Project Title:	Baltic+ SeaLaBio
Document Title:	Scientific Roadmap
Version:	2.1 Final version
Author(s) and affiliation(s):	Petra Philipson, Susanne Thulin, BG Sampsa Koponen, Jenni Attila, Kari Kallio, SYKE Carsten Brockmann, Dagmar Müller, BC Thomas Neumann, IOW Constant Mazeran, SOLVO
Version history:	1.0 Sept. 30, 2020. Version for ESA review 2.0 Nov. 10, 2020. Version for ESA review 2.1 Dec. 10, 2020. Final version
Distribution:	Public

Contents

Glossary ... 3
1 Introduction ... 4
2 Main scientific challenges ... 4
3 Baltic+ SeaLaBio project results ... 5
3.1 Atmospheric correction ... 5
3.1.1 Main results ... 5
3.1.2 Future work ... 6
3.2 In-water retrieval ... 8
3.2.1 Main results ... 8
3.2.2 Future work ... 8
3.3 Biogeochemical modelling ... 9
3.3.1 Main results ... 9
3.3.2 Future work ... 10
3.4 In situ data ... 11
3.4.1 Observations ... 11
3.4.2 Recommendations ... 11
4 Carbon cycle understanding ... 12
5 Future steps for SeaLaBio ... 13
6 A vision for the Baltic Sea – the LEGO Baltic Sea Initiative ... 14
References ... 17

Abstract

This document is a summary of the developments and achievements within the ESA Baltic+ SeaLaBio project. Main results are presented together with proposed future improvements and next developments steps, as well as areas of research for improved carbon cycle understanding.

Glossary

AC	Atmospheric correction
AERONET	Aerosol Robotic Network
BGC	Bio Geo Chemical
C2RCC	Care 2 Regional Coast Color
CDOM	Coloured Dissolved Organic Matter
Chl a	Chlorophyll a
CMEMS	Copernicus Marine Environment Monitoring Service
CO_2	Carbon dioxide
DOC	Dissolved Organic Carbon
EO	Earth Observation
ERGOM	Ecological Regional Ocean Model
HELCOM	Baltic Marine Environment Protection Commission
ICES	International Council for the Exploration of the Sea
IOP	Inherent Optical Properties
MERIS	Medium Resolution Imaging Spectrometer
MSI	MultiSpectral Instrument
NIR	Near Infrared
NN	Neural Network
OLCI	Ocean and Land Color Imager
OC	Ocean Color
pCO_2	Partial pressure of carbon dioxide
POC	Particulate Organic Carbon
POLYMER	POLYnomial based algorithm applied to MERIS
S2	Sentinel-2
S3	Sentinel-3
SAG	Scientific Advisory Group
TOA	Top of Atmosphere
TOC	Total Organic Carbon
TSM	Total Suspended Matter

1 Introduction

In WP5 of the ESA Baltic+ SeaLaBio project, the objective was to revisit all steps of the development work and to summarize the developments, consolidate the lessons learnt and to identify next developments steps and main road ahead with respect to improved carbon cycle understanding. The results and conclusions have been communicated to the SeaLaBio Scientific Advisory Group and other stakeholders in the Baltic community, through user consultation meetings and conference presentations, and their feedback have been included in the lists of proposed future work and statements regarding the main scientific roadmap for further research.

The Baltic+ SeaLaBio project aim:

Develop methods for assessing carbon dynamics and eutrophication in the Baltic Sea through integrated use of EO, models and ground based data.

Targeting the following research questions:

"Can we quantify the carbon flux from land to sea with Sentinel-3 and Sentinel-2 data in the Baltic Sea region? If not, what are the main obstacles and potential solutions to be addressed in the future?"

2 Main scientific challenges

Three scientific fields have been in focus for the SeaLaBio project and further development in these areas is expected to contribute to the carbon cycle analysis.

Atmospheric correction (AC) is the first step to estimate water quality from satellite data. The goal is to remove the effects of scattering and absorption by atmospheric molecules, aerosols and gasses from the signal and to correct for measurement geometry (solar and viewing angles). The result of this process is the normalised marine reflectance at various spectral bands in the visible and near infrared (NIR) domain. AC is the most critical step in the sense that if it fails, the estimation of water absorption and scattering is not possible even with a perfect in-water algorithm. In practice, even after many years of research and development in ocean colour, AC still remains a main issue in optical remote-sensing over complex waters. The reason is the requirement of a very low error of the AC in order to be able to retrieve water constituents, the variability of aerosols in the atmosphere (unknown concentration and type) and the relatively weak contribution of the marine signal to the total Top-Of-Atmosphere (TOA) radiometry. This is most amplified in the Baltic Sea as it is a land enclosed basin with aerosols varying between maritime and rural types of different optical depths, and a very low water signal due to the high CDOM absorption and at times low TSM. Hence, the key to success for a good retrieval of the chlorophyll concentration and CDOM absorption, as required to assess eutrophication and carbon dynamics in the Baltic area (coastal and inland waters), is a robust atmospheric correction visible and NIR part of the spectrum.

In-water inversion is the second step in the satellite data processing. It converts the water leaving reflectance into information about the absorption and scattering properties of water and the substances suspended or dissolved in it. The absorption and scattering properties can then be converted into concentrations if the conversion factors – which can differ between locations and seasons – are known. One of the problems in assessing the performance of the in-water retrieval algorithms is that the poor performance (i.e. that the processed result does not match the in-situ values) can be due to the in-water method itself, or the errors introduced in the atmospheric correction part. Thus, it is not straightforward to make conclusions about the suitability of the algorithm without simultaneously assessing the performance of the AC part.

The processes and soil affecting the water properties in the lakes and rivers have a strong influence on the coastal waters of the Baltic Sea. Thus, accurate quantification of these riverine fluxes is prerequisite to **model the carbon balance of the Baltic Sea**, and to determine its role as a sink or a source of atmospheric carbon. If significant correlation between CDOM and carbon loads exists, CDOM concentrations derived from EO data in riverine waters could allow for a quantification of carbon loads, and support the improvement of boundary conditions for biogeochemical (BGC) models in the future. EO based estimations of light attenuation can provide useful information for the calibration of radiation models used in BGC modelling, in addition to use in other applications such as vegetation growth limitation and fish ecology evaluation.

3 Baltic+ SeaLaBio project results

With specific focus on the Baltic Sea and Copernicus Sentinel-2 and Sentinel-3 data, the project had the ambition to advance the current state of the art in three important scientific themes:

- Atmospheric correction
- In-water retrievals
- Modelling of biogeochemical linkages between land and sea

All three support the overarching aim to better assess carbon dynamics and eutrophication in the Baltic Sea through integrated use of Earth Observation, models and ground-based data.

The main results, conclusions and recommendations for future work for each topic are summarized in chapter 3.1-3.2.2.2 below. In addition, a chapter with observations related to available in situ data in the Baltic Sea region has been included (Ch. 3.4).

3.1 Atmospheric correction

The prevailing uncertainty requirement for atmospherically corrected Sentinel-3 OLCI data over open waters is defined as 5.0 10^{-4}, in absolute value for the derived marine reflectance (Donlon, 2011), and this is achievable with standard AC routines over such waters. This number corresponds to 1.6 10^{-4} sr^{-1} in remote sensing reflectance, which seems to be relevant over the Baltic Sea, too, in view of retrieving the proper order of magnitude of the signal. However, this requirement, if applied to Baltic Sea data, needs much more robust AC due to the complexity of the signal over the full spectrum. Available methods generate relatively good results in open and less complex waters, but do not perform well in the most challenging optical conditions near the coast and towards the easternmost and northernmost parts of the Baltic Sea and need to be improved.

3.1.1 Main results

The scientific review presented in the Baltic+ Requirements Baseline (Attila et al., 2019) concluded that “spectral optimization” and “Neural Network” approaches, i.e. two types of ACs, are complementary and that there is an added value to combine both and benefit from the strengths of each approach. POLYMER (spectral optimization, Steinmetz et al., 2011) and C2RCC (Neural Network, Brockmann et al., 2016)) were identified as the currently best performing methods for Sentinel-2 and Sentinel-3 and it was decided to use these as starting points for the development of a dedicated Baltic+ Atmospheric Correction scheme. The marine model of C2RCC aggregates a large knowledge of natural variability, with realistic covariance in the IOPs (Inherent Optical Properties), thanks to training on simulated dataset based on the relevant bio-optical relationships. After the NN is trained, it is also very fast to process a whole scene at once. The spectral matching technique of POLYMER is known to be robust over many atmospheric conditions and perturbations (such as thin clouds, residual sun glint, adjacency effects), while its marine model with two components is not expected to work optimally over absorbing or very turbid waters. The key components and logic of the Baltic+ AC processor are listed below and the full description can be found in the Algorithm Theoretical Basis Document (Mazeran et al., 2020).

- The directional marine reflectance model is given as a function of five IOPs and three angles by a forward NN, specifically trained for conditions found over the Baltic.
- For a given combination of IOPs, the marine reflectance is computed, removed from the total signal, and the best atmospheric path reflectance is searched based on Polymer approach (spectral fitting).

- The discrepancy between the modeled marine reflectance and the one retrieved by removing the atmosphere is computed in a least-square sense and minimized with an optimization method on the five IOPs (Nelder-Mead simplex).
- The least-square minimization allows to rigorously compute the uncertainty of the retrieved IOPs at convergence, hence subsequently of the atmosphere and final marine reflectance, on a per-pixel basis.
- The directional marine reflectance is eventually normalised for Bidirectional Reflectance Distribution Function (BRDF) effects, as modelled with the forward NN.

Various analyses have been performed to justify the relevance of the developed Baltic+ AC scheme (see Validation Report for more details):

- Qualitative analysis of marine spectra and uncertainties on scenes
- Quantitative validation of marine spectra against in situ measurements
- Indirect validation of marine spectra through validation of CDOM-derived product against in situ measurements.

The processor has been applied primarily to OLCI 3A and 3B scenes covering various water conditions, i.e. very absorbing waters in the Northern end of Gulf of Bothnia, brighter waters and turbid plumes near the estuary of the Kokemäenjoki river and blue-green algae blooms in the Archipelago Sea and compared to POLYMER (Steinmetz et al., 2011), C2RCC (Brockmann et al., 2016) and standard (IPF) AC (Antoine and Morel, 1999) results. All results have been contrasted with in situ data (reflectances) from three AERONET-OC sites relevant for Baltic Sea waters. Looking at the amount of valid data generated, the fit in absolute levels and scattering of data, Baltic+ and POLYMER ACs generate best results, with advantage for Baltic+ at blue wavelengths and POLYMER at green-red wavelengths. In addition, the Baltic+ AC manages to provide a smooth map of reflectance with realistic amplitude. The indirect validation on CDOM product, after applying band ratio algorithm (see section3.2.1), has shown best result for the Baltic+ processor, compared to C2RCC, and has justified to use this processor for the S3 data production in the project.

The uncertainty map provided by the Baltic+ processor gives good insight in the trustworthiness of the results. Very high values appear in complex areas, like in the most turbid part of the river plume. Interestingly, the uncertainty is not systematically higher over undetected clouds (except very limited pixels), which shows that the correction is as much trustable here as over other clear-sky adjacent pixels, likely thanks to the POLYMER atmospheric model. In general, the uncertainty level follows the amplitude of the reflectance, typically with higher uncertainty in turbid waters, yet some cases of algal bloom present a remarkable low uncertainty. The uncertainty maps hence suggest trophic conditions where the model is best suited and where it could be improved.

In terms of technical developments, the Baltic+ AC is coded in Python, using the SNAP API for Python (snappy module). The input/output operations on S3 and S2 data, as well as the NN operators, are thus handled in a generic way. The algorithm can process both raster data (image) and text extractions (match-up data files). The NNs are handled with the TensorFlow framework. The source code is available on the public repository https://github.com/bcdev/baltic-scripts.

The adaptation of the promising S3-OLCI Baltic+ atmospheric correction to S2-MSI has been done, but is more challenging due to the different spectral properties and limited number of bands for S2. Validation of the results have been initiated with spectral evaluations and comparison of S3 and S2 reflectance outputs.

3.1.2 Future work

3.1.2.1 Technical improvements

- The implemented uncertainty estimate is a valuable metric to investigate the relevance of the marine model and the spectral fit and should, ideally, be used in the AC minimization, to constrain the bands with larger uncertainty. The current implementation technically allows for this capability but requires estimates of the forward model uncertainties, for a given set of optical properties (IOPs). Constraining the inversion with input uncertainties is expected to reduce the amplitude of the output uncertainty.

- The correction for NO2 in the pre-processing steps of the Baltic+ AC is based on climatology and should ideally be improved with concurrent NO2 data (see Tzortziou et al., 2018). To our knowledge, this is a general problem for any ocean colour processing, not solved to date. Using satellite observations of atmospheric NO2 (e.g., TEMPO, TROPOMI, GEMS, Sentinel-4, Sentinel-5) has been very recently proposed but never tested so far. Alternatively, the variability of NO2 could be added in the uncertainty formalism.
- The S3 OLCI 709 nm band is important for aquatic applications, but it can disturb the atmospheric correction due to strong influence of water vapour at these wavelengths. Hence, the band at 709 nm is not included in the aerosol estimation in the Baltic+ AC, but it is finally atmospherically corrected and included in the output of corrected reflectance bands. The use in the AC, with dedicated gaseous correction, would be inspected.
- Sun glint correction neglects effects of aerosols in the direct transmittance. This correction could be moved to the iterative retrieval of aerosol for improved performance.
- There are no quality flags implemented within the atmospheric correction processor. However, the neural network approach allows for simple definitions of quality flags. The input data could be checked against the range of the training datasets and only if the data values are inside the defined range, the NN can produce sensible results. In order to allow for meaningful output, the input data should be constrained to these ranges.
- The algorithm delivered to ESA is essentially based on a forward NN. Another version of the algorithm was proposed and tested in the early part of the project, using a backward NN for the IOP inversion and an overall minimization over the aerosol unknown. A crucial assumption of this approach was that the forward and backward NNs should be reciprocal (i.e. applying successively the backward and forward NNs to a given spectrum should give back this spectrum, at least on the training dataset). To that end, the training of a new backward NN was based on a minimisation of reflectance spectra using the newly trained forward NN. Results were not conclusive. Investigation of a so-called invertible neural network has been stopped for the moment, but this kind of network architecture might still be a candidate for future developments.

3.1.2.2 Validation

- As stated earlier, the S3 OLCI 709 nm band is important for aquatic applications and the performance of the atmospheric correction on this band needs further validation.
- Repetition of published studies with the new S3-OLCI Baltic+ processors would help to demonstrate its interest to the EO community as well as to identify further improvements.

3.1.2.3 Operationalisation

- The Baltic+ AC has been implemented as a plug-in to the ESA SNAP toolbox. This means that it will be available to the EO community in a more user-friendly way than the source code today delivered on github. However, the bio-optical model used in the Baltic+ AC is based on optical properties (SIOPs) and ranges prevailing in the Baltic Sea. Hence, the resulting forward Neural Net could work in waters which have similar characteristics (strong absorption, low scattering). For other cases (water types) a new model and a new ForwardNN is needed.
- The Baltic+ AC for S3 has shown the best results for CDOM product and encouraging results in general for the marine reflectance over the Baltic Sea. As such, it is a potential alternative for processing of data provided to the Baltic community via the Copernicus Marine

environment monitoring service (CMEMS). However, the processing speed is currently a concern for an operational use.

- Reprocessing historical data and thereby extending the available time series is considered valuable by the community, e.g. to support status and long-term trend assessments related to different European directives. Hence adaptation of the S3 OLCI Baltic+ processor to MERIS sensor would be required.

3.2 In-water retrieval

3.2.1 Main results

A Baltic Sea specific bio-optical model was developed to facilitate the development of a new in-water inversion algorithm. The model is based on parameter ranges (CDOM, Chl-a, TSM) determined according to data collected within Finnish, Swedish and German national and regional monitoring programs and campaigns. The model was used to create a simulated dataset which then was used for training a new neural network for the inversion. In parallel to this effort, an empirical band ratio was calibrated with in situ data, for direct application to the Baltic+ AC reflectance output.

- The Forward Neural net, going from IOPs (i.e. absorption and backscattering properties of the water constituents) to reflectance, has been trained with good results and proved improvements compared to earlier versions. A significant effort was put on training of a backwardNN to properly derive IOPs and concentrations from EO reflectances. However, the results do not reach the level of the band ratio method. Neural Network algorithms typically utilise all available bands which can make them more susceptible to errors in the AC. A test with a reduced number of bands did not lead to improvements.
- A band ratio algorithm was developed for CDOM estimation based on Baltic+ AC Sentinel-3 OLCI band 8 (665 nm) and band 6 (556) nm reflectances. The band ratio used wavelengths in the red and green part of the spectrum, where the atmospheric correction is less prone to errors. The band ratio algorithm was applied to satellite data and the results have been compared to aCDOM (400 nm) m^{-1} measurements collected in the Baltic Proper, the bay of Bothnia, the Bothnian Sea, the eastern Gulf of Finland and the Kokemaenjoki estuary with good results. The validation results show that the band ratio estimation accuracy is better than with the other current processors. Thus, despite its simplicity it is a valuable method for carbon studies in the Baltic Sea.

3.2.2 Future work

3.2.2.1 Technical improvements

- No Neural Network based inversion algorithm was established.

3.2.2.2 Validation

- Validation of the developed band ratio algorithm for more CDOM rich waters (i.e. aCDOM(400nm?) > 7 m^{-1}). This could be tested in lakes, where there are more in situ data available. If the range of applicability for this band ratio algorithm is limited, the backwardNN will hopefully improve.
- Validation of Chl-a absorption (and concentrations) estimations and its implication for the aCDOM band ratio performance.

3.2.2.3 Other

- The results from the backwardNN should be explored and will hopefully provide good CDOM estimates. An option could be to adapt the “old” Boreal processor developed for MERIS data, which has provided good results with respect to CDOM estimates in the Baltic Sea, to Sentinel-3 sensor specifics.

- CDOM band ratio algorithms developed and tested within SeaLaBio can support the investigation of different TOC-DOC-CDOM relationships and conditions in different geographic regions as well as different temporal/seasonal patterns.

3.3 Biogeochemical modelling

The BGC-model ERGOM (**E**cological **ReG**ional **O**cean **M**odel) used in the SeaLaBio project is the biogeochemical part of a 3-dimensional ecosystem model of the Baltic Sea developed by IOW. The code and the development tool CGT (code generation tool) is available from www.ergom.net.The hydrodynamic part of the model is based on MOM5.1 (www.gfdl.noaa.gov/mom-ocean-model). MOM is a widely used model for global and regional applications. Main inputs to the ERGOM model are meteorological forcing (i.e. dynamical downscaling of NCEP/NCAR reanalysis, Geyer and Rockel, 2013) and river runoff (derived from HELCOM data and assessments, Gustafsson et. al., 2012). The model simulates the marine nitrogen, phosphorous, and carbon cycles.

3.3.1 Main results

Main modifications to ERGOM during the SeaLaBio project consist of:

1. New improved model setup with higher horizontal resolution
2. Implementation of a CDOM state variable. This allows CDOM values of river water to be used as input data in addition to river runoff.
3. Introduction of photo bleaching of CDOM

The main results from the ERGOM modifications include:

- Improved representation of small-scale features by applying a higher horizontal model resolution, i.e. higher spatial resolution of output results (1 nautical mile vs 3 nautical miles) for data from 2017-2019.
    - Strong coast-open sea gradients are reproduced more realistically and improve the model performance in coastal areas.
    - Results based with 1 n.m. resolution show that fine structures resolved in simulated CDOM absorption can be clearly seen.
    - Added value is that the model results become more comparable with EO products, which is important for model validation and calibration with EO product.
    - The new approach allows for the estimate of land-sea carbon fluxes due to CDOM. Results showed a strong annual cycle that is caused by both the runoff cycle and the annual CDOM concentration (absorption) cycle.
    - The changed resolution made a re-calibration of sub-grid parametrizations (e.g. turbulence) necessary. Model performance for temperature and salinity at two stations in the northern Baltic, Bothnian Bay, were reasonably well reproduced by the model. However, bottom water temperature in winter was overestimated. The reasons underlying this were not evident. Salinity observations indicated a descending trend that was not reproduced by the model; the cause for this is unknown.
- Uncertainties in CDOM estimates owing to an absorption – salinity relationship have been eliminated by introducing an explicit CDOM state variable.
    - The CDOM state variable considers CDOM decay and uncertainties in the model error for salinity. For CDOM decay, “photobleaching” was implemented. A series of calibration simulations for the constant r0 have been performed and a reasonable r0 value is available.
    - A comprehensive dataset for riverine CDOM (n=80) from EO data that are used as CDOM-forcing data (river loads) has been prepared. The data provide essential boundary conditions for the CDOM state variable in the river mouths.
- The implementation of the EO based CDOM state variable improved model results.
    - The model input data was generated by deriving monthly mean CDOM values (2016-2019) from EO data (MSI-SYKE, Sentinel-2 and C2RCC) at 69 estuaries representing ERGOM load points around the Baltic Sea. Comparison of these CDOM values with

in situ data from a Flow-Through device (FT) and data from water samples (WS) show good correspondence.

- The previous version of the model (using salinity to estimate CDOM) produced R^2 = 0.16 when compared to in situ aCDOM values from monitoring stations in the Northern Baltic (n=578). This increased to R^2 = 0.61 when ERGOM simulated aCDOM estimates using the 75th percentile EO values was compared to in situ aCDOM . The same R^2 = (0.61) was achieved using the 95th percentile EO values albeit with slightly higher RMSE.
- Based on the results it is obvious that the BGC model benefits strongly from EO derived CDOM forcing data. The model can now provide more reliable estimates of light attenuation in water, which potentially provides more realistic simulations of several other state variables. This has consequences especially in the norther parts of the Baltic Sea where CDOM has a large effect on water transparency.

- The use of the EO based river aCDOM values and river runoff allowed us to estimate the TOC loading of eight major rivers in the Baltic Sea. The results show that when the EO data extraction area represents well the river water, the results are close to the estimate based on in situ measurements. In five of the eight rivers there were no in situ-based values available. Based on these results EO can be used to fill the gap in the reporting.

3.3.2 Future work

3.3.2.1 Technical improvements

- Add TOC and DOC components to improve ERGOM (requires proper conversion from CDOM to DOC)
  - Investigate conditions in different geographic regions as well as different temporal/seasonal patterns
  - Requires more in situ measurements of TOC and CDOM in different river types
- Consider effects of CDOM decay and dilution (bacterial breakdown process and effect on CDOM concentration is probably important to take into account in addition to the photochemical degradation)
- Recalibration of ERGOM internally dependent processes after introducing CDOM as boundary condition instead of salinity.
- Recalibration of ERGOM dependent processes, e.g. better representation of light climate (CDOM). The changed and improved PAR climate requires a revision of the parametrization of the impacted biogeochemical processes, starting with primary production.
- In general, review BGC model requirements for improvements possible by EO.
  - E.g. the EO based CDOM forcing data provision allows more rivers to be included in the model. This would lead to better characteristics of CDOM concentrations in coastal areas.
  - The 1 nm horizontal resolution might be too coarse for some river gradients when comparing EO results (e.g. in test areas). An even finer resolution would be a good improvement.
- Owing to the high computational effort, test, calibration, and scenario simulations require HPC facilities.

3.3.2.2 Validation

- Collect pCO_2 values from other sources, e.g. buoys and ships, and be used as validation
- Other EO based products (e.g. Chl a) could be used to validate the ERGOM model performance. Identification of deviations between model results and EO data could point to potential areas for improvement. A mismatch analysis between e.g. modelled Chl a and EO observations of Chl a concentration can provide better knowledge regarding potential areas for model improvement.

3.3.2.3 Other

- More basic research related to terrestrial TOC vs. DOC vs. CDOM relationships is needed to allow the EO based TOC loading monitoring to be utilized.
- Is the decrease of terrestrial CDOM mainly due to respiration (CO_2) or is it breaking down to other carbon components due to the light environment?
    - Respiration is temperature dependent. Can it be tuned to measured quantities? Better information can help constrain the BGC model.
    - CDOM is breaking down because of light climate. There is less PP when there is more CDOM.
      Can EO products be used to study these relationships?
    - Importance of autochthonous CDOM sources

3.4 In situ data

In situ data from April-October, 2015–2018, from several test areas, was collected and used for the project developments. The selection of test areas was made to cover different geographical areas and water quality characteristics and with respect to the availability of in situ data. The data consisted of water quality data from routinely sampled monitoring stations in coastal regions and open Baltic Sea areas, from measurement campaigns related to EO validation and from Ship-of-opportunity (bottle samples and continuous instrument data). In addition, reflectance data from AERONET-OC stations and RFEX Ship-of-opportunity was included. A full description of the test areas and data can be found in the Requirement Baseline document and the Dataset Description Document (Kallio et. al., 2019).

3.4.1 Observations

With respect to the collected in situ data set, some comments were made during the consultation events:

- In general, all recommendations on seasonal, spatial and temporal strategies are valuable, as well as, information on what variables are needed to evolve on-going EO related developments. Are more observations from the open sea a strong contribution or is an emphasis on coastal zones more urgent?
- Error bars/uncertainties related to the EO NN based estimations can be provided, but should also be further developed in relation to the in situ data. It was confirmed by several participants that error budgets for in situ data is indeed needed and work is in progress.
- The handling and reporting of nominal sampling coordinates vs. actual station sampling coordinates needs to be improved, as well as provision of correct time stamps for the sample, to support validation efforts.

Measurement uncertainties and determination methods of the laboratory measurements were not available in the ICES databases. In Finland, measurement uncertainties (for the recent years) and determination methods (for all measurements) are available in the national water quality database. In the Swedish national database this information is available for some measurements, but the way they are expressed differs from the Finnish system. The availability of the measurement uncertainties and determination methods should be discussed in the HELCOM and ICES communities.

3.4.2 Recommendations

The availability of appropriate in situ data is important for comparative analysis with EO data, for definition of areas appropriate for EO based monitoring and where further development is needed to obtain high quality estimates. This availability requires sampling efforts as well as definition of optimal sampling and determination methods for this objective. Timeliness and frequency are not fully considered in HELCOM monitoring as the assessment of the ecological status of the Baltic Sea, as well as of coastal water bodies, is based on a few summer months. During the SeaLaBio consultations, requests to provide guidance to e.g. national authorities regarding in situ data were expressed. This concerned:

- Parameters to be measured and related methodology
- Identification of regions with poor availability of in situ data.
- Location of (new) sampling stations for improved representation of the diversity of the Baltic Sea water types
- Guidance on timeliness and frequency, e.g. related to individual river flux data (capturing meteorological events causing high river discharge, such as heavy rains, snow melts in spring)

With respect to different parameters, measurements of NAP (non-algal particles) were mentioned to better define the contribution from land, resolvability and conversion of EO based a_{NAP} to carbon. More measurement of aCDOM and DOC are needed for the validation of the EO products and to calculate organic carbon from the EO based aCDOM. Simultaneous measurements of TOC and DOC would make possible the division of TOC to dissolved (DOC) and particulate organic carbon (POC). These measurements would also be valuable in the river mouths for the needs of estimating DOC/TOC fluxes via EO.

This availability requires sampling efforts as well as definition of optimal sampling and determination methods for this objective. With respect to eutrophication monitoring and measures HELCOM have issued "Guidelines for monitoring of chlorophyll a" with recommendations on how to sample and methodologies for laboratory determinations. Other HELCOM guidelines directly related to the variables that can be estimated by EO are: Guidelines for measuring turbidity and Guidelines for measuring Secchi depth. All manuals can be downloaded from https://helcom.fi/action-areas/monitoring-and-assessment/monitoring-guidelines/. The DOC and POC determination methods are described in the Combine Manual as technical notes. In addition, "Guidelines for monitoring of chlorophyll a via EO" has been published. These guidelines presently focused on ENVISAT MERIS data and should be updated with respect to Sentinel-3 OLCI. Potentially with strong contribution from the SeaLaBio project results. However, the in situ related guidelines are not made with reference to comparative analysis with EO data and a modified annex could be written with specific considerations for this purpose. In addition, similar guidelines for measuring aCDOM and Total Suspended Matter (TSM) etc. should be issued and presented to the community with the recommendation to include them in the monitoring programmes of the Baltic Sea.

A recommendation to ESA to support optical campaigns was raised, but such efforts are difficult to know the benefit from due to potential cloudiness and limited temporal and spatial extent. The increased use of fixed buoys and automatic stations was lifted and there are already a number of instruments available in the Baltic Sea. Such stations provide a good knowledge about regional, seasonal and diurnal changes. The diurnal variability provides a good support in understanding potential discrepancies between EO and in situ data if collected at different times during the day. These stations should be placed so that also non-complex situations are covered. Also gliders can be used, but both automatic station instruments and gliders need to be calibrated to provide good data. For open waters, an increased use of data from research vessels such as RV Aranda, RV Svea and RV Oceania, and "ships of opportunity" was recommended.

4 Carbon cycle understanding

On the global scale, there are large gaps in current coverage of environmental observations along the aquatic continuum. Rivers and near-shore coastal regions around river plumes are often left out of carbon budgets due to a combination of methodological constraints and poor data coverage. Closing these gaps could potentially alter global estimates of CO_2 outgassing from surface waters to the atmosphere by several-fold (Ward et. al, 2017). In addition, marginal seas like the Baltic Sea represent a dynamic and to large extent still highly uncertain component of the global carbon cycle. The often large temporal and spatial variations of sea-surface partial pressure of carbon dioxide (pCO_2) in such areas are driven by multiple complex mechanisms (Shuping et. al., manuscript). In order to identify and embrace uncertainties in global carbon budget estimations it is important to further adopt statistical and modeling approaches in these regions.

With respect to EO data, the SeaLaBio project has concentrated on improving the estimation of colored dissolved organic matter (CDOM) in the Baltic Sea to support the overarching carbon cycle analysis and understanding. To accomplish this, a new atmospheric correction (AC) algorithm has been developed. The Baltic+ AC and additional in-water estimation methods have improved CDOM estimation accuracy significantly. Other water quality parameters have not yet been validated and high concentrations of CDOM have caused challenges for the estimation of other parameters such as Chl-a in the past. However, the developments within SeaLaBio are expected to improve the reliability of Chl-a estimations especially in coastal regions with high CDOM. This in turn is also expected to lead to more extensive information about processes such as eutrophication along the vulnerable coastline of the Baltic Sea.

Data fusion was shown to be useful for the synergetic use of S2-MSI and S3-OLCI data. The high resolution of S2 allows estimations closer to shore while OLCI has daily coverage (cloud cover permitting). The method was also able to fill gaps caused by shallow areas masked in the EO results. The quality of the data fusion result depends on the quality of the input data and as discussed above, more can be done to further increase this level. Furthermore, the optimal parameterization of the fusion method (uncertainties, correlation lengths etc.) must be estimated.

We demonstrated that EO data can be used to estimate annual TOC loads from rivers to the Baltic Sea. The method still requires refinements and additional in situ data to confirm the conversion factors between CDOM and TOC in different areas and seasons. Once these are done, EO can provide additional information about loading from rivers. This is especially valuable for rivers that are currently not reported according to the HELCOM recommendations.

The carbon modelling effort within SeaLaBio has been focused on land to sea fluxes of CDOM. It is a feasibility study demonstrating the added value of combining EO and modelling for increasing the predictive capability of ecosystem models. The impact of CDOM on the marine carbon cycle is mainly indirect due to controlling sub-surface PAR. As indicated above, this project also showed that other EO based products of importance for the land-sea carbon flux, e.g. TOC and DOC, can be produced and used in models.

Air-sea fluxes (pCO_2 estimation) were not in focus for this Baltic+ theme but is of course important for carbon cycle understanding. Presently, most pCO_2-studies are based on in situ measurements (Löffler et al., 2012, Rutgersson et al., 2020 and Wesslander, 2011), but some studies (Parard et.al., 2016, Shuping et. al., manuscript) have estimated the sea surface pCO2 for the Baltic Sea using remote sensing supported approaches. Promising results were obtained and indicate that EO data also can contribute to this field of carbon cycle research.

There is a lot of uncertainty relating to the air-sea CO_2 exchange, especially in the coastal regions (river mouths, river plumes), and remote sensing using EO data with high spatial resolution should be able to contribute to better understanding in these areas.

5 Future steps for SeaLaBio

To define a road map in its true sense requires a well-defined target in relation to a larger vision and identification of consecutive steps to reach the goals. Assessing carbon dynamics and eutrophication in the Baltic Sea through integrated use of EO, biogeochemical models and ground based data is a diverse target and multiple parallel steps are possible and needed to achieve significant advances in the understanding of carbon dynamics and eutrophication related processes. In this chapter we have summarized the project outcome and the small scale "road map" with direct connection to the SeaLaBio achievements. The developments done in the SeaLaBio project have advanced the state-of-the-art in three important fields:

1. **EO data processing:** A new method for **atmospheric correction** of satellite images – based on combining the advantages of Polymer & C2RCC – can now provide more reliable water leaving reflectance values. This is a major step towards the formulation of an optimal AC for

the Baltic Sea. The NN based algorithm developments did not lead to qualitative results for **in water retrieval**. However, a band ratio algorithm based on Baltic+ AC results was established and provided better aCDOM values than the other current processors.

2. **Biogeochemical modelling:** The ERGOM model can now utilize EO based aCDOM values as input data and, as a result, provide more reliable estimates of light attenuation in water, which potentially provides more realistic simulations of several other state variables. This has consequences especially in the norther parts of the Baltic Sea where CDOM has a large effect on light available for primary production.
3. **Use of EO for monitoring carbon fluxes:** EO based data can e.g. provide information about the Total Organic Carbon loads from rivers.

Further improvements in all fields are still possible and even required, and identified technical possibilities have been listed in each specific chapter above (3.1-3.4). More extensive and dedicated efforts with respect to carbon cycle analysis and understanding are related to:

- Baltic-wide relationships CDOM-TOC, CDOM-DOC, CDOM-POC in river outlets (terrestrial sources)
- Separation of different sources of CDOM in the sea, those from land and those from production in the sediments to be able to better trace the land derived CDOM in the sea to conclude about its transformation. Experimental studies could help here (different optical properties, different isotopic and chemical composition).
- EO based data sets for terrestrial loads of TOC, DOC, POC
- EO based data to help increase understanding of quantitative primary production (or ecosystem production), which currently largely fails, particularly in the Gulf of Bothnia where the classical approach of C:N:P ratios does not work.
- Dynamics of terrestrial organic carbon in marine environment (some studies exist, maybe sufficient as a start)
- Study impact of terrestrial organic carbon on marine carbon cycle by means of models
- Research related to air-sea fluxes (pCO_2 estimation) where there is large uncertainty especially in the coastal regions.
- New EO sensors suitable for water quality are coming (e.g. CHIME, FLEX) and may create possibilities for continuing the AC development. There was interest for this among the SAG and the team.

6 A vision for the Baltic Sea – the LEGO Baltic Sea Initiative

The vision of HELCOM, as presented in the Baltic Sea Action Plan, is "*a healthy Baltic Sea environment, with diverse biological components functioning in balance, resulting in good environmental/ecological status and supporting a wide range of sustainable human economic and social activities*". To reach this vision a joint effort by many actors is required and a need for environmental data that can support the work by a large variety of decision makers, experts, research groups, businesses, national/regional/local stakeholders and administrators. Several initiatives and R&D projects are already running so there is plenty of momentum and energy around, which if appropriately harnessed and streamed has great potential to push things forward. Hence, there is a need to build a holistic approach to studying, monitoring and managing of the Baltic Sea. Continued and new ESA initiatives can support linking of all stages, of the work of remote sensing experts, biogeochemical modelers and experimentalists (in situ observations), and generate synergy.

EO based environmental information should be an increasingly important source of information to support these efforts and it should be easily available for relevant work, projects and initiatives, also those not directly supported by EO experts.

CMEMS, the Copernicus Marine Environmental Monitoring Service, is currently producing and providing several EO based data sets, which can freely be used to support status and trend assessments and research and development initiatives in the Baltic region and globally. However, the spatial and temporal resolution is limited and the end user uptake of the data is still relatively low. From Q1 2021 and onwards, CMEMS will provide 100 m (Sentinel-2) daily and monthly coastal water quality products, which should be sufficient for many coastal applications, but they will not be produced for the open Baltic. The SeaLaBio project has contributed with a new atmospheric correction (AC) algorithm and a regional algorithm for estimation of colored dissolved organic matter (CDOM) in the Baltic Sea, to support the overarching carbon cycle analysis and understanding. The Baltic+ AC and additional in-water estimation method have improved CDOM estimation accuracy significantly. After further validation, the CDOM product could be added to a publicly available service portfolio and made accessible to the Baltic user community. However, despite the improved availability and accessibility of data and information, there are still gaps in terms of temporal and spatial resolutions, type of products and, very important, quality of products.

Besides the data limitations, it might also be difficult to find, understand, download and analyse the products for users with no or little experience in EO data. To address this, ESA has initiated activities in the application domains of; "Integrated Maritime and Territorial Spatial Planning" and "Emerging HELCOM Monitoring and Assessment Priorities". These projects aims to identify suitable environmental data (EO, in situ, models) and GIS materials, integrate, process and store thematic information and create the data access, visualization and analysis systems and tools. An information service will be created that provide the relevant EO and non-EO data together through contemporary interactive apps, or by linking the machine-to-machine interfaces to the GIS tools of the user organizations (e.g. HELCOM and national user organisations). It is also envisaged that identified gaps in product availability, in terms of spatial and temporal resolution of publicly available EO products, could be overcome and provided within the framework of these initiatives.

In addition to the EO based services described above, there are other initiatives focused on other types of environmental data. The HELCOM Baltic Data Flows (BDF) project, co-financed by the Connecting Europe Facility of the European Union, seeks to enhance the sharing and harmonisation of data on marine environment originating from existing sea monitoring programmes, and to move towards service-based data sharing. In particular, open datasets will be made available by HELCOM to a wider community, such as European open data ecosystem, researchers, NGOs and private sector, in order to benefit from the availability of harmonised environmental data. The project will run from October 2020 to October 2023 and efforts between BDF and coming ESA initiatives should be coordinated.

The collection of existing data is one thing, but for many applications, including development and quality assessment of existing and new EO products, additional data is needed. More and strategically located samples, especially for some parameters and regions, would be an important contribution to coming developments and assessments. Another added value related to an extended effort to collect appropriate in situ measurements for validation of EO products is the potential to increase confidence in and acceptance of EO products as appropriate to complement current methods of monitoring of water quality and ecological status.

WHAT CAN ESA SUPPORT WITH A NEW INITIATIVE?

The objective of the current ESA Baltic Regional Initiative is to enhance the use of state-of-the art European satellite missions to support definition and cooperative implementation of regional priorities in the Baltic, in line with national policy objectives in the context of the EU legislation, national legislation as well as the HELCOM framework. We propose to support:

- Definition of parameters and locations for complementary in situ data collection
- Validation of existing open products with publicly available in situ data sets and formulate quality/uncertainty labels for easy access by potential end users
- Further development of EO products that currently have low quality
- Development of new regional products if non-existing
- Review of the temporal and spatial resolutions of provided and developed products

- Production and provision of missing temporal and spatial products
- Discussions and active lobbying to explicitly include EO based environmental data in national and regional guidelines and directives in response to national legislations and European directives
- Devoted training of a critical mass, at all levels, within the Baltic Sea stakeholder group
- Formulation of a recurring revision program to keep products and processes up to date
- And in general, at all stages, to link the efforts by experimentalists, bio-geochemical modelers and the remote sensing community to increase synergy

Despite the many ongoing initiatives discussed above we foresee that there are still missing pieces, i.e. "EO building blocks" to support all R&D needs to fulfill the HELCOM vision. With the list of actions presented above we propose to create the LEGO Baltic Sea Initiative. The LEGO Baltic Sea Initiative stands for the Leveraged use of Earth and Ground Observations for Baltic Sea R&D Initiative. To utilize Earth System Science (ESS), there needs to be possibilities to choose and combine alternative input data, to test different approaches, especially as new findings and knowledge become available. When data from both ground, space and models are collated and combined appropriately, information can be produced that make it possible to take different and new steps to ameliorate undesirable developments, monitor status and trends, and to reach good environmental status. Appropriate "building blocks" of data and information to support different applications and changing end user needs is therefore essential. This is the reason we suggest moving towards a "LEGO" based approach.

In general, products derived from EO data using different methods and algorithms, e.g. for chlorophyll estimation, can be used in models developed for different purposes that produce different output values. No one product or model will satisfy all needs, neither in space nor over time, nor will one solution be accepted by all parts of the Baltic community (political, science, business or the general public), partly because of traditional or institutional barriers. Spatial, temporal and thematic flexibility of information products is needed in combination with guidance sourced from well documented data and data products, and preferably endorsed by an appointed long-term Baltic panel of experts. Work by such a panel could also include establishment of standards (methods and products) and harmonization of metrics to ensure that appropriate comparisons between member states and regions can be done reliably. Consequently, in addition to the actions presented above, the LEGO initiative should bring together experts in monitoring and in situ, EO and modelling, as well as, representatives from e.g. HELCOM working groups and projects, to regularly produce new and revised guidelines to support WFD and MSFD and HELCOM Action Plan revision cycles.

As mentioned above, some areas with missing components could, if filled, greatly advance developments towards the HELCOM vision and global ESS. These include carbon cycling and improved knowledge of the function, structure and health of both ecosystems and socioeconomic sectors. "Building blocks" procured through the LEGO concept could also support the creation of combined bio-geo-chemical models for the Baltic region that takes into account the atmosphere, the terrestrial and marine/freshwater components and utilizes all state of the art remotely sensed products at different spatial and temporal resolutions and in situ data becoming available at a pan-Baltic level as earlier discussed within SeaLaBio. In a broader perspective, outputs from such models could enhance the global carbon budget models and support global monitoring of climate change, adaptation and mitigation efforts.

References

Antoine, D. and A. Morel (1999). A multiple scattering algorithm for atmospheric correction of remotely-sensed ocean colour (MERIS instrument): principle and implementation for atmospheres carrying various aerosols including absorbing ones, IJRS, 20, 1875-1916.

Attila, J., Koponen, S., Kallio, K., Lebreton, C., Brockmann, C., Müller, D., Mazeran, C., Philipson, P., Neumann, T. (2019). Baltic+ Sealabio Requirement Baseline, v2.0, April 2019.

Brockmann, C., Doerffer, R., Peters, M., Stelzer, K., Embacher, S., Ruescas, A. (2016). Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. European Space Agency, (Special Publication) ESA SP, SP-740

Donlon, C. (2011). Sentinel-3 Mission Requirements. Traceability Document. EOP-SM/2184, CD-cd, Issue 1.

EUMETSAT Recommendations for Sentinel-3 OLCI Ocean Colour product validations in comparison with in situ measurements – Matchup Protocols: http://www.eumetsat.int/website/wcm/idc/idcplg?IdcService=GET_FILE&dDocName=PDF_S3_OLCI_MATCHUP_PROTOCOLS_RECC&RevisionSelectionMethod=LatestReleased&Rendition=Web

Geyer, B. and Rockel, B.: coastDat-2 COSMO-CLM Atmospheric Reconstruction, https://doi.org/10.1594/WDCC/coastDat-2_COSMOCLM, https://doi.org/10.1594/WDCC/coastDat-2_COSMO-CLM, 2013.

Gustafsson, B. G., Schenk, F., Blenckner, T., Eilola, K., Meier, H. E. M., Müller-Karulis, B., Neumann, T., Ruoho-Airola, T., Savchuk, O. P., and Zorita, E.: Reconstructing the Development of Baltic Sea Eutrophication 1850–2006, AMBIO, 41, https://doi.org/10.1007/s13280-012-0318-x, https://doi.org/10.1007/s13280-012-0318-x, G02023, 2012.

HELCOM Monitoring Guidelines: https://helcom.fi/action-areas/monitoring-and-assessment/monitoring-guidelines/

Kallio, K., Attila, J., Koponen, S., Keto, V., Stelzer, K., Brockmann, C. and Philipson, P. (2019). Baltic+ Sealabio Dataset Description, v1.1, Aug 2019.

Koponen, K., Attila, J., Kallio, K., Väkevä, S., Mazeran, C., Müller, D., Brockmann, C., Lebreton, C., Philipson and P., Neumann, T. (2020). Baltic+ Sealabio Validation Report, v2.2, Aug 2020.

Löffler, A., Schneider, B., Perttilä, M., Rehder, G., 2012. Air-sea CO 2 exchange in the Gulf of Bothnia, Baltic Sea. Cont. Shelf Res. 37, 46–56. https://doi.org/10.1016/j.csr.2012.02.002

Mazeran, C., Müller, D., Lebreton, C., Brockmann, C., Koponen, S., Attila, J., Kallio, K., Väkevä, S., Neumann, T. (2020). Baltic+ Sealabio Algorithm Theoretical Document Basis, v2.2, June 2020.

Parard, G., Charantonis, A.A., Rutgersson, A., 2016. Using satellite data to estimate partial pressure of CO2 in the Baltic Sea. J. Geophys. Res. G Biogeosciences 121, 1002–1015. https://doi.org/10.1002/2015JG003064

Rutgersson, A., Pettersson, H., Nilsson, E., Bergström, H., Wallin, M.B., Nilsson, E.D., Sahlée, E., Wu, L., Mårtensson, E.M., 2020. Using land-based stations for air–sea interaction studies. Tellus A Dyn. Meteorol. Oceanogr. 72, 1–23. https://doi.org/10.1080/16000870.2019.1697601

Shuping, Z., Rutgersson, A., Philipson, P., Wallin, M.B., Schneider, B. and Perttilä, M. (Manuscript) Remote sensing supported sea surface pCO2 estimation and variable analysis in the Baltic Sea.

Steinmetz, F., Deschamps, P.-Y., and Ramon, D. (2011). Atmospheric correction in presence of sun glint: application to MERIS. Optics Express, Vol. 19, Issue 10: 9783-9800.

Tzortziou, M. O. Parker, B. Lamb, J. R. Herman, L. Lamsal, R. Stauffer, N. Abuhassan (2018). Atmospheric trace gas (NO2 and O3) variability in South Korean coastal waters, and implications for remote Sensing of coastal ocean color dynamics. Remote Sensing, 10, 1587.

Wesslander, K., Hall, P., Hjalmarsson, S., Lefevre, D., Omstedt, A., Rutgersson, A., Sahlée, E., Tengberg, A., 2011. Observed carbon dioxide and oxygen dynamics in a Baltic Sea coastal region. J. Mar. Syst. 86, 1–9. https://doi.org/10.1016/j.jmarsys.2011.01.001