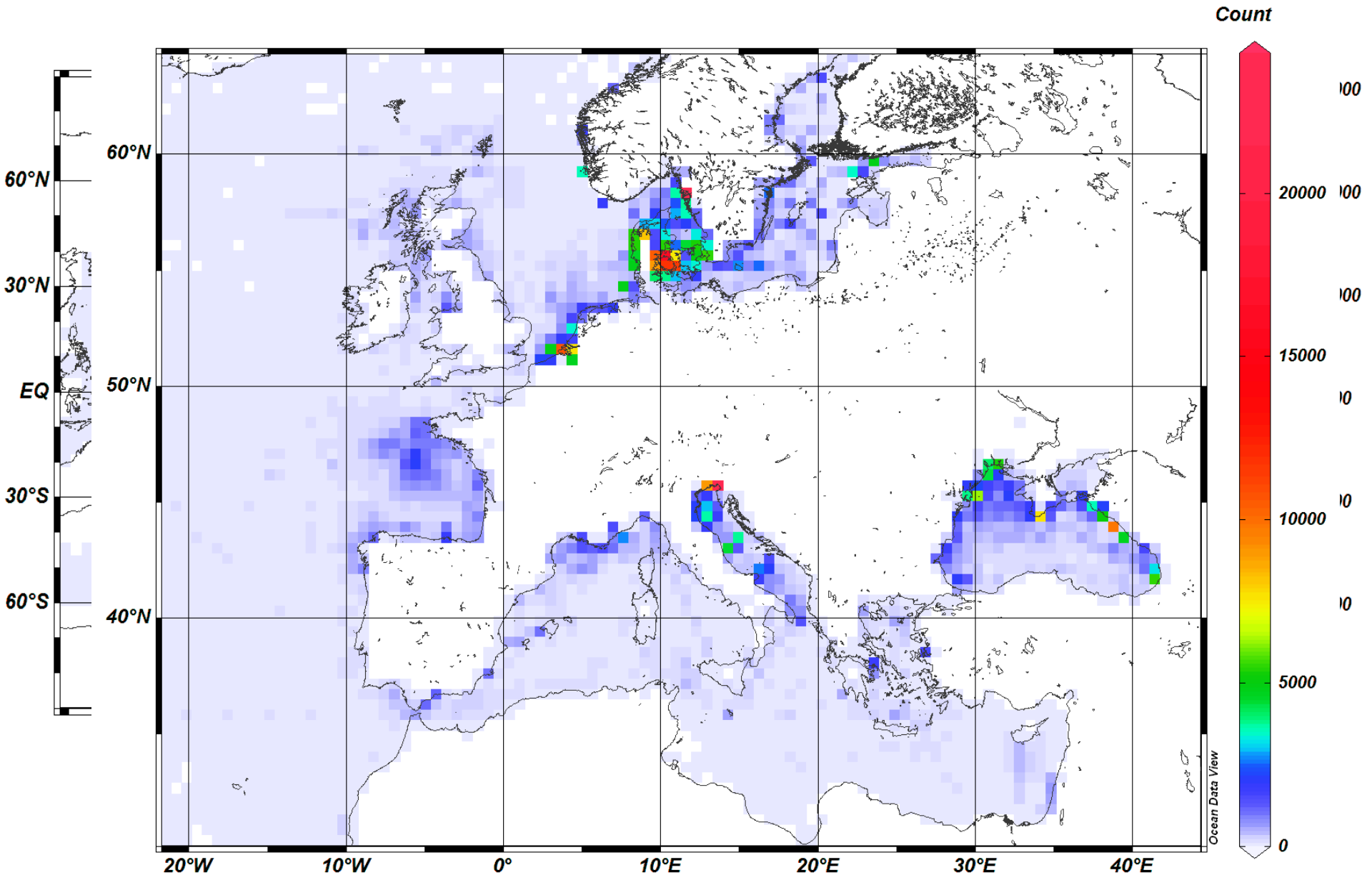
Duplicate Station Detection with ODV

Reiner Schlitzer, Alfred Wegener Institute, Bremerhaven, Germany

CDI catalogue as of July 2012

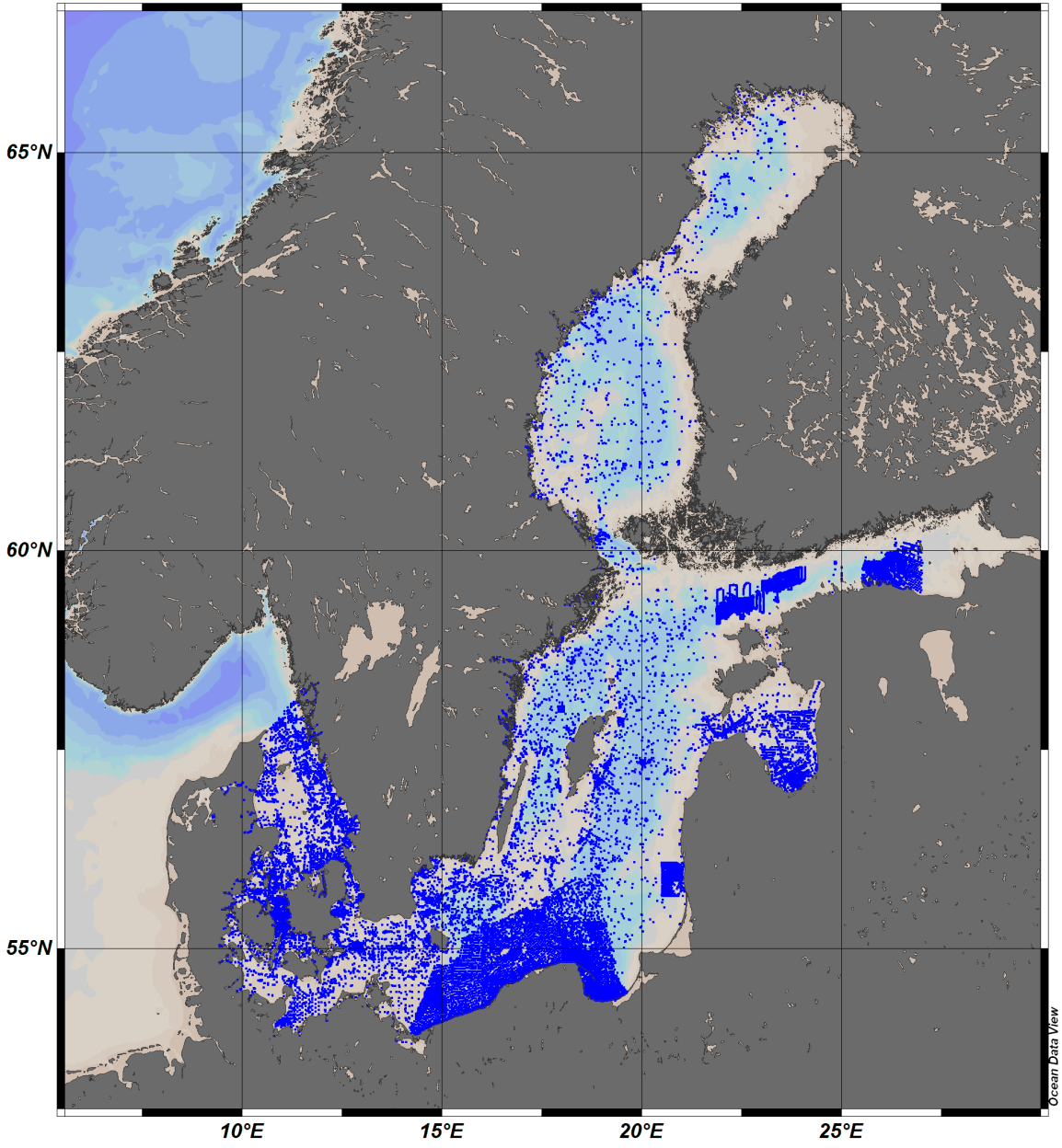


Duplicates Criteria:

Two stations are considered duplicates if

- **Position longitude and latitude agree within 0.001° (ca. 100 m)**
- **Station date/time agree within 3 minutes (to avoid time series data that sometimes have 5 minute cycle)**
- **CDI instrument type aggrees (to avoid CTD and bottle data as duplicates)**

These are very strict criteria. Matching stations have high probability of being true duplicates!

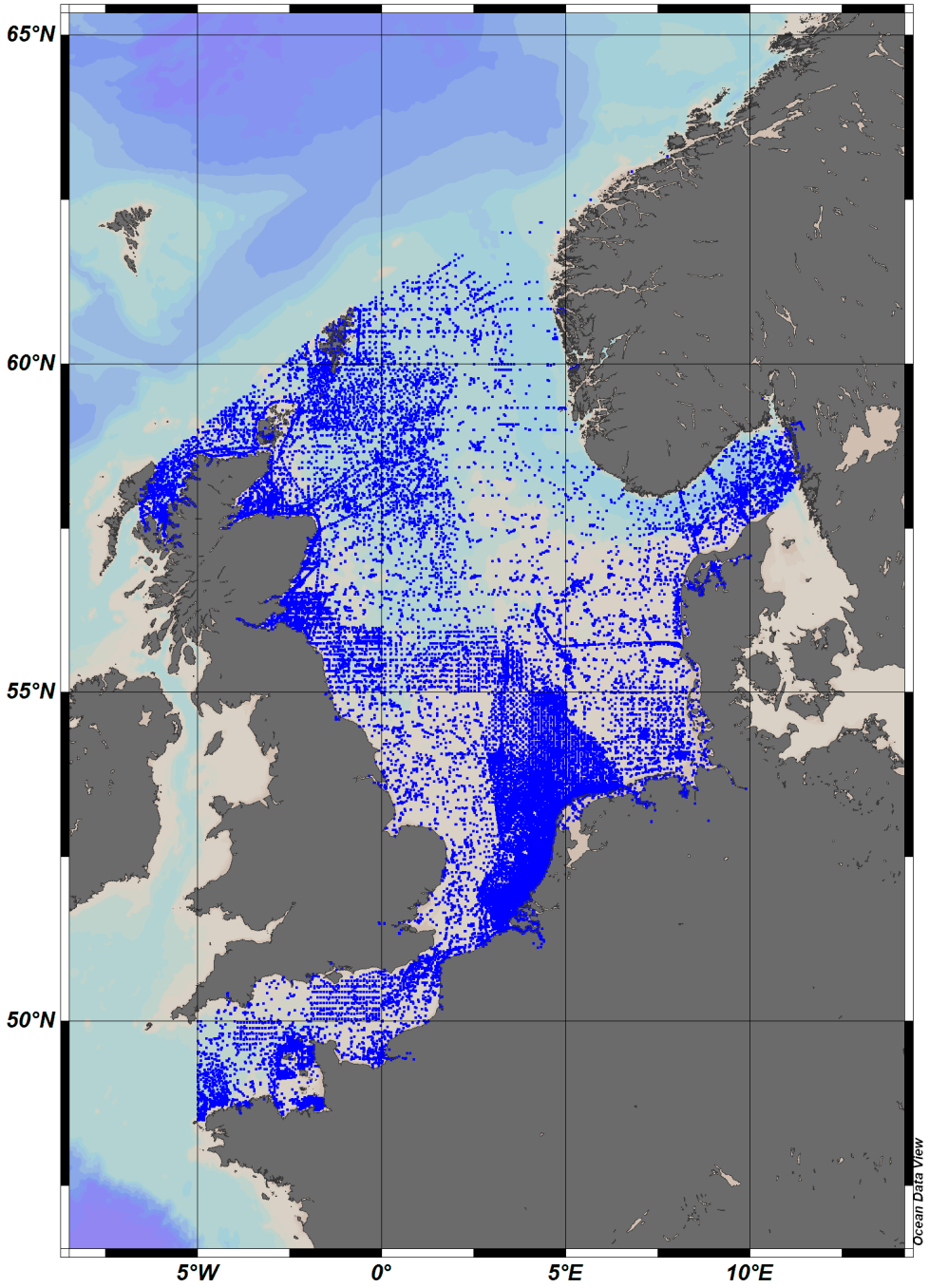


Baltic:

242,276 Stations

7,203 Duplicates

3 %



North Sea:

158,612 Stations

18,333 Duplicates

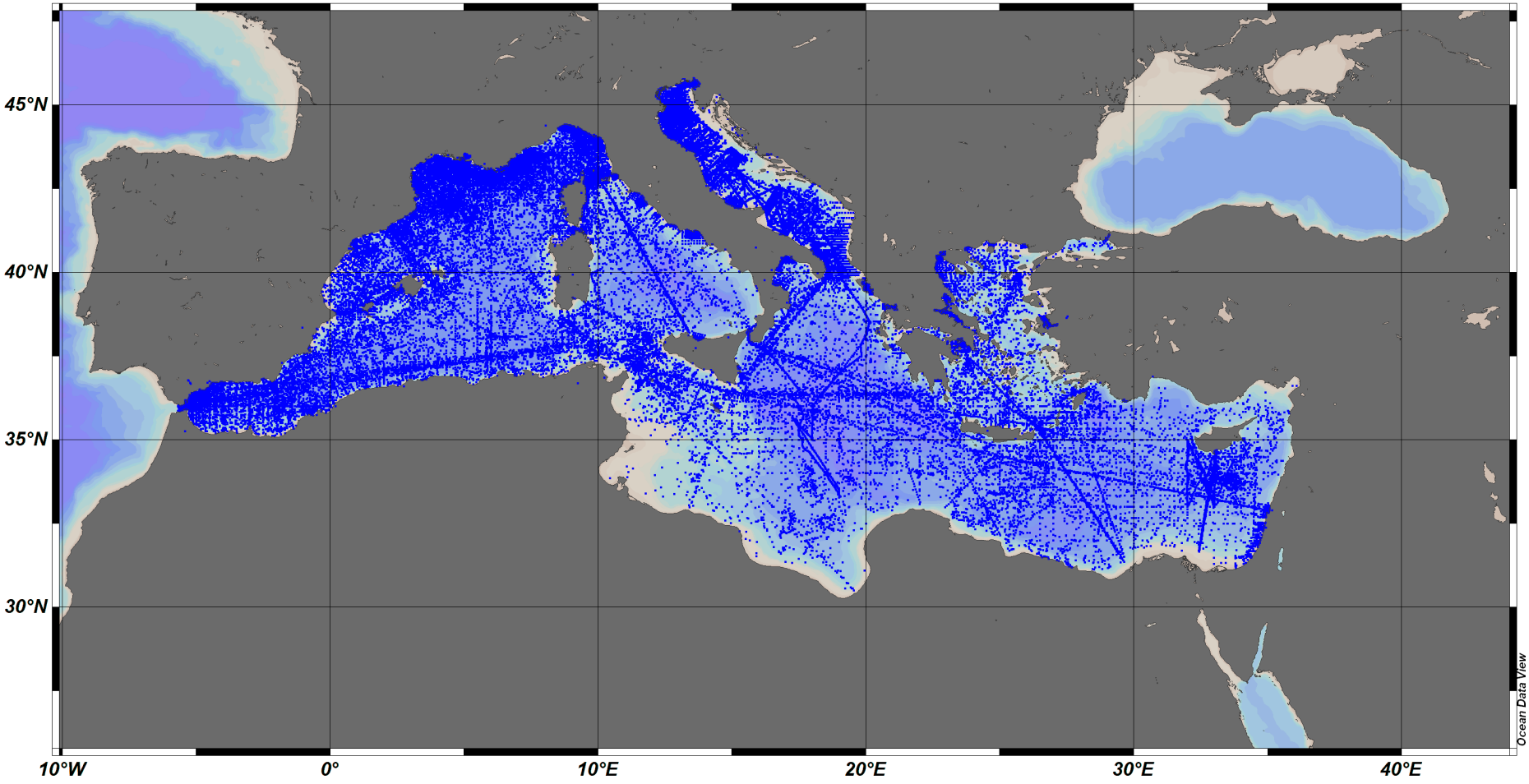
11.5%

Mediterranian:

17.5%

177,678 Stations

31,146 Duplicates

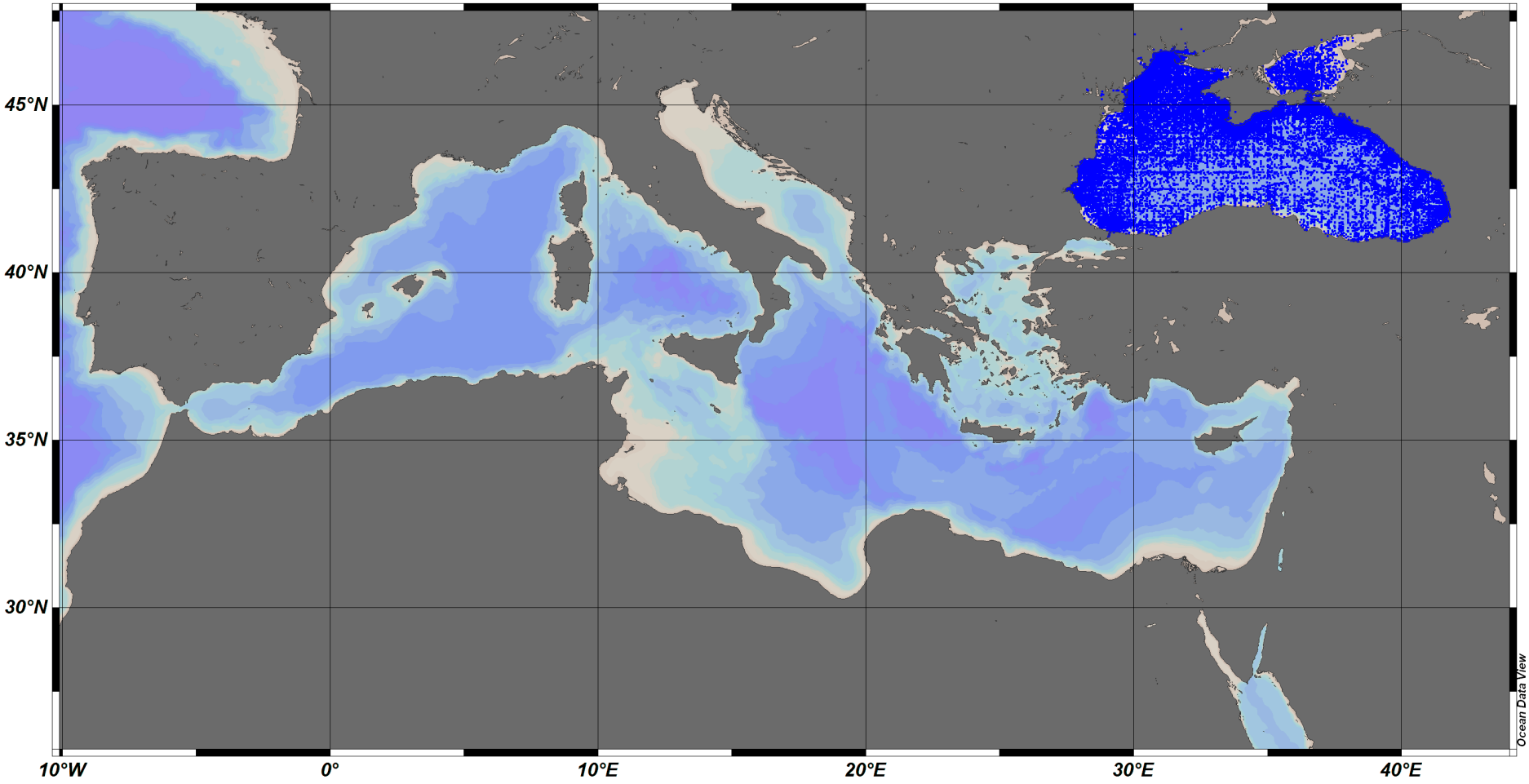


Black Sea:

21.7 %

159,788 Stations

34,789 Duplicates



Examples

cdi_identifier	author_EDMO	originator_EDI	custodium_EI	parameters_P021	instrument_L	platform_L06	min_instrume	max_instrum	distributor_EDMO
77AR1981_00053_H09	545	545	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,SLCA,TDNT,TDPX,TEMP	30	31	-9999.00	-9999.00	545
1022097	729	729	729	AMON,DOXY,NTOT,NTRA,NTRI,PHOS,PSAL,SLCA,TPHS	30	31			729
77AR1981_00053_H10	545	545	545	PSAL,TEMP	130	31	-9999.00	-9999.00	545
22097	729	729	729	TEMP	130	31			729
77AR1981_00216_H09	545	545	545	CPWC,PHOS,TDNT,TDPX,TEMP	30	31	-9999.00	-9999.00	545
77AR1981_00217_H09	545	545	545	CPWC,PHOS,TDNT,TDPX,TEMP	30	31	-9999.00	-9999.00	545
77AR1982_00015_H09	545	545	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,SLCA,SPWC,TDNT,TDPX,TEMP	30	31	-9999.00	-9999.00	545
1022112	729	729	729	AMON,DOXY,NTOT,NTRA,NTRI,PHOS,PSAL,SLCA,TPHS	30	31			729
77AR1984_00365_H09	545	545	545	ALKY,AMON,CPWC,DOXY,NTRA,NTRI,PHOS,PSAL,SLCA,TDNT,TDPX,TEMP	30	31	-9999.00	-9999.00	545
1022147	729	729	729	AMON,DOXY,NTOT,NTRA,NTRI,PHOS,PSAL,SLCA,TPHS	30	31			729
77AR1984_00365_H10	545	545	545	PSAL,TEMP	130	31	-9999.00	-9999.00	545
22147	729	729	729	TEMP	130	31			729
77AR1984_00518_H09	545	545	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,SLCA,SPWC,TDNT,TDPX,TEMP	30	31	-9999.00	-9999.00	545
1022155	729	729	729	AMON,DOXY,NTOT,NTRA,NTRI,PHOS,PSAL,SLCA,TPHS	30	31			729

Two data centers providing different subsets of the data.

Examples

cdi_identifier	author_EDMO	originator_EDI	custodium_EI	parameters_P021	instrument_L	platform_L06	min_instrume	max_instrum	distrib
77K21989_00004_H09	545	622	545	ALKY,AMON,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30		-9999.00	-9999.00	545
77VA1989_00004_H09	545	622	545	ALKY,AMON,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30	37	-9999.00	-9999.00	545
77K21989_00005_H09	545	622	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30		-9999.00	-9999.00	545
77VA1989_00005_H09	545	622	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30	37	-9999.00	-9999.00	545
77K21989_00006_H09	545	622	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30		-9999.00	-9999.00	545
77VA1989_00006_H09	545	622	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30	37	-9999.00	-9999.00	545
77K21989_00007_H09	545	622	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30		-9999.00	-9999.00	545
77VA1989_00007_H09	545	622	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30	37	-9999.00	-9999.00	545
77K21989_00008_H09	545	622	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30		-9999.00	-9999.00	545
77VA1989_00008_H09	545	622	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30	37	-9999.00	-9999.00	545
77K21989_00009_H09	545	622	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30		-9999.00	-9999.00	545
77VA1989_00009_H09	545	622	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30	37	-9999.00	-9999.00	545
77K21989_00010_H09	545	622	545	ALKY,AMON,DOXY,PHOS,PSAL,TDNT,TDPX,TEMP	30		-9999.00	-9999.00	545
77VA1989_00010_H09	545	622	545	ALKY,AMON,DOXY,PHOS,PSAL,TDNT,TDPX,TEMP	30	37	-9999.00	-9999.00	545
77K21989_00011_H09	545	622	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30		-9999.00	-9999.00	545
77VA1989_00011_H09	545	622	545	ALKY,AMON,DOXY,NTRA,NTRI,PHOS,PSAL,TDNT,TDPX,TEMP	30	37	-9999.00	-9999.00	545

Identical copies?

Examples

cdi_identifier	author_EDM(	originator_ED	custodium_E	parameters_P021	instrument_L	platform_L06	min_instrum	max_instrum	distribu
269782	1543	1543	1543	LITH	51	31	0.00	9.60	1543
269783	1543	1543	1543	LITH	51	31	0.00	9.00	1543
269784	1543	1543	1543	LITH	51	31	0.00	7.60	1543
269785	1543	1543	1543	LITH	51	31	0.00	9.00	1543
269786	1543	1543	1543	LITH	51	31	0.00	8.50	1543
269787	1543	1543	1543	LITH	51	31	0.00	9.30	1543
269788	1543	1543	1543	LITH	51	31	0.00	8.50	1543
269789	1543	1543	1543	LITH	51	31	0.00	12.80	1543
269790	1543	1543	1543	LITH	51	31	0.00	11.00	1543
269795	1543	1543	1543	LITH	51	31	0.00	22.00	1543
269796	1543	1543	1543	LITH	51	31	0.00	23.00	1543
269791	1543	1543	1543	LITH	51	31	0.00	16.00	1543
269792	1543	1543	1543	LITH	51	31	0.00	14.50	1543
269794	1543	1543	1543	LITH	51	31	0.00	27.00	1543
269798	1543	1543	1543	LITH	51	31	0.00	10.00	1543
269991	1543	1543	1543	LITH	51	31	0.00	10.00	1543
270040	1543	1543	1543	LITH	999	31	0.00	11.30	1543
270041	1543	1543	1543	LITH	999	31	0.00	14.60	1543
270043	1543	1543	1543	LITH	999	31	0.00	5.50	1543
270045	1543	1543	1543	LITH	999	31	0.00	12.50	1543

Different parts of sediment core?

Examples

cdi_identifier	author_EDM(	originator_ED	custodium_E	parameters_P021	platform_L06	min_instrume	max_instrum	distribut
Tua_2001_2001__001160_Tuapse-005_AMON_H10	685	723	723	AHGT,AMON	30			685
Tua_2001_2001__001164_Tuapse-005_NTRI_H10	685	723	723	AHGT,NTRI	30			685
Tua_2001_2001__001168_Tuapse-005_PHOS_H10	685	723	723	AHGT,PHOS	30			685
Tua_2001_2001__001172_Tuapse-005_SLCA_H10	685	723	723	AHGT,SLCA	30			685
Tua_2001_2001__001177_Tuapse-005_DOXY_H10	685	723	723	AHGT,DOXY	30			685
Tua_2001_2001__001134_Tuapse-005_TEMP_H10	685	723	723	AHGT,TEMP	30			685
Tua_2001_2001__001138_Tuapse-005_PSAL_H10	685	723	723	AHGT,PSAL	30			685
Tua_2001_2001__001140_Tuapse-005_ALKY_H10	685	723	723	AHGT,ALKY	30			685
Tua_2001_2001__001146_Tuapse-005_ALKY_H10	685	723	723	AHGT,ALKY	30			685
Tua_2001_2001__001148_Tuapse-005_DOXY_H10	685	723	723	AHGT,DOXY	30			685
Tua_2001_2001__001153_Tuapse-005_DOXY_H10	685	723	723	AHGT,DOXY	30			685
Tua_2001_2001__001156_Tuapse-005_WCOC_H10	685	723	723	AHGT,WCOC	30			685
Tua_2001_2001__001165_Tuapse-005_NTRI_H10	685	723	723	AHGT,NTRI	30			685
Tua_2001_2001__001166_Tuapse-005_PHOS_H10	685	723	723	AHGT,PHOS	30			685
Tua_2001_2001__001173_Tuapse-005_SLCA_H10	685	723	723	AHGT,SLCA	30			685
Tua_2001_2001__001174_Tuapse-005_DOXY_H10	685	723	723	AHGT,DOXY	30			685
Tua_2001_2001__001178_Tuapse-003_TEMP_H10	685	723	723	AHGT,TEMP	30			685
Tua_2001_2001__001184_Tuapse-003_PSAL_H10	685	723	723	AHGT,PSAL	30			685
Tua_2001_2001__001187_Tuapse-003_ALKY_H10	685	723	723	AHGT,ALKY	30			685
Tua_2001_2001__001191_Tuapse-003_ALKY_H10	685	723	723	AHGT,ALKY	30			685
Tua_2001_2001__001195_Tuapse-003_DOXY_H10	685	723	723	AHGT,DOXY	30			685
Tua_2001_2001__001200_Tuapse-003_DOXY_H10	685	723	723	AHGT,DOXY	30			685
Tua_2001_2001__001203_Tuapse-003_WCOC_H10	685	723	723	AHGT,WCOC	30			685
Tua_2001_2001__001207_Tuapse-003_AMON_H10	685	723	723	AHGT,AMON	30			685
Tua_2001_2001__001209_Tuapse-003_NTRI_H10	685	723	723	AHGT,NTRI	30			685
Tua_2001_2001__001213_Tuapse-003_PHOS_H10	685	723	723	AHGT,PHOS	30			685

duplicates_matching-instrument

One CDI per parameter. Shouldn‘t all parameters be merged into ONE station?

Examples

cdi_identifier	author_EDM(	originator_EC	custodium_E	parameters_P021	instrument_L	platform_L06	min_instrume	max_instrum	distrib
GN36200101001000_00170_269_H10	269	164	269	AHGT,PSAL,TEMP,DOXY,ZZZZ	130	31	1.00	3201.00	269
GN36200101001000_117B0_269_H10	269	164	269	PSAL,TEMP,DOXY,ZZZZ	130	31	1.00	201.00	269
GN36200101001000_137B0_269_H10	269	164	269	AHGT,PSAL,TEMP,DOXY,ZZZZ	130	31	1.00	76.00	269
GN36200101001000_00220_269_H10	269	164	269	PSAL,TEMP,DOXY,ZZZZ	130	31	1.00	2671.00	269
GN36200101001000_122B0_269_H10	269	164	269	AHGT,PSAL,TEMP,DOXY,ZZZZ	130	31	2.00	51.00	269
GN36200101001000_00A10_269_H10	269	164	269	AHGT,PSAL,TEMP,DOXY,ZZZZ	130	31	3.00	2842.00	269
GN36200101001000_00A20_269_H10	269	164	269	AHGT,PSAL,TEMP,DOXY,ZZZZ	130	31	1.00	1001.00	269
GN36200101001000_00B10_269_H10	269	164	269	AHGT,PSAL,TEMP,DOXY,ZZZZ	130	31	2.00	2800.00	269
GN36200101001000_00B20_269_H10	269	164	269	PSAL,TEMP,DOXY,ZZZZ	130	31	2.00	1001.00	269
GN36200101001000_00B30_269_H10	269	164	269	AHGT,PSAL,TEMP,DOXY,ZZZZ	130	31	2.00	151.00	269
SI2919220000300070_353_H09	353	353	353	AHGT,DOXY,PSAL,TEMP	30	31			353
SI2919220000300080_353_H09	353	353	353	AHGT,DOXY,PSAL,TEMP	30	31			353
SI2919220000300090_353_H09	353	353	353	AHGT,DOXY,PSAL,TEMP	30	31			353
SI2919220000300100_353_H09	353	353	353	AHGT,PSAL,TEMP	30	31			353
SI2919220000300110_353_H09	353	353	353	AHGT,DOXY,PSAL,TEMP	30	31			353
SI2919220000300120_353_H09	353	353	353	AHGT,PSAL,TEMP	30	31			353
SI2919220000300130_353_H09	353	353	353	AHGT,DOXY,PSAL,TEMP	30	31			353
SI2919220000300140_353_H09	353	353	353	AHGT,PSAL,TEMP	30	31			353
SI2919220000300150_353_H09	353	353	353	AHGT,DOXY,PSAL,TEMP	30	31			353
SI2919220000300160_353_H09	353	353	353	AHGT,PSAL,TEMP	30	31			353
SI2919220000300170_353_H09	353	353	353	AHGT,DOXY,PSAL,TEMP	30	31			353

Different depth coverage; different parameter sets.

Timeline:

Today: ***Potential duplicate lists, maps and Readme document ready to be put on SDN2 extranet.***

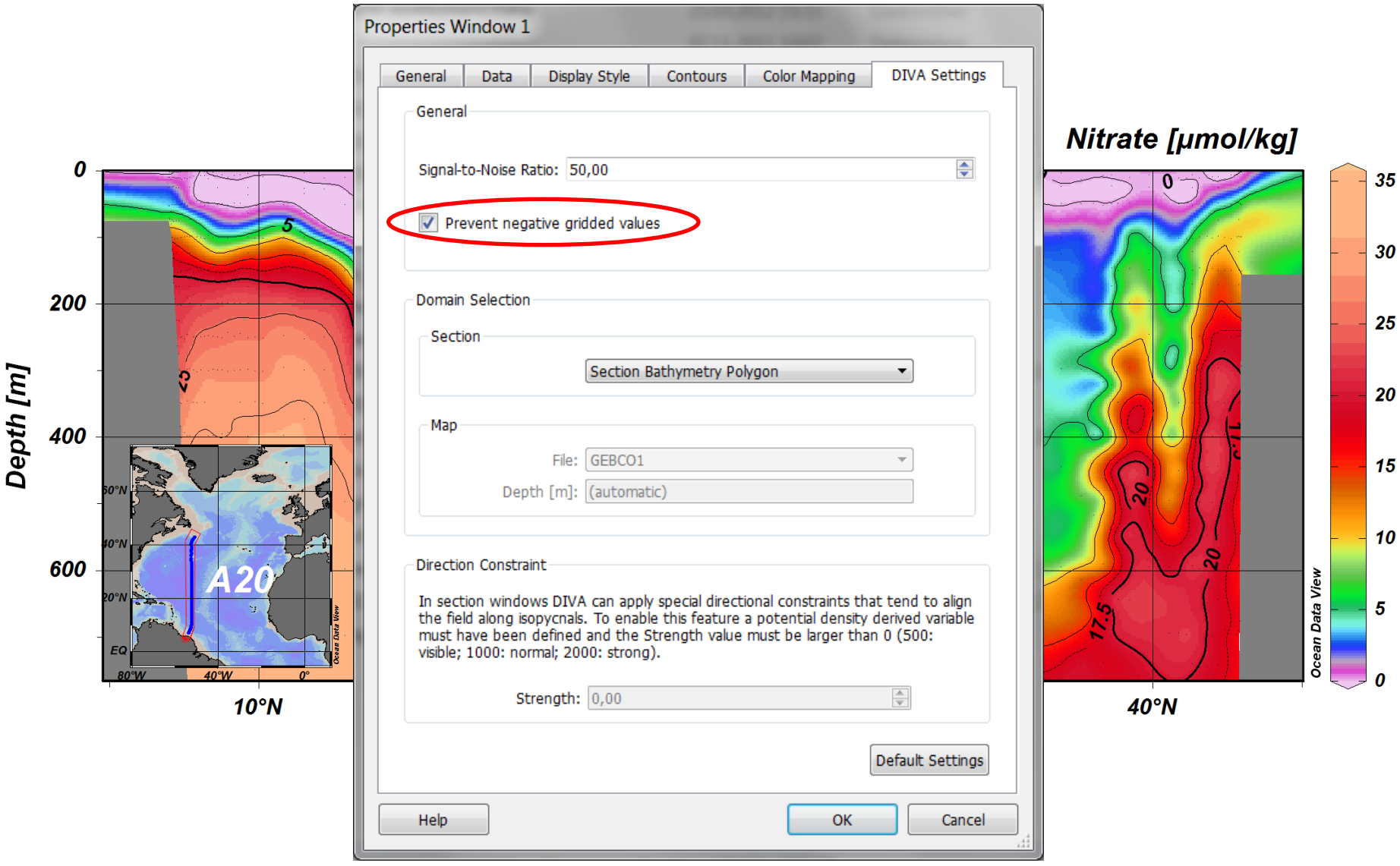
2012-10-01: ***Email from Sissy and Michele to all SDN data centres asking for duplicate station action. Coordination between data center and central catalog (MARIS) required.***

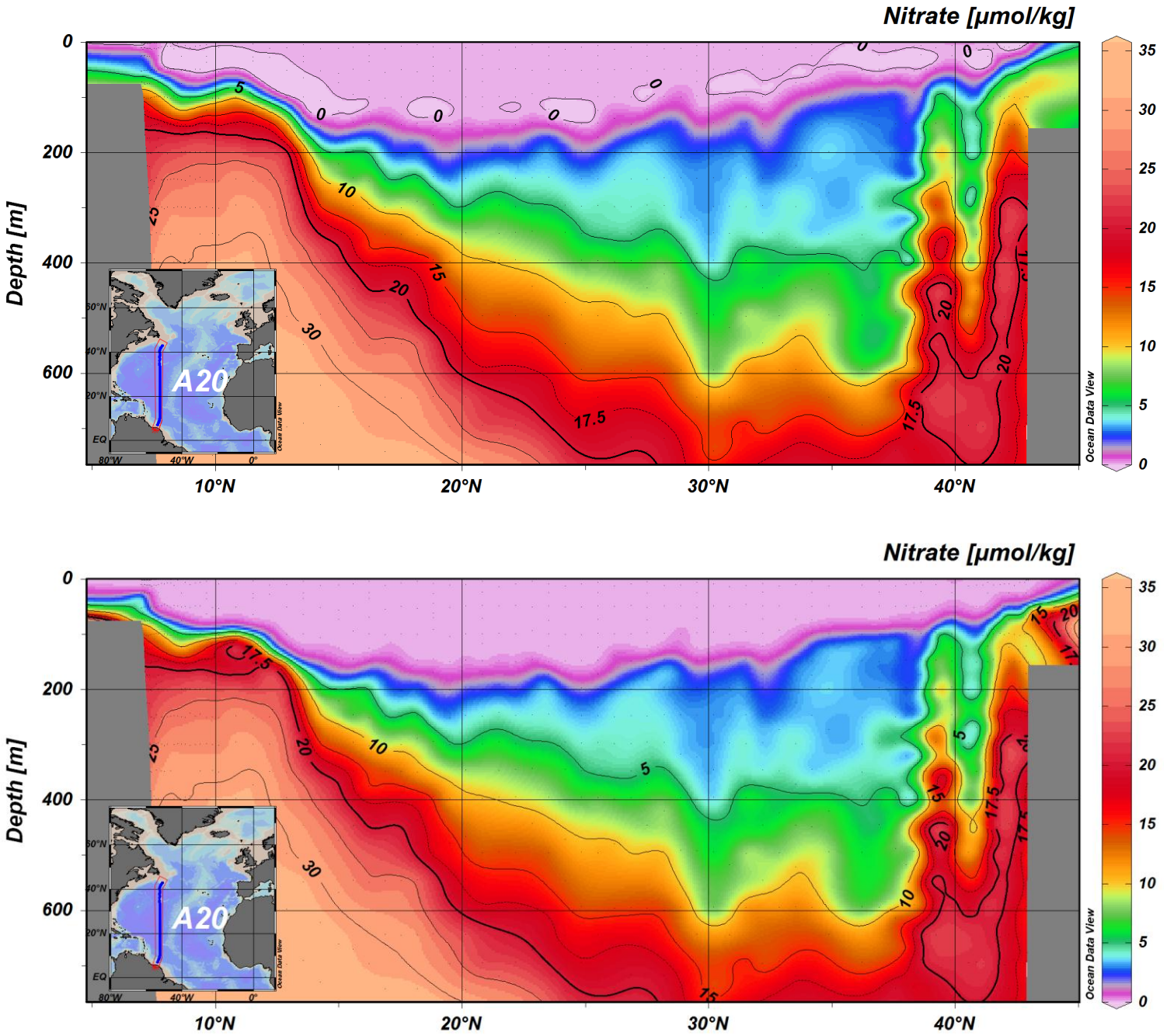
2012-12-01: ***Report from data centers to Sissy, Michele, and Dick on actions taken (e.g., changes to data files, coupling table, and CDI inventory).***

ODV Recent Developments

Reiner Schlitzer, Alfred Wegener Institute, Bremerhaven, Germany

1. DIVA positivity constraint



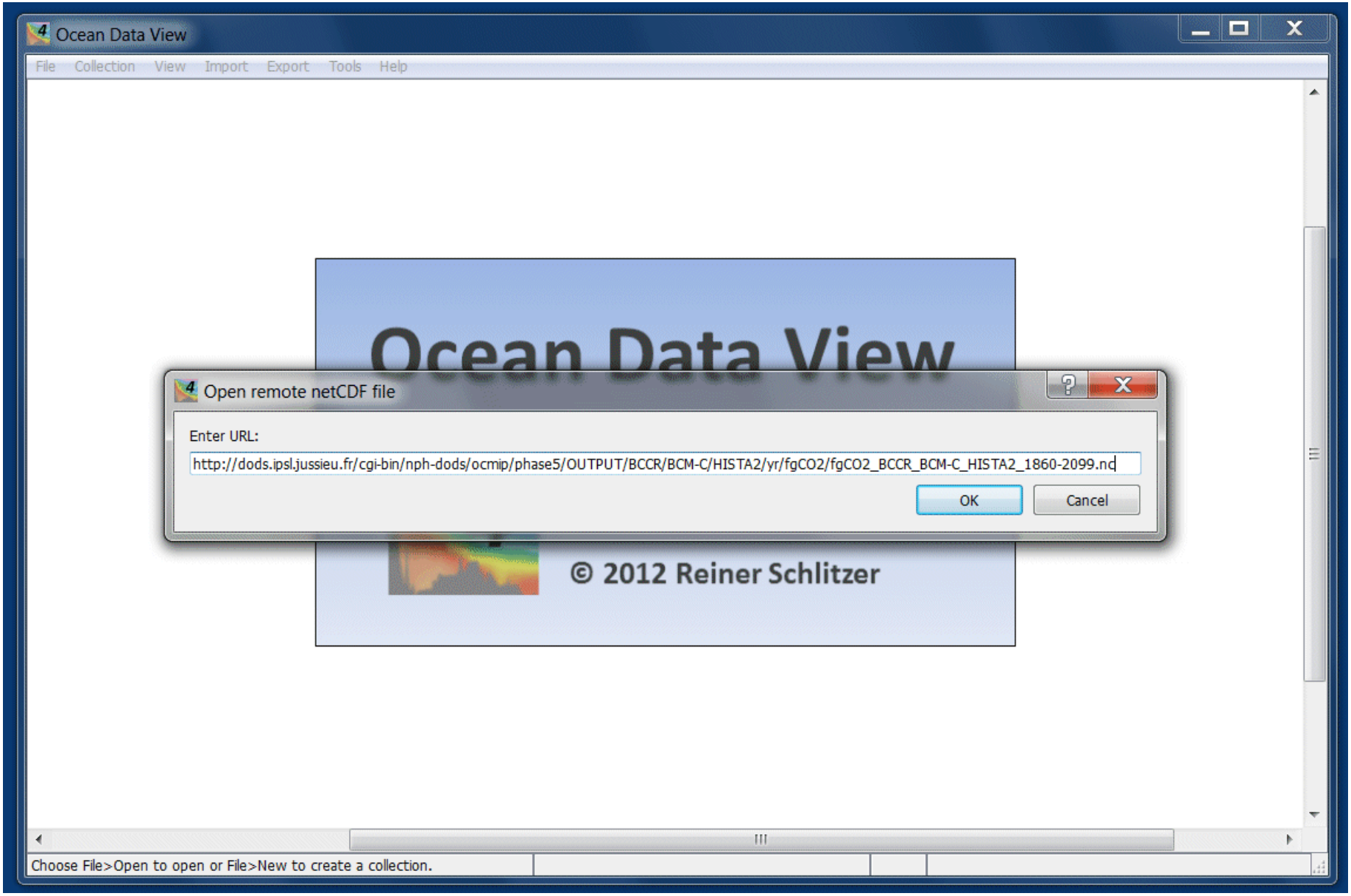

Nitrate [µmol/kg]
Depth [m]
0
200
400
600
10°N
20°N
30°N
40°N
A20
Ocean Data View
35
30
25
20
15
10
5
0

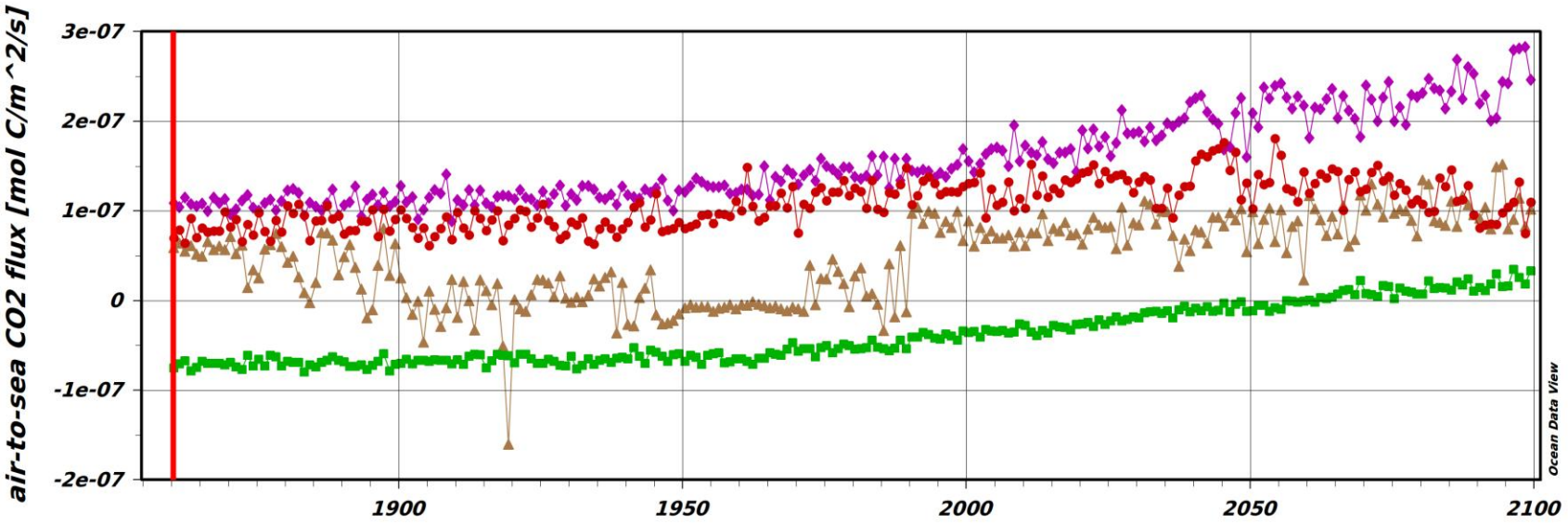
2. Switch to netCDF4 and remote access

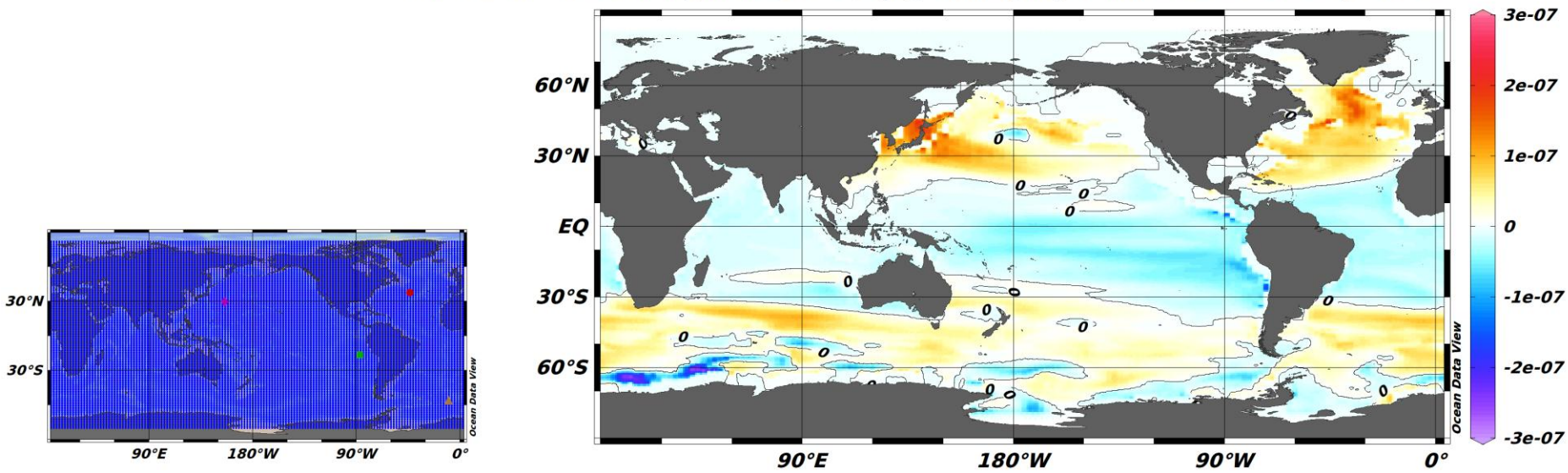
- **File size often very large – copying to local machine impractical**
- **Often only data from small domain or time interval needed**

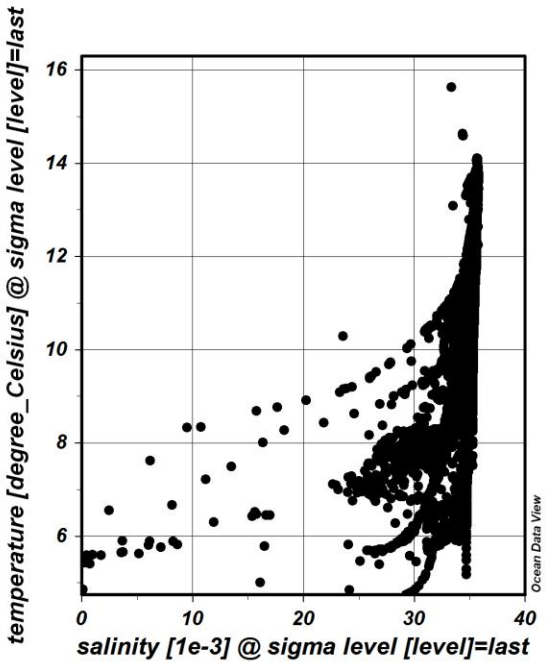
Solution:

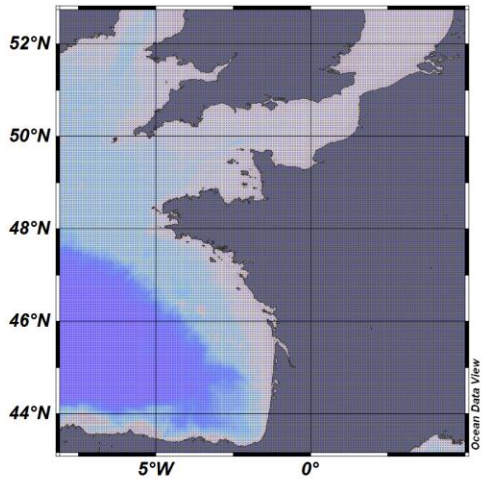
- **Access over the Internet via OPeNDAP**
- **Only user requested subset of data transported**

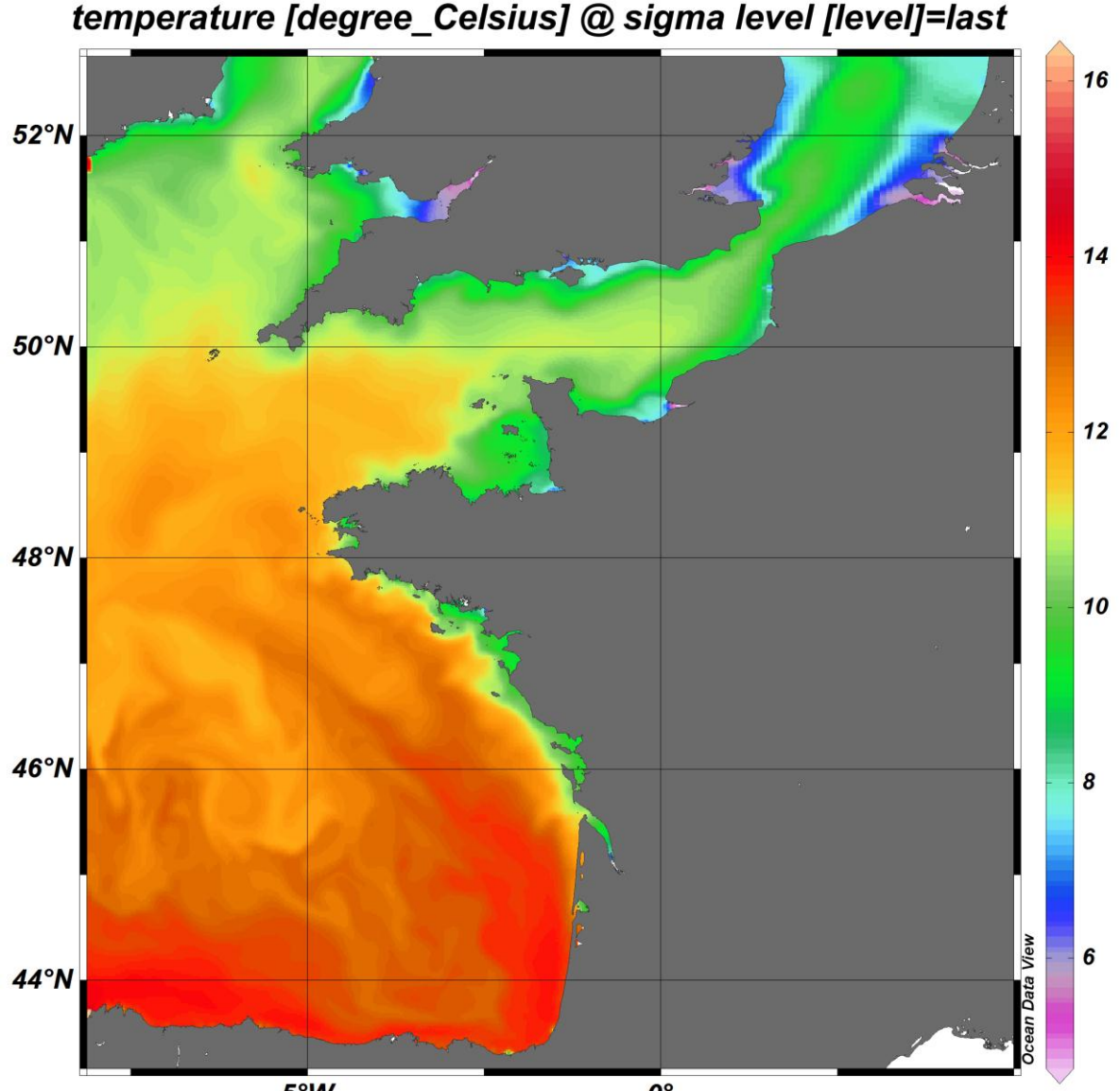

Ocean Data View
File Collection View Import Export Tools Help
Ocean Data View
Open remote netCDF file
Enter URL:
http://dods.ipsl.jussieu.fr/cgi-bin/nph-dods/ocmip/phase5/OUTPUT/BCCR/BCM-C/HISTA2/yr/fgCO2/fgCO2_BCCR_BCM-C_HISTA2_1860-2099.nc
OK
Cancel
© 2012 Reiner Schlitzer
Choose File>Open to open or File>New to create a collection.


air-to-sea CO2 flux [mol C/m^2/s]
3e-07
2e-07
1e-07
0
-1e-07
-2e-07
1900
1950
2000
2050
2100
Ocean Data View

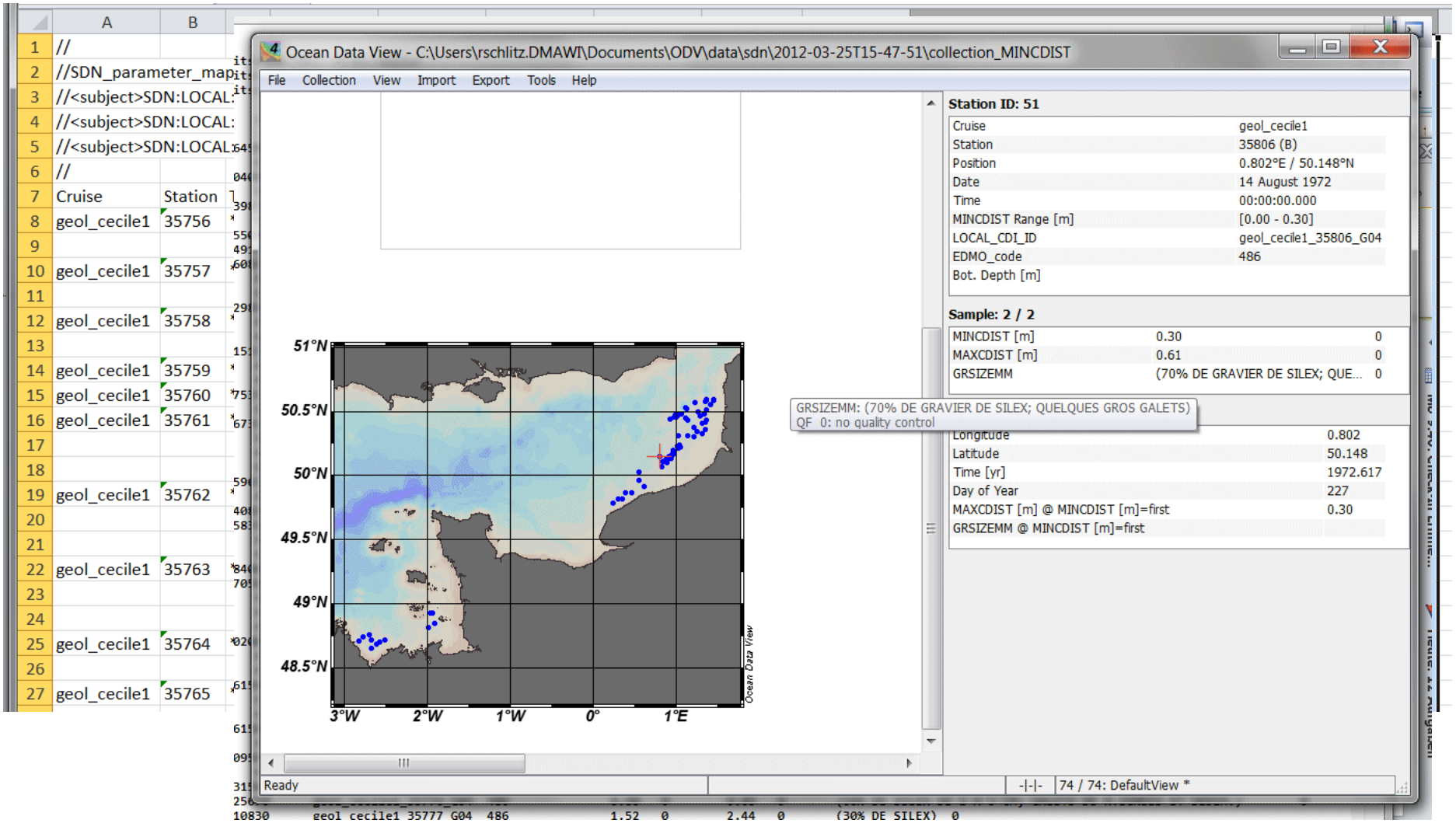

air-to-sea CO2 flux [mol C/m^2/s] @ time [days since 1800-01-01]=first
60°N
30°N
EQ
30°S
60°S
90°E
180°W
90°W
0°
3e-07
2e-07
1e-07
0
-1e-07
-2e-07
-3e-07
Ocean Data View
30°N
30°S
90°E
180°W
90°W
0°
Ocean Data View


temperature [degree_Celsius] @ sigma level [level]=last
16
14
12
10
8
6
0
10
20
30
40
salinity [1e-3] @ sigma level [level]=last
Ocean Data View


52°N
50°N
48°N
46°N
44°N
5°W
0°
Ocean Data View


temperature [degree_Celsius] @ sigma level [level]=last
52°N
50°N
48°N
46°N
44°N
5°W
0°
16
14
12
10
8
6
Ocean Data View

3. Non-numeric data variables



Coming soon …

Aggregated Variables

File Analysis…

```
//
//SDN_parameter_mapping
//<subject>SDN:LOCAL:DEPH</subject><object>SDN:P011::ADEPZZ01</object><units>SDN:P061::ULAA</units>
//<subject>SDN:LOCAL:PRES</subject><object>SDN:P011::PRESPR01</object><units>SDN:P061::UPDB</units>
//<subject>SDN:LOCAL:TEMP</subject><object>SDN:P011::TEMPPR01</object><units>SDN:P061::UPAA</units>
//<subject>SDN:LOCAL:PSAL</subject><object>SDN:P011::PSLTZZ01</object><units>SDN:P061::UUUU</units>
//<subject>SDN:LOCAL:DOX1</subject><object>SDN:P011::DOXYZZXX</object><units>SDN:P061::UMLL</units>
//<subject>SDN:LOCAL:DOXY</subject><object>SDN:P011::DOXYZZXX</object><units>SDN:P061::UPOX</units>
//<subject>SDN:LOCAL:TUR4</subject><object>SDN:P011::TURBXXXX</object><units>SDN:P061::USTU</units>
//<subject>SDN:LOCAL:NTRA</subject><object>SDN:P011::NTRAZZXX</object><units>SDN:P061::UPOX</units>
//<subject>SDN:LOCAL:SLCA</subject><object>SDN:P011::SLCAZZXX</object><units>SDN:P061::UPOX</units>
//<subject>SDN:LOCAL:PHPH</subject><object>SDN:P011::PHXXZZXX</object><units>SDN:P061::UUPH</units>
//<subject>SDN:LOCAL:CORG</subject><object>SDN:P011::CORGZZZX</object><units>SDN:P061::UPOX</units>
//<subject>SDN:LOCAL:POCP</subject><object>SDN:P011::MDMAP010</object><units>SDN:P061::UMMC</units>
//<subject>SDN:LOCAL:TSMP</subject><object>SDN:P011::TSEDZZZZ</object><units>SDN:P061::UMGL</units>
//



//
//SDN_parameter_mapping
//<subject>SDN:LOCAL:DEPTH</subject><object>SDN:P011::ADEPZZ01</object><units>SDN:P061::ULAA</units>
//<subject>SDN:LOCAL:Temperature</subject><object>SDN:P011::TEMPS901</object><units>SDN:P061::UPAA</units>
//<subject>SDN:LOCAL:Salinity</subject><object>SDN:P011::ODSDM021</object><units>SDN:P061::UUUU</units>
//<subject>SDN:LOCAL:Oxygen</subject><object>SDN:P011::DOXYCZ01</object><units>SDN:P061::KGUM</units>
//
```

When aggregating SDN files different BODC parameter/unit codes are maintained as separate variables (e.g., 3 oxygen variables).

User may combine separate variables by means of *aggregated derived variables* (e.g., 1 oxygen variable for scientific usage).

Aggregated derived variable

```
Oxygen [ml/l] =
        1 *       (DOX1 DISSOLVED OXYGEN ml/l)                    + 0
        0.02297 * (DOX2 DISSOLVED OXYGEN micromole/kg)            + 0
        0.02241 * (DOXY DISSOLVED OXYGEN millimole/m3)            + 0
```

Merging Principles:

- Merging Type (*exclusive* use of first available value, *average* over all available values)
- Order of input variables matters (in case of *eclusive* merge)
- Standard or user-specified conversion formulas
- Quality flag inheritence

Many small improvements:

- **More flexible duplicate station detection method**
- **…**