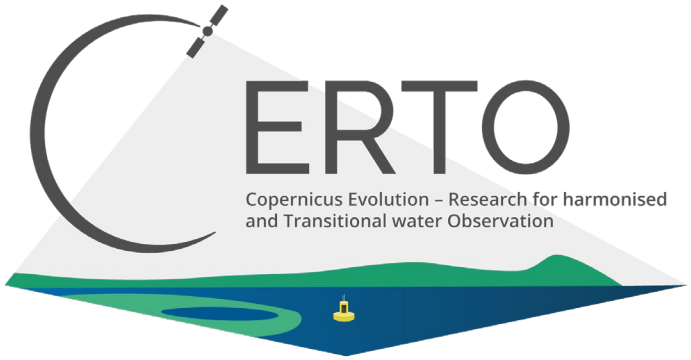

Horizon 2020 LC-SPACE-04-EO-2019-2020

Copernicus Evolution – Research for harmonised and Transitional-water Observation (CERTO)

Project Number: 870349

Deliverable No: D4.1		Work Package: 4	
Date:	08-JAN-2021	Contract delivery due date	31-DEC-2020
Title:	WP4 Optical water type classification Progress Report		
Lead Partner for Deliverable	PML		
Author(s):	Thomas Jackson, PML, Elizabeth Atwood, PML, Angus Laurenson, PML		
Dissemination level (PU=public, RE=restricted, CO=confidential)			PU
Report Status (DR = Draft, FI = FINAL)			DR

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme grant agreement N° 870349

Table of Contents

1 Glossary

C3S	Copernicus Climate Change Service
CERTO	Copernicus Evolution – Research for Transitional-water Observation
CLMS	Copernicus Land Monitoring Service
CMEMS	Copernicus Marine Environment Monitoring Service
c-means	Soft clustering scheme in which a data point can belong to more than one cluster
EO	Earth observation
ESA	European Space Agency
k-means	Hard clustering scheme in which each datapoint belongs to a single cluster.
OC-CCI	Ocean Colour Climate Change Initiative
OLCI	Ocean and Land Colour Instrument
OWT	Optical water type
PCA	Principal Component Analysis
R_{rs}	Remote sensing reflectance
WP	Work Package

2 Executive Summary

This short report covers the progress of WP4 (Optical water type classification) of the CERTO project within the first year. It contains sections that:

1) Summarise the findings of the review of the state of the art, giving suggestions as to what components need to be included in the final code package for release.
2) Review the current status of the codebase that will be released as a deliverable later in the project.
3) Give an overview of the optical water classes that have been created to date and how they were generated.

The report also contains a summary timeline of the developments/work that will take place between now and the end of the project in the form of a Gannt chart and associated text.

3 Introduction

This report covers the progress relating to CERTO work package 4: Optical water type classification, that has taken place within the first 12 months of the project. This progress has been slightly delayed due to the COVID-19 pandemic of 2020, but a significant amount of work has still been completed.

Over the course of the CERTO project, work package 4 aims to:

- Assess the current capabilities and limitations of optical water classification frameworks as utilised in the Copernicus services (C3S, CMEMS and CLMS).
- Research and improve the key steps of classification methodology: classification metrics, clustering approach, pre-processing and standardization of spectra, qualitative and quantitative masking of outputs, algorithm output blending, uncertainty assignment and overall algorithm optimisation.
- Create a multi-sensor optical classification scheme across complex coastal and transitional waters, allowing seamless switching between sensor inputs within the uncertainty limits defined by the user or downstream product requirements (e.g. indicators).

The main body of the report is split into three main sections. These sections (4, 5 and 6) review the state of the art and how this has impacted development of the CERTO code, outline the growing codebase that will later be a WP deliverable, and discuss initial cluster set generation respectively.

4 Components of an optical cluster set – state of the art

The division of waters into optical types is an established concept in marine sciences. Morel and Prieur (1977) distinguished two water types, those where bulk optical properties are dominated by phytoplankton and (Case-1) and those where bulk optical properties are uncoupled from phytoplankton (Case-2). With the introduction of multidimensional clustering techniques applied to remote sensing data (Moore et al 2009), this classification of waters has become non-binary with >10 optical water types (OWT) identified in both open-ocean (Jackson et al., 2017) and inland-water (Spyrakos et al., 2018).

Within the context of this document optical water class sets use fuzzy membership (c-means) rather than hard membership (k-means) classification schemes, meaning that classified spectra are assigned a membership between 0 and 1 for each class in the set.

Class sets are defined by a number of wavelengths (dimensions), with each class having a centre and domain. In 3-dimensional space this is easy to visualise as a centroid and volume. There is also the requirement to define a distance metric with which to compare a given spectrum to each class in the set. This will be discussed in section 4.1.1.

The task of reviewing all components of the clustering and classification methodologies is a significant undertaking and was the major component of Task 4.1. In order to be of operational use for global remote-sensing any techniques and approaches used must be robust enough for dealing with high volumes of multidimensional data. The generation of optical water class sets from remote sensing derived training datasets, which can contain millions of spectra, imposes computational limitations on the suite of clustering metrics that can be implemented. It is also important that the use of metrics for the creation and application of a class set is sufficiently automated and reproducible that it requires minimal expert input.

4.1.1 Similarity metrics between observations and type spectra

Clustering algorithms generally aim to partition data into groups (or clusters) such that similarity of data objects within each cluster is maximized while at the same time minimizing the similarity of data objects among clusters. Selection of a quantification method to measure how similar or dissimilar data objects are from one another is a fundamental step in this process. The fuzzy c-mean algorithm employed by Moore *et al.* (2001) and Jackson *et al.* (2017) to create optical water type classes from Ocean Colour data have relied upon the Euclidean distance metric during cluster formation optimization, with membership values being calculated after cluster formation using the Mahalanobis distance. Fuzzy c-means clustering on Euclidean distance has thus far proved to be the computationally lightest solution but limits the possible cluster shapes in multidimensional space, thus potentially hindering full optimization of data object cluster formation (Gueorguieva *et al.*, 2017). Fuzzy c-means clustering using other distance metrics have been proposed in recent years (Cebeci, 2020), and will be explored to test for improved algorithm performance. An overview of some of the most common distance metrics follows. Note that ‘cluster prototype’ is used to refer to test clusters formed during the cluster formation optimization process.

Euclidean distance

This metric is a special instance of the Minkowski distance. Setting p=2 in the following general equation:

$$D_{euc}(X, M) = \left(\sum_{i=1}^{n} |x_i - \mu_i|^p \right)^{1/p} \quad (1)$$

where $X = (x_1, x_2, \dots, x_n)$ is a data object in n-dimensional space (e.g. a given raster pixel of multispectral reflectance data with *n* bands or channels) and $M_c = (\mu_{c1}, \mu_{c2}, \dots, \mu_{cn})$ is the centroid of cluster prototype *c* formed during the optimization process and $D_{euc}(X, M)$ then represents the Euclidean distance (a simplified example would be the well-known Pythagorean distance between two points in a 2-dimensional space) between the data object and the cluster prototype centroid. Note that setting p=1 will provide the Manhattan (or city-block) distance, but this metric is not presented in any more detail here. Utilizing Euclidean distance while optimizing cluster formation inherently assumes hyperspherical cluster shapes (Gueorguieva *et al.*, 2017), which may not be the case across different water classes.

Mahalanobis distance

Here the metric can be understood as representing the distance between a data object and a cluster prototype centre in terms of standard deviations for that particular cluster prototype. Taking again $\vec{x} = (x_1, x_2, \dots, x_n)^T$ a data object in n-dimensional space (this time represented using vector notation) and $\overrightarrow{\mu_c} = (\mu_{c1}, \mu_{c2}, \dots, \mu_{cn})^T$ the center of cluster prototype *c*

$$D_{mah}(\vec{x}, \vec{\mu}) = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})} \quad (2)$$

where $S^{-1}$is the inverse of the covariance matrix

$$S = \begin{pmatrix} cov(x_1, x_1) & \cdots & cov(x_1, x_n) \\ \vdots & \ddots & \vdots \\ cov(x_n, x_1) & \cdots & cov(x_n, x_n) \end{pmatrix} \quad (3)$$

$D_{mah}(\vec{x}, \vec{\mu})$ then represents the mahalanobis distance. As mentioned above, an intuitive understanding of this metric is that similarity to the cluster prototype centre is represented in terms of standard deviation distance along each of the multidimensional axes. This allows for elliptical cluster shapes during cluster optimization, which may better capture large variations in certain reflectance bands for some nearshore or river water optical water type classes.

Angular distance (and the closely related cosine similarity metric)

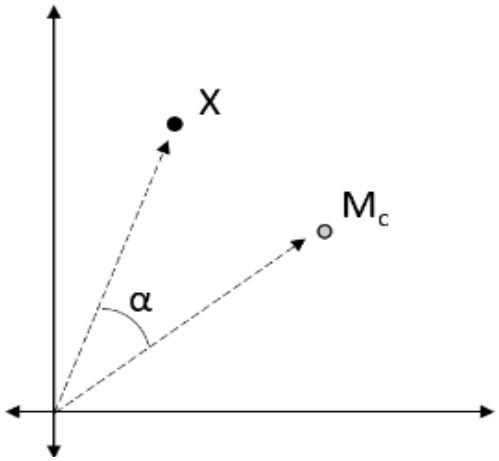

Figure 1: Diagrammatic illustration of angular distance in 2-dimensional space.

This metric is based on the angle formed between the multidimensional vectors of a data object and a cluster prototype centre. An example in 2-dimensions is shown in Figure 1, where $X = (x_1, x_2)$ is the data object, $M_c = (\mu_{c1}, \mu_{c2})$ the centroid of cluster prototype *c* formed during the optimization process and *α* the angle between these two vectors. Taking $X = (x_1, x_2, \dots, x_n)$ a data object in n-dimensional space and $M_c = (\mu_{c1}, \mu_{c2}, \dots, \mu_{cn})$ the cluster prototype *c* centre

$$D_{ang}(X, M) = 1 - \frac{\cos^{-1}(cosine\ similarity)}{\pi} \tag{4}$$

Where

$$cosine\ similarity(X, M) = \cos(\alpha) = \frac{\sum_{i=1}^{n} x_i \mu_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} \mu_i^2}} \tag{5}$$

$D_{ang}(X, M)$ then represents the angular distance as a proper distance metric that meets the same mathematical criteria for the other metrics listed here. This metric would have the benefit of allowing for even more flexibility in terms of possible cluster shape, since data objects lying along a similar multidimensional vector would be clustered similarly regardless of their individual vector magnitude (which may often be the case in more highly variable nearshore or river water reflectance signatures). While this flexibility is likely quite desirable, it does mean that there is a stage of information loss in the clustering (total reflectance, or integrated brightness) plays no role in the clustering.

4.1.2 Level of fuzziness used for classification

The level of fuzziness (m) used in the clustering of multidimensional data is a parameter that controls the relative weighting placed on the squared distance from each point to the centre of the cluster. The fuzziness level can be any value ≥1 where a value of 1 produces fully 'hard' clustering (k-means) and increasing fuzziness levels will degrade (blur, defocus) membership towards the fuzziest state. No theoretical or computational evidence distinguishes a single, optimal fuzziness level and recent studies have looked to optimize this parameter on an application specific basis (Bi et al 2019). The broad range of useful values for fuzziness seems to be 1 to 30 or so (Bezdek, 1984) and in the context of ocean-colour remote sensing it is considered to lie between 1.5 and 3.0 (Bi et al, 2019). Fundamentally the value used should be tight

enough to allow interpretation of different and distinct clusters while being fuzzy enough to allow coverage of the global optical diversity.

There has been some discussion of automatically/computationally optimising fuzziness factors within the genetics community (Dembele, 2008) and this was trialled in the water-colour realm by Bi et al. (2019) showing that for multiple inland Chinese water bodies, an m value of 1.36 gave a satisfactory set of seven distinct clusters. The code associated with this automated fuzziness determination has been made openly available and we aim to include this functionality within the CERTO system. As shown in section 5.1, the current code base can perform analyses with variable fuzziness factors, but the automated assignment of this value is not completed yet.

4.1.3 Spectral transformation

Modern remote sensing satellites such as OLCI have many wavebands in the visible range of the spectrum. One issue with c-means clustering is that it can become computationally prohibitive to cluster high dimensionality data for millions of pixels in search of a large (10 or more) clusters. One approach used to reduce the computational expense is to perform transformations to the R_{rs} spectra before clustering. One example of this is normalisation, which pushes the focus of the optical clustering to differentiate based on spectral shape rather than reflectance magnitude. Mélin et al. (2015) reduced the first order variability in reflectance to focus on the reflectance spectral shape by using the integrated value (i.e., the surface below the spectrum, Lubac & Loisel, 2007; Vantrepotte et al., 2012) following the formula (for each wavelength λ):

$$r_n(\lambda) = \frac{R_{rs}(\lambda)}{\int_{\lambda 1}^{\lambda 2} R_{rs}(\lambda)\, \delta\lambda} \qquad (6)$$

Where $r_n(\lambda)$ (in units of nm^{-1}) indicates the normalized spectrum obtained by integration between $\lambda 1$ (412 nm) and $\lambda 2$ (670 nm) computed by trapezoidal integration. The latest OC-CCI reprocessing version has moved from non-normalised (Jackson, et al 2017) to normalized optical waters classes (Jackson et al. 2020) for the global ocean (14 classes in this case) using six optical wavelengths. It has not yet been determined if this is suitable for data streams that are composed of many more wavelengths (such as the Oa1 to Oa11 bands of OLCI). We intend to allow normalisation to be an option within the CERTO classification scheme such that it can be turned on or off. This functionality has already been codified (using the spectral integral) and will be included in the final release.

One issue with normalisation is that there can be a loss of information at the normalisation stage. An alternative transformation would be to use PCA analysis to create a new set of orthogonal vectors (with a lower dimensionality than the R_{rs} bands) which can be used to cluster the data. The advantage of using the PCA transformation is that is minimises information loss (if a sufficient number of PCA components are included) and allows a relatively simple reverse transformation back to R_{rs} values. This transformation has also been codified and will be included in the final release.

4.1.4 Hybrid or hierarchical clustering methods

The use of hierarchical, partitional and hybrid clustering techniques present a possible solution to creating comparable optical water classes across different sensors. The use of a hierarchical scheme is currently being tested for implementation within the CMEMS processing chain for the Mediterranean waters. In previous processing versions the CMEMS Mediterranean processing chain has used a relatively simple portioning of waters into case I and case II with a chlorophyll-a algorithm designated for each case (Volpe et al., 2019). The new "CMEMS Mediterranean" approach follows Mélin et al. (2015) by using normalized by the integrated spectral values. The optical water class set consists of 6 classes (created using k-means clustering) with one class composed of complex water spectra characterized by the R_{rs} peak at 55x nm. The others five classes are considered open ocean waters for which the Max Band Ratio (MBR) approach is applicable. This technique shows excellent results discriminating complex waters from Case I waters, but leads to some misclassification in areas with high chlorophyll concentration due to phytoplankton blooms (e.g. Gulf of Lions) or mixing (e.g. Alboran Sea) that can be erroneously identified as Case II waters. In fact, comparing normalized R_{rs} spectra of a bloom area with normalized R_{rs} spectra of a complex coastal area, gives results very similar in the shape and values. On the other hand, without any normalization, R_{rs} spectra of a bloom area show low values with respect R_{rs} spectra of a complex costal area. Hence an additional classification step is added in which the Euclidian distance between measured R_{rs} spectra and non-normalized class spectra determines if the water is truly a member of the "complex" water class (if this Euclidean distance is lower than a threshold value, i.e 1.5).

This hybrid approach therefore uses both normalised and non-normalised spectra within its cluster designation and can be thought of as a hierarchical cluster approach. This approach of using additional spectral properties (others could be total spectral brightness or spectral angle) to distinguish between broad classifications such as bright and dark waters (which can be done irrespective of the number of visible bands available) allows the use of a hierarchy of classes. This broad separation of waters is a tactic that can be harmonised across multiple sensors.

We have not yet codified hierarchical clustering into the CERTO code base, but we do intend to implement this soon, especially as we move towards the multi-sensor (differing wave band sets) analysis.

4.1.5 Including uncorrected top-of-atmosphere and Rayleigh-corrected radiance alongside atmospherically corrected data.

Currently the optical water class generation and application is performed after the remote sensing data have undergone a number of processing stages such as atmospheric correction and cloud masking. It is technically possible to perform cluster analysis and classification at any stage in the processing chain and one candidate that has been posited is to generate clusters at the top-of-atmosphere (TOA) stage. This would mean that the cluster set would have to capture a greater level of spectral variance, which increase the complexity of the cluster generation. Though a benefit of this might be that we would be able to assign optical classes to 'components' such as clouds, sea ice, haze etc it also means that for a given number of classes we would have less classes to distinguish between water types as a

number would likely be dominated by atmospheric components (as the atmospheric signal typically accounts for around 90% of the TOA signal. Currently we would not recommend trying to identify optical water types using TOA data, though it might be of use to have two stages of optical classification: one at the TOA stage to distinguish atmospheric optical classes and one at the post-atmospheric correction stage to identify optical water classes. This has not yet been codified into the CERTO codebase and though we may test this sort of dual stage classification within CERTO it is unlikely to be sufficiently mature to be released in the final code base.

5 Code developments

The codebase for WP 4 has seen development in two main areas. The first is the creation of a python package that will be used for the generation and application of optical water types to remote-sensing reflectance data. This package is designed to have a structure similar to the packages of scikit and this will allow users to create and apply OWT sets with relative ease. The second area of development has been on automated regional summary report generation in the context of optical diversity. This means that given a bounding region of interest (such as for one of the six CERTO user case study areas) a user can automatically generate a summary report of the optical diversity of the region, in a standardised format, with figures showing geographic and temporal features in the optical characteristics of the area.

5.1 OWT package

The OWT package is built to follow the scikit-learn syntax. This means that it has a ‘generate’ and ‘apply’ functionality and can be easily integrated into scikit-learn workflows. This gives us a consistent, simple and potentially even familiar API for users. In addition, it enables us to easily vary the fitting parameters and normalisation schemes and compare them against one another in a consistent manner.

In this manner, users can build their own pipelines in order to cluster their data (this is the generation functionality). Alternatively, they can load a pre-fitted pipeline and predict the membership of new data to the existing clusters (the application functionality).

Persistence of models can be achieved through storing the fitted pipelines onto disk. In addition, the model parameters used to generate a given dataset will be saved inside the netcdf as metadata. This includes fitting parameters as well as fitted parameters of intermediate steps, such as normalisation values and principal components.

The generator creates a cluster set using a multi-metric assessment and a configuration defined by the user (normalisation or not, wave band set etc). The apply function then uses a generated cluster set to classify input data.

Below is a figure showing the example output from the cluster generation process (Figure 2). The figure shows four panels with metric scores against possible number of clusters. The four panels from left to right correspond to four different fuzziness factor values (as shown on the top bar from 1.5 to 5). The four coloured sets of points

in each panel correspond to four different metrics of cluster set performance (Xie beni index, Fuzzy partition coefficient, Silhouette score and the Davies Bouldin metric) with the error bars showing the variance generated by bootstrapping the analysis with 5 subsets of the total sample. For a set of K clusters and a dataset consisting of N data points the Xie Beni index is defined as:

$$\mathrm{C}= \frac{1}{N}\frac{WGSS}{\min_{k<k'} \delta_1(C_k\,,\ C_{k'})^2} \tag{7}$$

where $WGSS$ is the pooled within-cluster sum of squares (the sum of within-cluster dispersions for all clusters), C_k and $C_{k'}$ refer to clusters and δ_1 is the single-linkage distance, defined as:

$$\delta_1 = \min_{\substack{i\in I_k \\ j\in I_{k'}}} d(M_i,\ M_j) \tag{8}$$

with $M_i,\ M_j$ being pairs of observation points.

The Fuzzy Partition coefficient (F) was defined by Bezdek (1981):

$$F = \frac{1}{N}\sum_{j=1}^{K}\sum_{i=1}^{N}\left(\mu_{ij}\right)^m \tag{9}$$

The Silhouette score is calculated as:

$$\mathcal{C} = \frac{1}{K}\sum_{k=1}^{K} \mathfrak{s}_k \tag{10}$$

where $\mathfrak{s}_k$ is the per-cluster mean silhouette derived from all silhouette widths (s) :

$$\mathfrak{s}_k = \frac{1}{n_k}\sum_{i\in I_k} s(i) \tag{11}$$

The Davies Bouldin index is calculated as:

$$DB = \frac{1}{K}\sum_{k=1}^{K} M_k = \frac{1}{K}\sum_{k=1}^{K}\max_{k'\neq k}\left(\frac{\delta_k+\delta_{k'}}{\Delta_{kk'}}\right) \tag{12}$$

where $\delta_k$is the mean distance between the points belonging to cluster k and the cluster barycentre.

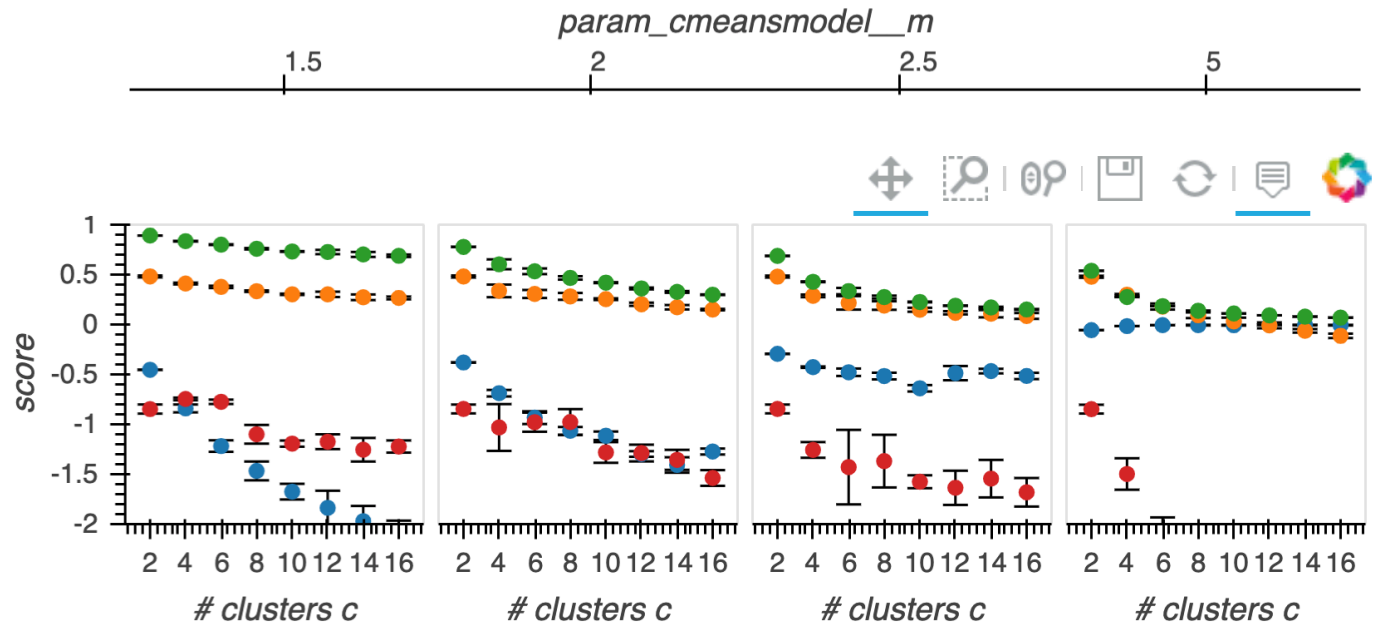

Figure 2: Cluster set scoring across a range of metrics, fuzziness values and number of clusters. Metrics are shown by colour Xie-Beni index (blue), Silhouette index (orange), Fuzzy partition coefficient (green) and Davies Bouldin index (red).

A cluster set generated by this code for OLCI is shown in section 6 alongside some simple maps of dominant clusters.

5.2 Regional characterisation reports

Work is also ongoing to develop a scheme for the automated generation of regional characterisation reports. These reports will allow the rapid assessment of a region of interest, such as the case study regions of CERTO. Though the final report structure is not yet precisely confirmed an example outline is given below.

The top-level structure of the report is organised such that the user defines the area of interest and then runs the report template. This will then perform a set of processing and plotting operations to outline the optical characteristics of the region in a temporal and geographic context.

These reports will allow more informed and targeted sampling within field campaigns and fixed moorings. The standardised format will also make it simpler to compare across multiple regions for a given trait or set of traits. The reports are built within a jupyter notebook which allows both the text and code sections to be version controlled, while keeping the report generator open and simple to share.

6 Initial cluster sets

We have begun generating potential water class sets for testing at the global and regional level using the current developmental version of the codebase. An example of such a class set for OLCI is shown below alongside some examples of applying the cluster sets at the regional level.

The latest optical water type set generated for OLCI makes use of the full range of optical wavebands (11 in total) and was created by subsampling from 6000 OLCI granules taken from across the globe and spanning the five-year OLCI archive. A random subset (10,000 spectra) was then sampled from the training dataset and was used to create a set of ten clusters. We intend to increase this number in the future but for now this is a sufficient balance between capturing spectral variation and computation time.

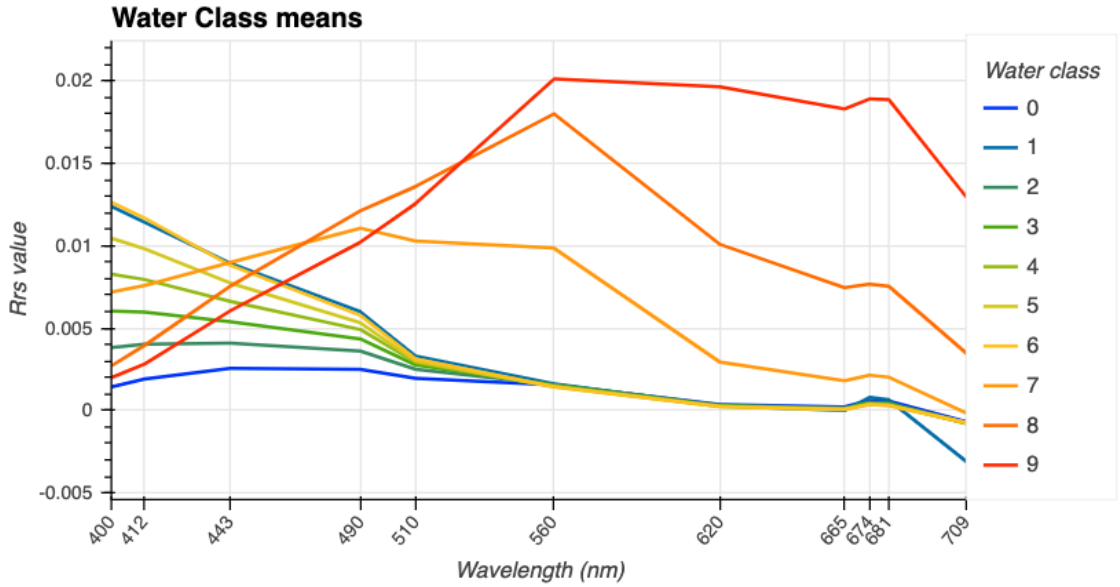

Figure 3: OLCI optical water class set generated from global dataset of 6000 OLCI tiles using 11 visible wavebands.

The clusters are actually defined in principle-component-space but as the principal components are known (and defined within the cluster set file) we can simply reconstruct the corresponding R_{rs} spectra for each cluster (as shown in Figure 3).
As expected, the primary transition across the optical water class set is from blue dominated to green dominated spectra (comparing water classes 6 and 8 for example). The principal component analysis shows that >95% of the sample variance is contained within 3 principal components, which follows the paradigm of considering ocean optical properties as a result of the relative influence of algal absorption, non-algal absorption and particulate scattering. However, this cluster set does appear to show a stronger distinction in the red bands than seen in previous cluster sets such as those used in OC-CCI. This could be due to the high density of wavebands in the 665-709 region.

Applying this class set to an example OLCI granule in the Indian Ocean (Figure 4) we can see that the dominant optical water class patterns highlight ocean features such as eddies and currents, coastal 'hotspots' and the impact of islands such as the Maldives.

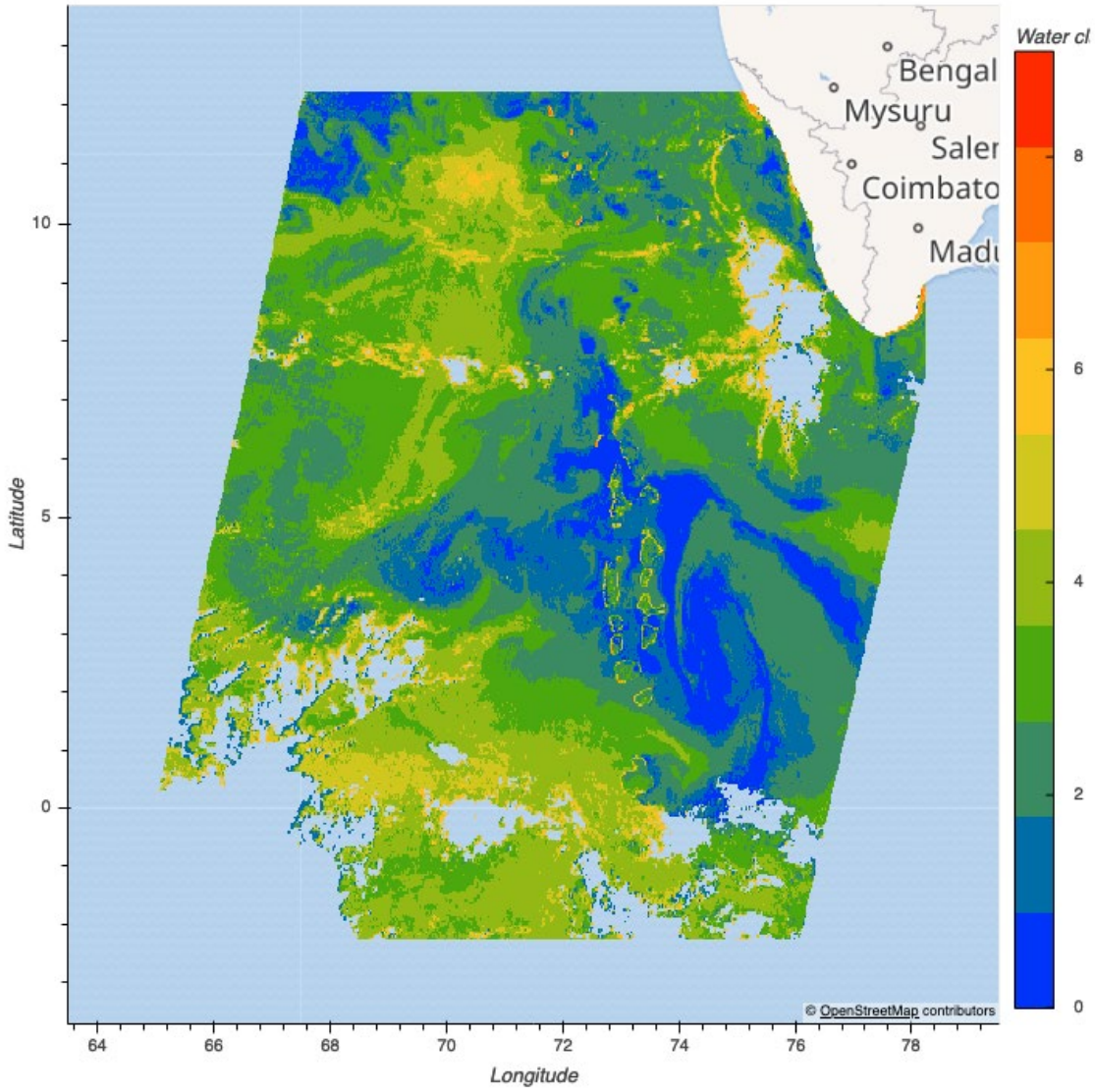

Figure 4: Map of the dominant optical water class per pixel when the example OLCI cluster set is applied to a series of OLCI granules taken on 2017-01-05 and processed using POLYMER atmospheric correction.

7 Development timeline

Now that we have an operational scikit-learn style operator and have generated OLCI classes using it we will begin to test the cluster generation with additional sensors (such as sentinel 2) and provide regional cluster sets for each of the case study areas.

This should be completed within the next 4-6 weeks and then we will aim to interpret the water class sets in a water quality context with each of the case study area experts/partners.

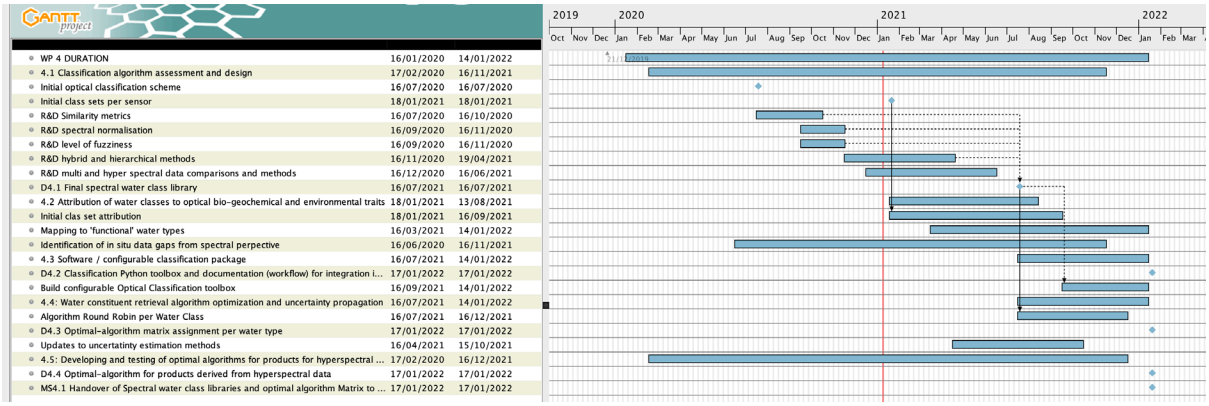

Figure 5: Gantt chart showing developement timeline of WP4 over the rest of the CERTO project.

8 References

Bezdek, JC. (1981) Pattern recognition with Fuzzy Objective Function Algorithms. New York: Plenum Press; DOI 10.1007/978-1-4757-0450-1

Cebeci, Z., & Cebeci, C. (2020). A fast algorithm to initialize cluster centroids in fuzzy clustering applications. *Information*, *11*(9),[446] https://doi.org/10.3390/info11090446

Dembele, D. (2008) "Multi-objective optimization for clustering 3-way gene expression data," Adv. Data Anal. Classif. 2(3), 211–225

Gueorguieva, Natacha and Valova, Iren and Georgiev, George (2017), M&MFCM: Fuzzy C-means Clustering with Mahalanobis and Minkowski Distance Metrics, Procedia Computer Science, [114], 224 – 233, doi: 10.1016/j.procs.2017.09.064

Lubac, B., and Loisel, H., (2007) Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea, Remote Sensing of Environment, Vol 110, Issue 1, 45-58, https://doi.org/10.1016/j.rse.2007.02.012.

Jackson, T., Sathyendranath, S., and Mélin, F. (2017). An improved optical classification scheme for the ocean colour essential climate variable and its applications. *Remote Sens. Environ*. 203, 152–161. doi: 10.1016/j.rse.2017.03.036

Jackson, T., Chuprin, A., Sathyendranth, S., Calton, B., (2020) OC-CCI Product User Guide for v5.0 Dataset, https://docs.pml.space/share/s/okB2fOuPT7Cj2r4C5sppDg

F. Mélin, V. Vantrepotte,(2015) How optically diverse is the coastal ocean?, Remote Sensing of Environment, Vol 160, 235-251, https://doi.org/10.1016/j.rse.2015.01.023.

Moore, T. S., Campbell, J. W., and Hui Feng, (2001) A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 8, pp. 1764-1776, doi: 10.1109/36.942555.

Moore, T.S., Campbell, J.W., Dowell, M.D., 2009. A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product.Remote Sens. Environ. 113, 2424–2430.

Morel, A. and Prieur, L. (1977). Analysis of variations in ocean color. *Limnol. Oceanogr.,* 22: 709-722.

Shun Bi, Yunmei Li, Jie Xu, Ge Liu, Kaishan Song, Meng Mu, Heng Lyu, Song Miao, and Jiafeng Xu, (2019) "Optical classification of inland waters based on an improved Fuzzy C-Means method," Opt. Express 27, 34838-34856

Spyrakos, E., O'Donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis, S.G.H., Neil, C., Barbosa, C.C.F., Binding, C.E., Bradt, S., Bresciani, M., Dall'Olmo, G., Giardino, C., Gitelson, A.A., Kutser, T., Li, L., Matsushita, B., Martinez-Vicente, V., Matthews, M.W., Ogashawara, I., Ruiz-Verdú, A., Schalles, J.F., Tebbs, E., Zhang, Y. and Tyler, A.N. (2018), Optical types of inland and coastal waters. Limnol. Oceanogr., 63: 846-870.

Vantrepotte, V., Loisel, H., Dessailly, D., Mériaux, X. (2012) Optical classification of contrasted coastal waters Remote Sensing of Environment, 123, pp. 306-323

Volpe, G., Colella, S., Brando, V. E., Forneris, V., La Padula, F., Di Cicco, A., Sammartino, M., Bracaglia, M., Artuso, F., and Santoleri, R. (2019) Mediterranean ocean colour Level 3 operational multi-sensor processing, Ocean Sci., 15, 127–146, https://doi.org/10.5194/os-15-127-2019