RECONSTRUCTION OF SEAWIFS CHLOROPHYLL DATA FOR THE

BLACK SEA

A THESIS SUBMITED TO THE GRADUATE SCHOOL OF MARINE SCIENCES OF MIDDLE EAST TECHNICAL UNIVERSITY

ΒY

SERKAN SANCAK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF PHILOSOPHY OF DOCTORATE IN THE DEPARTMENT OF PHYSICAL OCEANOGRAPHY

June 2011

RECONSTRUCTION OF SEAWIFS CHLOROPHYLL DATA FOR THE BLACK SEA

submitted by SERKAN SANCAK in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Physical Oceanography Department, Middle East Technical University by,

Prof. Dr. Ferit Bingel Director, Graduate School of Marine Sciences	
Prof. Dr. Emin Özsoy Head of Department, Physical Oceanography	
Prof. Dr. İ. Temel Oğuz Supervisor, Physical Oceanography Dept., METU	
Examining Committee Members:	
Prof. Dr. Zahit Uysal Graduate School of Marine Sciences, METU	
Prof. Dr. İ. Temel Oğuz Graduate School of Marine Sciences, METU	
Assoc. Prof. Dr. Cüneyt Güler Dept. of Geological Engineering, University of Mersin	
Prof. Dr. Ferit Bingel Graduate School of Marine Sciences, METU	
Prof. Dr. Emin Özsoy Graduate School of Marine Sciences, METU	

Date: 30 June 2011

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Serkan Sancak

Signature :

ABSTRACT

RECONSTRUCTION OF SEAWIFS CHLOROPHYLL DATA FOR THE BLACK SEA

Sancak, Serkan Ph.D., Department of Physical Oceanography Supervisor : Prof. Dr. İ. Temel Oğuz

June 2011, 194 pages

SeaWiFS was collecting ocean color data since 1997. This means chlorophyll-a data for more than ten years. Since, SeaWiFS Chl-a data is validated for Black Sea this data set can be used for analysis. Nevertheless, the data is not gap free due to cloud effect. One of the main objectives of this work is to obtain a gap free, complete Chl-a data set for the Black Sea. For this purpose DINEOF method will be used. EOF analysis is a by-product of this method and the results are used to summarize some temporal and spatial Chl-a behavior for the time period of data set. Some major findings General low productive phase with relatively low chlorophyll are: concentrations over the basin starts by the beginning of spring season and ends by the beginning of autumn except year 1999, in which high chlorophyll period begins by the end of summer. There is a gradual increase of Chl-a concentration towards east in winter and spring, and a similar change towards south in spring and summer. According to temporal analysis, there is an absence of a typical annual pattern repeating every year. Instead, there are appreciable interannual variability with some years having maximum concentrations in spring and summer, that may be even higher than the basin averaged values of winter and autumn seasons, whereas some years possess no definitive spring - summer maximum. Another major finding is that, the long-term mean chlorophyll distribution reveals an almost

iv

horizontally uniform interior basin with concentrations of about 1.0 mg/m³ that increases to 1.5 mg/m³ range along a narrow coastal band and exceeds 3 mg/m³ along the western coastal zone. A switch to lower concentrations (up to 0.5 mg/m³) within the interior basin takes place in the spring and summer months, thus giving rise to more pronounced onshore-offshore gradients around the basin.

The other main objective of this study is to regionalize the Black Sea basin according to its Chl-a content. The determined provinces will also be stated as bioms throughout the thesis, since, Chl-a is an indicator of primary production. Separating the oceans into its bioms using ocean color (Chl-a) data is applied to various seas so far. Though, the studies that include Black Sea, define the basin as a single province. This is the main motivation for the objective. By using more than ten years of data, the Black Sea is separated into nine provinces according to its Chl-a content. Nevertheless, the main reference region map includes only six provinces. A major outcome is that, the interior basin is formed by one composite cell contrary to the two separated gyres for the case of circulation dynamics.

Keywords: Black Sea, SeaWiFS, Chlorophyll-a, Ocean Color, DINEOF, EOF

SEAWİFS KLOROFİL VERİSİNİN KARADENİZ İÇİN YENİDEN YAPILANDIRILMASI

ÖΖ

Sancak, Serkan Doktora, Fiziksel Oşinografi Bölümü Tez Yöneticisi : Prof. Dr. İ. Temel Oğuz

Haziran 2011, 194 sayfa

SeaWiFS 1997'den beri deniz rengi verisi toplamaktadır. Bu on yılı aşan klorofil-a verisi demektir. SeaWiFS klorofil-a verisi Karadeniz için tasdiklenmiş oluğundan bu veri analizler için kullanılabilirdir. Yine de bulut etkisinden dolayı veri bütünlüğü yoktur. Bu çalışmanın temel amaçlarından biri Karadeniz için bütünlüklü ve eksiksiz bir klorofil-a veri setinin oluşturulmasıdır. Bu amaçla DINEOF metodu kullanılmıştır. EOF analizi bu metodun bir yan ürünüdür ve sonuçları zamansal ve bölgesel klorofil-a davranışlarının özetlenmesi için kullanılmıştır. Bazı ana bulgular şunlardır: Göreceli düşük klorofile sahip, basenin genel düşük üretim dönemi ilkbahar başında başlayıp sonbahar başında sona ermektedir. Yüksek klorofil başladığı 1999 yılı döneminin yaz sonu buna tezattır. Klorofil-a konsantrasyonunda kış ve ilkbaharda doğuya doğru aşamalı bir artış gözlenmektedir. Güneye doğru olmak üzere benzer bir durum ise ilkbahar ve yaz mevsimlerinde mevcuttur. Zamansal analizler sonucuna göre her yıl tekrarlanan tipik bir yıllık seyir mevcut değildir. Bunun yerine, kayda değer yıllar arası bir değişkenlik vardır. Bazı yıllarda en yüksek konsantrasyon ilkbahar ve yazın olmakta, hatta bu değer kış ve sonbahar basen ortalamasının da üzerinde olmaktadır. Bazı yıllarda ise belirgin bir ilkbahar yaz maksimumu oluşmamaktadır. Bir diğer ana bulgu, uzun dönem klorofil ortalaması dağılımının iç basende neredeyse birörnek yatay bir dağılımla 1.0

vi

mg/m³ civarında, dar kıyı şeridinde 1.5 mg/m³ civarında ve batı kıyıları bölgesi boyunca 3 mg/m³,'ün üzerinde oluşudur. İlkbahar ve yaz aylarında iç basendeki konsantrasyonda bir düşüş olmakta (0.5 mg/m³'e kadar) ve bu basen genelinde daha belirgin bir kıyı – açık deniz farkına yol açmaktadır.

Bu çalışmanın bir başka temel amacı ise Karadeniz baseninin klorofil-a muhteviyatına göre bölgelere ayrılmasıdır. Klorofil-a birincil üretimin bir göstergesi olduğundan, belirlenen bölgeler tez boyunca biomlar olarak da adlandırılacaktır. Şimdiye kadar deniz rengi (klorofil-a) kullanılarak denizlerin biomlara ayrılması çeşitli denizlere uygulanmıştır. Yine de, Karadeniz'i içeren çalışmalar baseni tek bir bölge olarak tanımlamışlardır. Bu çalışmanın esas motivasyonunu budur. On yılı aşkın bir veri kullanılarak, klorofil-a muhteviyatına göre Karadeniz dokuz bölgeye ayrılmıştır. Buna rağmen ana referans olan bölgesel haritada sadece altı bölge yer almaktadır. Önemli bir diğer sonuç ise iç basenin tek bir bölgeden oluşmasıdır, ki bu, akıntı dinamiklerinde mevcut olan iki ayrı döngü bölgesine ters bir bulgudur.

Anahtar Kelimeler: Karadeniz, SeaWiFS, Klorofil-a, Deniz Rengi, DINEOF, EOF

To Mom

ACKNOWLEDGEMENTS

The author wishes to express his deepest gratitude to his supervisor Prof. Dr. Temel Oğuz and previous supervisor Prof. Dr. Şükrü Beşiktepe for their guidance, advice, criticism, encouragements and insight throughout the research.

The author would also like to thank Prof. Dr. Emin Özsoy and Assoc. Prof. Dr. Barış Salihoğlu for their suggestions and comments.

The technical assistance of Dr. Murat Gündüz and Mr. Ersin Tutsak are gratefully acknowledged.

The author also wishes to thank to his mentor Akçakoca Ufuk Özgöz for all his eternal support and guidance.

TABLE OF CONTENTS

Abstract	iv-v
Öz	vi-vii
Acknowledgements	ix
Table of Contents	x-xi
List of Tables	xii
List of Figures	kiii-xvi
1. Chapter 1: INTRODUCTION	1
1.1. Introduction	1
1.2. Area of Interest	2
1.3. Studies on the Analysis of Long Term Satellite Observations	5
1.4. Purpose, Objectives and Expected Outcomes	9
2. Chapter 2: MATERIAL AND METHOD	11
2.1. Retrieving and Processing the SeaWiFS Data	11
2.2. Statistical Method: EOF	18
2.3. Filling the Gaps with Statistical Method DINEOF	20
2.4. Regionalization Method	22
3. Chapter 3: DINEOF RECONSTRUCTION OF THE DATA GAPS	24
3.1. Setting Up the Method by Sensitivity Tests	24
3.1.1. Test 1	33
3.1.2. Test 2	39
3.1.3. Test 3	41
3.1.4. Test 4	44
3.2. Construction of the Gap Free Data Set	46
3.3.EOF Analysis Result	47
4. Chapter 4: PROPERTIES OF SURFACE CHLOROPHYLL	
CONCENTRATION AS INFERRED FROM THE	
RECONSTRUCTED DATA	56
4.1.Long -term Average Properties	56
4.2. Annual Mean Properties	61

4.3. Regionalization of Chlorophyll Data	70
5. Chapter 5: DISCUSSION AND CONCLUSION	78
5.1. Discussion and Conclusion	78
References	82
Appendix A	90
Appendix B	93
Appendix C	167
Curriculum Vitae	192

LIST OF TABLES

TABLES

Table B.1. Table showing number of weeks according to seasons	94
Table B.2. Table showing number of weeks according to years	94

LIST OF FIGURES

FIGURES

	Figure 1.1. Black Sea basin's layout and bathymetry. Depth contours
3	are labeled in meter. (Oğuz <i>et. al.</i> , 2002)
	Figure 1.2. Black Sea Circulation. The schematic diagram for the
	main features of the upper layer circulation. (Korotaev et. al.,
3	2003)
	Figure 1.3. Schematic diagram of main DINEOF inputs and outputs
8	as applied to the North Sea (taken from Sirjacobs <i>et. al.</i> , 2008).
	Figure 2.1. Comparison of daily, weekly and monthly SeaWiFS
	derived Chl-a images from October 1997. Row (a) is daily
	images, 7 th , 11 th , 29 th , 30 th October. Row (b) is weekly images,
	4 consecutive weeks of October. Row (c) is monthly binned
15	October image
	Figure 2.2. Daily images of the same day. 30 th October 1997.
	Satellite passes at: (a) 08:51 a.m., (b) 10:31 a.m., (c) 12:09
15	p.m
	Figure 2.3. Weekly binned Chl-a images of the year 1999. Satellite
	imagery assumes a week as 8 days, resulting, 46 weeks in a
16-17	year.
	Figure 3.1. The map used for land masking of the data. 40 – 47 North
25	and 26 – 42 East
	Figure 3.2. Reconstructed weekly-binned Chl-a images of the year
26-27	1999. Only missing points are reconstructed
	Figure 3.3. Reconstructed weekly-binned Chl-a images of the year
28-29	1999. Complete reconstruction of the data
	Figure 3.4. Selected over estimated reconstructions. Only missing
	points reconstructions and complete reconstructions are
30	matched for random 3 weeks

Figure 3.5. Annual mean chl-a maps of year 1999. (a) is constructed	
from original data, (b) is constructed from output of only	
missing points reconstruction, (c) is constructed from output of	
complete reconstruction	31
Figure 3.6. Under and over-estimates of the reconstructions in Figure	
3.5. (a) original data – only missing points reconstruction, (b)	
original data – complete reconstruction	32
Figure 3.7. Time versus spatial basin average Chl-a graphic. Data	
from year 1999	32
Figure 3.8. Artificial dummy data set constructed for the DINEOF	
tests. Total 27 weeks in the set	35-36
Figure 3.9. Percentage errors maps for test 1. (a) is nev=2, ncv=8;	
nev=10, ncv=16 and nev=15, ncv=25, (b) is nev=1, ncv=7 and	
nev=1, ncv=25, (c) is nev=20, ncv=27 and nev=20, ncv=30	37
Figure 3.10. (a) Week 14 of the "dummy set", (b) map with cropped	
200x300 km artificial cloud, (c) reconstructed result of the test	
1, run 3	38
Figure 3.11. Original and reconstructed Chl-a maps together with	
related percentage error maps for test 2	40
Figure 3.12. Test 3 results. Original, cropped & reconstructed Chl-a	
images	42
Figure 3.13. Percentage error maps for test 3	43
Figure 3.14. Test 4 results. Original images (left) versus	
reconstructed (right) images	44
Figure 3.15. Percentage error maps for test 4	45
Figure 3.16. EOF results. First 6 modes for spatial component.	
Colorbar has arbitrary units	50-52
Figure 3.17. EOF results. First 6 modes for temporal component. Y	
axis has arbitrary units	. 53-55
Figure 4.1. Temporally averaged chlorophyll-a data. Data between 1	
January 1998 and 31 December 2007 are used	57
Figure 4.2. Spatially basin averaged chlorophyll-a data. Weekly-	
binned Chl-a maps between 14 September 1997 and 31	
December 2007 are used for basin averaged Chl-a values	58

Figure 4.3. Temporally averaged, seasonal chlorophyll-a maps,	
1998–2007	60
Figure 4.4. (a) Temporally averaged, annual chlorophyll-a maps,	
1998 – 2007	63
Figure 4.4. (b) Temporally averaged, seasonal chlorophyll-a maps	
according to years, 1998 – 2007. Autumn	64
Figure 4.4. (c) Temporally averaged, seasonal chlorophyll-a maps	
according to years, 1998 – 2007. Winter	65
Figure 4.4. (d) Temporally averaged, seasonal chlorophyll-a maps	
according to years, 1998 – 2007. Spring	66
Figure 4.4. (e) Temporally averaged, seasonal chlorophyll-a maps	
according to years, 1998 – 2007. Summer	67
Figure 4.5. Temporal plots of basin averaged Chl-a values, according	
to years, 1998 – 2007. Weekly-binned maps from Figure B.3.	
are used for basin averaged values	68-70
Figure 4.6. The Black Sea separated into 9 regions according to the	
SeaWiFS derived ocean color data collected between 1998 –	
2008. Though, only 6 regions can appear	72
Figure 4.7. Regionalized seasonal maps of the Black Sea. 1998 -	
2007 data are used to obtain the maps	73
Figure 4.8. Regionalized annual maps of the Black Sea. 1998 – 2007	
data are used to obtain the maps	75
Figure 4.9. Regions change in time. 14 September 1997 – 31	
December 2007. Number of pixels represent total area of a	
region in km ²	77
Figure B.1. Weekly-binned Chl-a images. 14 September 1997–31	
December 2007. Each week binned from 8 days SeaWiFS	
data	95-118
Figure B.2 Reconstructed weekly-binned Chl-a images 14	
September 1997–31 December 2007 Fach week binned from	
8 days SeaWiES data	10_1/2
0 uays Stavvii S uala I	13-142

CHAPTER 1

INTRODUCTION

1.1. Introduction

In many oceanographic applications, satellite data analysis suffers from both noise in the data and missing data points due to malfunctions of instruments, inaccessible data (e.g., due to cloud coverage), instrumental errors. Remote sensing (RS) products are particularly biased towards "good weather" conditions, mainly due to the cloud cover (Fettweis and Nechad 2011), and frequent gaps may occur in the time series of a RS product. In the past, several methods have been used to fill the gaps in the data retrieved from the satellites and recovery of missing data. For parameters exhibiting strong spatial or temporal correlations, the missing data can be estimated by use of statistical techniques. A traditional approach to fill in the missing data is the optimal interpolation as widely used in the reconstruction of Sea Surface Temperature (SST) and the Sea Level Anomaly (SLA) data sets all over the global oceans.

An alternative to the optimal interpolation is proposed by Beckers and Rixen (2003) and its modified version by Beckers *et. al.* (2006). This Data Interpolating Empirical Orthogonal Functions (DINEOF) technique based on the univariate methodology offers an improvement in terms of accuracy of the results as well as the computational speed (Alvera-Azcarate *et. al.*, 2005). DINEOF uses, through an iterative procedure, a truncated empirical orthogonal function (EOF) basis to calculate the value of the missing data. The procedure starts by removing the temporal and spatial mean from the original data and initializing the missing values to zero (i.e. to the mean). Then, at each iteration, the EOF basis is used to infer the missing data, and a new EOF basis is calculated using the improved dataset. Alvera-Azcarate *et. al.* (2007) further show that SST reconstructions can be improved by using a

multivariate approach in which SST, chlorophyll and wind fields are taken into account together for the analyses.

In the present paper, this technique is applied to the long-term SeaWiFS chlorophyll data in the Black Sea. Below, it is first provided in section 1.2 salient aspects of the Black Sea oceanographic characteristics relevant to the present study. Then it is briefly provided in section 1.3 a background material and current advances achieved on the data filling techniques. It is followed by purpose, objectives and expected outcomes of the present work in section 1.4.

1.2. Area of Interest

The Black Sea (Figure 1.1) is a unique marine environment, representing the largest land-locked basin in the world. It has a surface area of 4.2×10^5 km³ and a volume of 5.3×10^5 km³ with the maximum depth of about 2200 m. Its waters are in a state of almost complete isolation from the world ocean, excluding the restricted exchange with the Mediterranean Sea through the Turkish Straits System (the Bosphorus, Dardanelles Straits and the Sea of Marmara). As a result, the basin is almost completely anoxic, containing oxygen only in the upper 100 - 150 m of the water column (13% of the sea volume) and hydrogen sulphide in the deep waters. A permanent halocline separates the oxic and anoxic waters. (Oğuz *et. al.*, 1992; Özsoy and Ünlüata, 1997).

General circulation patterns in the basin is discussed by various observational and modeling studies. A schematic of the main circulation features is depicted in Figure 1.2., (Korotaev *et. al.*, 2003) as an updated version of the former one by Oğuz *et. al.* (1993). Various structural combinations of these features have been observed using remotely sensed and model-derived data sets or *in situ* measurements or their combinations. The major vortices (eddies and/or gyres) like the eastern and western gyres, the anticyclonic Batumi Gyre, and the coastally-attached Crimea, Sevastopol, Sinop, Sakarya and Bosphorus anticyclonic eddies are recurrently detected by remotely-sensed data. Other features like Rim Current and the filaments and dipole eddies accompanying with the Rim Current and coastal eddies



Figure 1.1. Black Sea basin's layout and bathymetry. Depth contours are labeled in meter. (Oğuz *et. al.*, 2002)



Figure 1.2. Black Sea Circulation. The schematic diagram for the main features of the upper layer circulation. (Korotaev *et. al.*, 2003)

are also identified clearly by satellite data (Sur *et. al.* 1994 and 1996, Oğuz *et. al.*, 2002).

The quasi-synoptic mesoscale surveys have been conducted over the entire Black Sea in 1990, 1991 and 1992 with participation of various institutions around the Black Sea (Oğuz *et. al.*, 1993, 1994, 1998), and depicted many details of the major circulation structures within the upper and intermediate layers such as the meandering cyclonic Rim Current, basin-wide, multi-centered cyclonic cell in interior part of the basin and a series of anticyclonic eddies between the coast and the Rim Current. Oğuz and Beşiktepe, (1999), further provided a detailed presentation of the Cold Intermediate Water mass formation characteristics within the Western Black Sea in April 1993.

Sur et. al., (1994 and 1996) focused on the mesoscale dynamics of the Black Sea boundary current system using the Coastal Zone Color Scanner (CZCS) and the Advanced Very High Resolution Radiometer (AVHRR) data in combination with the available in situ data. It provided a first detailed account of the mesoscale features, such as mesoscale eddies, jet instabilities, filaments, dipole eddies, river plumes and mushroom-like structures in addition to the Rim Current, cyclonic eastern and western gyres. An important implication of Sur et. al. (1994 and 1996) was to point how the ocean color (CZCS derived Chl-a) data provide an opportunity to study mesoscale circulation features in addition to the distributions of phytoplankton pigments. Sur and Ilyin (1997) extended these studies using additional AVHRR – infrared images combined with in situ measurements to analyze temporal evolution some mesoscale thermal patterns of the basin during 1991 – 1993. The different eddies, river influence on coastal current systems and coastal wave patterns propagating cyclonically around the basin have been detected. These studies further indicated importance of onshoreoffshore material transfer across the Rim Current zone and the shelf - deep basin interactions through jet instabilities, mesoscale eddies, filaments, mushroom-like structures. Spatial and temporal scales of these phenomena vary within the ranges of 10 - 100 km and 1 - 10 days. For further details on general oceanography of the Black Sea, it is referred to the comprehensive reviews provided by Özsoy and Ünlüata (1997) and Oğuz et. al. (1993).

1.3. Studies on the Analysis of Long Term Satellite Observations

As the SeaWiFS data stored for the past years since its launch in April 1997, the archive has become long enough to perform decadal-to-interannual scale time series analyses. The EOF analysis (the details of which is provided in Chapter 2) is one of the common tools used for processing only cloud-free satellite data. It is a mathematical procedure introduced into the Earth System Sciences by Obukhov (1947) and Lorenz (1956) and then often used in different disciplines (Sirovich and Everson, 1992). It provides an efficient way to synthesize the information contained in terms of a collection of empirical orthogonal functions (EOFs, also called principal components). These functions have some interesting properties: when only one EOF is used, this EOF is on average the closest to all images, when multiplied for each image by appropriate amplitude. Hence it is the best possible approximation of all images using only one spatial pattern (or EOF) and an amplitude for each image. With two EOFs, it can be shown that no other combination of two patterns can provide a better approximation to all images than these two. In general the first N EOFs are therefore the best way to summarize the information content of all images if only N patterns can be stored. Each image is then replaced by a filtered version in which the basic patterns are linearly combined with amplitudes corresponding to each image. Thus for the case of cloud-free data sets with no clouds and missing points, EOFs can be calculated straight forwardly to obtain an approximate representation of each data set (or image) as a truncated combination of a few EOFs. To date, it has been widely used for the analysis of AVHRR - SST data sets for different regions of the World's oceans. For example, Toumazou and Cretaux (2001) showed that the method is reliable and efficient for the oceanographic data sets (Sea Surface Height-SSH and Sea Surface Temperature-SST). Satellite-measured seasonal and interannual chlorophyll variability in the Northeast Pacific and Coastal Gulf of Alaska during September 1997 – August 2001 was studied by Brickley and Thomas (2004) to present the first synoptic quantification of chlorophyll variability and the most comprehensive view of the large-scale seasonal and interannual variability over the area by using the EOF decomposition as a statistical tool.

Vargas et. al. (2003) used the EOF analysis method for 7 years long (1993 - 1999) SST data derived from NOAA - AVHRR sensor data to identify the main SST spatial structure in the Gulf of Cadiz. The first 3 EOF modes managed to explain 79% of the temperature variance. While the first 2 modes dominate the temperature spatial variance in the area, the third mode has been identified as the main response of the SST field to the local zonal wind. Yoder et. al. (2001) used the same method to analyze the chlorophyll variability from the complete 7.5 years CZCS data along the US East Coast. They separated the study area into 21 smaller areas while applying the method to set the boundaries for averaging and subsequent analyses. Bisagni et. al. (2001) analyzed five-daily-averaged SST data for the Gulf of Maine and the Georges Bank region during 1993-1996. They used covariance and gradient EOF analyses as a tool to analyze sea-surface temperature variability in the area. Tseng et. al. (2000) employed the method for 2 years time series of monthly SST data to determine the major spatial and temporal variations in the East China Sea.

Baldacci et. al. (2001) presented the results of a combined empirical orthogonal function EOF analysis of AVHRR sea surface temperature data and SeaWiFS chlorophyll concentration data over the Alboran Sea (Western Mediterranean) for November 1997 – October 1998 period. Their objective was to gain a deeper insight on the interactions between the circulation and biological activity by inferring temporal and spatial variabilities of the local circulation system from long temporal series of satellite observations. They conclude that the EOF decomposition approach has proved to be a valid technique to obtain a concise description of the large volume of data in terms of trends and statistics. Similarly, Stanev and Peneva (2002) used the EOF method to study sea level response of Black Sea to the climate-induced changes using the SSH data sets (1993 - 1998). Nezlin and Dyakonov (1998) and Nezlin (2000) used CZCS data for 1978 – 1986 and the SeaWiFS data for 1997 – 1998, respectively, to monitor phytoplankton blooms in the Black Sea using Chl-a as an indicator, and their connection to regional eutrophication and long-term climatic oscillations. Using the AVHRR data from 1993 and 1996 – 1998, together with meteorological data, Ginzburg et. al. (2000) investigated the role of mesoscale structures (eddies, dipoles, jets)

in coastal – deep basin water exchange and mixing in the Black Sea. Long term satellite altimetry data of TOPEX/Poseidon and ERS-1 for the Black Sea are further used by Korotaev et. al. (2001 and 2003) to delineate the seasonal and mesoscale sea level variability of the Black Sea. They showed explicitly a distinct seasonality of the circulation, with major characteristic features repeating every year with some year-to-year variability. In a study performed by Oğuz et. al., (2002), Chl-a concentrations derived from SeaWiFS data between 1997 and 2000 is analyzed for elaborating the role of mesoscale processes controlling biological variability in the basin coastal waters. Daily, weekly and monthly images were interpreted in terms of documenting a close link between biological production and physical features of the boundary current system, and of emphasizing the role of eddy processes on controlling mesoscale chlorophyll distributions. Some major results can be summarized as follows: "the western coastal waters of the Black Sea are characterized by permanently high chlorophyll concentrations of more than 4.0 mg/m³, often subject to considerable dynamical transformations, and modulated by mesoscale structures including southward elongated filament-like features extending up to 100 km offshore with a lifetime of up to a month. Filaments, meanders, offshore jets and other forms of mesoscale and sub-mesoscale structures appear as a common signature of the system along the Anatolian coast. They play an important role in the cross-stream transport of biota and chemical constituents, and thus supporting productivity within the interior parts of the basin. The images reveal phytoplankton blooms lasting several months in all three autumn seasons", (Oğuz et. al., 2002).

Beckers and Rixen (2003) further extended EOF method to infer missing data from oceanographic data series and to extract the relevant empirical orthogonal functions. In simple terms it is an extension of the former EOF methodology to a case of an iterative method. It sets a first guess of zero anomalies in the missing data points, while the number of retained EOFs is fixed by a cross-validation technique based on the minimization of the root-mean-square (RMS) misfit between some data set aside and its reconstruction. For more details we refer to Alvera-Azcarate *et. al.* (2005), Beckers *et. al.* (2006). A scheme of the main inputs and outputs of

the DINEOF analysis is given in Figure 1.3. As a by-product, the method allows to detect the number of statistically significant EOFs by a crossvalidation procedure for a complete or incomplete data set. Since for the proposed filling and analysis method there is no need for a priori information about the error covariance structure and, the method is therefore selfconsistent and parameter free. This method of EOF analysis manage to interpolate missing points more than 75 % with acceptable success when adopted for a small region in the Adriatic Sea. This work constitutes one of the first attempts to complete the gaps in satellite data that then allowed analyzing a long term archive data.



DINEOF: principle

Figure 1.3. Schematic diagram of main DINEOF inputs and outputs as applied to the North Sea (taken from Sirjacobs et. al., 2008).

Alvera-Azcarate et. al. (2005) applied a modified version of this method for the Adriatic Sea surface temperature. The new version was named as the Data Interpolating Empirical Orthogonal Functions (DINEOF). This study proved a successful implementation of the DINEOF method to

satellite derived oceanographic data sets for different geographical region with missing data points. Miles *et. al.* (2009) also implemented DINEOF method to produce daily, 4-km, cloud-free SST and Chl-a analyses for the South Atlantic Bight in 2003 and compared the SST products with *in situ* buoy temperature measurements to demonstrate the utility of DINEOF method. This DINEOF method will also be used in the present study.

1.4. Purpose, Objectives and Expected Outcomes

The main theme of the present work is to (i) develop the Data INterpolating Empirical Orthogonal Function (DINEOF) method as a tool to fill the gaps in the satellite data for the Black Sea, (ii) apply the regionalization approach to the long-term Chlorophyll concentration data, and (iii) to use the derived products for elaborating our present understanding of the Black Sea circulation and ecosystem dynamics. As the methodology to confront with large data sets having large number of empty points is new and an open research problem, first part of the present study deals with implementation of an improved statistical method to cover successfully the gappy data caused by the clouds.

The data sets used during this study covers the weekly, seasonal and annual composite chlorophyll concentration for 1997 – 2007. The weekly composites will however constitute the main data set for most of the analyses, although others are also used for specific purposes. The DINEOF is a self-consistent filling mechanism which can interpolate even 50 % missing data surprisingly well. The method introduced to the scientific community by Beckers and Rixen (2003) and it's been showed that it is applicable to the satellite data sets by a study by Alvera-Azcarate *et. al.* (2005). DINEOF method will be explained in details in Chapter 2.

As the remotely sensed oceanic data becomes widely available to the oceanographic community, studies on categorizing marine systems into provinces having different biological-physical characteristics with the help of satellite data began to emerge in the oceanographic community (e.g. Yentsch, 1965; Platt and Sathyendranath, 1988; Longhurst *et. al.*, 1995).

Using CZCS Chl-a data collected between 1979 – 1986, Longhurst et. al. (1995) separated the World Oceans into 57 biogeochemical provinces. Black Sea was a part of the province named MEDI, together with Mediterranean Sea. The study also used a model to convert the Chl-a data into global net primary production. Longhurst (1998) further improved the regionalization approach in the book "Ecological Geography of the Sea". The Black Sea was again considered together with the Mediterranean, Marmara and Aegean Seas. Further details on more localized regionalization of broad areas such as the Mediterranean-Aegean-Black Sea system were left for future studies. Platt and Sathyendranath (2008) used SeaWiFS derived Chl-a data (1998 – 2003) as an ecological indicator for local regionalization of the pelagic zone of oceans. This study also presented some new methodology for data extraction and analysis. In the second part of this study a regionalization process is adopted to represent the basin in the form of distinctly different provinces in terms of their biological characteristics, following the work by Longhurst (1998). This approach is new and forms a step forward with respect to the previous studies in which the Black Sea was regionalized on the basis of its physical characteristics. The new approach is considered to be critically important for a more realistic identification of its primary production characteristics.

The expected outcomes of the present study are two-folds. The first one concerns with the use of the method developed here for other studies conducted around Turkey by different research groups. The ready-to-use form of the software with some documentation should be sufficient for this purpose provided that the user community has a sufficient background on the use of statistical methods and the EOF method. For example, the analysis can be easily extended to other Turkish Seas for the analysis of both chlorophyll concentration and sea surface temperature data.

The second outcome is to elaborate our understanding of the Black Sea biological production characteristics in terms of their mesoscale variability both in space and time. Filling the spaces with cloud coverage allow a better recognition of mesoscale dynamics. Besides, the products can form a basis for other studies such as boundary, initial conditions and/or data assimilation in the ecological models.

CHAPTER 2

MATERIAL AND METHOD

Handling huge data size constitutes a challenge that should be confronted while setting up the processing and analysis methods. This problem arises due to the use of sufficiently long time series (11 years for 1997 – 2007; 474 weakly images) for elaborating the observed circulation and/or biogeochemical features as detailed as possible with sufficient resolution (weekly in time and 9 km in space). Another problem is to confront with the data gaps due to the cloud coverage and to obtain temporally and spatially complete and continuous data sets. Below, technical and methodological details are described in terms of data retrieval, processing and analysis. We also provide main aspects of the EOF method followed by a brief description of the DINEOF method for filling the data gaps and its application, and finally the regionalization approach to the data.

2.1. Retrieving and Processing the SeaWiFS Data

The data corresponding to summer months have relatively low cloud coverage, but it possesses a major problem for the rest of the year. Figure 2.1 displays a set of daily, weekly and monthly images to show how the error due to missing points on daily data files propagate on weekly and monthly composites. The daily images in Figure 2.1 are randomly chosen to represent different percentage of gaps in the data. As expected, they always involve more data gaps with respect to others. In the case of multiple pass of the satellite from the area of interest, binning these data sets reduces the number of missing points. For example, in Figure 2.2, three images from the same day with different passes of the satellite may be combined to reduce slightly the number of missing points. However, dealing with the daily images

for the multi-yearly time series is impractical and the analysis is based on 8day binned data sets (hereinafter also called as weekly-binned data sets).

The SeaWiFS data are generally presented in the form of Local Area Coverage (LAC) and Global Area Coverage (GAC). They are provided either as "regular" or "merged" data sets. Merged LAC (MLAC) and merged GAC (MGAC) involve data from ground stations with more spatial coverage. Therefore, we prefer to use the merged LAC and GAC data sets which will be simply called LAC and GAC henceforth.

LAC has 1x1 km resolution that is the maximum resolution of the sensor. GAC can have 4x4 km or 9x9 km resolutions. But, for the case of binned data, NASA (National Aeronautics and Space Administration) distributes only 9 km resolution version. Therefore, when we compare the GAC data with the LAC data here; we employ its 9 km resolution version. In Figure 2.4, the weekly-binned LAC data from September 1999 is presented. In these images cloud coverage is also noted. The LAC data also involve some noise in the form of a red line crossing over the Sea of Azov along the east-west direction (see Figure 2.4 (a), (c)). The noise however disappears in the case of weekly-binned GAC data (Figure 2.5). To further explore how these two different data sets with different resolutions are good for representing the Chl-a features, the cloud-free monthly binned images for September, 1999, are used. The GAC image having a coarse resolution captures almost all features shown in the fine resolution LAC image except more pronounced spatial gradients along the boundaries of filaments, eddies and locally along offshore boundary of the Rim Current. The GAC image also includes some noise patterns, as spikes, generally in the size of data pixel. On the other hand, in the weekly images, the LAC images have more noise which GAC images do not have. Therefore weekly-binned 9 km spatial resolution GAC images seem to be the most appropriate choice for the study.

The SeaDAS software (version 4.9.4) provided by NASA is used for processing the data. It is a comprehensive image analysis package for the processing, display, analysis, and quality control of ocean color data using IDL (version 6.1) and Linux Fedora Core 2 platforms. For processing the raw data, the SeaDAS uses official algorithms developed by NASA-GSFC (Goddard Space Flight Center). The raw data include water leaving radiance

(Lw) at 412, 443, 490, 510, 555, 670, 765 and 865 nm wavelengths, and all other products that are produced from these data by empirical algorithms. An archive of *in situ* oceanographic and atmospheric data, called SeaBASS (SeaWiFS Bio-optical Archive and Storage System), are used for the algorithm development and product validation.

For the chlorophyll-a product the algorithm used is called OC4v4 (Ocean Color 4 version 4). OC4v4 is produced by the group at GSFC and is a fourth order polynomial equation. This equation gives Chl-a concentration in mg/m^3 by the empirical formula:

$$Chl - a = 10.0^{(0.366 - 3.067 R_{4S} + 1.930 R_{4S}^2 + 0.649 R_{4S}^3 - 1.532 R_{4S}^4)}$$

where $R_{4S} = \log_{10}(R_{555}^{443} > R_{555}^{490} > R_{555}^{510})$ and R_{4s} denoting the remote sensing reflectance and the "s" representing the code for specific satellite sensor, which in this case is SeaWiFS. $R_{\lambda_j}^{\lambda_i}$ represents a compact notation for the $R_{rs}(\lambda_i) / R_{rs}(\lambda_j)$ band ratio with the number part denoting the specific bands used (O'Reilly *et. al.*, 2000). The OC4v4 algorithm is previously tested for Black Sea and validated by the *in situ* data and proved to be working in the limits suggested for SeaWiFS data (Sancak *et. al.*, 2005).

As explained before, the weekly-binned merged SeaWiFS data (version 5.2) provided at the FTP site, (http://oceancolor.gsfc.nasa.gov/REPROCESSING/SeaWiFS/R5.2/) are used by cropping the Black Sea region for the longitude band of $26^{\circ} - 42^{\circ}$ E and the latitude band of $40^{\circ} - 47^{\circ}$ N. This gives an area with the size of 1290 km by 779 km. The full data set comprises 474 weekly-binned data covering eleven year period from the launch of the SeaWiFS sensor (September 1997) until the end of 2007.

The data is then transformed to the Mercator projection with SeaDAS, in order to put equal weight on each pixel (1x1 km) in the analysis. The maps obtained by this procedure have dimensions of 1290 km along the west – east direction and 779 km along the south – north direction. The processed data are retrieved in ascii form for further analysis performed with the Fortran code and the Matlab software. Each single matrix with dimensions of 1290x779 contains the weekly-binned GAC SeaWiFS Chl-a data. The ascii

files put together to form a 3D matrix with spatial dimension of 1290 x 779 and temporal dimension of 474. Using Matlab program (version 7.0.4.352 R14 Service Pack 2 running under Linux Fedora Core 2). this matrix is first converted to a format that is required for running the DINEOF code in Fortran 90. Resulting reconstructed matrix and the EOF analysis is then put back into the Matlab program for further processing and visualizations.

Figure 2.3 represents the 46, 8-days binned, images for the year 1999. The color-plate used in these images represents the chlorophyll concentration in the range $0 - 4 \text{ mg/m}^3$ varying linearly contrary to the default broader logarithmic range between $0 - 64 \text{ mg/m}^3$ in the SeaWiFS Data Analysis System (SeaDAS).



Figure 2. 1. Comparison of daily, weekly and monthly SeaWiFS derived Chla images from October 1997. Row (a) is daily images, 7th, 11th, 29th, 30th October. Row (b) is weekly images, 4 consecutive weeks of October. Row (c) is monthly binned October image.



Figure 2. 2. Daily images of the same day. 30th October 1997. Satellite passes at: (a) 08:51 a.m., (b) 10:31 a.m., (c) 12:09 p.m..



Figure 2. 3. Weekly-binned Chl-a images of the year 1999. Satellite imagery assumes a week as 8 days, resulting, 46 weeks in a year.





2.2. Statistical Method: EOF

The Empirical Orthogonal Functions (EOF), which is also known as the Principal Component Analysis (PCA), is a tool for spatial-temporal analysis of data sets. It can be used to compress and interpret the large amount of spatial-temporal information. This method can be summarized as follows.

M x N data Matrix F is:

$$F = \begin{bmatrix} F_1(1) & F_1(2) & \dots & F_1(N) \\ F_2(1) & F_2(2) & \dots & F_2(N) \\ \dots & \dots & \dots & \dots \\ F_M(1) & F_M(2) & \dots & F_M(N) \end{bmatrix} \quad \text{Location (M)}$$

Spatial covariance matrix of the field $F_m(t)$ is:

$$\mathbf{R}_{FF} = \mathbf{F} * \mathbf{F}^{\dagger}$$

$$R_{FF} = \begin{bmatrix} \langle F_1 F_1 \rangle & \langle F_1 F_2 \rangle & \dots & \langle F_1 F_M \rangle \\ \langle F_2 F_1 \rangle & \langle F_2 F_2 \rangle & \dots & \langle F_2 F_M \rangle \\ \dots & \dots & \dots & \dots \\ \langle F_M F_1 \rangle & \langle F_M F_2 \rangle & \dots & \langle F_M F_M \rangle \end{bmatrix}$$

where, $\langle F_i F_j \rangle$ is the covariance between time series F_i and F_j (F at locations i and j) defined as:

$$\langle F_i F_j \rangle = \langle F_j F_i \rangle = \frac{1}{N-1} \sum_{t=1}^N F_i(t) F_j(t)$$

where, i,j = 1...M

The matrix product R_{FF} is symmetric and square, even if F itself is not square. It is important to note that, if the data series in F are normalized by the standard deviation, the matrix R_{FF} is formally the correlation matrix instead of the covariance matrix. After the covariance matrix has been calculated from the data we solve the eigenproblem:

$$R_{FF} * E = E * \Lambda$$

Here Λ is the M x M diagonal matrix containing the eigenvalues λ_k of R_{FF} :

	$\lceil \lambda_1 \rceil$	0	 0]
Λ —	0	λ_2	 0
Λ –			
	0	0	 λ_{M}

The square matrix E (the eigenvector matrix) has dimension M x M; its column vectors E^k are the eigenvectors of R_{FF} corresponding to eigenvalues λ_k :

There is K non-zero eigenvalues and, hence, only K eigenvectors are used in the decomposition; so, the effective dimension of matrix E is M x K, where M are the spatial locations and K are the modes of the EOF decomposition. Each eigenvector E^k represents the spatial EOF pattern of mode k (it has dimension M, that is, the number of locations in the original maps). Spatial EOF patterns referred as the *Empirical Orthogonal Functions* or Simply *EOFs* (also called *Principal Vectors or Loadings*).

The time evolution of the k^{th} EOF is given by the time series $A^{k}(t)$, which is obtained by projecting the original data series $F_{m}(t)$ onto eigenvector E^{k} and summing over all locations m:

$$A^{k}(t) = \sum_{m=1}^{M} E_{m}^{k} F_{m}(t)$$

where, m=1...M counts the locations, t=1...N counts the time steps and k=1...K counts the EOF modes. In matrix notation, matrix A is obtained by multiplying matrices E^{\dagger} and F:

$$A = E^{\dagger} * F$$

where E^{\dagger} is K x M and F is M x N making A, which is a K x N matrix. Rows in matrix A are time series of length N that is the number of time steps in the original time series. They are referred as the *Principal Components* or *PCs* (also called *time series of Expansion Coefficients, Time Coefficients, Eigenvector time series* or *Scores*).

Each eigenvalue λ_k is proportional to the percentage of the variance of the field F that is accounted for by mode k. This percentage is calculated as:

% variance Mode k =
$$\frac{\lambda_k}{\sum_{i=1}^{K} \lambda_i} * 100$$

It is very important to note that, eigenvectors, E^k , is orthogonal in space and PCs, A^k , is orthogonal in time. E^k are uncorrelated over space and A^k are uncorrelated over time.

2.3. Filling the Gaps with the Statistical Method DINEOF

Data Interpolating Empirical Orthogonal Functions, DINEOF, is a parameter free EOF-based method for the reconstruction of missing data. It is parameter free because the necessary parameters are derived from the existing data, as explained by Beckers and Rixen (2003) and Alvera-Azcarate *et. al.* (2005). Since it uses EOF for the reconstruction, EOF analysis of the data is its by-product. The way in which the DINEOF method works is outlined below, following Alvera-Azcarate *et. al.* (2007).

Data are stored in a matrix **X** with the temporal and spatial average subtracted *a priori*. Missing data are initialized to zero, to guarantee that they are unbiased with respect to **X**. With this first guess, a first Singular Value Decomposition (SVD) is realized with one (k = 1) EOF. The missing data are then replaced using the obtained EOF:

$$\mathbf{X}_{i,j} = \sum_{p=1}^{k} \rho_p \left(\mathbf{u}_p \right)_i (\mathbf{v}_p^T)_j$$

where *i*, *j* are the spatial and temporal indexes of the missing data in matrix **X**; u_p and v_p are the p^{th} column of the spatial and temporal EOF U and V, respectively; and ρ_p is the corresponding singular value, with p = 1...k. With
the new values for the missing data, the SVD decomposition is performed again. The last two steps are repeated until convergence is obtained for the missing values. Then, the whole iterative procedure is performed for k = 2, 3, ..., k_{max} EOFs, where k_{max} is a predefined number that should vary according to the initial matrix characteristics. For each k, estimation for the missing values is obtained. The optimal number of EOFs retained for the reconstruction is determined by cross-validation: a number of points (typically 1% of the initial data) are set aside and considered as missing. For each EOF estimation, the error between those initial points and their reconstruction is calculated, so the optimal number of EOFs minimizing this error can be determined. DINEOF calculates the EOFs using a Lanczos method (Toumazou and Cretaux, 2001), which reduces the calculation time, thus, allowing the method to be applicable to large data sets, (Alvera-Azcarate et. al., 2005). More detailed formulation of the method is provided in Appendix A. as well as Beckers and Rixen (2003), Alvera-Azcarate et. al. (2005, 2007, 2009), Beckers et. al. (2006).

When compared with the Optimal Interpolation (OI) method, main advantages of the DINEOF method are that (i) it is faster and computational cost of the OI method may be prohibitive when using large matrices; (ii) contrary to other methods it does not require any a priori information of the statistics of the reconstructed data set (such as covariance or correlation length) as in the case of Optimal Interpolation method, thus reducing the subjectivity of the analysis; (iii) it is more accurate, when compared to *in situ* data, by an order of ~2.50 (using 452 independent *in situ* observations for validation), and DINEOF performs reconstruction of the data set nearly 30 times faster than with OI. According to Beckers *et. al.*, (2006), the DINEOF interpolation was shown to provide better results than optimal interpolation. The DINEOF method has also been compared to the Kriging method and was found to be more accurate for high temporal resolution and for typically 50 % data gaps (Gunes *et. al.*, 2006).

The code of DINEOF method is written in Fortran 90 by GeoHydrodynamics and Environment Research (GHER) group, University of Liege, led by Prof. Jean-Marie Beckers. The first version of the code was kindly provided by Prof. Jean-Marie Beckers, and its updated version later by

Dr. Aida Alvera-Azcarate. The final version of the program is now available on a web site, http://modb.oce.ulg.ac.be/mediawiki/index.php/DINEOF. The calculations for the EOF mode are performed using the Lanczos solver, which is provided by Toumazou and Cretaux (2001). This part uses the ARPACK freeware (Lehoucq *et. al.*, 1997) in the DINEOF Fortran 90 program.

2.4. Regionalization Method

In this study the term regionalization will stand for separating the Black Sea into its provinces according to their chlorophyll concentrations. The provinces are also termed as bioms following its definition by Campbell, (1996) as "the world's major communities, classified according to the predominant vegetation and characterized by adaptations of organisms to that particular environment". The first complete biom definition applied into marine environment is provided by Alan Longhurst (1998). Longhurst considered Chl-a as the basic indicator of primary production and used it for the regionalization of the seas on the global scale and considering regional seas as single bioms. In Longhurst (1998), the Black Sea was a single region together with Sea of Marmara, Aegean and Mediterranean Seas.

So far there are some regionalization studies applied on the Black Sea. Nezlin and Dyakonov (1998) separated the basin into eight regions on the basis of topographic characteristics of the basin to study the Chl-a concentration distribution obtained by the CZCS sensor for 1978 – 1986. Çokacar and Özsoy (1998) also identified twelve different regions depending on the topography and circulation patterns. Kopelevich *et. al.* (2002 and 2004), compared *in situ* and CZCS derived Chl-a data (1978 – 1986) and SeaWiFS Chl-a data (1998 – 2001) in eight subregions which were distinguished with respect to bathymetry, spatial distributions of CZCS values, surface currents and anthropogenic impacts.

The regionalization is later improved by Platt and Sathyendranath, (2008) in order to handle local applications in more detail. The method is simple, yet powerful enough to make bioms representative for the ecosystem. The bioms are determined simply by the sea surface Chl-a

concentration in the region. The concentration limits for each sub-region are set according to the regional characteristics.

CHAPTER 3

DINEOF RECONSTRUCTION OF THE DATA GAPS

3.1. Setting Up the Method by Sensitivity Tests

Weekly-binned Chl-a data of the year 1999 presented partly earlier in Figure 2.3 are used to explain basic features of the DINEOF method here. The first run with the DINEOF uses the default setting of the parameter values. The resulting map (Figure 3.1.) removes the land points from the original data and the remaining points form the 3D matrix for the subsequent analysis. The default settings are listed as follows:

- nev value = 20, sets the numerical variables for the computation of the required singular values and associated modes, or simply the number of modes calculated, (Toumazou and Cretaux, 2001).
- ncv value = 30, sets the maximal size for the Krylov subspace and cannot be changed as long as ncv > nev+5, (Toumazou and Cretaux, 2001).
- The cross-validation points are set to be chosen internally, (Beckers and Rixen, 2003; Alvera-Azcarate *et. al.*, 2005, 2007; Beckers *et. al.*, 2006).

The reconstructed 3D matrix constitutes a completely new data set, both in terms of missing parts in the original matrix and the original data itself. The code offers an option whether to include only missing values computed by the algorithm into the original data set or to form a new data set for the entire basin as a product of the DINEOF analysis. In complete reconstruction, the data is naturally more smoothed avoiding spikes at the cloud edges and eliminating any noise in the initial matrix. The results of two options are presented in Figure 3.2 for the reconstruction of only missing points and Figure 3.3 for a complete reconstruction over the entire basin. It is referred to Figure 2.3 for the original data with missing areas for further comparison of the reconstructed areas. It seems that the reconstructed areas are separated from their surroundings by spikes in Figure 3.2. The difference being less than 1 mg/m³ seems to be high and not realistic. The method in fact tends to estimate the missing values close to the overall mean of the 3D matrix, and the overall mean of the basin can be higher than the points reconstructed. This will be discussed in details in the following sections.



Figure 3. 1. The map used for land masking of the data. 40 - 47 North and 26 - 42 East.

When a complete reconstruction applied to the initial data matrix, the overestimation mentioned above is observed to spread over the whole basin, causing a more smoothed product with higher than the initial values. Some samples of such weeks are presented in Figure 3.4. It should also be noted that in some rare cases (e.g. the weeks number 3, 6 and 9) reconstructed data yield an underestimate with a difference of less than 1 mg/m³. Therefore, it may be concluded that it would be better to limit for the reconstruction of only missing points. However, more tests will be done to elaborate this conclusion further in the following sections.



Figure 3. 2. Reconstructed weekly-binned Chl-a images of the year 1999. Only missing points are reconstructed.



Figure 3. 2. (cont.) Reconstructed weekly-binned Chl-a images of the year 1999. Only missing points are reconstructed.



Figure 3. 3. Reconstructed weekly-binned Chl-a images of the year 1999. Complete reconstruciton of the data.



Figure 3. 3. (cont.) Reconstructed weekly-binned Chl-a images of the year 1999. Complete reconstruction of the data.



Figure 3. 4. Selected over estimated reconstructions. Only missing points reconstructions and complete reconstructions are matched for random 3 weeks.

To understand the effects of reconstructions on the overall data, the data for 46 weeks are averaged to form the annual mean Chl-a map of year 1999, (Figure 3.5). Only missing points reconstruction and complete reconstruction can be compared with the original data. A general overestimation in the case of complete reconstruction is shown in Figure 3.5c. On the contrary, the results for the case of reconstruction only for missing points represent features of the original data better (Figure 3.5b). It is obvious that the overestimation by the complete reconstruction is much higher than the one resulting from only missing points reconstruction, especially at the interior basin with lower Chl-a values.

The under (and over) estimated points in the reconstructed images are presented in Figure 3.6. To obtain these maps, the annual-mean

reconstructed data is subtracted from the original annual mean. Both under and over estimates are found to be less than 1 mg/m³, as mentioned before. In complete reconstruction both over and under estimates exceeds the case for the reconstruction of only missing points; hence, the latter option provides a better choice. The comparison of products in Figure 3.6 also delineates that over-estimates generally correspond to the regions of high Chl-a regions (i.e. Azov, North Western Shelf, Rim Current) whereas the under-estimates are more dominated in the low Chl-a regions (i.e. interior region). The "no difference" points seem to be situated between these high and low Chl-a areas (see Figure 3.6a). The superiority of the missing point reconstructions with respect to the overall basinwide reconstruction can be further seen in the basin averaged Chl-a values for the entire 1999 period in Figure 3.7.



Figure 3. 5. Annual mean Chl-a maps of year 1999. (a) is constructed from original data, (b) is constructed from output of only missing points reconstruction, (c) is constructed from output of complete reconstruction.



Figure 3. 6. Under and over-estimates of the reconstructions in Figure 3.5..
(a) original data – only missing points reconstruction, (b) original data – complete reconstruction.



Figure 3. 7. Time versus spatial basin average Chl-a graphic. Data from year 1999.

3.1.1. Test 1

The advantages of reconstructing only the missing points are shown in the previous section. For further tests, a new data set that is smaller in size for practical reasons is formed from the original 1999 data set (see Figure 3.2). It consists of 27 images that are assumed to represent 27 consecutive weeks (Figure 3.8). The new matrix assumed to be a real "data complete" set to perform further sensitivity tests. The images with spikes at the edge of reconstructed areas, like the ones in Figure 3.4, are discarded and the remaining weekly Chl-a images are used to re-construct a new 3D matrix. In all DINEOF runs for all tests, the reconstruction is only made for missing points.

The aim of the first test is to find out the best values for nev (the number of EOF modes wanted to be calculated) and ncv (the maximal size for the Krylov subspace) values described in section 3.1. Since, ncv > nev+5 should be satisfied, all the values we assign satisfy this condition. Also, ncv cannot be greater than the temporal dimension of the matrix, so the maximum ncv tried in this test is 27, (Toumazou and Cretaux, 2001). Only a single run made with ncv = 30 for a comparison with ncv = 27. In total, 7 different runs made with 7 different nev, ncv values.

An artificial cloud with a size of 200x300 pixels (or equivalently kilometers) was put over the North Western Shelf (NWS) in the middle of 14th week of the 27 weeks dummy set. Then a series of different runs with different combinations of the parameters "nev" and "ncv" as listed below are performed.

Run 1.	nev=1 , ncv=7
Run 2.	nev=1 , ncv=25
Run 3.	nev=2 , ncv=8
Run 4.	nev=10 , ncv=16
Run 5.	nev=15 , ncv=25
Run 6.	nev=20 , ncv=27
Run 7.	nev=20, ncv=30

For each run, a reconstructed 200x300 image of the Chl-a values are calculated and compared with the original matrix without cloud, and percentage error maps are obtained. The resulting maps are almost identical and may be put into three groups (Figure 3.9).

Run 7 was made to see if the percentage error in Run 6 is due to setting an ncv value equal to the temporal dimension of the matrix, as ncv needs to be less than or equal to the temporal dimension of the matrix. It seems that even if we assign ncv a value greater than the temporal dimension of the matrix, the percentage error map is same as ncv = temporal dimension of the matrix. This error map is the highest off all three maps in Figure 3.9. For presentation purposes the color scale in Figure 3.9 is set to 0 – 1, but for Figure 3.9(c) the maximum percentage error increases to 6, indicating 6 times higher than the original value. This therefore demonstrates that ncv cannot be set greater or equal to temporal dimension.

Run 1 and Run 2 show that choosing one EOF mode is not a good choice to compute the reconstructed values efficiently. It can be concluded that for Black Sea the value of nev can be anything as long as ncv > nev+5 and $ncv \neq 1$. For optimizing computation time, the sensitivity experiments suggests the choice of their values as nev=2, ncv=8. For the corresponding reconstructed image and its comparison with the original image it is referred to Figure 3.10. Additional sensitivity runs are performed to see if the percentage error map is changed with the different combinations of these two parameter values. It is found that as long as they satisfy the conditions given above, the error maps are identically same; thus, they are not shown once again.

Figure 3. 8. Artificial dummy data set constructed for the DINEOF tests. Total 27 weeks in the set.









Figure 3. 9. Percentage errors maps for test 1. (a) is nev=2, ncv=8; nev=10, ncv=16 and nev=15, ncv=25, (b) is nev=1, ncv=7 and nev=1, ncv=25, (c) is nev=20, ncv=27 and nev=20, ncv=30.



Figure 3. 10. (a) Week 14 of the "dummy set", (b) map with cropped 200x300 km artificial cloud, (c) reconstructed result of the test 1, run 3.

3.1.2. Test 2

The second test repeats the first one by putting the artificial cloud used in the previous test (see Figure 3.10b) to five different weeks sequentially, i.e. one at a time. The selected weeks are 1st, 8th, 14th, 20th and 27th weeks. Five different run has been made, each of which involves only one missing data. The original and reconstructed matrixes (200x300) are presented in Figure 3.11, together with the corresponding percentage error maps. The maximum percentage error is around 1 in all five weeks. Although, the values might be different in some parts of the reconstructed images, the features are correctly reproduced.





Figure 3. 11. Original and reconstructed Chl-a maps together with related percentage error maps for test 2.

3.1.3. Test 3

It is obvious that the gaps on data due to the clouds or for other reasons could occupy much bigger areas in real situations than the 200x300 matrix simulated in the test 1 and 2. Therefore, the size of the artificial cloud is enlarged to see how it will affect the reconstruction. Three different artificial cloud regions with the sizes of 350 x 400, 500 x 500, and 600 x 800 km² are put into the 14th week image. The reconstructions for each of these cloudy images are presented in Figure 3.12. As in the case of test 1 and test 2, the maximum percentage error is around 1 (Figure 3.13). This maximum value occurs in the regions where the Chl-a values are the greatest. This is similar to the results of the first two tests.







Figure 3. 13. Percentage error maps for test 3.

3.1.4. Test 4

Each of the previous test runs is performed for a single day corresponding to a single image. In this test, the artificial clouds for test 1 and test 2 are applied to five weeks, (1st, 8th, 14th, 20th, 27th) forming a data set with five image with missing data at the same location. The reconstructed images are presented in Figure 3.14, and the related percentage error maps are presented in Figure 3.15. The maximum percentage error is around 2, which is higher than the value obtained in the previous tests. However, the features under the cropped parts of the original images are reconstructed with reasonable accuracy.



Figure 3. 14. Test 4 results. Original images (left) versus reconstructed (right) images.





Figure 3. 15. Percentage error maps for test 4.

On the basis of these test runs, it is therefore concluded that the size of the missing area is not critical for the quality of the DINEOF based reconstructions. Even if 5 weeks out of 27 weeks (~20%) have the same cloud at the same area, the results are acceptable and percentage error is around 2 at most. Hence, the important point is that the features reconstructed are not artifacts of the DINEOD methodology. In these tests, since the clouds are artificial and we know what they are hiding, the reconstruction results can be compared with the real data. The features that are removed are recovered quite efficiently after the reconstruction.

3.2. Construction of the Gap Free Data Sets

As mentioned earlier the complete data set consists of 474 consecutive weeks between 14 September 1997 and 31 December 2007. The complete set prior to the DINEOF reconstructions is presented in the Appendix B, Figure B.1. The data sets with large gaps start around the mid-autumn and end by the early spring encompassing roughly 20 weeks. The remaining 25 weeks involve nearly full data coverage. To track the seasons and years of the weekly-binned images we refer to Table B.1 and Table B.2 in Appendix B. The former one shows the seasons with their relevant week numbers and Table B.2 yields the years with their relevant week numbers.

The data set presented in Figure B.1 used the parameterization values that were described earlier in this section. The resulting reconstructed 474 consecutive gap free images are presented in Figure B.2. It seems that general low productive phase with relatively low chlorophyll concentrations over the basin starts by the beginning of spring season and ends by the beginning of autumn. Year 1999 is an exception for that, in which high chlorophyll period begins by the end of summer. An early bloom starts by week 85 and ends by week 92, affecting the western and southern part of the basin reaching to its maximum at week 90. Then, the normal high chlorophyll period begins at week 94, resulting higher Chl-a all around the basin.

Another exceptional case is detected during 2000 – 2002. Low chlorophyll period ends around the week 139 and the bloom period continued afterwards until the week 206. Therefore, as shown for the year 2001, the relatively poor chlorophyll periods never occurred in 2001. Hence, an extraordinary situation occurs for this year and we observe the maximum Chl-a concentrations of the ten years of the data coverage at week 173 in summer 2001. Starting by the week 171 a very strong bloom occurred at the eastern cost and expanded the whole basin. Reaching at a maximum level at week 173 then switched into a decreasing trend. The post bloom signal is still detectable at week 179, around 50 days after the bloom maximum. Then another bloom dominated the centre of the basin as will be analyzed more deeply below.

When the reconstructed images are studied in details, some of the images show that the reconstructed parts are separated from the original

data with sharp gradients (e.g. the weeks 341, 342 and 344). Although the estimated values after reconstruction can be either over- or under-estimated, it is more likely to be overestimated for the weeks 341, 342 and 344. This arise a question whether the code tends to fill the gaps towards the overall mean value of the data set, which might be a natural outcome of the nature of the method. The overall mean of the data set is higher than the Chl-a values of the interior basin. This appears to be caused by very high coastal values and in the North Western Shelf ($\sim 3 - 10 \text{ mg/m}^3$), involving much higher concentrations than the values at interior parts (<1 mg/m³). Interestingly the spikes detected within the interior parts may end up with relatively high concentration values with respect to their relatively low original values.

Next, it is analyzed if this effect can be eliminated by excluding the Sea of Azov and the Marmara Sea from the original data sets, and by removing all the noises introduced by lakes and rivers. It is obvious that the Chl-a contents of the Azov Sea and the Marmara Sea are higher than the Black Sea. By removing these data the overall mean of the data set is expected to decrease. When this operation is fulfilled and the DINEOF code was run again using the same parameter setting, the resulting maps of 474 weeks after the reconstruction are presented in Figure B.3. It seems that the assumption made above was a valid one. When the cloudy weeks 341, 342 and 344 are checked again and compared with the results of the previous tests, it is observed that this problem is no more present.

3.3. EOF Analysis Result

This section briefly describes the spatial and temporal characteristics of the EOF modes. As mentioned earlier, each EOF mode has a certain percentage of weight. This percentage indicates how much of the data can be explained by each mode. It is found that mode1 contributes 76.6854 %, of the whole data set whereas mode 2 and mode 3 contributions are limited to 8.8578 % and 3.3329 %. The others are as follows: mode 4 = 2.9569 %, mode 5 = 2.5395 %, mode 6 = 1.7176 %. Thus the first six mode explains, 96,0901 % of the data set, but 76% of it comes from only the first mode. Trend removal is not applied to the data set. The reason for this is to detect any possible trend while analyzing long term behaviour of the data set. Hence, the annual cycle is the most powerful signal of the data.

The first two spatial modes (Figure 3.16) have much in common with the general Chl-a pattern of the basin. Since, no trend removal made on the data set, this is expected and the signal explains the general Chl-a distribution of the basin. In mode 2, in the negative region located on the north-western part of the basin, there is a small positive area. This area is located on the Danube discharge and probably a signal arises from it. The rest of the basin has no distinctive signal. In mode 3, the distinctive region mentioned in mode 2 is nearly zero while north of it has negative and south positive values. The negative region is weakly connected with the pattern in the NWS. Having both negative and positive signals in the western coast, suggests that the high Chl-a region has its own dynamics that may be decoupled from the rest. In mode 4, there is a positive signal along the Rim Current, apparently caused by higher Chl-a along the current. Although, some signals may be detected in mode 2 and 3 along the current zone, it turns out to be a dominant one in mode 4. This mode has also a different behaviour along the western coast. While a positive – negative switch takes place in the north – south direction in other modes, the signal changes from east to west in mode 4 and this change becomes clearer around the Danube discharge. This signal might be due to the widening and broadening of high Chl-a area along the coast in different seasons. It is therefore obvious that the western coastal zone shows dynamic changes both in north – south and east – west directions. As in other modes, a strong signal also appears along western coast in mode 5 such as a distinct positive signal located around the Danube discharge and a negative signal on both southern and northern side of the discharge. The same region has a more complex signal in mode 6. A small negative zone on the discharge surrounded by a positive area and a negative area begins on southern edge of this area going along the Rim Current. The signal on the Rim Current in mode 6 is similar to the signal on mode 4.

In general three major features emerge from the EOF spatial modes. The first one is the Rim Current as a unique and permanent feature of the basin. The second feature is distinctly different dynamics along the highly productive western coastal zone. The third feature is the lack of eastern and western gyres that are contrary to the findings of the circulation dynamics. Instead, the interior basin biologically has one cellular structure. The first three temporal EOF modes (Figure 3.30) on the other hand reflect the most dominant annual (Figure 3.17 Temporal Mode 1) and sub-annual fluctuations (Figure 3.17 Temporal Mode 2 and Temporal Mode 3) of the data.



Figure 3. 16. EOF results. First 6 mode for spatial component. Colorbar has arbitrary units.



Figure 3. 16. (cont.) EOF results. First 6 mode for spatial component. Colorbar has arbitrary units.



Figure 3. 16. (cont.) EOF results. First 6 mode for spatial component. Colorbar has arbitrary units.



Figure 3. 17. EOF results. First 6 mode for temporal component. Y axis has arbitrary units.



Figure 3. 17. (cont.) EOF results. First 6 mode for temporal component. Y axis has arbitrary units.



Figure 3. 17. (cont.) EOF results. First 6 mode for temporal component. Y axis has arbitrary units.

CHAPTER 4

PROPERTIES OF SURFACE CHLOROPHYLL CONCENTRATION AS INFERRED FROM THE RECONSTRUCTED DATA

4.1. Long-term Average Properties

To observe the general structure of the Chl-a concentration in the basin, the processed data which excludes the Sea of Azov and the Marmara Sea (as shown in Figure 3.18) are averaged temporally for 10 years period from 1 January 1998 to 31 December 2007 to provide the long-term mean chlorophyll concentration distribution (Figure 4.1). When compared with the temporally-averaged original data set in which the Sea of Azov and the Marmara Sea are included, no appreciable differences are observed between these maps. Therefore, only the temporally mean map obtained from reconstructed data set without Azov and Marmara Seas is presented here (Figure 4.1). It reveals an almost horizontally uniform interior basin chlorophyll concentration of about 1.0 mg/m³ that increases to 1.5 m/m³ range along a narrow coastal band and exceeds 3 mg/m³ along the western coastal zone.

Figure 4.2 provides the long-term changes of the basin-averaged chlorophyll variations obtained from the weekly-binned reconstructed maps between 14 September 1997 and 31 December 2007. The basin-averaged Chl-a concentration attains a minimum of about 0.8 mg/m³ in 2003 and peaks of the order of 3.0 mg/m³ during 1998 – 2001 period. In the years that follow the low concentration period of 2003, Chl-a concentration remained at intermediate values of $1.0 - 2.5 \text{ mg/m}^3$. Comparison of Figure 4.1 and 4.2 points to a rather interesting difference. As shown from Figure 4.1, except a coastal zone along the western part of the sea, concentration varies $0.5 - 1.0 \text{ mg/m}^3$ range whereas the basin-averaged concentrations in Figure 4.2 attain
the values around 1.5 mg/m³. This implies that relatively high concentrations confined along the coastal zone have an important contribution to the basin-averaged properties although this zone has a rather small areal coverage with respect to the remaining part of the basin.



Figure 4. 1. Temporally averaged chlorophyll-a data. Data between 1 January 1998 and 31 December 2007 are used.

Figure 4.3 depicts the seasonally-averaged Chl-a patterns obtained from the reconstructed data set for 1998 – 2007. Chl-a content in the basin is seen to be particularly high during autumn and winter while spring and summer seasons are characterized by relatively low Chl-a content. As compared to other seasons, Chl-a concentration is more evenly distributed within the interior basin during autumn, but relatively high Chl-a zone of the interior basin is limited to the eastern basin, particularly along the eastern cost, leaving the western half of the interior basin decreases after winter season, especially within the central basin, while the peripheral zone continues to attain slightly higher values. In summer, low Chl-a concentration



Figure 4. 2. Spatially basin averaged chlorophyll-a data. Weekly-binned Chl-a maps between 14 September 1997 and 31 December 2007 are used for basin averaged Chl-a values.

is evenly distributed throughout the interior basin and higher Chl-a area is limited solely to the Northwestern Shelf (NWS) and along the western coastal zone. In fact, as noted by the seasonal maps Chl-a concentration in the NWS is almost unchanged throughout the years without appreciable seasonal changes. The major seasonal differences, which may be inferred from the maps in Figure 4.3, are a gradual increase of concentration towards east in winter and spring, and a similar change towards south in spring and summer. Some seasonal changes are also noted at the coastal regions.





4.2. Annual Mean Properties

The annually-averaged spatial maps for the years 1998 – 2007 are presented in Figure 4.4a. The trend described earlier in Figure 4.2 is also observed in the annual-mean maps. Year 1998, 1999 and 2000 are characterized by relatively high Chl-a values with the year 2001 being the highest. We recall that summer of this year was abnormally productive. Then a decrease in concentrations starts with the year 2002 that attains its lowest values in 2003. For example, the high concentration productive zone in the NWS even attains a very narrow form in 2003. There is a slight increase in 2004 and 2005 but they are still lower than the years 1998, 1999 and 2000. Relative Chl-a increases in year 2006 and 2007 are at the intermediate levels with respect to high concentration years of the late 1980s and low concentration years of 2003 - 2005 period. In 1999, at the southwestern cost, waters with high Chl-a content apparently flow southward from the NWS towards the Bosphorus entrance region and continues eastward along the Turkish coast. The southwestern region in the vicinity of the Bosphorus entrance is characterized by relatively high Chl-a values which are even higher that those observed in 2001. Other years do not show such an anomalous feature.

In Figure 4.4b – e, further displayed the seasonal Chl-a maps complementing the annual-mean distributions presented previously in Figure 4.4a. Their general features are pointed out previously and therefore, only to unusual features which are not mentioned before are focused. Summer 2001 has more dominant contribution to the annual mean as compared to other years. Especially the western and central basins acquired unexpectedly high Chl-a values. In Summer 2007 the high Chl-a in the NWS is trapped in the northernmost part. However, other seasons of the year 2007 have similar Chl-a patterns. In general, high Chl-a concentrations in the NWS and western coasts seem to occupy relatively small area in autumn than other seasons. This situation however changes gradually in winter and later in spring as the area of high Chl-a values enlarges. The seasonal differences in Chl-a concentrations are clearly related to discharges from the Danube and other northwestern rivers.

The basin-averaged version of these temporal plots is presented in Figure 4.5. The main finding of the plots shown in this figure is the absence of a typical annual pattern repeating every year. Instead, there are appreciable interannual variability with some years having maximum concentrations in spring and summer, that may be even higher than the basin averaged values of winter and autumn seasons, whereas some years possess no definitive spring – summer maximum.







Figure 4. 4. (b) Temporally averaged, seasonal chlorophyll-a maps according to years, 1998 – 2007. Autumn.



Figure 4. 4. (c) Temporally averaged, seasonal chlorophyll-a maps according to years, 1998 – 2007. Winter.











Figure 4. 5. Temporal plots of basin averaged Chl-a values, according to years, 1998 – 2007. Weekly-binned maps from Figure B.3. are used for basin averaged values.



Figure 4. 5. (cont.) Temporal plots of basin averaged Chl-a values, according to years, 1998 – 2007. Weekly-binned maps from Figure B.3. are used for basin averaged values.



Figure 4. 5. (cont.) Temporal plots of basin averaged Chl-a values, according to years, 1998 – 2007. Weekly-binned maps from Figure B.3. are used for basin averaged values.

4.3. Regionalization of Chlorophyll Data

As remarked earlier in the preceding chapters, Chl-a concentration may be used to identify different provinces having different primary production characteristics as suggested by Longhurst (1998) and Platt and Sathyendranath (2008). The most critical aspect of such regionalization concept is to be able to identify the boundaries that are meaningful for the region. Within the scope of the present work, different intervals are assigned for the Black Sea spanning from very fine to coarse intervals. It turns out that, as expected, the choice of intervals are not very important at higher Chl-a values because Chl-a containing biota composition is not changing much at such high levels of Chl-a values, (Longhurst, 1998; Platt and Sathyendranath, 2008). The difficult part is at low level of Chl-a values, which is generally located at the interior basin. In order to precisely choose the intervals, a very fine regionalization with 0.1 mg/m³ intervals are tested first for the concentration range of 0 - 1 mg/m³ and with 0.25 mg/m³ at higher concentrations On the basis of this fine scale approach, the intervals are finally optimized as follows (in mg/m³):

 $0.00 \le \text{Region } 1 \le 0.25$ $0.25 < \text{Region } 2 \le 0.50$ $0.50 < \text{Region } 3 \le 0.70$ $0.70 < \text{Region } 4 \le 0.80$ $0.80 < \text{Region } 5 \le 0.90$ $0.90 < \text{Region } 6 \le 1.00$ $1.00 < \text{Region } 7 \le 2.00$ $2.00 < \text{Region } 8 \le 3.00$ 3.00 < Region 9

All weekly binned images were used in the set up for the intervals. Both incomplete and reconstructed data were used. Therefore, some regions may not occur in some maps. For example, the regionalized map obtained from complete 10 years data contains only 6 regions. This regionalization is applied to 10 years averaged, seasonal, annual and weekly data and seasonal data that is separated by years. The 9 regions are color coded in the regional maps.

First, temporally-averaged Chl-a map (Figure 4.6) obtained from 10 years long data that was presented earlier in Figure 4.1 reveals most of the main features of the regionalization and forms a reference for the others for comparison purposes. Figure 4.6 shows six regions of different chlorophyll concentrations starting with the Region 4 (0,70 - 0,80 mg/m³ range). The first three intervals with lower concentrations do not show a distinct signal for this long-term averaged data set. This map indicates that the interior basin is formed by one composite cell contrary to two separated gyres inferred from the circulation pictures. Therefore, the circulation and ecosystem dynamics of the interior basin may not necessarily match in terms of spatial variability of the physical and biological processes. This does not however mean that the biological processes are decoupled from the physical processes.



Figure 4. 6. The Black Sea separated into 9 regions according to the SeaWiFS derived ocean color data collected between 1998 – 2008. Though, only 6 regions can appear.

Data prepared for the ocean color maps presented in Figure 4.3 are used to produce the seasonal regionalized maps (Figure 4.7). The autumn map reveals only four regions (Region 6 - 9) starting with concentrations greater than 0.9 mg/m³. Lower values are not common except small number of outliers. In winter, the range shifts to Region 5 - 9 that implies a slightly lower concentration of 0.8 mg/m³ within the interior basin. The rest is almost similar with the autumn map. A switch to lower concentrations (up to 0.5 mg/m³) within the interior basin takes place in the spring and summer months, thus giving rise to more pronounced onshore-offshore gradients around the basin. As the spring map is mainly characterized by the Regions 4 – 9 with only small contribution from the Region 3, the summer map involves a much broader Region 3 extending almost entire interior part. Regions 1 and 2 do not appear in any of these seasonal maps. The Regions 9 and 8 are more or less the same in the autumn, winter and spring maps. Also, the outer border of the Region 7, along the coasts is stable in these seasons. However, these regions experience more changes in the summer. The Region 7 mostly disappears, leaving its space to regions with lower Chla values, while the Regions 9 and 8 are mostly confined into the northern



Figure 4. 7. Regionalized seasonal maps of the Black Sea. 1998 – 2007 data are used to obtain the maps.

basin, and is replaced by the Region 7 along the south coast.

The regionalization version of the annually-averaged Chl-a concentration data given previously in Figure 4.4 is shown in Figure 4.8. The regions obtained using the complete 10 years of data set has been called the reference map in Figure 4.6. Using this as a reference, the annual maps can be analyzed. Years 1998, 1999, 2000, 2002, 2006 and 2007 are alike with the reference map. In years 1998, 1999 and 2000 the Regions 5 and 6 are more dominant in the interior basin with respect to the Region 4 with lower Chl-a values. However, in years 2002, 2006 and 2007 the Region 4 seems to be more dominant than the Regions 5 and 6 in the interior basin. Nevertheless, all these years have the same provinces with the reference map. In years 2003, 2005 characterizing less productive phase of the Black Sea, the Region 3, which does not take place in the reference map, appears in the interior region with its lower Chl-a values.

Year 2001 is completely different than the other years, as remarked before. While other years contain 6 or 7 regions, this year has only Regions 7, 8 and 9 characterized by concentrations greater than 1 mg/m³ for the whole basin. As discussed earlier, this feature is dominated by the unexpected high spring and summer Chl-a values. However, an interesting feature is the relatively low concentrations within nearly half of the NWS whereas it is typically characterized by the Regions 8 and 9 with their higher Chl-a limits for the entire data set. A possible explanation for this phenomenon could be blooming of rather different algae species having low chlorophyll property (e.g. dinoflagellates) that dominated not only the interior basin but also the NWS in 2001 and caused a major reduction in areal coverages of the Regions 8 and 9.

For detailed weekly maps of regions Appendix C can be checked.

To elucidate temporal evolution of these nine provinces and compare them with each other, the number of pixels belonging to a certain region is plotted versus time (Figure 4.9). It should be remembered that each pixel has been chosen to represent 1 km². Therefore, the total number of pixels of a region points to the total area of that region in km². Firstly, it can be seen that the Region 1 has a negligible areal coverage. The Regions 8 and 9 with the highest Chl-a values also have small areal coverage apart from the Region 1.



Figure 4. 8. Regionalized annual maps of the Black Sea. 1998 – 2007 data are used to obtain the maps.

The most widely distributed province is the Region 7 with a broad Chla range of $1 - 2 \text{ mg/m}^3$ that is confined between relatively poor interior domain and rich coastal zone throughout the year. The Regions 3 and 4 follow the Region 7 in terms of concentrations. The main input for these regions comes from spring – summer productions. Winter season is more densely represented by the Region 5 and autumn by the Region 6, and their contributions exceed those from the Regions 8, 9 and 1, but are less than the Regions 3, 4 and 7. The Region 2 has a different behaviour with respect to other regions. It has less relatively low areal and temporal coverage except sudden isolated peaks mostly observed during the summer – autumn period. Hence, these higher values of Region 2 is absent in years 1999 and 2001.





CHAPTER 5

DISCUSSION AND CONCLUSION

5.1. Discussion and Conclusion

A reconstruction method called Interpolating Empirical 'Data Orthogonal Functions (DINEOF) with local error estimates has been successfully applied for a cloud filling process to a large matrix of the SeaWiFS- derived chlorophyll concentration data of the Black Sea. The main objective is to explore the feasibility of the use of this technique to improve the remote sensing operational system capability of the IMS-METU through the reconstructions of the gappy satellite data. The data used in this study are weekly-binned merged chlorophyll concentration data extracted from the database http://oceancolor.gsfc.nasa.gov/REPROCESSING/SeaWiFS/R5.2/ by cropping the Black Sea region for the longitude band of $26^{\circ} - 42^{\circ}$ E and the latitude band of 40° – 47° N. This gives an area with the size of 1290 km by 779 km. The full data set comprises 474 weekly-binned data covering eleven year period from the launch of the SeaWiFS sensor (September 1997) until the end of 2007.

In the DINEOF method, time series of images provide a mean to calculate principal components of incomplete data as eigenvectors of a covariance matrix and simultaneously fills in the missing data. In other words, the DINEOF method uses, through an iterative procedure, a truncated empirical orthogonal function (EOF) basis to calculate the value of the missing data. The procedure starts by removing the temporal and spatial mean from the original data and initializing the missing values to zero (i.e. to the mean). Then, at each iteration, the EOF basis is used to infer the missing data and a new EOF basis is calculated using the improved data set. Furthermore, based on the outputs of DINEOF analysis, the method allows to produce error maps associated to every reconstruction, providing thus

important complementary information for the practical interpretation or exploitation of the analyzed maps. Benefiting from the existing postprocessing calculation scheme for the error maps, a methodology classifies original pixels on a scale expressing the "outlying" character of each local data. This method makes use of (1) overall statistics synthesized by DINEOF, (2) the correlation length objectively estimated from the data, and (3) the value of the pixel relative to the distribution of surrounding pixels. Maps of outliers allow to visualize how unusual or suspicious are some pixels and patches in regards to the global dynamic of the dataset. The temporal deviation of the mean outlier factor from its global mean can serve as a good indicator pointing towards suspicious images or periods of unusual events, a rather efficient help when analyzing results over thousands of images.

Its implementation to the Black Sea chlorophyll data turns out to be robust, simple to use and does not need any a priori information about the error statistics of the data. The combination a Lanczos solver with DINEOF allows to calculate the EOFs in an optimized way. Furthermore, the time and computational resources needed for this case study are not very high and the computations can be carried out in a desktop PC.

When dealing with chlorophyll fields, the patchiness that usually characterizes their distribution makes the reconstruction of these data a difficult task. Chlorophyll fields are very decorrelated, however the characteristics of DINEOF make it suitable for reconstruction of such data. Some artificial noisy features present in the initial data are also filtered out of the final result using the reconstruction method. This is highly interesting when dealing with data sets that have not been properly treated and present abrupt changes in the vicinity of clouds. As an important feature of the method, the spatial and temporal correlation scales are computed internally in DINEOF, a useful feature to apply to a complex coastal ocean studies, where a priori information about temporal and spatial decorrelation scales is not available.

The results obtained have been analyzed giving an example of their reliability and usefulness. The results of the reconstructions are reasonably accurate, as seen in the visual examples given and in the validation studies made with a small sample data set. The validation was carried out in a data

set with increasing amounts of missing data. The comparison of the reconstructed fields with the original one reveals that the error is small. When visually checking the reconstruction results, the main biogeochemical features are recovered in the final results. While no *in situ* Chl-a data were available, favourable comparisons using twin experiments with artificially created data gaps show the reconstruction is successful and the features that are removed artificially are recovered quite efficiently after the reconstruction. The test runs show that the size of the missing area is not critical for the quality of the DINEOF based reconstructions. Even if 5 weeks out of 27 weeks (~20%) have the same cloud at the same area, the results are acceptable in terms of percentage error, and the features reconstructed are not artifacts of the DINEOF methodology. Removing the Sea of Azov and the Marmara Sea from the analysis which are highly eutrophic and attain relatively high chlorophyll concentrations improves reconstruction analysis.

The present study synthesizes 10 years long chlorophyll data in terms of their time-averaged fields at seasonal, annual, and interannual time scales. The long-term mean chlorophyll distribution reveals an almost horizontally uniform interior basin with concentrations of about 1.0 mg/m³ that increases to 1.5 mg/m³ range along a narrow coastal band and exceeds 3 mg/m³ along the western coastal zone. The interior basin is formed by one composite cell contrary to the two separated gyres for the case of circulation dynamics. Concentration is particularly high during autumn and winter with respect to other seasons. The year 2001 appears as an anomalous year with very high concentrations with respect to all the other years. In general, concentrations are relatively high during 1998 – 2002 period, low during the 2003 – 2005, and intermediate during 2006 – 2007.

Filling the spaces with cloud coverage allows a better recognition of mesoscale dynamics and elaborates the understanding of the Black Sea biological characteristics in terms of their variability both in space and time. Besides, the products can form a basis for other studies such as boundary, initial conditions and/or data assimilation in the ecological models. Further work may be used for reconstructing the sea surface temperature. The analysis can be extended to other Turkish Seas for the analysis of chlorophyll concentration and/or sea surface temperature data. The ready-to-use form of

the software with some documentation should allow any user to implement this tool for any variable and for any region.

REFERENCES

- Alvera-Azcarate, A., Barth, A., Rixen, M. and Beckers, J.-M., 2005: Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature, Ocean Modeling, Volume 9, Issue 4, 325-346.
- Alvera-Azcarate, A., Barth, A., Beckers, J.-M. and Weisberg, R.H., 2007: Multivariate reconstruction of missing data in sea surface temperature, chlorophyll and wind satellite fields, Journal of Geophysical Research, 112, C03008.
- Alvera-Azcarate, A., Barth, A., Sirjacobs, D. and Beckers, J.-M., 2009: Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF, Ocean Science, 5, 475-485.
- Baldacci, A., Corsini, G., Grasso, R., Manzella, G., Allen, J.T., Cipollini, P., Guymer, T.H. and Snaith, H.M., 2001: A study of the Alboran sea mesoscale system by means of empirical orthogonal function decomposition of satellite data, Journal of Marine Systems, Volume 29, Issues 1-4, May 2001, 293-311.
- Beckers, J.-M. and Rixen, M., 2003: Eof calculations and data filling from incomplete oceanographic data sets, Journal of Atmospheric and Oceanic Technology, 20(12), 1839-1856.
- Beckers, J.-M., Barth, A. and Alvera-Azcarate, A., 2006: DINEOF reconstruction of clouded images including error maps: Application to the Sea Surface Temperature around Corsican Island, Ocean Science Discussions, 3(4), 735-776.

- Bennani, M. and Braconnier, T., 1994: Comparative behaviour of eigensolvers on highly nonnormal matrices, Technical Report TR/PA/94/23, CERFACS, Toulouse, France, 20pp.
- Bisagni, J.J., Seemann, K.W. and Mavor, T.P., 2001: High-resolution satellite-derived sea-surface temperature variability over the Gulf of Maine and Georges Bank region: 1993-1996, Deep Sea Research Part II: Topical Studies in Oceanography, Volume 48, Issues 1-3, Coupled biological and physical studies of plankton populations on Georges Bank and related North Atlantic regions, 71-94.
- Brankart, J.-M. and Brasseur, P., 1996: Optimal analysis of *in situ* data in the Western Mediterranean using statistics and cross-validation, Journal of Atmospheric and Oceanic Technology, 16(2), 477-491.
- Brickley, P.J. and Thomas, A.C., 2004: Satellite-measured seasonal and inter-annual chlorophyll variability in the Northeast Pacific and Coastal Gulf of Alaska, Deep Sea Research Part II: Topical Studies in Oceanography, Volume 51, Issues 1-3, Views of Ocean Processes from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Mission: Volume 1, January-February 2004, 229-245.
- Campbell, J.B., 1996: Introduction to Remote Sensing, Taylor & Francis, London, 622 pp.
- Chatelin, F., 1993: Eigenvalues of Matrices, Wiley, John Wiley & Sons, 1st Edition, 400 pp.
- Çokacar, T. and Özsoy, E., 1998: Comparative Analyses and Modelling for Regional Ecosystems of the Black Sea, in: L. Ivanov and T. Oğuz (editors), Ecosystem Modeling as a Management Tool for the Black Sea, NATO Science Series 2: Environmental Security 47, Kluwer Academic Publishers, Dordrecht, vol. 1, 367 pp + vol. 2, 385 pp.

- Fettweis, M. and Nechad, B., 2011: Evaluation of *in situ* and remote sensing sampling methods of SPM concentration, Belgian continental shelf (southern North Sea), Ocean Dynamics, 61(2-3), 157-171.
- Ginzburg, A.I., Kostianoy, A.G., Soloviev, D.M. and Stanichny, S.V., 2000: Chapter 15 Remotely sensed coastal/deep-basin water exchange processes in the Black Sea surface layer, in: David Halpern (editor), Elsevier Oceanography Series, Elsevier, Volume 63, Satellites, oceanography and society, 273-287.
- Gunes, H., Sirisup, S. and Karniadakis, G.E., 2006: Gappy data: To Krig or not to Krig?, Journal of Computational Physics, Volume 212, Issue 1, 2006, 358-382.
- Kopelevich, O.V., Sheberstov, S.V., Yunev, O., Basturk, O., Finenko, Z.Z., Nikonov, S. and Vedernikov, V.I., 2002: Surface chlorophyll in the Black Sea over 1978-1986 derived from satellite and *in situ* data, Journal of Marine Systems, Volume 36, Issues 3-4, 145-160.
- Kopelevich, O.V., Burenkov, V.I., Ershova, S.V., Sheberstov, S.V. and Evdoshenko, M.A., 2004: Application of SeaWiFS data for studying variability of bio-optical characteristics in the Barents, Black and Caspian Seas, Deep Sea Research Part II: Topical Studies in Oceanography, Volume 51, Issues 10-11, Views of Ocean Processes from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Mission: Volume 2, May-June 2004, 1063-1091.
- Korotaev, G.K., Saenko, O.A. and Koblinsky, C.J., 2001: Satellite altimetry observations of the Black Sea level, Journal of Geophysical Research, 106, N C1, 917-933.
- Korotaev, G.K., Oguz T., Nikiforov, A.A. and Koblinsky, C.J., 2003: Seasonal, interannual and mesoscale variability of the Black Sea upper layer

circulation derived from altimeter data, Journal of Geophysical Research, 108 N C4, 3122.

- Lehoucq, R.B., Sorensen, D.C. and Yang, C., 1997: ARPACK User's Guide: Solution of large eigenvalue problems with implicitly restarted Arnoldi methods, Philadelphia, PA: Society for Industrial and Applied Mathematics, (SIAM), 160pp.
- Longhurst, A., 1998: Ecological geography of the sea, Academic Press, San Diego, 398pp.
- Longhurst, A., Sathyendranath, S., Platt, T. and Caverhill, C., 1995: An estimate of global primary production in the ocean from satellite radiometer data, Journal of Plankton Research, 17(6), 1245-1271.
- Lorenz, E.N., 1956: Empirical orthogonal functions and statistical weather prediction, Technical Report, Statistical Forecast Project Report 1, Department of Meteorology, MIT, 49pp.
- Miles, T.N., He, R. and Li, M., 2009: Characterizing the South Atlantic Bight seasonal variability and cold-water event in 2003 using a daily cloud-free SST and chlorophyll analysis, Geophysical Research Letters, 36, L02604.
- Nezlin, N.P., 2000: Chapter 14: Remote-sensing studies of seasonal variations of surface chlorophyll-a concentration in the Black Sea, in: David Halpern (editor), Elsevier Oceanography Series, Elsevier, Volume 63, Satellites, oceanography and society, 257-271.
- Nezlin, N.P. and Dyakonov V.Y., 1998: Seasonal and interannual variations of surface chlorophyll concentration in the Black Sea on CZCS data, in: L.I. Ivanov and T. Oğuz (editors), Ecosystem Modeling as a

Management Tool for the Black Sea, NATO ASI Series, Vol. 1, Kluwer Academic Publishers, Dordrecht, 137-150.

- O'Reilly, J.E., Maritorena, S., O'Brien, M.C., Siegel, D.A., Toole, D., Menzies, D., Smith, R.C., Mueller, J.L., Mitchell, B.G., Kahru, M., Chavez, F. P., Strutton, P., Cota, C.F., Hooker, S.B., McClain, C.R., Carder, K.L., Müller-Karger, F.E., Harding, L., Magnuson, A., Phinney, D., Moore, G.F., Aiken, J., Arrigo, K.R., Letelier, R. and Culver, M., 2000: Ocean Chlorophyll a Algorithms for SeaWiFS, OC2, and OC4: Version 4, in: S. B. Firestone and E.R. Firestone (editors), SeaWiFS postlaunch calibration and validation analyses, Part 3. SeaWiFS postlaunch technical report series, Vol. 11., NASA Technical Memorandum 2000-206892, 49 pp.
- Obukhov, A. M., 1947: Statistically homogeneous fields on a sphere, Uspekhi Matematicheskikh Nauk, Vol. 2, 196-198.
- Oğuz, T. and Beşiktepe, Ş., 1999: Observations on the Rim Current structure, CIW formation and transport in the Western Black Sea, Deep Sea Research, I, 46, 1733-1753.
- Oğuz, T., Violette, P.L. and Ünlüata, Ü., 1992: Upper layer circulation of the southern Black Sea: Its variability as inferred from hydrographic and satellite observations, Journal of Geophysical Research, 97, C8, 12569-12584.
- Oğuz, T., Latun, V., Latif, M.A., Vladimirov, V., Sur, H.İ., Markov, A., Özsoy, E., Kotovshchikov, B., Eremeev, V. and Ünlüata, Ü., 1993: Circulation in the surface and intermediate layers of the Black Sea, Deep Sea Research I, 40, 1597-1612.
- Oğuz, T., Aubrey, D.G., Latun, V.S., Demirov, E., Koveshnikov, L., Sur, H.İ., Diacanu, V., Beşiktepe, Ş., Duman, M., Limeburner, R. and Eremeev,

V., 1994: Mesoscale circulation and thermohaline structure of the Black sea observed during HydroBlack'91, Deep Sea Research I, 41, 603-628.

- Oğuz. T, Ivanov, L.I. and Beşiktepe, Ş., 1998: Circulation and hydrographic characteristics of the Black Sea during July 1992, in: L. Ivanov and T. Oğuz (editors), Ecosystem Modeling as a Management Tool for the Black Sea, NATO ASI Series, Environmental Security, Vol. 47, Kluwer Academic Publishers, Vol. 2, 69-92.
- Oğuz, T., Deshpande, A.G. and Malanotte-Rizzoli, P., 2002: On the Role of Mesoscale Processes Controlling Biological Variability in the Black Sea: Inferrences From SeaWiFS-derived Surface Chlorophyll Field, Continental Shelf Research, 22, 1477-1492.
- Özsoy, E. and Ünlüata Ü., 1997: Oceanography of the Black Sea: A Review of Some Recent Results, Earth Science Reviews, 42(4), 231-272.
- Platt, T. and Sathyendranath, S., 1988: Oceanic primary production: estimation by remote sensing at local and regional scales, Science, 241, 1613-1620.
- Platt, T. and Sathyendranath, S., 2008: Ecological indicators for the pelagic zone of the ocean from remote sensing, Remote Sensing of the Environment, 112(8), 3426-3436.
- Sancak, S., Beşiktepe, Ş.T., Yılmaz, A., Lee, M. and Frouin, R., 2005: Evaluation of SeaWiFS chlorophyll-a in the Black and Mediterranean Seas, International Journal of Remote Sensing, 26, 2045–2060.
- Sirjacobs, D., Alvera, A., Barth, A., Lacroix, G., Park, Y., Nechad, B., Ruddick, K. and Beckers, J.-M., 2008: Reconstruction of Missing Satellite Total Suspended Matter Data over the Southern North Sea and English Channel using Empirical Orthogonal Function

Decomposition of Satellite Imagery and Hydrodynamical Modelling, in: Ocean Optics 2008, 5-10 October, 2008, Braga, Italy, Centre Interfacultaire de Recherches en Océanologie - MARE - GHER, conference lecture.

- Sirovich, L. and Everson, R., 1992: Analysis and management of large scientific databases, International Journal of Supercomputing Applications, 6, 50-68.
- Stanev, E. V. and Peneva, E.L., 2002: Regional sea level response to global climatic change: Black Sea examples, Global and Planetary Change, 32, 33-47.
- Sur, H.İ. and Ilyin, Y.P., 1997: Evolution of satellite derived mesoscale thermal patterns in the Black Sea, Progress in Oceanography, Vol. 39, 109-151.
- Sur, H.İ., Özsoy, E. and Ünlüata, Ü., 1994: Boundary current instabilities, upwelling, shelf mixing, and eutrophication processes in the Black Sea, Progress in Oceanography, 33(4), 249-302.
- Sur, H.İ., Özsoy, E., Ilyin, Y.P. and Ünlüata, Ü., 1996: Coastal/deep ocean interactions in the Black Sea and their ecological/environmental impacts, Journal of Marine Systems, 7(1), 293-320.
- Toumazou, V. and Cretaux, J., 2001: Using a Lanczos eigensolver in the computation of empirical orthogonal functions, Monthly Weather Review 125 (5), 1243-1250.
- Tseng, C., Lin, C., Chen, S. and Shyu, C., 2000: Temporal and spatial variations of sea surface temperature in the East China Sea, Continental Shelf Research, Vol. 20, Issues 4-5, 373-387.

- Vargas, J.M., Garcia-Lafuente, J., Delgado, J. and Criado, F., 2003: Seasonal and wind-induced variability of Sea Surface Temperature patterns in the Gulf of Cadiz, Journal of Marine Systems, Vol. 38, Issues 3-4, 205-219.
- von Storch, H. and Zwiers, F.W., 1999: Statistical Analysis in Climate Research, Cambridge University Press, Cambridge, x+484pp.
- Wilks, D., 1995: Statistical Methods in the Atmospheric Sciences, Academic Press, Burlington, MA, 467pp.
- Yentsch, C.S., 1965: Distribution of chlorophyll and phaeophytin in the open ocean, Deep Sea Research and Oceanographic Abstracts, Vol. 12, Issue 5, 653-666.
- Yoder, J.A., O'Reilly, J.E., Barnard, A.H., Moore, T.S. and Ruhsam, C.M., 2001: Variability in coastal zone color scanner (CZCS) chlorophyll imagery of ocean margin waters off the US East Coast, Continental Shelf Research, Vol. 21, Issues 11-12, 1191-1218.

Data Interpolating Empirical Orthogonal Functions, DINEOF, Method

(Alvera-Azcarate et. al., 2005)

DINEOF is a self-consistent method for the reconstruction of missing data in oceanographic data sets and have been presented by Beckers and Rixen (2003).

Consider that X is the initial matrix of dimensions $m \ge n$, $m \ge n$ (with m the spatial dimension and n the temporal dimension), containing the observations. It may also contain some unknown values corresponding to the missing data.

For the reconstruction of these data, a Singular Value Decomposition (SVD) technique is used to compute the EOFs of the matrix, in which a first guess has been introduced for the missing data. The equation:

$$X = USV^{T}$$
(1)

allows calculating the spatial EOFs, U, with dimension m x r, the temporal EOFs, V, with dimension n x r, and their singular values S, with dimension r x r. The value r is the rank of the matrix, with $r \le \min(m,n)$. Only the most significant spatial and temporal EOFs are necessary for the reconstruction method. The k largest singular values and singular vectors can also be calculated by eigenvector decomposition:

$$XX^{T}u_{i} = \rho_{i}^{2}u_{i}$$
 (2)

with u_i the ith column of U and ρ_i the corresponding singular value, i = 1, ..., k. To avoid using equation (2), which implies the m x m matrix XX^T , we can rather use:

$$Av_i = \rho_i^2 v_i \tag{3}$$

$$u_i = \frac{x_{v_i}}{\rho_i} \tag{4}$$

where $A = X^T X$ is a real symmetric $n \times n$ matrix.

DINEOF can be explained as follows:

• The average value of the matrix is subtracted once for the entire procedure and the missing data points are initialized to zero in order to have an unbiased first guess. The missing data, however, are 'flagged' to differentiate them from those existing points on the mean. This demeaned matrix is used throughout the whole procedure.

• Two steps are then repeated for a given k until convergence:

An EOF decomposition of the matrix is realized, with only the first k
EOFs, to obtain a first estimate of the singular values and singular vectors.

– The elements $X_{i,j}$ corresponding to the flagged missing data are now replaced by the value obtained with the EOF series:

$$X_{i,j} = \sum_{p=0}^{k} \rho_p (u_p)_i (v_p^T)_j$$
 (5)

An improved guess has thus been introduced for the missing data, so we recompute the EOFs and obtain a new value for the missing data.

• Once the convergence is reached, the number of computed EOFs is increased, from $k = 1, ..., k_{max}$, so at the end we have k_{max} estimates for the missing data reconstructed with 1, 2, ..., k_{max} EOFs. But which estimate is the best? The answer is obtained by cross-validation.

• We calculate the optimal number of EOFs from the series of k_{max} EOFs. To do so, a random set of data is initially set aside from the valid data to apply a cross-validation technique, as described in, e.g., Wilks (1995, Chapter 6), Brankart and Brasseur (1996), von Storch and Zwiers (1999, Chapter 18), Beckers and Rixen (2003). This data set has a size of min(0.01 x m x n + 40, 0.03 x m x n). The optimal number of EOFs is the one that minimizes the error between the data set aside and the values obtained at these points with the reconstruction method.

 Once the optimal number N of EOFs is known, the whole procedure is repeated, now including the data set aside for cross-validation, but only with the N first EOFs considered as optimal. Final values for the missing data are then computed.

This is a general description of how the method works. For even more detailed description, see Beckers and Rixen (2003).

In the thesis, the reconstruction method for missing data DINEOF is applied to a large matrix. Lanczos method (see, e.g. Chatelin, 1993; Toumazou and Cretaux, 2001) is used for the EOF decomposition phase to make the application of DINEOF effective when working with large matrices. The desired characteristic of the EOF decomposition algorithm is the possibility to compute only the k largest EOFs at a small computational cost, since it must be used several times during the DINEOF iterations.

Toumazou and Cretaux (2001) have compared three different EOF decomposition methods, one based on the SVD algorithm, and two that express equation (1) as an eigenvalue problem: the QR strategy and the Lanczos method. They have shown that a Krylov-type method, called the Lanczos method, is a good choice when using large matrices. The EOF analysis performed by the mentioned methods gives similar results, although the Lanczos method requires half the storage memory than the others, and up to 22 times less computational time for large matrices. Another attractive characteristic of the Lanczos method is that it does not need to compute all the singular values, only the k largest ones. For these reasons, the Lanczos solver provided by Toumazou and Cretaux (2001), which uses the ARPACK freeware (Lehoucq *et. al.*, 1997), has been implemented in DINEOF.

The main characteristic that makes the Lanczos method suitable when dealing with a large matrix is that, instead of working with the n x n matrix A of equation (3), a p x p (p \ll n) tridiagonal matrix is used, obtained by the projection of A onto the sub-space Krylov $K_p(A,q)$ (Chatelin, 1993). The eigenvalues are calculated in this reduced matrix, in an iterative way until a convergence criterion is satisfied. This stopping criterion is based on the backward error (Bennani and Braconnier, 1994). For a more detailed explanation of the Lanczos method, see, e.g., Chatelin (1993) and (Toumazou and Cretaux, 2001).
APPENDIX B

Weekly Chl-a Images

Week numbers according to seasons will be presented in Table B.1. Week numbers according to years will be presented in Table B.2.

Original Chl-a data with gaps will be presented in Figure B.1.

Gap free reconstructed Chl-a data will be presented in Figure B.2.

Gap free reconstructed Chl-a data with removed Azov and Marmara Seas will be presented in Figure B.3.

(Hard copy of the images is in gray scale. The coloured versions can be found on the CD attached to the thesis.)

<u>Autumn</u>	<u>Winter</u>	<u>Spring</u>	<u>Summer</u>
1-10	11-21	22-33	34-44
45-56	57-67	68-79	80-90
91-102	103-113	114-125	126-136
137-148	149-159	160-171	172-182
183-194	195-205	206-217	218-228
229-240	241-251	252-263	264-274
275-286	287-297	298-309	310-320
321-332	333-343	344-355	356-366
367-378	379-389	390-401	402-412
413-424	425-435	436-447	448-458
459-470			

 Table B. 1. Table showing number of weeks according to seasons.

 Table B. 2. Table showing number of weeks according to years.

1997	1-14	
1998	15-60	
1999	61-106	
2000	107-152	
2001	153-198	
2002	199-244	
2003	245-290	
2004	291-336	
2005	337-382	
2006	383-428	
2007	429-474	



















































































































































APPENDIX C

Regionalized Weekly Maps

Weekly binned ocean color data presented in Figure B.3. are used to obtain the regionalized maps presented in Figure C.1.

The general regions of the basin obtained using 10 years data set has been called the reference map and was presented in Figure 4.6. This map can be used as a reference to see the status of the regions at a certain week as compared to the general tendency of the data.

(Hard copy of the images is in gray scale. The coloured versions can be found on the CD attached to the thesis.)

















































CURRICULUM VITAE

Serkan SANCAK

Institute of Marine Sciences, Middle East Technical University P.O. Box 28, 33731 Erdemli, Mersin/Turkey Phone: (+90) 324-521-2406, Fax: (+90) 324-521-2327 E-mail: sancak@ims.metu.edu.tr

Date and Place of Birth: May 19, 1975. Nicosia, Cyprus.

• EDUCATION

• Ph.D. in Physical Oceanography

Institute of Marine Sciences	
Middle East Technical University (METU), Turkey	2002-2011
(quit for two years due to health problems)	

M.Sc. in Physical Oceanography

Institute of Marine Sciences	
Middle East Technical University (METU), Turkey	1998-2001
B.Sc. in Physics	

Ankara University, Turkey 1992-1998

• RESEARCH TOPICS

- *M.Sc. Thesis:* Optical properties of the Turkish Sea and validation of the SeaWiFS.
- Ocean optics
- Remote sensing
- Mediterranean and Black Sea observations
- Long-term spatiotemporal data analysis

• LANGUAGES

Turkish (native), English (fluent), French (primitive).

• **PUBLICATIONS**

(Not including oral and poster presentations)

- B.Sc. Thesis: Denizlerin optik özellikleri ve bu özelliklerin tespitinde bir yöntem belirlenmesi, Ankara University, Ankara, June 1998, 38 pages, (in Turkish).
- M.Sc Thesis: Optical properties of the Turkish Sea and validation of the SeaWiFS, METU, Institute of Marine Sciences, Erdemli/İçel, September, 2001, 114 pages.

- Emin ÖZSOY, Mohammed A. LATİF, Şükrü T. BEŞİKTEPE, Vladimir BELOKOPYTOV, Nilgün ÇETİN, Serkan SANCAK and Adil SÖZER; 2000: İstanbul Boğazı'nda üç boyutlu akıntılar, su seviyesi değişimleri ve çevre denizler ile ilişkileri. 1. Ulusal Deniz Bilimleri Konferansı Bildiri Özetleri Kitapçığı; editör: Zahit UYSAL; 2000, s.6, (in Turkish).
- Serkan SANČAK, Şükrü T. BEŞİKTEPE ve Ayşen YILMAZ; 2000: Türkiye'yi çevreleyen denizlerin spektral optiksel özelliklerinin yerinde ölçüm ve uzaktan algılama ile belirlenmesi. 1. Ulusal Deniz Bilimleri Konferansı Bildiriler ve Poster Özetleri; editörler: Zahit UYSAL ve İlkay SALİHOĞLU; 2000, s. 26-29, (in Turkish).
- Serkan SANCAK, Şükrü T. BEŞİKTEPE, Ayşen YILMAZ, Michael LEE and Robert FROUIN, 2005: Evaluation of SeaWiFS chlorophyll-a in the Black and Mediterranean Seas. International Journal of Remote Sensing Volume 26, Number 10, pp. 2045-2060.

• INTERNATIONAL COURSES, TRAININGS, SERTIFICATES and SCHOLARSHIPS

- 1999 Using GIS as a Tool to Communicate with Decision Makers Turkish Ministry of Environment and Forestry.
- 2000 Training Course on Remote Sensing of Ocean Color IOCCG (International Ocean-Color Coordinating Group) – NATO and METU.
- 2002 NATO-ASI (Advanced Study Institute) Ocean Carbon Cycle and Climate Change Summer School – IOC, NATO and METU.
- 2002 Summer School on Combining Ocean Data and Dynamics NATO and METU.
- 2003 Special Training on Satellite Oceanography Laboratory MHI (Marine Hydrophysical Institute) Ukraine and METU.
- 2003 Spatio-Temporal Statistical Analysis of Multi-Platform Optical Ocean Observations Summer School – NSF (National Science Foundation, USA) and University of Maine.
- 2003 Onboard Training Scholarship for Hands on Bio-optical measurements, BEAGLE (Blue EArth GLobal Expedition) 2003 Cruise, - IOCCG, POGO (Partnership for Observation of the Global Oceans), IOC (Intergovernmental Oceanographic Commission) and JAMSTEC (Japan Agency for Marine-earth Science and TEChnology).
- 2006 Climate change impacts on marine Ecosystems Eur-Oceans, NATO, Tübitak and METU.
- 2008 E2E EcoModel, Analysis of End-To-End Food Webs and Biogeochemical Cycles – Supported by Eur-Oceans, IMBER and METU.

• **PROJECTS INVOLVED**

(Only last 3 named)

 Numerous projects carried out by METU-IMS between 1998-2010. (Mainly: NATO, European Union, NSF (USA) and TÜBİTAK).

- SeaDataNet Integrated Infrastructure Initiative of the EU Sixth Framework Programme coordinated by IFREMER (Institut Français de Recherche pour l'Exploitation de la Mer).
- Black Sea SCENE European Union FP6 Project.
- ECOOP European Coastal Sea Operational Observing and Forecasting System, European Commission's Sixth Framework Programme.

• ONBOARD and LABORATORY EXPERIENCE

- Participated to METU-IMS cruises between 1998-2010, on Black, Aegean, Marmara and Mediterranean Seas. (Over 500 days).
- Participated to daily cruises on Damariscotta River and Atlantic Ocean Cost of Gulf Of Maine.
- Participated to BEAGLE 2003 cruise on Pacific Ocean.
- Able to operate all kinds of ADCP, CTD and various bio-optical measurement devices onboard.
- Chemical and biological laboratory analysis for bio-optical measurements
- Also trained for wet-lab, water sampling and rosette operation.
- Remote sensing laboratory experience with AVHRR, SeaWiFS and MODIS satellites and Quorum antenna system.