set.

Ferrybox data

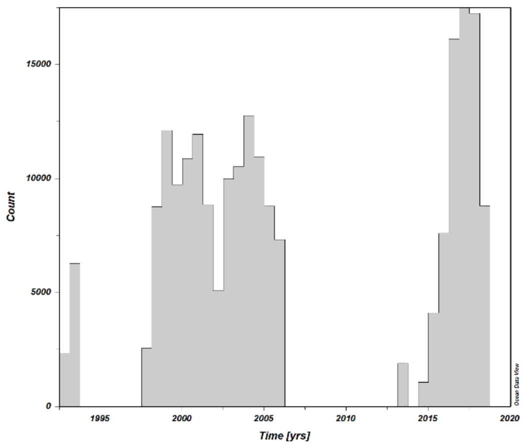


(a)

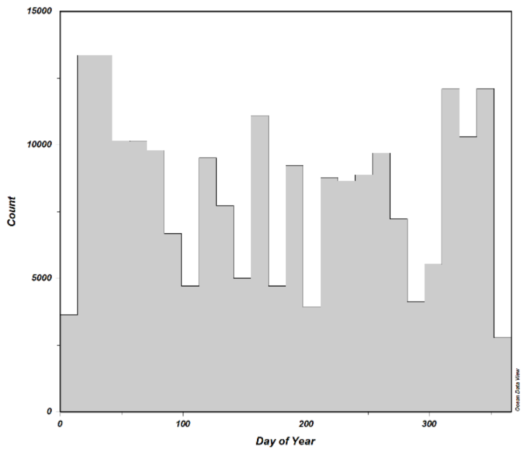


(b)

Figure 3. Map of ferrybox data (a) and data density map (b) showing where most measurements have been sampled.



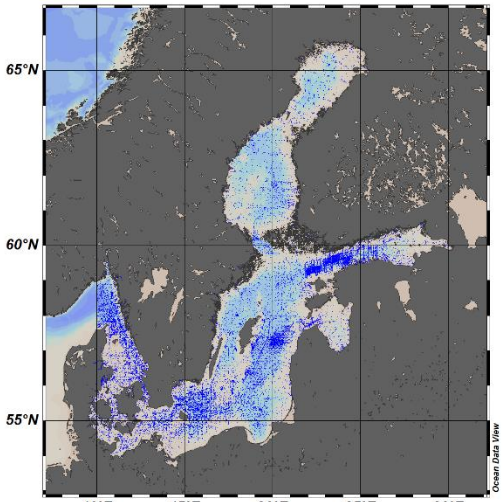
(a)



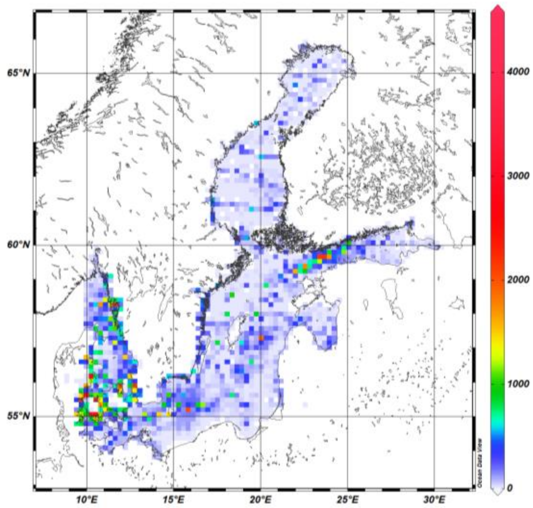
(b)

Figure 4. Temporal distribution (a) and seasonal distribution over the year (b) for the ferrybox data.

Profile data

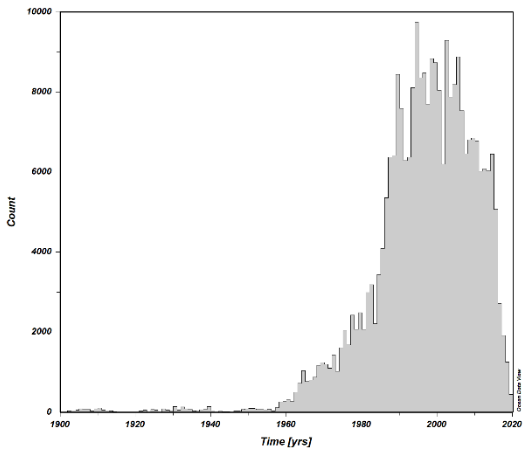


(a)

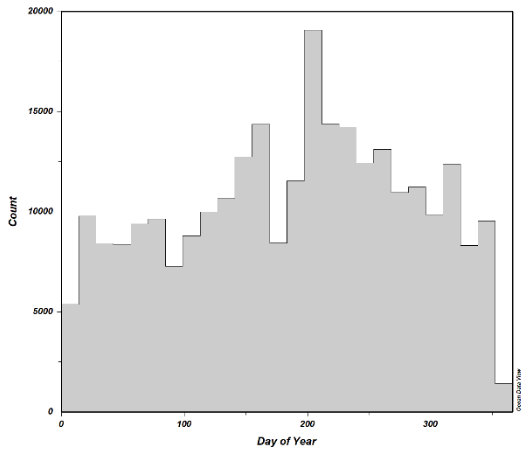


(b)

Figure 5. (a) Map with all profile data; (b) data density map showing where most measurements have been sampled.



(a)



(b)

Figure 6. Temporal distribution (a) and seasonal distribution over the year (b) for the profile data.

2. Quality Control Procedure

Salinity field in the Baltic Sea has a large geographical variability, from 0 in the north up to 36 in the southwest, Figure 7(a). To handle this, and make quality control analysis more efficient, salinity data were divided into sub-regions, Figure 7(b), and visually inspected. Density was calculated and plotted to find unstable profiles. The same procedure was applied for all data, not considering a difference between quality flags 0 and 1; since it is well known that quality controlled data still can contain errors.

With the large seasonal variability of the temperature field (below 0°C in the surface during winter, and up to above 25°C during summer) it is hard to find suspicious data using range checks or by visually inspecting all data. A solution is to make subsets of data for a season or a month at the time.

Checking bottom water is easier because it is not affected by the seasonal temperature cycle as much and it also has more stable salinity concentrations, especially in the Baltic Proper.

The quality control work followed the best practices that were defined during the project SeaDataNet 2, see Annex 1 - QC Best Practices.

The most powerful and useful quality control tool used was visual inspection of subsets of data in ODV to discover spikes, outliers, unstable profiles and stations on land.

This time only new and updated data since the previous historical TS product, version 1 [6], was included in the quality control process. After the quality control was finished, this data were merged together with the previous historical data product, making a complete data product and keeping quality information of older not updated data. Updated newer data replaced older data in the merging process.

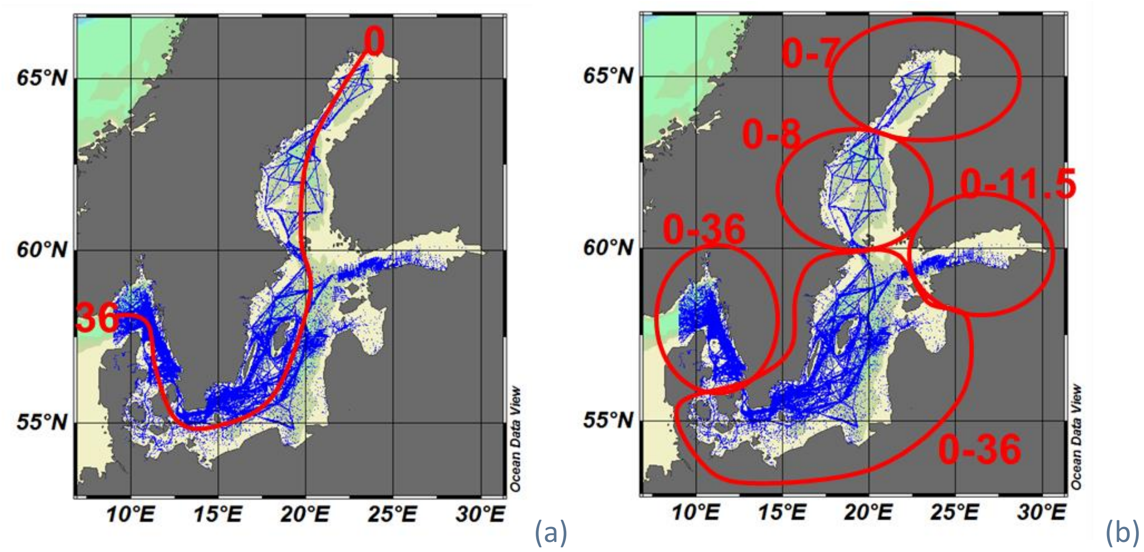


Figure 7. Salinity variability in The Baltic Sea (a), sub-regions used to make the quality control more efficient (b).

3. Quality assessment results

All new and updated data since November 2017 have been quality controlled according to chapter 2 Quality Control Procedure. 39,264 salinity values (0.81% of the total) and 6,269 temperature values (0.14% of the total) were flagged as suspicious/bad, quality flag 3 and 4, Table 2.

The suspicious values consist mainly of spikes, outliers and unstable density profiles, but there are also a few other problems:

- 2,511 profiles lack both temperature and salinity data, probably wrongly having the P02 terms for these parameters in the metadata (CDI file) when there actually are no data present in the data file. There were also 206 profiles lacking data in the previous data release (V1), these 2,817 stations have all been removed from the dataset.

- Some suspicious duplicates and some data that seem to lack time information which make them look like potential duplicates. 31 confirmed duplicate stations have been removed.

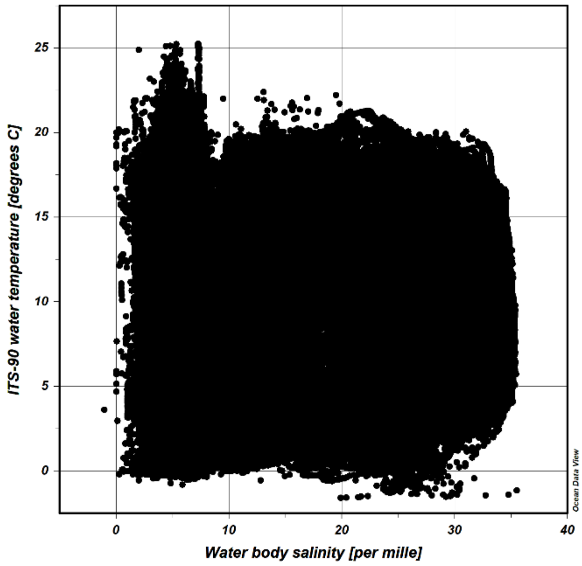
- Some data seem to have been split into two separate data files for unknown reasons.

Overall this dataset contained less obvious errors and extreme outliers. More time could be spent on finding suspicious data leading to an increase of data points flagged as such.

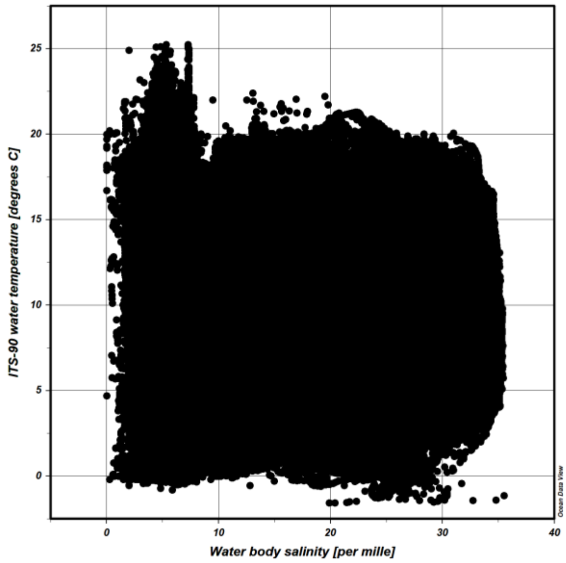
Data that was flagged as suspicious or bad in the previous quality control [1] which were not updated in the Seadatacloud portal until July 2019 have kept their quality flag also in this version 2 of the historical dataset.

Quality flag	0	1	2	3	4	5	6	7	8	A	Total
Salinity	127,553	2,020,321	2,636,065	43,469	384	1,120	19	0	1,719	0	4,830,650
% of total	2.6	41.8	54.6	0.9	<0.1	<0.1	<0.1	0	<0.1	0	
Temperature	157,266	1,801,252	2,636,406	13,212	329	26	0	0	1,705	0	4,610,196
% of total	3.4	39.1	57.2	0.3	<0.1	<0.1	0	0	<0.1	0	

Table 2. Number of values for each quality flag for the new and updated data subset that was quality controlled, 1900-2019.



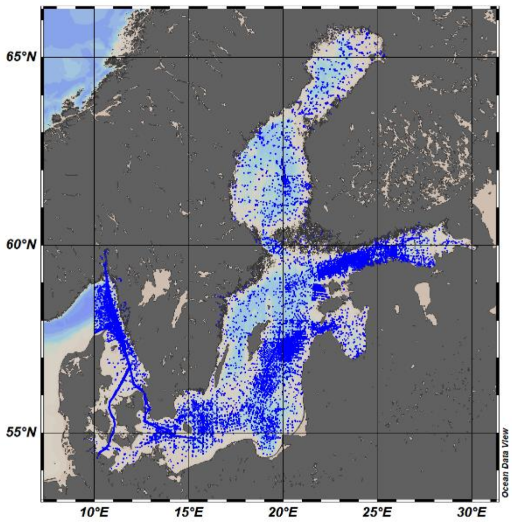
(a)



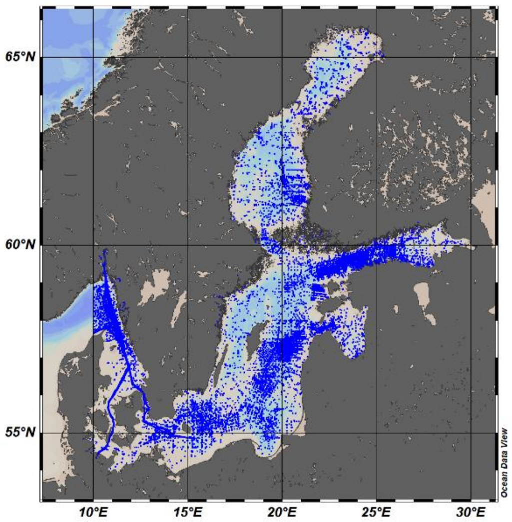
(b)

Figure 8. TS diagram showing data before (a) and after (b) quality control of the new and updated data.

The following figures are all after quality control where suspicious data have been removed if not otherwise clearly stated. Data are only from the new and updated data subset.



(a)



(b)

Figure 9. Spatial distribution of temperature (a) and salinity (b).

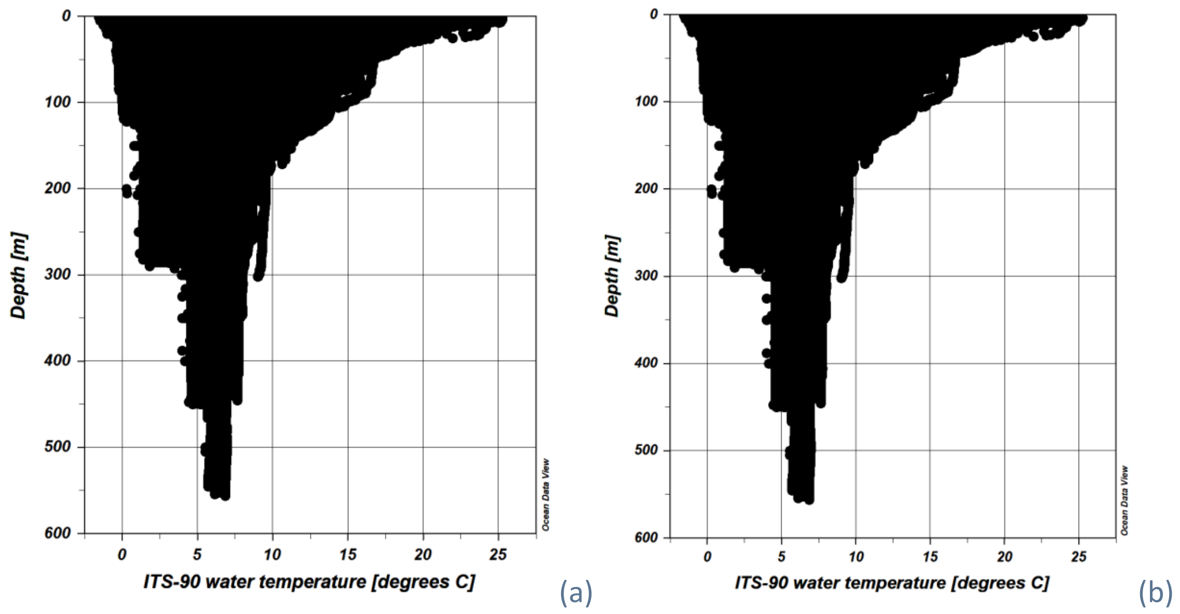


Figure 10. Temperature observations before QC (a) and after QC (b).

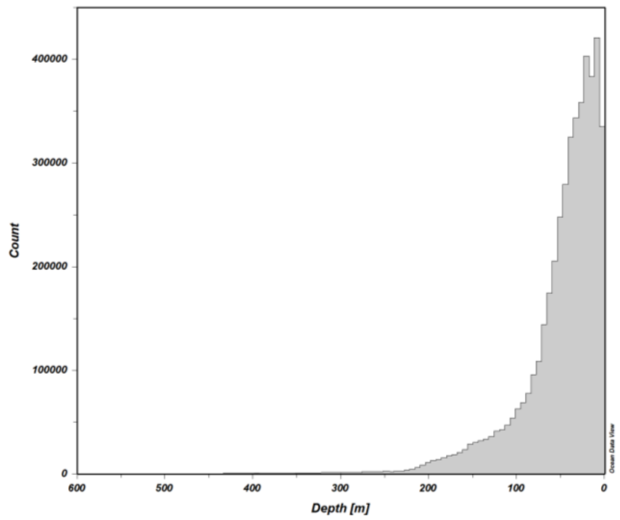


Figure 11. Temperature distribution over depth.

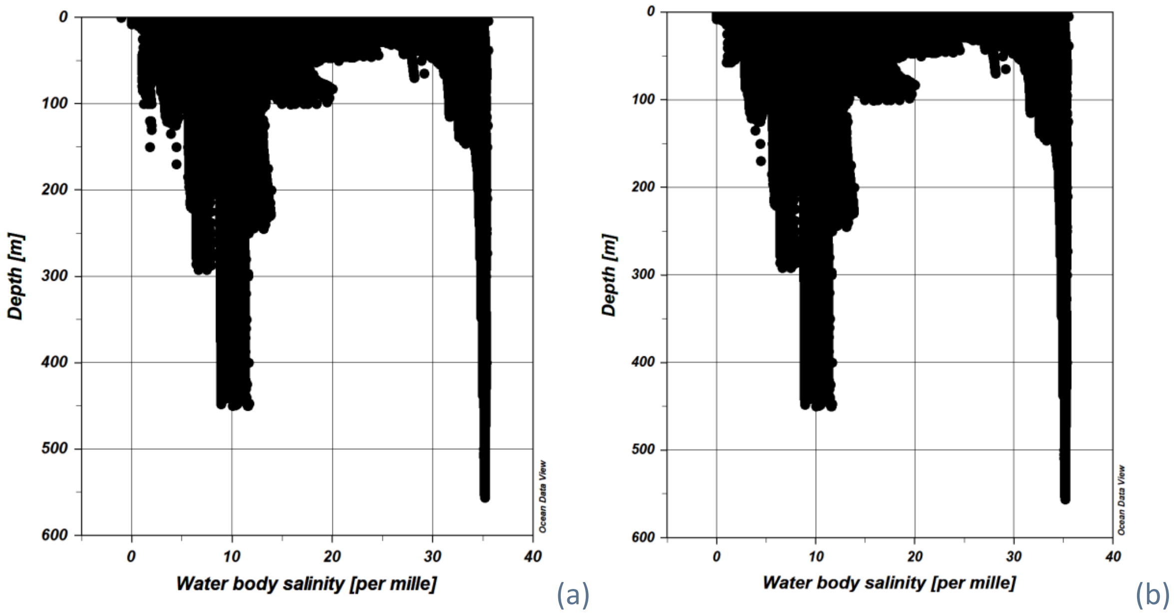


Figure 12. Salinity observations before QC (a) and after QC (b).

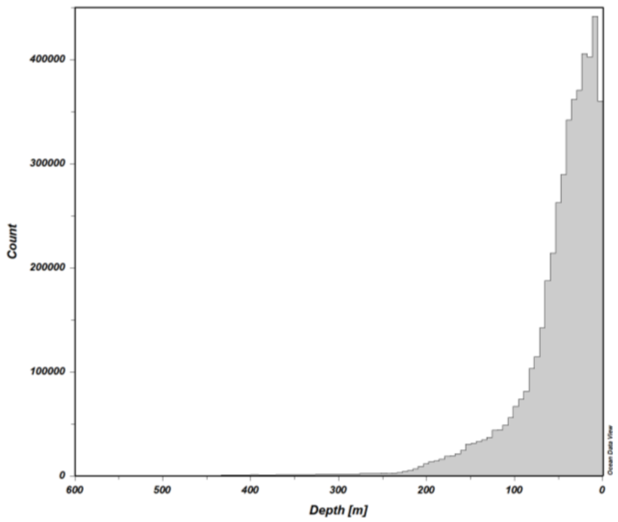


Figure 13. Salinity distribution over depth.

4. Technical Specifications

Product Format

Ocean Data View (ODV) collection.

Data Policy

No limitations on usage, however for data access registration is required at http://www.marine-id.org.

Product Usability

The collection contains unique large array of quality controlled data on temperature and salinity of Baltic Sea waters for the period 1900 – 2019 that can be used to support general oceanographic studies, such as investigation of variability of physical properties or long time trends, as well as applications such as circulation models.

Whenever SDC_BAL_DATA_TS_V2 product is used, this PIDoc should be cited in any publication. We also ask users to remember that hard-working scientists made these measurements, often under severe conditions. Further, the data providers normally possess insight on the quality and context of the data not always shared with the SeaDataCloud team. Hence, inviting data providers and product leaders to collaborate in scientific investigations that depend on their data and data products is considered good and fair practice. Importantly, this will promote further sharing of data and will be beneficial to science.

Changes since previous version

The previous version of the product was released earlier in the SeaDataCloud project and it is available in SEXTANT product catalogue [5] under the name "Baltic Sea - Temperature and salinity Historical Data collection SeaDataCloud V1" (SDC_V1) [6]. Compared to the SDC_V1 collection there is an increase of data in the current collection (SDC_V2), Table 3. The spatial domain is the same but the time coverage extended from 1900 – 2017 to 1900 – 2019.

	SDC_V1	% of total	SDC_V2	% of total	% increase
Total	14,038,820		14,753,042		5.1
Temperature	13,683,395	97.5	14,384,851	97.5	5.1
Salinity	13,709,759	97.8	14,425,429	97.8	5.2
Both T and S	13,473,081	96.0	14,174,006	96.1	5.2

Table 3. Number of values for Temperature, Salinity and TS couples for the Baltic Sea. All data have been considered in order to compare SDC_V1 and SDC_V2 datasets. 55% of the increase in data originates from 2017-2019 compared to 45% from the period 1900-2016.

Annex 1 - QC Best Practices

- The basic QC analysis steps applied during SeaDataNet2 Project using ODV were:

- **Data coverage**;
- **Data distribution maps** per Temperature, Salinity and TS couples;
- **Data density maps** (domain binning);
- **Time coverage and time distribution** → histograms with annual, seasonal and monthly data distribution;
- TS scatter plots of the entire dataset;
- Scatter plot of observations with QF=1 (good) and QF=2 (probably good);
- Scatter plot observations with QF=0 (no quality check);
- **Gross range check** to detect observations with temperature and salinity out of reasonable values;
- Visual control of scatter-plots to identify wrong profiles (outliers);
- Identification of stations falling on land;
- Identification of stations having unreal depth;
- Identification of wrong or missing data;
- Stability check on density

Additional checks are advisable per specific:

- **areas** with similar hydrodynamic characteristics;
- **layers** (surface, intermediate, bottom);
- **time periods** (decades, or specific periods i.e. Eastern Mediterranean Transient, Western Mediterranean Transition, Norther Ionian Reversal);
- **Instrument type** → consistency issue of historical data;

Duplicate Check is another important step when performing SDC data integration with external data sources for climatologies and new data products generation.

Table 4 lists the Quality Flags (QF) adopted by SeaDataNet and their definition [7]. QF assigned by the data centers are modified by the regional products' leaders when/if a data anomaly is detected. The data anomaly is reported to the data center asking for correction in the central CDI.

Key	Entry Term	Abbreviated term	Term definition
0	no quality control	none	No quality control procedures have been applied to the data value. This is the initial status for all data values entering the working archive.
1	good value	good	Good quality data value that is not part of any identified malfunction and has been verified as consistent with real phenomena during the quality control process.
2	probably good value	probably_good	Data value that is probably consistent with real phenomena but this is unconfirmed or data value forming part of a malfunction that is considered too small to affect the overall quality of the data object of which it is a part.
3	probably bad value	probably_bad	Data value recognised as unusual during quality control that forms part of a feature that is probably inconsistent with real phenomena.
4	bad value	bad	An obviously erroneous data value.
5	changed value	changed	Data value adjusted during quality control. Best practice strongly recommends that the value before the change be preserved in the data or its accompanying metadata.
6	value below detection	BD	The level of the measured phenomenon was too small to be quantified by the technique employed to measure it. The accompanying value is the detection limit for the technique or zero if that value is unknown.
7	value in excess	excess	The level of the measured phenomenon was too large to be quantified by the technique employed to measure it. The accompanying value is the measurement limit for the technique.
8	interpolated value	interpolated	This value has been derived by interpolation from other values in the data object.
9	missing value	missing	The data value is missing. Any accompanying value will be a magic number representing absent data.
A	value phenomenon uncertain	ID_uncertain	There is uncertainty in the description of the measured phenomenon associated with the value such as chemical species or biological entity.

Table 4. List of SeaDataNet Quality Flags (QF) used to describe the data value; no changes are made to the original data values.

Annex 2 – Data providers

EDMO	Originator/Collating centre
43	British Oceanographic Data Centre
100	Baltic Sea Research Institute Warnemuende (IOW)
117	Federal Maritime and Hydrographic Agency, Dept. Oceanography
193	Institute of Meteorology and Water Management National Research Institute, Maritime Branch in Gdynia (IMWM MB)
195	Institute of Oceanology, Polish Academy of Sciences (IO PAS)
396	Marine Institute
486	IFREMER / IDM / SISMER - Scientific Information Systems for the SEA
540	Shom
545	Swedish Meteorological and Hydrological Institute
618	Stockholm Marine Research Centre, SMF
622	Umea Marine Sciences Centre, UMF
628	Finnish Institute of Marine Research (FIMR)
630	NIOZ Royal Netherlands Institute for Sea Research
680	Department of Marine Research of the Environmental Protection Agency
681	All-Russia Research Institute of Hydrometeorological Information - World Data Centre (RIHMI-WDC) National Oceanographic Data Centre (NODC)
682	Atlantic Scientific Research Institute for Marine Fishery and Oceanography
684	Arctic and Antarctic Research Institute, Roshydromet (Saint-Petersburg)
685	P.P.Shirshov Institute of Oceanology, RAS
698	Latvian Institute of Aquatic Ecology
713	Marine Systems Institute at Tallinn University of Technology
729	Aarhus University, Department of Bioscience, Marine Ecology Roskilde
867	Baltic Research Institute of Fishery of Latvia (BaltNIIRH)
890	Krondshtadt Hydrometeorological Observatory, North-West HMS
894	Lithuanian Hydrometeorological Service under the Ministry of Environment
911	North-Western Regional Administration of Hydrometeorology of Roshydromet
924	Russian State Hydrometeorological University, St-Petersburg
931	Odessa Branch of SOI (State Oceanographic Institute)
989	Federal Research Centre for Fisheries (Cuxhaven)
991	Federal Research Centre for Fisheries Institute for Baltic Sea Fishery
993	State Agency for Environment, Nature and Geology, Mecklenburg-Vorpommern
1051	UNKNOWN
1104	Finnish Environment Institute
1181	State Agency for Nature and Environment of Schleswig Holstein (LANU)
1334	University of Gothenburg, Faculty of Science
1351	Institute of Marine Research
1353	Swedish Environmental Protection Agency

1354	The Swedish Board of Fisheries
1368	Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research
1417	Norwegian Institute of Water Research (NIVA)
1507	Carl von Ossietzky University of Oldenburg - Institute for Chemistry and Biology of the Marine Environment (ICBM)
1570	Thünen-Institute of Sea Fisheries (TI-SF)
1571	Thünen-Institute of Fisheries Ecology (TI-FI)
1574	Thünen-Institute of Baltic Sea Fisheries (TI-OF)
1575	Federal Research Institute for Rural Areas, Forestry and Fisheries (VTI)
1576	Institute of Fisheries Ecology - Cuxhaven (VTI-CUX)
1725	Finnish Meteorological Institute
1850	Federal Maritime and Hydrographic Agency
2108	Environmental Protection Agency (EPA)
2195	DTU Aqua – National Institute of Aquatic Resources, Technical University of Denmark
2424	National Oceanography Centre, Liverpool
2537	State Office for Agriculture, Environment and Rural Areas of Schleswig Holstein (LLUR)
2561	National Marine Facilities Sea Systems
2947	GEOMAR Helmholtz Centre for Ocean Research Kiel
3512	The Swedish Agency for Marine and Water Management
4614	ERIC Euro-Argo
4768	County Administrative Board of Vastra Gotaland
4859	Oresund Water Conservation Association
4916	Holmen Papper Ab

References

1. Bäck Örjan, Simoncelli Simona, Coatanoan Christine (2018). SeaDataCloud Temperature and Salinity Historical Data Collection for the Baltic Sea (Version 1). Product Information Document (PIDoc). *https://doi.org/10.13155/57039*
2. SeaDataCloud project (2016-2020), grant agreement 730960, EU H2020 programme, *www.seadatanet.org/About-us/SeaDataCloud*
3. Seadatanet 2 project (2011-2015), grant agreement 283607, EU Seventh Framework Programme, *https://www.seadatanet.org/About-us/SeaDataNet-2*
4. Schlitzer, R., Ocean Data View, odv.awi.de, 2019
5. Sextant catalogue, Seadatanet products *http://sextant.ifremer.fr/en/web/seadatanet*
6. Digital Object Identifier (DOI, *https://www.doi.org/*) for the SeaDataCloud historical data collection V1, *https://doi.org/10.12770/afcaabb7-0618-4bf6-9e8c-4cdd3da51fab*
7. Controlled vocabulary for Seadatanet Measurand Qualifier Flags, L20, *http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=L20*

List of acronyms

Acronym	Definition
ARC	Arctic ocean
BAL	Baltic Sea
BLS	Black Sea
CDI	Common Data Index
CLIM	Climatology
CMEMS	Copernicus Marine Environment Monitoring Service
DATA	Aggregated Dataset
DIVA	Data-Interpolating Variational Analysis (software)
DOI	Digital Object Identifier
EC	European Commission
EDMO	European Directory of Marine Organisations (SeaDataNet catalogue)
GLO	GLobal Ocean
IOC	Intergovernmental Oceanographic Commission
IODE	International Oceanographic Data and Information Exchange (IOC)
MED	Mediterranean Sea
NAT	North Atlantic Ocean
NWS	North West Shelf
ODV	Ocean Data View Software
QC	Quality Checks
QF	Quality Flags
SDC	SeaDataCloud
SDN	SeaDataNet
TS	Temperature and Salinity
WOA	World Ocean Atlas
WP	Work Package