| | |
|---|---|
| **Project Title:** | Baltic+ SeaLaBio |
| **Document Title:** | Dataset Description |
| **Version:** | 1.2 Public version |
| **Author(s) and affiliation(s):** | Kari Kallio, Jenni Attila, Sampsa Koponen, Vesa Keto, SYKE<br>Kerstin Stelzer, Carsten Brockmann, BC<br>Petra Philipson, BG |
| **Version history:** | 0.1 Apr. 02, 2019. Template<br>0.5 May 28, 2019. 1st draft<br>1.0 July 3, 2019. Version sent to ESA for review<br>1.1 Aug 30, 2019. Improved after ESA comments<br>1.2 Nov 20, 2019. Public version |
| **Distribution:** | Public |

# Contents

**Abstract**

This document describes the EO and in situ data collected during WP2 of the Baltic+ SeaLaBio project. It also includes instructions where the open access data used in the project can be obtained. Some of the in situ data is directly accessible through the project web page.

**Glossary**

| | |
|---|---|
| aCDOM | absorption coefficient of CDOM |
| AERONET | Aerosol Robotic Network |
| CDOM | Coloured Dissolved Organic Matter |
| Chl-a | Chlorophyll *a* |
| CODA | Copernicus Online Data Access (EUMETSAT) |
| CTD | Instrument that measures conductivity, temperature, and pressure in water |
| DOC | Dissolved Organic Carbon |
| DOM | Dissolved Organic Matter |
| EFR | TOA radiances at full resolution |
| EO | Earth Observation |
| ERR | TOA radiances at reduced resolution |
| HELCOM | Helsinki Commission |
| ICES | International Council for the Exploration of the Sea |
| L1C | Level 1C, Ortho-rectified TOA reflectance image |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| MSI | MultiSpectral Instrument |
| NR | Near-Real-Time Non-Time-Critical |
| NT | Non-Time-Critical |
| OC | Ocean Color |
| OLCI | Ocean and Land Color Imager |
| RFLEX | Unattended shipborne hyperspectral reflectance measurement system |
| Rrs | Remote sensing reflectance |
| S2(A/B) | Sentinel-2 (A or B) |
| S3(A/B) | Sentinel-3 (A or B) |
| SCIHUB | ESA's Scientific Data Hub |
| SST | Sea Surface Temperature |
| TSM | Total suspended matter |
| TOA | Top of Atmosphere |
| TOC | Total Organic Carbon |
| TSM | Total Suspended Matter |
| VIIRS | Visible Infrared Imaging Radiometer Suite |

# 1 Introduction

The data used in the SeaLaBio project consists of various types of in situ data and satellite data (S2 and S3). The data available for the project was preliminary described in the Requirement Baseline document (RD, Section 3.2). All the data presented in the Requirement Baseline document (Table 3) is utilized in the project except data from the Swedish marine buoy system and the NorSOOP ferrybox ships. The NorSOOP ferrybox routes are outside the SeaLaBio study areas. Tree of the marine buoys are located on the East coast of Sweden and two of them are in the SeaLaBio study areas. Unfortunately, we found out that the buoy data (e.g. Chl-a, turbidity) is not corrected with control sample measurements and are therefore not suitable for the EO validation in its present format. However, there are data from monitoring stations close to these buoys and we will investigate the possibility to evaluate the quality of the buoy data, using the monitoring data, and include it if good results are obtained.

The selection of test areas was based on covering different geographical areas and water quality characteristics and on the availability of in situ data. These are areas where large errors are reported in atmospheric correction and in-water processing. The geographical areas selected for the testing and validation are presented in the Requirement Baseline document (Section 5.2, Table 6). The test years were 2015 – 2018 and season was limited to April-October.

The open access in situ data is accessible through the SeaLaBio project website. All S2 and S3 data is publicly available through the Copernicus services and thus not replicated here.

# 2 EO data

## 2.1 Sentinel data

SYKE has downloaded systematically all Sentinel-2 (starting from 04.07.2015) and Sentinel-3 (starting from 25.04.2016) images from the Baltic Sea area (approx. 5 – 35 deg. E, 52 – 70.5 deg. N).

Sentinel-2 MSI L1C data were downloaded from the Finnish Collaborative Ground Station operated by the Finnish Meteorological Institute (FMI). The same data are publicly available from the SCIHUB service:

https://scihub.copernicus.eu/

In order to use this service one has to Sign up and follow the instructions. Below are example dataset names from the Baltic Sea:

S2A_MSIL1C_20190525T094031_N0207_R036_T36WVA_20190525T101151.zip
S2B_MSIL1C_20190525T103029_N0207_R108_T36WVD_20190525T122847.zip

Sentinel-3 OLCI Level 1b data were downloaded from EUMETSAT CODA. The same data are available from SCIHUB (link above). Below are example dataset names. The abbreviation EFR refers to TOA (Top of Atmosphere) radiances at full resolution (~300 m on ground). The alternative to it is ERR (TOA radiances at reduced resolution or 1200 m on ground). The time stamps (in format YYYYMMDDTHHMMSS) are in the order: sensing start time, sensing stop time and product creation date. The abbreviations NR and NT refer to Near-Real-Time and Non-Time-Critical, respectively. The former are available within a few hours of the satellite over pass. The latter come during the next day and include consolidated auxiliary data. In the processing we will use NT data.

S3B_OL_1_EFR____20190525T081543_20190525T081843_20190525T101852_0179_025_349_1980_MAR_O_NR_002.zip
S3A_OL_1_EFR____20190525T103318_20190525T103618_20190526T173353_0179_045_108_1800_MAR_O_NT_002.zip

The full description of the naming convention and other useful information can be found here:

https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-olci/naming-convention

For the processing of the data SYKE utilizes the FinCal massive parallel processing system. When processing matchups the system goes through the whole image archive and checks if the coordinate and date information of an in situ stations matches any of the images. For processing of product images (maps) a number of representative days will be selected based on cloudiness, availability of in situ data and visibility of interesting features in water.

## 2.2 MODIS and VIIRS data

The SeaLaBio team plans to intercompare the water products derived with SeaLaBio algorithms with those from other satellite missions such as MODIS and VIIRS. The datasets for those instruments are available through:

https://ladsweb.modaps.eosdis.nasa.gov/search/

Here we will use images from the same days as the Sentinel product images described at the end of Chapter 2.1.

# 3 In situ data

In situ data used in the SeaLaBio project can be divided into:

- Water quality data of routine monitoring stations from coastal regions (local and regional), and from open Baltic Sea areas. This monitoring is mainly based on water sampling.
- Water quality data from measurement campaigns related to EO validation
- Water quality data from Ship-of-opportunity (bottle samples and continuous instrument data)
- Reflectance data (AERONET-OC and RFEX Ship-of-opportunity)

The main characteristics of the in situ data used in SeaLaBio, including the availability of the data, are presented in Table 1.

**Table 1.** Description and availability of the SeaLaBio in situ data. Variable numbers refer to the first column (Index) of Table 2.

| **Data type** | **Area** | **Data source** | **Variables and years** | **Availability of data used in SeaLaBio** |
|---|---|---|---|---|
| Water quality data of coastal (local and regional) monitoring | Finnish coast: Bay of Bothnia and Qvark, Kokemäenjoki river estuary, Eastern Gulf of Finland | VESLA national database of Finland | 1,2,4,5,6,8,9,10, 11,12,13,14,15<br><br>2015-2018 | Available at: http://eo.ymparisto.fi/data/water/Baltic_SeaLaBio/ File: BalticPlus_insitu_measurements_FIN.xlsx* |
| | Swedish coast: Bay of Bothnia, Archipelago of Stockholm and Åland Sea | National data bases of Sweden | 1,2,3,4,5,6,8,9, 10,11,12,13,14<br><br>2015-2018 | http://eo.ymparisto.fi/data/water/Baltic_SeaLaBio/ File: BalticPlus_insitu_measurements_SWE.xlsx |
| | German coast | ICES database | 2,4,5<br><br>2015-2017 | http://eo.ymparisto.fi/data/water/Baltic_SeaLaBio/ File: BalticPlus_insitu_measurements_GER.xlsx |
| ICES water quality | Especially open sea areas next to the SeaLaBio pilot areas and Baltic Proper pilot area. | ICES database | 1,2,4,5,8<br><br>2015-2017 | http://eo.ymparisto.fi/data/water/Baltic_SeaLaBio/ File: BalticPlus_insitu_measuremets_ICES.xlsx |
| Water quality from campaigns | Kokemäenjoki and Kyrönjoki estuaries (Gulf of Bothnia, Finnish coast) | Research data (SYKE) | 1,2,5,6,8,9,12,13<br><br>2017 (Kokemäenjoki) and 2018 (Kyrönjoki) | http://eo.ymparisto.fi/data/water/Baltic_SeaLaBio/ Files: FIN_KokemaenjokiEstuary2017_WaterQuality.xlsx FIN_KyronjokiEstuary2018_WaterQuality.xlsx |
| | | | | |
| Water quality data from Ship-of-opportunity (bottle samples) | Alg@line** Finnmaid ship route: Helsinki–Travemünde | Research data (SYKE) | 5<br><br>2016-2018 | http://eo.ymparisto.fi/data/water/Baltic_SeaLaBio/ File: BalticPlus_Algaline_bottle.xlsx |
| Water quality data from Ship-of-opportunity, instrument data | Alg@line* Finnmaid ship route: Helsinki–Travemünde | Research data (SYKE) | 3, 5 (quality checked data)<br><br>2015-2016 | http://eo.ymparisto.fi/data/water/Baltic_SeaLaBio/ File: BalticPlus_Algaline_line.xlsx |
| Reflectance | Alg@line RFLEX | Research data (SYKE) | Rrs 2016 | Contact SYKE (sampsa.koponen@ymparisto.fi) |
| | AERONET-OC stations, Helsinki Lighthouse and Gustav Dahlen | AERONET-OC | Rrs 2015-2017 | AERONET-OC web site (https://aeronet.gsfc.nasa.gov/new_web/ocean_color.html) |

* Uncertainties are available in BalticPlus_insitu_uncertainties_FIN.xlsx
** https://www.finmari-infrastructure.fi/ferrybox/

The water constituents included in the water quality data (coastal routine monitoring, open Baltic sea routine monitoring, campaigns, Alg@line ) are listed in Table 2. The wavelength(s) used for aCDOM determination varies between countries. The number of aCDOM measurements is low compared e.g. to Chl-a. Therefore we also included PtColor measured routinely in Finnish coastal waters in the SeaLaBio dataset.

Table 2. Water constituents included in the SeaLaBio water quality data (coastal and open Baltic sea routine monitoring, campaigns). Water constituents available in the datasets vary by region and station.

| **Index** | **Short name** | **Explanation** | **Unit** |
|---|---|---|---|
| 1 | Secchi | Secchi depth | m |
| 2 | SST_Sampler | Temperature measured using a sampler | $^oC$ |
| 3 | SST_CTD | Temperature measured by CTD | $^oC$ |
| 4 | Salinity | | PSU |
| 5 | Chl-a | Chlorophyll *a* | µg/l |
| 6 | TSM(SPM) | Total suspended matter (suspended particulate matter) | mg/l |
| 7 | InorganicSM | Inorganic suspended matter | mg/l |
| 8 | Turbidity | | FNU or FTU |
| 9 | aCDOM400 | absorption coefficient of CDOM at 400 nm | 1/m |
| 10 | aCDOM440 | absorption coefficient of CDOM at 440 nm | 1/m |
| 11 | aCDOM700 | absorption coefficient of CDOM at 700 nm | 1/m |
| 12 | aCDOM750 | absorption coefficient of CDOM at 750 nm | 1/m |
| 13 | DOC | Dissolved organic carbon | mg/l |
| 14 | TOC | Total organic carbon | mg/l |
| 15 | PtColor | Indirect measure of CDOM through comparison of water samples with standard cobalt chloride disks | mgPt/l |

RFEX is an unattended shipborne hyperspectral reflectance measurement system (Simis & Olsson 2013), which was installed on two Alg@line ships.

The ICES data was requested in two phases. First the data from defined HELCOM stations were downloaded trough https://ocean.ices.dk/Helcom/Helcom.aspx?Mode=1. This dataset included T, salinity, Chl-a and Secchi, which belong to the ICES oceanographic data variables. We also checked the Ecosystem database of ICES (http://ecosystemdata.ices.dk/), and downloaded a few turbidity observation from it. In the eastern Gulf of Finland, we did not use the ICES data, because the removal of duplicate observation (in ICES and VESLA database) was challenging due to differences in metadata. Most of the observation made in the central Gulf of Finland are, however, in the VESLA database and thus included in the SeaLabio dataset.

The coastal routine data files of Finland and Sweden have more data than what is available from these areas in the ICES database. This is due to the fact that not all local and regional data is sent to ICES. The German coastal data was downloaded from ICES database.

Only measurements of the surface layer (0 - 15 m) were accepted in the SeaLaBio dataset.
The routine monitoring data files include the 'Region'-column, which can be used to classify the data by area. Following areal division was used here: Bay of Bothnia and Qvark, Kokemäenjoki Estuary (Finnish coast), Bothnian Sea, Eastern Gulf of Finland, Stockholm archipelago, Åland Sea, German coast and Baltic Proper.

The SeaLaBio in situ data (routine monitoring measurements based on sampling at fixed stations) is here visualized in four ways: histogram of the total number of observations (Fig. 1), time series of the number of observations (Fig. 2), location of in situ stations on map (Fig. 3) and spatial representation of the number of observations (Fig. 4). **In the calculations of the number of observations multiple sampling depths were not consider, i.e. if a variable was measured from more than one depth during one station visit the number of observation increased only by one for that variable**.

The total number of observation varied considerably by variable (Fig. **1**). SDT, Chl-a and turbidity were measured more often than the variables related to CDOM (absorption coefficients, DOC, TOC, PtColor), when comparing the variables directly linked to EO estimation of water quality. The number of TSM observations was lower than that of TSM.

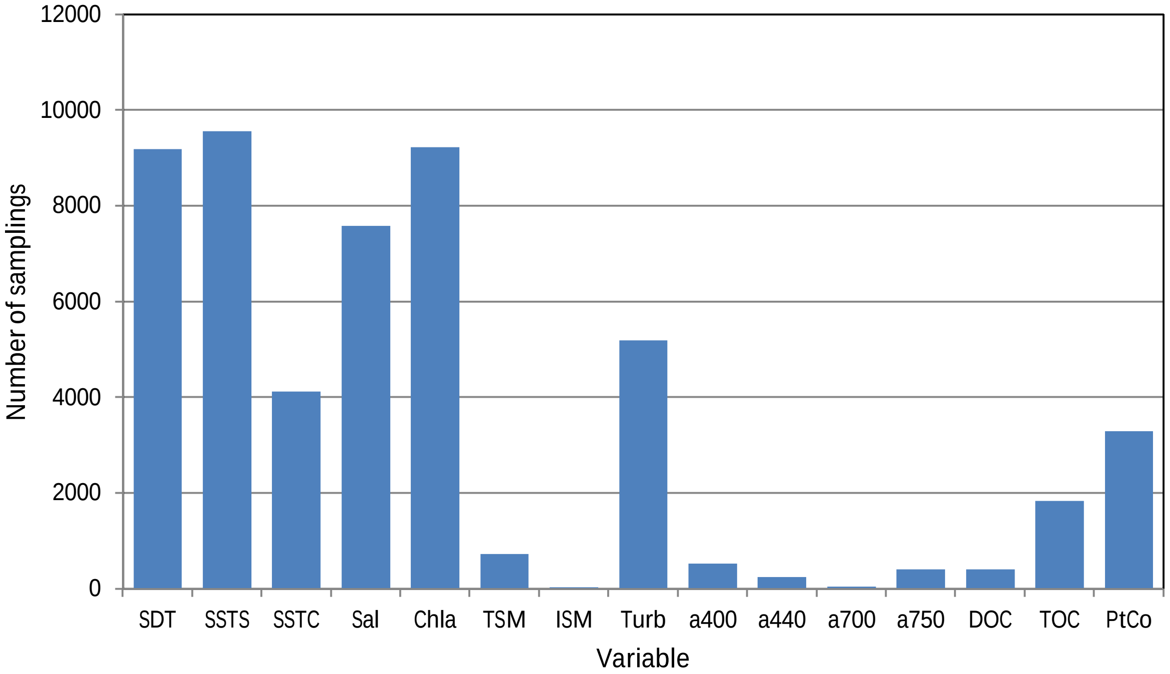



Fig. 1. Total number of observations in the SeaLaBio in situ dataset (2015 – 2018, April-October, routine monitoring measurements based on sampling at fixed stations).

The time series of the number of observations, integrated over four years, were similar for all variables: the number of samples per week was typically higher during summer months (June-August) than in spring and autumn. Here the time series are presented for chl-a, turbidity, acdom400 and TOC (Fig. 2).

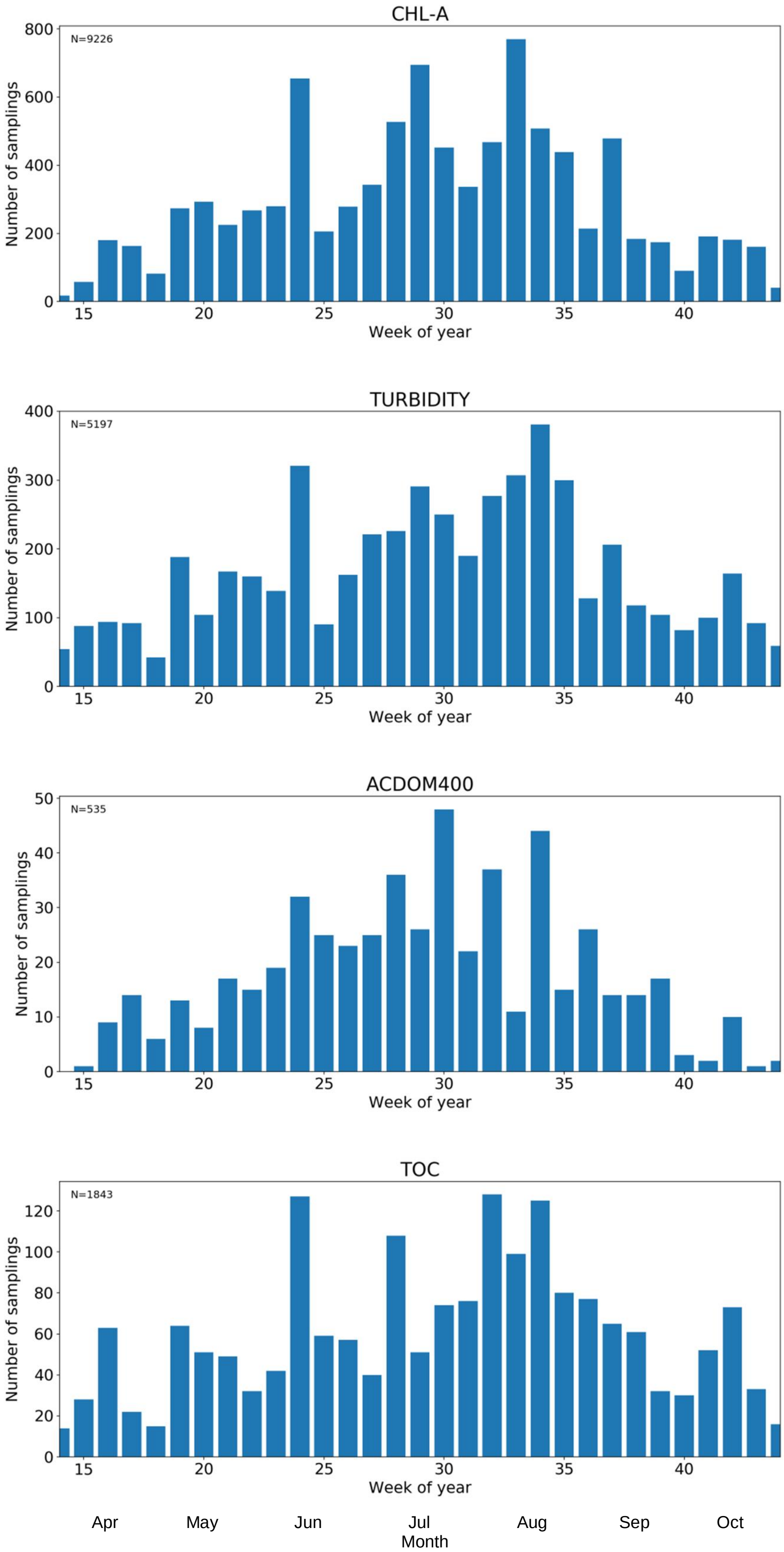



Fig. 2. Time series of the number of in situ observations for Chl-a, turbidity, aCDOM400 and TOC during April-October in 2015-2018 in the SeaLaBio dataset (routine monitoring measurements based on sampling at fixed stations). N is the total number of observations.

Locations of the measurement stations of the SeaLaBio database are presented in Fig. 3. The map of the number of observations of Chl-a in the HELCOM 20 km x 20 km grid is presented in Fig. 4. The corresponding maps for other variables are included in the appendix. The availability of different variable measurements varied considerably regionally. For example, SDT and Chl-a were usually measured in all SeaLaBio sub-areas, while CDOM related variables were often measured in the northern Baltic sea, but not in southern parts. Inorganic SM was measured only the northern parts of the Swedish coast.

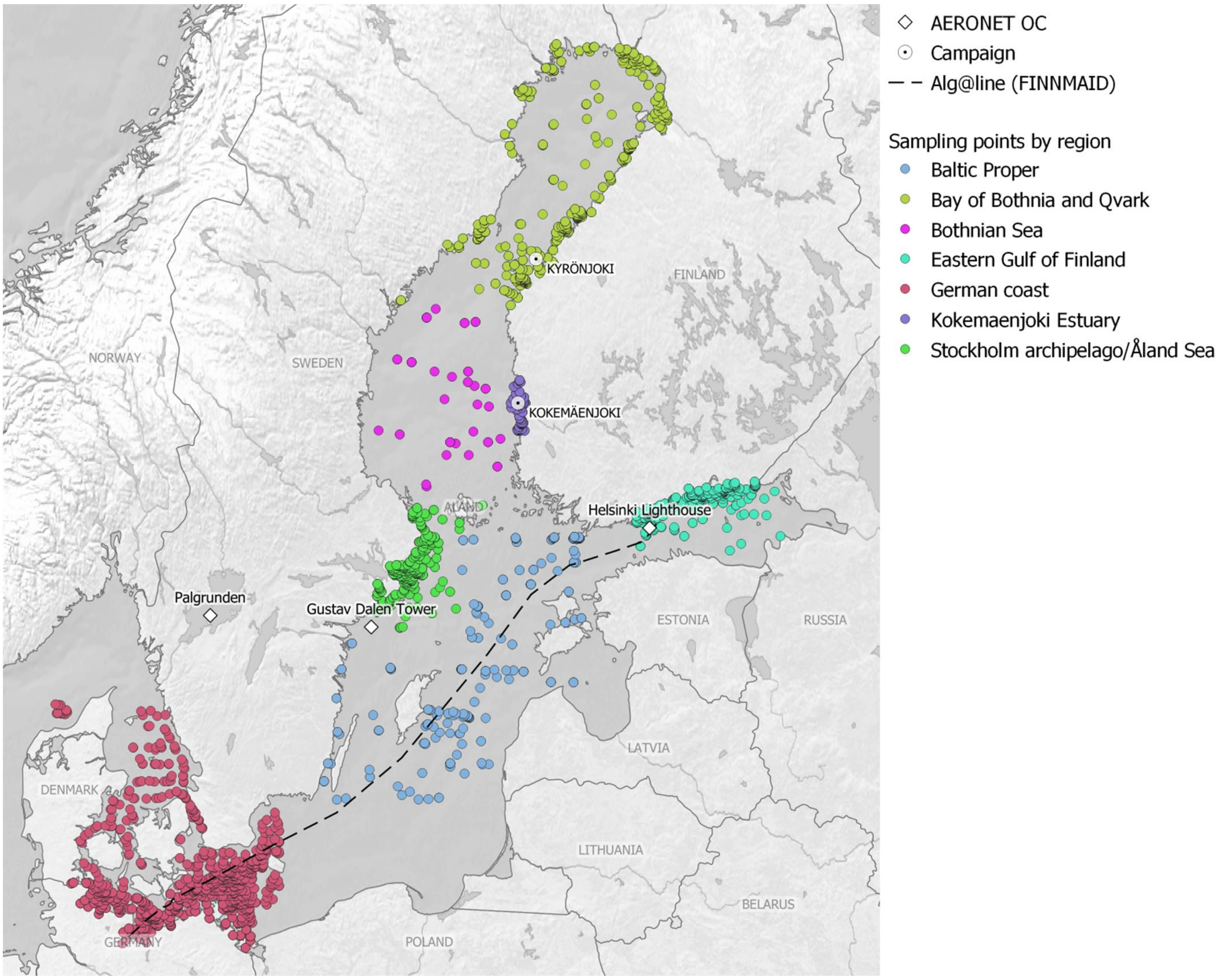



Fig. 3. Location of in situ stations included in the SeaLaBio dataset. These stations had at least one measurement of at least one of the SeaLaBio variables.

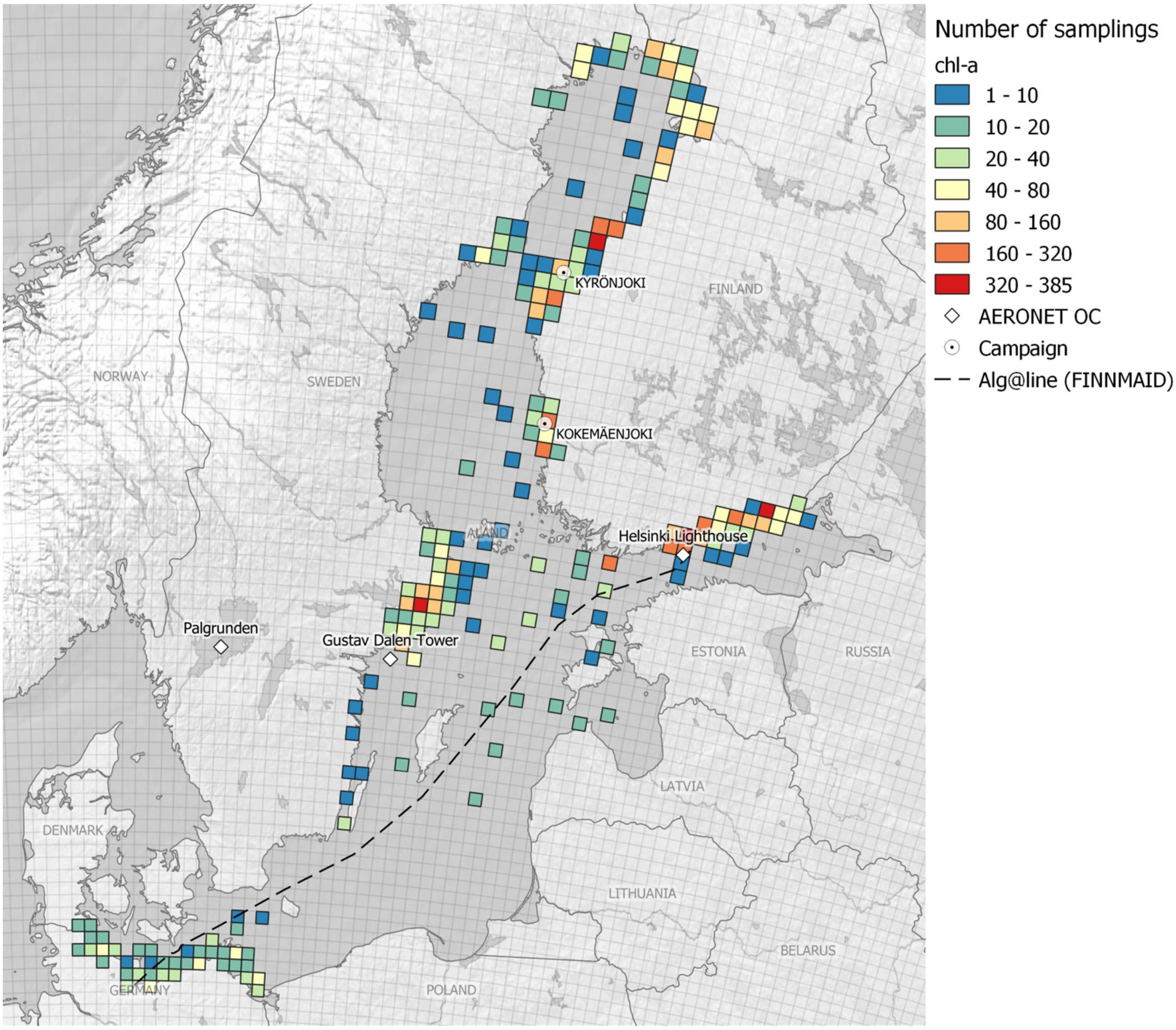



Fig. 4. The number of in situ Chl-a observations (routine monitoring measurements based on sampling at fixed stations) in the SeaLabio dataset presented in the HELCOM 20 km x 20 km grid. The locations of the FINNMAID ferrybox route, AERONET OC stations and campaign locations are also indicated. Results for the rest of the variables are presented in the Appendix.

The determination methods and uncertainties of laboratory measurements were available for part of the downloaded in situ water quality data (Finnish and Swedish coastal data). The way this information is expressed in the databases varies between countries, and therefore their harmonization is laborious. The uncertainties for each measurement (whenever available in the VESLA database) in the Finnish coastal data are presented in a separate file (BalticPlus_insitu_uncertainties_FIN.xlsx).

The number of actual matchups (with satellite and in situ data) was not analyzed in this document. This will be done in Version 2 of the Validation Report.

# 4 Availability of the SeaLaBio dataset

This section summarizes the availability of the data used in SeaLaBio.

EO data

- Sentinel data used in SeaLaBio is publicly available from the Copernicus services. For downloading the data see instructions in Table 1.

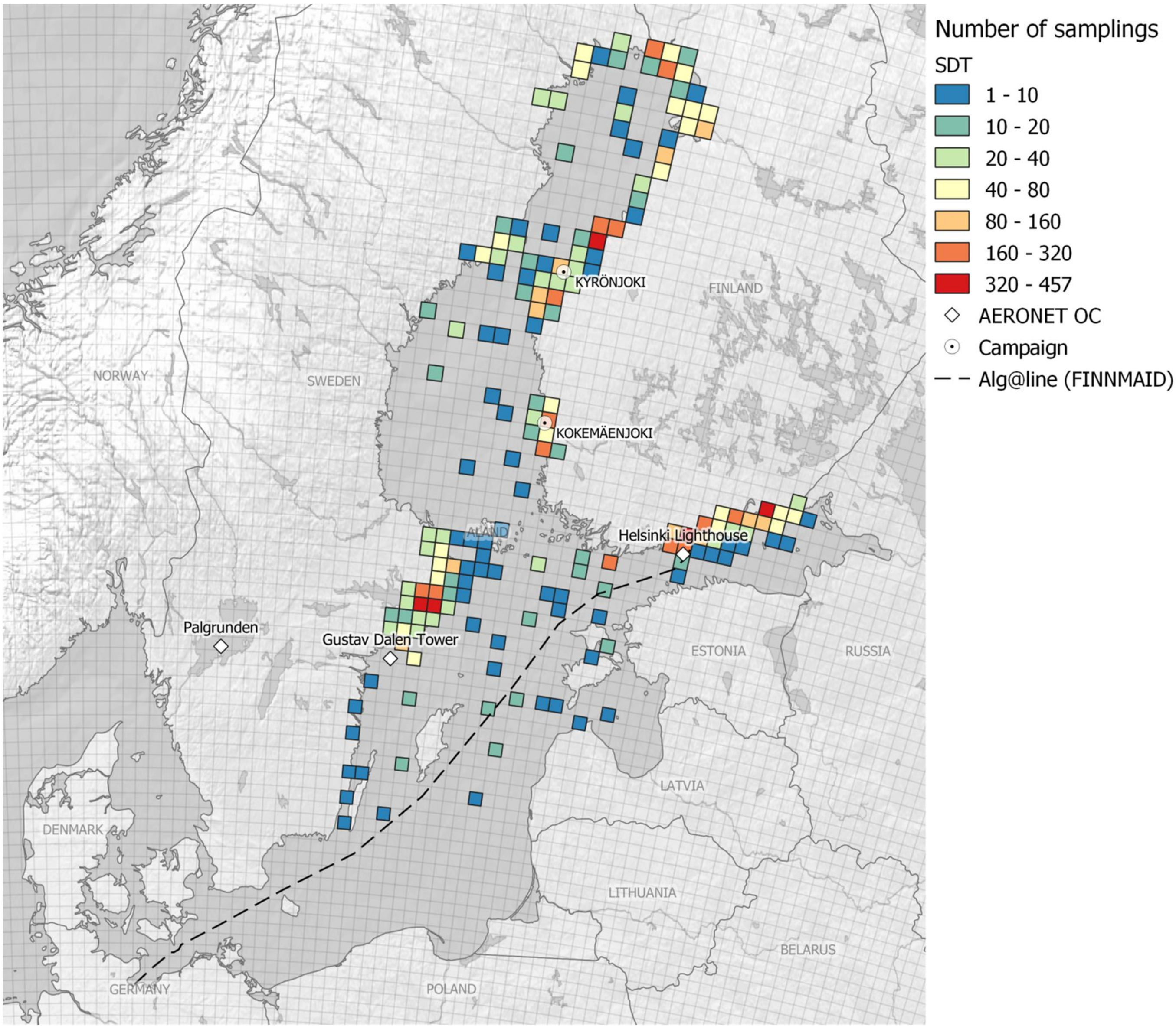
In situ data

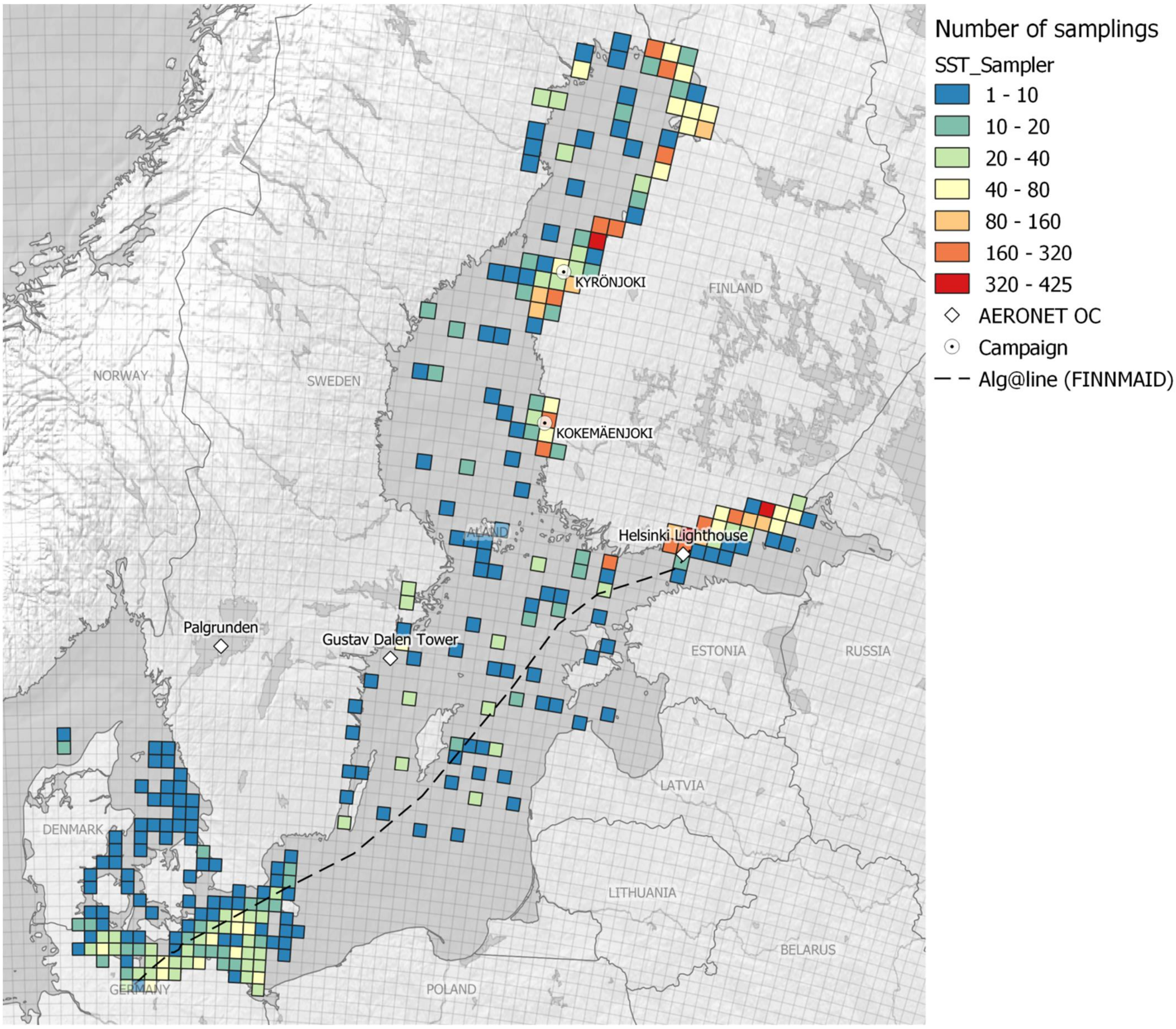
- Most of the openly available in situ data used in the SeaLaBio can be downloaded as excel files from http://eo.ymparisto.fi/data/water/Baltic_SeaLaBio/ . See Table 1 for details.
- Information on the availability of other openly available in situ data (e.g. AERONET-OC) can be found in Table 1.

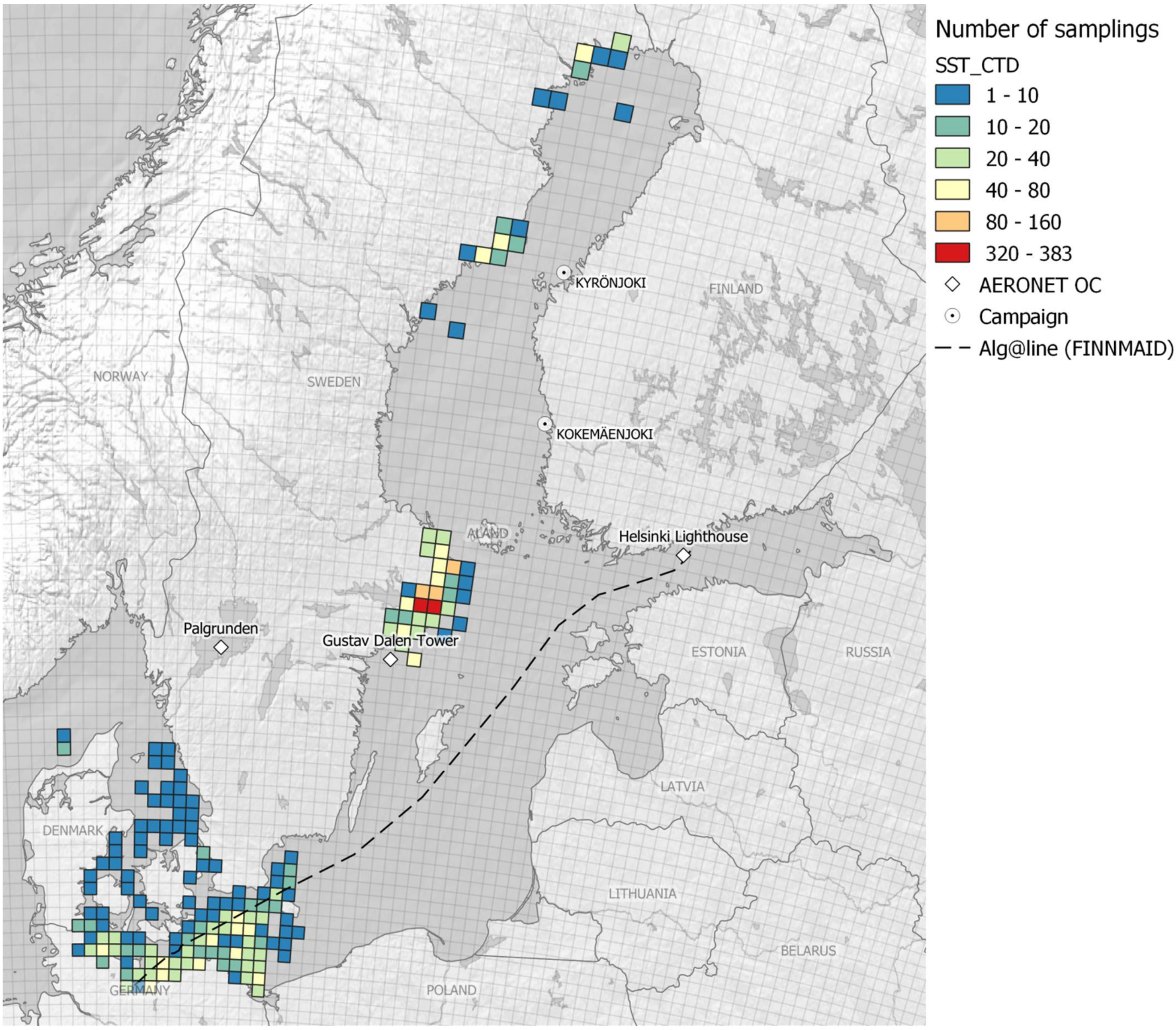
# 5 References

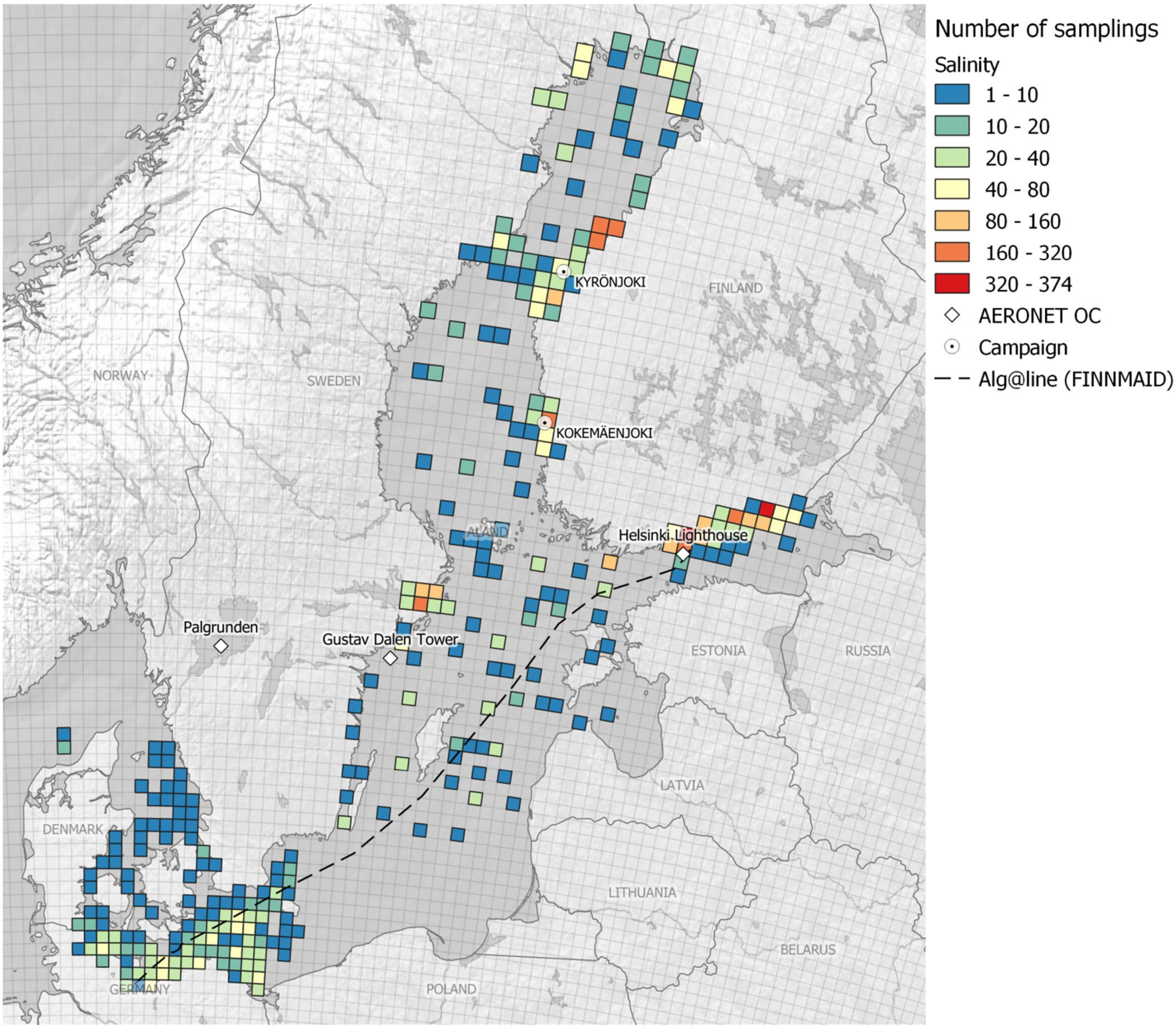
Simis, S. G.H. & Olsson, J. 2013. Unattended processing of shipborne hyperspectral reflectance measurements, Remote Sensing of Environment 135: 202-212, https://doi.org/10.1016/j.rse.2013.04.001.

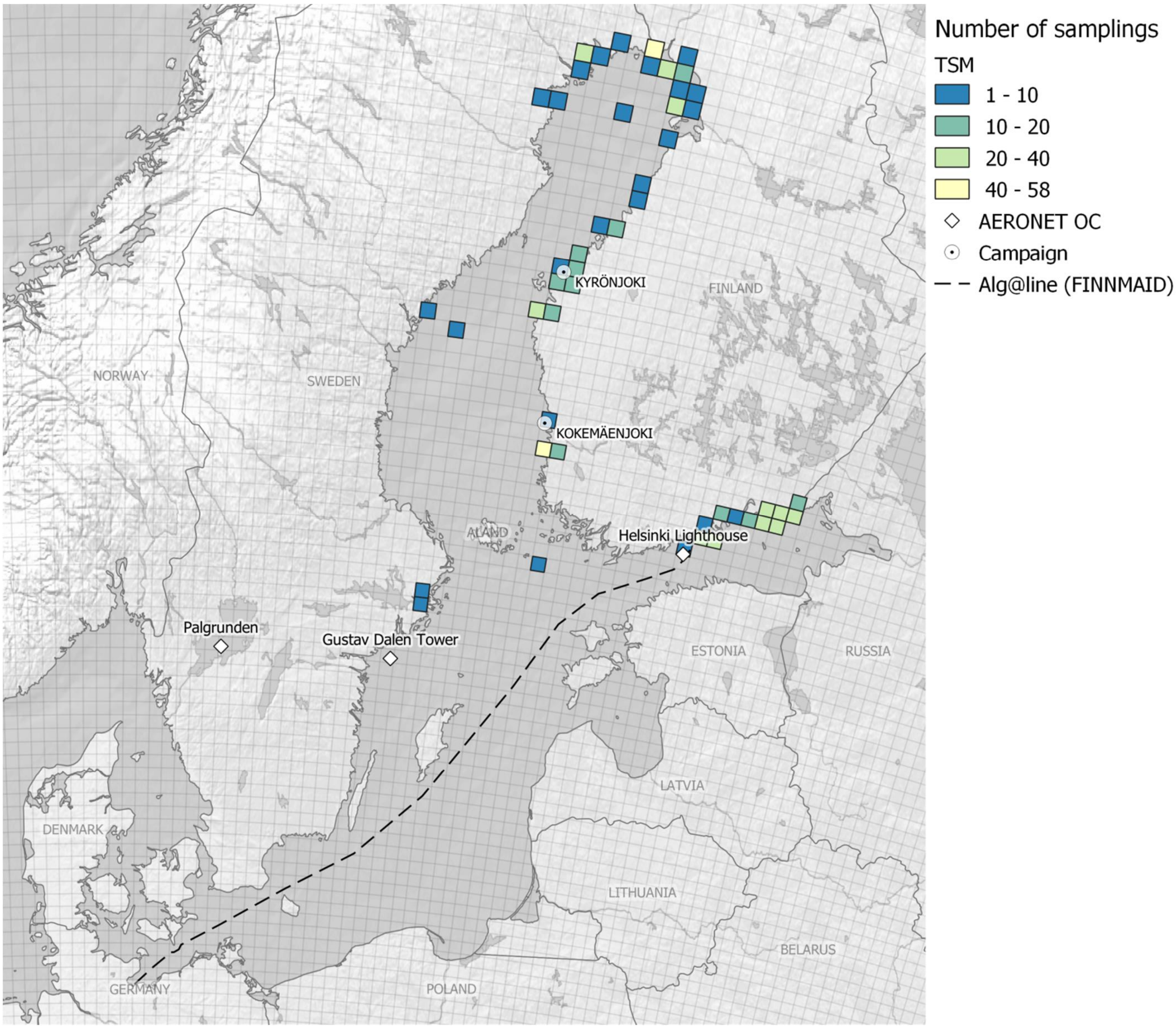
# 6 Appendix: The number of observations by variable in the HELCOM 20 km x 20 km grid

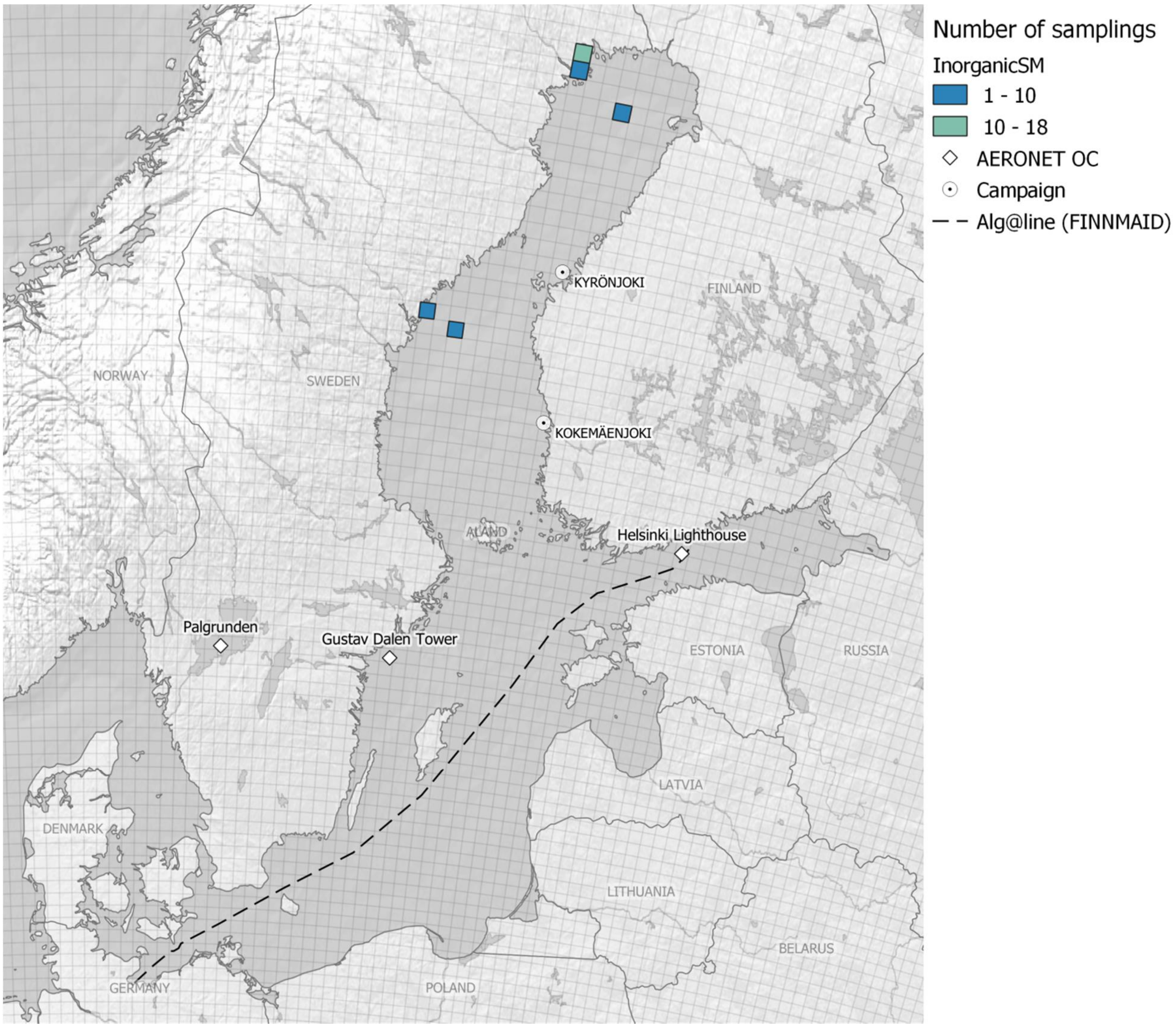


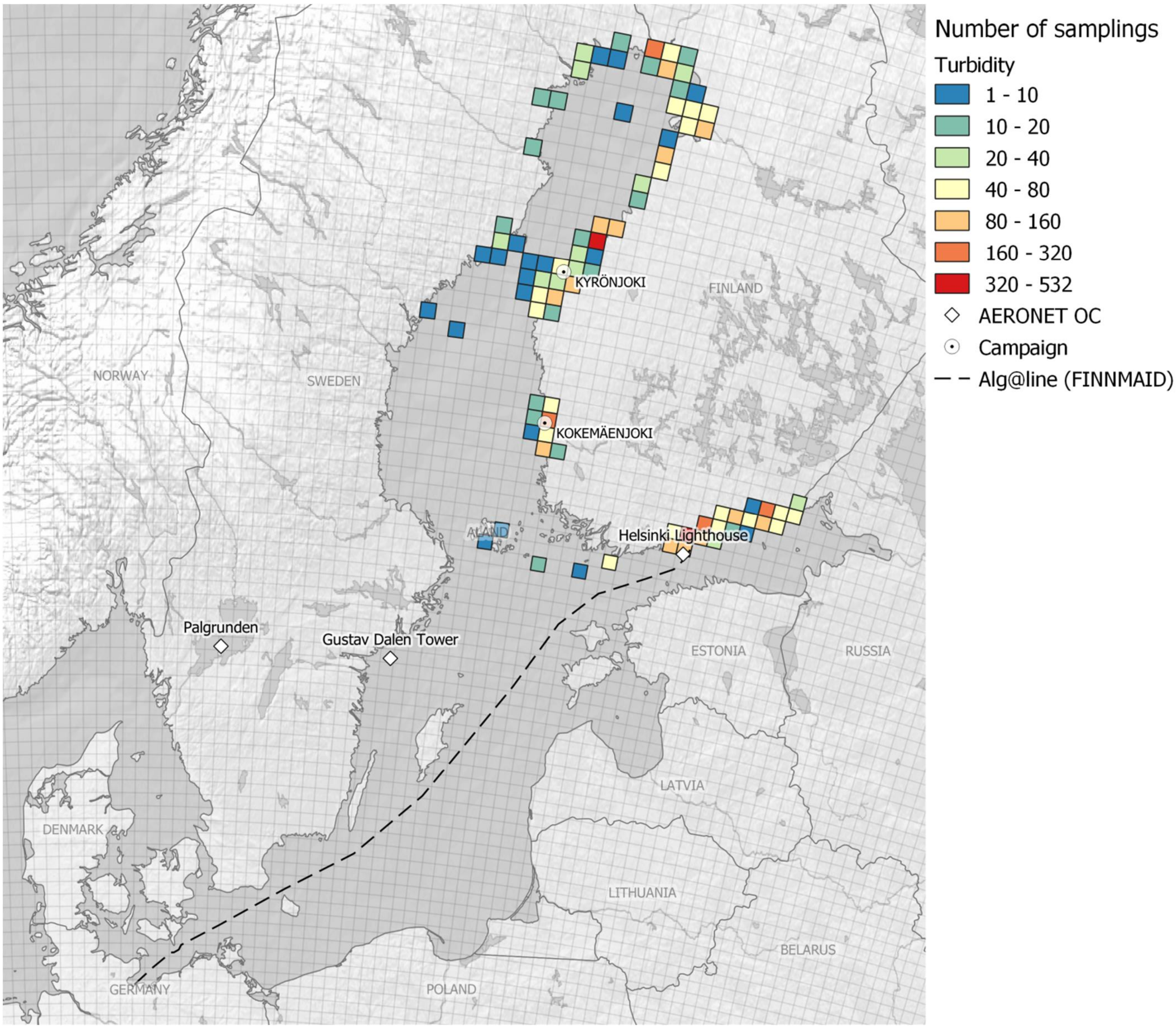

Number of samplings
SST_Sampler
1 - 10
10 - 20
20 - 40
40 - 80
80 - 160
160 - 320
320 - 425
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
NORWAY
SWEDEN
FINLAND
ÅLAND
ESTONIA
RUSSIA
LATVIA
LITHUANIA
BELARUS
DENMARK
GERMANY
POLAND

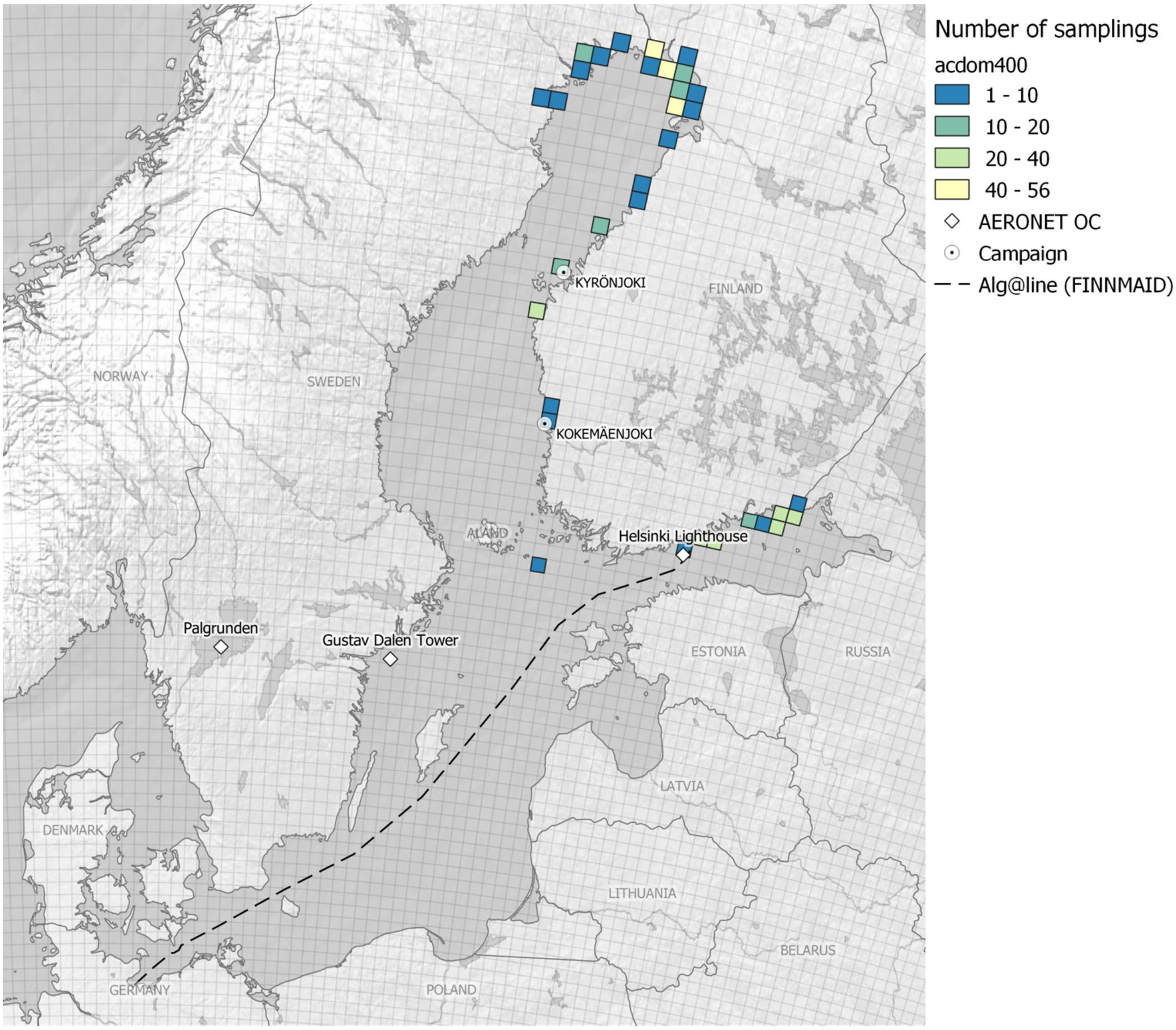

Number of samplings
SST_CTD
1 - 10
10 - 20
20 - 40
40 - 80
80 - 160
320 - 383
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
NORWAY
SWEDEN
FINLAND
ALAND
ESTONIA
RUSSIA
LATVIA
LITHUANIA
BELARUS
DENMARK
GERMANY
POLAND

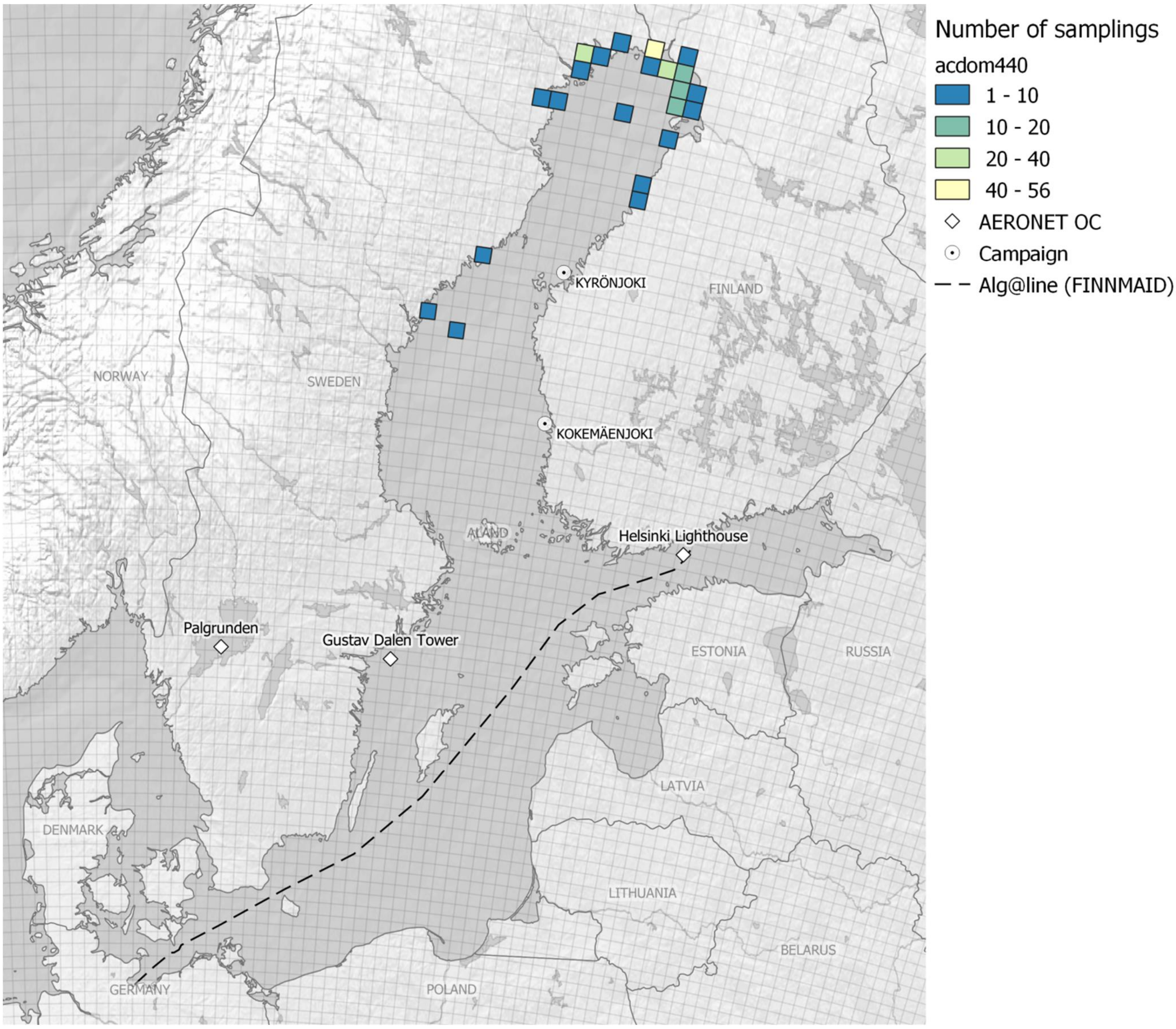

Number of samplings
Salinity
1 - 10
10 - 20
20 - 40
40 - 80
80 - 160
160 - 320
320 - 374
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
NORWAY
SWEDEN
FINLAND
ESTONIA
RUSSIA
LATVIA
LITHUANIA
BELARUS
POLAND
GERMANY
DENMARK
ÅLAND

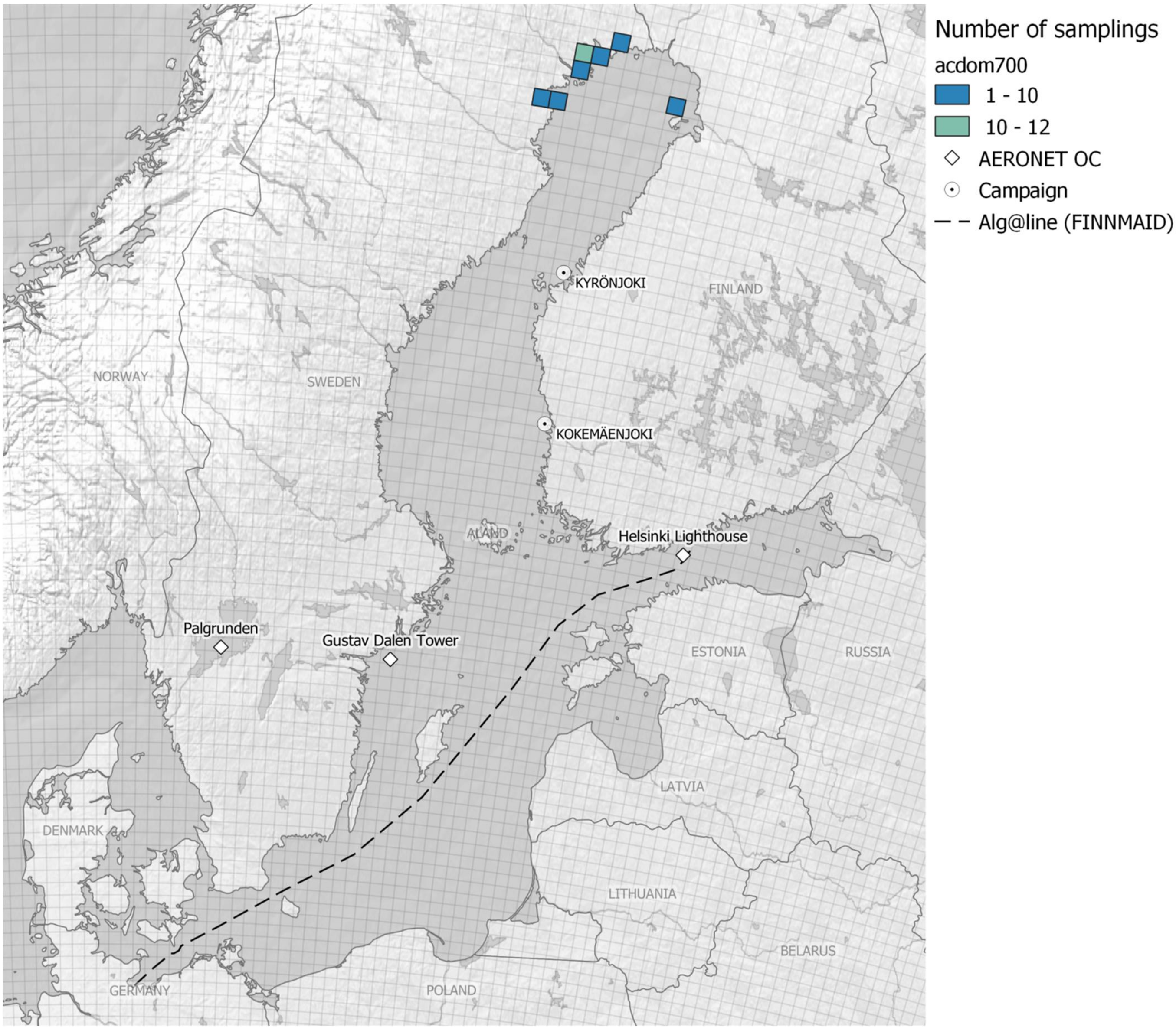

Number of samplings
TSM
1 - 10
10 - 20
20 - 40
40 - 58
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
NORWAY
SWEDEN
FINLAND
ALAND
ESTONIA
RUSSIA
LATVIA
DENMARK
LITHUANIA
BELARUS
GERMANY
POLAND

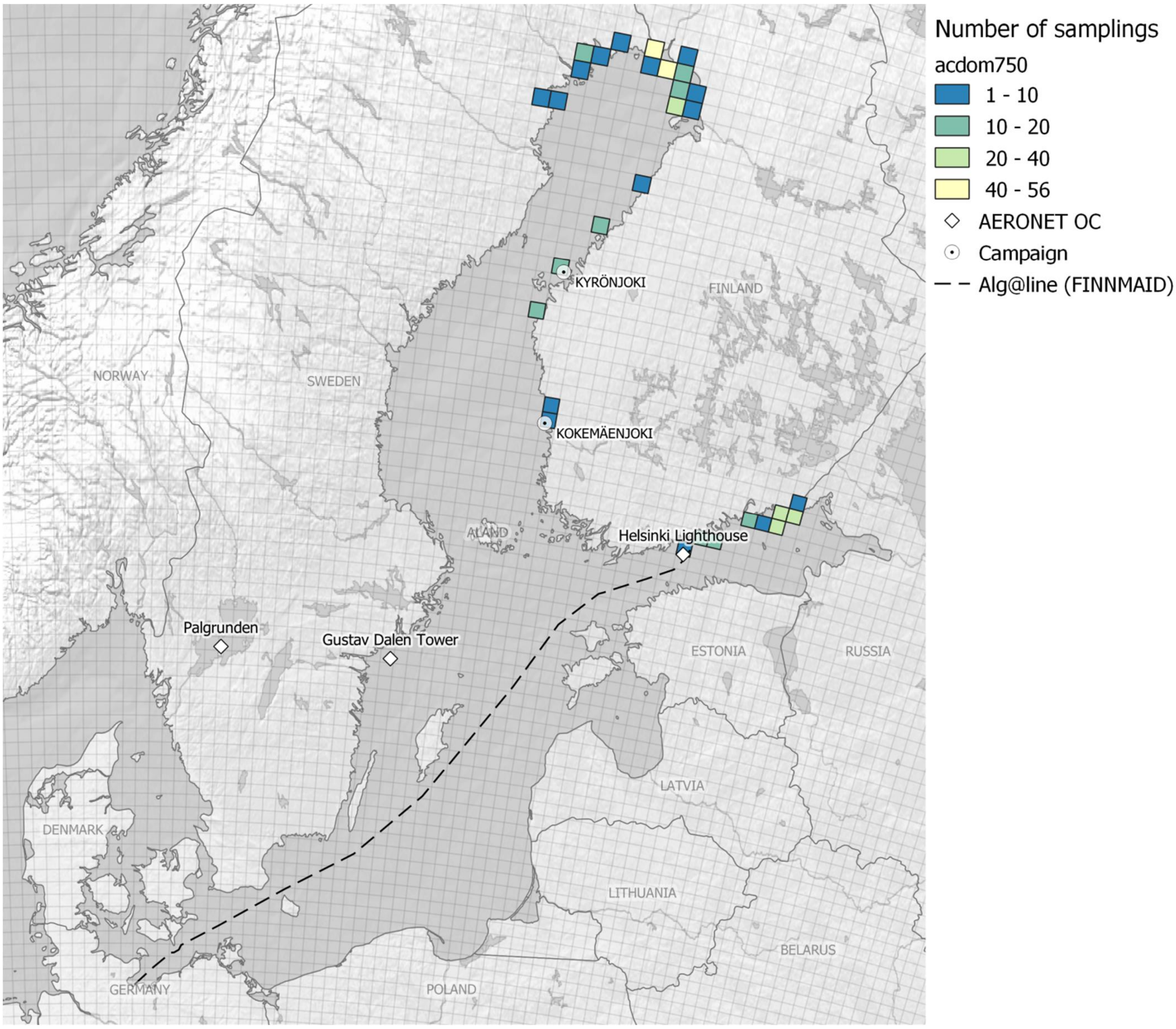

Number of samplings
InorganicSM
1 - 10
10 - 18
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
FINLAND
NORWAY
SWEDEN
ÅLAND
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
ESTONIA
RUSSIA
LATVIA
DENMARK
LITHUANIA
BELARUS
GERMANY
POLAND

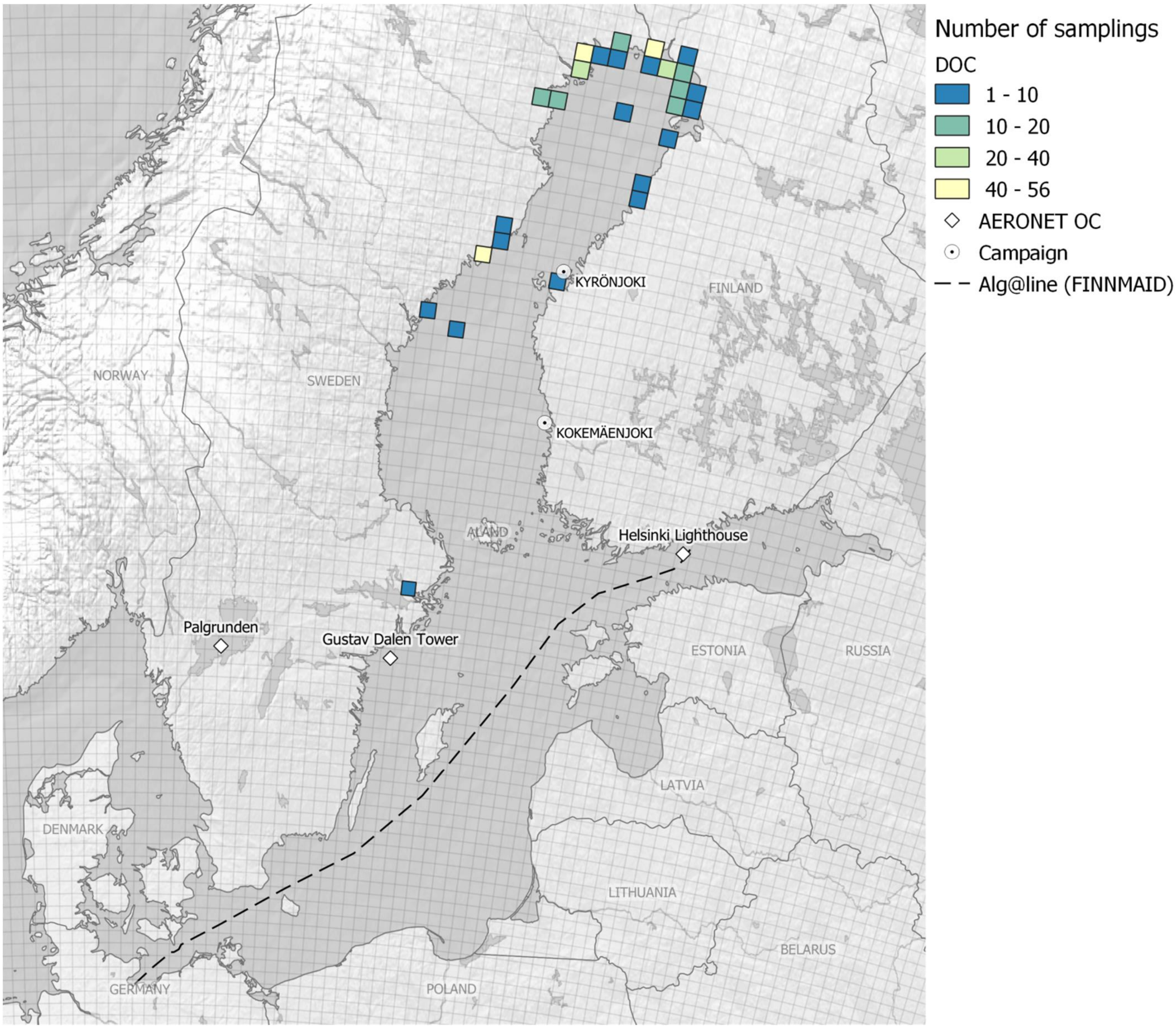

Number of samplings
Turbidity
1 - 10
10 - 20
20 - 40
40 - 80
80 - 160
160 - 320
320 - 532
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
NORWAY
SWEDEN
FINLAND
ESTONIA
RUSSIA
LATVIA
LITHUANIA
DENMARK
BELARUS
GERMANY
POLAND

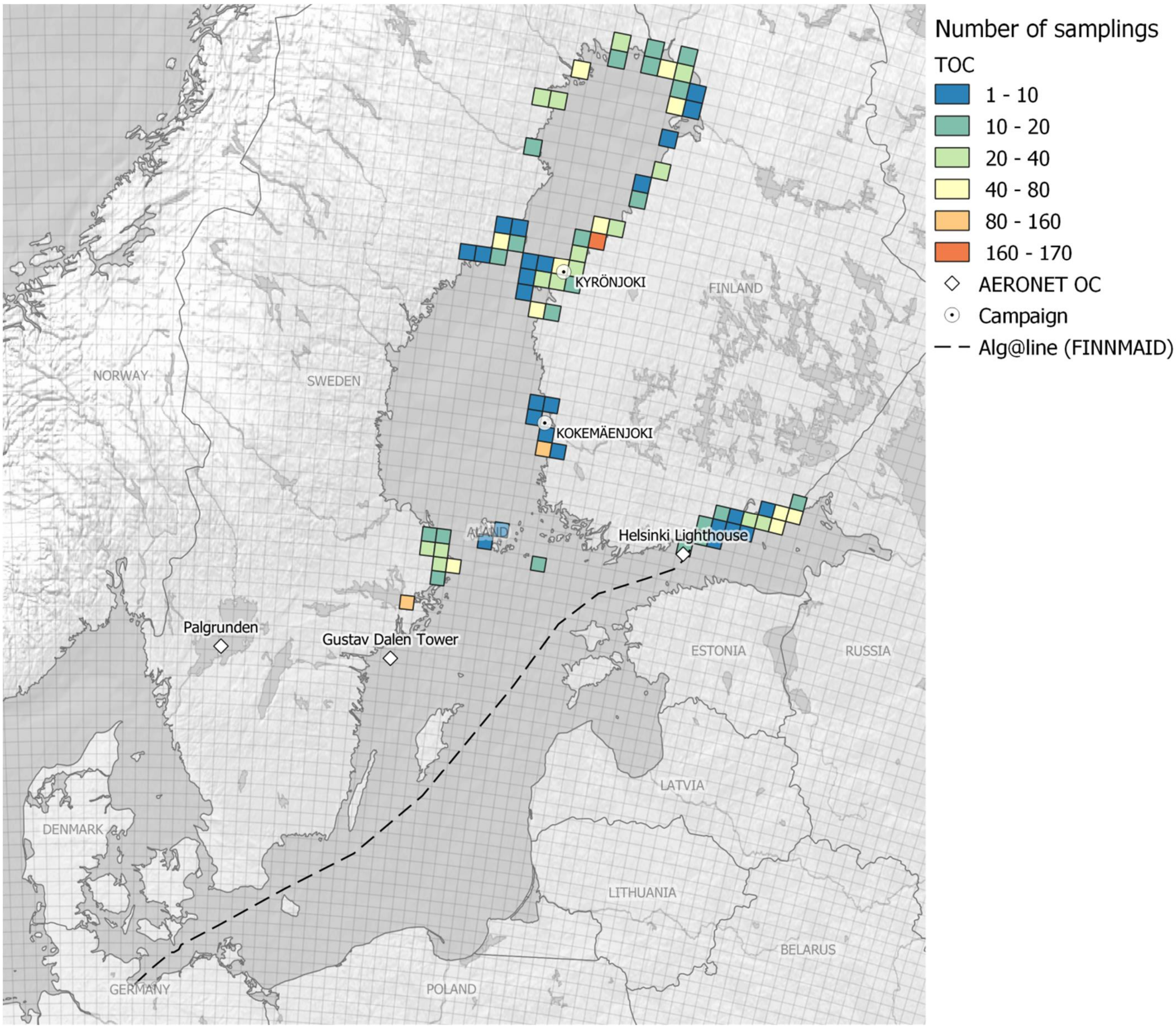

Number of samplings
acdom400
1 - 10
10 - 20
20 - 40
40 - 56
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
FINLAND
SWEDEN
NORWAY
ÅLAND
ESTONIA
RUSSIA
LATVIA
LITHUANIA
DENMARK
BELARUS
GERMANY
POLAND

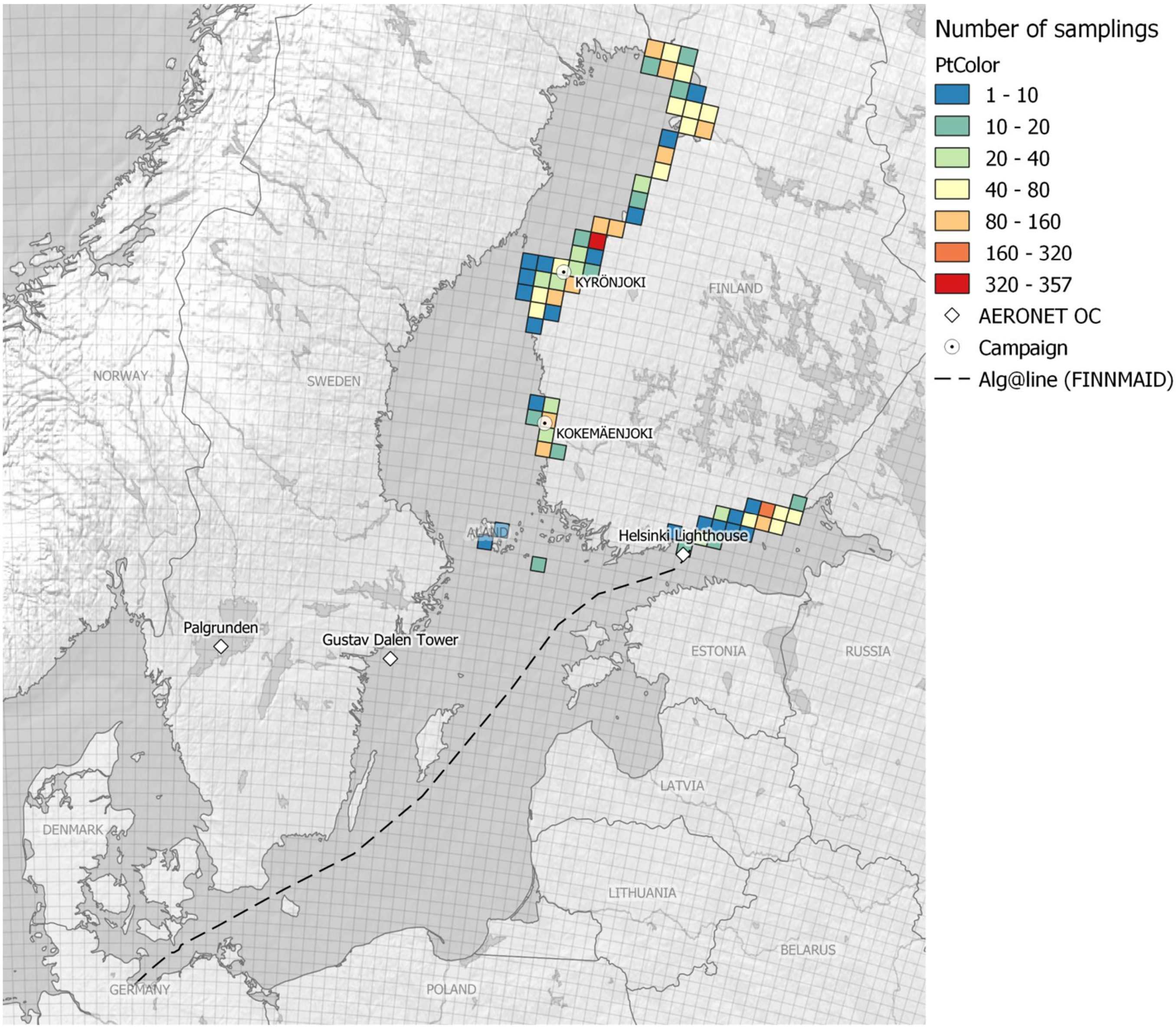

Number of samplings
acdom440
1 - 10
10 - 20
20 - 40
40 - 56
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
FINLAND
NORWAY
SWEDEN
ÅLAND
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
ESTONIA
RUSSIA
LATVIA
DENMARK
LITHUANIA
BELARUS
GERMANY
POLAND


Number of samplings
acdom700
1 - 10
10 - 12
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
FINLAND
NORWAY
SWEDEN
ÅLAND
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
ESTONIA
RUSSIA
LATVIA
DENMARK
LITHUANIA
BELARUS
GERMANY
POLAND


Number of samplings
acdom750
1 - 10
10 - 20
20 - 40
40 - 56
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
NORWAY
SWEDEN
FINLAND
ALAND
ESTONIA
RUSSIA
LATVIA
LITHUANIA
BELARUS
DENMARK
GERMANY
POLAND


Number of samplings
DOC
1 - 10
10 - 20
20 - 40
40 - 56
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
NORWAY
SWEDEN
FINLAND
ALAND
ESTONIA
RUSSIA
LATVIA
DENMARK
LITHUANIA
BELARUS
GERMANY
POLAND


Number of samplings
TOC
1 - 10
10 - 20
20 - 40
40 - 80
80 - 160
160 - 170
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
NORWAY
SWEDEN
FINLAND
ALAND
ESTONIA
RUSSIA
LATVIA
DENMARK
LITHUANIA
BELARUS
GERMANY
POLAND


Number of samplings
PtColor
1 - 10
10 - 20
20 - 40
40 - 80
80 - 160
160 - 320
320 - 357
AERONET OC
Campaign
Alg@line (FINNMAID)
KYRÖNJOKI
KOKEMÄENJOKI
Helsinki Lighthouse
Palgrunden
Gustav Dalen Tower
ÅLAND
FINLAND
NORWAY
SWEDEN
ESTONIA
RUSSIA
LATVIA
DENMARK
LITHUANIA
BELARUS
GERMANY
POLAND