Ecological Indicators 95 (2018) 794–804

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original Articles

Applying MERIS time series and dynamic time warping for delineating areas with similar temporal behaviour in the northern Baltic Sea

Tapio Suominen

Department of Geography and Geology, University of Turku, FI-20014, Finland

ARTICLE INFO

Keywords:
Rremote sensing
Time series
Dynamic time warping
Coastal waters
The Baltic Sea
Ocean colour

ABSTRACT

Coastal waters are subject to ongoing long-term developments, cycles of varying lengths and random variations. Assessments of water quality should not be based only on temporally sparse sampling and inter-annual comparisons of periodical data, but also on their temporal behaviour as an entity. In the latter approach, regions having similar inter- and intra-annual temporal pattern are classified together, regardless of the differing levels of the observed parameter.

Altogether 4602 time series, containing Medium Resolution Imaging Spectrometer (MERIS) remote sensing reflectance data $R_{rs}(443)$, $R_{rs}(560)$ and $R_{rs}(665)$ and one band ratio $R_{rs}(681)/R_{rs}(620)$, were formed for the Gulf of Bothnia in the Northern Baltic Sea, covering the ice free periods of 1.6.–30.9. in 2004–2011. I was interested in the similarities in temporal shapes of the reflectance time series, and the time series were standardized by their annual means and standard deviation before further processing. Dynamic time warping (DTW) was used to measure the dissimilarity between the standardized time series, which were clustered in order to find areas having similar temporal character. In partitional clustering, a random initial centroid time series is selected for each cluster, which is then adjusted according to the selected centroid function to find coherent clusters. With DTW, a specific DTW barycentric averaging (DBA) is commonly used for this purpose.

Appropriate configurations for DTW and clustering were searched by evaluating the clustering results with two cluster validity indices. Two key settings in DTW are 1) step pattern, which defines how the minimum distances between the observations are searched, and 2) window constraint, which constrain the allowed time difference between the observations to be compared. The performance of two step patterns, symmetricP0 and symmetricP1, and three window constraints, $+-1$, $+-7$ and $+-21$ days, were tested. The partitions may vary due to randomness in the selection of the initial centroid time series. The stability of the repeated clustering was evaluated by Variation of Information–index (VI). With this metric, symmetricP1 step pattern performed slightly better than symmetricP0. Longer window constraints produced more labile partitions. Allowing a certain amount of temporal distortion is however desirable, and because the VI showed satisfactory results also with window constraint of $+-7$ days, it was selected for further computation. A Silhouette index was used to evaluate the appropriate number of clusters (k). Regardless the number of k, the clusters were neither internally coherent nor clearly deviating from each other.

Although the time series did not form strong clusters in the terms of clustering, they formed spatially distinctive and coherent groups on a map. The groups reflected the surface layer circulation pattern of the Gulf of Bothnia, rivers with fresh water input and terrestrial washed-out materials being among the most recognisable phenomena. Hard partitions gave, however, too simplified view on spatiality of temporal patterns. To avoid visual mis-interpretations, the prototypes of each cluster were calculated with DBA and distances in DTW space from these prototypes to all the other time series were visualized. In this way, clustering was used to define the macro-areas of similar temporal pattern, but similarities were defined in continuous DTW scale. This allowed more precise evaluation of the spatial–temporal relationships.

1. Introduction

As fresh waters end up in the coastal seas, the substances it contain start their four-dimensional cycle, in which they are altered by physical, chemical or biological processes. At any given moment the waters are subject to ongoing long-term developments, cycles of varying lengths

E-mail address: tapio.suominen@utu.fi.

https://doi.org/10.1016/j.ecolind.2018.08.023
Received 17 May 2018; Received in revised form 3 August 2018; Accepted 12 August 2018
1470-160X/

and random variations. With time some patterns or trends may become detectable, allowing future predictions and making it possible to react to unwanted developments. These developments occur mainly due to local anthropogenic activities, which harmful effects are adapted to natural seasonal processes of the region (e.g. natural algal blooms reinforced by human induced eutrophication). Not only the levels but also the timing of the seasonal patterns may alter due to large scale changes, climate change in particular (BACC, 2008; HELCOM, 2013), which will have unpredictable effects on the aquatic environment.

There is no single method which meets the challenge of collecting data in a coastal sea with varying water properties (Gohin et al., 2008; Petersen et al., 2008). In situ sampling offers a wide selection of water quality parameters and highly reliable data. The downside is that an in situ sample represents only a small volume of water at a certain moment in time, while variables like chlorophyll-*a* are unevenly distributed (Rantajärvi et al., 1998). A sparse or periodic sampling scheme may be adequate in stable or in annually repetitive conditions, although ideally samples would be taken year round to observe also the baseline (Ferreira et al., 2011). In a worst case, a spatial view on water properties is biased due to insufficient temporal coverage. Remote sensing facilitate a multisource approach to data gathering, with an extensive spatial and temporal coverage. Remote sensing data contains already a vast reserve for coastal sea research and due to new instruments, this reserve is growing extensively (Donlon et al., 2012; Malenovský et al., 2012). Disadvantages of using remote sensing in operational coastal monitoring are the limited range of water properties that can be observed and the reliability of processed geophysical data, especially in optically complex coastal waters (Beltrán-Abaunza et al., 2014; Attila et al., 2013; Smith et al., 2013).

Krug et al. (2017) reviewed the ocean surface partitioning strategies using remote sensing, and listed several applications of partitioning of ocean surface into functional units, including enhanced sampling designs, estimating biogeochemical fluxes in different spatial scales (e.g. Fay and McKinley, 2014) and estimating the connections between biological processes and environmental forcing. Classifying areas by their phenology with remote sensing data has extensively been done in terrestrial environments (Jönsson and Eklundh, 2004; Verbesselt et al., 2010; Lhermitte et al., 2011; Maus et al., 2016). Extending a similar approach to aquatic environments may require revised methods due to different dynamics of initial data. Temporal analysis of optical water properties has typically applied in studies concerning phytoplankton phenology (Platt et al., 2009, Racault et al., 2012, Loisel et al. 2017). Mélin and Vantrepotte (2015) and Trochta et al. (2015) approached the topic by first classifying the water bodies to optical water types and then examining their spatial and temporal patterns by class memberships. According to them, optical diversity (i.e. areas belonging to multiple optical water type during a certain period) is highest in the areas between coastal domain and open ocean, whereas e.g. closed seas (like the Baltic Sea) and inland water bodies show lower optical diversity in temporal terms.

In the Baltic Sea the optical properties of the waters are driven by the abundance phytoplankton, particulate suspended matter and coloured dissolved organic matter (CDOM) (Attila et al. 2013; Kratzer et al. 2008). Temporal patterns of these properties are forced by the physical environment, especially climate and prevailing weather (temperature, solar radiation, rain fronts and runoffs). An effective way to explore the temporal patterns is decomposing the time series to seasonal, long term and stochastic components (e.g. Loisel et al., 2017, Vantrepotte and Mélin, 2011). However, in this study I aim to analyse the spatiality of similarly behaving sea areas in a long term, not only those areas whose temporal pattern of the optical properties recurs annually or within some other period. From a functional point of view, these areas could be tied together; the drivers behind the behaviour may be similar, even if their optical outcomes in coastal waters are not periodically recurred. Thus, identifying such areas might be useful in the sense of practical water management.

Time series clustering is a process of unsupervised partitioning in such a way that homogeneous time series are grouped together based on certain similarity measure (Aghabozorgi et al., 2015). Challenges in time series data mining are that it is computationally expensive due to massive and high dimensional data sets. Secondly, time series data is typically noisy, it may include temporal shifts and data gaps, or time series may have varying lengths (Lin et al., 2004; Aghabozorgi et al., 2015). The approaches how the similarity between the time series are measured are numerous (Lhermitte et al., 2011; Serrà and Arcos, 2014; Aghabozorgi et al., 2015). The choice of a proper distance approach depends on the characteristic and length of time-series, and on the objective of clustering time series to a high extent (Aghabozorgi et al., 2015).

Dynamic time warping (DTW) is among the most used time series dissimilarity measures. DTW was originally developed for speech recognition (Sakoe and Chiba, 1978), but it has been applied to various tasks, including remote sensing (e.g. Petitjean et al., 2012; Maus et al., 2016). The method is sensitive to a spiky noise and it is usually applied in satellite image time series, which are obtained from relatively gently varying objects, like vegetation indexes. In the aquatic environments this might be the case in the open sea, but in many cases the coastal areas with anisotropy and temporal irregularity are of interest. DTW has nevertheless some advantageous properties considering especially coastal sea monitoring. It is able to handle temporal deviations in trajectories (Keogh and Ratanamahatana, 2005; Petitjean et al., 2011), which is beneficial in aquatic environments, where water properties are spread due to water movements. DTW has also been applied in irregularly sampled time series, which is usually the case with satellite image time series (SITS). Further, it is able to use multiband data (Petitjean et al., 2012), although this property has not been utilized yet in this paper.

The research question is formulated as follows: can the optically complex coastal waters be classified by similarities on their temporal behaviour in reflectance data. I hypothesise that assessments of water quality should not be based only on temporally sparse sampling and inter-annual comparisons of periodical data, but also on their temporal behaviour as an entity. The challenge in partitioning the satellite image time series is the extensive data, since the time series partitioning is typically computationally expensive. Thus, this paper aims also to find appropriate configurations for chosen methods and physical environment, and to facilitate later studies with similar starting point. To answer these question, satellite image time series of eight years from June to September were retrieved for the Gulf of Bothnia, the northern Baltic Sea. From the imagery, I extracted 4602 time series of normalized reflectance $R_{rs}(443)$, $R_{rs}(560)$ and $R_{rs}(665)$, and the ratio $R_{rs}(681)/R_{rs}(620)$. The band ratio $R_{rs}(681)/R_{rs}(620)$ has been used for CDOM estimations and the ratio has performed in some conditions better than more advanced algorithms like C2R (Case II Regional, Doerffer and Schiller, 2007) (Attila et al., 2013). Since I wanted to separate the areas by their divergent temporal patterns, not by differing levels of reflectance, I standardized the reflectance time series by their annual mean and standard deviation prior to analysis. The time series were clustered with partitional clustering, by using DTW barycentric averaging (DBA) as a centroid averaging method, and DTW distances as a dissimilarity measure between the time series. Validity of repeated partitions with different configurations of DTW were evaluated with two cluster validity indices (CVI), Variation of Information (VI) and Silhouette.

2. Material and methods

2.1. Study area

The Baltic Sea (Fig. 1) is a brackish marginal sea, with one outlet via the Danish Straits in the south. Horizontally, the surface layer salinity gradually decreases to almost zero at the end of the northernmost basin,