Thèse de Doctorat

Sciences de la terre de l'univers et de l'espace
Terre et enveloppes fluides

présentée à ***l'Ecole Doctorale en Sciences Technologie et Santé (ED 585)***

de l'Université du Littoral Côte d'Opale

par

Ana Gabriela BONELLI

pour obtenir le grade de Docteur de l'Université du Littoral Côte d'Opale

Estimation et analyse spatio-temporelle du carbone organique dissous pour l'océan mondial à partir de la télédétection de la couleur de l'eau

Soutenue le 16/06/2021, après avis des rapporteurs, devant le jury d'examen :

Mme. Ana Laura Delgado, Chargée de recherche, IADO	**Président du jury**
M. Emmanuel Boss, Professeure, UMaine	**Rapporteur**
M. Emmanuel Devred, Directeur de recherche, BIO	**Rapporteur**
Mme. Séverine Alvain, Chargée de recherche HDR, CNRS	**Examinatrice**
M. Olivier Aumont, Chargée de recherche HDR, LOCEAN	**Examinateur**
Hubert Loisel, Professeure, ULCO	**Directeur de thèse**
Vincent Vantrepotte, Chargée de recherche, LOG	**Co-encadrant**
Antoine Mangin, Directeur de recherche, ACRI-ST	**Co-encadrant**

PhD thesis

Earth, universe and space sciences
Earth and superficial envelopes

presented at the ***Doctoral School of Science, Technology and Health (ED 585)***

of the Université du Littoral Côte d'Opale

by

Ana Gabriela BONELLI

To obtain the degree of Doctor of the Université du Littoral Côte d'Opale

Estimation and spatio-temporal analysis of dissolved organic carbon for the global ocean from remote sensing of water color

Defended on the 06/16/2021 after approval of the reviewers, in front of the jury:

Mrs. Ana Laura Delgado, Researcher CONICET, IADO	**President of the jury**
Mr. Emmanuel Boss, Professor, UMaine	**Reviewer**
Mr. Emmanuel Devred, Research Director, BIO	**Reviewer**
Mrs. Séverine Alvain, Researcher HDR, CNRS	**Examiner**
Mr. Olivier Aumont, Researcher HDR, LOCEAN	**Examiner**
Mr. Hubert Loisel, Professor ,ULCO	**Thesis supervisor**
Mr. Vincent Vantrepotte, Researcher, LOG	**Co-supervisor**
Mr. Antoine Mangin, , Research Director, ACRI-ST	**Co-supervisor**

Acknowledgements

After years of hard work I finally get to write the last few words of this report. While this chapter of my life is reaching the end, I want to express my sincere gratitude to all who have directly or indirectly made it possible.

First, I want to show my appreciation to my supervisors, Hubert Loisel, Vincent Vantrepotte and Antoine Mangin, for giving me the opportunity to do this PhD. Their guidance and support, including the attention to all of the little details, helped me to progress on my work always looking for better results.

I would also like to thank Olivier Aumont for his insights about my work which helped me to better understand and interpret my results. I am also grateful to the rest of members of the jury who have accepted to evaluate my thesis work: Emmanuel Boss, Emmanuel Devred, Séverine Alvain and Ana Laura Delgado.

I want to express my acknowledgement to the French Spatial Agency (CNES) and ACRI-ST for providing the financial support needed for this PhD research work, and to the Laboratory of Oceanology and Geoscience (LOG) and its director, François Schmidt and now Hubert Loisel, for hosting me. Further, I would like to show my sincere gratitude to the administrative staff of CNES, ACRI-ST and LOG for their helpful support with the paperwork required to do a PhD in France.

My most sincere thanks to David Dessailly for his collaboration processing data and to Cédirc Jamet for shearing his knowledge about neural networks. Further thanks to Julien Demaria for his patience and help on my first steps into Python coding and his hard work over the match-up datasets. I also want to thank him and Christophe 'Gaucho' Lerebourg for shearing some *mate(tea)* with me and swimming pool training days.

I want to especially thank Alberto Piola for not losing his hopes on me, encouraging me to pursue a research career. I owe him a lot, starting with a scientific publication that I promise it will be done.

Doing a PhD so far from home is not easy. I am very grateful to Odette and Michel for welcoming me at their house at the very beginning, making me feel like home. Likewise, I am thankful for the good friends I made in LOG: Dani, Sara, Kien, Manh, Cuong, Yang, Roy and Dat. They made my days much more interesting and fun. Side by side we learned the meaning of 'No pain, no gain' and 'Match-up is a b#@$%'.

I also counted with the support of my pre-thesis friends: Anita, Noe and *Las Batichicas*. Strong, independent and beautiful women who, from their respective fields, make the difference looking for a better world with gender equality.

I could have never achieved this PhD without the support that my parents, Millie and Pablo, constantly show me. Long time ago they gave up the idea of me having a standard life—if they ever had it—and encouraged me to always go for more. I hope to make you proud.

Also thanks to my elder brothers, Marcos and Lucas. They have bugged me since day one and that made me strong enough to face this kind of challenges. And thanks to all my extended family for their messages, video calls, impromptu meetings at airports and at half way vacations. Basically, thank you all for being there for me.

At last but definitely not least, I want to thank my best friend, adventures partner, forever *compinche* and husband, Lucas Blanco. He always believes in me, even when I don't. I am, have been and will always be your number one fan.

Table of content

Acknowledgements i
TABLE OF CONTENTS iii
List of symbols.....vii
List of Figures ix
List of Tables.....xvii
Introduction 1
1. Chapter 1: State of the Art.....9
1.1 Dissolved Organic Matter (DOM) 9
1.2 Dissolved Organic Carbon (DOC).....10
1.2.1 Labile DOC.....11
1.2.2 Semi-labile DOC.....11
1.2.3 Refractory DOC.....11
1.3 DOC sources, sinks and processes.....13
1.3.1 DOC source.....13
1.3.2 DOC sinks, export and sequestration.....16
1.4 Colored Dissolved Organic Matter (CDOM) 19
1.5 CDOM to DOC relationships.....21
1.6 Current assessment of CDOM and DOC distribution at global scale.....23
1.6.1 CDOM in situ.....23
1.6.2 CDOM from space.....24
1.7 DOC global distribution.....25
1.7.1 The different approaches providing the DOC spatial distribution at global scale.....25
1.7.1.1 In situ.....25
1.7.1.2 DOC from remote sensing.....26
1.7.1.3 DOC from coupled bio-physical models: the example of the PISCES model.....28
2. Chapter 2: CDOM estimation at global scale, spatio-temporal variability and contribution to the total absorption budget.....31
2.1 Colored dissolved organic matter absorption at global scale from ocean color radiometry observation: spatio-temporal variability and contribution to the absorption budget.....31

2.1.1 Introduction.....31
2.1.2 Materials and methods.....34
2.1.2.1 Datasets description.....34
2.1.2.1.1 Optical typology.....34
2.1.2.1.2 In situ and matchup data sets used for validation.....34
2.1.2.1.3 Satellite Data used for global CDOM spatio-temporal variability.....37
2.1.2.2 Statistical indicators 38
2.1.2.3 Models description.....38
2.1.2.3.1 Arin et al. (2018) 38
2.1.2.3.2 Chen et al. (2017).....39
2.1.2.3.3 Shanmugam (2011).....39
2.1.2.3.4 Loisel et al. (2014).....40
2.1.3 Results and discussion.....41
2.1.3.1 Adaptation of the Loisel et al. (2014) algorithm for estimating $a_{cdom}(443)$ over the global ocean 41
2.1.3.1.1 CDOM-KD2 parameterization 41
2.1.3.1.2 Model development in the context of satellite application.....42
2.1.3.2 Intercomparison and validation of $a_{cdom}(443)$ inversion models.....44
2.1.3.2.1 Performance and inter-comparison of the different $a_{cdom}(443)$ inversion models over the in situ and matchup data sets.....44
2.1.3.2.2 Comparison of the $a_{cdom}(443)$ models on moderate to non-turbid waters.....47
2.1.3.2.3 Global $a_{cdom}(443)$ spatio-temporal patterns.....49
2.1.3.2.4 Global scale covariation between a_{cdom}, a_{cdm} and Chl-a dynamics.....51
2.1.3.2.5 Global $a_{cdom}(443)/a_{cdm}(443)$ ratio spatio-temporal patterns.....54
2.1.3.2.6 Global $a_{cdom}(443)/a_{nw}(443)$ ratio spatio-temporal patterns.....56
2.1.4 Conclusions 58
2.2 Temporal variability of $a_{cdom}(443)$ over the global ocean.....60
2.2.1 Census X-11 and trend analysis.....60
2.2.2 $a_{cdom}(443)$ temporal schemes of variability.....61
2.2.3 Interannual changes in $a_{cdom}(443)$ values.....62
2.3 Conclusion and perspectives …….....65
3. Chapter 3: Estimation of Dissolved Organic Carbon in global scale from satellite data.....…… 67

3.1 In situ and satellite data sets used for the development and validations……67
3.1.1 The whole DOC and ancillary variables data…….....67
3.1.2 The different sub data sets used for development and validation……71
3.1.3 Roshan and DeVries (2017) global annual mean……..……...75
3.2 Existing satellite models for open ocean waters……..……75
3.2.1 Aurin et al. (2018)…….....75
3.2.2 Siegel et al. (2002)……...76
3.3 Development of the algorithm……..……..76
3.3.1 Input variable selection……...76
3.3.2 Combination and time lag selection……78
3.3.3 Structure selection……...83
3.3.4 Water type model dependency……85
3.4 Evaluation of the performance of NN29s on climatological data and comparison with other models performance on climatological data……..….88
3.4.1 NN29s comparison with Siegel et al. (2002) and Aurin et al. (2018)..…….88
3.4.2 DOC global distribution and temporal variability……...91
3.4.3 Global Distribution: comparison with Roshan and DeVries (2017)..…….94
3.5 Comparison with PISCES…..…….98
3.6 Conclusions and perspectives...……..……....100
4. General conclusions and perspectives……..……103
References …...…….……..…….…107
Résumé Étendu.…..……...…….……..…….…139
Résumé …….……...…….……..……153
Abstract.....155

List of symbols

Parameter	Description	Unit
a_{cdom}	Colored dissolved organic matter absorption	m^{-1}
a_{cdm}	Colored dissolved and detrital matter absorption	m^{-1}
a_{nap}	Non-algal particles absorption	m^{-1}
a_{ph}	Phytoplankton absorption	m^{-1}
a_{nw}	Non-water absorption	m^{-1}
R_{rs}	Remote Sensing Reflectance	sr^{-1}
λ	Wavelength	nm
Chl-a	Chlorophyll-*a*	mg/m^3
DOC	Dissolved organic carbon	μmol/L
POC	Particulate Organic Carbon	mg/L
TOC	Total Organic Carbon	mg/L
SSS	Sea surface salinity	psu
SST	Sea Surface Temperature	ºC
PAR	Photosynthetically Active Radiation	$Em^{-2}d^{-1}$
MLD	Mixed Layer Depth	m
DOM	Dissolved organic mater	
CDOM	Colored dissolved organic matter	
CDM	Colored dissolved and detrital matter	
DIC	Dissolved Organic carbon	
TEP	Transparent exopolymer particles	
NAP	Non-algal particles	
ANN	Artificial Neural Network	
OCR	Ocean Color Radiometry	

List of Figures

Fig 1 Variations in the atmospheric CO_2 concentration over time based on ice core reconstructions and direct observations since 1958. (a) CO2 variations over the last 400 kyr (1 kyr = 1000 years). (b) CO2 variations from 25,000BC to present. (c) CO2 variations during the last 1000 years as reconstructed from Antarctic ice cores. (d) CO2 variations during the last 50 years as directly measured in the atmosphere at Mauna Loa, Hawaii. The Mauna Loa data are from Keeling and Whorf (1998), the Law Dome, Antarctica, data from Etheridge et al. (1996), the Taylor Dome, Antarctica, data from Indermühle et al. (1999, 2000), the Dome C data from Monnin (2001), and the Vostok, Antarctica, data from Petit et al., (1999). Figure from Sarmiento and Gruber (2006)……..……2

Fig 2 Representation of Dissolved Organic Mater (DOM) composition including the Dissolved Organic Carbon (DOC) and its fraction in the Total organic Carbon (TOC) bulk, the Dissolved Organic Nitrate (DON) and Phosphate (DOP) and the Colored Dissolved Organic Mater (CDOM).……..……6

Fig 1.1 Size range of particulate (POM) and dissolved organic matter (DOM) and organic compounds in natural waters. AA, amino acids; CHO, carbohydrates; CPOM, coarse particulate organic matter; FA, fatty acids, FPOM, fine particulate organic matter; HA, hydrophilic acids; HC, hydrocarbon; VPOM, very fine particulate organic matter. From Nebbioso and Piccolo (2013)..……..……9

Fig 1.2 Vertical profile of DOC and its different fractions, refractory, semi-refractory, semi-labile adapted from Hansell (2013) (a). Distribution of ANN-derived (Color map) and observational DOC (colored dots) at 600m from Roshan and DeVries (2017)..……..……..……12

Fig 1.3 Carbon pump schema representing the fixation, of CO_2 by phytoplankton and posterior degradation mineralization and sequestration of DOC into the deep ocean. Diagram from Buchan et al., (2014).…..……..……..……13

Fig 1.4 Simplified diagram of the global carbon cycle. Numbers denote reservoir mass in Pg C (1 Pg C = 10 15 g C) and annual carbon fluxes in Pg C yr $^{-1}$ between the atmosphere and the land and ocean. Black numbers and arrows indicate reservoir mass and exchange fluxes estimated for the time prior to the Industrial Era, about

1750. Red arrows and numbers indicate annual " anthropogenic " fluxes averaged over the 2000–2009 time period. Red numbers in the reservoirs represent cumulative changes of anthropogenic carbon over the Industrial Period 1750–2011. Diagram from Kandasamy and Nath (2016)……...……15

Fig 1.5 Schema of sources and sinks of CDOM to the ocean from Coble (2007)…….……21

Fig 1.6 (a) Relationship of $a_{cdom}(355)$ in the Chesapeake Bay mouth and plume region for the 2004–2006 research cruises and Delaware Bay mouth from Mannino et al. (2008). (b) Relationship of DOC and $a_{cdom}(325)$ from samples taken on BATS cruises in the Sargasso Sea from spring of 1994 through the end of 2000 (Nelson and Siegel, 2002)…...…….……22

Fig 1.7 Global distribution of GOCAD and NOMAD field stations for CDOM concentration from Aurin et al. (2018)……. …….……23

Fig 1.8 Scheme of absorption spectra of phytoplankton (a_{phy}), CDOM (a_{cdom}) and NAP (a_{nap})…...……...…….…… 24

Fig 1.9 Distribution of ANN-derived and observational DOC. Color map is the artificial neural network (ANN)-derived dissolved organic carbon (DOC) concentration, and colored dots are the observed DOC concentration at 20m. Figure from Roshan and DeVries(2017).…….……..……26

Fig 1.10 Climatological DOC distribution from a regression analysis based upon wintertime SST values. Figure from Siegel et al. (2002)…..…….……27

Fig 1.11 Retrieved three-year mean, 9 km nominal resolution DOC from Aquarius and MODIS Aqua using the MLR2 inversion from Aurin et al. (2018). Figure from Aurin et al. (2018)……..…… …….……..……...……28

Fig 1.12 DOC annual average concentration simulated with PSCES biogeochemical model…….……..……..……...……..……29

Fig 2.1 From top to bottom, $a_{cdom}(443)$ absolute frequency distribution histograms for the DS1 (a, b, c), DS2 (d, e, f) and DS3 (g, h, i) datasets. The histograms corresponding to the complete datasets are in grey, those for the water classes 1 and 2 subsets are in red, while the subsets gathering the water classes 3 to 17 are in blue. N, $\overline{X}$, m, std,

q1, q3 correspond to the number of datapoints, mean, median, standard deviation and first and third quantiles values, respectively..……..…….……..……35

Fig 2.2 Distribution of DS2 (circles) and DS3 (crosses) data points used for comparison of the four models. Red dots represent the highly turbid data points strongly influenced by terrestrial inputs (classes 1 and 2 from the Mélin and Vantrepotte (2015) optical typology) while the blue dots are associated with marine waters belonging to other optical classes…..…….……36

Fig 2.3 Satellite R_{rs} (R_{rs} SAT) vs in situ R_{rs} (R_{rs} IS) from DS3 at (a) 412, (b) 443, (c) 490, (d) 510, (e) 560 and (f) 670 nm. Black solid line represents the 1:1 line, colored solid lines are the threshold limit, where $|R_{rs}$ SAT- R_{rs} IS$| > R_{rs}$ IS * 0.75. Black circles represent the data points for which any of the bands overcomes the settled threshold. This data points where labeled as outliers and not further considered in the analysis…….……...…….……37

Fig 2.4 Different steps of the CDOM-KD2 inversion model parameterized from the IOCCG dataset: (a) Δ_p (443 - 560) as a function of ΔK_d (443 - 560) (Eq 2.23) (b) X as a function of ΔK_d (443 - 560) - Δ_p (443 - 560) (Eq 2.21), and (c) a_{cdom}(443) as a function of X (Eq 2.20)….…….……..…42

Fig 2.5 Performance of the CDOM-KD2 inversion model considering a NN based inversion for the calculation of the ΔK_d term (Eq 2.20) and using the DS1 data set….…….…….…….…….…….. 43

Fig 2.6 Validation scatter plots of the four tested models (CDOM-KD2, A2018, C2017 and S2011) using the DS2 (a, c, e, g) and DS3 (b, d, f, h) complete data sets (classes 1 to 17). Radar plots of the statistics used for evaluating the four models for DS2 (i) and DS3 (j)…….……...…….……45

Fig 2.7 a_{cdom}(443) absolute frequency distribution histograms for the DS2 (a, c, e, g) and DS3 (b, d, f, h) complete data sets for the in situ (grey) and modeled values by the four tested models (CDOM-KD2: purple, A2018 : red,, C2017; green and S2011; blue). N, $\overline{X}$, m, std correspond to the number of data points, mean, median and standard deviation respectively…..…….……46

Fig 2.8 Validation scatter plots of the four tested models (CDOM-KD2, A2018, C2017 and S2011) over DS2 (a, c, e, g) and DS3 (b, d, f, h) non-turbid subset (classes 2 to

17). Radar plots of the statistics used for evaluating the four models in DS2 (i) and DS3 (j)…..……..48

Fig 2.9 $a_{cdom}(443)$ absolute frequency distribution histograms over DS2 (a, c, e, g) and DS3 (b, d, f, h) non-turbid subsets (classes 2 to 17) for the in situ (grey) and modeled values by the four tested models (CDOM-KD2: purple, A2018: red, C2017: green and S2011: blue). N, $\overline{X}$, m, std correspond to the number of datapoints, mean, median and standard deviation respectively...…….……..49

Fig 2.10 Global average $a_{cdom}(443)$ [m^{-1}] map produced with GlobColour L3 merged 25km 8 days composite data from 23rd April 2002 to 13th April 2012 and CDOM-KD2 model (a) with its coefficient of variation (%) (c) and the global $a_{cdom}(443)$ [m^{-1}] average maps produced from the 4 years archive of GlobColour OLCI L3 25km 8 days composite data from 22nd April 2016 to 16th June 2020 (b) with its coefficient of variation (%) (d).…….…….51

Fig 2.11 Correlation maps between (a) CDOM-KD2 and a_{cdm}, (b) CDOM-KD2 and *Chl-a*, (c) a_{cdm} and *Chl-a*. White areas represent pixels where either there is no data available or where the correlation is not significant (p>0.05)…….……..52

Fig 2.12 (a) Location of the stations considered for the two time series plotted in panels (b) and (c) (red circles). At these two stations the correlation between a_{cdom} and both *Chl-a* and a_{cdm} is minimum (South Pacific Gyre : SPG) and maximum (North Atlantic, NA). Time series of $a_{cdom}(443)$, $a_{cdm}(443)$ and *Chl-a* at SPG (b) and NA (c)……..…….…….……54

Fig 2.13 Global distribution of the average $a_{cdom}(443)/a_{cdm}(443)$ ratio (a) and variation coefficient (%) (b) for the GlobColour L3 merged 25km, 8 days composite data from 23rd April 2002 to 13th April 2012…….…….55

Fig 2.14. Global distribution of the average $a_{cdom}(443)/a_{nw}(443)$ ratio (a) and variation coefficient (b) for the GlobColour L3 merged 25km, 8-day composite data from 23rd April 2002 to 13th April 2012……..…….……57

Fig 2.15 Relative contribution of the (a) seasonal (S(t)), (b) trend cycle (T(t)) and (c) irregular (I(t)) Census X-11 components to the total variance of $a_{cdom}(443)$ signal over the time period 2002-2012…….……62

Fig 2.16 Global rate of change (RC; %/year) for $a_{cdom}(443)$ monthly time series (CDOM-KD2 algorithm and R_{rs} from GlobColour L3 merged data) between 1997 to 2012 (a) and 2002 to 2012 (b) calculated with CDOM-KD2 algorithm and R_{rs} from GlobColour L3 merged data. Black circles show the locations chosen for a time series extraction……63

Fig 2.17 Time series extraction of $a_{cdom}(443)$ (X(t)), and corresponding X11 I(t), S(t) and T(t) component in two regions of the Pacific ocean showing positive trend in CDOM over the period 2001-2012, located in the northern, (a), the southern (b) Pacific oligotrophic gyres waters (see location in Fig 2.16) ……64

Fig 3.1 In situ DOC measurements from DS1. N, $\overline{X}$, m and std correspond to the number of data points, mean, median and standard deviation, respectively……68

Fig 3.2 Representation of the time lags methodology used for the match-up of in situ DOC with other variables…….70

Fig 3.3 Timeline illustrating past, current, and future global ocean-color satellite missions adapted from Blondeau-Patissier et al. (2014)……71

Fig 3.4 Location of DD-MLR data points (a) and the corresponding histogram of the in situ DOC gathered (b). N, $\overline{X}$, m and std correspond to the number of data points, mean, median and standard deviation, respectively……71

Fig 3.5 Global distribution of DD-NN data points (a) showing the training subset in blue and the validation subset in red, and the histogram of the in situ DOC gathered for each subset (b). N, $\overline{X}$, m and std correspond to the number of data points, mean, median and standard deviation, respectively…… ……72

Fig 3.6 Global distribution of DD-NNCHL data points (a) showing the training subset in blue and the validation subset in red, and the histogram of the in situ DOC gathered for each subset (b). N, $\overline{X}$, m and std correspond to the number of data points, mean, median and standard deviation, respectively. ……72

Fig 3.7 Global distribution of DV1 data points (a) and the corresponding histogram of the in situ DOC (b). N, $\overline{X}$, m and std correspond to the number of data points, mean, median and standard deviation, respectively……73

Fig 3.8 Global distribution of DV2 data points (a) and the corresponding histogram of the in situ DOC (b). N, $\overline{X}$, m and std correspond to the number of data points, mean, median and standard deviation, respectively.……73

Fig 3.9 Direct correlation between the in situ DOC data from DD-MLR and the corresponding SST, SSS, PAR, MLD, $a_{cdom}(443)$, *Chl-a* and R_{rs} (412, 443, 490, 510, 560, 670) at time lag 0 (panels a to i respectively). Note that the *y* axes range changes depending on the variable used for the correlation...……77

Fig 3.10 Example of multi-linear regression between in situ DOC from DD-MLR and estimated DOC from different combinations of SST, SSS, PAR, MLD, $a_{cdom}(443)$, *Chl-a* and R_{rs} (412, 443, 490, 510, 560) at time lag 0……78

Fig 3.11 (a) Scheme of initial ANN structure with one input layer (IL), one hidden layer (HL) and one output layer (OL), the respective activation function (AF1 and AF2) for the hidden layer and the output layer (ReLU and Linear, respectively), and the Adam optimization technique (O). (**b**) Flow chart of NN29b structure with one input layer (IL), two hidden layer (HL1 and HL2) and one output layer (OL), the respective activation functions (AF1, AF2 and AF23) for each hidden layer and for the output layer (softsign, exponential and exponential, respectively), and the RMSprop optimization technique (O)……80

Fig 3.12 Scheme of the decision system followed for the selection of the best input variables and structure used in the ANN to estimate DOC……81

Fig 3.13 AIC decline in the progress of the ANN input data selection. Orange vertical lines indicate the point of increase of number of input variables. ……82

Fig 3.14 Comparison of the model-derived and measured DOC from DD-NN for NN29 (a) and NN29b (b) for the training (blue) and validation (red) data sets, with their respective histograms of DOC measured in situ (grey) or estimated from NN29 (c) and NN29b (d). N, RMSD, MAPD, MB, std and r correspond to the number of data points, the root mean square deviation, median absolute percentage deviation, mean bias, standard deviation and the correlation coefficient, respectively (top panels); and $\overline{X}$ and m correspond mean and median, respectively (bottom panels)…....…84

Fig 3.15 Comparison of the model-derived and measured DOC from DD-NNCHL for NN29b (a) and NN29bCHL (b) and their respective histograms in pink and cyan (c

and d) over the in situ DOC concentration histogram (gray). N, RMSD, MAPD, MB, std and r correspond to the number of data points, the root mean standard deviation, median absolute percentage deviation, mean bias, standard deviation and the correlation coefficient, respectively (top panels); and $\overline{X}$ and m correspond mean and median respectively (bottom panels).……...…..……...……85

Fig 3.16 Comparison of the model-derived and measured DOC for the NN29b (pink) and NN29bCHL (cyan) models from the DD-NNCHL data set and for each water classes 1 to 17. The solid line represents the 1:1 line, and the slope value of the best fit linear regression type-II is provided.. ……..…….……86

Fig 3.17 Comparison of the model-derived and measured DOC from DD-NNCHL for NN29s (a), its respective histogram (purple) over the in situ DOC concentration histogram in gray (b) and the radar plot comparing the performance of NN29b (pink), NN29bCHL (cyan) and NN29s (purple). N, RMSD, MAPD, MB, std and r correspond to the number of data points, the root mean square deviation, median absolute percentage deviation, mean bias, standard deviation and the correlation coefficient, respectively (left) and $\overline{X}$ and m correspond mean and median respectively (right)..……...…….……87

Fig 3.18 Scheme of NN29s model functionality, starting with water classification which is used for the decision of which ANN will be applied (NN29b or NN29CHL) to estimate DOC…….……88

Fig 3.19 Comparison of the model-derived and measured DOC from DV1 for NN29s (a), Siegel et al. (2002) (b) and Aurin et al. (2018) (c) from DV1, and their respective histograms in purple (d), yellow (e) and green (f) over the in situ DOC histogram (gray). N, RMSD, MAPD, MB, std and r correspond to the number of data points, the root mean squared deviation, median absolute percentage deviation, mean bias, standard deviation and the correlation coefficient, respectively (top panels). $\overline{X}$ and m correspond to the mean and median respectively (bottom panels).…….…….…….……89

Fig 3.20 Comparison of the model-derived and measured DOC from DV2 for NN29s from 10 years a weekly time series (2002 - 2012) match-up with in situ DOC monthly climatology from DV2. N, RMSD, MAPD, MB, std and r correspond to the number of data points, the root mean standard deviation, median absolute

percentage deviation, mean bias, standard deviation and the correlation coefficient respectively…..……..……..…… 90

Fig 3.21 Monthly climatology of DOC concentration generated with NN29s from 10 years of weekly time series (2002 - 2012) with in situ DOC monthly climatology from DV2 over plotted…….……..……..……..……93

Fig 3.22 ANN-derived annual average DOC concentration from Roshan and DeVries (2017) (a) and from NN29s (b) and the MAPD ((XX-YY)/(XX+YY) * 200) map comparing the two models (c)…….……95

Fig 3.23 a) DOC concentration map estimated with NN29s for the period between 19th of December 2007 to 8th of January 2008 in the equatorial Pacific and (b) between the 1st and 31 of March 2006 in the NW Pacific. The colored dots show the in situ DOC measured within the maps period.……..……96

Fig 3.24 a) DOC concentration map estimated with NN29s for the period of Septenber-2007 to February-2008 when La Niña event developed. MAPD ((XX-YY)/(XX+YY) * 200) maps of DOC (b), SST (c), MLD (d), *Chl-a* (e) and $a_{cdom}(443)$ (f) calculated for the same period versus the annual average of each variable (2002 to 2012). The dots show the in situ DOC measured took within the La Niña event in (a) and only the location of the measurements in panels e to f.....…… 97

Fig 3.25 Annual average of DOC concentrated generated with PISCES model (a), and with NN29s (b), the density plot (c) and the MAPD ((XX-YY)/(XX+YY) * 200) map comparing them (d)...…….....98

Fig 3.26 Monthly climatology of DOC concentration in the surface ocean generated with PISCES with in situ DOC monthly climatology from DV2……...…….....99

Fig 4.1 Global map of the annual average relative contribution of POC to TOC produced with 8 days composite data from 2002 to 2012, with a spatial resolution of 25km……..…106

List of tables

Table 3.1 Summary of the data sets for the development and validation of the different algorithms (MLR, NN29b, NN29bChl, NN29s, etc……73

Table 3.2 Number of possible combinations calculated according to the amount of input variables used (SSS, SST, PAR, MLD, $a_{cdom}(443)$ and *Chl-a*) considering that each variable is tested at 5 different time lags. …..…….……...……79

Table 3.3 Sample size with increasing dimensionality required to maintain a constant standard error of the probability of an input estimated in the ANN pattern layer (Silverman, 1986).…….……...…..79

Introduction

The global carbon cycle traces the exchange and storage of carbon among numerous different reservoirs of the Earth system (Ito et al., 2020). This consists of two domains: the fast domain with relatively rapid turnovers time (0 to ~12000 years) represents only 0.3% of the total carbon but presents high exchange flux ranging from ~10 to 100 Pg C yr^{-1}; and the slow domain with turnovers time > 12000 years, containing 99.7% of the total carbon but with an exchange flux of only ~0.01 to 0.1 Pg C yr^{-1} (https://earthobservatory.nasa.gov/features/CarbonCycle; Kandasamy and Nath, 2016).

The fast domain is characterized by fast exchange fluxes between the different sub-domain (air, land, ocean, soil, sediments, freshwater). The carbon of this latter domain is composed by about 1.67% of atmospheric carbon (730 Pg C), 88.6% of oceanic carbon (38700 Pg C), 4% of oceanic surface sediments (1750 Pg C), and on land 1.26% of carbon from the vegetation (550 Pg C), 4.46% from soils (1950 Pg C), and 0.004% from freshwaters (1.7 Pg C). In contrast, the slow domain contains 15 x 10^6 Pg C localized in rocks and deep sediments (Sundquist, 1986).

The two domains are not independent, thus the fast domain receives carbon from the slow domain through volcanic emissions of CO_2 (0.1 Pg C), chemical weathering (0.3 Pg C), and erosion and sediment formation on the sea floor. The natural exchange fluxes between the two domains are relatively small (<0.4 Pg C yr^{-1}) and constant over the last few centuries (Kandasamy and Nath, 2016).

Along the carbon cycle, atmospheric CO_2 is reduced through photosynthesis on land and in the ocean to be later oxidized back to CO_2 through natural processes such as biological growth, respiration, ecological dynamics (competition of organisms and fire disturbance), gas solubility, atmospheric transport, and anthropogenic activities, including fossil fuel and biomass combustion and land use change (Kandasamy and Nath, 2016).

Thanks to the preservation of air bubbles in ice cores, a reconstruction of the historical atmospheric CO_2 concentrations has been done (Indermühle et al., 1999; Petit et al., 1999). This showed that CO_2 has varied quasi-periodically over the last 400000 years oscillating between approximately 180 ppm and 280 ppm (Fig 1 a). Lowest values of CO_2 concentration coincided with fully glacial conditions, while highest concentrations

coincide with interglacial conditions. Thus, it is thought that these changes are connected with the oceanic carbon cycle, which controls the atmospheric CO_2 on timescales longer than a few hundred years (Sarmiento and Gruber, 2006).

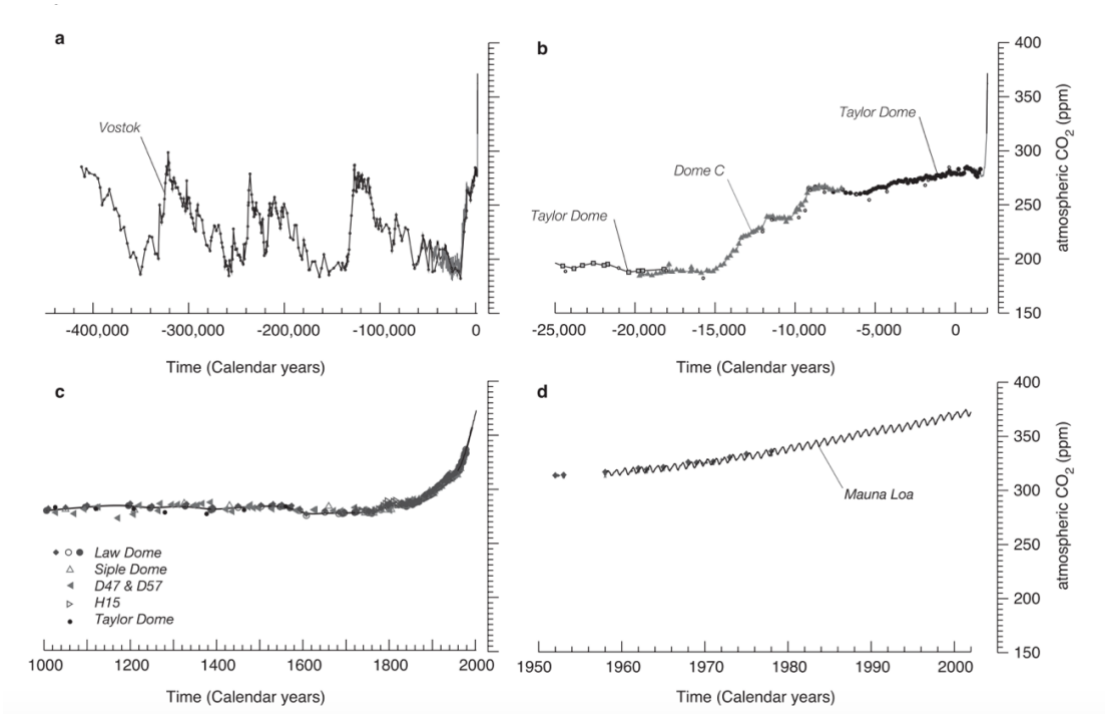


Fig 1 Variations in the atmospheric CO_2 concentration over time based on ice core reconstructions and direct observations since 1958. (a) CO2 variations over the last 400 kyr (1 kyr = 1000 years). (b) CO2 variations from 25,000BC to present. (c) CO2 variations during the last 1000 years as reconstructed from Antarctic ice cores. (d) CO2 variations during the last 50 years as directly measured in the atmosphere at Mauna Loa, Hawaii. The Mauna Loa data are from Keeling and Whorf (1998), the Law Dome, Antarctica, data from Etheridge et al. (1996), the Taylor Dome, Antarctica, data from Indermühle et al. (1999, 2000), the Dome C data from Monnin (2001), and the Vostok, Antarctica, data from Petit et al., (1999). Figure from Sarmiento and Gruber (2006).

In the last 200 years, since the beginning of the industrial revolution, a drastic increase of atmospheric CO_2 is observed (Fig 1; Crutzen and Stoermer 2000) as a consequence of the production of energy by burning the fossil fuels (coal, oil and gas), along with the conversion of forests and other pristine areas into lands for agricultural and other human use (Sarmiento and Gruber, 2006). This rising atmospheric CO_2 content seems to induce an effective exchange of fluxes between the atmosphere and its two major sinks, the land and oceans (Kandasamy and Nath, 2016).

The ocean reservoir of carbon can be divided into two groups of compartments: inorganic (~37,100 Pg C) and organic (~700 Pg C) (Kandasamy and Nath, 2016, Hansell et al., 2009). The total organic carbon (TOC) in the ocean can be found in a particulate (POC) or dissolved (DOC) states, this later component being the principal component of TOC (Sharp, 2002). For example, in situ measurements of surface waters of the Atlantic ocean have shown that only 10 % of TOC is POC (Kumari and Mohan, 2018; Santana-Falcón et al., 2017), a similar percentage (11%) is found for the Baltic sea (Maciejewska and Pempkowiak, 2014) and a much lower contribution of POC to TOC (0.02 to 5 %) has been observed in the NE Pacific (Kumari and Mohan, 2018) or in the Mediterranean Sea (1.3 to 3.7%; Seritti et al. 2003). Due to their different role in the carbon cycle, as well as their different carbon export pathways toward the deep ocean, the spatio-temporal distribution of POC and DOC as well as their relative contributions to the TOC have to be better characterized over the global ocean.

Particulate organic carbon (POC) gathers organic carbon particles with a diameter > 0.4 μm (Duforêt-Gaurier et al., 2010). It can be locally produced by phytoplankton, bacteria, zooplankton, and organic detritus (e.g. fecal pellets and marine snow), or may be transported to a certain location of the ocean from distant sources by oceanic horizontal currents, as well as by river outflow (Evers-King et al., 2017; and references therein). After its generation in the euphotic zone of the ocean, part of the POC is exported to the deep ocean via the "carbon biological pump" (CBP; Anderson and Ducklow, 2001; Volk and Hoffert, 1985).

The CBP comprises all processes through which biogenic carbon from the euphotic zone is sequestrated in the deep ocean to be mineralized, maintaining the strong vertical gradients of oceanic inorganic carbon (Ducklow et al., 2001). The export of POC is driven by passive sinking and active transport by planktonic migrations (Ducklowet al., 2001; Sanders et al. 2014). The passive sinking is strongly related to the production of transparent exopolymer particles (TEP) and biominerals (opal and calcite) by phytoplankton (Armstrong et al., 2002; De La Rocha and Passow, 2007). The TEP leads to the formation of sinking detrital particles that transport phytoplankton material from surface waters to the deep ocean (Passow, 2002). On the other hand, zooplankton heterotrophic activity contributes to the acceleration of the sinking speed of the organic

material through the repackaging of organic carbon produced via photosynthesis into fecal pellets (Turner, 2002).

Nevertheless, only 1 % of surface primary production is thought to be sequestered in the deep ocean (Ducklow et al., 2001; Poulton et al. 2006). The portion of POC that is not exported to the deep ocean can be either transferred to higher trophic levels through the food chain, transformed into detritus, or recycled via the microbial loop, with some of it going into the pool of dissolved organic (DOC) and inorganic carbon (DIC). Therefore, POC is involved in two important carbon fluxes in the ocean, primary production and export to either the deep ocean or the dissolved organic and inorganic carbon pools, DOC and DIC respectively (Evers-King et al., 2017).

The DOC is the largest organic carbon reservoir in the ocean. Depending on how biologically and photochemically available is it can be categorized in labile, semi-labile and refractory (Carlson, 2002; Jiao et al., 2010; Sarmiento and Gruber, 2006). These three categories present different distribution and turnover time. The labile DOC constitutes 1% of the total DOC bulk in the ocean, being found up to 300 m depth with a turnover time of minutes to days (Fuhrman and Ferguson, 1986; Hansell and Carlson, 1998b; Keil and Kirchman, 1999). The semi-labile DOC represents 15-20% of the net production in the euphotic zone (Hansell and Carlson, 1998b). For its resistance to rapid microbial degradation this DOC turnover time ranges from months to years (Carlson, 2002; Hansell, 2002). Consequently, it accumulates in the surface and can be transported horizontally by wind driven currents or exported to deep water via meridional overturning circulation and ventilation (Carlson et al., 1994; Copin-Montégut and Avril 1993; Hansell et al., 2002, 2009; Hansell and Carlson 2001; Hopkinson and Vallino, 2005). At last, the refractory DOC is the most resistant fraction to microbial remineralization (Barber, 1968), being reactive only at a multi-millennial time scale with residence time reaching up to 12,500 years. Thus, it represents the greatest percentage of oceanic DOC pool (94%) distributed at all depths (Hansell et al., 2009).

DOC is mostly locally produced in the euphotic zone by all the components of the food web, starting with the fixation of atmospheric CO_2 by phytoplankton. Subsequently, is partially consumed by the heterotrophic bacterial community at the beginning of the microbial loop (Sarmiento and Gruber, 2006) or degraded due to photolysis by ultraviolet (UV) irradiation at the ocean surface (Mopper et al., 1991). The most recalcitrant fractions

of DOC (semi-labile and refractory DOC) escape the rapid mineralization processes. These fractions can be then exported out of the euphotic zone by transport and mixing processes such as subduction, convection and diffusion contributing to the biological carbon bump and its deep ocean sinks (Hansell et al., 2009). Its export through overturn of the ocean water column plays a central role in the carbon biological pump (Hansell et al., 2009).

The processes leading the carbon pump have been studied, but due to the complexity of the in situ measurements needed for the evaluation of DOC dynamics, only limited samples of DOC in localized areas for some specific period of time are only available (Aurinet al., 2018; Fichot and Benner, 2012; Letscher and Moore, 2015). Even though worldwide distributed databases have been built (e.g. GOCAD, NOAA, etc) gathering data from different missions, they do not provide enough information to fully understand the temporal variability of DOC in the global ocean surface and water column (Roshan and DeVries, 2017).

Satellite remote sensing is a powerful tool providing a global view of optically significant components in the ocean, such as *Chl-a* concentration, dissolved organic matter (CDOM) concentration or suspended particulate matter. While POC has been successfully estimated from ocean color algorithms in open ocean (Gardner et al., 2006; Kostadinov et al. 2016; Loisel et al., 2002; Stramski et al., 1999, 2008), the DOC estimation in open waters at global scale is still challenging, despite some few relatively recent attempts using sea surface temperature (Siegel et al., 2002) or ocean color radiometry (Aurin et al., 2018). Because CDOM represents the colored portion of DOM (Fig. 2), which gathers dissolved organic carbon, nitrogen and phosphorus (Sharp, 2002), it has been explored the possibility use its absorption property (a_{cdom}) to estimate DOC concentration. In coastal waters, the estimation from remote sensing observation of DOC is made possible due to 1) the accuracy of CDOM inversion from space in these environments (e.g. Loisel et al. 2014; Mannino et al., 2008) and 2) to the presence of strong CDOM-DOC relationships which are following in same dilution patterns (e.g. Mannino et al., 2008; Matsuoka et al., 2017; Vantrepotte et al., 2015).

Nevertheless, in open ocean their dynamics are temporally decoupled (Nelson et al., 1998; Nelson et al., 2010; Nelson and Siegel, 2002), since the kinetic of the processes driving the distribution of DOC are fundamentally different from those driving CDOM

distribution (Nelson and Siegel, 2002). Furthermore, in these areas where CDOM concentration is relatively low when compared to coastal waters, the similarity of the absorption of CDOM and non-algal particles (NAP) tends to further complicate the specific estimation of the absorption signals of CDOM. Therefore most of the existing inversion models aim to estimate the sum of the CDOM and NAP (referred as Colored Detrital Matter absorption, Maritorena et al., 2002). Therefore, it was assumed that other parameters than a_{cdom} allowing to impact the DOC variability should be used as possible predictors to estimate DOC from remote sensing. As follows, some models have been developed using sea surface temperature (Siegel et al., 2002) and the combination of a_{cdom} and sea surface salinity (Aurin et al., 2018), but with lukewarm results and sometimes erroneous global patterns. For instance, the SSS and ocean color radiometry (OCR) based approach (Aurin et al., 2018) drives the lowest DOC values in the gyre areas where DOC is supposed to present high values (Roshan and DeVries, 2017). In the same way, the DOC spatial distribution is too constrained latitudinally by the SST based approach (Siegel et al., 2002).

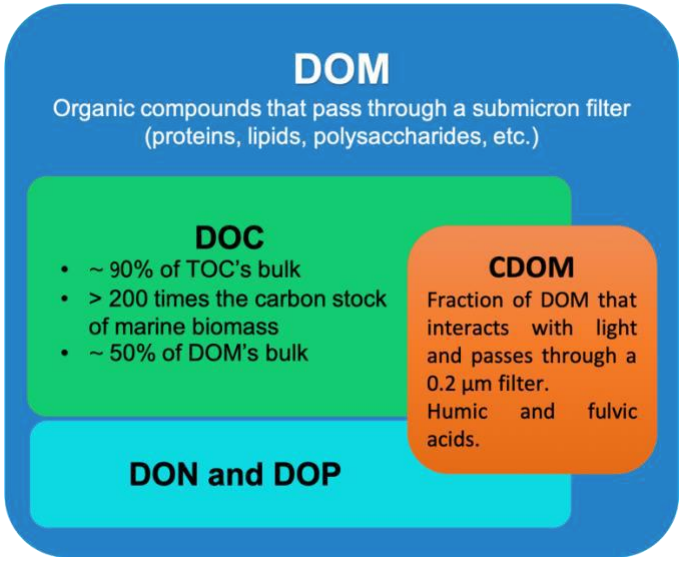


Fig 2 Representation of Dissolved Organic Mater (DOM) composition including the Dissolved Organic Carbon (DOC) and its fraction in the Total organic Carbon (TOC) bulk, the Dissolved Organic Nitrate (DON) and Phosphate (DOP) and the Colored Dissolved Organic Mater (CDOM).

It is in this framework that this PhD arises with the main objective of developing an algorithm to estimate the concentration of DOC in the open ocean from spatial remote sensing radiometry. For that purpose, the novelty of the proposed approach is based on the involvement of information on the bio-optical and physical status of the waters mass

examined at a given time. Because of the strong decoupling between DOC and $a_{cdom}(\lambda)$ we decide to introduce in this algorithm a temporal dimension allowing to consider "the history" of the water mass. To achieve this, first it is necessary to defined the best algorithm to estimate a_{cdom} from satellite ocean color radiometry observations over open ocean waters. Then, it is necessary to define first what are the pertinent variables, besides $a_{cdom}(\lambda)$, which provide the best performance in the estimation of DOC concentration and second at which time lag these later variables have to be taken into account. At last, the model structure (Neural Net, Multi-Linear Regression, etc.) to produce accurate outputs has to be defined.

The structure of the manuscript is as follows. The first chapter presents the knowledge of DOC sources, sinks and distribution in the open ocean, along with CDOM dynamics information. This chapter gathers information from previous studies based on in situ measurements, ocean color radiometry and biogeochemical models.

In the second chapter, different ocean color models developed to estimate a_{cdom} in open ocean waters from remote sensing are tested and compared to a new one developed in the frame of this PhD. The best performing algorithm is selected and implemented to process ocean color data collected over 10 years to characterize the spatio-temporal variability of $a_{cdom}(\lambda)$ over open ocean waters. These patterns are then discussed with regards to *Chl-a* and the absorption of colored dissolved and detrital material.

In the third chapter, the algorithm to estimate DOC concentration over open ocean waters from satellite images is described. For this purpose, different variables are tested at different time lags as possible estimators of DOC concentration. The performance of the new model is then validated with in situ data gathered from different missions and sampling stations. The temporal and spatial variability previously published of the estimated DOC is characterized and compared to results obtained with models based on satellite data (Aurin et al., 2018; Siegel et al., 2002). The results are also compared to the annual average picture generated by Roshan and DeVries (2017) through the implementation of a neural network to extrapolate global in situ data filling the gaps. At last, the outputs of the present algorithm are compared with the DOC concentration derived from the biogeochemical model PISCES (Aumont et al., 2003, 2015) to better understand the DOC spatio-temporal patterns observed.

1. Chapter 1: State of the Art

1.1 Dissolved Organic Matter (DOM)

The Dissolved Organic Matter (DOM) is a complex compound operationally defined as the filtrate passing through a 0.45-μm filter pore whereas the particulate organic matter (POM) corresponds to the matter gathered on the corresponding filters (Schnitzer and Kahn, 1972; Fig 1.1). It most likely corresponds to a very fine colloidal suspension rather than a chemical solution. DOM's composition is described as a complex mixture of low-molecular weight substances (carbohydrates, amino acids, peptides, small carboxylic acids and alcohols) which constitute the easily decomposed labile DOM pool (Kulovaara et al., 1996; Sachse et al., 2005) and a refractory pool formed by higher molecular weights components, aromatic elements, lignin, humic compounds (Tranvik et al., 2009), and even bacteria-derived organic matter, for example recalcitrant peptidoglycan from cell walls (Keil et al., 2000; Yamada and Tanoue 2006). This refractory DOM pool also includes the black carbon (BC), formed by highly unsaturated organic recalcitrant compounds. The lifetimes of DOM constituents range from minutes to millennia according to their level of lability (Carlson et al., 1994, 1996; Hedges, 2002).

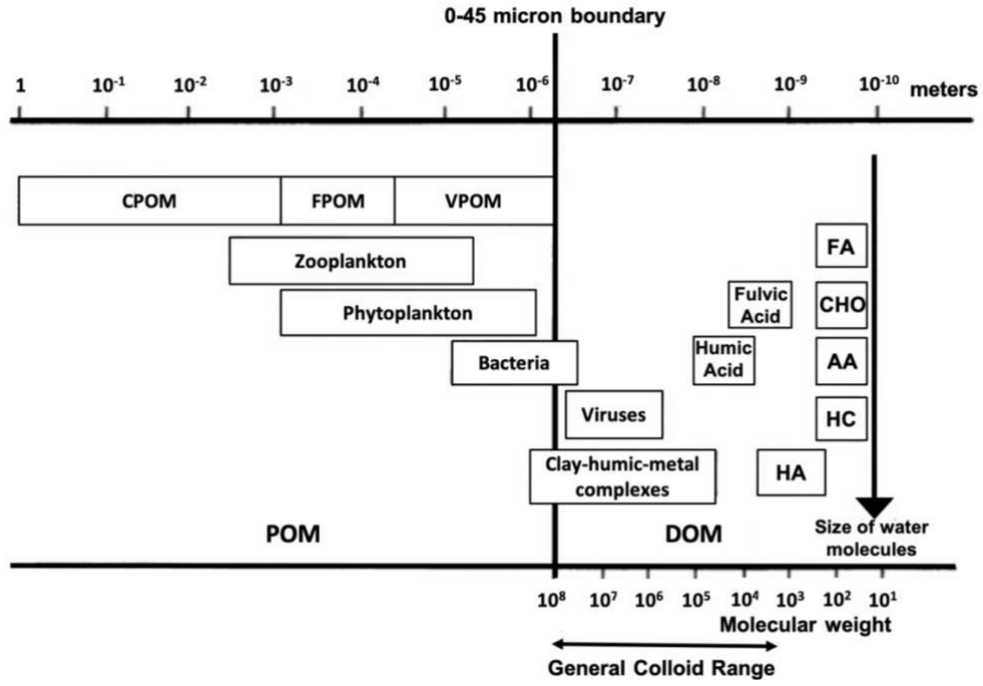


Fig 1.1 Size range of particulate (POM) and dissolved organic matter (DOM) and organic compounds in natural waters. AA, amino acids; CHO, carbohydrates; CPOM, coarse particulate organic matter; FA, fatty acids, FPOM, fine particulate organic matter; HA, hydrophilic acids; HC, hydrocarbon; VPOM, very fine particulate organic matter. From Nebbioso and Piccolo (2013).

DOM therefore includes a variety of chemical elements including mainly carbon plus a mixture that includes nitrogen, phosphate, oxygen, hydrogen, and trace amounts of other elements (Moody and Worrall, 2017).

The importance of characterizing DOM variability and composition in natural-water ecosystems lies in the number of processes in which it is involved (Nebbioso and Piccolo, 2013). DOM is a strong chelating agent for metals, thus affecting their solubility, transport, and toxicity (Schnitzer and Kahn, 1972). It is also involved in the transport of organic pollutants (Carter and Suffet, 1982) formation of colloidal particles (Tipping, 1986), pH-buffering (Oliver et al., 1983) and the distribution of ions between aqueous and solid phases (Jenne, 1975). It further serves as substrate to heterotrophic microbial populations and as a source of nitrogen and phosphorus to nutrient-starved autotrophs (Hansell et al., 2009).

The colored part of the DOM (referred as CDOM: Colored Dissolved Organic Matter) is acting as a controlling parameter of numbers of photo-dependent processes (Zafiriou et al., 1984). Due to its interaction with light, CDOM can be detected optically from its absorption coefficient, and then can potentially be estimated from ocean color remote sensing observations. The carbon part of DOM referred as DOC (Dissolved Organic Carbon) represents the largest reservoir of organic carbon in the ocean (Hansell et al., 2009; Nebbioso and Piccolo, 2013). In this PhD, a focus is performed on the two latter components of the DOM which are detailed in the further sections.

1.2 Dissolved Organic Carbon (DOC)

The Dissolved Organic Carbon (DOC) constitutes approximately half of DOM's bulk (Moody and Worrall, 2017). This carbon reservoir estimated to 662 Pg C is the largest reservoir of organic carbon in the ocean with more than 200 times the carbon budget of marine biomass (Hansell et al., 2009; Nebbioso and Piccolo, 2013).

As briefly mentioned previously for DOM, the DOC can be categorized in three major pool according to its availability to biological and photochemical degradation processes: labile, semi labile, and refractory (Carlson, 2002; Jiao et al., 2010; Sarmiento and Gruber, 2006).

1.2.1 Labile DOC

The biologically labile DOC is the fraction available for microbial consumption (Zweifel, 1999), consequently its turnover time ranges from minutes to days (Fuhrman and Ferguson, 1986; Keil and Kirchman, 1999). Labile DOC composition consists on a mix of high molecular weight and low molecular weight compounds found in the euphotic zone (Fig 1.2 a). It is mainly produced by autotrophic organisms in the euphotic zone of productive areas of the ocean, such as: the tropical open ocean upwelling systems (mostly the Equatorial Pacific), the Southern Ocean, the sites of coastal upwelling, western boundary currents, and estuarine systems. Even though labile DOC represents a large flux of carbon in the ocean that supports large portion of heterotrophic bacterial growth, it only constitutes 1% of the total DOC bulk in the ocean because of its fast turnover (Hansell and Carlson, 1998b).

1.2.2 Semi-labile DOC

Semi-labile DOC is a more biologically resistant fraction with a turnover on time scales of months to years. It is constituted by a mix of high and low molecular weight carbohydrates that have strong conservative spectroscopic and chemical properties throughout the global ocean (Aluwihareet al., 1997; Amon and Benner, 1996; Benner et al., 1992). This fraction represents 15–20% of net community productivity in the euphotic zone (~ 1.8 Pg C yr^{-1}; Hansell and Carlson, 1998b). Being resistant to rapid microbial degradation, semi-labile DOC accumulates in the surface ocean (Carlson, 2002; Hansell, 2002), and can be transported by currents to be relocated in different areas of the global ocean. Due to its resistance to rapid degradation it can be dragged to depths up to 500 m by Ekman transport (Fig 1.2 a; Goldberg et al., 2009; Skoog and Benner, 1997;) or to deeper regions via meridional overturning circulation and ventilation (Carlson et al., 1994; Copin-Montégut and Avril, 1993; Hansell, 2002; Hansell et al., 2009; Hansell and Carlson, 2001; Hopkinson and Vallino, 2005).

1.2.3 Refractory DOC

The refractory DOC is the most resistant fraction to microbial remineralization (Barber, 1968) being only reactive at a multi-millennial time scale. Some works have documented a residence time reaching up to 12,500 years. Accordingly this part of the DOC is assumed to represent 94% of the oceanic DOC pool (624 Pg C; Hansell et al., 2009).

Because the average age of the deep DOC is greater than the time scale of thermohaline circulation, refractory DOC is reintroduced to the surface waters as it follows the path of ocean circulation (Carlson, 2002).

The refractory DOC pool is dominated by low molecular weight DOM (Amon and Benner, 1996; Benner et al., 1992; Skoog and Benner, 1997) and is assumed to represent the bulk of deep (>1000 m) DOC stocks (Bauer et al., 1992; Williams and Druffel, 1987). However, Hansell and Carlson (1998a) examined the change of DOC along the deep-ocean and suggested that part of the deep DOC pool is slowly remineralized over time scales of decades to centuries. This fraction of DOC is more recalcitrant than the semi-labile DOC, but more labile than the refractory. Carlson et al. (2010) refers it as semi-refractory DOC. The semi-refractory DOC pool presents annual to multi-decadal time scale of removal (Druffel et al., 1989; Bauer et al., 1992).

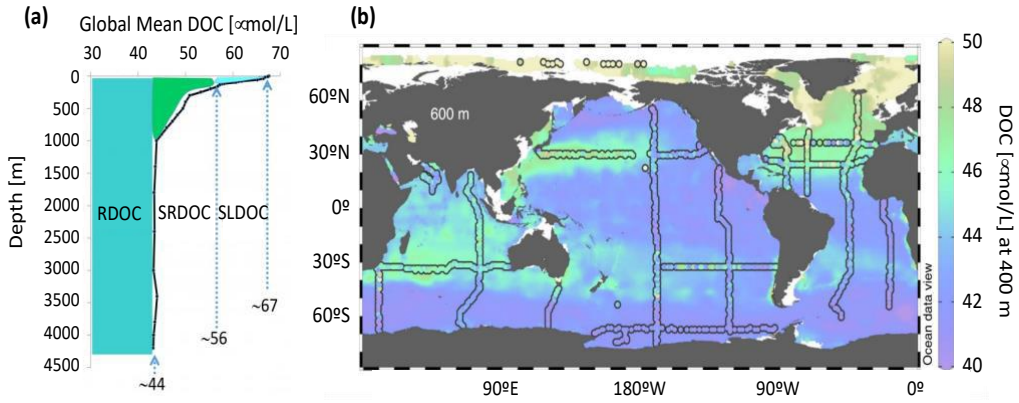


Fig 1.2 Vertical profile of DOC and its different fractions, refractory (RDOC), semi-refractory (SRDOC), semi-labile (SLDOC) adapted from Hansell (2013) (a). Distribution of ANN-derived (Color map) and observational DOC (colored dots) at 600m from Roshan and DeVries (2017).

The refractory DOC can be found at all depths and due to its multi-millennial residence time in studies focused in shorter time periods it can be consider constant. Roshan and DeVries (2017) simulated a DOC concentration map at 600 m depth, under the pycnocline, applying an artificial neural network and compared their results with in situ measurements of DOC (Fig 1.2 b). This map shows in situ concentrations ranging between 40 and >50 μmol/L, in accordance with the annual average concentration provided by Hansell (2013) considering that at such depth it is expected to find a mixture of refractory and semi-refractory DOC for which the mean global concentration variates from 44 to 56 μmol/L (Fig 1.2 a)

1.3 DOC sources, sinks and processes

DOC has an important role in the long-term sequestration (decades to centuries) of carbon, with a great impact over the biological pump (Boyd et al., 2019; Hansell et al., 2009). The biological pump is the sum of processes that transport biogenic carbon from the surface euphotic zone to the ocean's interior where the material is mineralized, driving respiration in the ocean and maintaining the ocean's strong vertical gradients of inorganic carbon and nutrient concentrations (Fig 1.3). The main components of the pump are passive particulate carbon sink, active vertical migration by zooplankton, and DOC mixed downward from the surface (Ducklow et al., 2001).

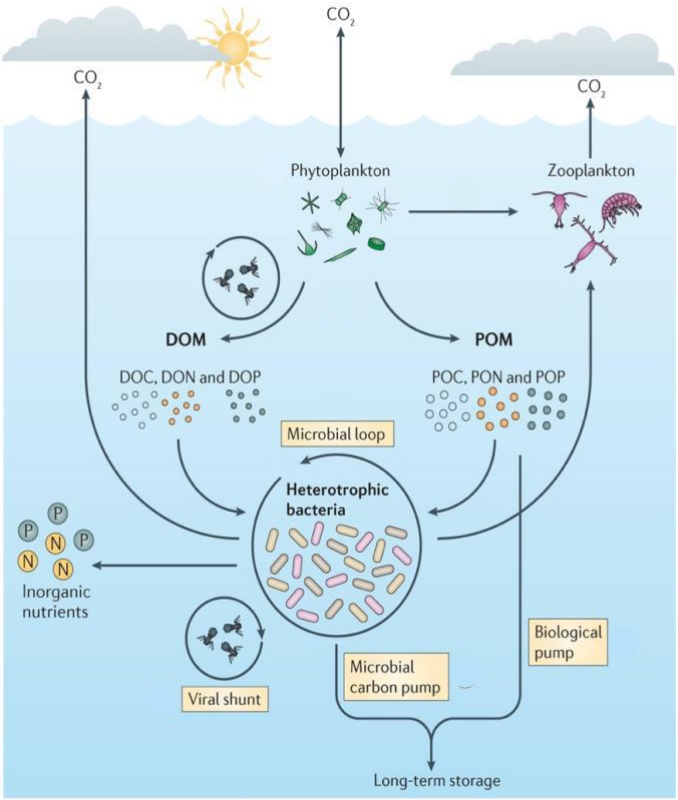


Fig 1.3 Carbon pump schema representing the fixation, of CO_2 by phytoplankton and posterior degradation mineralization and sequestration of DOC into the deep ocean. Diagram from Buchan et al., (2014).

1.3.1 DOC sources

Most marine DOC is produced autochthonously in the euphotic zone of the oceans resulting from the temporal and spatial uncoupling of in situ biological production and consumption processes led by photosynthetic plankton (e.g. diatoms, bacteria, algae) or microfauna (Fig 1.3; Hansell et al., 2002; Hansell and Carlson, 2015; Nebbioso and

Piccolo, 2013). The quantity and quality of DOC produced during these bloom events is conditioned by numbers of biological, chemical, and physical parameters. DOC production is constrained by the magnitude of primary production, and several mechanisms related to this: phytoplankton extracellular release, grazer mediated release and excretion, release via cell lysis (both viral and bacterial), solubilization of particles, and bacterial transformation and release (Carlson, 2002).

The DOC production from autotrophic organisms in the upper layer of the ocean related to the increased primary production, is partially consumed after the phytoplankton bloom period. The more recalcitrant DOC, which is not rapidly mineralized, is exported to the deep ocean (Carlson et al., 1994). The amount of DOC that escapes rapid remineralization varies across environments and seasons (Hansell and Carlson, 1998b; Hansell and Peltzer, 1998; Romera-Castillo et al., 2016).

Semi-labile DOC accumulates in the surface and is transported by currents to be relocated in different areas of the global ocean. For example, in the subtropical gyres where the rates of primary production and of POC export are the lowest, it is possible to find high production of DOC (~15 $gCm^{-2}yr^{-1}$; Goldberg et al., 2009; Roshan and DeVries, 2017; Skoog and Benner, 1997), suggesting a decoupling between net DOC production and nutrient availability. This is caused by the effect of large-scale wind-driven circulation in the low latitudes which generate upwelling in the equatorial region that prevents DOC export to deeper waters, while poleward surface flows transport DOC accumulated in the tropics to the subtropics where it is subducted (Roshan and DeVries, 2017).

Besides the incoming of DOC produced in other oceanic regions, the nutrient-depleted stably stratified ecosystems of the gyres are conducive to the presence of picoplankton (Roshan and DeVries, 2017). These small plankton (diameter < 2 μm) are adapted to this impoverished environment (Bragg et al., 2010) and remain efficient producers of many types of DOM compounds (Zhao et al., 2017), enhancing the production of DOC. In addition to this, the system presents an efficient microbial loop, which processes the net primary production through the marine food web in the euphotic zone, accumulating refractory DOC as fresh organic matter is degraded (Jiao et al., 2010). This refractory DOC pool only becomes available to marine microbes after being exported to deeper waters, out of the euphotic zone (Carlson et al., 1994), where the formation of organic gel (Verdugo & Santschi 2010, Verdugo 2012) by deep-ocean DOC generates aggregates

which adsorption into suspended and sinking particles reduces DOC concentration (Hansell, 2013).

In the tropics, the upwelling systems controlled by variations in wind, regulate the uplift of nutrient-enriched surface waters stimulating an active primary production. The supply of nutrient to the euphotic zone depends on the intensity of the upwelling, which reshapes the community structure (Abbott and Zion, 1985; Hanson et al., 2005; Loureiro et al., 2011). Thereby DOC accumulation in such areas are strongly dependent on the physical conditions driving the intensity of the upwelling. Nevertheless, the quantity and quality of its production is ultimately constrained by the amount of primary production and varies considerably depending on the magnitude of the generated bloom (Carlson, 2002).

In the tropical and subpolar regions where the availability of nutrients is greater the net DOC production is low. The plankton communities are constituted by larger phytoplankton cells (nano- and microplankton), which promote the formation of fast-sinking particle aggregates and fecal pellets (Hirata et al., 2011) and a less-efficient food web. As well, the greater availability of nutrients facilitates the microbial consumption of DOC (Letscher et al., 2015).

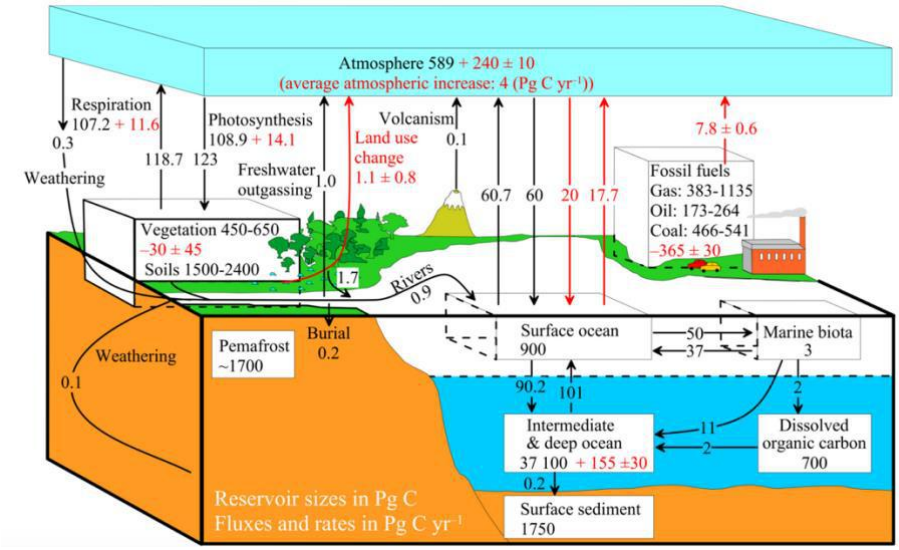


Fig 1.4 Simplified diagram of the global carbon cycle. Numbers denote reservoir mass in Pg C (1 Pg C = 10 15 g C) and annual carbon fluxes in Pg C yr $^{-1}$ between the atmosphere and the land and ocean. Black numbers and arrows indicate reservoir mass and exchange fluxes estimated for the time prior to the Industrial Era, about 1750. Red arrows and numbers indicate annual " anthropogenic " fluxes averaged over the 2000–2009 time period. Red numbers in the reservoirs represent cumulative changes of anthropogenic carbon over the Industrial Period 1750–2011. Diagram from Kandasamy and Nath (2016).

Alongside the autochthonous production (Fig 1.3), a part of the marine DOM is a by-product of the biological production over terrestrial ecosystems This allochthonous fraction of marine DOC is transported to the ocean waters by land washing, rivers, lakes, glaciers and other natural sources (Fig 1.4; Griffith et al. 2012; Schlesinger and Melack, 1981). The transfer of terrestrial carbon into the sea is an important component of the global carbon cycle (Griffith et al., 2012; Schlesinger and Melack, 1981). The highly stratified surface of the Arctic Ocean is for instance enriched in DOC by the input of terrigenous organic matter via high fluvial fluxes to the system (Dittmar and Kattner, 2003), and by redistribution to higher latitudes with the wind-driven circulation (Hansell, 2002).

1.3.2 DOC sinks, export and sequestration

The main loss mechanisms for DOC in the surface of the ocean include biotic and abiotic processes. The first comprise the uptake of DOC by heterotrophic bacteria which respiration accounts for a large fraction of primary production in most oceanic ecosystems (Ducklow, 1999). The second sink process is equally important and consist in the photooxidation of DOC by solar ultraviolet radiation (Christian and Anderson, 2002). On the other hand, DOC is also removed from the surface ocean by export to the mesopelagic zone by Ekman transport in the subtropical region (Goldberg et al., 2009; Skoog and Benner, 1997) and by sequestration into the deep ocean due to ventilation (Carlson et al., 1994; Copin-Montégut and Avril, 1993; Hansell, 2002; Hansell et al., 2009; Hansell and Carlson, 2001; Hopkinson and Vallino, 2005; Fig 1.3).

DOC concentration in the ocean is very low and variable depending on depth and latitude (Hansell et al., 2009). Mean DOC concentrations present systematic meridional trends within depth horizons characteristic of the epipelagic zone (0–200 m), mesopelagic zone (200–1000 m) and bathypelagic zone (1000–4000 m). DOC export from the surface ocean to the deep ocean is a process that includes its accumulation in the euphotic zone in the tropics, redistribution to higher latitudes with the wind-driven circulation, and eventual transport to depth with the overturning circulation at high latitudes and subduction in the subtropical gyres (Hansell, 2002).

In the euphotic zone mean DOC concentration varies between 40 and 80 μmol/L. Maximum values of 70–80 μmol/L are observed in the tropical and subtropical systems

(40°N to 40°S) where vertical stratification of the upper water column favors the accumulation of organic matter (Hansell et al., 2009) along with poleward surface flows transport DOC accumulated in the tropics to the subtropics (Roshan and DeVries, 2017). Here part of the accumulated DOC can be exported to depths of a few hundred meters due to Ekman convergence of surface waters where it is mineralized. Most of the DOC that transits along this path is returned for exchange with the atmosphere within months to years (Hansell et al., 2009).

On the contrary, lowest concentrations of DOC ~ 40–50 μmol/L in the surface are observed in the subpolar region and in the circumpolar Southern Ocean (> 50°S; Sarmiento and Gruber, 2006) where transported with the wind-driven surface currents from low to high latitudes DOC is exported to depths >1000 m via meridional overturning circulation and ventilation to be long-term sequestrated in the interior of the ocean (Carlson et al., 1994, 2010; Copin-Montégut and Avril, 1993; Hansell, 2002; Hansell et al., 2009; Hansell and Carlson 2001; Hopkinson and Vallino, 2005). This consists on deep-water formation with low-DOC which is brought to the surface, diluting the near-surface DOC concentrations and exporting DOC into the deep ocean (Carlson et al., 2010). In the deep ocean DOC is remineralized and the vertical gradient is regenerated with a great impact over the total carbon export (Hansell, 2001).

In the global ocean, semi-labile DOC net export represents approximately 20% of global export production (1.8 Pg C yr^{-1}), nonetheless this process is mainly located in the low latitude where the export of DOC does not go much deeper than 100 m. Only 0.3 Pg C yr^{-1} reaches depths > 500 m decreasing the DOC export - POC export rates with depth (Carlson et al., 2010). As such, DOC mineralization makes its greatest contribution to oxygen consumption in the upper ocean (up to 70% of oxygen consumption at < 400 m; Abell et al., 2000; Doval and Hansell, 2000).

Refractory DOC removal in the upper euphotic zone has been attributed to photolysis by ultraviolet (UV) irradiation at the ocean surface (Mopper et al., 1991 and transformation to suspended particles and/or interaction with them (Druffel et al., 1992). UV photolysis in the surface layer can oxidize refractory organic matter (Mopper and Kieber, 2002), transforming 'refractory' DOC to a biologically available form (Benner and Biddanda, 1998).

In the deep ocean (meso and bathypelagic zones) the distribution of DOC presents a ~ 29% decrease in concentration along the path of the deep global thermohaline circulation (from the deep North Atlantic to the deep North Pacific; Hansell and Carlson, 1998a; Hansell et al., 2009). Bauer et al. (1992) showed there is a DOC age difference of 1600 years between the Sargasso Sea and the Southern Ocean, suggesting that a portion of the refractory fraction (and/or semi-refractory) might be consumed along the deep ocean circulation across the Atlantic (Hansell et al., 2009).

Hansell et al. (2009) estimated that the DOC concentrations in the deep ocean over a single circulation of the abyss ranges from 34 to ~ 50 μmol C kg^{-1}. The largest deep ocean DOC gradients along intermediate and deep ventilation pathways (~ 12 μmol C kg^{-1}) are observed in the North Atlantic basin. Vertical input from North Atlantic Deep Water formation results in bathypelagic DOC concentrations > 50 μmol/L north of 50°N. Meanwhile in the equator the DOC concentration decreases to 40–45 μmol/L and reaches the lowest concentrations of about 39 μmol/L in the deep south Atlantic at 25°–50°S (Hansell et al., 2009). The great gradient observed in the deep Atlantic Ocean is a product of the biotic remineralization of DOC as well as dilution produced by the mixing with DOC-impoverished water from the Antarctic at intermediate and bottom depths. The deep Atlantic represents a DOC sink of ~ 86 Tg C yr^{-1}, calculated as water mass formation rates times DOC concentration gradients (Hansell et al., 2009).

In the deep Pacific Ocean, the DOC concentration is lower than in the Atlantic with a sink of ~ 43 Tg C yr^{-1} of DOC (Hansell et al., 2009). It presents an isopycnal gradient as a result of the injection of relatively DOC-enriched waters from the circumpolar deep layer of the Southern Ocean, due to deep ventilation. The near-bottom water mass is transported northward along the deep Pacific while DOC is gradually mineralized. In transit the path DOC concentration decreases from ~ 42 μmol/L in the deep south Pacific to ~ 36 μmol/L in the deep North Pacific. The near-bottom water mass that enters into the North Pacific from the south, gains buoyancy via vertical mixing and rises to the mid water column where it returns south as Pacific Deep Water (PDW). During southward transit of PDW, DOC continues to decline, reaching a concentration of~ 34 μmol/L at mid depth in the South Pacific (Hansell et al., 2009).

The removal of DOC in the deep ocean has been related to abiotic interactions with particles (Hansell et al., 2009), where biopolymers imbedded in seawater, such as gels

and transparent exopolymers (Carlson, 2002; Passow and Alldredge, 1994; Wells, 1998), capture organic molecules rising them up to the particle size spectrum (Engel et al., 2004; Verdugo and Santschi, 2010; Verdugo et al., 2004). This establishes that 15% of the sinking POC that reaches the deep ocean is originated from adsorption (or addition via gel formation) of recalcitrant DOC (Druffel and Williams, 1990), this represents four times the rate of organic carbon sequestration in deep ocean sediments (Dunne et al., 2007; Lochte et al., 2003). Thus, most of the DOC-derived abiotically formed particles are mineralized while still suspended in the water column or after falling to the ocean bottom. Thus, that refractory DOC conversion to particles is a mechanism through which recalcitrant organic matter is transformed to a more biologically available form (Hansell et al., 2009).

1.4 Colored Dissolved Organic Matter (CDOM)

The colored dissolved organic matter (CDOM), also known as gelbstoff, gilvin and yellow substance, is the colored fraction of the total dissolved organic material (DOM) (Coble, 2007) present in all natural waters (e.g., Siegel et al., 2002).

The chemical composition of CDOM is very diverse gathering many humic substances, including lignins, phenols, and other plant degradation products (Thurman, 1985), along with sugars, amino acids, and other small molecules polymerized in the ocean due to UV radiation (Harvey and Boran, 1985; Harvey et al., 1983). Among the non-humic components of marine CDOM are pigment-like components and amino acid or protein-like components (Coble, 1996; Coble et al., 1990, 1998; Mayer, 1999; Mopper and Schultz, 1993;) which provide evidence of CDOM production autochthonously in the ocean (Coble, 2007).

Operationally it is defined as material that passes through a submicron filter (usually 0.2μm) and appreciably absorbs light in the UV and short visible wavelengths (Nelson and Siegel, 2013). CDOM absorption presents a decreasing exponential shape towards long wavelengths with strong absorption in the UV and blue spectral domains, being usually modeled as an exponential function over a short wavelength interval (Jerlov, 1976; Bricaud et al., 1981).

$$a_{cdom}(\lambda) = a_{cdom}(\lambda_0)e^{-S(\lambda-\lambda_0)} \quad \text{(Eq 1.1)}$$

where λ_0 is the reference wavelength, and S (nm^{-1}) the CDOM absorption spectral slope.

While CDOM quantification is assessed from its absorption (a_{cdom}) at a defined wavelength, information on the CDOM quality or origin can be examined though its spectral slope (Carder et al., 1989; Lee et al., 2010). Freshwater from coastal environments generally present lower a_{cdom} slopes than oceanic environments (Blough and Del Vecchio, 2002; Bricaud et al., 1981; Coble, 2007; Nelson et al., 1998, 2004, 2007, 2010; Nelson and Siegel, 2013; Twardowski et al., 2004).

Because of its absorbing spectrum increasing exponentially with decreasing wavelength, CDOM strongly contributes to the regulation of UV penetration into the ocean and mediates numbers of photochemical reactions. CDOM plays a crucial role in many biogeochemical processes at the surface ocean including primary productivity and the air–sea exchange of radiatively important trace gases (e.g., Arrigo and Brown, 1996; Mopper et al., 1991; Toole et al., 2006; Toole and Siegel, 2004; Zepp et al., 1998). Moreover, for estuarine waters and for coastal waters strongly influenced by river inputs, light absorption by CDOM often dominates the absorption by phytoplankton in the blue portion of the visible spectrum with direct impact over the primary production and ecosystem structure (Blough and Del Vecchio, 2002).

Light absorption measurements in different areas and depths of the ocean have shown that CDOM is always present in the marine environment (Nelson et al., 2007, 2010; Swan et al., 2009). The presence of CDOM in the deep ocean suggests that a fraction of oceanic CDOM is biologically refractory (Nelson and Siegel, 2013). CDOM is very complex and divers, depending on its labile fraction, age, origin and whether it has transitioned from freshwaters to marine (Coble, 2007). It gathers humic substances of high molecular weight (HMW, the more labile ones) and of low molecular weight (LMW) such as fulvic acids (Carder et al., 1989; Sempéré and Cauwet, 1995).

As pointed out for DOC, the sources of CDOM are variable depending on the region of the oceans (Fig 1.5). In estuaries and coastal waters CDOM concentration is very high and mainly of terrestrial in origin. The principal source of CDOM comes from rivers discharge followed by land washing (Blough and Del Vecchio, 2002). Other possible sources are benthic inputs from seagrass and corals, resuspension events caused by storms which introduce porewater CDOM in coastal waters (Boss et al., 2001).

In open ocean, on the other hand, in areas where DOM of terrestrial origin is not advected, CDOM is locally produced and is directly related to biological activity in the water column (Mobley et al., 2004) along with upwelling and convective export (Nelson and Siegel, 2002). The biological processes involved in the production of CDOM include release or excretion by organisms and lysis of cells by viruses. CDOM composition is dominated by marine humic compounds and new CDOM of biological origin, residual products of phytoplankton and other organic particles degradation (Bricaud et al., 1981; Coble, 2007; Nelson et al., 1998, 2010; Nelson and Siegel, 2002; Prieur and Sathyendranath, 1981).

Although CDOM in oceanic water can be considered as a by-product of phytoplankton and associated organic matter no direct correlation between CDOM and phytoplankton is observed (Bricaud et al., 1981). This decoupling can be explained by the difference in kinetic between phytoplankton biomass and CDOM concentration (which rely on past phytoplankton concentrations; Mobley et al., 2004) as well as to the impact of the different sink processes driving the net CDOM balance, such as microbial activity, photooxidation, and other abiotic processes.

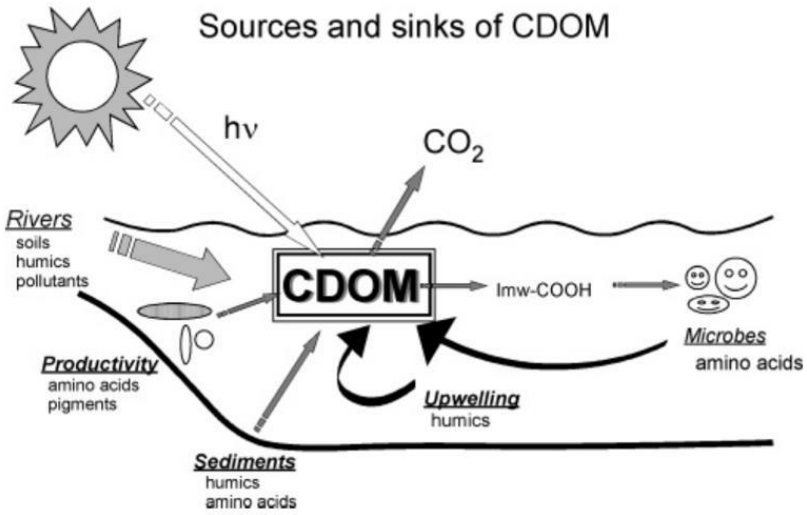


Fig 1.5 Schema of sources and sinks of CDOM to the ocean from Coble (2007)

1.5 CDOM to DOC relationships

Over water masses influenced by terrestrial inputs of DOM, CDOM and DOC distribution follow the same dilution pattern leading to the presence of a significant CDOM-DOC relationships often describe by a linear function. Nevertheless, the relationship between

DOC and CDOM is highly variable in time and space due to variation in the origin and quality of the DOM (Mannino et al., 2008; Fig 1.6 a). Numerous studies (e.g. Fichot and Benner, 2012; Vantrepotte et al., 2015) have demonstrated the potential of using the CDOM slope in the UV domain for constraining this seasonal or inter regional variability and thus for depicting DOC content over contrasted coastal regions.

On the contrary, such direct relationships are not observed in open ocean (Fig 1.6 b) other than the Arctic, where DOC concentrations is tightly correlated with CDOM absorption (r2 = 0.97; Matsuoka et al., 2012). This lack of correlation is due to the temporal decoupling of the different processes driving CDOM and DOC dynamics (Nelson et al., 1998, 2010 ; Nelson and Siegel, 2002). In the subtropical Sargasso Sea for instance, no correlation exists between DOC and CDOM in the upper water column (Nelson et al., 1998). This appears to be a result of the summertime photobleaching of the surface CDOM which does not have a noticeable impact upon the concentration of DOC (Siegel and Michaels, 1996). In some areas CDOM and DOC can even be negatively correlated, situation observed in the Southern Ocean (Nelson and Siegel, 2013; Weishaar et al., 2003).

This suggests that in open ocean the kinetic of the processes driving the surface DOC distribution are fundamentally different from those driving the CDOM distribution (Nelson and Siegel, 2002). Because a_{cdom} is the only way in which ocean color is impacted by DOC, some other independent knowledge of water type is needed for retrieval of DOC from space (Aurin et al., 2018).

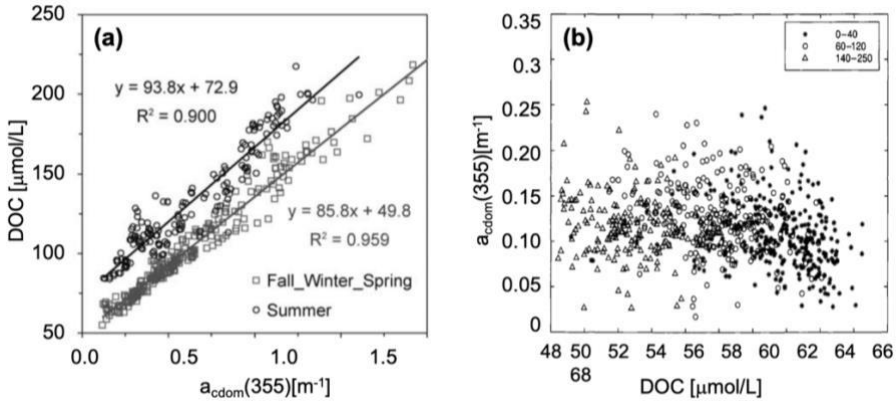


Fig 1.6 (a) Relationship of $a_{cdom}(355)$ in the Chesapeake Bay mouth and plume region for the 2004–2006 research cruises and Delaware Bay mouth from Mannino et al. (2008). (b) Relationship of DOC and $a_{cdom}(325)$ from samples taken on BATS cruises in the Sargasso Sea from spring of 1994 through the end of 2000 (Nelson and Siegel, 2002).

1.6 Current assessment of CDOM and DOC distribution at global scale

1.6.1 CDOM in situ

Various methods allow to measure CDOM absorption coefficient which mainly include spectroscopy based methods commonly used for discrete samples analysis and fluorescence based techniques (Bricaud et al., 1981¸ 2010; Carder et al., 1989; D'Sa et al., 1999; Green and Blough, 1994) which able continuous measurements. This methods provide a good description of CDOM status in a particular location. Global databases such as GOCAD and NOMAD (Aurin et al., 2018; Werdell and Bailey, 2005) have been built by gathering discrete measurements from diverse missions and stations (Fig 1.7). These databases with over 48000 data points provide a good overview of global DOC distribution in space, but yet not in time. A possible solution for this comes along with the use of innovative platforms (Bio-Argo) that allow to have a larger description of the CDOM distribution at the sea surface and along the water column. Anyway, this platform measures the fluorescence, a limited parameter to characterize CDOM in open-ocean due to its low concentration (Nelson and Siegel, 2002).

It is because of this that even when the in situ data allows a partial description of the CDOM distribution in terms of temporal and spatial coverage, it is limited to capture the parameters driving CDOM dynamics at global scale. In that respect, OCR represents a valuable tool for obtaining a synoptic view of CDOM distribution at global scale.

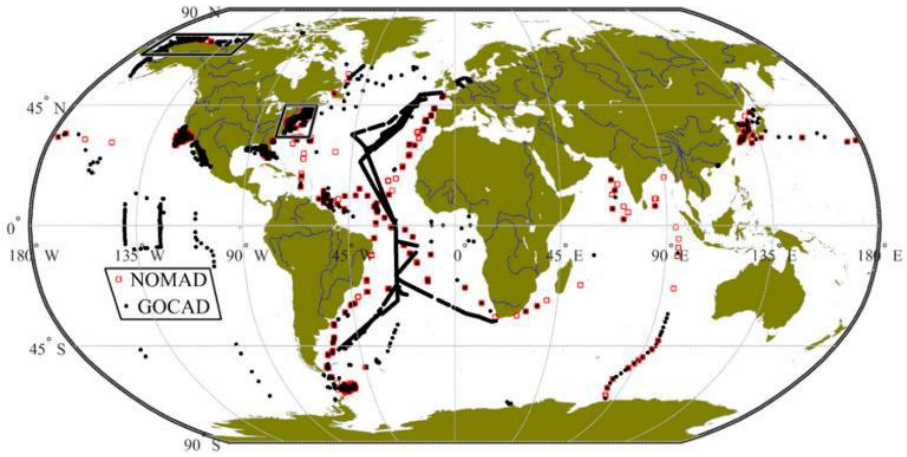


Fig 1.7 Global distribution of GOCAD and NOMAD field stations for CDOM concentration from Aurin et al. (2018)

1.6.2 CDOM from space

The absorption of CDOM (a_{cdom}) represents up to 90% of the non-water UV absorption in the water (Johannessen et al., 2003; Nelson et al., 1998; Zepp et al., 2007) being the main controller of the UV radiation (280 to 400 nm) penetration into the open ocean. At 443 nm and at the surface of the oceans, a_{cdom} is assumed to contribute more than 50% to the light absorption (Babin et al.; 2003; Bricaud et al., 2002, 2010; Siegel et al., 2002;). Due to its strong impact on the water masses absorption budget, CDOM can therefore be detected from OCR. It is worth noticing that this strong absorption in the domain might represent an issue for optically estimating the chlorophyll-a concentration (Carder et al. 1991; Loisel et al., 2010; Siegel et al., 2005a, 2005b, 2013).

While the shape of the absorption properties of a_{cdom} and phytoplankton (a_{ph}) strongly differ, high similarity exists between CDOM and non-algal particles (a_{nap}) spectral shape (Lee et al., 2002; Loisel et al., 2010; Morel and Gentili, 2009; Fig 1.8).

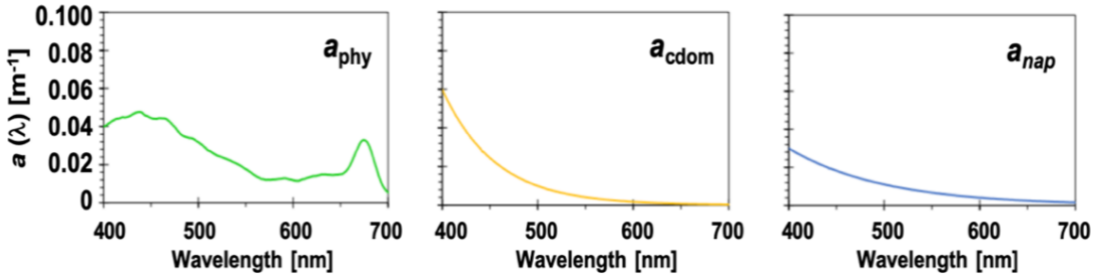


Fig 1.8 Scheme of absorption spectra of phytoplankton (a_{phy}), CDOM (a_{cdom}) and NAP (a_{nap}).

Consequently a_{cdom} and a_{nap} are not easily distinguishable from OCR observation (Coble, 2007). In coastal waters, which contain a considerable amount of CDOM and CDM of terrestrial origin and where CDOM and NAP dynamics can be decoupled, specific CDOM inversion models have been developed (e.g. Cao et al., 2018; D'Sa et al., 2002; Johannessen et al., 2003; Kahru and Mitchell, 2001; Loisel et al., 2014; Mannino et al., 2008; Matthews, 2011; Tehrani et al., 2013).

Conversely, most of the existing ocean color algorithms over open ocean waters, where CDOM concentration is very low, were initially developed for estimating the absorption coefficient of Colored Dissolved and Detrital Matter (CDM) which is defined as the addition of CDOM and NAP absorption (Siegel et al., 2002; Organelli et al., 2016). The first Ocean Color Radiometry (OCR) algorithm dedicated to the estimation of $a_{cdom}(\lambda)$ was based on the use of variable $a_{cdom}(443)$ vs *Chl-a* relationships (Morel and Gentili,

2009) providing only a relative estimation of $a_{cdom}(\lambda)$, since it is calculated in reference to a standard Chlorophyll content. Later, two purely empirical approaches based on a blue-to-green reflectance ratios (Shanmugam, 2011), or multi linear relationships (MLR) between $a_{cdom}(\lambda)$ and $R_{rs}(\lambda)$ at different wavelengths (Aurin et al., 2018) have been proposed to assess $a_{cdom}(\lambda)$ from ocean color observation. Because these two algorithms are purely empirical, it is thought that their performance might be highly dependent on the dataset used for their development. Also, because the model of Shanmugam (2011) uses the blue to green R_{rs} band ratio, also used for the estimation of satellite Chl-*a*, it is invalid to evaluate the how the two products correlate since their estimation is not in independent.

Two semi-analytical approaches have also been proposed, this is the case of the one based on matchup between satellite GSM-derived $a_{cdm}(443)$ values (Maritorena et al., 2002) and *in situ* $a_{cdom}(\lambda)$ measurements presented by Swan et al. (2012) to assess $a_{cdom}(\lambda)$ from $a_{cdm}(443)$. This model presents good estimate of a_{cdom} in open ocean within the UV spectrum, but as $a_{cdm}(443)$ is used as input, it is not suitable to evaluate how they covary. The second semi-analytical approach developed by Chen et al. (2017) involves other inherent optical properties (IOPs) to assess $a_{cdom}(443)$ in coastal and open ocean water.

Based on validation exercises performed for each of the mentioned models, these have shown to retrieve good estimates of $a_{cdom}(\lambda)$ in the open ocean. Thus it is necessary to perform a validation exercise over a common dataset, to be able to propose which of them performs the best approach to adequately assess $a_{cdom}(443)$ over open ocean, calculating the $a_{cdom}(443)$ variability with regards to the chlorophyll concentration and $a_{cdm}(443)$ and quantifying $a_{cdom}(\lambda)$ contribution to the total absorption budget.

1.7 DOC global distribution

1.7.1 The different approaches providing the DOC spatial distribution at global scale

1.7.1.1 In situ

A global picture of the DOC distribution at different depth (20, 300 and 600 meters) has been generated by Roshan and DeVries (2017) from an artificial neural network applied to a DOC in situ data base (see the DOC transects in Fig 1.9) gathering data collected at from a recent compilation (Letscher and Moore, 2015) plus CLIVAR repeat section A10 (completed in 2011). In situ data set of salinity, temperature, macronutrients, chlorophyll, light penetration, and dissolved oxygen are used as input parameter of the

ANN. The annual average presented by the authors follows the expected patterns, but this method does not allow to analyze the temporal variability of DOC (due to the limitation of the in situ DOC data base, despite its relatively large spatial coverage).

The in situ measurement of DOC can be only performed from discrete water samples limited in time and space and the analysis of DOC is highly time consuming (Fichot and Benner, 2012).

In this context, it has been explored the possibility of estimating DOC concentrations through the measurement of the optical properties of dissolved organic matter (DOM) (absorption and fluorescence) what represents a more efficient alternative since it can be rapidly and continuously acquired in situ (Vodacek et al., 1997; Hitchcock et al., 2004). Nevertheless, as it has been already mentioned, the relationship between DOC and DOM absorption (a_{cdom}) varies among geographical regions and seasons (Blough and Del Vecchio, 2002) limiting the capability to predict DOC concentration from simple linear relationships with CDOM. Therefore it is imperative to find a proper optical estimator to asses DOC concentration. This would enable the estimation of DOC concentration with satellite data giving the possibility to analyze its global variability in time and space.

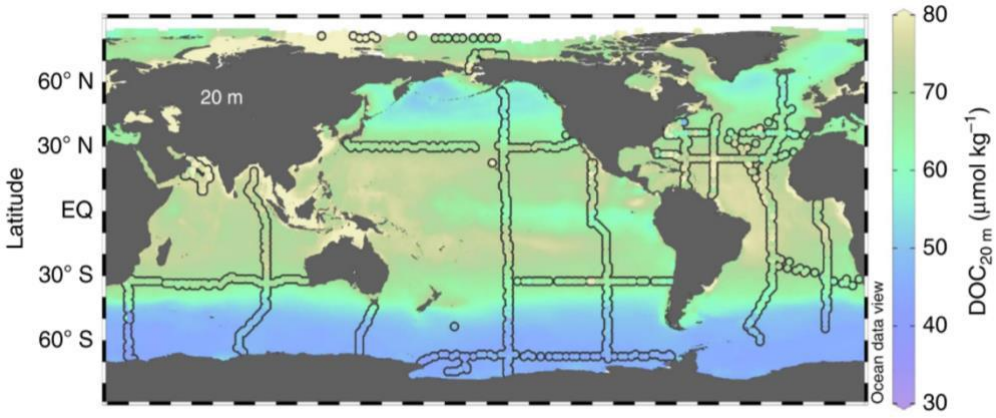


Fig 1.9 Distribution of ANN-derived and observational DOC. Color map is the artificial neural network (ANN)-derived dissolved organic carbon (DOC) concentration, and colored dots are the observed DOC concentration at 20m. Figure from Roshan and DeVries (2017)

1.7.1.2 DOC from remote sensing

There have been a few algorithms developed to estimate DOC in global scale from ocean color remote sensing based on different premises. Siegel et al. (2002) presented a model to estimate DOC from sea surface temperature (SST). This model is based on the linear

correlation calculated from in situ measurements of both, DOC and SST, in the Pacific, Atlantic, Indian and Southern oceans. Their results do not clearly show the expected patterns described by in situ DOC observations with high concentrations (~80 μmol/L) in the subtropical gyres. Instead the DOC distribution presents high concentration over quasi-uniform bands between the tropics that smoothly decreases towards higher latitudes (Fig 1.10), following the well know spatial pattern of SST, the unique input variable of this algorithm. The values estimated between the tropics range between 75 and 80 μmol/L in the Pacific basin, >80 μmol/L in the Atlantic and around 70 μmol/L in the Indian ocean. These results show high overestimation of DOC concentration in the equatorial band and especially in the tropical Atlantic, and a slight underestimation in the North Atlantic where higher concentrations are expected (~50 μmol/L) due to DOC accumulation before being exported to the deep ocean due to ocean ventilation.

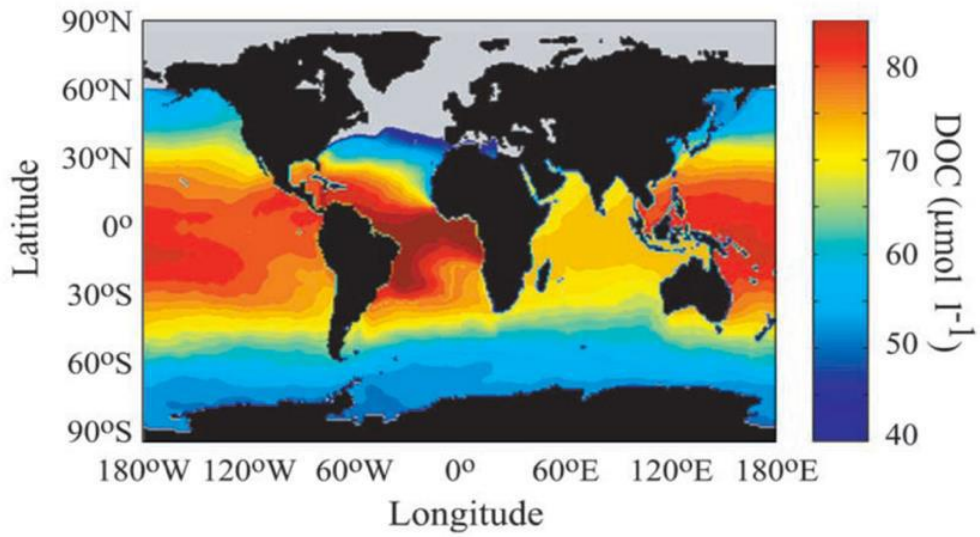


Fig 1.10 Climatological DOC distribution from a regression analysis based upon wintertime SST values. Figure from Siegel et al. (2002).

Another model proposed by Aurin et al. (2018) estimates DOC from a_{cdom} (355) and Sea Surface Salinity (SSS). This model tries to compensate the lack of direct correlation between DOC and a_{cdom} in open ocean by using also SSS data as input with the objective of tracing DOM distribution that is not detected by ocean color sensors. Nevertheless, this task is not accomplished. The authors highlighted the weak performance of the model to estimate DOC in global scale with large differences between estimated values and in situ measurements in several regions (±~50% – ~100%) indicating fundamental

weaknesses in the algorithm (Fig 1.11). For instance, the minimum of DOC concentration are here observed in the subtropical gyres, which is not consistent with current knowledge, as discussed previously. In the same way DOC is highly overestimated at high latitudes with concentrations higher than 70 µmol/L.

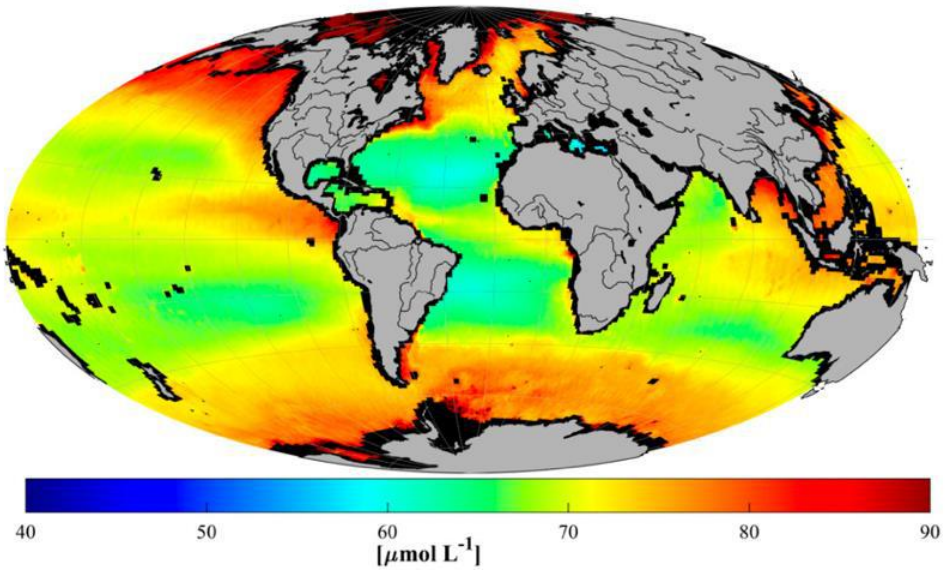


Fig 1.11 Retrieved three-year mean, 9 km nominal resolution DOC from Aquarius and MODIS Aqua using the MLR2 inversion from Aurin et al. (2018). Figure from Aurin et al. (2018).

1.7.1.3 DOC from coupled bio-physical models: the example of the PISCES model

PISCES (Pelagic Interactions Scheme for Carbon and Ecosystem Studies) is a biogeochemical model which simulates marine biological productivity and describes the biogeochemical cycles of carbon and of the main nutrients (P, N, Si, Fe) (Aumont et al., 2003, 2015). This model retrieves the simulated semi-labile DOC taking in consideration several assumptions that include types of zooplankton, phytoplankton and degradation rates. In general the simulations obtain by this model respect the global patterns observed with in situ measurements, with the exception of some areas (Fig 1.12). For example, the Arctic is underestimated by PISCES DOC which annual mean presents concentrations of ~50 µmol/L, while it is known that due to the great influence of terrestrial inputs DOC concentration is >55 µmol/L (Matsuoka et al., 2013). Thus, even though this model retrieves very good simulations of DOC concentration, it still could be improved and for this needs to be validated with high spatial and temporal resolution observations, what could be achieved with an ocean color model capable of estimating DOC concentration from satellite products.

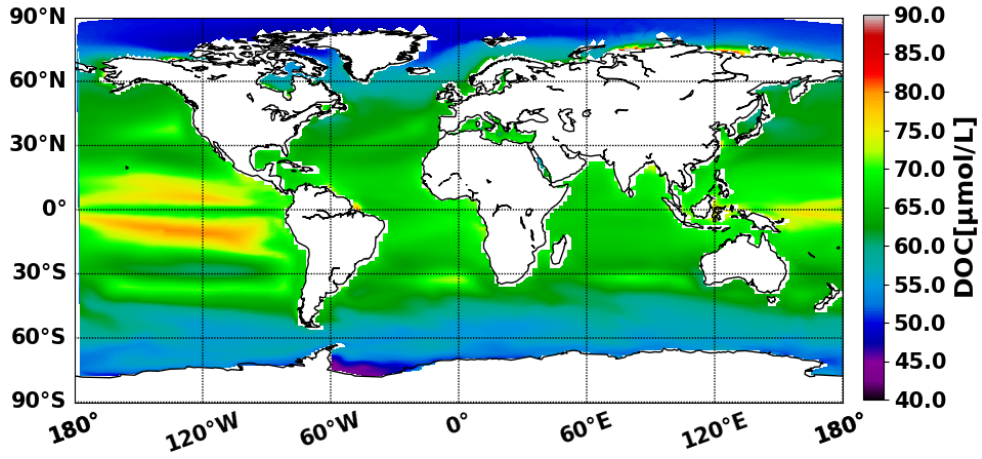


Fig 1.12 DOC annual average concentration simulated with PISCES biogeochemical model.

2. Chapter 2: CDOM estimation at global scale, spatio-temporal variability and contribution to the total absorption budget

Knowledge about the colored part of the DOM distribution and variability represents a prerequisite for further estimate its organic carbon content (DOC). Historical methods allowing to map the absorption coefficient of colored detrital matter (a_{cdm}) from space were not making the distinction between dissolved and particulate detritus.

A first objective of this PhD was thus to propose a new model for estimating the absorption coefficient of CDOM at 443 nm ($a_{cdom}(443)$, m^{-1}) and to compare the performance of this new model with that of the few recent algorithms proposed for specifically quantifying CDOM absorption at global scale from OCR. The generated archive has been further used for describing 1) the main spatial CDOM distribution pattern and range of variability at global scale 2) CDOM covariation with *Chl-a* and CDM 3) to deliver updated information on the relative contribution in the total absorption budget. The obtained results were included in publication submitted to Remote Sensing of Environment entitled '*Colored dissolved organic matter absorption at global scale from ocean color radiometry observation: spatio-temporal variability and contribution to the absorption budget*' which manuscript is provided below.

A second part of this Chapter further details the temporal variability of $a_{cdom}(443)$ over the global ocean performed applying adapted statistical analyses to long lasting Globcolour time series.

2.1 Colored dissolved organic matter absorption at global scale from ocean color radiometry observation: spatio-temporal variability and contribution to the absorption budget

2.1.1 Introduction

Colored dissolved organic matter (CDOM), also known as gelbstoff, gilvin and yellow substance, is the colored fraction of the total dissolved organic material (DOM) (Coble, 2007). Although it represents a small part of the total DOM in the open sea (Nelson et al., 1998, 2010; Nelson and Siegel, 2002; Siegel et al., 2002), CDOM plays a significant role in aquatic photochemistry and photobiology, interfering in various biogeochemical cycles as it absorbs light over a broad spectral range covering visible and UV domains (Aurin et

al., 2018; Blough and Del Vecchio, 2002; Coble, 2007; Kieber et al., 1996; Toole et al., 2006). The absorption of CDOM (a_{cdom}) has been used to estimate CDOM concentration in water (Coble et al., 2004), and is a privileged path to assess Dissolved Organic Carbon, at least in coastal waters (Del Castillo and Miller, 2008; Fichot and Benner, 2011; Mannino et al., 2008; Vantrepotte et al., 2015). The composition of CDOM is very complex and diverse, depending on its origin, labile fraction, age, and whether it has transitioned from fresh waters to marine environment. Coastal waters generally present high concentration of CDOM, mainly of terrestrial origin, introduced to the oceanic system through rivers discharge and land washing (Coble et al., 1998; Tzortziou et al., 2015). For this reason, $a_{cdom}(\lambda)$ can be used as a good tracer of inland waters dispersion in coastal areas (Fichot and Benner, 2012). In contrast to coastal waters, CDOM in open water is dominated by new CDOM of biological origin. In this type of water, and out of areas affected by advection of coastal waters, CDOM is generally considered as a residual product of phytoplankton and other organic particles generated during degradation processes (Bricaud et al., 1981; Coble, 2007; Nelson et al., 1998, 2010; Nelson and Siegel, 2002; Prieur and Sathyendranath, 1981). These latter processes controlling the dynamic of CDOM in open ocean waters are strongly dependent on the coupling between physical and biogeochemical processes which rely on forcing parameters such as the light availability and vertical mixing. The great diversity of the processes controlling a_{cdom} variability make its dynamics over open ocean still not well characterized. Besides the necessity to improve our knowledge on the oceanic spatio-temporal distribution of $a_{cdom}(\lambda)$, the great absorption level of CDOM in the blue spectral domain does represent an issue for estimating the chlorophyll-a concentration, *Chl-a*, from ocean color observation (Carder et al., 1991; Loisel et al., 2010; Siegel et al., 2005a, 2005b, 2013).

The similar spectral behaviors between $a_{cdom}(\lambda)$ and the absorption coefficient by non-algal particles, $a_{nap}(\lambda)$, makes these two absorption coefficients difficult to distinguish from ocean color inverse algorithms. For this reason, the ocean color community has historically focused on the development of inverse algorithms to assess the colored detrital matter absorption coefficient, $a_{cdm}(\lambda)$, which combines the contributions of non-algal particles and CDOM (eg. Boss and Roesler, 2006; Ciotti and Bricaud, 2006; Lee et al., 2002; Maritorena et al., 2002). The great CDOM concentration of surface coastal waters, making its presence easier to detect, has however stimulated the development of empirical or semi-analytical approaches to assess $a_{cdom}(\lambda)$ in coastal waters (e.g. Cao et

al., 2018; Loisel et al., 2014; Mannino et al., 2008). In open ocean waters, where CDOM is present in a much lower concentration than in coastal waters, the first Ocean Color Radiometry (OCR) algorithm dedicated to the estimation of $a_{cdom}(\lambda)$ was based on the use of variable $a_{cdom}(443)$ vs *Chl-a* relationships (Morel and Gentili, 2009). However, as mentioned by the latter authors, this algorithm only provides a relative estimate of $a_{cdom}(\lambda)$, since it is calculated "in reference to a standard Chlorophyll content". More recently, purely empirical approaches based on a blue-to-green reflectance ratios (Shanmugam, 2011), or multi linear relationships (MLR) between $a_{cdom}(\lambda)$ and $R_{rs}(\lambda)$ at different wavelengths (Aurin et al., 2018) have been proposed to assess $a_{cdom}(\lambda)$ from ocean color observation. Based on matchup between satellite GSM-derived $a_{cdm}(443)$ values (Maritorena et al., 2002) and *in situ* $a_{cdom}(\lambda)$ measurements a semi-analytical approach has also been proposed by Swan et al. (2013) to assess $a_{cdom}(\lambda)$ from $a_{cdm}(443)$. Very recently, a semi-analytical approach involving other inherent optical properties (IOPs) has been developed to assess $a_{cdom}(443)$ in coastal and open ocean water (Chen et al., 2017).

This study emerged in this context and aims at i) proposing the best approach to adequately assess $a_{cdom}(443)$ over oceanic areas, ii) assessing the $a_{cdom}(443)$ variability with regards to the chlorophyll concentration and $a_{cdm}(443)$ and iii) quantifying the contribution of $a_{cdom}(443)$ to $a_{cdm}(443)$ and the non-water absorption coefficients, $a_{nw}(443)$, over the global ocean. For that purpose, the performance of different algorithms, including a new one and three previously published algorithms, is evaluated using a large set of *in situ* and matchup data points. The description of these *in situ* and satellite data are first provided. The different selected algorithms are then presented, and the adaptation of a previously published algorithm dedicated to the estimation of $a_{cdom}(412)$ in coastal waters is described. The description of the $a_{cdom}(443)$ spatio-temporal patterns, as well as of its relative contribution to $a_{cdm}(443)$ and non-water absorption coefficient, $a_{nw}(443)$ are then provided.

2.1.2 Materials and methods

2.1.2.1 Datasets description

2.1.2.1.1 Optical typology

In order to evaluate the performance of the different $a_{cdom}(443)$ inversion models considered according to the optical water type characteristics, each sample available was associated with the 16 optical classes defined by Mélin and Vantrepotte (2015) defined from a global classification of the R_{rs} spectral shape (normalized reflectance spectra). An additional class, numbered as 17, has been added to this latter classification to consider the most oligotrophic waters initially not represented in Mélin and Vantrepotte (2015). Data belonging to classes 1 and 2 can be considered as representing turbid water masses strongly impacted by terrestrial inputs. In contrast, samples associated with classes 9 to 17 correspond to waters where the reflectance spectra are well represented by the Case 1 reflectance model by Morel and Maritorena, (2001) while samples for classes 8 to 3 are more likely related to diverse types of Case 2 waters which spectral shape are increasingly departing from the Case 1 modeled spectra.

2.1.2.1.2 In situ and matchup data sets used for validation

Three different datasets were defined for the development, validation and inter-comparison exercises presented in this study. The first dataset (DS1) corresponds to the synthetic ocean color dataset developed by the International Ocean Color Coordinating Group (IOCCG) working group dedicated to inverse algorithm development (IOCCG, 2006). This dataset gathers 500 data points of inherent optical properties (IOPs) and remote sensing reflectance, $R_{rs}(\lambda)$, computed from radiative transfer simulations every 3 nm from 400 to 700 nm for each IOPs combination. The $a_{cdom}(443)$ values for the whole DS1 dataset range between 0.0025 m^{-1} and 2.37 m^{-1}, with a median value of 0.12 m^{-1} (Fig 2.1a). This data set is used for the development of the new algorithm, hereafter referred as CDOM-KD2. The evaluation of the relative performances of the different considered algorithms has been performed using an *in situ* validation dataset (DS2) and a matchup dataset (DS3).

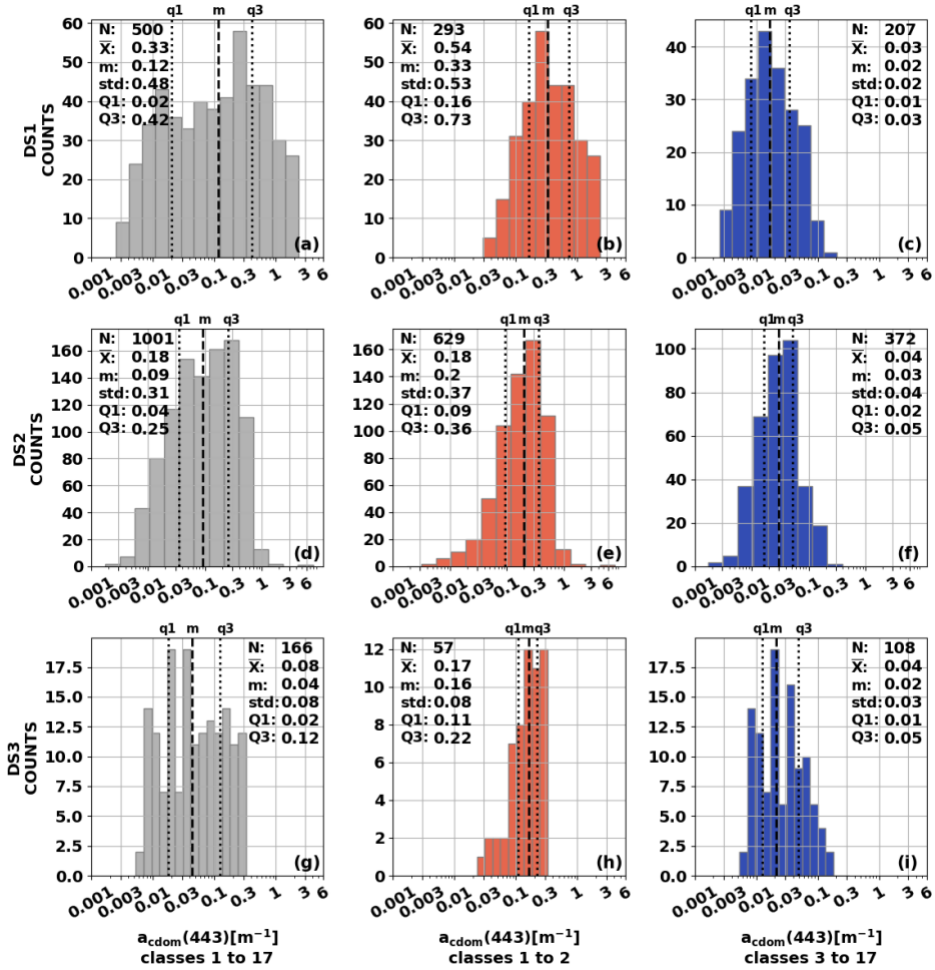


Fig 2.1 From top to bottom, $a_{cdom}(443)$ frequency distribution histograms for the DS1 (a, b, c), DS2 (d, e, f) and DS3 (g, h, i) datasets. The histograms corresponding to the complete datasets are in grey, those for the water classes 1 and 2 subsets are in red, while the subsets gathering the water classes 3 to 17 are in blue. N, $\overline{X}$, m, std, q1, q3 correspond to the number of datapoints, mean, median, standard deviation and first and third quantiles values, respectively.

The DS2 (Fig 2.1d, e, f) data set includes 1001 in situ $R_{rs}(\lambda)$ and $a_{cdom}(443)$ measurements worldwide distributed (Fig 2.2). It gathers data collected from diverse cruises previously presented in Loisel et al. (2018), and other additional data collected within the NOMAD (Werdell and Bailey, 2005) and Plumes and Blooms (https://seabass.gsfc.nasa.gov/experiment/Plumes_and_Blooms) projects not included in Loisel et al. (2018). The $a_{cdom}(443)$ range of variability in DS2 is [0.002; 7.84] m^{-1}, with a median value of 0.094 m^{-1}.

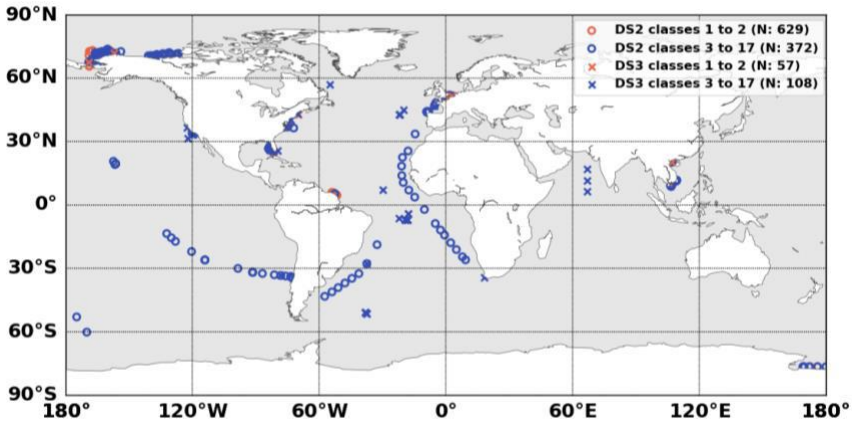


Fig 2.2 Distribution of DS2 (circles) and DS3 (crosses) data points used for comparison of the four models. Red dots represent the highly turbid data points strongly influenced by terrestrial inputs (classes 1 and 2 from the Mélin and Vantrepotte (2015) optical typology) while the blue dots are associated with marine waters belonging to other optical classes.

The DS3 matchup dataset was built from two distinct data sets. First, the GlobColour daily merged L3 Ocean Colour products at 4 km^2 of spatial resolution (http://www.globcolour.info/CDR_Docs/GlobCOLOUR_PUG.pdf) were matched with the *in situ* GOCAD (Aurin et al., 2018) data set and covers the September 1997- August 2012 time period. The matchups were computed following the MERMAID tools protocol (http://mermaid.acri.fr/dataproto/dataproto.php) which is based on the NASA Ocean Color protocol (Bailey and Wang, 2001). In practice, daily matchups (with a 3-hour time window) were produced using a 3x3 pixel window, in which the coefficient of variation of $R_{rs}(\lambda)$ needs to be below 0.15 while the number of valid pixels needs to be above 50% (implying a minimum of 5 valid pixels). The second matchup data set is the NOMAD matchup dataset (Werdell and Bailey, 2005) based only on SeaWiFS observations and gathering data collected between October 1997 and March 1999. To limit the impact of the propagation of $R_{rs}(\lambda)$ errors, due to imperfect atmospheric corrections on the $a_{cdom}(443)$ retrieval accuracy, an additional condition was applied on the selection of the matchup data points for the two matchup data sets. In practice, a matchup data was considered invalid if, for any visible wavelengths used in the algorithms, the absolute difference between satellite $R_{rs}(\lambda)$ and *in situ* $R_{rs}(\lambda)$ was greater than $0.75 * \textit{in situ}\ R_{rs}(\lambda)$ (Fig 2.3). The application of this criterion results to 166 final matchup data points, out of 399 data points satisfying to the first selection criteria. The $a_{cdom}(443)$ range of variability in DS3 is [0.0052; 0.33] m^{-1}, with a median value of 0.08 m^{-1}. Even if DS3 covers the time period of the last ocean color sensors which have been recently launched, OLCI A

and B, matchup for these sensors are unfortunately not available in our present data base of $a_{cdom}(443)$ *in situ* measurements.

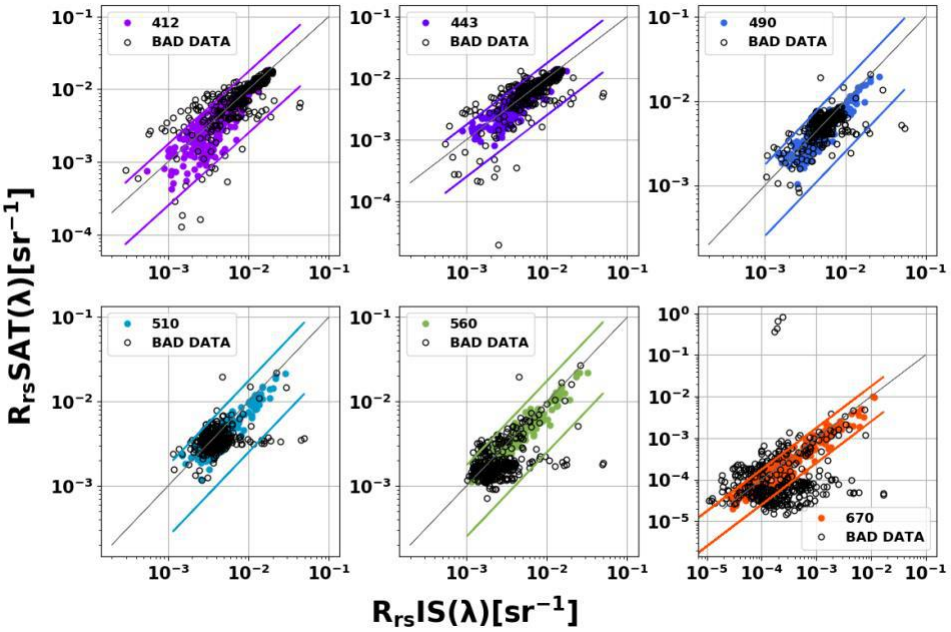


Fig 2.3 Satellite R_{rs} (R_{rs} SAT) vs in situ R_{rs} (R_{rs} IS) from DS3 at (a) 412, (b) 443, (c) 490, (d) 510, (e) 560 and (f) 670 nm. Black solid line represents the 1:1 line, colored solid lines are the threshold limit, where $|R_{rs}$ SAT- R_{rs} IS$| > R_{rs}$ IS * 0.75. Black circles represent the data points for which any of the bands overcomes the settled threshold. This data points where labeled as outliers and not further considered in the analysis.

2.1.2.1.3 Satellite Data used for global CDOM spatio-temporal variability

The global spatio-temporal dynamics of satellite derived of $a_{cdom}(443)$ was assessed from GlobColour L3 merged and OLCI $R_{rs}(\lambda)$ data. GlobColour L3 merged products include satellite observations from SeaWiFS, MERIS, MODIS Aqua and VIIRS NPP sensors. These merged products ($R_{rs}(\lambda)$, *Chl-a*, and $a_{cdm}(\lambda)$) are generated by simple averaging or weighted averaging, depending on the conditions (water types, region, glint/aerosol conditions, etc.). Both *Chl-a*, and $a_{cdm}(\lambda)$ are estimated by the GSM model (Maritorena and Siegel, 2005). Global maps and the time series extraction were produced with GlobColour L3 merged 25 km resolution and 8 days composite data from 23rd April 2002 to 13th April 2012. Due to the lack of matchup data points for OLCI-A/B sensors, the $a_{cdom}(443)$ merged products will be compared to the ones produced by applying the selected algorithm to the OLCI-A/B $R_{rs}(\lambda)$) data.

2.1.2.2 Statistical indicators

The performance of the $a_{cdom}(443)$ inversion models was evaluated from a graphical comparison sustained with quantitative statistical metrics including: the root mean the square deviation (RMSD, Eq 2.1), the median ratio (MR, Eq 2.2), the median absolute percent difference (MAPD, Eq 2.3) and the Pearson correlation coefficient (r).

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(yi-xi)2}{N}} \quad (2.1)$$

$$\mathrm{MR} = \frac{median(yi)}{median(xi)} \quad (2.2)$$

$$MAPD = median \sum_{i=1}^{N} \frac{|yi-xi|}{xi} \times 100 \quad (2.3)$$

where y_i and x_i are the estimated and the in-situ values, respectively.

The MAPD has been calculated considering the median of the individual absolute percent differences between the modeled and measured data instead of the mean to minimize the impact of potential outliers (Loisel et al., 2018).

This statistical parameters were summarized in radar plots where the smallest the area of the polygon is the better is the performance of the model.

2.1.2.3 Models description

In the present paper, four different models are evaluated for estimating $a_{cdom}(443)$ from OCR. These general models, which are based on different assumptions, include two empirical methods recently defined by Aurin et al. (2018) and Shanmugam (2011), and two semi-analytical approaches proposed by Chen et al. (2017) and Loisel et al. (2014). This latter model, dedicated to the estimation of $a_{cdom}(412)$, is here modified and improved to assess $a_{cdom}(443)$.

2.1.2.3.1 Aurin et al. (2018)

Aurin et al. (2018) (further referred to as A2018) recently developed an empirical model for estimating $a_{cdom}(\lambda)$ at global scale. This model is based on a multiple linear regression (MLR) between the natural logarithm of $R_{rs}(\lambda)$ at four different visible wavelengths and the natural logarithm of $a_{cdom}(\lambda)$. It can be described as follows:

$$ln(a_{cdom}(\lambda)) = \begin{bmatrix} \beta 0 + \beta 1 * ln\left(\mathrm{R}_{rs}(\lambda_1)\right) + \beta 2 * ln\left(\mathrm{R}_{rs}(\lambda_2)\right) + \\ \beta 3 * ln\left(\mathrm{R}_{rs}(\lambda_3)\right) + \beta 4 * ln\left(\mathrm{R}_{rs}(\lambda_4)\right) \end{bmatrix} \quad (2.4)$$

where λ_1 to λ_4 are the sensor-specific wavelengths (i.e., 443, 488, 531, and 547 nm for MODIS, 443, 490, 510, and 555 nm for SeaWiFS). Here SeaWiFS bands were used to be consistent with the other models tested. β0 to β4 are the regression coefficients for estimating $a_{\mathrm{cdom}}(443)$ (here β0=-6.41; β1=-0.743; β2=-0.145; β3=-0.367; and β4=0.547).

2.1.2.3.2 Chen et al. (2017)

Chen et al. (2017) (further referred to as C2017) recently developed a semi-analytical model aiming at estimating $a_{\mathrm{cdom}}(443)$ at global scale from the particulate backscattering coefficients, $b_{\mathrm{bp}}(443)$, and the absorption coefficients of phytoplankton, $a_{\mathrm{phy}}(443)$, and colored detrital matter, $a_{\mathrm{cdm}}(443)$, as follows:

$$a_{\mathrm{cdom}}(443) = \chi * a_{\mathrm{cdm}}(443) + \gamma * b_{\mathrm{bp}}(555) + \kappa * a_{\mathrm{phy}}(443) \quad (2.5)$$

where χ, γ and κ are three independent empirical parameters covarying with the water optical properties estimated from the NQAA algorithm (Chen et al., 2016). χ is a function of the $a(\lambda)$-based triangle area index of the total absorption coefficient (TAI). The a(λ)-based TAI is defined as follows:

$$TAI = a(\lambda) - \frac{555-490}{555-443} * a(\lambda_0) - \frac{490-443}{555-443} * a(\lambda_2) \quad (2.6)$$

where a is the total absorption coefficient calculated by NQAA (Chen et al., 2016).

2.1.2.3.3 Shanmugam (2011)

This model, developed for coastal and ocean waters, uses two slope parameters to describe $a_{\mathrm{cdom}}(\lambda)$ in the UV and visible spectral domain as follows:

$$a_{cdom}(\lambda) = a_{cdom}(350) * e^{(-S(\lambda-350)-\gamma^0)} \quad (2.7)$$

where $a_{\mathrm{cdom}}(350)$ is estimated from the blue to green reflectance ratio (Eq 2.8):

$$a_{cdom}(350) = 0.5567 * \left(\frac{\mathrm{R}_{rs}(443)}{\mathrm{R}_{rs}(555)}\right)^{(-2.0421)} \quad (2.8)$$

The spectral slope S is estimated from $a_{cdom}(350)$ and $a_{cdom}(412)$ (Eq 2.10), this latter being also calculated from the blue to green reflectance ratio (Eq 2.9):

$$a_{cdom}(412) = X * \left(\frac{R_{rs}(443)}{R_{rs}(555)}\right)^{(Y)} \quad (2.9)$$

$$S = 0.0058 * \left(\frac{a_{cdom}(412)}{a_{cdom}(350)}\right)^{(-0.9677)} \quad (2.10)$$

The parameter γ^0 in Eq 2.7 takes into account the large variability of CDOM in coastal and ocean waters and is calculated as follows:

$$\gamma^0 = \frac{a_{cdom}(350) - (1/\gamma)}{a_{cdom}(350) + (1/\gamma)} \quad (2.11)$$

where γ is the slope of the hyperbolic model to estimate CDOM (Twardowski et al., 2004):

$$\gamma = 2.9332 * \left(\frac{a_{cdom}(412)}{a_{cdom}(350)}\right)^{(-0.7506)} \quad (2.12)$$

2.1.2.3.4 Loisel et al. (2014)

Loisel et al. (2014) developed a semi-analytical model for estimating $a_{cdom}(412)$ in coastal waters from ocean color remote sensing observations (CDOM-KD1). This model is based on the theoretical link between the vertical attenuation coefficient, $K_d(\lambda)$ and IOPs which has been reformulated as follows:

$$K_d(\lambda) = K_w(\lambda) + f\left(a_{cdom}(\lambda)\right) + \Delta_p(\lambda) \quad (2.13)$$

where K_w is the diffuse attenuation coefficient for pure see water, $f\left(a_{cdom}(\lambda)\right)$ is a function that depends exclusively on the absorption coefficient of CDOM and $\Delta_p(\lambda)$ is the contribution of particles in the attenuation process. To minimize the impact of scattering on the retrieval of a_{cdom} at 412 nm, the model involves the difference of $K_d(\lambda)$ – $K_w(\lambda)$ at two specific wavelengths. Based on these different considerations the model formalism is expressed as follows:

$$a_{cdom}(412) = 10^{[0.15482*(X)^2 + 1.1939*(X) + 0.0689]} \quad (2.14)$$

where $X = \Delta K_d(412-560) - \Delta_p(412-560)$ (2.15)

with $\Delta K_d(412-560) = ((K_d(412) - K_w(412) - (K_d(560) - K_w(560))$ (2.16)

and $\Delta_p(412-560) = 10^{\left[-0.009 * \left(\mathrm{Log}_{10}(\Delta K_d)\right)^2 + 1.147 * \mathrm{Log}_{10}(\Delta K_d) - 0.26\right]}$ (2.17)

In the context of remote sensing applications, ΔK_d is directly estimated from the R_{rs} using a parametrization developed from the IOCCG (2006) data set:

$$\Delta K_d(412-560) = 10^{\left[\begin{array}{c} A * \mathrm{Log}_{10}\left(\frac{R_{rs}(412)}{R_{rs}(560)}\right)^3 + B * \mathrm{Log}_{10}\left(\frac{R_{rs}(412)}{R_{rs}(560)}\right)^2 \\ + C * \mathrm{Log}_{10}\left(\frac{R_{rs}(412)}{R_{rs}(560)}\right) + D \end{array}\right]} \quad (2.18)$$

where A, B, C and D coefficients are -0.12484, 0.160857, -1.2292 and -0.886471, respectively, for a sun angle (θ_s) of 30º.

2.1.3 Results and discussion

2.1.3.1 Adaptation of the Loisel et al. (2014) algorithm for estimating $a_{cdom}(443)$ over the global ocean

A new model, referred to as CDOM-KD2, which consists of an adaptation of the general semi-empirical coastal model published by Loisel et al. (2014), has been developed for estimating $a_{cdom}(443)$ over the global ocean. This adaptation was developed considering the synthetic DS1 data set.

2.1.3.1.1 CDOM-KD2 parameterization

Assuming a restricted CDOM absorption at 560 nm (see section 2.1.2.4.4) $a_{cdom}(443)$ can be expressed as follows:

$$a_{cdom}(443) = \mathrm{f}\left[\begin{array}{c} ((K_d(443) - K_w(443)) - (K_d(560) - K_w(560)) - \\ \Delta_p(443-560) \end{array}\right] \quad (2.19)$$

The attenuation coefficient of light due to pure seawater, $K_w(\lambda)$, has been vastly documented (Morel and Maritorena, 2001; Morel et al., 2007). From literature K_w at 443 nm and 560 nm were set to 0.00948 and 0.0645 m^{-1} respectively (Morel et al., 2007; Loisel et al., 2014). Following the same approach of Loisel et al. (2014), $a_{cdom}(443)$ was empirically modelled as follows:

$$a_{cdom}(443) = 10^{[0.9902 * X - 0.0522]} \quad (2.20)$$

where $X = \Delta K_d\ (443\text{-}560) - \Delta_p\ (443\text{-}560)$ (2.21)

with $\Delta K_d(443\text{-}560) = (K_d\ (443) - K_w\ (443)) - (K_d\ (560) - K_w\ (560))$ (2.22)

Δ_p (443-560), which takes into account the contribution of particulate matter to the attenuation of light, was parameterized from ΔK_d (Fig 2.4 a) as follows:

$$\Delta_p(443 - 560) = 10^{[0.906 * \mathrm{Log}_{10}(\Delta K_d) - 0.526]} \quad (2.23)$$

The model shows good accuracy over the whole range of $a_{cdom}(443)$ in DS1 (RMSD=0.11, MAPD=15.06 % and MB=0.01 m-1). The highest uncertainty is in the parameterization of X (Fig 2.4b).

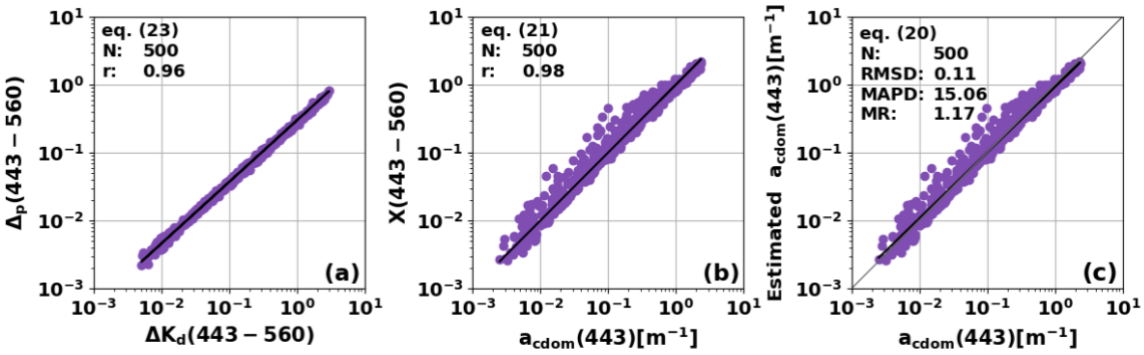


Fig 2.4 Different steps of the CDOM-KD2 inversion model parameterized from the IOCCG dataset: (a) Δ_p (443 - 560) as a function of ΔK_d (443 - 560) (Eq 2.23) (b) X as a function of ΔK_d (443 - 560) - Δ_p (443 - 560) (Eq 2.21), and (c) $a_{cdom}(443)$ as a function of X (Eq 2.20).

2.1.3.1.2 Model development in the context of satellite application

In order to avoid the cumulative impact of the relative errors associated with the K_d estimation performed at each individual wavelength considered in CDOM-KD1, ΔK_d was assessed directly from $R_{rs}(\lambda)$ using empirical formulations based on Hydrolight simulations (Loisel et al., 2014). In the present study, the calculation of ΔK_d is now performed through a Neural Network (NN) approach, following the same training data set and protocol of the NN originally developed in (Jamet et al., 2012) for estimating $K_d(\lambda)$ in the visible domain.

This NN consists in a Multi-Layer Perceptron model (MLP, (Rumelhart et al., 1986)) based on 7 possible input parameters including the R_{rs} at 412, 443, 490, 510, 560 and 670 nm and the sun angle, θ_s. In practice, $R_{rs}(412)$ was not used in the model definition due to the general high uncertainty level associated with the satellite R_{rs} signal at this spectral band (Goyens et al., 2013; Jamet et al., 2012; Mélin et al., 2007; Zibordi et al., 2006). The reflectance in the red part of the spectrum ($R_{rs}(670)$) was considered as an input of the model depending on the relative level of turbidity of the water. This selective definition was performed considering the impact of the low signal to noise ratio on the satellite R_{rs} data validity in the red spectral domain, this issue being particularly relevant when dealing with non-turbid water environments (e.g., Hu et al., 2012). Following the recent works by Loisel et al. (2018), a switch criterion was therefore used to differentiate non-turbid and turbid waters. In practice, data showing a $R_{rs}(490)/ R_{rs}(560)$ ratio lower or equal to 0.85, were considered as turbid. In this case the R_{rs} input for the NN were restricted to the range 443 – 670 nm, while the model considered has two hidden layers with five neurons for each layer. On the other hand, if the previous ratio was higher than 0.85, emphasizing the presence of non-turbid water, the input R_{rs} values were ranging from 443 to 560 nm and the NN has two hidden layers with four neurons for each layer.

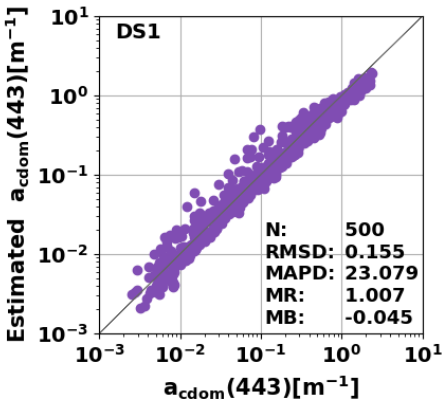


Fig 2.5 Performance of the CDOM-KD2 inversion model considering a NN based inversion for the calculation of the ΔK_d term (Eq 2.20) and using the DS1 data set.

The performance of the $a_{cdom}(443)$ inversion model (CDOM-KD2) based on ΔK_d estimates computed from the latter NN is presented in Fig 2.5 for the DS1 synthetic dataset considering a sun zenith angle of 30º. Globally, a relevant retrieval of $a_{cdom}(443)$ is obtained from this NN based model (MAPD=23%) although a higher scatter when comparing these results with those obtained when the model used the true K_d (MAPD=15%, Fig 2.4-c).

2.1.3.2 Intercomparison and validation of $a_{cdom}(443)$ inversion models

2.1.3.2.1 Performance and inter-comparison of the different $a_{cdom}(443)$ inversion models over the in situ and matchup data sets

The performance of the A2018, C2017, S2011 and CDOM-KD2 models have been first evaluated for global scale application (including coastal and open ocean waters) from the DS2 *in situ* validation data set (Fig 2.6). The models C2017, S2011 and CDOM-KD2, which are based on distinct formalisms and assumptions, show an overall general satisfying accuracy in the retrieved $a_{cdom}(443)$ over the 3 orders of magnitude covered in DS2 (e.g. MAPD of 49.22, 39.57 and 36.95%, respectively). This general feature is also underlined by the overall agreement between the distribution histograms shapes and statistics reported in the Fig 2.7 for the latter three models. The global performance of A2018 for estimating $a_{cdom}(443)$ significantly departs from the others. The range of $a_{cdom}(443)$ values retrieved from the global empirical model by Aurin et al. (2018) being much narrower than that for the other three models (Fig 2.7c) as a consequence of the large overestimation of A2018 derived $a_{cdom}(443)$ values for low and moderate a_{cdom} loads ($a_{cdom}(443) < 0.1$ m^{-1}, Fig 2.6b) as well as due to the sharp underestimation of the highest A2018-derived $a_{cdom}(443)$ values ($a_{cdom}(443) > 0.1$ m^{-1}, Fig 2.6b). A lower performance of the MLR model by Aurin et al. (2018) at 443 nm was already documented by the latter authors who emphasized the lower performance of this empirical model for estimating a_{cdom} at wavelengths > 412 nm especially when using SeaWIFS bands as input values.

The use of the optical typology provided by Mélin and Vantrepotte (2015) provides a finer characterization of the model performances. The class-based distribution further confirms the global relevance of C2017, S2011 and CDOM-KD2 derived $a_{cdom}(443)$ values with a general satisfying accuracy over the 17 waters types considered. A lower precision in the retrieved $a_{cdom}(443)$ is however observed for the three latter models in the most turbid waters (Class 1) as underlined by the higher scatter in the Figs 2.6 a, e and g for the corresponding samples. Further, a slight overestimation of the highest $a_{cdom}(443)$ values is found for the $a_{cdom}(443)$ values derived from CDOM-KD2 and C2018 for the samples associated with the Class 1. The difficulty to estimate CDOM in such highly turbid environments from general formulations requires the development of specific inversion models. While few formulations have been proposed for estimating $a_{cdom}(\lambda)$ in optically complex waters (e.g. Cao et al., 2018 ; Loisel et al., 2014) CDOM estimates in coastal environments are often derived from regional models (Cao et al., 2018; Mannino

et al., 2014; Matsuoka et al., 2013).

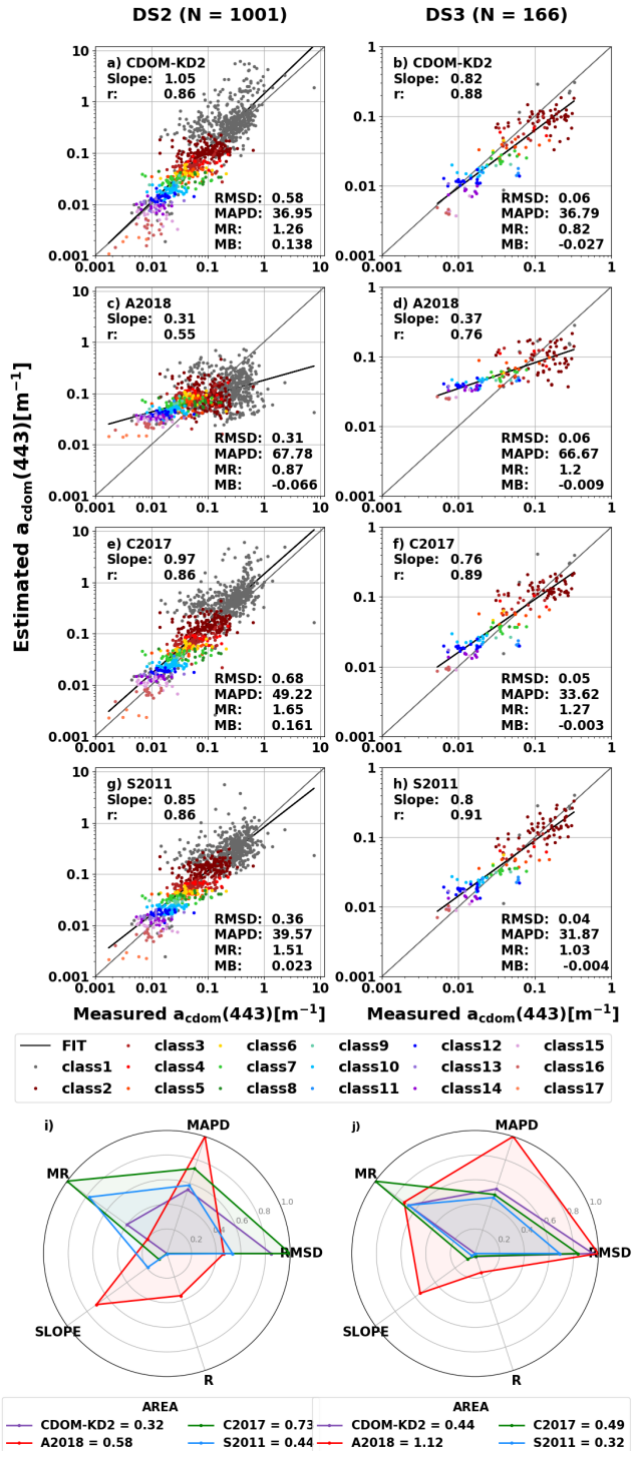


Fig 2.6 Validation scatter plots of the four tested models (CDOM-KD2, A2018, C2017 and S2011) using the DS2 (a, c, e, g) and DS3 (b, d, f, h) complete data sets (classes 1 to 17). Radar plots of the statistics used for evaluating the four models for DS2 (i) and DS3 (j).

The radar plot for the whole data set DS2, which provides a synthetic view of the accuracy of the different models considered for estimating $a_{cdom}(443)$, confirms the previous results further underlining the vicinity in the performance of the C2017, S2011 and CDOM-KD2

with a slightly better general performance for the CDOM-KD2 method. Results obtained considering the whole DS3 matchup data set are globally in line with those derived from the DS2 validation data with a general satisfying and comparable accuracy for C2017, S2011 and CDOM-KD2 (MAPD of 33.62, 31.87 and 36.79 %, Fig 2.6b, d, f, h and j, Fig 2.7) and lower general performance of the A2018 model (MAPD 66.67 %). The differences in the coverage provided by DS2 and DS3 induced slight modulations in the finer patterns. An underestimation of the highest $a_{cdom}(443)$ values in DS3 is for instance observed for the CDOM-KD2 model while the reversal situation was found for DS2 which however accounts for a higher amount of CDOM rich waters (> 0.5 m^{-1}).

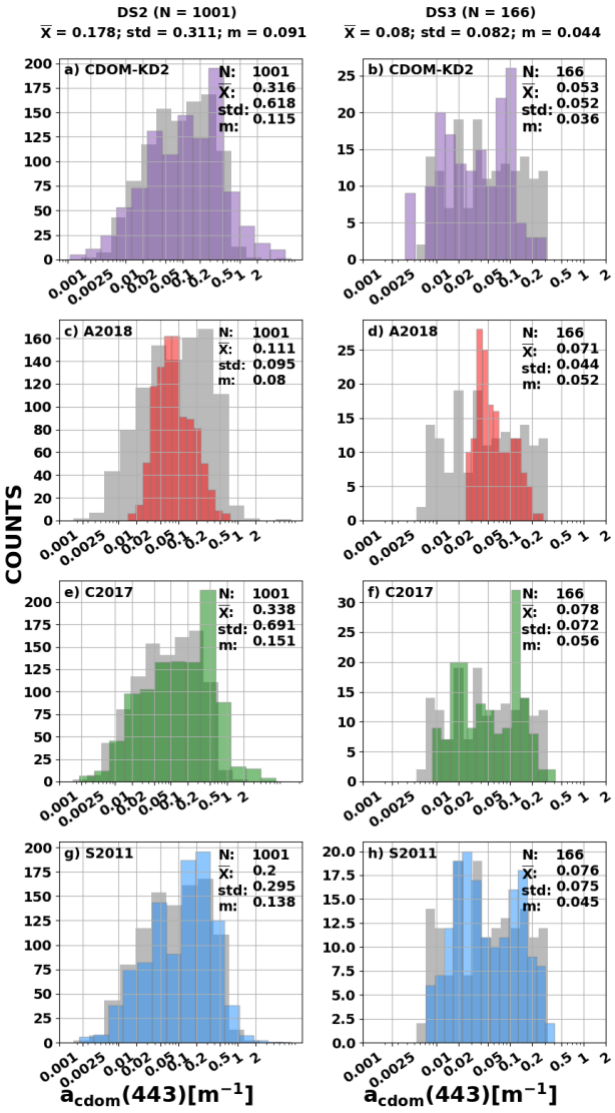


Fig 2.7 $a_{cdom}(443)$ absolute frequency distribution histograms for the DS2 (a, c, e, g) and DS3 (b, d, f, h) complete data sets for the in situ (grey) and modeled values by the four tested models (CDOM-KD2: purple, A2018 : red,, C2017; green and S2011; blue). N, $\overline{X}$, m, std correspond to the number of data points, mean, median and standard deviation respectively.

2.1.3.2.2 Comparison of the $a_{cdom}(443)$ models on moderate to non-turbid waters

A focus on the performance of the different models was performed on the moderate to non-turbid data to evaluate the applicability of the different methods considered for open ocean waters. For that purpose, the validation exercise is now performed excluding from DS2 and DS3 the ultra-turbid samples corresponding to the Mélin and Vantrepotte (2015) Class 1 and Class 2 waters. The new DS2 and DS3 data sets are now composed by 373 (instead of 1001) and 108 (instead of 166) data points, respectively. The main features described previously on the whole data sets are globally observed from these restricted datasets with an overall satisfying performance of the inversion of $a_{cdom}(443)$ values in oceanic waters. For instance, the MAPD (and slope) values for C2017, S2011 and CDOM-KD2 are 34.77 % (0.74), 28.11 % (0.69) and 27.42% (0.83) for DS2 and 33.92 % (0.68), 27.98 % (0.63) and 30.85 % (0.79) for DS3, respectively). As with previous results obtained on the whole data sets (Fig 2.6 and 8) the performances of these three models overcome that of the A2018 inversion algorithm (MAPD of 97% and 103.18 %, for DS2 and DS3, respectively).

The comparison of the overall statistics for the C2017, S2011 and CDOM-KD2 models further confirms the general consistency in the $a_{cdom}(443)$ retrieval from these three different approaches, which precision is generally increased when excluding the most turbid environments. Among the three latter formulations, based on the statistics, CDOM-KD2 model shows slightly better overall performance considering both *in situ* (Fig 2.8 i) and matchup (Fig 2.8 j) data sets, although a slight underestimation of the $a_{cdom}(443)$ in the ultra-oligotrophic waters associated with classes 16 and 17 can be observable (but should be confirmed according to the low number of data points). It is worth to notice that the S2011 model strongly depends on the blue to green reflectance ratio which is also used for estimating *Chl-a* concentration in offshore waters. The use of common inputs for assessing both CDOM and *Chl-a* might therefore tend to artificially strengthen the co-variation between these two variables making difficult the assessment of their specific dynamics using the latter CDOM inversion method. Moreover, the $a_{cdom}(443)/a_{cdm}(443)$ ratio values estimated at global scale using S2011 (for $a_{cdom}(443)$) and GSM (for $a_{cdm}(443)$) present numerous unrealistic values, the mean and standard variation values being of 1.0 ± 0.63 over the 10-years GlobColour climatology. The CDOM-KD2 model was therefore selected for describing the global scale spatio-temporal variability of

$a_{cdom}(443)$ and of the relative contribution of CDOM to the absorption of the whole detrital matter pool as depicted by the ratio $a_{cdom}(443)/a_{cdm}(443)$.

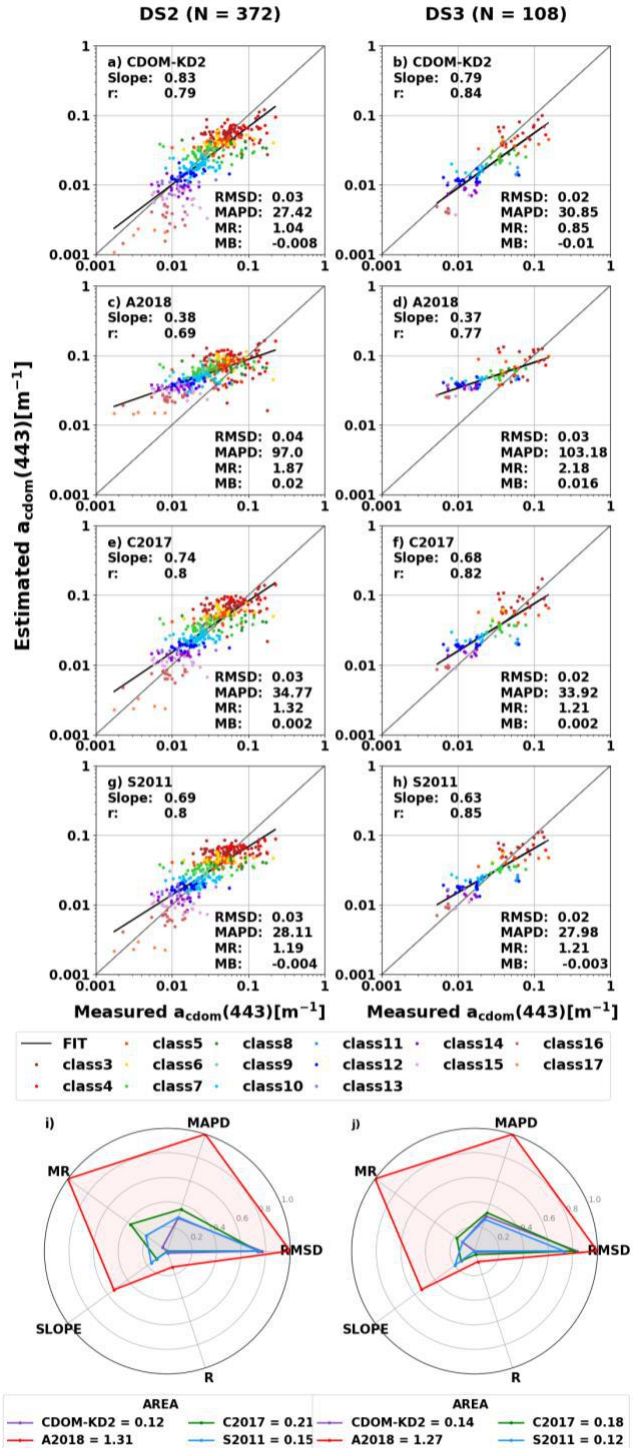


Fig 2.8 Validation scatter plots of the four tested models (CDOM-KD2, A2018, C2017 and S2011) over DS2 (a, c, e, g) and DS3 (b, d, f, h) non-turbid subset (classes 2 to 17). Radar plots of the statistics used for evaluating the four models in DS2 (i) and DS3 (j).

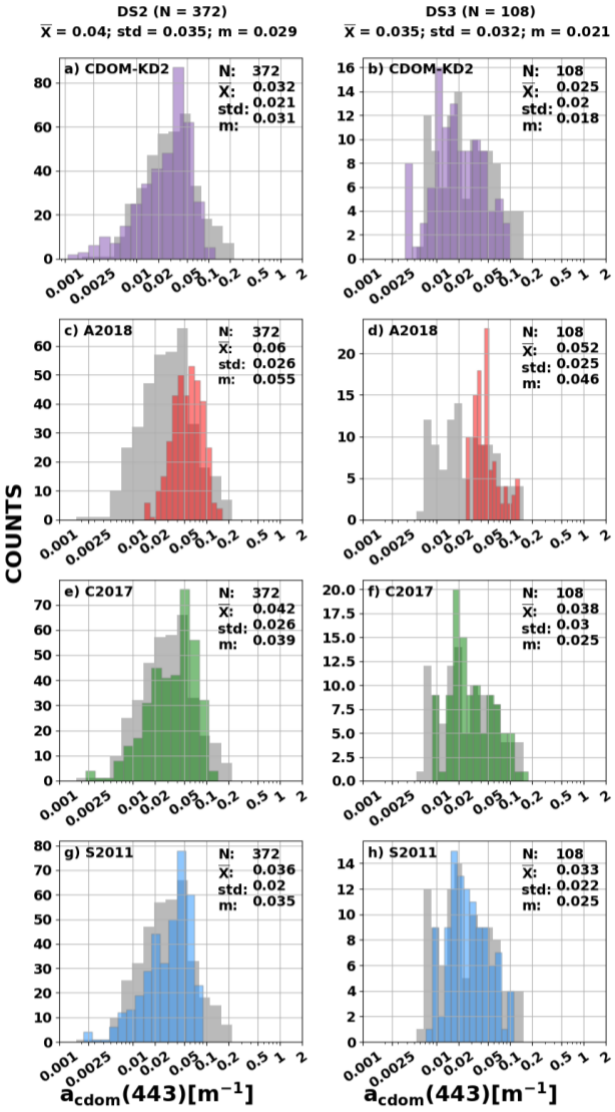


Fig 2.9 $a_{cdom}(443)$ absolute frequency distribution histograms over DS2 (a, c, e, g) and DS3 (b, d, f, h) non-turbid subsets (classes 2 to 17) for the in situ (grey) and modeled values by the four tested models (CDOM-KD2: purple, A2018: red, C2017: green and S2011: blue). N, $\overline{X}$, m, std correspond to the number of datapoints, mean, median and standard deviation respectively.

2.1.3.2.3 Global $a_{cdom}(443)$ spatio-temporal patterns

The global scale spatial distribution of CDOM-KD2 derived $a_{cdom}(443)$ is depicted on Fig 2.10a from the overall GlobColour L3 10-year archive average map (April 2002-April 2012). $a_{cdom}(443)$ shows a high spatial dynamic with values ranging over 3 orders of magnitude ($a_{cdom}(443) < 0.001$ m^{-1} to > 2 m^{-1}). $a_{cdom}(443)$-CDOM-KD2 general global spatial patterns agree with the distribution patterns expected and previously described by other authors. High CDOM values are found in coastal waters and in the sub-polar and equatorial areas, while low values are located throughout the subtropics (Siegel et al.,

2005a; Nelson and Siegel, 2013). Lowest values are found in the oligotrophic gyre areas such as the South Pacific Gyre, where estimated values are in line with in situ observations performed in the area (Bricaud et al., 2010, minimum $a_{cdom}(440) \approx 0.001$ m^{-1}). In agreement with Bricaud et al. (2012) for $a_{cdm}(443)$, an evident asymmetry is observed between the northern and the southern hemispheres along the year, the northern hemisphere oceanic waters being richer in CDOM.

The $a_{cdom}(443)$ coefficient of variation (CV, %) map computed from the GlobColour merged archive (Fig 2.10c) illustrates the high spatial heterogeneity in the temporal dynamics of $a_{cdom}(443)$ at global scale. Very stable areas (CV < 10%) are located mainly in the oceanic gyres, in the waters located within [40º-60º] North and South latitudinal layers and in the northern Indian ocean. The latter areas coincide with the poorest regions of the ocean (gyres) as well as with oceanic regions located between the main oceanic currents. On the other hand, areas showing the highest temporal dynamics (CV > 60%) are those strongly influenced by main oceanic currents, upwelling areas and regions strongly impacted by terrestrial inputs of DOM, such as the oceanic area impacted by the Amazon plume during the retroflection of the North Brazilian Current (Salisbury et al., 2011).

The impact of the main oceanic circulation patterns on $a_{cdom}(443)$ temporal dynamics in the open ocean is particularly visible within water masses surrounding oceanic gyres as well as within three latitudinal bands located around 0º, 30º N and 30º S where $a_{cdom}(443)$ CV ranges from 40% to 70% over the 10-year GlobColour time period. Strong temporal dynamics area are also clearly visible along the Antarctic Polar Frontal Zone (APFZ) (CV from 50 to 70%) characterized by the presence of a marked seasonality in the water masses characteristics due to the occurrence of a strong phytoplankton spring bloom (Abbott et al., 2000; Tremblay et al., 2002).

The highest $a_{cdom}(443)$ temporal dynamics (CV > 80%) is observed mostly in areas influenced by terrestrial inputs of dissolved matter. Such high temporal variations are found for instance within the oceanic water influenced by the Amazon - Orinoco systems where the observed strong modulation in the surface CDOM loads are related to the combined effect of the highly variable regional currents system and the seasonal dynamics of the terrestrial inputs associated with these two large river systems (López et al., 2012; Salisbury et al., 2011). The high $a_{cdom}(443)$ temporal variability found in the

western Africa and Arabian sea waters can be more likely related to the influence of desert dusts on the temporal coverage and radiometric quality of the OCR observations.

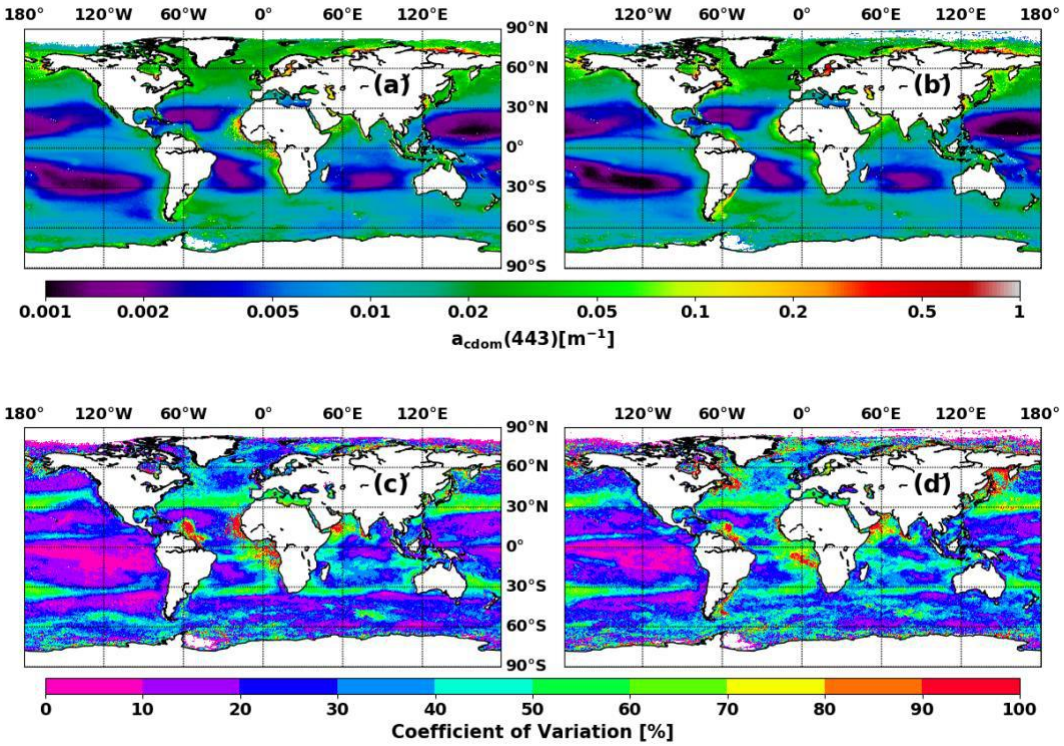


Fig 2.10 Global average $a_{cdom}(443)$ [m^{-1}] map produced with GlobColour L3 merged 25km 8 days composite data from 23rd April 2002 to 13th April 2012 and CDOM-KD2 model (a) with its coefficient of variation (%) (c) and the global $a_{cdom}(443)$ [m^{-1}] average maps produced from the 4 years archive of GlobColour OLCI L3 25km 8 days composite data from 22nd April 2016 to 16th June 2020 (b) with its coefficient of variation (%) (d).

The OLCI L3 4-year $a_{cdom}(443)$ average and variation coefficient maps are presented in Fig 2.10b and d, respectively. The general patterns observed for $a_{cdom}(443)$ spatial distribution and dynamics from OLCI data are in agreement with those previously depicted from the GlobColour merged archive. OLCI L3 4-year average $a_{cdom}(443)$ however reaches more extreme end-member values than the GlobColour 10-year average merged data (lower values in the ultra-oligotrophic and higher ones in eutrophic waters, respectively, Fig 2.10b).

2.1.3.2.4 Global scale covariation between a_{cdom}, a_{cdm} and Chl-a dynamics

Over oceanic waters the temporal dynamics of CDM is assumed to be primarily driven by phytoplankton and associated by products variability (Bricaud et al., 2012; Siegel et

al., 2005b). In open ocean, CDOM is the dominant component of CDM (Kopelevich and Burenkov, 1977; Siegel et al., 2002) being therefore also expected to be highly related to CDM and *Chl-a* dynamics excepted in some specific environments, such as upwelling regions, where phytoplankton, particulate and dissolved matter dynamics are assumed to be decoupled (Siegel et al., 2002).

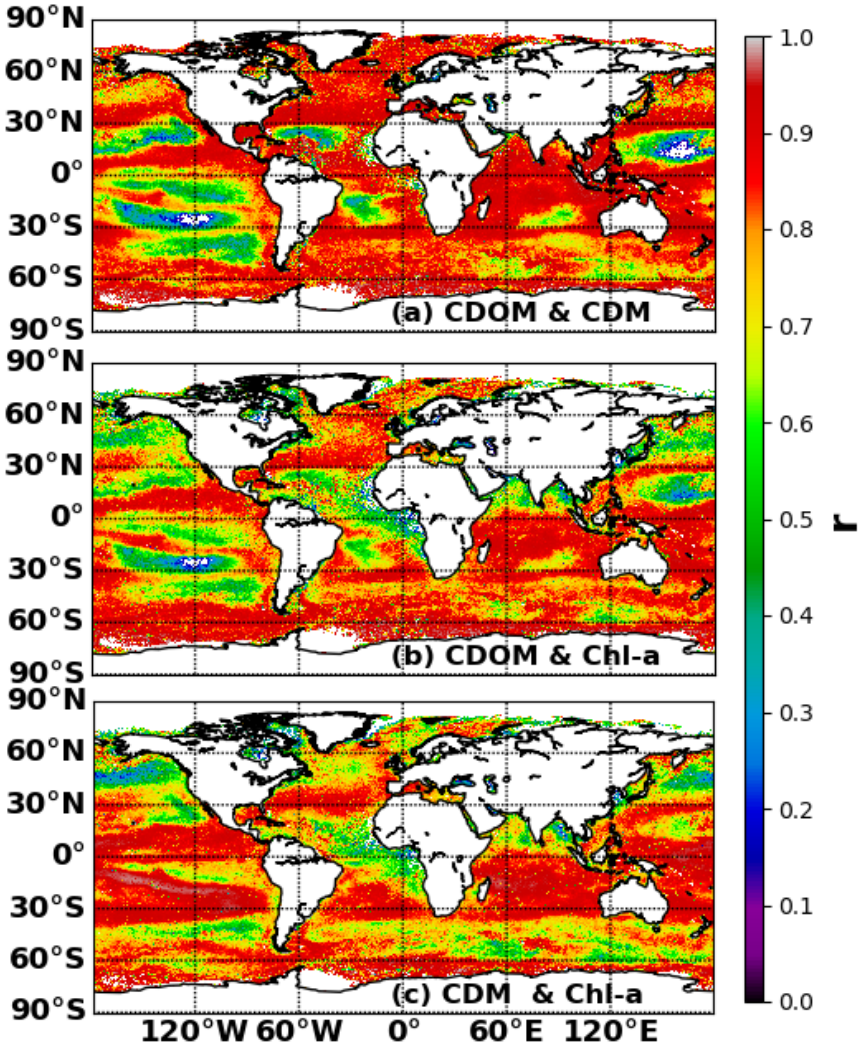


Fig 2.11 Correlation maps between (a) CDOM-KD2 and $a_{cdm}(443)$, (b) CDOM-KD2 and *Chl-a*, (c) a_{cdm} and *Chl-a*. White areas represent pixels where either there is no data available or where the correlation is not significant ($p>0.05$).

The covariation of CDOM (CDOM-KD2 derived), CDM and *Chl-a* (GSM estimates) was here examined through correlation maps based on the 10-year Globcolour 8 days composite archive (Fig 2.11). A general very high positive correlation between CDOM and CDM prevails over a large part of the global ocean ($r > 0.9$). However, lower correlation level ($r <0.6$) are observed over large areas. This is specifically the case of all

oceanic gyre regions where the correlation between $a_{cdom}(443)$ and $a_{cdm}(443)$ decrease towards the gyre center reaching values close to zero in the North and South Pacific gyres. Relatively low CDOM to CDM correlation levels (r [0.5-0.6]) are also observed in the Pacific sub-arctic and sub-Antarctic gyre waters, Pacific equatorial divergence area and patchy regions of the Antarctic waters.

Globally, oceanic areas where CDOM and CDM exhibit a lower correlation, show contrasted patterns with respect to the correlation patterns observed for the two latter parameters with *Chl-a* (Fig 2.11b and c). The divergence in the relationships between the temporal patterns of CDOM and CDM with respect to phytoplankton dynamics is particularly marked over gyre areas. In these very oligotrophic environment, the CDOM-*Chl-a* correlation levels are generally lower than those observed for CDM-*Chl-a* correlation, suggesting differences in the processes driving dissolved and particulate detrital matter in the corresponding areas. This apparent heterogeneity in the temporal dynamics of CDOM with respect to that of particulate detrital matter and phytoplankton can be related to the processes driving CDOM dynamics.

This feature is illustrated (Fig 2.12) with the $a_{cdom}(443)$, $a_{cdm}(443)$, and *Chl-a* time series extractions over two contrasted oceanic areas, that is the South Pacific Gyre (SPG) and North Atlantic (NA).

In the very clear waters of the South Pacific Gyre (Fig 2.12b) the CDOM time series exhibits a higher level of noise without real seasonal pattern. In this areas, CDOM degradation and production is strongly driven by photo-degradation processes (Chen and Bada, 1992; Siegel et al., 2005b) but also by bacteria activity which generates great fluctuations in periods of days (Nelson et al., 2004) in these nutrient depleted areas (Raimbault et al., 2008) which present an efficient microbial loop. A rapid degradation of the CDOM produced from phytoplankton and associated by-products might explain the absence of seasonality observed for CDOM in this area. The latter result is in line with in situ observations by Bricaud et al. (2010) along the BIOSOPE transect. They reported that small-scale changes in the phytoplankton biomass in the most oligotrophic waters of the SPG (*Chl-a* < 0.1 mg m^{-3}) do not induce significant variation in the CDOM content (their Fig 2.14), in contrast to the non-algal particles (a_{nap}) which shows a high correlation with *Chl-a* (their Fig 2.10).

In contrast to the SPG station, a strong co-variation between *Chl-a*, $a_{cdom}(443)$, and $a_{cdm}(443)$ characterizes the times series of the NA station (Fig 2.12c), with the clear presence of a spring maximum for all of the parameters. The phytoplankton bloom in this area (Dutkiewicz et al., 2001; Lévy et al., 2005) is therefore the main driver of both the particulate and dissolved detrital matter dynamics with no apparent lag in the CDOM and CDM (on 8 days composite data) dynamics and thus in the impact of corresponding source and sink controlling processes. At this area, CDOM is locally produced by phytoplankton excretions and lysis (Nelson and Siegel, 2002).

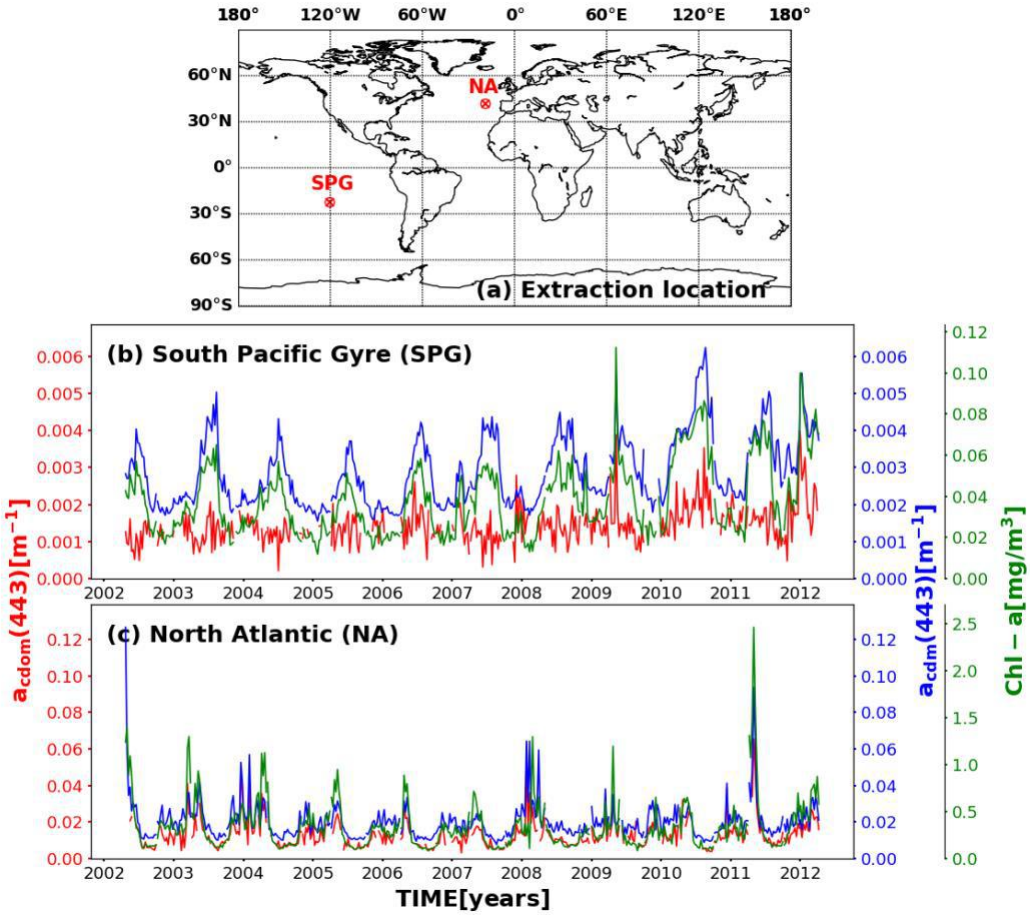


Fig 2.12 (a) Location of the stations considered for the two time series plotted in panels (b) and (c) (red circles). At these two stations the correlation between a_{cdom} and both *Chl-a* and a_{cdm} is minimum (South Pacific Gyre : SPG) and maximum (North Atlantic, NA). Time series of $a_{cdom}(443)$, $a_{cdm}(443)$ and *Chl-a* at SPG (b) and NA (c).

2.1.3.2.5 Global $a_{cdom}(443)$/ $a_{cdm}(443)$ ratio spatio-temporal patterns

The previously presented results have demonstrated the potential of the CDOM-KD2 algorithm for accurately estimating $a_{cdom}(443)$ at global scale including the most

oligotrophic waters. The $a_{cdom}(443)$ derived maps is now used for assessing the spatio-temporal variability of the relative importance of the dissolved matter in the total detrital matter absorption through the $a_{cdom}(443)/a_{cdm}(443)$ ratio.

Numerous works have considered the $a_{cdom}(443)/a_{cdm}(443)$ ratio as spatially and temporally quasi invariant or constant assuming CDOM to be the major contributor (> 80%) to CDM in the blue spectra domain (Nelson et al., 1998, Swan et al., 2009).

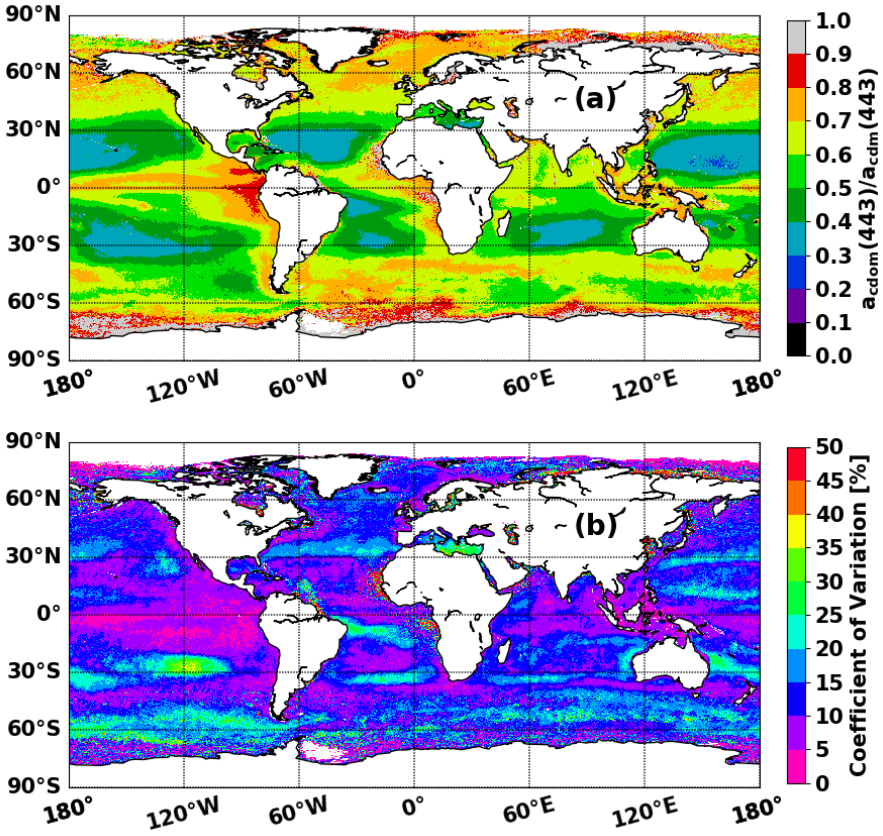


Fig 2.13 Global distribution of the average $a_{cdom}(443)/a_{cdm}(443)$ ratio (a) and variation coefficient (%) (b) for the GlobColour L3 merged 25km, 8 days composite data from 23rd April 2002 to 13th April 2012.

A very high spatial dynamics is observed for the $a_{cdom}(443)/a_{cdm}(443)$ ratio distribution at global scale with values ranging from about 0.2 to almost 1 (Fig 2.13a). The overall global average $a_{cdom}(443)/a_{cdm}(443)$ ratio for the GlobColour 10-year archive reaches 0.61 ± 0.14 but with a wide variability on spatial scale (standard deviation 0.14). The lowest ratio values (0.2 to 0.5) are located in the oligotrophic gyre waters. The presence of low $a_{cdom}(443)/a_{cdm}(443)$ values in the oligotrophic ecosystems is in line with results in Fig 2.12b where the restricted contribution of CDOM to CDM is also clearly illustrated, further confirming the potential high rate of degradation of the dissolved material due to

active photochemical or biological processes in gyre waters. Besides the expected presence of high $a_{cdom}(443)/a_{cdm}(443)$ values over some coastal environments (e.g. Baltic and Black seas), the relative contribution of CDOM to CDM generally tends to increase in oceanic waters from mid to high latitudes (> 30°) reaching maximum values (around 0.9) in polar regions. High $a_{cdom}(443)/a_{cdm}(443)$ values are also found in the equatorial Pacific (>0.7) and Atlantic (>0.6) waters.

The global temporal variability in the $a_{cdom}(443)/a_{cdm}(443)$ is illustrated from the variation coefficient map computed over the 10-year GlobColour archive (Fig 2.13b). The overall average $a_{cdom}(443)/a_{cdm}(443)$ variation coefficient is 15 % (with a standard deviation of 15%) and ranged from 2% in the Equatorial Pacific to 50% in Arctic waters emphasizing sharp spatial disparities in the global distribution for this parameter. A larger temporal variability in $a_{cdom}(443)/a_{cdm}(443)$ ratio value is for instance found in oceanic areas surrounding the main oceanic gyres (CV ranging 15 to 25%) when compared to the central gyre areas where an overall higher temporal stability is found (CV < 10 %). This general pattern is however not noticed in the SPG which shows the highest temporal dynamics in $a_{cdom}(443)/a_{cdm}(443)$ with CV > 25% in the most oligotrophic waters. Similar values are found in the eastern Mediterranean Sea which original character in terms of optical properties has been pointed out by several authors (Claustre et al., 2002; Loisel et al., 2011; Morel and Gentili, 2009). Finally, an overall high temporal dynamic of the relative importance of CDOM to CDM is also observed in the Antarctic circumpolar current waters with CV values oscillating around 30%.

While explaining the apparent decoupling between a_{cdom} and a_{nap} at global scale is beyond the scope of the present work, our results do not support the widespread assumption of an overall global dominant and temporally slightly variable contribution of CDOM to CDM. This pattern further emphasizes the need to further investigate the dynamics and environmental factors controlling the dissolved and particulate components of the ocean detrital matter pool.

2.1.3.2.6 Global $a_{cdom}(443)/a_{nw}(443)$ ratio spatio-temporal patterns

The global scale contribution of $a_{cdom}(443)$ to $a_{nw}(443)$ ranges from 0.1 to 0.9 with an average value of 0.42 (±0.29) considering the 10 year GlobColour data set (Fig 2.14). Spatial patterns for the ratio $a_{cdom}(443)$ to $a_{tot}(443)$ (not shown) are very similar to those

reported for $a_{cdom}(443)$ to $a_{nw}(443)$ the global average for this parameter being of 0.35 (±0.26).

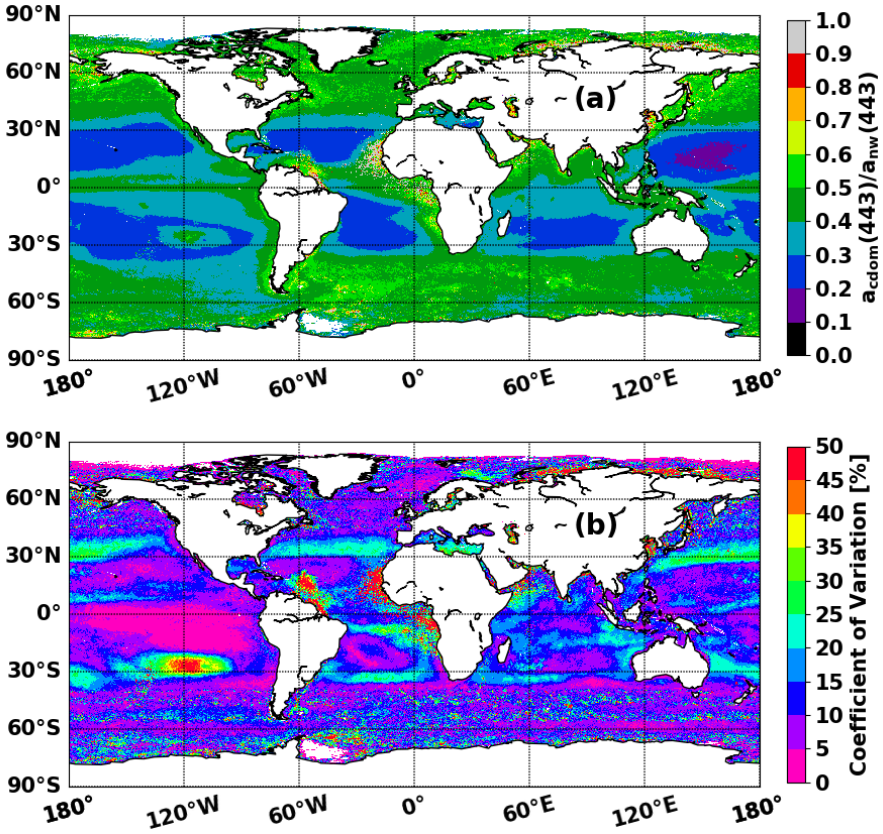


Fig 2.14. Global distribution of the average $a_{cdom}(443)/a_{nw}(443)$ ratio (a) and variation coefficient (b) for the GlobColour L3 merged 25km, 8-day composite data from 23rd April 2002 to 13th April 2012.

It is worth noting that Siegel et al. (2005a) reported a global average $a_{cdm}(440)/a_{nw}(440)$ ratio value of 0.46. The $a_{cdom}(443)/a_{nw}(443)$ ratio shows high spatial variability at global scale. Lowest $a_{cdom}(443)/a_{nw}(443)$ are found in gyre waters with an $a_{cdom}(443)/a_{nw}(443)$ average ratio of 0.31 (±0.13). In these oligotrophic waters $a_{cdom}(443)/a_{nw}(443)$ is globally ranging from 0.2 to 0.4 with slightly lower minima in the North Pacific Gyre (0.1) and higher maxima (reaching locally up to 0.5-0.6) in the eastern part of the SPG. In the SPG the mean $a_{cdom}(443)/ a_{nw}(443)$ observed here (0.32 ± 0.15) is in line with the one reported by Bricaud et al. (2010) from the BIOSOPE study (0.45) based on *in situ* measurements from the surface to the deep chlorophyll maximum. A larger range of variation [0.2-0.6] is however found in this work when compared with the BIOSOPE results ([0.3-0.5] over the whole area). In equatorial and high latitudes a higher contribution of $a_{cdom}(443)$ to $a_{nw}(443)$ is globally found with values ranging between 0.3 to 0.7 this maximum being reached in terrestrial influenced water masses. This result agrees with former works which have also reported a CDOM contribution around 0.7 for the latter environments

(e.g. Bélanger et al., 2006; Matsuoka et al., 2007, 2009, 2013, in Arctic waters). In the Mediterranean Sea, CDOM contributes for 0.2 to 0.6 (average of 0.42 ± 12) of the non-water absorption at 443 nm with minimum values located in the eastern Mediterranean waters. At the BOUSSOLE site in the NW Mediterranean Sea, the mean and standard deviation values of the $a_{cdom}(443)/a_{nw}(443)$ ratio are 0.46 ± 0.10 in good agreement with those reported by Organelli et al. (2014), who observed a yearly mean $a_{cdom}(443)/a_{nw}(443)$ ratio values around 0.5 over the first attenuation layer.

The coefficient of variation map underlines that temporal variations in the $a_{cdom}(443)$ to $a_{nw}(443)$ ratio are generally relatively small (<15%), regions showing most variable $a_{cdom}(443)$ to $a_{nw}(443)$ (> 30%) ratio corresponding to the areas where the $a_{cdom}(443)/a_{cdm}(443)$ ratio shows the highest dynamics (e.g. latitudinal bands around 30°N and 30°S, central SPG waters, and areas impacted by terrestrial matters).

2.1.4 Conclusions

A new model for assessing the $a_{cdom}(443)$ global distribution from OCR (CDOM-KD2) has been developed adapting an existing semi-analytical formalism (Loisel et al., 2014) based on the use of the vertical attenuation coefficient of the downwelling irradiance, K_d. Among the four models evaluated: Aurin et al. (2018), Shanmugam (2011), Chen et al. (2017) and CDOM-KD2, the last three methods, although based on different assumptions, show consistent performances at estimating surface $a_{cdom}(443)$ values at global scale. The CDOM-KD2 inversion model here proposed performs slightly better when considering both the in situ (DS2) and matchup (DS3) data sets used in the frame of this study, especially over open ocean waters. These results clearly underline the actual possibility to specifically estimate $a_{cdom}(443)$ at global scale and to overcome limitations related to the use of $a_{cdm}(443)$ especially for open ocean dedicated studies related to the DOC dynamics analysis.

The CDOM-KD2 model was applied to global satellite archives of merged (GlobColour) or individual recent satellite (OLCI) to characterize the $a_{cdom}(443)$ spatio-temporal patterns of variability as well as that of the contribution of CDOM to CDM and of CDOM to the non-water absorption. While the $a_{cdom}(443)$ as well as CDOM relative contribution in both CDM and total absorption spatial variability are particularly marked between terrestrial influenced water masses and oceanic gyres end-members, a relative restricted

temporal variability (10 year CV <50%) is in contrast generally observed in most of the oceanic domains.

Globally, in oceanic gyres, where CDOM loads are the lowest ($a_{cdom}(443)$ <0.002 m^{-1}), CDOM is not dominant in the total detrital matter absorption budget (<40%) representing also a reduced fraction of the total water absorption (<30%), these general features being slightly variable in time (CV < 10%). In these oceanic regions, correlation analysis reveals that CDOM dynamics is generally slightly coupled with that of CDM and *Chl-a* which both conversely show a strong co-variation. This tends to indicate that phytoplankton dynamics is the main driver of the particulate detrital matter variability in gyre systems whereas dissolved organic matter dynamics cannot be considered as a direct function of phytoplankton and phytoplankton by-products. This further underlines that other forcing parameters such as microbial and light dependent processes act as the main controlling factors explaining CDOM dynamics in these gyre systems. An exception to the previous general patterns is however observed in the most oligotrophic waters of the eastern SPG where a highest temporal variability (CV> 35%) is found for the $a_{cdom}(443)/a_{cdm}(443)$ and $a_{cdom}(443)/a_{nw}(443)$ ratios when compared to the other gyre waters. Further, the contribution of CDOM to the total absorption in the latter area is also higher than that for the other gyre waters (>40%) suggesting the presence of a higher decoupling between particulate and dissolved matter dynamics for that region.

In contrast with gyres areas, polar and oceanic waters influenced by large river inputs globally show the highest values and a high temporal variability for $a_{cdom}(443)$, $a_{cdom}(443)/a_{cdm}(443)$ and $a_{cdom}(443)/a_{nw}(443)$ ratios. In the corresponding regions CDOM represents 60% or more of CDM while a general high coupling in the dynamics of the dissolved and particulate detrital matter prevails. The later components do not necessarily covary with phytoplankton dynamics especially in areas significantly impacted by terrestrial inputs. Subtropical (around 30°N and S) and Equatorial regions show and intermediate situation with an overall moderate level of temporal variability for $a_{cdom}(443)$, $a_{cdom}(443)/ a_{cdm}(443)$ and $a_{cdom}(443)/a_{nw}(443)$ ratios.

This apparent heterogeneity in the CDOM, CDM and *Chl-a* dynamics, and thus in the factors controlling both dissolved and particulate matter variability in the global ocean should be further investigated. Further, the high variability observed in the relative contribution of CDOM to the total absorption might be considered in future works for

more precisely quantifying the impact of CDOM on *Chl-a* estimates over oceanic waters from OCR.

2.2 Temporal variability of $a_{cdom}(443)$ over the global ocean

2.2.1 Census X-11 and trend analysis

To complement the results presented on section 2.1 where general $a_{cdom}(443)$ global spatio-temporal patterns are provided, an additional analysis was performed applying the Census X-11 procedure (Shiskin et al., 1967) to the monthly GlobColour L3 merged time series from 1997 to 2012, the longest time series of CDOM-KD2 available, and from 2002 to 2012, the same period used in section 2.1.

The Census X-11 method is a standard tool in economics (Findley et al., 1998) that has been adapted for a variety of applications. Census X-11 method was first adapted by Pezzulli et al. (2005) for the decomposition of sea surface temperature (SST) time series at global scale. The X-11 algorithm assumes that any time series X(t) of specific periodicity *p* can be decomposed into three terms: the irregular component, I(t), the seasonal component, S(t), and the trend cycle component, T(t):

$$X(t) = I(t) + S(t) + T(t) \quad (2.24)$$

The main result of Census X-11 algorithm is an improved estimation of both seasonal and trend terms which are computed alternatively, allowing a proper separation of these two signals (Vantrepotte and Mélin, 2009, 2011). Here the seasonal term is determined locally in time, ensuring that the annual cycle of each year is not biased by uncorrelated events occurring at other times. Simultaneously, this allows inter-annual variations in the shape of the annual cycle contrarily to a climatological description of seasonality (i.e., a fixed annual cycle, Vantrepotte and Mélin, 2009, 2011). Consequently, the X-11 trend-cycle component is able to reproduce more properly long-term evolution in the mean level of the variable under study, and it has been shown to be particularly well adapted to model climate variations (Pezzulli et al., 2005).

The outputs obtained by applying the Census X-11 method are:

1) Time series of each component: S(t), I(t), and T(t)
2) Maps of the relative contribution of each component (in %) to the total variance of the original time series, t

In addition to the latter analyses, the presence of monotonic trend in the $a_{cdom}(443)$ time series was evaluated computing seasonal Mann-Kendall statistics used to compute the Rate of Change (RC, in %/year) and the p-value for each time series

The latter methods were first applied to the Globcolour monthly data of $a_{cdom}(443)$ computed using the CDOM-KD2 model

2.2.2 $a_{cdom}(443)$ temporal schemes of variability

The maps provided in the Fig 2.15 show the distribution of the relative contribution of each X-11 term to the total variance of $a_{cdom}(443)$ from 2002-2012. While the variation coefficient maps provided in the Fig 2.16 were describing the amplitude of the temporal variability of each pixel, these maps provided deeper information on the origin of the observed variability.

Seasonal oscillation is explaining the vast majority (>80%) of the CDOM variation in the two latitudinal layers around 30°N and S where strong CDOM variation were observed (Fig 2.15 a). Such strong importance of the seasonality is also found in the southern Atlantic waters, in the Arabian Sea, in patchy areas of the Southern Ocean and northern Atlantic as well and in water masses influenced by terrestrial inputs. Note that similar global patterns have been depicted for *Chl-a* (Vantrepotte and Mélin, 2011). As a matter of fact, areas where S(t) is dominating for both CDOM and *Chl-a* are showing a high correlation coefficient between these two parameters (Fig 2.11). In other words, CDOM and *Chl-a* strongly covary when the seasonal variation is explaining the major part of the temporal variation underlining the tight link between phytoplankton blooms dynamics and CDOM behavior in these areas.

Conversely, areas where a lower covariation was found between *Chl-a* and CDOM are corresponding to water masses where the seasonality shows a lower contribution to the variance of CDOM. This is specifically the case in the gyre areas (except the South Atlantic gyre) where S(t) contribution is representing less than 50 % of the total variance of $a_{cdom}(443)$ (Fig 2.15 a) and where instead strong contribution of the irregular ([10-40%]; Fig 2.13.15 c) and trend-cycle ([40-80%] ; Fig 2.13.15 b) terms are observed. This further underlines that the apparent relative uncoupling between *Chl-a* and CDOM in the ultra-oligotrophic waters (Fig 2.11) occurs in areas where the amplitude of the temporal

variation is generally low (Fig 2.13.15) and mainly driven by long term oscillation and/or sub-annual processes with a less marked seasonality.

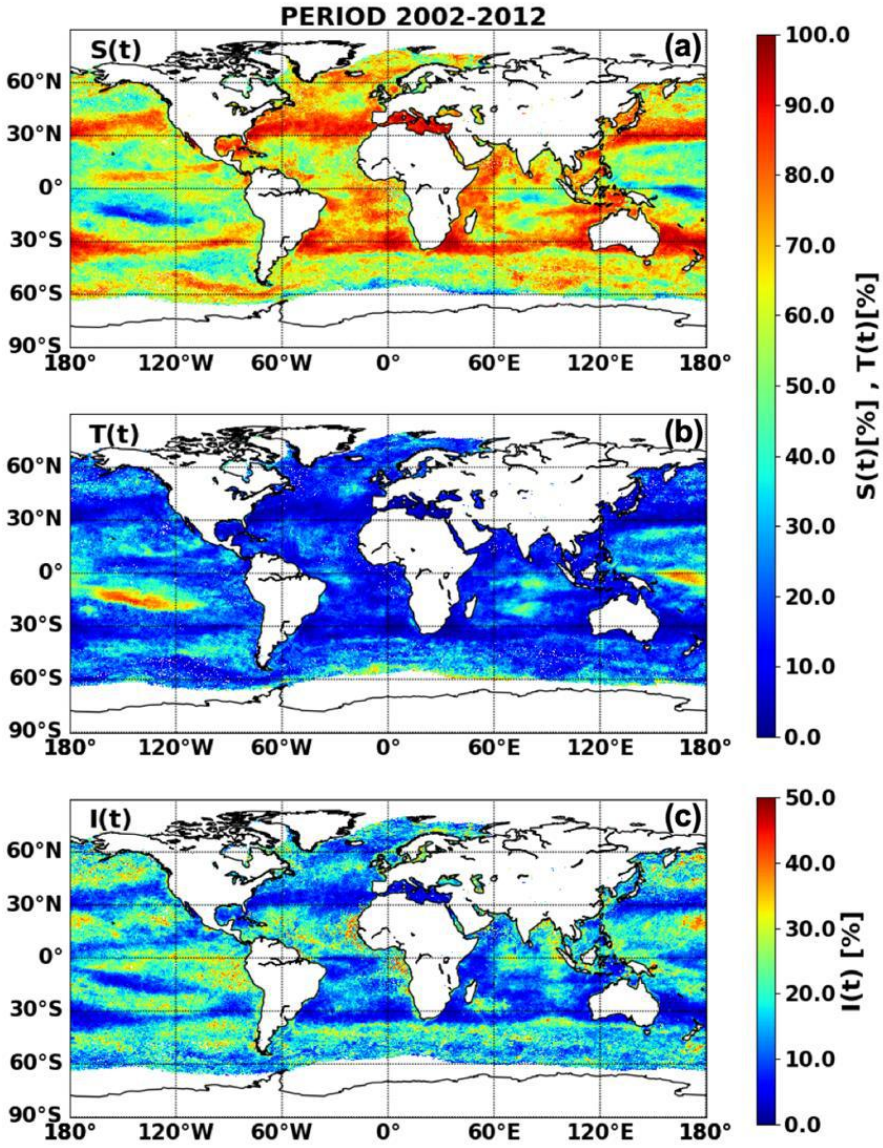


Fig 2.15 Relative contribution of the (a) seasonal (S(t)), (b) trend cycle (T(t)) and (c) irregular (I(t)) Census X-11 components to the total variance of $a_{cdom}(443)$ signal over the time period 2002-2012.

2.2.3 Interannual changes in $a_{cdom}(443)$ values

Two trend analyses were performed in order to evaluate the presence of significant changes in CDOM loads at global scale. One analysis was performed from 2002 to 2012 in order to consider the same time period than the one previously considered in this Chapter and another one extending the time window from 1997 to 2012 and asses longer term CDOM evolution. Note that an extended analyses on the 1997-2020 using the GlobColour data set time was not possible due the absence of R_{rs} data at 510 nm between the time period April-2012 to April-2016.

Interestingly, very different patterns are observed in the trends calculated during the 2002-2012 and 1997-2012 time periods considering the SeaWiFS data. When considering 1997- to 2012, negative trends of small amplitude (< 3%/year in absolute value) are observed over the vast part of the global ocean including Gyre areas whereas reduced increase in CDOM are observed over patchy region of the eastern and western south America, southern Atlantic, western US coast.

A totally different situation exists when looking to the 2002-2012 time period for which a reversal pattern is found especially in the Pacific Gyre regions with a sharp increase (>5%/year) in the CDOM level. It is worth noticing that the increase in CDOM observed in these ultraoligotrophic regions of the Pacific are corresponding to regions where a strong decrease in *Chl-a* was pointed out by several authors over different time periods (Vantrepotte and Mélin, 2011; Polovina et al., 2008) who argued the presence of a sharper desertification of these oceanic deserts. The situation observed here for CDOM is therefore strongly contrasting with these previous works.

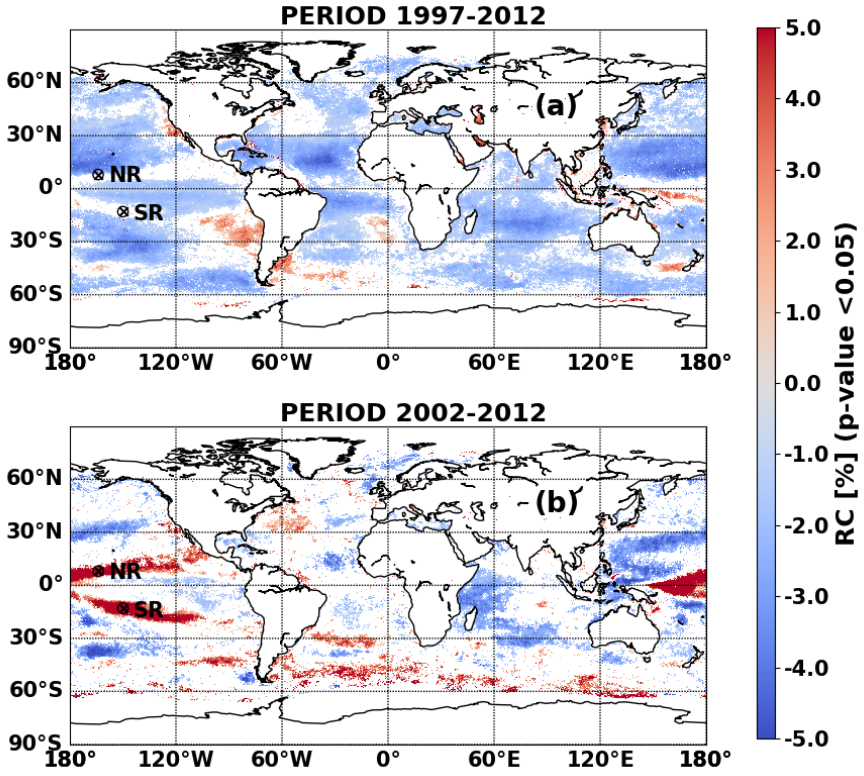


Fig 2.16 Global rate of change (RC; %/year) for $a_{cdom}(443)$ monthly time series (CDOM-KD2 algorithm and R_{rs} from GlobColour L3 merged data) between 1997 to 2012 (a) and 2002 to 2012 (b) calculated with CDOM-KD2 algorithm and R_{rs} from GlobColour L3 merged data. Black circles show the locations chosen for a time series extraction.

The Fig 2.17 allow a more detailed view of the actual shape of interannual evolution in CDOM in the northern and southern oligotrophic regions (8ºN, 164ºW and 13ºS, 150ºW, respectively).

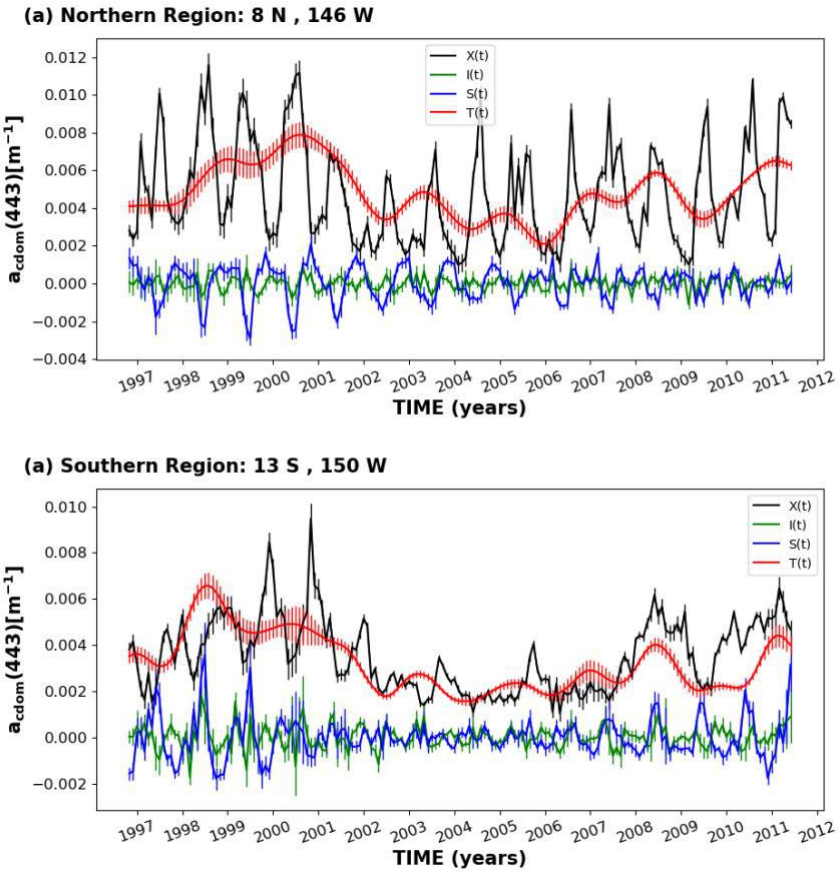


Fig 2.17 Time series extraction of $a_{cdom}(443)$ (X(t)), and corresponding X11 I(t), S(t) and T(t) component in two regions of the Pacific ocean showing positive trend in CDOM over the period 2001-2012, located in the northern, (a), the southern (b) Pacific oligotrophic gyres waters (see location in Fig 2.16).

These extractions emphasize that the use of monotonic trend detection tend to mask the interannual changes in CDOM which are actually occurring from 1997 to 2012. For both regions a "bump" in the series is observed from 1997 to 2003 with first an increase in CDOM from 1997 to 1998, a conservation of high CDOM values 1998 to 2001 and then a decrease from 2001 to 2003. CDOM values are then (2003-2006) lower than those observed at the beginning of the series (1997), eventually CDOM is increasing again from 2007 to 2012. This final increase is explaining the results obtained on the maps for the 2002-2012 time period.

Note that the temporal patterns observed for CDOM over the SeaWiFs time period (1997-2007) are in line with those reported for *Chl-a* for the same temporal window (Vantrepotte

and Mélin, 2011). The latter authors have clearly related the interannual changes in the *Chl-a* levels for this period over the Pacific gyre areas to climate indices.

These first results that should be further performed on longer time series underline the need to have a deeper look into the shape of the interannual changes considering its non-linear characteristics.

2.3 Conclusion and perspectives

In this chapter a new model to estimate the absorption of CDOM in open ocean from $R_{rs}(\lambda)$ (CDOM-KD2) is presented. This is validated and compared to other previously published (Aurin et al., 2018; Chen et al., 2017; Shanmugam, 2011). Results showed slightly better performance of CDOM-KD2 at estimating surface $a_{cdom}(443)$ particularly in open ocean.

The CDOM-KD2 model presents the great advantage of the possibility to estimate $a_{cdom}(443)$ at global scale independently of $a_{cdm}(443)$. Therefore, CDOM-KD2 allows the analysis of the variability of the contribution of CDOM absorption to CDM and to non-water absorptions. The results of this exercise shows that while the spatial variability of the contribution of CDOM absorption to CDM and to non-water absorptions is very marked in the global oceans, the patterns temporal variability are relatively smooth. It is observed that only a few regions of the global ocean present relatively high temporal variability in the ratio $a_{cdom}(443)/a_{cdm}(443)$ and $a_{cdom}(443)/a_{nw}(443)$, along with high correlation between CDOM and CDM and *Chl-a*. In this areas, more specifically the polar regions, the oceanic gyres end-members and oceanic waters influenced by large river inputs, the greatest part of the temporal variability of CDOM and *Chl-a* is due to seasonal variation. All of this implies that in the later regions phytoplankton is the main driver of CDOM dynamics.

On the other hand, in the gyres, where $a_{cdom}(443)/a_{cdm}(443)$ presents the lowest values, correlation analysis reveals that CDOM dynamics is poorly coupled with that of CDM and *Chl-a* which both conversely show a strong co-variation. Meaning that in this oligotrophic regions phytoplankton is not the main controller of CDOM dynamics, while it is he main driver of the particulate detrital matter variability. Hence, other forcing parameters like microbial activity and light dependent processes are controlling CDOM's concentration. This pattern is observed in all gyres, except for the eastern SPG in which

the decoupling between particulate and dissolved matter dynamics is very strong and the contribution of CDOM to the total absorption is higher than the observed for the other gyre waters. In the latter area the temporal variability found for the $a_{cdom}(443)/a_{cdm}(443)$ and $a_{cdom}(443)/a_{nw}(443)$ ratios is also much higher than in the other gyre waters.

The analysis of the temporal variability shows that the mentioned uncoupling between *Chl-a* and CDOM in ultra-oligotrophic waters, especially marked in the SPG but not so much in the South Atlantic gyre, occurs in areas where the amplitude of the temporal variation is low and mainly driven by long term oscillation and/or sub-annual processes with a less marked seasonality.

The different patterns found for the interannual rate of change of CDOM when two different time periods are used (1997-2012 and 2002-2012), demonstrate the need to make a deeper analysis into the shape of the interannual changes of the analyzed variables taking into account non-linear variations which are not represented by classical monotonic trend analysis.

In the future, the analysis of the temporal variability of CDOM and *Chl-a* should be extended to a longer time period (from 1997 to 2020) to better evaluate the impact of CDOM over *Chl-a* and determined the bias that this generates on the estimates of *Chl-a* in the open ocean from OCR. This should be paired with the analysis of environmental data to evaluate the physical conditions leading to changes in the phytoplankton community, such as variation in PAR that would directly affect the primary production, changes in SST which indicate mixing prosses, upwelling and other changes in the water masses, or the occurrence of ENSO events leading to strong environmental changes with a global affect.

3. Chapter 3: Estimation of Dissolved Organic Carbon in global scale from satellite data

In this chapter, a new approach to estimate DOC over open ocean water based on an Artificial Neural Network (ANN) algorithm combining various satellite products is presented. The estimation of DOC from ocean color radiometry is a very challenging task considering that 1) CDOM is the only optical parameter able to trace DOM from space and 2) the relationship between CDOM and DOC is highly variable in open ocean waters due to different CDOM and DOC kinetics (Aurin et al., 2018). It is therefore necessary to take into account additional information allowing, for instance, the water masses history to be considered. For that reason, the development of the ANN was performed testing different input parameters and considering different time lags for these input variables. In other terms, the rationale of the ANN was developed taking not only into consideration the water masses situation at the defined moment of the estimation, but also including the processes that have led to the DOC concentration observed.

The development of the DOC algorithm using in situ data only was not feasible due to the limited in situ DOC and ancillary in situ coincident variables (such as *Chl-a*, $a_{cdom}(\lambda)$, SST, SSS, MLD). This limitation is even higher when time lags between DOC and the ancillary variables will be considered in the DOC model development. Considering the latter feature, the development data set was built by matching the DOC in situ data with a list of potential input parameters for the models and associated time lags.

3.1 In situ and satellite data sets used for the development and validations

3.1.1 The whole DOC and ancillary variables data

The *in situ* DOC dataset gathers worldwide distributed data (Fig 3.1) from different missions and databases completing a total of 4343 DOC data points covering the time period 1991-2015. The used databases include:

- ANTARES (Lefevre et al., 2016)
- Bermuda Atlantic Time-Series study (BATS) site (Hansell and Carlson, 2001)
- GLobal Ocean Data Analysis Project (GLODAP) database (Key et al., 2004)
- Global Ocean Carbon Algorithm Database (GOCAD) (Aurin et al., 2018)

- Gulf of Mexico NACP-OCB Coastal Synthesis (GoMX - NACP-OCB) (Osburn et al., 2011)
- Hawaii Ocean Time-series (HOTS) (Karl and Lukas, 1996; Church et al., 2013)
- K2S1 (Honda et al., 2017)
- RV Polarstern cruise ARKTIS-XXVII/2 (Schauer, 2008)
- TRANSDRIFT (Juhls et al., 2019)

For the ANN training data set in a context of remote sensing application, surface samples (depth > 50 m) were only considered (68% of the data points). Further, DOC samples with a concentration lower than < 44 µmol/L (0.98% of the superficial data points) were also excluded since they represent the refractory DOC having a residence time of thousands of years (section 1.2.3, Hansell et al., 2009). Hence, this background DOC signal does not present any variability in the time period of this study being a potential source of bias for the model.

The remaining DOC data points were matched with the different satellite and Argo floats (for MLD) data (further detailed) considered as relevant for the estimation of DOC concentration (see section 1.3). This resulting data set (DS1) gathers a total of 2895 DOC in situ measurements from 1991 to 2015 keeping a worldwide distribution (Fig 3.1).

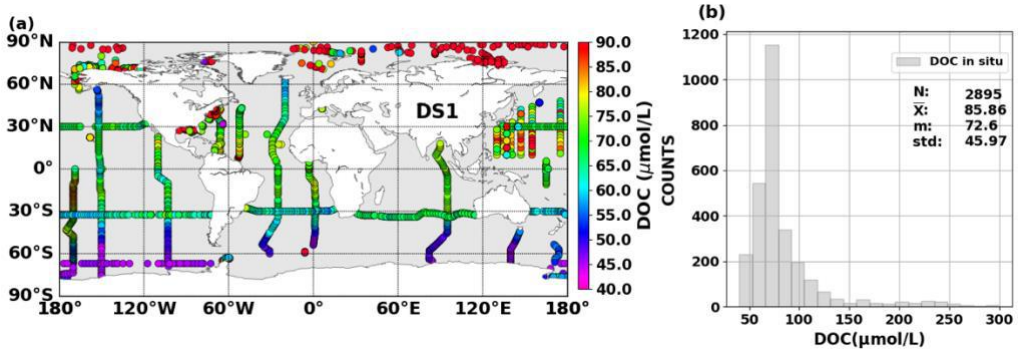


Fig 3.1 In situ DOC measurements from DS1. N, $\overline{X}$, m and std correspond to the number of data points, mean, median and standard deviation, respectively.

Starting from this DOC data set, other ancillary variables were further added. This includes:

- Remote sensing reflectance (R_{rs}) from GlobColour L3 merged 8-day composites data with a spatial resolution of 4 km^2.
- Chlorophyll-*a* (*Chl-a*) concentration GlobColour estimates computed applying the ocean chlorophyll 4-band (OC4) algorithm (O'Reilly et al., 1998) which relates R_{rs}

band ratios to *Chl-a* through a single polynomial function. In practice, OC4 considers as input the maximum band ratio (MBR) determined as the greater value among the $R_{rs}(443)/R_{rs}(555)$, $R_{rs}(490)/R_{rs}(555)$ and $R_{rs}(510)/R_{rs}(555)$ ratios.

- Absorption of colored dissolved organic matter at 443 nm $a_{cdom}(443)$ also calculated from GlobColour R_{rs} following the methodology described in Chapter 1 (CDOM-KD2 model).
- Photosynthetically active radiation (PAR) from GlobColour L3 merged 8-day composites data with a spatial resolution of 4 km. The missing values of PAR were fulfilled with a monthly climatology from European Union Open Data Portal (ODP) GMIS - SeaWiFS with a spatial resolution of 9 km (https://data.europa.eu/euodp/en/data/dataset/03de7eee-495f-480e-b028-03f84f947b19).
- Mixed layer depth (MLD) from Argo JAMSTEC data base with of 10-day average data from January 2001 to present and a spatial resolution of 1 degree in the global ocean calculated from Argo floats temperature and salinity profiles. The missing values of MLD were fulfilled with a monthly MLD climatology (described in Holte et al., 2017) generated from Argo profiles with an hybrid method (Holte and Talley, 2009) and a spatial resolution of 1 degree from the average over the entire Argo record.
- Sea surface temperature (SST) from NOAA Optimum Interpolation (OI) Sea Surface Temperature (SST) V2 weekly product (Reynolds and Smith, 1994) with data collected between 1991 and 2020 and a spatial resolution of 1 degree. The missing values of SST were filled interpolating monthly data between 2002 and 2015 from ISAS-15 (In Situ Analysis System; Gaillard et al., 2016; Kolodziejczyk et al., 2017) with a spatial resolution of 0.5 degree.
- Sea surface salinity (SSS) from the ESA Sea Surface Salinity Climate Change Initiative (Boutin et al., 2019) version 1.8 one week running mean product with data between 2010 and 2017 and a spatial resolution of 25 km (evaluation product). The missing values of SSS were filled by interpolating monthly data between 2002 and 2015 from ISAS-15 (In Situ Analysis System; Kolodziejczyk et al., 2017, spatial resolution of 0.5 degree.

The 8 days match-ups between DOC in situ and GlobColour data were performed using a 3x3 pixel window centered on the position of the DOC samples, in which the coefficient of variation of $R_{rs}(\lambda)$ needs to be below 0.15 while the number of valid pixels needs to be above 50% (implying a minimum of 5 valid pixels). For SSS the mean of a 3x3 pixel

window was also used. For SST and MLD that have a spatial resolution of 1º the direct pixel containing the in situ DOC measurement location was considered.

The later production of global maps with the new ANN is made with a spatial resolution of 25 km. If the input variable used does not have this resolution it is reshaped to accomplish this requirement. For this an empty grid of the desired size was generated and the pixels were filled with the data located in the closest pixel of the original matrix.

In all cases, when the gaps were filled by climatological data, the followed protocol was the same as for the respective non-climatological data of the products.

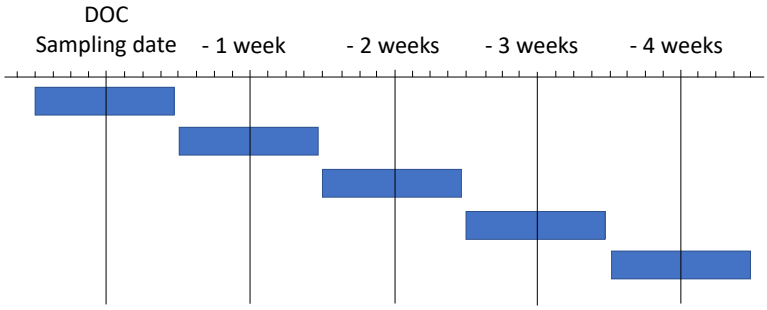


Fig 3.2 Representation of the time lags methodology used for the match-up of in situ DOC with other variables.

In practice, the global dataset was built between the in situ DOC data and the products enumerated above at different time lags. More precisely, the ancillary variables were considered at the same week of the in situ DOC measurement and also at 1, 2, 3 and 4 weeks before (Fig 3.2). In the case of MLD, where the data is a 10 days mean, the closest mean to the corresponding week was used, the same protocol was followed for the monthly and climatological data. The use of time lags in the training data set intends to consider not only the state of the water at the moment of the in situ measurements, but also the processes that lead to it.

As the origin of the DOC coincident (and lagged) products is from different sensors, the coverage and spatial and temporal resolution are not the same, thus the amount of matchups for each of them varies.

Two factors reduced the amount of data available for the development of the ANN, the lack of satellite data in the polar regions, and the fact that between 1991 and 1997 there were no operating ocean color satellite missions (Fig 3.3). These two latter features have

therefore tended to reduce the number of DOC/input variables coupled data points needed for the training and validation of the new algorithm.

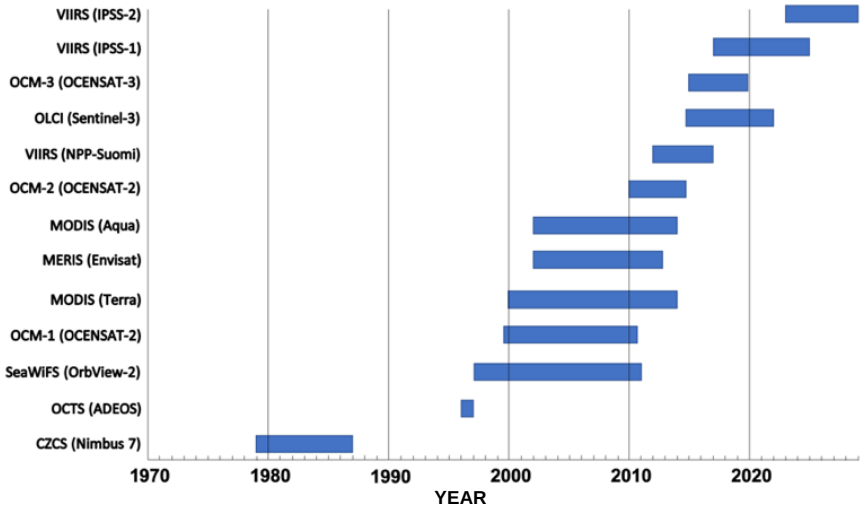


Fig 3.3 Timeline illustrating past, current, and future global ocean-color satellite missions adapted from Blondeau-Patissier et al. (2014).

3.1.2 The different sub data sets used for development and validation

The development of the ANN follows several steps starting with the linear correlation and multi linear regression performed to select the potential input data, followed by the iterative training and validation of different ANN configurations. Thus, three development data sets were used, DD-MLR (Fig 3.4), DD-NN (Fig 3.5) and DD-NNCHL (Fig 3.6) (Table 4.1).

DD-MLR was used for the calculation of the linear regressions and multi linear regressions of the products listed in section 3.1.1 in order to proceed to the first step of the model that consists in choosing the best descriptive variables.

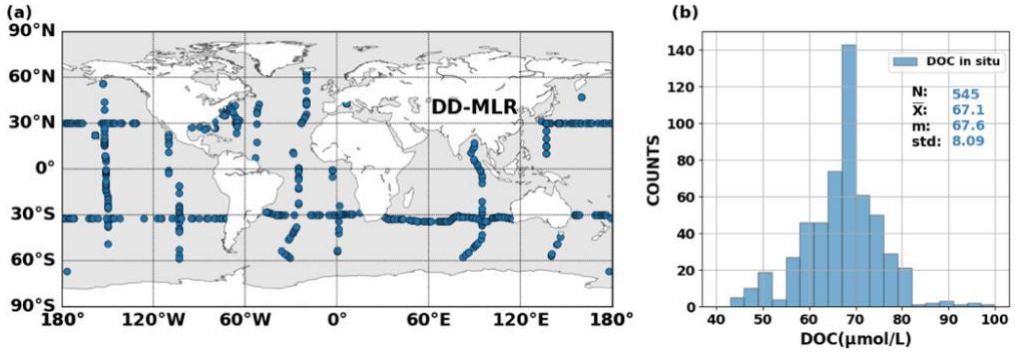


Fig 3.4 Location of DD-MLR data points (a) and the corresponding histogram of the in situ DOC gathered (b). N, $\overline{X}$, m and std correspond to the number of data points, mean, median and standard deviation, respectively.

For training the ANN two different development data sets were used DD-NN and DD-NNCHL (Table 4.1), which were randomly divided into a training (70%) and a validation (30%) subsets which are defined within the ANN development process. These validation subsets are used by the ANN to test the results of the training.

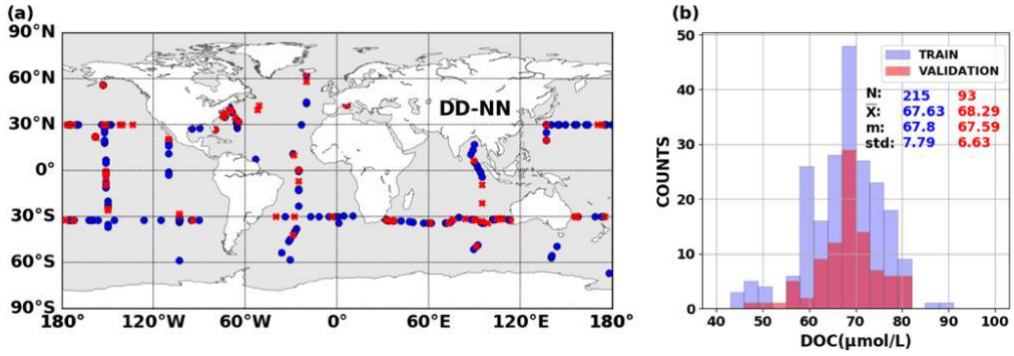


Fig 3.5 Global distribution of DD-NN data points (a) showing the training subset in blue and the validation subset in red, and the histogram of the in situ DOC gathered for each subset (b). N, $\overline{X}$, m and std correspond to the number of data points, mean, median and standard deviation, respectively.

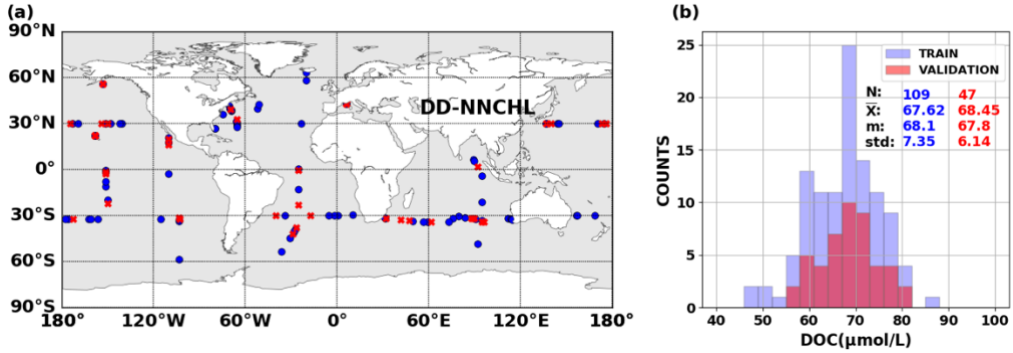


Fig 3.6 Global distribution of DD-NNCHL data points (a) showing the training subset in blue and the validation subset in red, and the histogram of the in situ DOC gathered for each subset (b). N, $\overline{X}$, m and std correspond to the number of data points, mean, median and standard deviation, respectively.

Finally, for validation purposes, two validation datasets were built: DV1 (Fig 3.7) and DV2 (Fig 3.8). They both gather the extracted satellite climatological DOC data for the points where in situ DOC is available. DV1 was used for testing three satellite models developed to esimated DOC: Siegel et al. (2002), Aurin et al. (2018) and the new ANN presented later in this chapter. While DV2 was used to make a deeper evaluation of the new model perfomance.

The validation datasets were built with monthly climatology of the estimated DOC to have a better global coverage and increase the amount of matching points with in situ DOC.

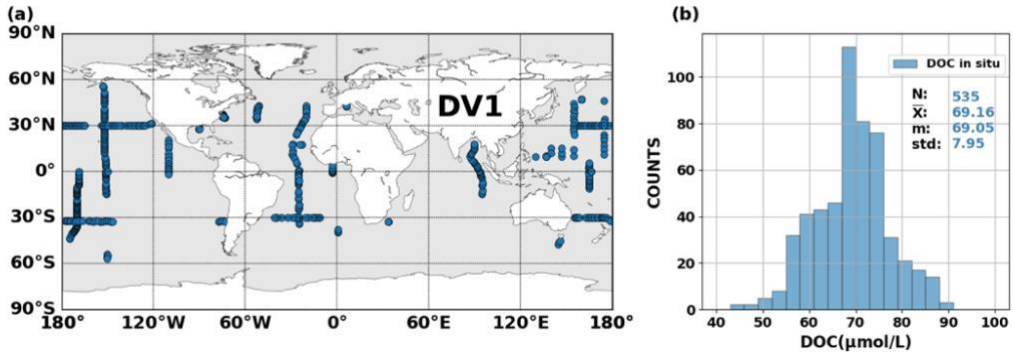


Fig 3.7 Global distribution of DV1 data points (a) and the corresponding histogram of the in situ DOC (b). N, $\overline{X}$, m and std correspond to the number of data points, mean, median and standard deviation, respectively.

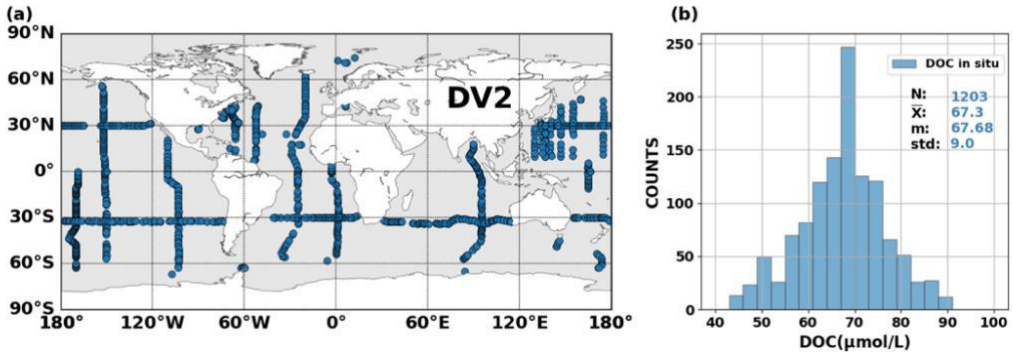


Fig 3.8 Global distribution of DV2 data points (a) and the corresponding histogram of the in situ DOC (b). N, $\overline{X}$, m and std correspond to the number of data points, mean, median and standard deviation, respectively.

Table 3.1 Summary of the data sets for the development and validation of the different algorithms (MLR, NN29b, NN29bChl, NN29s, etc).

Dataset	Variables included	N	DOC range	Mean	Median	std
DS1						
The whole dataset	in situ DOC [1991 - 2015]	2895	[44 – 425.2]	85.86	72.6	45.97
	Satellite SST, SST,PAR, *Chl-a*, $a_{cdom}(443)$ and R_{rs} (412, 443, 490, 510, 560, 670) and ARGO floats MLD at lag 0, -1,-2, -3, and -4 weeks respect to in situ DOC sampling date.					

DD-MLR						
Development data set used for linear correlation and for the MLR algorithms.	In situ DOC at the minimum depth (<50 m) [2002 - 2012] matching with satellite SST, SST,PAR, *Chl-a*, $a_{cdom}(443)$ and R_{rs} (412, 443, 490, 510, 560, 670) and ARGO floats MLD at lag 0 respect to in situ DOC sampling date.	545	[44.1 – 98.2]	67.1	67.6	8.08
DD-NN						
Development data set used to build NN29 and NN29b.	In situ DOC at the minimum depth (<50 m) [2002 - 2012] matching with DOC estimated with NN29 and NN29b from: ➢ Satellite SST (-1 week) ➢ $a_{cdom}(443)$ (-2 weeks) ➢ ARGO floats MLD (-1 week)	308	[44.84 - 89.8]	67.83	67.74	7.46
DD-NNCHL						
Development data set used to build NN29bCHL.	DOC in situ at the minimum depth (<50 m) [2002 - 2012] matching with DOC estimated with NN29b and NN29bCHL from weekly data of: ➢ Satellite SST (-1 week) ➢ *Chl-a* (-1 week) ➢ $a_{cdom}(443)$ (-2 weeks) ➢ ARGO floats MLD (-1 week)	156	[47.23 - 86.9]	67.87	67.86	7.02
DV1						
Climatological data set used for the validation and compaison of Siegel et al. (2002), Aurin et al. (2018) and NN29s (presented here.	Monthly climatology of in situ DOC at the minimum depth (<50 m) [1996 - 2009] and extracted monthly climatology of DOC (2002-2012) at the same location estimated with: ➢ NN29s from weekly data of SST (-1 week), MLD(-1 week), *Chl-a*(-1 week) and $a_{cdom}(443)$ (-2 weeks). ➢ Siegel et al. (2002) from the monthly climatology of SST. ➢ Aurin et al. (2018) from the monthly climatology of $R_{rs}(\lambda)$ and SSS.	535	[44 - 89.4]	69.31	69	8.15
DV2						
Climatological data set used for the monthly climatological validation of estimated DOC calculated with NN29s.	Monthly climatology of in situ DOC at the minimum depth (<50 m) [1994 - 2014] and extracted monthly climatology of DOC, at the same location, estimated with NN29s from weekly data from 2002 to 2012 of: ➢ Satellite SST (-1 week) ➢ *Chl-a* (-1 week) ➢ $a_{cdom}(443)$ (-2 weeks) ➢ ARGO floats MLD (-1 week)	1203	[44.0 - 89.95]	67.22	67.59	9.26

3.1.3 Roshan and DeVries (2017) global annual mean

The annual mean calculated with DOC derived from the final ANN will be compared with the annual picture generated by Roshan and DeVries (2017).

Roshan and DeVries (2017) built a neural network to reconstruct a global annual average DOC distribution at different depths. The in situ input data for this ANN include nitrate, phosphate, apparent oxygen utilization, silicate, dissolved oxygen, salinity, temperature, potential density anomaly, depth, bottom depth, depth of the euphotic zone, and chlorophyll concentration, each provided on a 1° × 1° grid along with the DOC observations between January 1995 and May 2014. The data was randomly split into 70% for training and 30% for validation.

The ANN structure for this model was able to reproduce validation data sets with acceptable correlation metrics and with no sign of overfitting. It consisted of one hidden layer with 10 to 20 neurons fully connected, in a feed-forward architecture, to a single-node output layer. They used a sigmoid activation function for the hidden layer, and a linear activation function for the output layer. Levenberg–Marquardt and Bayesian regularization methods were used for back propagation.

3.2 Existing satellite models for open ocean waters

In the present chapter two existing empirical models based on the use of satellite data to estimate DOC concentration were also evaluated. One published by Aurin et al. (2018) who used $a_{cdom}(355)$ and SSS as inputs values and one by Siegel et al. (2002) which is based on a relationship between DOC and SST.

3.2.1 Aurin et al. (2018)

Aurin et al. (2018) estimated DOC from satellite derived $a_{cdom}(355)$ (using the model described in section 2.2.3.1) and SSS (from Aquarius mission records) through a multi-linear regression approach, described as follows:

$$DOC = \beta 0 + \beta 1 * a_{cdom}(355) + \beta 2 * SSS \quad (3.1)$$

where β0 to β2 are the regression coefficients for estimating DOC (here β0=192.718; β1=26.790; β2=-3.555) and $a_{cdom}(355)$ is computed as follows:

$$ln(a_{cdom}(\lambda)) = \begin{bmatrix} \beta 0 + \beta 1 * ln\left(\mathrm{R}_{rs}(\lambda_1)\right) + \beta 2 * ln\left(\mathrm{R}_{rs}(\lambda_2)\right) + \\ \beta 3 * ln\left(\mathrm{R}_{rs}(\lambda_3)\right) + \beta 4 * ln\left(\mathrm{R}_{rs}(\lambda_4)\right) \end{bmatrix} \quad (3.2)$$

where λ_1 to λ_4 are the sensor-specific wavelengths (443, 490, 510, and 555 nm) and β0 to β4 are the regression coefficients (here β0 = −4.199, β1 = −2.563, β2 = 1.214, β3 = 0.955 and β4 = −0.040).

3.2.2 Siegel et al. (2002)

Siegel et al. (2002) constructed a DOC climatology using in situ observations collected from large-scale hydrographic transect cruises. These data were correlated with climatological winter sea surface temperature (SST) from NOAA database (National Oceanic and Atmospheric Administration, 1998). DOC is calculated in this model using an individual linear regression relationship for each ocean oceanic basin presented below:

Atlantic ocean

$$DOC = \begin{cases} 3.493 * SST - 9.79, & DOC < 85 \\ 85, & DOC \geq 85 \end{cases} \quad (3.3)$$

Indian Ocean

$$DOC = 0.795 * SST + 48.58 \quad (3.4)$$

Pacific and Southern Ocean

$$DOC = 0.993 * SST + 52.05 \quad (3.5)$$

3.3 Development of the algorithm

The development of the artificial neural network (here referred as NN29s) to estimate DOC follows a two steps process in which the most relevant input variables from the list mentioned in section 3.1.1 has first to be selected (ex. SST, SSS, etc.), while in a second step, the best ANN structure using the chosen input data has to be defined.

3.3.1 Input variable selection

The development data set DD-MLR (Fig 3.4; Table 4.1) was used to select the most valuable input variables for estimating DOC. In order to do so, individual type II linear

correlations between in situ DOC and SST, SSS, PAR, MLD, $a_{cdom}(443)$, *Chl-a* and R_{rs} (412, 443, 490, 510, 560, 670) at time lag 0 were first performed. This first test showed that in general no good correlation exists between DOC and any of the tested products when these describing variables were considered individually (Fig 3.9). As a matter of fact, the highest DOC correlation observed was with SST with R^2 of 0.292 only (Fig 3.9 a).

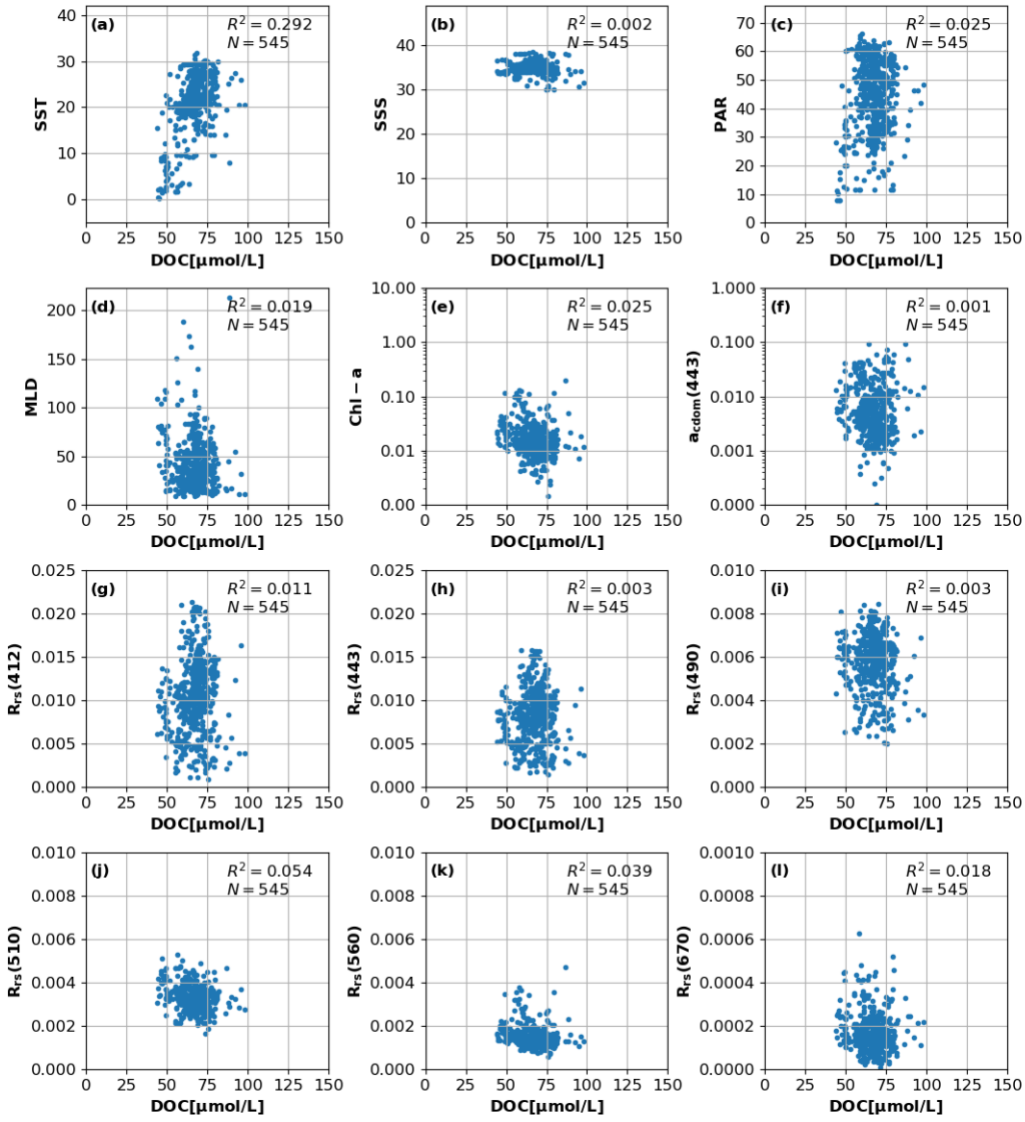


Fig 3.9 Direct correlation between the in situ DOC data from DD-MLR and the corresponding SST, SSS, PAR, MLD, $a_{cdom}(443)$, *Chl-a* and R_{rs} (412, 443, 490, 510, 560, 670) at time lag 0 (panels a to i respectively). Note that the *y* axes range changes depending on the variable used for the correlation.

Thus, a second test was carried out using DD-MLR data set (Table 4.1) implementing several multi linear regressions with different combinations of input variables at time lag 0, showing that SST, SSS, PAR, MLD, $a_{cdom}(443)$ and *Chl-a* could possibly represent

relevant predicting variables of DOC concentration. Indeed, when considering the combination of the latter variables, a significant correlation with DOC is observed (Fig 3.10 a, b), although a lack of correlation at the highest DOC values can be noticed.

On the other hand, when testing the interest of using only $R_{rs}(\lambda)$ as input parameter, it appears that DOC estimates are very constrained within a narrow range of variation (Fig 3.10 c). The same pattern is observed even when excluding the band at 670 nm which is known to present a relative high level of noise. This result was a surprising as it was expected that using R_{rs} as input parameter would have led to consider a "raw" optical information potentially less affected by noise when compared to R_{rs} derived *Chl-a* and $a_{cdom}(443)$ values which precision depends on the bio-optical algorithms related uncertainties

Based on the latter features, SSS, SST, PAR, MLD, $a_{cdom}(443)$ and *Chl-a* were selected as a starting subset of input variables for further evaluating models based on different combinations of these descriptors, which have been included considering different time lags.

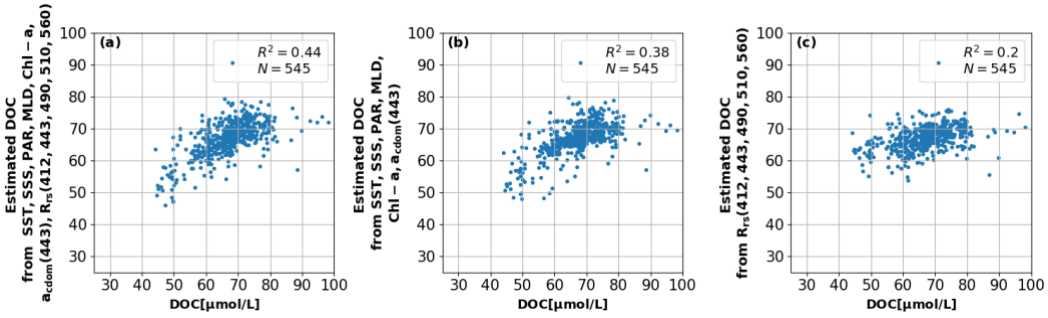


Fig 3.10 Example of multi-linear regression between in situ DOC from DD-MLR and estimated DOC from different combinations of SST, SSS, PAR, MLD, $a_{cdom}(443)$, *Chl-a* and R_{rs} (412, 443, 490, 510, 560) at time lag 0.

3.3.2 Combination and time lag selection

The performance of the ANN was evaluated considering different combination of the set of input variables previously defined (SSS, SST, PAR, MLD, $a_{cdom}(443)$ and *Chl-a*) as well as different time lags between the latter descriptors and DOC. The total amount of possible combinations of the six products at five different time lags (0, -1 week, -2 weeks, -3 weeks, -4 weeks) is reaching a total of 768211 (Table 4.2).

When building an ANN, it is necessary to take into consideration the number of the available data points to be aware of what is the maximum amount of input variables that can be used (May et al., 2011). In the ANN while the amount of inputs variables increases linearly, the total error of the model increases exponentially. Hence, in order to map a given function over the model with sufficient confidence, an exponentially increasing number of samples is required (Scott, 1992). Since the number of data points available for mapping a function in the model is normally finite, the amount of input variables that can be used with good confidence in the result is limited. Therefore, Silverman (1986) established the growth of the minimum sample size required to maintain a constant error associated with estimates of the input probability, as determined by the pattern layer of a generalized regression neural network (Table 3.3). Considering the statistics by Silverman (1986) and that DS1 contains 2895 data points, the maximum number of input variables that could be used in the ANN was set to 6.

Table 3.2 Number of possible combinations calculated according to the amount of input variables used (SSS, SST, PAR, MLD, $a_{cdom}(443)$ and *Chl-a*) considering that each variable is tested at 5 different time lags.

# inputs	# combinations
1	30
2	435
3	4060
4	27405
5	142506
6	593775
TOTAL	768211

Table 3.3 Sample size with increasing dimensionality required to maintain a constant standard error of the probability of an input estimated in the ANN pattern layer (Silverman, 1986).

# inputs	Sample size
1	4
2	19
3	67
4	223
5	768
6	2790
7	10700

Models with large number of input variables tend to be biased as a consequence of over-fitting, therefore the best model is not always the one with lowest mean square error (MSE, May et al., 2011). Hence, to avoid the use of an over-fitted model it is more

appropriate to base the selection on the Akaike information criterion, AIC, (Akaike, 1974), which penalizes overfitting. The lowest the AIC is, the better the model performs. The AIC calculation is performed as follows:

$$AIC = n * \ln\ (MSE) + \frac{n+p}{1-(p+2)/n} \tag{3.5}$$

where n is number of data points in the training dataset, MSE is the mean square error of the estimation and p is the number of weights and bias in the ANN calculations.

The test of the combination of input variables and time lags was performed using a standard fully connected ANN structure (Fig 3.11 a). This consisted in one input layer with 1 to 6 Input Nodes (IN), where each node corresponds to one input variables (Table 4.3), one hidden layer with $2 * IN$ and one output layer with one output node, here the DOC concentration. The activation functions used were the rectified linear unit (ReLU) for the hidden layer, widely used non-linear for being more efficient than others (e.g. sigmoid; Sharma et al., 2020), and a linear activation function for the output layer. At last, the optimization technique used to reduce the errors was the adaptive moment estimation (Adam; Kingma and Ba, 2015).

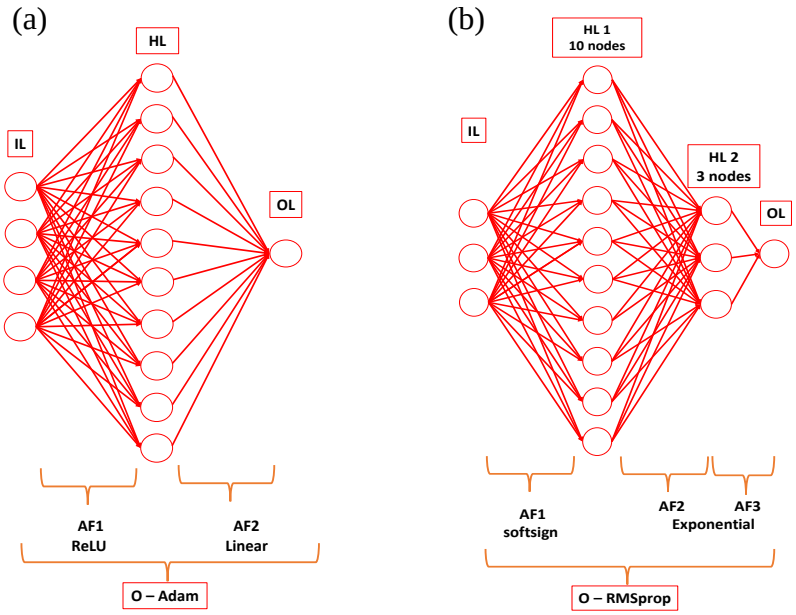


Fig 3.11 (a) Scheme of initial ANN structure with one input layer (IL), one hidden layer (HL) and one output layer (OL), the respective activation function (AF1 and AF2) for the hidden layer and the output layer (ReLU and Linear, respectively), and the Adam optimization technique (O). (**b**) Flow chart of NN29b structure with one input layer (IL), two hidden layer (HL1 and HL2) and one output layer (OL), the

respective activation functions (AF1, AF2 and AF23) for each hidden layer and for the output layer (softsign, exponential and exponential, respectively), and the RMSprop optimization technique (O).

Before starting the test of the different estimator combinations a reference AIC, AICr, was initialized as infinite. Once the test was launched, for each combination of input variables it was checked that the sample size was big enough to train and test the ANN considering to the number of inputs used (Table 4.3). If it was not, the input combination was automatically discarded. On the contrary, if the amount of data points was enough, the data set was normalized, by subtracting the mean and dividing by the standard deviation, and randomly split into two subsets, one for training (70% of the data) and other for validating (30% of the data) the ANN.

After the training for a defined configuration, the AIC was calculated and compared to AICr, if it was smaller the ANN was saved, and its AIC value was set as the new AICr, if not both were discarded (Fig 3.12).

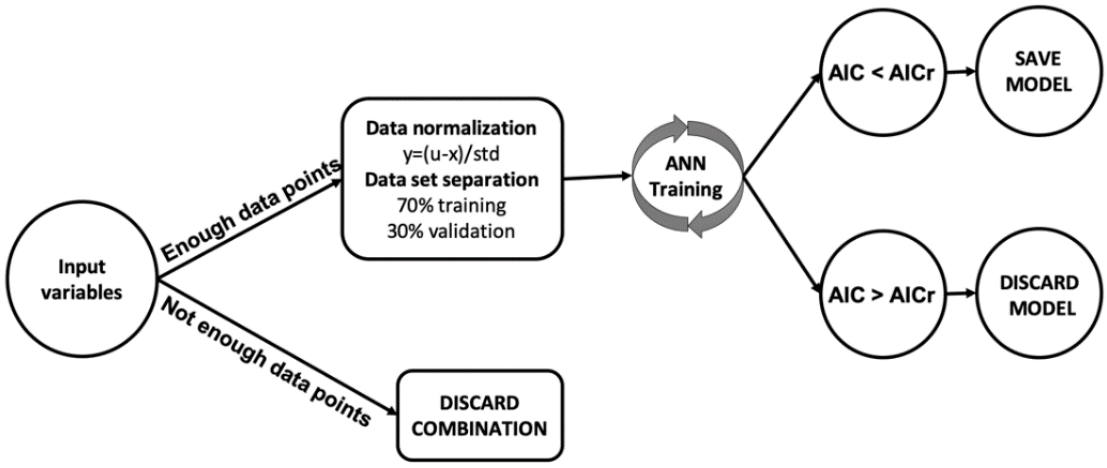


Fig 3.12 Scheme of the decision system followed for the selection of the best input variables and structure used in the ANN to estimate DOC.

Originally, the selection process was planned to be applied to all the possible combinations with 1 to 6 input variables. Nevertheless, because the process was very time consuming and considering that by adding more input variables there was no evident improvement, the process was forced to stop while running the 5-inputs loop, the last ANN saved being one with only 3 input variables.

After many simulations, a total of 29 artificial neural networks were saved with a decrease of AIC from >16000 to 1059.89 (Fig 3.13). This final value corresponds to the last

simulation saved. This model, which shows the best performance among the different configurations tested, is referred as NN29 and was chosen as a first ANN for estimating DOC. It considers 3 input parameters: SST (-1 week), $a_{cdom}(443)$ (-2 weeks) and MLD (-1 week).

The data set used for training and validating (DD-NN) the model NN29 counts with a total of 308 data points, 215 for training and 93 for validation (Fig 3.5).

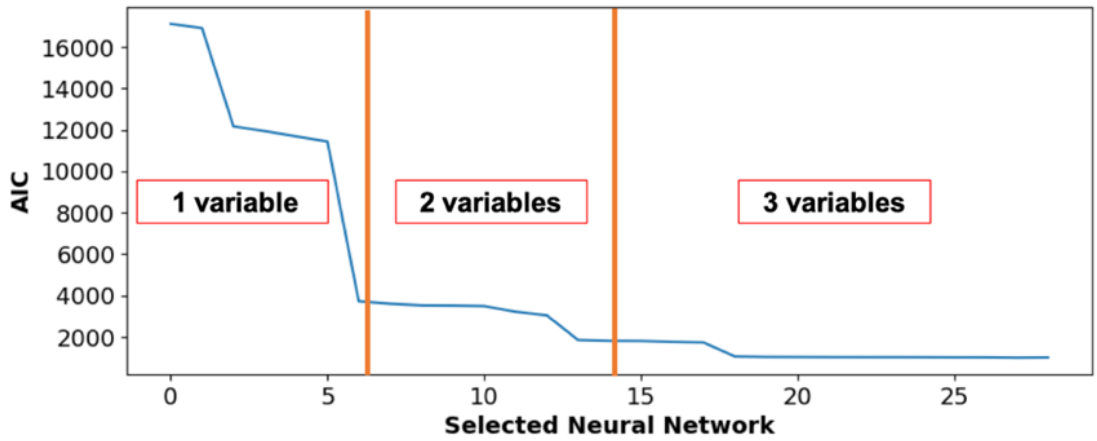


Fig 3.13 AIC decline in the progress of the ANN input data selection. Orange vertical lines indicate the point of increase of number of input variables.

The estimations retrieved by NN29 presents DOC concentration ranging between 50 and 85 µmol/L (Fig 3.14 a), while in situ DOC range was 44.8 to 89.8 µmol/L. However low slope (0.43, 0.2 and 0.38 for the training, validation and complete dataset respectively) and *r* (0.64, 0.24 and 0.54 for the training, validation and complete dataset respectively) values obtained in the linear correlation between in situ DOC vs estimated DOC emphasize the limited performance of this first model (Fig 3.14 a). It is evident that NN29 is slightly overestimating DOC in the lowest values and underestimating it in the highest ones (Fig 3.14 c). Despite this, the selection of the inputs is in agreement with what has been observed by Siegel et al. (2002) at the BATS station, where changes in the water temperature and mixed layer depth are immediately followed by a direct response of DOC concentration, while $a_{cdom}(443)$ has an opposite behavior with a larger time lag (see their Fig 8). Therefore, it is thought that the poor performance of NN29 may be caused by a non-optimal structure of the neural network.

3.3.3 Structure selection

To increase the performance of NN29, the same input variables and dataset (DD-NN) were used to train a new ANN with different structure. The new version of the ANN was named NN29b.

The restructuration of the model was performed changing the number of nodes and hidden layers, testing different activation functions and optimization techniques. Thus, the number of hidden layers ranged between 1 and 2 with a number of neurons ranging between 1 and 10. Nine activation functions were tested: ReLU, sigmoid, softmax, softplus, softsign, tanh, selu, elu and exponential. And seven different optimization techeniques were performed: Adam (Kingma and Ba, 2015), RMSprop (Ruder, 2016), Adadelta (Zeiler, 2012), Adagrad (Kingma and Ba, 2015), Adamax (Ruder, 2016), Nadam (Ruder, 2016), Ftrl (Shalev-shwartz 2007).

Following the same protocol described in the previous section, different combinations of structures were tested with the difference that in this case the initial AICr was set to 1059.89, the AIC value obtained for NN29.

The best structure selected for the estimation of DOC, with AIC of 1009.41, is formed by 2 hidden layers, the first one with 10 nodes and the second one with 3. With softsign activation functions for the first hidden layer and exponential for the second hidden layer and output layer, and RMSprop optimization technique (Fig 3.11 b).

The restructured model (NN29b) shows a general improvement in the estimation of DOC, with slope from the linear regression between in situ DOC and estimated DOC reaching values of 0.63, 0.45 and 0.59 for the training, validation and complete dataset respectively (representing an increase of 46, 40 and 120 % when compared to NN29). Also *r* values increased by 12, 112 and 26% with values reaching 0.72, 0.51 and 0.68 for the training, validation and complete dataset, respectively. The RMSD for the NN29b is slightly lower than in NN29, the MAPD is quite similar between both models whereas the calculated MB was higher for NN29b (Fig 3.14. a, b). It is worth noticing that NN29b retrieves DOC varying over a wider range than NN29 leading to a better coverage of the observed range in the in situ DOC measurements (Fig 3.14. c and d).

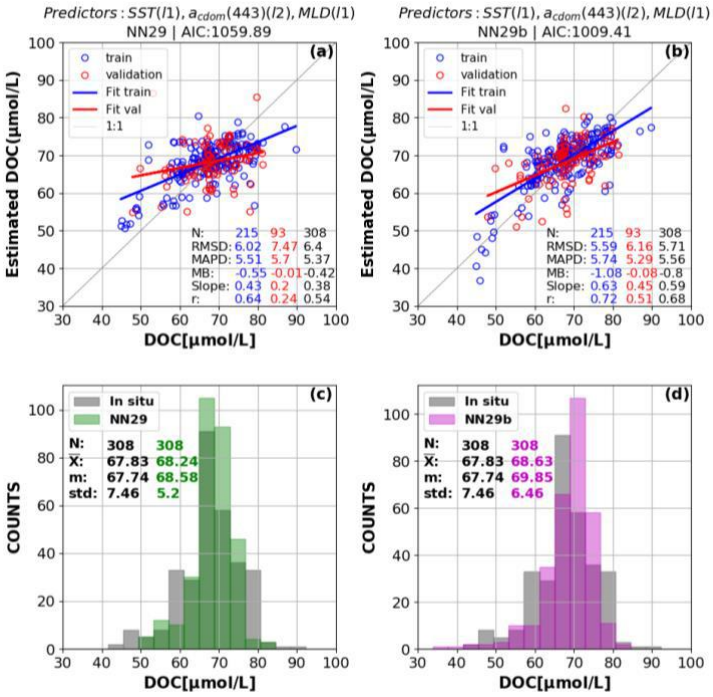


Fig 3.14 Comparison of the model-derived and measured DOC from DD-NN for NN29 (a) and NN29b (b) for the training (blue) and validation (red) data sets, with their respective histograms of DOC measured in situ (grey) or estimated from NN29 (c) and NN29b (d). N, RMSD, MAPD, MB, std and r correspond to the number of data points, the root mean square deviation, median absolute percentage deviation, mean bias, standard deviation and the correlation coefficient, respectively (top panels); and $\overline{X}$ and m correspond mean and median, respectively (bottom panels).

The effect of the addition of *Chl-a* on the DOC estimation was re-evaluated to make sure that this input variable was not overlooked during the selection process by a poor performance of the original structure. Therefore, a new ANN with the same structure of NN29b was tested with *Chl-a* as additional input variable evaluated individually at the 5 different time lags, 0, -1 week, -2 weeks, -3 weeks and -4 weeks. The results of this exercise showed an improvement in the performance of the ANN when *Chl-a* at a time lag of -1 week is included as a predictor with an AIC 44% lower than NN29b (567.35 vs 1009.42 respectively). This new ANN referred to as NN29bCHL, was trained with SST (-1 week), $a_{cdom}(443)$ (-2 weeks), MLD (-1 week) and *Chl-a* (-1 week) from DD-NNCHL (Fig 3.6).

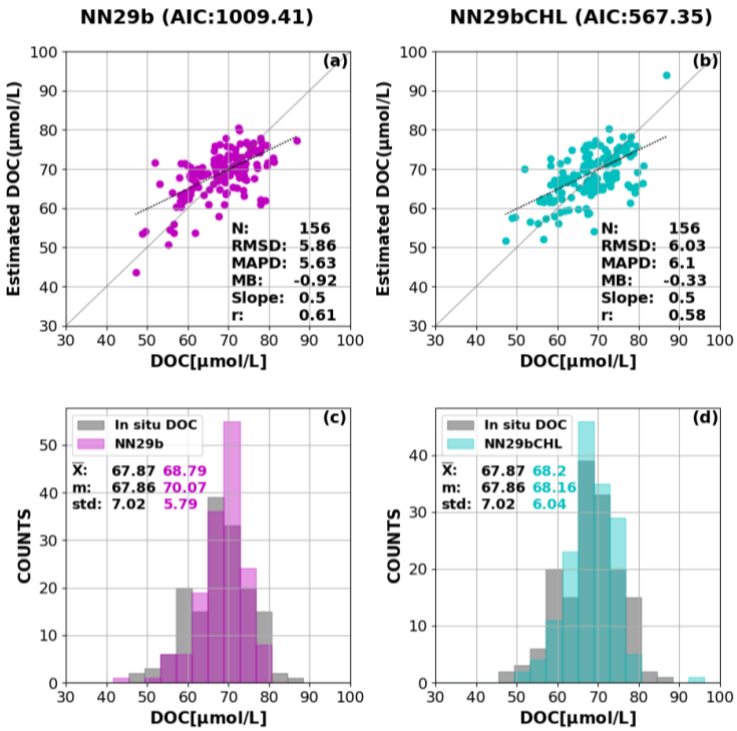


Fig 3.15 Comparison of the model-derived and measured DOC from DD-NNCHL for NN29b (a) and NN29bCHL (b) and their respective histograms in pink and cyan (c and d) over the in situ DOC concentration histogram (gray). N, RMSD, MAPD, MB, std and r correspond to the number of data points, the root mean standard deviation, median absolute percentage deviation, mean bias, standard deviation and the correlation coefficient, respectively (top panels); and $\overline{X}$ and m correspond mean and median respectively (bottom panels).

3.3.4 Water type model dependency

The previous models were developed considering the whole data set available for each specific configuration. An evaluation of the performance of NN29b and NN29bCHL according to the different optical water types was further performed considering the DD-NNCHL data set which has been categorized according to the optical water types defined in Mélin and Vantrepotte (2015) described in section 2.1.2.1.1.

Applying NN29b and NN29bCHL to each individual OWT subset an apparent better estimation of DOC concentration by NN29bCHL is found for water classes 1 to 9, while a better DOC estimation by the model NN29b is found for the water classes 10 to 17 (Fig 3.16). Note that classes 1 to 9 belong to Case 2 waters, while classes 10 to 17 belong to Case 1 caters according to the definition of Morel and Prieur (1977).

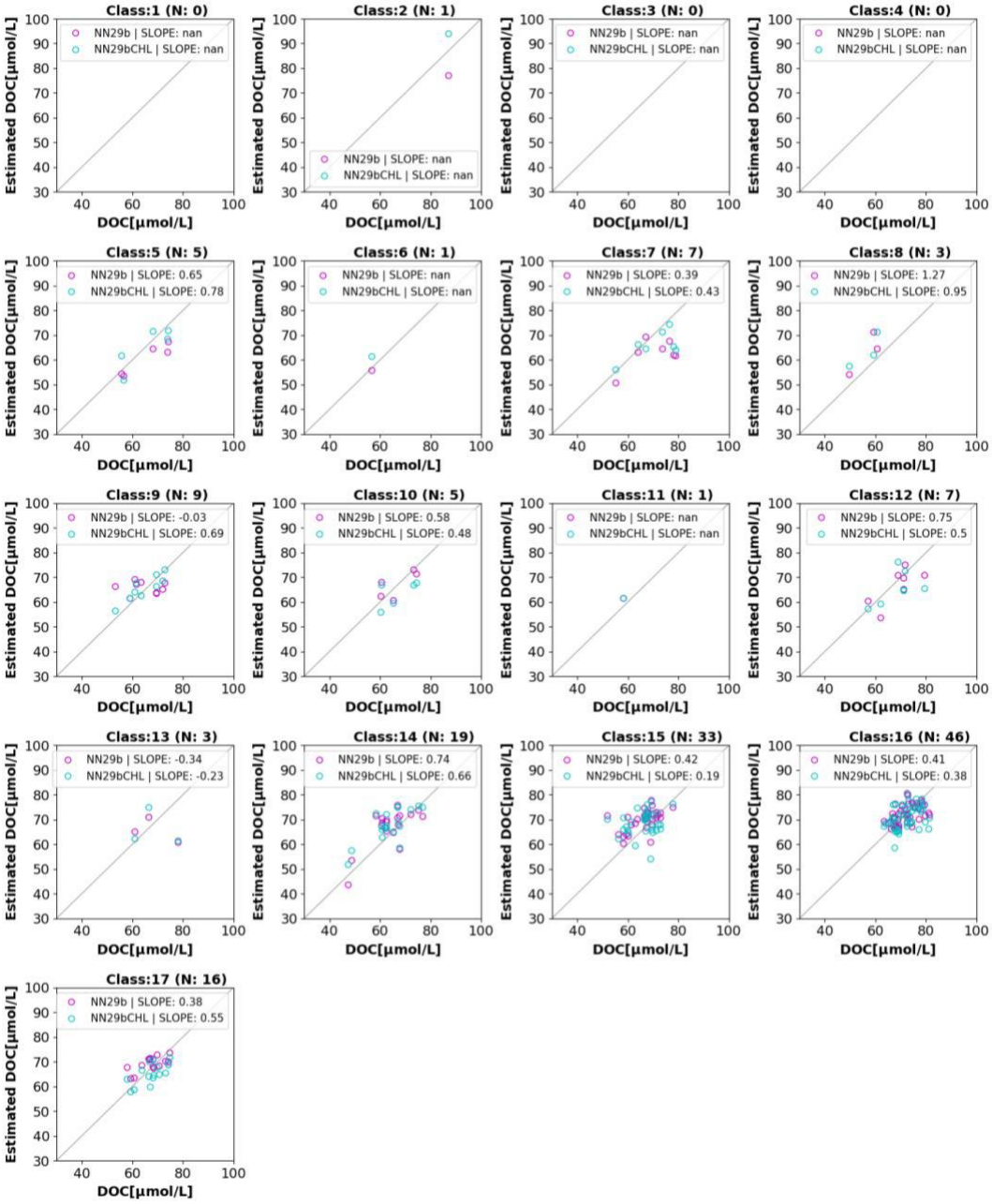


Fig 3.16 Comparison of the model-derived and measured DOC for the NN29b (pink) and NN29bCHL (cyan) models from the DD-NNCHL data set and for each water classes 1 to 17. The solid line represents the 1:1 line, and the slope value of the best fit linear regression type-II is provided.

Taking this result into consideration, the NN29b and NN29bCHL were applied to their best domain of applicability in order to combine both methods on the corresponding groups of classes. This class dependent combination of NN29b and NN29bCHL allow an improvement of the DOC retrieval (Fig 3.17 a, b) as illustrated by the better slope of the linear regression between in situ and estimated DOC (0.57 for the mixed model vs 0.5 for both NN29b and NN29bCHL), the higher r for this relationship (0.68 for the mixed model

vs 0.61 for the NN29b and 0.58 for NN29bCHL models) and the lower average error in DOC estimates (RMSD = 5.43 µmol/L for the mixed model vs 5.86 µmol/L for NN29b and 6.03 µmol/L for NN29bCHL). The better performance of the combination of the latter two modes is also sustained by the radar plot where the mixed method presents an area 15% and 17% smaller than the one obtained considering NN29bCHL and NN29b, respectively (Fig 3.17 c).

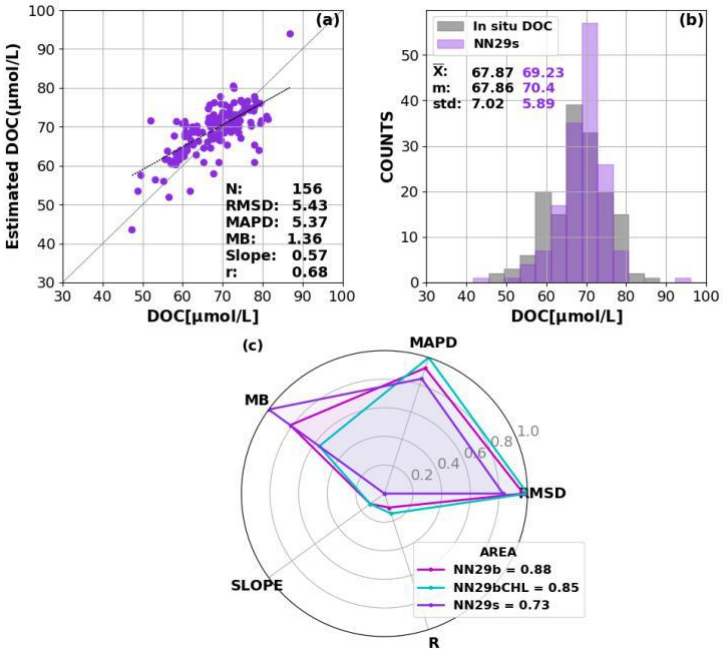


Fig 3.17 Comparison of the model-derived and measured DOC from DD-NNCHL for NN29s (a), its respective histogram (purple) over the in situ DOC concentration histogram in gray (b) and the radar plot comparing the performance of NN29b (pink), NN29bCHL (cyan) and NN29s (purple). N, RMSD, MAPD, MB, std and r correspond to the number of data points, the root mean square deviation, median absolute percentage deviation, mean bias, standard deviation and the correlation coefficient, respectively (left) and $\overline{X}$ and m correspond mean and median respectively (right).

A possible explanation of the different performances of the two ANNs according to the OWT could be that in Case 2 waters *Chl-a* and $a_{cdom}(443)$ are the main drivers of DOC dynamics, while in Case 1 water the physical forcings such as SST and MLD are the predominant driving parameters.

From these results, a new model called NN29s that combines NN29b and NN29bCHL by using a switch activated by the water class status was built. If the water is categorized as class 1 to 9 NN29s consist in applying NN29bCHL model to estimate DOC, while on the

contrary, if the water classification for a defined pixel is between 10 and 17, then NN29s uses the NN29b model for deriving DOC content (Fig 3.18).

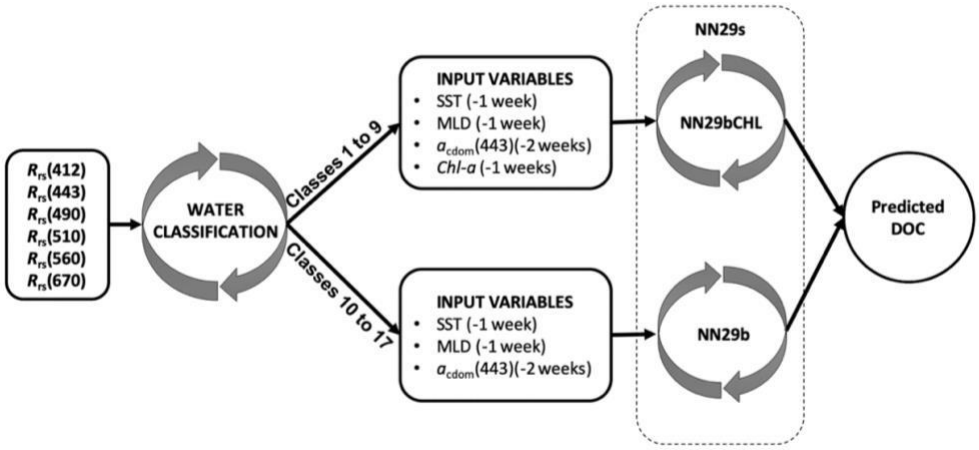


Fig 3.18 Scheme of NN29s model functionality, starting with water classification which is used for the decision of which ANN will be applied (NN29b or NN29CHL) to estimate DOC.

In order to test the relevance of using temporally lagged input data, the same structure of the NN29s model has been re-trained with the same input variables measured simultaneously with the in situ DOC. The validation exercise of this alternative model with the DD-NNCHL subset shows great decrease of the performance. In general the use of non-temporally lagged data presents less precision reflected on higher RMSD (10.46 μmol/L *vs* 5.43 μmol/L of NN29s) and MAPD (6.01 % *vs* 5.37 % of NN29s) and lower slope (0.36 *vs* 0.57 of NN29s) and r (0.26 *vs* 0.68 of NN29s). Results that evidence the importance of using temporally lagged input data.

3.4 Evaluation of the performance of NN29s on climatological data and comparison with other models performance on climatological data

3.4.1 NN29s comparison with Siegel et al. (2002) and Aurin et al. (2018)

Monthly climatology for the DV1 data set (Fig 3. 7; Table 4.1) of in situ and estimated DOC computed with 8 days composite data by the NN29s model, and using the formulations proposed by Siegel et al. (2002) and Aurin et al. (2018) were used for the comparison and validation of the three mentioned models (Fig 3.19). Monthly climatologies, with all their inherent limitation (especially for the inter-annual variability), were considered for this exercise to increase the amount of available points for the models comparison.

The comparison with climatological in situ data shows that the model NN29s provides the most accurate estimation of DOC when compared to the other two models, presenting the lowest RMSD, MAPD and MB (7.02 μmol/L, 6.86 % and 0.82 μmol/L, respectively).

On the other hand, the model of Siegel et al. (2002), although covering the general range of DOC values observed in situ (44 – 89.4 μmol/L), shows a bimodal distribution in the estimated DOC as illustrated from both scatterplot and histogram representations reported in Figs 3.17 b and e, respectively. The model by Siegel et al. (2002) is based on three different equations for the estimation of DOC depending on the temperature of oceanic basin. The points that are departing from the general pattern in the scatterplot reported in the Fig 3.19 b are indeed mainly related to an apparent failure of the equation used to calculate DOC in the Atlantic Ocean. This is not surprising since the annual climatology presented by the authors already showed very high DOC values in this basin (Fig 1.10).

At last, the model presented by Aurin et al. (2018) shows a high bias in the DOC estimates with a great overestimation reflected by a RMSD of 37,8 μmol/L, a MAPD of 32,53 % and a MB of 31.47 μmol/L (Fig 3.19 c). This bias is further illustrated by the mean and median DOC values for that model which are both 43% higher than the ones for in situ data (Fig 3.19 f). This result is not surprising since the global annual average map presented by the authors (Fig 1.11) exhibits a global distribution opposite to what has been described based on in situ observations (sections 1.7 and 1.3).

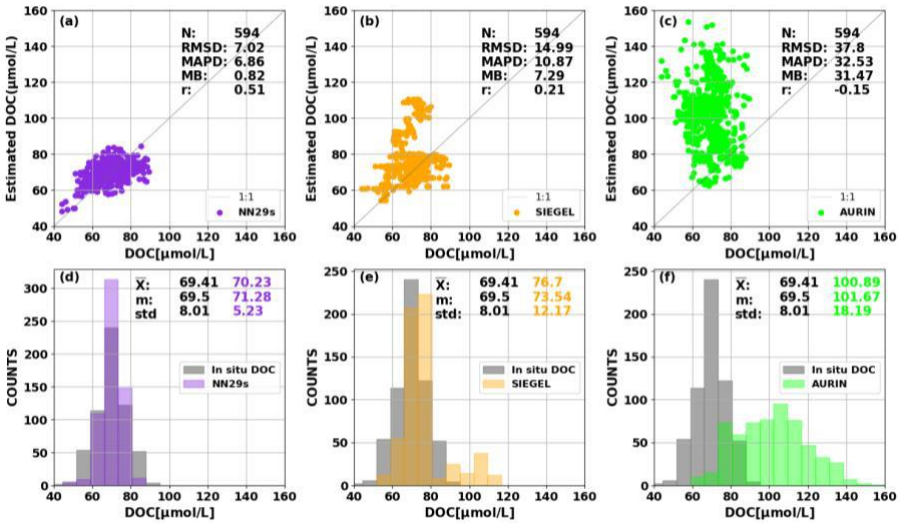


Fig 3.19 Comparison of the model-derived and measured DOC from DV1 for NN29s (a), Siegel et al. (2002) (b) and Aurin et al. (2018) (c) from DV1, and their respective histograms in purple (d), yellow (e) and green (f) over the in situ DOC histogram (gray). N, RMSD, MAPD, MB, std and r correspond to the

number of data points, the root mean squared deviation, median absolute percentage deviation, mean bias, standard deviation and the correlation coefficient, respectively (top panels). $\overline{X}$ and m correspond to the mean and median respectively (bottom panels).

To better analyze the performance of NN29s, the validation exercise was repeated comparing, for each individual month, the in situ DOC with the extraction of NN29s-derived DOC from the monthly climatology gathered in DV2 (Fig 3.20; Table 4.1). In this case to limit the impact of inter-annual variability, only data collected between 2002 and 2012, corresponding to the satellite time period considered, have been used to generate the in situ climatology.

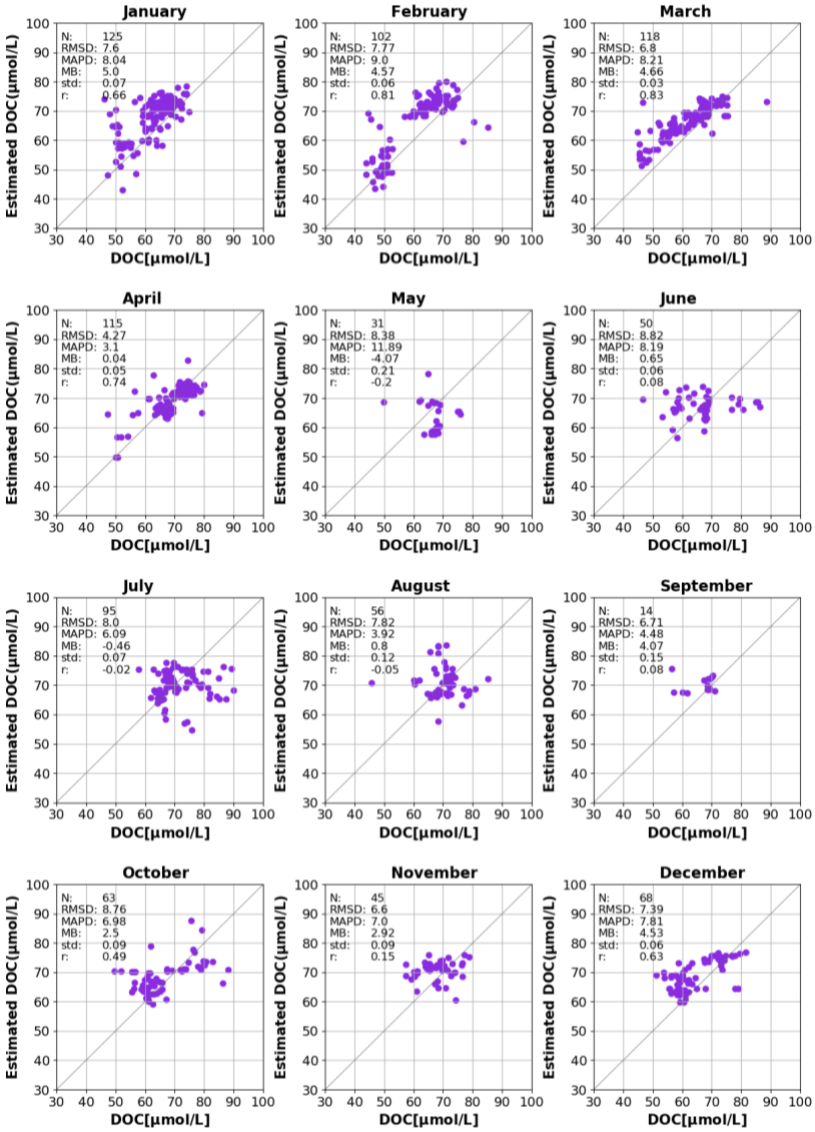


Fig 3.20 Comparison of the model-derived and measured DOC from DV2 for NN29s from 10 years a weekly time series (2002 - 2012) match-up with in situ DOC monthly climatology from DV2. N, RMSD, MAPD, MB, std and r correspond to the number of data points, the root mean standard deviation, median absolute percentage deviation, mean bias, standard deviation and the correlation coefficient respectively.

The results tend to show that no drastic bias appears according to the considered months, even if the number of points and the range of variability for some specific months do not still allow to draw a definitive conclusion. This exercise should therefore be repeated in the future with a more complete data set.

From January to April, as well as for the months of October to December, that is for months at which the number and range of variability of data points are significantly higher, DOC is retrieved more accurately as emphasized by the relative low RMSD (≤ 10 μmol/L), MAPD (≤ 9 %) and std values (between 0.03 and 0.09 μmol/L), and relatively good r values except in November (r=0.19 in November and r > 0.48 for the other months; Fig 3.20). In both June and July, some data points depict from the general dataset at DOC concentration higher than 80 μmol/L. These data points come from very coastal sampling stations for which the present algorithm may not be well suitable (coastal dedicated algorithms, as the one of Vantrepotte et al. (2015), can be used for that purpose).

3.4.2 DOC global distribution and temporal variability

In order to compare the consistency in the distribution of the measured and modelled DOC values, in situ DOC monthly climatology data points corresponding to the validation data set DV2 were over plotted on the global monthly climatology maps of the NN29s derived surface DOC (Fig 3.21).

From these figures, it appears that highly biased DOC estimates are mainly located in the east coast of USA between May and August, in the western Pacific between July and October and more locally in the Arctic region in July (Fig 3.20).

The observed high discrepancies between in situ and estimated DOC in coastal areas and in the Arctic domain are not surprising since those regions were slightly represented in the training data set used for developing the model NN29s (Fig 3.5, Fig 3.6).

The discrepancies found in the western Pacific could more likely be attributed to anomalous years, which might have been smoothed when in the climatology data. Specifically, the highest biased predictions in the latter area indeed correspond to in situ DOC concentrations measured between 1998 and 2001, period of time which includes the strongest El Niño/Southern Oscillation (ENSO) event registered in the twentieth century followed by a 2-year-long moderate-to-strong La Niña event (Shabbar and Yu,

2009). Moreover, these latter years are not included in the temporal time series of estimated DOC used to generate the monthly climatology (from April 2002 to April 2012).

If the in situ data points sampled between 1998 and 2001 are dismissed, an overall good consistency is found between in situ measurements and estimated DOC as it happens in November (on the years 1997, 2003, 2009, 2010) and December (on the years 1997, 2009) (Fig 3.21).

Furthermore, the DOC monthly climatology maps derived from the NN29s model exhibit an annual cycle dynamic that agrees with previous works. These results stated that the DOC locally produced in the euphotic zone is conditioned by the intensity of the primary production and related processes (Hansell et al., 2002). Specifically, it is known that DOC produced at the equator is exported toward the gyre areas by divergent currents where it accumulates (Roshan and DeVries, 2017). It appears therefore logical to observe that the greatest DOC concentrations are located in the subtropical areas.

It is evident that DOC concentration decreases with increasing latitude from a maximum of approximately 80-85 μmol/L in the subtropics to 50 μmol/L in the Arctic subpolar area and to approximately 45 μmol/L in the Southern Ocean. This is consistent with Sarmiento and Gruber (2006) who reported that the lowest concentrations of DOC can be found in the Southern Ocean with values ranging between 40 and 50 μmol/L.

In the Pacific Ocean, high DOC concentrations throughout the year are localized in the gyres divided by a low-DOC concentration band in the equator coinciding with the upwelling of Peru region and divergent currents that drag the DOC poleward. In the North Pacific, the highest DOC concentrations (~85 μmol/L) are found in May, at the end of the northern spring. Afterwards starts a decrease at the same time that the South Pacific increase the concentration reaching a maximum (~80 μmol/L) by November/December, corresponding to the end of the southern spring.

During the period of maximum DOC in the South Pacific, the division observed by the low-DOC band at the equator appears to be more diffuse, becoming more evident when the northern Pacific reaches its highest DOC concentration.

On the other hand, in the subtropical Atlantic a similar, yet a lower, seasonality is observed. A wide latitudinal band from 30ºN to 30ºS of high DOC is strongly marked in January, getting narrower by August when it starts to expand again.

On the contrary, such seasonal variations are not found in the Indian ocean where no clear monthly dynamics can be observed with DOC values rounding 73 µmol/L.

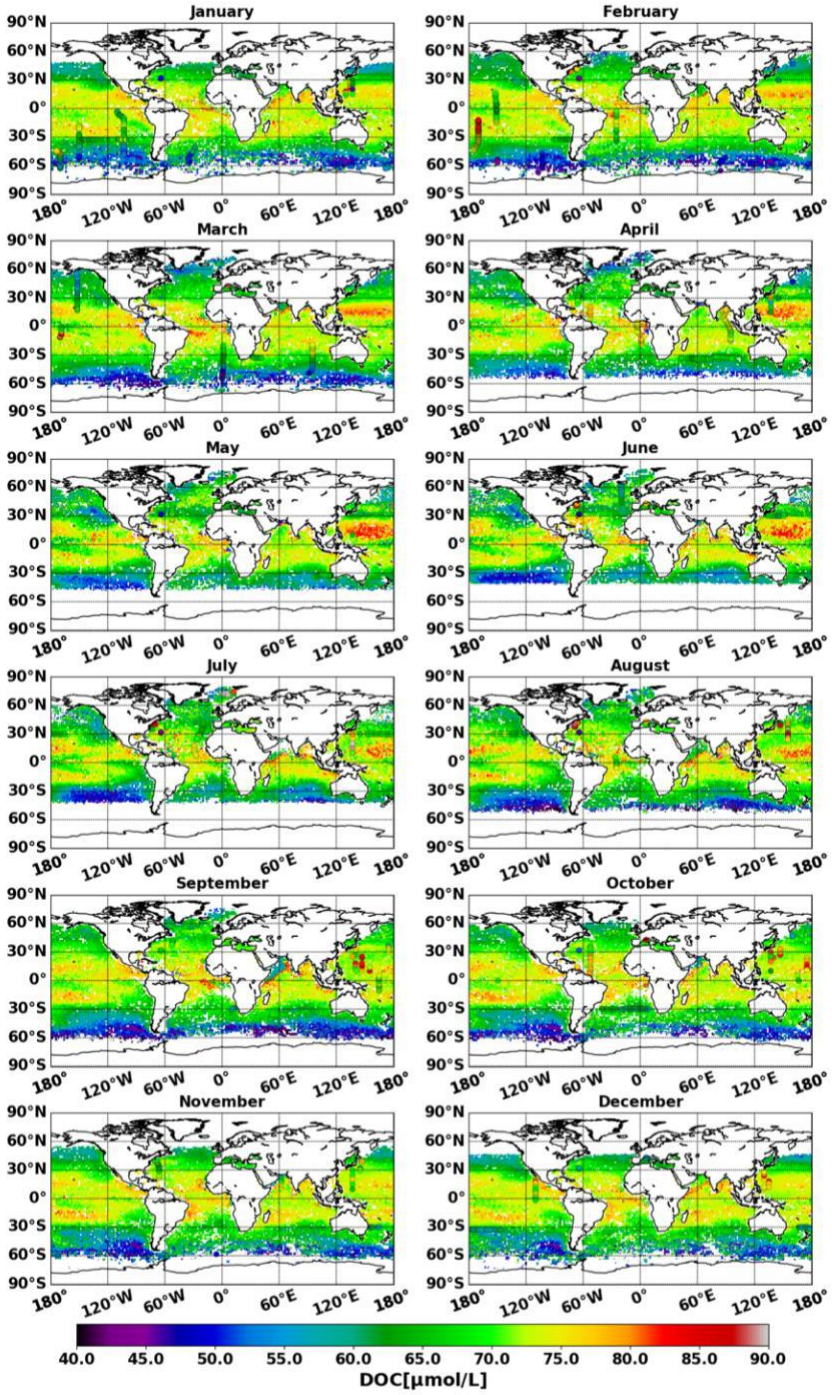


Fig 3.21 Monthly climatology of DOC concentration generated with NN29s from 10 years of weekly time series (2002 - 2012) with in situ DOC monthly climatology from DV2 over plotted.

In the North Atlantic, it has been shown that DOC is brought to the region by wind-driven surface currents from lower latitudes, accumulates there during summer and is exported in winter to the deep ocean (depths >1000 m) via meridional overturning circulation and ventilation to be long-term sequestrated in the ocean interior (Carlson et al., 1994, 2010; Copin-Montégut and Avril, 1993; Hansell et al., 2002, 2009; Hansell and Carlson, 2001; Hopkinson and Vallino, 2005). In line with this, the monthly climatology of estimated DOC show that in the North Atlantic the low DOC concentration found (50 μmol/L) in March and April tends to increase towards the summer season due to an accumulation of DOC favored by the water stratification, reaching a maximal concentration of 75 μmol/L in June. From June to September, DOC concentration remains generally stable in this oceanic region.

3.4.3 Global Distribution: comparison with Roshan and DeVries (2017)

The comparison of the annual average map (Fig 3.22 a, b) documented by Roshan and DeVries (2017) and the one computed from the NN29s model shows consistency in the estimated DOC distribution. It is worth noting that only 6% of the data used to develop NN29s is included in the dataset used by Roshan and DeVries (2017). For both models, the highest DOC values are located in a wide band in the subtropical area. In a consistent way, these high DOC waters split, as mentioned previously, into two waters bands in the Pacific Ocean with a low-DOC band at the equator due to the upwelling and divergent currents that export the DOC poleward. Further, both models show that the western Pacific waters has higher concentration of DOC than the eastern Pacific ones. In addition, both show concentrations ranging between 50 and 70 μmol/L in the North Atlantic and between 45 and 50 μmol/L in the Southern Ocean.

Despite the localized discrepancies between both models estimates, 75% of the DOC values modeled from Roshan and DeVries (2017) and NN29s show MAPD lower than 15% (Fig 3.22 c). Therefore, considering that the general good performance of the DOC estimated from the model of Roshan and DeVries (2017), with a documented R^2 ranging between 0.8 and > 0.9 depending on the basin, the similarity of the results obtained with the NN29s model are encouraging, as these would confirm the robustness of the new method presented here.

Few oceanic areas show however relevant discrepancies between the two models. This is the case of the northwest Pacific and eastern equatorial Pacific where the model by Roshan and DeVries (2017) deliver DOC values between 20 to 30 % lower than those estimated from the NN29s model.

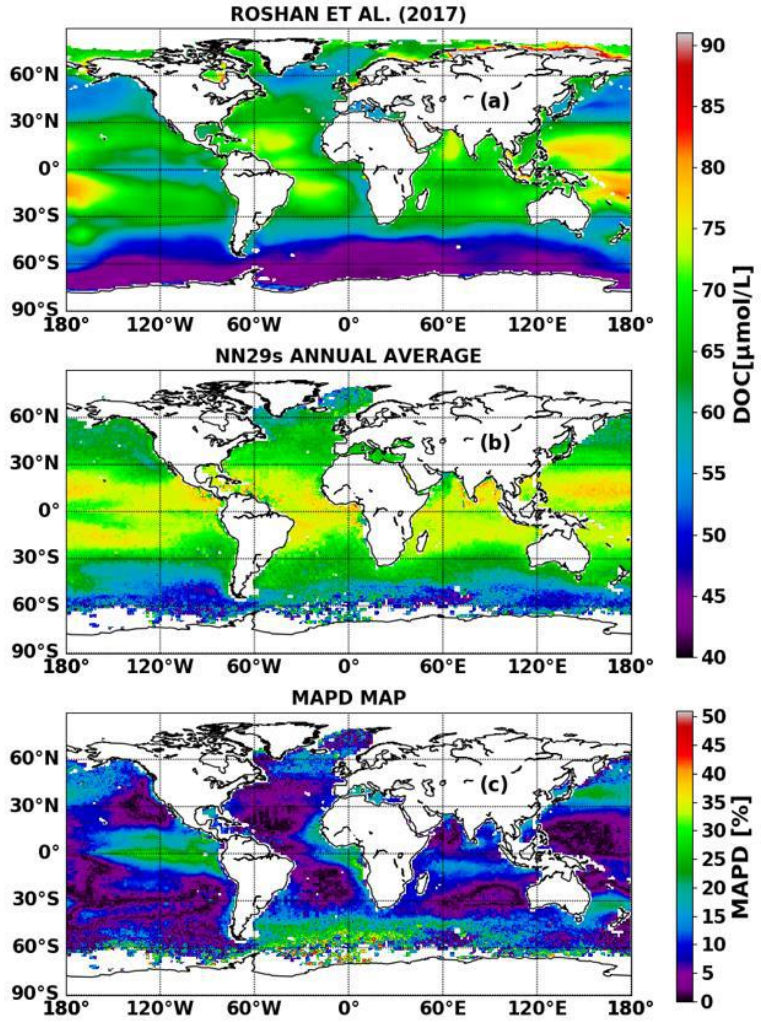


Fig 3.22 ANN-derived annual average DOC concentration from Roshan and DeVries (2017) (a) and from NN29s (b) and the MAPD ((XX-YY)/(XX+YY) * 200) map comparing the two models (c).

The difference between both models in the equatorial Pacific is thought to be caused by the fact that the data used by Roshan and DeVries (2017) to generate the annual average of DOC was collected during very short time period, from the 24 of December 2007 and the 06 of January 2008 in the equatorial Pacific. This sampling period coincides with an El Niño/Southern Oscillation (ENSO) event which started in September 2007 reaching a peak in February 2008. It is known that this phenomenon might conduct to the presence of anomalous biophysical conditions in the area (e.g. negative SST anomaly situation, https://www.ncdc.noaa.gov/sotc/enso/200813). Thus, the temporal representativeness of

the data used by Roshan and DeVries (2017) could be questioned and might explain the deviation of this models when compared with the NN29s outputs.

Nevertheless, to evaluate the capability of NN29s to reproduce the ENSO events the weekly and monthly maps generated with NN29s at the equatorial Pacific were analyzed in comparison to the in situ data used by Roshan and DeVries (2017). The map of the estimation produced with NN29s for the period of in situ sampling (Fig 3.23 a) confirms that NN29s does not achieve the estimation of the lowest concentrations registered in the area. It is observed that while the in situ DOC measurements during the ENSO event range between 46 and 75 μmol/L, the NN29s-derivide DOC in the region exhibits concentrations between 65 and 75 μmol/L . Therefore, even when NN29s does not get to estimate the lowest concentrations of DOC observed in the equatorial Pacific during the ENSO event, at least its estimation does not overpass the range of DOC concentrations observed in situ.

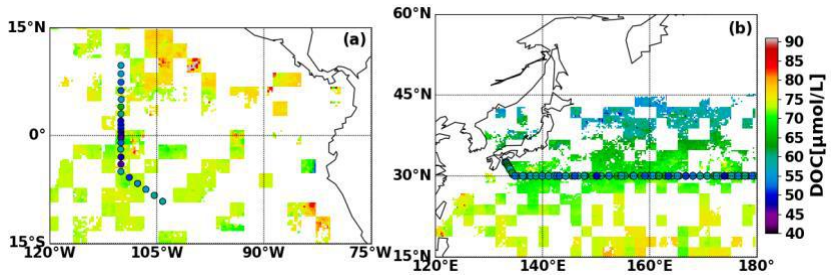


Fig 3.23 a) DOC concentration map estimated with NN29s for the period between 19th of December 2007 to 8th of January 2008 in the equatorial Pacific and (b) between the 1st and 31 of March 2006 in the NW Pacific. The colored dots show the in situ DOC measured within the maps period.

Furthermore, the estimation of DOC concentration during the complete period when La Niña event peaked (September-2007 to February-2008) does not show anomalies in respect to the climatology (Fig 3.24 b). It is noticed that from the input variables used to calculate DOC with NN29s (SST, *Chl-a*, MLD and $a_{cdom}(443)$) only SST showed clear variability during this event (Fig 3.24 b), with MAPD of 10% in respect to the annual average. Thus, it is thought that the sensitivity of NN29s to such change of SST is not enough to register this kind of events.

A different situation is observed in the NW Pacific where the MAPD between the two models is about 25% (Fig 3.22). Roshan and DeVries (2017) calculation of DOC concentrations ranges between 50 and 60 μmol/L, while the in situ data concentration

presents a slightly wider range, from 48.4 to 70.1 μmol/L. Meanwhile, NN29s annual average DOC estimates range from 60 to 70 μmol/L.

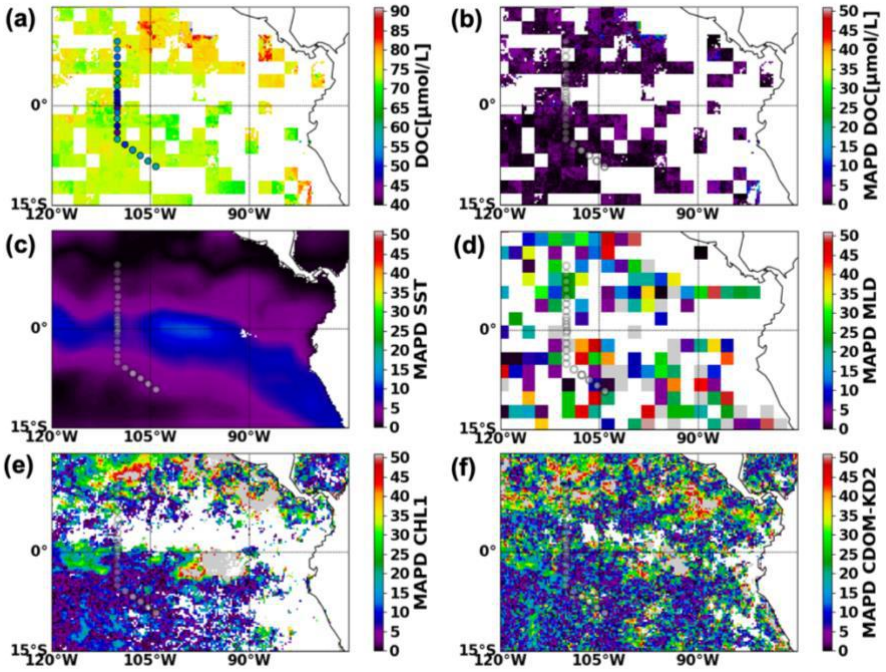


Fig 3.24 a) DOC concentration map estimated with NN29s for the period of Septenber-2007 to February-2008 when La Niña event developed. MAPD ((XX-YY)/(XX+YY) * 200) maps of DOC (b), SST (c), MLD (d), *Chl-a* (e) and $a_{cdom}(443)$ (f) calculated for the same period versus the annual average of each variable (2002 to 2012). The dots show the in situ DOC measured took within the La Niña event in (a) and only the location of the measurements in panels e to f.

Anyhow, the weekly mean maps generated with NN29s (Fig 3.23 b) show that in the NW Pacific between 30ºN and 60ºN the DOC ranges between 65 μmol/L to 50 μmol/L, with the lowest values register at 45ºN. According to the latest results, it is thought that the high values estimated by NN29s in the annual mean are an artifact generated by averaging the data. Also it is thought that the low concentration registered by Roshan and DeVries (2017) in the area might also be an artifact caused by the extrapolation method that they use to fill the gaps, since there is no in situ data measured between 30 ºN and 60 ºN.

To conclude, in the NW Pacific there are two factors that might generate the discrepancies observed between the two models: a slight underestimation of DOC concentration by Roshan and DeVries (2017) due to the extrapolation method, and a slight overestimation by NN29s probably caused by the use of the annual average.

A different situation happens in the southern Ocean (south of 40ºS) where the DOC from Roshan and DeVries (2017) is 30% lower than the one delivered from the NN29s model.

In this region, the satellite data temporal coverage is limited, with acquisitions mainly performed during the southern spring and summer seasons. Thus, while DOC concentration is well represented between September and March, it scarce during the rest of the year. Counting with only 7 months per year for the calculation of the annual average of DOC from NN29s might therefore be the source of bias that produces the differences observed between the two models in the mentioned area.

3.5 Comparison with PISCES

NN29s outputs were also compared with the ones of the biogeochemical model PISCES. Both approaches show globally similar results in the DOC distribution as confirmed by the mean MAPD of 7.16 ± 5 % (Fig 3.25) between the two models. Discrepancies mainly appear for the lowest DOC concentration (< 60 µmol/L for NN29s estimates) for which PISCES tends to retrieve higher DOC values (Fig 3.25 c). This correspond for instance to the situation found in the South Pacific (between 30 and 60ºS) where the PISCES modeled DOC are diverging by 20% from the NN29s estimates (Fig 3.25 d). This area is however characterized by presenting the lowest DOC concentrations observed in the upper ocean (~ 40–50 µmol/L; Sarmiento and Gruber, 2006), therefore suggesting that the PISCES model is failing to estimate DOC concentration in this oceanic region. This statement is confirmed from the comparison between the PISCES DOC monthly climatology with the in situ DOC climatology from DV2 (Fig 3.26).

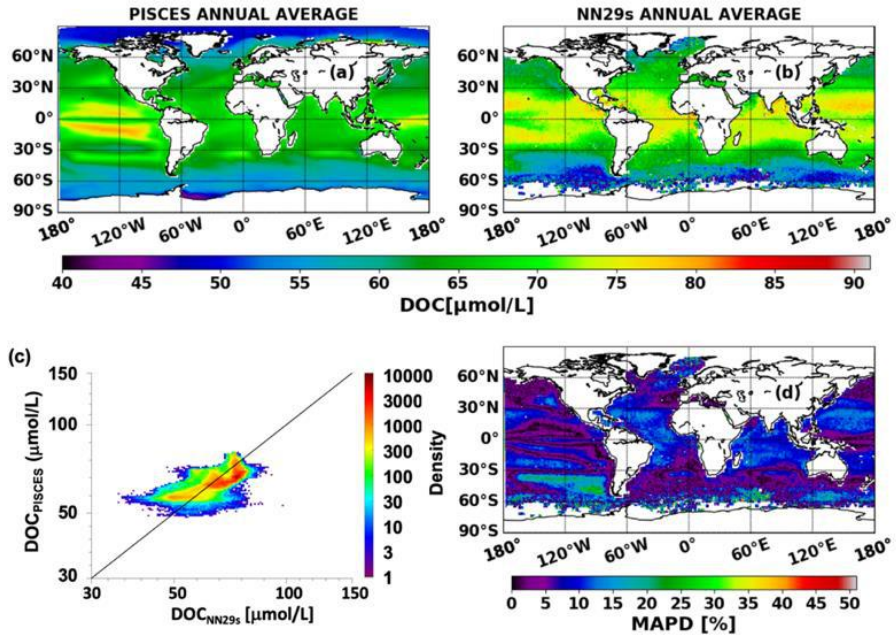


Fig 3.25 Annual average of DOC concentrated generated with PISCES model (a), and with NN29s (b), the density plot (c) and the MAPD ((XX-YY)/(XX+YY) * 200) map comparing them (d).

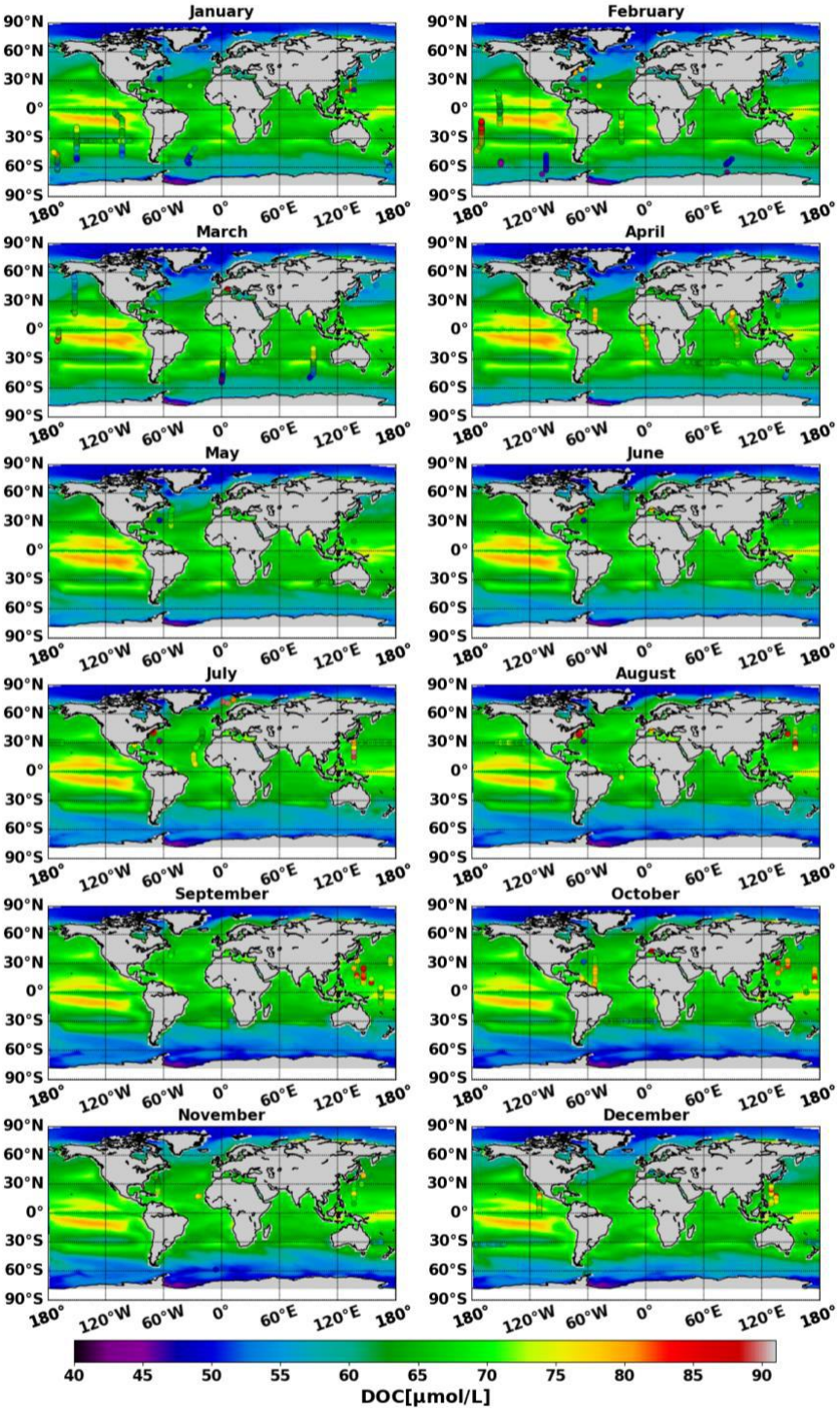


Fig 3.26 Monthly climatology of DOC concentration in the surface ocean generated with PISCES with in situ DOC monthly climatology from DV2.

In the tropical Atlantic ocean, on the other hand, PISCES estimates show DOC concentration ranging 60 to 70 μmol/L (Fig 3.26), while NN29s-derived DOC oscillates between 60 and 80 μmol/L reaching higher values as it is also observed from the in situ observations in that oceanic region (Fig 3.21).

A sharp difference evidenced by the comparison of the monthly climatologies is observed at the southern Ocean where the PISCES model consistently overestimates DOC concentration, while NN29s DOC estimates are globally in line with in situ observations (Fig 3.26).

3.6 Conclusions and perspectives

In this chapter, the development of a new model (NN29s) to estimate DOC in the open ocean through an ANN has been presented. This model uses as input the MLD, SST, and *Chl-a* 1 week before the estimation date, and $a_{cdom}(443)$ 2 weeks before. NN29s has been vastly tested through performance analyses which have diversely taken into account optical water typology, match-up exercises (from monthly climatology) and comparisons with other existing DOC inversion models.

The performance of the new model proposed here has shown its great potential, as it globally depicts the expected features for DOC in terms of spatial distribution and temporal dynamics which are globally in agreement with the patterns observed in situ.

The DOC annual average calculated with the new model showed great similarities with the one presented by Roshan and DeVries (2017). This suggests the great robustness of NN29s since the annual picture produced by the later authors proved to have great accuracy with in situ DOC measurements. This consistency along with the observed good representation of the DOC annual dynamics on the monthly climatology provides confidence to the performance of NN29s. Such result is very promising as NN29s can also be used to produce high temporal resolution estimations.

However, some problems have been found. First, NN29s could not be properly tested estimating DOC concentrations higher than 85μmol/L due to lack of in situ data in this range. Normally such high values are not found in open ocean waters, where the maximum average rounds the 80 μmol/L. Yet in the western Pacific DOC concentrations close to 90 μmol/L have been registered and NN29s could not reproduce them correctly. It would be necessary to gather more in situ data with DOC concentration >85 μmol/L worldwide distributed to be able to perform a better validation of NN29s in the mentioned range.

A greater issue which is limiting the applicability of the model is related to the lack of available data of the mixed layer depth drove to the production of low coverage maps, within 0.4 - 14.6 % coverage for the 8 days composite global maps. This possibly leads to the addition of noise to the estimation. A possible solution to overcome this issue would consist in testing the performance of NN29s with a different source of MLD in addition to ARGO floats data. It would also be relevant to better explore how much is the performance affected by removing MLD as input parameters of the model, reshaping the structure of the ANN.

Another possible improvement for the NN29s model would be to mix the two original ANNs (NN29b and NN29bCHL) by taking in consideration the belonging probability of each pixel to the two groups of optical classes used for the activation of the switch. This weighted approach would be useful to avoid any spatial artifact in the DOC distribution, providing a smooth transition between the values estimated by the two different sub-models.

In addition, in future studies it would be necessary to pursue a deeper analysis of the covariation of DOC and the variables used for its estimation, to better understand their respective weight and action on the estimation of DOC.

4. General conclusions and perspectives

This thesis arises from the need of better understanding the temporal and spatial variability of the dissolved organic carbon (DOC) in the global ocean. In this frame, new developments were performed to deliver innovative information on DOC distribution at global scale through maps of DOC distribution at global scale estimated from satellite observation.

On this basis the first algorithm capable of estimating DOC on global scale from satellite and ARGO floats data was developed. For that purpose, the three main following tasks have been accomplished:

1. Define the best algorithm to estimate a_{cdom} from satellite data in open ocean.
2. Determine the variables and time lag that allow the estimation of DOC concentration taking in consideration the water mass history.
3. Define the model structure to produce accurate DOC estimations.

The development of a new model to estimate $a_{cdom}(443)$ at global scale (CDOM-KD2) provides slightly better estimations compared to other previously published algorithms (Aurin et al., 2018; Chen et al., 2017; Shanmugam, 2011). The CDOM-KD2 model validation with both, in situ and match up data, retrieved more accurate estimation of $a_{cdom}(443)$ especially in open ocean waters.

The new model to estimate $a_{cdom}(443)$ presents the great advantage of retrieving an estimation completely independent from the absorption of the non-algal particles (a_{nap}). Since a_{nap} and a_{cdom} have similar spectral shape CDOM and NAP were usually estimated simultaneously and very few models were available until now to estimate specifically CDOM absorption coefficient over open ocean waters..

The resulting estimated $a_{cdom}(443)$ from CDOM-KD2 along with satellite SST and *Chl-a* and MLD from ARGO floats, were used in the development of the new model to estimate DOC concentration. The mentioned input variables were used temporally lagged in respect to the DOC estimation date. In agreement with previous reports based on in situ observations, the $a_{cdom}(443)$ input was taken from two weeks before the DOC estimation date, while the rest of the variables were taken from 1 week before.

In contrast with models previously published to estimate DOC in global scale, the new model takes in consideration the optical water classes as defined by Mélin and Vantrepotte (2015). In practice two formulation of ANN were considered for retrieving DOC over two groups of Classes basically depicting Class 1 and Class 2 waters. The first model considers classes 1 to 9 while the second one considers classes 10 to 17. Then, by gathering two artificial neural networks that are alternatively used depending on the water type classification, the new model adjusts the estimation to the environmental conditions retrieving more accurate results. Future improvement could be achieved using the probability of belonging of a given pixel to a given class to mix the latter algorithms and provide smoother map of DOC (Mélin and Vantrepotte, 2015; Vantrepotte et al., 2012)

The performance of the new model was validated with good accuracy at global scale. The global monthly climatology properly replicates the expected features according to seasonal in situ observations. As well, the annual average presents consistent results with the accurate annual "picture" presented by Roshan and DeVries (2017). Some discrepancies have been observed between the monthly satellite climatology obtained by the present algorithm and the one provided by Roshan and DeVries (2017) in the eastern equatorial Pacific that might be related to specific situation related to anomalous situation due to exceptional climatic events (e.g. La Nina).

During La Niña events the water temperature in the equatorial Pacific decreases due to a rise of the upwelling of Peru (Mann and Kump, 2015), affecting the productivity of the equatorial Pacific along with all the processes linked to it. Thus changes in the DOC concentration would be expected but not observed in the estimated DOC from the new model. Hence it is necessary to better study the sensitivity of the new model to changes in the input variables to be able to improve its performance in the reproduction of changes induced by this event.

To conclude, the results presented in this PhD sustain that the information retrieved by the new DOC model represents a great step towards the comprehension of the dynamics and distribution of dissolved organic carbon in the open ocean. The results presented proved that the sea surface temperature, salinity and CDOM concentration are not enough to estimate DOC concentration in the open ocean. Thus, other variables such us the Chlorophyll-a concentration and the mixed layer depth should be taken in consideration, as well as the temporal lag between the estimator variation and the respective response of

DOC concentration. Further analyses should be provided to better assess the sensitivity of the model to the different inputs. Preliminary studies tend to demonstrate that MLD seems to be a key factor in the DOC estimation. It should be worth to couple our model with outputs of global physical models which provide MLD at the required time steps.

The estimation of DOC with the new model has been proved to be good in open ocean, while failing in coastal waters, such as in the east coast of USA, where NN29s seem to underestimate the concentration of DOC. Nevertheless, several models have been developed to estimate DOC concentration in open ocean. For example, the model developed by Vantrepotte et al. (2015) which retrieves accurate results in global coastal areas. Hence, by merging both models a precise estimation of DOC in global scale on both environments would be obtained.

An interesting further outcome from this thesis consists in describing the relative contribution of DOC and particulate organic carbon (POC) to the total organic carbon budget (TOC). A few studies have reported that the average relative fraction of POC to to TOC is around 1 to 10% (Kumari and Mohan, 2018; Maciejewska and Pempkowiak 2014; Sanders et al., 2014; Santana-Falcón et al., 2017), depending on the oceanic basins. Therewith, the calculations based on the annual average of DOC estimated with the new model and of POC estimated with the model proposed by Loisel et al. (2002) using the particulate backscattering coefficient as estimated in Loisel et al. (2018) show that the mentioned percentages are even more variable depending on the location (Fig 4.1). It is observed that POC represents on average 5.32 ± 3.75 %, reaching ≥ 10 % only in the coastal regions and in the frontal area that delimits the Southern Ocean. Thus, while the greatest fraction of the organic carbon in the surface of the open ocean is dissolved, its contribution to TOC is spatially variable. The origin of such variability, as well as the temporal variability, should be further analyzed. The North Atlantic area, characterized by a strong winter mixing as well as an intense spring phytoplankton bloom, would represent a good working area to start analyzing the respective dynamics between POC and DOC. This area is also a greatly sampled by previous and present large oceanic programs.

Built on this, a following step would be to obtain the integrate DOC concentration over the euphotic layer (DOC_{zeu}) through the determination of its relationship with surface

DOC (DOC_{surf}). This exercise has been performed for POC concentrations by Duforêt-Gaurier et al. (2010) based on statistical analysis of vertical profiles.

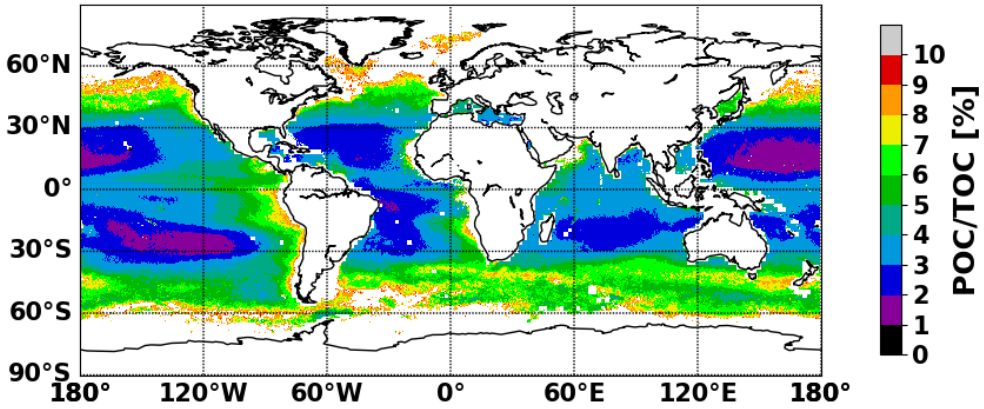


Fig 4.1 Global map of the annual average relative contribution of POC to TOC produced with 8 days composite data from 2002 to 2012, with a spatial resolution of 25 km.

A fraction of the DOC that accumulates in the gyres and poles is an indirect product of primary production. Meanwhile, the primary producers are the main direct source of POC in the ocean, reason why POC and chlorophyll a distribution present similar features. Consequently, POC/TOC ratio presents the lowest values in the subtropical gyres. In the poles on the other hand, where high concentrations of DOC and POC are expected, POC seems to take more importance in TOC, representing more than 10% of it. Thus, it would be interesting to inquire into the relationship between DOC and chlorophyll a, focusing specifically on its variability in space and time. This is of special interest since chlorophyll a is the extra input variable required to estimate DOC when the water mass belong to classes 1 to 9.

References

Abbott, M. R., J. G. Richman, R. M. Letelier, and J. S. Bartlett. 2000. “The Spring Bloom in the Antarctic Polar Frontal Zone as Observed from a Mesoscale Array of Bio-Optical Sensors.” *Deep-Sea Research Part II: Topical Studies in Oceanography* 47 (15–16): 3285–3314. doi:10.1016/S0967-0645(00)00069-2.

Abbott, Mark R., and Philip M. Zion. 1985. “Satellite Observations of Phytoplankton Variability during an Upwelling Event.” *Continental Shelf Research* 4 (6): 661–680. doi:10.1016/0278-4343(85)90035-4.

Abell, Jeffrey, Steve Emerson, and Paul Renaud. 2000. “Distributions of TOP, TON and TOC in the North Pacific Subtropical Gyre: Implications for Nutrient Supply in the Surface Ocean and Remineralization in the Upper Thermocline.” *Journal of Marine Research* 58 (2): 203–222. doi:10.1357/002224000321511142.

Akaike, H. 1974. “A New Look at the Statistical Model Identification.” *IEEE Transactions on Automatic Control* 19 (6): 716–723. doi:10.1109/TAC.1974.1100705.

Aluwihare, Lihini I., Daniel J. Repeta, and Robert F. Chen. 1997. “A Major Biopolymeric Component to Dissolved Organic Carbon in Surface Sea Water.” *Nature* 387 (6629): 166–169. doi:10.1038/387166a0.

Amon, Rainer M.W., and Ronald Benner. 1996. “Bacterial Utilization of Different Size Classes of Dissolved Organic Matter.” *Limnology and Oceanography* 41 (1): 41–51. doi:10.4319/lo.1996.41.1.0041.

Anderson, Thomas R., and Hugh W. Ducklow. 2001. “Microbial Loop Carbon Cycling in Ocean Environments Studied Using a Simple Steady-State Model.” *Aquatic Microbial Ecology* 26 (1): 37–49. doi:10.3354/ame026037.

Armstrong, Robert A., Cindy Lee, John I. Hedges, Susumu Honjo, and Stuart G. Wakeham. 2002. “A New, Mechanistic Model for Organic Carbon Fluxes in the Ocean Based on the Quantitative Association of POC with Ballast Minerals.” *Deep*

Sea Research Part II: Topical Studies in Oceanography 49 (1–3): 219–236. doi:10.1016/S0967-0645(01)00101-1.

Arrigo, Kevin R., and Christopher W. Brown. 1996. "Impact of Chromophoric Dissolved Organic Matter on UV Inhibition of Primary Productivity in the Sea." *Marine Ecology Progress Series* 140: 207–216.

Aumont, O., C. Ethé, A. Tagliabue, L. Bopp, and M. Gehlen. 2015. "PISCES-v2: An Ocean Biogeochemical Model for Carbon and Ecosystem Studies." *Geoscientific Model Development* 8 (8): 2465–2513. doi:10.5194/gmd-8-2465-2015.

Aurin, Dirk, Antonio Mannino, and David Lary. 2018. "Remote Sensing of CDOM, CDOM Spectral Slope, and Dissolved Organic Carbon in the Global Ocean." *Applied Sciences* 8 (12): 2687. doi:10.3390/app8122687.

Babin, Marcel, Dariusz Stramski, Giovanni M. Ferrari, Herve Claustre, Annick Bricaud, Grigor Obolensky, and Nicolas Hoepffner. 2003. "Variations in the Light Absorption Coefficients of Phytoplankton, Nonalgal Particles, and Dissolved Organic Matter in Coastal Waters around Europe." *Journal of Geophysical Research: Oceans* 108 (7). doi:10.1029/2001jc000882.

Bailey, Sean, and Menghua Wang. 2001. "Satellite Aerosol Optical Thickness Match-Up Procedures." In *NASA Tech. Memo., 2001-209982,* edited by Giulietta S. Fargion, Robert Barnes, and Charles McClain, 70–72. Greenbelt: Goddard Space Flight Space Center.

Balch, W. M., D. T. Drapeau, B. C. Bowler, E. R. Lyczkowski, L. C. Lubelczyk, S. C. Painter, and A. J. Poulton. 2014. "Surface Biological, Chemical, and Optical Properties of the Patagonian Shelf Coccolithophore Bloom, the Brightest Waters of the Great Calcite Belt." *Limnology and Oceanography* 59 (5): 1715–1732. doi:10.4319/lo.2014.59.5.1715.

Barber, Richard T. 1968. "Dissolved Organic Carbon from Deep Water Resists Microbial Oxidation." *Nature* 220: 274–275.

Bauer, James E., Peter M. Williams, and Ellen R.M. Druffel. 1992. "14C Activity of Dissolved Organic Carbon Fractions in the North-Central Pacific and Sargasso Sea."

Nature 357 (6380): 667–670. doi:10.1038/357667a0.

Bélanger, Simon, Huixiang Xie, Nickolay Krotkov, Pierre Larouche, Warwick F. Vincent, and Marcel Babin. 2006. "Photomineralization of Terrigenous Dissolved Organic Matter in Arctic Coastal Waters from 1979 to 2003: Interannual Variability and Implications of Climate Change." *Global Biogeochemical Cycles* 20 (4): n/a-n/a. doi:10.1029/2006GB002708.

Benner, R., J. D. Pakulski, M. McCarthy, J. I. Hedges, and P. G. Hatcher. 1992. "Bulk Chemical Characteristics of Dissolved Organic Matter in the Ocean." *Science* 255 (5051): 1561–1564. doi:10.1126/science.255.5051.1561.

Benner, Ronald, and Bopaiah Biddanda. 1998. "Photochemical Transformations of Surface and Deep Marine Dissolved Organic Matter: Effects on Bacterial Growth." *Limnology and Oceanography* 43 (6): 1373–1378. doi:10.4319/lo.1998.43.6.1373.

Blough, Niel V., and Rossana Del Vecchio. 2002. "Chromophoric DOM in the Coastal Environment." In *Biogeochemistry of Marine Dissolved Organic Matter*, edited by Dennis A. Hansell and Craig A. Carlson, 509–546. San Diego.

Boss, Emmanuel, W. Scott Pegau, J. Ron V. Zaneveld, and Andrew H. Barnard. 2001. "Spatial and Temporal Variability of Absorption by Dissolved Material at a Continental Shelf." *Journal of Geophysical Research: Oceans*. doi:10.1029/2000jc900008.

Boss, Emmanuel, and Collin Roesler. 2006. "Over Constrained Linear Matrix Inversion with Statistical Selection." In *Remote Sensing of Inherent Optical Properties: Fundamentals, Test of Algorithms, and Applications, Reports of the International Ocean-Colour Coor- Dinating Groups*, edited by Z.P. Lee, 5:57–62. Dartmouth: IOCCG.

Boutin, J., J.-L. Vergely, J. Koehler, F. Rouffi, and N. Reul. 2019. "ESA Sea Surface Salinity Climate Change Initiative (Sea_Surface_Salinity_cci): Monthly Sea Surface Salinity Product v1.8." Centre for Environmental Data Analysis. doi:10.5285/9ef0ebf847564c2eabe62cac4899ec41.

Bragg, Jason G., Stephanie Dutkiewicz, Oliver Jahn, Michael J. Follows, and Sallie W.

Chisholm. 2010. “Modeling Selective Pressures on Phytoplankton in the Global Ocean.” *PLoS ONE* 5 (3): 1–6. doi:10.1371/journal.pone.0009569.

Bricaud, A., A. M. Ciotti, and B. Gentili. 2012. “Spatial-Temporal Variations in Phytoplankton Size and Colored Detrital Matter Absorption at Global and Regional Scales, as Derived from Twelve Years of SeaWiFS Data (1998-2009).” *Global Biogeochemical Cycles* 26 (1): 1–17. doi:10.1029/2010GB003952.

Bricaud, A., C. S. Roesler, J. S. Parslow, and J. Ishizaka. 2002. “Bio-Optical Studies during the JGOFS-Equatorial Pacific Program: A Contribution to the Knowledge of the Equatorial System.” *Deep-Sea Research Part II: Topical Studies in Oceanography* 49 (13–14): 2583–2599. doi:10.1016/S0967-0645(02)00049-8.

Bricaud, Annick, Marcel Babin, Hervé Claustre, Joséphine Ras, and Fanny Tièche. 2010. “Light Absorption Properties and Absorption Budget of Southeast Pacific Waters.” *Journal of Geophysical Research: Oceans* 115 (8): 1–19. doi:10.1029/2009JC005517.

Bricaud, Annick, Andre Morel, and Louis Prieur. 1981. “Absorption by Dissolved Organic Matter of the Sea (Yellow Substance) in the UV and Visible Domains.” *Limnology and Oceanography* 26 (1): 43–53. doi:10.4319/lo.1981.26.1.0043.

Buchan, Alison, Gary R. LeCleir, Christopher A. Gulvik, and José M. González. 2014. “Master Recyclers: Features and Functions of Bacteria Associated with Phytoplankton Blooms.” *Nature Reviews. Microbiology* 12 (10). Nature Publishing Group: 686–698. doi:10.1038/nrmicro3326.

Cao, Fang, Maria Tzortziou, Chuanmin Hu, Antonio Mannino, Cédric G. Fichot, Rossana Del Vecchio, Raymond G. Najjar, and Michael Novak. 2018. “Remote Sensing Retrievals of Colored Dissolved Organic Matter and Dissolved Organic Carbon Dynamics in North American Estuaries and Their Margins.” *Remote Sensing of Environment* 205 (February). Elsevier: 151–165. doi:10.1016/j.rse.2017.11.014.

Carder, K. L., S. K. Hawes, K. A. Baker, R. C. Smith, R. G. Steward, and B. G. Mitchell. 1991. “Reflectance Model for Quantifying Chlorophyll a in the Presence of Productivity Degradation Products.” *Journal of Geophysical Research* 96 (C11). doi:10.1029/91jc02117.

Carder, Kendall L., Robert G. Steward, George R. Harvey, and Peter B. Ortner. 1989. "Marine Humic and Fulvic Acids: Their Effects on Remote Sensing of Ocean Chlorophyll." *Limnology and Oceanography* 34 (1): 68–81. doi:10.4319/lo.1989.34.1.0068.

Carlson, Craig A. 2002. "Production and Removal Processes." In *Biogeochemistry of Marine Dissolved Organic Matter*, edited by Dennis A. Hansell and Craig A. Carlson, 9–151. San Diego: Academic Press.

Carlson, Craig A., Hugh W. Ducklow, and Thomas D. Sleeter. 1996. "Stocks and Dynamics of Bacterioplankton in the Northwestern Sargasso Sea." *Deep-Sea Research Part II: Topical Studies in Oceanography* 43 (2–3): 491–515. doi:10.1016/0967-0645(95)00101-8.

Carlson, Craig A., Dennis A. Hansell, Norman B. Nelson, David A. Siegel, William M. Smethie, Samar Khatiwala, Meredith M. Meyers, and Elisa Halewood. 2010. "Dissolved Organic Carbon Export and Subsequent Remineralization in the Mesopelagic and Bathypelagic Realms of the North Atlantic Basin." *Deep-Sea Research Part II: Topical Studies in Oceanography* 57 (16). Elsevier: 1433–1445. doi:10.1016/j.dsr2.2010.02.013.

Carlson, Craig A, Hugh W Ducklowt, and Anthony F Michaelst. 1994. "Annual Flux of Dissolved Organic Carbon from the Euphotic Zone in the Northwestern Sargasso Sea" 371 (September): 405–408.

Carter, Charles W., and Irwin H. Suffet. 1982. "Binding of DDT to Dissolved Humic Materials." *Environmental Science & Technology* 16 (11): 735–740. doi:10.1021/es00105a003.

Chen, Jun, Xianqiang He, Bin Zhou, and Delu Pan. 2017. "Deriving Colored Dissolved Organic Matter Absorption Coefficient from Ocean Color with a Neural Quasi-Analytical Algorithm." *Journal of Geophysical Research: Oceans* 122 (11): 8543–8556. doi:10.1002/2017JC013115.

Chen, Jun, Zhongping Lee, Chuanmin Hu, and Jianwei Wei. 2016. "Improving Satellite Data Products for Open Oceans with a Scheme to Correct the Residual Errors in Remote Sensing Reflectance." *Journal of Geophysical Research: Oceans* 121 (6):

3866–3886. doi:10.1002/2016JC011673.

Chen, Robert F, and Jeffrey L Bada. 1992. “The Fluorescence of Dissolved Organic Matter in Seawater.” *Marine Chemistry* 37 (3–4): 191–221. doi:10.1016/0304-4203(92)90078-O.

Christian, James R., and Thomas R. Anderson. 2002. “Modeling DOM Biogeochemistry.” In *Biogeochemistry of Marine Dissolved Organic Matter*, edited by Dennis A. Hansell and Craig A. Carlson, 717–755. San Diego: Academic Press.

Church, Matthew J., Michael W. Lomas, and Frank Muller-Karger. 2013. “Sea Change: Charting the Course for Biogeochemical Ocean Time-Series Research in a New Millennium.” *Deep Sea Research Part II: Topical Studies in Oceanography* 93 (September): 2–15. doi:10.1016/j.dsr2.2013.01.035.

Ciotti, Aurea M., and Annick Bricaud. 2006. “Retrievals of a Size Parameter for Phytoplankton and Spectral Light Absorption by Colored Detrital Matter from Water-Leaving Radiances at SeaWiFS Channels in a Continental Shelf Region off Brazil.” *Limnology and Oceanography: Methods* 4 (7): 237–253. doi:10.4319/lom.2006.4.237.

Claustre, H., A. Morel, S. B. Hooker, M. Babin, D. Antoine, K. Oubelkheir, A. Bricaud, K. Leblanc, B. Quéguiner, and S. Maritorena. 2002. “Is Desert Dust Making Oligotrophic Waters Greener?” *Geophysical Research Letters* 29 (10): 107-1-107–4. doi:10.1029/2001gl014056.

Coble, Paula G. 1996. “Characterization of Marine and Terrestrial DOM in Seawater Using Excitation-Emission Matrix Spectroscopy.” *Marine Chemistry* 51 (4): 325–346. doi:10.1016/0304-4203(95)00062-3.

Coble, Paula G. 2007. “Marine Optical Biogeochemistry: The Chemistry of Ocean Color.” *Chemical Reviews* 107 (2): 402–418. doi:10.1021/cr050350+.

Coble, Paula G., Carlos E. Del Castillo, and Bernard Avril. 1998. “Distribution and Optical Properties of CDOM in the Arabian Sea during the 1995 Southwest Monsoon.” *Deep Sea Research Part II: Topical Studies in Oceanography* 45 (10–11): 2195–2223. doi:10.1016/S0967-0645(98)00068-X.

Coble, Paula G., Sarah A. Green, Neil V. Blough, and Robert B. Gagosian. 1990. "Characterization of Dissolved Organic Matter in the Black Sea by Fluorescence Spectroscopy." *Nature* 348 (6300): 432–435. doi:10.1038/348432a0.

Coble, Paula, Chuanmin Hu, Richard W. Gould, Grace Chang, and A. Michelle Wood. 2004. "Colored Dissolved Organic Matter in the Coastal Ocean: An Optical Tool for Coastal Zone Environmental Assessment and Management." *Oceanography* 17 (SPL.ISS. 2): 50–59. doi:10.5670/oceanog.2004.47.

Copin-Montégut, Gérard, and Bernard Avril. 1993. "Vertical Distribution and Temporal Variation of Dissolved Organic Carbon in the North-Western Mediterranean Sea." *Deep-Sea Research Part I* 40 (10): 1963–1972. doi:10.1016/0967-0637(93)90041-Z.

Crutzen, P.J., and E.F. Stoermer. 2000. "The 'Anthropocene.'" *IGBP Newsletter* 41: 17–18.

D'Sa, E. J., C. Hu, F. E. Muller-Karger, and K. L. Carder. 2002. "Estimation of Colored Dissolved Organic Matter and Salinity Fields in Case 2 Waters Using Sea WiFS: Examples from Florida Bay and Florida Shelf." *Proceedings of the Indian Academy of Sciences, Earth and Planetary Sciences* 111 (3): 197–207. doi:10.1007/BF02701966.

D'Sa, Eurico J., Robert G. Steward, Anthony Vodacek, Neil V. Blough, and Dave Phinney. 1999. "Determining Optical Absorption of Colored Dissolved Organic Matter in Seawater with a Liquid Capillary Waveguide." *Limnology and Oceanography* 44 (4): 1142–1148. doi:10.4319/lo.1999.44.4.1142.

De La Rocha, Christina L., and Uta Passow. 2007. "Factors Influencing the Sinking of POC and the Efficiency of the Biological Carbon Pump." *Deep Sea Research Part II: Topical Studies in Oceanography* 54 (5–7): 639–658. doi:10.1016/j.dsr2.2007.01.004.

Del Castillo, Carlos E., and Richard L. Miller. 2008. "On the Use of Ocean Color Remote Sensing to Measure the Transport of Dissolved Organic Carbon by the Mississippi River Plume." *Remote Sensing of Environment* 112 (3): 836–844. doi:10.1016/j.rse.2007.06.015.

Dittmar, Thorsten, and Gerhard Kattner. 2003. “Recalcitrant Dissolved Organic Matter in the Ocean: Major Contribution of Small Amphiphilics.” *Marine Chemistry* 82 (1–2): 115–123. doi:10.1016/S0304-4203(03)00068-9.

Doval, M.D, and Dennis A Hansell. 2000. “Organic Carbon and Apparent Oxygen Utilization in the Western South Pacific and the Central Indian Oceans.” *Marine Chemistry* 68 (3): 249–264. doi:10.1016/S0304-4203(99)00081-X.

Druffel, Ellen R. M., Peter M. Williams, James E. Bauer, and John R. Ertel. 1992. “Cycling of Dissolved and Particulate Organic Matter in the Open Ocean.” *Journal of Geophysical Research* 97 (C10): 15639. doi:10.1029/92JC01511.

Druffel, Ellen R.M., and Peter M. Williams. 1990. “Identification of a Deep Marine Source of Particulate Organic Carbon Using Bomb 14C.” *Nature* 347 (6289): 172–174. doi:10.1038/347172a0.

Druffel, Ellen R M, Peter M Williams, K E N Robertson, Sheila Griffin, A J T Jullt, Douglas Donahuet, Lawrence Toolint, and T W Linickt. 1989. “Radiocarbon in Dissolved Organic and Inorganic.” *Radiocarbon* 31 (3): 523–532.

Ducklow, Hugh W. 1999. “The Bacterial Component of the Oceanic Euphotic Zone.” *FEMS Microbiology Ecology* 30 (1): 1–10. doi:10.1111/j.1574-6941.1999.tb00630.x.

Ducklow, Hugh W., Deborah K. Steinberg, and Ken O. Buesseler. 2001. “Upper Ocean Carbon Export and the Biological Pump.” *Oceanography* 14 (SPL.ISS. 4): 50–58. doi:10.5670/oceanog.2001.06.

Duforêt-Gaurier, L., H. Loisel, D. Dessailly, K. Nordkvist, and S. Alvain. 2010. “Estimates of Particulate Organic Carbon over the Euphotic Depth from in Situ Measurements. Application to Satellite Data over the Global Ocean.” *Deep-Sea Research Part I: Oceanographic Research Papers* 57 (3): 351–367. doi:10.1016/j.dsr.2009.12.007.

Dunne, John P., Jorge L. Sarmiento, and Anand Gnanadesikan. 2007. “A Synthesis of Global Particle Export from the Surface Ocean and Cycling through the Ocean Interior and on the Seafloor.” *Global Biogeochemical Cycles* 21 (4): n/a-n/a.

doi:10.1029/2006GB002907.

Dutkiewicz, Stephanie, Mick Follows, John Marshall, and Watson W. Gregg. 2001. "Interannual Variability of Phytoplankton Abundances in the North Atlantic." *Deep-Sea Research Part II: Topical Studies in Oceanography* 48 (10): 2323–2344. doi:10.1016/S0967-0645(00)00178-8.

Engel, Anja, Slike Thoms, Ulf Riabesell, Emma Rochelle-Newall, and Ingrid Zondervan. 2004. "Polysaccharide Aggregation as a Potential Sink of Marine Dissolved Organic Carbon." *Nature* 428 (6986): 929–932. doi:10.1038/nature02453.

Etheridge, D M, L P Steele, R L Langenfelds, and R J Francey. 1996. "Natural and Anthropogenic Changes in Atmospheric CO 2 Over." *Journal of Geophysical Research* 101 (95): 4115–4128.

Evers-King, Hayley, Victor Martinez-Vicente, Robert J.W. Brewin, Giorgio Dall'Olmo, Anna E. Hickman, Thomas Jackson, Tihomir S. Kostadinov, et al. 2017. "Validation and Intercomparison of Ocean Color Algorithms for Estimating Particulate Organic Carbon in the Oceans." *Frontiers in Marine Science* 4 (AUG): 1–20. doi:10.3389/fmars.2017.00251.

Fichot, Cédric G., and Ronald Benner. 2011. "A Novel Method to Estimate DOC Concentrations from CDOM Absorption Coefficients in Coastal Waters." *Geophysical Research Letters* 38 (3). doi:10.1029/2010GL046152.

Fichot, Cédric G., and Ronald Benner. 2012. " The Spectral Slope Coefficient of Chromophoric Dissolved Organic Matter (S 275-295) as a Tracer of Terrigenous Dissolved Organic Carbon in River-Influenced Ocean Margins ." *Limnology and Oceanography* 57 (5): 1453–1466. doi:10.4319/lo.2012.57.5.1453.

Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto, and Bor-Chung Chen. 1998. "New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program." *Journal of Business & Economic Statistics* 16 (2): 127. doi:10.2307/1392565.

Fuhrman, JA, and RL Ferguson. 1986. "Nanomolar Concentrations and Rapid Turnover of Dissolved Free Amino Acids in Seawater: Agreement between Chemical and

Microbiological Measurements." *Marine Ecology Progress Series* 33: 237–242. doi:10.3354/meps033237.

Gaillard, Fabienne, Thierry Reynaud, Virginie Thierry, Nicolas Kolodziejczyk, and Karina von Schuckmann. 2016. "In Situ–Based Reanalysis of the Global Ocean Temperature and Salinity with ISAS: Variability of the Heat Content and Steric Height." *Journal of Climate* 29 (4): 1305–1323. doi:10.1175/JCLI-D-15-0028.1.

Gardner, W. D., A. V. Mishonov, and M. J. Richardson. 2006. "Global POC Concentrations from In-Situ and Satellite Data." *Deep-Sea Research Part II: Topical Studies in Oceanography* 53 (5–7): 718–740. doi:10.1016/j.dsr2.2006.01.029.

Goldberg, Stuart J., Craig A. Carlson, Dennis A. Hansell, Norm B. Nelson, and David A. Siegel. 2009. "Temporal Dynamics of Dissolved Combined Neutral Sugars and the Quality of Dissolved Organic Matter in the Northwestern Sargasso Sea." *Deep-Sea Research Part I: Oceanographic Research Papers* 56 (5): 672–685. doi:10.1016/j.dsr.2008.12.013.

Goyens, C., C. Jamet, and K. G. Ruddick. 2013. "Spectral Relationships for Atmospheric Correction I Validation of Red and near Infra-Red Marine Reflectance Relationships." *Optics Express* 21 (18): 21162. doi:10.1364/oe.21.021162.

Green, Sarah A., and Neil V. Blough. 1994. "Optical Absorption and Fluorescence Properties of Chromophoric Dissolved Organic Matter in Natural Waters." *Limnology and Oceanography* 39 (8): 1903–1916. doi:10.4319/lo.1994.39.8.1903.

Griffith, D. R., A. P. McNichol, L. Xu, F. A. McLaughlin, R. W. Macdonald, K. A. Brown, and T. I. Eglinton. 2012. "Carbon Dynamics in the Western Arctic Ocean: Insights from Full-Depth Carbon Isotope Profiles of DIC, DOC, and POC." *Biogeosciences* 9 (3): 1217–1224. doi:10.5194/bg-9-1217-2012.

Hansell, D. A., and C. A. Carlson. 2015. "Dissolved Organic Matter in the Ocean Carbon Cycle." *Eos (United States)* 96 (15): 8–12. doi:10.1029/2015eo033011.

Hansell, Dennis. 2001. "Marine Dissolved Organic Matter and the Carbon Cycle." *Oceanography* 14 (4): 41–49. doi:10.5670/oceanog.2001.05.

Hansell, Dennis A. 2002. “DOC in the Global Ocean Carbon Cycle.” In *Biogeochemistry of Marine Dissolved Organic Matter*, edited by Dennis A. Hansell and Craig A. Carlson, 685–715. San Diego: Academic Press.

Hansell, Dennis A. 2013. “Recalcitrant Dissolved Organic Carbon Fractions.” *Annual Review of Marine Science* 5: 421–445. doi:10.1146/annurev-marine-120710-100757.

Hansell, Dennis A., and Craig A. Carlson. 1998a. “Deep-Ocean Gradients in the Concentration of Dissolved Organic Carbon.” *Nature* 395 (6699): 263–266. doi:10.1038/26200.

Hansell, Dennis A., and Craig A. Carlson. 1998b. “Net Community Production of Dissolved Organic Carbon.” *Global Biogeochemical Cycles* 12 (3): 443–453. doi:10.1029/98GB01928.

Hansell, Dennis A., and Craig A. Carlson. 2001. “Biogeochemistry of Total Organic Carbon and Nitrogen in the Sargasso Sea: Control by Convective Overturn.” *Deep-Sea Research Part II: Topical Studies in Oceanography* 48 (8–9): 1649–1667. doi:10.1016/S0967-0645(00)00153-3.

Hansell, Dennis A., Craig A. Carlson, Daniel J. Repeta, and Reiner Schlitzer. 2009. “Dissolved Organic Matter in the Ocean a Controversy Stim Ulates New Insights.” *Oceanography* 22 (SPL.ISS. 4): 202–211. doi:10.5670/oceanog.2009.109.

Hansell, Dennis A., Craig A. Carlson, and Yoshimi Suzuki. 2002. “Dissolved Organic Carbon Export with North Pacific Intermediate Water Formation.” *Global Biogeochemical Cycles*. doi:10.1029/2000gb001361.

Hansell, Dennis A., and Edward T. Peltzer. 1998. “Spatial and Temporal Variations of Total Organic Carbon in the Arabian Sea.” *Deep Sea Research Part II: Topical Studies in Oceanography* 45 (10–11): 2171–2193. doi:10.1016/S0967-0645(98)00067-8.

Hansell, Dennis A, Craig A Carlson, and Yoshimi Suzuki. 2002. “Dissolved Organic Carbon Export with North Pacific Intermediate Water Formation.” doi:10.1029/2000GB001361.

Hansell, Dennis, Craig Carlson, Daniel Repeta, and Reiner Schlitzer. 2009. "Dissolved Organic Matter in the Ocean: A Controversy Stimulates New Insights." *Oceanography* 22 (4): 202–211. doi:10.5670/oceanog.2009.109.

Hanson, C.E., C.B. Pattiaratchi, and A.M. Waite. 2005. "Sporadic Upwelling on a Downwelling Coast: Phytoplankton Responses to Spatially Variable Nutrient Dynamics off the Gascoyne Region of Western Australia." *Continental Shelf Research* 25 (12–13): 1561–1582. doi:10.1016/j.csr.2005.04.003.

Harvey, George R., and Deborah A. Boran. 1985. "Geochemistry of Humic Substances in Seawater." In *Humic Substances in Soil Sediment and Water*, edited by G. R. Aiken, D. M. McKnight, R. L. Wershaw, and R MacCarthy, 223–247. Wiley.

Harvey, George R., Deborah A. Boran, Larry A. Chesal, and John M. Tokar. 1983. "The Structure of Marine Fulvic and Humic Acids." *Marine Chemistry* 12 (2–3): 119–132. doi:10.1016/0304-4203(83)90075-0.

Hedges, J.I. 2002. "Why Dissolved Organic Matter?" In *Biogeochemistry of Marine Dissolved Organic Matter*, edited by D. A. Hansell and Craig A. Carlson. San Diego: Academic Press.

Hirata, T., N. J. Hardman-Mountford, R. J.W. Brewin, J. Aiken, R. Barlow, K. Suzuki, T. Isada, et al. 2011. "Synoptic Relationships between Surface Chlorophyll-a and Diagnostic Pigments Specific to Phytoplankton Functional Types." *Biogeosciences* 8 (2): 311–327. doi:10.5194/bg-8-311-2011.

Hitchcock, Gary L, Robert F Chen, G.Bernard Gardner, and William J Wiseman. 2004. "A Lagrangian View of Fluorescent Chromophoric Dissolved Organic Matter Distributions in the Mississippi River Plume." *Marine Chemistry* 89 (1–4): 225–239. doi:10.1016/j.marchem.2004.03.018.

Holte, James, and Lynne Talley. 2009. "A New Algorithm for Finding Mixed Layer Depths with Applications to Argo Data and Subantarctic Mode Water Formation*." *Journal of Atmospheric and Oceanic Technology* 26 (9): 1920–1939. doi:10.1175/2009JTECHO543.1.

Holte, James, Lynne D. Talley, John Gilson, and Dean Roemmich. 2017. "An Argo

Mixed Layer Climatology and Database." *Geophysical Research Letters* 44 (11): 5618–5626. doi:10.1002/2017GL073426.

Honda, Makio C., Masahide Wakita, Kazuhiko Matsumoto, Tetsuichi Fujiki, Eko Siswanto, Kosei Sasaoka, Hajime Kawakami, et al. 2017. "Comparison of Carbon Cycle between the Western Pacific Subarctic and Subtropical Time-Series Stations: Highlights of the K2S1 Project." *Journal of Oceanography* 73 (5). Springer Japan: 647–667. doi:10.1007/s10872-017-0423-3.

Hopkinson, Charles S., and Joseph J. Vallino. 2005. "Efficient Export of Carbon to the Deep Ocean through Dissolved Organic Matter." *Nature* 433 (7022): 142–145. doi:10.1038/nature03191.

Hu, Chuanmin, Zhongping Lee, and Bryan Franz. 2012. "Chlorophyll a Algorithms for Oligotrophic Oceans: A Novel Approach Based on Three-Band Reflectance Difference." *Journal of Geophysical Research: Oceans* 117 (1): 1–25. doi:10.1029/2011JC007395.

Indermühle, A., T. F. Stocker, F. Joos, H. Fischer, H. J. Smith, M. Wahlen, B. Deck, et al. 1999. "Holocene Carbon-Cycle Dynamics Based on CO2 Trapped in Ice at Taylor Dome, Antarctica." *Nature* 398 (6723): 121–126. doi:10.1038/18158.

Indermühle, Andreas, Eric Monnin, Bernhard Stauffer, Thomas F. Stocker, and Martin Wahlen. 2000. "Atmospheric CO 2 Concentration from 60 to 20 Kyr BP from the Taylor Dome Ice Core, Antarctica." *Geophysical Research Letters* 27 (5): 735–738. doi:10.1029/1999GL010960.

IOCCG. 2006. *IOCCG Report Number 05: Reports of the International Ocean-Colour Coordinating Group Remote Sensing of Inherent Optical Properties : Fundamentals, Tests of Algorithms, and Applications*. *IOCCG Report 5*. Vol. 5.

Ito, Gen, Anastasia Romanou, Nancy Y. Kiang, Gregory Faluvegi, Igor Aleinov, Reto Ruedy, Gary Russell, Paul Lerner, Maxwell Kelley, and Ken Lo. 2020. "Global Carbon Cycle and Climate Feedbacks in the NASA GISS ModelE2.1." *Journal of Advances in Modeling Earth Systems* 12 (10): 1–44. doi:10.1029/2019MS002030.

Jamet, C., H. Loisel, and D. Dessailly. 2012. " Retrieval of the Spectral Diffuse

Attenuation Coefficient K d (λ) in Open and Coastal Ocean Waters Using a Neural Network Inversion .” *Journal of Geophysical Research: Oceans* 117 (C10): n/a-n/a. doi:10.1029/2012jc008076.

Jenne, E.A. 1975. “Sorption Phenomenon in Organics of Bottom Sediments (Reimers et Al.).” In Heavy Metals in the Aquatic Environment, 131–136. Elsevier. doi:10.1016/B978-0-08-018068-7.50023-X.

Jerlov, N.G. 1976. “Marine Optics.” Elsevier Oceanography Series, Elsevier, Amsterdam, 230.

Jiao, Nianzhi, Gerhard J. Herndl, Dennis A. Hansell, Ronald Benner, Gerhard Kattner, Steven W. Wilhelm, David L. Kirchman, et al. 2010. “Microbial Production of Recalcitrant Dissolved Organic Matter: Long-Term Carbon Storage in the Global Ocean.” *Nature Reviews Microbiology* 8 (8). Nature Publishing Group: 593–599. doi:10.1038/nrmicro2386.

Johannessen, S. C., W. L. Miller, and J. J. Cullen. 2003. “Calculation of UV Attenuation and Colored Dissolved Organic Matter Absorption Spectra from Measurements of Ocean Color.” *Journal of Geophysical Research: Oceans* 108 (9). doi:10.1029/2000jc000514.

Juhls, Bennet, Pier Paul Overduin, Jens Hölemann, Martin Hieronymi, Atsushi Matsuoka, Birgit Heim, and Jürgen Fischer. 2019. “Dissolved Organic Matter at the Fluvial-Marine Transition in the Laptev Sea Using in Situ Data and Ocean Colour Remote Sensing.” *Biogeosciences* 16 (13): 2693–2713. doi:10.5194/bg-16-2693-2019.

Kahru, Mati, and B. Greg Mitchell. 2001. “Seasonal and Nonseasonal Variability of Satellite-Derived Chlorophyll and Colored Dissolved Organic Matter Concentration in the California Current.” *Journal of Geophysical Research: Oceans* 106 (C2): 2517–2529. doi:10.1029/1999JC000094.

Kandasamy, Selvaraj, and Bejugam Nagender Nath. 2016. “Perspectives on the Terrestrial Organic Matter Transport and Burial along the Land-Deep Sea Continuum: Caveats in Our Understanding of Biogeochemical Processes and Future Needs.” *Frontiers in Marine Science* 3 (DEC). doi:10.3389/fmars.2016.00259.

Karl, David M., and Roger Lukas. 1996. “The Hawaii Ocean Time-Series (HOT) Program: Background, Rationale and Field Implementation.” *Deep-Sea Research Part II: Topical Studies in Oceanography* 43 (2–3): 129–156. doi:10.1016/0967-0645(96)00005-7.

Keeling, C. D., and T. P. Whorf. 1998. *Atmospheric CO2 Concentration—Mauna Loa Observatory, Hawaii, 1958–1997*. Oak Ridge, TN.

Keil, R.G., and D.L. Kirchman. 1999. “Utilization of Dissolved Protein and Amino Acids in the Northern Sargasso Sea.” *Aquatic Microbial Ecology* 18: 293–300.

Keil, Richard G., E. Tsamakis, and J. I. Hedges. 2000. “Early Diagenesis of Particulate Amino Acids in Marine Systems.” In *Perspectives in Amino Acid and Protein Geochemistry*, edited by G. A. Goodfriend, M. J. Collins, M. L. Fogel, S. A. Macko, and J. F. Wehmiller, 69–82. Oxford University Press.

Key, R. M., A. Kozyr, C. L. Sabine, K. Lee, R. Wanninkhof, J. L. Bullister, R. A. Feely, F. J. Millero, C. Mordy, and T.-H. Peng. 2004. “A Global Ocean Carbon Climatology: Results from Global Data Analysis Project (GLODAP).” *Global Biogeochemical Cycles* 18 (4): n/a-n/a. doi:10.1029/2004GB002247.

Kieber, David J., Jianfu Jiao, Ronald P. Kiene, and Timothy S. Bates. 1996. “Impact of Dimethylsulfide Photochemistry on Methyl Sulfur Cycling in the Equatorial Pacific Ocean.” *Journal of Geophysical Research: Oceans* 101 (C2): 3715–3722. doi:10.1029/95JC03624.

Kingma, Diederik P., and Jimmy Lei Ba. 2015. “Adam: A Method for Stochastic Optimization.” *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.

Kolodziejczyk, Nicolas, Annaig Prigent-Mazella, and Fabienne Gaillard. 2017. “ISAS-15 Temperature and Salinity Gridded Fields.” SEANOE. doi:https://doi.org/10.17882/52367.

Kopelevich, O. V., and V. I. Burenkov. 1977. “Relation between the Spectral Values of the Light Absorption Coefficients of Sea Water, Phytoplanktonic Pigments, and the Yellow Substance.” *Oceanology* 17: 278–282.

Kostadinov, Tihomir S., Svetlana Milutinović, Irina Marinov, and Anna Cabré. 2016. "Carbon-Based Phytoplankton Size Classes Retrieved via Ocean Color Estimates of the Particle Size Distribution." *Ocean Science* 12 (2): 561–575. doi:10.5194/os-12-561-2016.

Kulovaara, Maaret, Nina Corin, Peter Backlund, and Jari Tervo. 1996. "Impact of UV254-Radiation on Aquatic Humic Substances." *Chemosphere* 33 (5): 783–790. doi:10.1016/0045-6535(96)00233-0.

Kumari, Radha Karuna, and P. M. Mohan. 2018. "Review on Dissolved Organic Carbon and Particulate Organic Carbon in Marine Environment." *ILMU KELAUTAN: Indonesian Journal of Marine Sciences* 23 (1): 25. doi:10.14710/ik.ijms.23.1.25-36.

Lee, ZhongPing, Kendall L. Carder, and Robert A. Arnone. 2002. "Deriving Inherent Optical Properties from Water Color: A Multiband Quasi-Analytical Algorithm for Optically Deep Waters." *Applied Optics* 41 (27): 5755. doi:10.1364/AO.41.005755.

Lee, ZhongPing, Shaoling Shang, Chuanmin Hu, Marlon Lewis, Robert Arnone, Yonghong Li, and Bertrand Lubac. 2010. "Time Series of Bio-Optical Properties in a Subtropical Gyre: Implications for the Evaluation of Interannual Trends of Biogeochemical Properties." *Journal of Geophysical Research* 115 (C9): C09012. doi:10.1029/2009JC005865.

Lefevre, D., Tamburini C., N. Bhairy, L. Chirurgien, C. Gojak, K. Bernardet, Z. Hafidi, et al. 2016. "ANTARES Observatory Data (Mooring IL07)." SEANOE. doi:https://doi.org/10.17882/47130.

Letscher, Robert T., Angela N. Knapp, Anna K. James, Craig A. Carlson, Alyson E. Santoro, and Dennis A. Hansell. 2015. "Microbial Community Composition and Nitrogen Availability Influence DOC Remineralization in the South Pacific Gyre." *Marine Chemistry* 177: 325–334. doi:10.1016/j.marchem.2015.06.024.

Letscher, Robert T., and J. Keith Moore. 2015. "Preferential Remineralization of Dissolved Organic Phosphorus and Non-Redfield DOM Dynamics in the Global Ocean: Impacts on Marine Productivity, Nitrogen Fixation, and Carbon Export." *Global Biogeochemical Cycles* 29 (3): 325–340. doi:10.1002/2014GB004904.

Lévy, Marina, Yoav Lehahn, Jean Michel André, Laurent Mémery, Hubert Loisel, and Eyal Heifetz. 2005. "Production Regimes in the Northeast Atlantic: A Study Based on Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) Chlorophyll and Ocean General Circulation Model Mixed Layer Depth." *Journal of Geophysical Research C: Oceans* 110 (7): 1–16. doi:10.1029/2004JC002771.

Lochte, Karin, Robert Anderson, Roger Francois, Richard A. Jahnke, Graham Shimmield, and Alexander Vetrov. 2003. "Benthic Processes and the Burial of Carbon." In *Ocean Biogeochemistry: The Role of the Ocean Carbon Cycle in Global Change*, edited by Michael J.R. Fasham, 195–216. Berlin Heidelberg: Springer Verlag.

Loisel, H., V. Vantrepotte, K. Norkvist, X. Mriaux, M. Kheireddine, J. Ras, M. Pujo-Pay, et al. 2011. "Characterization of the Bio-Optical Anomaly and Diurnal Variability of Particulate Matter, as Seen from Scattering and Backscattering Coefficients, in Ultra-Oligotrophic Eddies of the Mediterranean Sea." *Biogeosciences* 8 (11): 3295–3317. doi:10.5194/bg-8-3295-2011.

Loisel, Hubert, Bertrand Lubac, David Dessailly, Lucile Duforet-Gaurier, and Vincent Vantrepotte. 2010. "Effect of Inherent Optical Properties Variability on the Chlorophyll Retrieval from Ocean Color Remote Sensing: An in Situ Approach." *Optics Express* 18 (20): 20949. doi:10.1364/oe.18.020949.

Loisel, Hubert, Jean Marc Nicolas, Pierre Yves Deschamps, and Robert Frouin. 2002. "Seasonal and Inter-Annual Variability of Particulate Organic Matter in the Global Ocean." *Geophysical Research Letters* 29 (24): 10–13. doi:10.1029/2002GL015948.

Loisel, Hubert, Dariusz Stramski, David Dessailly, Cédric Jamet, Linhai Li, and Rick A. Reynolds. 2018. "An Inverse Model for Estimating the Optical Absorption and Backscattering Coefficients of Seawater From Remote-Sensing Reflectance Over a Broad Range of Oceanic and Coastal Marine Environments." *Journal of Geophysical Research: Oceans* 123 (3): 2141–2171. doi:10.1002/2017JC013632.

Loisel, Hubert, Vincent Vantrepotte, David Dessailly, and Xavier Mériaux. 2014. "Assessment of the Colored Dissolved Organic Matter in Coastal Waters from

Ocean Color Remote Sensing." *Optics Express* 22 (11): 13109. doi:10.1364/oe.22.013109.

López, Ramón, Carlos E. Del Castillo, Richard L. Miller, Joseph Salisbury, and Dominik Wisser. 2012. "Examining Organic Carbon Transport by the Orinoco River Using SeaWiFS Imagery." *Journal of Geophysical Research: Biogeosciences* 117 (3): 1–13. doi:10.1029/2012JG001986.

Loureiro, Sofia, Albert Reñé, Esther Garcés, Jordi Camp, and Dolors Vaqué. 2011. "Harmful Algal Blooms (HABs), Dissolved Organic Matter (DOM), and Planktonic Microbial Community Dynamics at a near-Shore and a Harbour Station Influenced by Upwelling (SW Iberian Peninsula)." *Journal of Sea Research* 65 (4): 401–413. doi:10.1016/j.seares.2011.03.004.

Maciejewska, Anna, and Janusz Pempkowiak. 2014a. "DOC and POC in the Water Column of the Southern Baltic. Part I. Evaluation of Factors Influencing Sources, Distribution and Concentration Dynamics of Organic Matter." *Oceanologia*. doi:10.5697/oc.56-3.523.

Maciejewska, Anna, and Janusz Pempkowiak. 2014b. "DOC and POC in the Water Column of the Southern Baltic. Part I. Evaluation of Factors Influencing Sources, Distribution and Concentration Dynamics of Organic Matter* Inflows from North Sea." *OCEANOLOGIA* 56 (3): 523–548. doi:10.5697/oc.56-3.523.

Mann, Michael E., and Lee R. Kump. 2015. *Dire Predictions: Understanding Climate Change*. 2nd ed. Pearson Education, Inc.

Mannino, Antonio, Michael G. Novak, Stanford B. Hooker, Kimberly Hyde, and Dirk Aurin. 2014. "Algorithm Development and Validation of CDOM Properties for Estuarine and Continental Shelf Waters along the Northeastern U.S. Coast." *Remote Sensing of Environment* 152 (2014). Elsevier B.V.: 576–602. doi:10.1016/j.rse.2014.06.027.

Mannino, Antonio, Mary E. Russ, and Stanford B. Hooker. 2008. "Algorithm Development and Validation for Satellite-Derived Distributions of DOC and CDOM in the U.S. Middle Atlantic Bight." *Journal of Geophysical Research: Oceans* 113 (7). doi:10.1029/2007JC004493.

Maritorena, Stéphane, and David A. Siegel. 2005. “Consistent Merging of Satellite Ocean Color Data Sets Using a Bio-Optical Model.” *Remote Sensing of Environment* 94 (4): 429–440. doi:10.1016/j.rse.2004.08.014.

Maritorena, Stéphane, David A. Siegel, and Alan R. Peterson. 2002. “Optimization of a Semianalytical Ocean Color Model for Global-Scale Applications.” *Applied Optics* 41 (15): 2705. doi:10.1364/ao.41.002705.

Matsuoka, A., Boss, E., Babin, M., Karp-Boss, L., Hafez, M., Chekalyuk, A., Proctor, C. W., Werdell, P.J., Bricaud, A., 2017. "Pan-Arctic optical characteristics of colored dissolved organic matter: Tracing dissolved organic carbon in changing Arctic waters using satellite ocean color data." Remote Sensing of Environment 200: 89-101. doi:10.1016/j.rse.2017.08.009.

Matsuoka, A., S. B. Hooker, A. Bricaud, B. Gentili, and M. Babin. 2013. “Estimating Absorption Coefficients of Colored Dissolved Organic Matter (CDOM) Using a Semi-Analytical Algorithm for Southern Beaufort Sea Waters: Application to Deriving Concentrations of Dissolved Organic Carbon from Space.” *Biogeosciences* 10 (2): 917–927. doi:10.5194/bg-10-917-2013.

Matsuoka, Atsushi, Yannick Huot, Koji Shimada, Sei Ichi Saitoh, and Marcel Babin. 2007. “Bio-Optical Characteristics of the Western Arctic Ocean: Implications for Ocean Color Algorithms.” *Canadian Journal of Remote Sensing* 33 (6): 503–518. doi:10.5589/m07-059.

Matsuoka, Atsushi, Pierre Larouche, Michel Poulin, Warwick Vincent, and Hiroshi Hattori. 2009. “Phytoplankton Community Adaptation to Changing Light Levels in the Southern Beaufort Sea, Canadian Arctic.” *Estuarine, Coastal and Shelf Science* 82 (3): 537–546. doi:10.1016/j.ecss.2009.02.024.

Matthews, Mark William. 2011. “A Current Review of Empirical Procedures of Remote Sensing in Inland and Near-Coastal Transitional Waters.” *International Journal of Remote Sensing* 32 (21): 6855–6899. doi:10.1080/01431161.2010.512947.

May, Robert, Graeme Dandy, and Holger Maier. 2011. “Review of Input Variable Selection Methods for Artificial Neural Networks.” In *Artificial Neural Networks - Methodological Advances and Biomedical Applications*. InTech.

doi:10.5772/16004.

Mayer, L.M. 1999. “Extent of Coverage of Mineral Surfaces by Organic Matter in Marine Sediments.” *Geochimica et Cosmochimica Acta* 63 (2): 207–215. doi:10.1016/S0016-7037(99)00028-9.

Mélin, F., and V. Vantrepotte. 2015. “How Optically Diverse Is the Coastal Ocean?” *Remote Sensing of Environment* 160 (February): 235–251. doi:10.1016/j.rse.2015.01.023.

Mélin, Frédéric, Giuseppe Zibordi, and Jean-François Berthon. 2007. “Assessment of Satellite Ocean Color Products at a Coastal Site.” *Remote Sensing of Environment* 110 (2): 192–215. doi:10.1016/j.rse.2007.02.026.

Mobley, Curtis, Dariusz Stramski, Paul Bissett, and Emmanuel Boss. 2004. “Optical Modeling of Ocean Waters: Is the Case 1 - Case 2 Classification Still Useful?” *Oceanography* 17 (2): 60–67. doi:10.5670/oceanog.2004.48.

Monnin, E. 2001. “Atmospheric CO2 Concentrations over the Last Glacial Termination.” *Science* 291 (5501): 112–114. doi:10.1126/science.291.5501.112.

Moody, C. S., and F. Worrall. 2017. “Modeling Rates of DOC Degradation Using DOM Composition and Hydroclimatic Variables.” *Journal of Geophysical Research: Biogeosciences* 122 (5): 1175–1191. doi:10.1002/2016JG003493.

Mopper, Kenneth, and David J. Kieber. 2002. “Photochemistry and Cycling of Carbon, Sulfur, Nitrogen and Phosphorus.” In *Biogeochemistry of Marine Dissolved Organic Matter*, edited by Dennis A. Hansell and Craig A. Carlson, 455–507. San Diego: Academic Press.

Mopper, Kenneth, and Christopher A. Schultz. 1993. “Fluorescence as a Possible Tool for Studying the Nature and Water Column Distribution of DOC Components.” *Marine Chemistry* 41 (1–3): 229–238. doi:10.1016/0304-4203(93)90124-7.

Mopper, Kenneth, Xianliang Zhou, Robert J. Kieber, David J. Kieber, Richard J. Sikorski, and Ronald D. Jones. 1991. “Photochemical Degradation of Dissolved Organic Carbon and Its Impact on the Oceanic Carbon Cycle.” *Nature* 353 (6339): 60–62.

doi:10.1038/353060a0.

Morel, A., H. Claustre, D. Antoine, and B. Gentili. 2007. "Natural Variability of Bio-Optical Properties in Case 1 Waters: Attenuation and Reflectance within the Visible and near-UV Spectral Domains, as Observed in South Pacific and Mediterranean Waters." *Biogeosciences* 4 (5): 913–925. doi:10.5194/bg-4-913-2007.

Morel, Anclré, and Louis Prieur. 1977. "Analysis of Variations in Ocean Color1." *Limnology and Oceanography* 22 (4): 709–722. doi:10.4319/lo.1977.22.4.0709.

Morel, André, and Bernard Gentili. 2009. "A Simple Band Ratio Technique to Quantify the Colored Dissolved and Detrital Organic Material from Ocean Color Remotely Sensed Data." *Remote Sensing of Environment* 113 (5): 998–1011. doi:10.1016/j.rse.2009.01.008.

Morel, André, and Stéphane Maritorena. 2001. "Bio-Optical Properties of Oceanic Waters: A Reappraisal." *Journal of Geophysical Research: Oceans* 106 (C4): 7163–7180. doi:10.1029/2000jc000319.

Nebbioso, Antonio, and Alessandro Piccolo. 2013. "Molecular Characterization of Dissolved Organic Matter (DOM): A Critical Review." *Analytical and Bioanalytical Chemistry* 405 (1): 109–124. doi:10.1007/s00216-012-6363-2.

Nelson, N. B., D. A. Siegel, and A. F. Michaels. 1998. "Seasonal Dynamics of Colored Dissolved Material in the Sargasso Sea." *Deep-Sea Research Part I: Oceanographic Research Papers* 45 (6): 931–957. doi:10.1016/S0967-0637(97)00106-4.

Nelson, Norman B., Craig A. Carlson, and Deborah K. Steinberg. 2004. "Production of Chromophoric Dissolved Organic Matter by Sargasso Sea Microbes." *Marine Chemistry* 89 (1–4): 273–287. doi:10.1016/j.marchem.2004.02.017.

Nelson, Norman B., and David A. Siegel. 2002. "Chromophoric DOM in the Open Ocean." In *Biogeochemistry of Marine Dissolved Organic Matter*, edited by Dennis A. Hansell and Craig A. Carlson. San Diego: Academic Press.

Nelson, Norman B., and David A. Siegel. 2013. "The Global Distribution and Dynamics of Chromophoric Dissolved Organic Matter." *Annual Review of Marine Science* 5:

447–476. doi:10.1146/annurev-marine-120710-100751.

Nelson, Norman B., David A. Siegel, Craig A. Carlson, and Chantal M. Swan. 2010. "Tracing Global Biogeochemical Cycles and Meridional Overturning Circulation Using Chromophoric Dissolved Organic Matter." *Geophysical Research Letters* 37 (3): n/a-n/a. doi:10.1029/2009gl042325.

Nelson, Norman B., David A. Siegel, Craig A. Carlson, Chantal Swan, William M. Smethie, and Samar Khatiwala. 2007. "Hydrography of Chromophoric Dissolved Organic Matter in the North Atlantic." *Deep-Sea Research Part I: Oceanographic Research Papers* 54 (5): 710–731. doi:10.1016/j.dsr.2007.02.006.

O'Reilly, J.E., S. Maritorena, B.G. Mitchell, D.A. Siegel, K.L. Carder, S.A. Garver, M. Kahru, and C. McClain. 1998. "Ocean Color Chlorophyll Algorithms for SeaWiFS." *Journal of Geophysical ResearchGeophys. Res.* 103 (24): 937–953.

Oliver, Barry G., Earl M. Thurman, and Ronald L. Malcolm. 1983. "The Contribution of Humic Substances to the Acidity of Colored Natural Waters." *Geochimica et Cosmochimica Acta* 47 (11): 2031–2035. doi:10.1016/0016-7037(83)90218-1.

Organelli, Emanuele, Annick Bricaud, David Antoine, and Atsushi Matsuoka. 2014. "Seasonal Dynamics of Light Absorption by Chromophoric Dissolved Organic Matter (CDOM) in the NW Mediterranean Sea (BOUSSOLE Site)." *Deep-Sea Research Part I: Oceanographic Research Papers* 91. Elsevier: 72–85. doi:10.1016/j.dsr.2014.05.003.

Organelli, Emanuele, Annick Bricaud, Bernard Gentili, David Antoine, and Vincenzo Vellucci. 2016. "Retrieval of Colored Detrital Matter (CDM) Light Absorption Coefficients in the Mediterranean Sea Using Field and Satellite Ocean Color Radiometry: Evaluation of Bio-Optical Inversion Models." *Remote Sensing of Environment* 186 (December). Elsevier Inc.: 297–310. doi:10.1016/j.rse.2016.08.028.

Osburn, Chris, Thomas Bianchi, Robert Chen, Paula Coble, Eurico D'Sa, George Gardner, and Xu-Chen Wang. 2011. "Absorption Spectra and DOC Concentrations from the NOAA Ship Ronald H. Brown Cruise RB-07-05, R/V Pelican Cruises PE07-32, PEJun2000 in the Gulf of Mexico (GoMX - NACP-OCB Project)."

Biological and Chemical Oceanography Data Management Office (BCO-DMO). http://lod.bco-dmo.org/id/dataset/3296.

Passow, Uta. 2002. "Transparent Exopolymer Particles (TEP) in Aquatic Environments." *Progress in Oceanography* 55 (3–4): 287–333.

Passow, Uta, and A.L. Alldredge. 1994. "Distribution, Size and Bacterial Colonization of Transparent Exopolymer Particles (TEP) in the Ocean." *Marine Ecology Progress Series* 113: 185–198. doi:10.3354/meps113185.

Petit, J. R., J. Jouzel, D. Raynaud, N. I. Barkov, J.-M. Barnola, I. Basile, M. Bender, et al. 1999. "Climate and Atmospheric History of the Past 420,000 Years from the Vostok Ice Core, Antarctica." *Nature* 399 (6735): 429–436. doi:10.1038/20859.

Pezzulli, S., D. B. Stephenson, and A. Hannachi. 2005. "The Variability of Seasonality." *Journal of Climate* 18 (1): 71–88. doi:10.1175/JCLI-3256.1.

Polovina, Jeffrey J., Evan A. Howell, and Melanie Abecassis. 2008. "Ocean's Least Productive Waters Are Expanding." *Geophysical Research Letters* 35 (3): L03618. doi:10.1029/2007GL031745.

Poulton, A. J., R. Sanders, P. M. Holligan, M. C. Stinchcombe, T. R. Adey, L. Brown, and K. Chamberlain. 2006. "Phytoplankton Mineralization in the Tropical and Subtropical Atlantic Ocean." *Global Biogeochemical Cycles* 20 (4): n/a-n/a. doi:10.1029/2006GB002712.

Prieur, Louis, and Shubha Sathyendranath. 1981. "An Optical Classification of Coastal and Oceanic Waters Based on the Specific Spectral Absorption Curves of Phytoplankton Pigments, Dissolved Organic Matter, and Other Particulate Materials." *Limonology And Oceanography* 26 (4): 671–689. http://www.jstor.org/stable/10.2307/2836033.

Raimbault, Patrik, Nicole Garcia, and Cerutti. 2008. "Distribution of Inorganic and Organic Nutrients in the South Pacific Ocean – Evidence for Long-Term Accumulation of Organic Matter in Nitrogen-Depleted Waters." *Biogeosciences* 5: 281–298. doi:10.5194/bgd-4-3041-2007.

Reynolds, Richard W., and Thomas M. Smith. 1994. “Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation.” *Journal of Climate* 7 (6): 929–948. doi:10.1175/1520-0442(1994)007<0929:IGSSTA>2.0.CO;2.

Romera-Castillo, Cristina, Robert T. Letscher, and Dennis A. Hansell. 2016. “New Nutrients Exert Fundamental Control on Dissolved Organic Carbon Accumulation in the Surface Atlantic Ocean.” *Proceedings of the National Academy of Sciences* 113 (38): 10497–10502. doi:10.1073/pnas.1605344113.

Roshan, Saeed, and Timothy DeVries. 2017. “Efficient Dissolved Organic Carbon Production and Export in the Oligotrophic Ocean.” *Nature Communications* 8 (1). Springer US. doi:10.1038/s41467-017-02227-3.

Ruder, Sebastian. 2016. “Enhanced Reader.Pdf.” *ArXiv Preprint ArXiv:1609.04747*.

Rumelhart, David E., G. E. Hinton, and R. J. Williams. 1986. “Learning Internal Representations by Error Propagation.” In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, edited by D. E. Rumelhart, J. L. McClennad, and P. R. Group, 318–362. Cambridge: MIT Press.

Sachse, A., R. Henrion, J. Gelbrecht, and C.E.W. Steinberg. 2005. “Classification of Dissolved Organic Carbon (DOC) in River Systems: Influence of Catchment Characteristics and Autochthonous Processes.” *Organic Geochemistry* 36 (6): 923–935. doi:10.1016/j.orggeochem.2004.12.008.

Salisbury, J., D. Vandemark, J. Campbell, C. Hunt, D. Wisser, N. Reul, and B. Chapron. 2011. “Spatial and Temporal Coherence between Amazon River Discharge, Salinity, and Light Absorption by Colored Organic Carbon in Western Tropical Atlantic Surface Waters.” *Journal of Geophysical Research: Oceans* 116 (7): 1–14. doi:10.1029/2011JC006989.

Sanders, Richard, Stephanie A. Henson, Marja Koski, Christina L. De La Rocha, Stuart C. Painter, Alex J. Poulton, Jennifer Riley, et al. 2014. “The Biological Carbon Pump in the North Atlantic.” *Progress in Oceanography* 129 (PB). Elsevier Ltd: 200–218. doi:10.1016/j.pocean.2014.05.005.

Santana-Falcón, Yeray, Xosé Antón Álvarez-Salgado, María Dolores Pérez-Hernández,

Alonso Hernández-Guerra, Evan Mason, and Javier Arístegui. 2017a. "Organic Carbon Budget for the Eastern Boundary of the North Atlantic Subtropical Gyre: Major Role of DOC in Mesopelagic Respiration." *Scientific Reports* 7 (1): 1–12. doi:10.1038/s41598-017-10974-y.

Santana-Falcón, Yeray, Xosé Antón Álvarez-Salgado, María Dolores Pérez-Hernández, Alonso Hernández-Guerra, Evan Mason, and Javier Arístegui. 2017b. "Organic Carbon Budget for the Eastern Boundary of the North Atlantic Subtropical Gyre: Major Role of DOC in Mesopelagic Respiration." *Scientific Reports* 7 (1). Nature Publishing Group. doi:10.1038/s41598-017-10974-y.

Sarmiento, Jorge L., and Nicolas Gruber. 2006. *Ocean Biogeochemical Dynamics*. *Princeton U. Press*. http://physicstoday.scitation.org/doi/10.1063/1.2754608.

Schauer, Ursula. 2008. *The Expedition ARKTIS-XXII/2 of the Research Vessel "Polarstern" in 2007, Berichte Zur Polar- Und Meeresforschung (Reports on Polar and Marine Research)*. Bremerhaven. doi:10.2312/BzPM_0579_2008.

Schlesinger, W. H., and J. M. Melack. 1981. "Transport of Organic Carbon in the World's Rivers." *Tellus* 33 (2): 172–187. doi:10.3402/tellusa.v33i2.10706.

Schnitzer, M., and S. U. Kahn. 1972. *Humic Substances in the Environment*. Edited by Marcel Dekker. *Journal of Environmental Quality*. New York, NY.

Scott, David W. 1992. *Multivariate Density Estimation*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc. doi:10.1002/9780470316849.

Sempéré, Richard, and Gustave Cauwet. 1995. "Occurrence of Organic Colloids in the Stratified Estuary of the Krka River (Croatia)." *Estuarine, Coastal and Shelf Science* 40 (1): 105–114. doi:10.1016/0272-7714(95)90016-0.

Seritti, Alfredo, Beniamino B. Manca, Chiara Santinelli, Ester Murru, Alfredo Boldrin, and Luciano Nannicini. 2003. "Relationships between Dissolved Organic Carbon (DOC) and Water Mass Structures in the Ionian Sea (Winter 1999)." *Journal of Geophysical Research* 108 (C9): 8112. doi:10.1029/2002JC001345.

Shabbar, Amir, and Bin Yu. 2009. "The 1998-2000 La Niña in the Context of Historically

Strong La Niña Events." *Journal of Geophysical Research Atmospheres* 114 (13). Blackwell Publishing Ltd: D13105. doi:10.1029/2008JD011185.

Shalev-shwartz, Shai. 2007. "Online Learning : Theory , Algorithms , and Applications," no. July.

Shanmugam, Palanisamy. 2011. "New Models for Retrieving and Partitioning the Colored Dissolved Organic Matter in the Global Ocean: Implications for Remote Sensing." *Remote Sensing of Environment* 115 (6): 1501–1521. doi:10.1016/j.rse.2011.02.009.

Sharma, Siddharth, Simone Sharma, and Anidhya Athaiya. 2020. "Activation Functions in Neural Networks." *International Journal of Engineering Applied Sciences and Technology* 04 (12): 310–316. doi:10.33564/ijeast.2020.v04i12.054.

Sharp, J.H. 2002. "Analytical Methods for Total DOM Pools." In *Biogeochemistry of Marine Dissolved Organic Matter*, edited by Dennis A. Hansell and Craig A. Carlson, 35–58. San Diegp: Academic Press.

Shiskin, J., A. Young, and J. Musgrave. 1967. *The X-11 Variant of the Census Method II Seasonal Adjustment Program*. U.S. Department of Commerce, Bureau of the Census.

Siegel, D. A., S. Maritorena, N. B. Nelson, D. A. Hansell, and M. Lorenzi-Kayser. 2002. "Global Distribution and Dynamics of Colored Dissolved and Detrital Organic Materials." *Journal of Geophysical Research: Oceans* 107 (12): 1–14. doi:10.1029/2001jc000965.

Siegel, D.A., M.J. Behrenfeld, S. Maritorena, C.R. McClain, D. Antoine, S.W. Bailey, P.S. Bontempi, et al. 2013. "Regional to Global Assessments of Phytoplankton Dynamics from the SeaWiFS Mission." *Remote Sensing of Environment* 135 (August): 77–91. doi:10.1016/j.rse.2013.03.025.

Siegel, Davey A., S. Maritorena, N. B. Nelson, M. J. Behrenfeld, and C. R. McClain. 2005. "Colored Dissolved Organic Matter and Its Influence on the Satellite-Based Characterization of the Ocean Biosphere." *Geophysical Research Letters* 32 (20): 1–4. doi:10.1029/2005GL024310.

Siegel, David A., Stéphane Maritorena, Norman B. Nelson, and Michael J. Behrenfeld. 2005. "Independence and Interdependencies among Global Ocean Color Properties: Reassessing the Bio-Optical Assumption." *Journal of Geophysical Research C: Oceans* 110 (7): 1–14. doi:10.1029/2004JC002527.

Siegel, David A., and Anthony F. Michaels. 1996. "Quantification of Non-Algal Light Attenuation in the Sargasso Sea: Implications for Biogeochemistry and Remote Sensing." *Deep Sea Research Part II: Topical Studies in Oceanography* 43 (2–3): 321–345. doi:10.1016/0967-0645(96)00088-4.

Silverman, B. W. 1986. "Density Estimation for Statistics and Data Analysis." In *Monographs on Statistics and Applied Probability*. London: Chapman and Hall.

Skoog, Annelie, and Ronald Benner. 1997. "Aldoses in Various Size Fractions of Marine Organic Matter: Implications for Carbon Cycling." *Limnology and Oceanography* 42 (8): 1803–1813. doi:10.4319/lo.1997.42.8.1803.

Stramski, D., R. A. Reynolds, M. Babin, S. Kaczmarek, M. R. Lewis, R. Röttgers, A. Sciandra, et al. 2008. "Relationships between the Surface Concentration of Particulate Organic Carbon and Optical Properties in the Eastern South Pacific and Eastern Atlantic Oceans." *Biogeosciences* 5 (1): 171–201. doi:10.5194/bg-5-171-2008.

Stramski, Dariusz, Rick A. Reynolds, Mati Kahru, and B. Greg Mitchell. 1999. "Estimation of Particulate Organic Carbon in the Ocean from Satellite Remote Sensing." *Science* 285 (5425): 239–242. doi:10.1126/science.285.5425.239.

Sundquist, Eric T. 1986. "Geologic Analogs: Their Value and Limitations in Carbon Dioxide Research." In *The Changing Carbon Cycle*, 371–402. New York, NY: Springer New York. doi:10.1007/978-1-4757-1915-4_19.

Swan, Chantal M., Norman B. Nelson, David A. Siegel, and Tihomir S. Kostadinov. 2012. "The Effect of Surface Irradiance on the Absorption Spectrum of Chromophoric Dissolved Organic Matter in the Global Ocean." *Deep-Sea Research Part I: Oceanographic Research Papers* 63 (May): 52–64. doi:10.1016/j.dsr.2012.01.008.

Swan, Chantal M., David A. Siegel, Norman B. Nelson, Craig A. Carlson, and Elora Nasir. 2009. "Biogeochemical and Hydrographic Controls on Chromophoric Dissolved Organic Matter Distribution in the Pacific Ocean." *Deep-Sea Research Part I: Oceanographic Research Papers* 56 (12). Elsevier: 2175–2192. doi:10.1016/j.dsr.2009.09.002.

Tehrani, Nazanin, Eurico D'Sa, Christopher Osburn, Thomas Bianchi, and Blake Schaeffer. 2013. "Chromophoric Dissolved Organic Matter and Dissolved Organic Carbon from Sea-Viewing Wide Field-of-View Sensor (SeaWiFS), Moderate Resolution Imaging Spectroradiometer (MODIS) and MERIS Sensors: Case Study for the Northern Gulf of Mexico." *Remote Sensing* 5 (3): 1439–1464. doi:10.3390/rs5031439.

Thurman, E. M. 1985. *Organic Geochemistry of Natural Waters*. Dordrecht: Springer Netherlands. doi:10.1007/978-94-009-5095-5.

Tipping, E. 1986. "Some Aspects of the Interactions between Particulate Oxides and Aquatic Humic Substances." *Marine Chemistry* 18 (2–4). Elsevier: 161–169. doi:10.1016/0304-4203(86)90005-8.

Toole, D. A., D. Slezak, R. P. Kiene, D. J. Kieber, and D. A. Siegel. 2006. "Effects of Solar Radiation on Dimethylsulfide Cycling in the Western Atlantic Ocean." *Deep-Sea Research Part I: Oceanographic Research Papers* 53 (1). Pergamon: 136–153. doi:10.1016/j.dsr.2005.09.003.

Toole, Dierdre A., and David A. Siegel. 2004. "Light-Driven Cycling of Dimethylsulfide (DMS) in the Sargasso Sea: Closing the Loop." *Geophysical Research Letters* 31 (9): n/a-n/a. doi:10.1029/2004GL019581.

Tranvik, Lars J., John A. Downing, James B. Cotner, Steven A. Loiselle, Robert G. Striegl, Thomas J. Ballatore, Peter Dillon, et al. 2009. "Lakes and Reservoirs as Regulators of Carbon Cycling and Climate." *Limnology and Oceanography* 54 (6 PART 2): 2298–2314. doi:10.4319/lo.2009.54.6_part_2.2298.

Tremblay, J.E., M.I. Lucas, G. Kattner, R. Pollard, V.H. Strass, U. Bathmann, and A. Bracher. 2002. "Significance of the Polar Frontal Zone for Large-Sized Diatoms and New Production during Summer in the Atlantic Sector of the Southern Ocean." *Deep*

Sea Research Part II: Topical Studies in Oceanography 49 (18): 3793–3811. doi:10.1016/S0967-0645(02)00111-X.

Turner, JT. 2002. "Zooplankton Fecal Pellets, Marine Snow and Sinking Phytoplankton Blooms." *Aquatic Microbial Ecology* 27: 57–102. doi:10.3354/ame027057.

Twardowski, Michael S., Emmanuel Boss, James M. Sullivan, and Percy L. Donaghay. 2004. "Modeling the Spectral Shape of Absorption by Chromophoric Dissolved Organic Matter." *Marine Chemistry* 89 (1–4): 69–88. doi:10.1016/j.marchem.2004.02.008.

Tzortziou, Maria, Christina Zeri, Elias Dimitriou, Yan Ding, Rudolf Jaffe, Emmanouil Anagnostou, Elli Pitta, and Angeliki Mentzafou. 2015. "Colored Dissolved Organic Matter Dynamics and Anthropogenic Influences in a Major Transboundary River and Its Coastal Wetland." *Limnology and Oceanography* 60 (4): 1222–1240. doi:10.1002/lno.10092.

Vantrepotte, V., H. Loisel, D. Dessailly, and X. Mériaux. 2012. "Optical Classification of Contrasted Coastal Waters." *Remote Sensing of Environment* 123 (August): 306–323. doi:10.1016/j.rse.2012.03.004.

Vantrepotte, V., and F. Mélin. 2009. "Temporal Variability of 10-Year Global SeaWiFS Time-Series of Phytoplankton Chlorophyll a Concentration." *ICES Journal of Marine Science* 66 (7): 1547–1556. doi:10.1093/icesjms/fsp107.

Vantrepotte, V., and F. Mélin. 2011. "Inter-Annual Variations in the SeaWiFS Global Chlorophyll a Concentration (1997-2007)." *Deep-Sea Research Part I: Oceanographic Research Papers* 58 (4): 429–441. doi:10.1016/j.dsr.2011.02.003.

Vantrepotte, Vincent, François-Pierre Danhiez, Hubert Loisel, Sylvain Ouillon, Xavier Mériaux, Arnaud Cauvin, and David Dessailly. 2015. "CDOM-DOC Relationship in Contrasted Coastal Waters: Implication for DOC Retrieval from Ocean Color Remote Sensing Observation." *Optics Express* 23 (1): 33. doi:10.1364/oe.23.000033.

Verdugo, Pedro, Alice L. Alldredge, Farooq Azam, David L. Kirchman, Uta Passow, and Peter H. Santschi. 2004. "The Oceanic Gel Phase: A Bridge in the DOM–POM

Continuum." *Marine Chemistry* 92 (1–4): 67–85. doi:10.1016/j.marchem.2004.06.017.

Verdugo, Pedro, and Peter H. Santschi. 2010. "Polymer Dynamics of DOC Networks and Gel Formation in Seawater." *Deep-Sea Research Part II: Topical Studies in Oceanography* 57 (16). Elsevier: 1486–1493. doi:10.1016/j.dsr2.2010.03.002.

Vodacek, Anthony, Neil V. Blough, Michael D. DeGrandpre, Michael D. DeGrandpre, and Robert K. Nelson. 1997. "Seasonal Variation of CDOM and DOC in the Middle Atlantic Bight: Terrestrial Inputs and Photooxidation." *Limnology and Oceanography* 42 (4): 674–686. doi:10.4319/lo.1997.42.4.0674.

Volk, Tyler, and Martin I. Hoffert. 1985. "Ocean Carbon Pumps: Analysis of Relative Strengths and Efficiencies in Ocean-Driven Atmospheric CO2 Changes." In *The Carbon Cycle and Atmospheric CO2: Natural Variations Archean to Present*, edited by E.T. Sundquist and W.S. Broecker, 99–110. American Geophysical Union; Geophysical Monograph 32. doi:10.1029/GM032p0099.

Weishaar, James L., George R. Aiken, Brian A. Bergamaschi, Miranda S. Fram, Roger Fujii, and Kenneth Mopper. 2003. "Evaluation of Specific Ultraviolet Absorbance as an Indicator of the Chemical Composition and Reactivity of Dissolved Organic Carbon." *Environmental Science & Technology* 37 (20): 4702–4708. doi:10.1021/es030360x.

Wells, Mark L. 1998. "Marine Colloids: A Neglected Dimension." *Nature* 391 (6667): 530–531. doi:10.1038/35248.

Werdell, P. Jeremy, and Sean W. Bailey. 2005. "An Improved In-Situ Bio-Optical Data Set for Ocean Color Algorithm Development and Satellite Data Product Validation." *Remote Sensing of Environment* 98 (1): 122–140. doi:10.1016/j.rse.2005.07.001.

Williams, Peter M., and Ellen R.M. Druffel. 1987. "Radiocarbon in Dissolved Organic Matter in the Central North Pacific Ocean." *Nature* 330 (6145): 246–248. doi:10.1038/330246a0.

Yamada, Namiha, and Eiichiro Tanoue. 2006. "The Inventory and Chemical Characterization of Dissolved Proteins in Oceanic Waters." *Progress in*

Oceanography 69 (1): 1–18. doi:10.1016/j.pocean.2005.11.001.

Zafiriou, Oliver C., Jacques Joussot-Dubien, Richard G. Zepp, and Rod G. Zika. 1984. "Photochemistry of Natural Waters." *Environmental Science & Technology* 18 (12): 358A-371A. doi:10.1021/es00130a001.

Zeiler, Matthew. 2012. "Adadelta: An Adaptive Learning Rate Method." *ArXiv Preprint ArXiv:1212.5701*.

Zepp, R. G., D. J. Erickson III, N. D. Paul, and B. Sulzberger. 2007. "Interactive Effects of Solar UV Radiation and Climate Change on Biogeochemical Cycling." *Photochemical & Photobiological Sciences* 6 (3): 286. doi:10.1039/b700021a.

Zepp, R.G., T.V. Callaghan, and D.J. Erickson. 1998. "Effects of Enhanced Solar Ultraviolet Radiation on Biogeochemical Cycles." *Journal of Photochemistry and Photobiology B: Biology* 46 (1–3): 69–82. doi:10.1016/S1011-1344(98)00186-9.

Zhao, Zhao, Michael Gonsior, Jenna Luek, Stephen Timko, Hope Ianiri, Norbert Hertkorn, Philippe Schmitt-Kopplin, et al. 2017. "Picocyanobacteria and Deep-Ocean Fluorescent Dissolved Organic Matter Share Similar Optical Properties." *Nature Communications* 8 (1): 15284. doi:10.1038/ncomms15284.

Zibordi, G., F. Melin, and J.-F. Berthon. 2006. "A Time-Series of Above-Water Radiometric Measurements for Coastal Water Monitoring and Remote Sensing Product Validation." *IEEE Geoscience and Remote Sensing Letters* 3 (1): 120–124. doi:10.1109/LGRS.2005.858486.

Zweifel, U. L. 1999. "Factors Controlling Accumulation of Labile Dissolved Organic Carbon in the Gulf of Riga." *Estuarine, Coastal and Shelf Science* 48 (3): 357–370. doi:10.1006/ecss.1998.0428.'

Résume étendu

Le cycle mondial du carbone retrace l'échange et le stockage du carbone entre de nombreux réservoirs du système terrestre. Il s'agit de deux domaines : le domaine rapide avec un temps de rotation relativement rapide (0 à 12000 ans) ne représente que 0,3 % du carbone total mais présente un flux d'échange élevé entre les différents sous-domaines allant de 10 à 100 Pg C ans^{-1}; et le domaine lent avec des temps de rotation > 12000 ans, contenant 99,7% du carbone total mais avec un flux d'échange de seulement 0,01 à 0,1 Pg C an^{-1}.

Le principal réservoir de carbone dans le domaine rapide est l'océan, contenant 88,6% de la masse totale. Le reste se trouve dans l'atmosphère (1,67 %), les sédiments océaniques de surface (4 %) et sur la végétation terrestre (1,26 %), les sols (4,46 %) et les eaux douces (0,004 %). En revanche, le domaine lent contient 15 x 106 Pg C localisés dans les roches et les sédiments profonds. Les deux domaines échangent naturellement du carbone avec des flux relativement faibles (0,4 Pg C an^{-1}) constants au cours des derniers siècles.

Au cours des 200 dernières années, depuis le début de la révolution industrielle, on observe une augmentation drastique du CO_2 atmosphérique. Cela semble induire un échange efficace de flux entre l'atmosphère et ses deux principaux puits, la terre et les océans.

Le réservoir océanique de carbone peut être divisé en deux groupes de compartiments : inorganique (37,100 Pg C) et organique (700 Pg C). Le carbone organique total (COT) dans l'océan peut être trouvé dans un état particulaire (POC) ou dissous (COD), ce composant ultérieur étant le composant principal du COT. Par exemple, les mesures in situ des eaux de surface de l'océan Atlantique ont montré que seulement 10 % du COT est POC, un pourcentage similaire (11 %) est trouvé pour la mer Baltique et une contribution beaucoup plus faible du POC au COT (0,02 à 5 %) a été observé dans le Pacifique NE ou dans la mer Méditerranée (1,3 à 3,7%). En raison de leur rôle différent dans le cycle du carbone, ainsi que de leurs différentes voies d'exportation du carbone vers les profondeurs océaniques, la distribution spatio-temporelle du POC et du COD ainsi que leurs contributions relatives au COT doivent être mieux caractérisées sur l'océan mondial.

Le carbone organique particulaire (POC) recueille des particules de carbone organique d'un diamètre compris entre 0,4 et 200μm. Il peut être produit localement par le phytoplancton, les bactéries, le zooplancton et les détritus organiques (p. ex., les granulés fécaux et la neige marine), ou peuvent être transportés à un certain endroit de l'océan à partir de sources lointaines par des courants horizontaux océaniques, ainsi que par le débit de la rivière. Après sa génération par photosynthèse dans la zone euphotique de l'océan, une partie du POC est exportée vers l'océan profond via la « pompe biologique du carbone » (CBP).

Le CBP comprend tous les processus par lesquels le carbone biogène de la zone euphotique est séquestré dans l'océan profond pour être minéralisé, maintenant ainsi les forts gradients verticaux du carbone inorganique océanique. L'exportation de PDC est attribuable au naufrage passif et au transport actif par migrations planctoniques. Le naufrage passif est fortement lié à la production de particules d'exopolymères transparents (TEP) et de biominerals (opale et calcite) par le phytoplancton. Le TEP entraîne la formation de particules détritiques qui coulent et transportent le phytoplancton des eaux de surface jusqu'aux profondeurs de l'océan. D'autre part, l'activité hétérotrophe du zooplancton contribue à l'accélération de la vitesse de naufrage de la matière organique par le reconditionnement du carbone organique produit par photosynthèse en granulés fécaux.

Néanmoins, on pense que seulement 1 % de la production primaire de surface est séquestrée dans l'océan profond. La portion de PDC qui n'est pas exportée vers les profondeurs océaniques peut être transférée à des niveaux trophiques plus élevés par la chaîne alimentaire, transformée en détritus ou recyclée par la boucle microbienne, et une partie de celle-ci peut être transférée dans le bassin de carbone organique dissous (COD) et inorganique (CID).. Par conséquent, le PDC est impliqué dans deux flux de carbone importants dans l'océan, la production primaire et l'exportation vers les grands fonds océaniques ou les bassins de carbone organique et inorganique dissous, le COD et le CID respectivement.

Selon la disponibilité biologique et photochimique, le COD peut être catégorisé en labiles, semi-labiles et réfractaires. Ces trois catégories présentent des temps de distribution et de rotation différents. Le COD labile représente 1 % du volume total de COD dans l'océan,

se trouvant jusqu'à 300 m de profondeur avec un temps de rotation de quelques minutes à quelques jours. Le COD semi-labile représente 15 à 20 % de la production nette dans la zone euphotique. Pour sa résistance à la dégradation microbienne rapide, ce temps de renouvellement du COD varie de mois à années. Par conséquent, il s'accumule en surface et peut être transporté horizontalement par les courants portés par le vent ou exporté en eau profonde par la circulation méridionale de renversement et de ventilation. Enfin, le COD réfractaire est la fraction la plus résistante à la reminéralisation microbienne, ne réagissant qu'à une échelle de temps de plusieurs millénaires avec un temps de séjour pouvant atteindre 12500 ans. Il représente donc le pourcentage le plus élevé de COD océaniques (94 %) répartis à toutes les profondeurs.

Sur les eaux océaniques, le COD est principalement produit localement dans la zone euphotique avec la fixation du CO_2 atmosphérique par le phytoplancton. Par la suite, est partiellement consommée par la communauté bactérienne hétérotrophe au début de la boucle microbienne ou dégradée par photolyse par irradiation ultraviolette (UV) à la surface de l'océan. Les fractions les plus récalcitrantes du COD (COD semi-labile et réfractaire) échappent aux processus de minéralisation rapide. Ces fractions peuvent ensuite être exportées hors de la zone euphotique par des processus de transport et de mélange tels que la subduction, la convection et la diffusion contribuant au choc biologique du carbone et à ses puits océaniques profonds. Son exportation par renversement de la colonne d'eau de l'océan joue un rôle central dans la pompe biologique du carbone.

Les processus conduisant la pompe à carbone ont été étudiés, mais en raison de la complexité des mesures in situ nécessaires à l'évaluation de la dynamique du COD, seuls des échantillons limités de COD dans des zones localisées pour une période donnée sont disponibles. Même si des bases de données distribuées dans le monde entier ont été créées (p. ex., GOCAD, NOAA, etc.) pour recueillir des données provenant de différentes missions, elles ne fournissent pas suffisamment d'information pour bien comprendre la variabilité temporelle du COD dans la surface et la colonne d'eau de l'océan mondial.

Les processus conduisant la pompe à carbone ont été étudiés, mais en raison de la complexité des mesures in situ nécessaires à l'évaluation de la dynamique du COD, seuls des échantillons limités de COD dans des zones localisées pour une période donnée sont

disponibles. Même si des bases de données distribuées dans le monde entier ont été créées (p. ex., GOCAD, NOAA, etc.) pour recueillir des données provenant de différentes missions, elles ne fournissent pas suffisamment d'information pour bien comprendre la variabilité temporelle du COD dans la surface et la colonne d'eau de l'océan mondial. La rareté actuelle des mesures in situ du COD sur l'océan mondial limite fortement la compréhension de la dynamique temporelle et spatiale du COD. Cela représente un véritable obstacle à notre capacité de considérer précisément la contribution de ce stock de carbone dans le budget global du cycle du carbone océanique et d'améliorer sa représentation dans les modèles biogéochimiques mondiaux.

La télédétection par satellite est un outil puissant pour décrire de façon synoptique la dynamique biogéochimique des océans. En fait, la possibilité d'évaluer le contenu en COD dans l'océan côtier à partir de l'observation par radiométrie couleur de l'océan (OCR) en utilisant les propriétés d'absorption de la matière organique dissoute colorée (CDOM) comme un proxy optique unique a été récemment démontrée. L'estimation du COD depuis l'espace dans les eaux océaniques est cependant plus complexe. Cela est dû, d'une part, à la difficulté d'évaluer spécifiquement l'absorption du CDOM sur des eaux claires où le CDOM est habituellement représenté dans un terme d'absorption détritique en vrac, y compris les matières particulaires et dissoutes. D'autre part, dans ces eaux où le CDOM et le COD présentent une cinétique différenciée, l'élaboration de méthodes d'inversion du COD exige 1) l'utilisation de variables descriptives physiques ou biologiques supplémentaires et 2) la prise en compte de l'historique des masses d'eau dans la procédure d'élaboration du modèle. Cette thèse se pose dans ce contexte et vise à réaliser des développements méthodologiques pour fournir des informations innovantes sur la distribution du COD à l'échelle mondiale, sur la base de l'exploitation de l'observation par satellite.

C'est dans ce cadre que ce doctorat se pose avec l'objectif principal de développer un algorithme pour estimer la concentration de COD dans l'océan à partir de la radiométrie de télédétection spatiale. À cette fin, la nouveauté de l'approche proposée repose sur la participation d'informations sur l'état bio-optique et physique de la masse d'eau examinée à un moment donné. En raison du fort découplage entre DOC et $a_{cdom}(\lambda)$, nous décidons d'introduire dans cet algorithme une dimension temporelle permettant de prendre en compte « l'histoire » de la masse d'eau. Pour y parvenir, il est d'abord

nécessaire de définir le meilleur algorithme pour estimer l'a_{cdom} à partir des observations de radiométrie de couleur océanique par satellite sur les eaux océaniques libres. Ensuite, il est nécessaire de définir en premier lieu quelles sont les variables pertinentes, outre $a_{cdom}(\lambda)$, qui fournissent les meilleures performances dans l'estimation de la concentration en COD et en second lieu, à partir de quel temps ces dernières variables doivent être prises en compte. Enfin, la structure du modèle (Neural Net, Multi-Linear Regression, etc.) pour produire des sorties précises doit être définie.

Une première réalisation de ce doctorat présentée au chapitre 2 consiste à développer une nouvelle méthode semi-analytique pour estimer spécifiquement l'absorption du CDOM par l'OCR sur l'océan mondial.

La matière organique dissoute chromophorique (CDOM), aussi appelée gelbstoff, gilvin et substance jaune, est la fraction colorée de la matière organique dissoute totale (DOM). Bien qu'il ne représente qu'une petite partie des DOM totaux en haute mer, le CDOM joue un rôle important en photochimie aquatique et en photobiologie, interférer dans divers cycles biogéochimiques en absorbant la lumière sur une vaste gamme spectrale couvrant les domaines visible et UV. La composition du CDOM est très complexe et diversifiée, en fonction de son origine, de sa fraction labile, de son âge et de son passage des eaux douces au milieu marin. Les eaux côtières présentent généralement une forte concentration de CDOM, principalement d'origine terrestre, introduit dans le système océanique par le rejet de rivières et le lavage des terres. En revanche, le CDOM en eau libre est dominé par le nouveau CDOM d'origine biologique. Dans ce type d'eau et à l'extérieur des zones touchées par l'advection des eaux côtières, le CDOM est généralement considéré comme un produit résiduel du phytoplancton et d'autres particules organiques produites pendant les processus de dégradation. Ces derniers processus contrôlant la dynamique du CDOM en haute mer dépendent fortement du couplage entre les processus physiques et biogéochimiques qui reposent sur des paramètres de forçage tels que la disponibilité de la lumière et le mélange vertical. La grande diversité des processus contrôlant la variabilité d'a_{cdom} rend sa dynamique sur l'océan ouvert encore mal caractérisée. Outre la nécessité d'améliorer nos connaissances sur la distribution spatio-temporelle océanique de l' $a_{cdom}(\lambda)$, le grand niveau d'absorption de CDOM dans le domaine spectral bleu représente un problème pour estimer la concentration de chlorophylle-a, Chl-a, à partir de l'observation de la couleur de l'océan.

Les comportements spectraux similaires entre $a_{cdom}(\lambda)$ et le coefficient d'absorption par des particules non algales, $a_{nap}(\lambda)$, rendent ces deux coefficients d'absorption difficiles à distinguer des algorithmes inverses de couleur de l'océan. Pour cette raison, la communauté des couleurs océaniques s'est historiquement concentrée sur le développement d'algorithmes inverses pour évaluer le coefficient d'absorption de la matière détritique colorée, $a_{cdm}(\lambda)$, qui combine les contributions des particules non algales et CDOM. La grande concentration de CDOM dans les eaux côtières de surface, qui rend sa présence plus facile à détecter, a toutefois stimulé l'élaboration d'approches empiriques ou semi-analytiques pour évaluer l'$a_{cdom}(\lambda)$ dans les eaux côtières. Dans les eaux océaniques libres, où le CDOM est présent dans une concentration beaucoup plus faible que dans les eaux côtières, le premier algorithme de radiométrie par couleur océanique (ROC) dédié à l'estimation de l'$a_{cdom}(\lambda)$ était basé sur l'utilisation de relations variables $a_{cdom}(443)$ vs Chl-a. Toutefois, tel que mentionné par ces derniers auteurs, cet algorithme ne fournit qu'une estimation relative de l'$a_{cdom}(\lambda)$, puisqu'il est calculé « en référence à une teneur standard en chlorophylle ». Plus récemment, des approches purement empiriques fondées sur des rapports de réflectance bleu-vert ou des relations multi-linéaires (MLR) entre $a_{cdom}(\lambda)$ et $R_{rs}(\lambda)$ à différentes longueurs d'onde ont été proposées pour évaluer $a_{cdom}(\lambda)$ à partir de l'observation de la couleur de l'océan. Sur la base du matchup entre les mesures $a_{cdm}(443)$ dérivées par satellite GSM et $a_{cdom}(\lambda)$ in situ, une approche semi-analytique pour évaluer $a_{cdom}(\lambda)$ à partir d'$a_{cdm}(443)$. Très récemment, une approche semi-analytique impliquant d'autres propriétés optiques inhérentes (IOP) a été développée pour évaluer l'$a_{cdom}(443)$ dans les eaux côtières et en haute mer.

Dans ce contexte, i) la meilleure approche pour évaluer adéquatement $a_{cdom}(443)$ sur les zones océaniques est proposée, ii) l'évaluation de la variabilité $a_{cdom}(443)$ en ce qui concerne la concentration de chlorophylle et $a_{cdm}(443)$ et iii) la quantification de la contribution d'$a_{cdom}(443)$ à $a_{cdm}(443)$ et les coefficients de non-absorption d'eau, $a_{nw}(443)$, sur l'océan global. À cette fin, la performance de différents algorithmes, dont un nouveau (CDOM-KD2) et trois algorithmes précédemment publiés (Aurin et coll., 2018; Chen et coll., 2017; Shanmugam, 2011) est évaluée. La description de ces données in situ et satellitaires est d'abord fournie. Les différents algorithmes sélectionnés sont ensuite présentés, et l'adaptation d'un algorithme précédemment publié dédié à

l'estimation de l' $a_{cdom}(412)$ dans les eaux côtières est décrite. La description des modèles spatio-temporels d' $a_{cdom}(443)$, ainsi que de sa contribution relative à l' $a_{cdm}(443)$ et au coefficient d'absorption non hydrique, $a_{nw}(443)$ sont alors fournis.

Le CDOM-KD2 consiste en une adaptation à l'application en haute mer du modèle côtier semi-empirique général publié par Loisel et al. (2014) élaboré pour estimer l' $a_{cdom}(443)$ sur les eaux côtières. Cette adaptation a été élaborée en tenant compte de l'ensemble de données synthétique sur la couleur des océans élaboré par le groupe de travail de l'International Ocean Color Coordinating Group (IOCCG) consacré au développement d'algorithmes inversés. Cet ensemble de données rassemble 500 points de données de propriétés optiques inhérentes (IOP) et de réflectance de télédétection, $R_{rs}(\lambda)$, calculés à partir de simulations de transfert radiatif tous les 3 nm de 400 à 700 nm pour chaque combinaison IOP.

La performance de CDOM-KD2 a été évaluée à partir d'un ensemble de données réparties dans le monde entier avec des données in situ provenant de diverses croisières présentées dans Loisel et al. (2018), NOMAD et Plumes and Blooms (https://seabass.gsfc.nasa.gov/experiment/Plumes_and_Blooms). des projets et un ensemble de données de couplage à partir de deux ensembles de données distincts. Tout d'abord, les produits GlobColour fusionnés quotidiennement L3 Ocean Colour à une résolution spatiale de 4 km^2 (http://www.globcolour.info/CDR_Docs/GlobCOLOUR_PUG.pdf) ont été appariés avec l'ensemble de données GOCAD in situ. Et deuxièmement, l'ensemble de données NOMAD match up basé uniquement sur des observations SeaWiFS.

Afin d'évaluer la performance des différents modèles d'inversion $a_{cdom}(443)$ considérés en fonction des caractéristiques optiques du type d'eau, les points de données ont été classés dans les 16 classes optiques définies par Mélin et Vantrepotte (2015) plus une classe supplémentaire, numéro 17, pour considérer les eaux les plus oligotrophes. Les 17 classes sont définies à partir d'une classification globale de la forme spectrale R_{rs} (spectres de réflectance normalisés). Les données des classes 1 et 2 peuvent être considérées comme représentant les masses d'eau turbides fortement touchées par les apports terrestres. En revanche, les échantillons associés aux classes 9 à 17 correspondent à des eaux où les spectres de réflectance sont bien représentés par le modèle de réflectance du

cas 1 de Morel et Maritorena, (2001) alors que les échantillons des classes 8 à 3 sont plus susceptibles d'être liés à divers types d'eaux du cas 2 dont la forme spectrale s'écarte de plus en plus des spectres modélisés du cas 1.

L'utilisation de la typologie optique fournie par Mélin et Vantrepotte (2015) permet une caractérisation plus fine des performances du modèle. La répartition par classe confirme en outre la pertinence globale des valeurs d'a_{cdom} dérivées de CDOM-KD2(443) avec une précision générale satisfaisante pour les 17 types d'eaux considérés et une plus grande précision sur les eaux les plus claires par rapport aux quelques méthodes existantes (Aurin et coll., 2018; Chen et coll., 2017; Shanmugam, 2011).

Le modèle CDOM-KD2 présente le grand avantage de la possibilité d'estimer l'$a_{cdom}(443)$ à l'échelle mondiale indépendamment de $a_{cdm}(443)$. Par conséquent, le modèle CDOM-KD2 a été appliqué aux archives satellitaires mondiales des satellites fusionnés (Globcolor) ou des satellites récents individuels (ICSA) pour caractériser les modèles spatio-temporels de variabilité de l'$a_{cdom}(443)$ ainsi que la contribution du CDOM au CDM et du CDOM aux modèles non spatio-temporels.absorption d'eau. Les résultats si cet exercice montre que si la variabilité spatiale de la contribution de l'absorption du CDOM au MDP et aux absorptions non hydriques est très marquée dans les océans mondiaux, la variabilité temporelle est relativement lisse. On observe que seules quelques régions de l'océan mondial (régions polaires, régions terminales des gyres océaniques et eaux océaniques influencées par de grands apports fluviaux) présentent une variabilité temporelle relativement élevée dans le rapport $a_{cdom}(443)/ a_{cdm}(443)$ et $a_{cdom}(443)/ a_{nw}(443)$, ainsi qu'une forte corrélation entre CDOM et CDM et Chl-a.

Les eaux polaires et océaniques influencées par les grands apports fluviaux dans le monde présentent les valeurs les plus élevées et une variabilité temporelle élevée pour les rapports $a_{cdom}(443)$, $a_{cdom}(443)/a_{cdm}(443)$ et $a_{cdom}(443)/ a_{nw}(443)$. Dans les régions correspondantes, le CDOM représente 60 % ou plus du MDP, tandis qu'un couplage général élevé dans la dynamique de la matière détritique dissoute et particulaire prévaut. Les composantes ultérieures ne se recoupent pas nécessairement avec la dynamique du phytoplancton, en particulier dans les zones fortement touchées par les apports terrestres. Les régions subtropicales (environ 30°N et S) et équatoriales présentent une situation

intermédiaire avec un niveau global modéré de variabilité temporelle pour les rapports $a_{cdom}(443)$, $a_{cdom}(443)/a_{cdm}(443)$ et $a_{cdom}(443)/ a_{nw}(443)$.

D'autre part, dans les gyres, où $a_{cdom}(443)/a_{cdm}(443)$ présente les valeurs les plus faibles, l'analyse de corrélation révèle que la dynamique du CDOM est mal couplée à celle du CDM et du Chl-a, qui tous deux montrent à l'inverse une forte co-variation. Cela tend à indiquer que la dynamique du phytoplancton est le principal facteur de la variabilité de la matière détritique particulaire dans les systèmes de tourbillons, alors que la dynamique de la matière organique dissoute ne peut être considérée comme une fonction directe du phytoplancton et des sous-produits du phytoplancton. Cela souligne également que d'autres paramètres de forçage comme l'activité microbienne et les processus dépendants de la lumière contrôlent la concentration du CDOM. Ce schéma est observé dans tous les gyres, à l'exception du SPG de l'est dans lequel le découplage entre la dynamique des particules et de la matière dissoute est très fort et la contribution du CDOM à l'absorption totale est plus élevée que celle observée pour les autres gyres. Dans le SPG, la variabilité temporelle constatée pour les rapports $a_{cdom}(443)/a_{cdm}(443)$ et $a_{cdom}(443)/a_{nw}(443)$ est également beaucoup plus élevée que dans les autres gyres. De plus, la contribution du CDOM à l'absorption totale dans cette dernière zone est également plus élevée que celle des autres eaux tourbillonnaires (>40 %), ce qui suggère la présence d'un découplage plus important entre la dynamique des particules et de la matière dissoute pour cette région.

Les différentes tendances observées pour le taux interannuel de changement du CDOM lorsque deux périodes différentes sont utilisées (1997-2012 et 2002-2012), démontrent la nécessité d'effectuer une analyse plus approfondie de la forme des changements interannuels des variables analysées en tenant compte des facteurs non annuels. les variations linéaires qui ne sont pas représentées par l'analyse classique des tendances monotoniques.

À l'avenir, l'analyse de la variabilité temporelle du CDOM et de l'Chl-a devrait être prolongée à une période plus longue (de 1997 à 2020) afin de mieux évaluer l'incidence du CDOM sur l'Chl-a et de déterminer le biais que cela génère sur les estimations de l'Chl-a en haute mer à partir du ROC. Cela devrait être jumelé à l'analyse des données environnementales pour évaluer les conditions physiques menant à des changements dans la communauté de phytoplancton, comme la variation du PAR qui aurait une incidence

directe sur la production primaire, les changements dans la SST qui indiquent le mélange de prosses, remontée et autres changements dans les masses d'eau, ou la survenance d'événements ENSO conduisant à de forts changements environnementaux avec un impact global.

Cette apparente hétérogénéité dans la dynamique du CDOM, du CDM et du Chl-a, et donc dans les facteurs contrôlant la variabilité des matières dissoutes et particulaires dans l'océan global, devrait faire l'objet d'une étude plus approfondie. De plus, la variabilité élevée observée dans la contribution relative du CDOM à l'absorption totale pourrait être prise en compte dans les travaux futurs pour quantifier plus précisément l'impact du CDOM sur les estimations du Chl-a sur les eaux océaniques du ROC.

La deuxième réalisation principale de cette thèse présenté au chapitre 3 a été de proposer une nouvelle méthode (modèle NN29s), basée sur une approche de réseau neuronal artificiel (ANN), pour estimer la concentration de COD en surface sur l'océan global. Ce modèle consiste en la combinaison de deux formulations en fonction des caractéristiques optiques/biogéochimiques des masses d'eau (NN29b et NN29bCHL). Les deux formulations diffèrent en termes de variables d'entrée qui sont également considérées à des décalages de temps différents (1 et 2 semaines avant), un des aspects innovants de cette méthode. Cela permet de prendre en compte les principaux paramètres de forçage physique et biologique des masses d'eau ainsi que l'historique spécifique des masses d'eau pour calculer des valeurs COD précises.

Le développement de l'algorithme DOC en utilisant uniquement des données in situ n'était pas possible en raison du DOC in situ limité et des variables concomitantes in situ auxiliaires (telles que Chl-a, $a_{cdom}(443)$, SST, SSS, MLD). Cette limite est encore plus élevée lorsque les décalages temporels entre le COD et les variables auxiliaires seront pris en compte dans le développement du modèle COD. Compte tenu de cette dernière caractéristique, l'ensemble de données de développement a été construit en faisant correspondre les données DOC in situ avec une liste de paramètres d'entrée potentiels pour les modèles et les décalages de temps associés.

L'ensemble de données DOC in situ rassemble des données de surface DOC distribuées dans le monde entier provenant de différentes missions et bases de données, y compris ANTARES, du site de l'étude Bermuda Atlantic Time-Series (BATS), du projet

d'analyse des données océaniques GLobal (GLODAP). base de données, Global Ocean Carbon Algorithm Database (GOCAD), Gulf of Mexico NACP-OCB Coastal Synthesis (GoMX - NACP-OCB), Hawaii Ocean Time-series (HOTS), K2S1, RV Polarstern cruise ARKTIS-XXVII/2 et TRANSDRIFT.

Les points de données du COD ont été appariés avec les différents satellites (R_{rs}, Chl-a, PAR, SSS, SST, CDOM-KD2) et les flotteurs ARGO (pour MLD), pertinents pour son estimation, examinés à la même semaine de la mesure du COD in situ et aussi à 1, 2, 3 et 4 semaines avant. Les dernières variables ont été testées et, finalement, les NN29s ont été élaborées à l'aide des satellites CDOM-KD2, SST et Chl-a et MLD des flotteurs ARGO.

Le rendement de ce nouveau modèle a été démontré à l'aide d'exercices de validation et de comparaisons avec d'autres méthodes.

Le modèle a été largement testé grâce à des analyses de performance qui ont diversement pris en compte la typologie optique de l'eau, des exercices d'appariement (à partir de la climatologie mensuelle) et des comparaisons avec d'autres modèles d'inversion de COD existants et avec le modèle biogéochimique PISCES.

Pour la classification de la typologie optique de l'eau, la classification de l'eau de Melin et Vantrepotte (2015) précédemment décrite a été appliquée. En référence à cela, deux formulations différentes ont été développées, l'une pour les eaux de classe 1 et l'autre pour la classe 2. La première utilise CDOM-KD2 (-2 semaines), SST (-1 semaine) et MLD (-1 semaine) comme estimateurs, et la seconde les mêmes variables d'entrée plus une autre, Chl-a (-1 semaine).

Les cartes climatologiques mensuelles mondiales produites et l'exercice de validation effectué avec l'ensemble de données climatologiques matchup généré avec les NN29s ont montré leur grand potentiel. Le modèle décrit globalement les caractéristiques attendues du COD en termes de distribution spatiale et de dynamique temporelle qui sont globalement en accord avec les modèles observés in situ.

La moyenne annuelle du COD calculée à l'aide du nouveau modèle a été comparée à celle présentée par Roshan et DeVries (2017). La comparaison des deux a montré de grandes similitudes, suggérant la grande robustesse des NN29s puisque l'image annuelle produite par les auteurs ultérieurs s'est avérée avoir une grande précision avec les mesures DOC

in situ. Cette cohérence ainsi que la bonne représentation observée de la dynamique annuelle du COD sur la climatologie mensuelle donnent confiance au rendement des NN29s. Un tel résultat est très prometteur car les NN29s peuvent également être utilisés pour produire des estimations à haute résolution temporelle.

Toutefois, en raison de l'absence de données in situ sur le COD, les N29 n'ont pas pu faire l'objet d'essais adéquats pour estimer les concentrations de COD supérieures à 85 μmol/L. Néanmoins, ces valeurs élevées ne se trouvent normalement pas dans les eaux en haute mer, où la moyenne maximale arrondit à 80 μmol/L. Pourtant, dans le Pacifique occidental, les concentrations de COD près de 90 μmol/L ont été enregistrées et les NN29s n'ont pas pu les reproduire correctement. Par conséquent, il serait nécessaire de recueillir davantage de données in situ avec une concentration de COD >85 μmol/L distribuée dans le monde entier pour être en mesure d'effectuer une meilleure validation des NN29s dans la gamme mentionnée.

Un problème plus important qui limite l'applicabilité du modèle est lié au manque de données disponibles sur la profondeur de la couche mixte. Ces données sont obtenues à partir de flotteurs ARGO, donc même lorsqu'ils fournissent des données réparties dans le monde entier, la couverture est faible. En raison de cette faible couverture, des cartes sont produites, à moins de 0,4 - 14,6 % du total des pixels présentent des données pour les cartes globales composites de 8 jours. Cela peut mener à l'ajout de bruit à l'estimation et le calcul des cartes moyennes, qui présentent des caractéristiques inégales. Une solution possible pour surmonter ce problème consisterait à tester les performances des NN29s avec une source différente de MLD en plus des données des flotteurs ARGO. Il serait également pertinent de mieux explorer dans quelle mesure la performance est affectée par la suppression de MLD comme paramètres d'entrée du modèle, remodelant la structure de l'ANN.

En outre, le modèle NN29s pourrait être amélioré en prenant en considération la probabilité d'appartenance de chaque pixel aux deux groupes de classes optiques lorsque le commutateur qui lance l'utilisation de l'une des deux formulations différentes, NN29b et NN29bCHL, est activé. Cette approche pondérée serait utile pour éviter tout artefact spatial dans la distribution du COD, fournissant une transition en douceur entre les valeurs estimées par les deux sous-modèles différents.

De plus, dans les études futures, il serait nécessaire de poursuivre une analyse plus approfondie de la covariable du COD et des variables utilisées pour son estimation, afin de mieux comprendre leur poids respectif et leur action sur l'estimation du COD.

Un autre résultat intéressant de cette thèse consiste à décrire la contribution relative du COD et du carbone organique particulaire (POC) au budget total de carbone organique (COT). Quelques études ont indiqué que la fraction relative moyenne du PDC à la COT est d'environ 1 à 10 % (Kumari et Mohan, 2018; Maciejewska et Pempkowiak, 2014; Sanders et coll., 2014; Santana-Falcón et coll., 2017), selon les bassins océaniques. Par conséquent, les calculs fondés sur la moyenne annuelle des COD estimés avec le nouveau modèle et des POC estimés avec le modèle proposé par Loisel et al. (2002) en utilisant le coefficient de rétrodiffusion des particules estimé dans Loisel et al. (2018) montrer que les pourcentages mentionnés sont encore plus variables selon l'emplacement. On observe que le PDC représente en moyenne 5,32 3,75 %, atteignant 10 % seulement dans les régions côtières et dans la zone frontale qui délimite l'océan Austral. Ainsi, alors que la plus grande fraction du carbone organique à la surface de l'océan est dissoute, sa contribution au COT est spatialement très variable. L'origine de cette variabilité, ainsi que la variabilité temporelle, devraient être analysées plus en détail. La région de l'Atlantique Nord, caractérisée par un fort mélange hivernal et une floraison printanière intense de phytoplancton, représenterait une bonne zone de travail pour commencer à analyser la dynamique respective entre le PDC et le COD. Cette zone est également largement échantillonnée par les programmes océaniques antérieurs et actuels.

L'étape suivante consisterait à obtenir la concentration de COD intégrée sur la couche euphotique (DOC_{zeu}) en déterminant sa relation avec le COD de surface (DOC_{surf}). Cet exercice a été effectué pour les concentrations de PDC par Duforêt-Gaurier et al. (2010) à partir d'une analyse statistique des profils verticaux.

Le COD est un produit indirect de la production primaire qui s'accumule dans les gyres et les pôles. En attendant, les producteurs primaires sont la principale source directe de POC dans l'océan, raison pour laquelle le POC et la chlorophylle une distribution présentent des caractéristiques similaires. Par conséquent, le rapport PDR/COT présente les valeurs les plus faibles dans les tourbillons subtropicaux. Dans les pôles, par contre, où l'on s'attend à des concentrations élevées de COD et de POC, le POC semble prendre

plus d'importance dans le COT, représentant plus de 10 % de celui-ci. Ainsi, il serait intéressant d'examiner la relation entre le COD et la chlorophylle a, en se concentrant spécifiquement sur sa variabilité dans l'espace et le temps. Cela est particulièrement intéressant, car la chlorophylle a est la variable d'entrée supplémentaire nécessaire pour estimer le COD lorsque la masse d'eau appartient aux classes 1 à 9.

Les tests et les résultats obtenus dans le cadre de ce doctorat se sont avérés être un grand pas vers une meilleure compréhension de la dynamique et de la distribution du CDOM et du DOC en pleine mer et apporteront de nouvelles informations sur la contribution du DOC au budget total du carbone organique et son rôle dans le cycle mondial du carbone.

Mots-clés: Carbone organique dissous, matière organique dissoute colorée, cycle du carbone, couleur des océans, télédétection, eaux océaniques libres.

Keywords: Dissolved organic carbon, colored dissolved organic matter, carbon cycle, ocean color, remote-sensing, open ocean waters.

Résumé

Le carbone organique dissous (COD) joue un rôle central dans la pompe à carbone biologique. Celles-ci est principalement produit à la surface de l'océan avec la fixation du CO_2 par le phytoplancton et une partie de celui-ci est exportée vers les grands fonds. Cependant, le manque de mesures in situ du COD sur l'océan mondial est un obstacle à l'examen de sa contribution au cycle du carbone océanique. La télédétection par satellite est un outil qui permet décrire la dynamique biogéochimique des océans. En fait, la teneur en COD dans les eaux côtières a été évaluée à partir des observations de radiométrie couleur de l'océan (OCR) par satellite en utilisant la matière organique dissoute colorée (CDOM) comme proxy. Cependant, l'estimation du COD en haute mer est plus complexe en raison de la difficulté de mesurer l'absorption de CDOM dans les eaux claires et des différentes cinétiques que le CDOM et le DOC présentent dans ces eaux. Par conséquent, les méthodes d'inversion du COD nécessitent l'utilisation de variables physiques ou biologiques supplémentaires et la prise en compte de l'historique des masses d'eau.

Dans ce contexte, une première réalisation de ce doctorat a été le développement d'un nouveau modèle semi-analytique (CDOM-KD2) pour estimer l'absorption du CDOM par le OCR au-dessus de l'océan mondial. Cela a fourni une description actualisée de la variabilité temporelle et spatiale du CDOM et de sa contribution à la dynamique du budget d'absorption d'eau à l'échelle mondiale. La deuxième réalisation de ce travail a été la proposition d'une méthode (modèle N29), basée sur une approche de réseau neuronal artificiel, pour estimer la concentration de COD en surface sur l'océan global. Ce modèle consiste en la combinaison de deux formulations en fonction des caractéristiques optiques/biogéochimiques des masses d'eau, différentes dans les variables d'entrée utilisées, qui sont également considérées à des décalages de temps différents.

Les résultats obtenus dans le cadre de ce doctorat représentent un grand pas vers une meilleure compréhension de la dynamique et de la distribution du CDOM et du DOC en pleine mer et apporteront de nouvelles informations sur la contribution du DOC au cycle du carbone mondial.

Mots-clés: Carbone organique dissous, matière organique dissoute colorée, cycle du carbone, couleur des océans, télédétection, eaux océaniques libres.

Abstract

The dissolved organic carbon (DOC) plays a central role in the biological carbon pump. This is mostly produced in the surface of the ocean with the fixation of CO_2 by phytoplankton and part of it is exported to the deep ocean. However, the lack of DOC in situ measurements over the global ocean is an impediment to examine its contribution to the ocean carbon cycle. Satellite remote sensing is a powerful tool to describe ocean biogeochemical dynamics. In fact, DOC content in coastal waters has been the assessed from the satellite ocean color radiometry (OCR) observations using the Colored Dissolved Organic Matter (CDOM) as a proxy. However, its estimation in open ocean is more complex due to the difficulty to measure CDOM absorption in clear waters and to the different kinetics that CDOM and DOC present in these waters. Hence the development of DOC inversion methods requires the use of additional physical or biological variables and the consideration of the water masses history.

In that context, a first achievement of this PhD was the development of a new semi-analytical model (CDOM-KD2) for estimating CDOM absorption from OCR over the global ocean. This provided an updated description of the CDOM temporal and spatial variability and of its contribution to the water absorption budget dynamics at global scale. The second accomplishment of this work was the proposal of a method (NN29s model), based on artificial neural network approach, for estimating surface DOC concentration over the global ocean. This model consists in the combination of two formulations depending on the water masses optical/biogeochemical characteristics, differing in the input variables used, which are also considered at different time lags.

The results obtained in this PhD represent a great step towards a better comprehension of the dynamics and distribution of CDOM and DOC in the open ocean and will bring new insights about the contribution of DOC to the global carbon cycle.

Keywords: Dissolved organic carbon, colored dissolved organic matter, carbon cycle, ocean color, remote-sensing, open ocean waters.