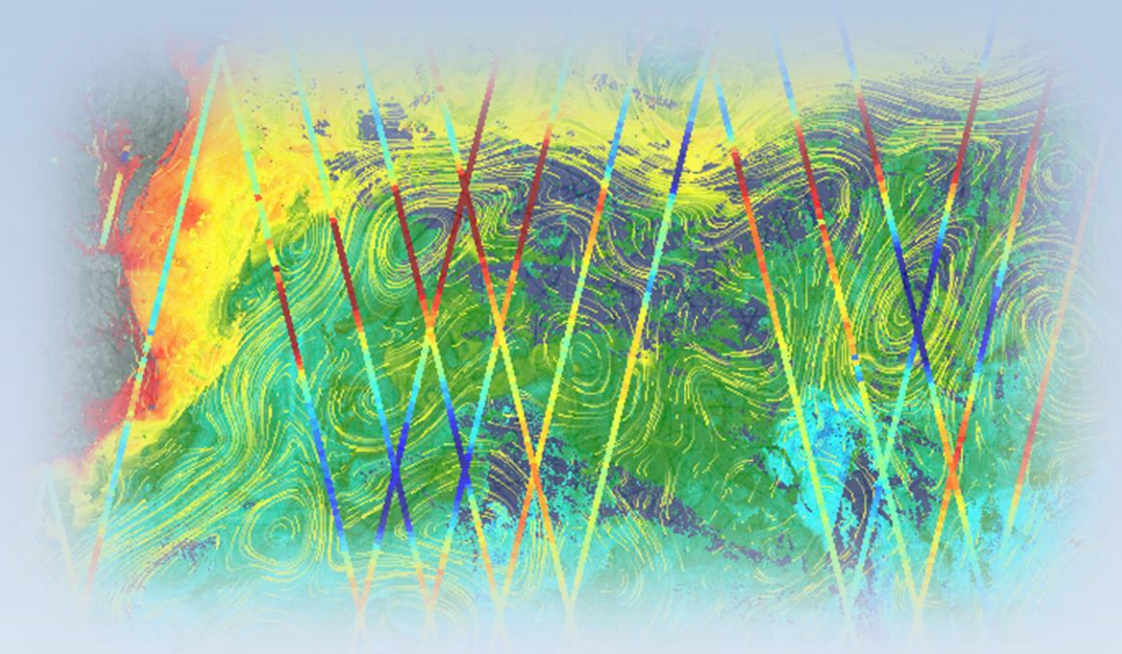
Investigation of meso-scale Sentinel-3 product along-track correlations and the potential of inter-track SSHA estimation using machine learning



Master of Science Thesis

Konstantinos Vlachos

TU Delft

Geoscience & Remote Sensing

Deltares
Enabling Delta Life

Investigation of meso-scale Sentinel-3 product along-track correlations and the potential of inter-track SSHA estimation using machine learning

By

Konstantinos Vlachos

in partial fulfilment of the requirements for the degree of

Master of Science

in Applied Earth Science, Geoscience & Remote Sensing track

at the Delft University of Technology,

to be defended publicly on Thursday December 19, 2019 at 2:30 PM.

Supervisor:	Dr. M. A. Eleveld	Deltares
Thesis committee:	Dr. P. Lopez-Dekker (Chairperson)	TU Delft
	Dr. I. Smal	TU Delft
	Prof. dr. M. Verlaan	TU Delft

This thesis is confidential and cannot be made public until December 20, 2019.

Cover page image source: http://www.globcurrent.org/

Abstract

Satellite altimetry is an important technology used to measure sea level with high spatial and temporal resolution. Sentinel-3, a Copernicus satellite mission, offers three types of variables captured simultaneously for the first time; sea level (SSH), sea surface temperature (SST) and ocean colour (OC) variables. Sea level is measured with SAR altimetry, a technique that considerably increases spatial resolution compared to other means of observation. Altimetry measures sea level across a line that coincides with the satellite ground track, whereas SST and OC are measured on a grid. What we lack are sea level observations in-between ground tracks that would better resolve meso-scale variability. This thesis is focused on two objectives, considering previous work that has indicated associations between those variables. The first objective was to investigate the correlations among SSH, SST and OC, while the second objective was to assess to what extent inter-track sea level can be estimated using SST and OC as predictors in machine learning algorithms.

Daily Sentinel-3 data over a period of eleven months were pre-processed and brought into a form that allowed for computation of metrics such as auto- and cross-correlations in the along-track direction. The focus was on the spatial scales that would enable to detect meso-scale features, such as eddies. With respect to the inter-track sea level estimation two paths were followed. In the first path, Random Forest (RF) and Multilayer Perceptron (MLP) were applied using the complete 11-month dataset as input. Moreover, RF was applied on input data that belong to each separate day. In the second path, 1D Convolutional Neural Network (CNN) was used on the complete 11-month dataset, which inherently honors the spatial dependency of the variables in contrast to the first path.

Generally, the correlations between the variables were found to exist in the meso-scale but were not always strong and they depend on several other factors, such as meteorological conditions, scales included in the analysis and techniques used. All three techniques -RF, MLP and 1D CNN- that were applied on the complete 11-month dataset gave poor results. On the contrary, when RF was applied on the per-day data gave promising results that are reliable mostly in the vicinity of the ground track, although they are not based on one single global model.

The results from this project suggest that there must be more research on the correlation analysis of Sentinel-3 data. It can be improved by using additional or similar techniques, such as localized cross-correlation metrics on various spatial scales. With respect to the inter-track sea level estimation, far more investigation is needed. However, there are indications that a machine learning data-driven approach could potentially work to some extent. Sentinel-3 data will become more abundant in the next years which will assist data science algorithms such as CNNs which require huge datasets.

Keywords: Sentinel-3, sea level, sea surface temperature, ocean colour, Convolutional Neural Network, Random Forest

Preface

Chapter **one** introduces the reader to oceanography in general, as well as in further relevant details. In particular, ocean features, sea level, ocean temperature and ocean colour variables are described with a focus on their remote sensing aspect; for instance, how those variables are measured and some physics behind them, as well as the associations among them. In addition, several techniques are described regarding ways to estimate sea level historically and currently by the use of different types of models. Chapter **two** describes the new Sentinel-3 satellite mission which is one of the triggers of the methodology of this thesis. Especially, chapters one and two add value to the thesis because they briefly make clear that certain aspects are missing regarding sea level estimation research and what is implemented here has not been done before. Chapter **three** summarizes in a more compact and structured way the motivation, the reasons for choosing the study areas, as well as the research objectives. The two sections of chapter **four** give literature information about various processing and analysis techniques, respectively, so the reader can understand basic concepts that were used such as (auto-/cross-) correlations and machine learning analysis using Random Forest and Artificial Neural Networks. Chapter **five** documents the complete methodology used from the data download to analysis and final map production stage. The first sub-section pertains to data themselves, while the second sub-section describes the techniques used to bring the data into an Analysis Ready form, as well as it records the machine learning methodologies that where applied to estimate the sea level and the various approaches chosen to handle the data. Chapter **six** gets into the presentation of the results of each methodology, as well as the findings and the techniques are discussed with a critical perspective that is necessary. In chapter **seven** a brief summary of the thesis is given and the main take home messages that the writer believes are important regarding pros and cons. In addition, possible suggestions on how would the result might become better, as well as future work which does not strictly follow the same course followed here regarding the approaches. Chapter **eight** records the bibliography used. Finally, chapter **nine** presents several additional information about various topics that refer to the thesis such as details about Sentinel-3 and different experimentations with the data that did not show anything significant. Also complementary graphs are included.

Acknowledgements

Any individual's achievement in life – regardless of how one defines achievement – is a result of countless variables that continuously act since, and even before, birth. Apart from undeniable core personal characteristics such as skills, mental toughness, hard work, curiosity, emotional intelligence etc. that play a big role in achievement, there is another underestimated component which is out of our control; chance. It might be underestimated because human mind has evolved to function in a certain way for surroundings that are more similar to a dangerous forest rather than a safer urban environment. Thus, mind wants to see patterns, make heuristic assumptions and build causal (and biased) models in order to survive. It chases information, knowledge and certainty.

However, in our modern age the technological advances have changed disproportionately fast compared to the maturity of societies, especially since our collective monkey mind changes extremely slow. Thus many contradictions that originate from the civilized/animalistic nature dipole become more apparent. Taking also into account the globalized nature of the problems, the state of one's living can be further degraded by the de facto uncertainties of present and future. The positive side is that regardless of whether the future of the individual or the humanity will lean towards a dystopia or eutopia, there is a common denominator for almost all of us. This is the social aspect of human nature and our interdependency. For this reason, I would like to express my thankfulness to several key people starting from the top of the pyramid.

First of all, I would like to deeply thank my daily supervisor Marieke for her advice, optimism and willingness to help at any time. I am grateful that she was supportive, believed in me and gave me the academic freedom and time I needed throughout every stage of the thesis. I also thank Martin for his crucial contribution especially at the very first stage when I tried to formulate the thesis proposal. Moreover, I highly appreciate the help from the two GRS members that were involved each one contributing with his point of view; firstly, Paco who not only offered assisting critical questions at our meetings which helped me reconsider certain aspects of the thesis but also believed in my work despite its initial vagueness and uncertainty, and secondly, Ihor who assisted me with very critical suggestions and ideas regarding the analysis approaches of the problem and was willing to help directly any time.

In addition, I must express my gratitude to the rest of the people that contributed to my MSc. studies and thesis to various extents and in different ways; people that were back home and here in the Netherlands. I need to attribute special thanks to my friend Kostas with whom together we densely experienced many aspects of the expat life which we consumed with a critical perspective, as well as Kosmas and Karel. Everyone that I met and is not mentioned contributed knowingly and unknowingly. Of course, nothing would be feasible without the altruistic support of my family and especially my sister who is always supportive emotionally and materially under any circumstances.

Last but not least, I would like to thank Deltares that provided me funding and useful resources, as well as my committee as a whole which showed enough flexibility for realizing this thesis.

Kostas Vlachos
December 2019

Contents

List of Acronyms and Abbreviations

AATSR	Advanced Along Track Scanning Radiometer
AMSRE	Advanced Microwave Scanning Radiometer-EOS
API	Application Programming Interface
AVHRR	Advanced Very High Resolution Radiometer
AVISO	Archiving, Validation and Interpretation of Satellite Oceanographic data
CDOM	Colour Dissolved Organic Matter
CODA	Copernicus Online Data Access
DAC	Dynamic Atmospheric Correction
DCSM	Dutch Continental Shelf Model
DHuS	Data Hub Software
DTC	Dry Tropospheric Correction
ENSO	El Nino Southern Oscillation
EOF	Empirical Orthogonal Function
EPSG	European Petroleum Survey Group
GMES	Global Monitoring for Environment and Security
GNU	GNU's Not Unix
GODAE	Global Ocean Data Assimilation Experiment
GOES	Geostationary Operational Environmental Satellite
GTSM	Global Tide and Surge Model
HTTPS	HyperText Transfer Protocol Secure
HYCOM	Hybrid Coordinate Ocean Model
IB	Inverse Barometer
IDW	Inverse Distance Weighted
IQR	Interquartile Range
LEO	Low Earth Orbit
Level-4	L4
MODIS	Moderate Resolution Imaging Spectroradiometer
MTSAT-1R	Multi-functional Transport Satellite 1R
MWR	Microwave Radiometer
NaN	Not-a-Number
NetCDF	Network Common Data Format
OGCM	Oceanic General Circulation Model
OI	Optimal Interpolation
OS	Operating System
PCA	Principal Component Analysis
PIES	Pressure sensor-equipped inverted echo sounder
POD	Precise Orbit Determination
ReLU	Rectified Linear Unit
RMSE	Root Mean Squared Error

ROI	Region of Interest
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SLA	Sea Level Anomaly
SLOSH	Sea, Lake and Overland Surges from Hurricanes
SSB	Sea State Bias
SSH	Sea Surface Height
SSHA	Sea Surface Height Anomaly
SWOT	Surface Water Ocean Topography
TRMM MI	Tropical Rainfall Measuring Mission Microwave Imager
TSM	Total Suspended Matter
VLM	Vertical Land Motion
WTC	Wet Tropospheric Correction

List of Tables

List of Figures

1. Introduction

First and foremost, a background in oceanography with an emphasis to the surface aspect of it is important to be presented. This is going to build the very basic understanding of what some of the important ocean features and interactions are which at the same time are related to the thesis topic. In particular, specific focus is put on the sea level and satellite oceanography which are the core concepts.

Sea level fluctuation components are numerous including both long- and short-term, some of them being global ocean circulation and Moon/Sun effects, respectively. Regarding the shorter timescale and spatial components (e.g. (sub-) meso-scale eddies, storm surge, estuarine water discharge), there might be an expected and known association between water temperature, (bio) geochemical content and sea level (Robinson, 2010; Shutler et al., 2016). In turn, the sea level is an important parameter for offshore activities either afar or close to the coast in respective fields which include maritime services, energy production (e.g. wind, oil drilling) and food production (e.g. aquacultures).

As a consequence, how efficiently those fields function can have potentially both short and long term impact on the local economy, hence local society itself. This is the point where satellite oceanography, namely the operational aspect of it, comes into play; in particular among others, all of the above problems could be combated by the ocean and coastal waters satellite monitoring, analyzing and modelling, hence leading towards exploiting the satellite data in various ways.

1.1. Surface Oceanography

1.1.1. Ocean Features and phenomena

The sea level, if one does not consider the tidal component, can be split into two categories. The one pertains to sea level variability greater than 500km which is considered the large-scale. An example of phenomena at these scales is the El Nino/La Nina. This type of variability is mainly owed to atmospheric forcings (e.g. wind stress). The second category concerns the spatial scales smaller than 500km where the sea level changes are owed to ocean currents and eddies (Fu & Cazenave, 2001). However, a very general classification is that below 10 km is considered micro-scale, < 1000 km meso-scale with its subdivisions (e.g. sub-meso-scale) and > 1000 km the mega- or synoptic scale. The large-scale is the most well studied due to the observation instruments which historically did not allow for a higher resolution. With respect to the temporal scale, sea level can vary from daily (e.g. internal waves) to hundreds or thousands of years (e.g. response to climate change). Figure 1 shows in more detail how different phenomena vary in different spatial and temporal scales. Generally, the smaller the spatial scale structures are the more ephemeral and less lifespan they have (Robinson, 2010).

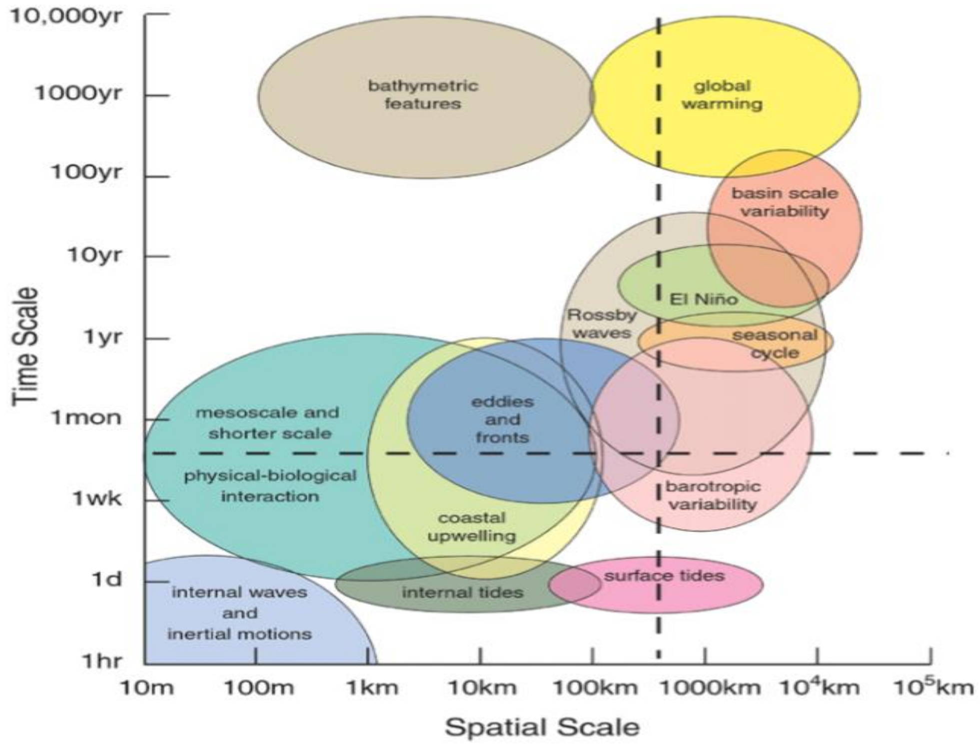


Figure 1: The graph shows the various kind of phenomena that can take place in different spatial and temporal scales. The meso-scale on which the thesis is focused is considered generally between 10 km and 1000 km, with some subdivisions such as the sub-meso-scale. (from Robinson, 2010)

Focusing on the meso-scale features on which the attention fell after the advent of satellite oceanography, phenomena such as eddies, rings, currents, fronts and filaments, are formed due to 1) tides 2) wind 3) thermohaline circulation 4) density gradients (geostrophy) that cause water pressure differences (Shutler et al., 2016). The main two features that are most important for this work are the eddies, currents and fronts. The eddies are water bodies that rotate with diameters from 50 km to 500 km. Filaments are smaller scale features that are generated by the effect of wind and other density differences when a current or eddy is decomposed (Robinson, 2010).

Eddies and sea level can be related to sea surface temperature (SST) in the sense that a difference in density is needed between the outer and inner part of the eddy to be formed. This difference in density is a result of temperature difference. However, in the cases that there are horizontal density gradients at depth instead of the surface one can still find an eddy without relation between SST and sea level (Robinson, 2010). For instance, there are the so-called meddies (i.e. Mediterranean eddies) (McDowell & Rossby, 1978) which do not show a strong sea surface temperature signal due to their genesis mechanism.

Moreover, variability can be spatial and temporal and usually many different components act together. For example, there is a different regime regarding North and South hemisphere, since cyclonic and anti-cyclonic spinning eddies are formed (Robinson, 2010). This means that there is different sea level variability pattern depending on whether we refer to North or South hemisphere. Also, the meso-scale ocean features are highly localized (Yelekci et al., 2001) in the sense that different relations between variables (e.g. sea level, temperature, coastal geometry etc.) occur at different sub-regions of a certain study area.

Regarding temporal variability, one can observe both quasi-permanent patterns of eddies as well as temporary (Robinson, 2010). Therefore, some of them might not be so easy to detect. A reason for this, is that the sea level variability phenomena act simultaneously at different scales thus hiding the eddy signal. However, depending on the region (mainly latitude) different scales dominate over others. When the meso-scale energy is not high enough, then the large scale sea level variability dominates and vice-versa (Kuragano & Kamachi, 2000). Therefore, due to short-lasting and intense meteorological events there might be a superpositioned SST signal which results in hiding the temperature differences of an eddy (Robinson, 2010).

Moreover, eddies and other meso-scale phenomena can also be detectable through the ocean colour. However, the relation between them is ephemeral and typically last a few days. In order to increase the detectability a coloured tracer must be inserted again into the water via rivers, or re-suspension of sea bottom sediment, among others. An example can be seen, where in Figure 2 at the red circle area there is a low (blue), high (cyan), low, high sequence from north to south and it coincides with a change in SST. Whereas, in Figure 3 this current is difficult to spot due to the absence of high concentration in a pigment, chlorophyll in this case.

Nonetheless, a difference in the relation of eddies/sea level/ocean colour between cyclonic and anti-cyclonic eddies has been observed. Namely, in the cyclonic eddies the chlorophyll does not collocate exactly with the eddy core, which is attributed to a combination of causes (Robinson, 2010). Thus, ocean colour is a variable that should be used with care and under specific conditions for meso-scale studies.

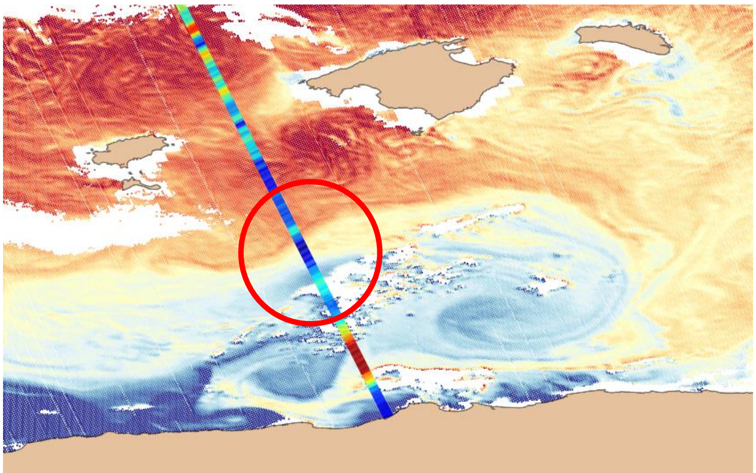


Figure 2: The SST map shows the meso-scale phenomena, i.e. the current and the clear transition from warm to less warm water which also has a signature on the altimeter track that passes on top

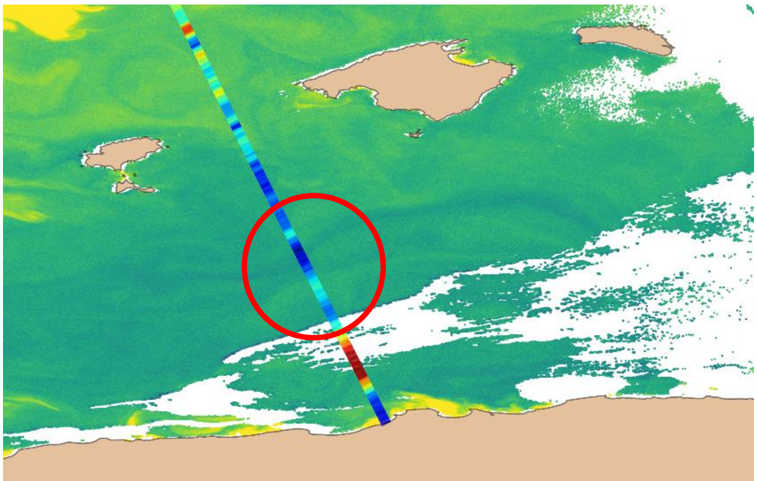


Figure 3: The map shows the same area as in Figure above, using the Chlorophyl variable of Sentinel-3 satellite. By looking at the circle the transition is not depicted as clearly as happens with the SST due to lack of a pigment (i.e. chlorophyl) in the water (Figure 2).

1.1.2. Sea level

Sea level change mechanisms can be split in two groups; steric (or baroclinic) and mass redistribution (or barotropic). According to Martin (2014) the former group of mechanisms is related to regional changes in temperature and salinity (e.g. seasonal variations), hence density and volume while the latter is related to water mass redistribution and other large scale phenomena. Some of the latter could be the tides owed to the Sun and the Moon, or the rigidity of the Earth.

Some examples of the former are the increase in global ocean temperature due to climate change, which in larger scales evokes sea level rise, depending always on the geographical region. For example, coastal sea levels might follow different patterns compared to the open sea and ocean (Robinson, 2010) because of higher complexity close to the coasts. For instance, the impact of discharge water may regulate the sea level compared to the open waters. Hence, the sea level components vary both in time and space and therefore have various ways to measure it.

1.1.2.1. Measuring sea level

Below some of the core measurement techniques along with their limitations are provided.

Tide gauge

Tide gauges are mechanisms that are put at the coastal areas which measure the sea level. The sea level is measured relative to the land and contains all of the sea level components, thus it needs corrections in order to derive the sea level component of interest, if possible. However, depending on the geographical location the land is more or less unstable throughout time due to, i.e. vertical land motion (VLM) (Frederikse et al., 2016) because of tectonics or generally land subsidence due to human impact (Conrad, 2013). For this reason, a precise positioning system is also located at the same spot, in order to measure the vertical as well as the horizontal motions (Frederikse, 2018). One major advantage of tide gauges is the extremely long historic record, although they are not spatially representative as well as they cannot measure in the middle of the oceans (Frederikse, 2018).

Pressure sensor-equipped Inverted Echo Sounder (PIES)

The two different components of the sea level can also be measured by using PIES (Baker-Yeboah et al., 2009). These are sensors that are positioned in the sea bottom which measure the vertical travel time of sound as well as hydrostatic pressure. On the one hand, by using a look-up table along with the travel time, the steric sea level can be estimated. On the other hand, the hydrostatic pressure can yield the sea level change due to mass redistribution. However, the use of the methodology is geographically restricted since i) it takes time to install the stations and ii) it needs quasi-stable and known oceanic conditions since it is based on a look-up table of the regional hydrography.

Altimetry

Altimetry is the satellite observation technique of measuring the sea level from a distance. Altimetry is an active remote sensing technique which utilizes the microwave spectrum. Using altimetric techniques, the two core variables that one can measure is the sea surface height (SSH) and the significant wave height (SWH). The SSH is the sea level, while the SWH represents the magnitude of the ocean waves, simply put.

The sea surface height in a given geographic region is defined as the distance between the sea surface and a reference ellipsoid. What the satellite does is to measure the distance between itself and the sea level, i.e. range. By subtracting the range from the distance between the satellite and the reference ellipsoid (i.e. altitude) which is calculated with a Global Navigation Satellite System (GNSS), one can get the SSH (Eq. (1)). However, in reality this is far more complex, since it essentially needs several corrections. Those corrections belong to four types and are related to 1) atmospheric propagation delay errors, i.e. Ionospheric correction, Wet tropospheric correction (WTC), Dry tropospheric correction (DTC) 2) instrument errors such as tracker bias, antenna gain pattern 3) sea surface errors, e.g. sea state bias, Inverse Barometric effect (IB) and 4) external geophysical adjustments like tides and atmospheric pressure loading.

$$SSH = Altitude - Range - Corrections \quad Eq.\ (1)$$

Finally, doing further correction by subtracting the mean sea surface (MSS) or geoid, the Sea surface height anomaly (SSHA) is derived (Eq. (2)). Also, depending on the application and the study area further corrections are needed such as Dynamic Atmospheric Correction (DAC) which are related to the Inverse Barometric effect and the sea level pressure and winds, as well as tides e.g. Ocean tide, Earth tide, Pole tide, among others.

$$SSHA = SSH - MSS \quad Eq.\ (2)$$

SSHA or Sea Level Anomaly (SLA) is usually used in long-term sea level trend studies, study of the meso-scale currents, eddies etc. The SSHA mostly contains the baroclinic effects that is why it is used in climate studies which are related to ocean temperature and salinity, however depending on the region and the geophysical corrections accuracy, they can still contain barotropic effects (Shao & Zhao, 2015).

Altimetry observations began ~30 years ago which restricts the temporal scale of observation to decadal instead of e.g. multi-decadal compared to tide gauges. Another restriction of altimetry is that the measurements are not reliable enough close to the coasts due to inaccurate geophysical corrections, as well as the contamination of the radar signal that originates from the land (Sterlini et al., 2017). In addition, the measurements are depicted as lines instead of images as is common in the remote sensing field. Still, despite those pitfalls it is considered the best method of measuring the sea level globally since it allows for both space and time sea level observations.

However, lately the problems in altimetry have started to being solved through various methods with respect to equipment and signal processing methodologies, among others. For instance, the last years a new approach in altimetry has been in rise and is going to stay, which is called Synthetic Aperture Radar (SAR) or delay-Doppler altimetry and is extremely promising (Cipollini et al., 2017) since it uses state-of-the art equipment and processing techniques, and it has far better results. Currently, only two satellite platforms work with SAR, being the pioneer Cryosat (Wingham et al., 2006) and Sentinel-3 which is the primary source of data of this thesis. SAR is exceptional in that it has allowed for high spatial resolution (close to 300 meters) as well as it has reduced the noise, e.g. close to the coastal regions, thus it allows for more coastal altimetry applications, such as detection of (sub) meso-scale eddies (Shutler et al., 2016). However, altimetric observations have become more accurate not only because of better retrieval techniques, but also improvements in atmospheric corrections (Bouffard et al., 2008) as well as tidal models (Stammer et al., 2014).

Finally, on 2021 a new altimetric mission is due which is going to depict sea level in an image form, thus offering high spatial coverage. This mission exploits a new technique called swath-altimetry. The name of the mission is Surface Water Ocean Topography (SWOT) (Fu et al., 2012), and it is about to open new doors in the scientific community for far more varying studies.

1.1.3. Surface Ocean Temperature

Sea Surface Temperature (SST) has been proven one of the most important ocean variables that is crucial in large-scale as well as smaller scale studies, since it is inherently connected with ocean phenomena and interactions with the oceanic conditions in depth, as well as the atmosphere and the land contributions.

The SST is affected primarily by two sources. The first source is the outer source, i.e. the Sun, the events on the surface and the effect of atmosphere. The second source is the inner one and it includes the ocean surface and near-surface interactions (Martin, 2014)

The near-surface temperature is a result originating from a combination of causes which in turn affect the ocean phenomena that occur. The main causes are 1) the sun incoming radiation absorption in the upper meters 2) the re-emission of thermal radiation from the ocean back to the atmosphere from the upper first millimeters and 3) the mixing of water in the subsurface due to turbulent phenomena (Minnett & Kaiser-Weiss, 2012).

Regarding the first micrometers which is the skin SST they are affected also by the diurnal temperature changes (Fig. 4), due to the differential incoming radiation as well as the deeper mixing. On the other hand, the so-called foundation SST represents the temperature of the well-mixed layer when the diurnal thermocline of the previous day is very weak (Robinson, 2010).

1.1.3.1. Measuring the SST

The SST measurements come from generally two different sources, namely in-situ and satellite sensors where every sensor (e.g. onboard buoy, boat front, infrared radiometer, microwave radiometer) measures different SST, i.e. from a different layer of the water. The two major groups are described below.

In-situ

Firstly, surface, as well as in depth temperature is observed by floats and drifters according to the ARGO project which started at around 1990 (Davis, 1991). This project is based on in-situ measurements using floats that circulate around most of the oceans. These floats sink in the water column and then they immerge in order to send the data to the data centers and they are designed for specific (large) depths. This leads to certain regions lack of floats, such as the North Sea, since their bathymetry is not deep enough. Also, as can be imagined the spatial resolution of these techniques is very low.

Satellites

The SST satellite measurements began about ~40 years ago. The SST is measured in the two broad categories of the EM spectrum; microwaves and thermal infrared. Thermal infrared depicts the very shallow first layers (Fig. 5) and has the advantage of giving very detailed spatial information but has the major drawback of cloud cover which obstructs the observations. Microwave, on the other hand, can depict the temperature regime from deeper layers (Fig. 5) and has the ability to not being affected by non-raining clouds, although its spatial resolution is significantly lower (Martin, 2014).

Concerning satellites and especially passive ones what thermal infrared sensors measure is a spatial average of the radiation that is emitted from the first 3-13 μm (Martin, 2014) of the water surface (Shutler et al., 2016), i.e. skin SST. However, if there is high enough mixing with the lower water layers (due to e.g. winds, waves), the estimated skin temperatures can be considered representative of a few meters below the surface (Shutler et al., 2016).

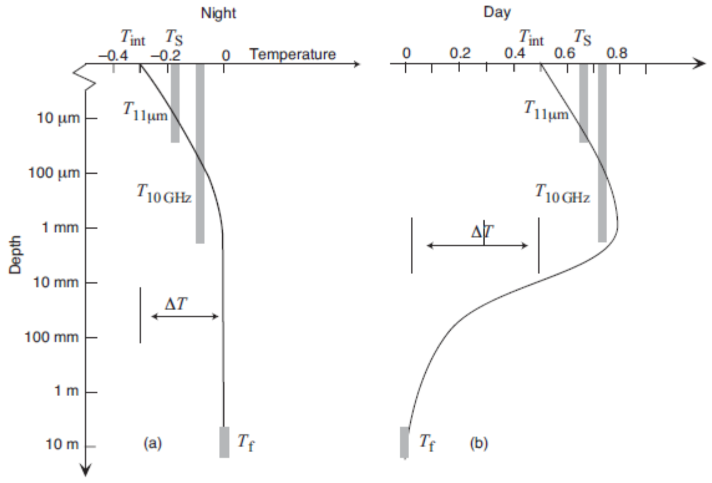


Figure 4: The diurnal change of the thermocline is depicted. In detail how the SSTskin is changing. The thermocline is stronger during the daylight due to larger incoming radiation absorption, while at night the thermocline almost evens out, making the SST representative of the more deep first layers (from Martin, 2014)

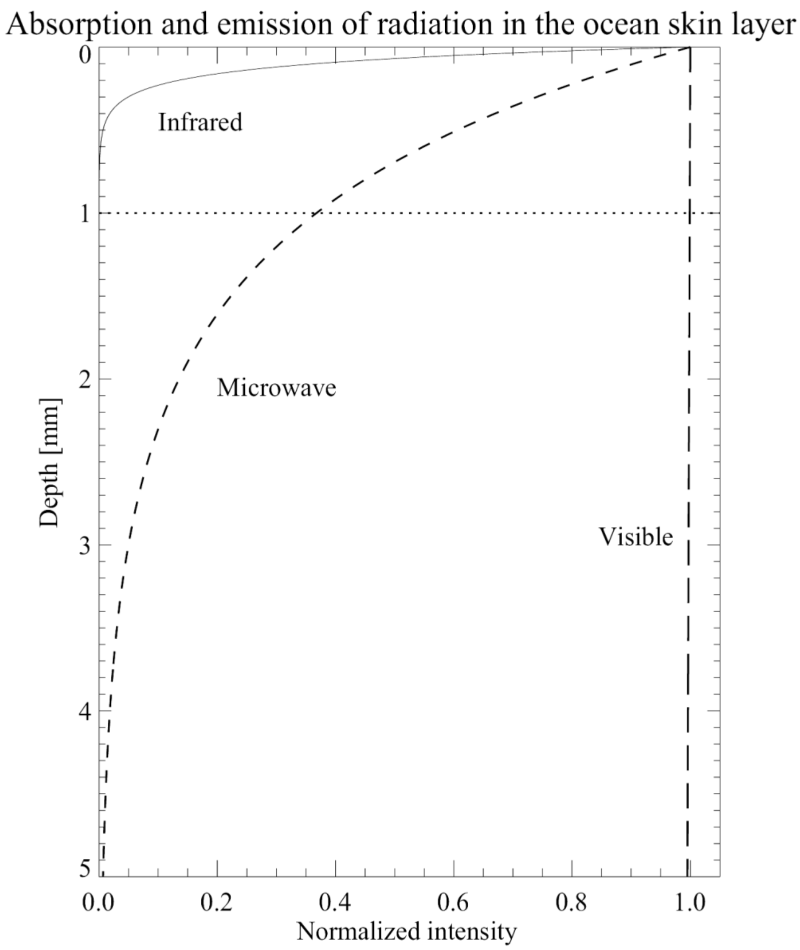


Figure 5: The figure shows the depths which the infrared and microwave radiation are related to, which are parts of the EM spectrum that the satellites measure, each one exhibiting pros and cons in the remote sensing SST earth observation. (from Minnett & Kaiser-Weiss, 2012)

1.1.4. Ocean Colour

By Ocean Colour (OC) we refer to the signature that the ocean or surface waters have when using optical EM sensors in the visible spectrum (OC radiometers). Using this information, several biotic and abiotic parameters/variables of great biogeochemical interest can be derived with some of the most important being chlorophyll-a, total suspended matter (TSM), colour dissolved organic matter (CDOM) and diffuse attenuation coefficient (Kd) (Le Traon et al., 2015).

Chlorophyll-a is the green pigment that is contained in the algae. TSM is considered as the matter including organic and inorganic that is suspended in the water (Eleveld et al., 2008). CDOM represents the dissolved organic material which are organic molecules that are produced from detritus and other organic matter (Hoge et al., 1975). Detritus is considered the dead particulate organic material (Patten, 1975). Kd represents the ability of the light to penetrate water (Valente et al., 2016).

Ocean colour is especially important for the coastal regions (Le Traon et al., 2015) because of the very high variability, e.g. on estuaries where the river discharge fresh water that is mixed with saline sea water. Some possible reasons are nutrients that come from human activities etc. (Robinson, 2010) which affect the primary production (i.e. phytoplankton biomass times phytoplankton growth rate) and dissolved minerals and dead diluted organic material that are imported through rivers etc. (Milliman & Meade, 1983).

Due to this complexity at coastal waters (also referred as complex waters) and the higher impact of CDOM, TSM and Chlorophyll on the backscattered signal the estimation of those parameters is trickier. On the other hand, signal at open waters and deep lakes is dominated by chlorophyll (Siegel et al., 2005) therefore the estimations of this variable is more reliable there. In turn, this complexity could insert difficulties in relating those biotic parameters with other abiotic ones, such as sea level.

Geophysical variables (or parameters) that are derived from OC radiometers can be estimated based on (semi) empirical data-driven approaches (Robinson, 2010). In addition, they can also be estimated by utilizing physical principles - as happens with SST and sea level - such as radiative transfer and optical modelling assuming that the algorithm is parametrized with Specific Inherent Optical Properties (SIOP) (Mobley, 1994). However, radiometers onboard satellites such as Sentinel-3, use (semi) empirical data-driven approaches which utilize univariate or multivariate data processing and analysis techniques.

1.1.5. Associations between ocean variables

Temperature and chlorophyll

The single-celled algae, phytoplankton, contains the chlorophyll pigment (chlorophyll-a), as a result the latter can give information about algae presence and concentration (Boyce et al., 2010). Algae can highly flourish in a mixed layer and reproduces fast under solar rich conditions due to high photosynthesis. In turn, this increases the SST further establishing a positive feedback loop (Shutler et al., 2016). Devi & Sarangi (2017) studied the Indian Ocean during El Nino and La Nina (i.e. large-scale phenomena) for a 12-year period and found both positive and negative correlation between SST anomalies/SSHA and chlorophyll-a concentrations. Although, these are relations that depend on local oceanographic conditions.

Chlorophyll and TSM/CDOM

Total Suspended Matter or CDOM could originate from algae, thus they can be correlated with chlorophyll-a (Robinson, 2010). However, if they originate from a different source such as re-suspension of sea bottom sediment (McCandliss et al., 2002) or river discharge, the correlation becomes more complex. Thus, in this case the two aforementioned variables are non-redundant (Robinson, 2010). This non-redundancy of information is especially true in the complex waters along coastal regions (Robinson, 2010) rather than open waters where chlorophyll dominates.

Temperature and SSH(A)

The relation between SST and SSH is the best studied and well established relation among all variable relations that are used in this thesis. It is evident that their relationship is not always straightforward as one can imagine by simplified physics, since the oceans are a complex system were cause-effect associations can be well hidden. Despite that, their spatiotemporal relationship has been formulated in the frequency (Fourier) domain based on Held et al. (1995) and it has been shown that it can be both linear (Hausmann & Czaja, 2012, Isern-Fontanet et al., 2008) and non-linear depending on the region and spatial scale (Tandeo et al., 2013a). Similar finding have been discovered by Casey & Adamec (2002) who suggest that there is a spatiotemporal large-scale correlation between SST and SSH based on Singular Value Decomposition (SVD) and Empirical Orthogonal Functions (EOF) analyses of satellite data, by studying patterns during El Nino-Southern Oscillation (ENSO) period.

In addition, regarding the meso-scale, there is an observed association between SSH and SST on the ocean/current fronts, as well as SST and the geostrophic flow. In turn the geostrophic flow is related to the SSH. This association between SST and SSH, for instance, is used in current "velocity projection" as is demonstrated by the GlobCurrent platform (Johannessen et al. 2015).

Moreover, a typical meso-scale example is that of eddies which can be clearly depicted by SSHA observations, as well as SST observations since, usually, the cores of the eddies have different temperature than the boundary (Robinson 2010). Nevertheless, the eddy can still occur without the SST spatial variability since there is no direct causal relationship. As a result, a correlation between SSHA and SST in eddies is expected although not always.

To elaborate further, Jones et al. (1998) described several possible mechanisms that result in correlation between SST and SSH(A) which are:

1. Relation between SST and temperature of the deeper ocean layers, which are associated with SSH(A) and water column volume.
2. Meridional SST gradient would induce a movement of warm water bodies towards the North and cold water towards the South. In this case, the SST would be associated with the SSH(A) gradient, instead of SSH(A).
3. Ocean-atmosphere coupling where high SST increases the temperature which is directly above the surface, which in turn leads to air mass movements, thus SSH(A) differences.

Furthermore, Behringer (1994) implied that a correlation exists between SST and SSH in the Atlantic by comparing Jason-1 SSH data with dynamic height estimates of Oceanic General Circulation Model (OGCM) assimilative ocean model. The model at that time assimilated SST satellite data to a high extent and temperature vertical profiles from bathythermographs to a lesser extent. However, their relations depend on the oceanic conditions, features, geography etc. instead of how much SST data were assimilated.

SSH(A) and chlorophyll-a

Based on monthly grouped satellite data during El Nino, algal blooms have been investigated and attributed to lowered SSHA, among other reasons (Murtugudde et al., 1999 in Wilson & Adamec, 2001). Wilson & Adamec (2001) also found negative correlations on large spatial scale (>500km) and weekly binned satellite data, which means that higher SSHA induces a deeper thermocline which in turn reduces the bioavailability of nutrients in the euphotic zone. However, the correlations are restricted to certain regions and latitudes with their own hydrological regimes. The investigated correlations pertained to the time series using Root-Mean-Squared-Errors (RMSEs) which gave direct spatial relationships, as well as EOFs which gave spatiotemporal relationships. Notably, EOFs are commonly used for large-scale spatiotemporal associations (Dawson, 2016).

SSH(A) and density/volume changes

Changes of SSH(A) can be also related directly with water density/volume changes. Two relevant mechanisms are 1) water fluxes on estuarine regions where fresh water is mixed with saline sea water and 2) vertical or horizontal advection of water (e.g. upwelling at coastal areas) (Jones et al., 1998). In this sense, SSHA can be associated not causally to e.g. SST, Chlorophyll and suspended matter because of the river discharge different temperature compare to the coastal waters.

1.2. Studies related to Sea Level estimation

The purpose of this chapter is to introduce the reader to the research on sea level estimation. In this way and in combination with the rest of the chapters it will become clear that the topic of this thesis has not been tried before, to the best of this author’s knowledge, as no similar scientific paper was found. Of course, in no way the information provided here is extensive, since the sea level research is enormous.

1.2.1. Numerical (assimilative) models

To begin with, historically the use of physical models was the main tool in predicting variables about the ocean state (Elsberry & Garwood, 1980), and later was combined with in-situ as well as satellite measurements by assimilating them (Cummings et al., 2009, Elsberry & Garwood, 1980). Global Ocean Data Assimilation Experiment (GODAE) is the major team working in global oceans forecasting and assimilation. The assimilation of data was imperative due to several restrictions that the purely physical model had, that originate from causes such as the physics themselves, the lateral boundary conditions, the grid resolution as well as atmospheric forcings (Cummings et al., 2009). However, the spatial resolution of the output sea level products in reality is worse than the nominal (original resolution, according to design), hence it does not contain smaller scale variabilities, which are crucial in depicting meso-scale phenomena. Therefore, large-scale (assimilative) models are not good in meso-scale variabilities. However, they can provide hourly, daily and monthly predictions. Some of the assimilations systems and methods used are summarized in the table below (*Tab. 1*):

Table 1: Record of some of the assimilative ocean models (from Cummings et al. ,2009)

System Name	Country	Reference
BODAS	Australia	Oke et al., 2008
ECCO-JPL	USA	Fukumori, 2002
FOAM	UK	Martin et al., 2007, Lea et al., 2008
Mercator	France	Brasseur et al., 2005
MOVE/MRIS.COM	Japan	Fujii & Kamachi, 2003
NCODA	USA	Cummings, 2005
NEMOVAR	EU	Weaver et al., 2005
TOPAZ	Norway	Evensen, 2006

Apart from the global ocean models, there are regional and continental shelf models, as well as others that are devoted to surge predictions. In the first group one of them is the Dutch Continental Shelf Model (DCSM) (Gerritsen et al., 1995) while in the second group some of them are Sea, Lake and Overland Surges from Hurricanes (SLOSH) (Jelesnianski et al., 1992) and Global Tide and Surge Model (GTSM) (Kernkamp et al., 2011, Muis et al., 2016). GTSM predicts tide surges and is operational to date, and is in the beginnings of assimilating satellite altimetry data. Notably, one of the innovations of GTSM is that it uses an irregular grid whose size is based on tessellations that get smaller towards the coast, in order to be able to approximate better the more complex hydrodynamics there (Kernkamp et al., 2011).

With respect to SLOSH, it is used to predict large scale phenomena related to hurricane storm surges. Due to the high complexity of the SSH(A) and the nature of the satellite altimetry data, there is variety of research that tries to combat various problems. Some of them are described briefly below.

Moreover, regarding assimilative models there is also Hybrid Coordinate Ocean Model (HYCOM) (Halliwell, 1998, Halliwel et al., 2000) which predicts sea level among others and can depict the sea level spatial pattern fairly accurately, however with not good sea level accuracy.

1.2.2. Data driven models

Proceeding, Optimal Interpolation (OI) is a geostatistical technique to predict SSHA in-between the altimetry tracks (Le Traon, 2003, Pujol et al., 2016) and tries to enhance the information we get regarding meso-scale variability. It is a spatiotemporal approach which utilizes many altimetry satellite missions (AVISO, 2018). However, despite being considered the best approach with the most widely used products today, Lguensat (2017) described two core drawbacks. Firstly, it works best only for large-scale structures and, secondly, quality highly depends on the statistical distribution of data (i.e. being Gaussian), as well as the spatial and temporal lags used to compute the auto-covariance function.

In addition, there is also a new approach that has been developed and tested lately, namely, the analog data assimilation approach (Tandeo, 2015). This approach forecasts sea level by using principally satellite altimetry measurements without the assistance of any physical principles. According to Lguensat (2017) the method works more or less better than the currently widely used OI and model-driven data assimilative models. Additionally, it uses less computational resources, as well as historic observations and simulations, and predicts daily in the meso-scale, locally as well as globally.

In conclusion, it is worth to mention that other data-driven approaches have been developed that aim to forecast phenomena related to SSH changes by using Artificial Neural Networks for the Mediterranean Sea (Rixen et al., 2001). Another example is the SSHA forecasting using polynomial-harmonic deterministic least squares, but also a combination of polynomial-harmonic with autoregressive models (Niedzielski & Kosek, 2009). Still these studies have restrictions regarding true spatial resolution, especially considering that they did not use up-to-date SSH measurement approaches such as SAR.

2. Sentinel-3 Remote Sensing

Copernicus (former GMES) is a European program that has been established in order to assist several stakeholders in decision making regarding fields such as environment, climate, society etc. (Aschbacher & Milagro-Perez, 2012). Sentinel 3 (A and B) is a twin satellite constellation that falls under the framework of Copernicus whose purpose is to measure and monitor the Earth's lands and oceans in both long and short time scales (Donlon et al., 2012). Sentinel 3-A and B are in orbit since February 2016 and April 2018, respectively. The pair follows Low Earth Orbit (LEO) sun-synchronous polar orbits with a 27-day revisit period. The inter-track distance (i.e. distance between ground tracks) is 104km (Le Roy et al., 2007) at the equator and lower towards the poles. Onboard, it carries many sensors with the interest focused on three of them, being SAR Radar Altimeter (SRAL), Sea and Land Surface Temperature Radiometer (SLSTR) and Ocean and Land Colour Instrument (OLCI) which are used to estimate variables such as ocean topography, temperature and Earth surface colour, respectively. In addition, it needs to be noted that Sentinel-3 is the first satellite mission that measures sea level, sea surface temperature and ocean colour variables simultaneously. The current thesis pertains to the ocean component, thus only the relevant products of Sentinel 3 are of interest.

2.1. Altimetry

The Sentinel-3 altimetry system is based on three core instruments to measure Sea Surface Height, among other variables. Namely, the instruments are the SRAL, the Microwave Radiometer (MWR) and the Precise Orbit Determination (POD) system. The SRAL (Le Roy et al., 2010) is a nadir-looking active microwave system (or radar altimeter) which sends pulses in two frequencies (Ku and C). Then, it measures the pulses that are back-scattered by the Earth's surface. MWR is a passive microwave instrument which is used to estimate the delay of the pulses caused during the atmospheric propagation. This is especially useful to estimations close to the coasts due to the higher atmospheric water content variability (Donlon et al., 2012). Finally, POD is a collection of measurement systems that aims to estimate the precise orbit.

In conclusion, SRAL offers sea level estimates, such as SSH with the highest possible along-track spatial resolution to date, which is about 300m. It achieves that by a combination of signal processing techniques and instrument characteristics which comprise SAR mode (i.e. delay-Doppler) altimetry on the contrary to Low Rate Mode (LRM) which is the conventional altimetry approach, though deactivated during the operational phase of Sentinel-3 (EUMETSAT, 2017). Its SSH nominal accuracy is ~3.4 cm.

2.2. Sea Surface Temperature

The SLSTR instrument is a radiometer devoted in measuring the Sea Surface Temperature. It is a dual-view instrument with a nadir-looking and an oblique-looking view. The oblique looking view points at the opposite direction of the satellite movement (EUMETSAT, 2018b). Therefore, it provides two swaths, the nadir-looking being ~1400km and the oblique-looking ~740km (Fig. 6). The average global revisit time is ~1.9 days at the equator, which of course becomes less towards the poles due to the overlap of the swaths of the neighboring relative orbits. Together Sentinel 3-A and B further decrease the revisit time to ~0.9 days.

In order to estimate the SST not only the three thermal bands are used but also VNIR and SWIR (App. Table 5) for certain geophysical corrections. The SST estimation is realized using radiative transfer models with linear regression algorithms (Závody et al., 1995, Merchant et al., 1999). The final SST product has a spatial resolution of 1km and a nominal accuracy less than 0.3 K (EUMETSAT, 2018b).

2.3. Ocean Colour

As far as the ocean colour is concerned, it is measured by the passive sensor, OLCI. It uses 21 spectral bands which fall in the VNIR/SWIR parts of the EM spectrum (App. Table 6). The instrument is a push-broom imaging spectrometer and is composed of five cameras whose field-of-view has overlap. The cameras are positioned in the shape of a fan and they point perpendicular to the flight direction (along-track) of the satellite. Moreover, they are tilted by some amount towards the west in order to alleviate the sun glint effect (EUMETSAT, 2018).

The revisit time is the same as SLSTR's, whereas the spatial resolution is higher and reaches ~300m ESA (2017). In addition, the swath of the OLCI is approximately 1200km and it has a full overlap with the SLSTR swath (Fig. 7).

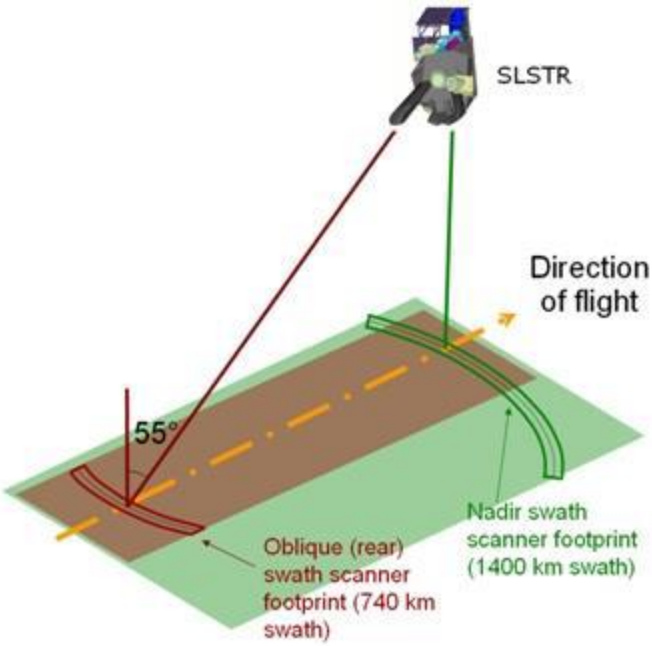


Figure 6: The SLSTR instrument on-board Sentinel-3 is depicted. It is a dual-view instrument, where the one view looks opposite from the direction of flight, while the other one looks towards the Earth. This results in a rear-swath smaller than the nadir-swath. (From EUMETSAT, 2018b)

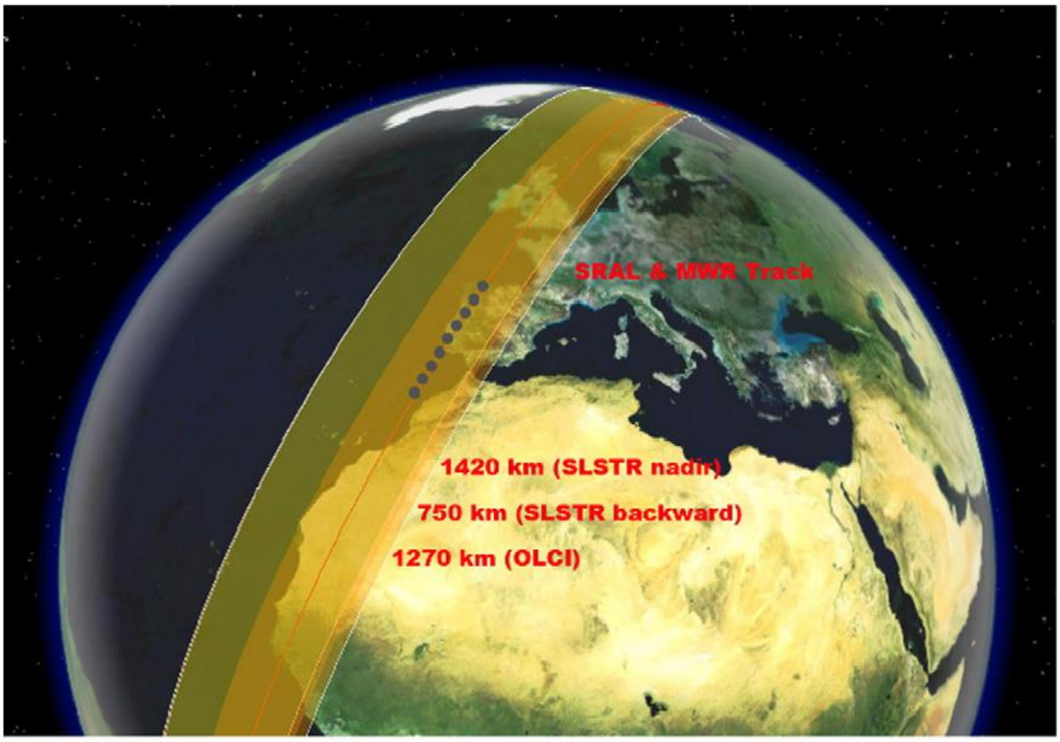


Figure 7: The footprint of the three sensors that are on-board Sentinel-3 are depicted. SLSTR, the instrument that measures the sea surface temperature, has the larger nadir-looking swath. The OLCI instruments which measures the ocean colour, has smaller swath which overlaps fully with the SLSTR nadir-looking one. The SRAL instrument which measures the sea level, does this on a line along-track. (From Donlon et al., 2012)

3. Motivation and Research Objectives

3.1. Motivation and New Opportunities

Putting things into perspective, sea level is mostly measured by tide gauges at the coasts and satellite altimetry offshore. Close to the coasts satellite altimetry shows higher errors due to more complex hydrodynamic processes and significant land/water signal interference, despite SAR approaches have helped in alleviating this effect. Tide gauges yield point (0D) observations while altimetry yields linear (1D) observations. On the other hand, physical-, data- or hybrid-based models do not currently offer high spatial resolving power in the meso-scale. However, on early 2021 the new satellite altimetry mission, SWOT will be launched, that is going to offer swath (2D) sea level products for the first time (Fu et al., 2012).

One of the main purposes of the aforementioned datasets is for climate studies which means long temporal and large spatial scales, instead of operational purposes; the only exception being the surge forecasting. This purpose is mainly achieved by using a lot of altimetry missions along with other datasets, e.g. SST. However, still because of the techniques used the rest of the products such as SST might give greater SSH accuracy, although they do not help in increasing the true spatial resolution. What we currently lack, are sea level products of large spatial and high temporal resolution especially with a meso-scale resolving capability.

The goal of this thesis is to investigate whether the SSHA spatial coverage can be expanded combining it with SST and OC which already have high resolving power with a focus on the meso-scale, since this will be extremely helpful for operational applications. The estimations can also be used as inputs to assimilative models (e.g. Numerical Ocean Prediction, Tide and Surge) and complementary to SWOT products or even be verified by SWOT which will have higher accuracy. In addition, by mapping meso-scale phenomena in-between the altimetry ground tracks could also be useful in quantifying kinetic energy of currents and eddies such as Ducet et al. (2000) with higher precision.

Also the fact that at least 3-4 altimeters are necessary for having an accurate picture of the meso-scale variability (Le Traon et al., 2015) is a drawback that can be potentially be solved by combining the different variables. In the future, this could lead to the decrease of separate satellite altimetry and SST/OC devoted missions, which would mean less cost. On top of that, decreasing the number of satellite missions in general could also be a temporary yet small relief of the space debris problem (NASA, 2016, NASA, 2008) until an agreement on the best solution is come and implemented. Therefore, smart data processing and analysis techniques are imperative to date.

3.1.1. Study Area criteria

Until this point, it has been argued that the relations between SSH(A), SST and OC variables are far from simple. Their relationship sometimes is causal and others simply correlational with high spatio-temporal complexity, from large- to meso-scale. In addition, certain variables and their relationships, as well as their relationships with ocean phenomena at different scales have well been studied compared to others. For instance, SSH(A) and SST, and Chlorophyll to a lesser extent, are well studied, especially in the large scale in phenomena such as El Nino. In addition, the different components of sea level can act simultaneously making the disclosure of relationships even more difficult. Apart from the biogeophysical aspect, there are certain restrictions that are imposed by the satellite mission itself with respect to the choice of the study area, such as the operational time span which is small, as well as the *a priori* decision to work with Sentinel-3 due to the new opportunities it offers.

The part of the Gulf Stream close to New York was chosen because this is a region which is characterized by a clear distinction of quasi-stable meso-scale phenomena such as ocean currents and eddies (Robinson, 2010) as opposed to North Sea which varies fast due to atmospheric and meteorological effects. In fact, atmospheric effects such as wind, as well as Rhine river discharge have a high impact on SSH which means that i) meso-scale phenomena would be not so clear (Pietrzak et al., 2011, Frederikse, 2018) or of smaller scales and ii) the impact of the SST signal-to-noise ratio on SSH(A) would be very small.

In conclusion, choosing a region such as the Gulf Stream would be a favourable option since most of the Sentinel-3 variables could be utilized, and especially SSHA, SST and Chlorophyll, and to a lesser extent TSM, CDOM and Kd. The primary focus was on SSHA and SST. Therefore, a basic assumption is that the relations between most of the variables in the meso-scale and under the time span that S3 operates would be clearer in the Gulf Stream. Although, a small part of the exploratory and correlational research was also realized at North Sea for comparison purposes, and some additional results are given in the Appendix.

3.2. Research Objectives

In summary, the two major research objectives along with some sub-objectives that this thesis deals with are:

- Investigation of possible meso-scale correlations between SSHA, SST and OC variables
    - Correlations
    - Auto- and cross- correlations
- Potential of machine learning techniques in expanding the SSH measurements spatially in the across-track direction
    - Spatial dependence not taken into account
    - Spatial dependence directly taken into account

4. Data Science Theoretical Background

4.1. Data Processing

4.1.1. Outlier detection

According –but not restricted- to Hawkings (1980), outliers of a certain dataset are considered those values that fall too far from most of the rest of measured values. Therefore, they are considered extreme values. However, one always needs to define what "too far" means, which is something that depends on the specific problem under tackle. There are different approaches as simple as using rules of thumb based on statistical metrics (e.g. divergence from the standard deviation), and more complex approaches such as machine learning (e.g. decision trees, PCA).

Regarding the nature of the outliers in real datasets, it can be either "artificial" or natural. An "artificial" outlier could be owed to measurement, instrument, sampling etc. errors. While a natural outlier means that there is an inherent peculiarity in the measurement which could potentially reveal something not common and of specific (high) interest. Therefore, it is subjective whether one needs to disregard the outliers or not and depends on the nature of the problem. In addition, it is very crucial since it will dictate what analysis techniques will be used later.

4.1.2. Moving average filtering

Assuming data on one dimensional (time) domain one can apply a 1D filter. In signal processing, a dataset being in the time-domain means that it belongs in the physical space. While a frequency-domain mean that the data are mathematically transformed in a space with no direct physical meaning. The moving average filter applied is a sliding window and the values of its elements are equal to one divided by the number of window elements (Booth et al., 2006). The new value of each element of the original 1D signal (variable) is equal to the convolution of the original signal and the filter. The moving average is considered a low-pass filter. A schematic example with an explanation can be seen in Figure 8:

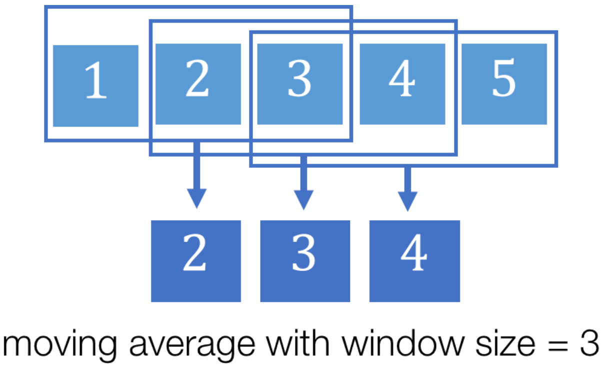


Figure 8: The graphs shows the way that a moving average filter is applied. The upper row depicts the vector that needs to be smoothened, and the lower row depicts the window of size 3. (from https://dawn.cs.stanford.edu/2017/08/07/asap/, Retrieved on: 01/2019)

The value of the first element of the array after convolving and zero-padding one element to its left, becomes

$$\frac{0 \cdot 2 + 1 \cdot 3 + 2 \cdot 4}{3}$$

The value of the second element of the array becomes

$$\frac{1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4}{3}$$

and so on

It needs to be noted that since moving average is based on the mean central tendency metric, can be biased by extreme values. For this reason, prior to applying this filter pre-processing is needed such as, for instance, removing outliers (Chou, 1975, Booth et al., 2006).

4.1.3. (Auto-/cross-) correlation

Assuming there are two variables each one consisted of consecutive measurements, their correlation can be computed by some metric. Depending on whether the variables change linearly or non-linearly together, different kinds of metrics can be used. In the former case, the most common metric is the Pearson's correlation coefficient which is one unitless number, can be used in robust statistics and defines as (Ott & Longnecker, 2010):

$$\rho = \frac{cov(x, y)}{\sigma_x \sigma_y}$$

$$where\ cov(x, y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)\ the\ covariance$$

$$\sigma\ is\ the\ standard\ deviation\ and\ \mu\ is\ the\ mean$$

Proceeding, one can also compare any variable with itself. For this purpose, the term autocorrelation is used and it has meaning on datasets that usually refer to time and/or space. Thus, a correlation metric can be computed between a variable and a lagged version of itself. To put it simply, measurements that are close together are related stronger than the measurements that are afar. Autocorrelation can be defined as (Gubner, 2006):

$$R_{xx} = \sum_{n} x_{n+k} \cdot x_n^*$$

$$where\ x^* is\ the\ complex\ conjugate\ of\ x\ and\ k\ is\ the\ lag$$

In the same manner, one can compute the cross-correlations which would yield the relation of one variable with lagged versions of the other variable. For example, two identical sinusoids with zero lag would give the highest possible absolute value of cross-correlation. Whereas, if one moves the one sinusoid to the positive or negative horizontal axis of the Cartesian 2D plane by half the wavelength, s/he would take an anti-cross-correlation, again, with the highest possible absolute value.

Also, in order to have comparable auto- and cross-correlation metrics, they can be normalized between -1 and 1 (Park, 2018). The two extreme values show the highest possible correlation, where the former shows positive while the latter shows negative correlation (anti-correlation). The normalization takes place before the cross-correlation computation based on the auto-correlation formula above which uses the z-scores.

$$z_x = \frac{x - \mu_x}{\sigma_x N} \ and \ z_y = \frac{y - \mu_y}{\sigma_y}$$

$$where\ N\ is\ the\ number\ of\ elements\ and\ could\ be\ used\ on\ z_y\ instead\ of\ z_x formula$$

4.1.4. Ordinary Least-squares (OLS) regression

OLS is a linear fitting technique which estimates the slope and the intercept parameters of a linear regression model based on given observations that are accompanied by errors

$$\boldsymbol{y} = a\boldsymbol{x} + \beta + \boldsymbol{e}$$

$$\boldsymbol{y}, the\ response\ variable$$
$$\boldsymbol{x}, the\ explanatory\ variable$$
$$a, \beta, the\ slope\ and\ intercept, respectively$$

$$\boldsymbol{e}, the\ errors$$

In the case that the slope is found to be equal or close to 1 and the intercept equal or close to zero, then this means that the response variable can perfectly predicted without any bias. However, always certain assumptions must hold and further post processing normally should be carried. For example, the variables must be random, linearly related, and usually normally distributed. However, violation of the latter does not always have major impact on the estimation. Also, the residuals (or errors) which is the difference between the response variable and its predicted version must be random, normally distributed with zero mean and constant variance (Ott & Longnecker, 2010).

4.1.5. Inverse Distance Weighted (IDW) Interpolation

Assume measurements of a given variable no matter if it has spatial or temporal components. The measurements fall in a defined domain restricted by the minimum and maximum values. To interpolate means that one needs to estimate what is the value of the variable in a position without measurement (query point) that lies in the aforementioned domain. For this reason, there are several techniques that have been devised. Some of them are functional while others are stochastic. IDW is a functional, precise interpolation method. The value at the unknown position is estimated based on the values of its neighbors weighted by the inverse distance. The formulation can be seen below (Shepard, 1968):

$$\hat{y} = \frac{\sum_{i=1}^{N} w_i y_i}{\sum_{i=1}^{N} w_i}\ if\ d > 0$$

$$\hat{y} = y_i\ if\ d = 0$$

$$where\ w_i = \frac{1}{d^p}$$

$$d, the\ distance\ between\ the\ query\ point\ and\ the\ ith\ measurement$$

$$p, the\ power\ which\ usually\ is\ 0{,}1{,}2\ or\ 3$$

$$N, the\ number\ of\ neighbors\ that\ are\ taken\ into\ account$$

4.2. Data Analysis

4.2.1. Machine Learning Techniques

Machine learning is a collection of computational techniques which uses different variables (i.e. features) in order to predict another variable of interest (i.e. label or target). As input, variables with direct physical meaning (e.g. temperature) and indirect physical meaning (e.g. n_{th} order derivatives) can be used which might be informative. The basic idea of machine learning is that it uses the current information that is contained in a measured dataset in order to predict something of interest. A major drawback of this approach is that the techniques cannot predict something that is not already contained in the given dataset. This means that in the case of extreme events presence it might be very difficult, if not impossible, to predict them with accuracy as good as with the rest of the dataset. Of course, researchers are currently working on these cases (e.g. Liu et al., 2016). Moreover, missing values and outliers are usually a problem since they can make the algorithms non-functional and biased, respectively.

On the other hand, a major advantage is that usually the ML algorithms keep the assumptions and part of the *a priori* knowledge to minimal, and they do not ask for specific statistical distribution assumptions, which makes them non-parametric. Especially, in the case of neural networks, one might need minimum pre-processing since the algorithm finds on its own whether there are associations or not. However, much time should be spent on exploring and understanding the data, in order to decide which of them should/should not be used or should be transformed into something more informative (Ketkar, 2017), thus making more wise decision regarding certain parameter values. Therefore, domain knowledge is often crucial. These are aspects of ML that offer lower computational complexity when big datasets are into play, compared to e.g. pure statistical inference.

The general workflow of a prediction scheme using ML is that after pre-processing the features and the label, is that the data need to be split into at least two parts; the train and test sets. Usually, the 80-20% rule applies, although one should experiment with additional proportions. The train set can be further split into two parts; the train and the validation sets. All of the sets, comprise both features and labels, while in all cases the algorithm uses the features to predict something which is compared to the labels.

The idea is that the algorithm learns based on the train set. The user or the algorithm itself (in case of Neural Networks) should use the validation set to check whether it has learnt correctly. If not learnt accurately enough, the so-called hyperparameters are tuned and the procedure is realized again. After arriving to the least-worst hyperparameter, the algorithm is applied on the test set. At the end of every prediction/comparison an error metric is used in order to decide to what extent the prediction is accurate. In turn, choosing an error metric is non-trivial since it depends on the type of dataset and whether it contains e.g. natural outliers, among other things.

4.2.2. Random Forest

Random Forest (either for classification or regression problems) is a machine learning technique based on ensemble of decision trees. It builds decision trees that are not correlated based on subsamples of the tabular data. After that the decision trees are averaged in order to get the final prediction, for regression problems (Breiman, 2001). Moreover, another aspect of the random forest is that it can yield by how much each feature contributed to the final prediction, i.e. feature importance. This is a procedure that is realized after the splitting criterion in each tree is updated. Two important aspects of Random Forest is

Regarding the decision tree itself, it is a decision scheme based on split rules which categorize the data many times until certain criteria are met. There are various parameters that can be tuned in order to achieve the best possible performance such as number of trees, depth of each tree, loss function threshold, among others. It is considered one of the simplest, yet most used and among the best ML techniques (Hastie, 2009). Usually in regression problems the split rule is the Mean Squared Error (MSE). A basic scheme of a decision tree along with its components can be seen in Figure 9.

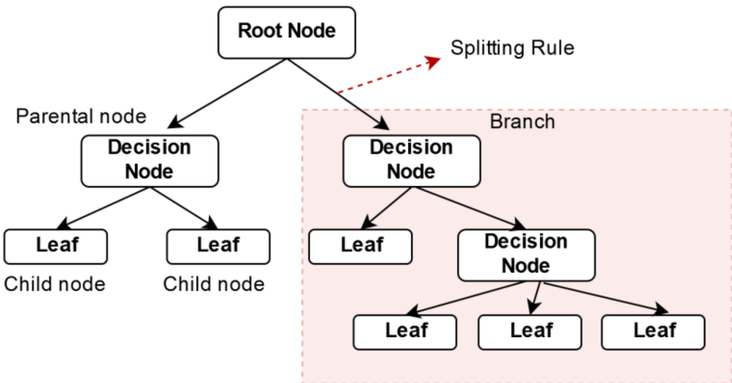


Figure 9: A basic Decision Tree scheme is presented showing its main components. The inputs are split based on different rules (e.g. greater or lower than threshold values) many times, until certain conditions are met. These conditions comprise the hyperparameters which can be number of leafs, number of nodes etc. among others.

4.2.3. Artificial Neural Networks

4.2.3.1. Multilayer Perceptron

The Single Layer Perceptron or simply Perceptron is the simplest form of an Artificial Neural Network (ANN) which comprises the input and output layers and is used in linear problems (Freund & Schapire, 1999). On the other hand, Multilayer Perceptron is a little bit more complex, is composed of an input, an output and at least one hidden layer and is used in both linear and non-linear problems (Hastie et al., 2009). Multilayer perceptron is elaborated further below.

The basic building block of an ANN in general is the so-called unit (or node or neuron). A unit can be considered as a transfer function which takes one or more vectors as input. The transfer function linearly transforms the input vector by applying a scale (weight) and an offset (bias) which are randomly chosen from a given probability distribution. The unit produces a scalar value and its output is generated after the application of the activation function (Hahnloser et al., 2000). The activation function is either linear or non-linear and which one should be chosen depends on the nature of the problem. However, most of the times a non-linear activation function is needed, since this is the approach to detect more complex associations between variables.

Moving forward, an ANN is structured in stack of layers and each layer is composed of units. The units of the first layer (or input layer) represent the input features. The units of the last layer (or output layer) represent the output (or estimate or prediction), while every layer in-between the first and the last are considered as the hidden layers. In turn, the hidden layers comprise one or several units which basically represent artificially created features with no direct physical meaning. As the network becomes deeper the artificially created features resemble even less the original input features. The number of units in each hidden layer is taken as the width of the layer, whereas the number of hidden layers as the depth of the network (Ketkar, 2017). An example of a basic ANN with fully-connected layers can be seen in Figure 10.

The number of the elements of every input feature corresponds to the label or target i.e. the target layer. The values of the target layer can have any value e.g. discrete or continuous. By the time the network predicts the values of the target layer, it needs to compare them with it in order to assess how good the prediction is. The comparison is realized by using a loss function which can be anything, depending on the nature of the problem (i.e. regression, classification). For regression problems some metrics that are commonly used are mean squared error (MSE) and mean absolute error (MAE) each one with its advantages and disadvantages. For example, MSE is more susceptible to outliers compared to MAE. Other useful loss function metrics are Huber Loss (Huber, 1964), which is robust to outliers and very asymmetrically distributed data, as well as Charbonnier loss (Barron, 2019) robust to outliers. Thus, the loss function one should use is not a trivial step. After applying the activation function one can start using other regularization layers such as e.g. Dropout which can prevent overfitting by dropping units during training (Srivastava et al., 2014).

To sum up, the general idea of an ANN is that it has some inputs. Then linear and non-linear transformations are applied which produce a prediction. The prediction is compared to the target (original) by using a loss metric. This procedure is repeated until the loss metric is minimized (Behzad, 2014). The minimization takes place by using an optimizer (usually some form of gradient descent).

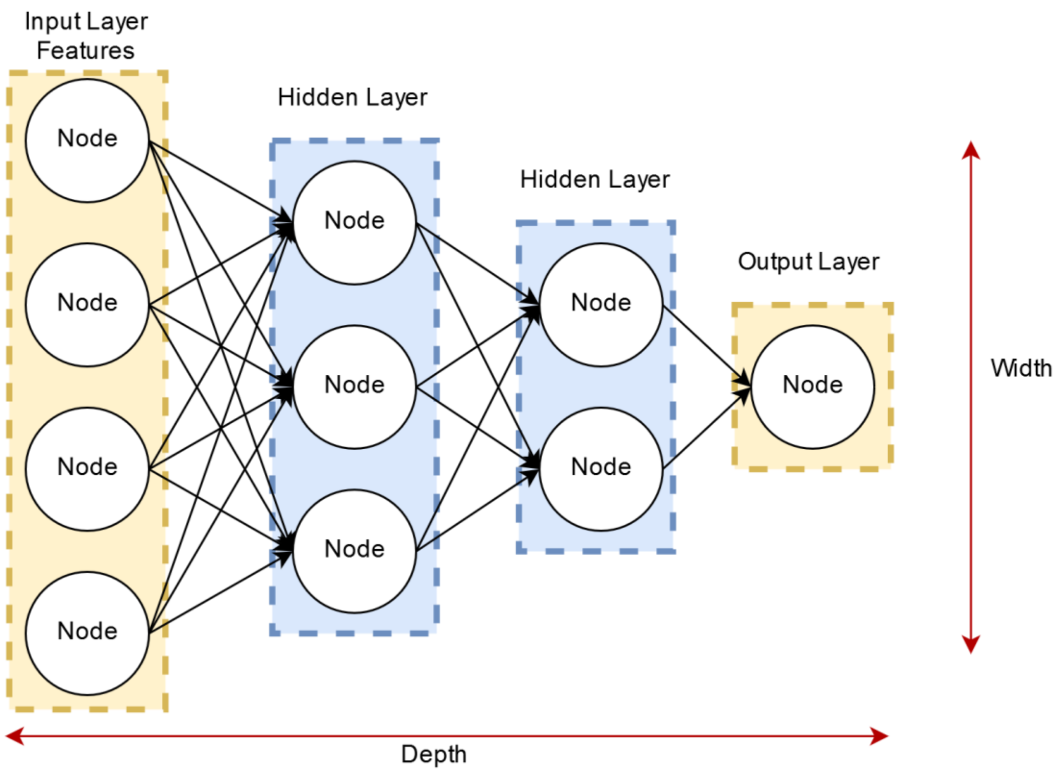


Figure 10: The above graph shows a multilayer perceptron architecture. By definition, the MLP is a fully-connected network which means that every node's output is fed to each node of the next layer. The input layer is composed of four features (nodes). Also there are two hidden layers where each layer consists of three and two nodes, respectively.

4.2.3.2. Convolutional Neural Networks (CNN)

The CNNs follow the same principles as the regular ANNs but in a somewhat different context, since each input is assumed to show auto-correlation to a certain extent. For this reason, the basic operation of convolution is exploited. The general architecture of a CNN from beginning to end is input layer, convolution/activation/pooling layers (hidden layers), fully connected layer (hidden layer), output layer (Ketkar, 2017). Of course, depending on the needs there can be more packs of convolution/activation/pooling and/or fully connected layers.

The convolution layer applies a predefined number of filters and filter (window) size on the inputs. The values of each filter are the weights and are defined in the same manner as in simple ANNs. After that, the output of the convolution layer is fed to a pooling layer. The pooling layer is a layer which uses a predefined window to reduce the size of its input based on a rule. The rule usually is an average (Average Pooling) or maximum (Max Pooling) value out of the values inside the window. This procedure is similar to binning and downsampling the given array (Yamashita et al., 2018).

5. Materials and Methods

5.1. Dataset

5.1.1. Data Description

Sentinel-3 offers various products which come from the three core instruments in the form of compressed (ZIP) NetCDF4. For the current study the Level-2 marine products were used. SRAL offers 3 data files, namely the "reduced", the "standard" and the "enhanced". The "enhanced" data file was chosen since it included complete altimetry information such as 1/20Hz, C/Ku-band as well as raw waveforms. In addition, SLSTR offers one data file which follows the GHRSST naming and structure directives. On the other hand, OLCI offers 24 data files and 7 annotation files. The 24 data files are composed of the surface reflectances and variables, while the 7 annotation files are composed of information such as geographical coordinates, flags etc. More detailed information about the products and the data files that are chosen can be seen in Table 2. Regarding the spatial resolution, the SLSTR 1km and OLCI 300m were chosen. It is worth noting that generally, the products which contain more details (i.e. "enhanced" SRAL and 300m OLCI) were chosen in case that some extra information was going to be necessary during the processing/analysis stages.

Also, Sea Surface Temperature Level-4 (SST L4) were used. The data are delivered as one netCDF file which includes the desirable time series of AOI and represents the foundational SST. The L4 product is gap-filled by using SST data from multiple sensors such as AVHRR, AATSR, SEVIRI, AMSRE, TRMM MI, MODIS, GOES Imager, MTSAT-1R as well as in-situ data from buoys (Chao et al., 2009). The temporal resolution of this product is 1 day while the spatial resolution is 1 km.

5.1.2. Data acquirement

The S3 marine products are disseminated by the Copernicus Online Data Access (CODA) web service. The dates that were chosen were between 01-05-2018 and 30-03-2019 and were queried and downloaded by gaining access to the Data Hub Software (DHuS). The access was made possible by taking advantage of the *dhus.sh* bash script that is offered by the services and uses the corresponding API. On top of that, an additional bash script was created in order to fully automate the query/download process. Working in a Windows 7 OS, additional tools were needed such as *MSYS* and *wget*; the former is a collection of GNU utilities (e.g. bash command line) while the latter is a free software package that retrieves data using various internet protocols (e.g. HTTPS). Further information about the online data access can be found at https://coda.eumetsat.int/manual/CODA-user-manual.pdf.

Regarding the SST L4 products, they are derived from NASA (JPL OurOcean Project, 2010, https://podaac.jpl.nasa.gov/dataset/JPL_OUROCEAN-L4UHfnd-GLOB-G1SST) by using the online spatial subset option. The spatial extents that were used for the North Sea and Gulf Stream cases are described by the bounding boxes (-3.11, 9.11, 51.10, 58.21) and (-80.00, -65.00, 31.00, 40.00), respectively. The coordinates are given in geographic coordinate reference system, namely World Geodetic System 1984 (EPSG: 4326).

5.1.3. Data preparation

After the data acquirement, the products were unzipped and subset based on the Region-Of-Interest (ROI) in geographical coordinates in order to reduce the file size. Particularly, for the SRAL coordinates needed to be adjusted since they are defined as degrees East relative to Greenwich meridian, which means that they are also represented by values greater than 180 degrees. Therefore, in those cases, a subtraction of 360 degrees was needed. Regarding the OLCI files, only the relevant ones where selected (e.g. chl_oc4me.nc, wqsf.nc) while the rest of them (e.g. iwv.nc) were deleted (Tab. 2). As a result of subset, the shape of the original data was changed from matrix to array. The choice of working with the data being in an array (i.e. considered as points) instead of a matrix (i.e. considered as image) form was driven by hardware (i.e. RAM) restrictions. Namely, matrix form would contain more elements including useless cloud covered points, hence it would need higher memory capacity. Throughout every query/download and unzip/subset stage the dataset was checked for any corrupted files and logged to a text file. In case of corrupted file presence the file was re-downloaded/subset.

Next, the flags of SLSTR and OLCI are extracted into bit values and then applied on the corresponding variables and their errors, i.e. SST, chlorophyll, total suspended matter concentration, diffuse attenuation coefficient and coloured detrital and dissolved material absorption. The flags and quality level thresholds that were applied can be seen in Table 2 and followed the suggestions of data providers (EUMETSAT, 2018; GHRSST Science Team, 2010). Notably, regarding the OLCI products the only flag that was not used is the High Solar Zenith (HISOLZEN). It was observed that by using this flag the remaining scenes consisted of considerably less data and this is why it was not used. The SST L4 data that are used are based on sea mask.

Table 2: The table below names the flags which were used in the preprocessing stage of the Sentinel-3 and SST Level-4 satellite data

Ocean Colour variables	**WQSF (flags)**			
Algal pigment concentration in open waters	CHL_OC4ME CHL_OC4ME_err	WATER CLOUD CLOUD_AMBIGUOUS CLOUD_MARGIN INVALID COSMETIC SATURATED SUSPECT HIGHGLINT SNOW_ICE	AC_FAIL WHITECAPS ANNOT_ABSO_D ANNOT_MIXR1 ANNOT_DROUT ANNOT_TAU06 RWNEG_O2 RWNEG_O3 RWNEG_O4 RWNEG_O5 RWNEG_O6 RWNEG_O7 RWNEG_O8	OC4ME_FAIL
Diffuse attenuation coefficient	KD490_M07 KD490_M07 _err			KDM_FAIL
Total suspended matter concentration	TSM_NN TSM_NN_err			OCNN_FAIL
Coloured Detrital and Dissolved Material absorption	ADG443_NN ADG443_NN_err			
Sea Surface Temperature	**l2p_flags**	**quality_level**		
sea_surface_temperature	land ice lake rive exception cloud sun_glint	>= 2 (worst quality of usable data)		
Sea Surface Temperature Level-4	**mask (flags)**			
analysed_sst	sea			
Sea Surface Height Anomaly	**flags**			

ssha_20_ku	surf_class_20_ku surf_type_20_ku surf_type_class_20_ku			

Another thing that needs to be highlighted, as far as the SST is concerned, the measurements with a quality level greater than or equal to 2 were used. This is the threshold of the usable data where 2 denotes the worst quality. It was noticed that in the Gulf Stream there were cloud edge effects on some of the scenes and specific regions of each scene (Fig. 11). For this reason, it was chosen to consider those values as outliers and disregard them.

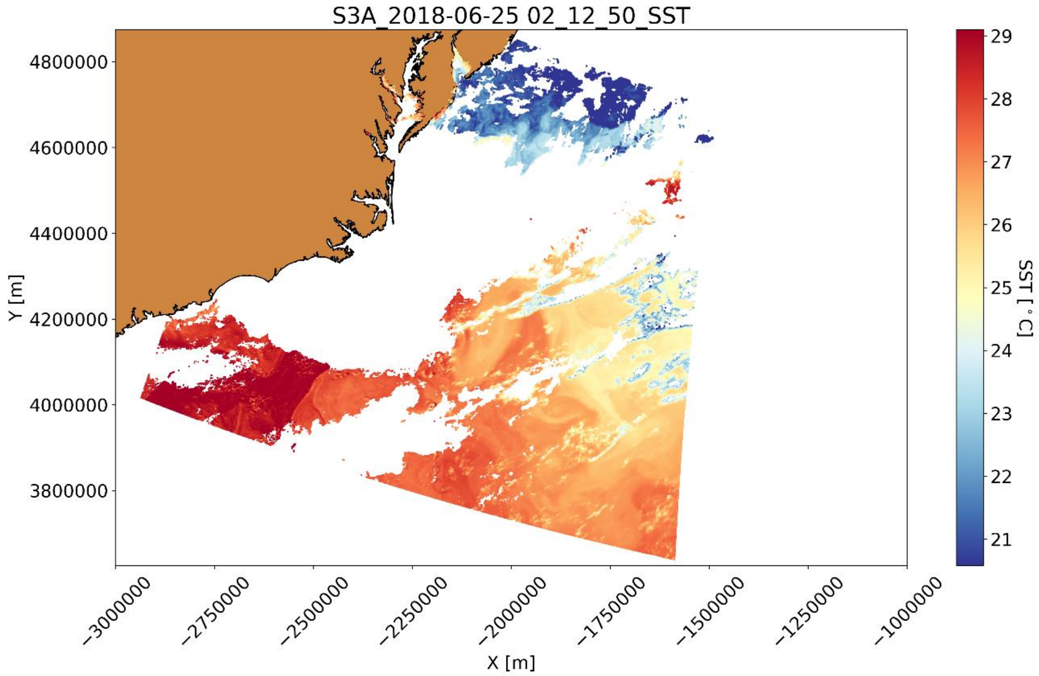


Figure 11: This is an image of the SST acquired by SLSTR on-board Sentinel-3. It is masked using the most conservative flag options possible. However, evidently there are still cloud edge effects at the Eastern side of the swath which are depicted as light blue colours which surround white regions. This is the region of Gulf Stream.

Furthermore, the choice of the chlorophyll product for open waters (CHL_OC4ME) instead of the complex coastal waters (CHL_NN) was based on the interest in the open water mostly. The reason for this is that one of the basic assumptions is that the effect of chlorophyll at open waters is assumed to be more clear, compared to the more complex signal at the coastal regions. The parameter ADG443_NN represents the absorption by detritus plus CDOM (Valente et al., 2016). KD490_M07 represents the diffuse attenuation coefficient for downward irradiance (Valente et al., 2016), which basically shows how easily the light penetrates the open waters. More information about the products can be derived from https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-3-slstr/level-2/sea-surface-temperature-ghrsst (Retrieved on: 04/2019) and EUMETSAT (2018a).

Finally, before moving to further processing the data needed to be transformed into some local projected coordinate reference system (CRS) since distance calculations were realized. The CRS of Gulf Stream and North Sea that were used are NAD83/UTM zone 23N (EPSG: 26923) and ETRS89/LAEA Europe (EPSG: 3035), respectively.

5.2. Data pre-processing and analysis

Before proceeding further, some of the techniques that were widely used must be highlighted and explained. Therefore, every time these concepts are mentioned later, the reader can refer to this section.

5.2.1. Processing

5.2.1.1. Interpolation of SST, SST L4 and OC

One of the very first steps used in bringing the data into an analysis ready form was the interpolation of the SST and OC variables on the SSHA points. This was achieved by using the IDW technique as described in the introduction (Chap. 4.1.5) using a power of 1 (p = 1). Theoretically, the closest 8 neighbours are used, although in practice due to missing data this reduces to 3-5 points that fell in a search radius, which was equal to:

$$radius = SR \cdot \sqrt{2}$$

$$SR, the\ spatial\ resolution\ of\ the\ product$$

$$i.e. 1km\ for\ the\ SST\ (and\ SST\ L4)\ and\ 300m\ for\ the\ OC\ variables$$

The reason for choosing this radius is in order to only include the points that are close enough, but not too far (all within the 3 x 3 window). An idealized sketch can be seen in (Fig. 12).

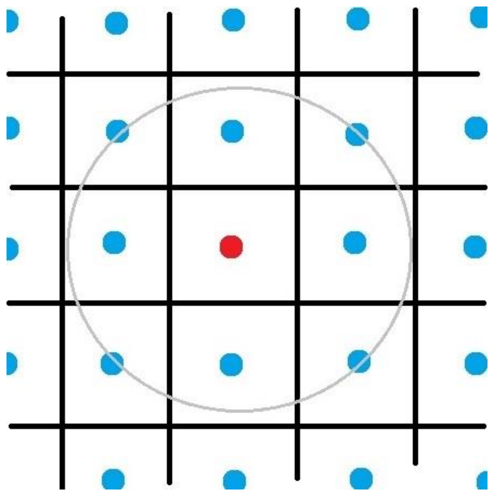


Figure 12: The red point is a point on which we need to interpolate a variable (e.g. SST). The circle denotes the search area and the blue points show the points of the variable. This is the search radius that was used in IDW interpolations on the along-track measurement points.

5.2.1.2. Outlier detection

In this thesis several ways to treat outliers and disregard them were used especially for visualization purposes. Some of them are using the percentiles (e.g. 1% and 99%) and 2.5 or 3 times the standard deviation of each dataset. However, the one that was widely used in order to bring the data into an Analysis Ready form, was the IQR Rule or Tukey's method (Tukey, 1977). Based on this rule of thumb, the outliers are considered those values that are greater than $Q_3 + 1.5 \cdot IQR$ and less than $Q_1 - 1.5 \cdot IQR$. The explanation of the expressions/symbols can be seen below:

$$Q_1, Q_3 = 1st\ and\ 3rd\ quartile$$

$$IQR = Q_3 - Q_1$$

This option was used on every variable.

5.2.1.3. Moving Average filtering

The use of a moving average low-pass filter was chosen because of its simplicity both in understanding and computational complexity, as well as its straightforward implementation. It was used for two purposes. The first purpose was to filter out the high frequency noise (< ~10km), while the second one was to approximate the general (non-linear) trends. The approaches were applied in an exploratory mode mainly on SSHA, SST and Chlorophyll, which were the core variables of this project and, later, on the rest of the variables.

Two moving average approaches were explored in order to arrive at the one that would represent the data better. The first approach did not take into account the missing values. This means that any elements of the arrays with missing values (i.e. Not-a-Number) were disregarded. In turn, this means that in presence of large number of consecutive missing values the elements at the edges would be affected by elements that are physically very far. This is illustrated in Fig. 13 and 14. In addition, zero-padding is necessary at the edges of the array.

The second approach deals with the missing values differently, in the sense that it imputes them instead of disregarding. The missing values were interpolated by using the value of the nearest neighbour that has a value. The same choice holds for the previously zero-padded elements. After imputation of the missing values, the moving average filtering is applied. This approach was applied using the ASTROPY Python package.

The second approach show advantages since the edge effects that take place using the first approach are absent. The edge effects are clear both at the right and left edges of the spatial series, as well as in the intermediate space where many consecutive missing values occur (Fig. 14)

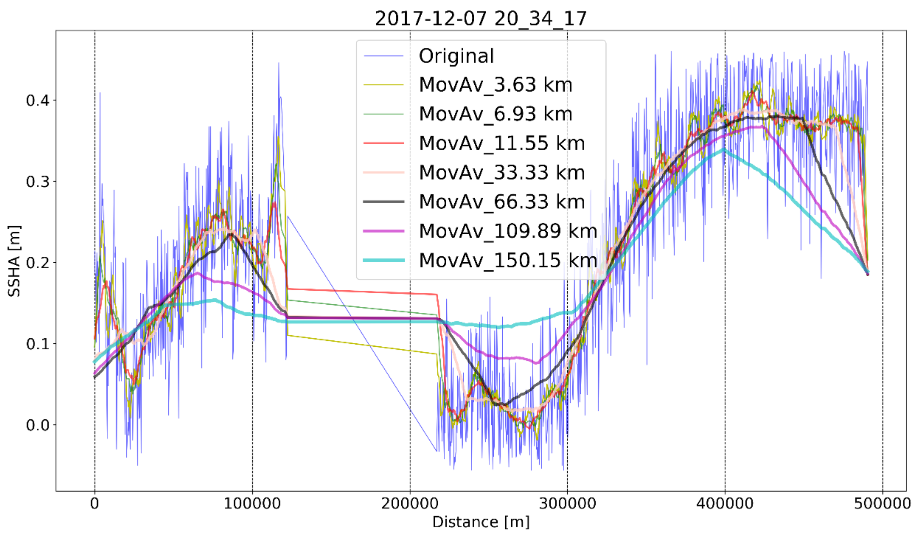


Figure 13: Common moving average with zero-padding at the edges

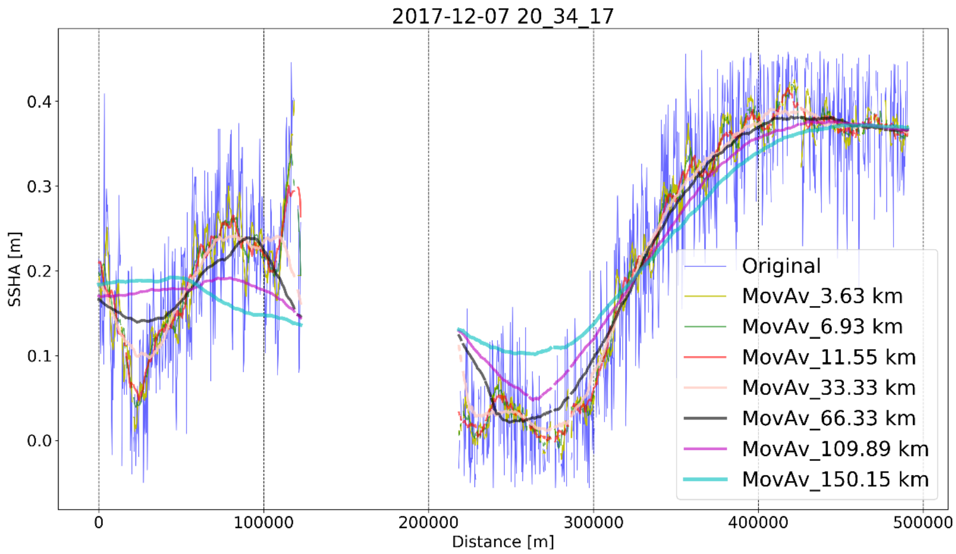


Figure 14: Moving average filtering using the Astropy approach, which treats the missing values differently by interpolating using nearest neighbour.

5.2.2. Remark on number of tracks

At this point, a remark which holds for every processing and analysis step, is that the number of tracks used depending whether it includes SSHA/SST or SSHA/SST/OC was different. In fact, the former dataset was far larger than the latter, because the tracks with common sensing dates needed to be queried. There was a deficit of tracks on specific dates due to exclusion because of e.g. total cloud cover or partial cloud cover that was right at the regions where the ground track is (hence the OC and/or SST did not have measurements there). Due to the differential data deficit, mainly, between SST and OC that are affected by the cloud cover, the chances of finding common sensing dates among SSHA/SST/OC were lower than SSHA/SST. Also, as a consequence of the above, not all of the pairs SSHA/SST and SSHA/OC correspond to the same dates.

5.2.3. Along-track correlations

First of all, the original SSHA data are cleaned from any outliers, and later are smoothed by removing the high-frequencies using the 1D moving average filter with a window size of 99km (or 300 array elements). After that, the large spatial trend is computed again using 1D moving average filter with a larger window size of 297km (or 900 array elements). The meso-scale variability is disclosed by subtracting the large spatial trends from the smoothed data. This is a technique that is has also been used slightly differently by Chelton et al. (2011) and AVISO (2019). Nevertheless, experiments were done on different combinations of window sizes.

A similar approach is followed for the SST. Firstly, the variables are interpolated using IDW on the SSHA along-track points. After that, they are smoothed from the high-frequency noise using a search radius of 50 km (i.e. diameter of 100 km). In addition, the large spatial trend is computed on the original product by using a 2D moving average with a search radius of 150 km (i.e. diameter of 300 km). In the same manner, as with SSHA, the large-scale trend is subtracted from the smoothed version. Two examples of the 2D moving average effect on SST can be seen in Fig. 15, 16 and 17, for the North Sea. Afterwards, the cross correlations of the products that have been acquired at the same days are computed and plotted.

The same approach was followed regarding Chlorophyll-a, TSM and CDM absorption coefficient with the difference of the search radius which was 100 km (i.e. diameter of 200 km). The reason for this choice came out of trial and error and it was noticed that there were less bad relations.

As far as the SST L4 is concerned, since it did not contain any outliers and missing values, as well as it is a spatially complete product it was decided to use it in a slightly different manner. Therefore, in this case, the SSHA data after the relevant pre-processing, they were de-noised by smoothing with a moving average filter using a window size of 11.5 km (or 35 array elements). The SST L4 were directly mapped by using IDW on to the SSHA along-track points and then smoothed using a 2D moving average filter with a search radius equal to 10km. After that, the autocorrelations of both variables were computed, as well as their cross-correlations using a maximum lag distance of 132 km (or 400 array elements). The choice of the maximum lag was based on the assumption that the meso-scale phenomena at the Gulf Stream are more or less stable, therefore even 104km is a conservative choice. The reason that the auto-correlations were computed only on this dataset is that the Sentinel-3 SST ~10 km smoothed version was dominated by missing values, which is a crucial problem for the autocorrelation algorithm. Interpolation would not solve the problem because it would import lots of "random-like" values and abnormalities which would result in weird patterns of abrupt highs and lows.

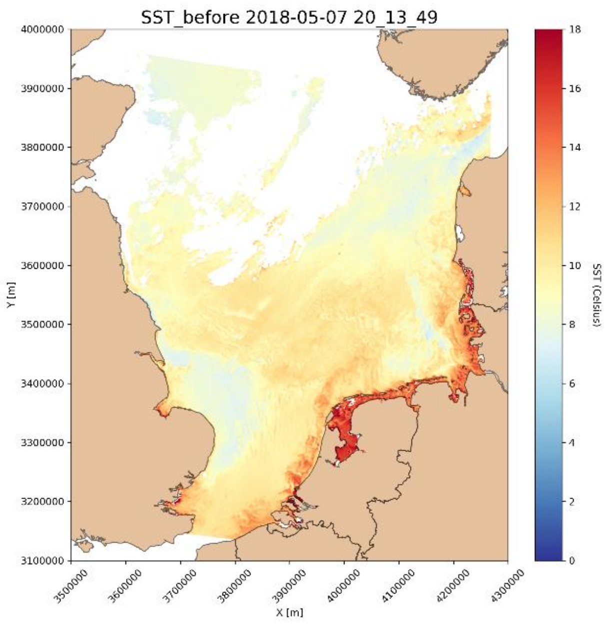


Figure 15: Original SST version at the North sea

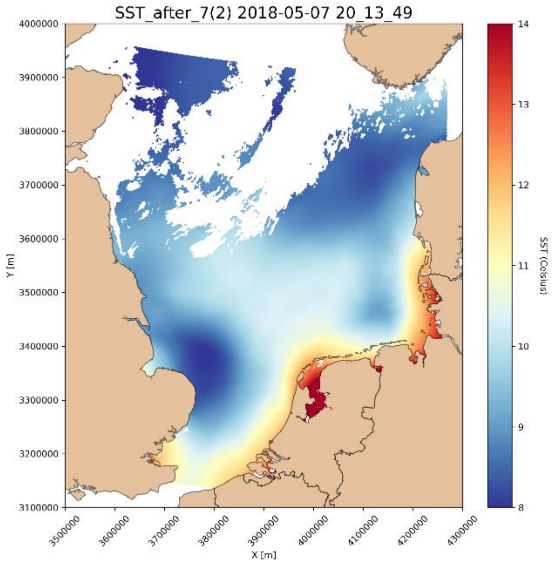


Figure 16: Same scene as in Figure 15, but this time it is 2D filtered using the moving average with a search radius of 50 km (or 100 km diameter)

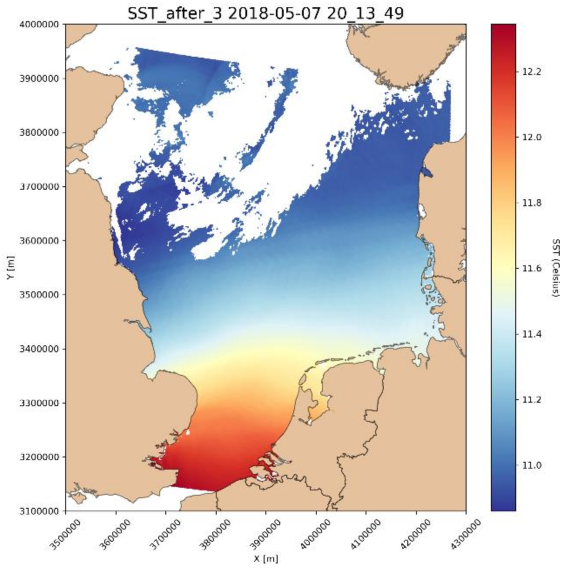


Figure 17: Same as in Figure 15, with a radius of 150 km (or 300 km diameter)

5.2.4. Spatially independent analysis

From now on more specialized machine learning terminology is used which brings the need to make the matches with the thesis dataset. First of all, the terms label and feature(s) are used interchangeably instead of SSHA and SST/OC variables, respectively. For example, the five different versions of SST comprise five features. In addition, the number of samples can be used instead of number of tracks. Eventually, the sample size can be used instead of vector size/length, track size/length and number of observations/measurements of sample.

Spatially independent analysis means that the techniques described below did not consider (at least directly) the fact that all of the variables of interest show spatial autocorrelation. The ML techniques used were two, namely, i) Random Forest (RF) and ii) Multilayer Perceptron (MLP). The preparation these approaches needed was minimal, especially, after bringing the data into along-track array form as described previously. Although the algorithm execution time was significant and imposed time restrictions of the practical part of this thesis.

5.2.4.1. Random Forest

5.2.4.1.1. Complete track archive approach

Firstly, the features needed to be constructed. The features that were chosen after the relatively shallow correlation analysis, were different smoothed versions of the original variables. Practically, this means that for SST and OC variables, 2D moving average was applied multiple times. Regarding the SST, the following search radii were used to generate the various versions; 5 km, 12.5 km, 32 km, 53 km, 95 km, 125 km and 150 km. As far as the OC variables are concerned, the radii used were 5 km, 12.5 km, 32 km, 95 km and 150 km. The reasoning behind the choice of the radii was that they needed to be able to represent the meso-scale phenomena and at the same time not require much computing resources. Therefore, the diameters needed were larger than 10 km which discard the high-frequency variations (micro-scale) and less than 300 km which is a critical landmark for the meso-scale variability and especially eddies (e.g. Fig. 1, AVISO, 2019). Moreover, especially for the OC variables, since the grid size was very dense (~300 m spatial resolution), they needed even more resources and this is why less smoothed versions of it were generated.

As it is highlighted, time required was a major drawback of this approach and owed to the feature generation. The reason is the 2D moving average filtering application with the use of various window sizes. A different approach would probably be faster, if the data were e.g. in a matrix form, so different and more sophisticated image/matrix operation techniques could be used.

After bringing the data into the along-track array form and before using them as input in the Random Forest algorithms, they were further processed. Since they still contained missing values in the middle, as well as at the edges of the timeseries, instead of discarding all of them at once, it was chosen to use one-dimensional (along-track) interpolation/extrapolation on them. The interpolation used was linear, while for the extrapolation Akima technique (Akima, 1970) was used. Both of them were more or less a heuristic choice. Although, Akima was assumed to work better due to its higher complexity and could extrapolate non-linearly since it is a spline based method. Higher complexity does not mean that it is always better, but in the specific case any potential Akima flaws would not have any noticeable impact. The reason for this is that very few missing values were extrapolated, which would not allow for significant impact of possible bad imputed values. After that, the remaining missing values that were not imputed (extra- or inter-polated) were discarded.

Before starting the training using RF, the features were min-max normalized between -1 and 1. The reason for this is that the data consisted of variables that were in different scales. Namely, the SST measurements were at about 20 ^{o}C, while the e.g. chlorophyll's range was between roughly -2 and 1.5 log(mg/m^3). The negative values are owed to the fact that the chlorophyll is an extremely skewed variable, due to its variability from coast to open ocean, therefore the product is offered log-normalized. However, it is still characterized by a highly skewed distribution which would be better to be eliminated before using the variable as input to the RF. Although, it was decided to not be log-normalized further in order to avoid additional errors, thus it was simply min-max normalized.

Regarding the training procedure, the dataset was split into train and test experimenting with the 70/30 and 80/20 proportions. The training features and the label were used to fit the model, and after that the model was used to predict on the training as well as the test features to predict the SSHA. The predicted SSHA was compared to the original (label) SSHA and a linear regression models was fitted and the RMSE was computed.

Furthermore, the most important part was to choose the best hyperparameter values in order to do the most accurate possible estimations. The choice was done by trial and error and the guide to this was the train and test errors, and the computation of the slope, the intercept and the RMSE of the linear regression model. Namely, the best RF model was found using the hyperparameters that are shown in the Table 3. Nevertheless, these are values that are indicative since there is always space for improvement.

Table 3: An indication of the hyperparameters used in the Random Forest complete track archive training are presented below.

Hyperparameter	**Value**
# of estimators (trees)	300
Split Criterion	Mean Squared Error
Maximum Depth	15
Minimum number of sample leaf	8

All in all, an approach similar to leave-one-out cross-validation was implemented. This means that, one track was kept out in the beginning, and the train/test split was implemented on the remaining data. This procedure ran iteratively, and every time the RMSE of train, test and the track that was out were recorded in order to be compared.

The whole procedure as described above, although less thoroughly regarding the hyperparameters, was realized for the SST data only. The only exception was that they were not normalized between -1 and 1 since there was no need.

5.2.4.1.2. Per-track approach

Moving on to the next approach, the same preparation as in the previous sub-chapter was followed in order to bring the data in a form ready for RF training. The same ranges of proportions regarding train/test sets were used. The values of the hyperparameters can be seen in Table 4.

A per-track approach means that a RF model was trained for every track (i.e. satellite pass). This means that the number of RF models is equal to the number of tracks with common sensing dates. The method was applied for both SSHA/SST/OC and SSHA/SST as they were described previously. As one can imagine, the number of observations in every track was far smaller compared to the one used in the previous section. This also explains the use of low values of the hyperparameters, since the RF model did not demand high complexity.

However, the number of features regarding SST and OC was smaller, because several features were observed to be redundant regarding the accuracy of the RF SSHA estimation, thus discarded. This redundancy was observed by looking at the summary statistics of the feature importances of all RF models that is computed after the RF training. This practically shows which of the features contributed the most.

Table 4: An indication of the hyper parameters used in the per-track approach

Hyperparameter	**Value**
# of estimators	6
Split Criterion	Mean Squared Error
Maximum Depth	9
Minimum number of sample leaf	3

5.2.4.2. Multilayer Perceptron

The features used for the MLP was the various SST smooth versions, i.e. different combinations of 5 km, 12.5 km, 16 km, 32 km, 53 km, 75 km, 95 km, 105 km, 125 km and 150 km. The reason for not using the OC variables was that there were memory errors and after the first 2-3 epochs the network ceased. Thus, the three additional smoothed versions were added as features with the hope of yielding any additional information.

The general mentality to building the network was a step-by-step adding of layers and then one-by-one change of some of the hyperparameters. The basic layers that were used were the fully-connected, the activation function and the dropout. The Leaky ReLU activation was used with the alpha parameter equal to 0.8. Other values also were tried, but generally values between 0.6 and 0.9 were enough.

As far as the network's structure is concerned, the MLP depth was increased gradually, whereas there was experimentation regarding the width. Specifically, a structure of a pyramid shape was used, as well as a combination of pyramid and reverse pyramid. A pyramid structure means that the width of the network starts being wide (i.e. many units) and gradually dissipates to one unit which predicts the output layer. The maximum number of units used was 256. The depth of the network was between 1 and 15 pairs of fully-connected/activation layers. Relative to the loss functions MSE, MAE, Huber and Charbonnier were used, with a focus on the last three due to their higher robustness to outliers. The optimizer used was the Adam optimizer (Kingma & Ba, 2014).

Finally, different combinations of proportions between train/test and train/validation sets were used that fall between the 70/30% and 90/10% proportions. The split of the samples into set was realized pseudo-randomly. Also, the number of epochs used was approximately 100.

5.2.5. Spatially dependent analysis using CNN

The spatially dependent analysis means that it takes into account the inherent spatial relationships of each one of the variables. This is achieved by using the CNNs in the one-dimensional domain which exploits the convolution operation. The analysis was carried out on both SSHA/SST and SSHA/SST/OC. Notably, the latter was able to work without any memory issues, the reason being the CNN technique itself.

First of all, before applying the CNN, some preparation was needed. The SSHA data were prepared by detection of outliers without removing them, smoothed with a window size of 11.5km. Same approach as before with the SST, as well as all of the OC variables, with their corresponding search radius regarding interpolation and 2D smoothing. The features (or SST and OC smoothed versions) used were of 5km, 50km and 150km radii. Since CNNs generate the features automatically via convolution and pooling operations, in principle they need few feature engineering procedures, if not at all. Although, in this case it was decided to create some, though very few, features that would be considered representative of every spatial scale.

In order to apply the 1D CNN all of the tracks (or samples) needed to be brought into the same size, since every track has slightly different number of elements (or measurements). The difference in size between them is owed to either masking/cleansing or because they simply represented different scenes, hence belong to a different relative orbit. For this reason, the statistics of all usable track sizes where calculated which can be seen in Figure 18. The statistics were used as a guide to what would be the common vector size to be used which in the end was equal to the maximum vector size (=3126). Practically, this means that a vector (tracks) with size less than 3126 was concatenated with NaNs in order to reach 3126. The same approach was followed for the features based on the maximum vector size.

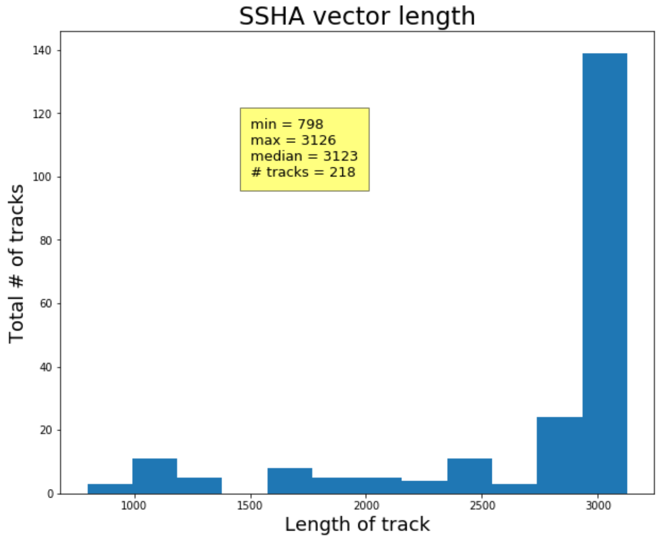


Figure 18: The graph on the left shows the vector size (track size in number of elements) statistical distribution of the satellite tracks. Most of the tracks have a size of over 2750, with a median of 3123. The median was used as a guide metric since the distribution is heavily asymmetric.

The NaNs of the features strictly, needed to be interpolated because the CNN is not robust to missing values. For this reason, several ways to impute them were tried such as zero, average, random from standard normal distribution and, finally, extreme negative values. The assumption behind those choices was that they would not have a high impact in the prediction, thus the result was expected to not be affected much.

The CNN was applied on different combinations of features, with a main focus on the original (~10 km smoothed) data since a basic assumption is that the CNN will create itself the features and choose which of them are important and which are not.

Regarding the structure and composition of the network, the layers that were used were Convolutional, Activation, Average Pooling, (spatial) Dropout and Fully-connected layer. The Average pooling was assumed to work better than the commonly used Max Pooling, because the problem we are dealing with is a regression problem and the variables are continuous, thus an average instead of a max was more appropriate. The spatial dropout differs from the simple dropout in that it discards complete samples (tracks or features) instead of units that is done in fully-connected layers as the functionality is far different between fully-connected and convolutional network types (Tompson et al., 2015).

Concerning the depth and the width, different combinations were tried but the reasoning was the same as with the MLP. Start adding the layers one-by-one and check for the outcome. Generally, the structures tried were a pyramid, i.e. wide network at the top which get narrower towards the bottom, as well as a same-width structure, i.e. all of the layers had the same width. Regarding the kernels used, they were much more in the first layers and get fewer at the bottom layers. The reasoning behind those choices where larger number of parameters is used at the top and less at the bottom, is that the features at the top would have more detailed and small-scale characteristics to be learnt, while at the bottom the larger characteristics such as trends would be learnt.

As far as the activation function is concerned, in the beginning Rectified Linear Unit (ReLU) was chosen (Nair & Hinton, 2010) since it is considered one of the most preferred activation functions (LeCun et al., 2015). It is suitable for regression problems also. However, it is only activated for values greater than zero, which means that it saturates the predictions that have negative values. For this reason, Leaky ReLU (Maas et al., 2013) was used instead, which is a variation of ReLU that can compensate for the aforementioned disadvantage. The only parameter that was tuned is the alpha parameter in the range between 0.5 and 0.8. The rest of its values was observed not to impose any significant impact on the outcome.

Regarding the train/validation/test split, the ratios were kept to higher values because of the very small data set, e.g. for the SSHA/SST it was ~200 samples. Thus, ratios as high as 95/5% were also tried apart from the golden 80/20% rule.

Moving on to the next dataset, the same procedures were applied in the SSHA/SST L4 pair. There was no problem of missing values, except for those that were artificially created with the procedure of bringing the tracks into the same size. Also, this time only the ~10 km data were used. This implies that the number of inputs (or features) was one (i.e. SST L4).

The relevant Python code along with a short manual can be found either on the GitHub repository (profile name: konstantinos2018) or by directly asking the author of this thesis.

6. Results & Discussion

This chapter aims to present the key results and discuss some probable causes why certain techniques and approaches worked or not. Also, out of those that worked it is investigated to what extent they did and which types of problems they can tackle.

6.1. Representativeness in capturing meso-scale phenomena

Before proceeding to the presentation and discussion of the correlations on the along-track between the different variables, it is imperative to discuss about how some meso-scale phenomena look like in the along-track sections and try to make some interpretations. Also, it is important because it will help the reader to be critical and understand about the significance of searching for correlations using the "correct" data, since the widely known saying "garbage in, garbage out" holds true.

In Figure 19 along-track sections with a ~10 km moving average filter for both SSHA and SST can be seen, regarding the Gulf Stream. On Figure 20 the SSHA is the same as before, while the SST has been de-trended with a smoothed version of itself with radius 50 km (or diameter 100 km). Figure 21 shows both SSHA and SST de-trended (between 100 km and 300 km) in order to depict the meso-scale eddies more clearly. The first thing that is noticed in Figure 19 is that without any deep processing the eddies and some smaller spikes that are better seen in a different along-track section (Fig. 22) are clearly visible. The eddies are represented by the sinusoidal-like SSHA signal. What can be noticed is that as one subtracts the larger trends, eddies become more distinguishable, while the 100 km/300 km also eliminates the intermediate meso-scale features (e.g. larger and smaller spikes) (Fig. 23).

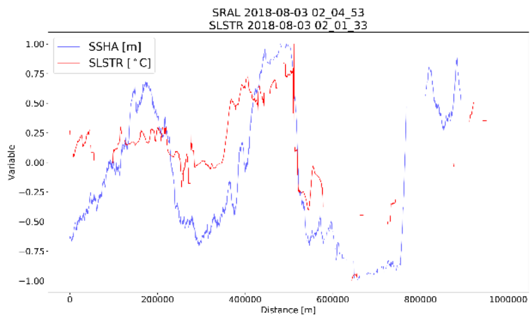


Figure 19: SSHA and SST ~10 km filtered

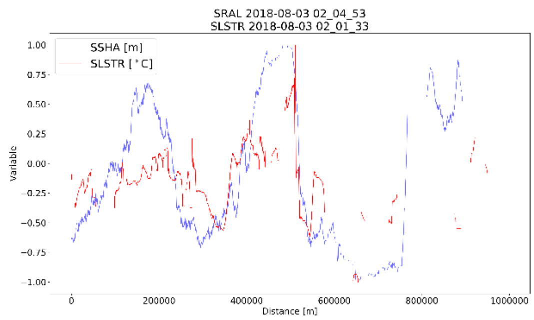


Figure 20: De-trending of SST. 100 km along-track is subtracted from 10km one. The SSHA is ~10 km

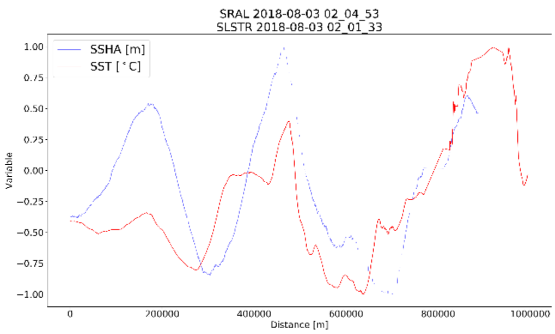


Figure 21: De-trending of SSHA and SST. 150 km along-track is subtracted from 10km one.

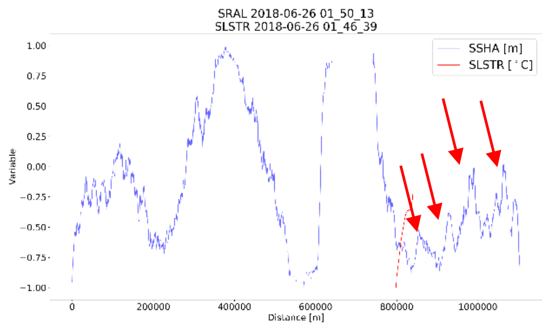


Figure 22: In the ~10km filtered SSHA some spikes in-between the eddies are observed (red arrows) in almost every along-track section.

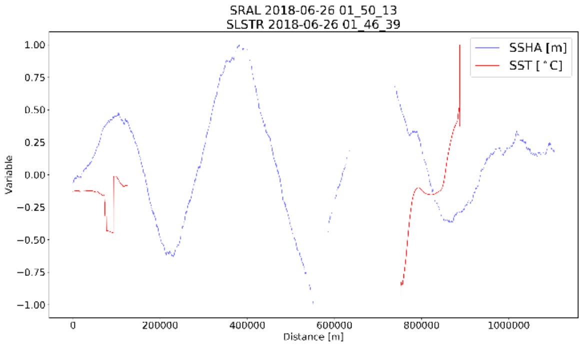


Figure 23: De-trending of SSHA and SST. 300 km along-track is subtracted from 100 km one. This time the absence of the small meso-scale spikes is evident, compared to Figure 22

However, if one observes some equivalent graphs for the North Sea, s/he will not see these patterns as easily (Fig. 24) without processing (Fig. 25). This means that the meso-scale phenomena are not always easy to detect without exploration in different scales and combinations of different scales, since various factors affect their visibility. For example, as mentioned before, in the North Sea the meteorological and atmospheric effects are high which means that the meso-scale signal may be hidden. This leads to further error coming from the SRAL itself, e.g. the Inverted Barometer correction needed might not be the default that SRAL provides. For instance, in the North Sea a different IB correction is needed (Premier, 2017).

With respect to the SST detrending, two probable sources of bias need to be mentioned. The SSHA is moving averaged in the along-track direction only, while the SST is averaged in every possible 2D direction. Thus, in a way the variables do not contain the same directional bias when compared. In addition, on the SST scenes normally a seasonal component should have been subtracted from the SST, however due to the less than one year timespan of the dataset (< 1 year) this was not chosen.

Adding to that, another thing that was noticed is the different sizes of those sinusoidal-like patterns, between Gulf Stream and North Sea, as well as at the very same region itself. This leads to another possible source of bias in the ability to truly depict meso-scale features using the along-track data, seems to be the geometry of the features relative to the satellite ground track. For example, usually eddies are cyclic/elliptic, which means that the ground-track can cross-cut them in different parts and directions. In particular, it plays a role whether it cross-cuts the eddy perpendicular to the major axis or not, or close at the ellipse's edge. Also, assuming that there is a narrow current, it matters whether it is cross-cut perpendicular or parallel to its flow, since the signature on the along-track section will be totally different. If it is cross-cut perpendicular to the flow, then it will be depicted as a relatively narrow area which could easily be an edge of an eddy.

Now, one might wonder why the above matters in the correlation analysis, since the SSHA/SST/OC are measured at the same time they should show the same signature in the along-track which is irrelevant to the geometry of the satellite and the ocean phenomena. The truth is that despite they are measured simultaneously, some of them were discarded due to not passing the quality flags because of the presence of clouds. The result of this was that part of the data were composed of pairs of SSHA/SST and SSHA/OC that belonged to the same sensing date but a different sensing time. To give an indication about the magnitude of the differences, a bar graph can be seen in Figure 26. In fact, 37% of the pairs SSHA/SST have a sensing time difference less than 5 minutes, 50% less than 60 minutes and 55% less than 70 minutes. The rest of them have a sensing time difference of almost 12 hours (Fig. 26).

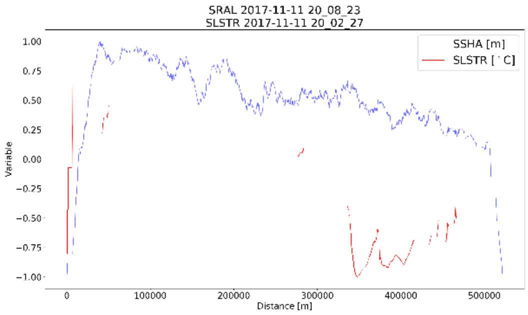


Figure 24: ~10 km smoothed SSHA and SST at North Sea. Eddies or current indications are not present

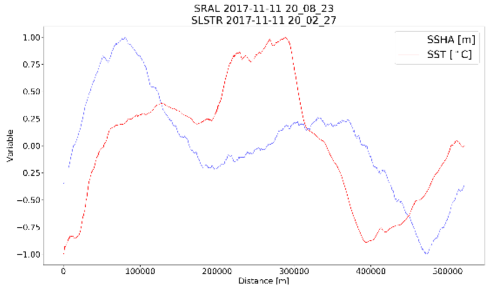


Figure 25: Detrended SSHA and SST (100 km – 300 km) SSHA and SST at North Sea, which reveal some meso-scale phenomena that were not possible to distinguish before (Fig. 24)

At this stage of the thesis, the assumption that this time discrepancy between the products used would not play a negative dominant role in the relation between the SSHA and SST (and OC) at the Gulf Stream, because of the relative stability of the ocean phenomena there such as the current that goes towards the North and turns to the East as well as the largest eddies. However, theoretically this always depends on meteorological effects such as winds, which would increase the speed of ocean flows thus hiding the SSHA/SST relation.

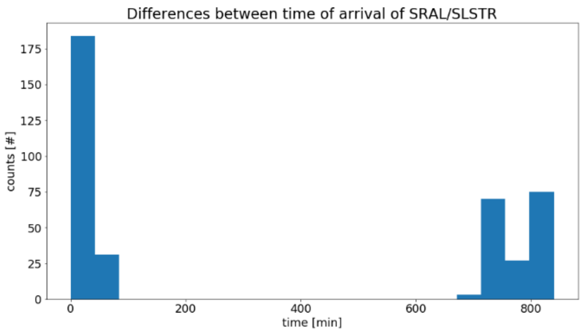


Figure 26: Time differences of satellite arrival time between SRAL (SSHA) and SLSTR (SST) products at the same dates. 55% of the tracks have a time difference less than 70 minutes.

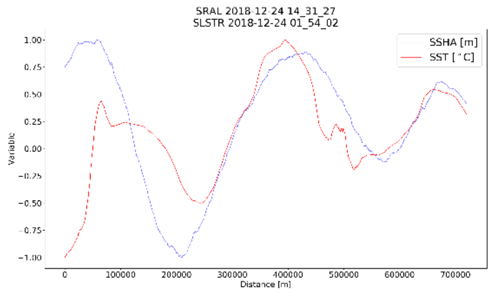


Figure 27: Along-track overlap between SSHA and SST. Remarkably the time difference between the two products is ~12 hours and still a correlation is apparent.

Furthermore, another point to be highlighted is that in the graphs Figure 21 and Figure 25 which depict detrended features, the SST gaps due to clouds and outlier removal that were present in the quasi-original products (Fig. 19, Fig. 24) are now absent. This is an effect of the larger windows (search radii) smoothing which includes values that exist and they are far from the altimetry track. In other words, it is the effect of the neighbours that are far. Therefore, the missing data are decreasing as larger smoothing search radii (or window sizes) are used - which is nicely demonstrated in Figure 52 which is later discussed. Thus, this fact by itself can also import some error, which cannot be quantified.

All in all, another characteristic that is seen in the Figure 22, are the sub-meso-scale spikes that are evident as described above. Those spikes might be either artificial due to the processing or natural. An artificial cause could be the moving average filtering that is generally known to possibly produce such features, while a natural cause could be phenomena such as filaments or other smaller meso-scale phenomena which also seems very probable. This is also one reason why it was chosen to use spatial scales that would eliminate such ambiguous phenomena (i.e. larger than 100 km) that were not always depicted in the SST as was the case with SSHA.

6.2. Along-track Correlations

Some of the pitfalls and the difficulties were discussed that arise from the approaches to depict the meso-scale phenomena as clear as possible in SSHA and SST. However it is logical to assume that the same is true for the rest of the variables. Moving on, the meso-scale relations at the scale between 100 km and 300 km were examined further, mostly qualitatively, with a more focus on the Gulf Stream from now on.

6.2.1. SSHA versus SST, OC

As has already been seen, most of the along-track plots show associations between SSHA and SST to a larger or smaller extent, even when the time difference is half a day (Fig. 27). The same thing holds true for the relationship between SSHA and Chlorophyll, although this is not the case for other variables which were less examined compared to the rest. For instance, in Figure 28 a relation between SSHA and Chlorophyll can be inferred. Although, in the cases where there seems to be a correlation there is a spatial lag between them, which may be owed to the difference in timing. The same can be seen with the TSM (Fig. 29), although, quantitatively speaking the overlap with zero or larger lag was almost absent. An issue that was also noticed regarding the OC variables is that the presence of clouds was higher which imposed problem with working deeply with them, especially in the analysis part.

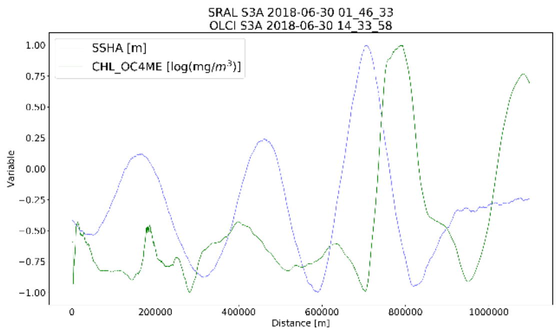


Figure 28: SSHA versus chlorophyll (100-300 km detrend) with half a day sensing time difference

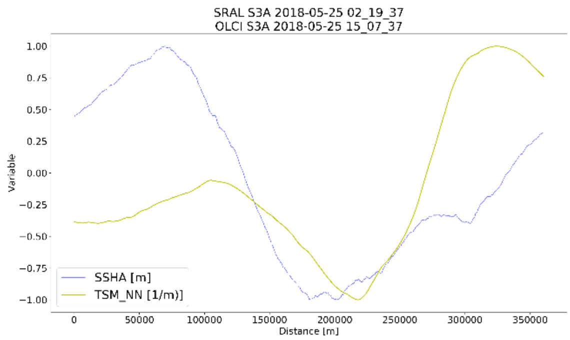


Figure 29: SSHA versus TSM (100-300km detrend) with half a day sensing time difference

Before moving further the scatter plots of SSHA and SST were examined. The majority of them showed weird nonsensical patterns (App. Chap. 9.4), which is something that was expected since the SSHA and SST do not show perfect pattern coincidence.

According to the above, while it is true that meso-scale phenomena correlations between SSHA and SST/Chl are there, a more quantitative approach was also investigated. Namely, the cross-correlations were computed regarding the SSHA and SST, while the OC was left out because of the high percentage of missing values. A pair of a correlogram (Fig. 31) accompanied by an along-track section (Fig. 30) can be seen. The absolute value of the cross-correlation showed that the 47% were above 0.5 which could be taken as not that bad if one thinks of all the possible sources of error and the signal components. In addition, 14% was between 0.5 and 0.6 which means that 33% was over 0.6. Thus, if one can trust the cross-correlations this means that 33% of the data are close to be usable for deeper analysis without considering that the lag is far from zero.

The cross-correlation technique itself that was used, as implied in the last previous sentence might not be trustable enough. This is due to the fact that it is a metric that can only detect linear correlations, and only shows global relationship between the timeseries. Simply put, global means that the algorithm takes the one vector, displaces it relative to the other and computes the cross-correlation metric. In the case of existence of cross-correlations in only a certain region of the time series, then the metric would fail since it would be affected by the rest non-existent correlations. Therefore, by only considering the last two mentioned peculiarities one can assume that these cross-correlation graphs tell us about a small part of the story.

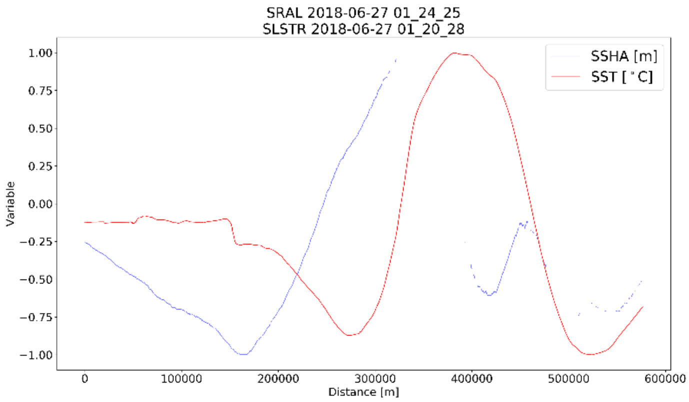


Figure 30: The figure shows the along-track SSHA, SST with the corresponding correlogram below (Fig. 31)

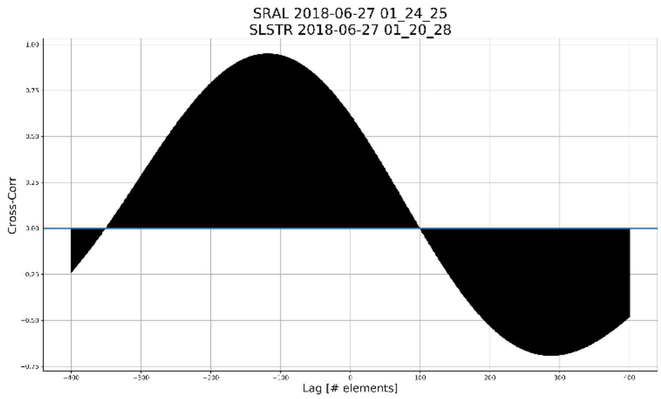


Figure 31: The figure shows the cross-correlation between the SSHA and SST detrended (100 km – 300 km)

6.2.2. SSHA versus SST L4

The importance of significant cloud presence (i.e. missing values) in the dataset is high. Hence, it was speculated that it would pose problems in the machine learning prediction, since it is a well-known obstacle before applying ML techniques which - depending on how one handles it - imports bias, larger or smaller, anyway. All of the ML techniques used here would not suffer from the missing value problem due to bias, however the Convolutional Neural Networks would not work at all, since at their core they are based on the continuity of the observations. This means that missing values cannot simply be omitted.

Despite one can think of many creative ways to deal with the missing values problem, the present work kept it simple and dealt with it with the two following ways concerning the inputs in the CNN. The first way, and the most common was to interpolate the missing values using various approaches, and the second way was to use a different SST dataset, namely, SST Level 4. Therefore, before proceeding to the ML analysis, a brief investigation on the correlations of SSHA and the SST L4 was conducted.

To begin with, two indicative along-track sections of SSHA/SST L4 can be seen in Figure 32, Figure 33. Generally, it was observed that the SST L4 were following the general trend of the SSHA, without excluding absence of collocation of patterns such as in Figure 34. Two things can be observed. The first one is that the SST L4 along-track data seem really smooth if one considers how the Sentinel-3 SST of the same smoothing window look like. The second thing is that in several along-track sections smaller scale periodic signal is observed which is also smooth. These two features might probably owed to the procedures held during the production of the Level-4 product, such as resampling of different products. In addition, by looking at the map (Fig. 35) and comparing it to a map of Sentinel-3 SST (Fig. 36), it can be inferred that the SST L4 does not have a truly 1km resolution, since it cannot resolve smaller scale features as SLSTR can.

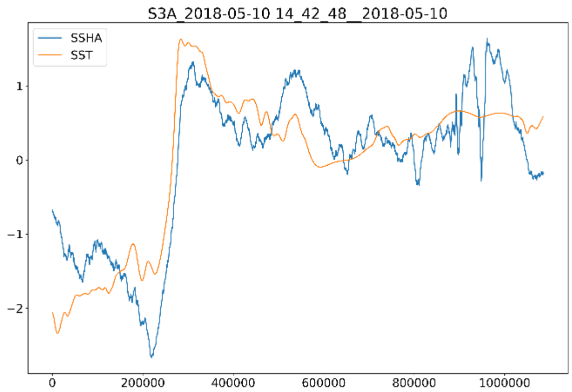


Figure 32: SSHA (~10 km smoothed) versus SST Level-4 original (1 km) along-track section

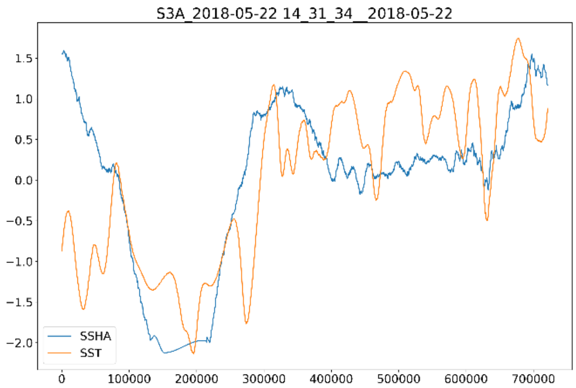


Figure 33: SSHA (~10 km smoothed) versus SST Level-4 original (1 km) along-track section. Despite SST L4 is in the 1 km, it seems rather smooth with a high frequency sinusoidal pattern in it. This is might owed to SST L4 processing schemes that are used in order to create it.

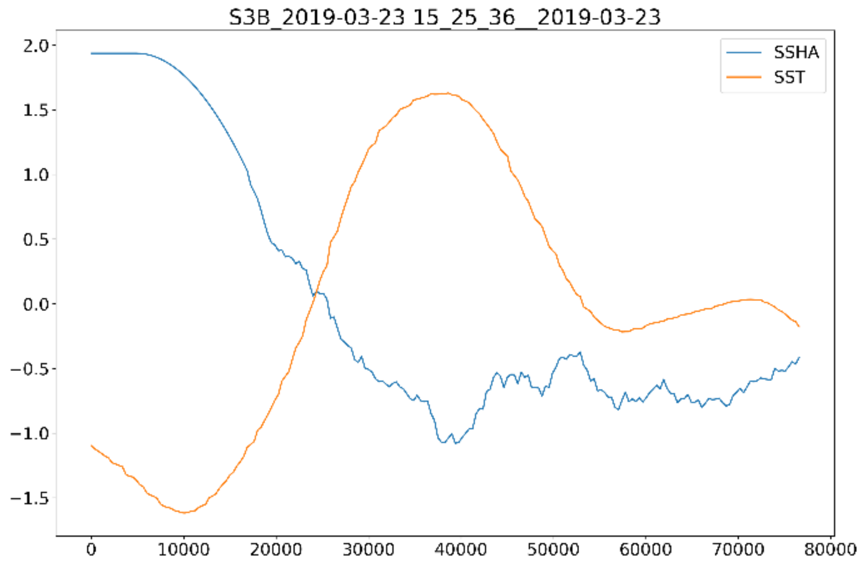


Figure 34: An example of SSHA and SST L4 with absence of collocation of pattern

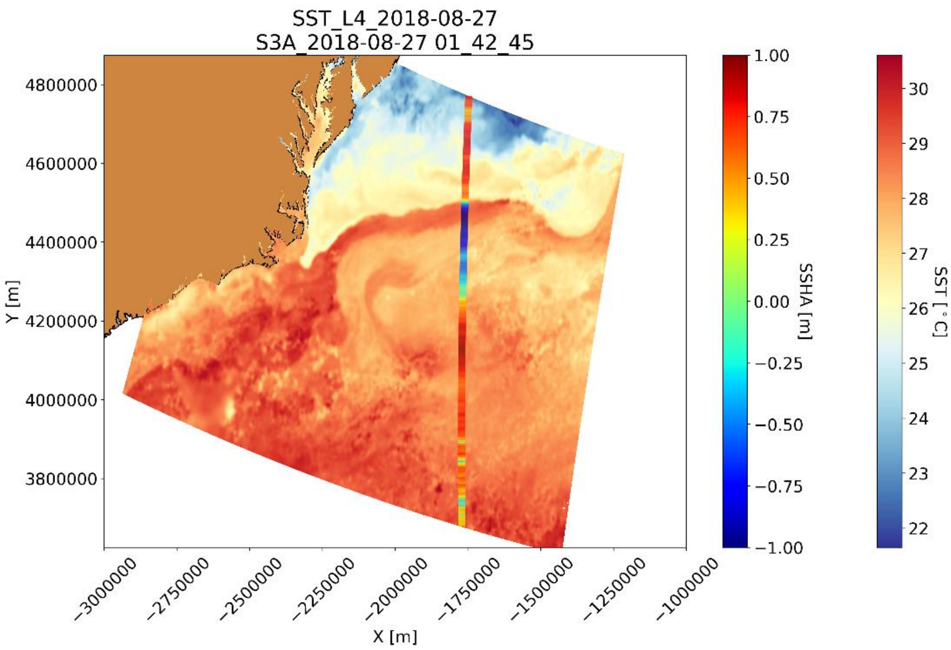


Figure 35: SST L4 in 1 km pixel size. This is a typical example of the difference between pixel size and spatial resolution. The SST L4 product despite having a 1 km grid, the resolving power is lower than this which can easily be compared to the figure below (Fig. 36)

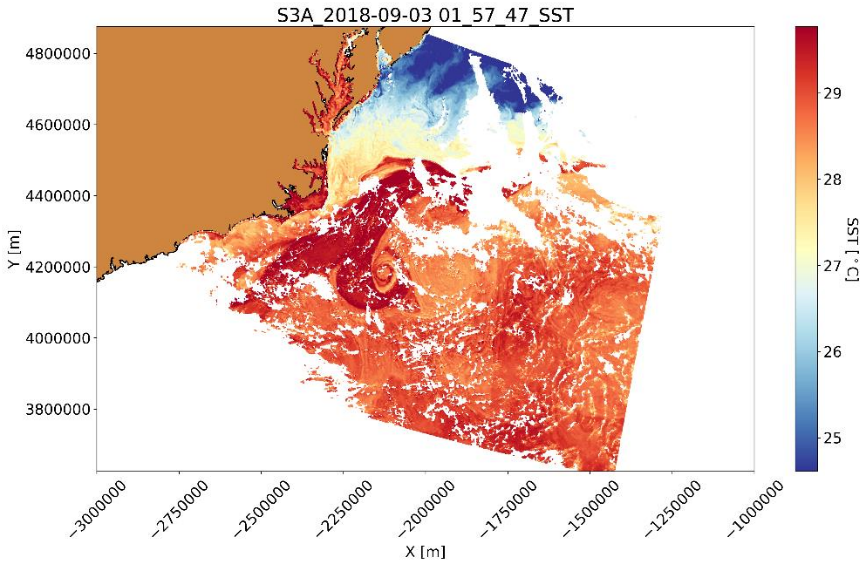


Figure 36: SST of Sentinel-3 three in 1 km resolution

In addition, the auto-correlations and cross-correlations on the ~10 km can be summarized in the Figure 37 and Figure 38, respectively. The auto-correlograms suggest the existence of autocorrelations of each variable, with the SSHA being clearly stronger than the SST L4. In addition, the SSHA show a dissipating periodic autocorrelation pattern which is the signature of the (quasi) periodicity of the eddies/current that was discussed before in the qualitative analysis. These datasets need to be autocorrelated like this since it is used in the CNN regression analysis, which is a prerequisite, although not strictly.

In regards of the cross-correlation a summary graph, an along-track section along with its corresponding cross-correlogram are shown (Fig. 38). In addition, a summary histogram of the maximum positive and maximum negative cross-correlations are also shown (Fig. 39).

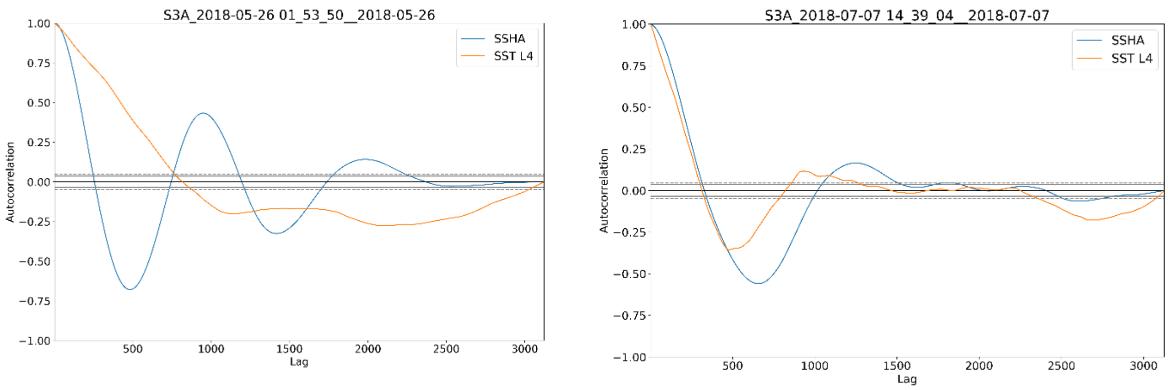


Figure 37: In the above two figures the autocorrelation of SSHA and SST L4 ~10 km smoothed versions can be seen. Both of them show autocorrelations, although the SSHA presents a more sinusoidal pattern which might imply that it can relate to phenomena such as eddies more closely.

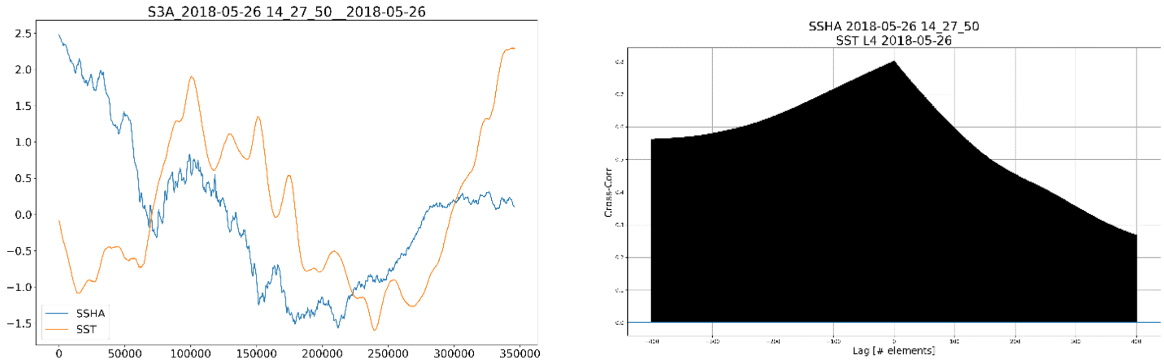


Figure 38: The along-track section of SSHA and SST L4 ~10 km smoothed versions (left), along with the corresponding cross-correlation (right) are depicted.

This histogram is bi-modal with the one mode being over zero while the other being roughly 0.7. This suggests that the dataset is in a way split into two parts. The one part which shows no cross-correlations between SSHA and SST L4, and the other part which shows significant cross-correlations.

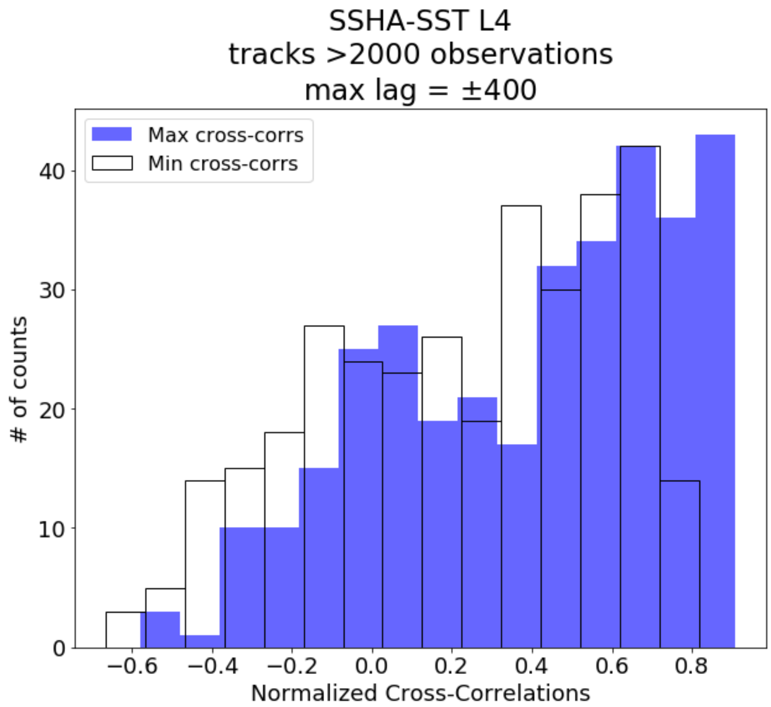


Figure 39: The histogram of maximum and minimum cross-correlations computed regarding the SSHA and SST L4 ~10 km smoothed data. The distribution is bimodal, where the higher mode is over ~0.7, while the lower mode is over 0 correlation. Therefore, most of the cross-correlations seem to be positive. The lag interval and maximum used were 1 element (~0.330 km) and 400 elements (132 km)

6.3. Spatially independent ML analysis

The spatially independent analysis techniques that were used are related to Random Forest and Multilayer Perceptron.

6.3.1. Random Forest

6.3.1.1. Complete track archive approach

The RF model was build based on the complete track archive and it is a global spatiotemporal model in the sense that it uses the complete time series (i.e. every track) where every timeseries refers to a different region of the study area. To make it clear, every single observation is considered to be independent from the other, thus no true spatial or temporal relation between the observations is assumed.

In Figure 40 the plot between the label SSHA and the predicted SSHA of the test dataset and the training dataset can be seen. Apart from that, there is the comparison between the prediction of SSHA for the track that was left out in order to do the leave-one-out cross validation. While in Figure 41 the corresponding along-track section of the SSHA and predicted SSHA can be seen. A very peculiar thing was observed which is illustrated further on a summary statistics histogram (Fig. 42). Despite the fact that 1) both the training and testing errors are acceptably low, 2) the RMSEs of the linear regression between predicted and labels are very low, too and 3) the slope and the intercept of the linear regression are proper, the relation between the predicted and label of the track that was left out seems to be random. This holds true for every such comparison, which raises questions about how this is possible.

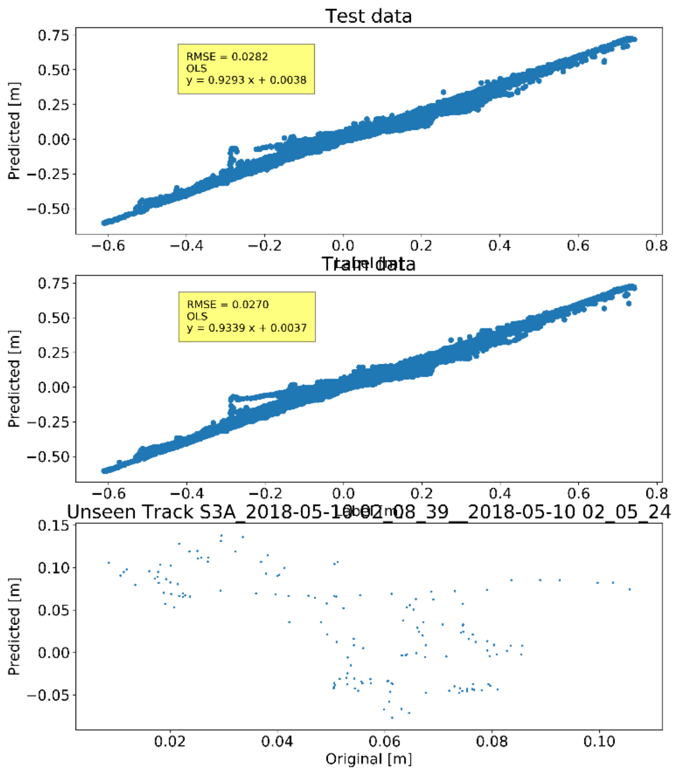


Figure 40: The uppermost and the middle figures, show the SSHA prediction versus the test and train corresponding labels. The methods used is the complete track archive approach using SST and OC smoothed versions as features. The lowermost figure shows the corresponding graph using the track that was left out of the training/testing during the leave-one-out cross-validation procedure. Surprisingly, despite the train and test errors are low without any possibility of overfitting, the unseen track cannot be predicted.

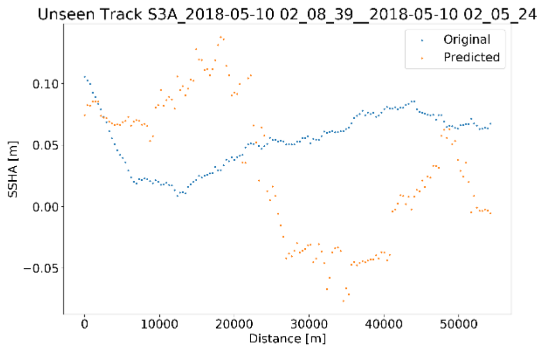


Figure 41: The figure shows the corresponding along-track section of predicted and label SSHA of the Figure 40 (lowermost)

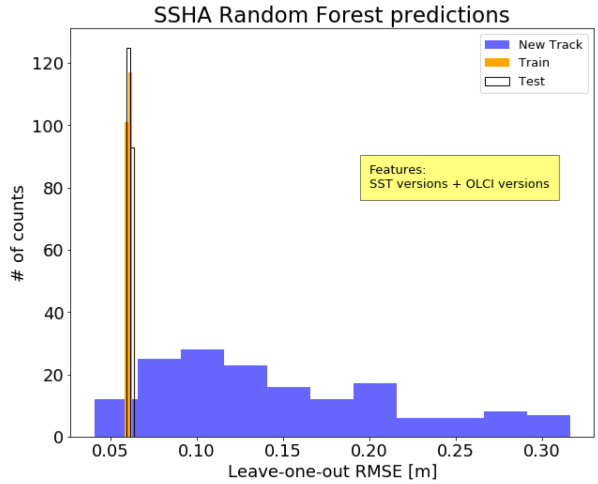


Figure 42: The histogram of training and test errors that were computed in every training session during the leave-one-out cross validation. The unseen track errors in blue, show a huge discrepancy compared to the train and test errors.

The same situation occurs when using the SST smoothed versions also. Although, in this case while the achieved train and test errors are optimal, the regression slope is lower close to 0.8, compared to SST/OC where the slope was ~0.9. Moreover, all of the predicted SSHA seem to follow a pattern and they seem to be far from random in nature.

Having excluded every possibility of technical mistake, there is a hypothesis about this peculiarity. The following facts may play a role with different weights in tricking the ML algorithm to learn things and patterns that are physically non-existent. Firstly, the sensing time difference between the SSHA and the rest of the features might import artificial (fake) relations to a certain extent, which of course the algorithm takes them for granted. Secondly, artificial patterns might be generated, due to smoothing. Thirdly, since there is a general notion that the ML algorithms tend to learn and find patterns which might not exist. In other words, the algorithm could probably detect patterns if one takes the complete data (pairs of SSHA/SST/OC) and shuffle them, without instructing it about the relationship of one observation relative to the other.

The same dataset was used in the MLP algorithm, however as already mentioned there were memory issues when using large feature dataset. However, even with the SST smoothed versions the training was not successful as the training/validation errors were very high with various NN structures. Some indicative prediction/label SSHA along-track sections can be seen, though (Fig. 43).

6.3.1.2. Per-track approach

Moving forward to the next approach, namely the per-track approach, by both using only SST or SST and OC smoothed versions, the models performed well. The difference maker in the accuracy of the results was the smoothed radius of the features.

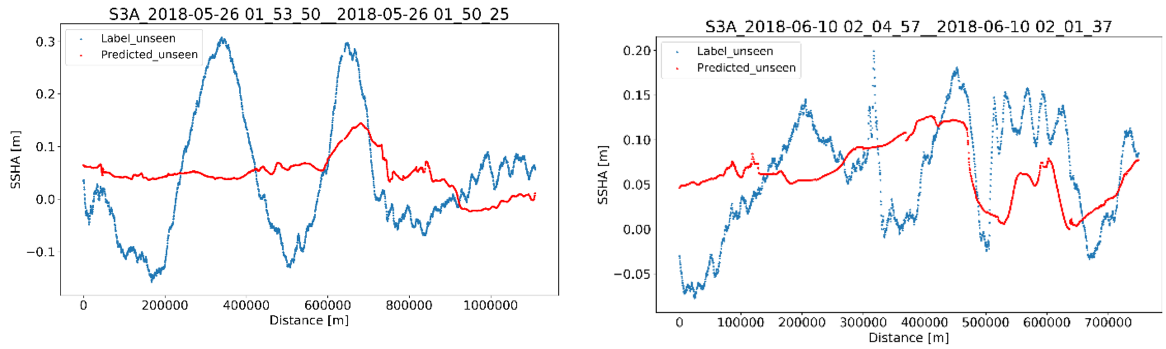


Figure 43: Multilayer perceptron predicted versus label SSHA along-track sections

In particular, beginning with the SSHA and SST dataset, the different smoothed SST versions were examined since it had been anticipated that some of them would be more informative. However, this is an approach that makes the model always depend on the ocean features and their scale that are present at the time of the satellite pass, thus not being very flexible. Consequently, each track (or corresponding model) would need its own features that represent the single situation better. Although, it was speculated that some of the features no matter what, would be more informative since they would represent phenomena of all scales, thus it would lead to a choice of non-redundant features. Indeed by trial and error and by primarily inspecting the feature importance graphs (e.g. Fig. 44), it was found that certain features had higher positive impact on the prediction's accuracy. The training and testing errors can be seen accumulatively in the histograms (Fig. 45) which show the impact of different feature combinations on the RMSEs.

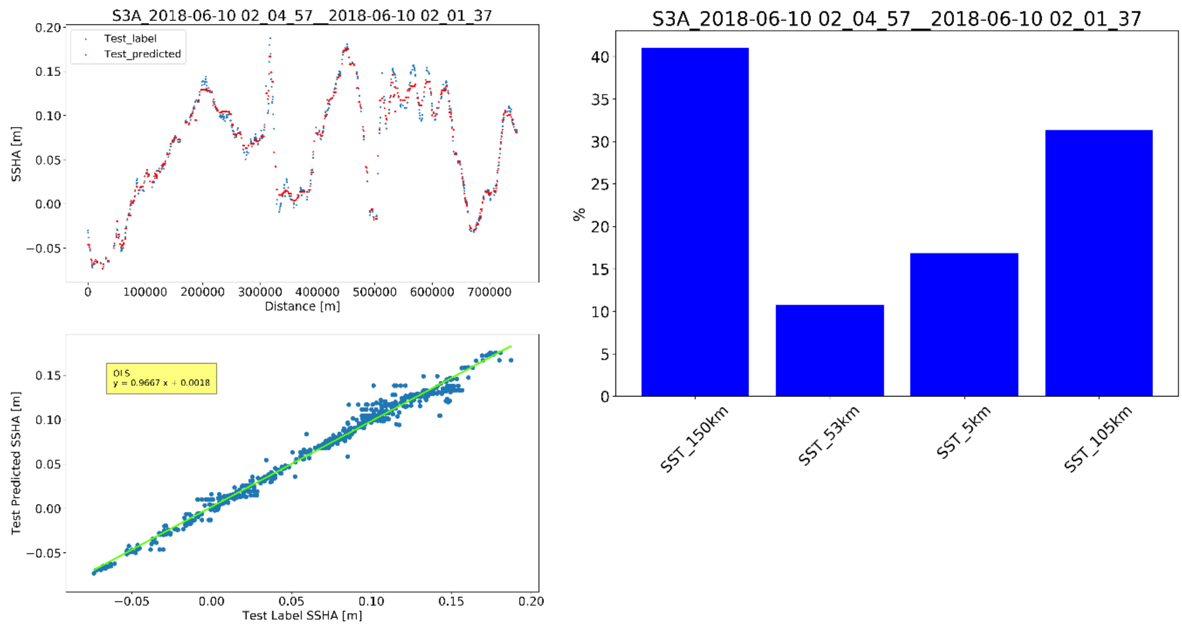


Figure 44: On the left upper graph the predicted test versus label test of SSHA using the RF per-track approach can be seen. On the left lower, the corresponding scatter plot shows a linear trend with a slope ~0.9 and an intercept close to 0. On the right the corresponding feature importance if the features to build the model is shown.

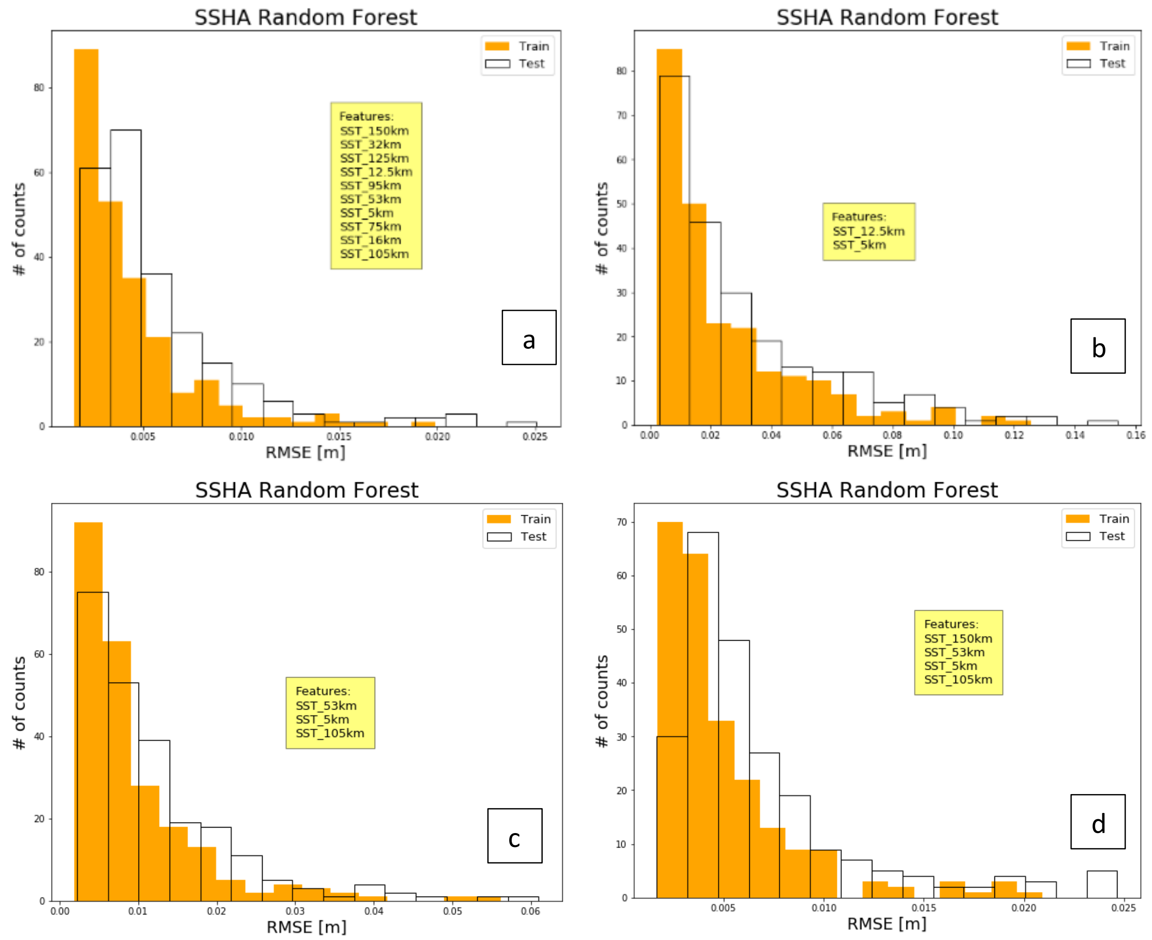


Figure 45: The histograms above show the statistical distribution of the RMSEs of train and test sets, when using different combinations of SST features.

In particular the Figure 45a shows the train/test RMSE distribution when all of the SST features were used which has an acceptably low value with a maximum value close to 2.5 cm. The Figure 45b shows what happens when only the small search radius SST smoothed versions are used. The impact is very big, since the maximum RMSE goes very high close to 16 cm which is not acceptable since the nominal accuracy error of the SSH is ~3.4 cm. The Figure 45c shows the RMSEs produced by using some small (~10 km diameter), medium (~100 km) and somewhat larger (~200 km) scales SST versions. The RMSEs falls compared to the previous histogram, however is still not acceptable since the maximum is about 6 cm. The Figure 45d graph shows the RMSEs produced out of the optimal number of features. This graphs can explain basically the importance of using all possible spatial scales that depict different spatial aspects of the sea level meso-scale phenomena in the range 10 km to 300 km, and at the same time not using redundant features. The RMSE maximum of the last graph is close to 2.5 cm, as a result it is acceptable.

However, this approach needs not to be taken lightly. It is limited in the sense that the relationships that the RF finds in order for them to be true, need the input data to be from the exact same relative orbit (satellite pass), i.e. the timing needs to be the same. Since it is very easy for the RF technique due to its well capabilities to multivariately fit the features successfully, then it means that practically what we see as good performance is a form of memorization. Another argument that might indicate that this is might be a form of memorization is that the RF algorithm finds relationships in every track, which obviously includes tracks of different timing. Additionally, in order to check qualitatively the magnitude of this controversial problem, some models were blindly chosen and applied in the rest of the tracks. What was noticed is that the predictions generally were extremely poor (Fig. 46), something that was expected, although at dates close to the model used the prediction were less poor (Fig. 47). This can be easily explained by the fact that some ocean phenomena at certain scales have various lifespans.

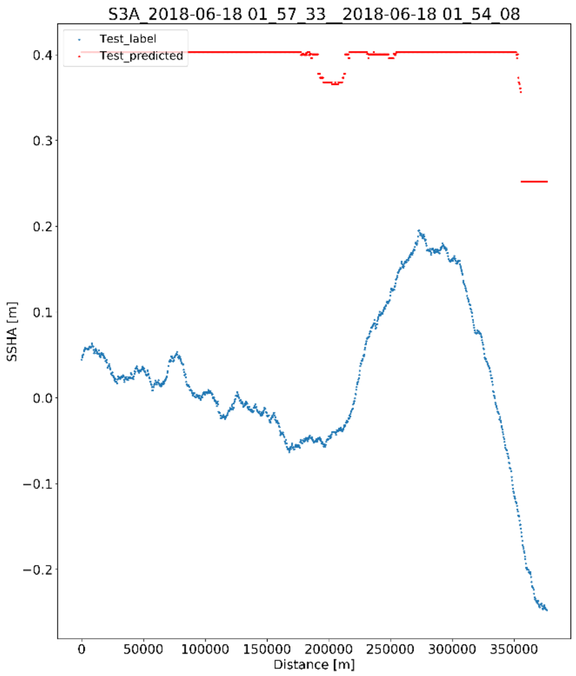


Figure 46: The model used here was built based on the 25-05-2018 track and applied on the date that is seen in the graph, i.e. 18-06-2018. The result seems very bad

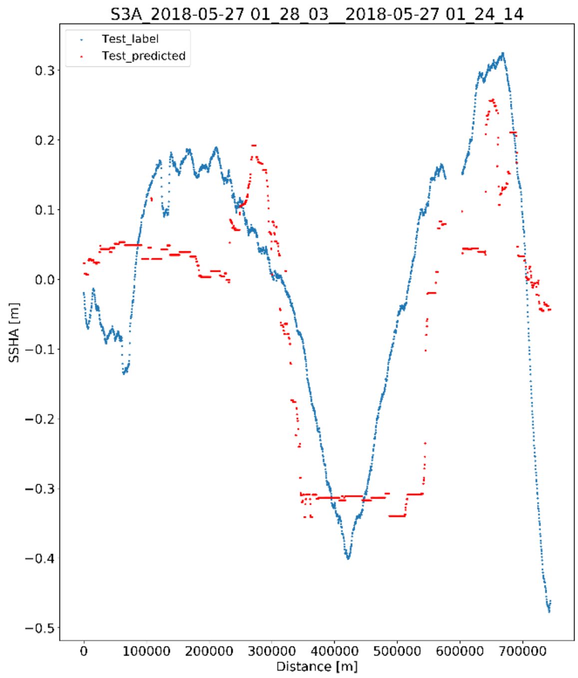


Figure 47: This is the same model as in Figure 46 while the application date that is seen in the graph is on 27-05-2018. The result does not seem too bad compared to Figure 46

Moreover, by looking at two of the produced SSHA maps computed for every track both for SST and SST/OC (Fig. 48, Fig. 49), in some of them one can notice certain saturation regions which can be explained as the inability of the model to extrapolate, since it is trained on non-representative data. This is illustrated in Figure 50 where one can observe that the lower SST values in the North-East part are not sampled at all, thus they are not included in the input features domain. As a result, the aforementioned saturation at that region occurs (Fig. 48). However, by looking at Fig. 49 the same region shows less saturation. This effect is owed to the use of chlorophyll apart from SST, which can be seen in Figure 51. The chlorophyll values at the saturation region can also be found at the area of the altimetry track footprint, which means that these are sampled values (i.e. inside the training domain).

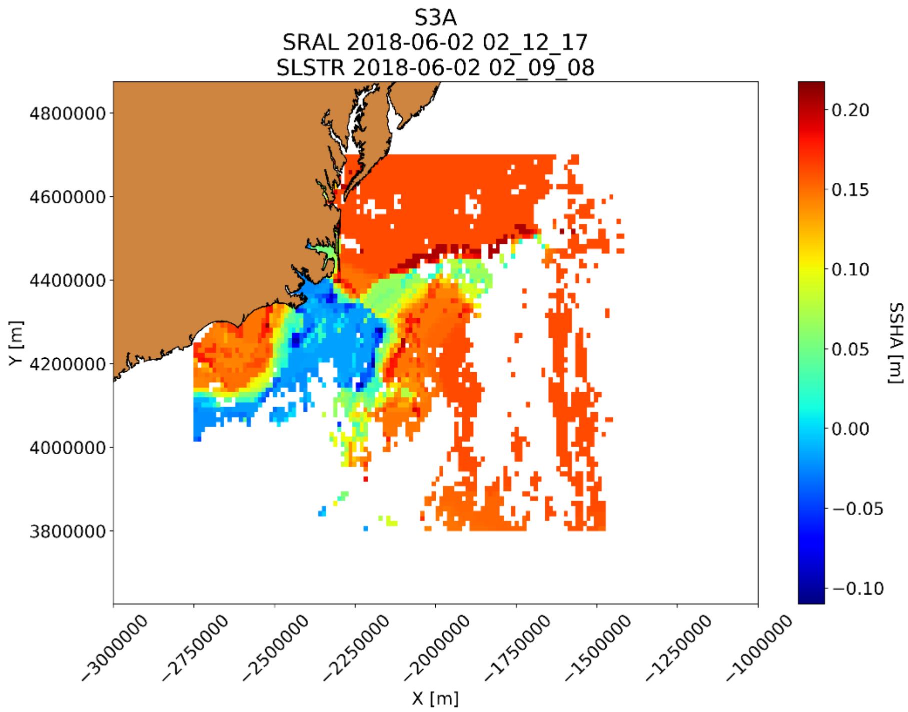


Figure 48: SSHA map prediction based on SST using the per-track RF approach. A saturation area can be noticed at the Northern coasts. This may be an effect of the misrepresentativeness of the data used to train the RF model, hence it is a demonstration of the inability of the model to generalize even on the specific study area at the same date and time.

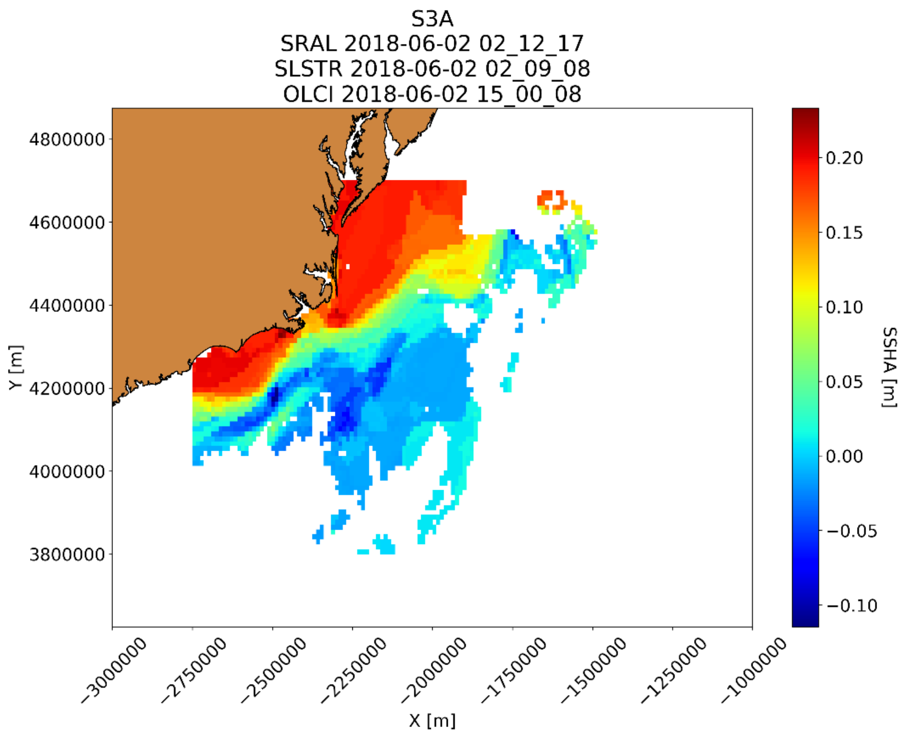


Figure 49: SSHA map prediction based on SST and OC variables using the per-track RF approach. Compared to the above map, the saturation areas are minimized which is a fact probably owed to the additional number of features used (OC variables).

On the other hand, two different arguments can give meaning to the above result instead of considering it as memorization. The first and most straightforward is that by looking at the train and test RMSEs and taking into account that the splitting procedures are random, as well as the test/train proportions are proper, then this cannot be explained as memorization or overfitting. The second argument can be understood better if could be thought of inside a different framework. In particular, if one aims to build a global spatial and temporal model, then this approach is poor. On the contrary, if one aims to create a local spatial and static model then this approach might not be in the wrong direction. For example, an idea is that this could be used for other purposes, e.g. lake ecological studies where the sea level (anomaly) might be of use where the sea level may have a faster response to temperature/evaporation, thus purely spatial approaches like this could be useful.

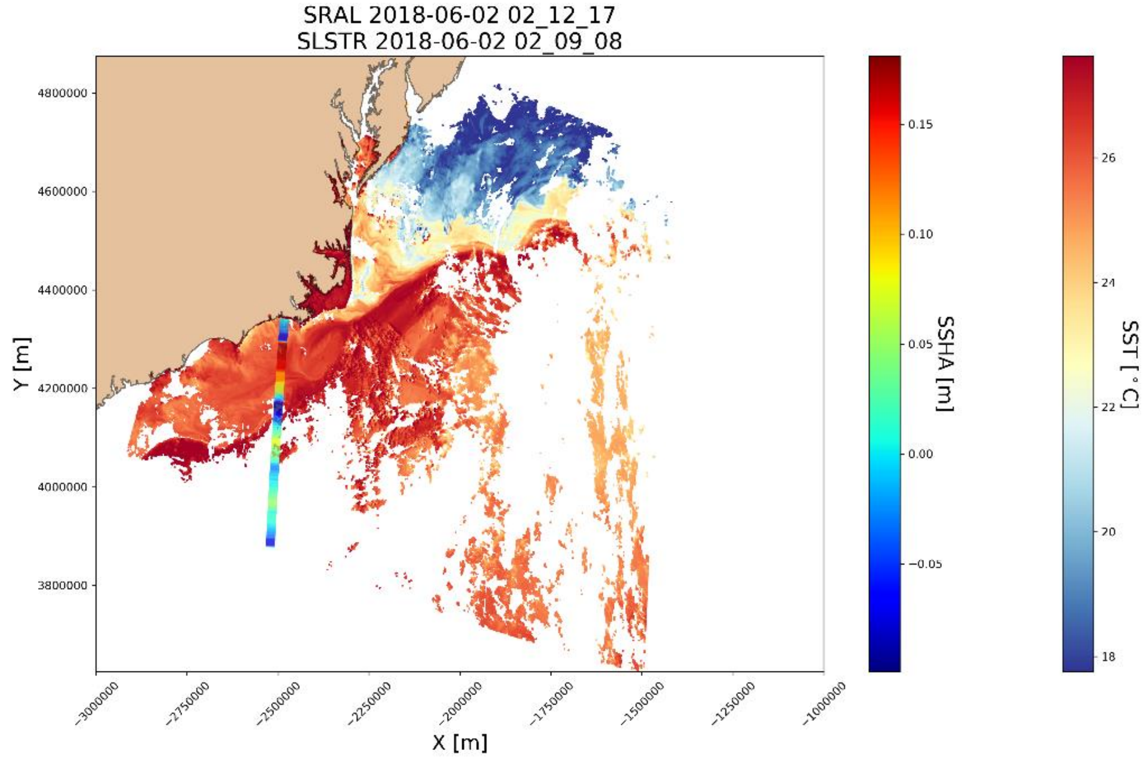


Figure 50: Sentinel-3 SST that was used to produce the SSHA predictions in Figure 48, Figure 49 using the RF per-track training approach.

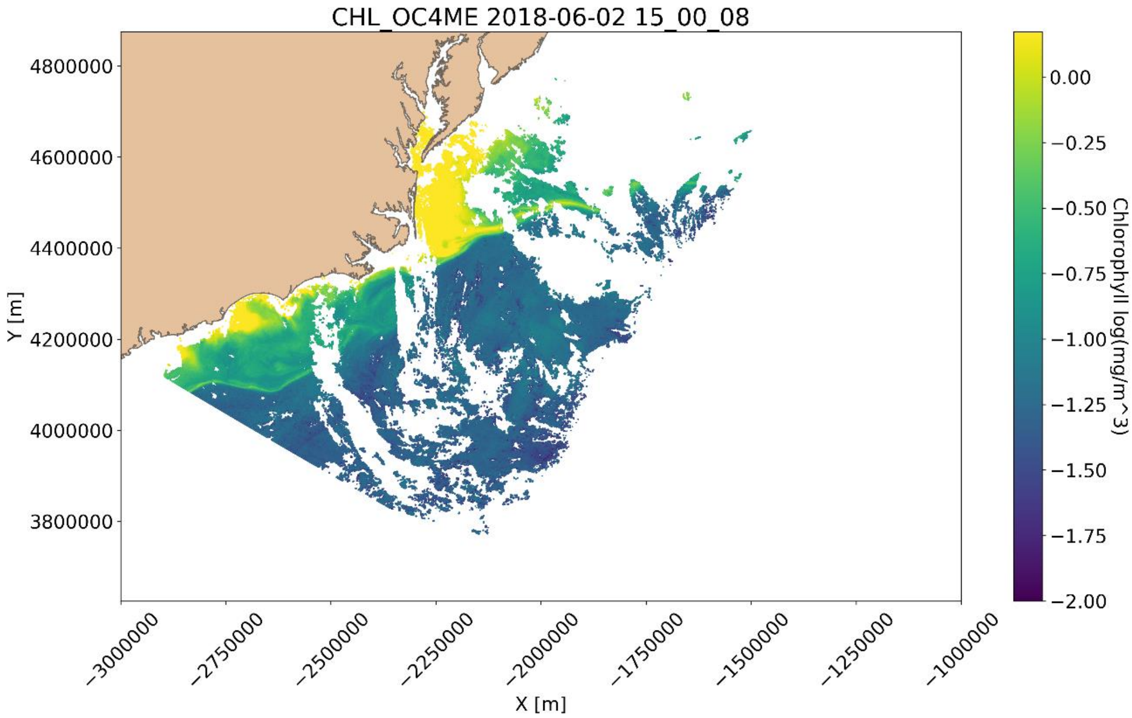


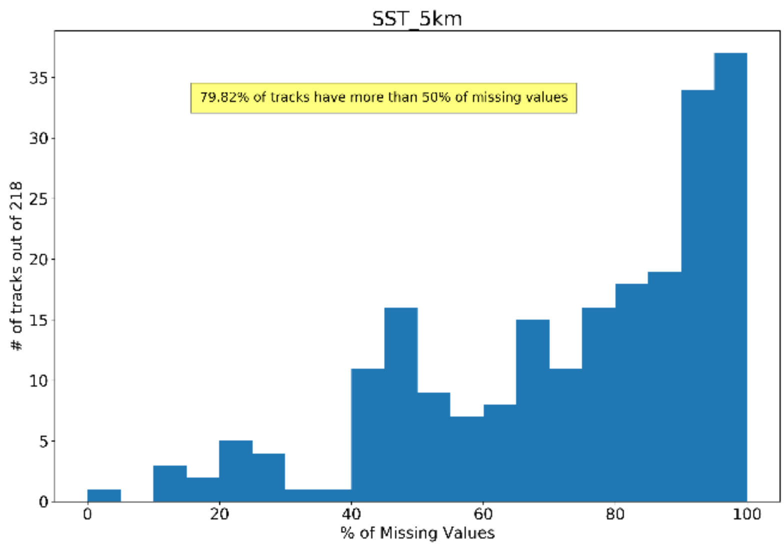
Figure 51: Sentinel-3 chlorophyll (CHL_OC4ME algorithm) that was used to produce the Figure 49 SSHA prediction using the RF per-track training approach.

All in all, in order for the approach to work, the following requirements are needed: 1) the SSHA and SST need to be of the same relative orbit and sensing time 2) there needs to be minimum cloud cover in order to use the best and as many data as possible due to the inherent data size restriction 3) the spatial ocean features signatures on the SSHA and SST do not need to be different very far from the track, otherwise the model cannot generalize to truly unknown data. However, there is one case where the model will always work, which is at the locations close to the ground track.

6.4. Spatially dependent analysis using CNN

The spatially dependent analysis is related to the use of the CNNs which take advantage of the spatial relationships. The 1D CNN was applied on Sentinel-3 data first and on to the SST L4 later.

Regarding Sentinel-3 data, CNN training used the ~10 km smoothed versions of SST and OC variables, as well as just SST by itself. Concerning, SST L4 the smoothed 10 km version was used. As mentioned before, there were many missing values on the SST and OC due to cloud cover. Those values almost exploded after concatenating the vectors with the NaNs in order to bring them at the same size. In order to get a more quantitative view of how big the problem is after concatenating with NaNs the histogram can be examined which pertain to SST (Fig. 52). Thus, the histograms reveal that 4/5 of the SST 5 km, 2/5 of the SST 50 km and 1/5 of the SST 150 km radii contain missing values over 50%. The first number is considered quite large if one needs to use this dataset as input in a CNN. The problem of course becomes larger when those missing values are in consecutive elements of the data array, which implies that any interpolation technique could not alleviate the problem, which is an option that was also tried. In the Appendix, the reader can find relevant histograms regarding the rest of the OC variables (Chap. 9.5).



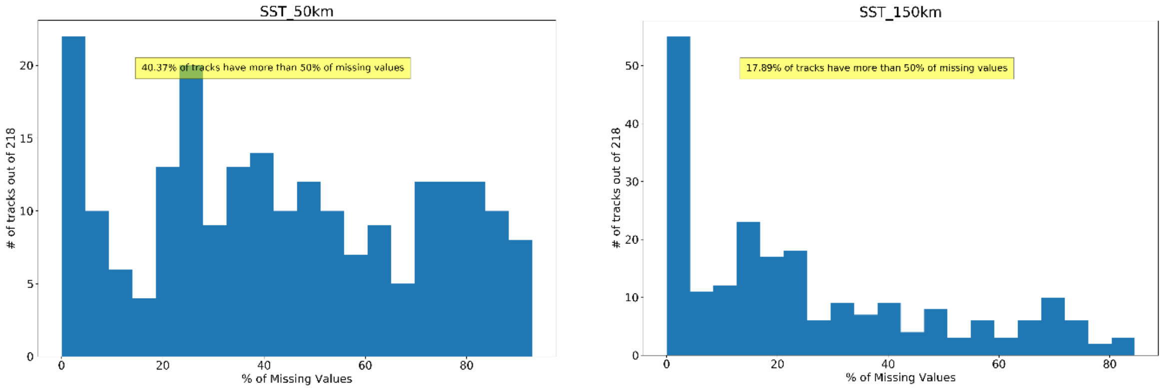


Figure 52: The above histograms depict a summary of the missing values of three smoothed versions of the SST, 5 km, 50 km, 150 km radius. The 5 km radius version show that 80% of the tracks have more than 50% missing values. While the smoothing radius increases, the missing values are get less which is an effect of the moving average filter. As a consequence, the 150 km radius contains 18 % of tracks with more than 50 % missing values.

On the contrary, the training was possible using the SST L4 only, as feature. Some of the train/validation errors with an example of along-track section of SSHA label/predicted can be seen in Figure 53. What is observed is that the validation error in all cases started very low compared to the training errors that for sure the algorithm is not overfitting since the validation error is lower than the training error. However, after a certain epoch the two errors seem to stabilize with the validation been slightly but not significantly lower than the training error. Some possible interpretations of this is that either the CNN architectures that were constructed were not complex enough or that there is too much noise in the feature and label or some other factor.

Finally, in the histograms in Figure 53 the train/test error distributions for the summary of the tracks was computed and plotted in every CNN architecture. These graphs show generally a mode over 13 cm at best, which is not acceptable since the nominal accuracy of the SSH is ~3.4 cm. All in all, it seems that currently the 1D CNN does not lead to many promising results, based solely on these datasets. Of course, there is always room for improvement such as deeper fine-tuning of the hyperparameters or allowing the CNN to run for a prolonged period of time.

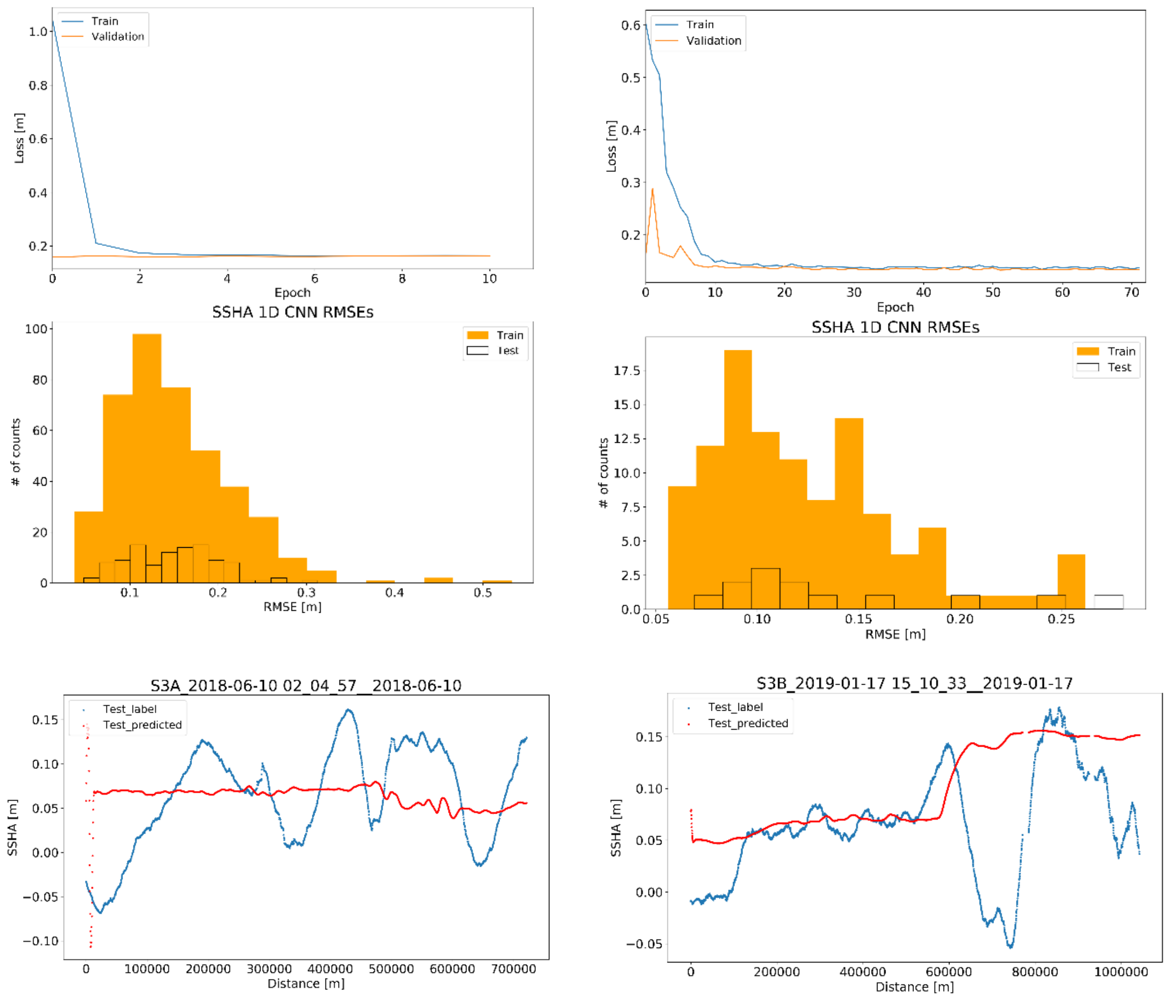


Figure 53: The two graphs in the first row show the train and validation loss of the 1D CNN for two max epochs. The graphs in the second row show the train vs. test RMSEs which are very high compared to the desired ~0.03m. The third row shows the corresponding along-track SSHA predictions.

7. Conclusions

7.1. Research objectives

The main research objectives of this thesis were to investigate the meso-scale correlations between SST, SSHA and OC variables up to a certain extent and, then, use them as input to machine learning algorithms in order to make SSHA estimation in the across-track direction of Sentinel-3. The focus was more on the SSHA and SST, to a lesser extent to chlorophyll and far less to the rest of the OC variables. It was found that the relationships are rather complex, however there seems to be a stronger relation between SSHA and SST, which was not so easy to quantify using the along-track sections.

Furthermore, with respect to the machine learning approaches used, some of them had very bad results and need more understanding on the physics as well as more devotion on the pre-processing. Some other techniques, showed controversial results which would need more investigation in order to understand them deeper. However, some hypotheses were formulated about possible reasons why they had this controversy. One of the techniques, namely the Random Forest in the per-track approach despite its controversy, if it is seen from a different prism it can be considered to work well. This means that if a problem is seen as a purely spatial problem assuming the system is not dynamic, then this approach might work, indeed.

7.2. Suggestions

To put things into perspective, the opinion of the author based on the literature review and the experimental research conducted in this thesis is that there are no shortcut paths in the study of SSHA spatial prediction. This means that machine learning, and especially neural network techniques are not magic wands that blindly solve problems without deep exploration and understanding of the data and their interrelations. For this reason, the following pre-processing approaches to the following problems are suggested.

- The effects of clouds are a big problem and can be partially be tackled by using DINEOF, since even the SST L4 did not seem to alleviate the situation on 1D CNN, although it showed greater associations between temperature and SSHA compared to the Sentinel-3 SST. DINEOF is a spatio-temporal interpolation approach widely used in the marine and atmospheric sciences for gap-filling (e.g. Beckers et al., 2006). In addition a combination of a longer Sentinel-3 data archive with a choice of an area with even less cloud cover and especially during the autumn and winter is suggested.
- The latter suggestion above inevitably leads to better SST pre-processing by subtracting the seasonal component. In general, though, more investigation is needed in spatial and temporal components of every variable.
- The cross-correlations in the way they were used, represented a global relationship along the space timeseries. Especially, with the knowledge that the complexity of the system under study is high, another approach can be followed. For instance, what is proposed is to use a more localized cross-correlation (window time lagged cross-correlation), i.e. define various windows with ascending number of sizes, run the window in the along-track and compute the cross-correlation at defined lags, so only the close by regions will be taken into account. A possible result would be

a coloured matrix which would indicate at which windows and which lags there is high correlation. However, this needs good pre-processing.

- Other sources of data could be used to further validate the possible SSHA prediction accuracies, such as tide gauges close to the coasts and other altimeters in the open ocean.

With regard to the machine learning analysis, the following suggestions are proposed:

- Further work for the per-track approach would be to take SSHA and SST randomly from different dates with a large time span and train a model based on them. Then it will be checked whether a model can be built as easily as happened in this thesis, thus it would indicate to what extent this is a form of memorization.
- The potential memorization extent can also be checked if more than one consecutive swaths were considered in order to train the model. For example, a footprint that would span from the equator towards the pole. In this way, a greater variability of the dataset would be imported.

7.3. Future work

- Possible future work in trying to predict SSHA more accurately could be based on the per-track spatially independent training. For example, the SSHA time series of each relative orbit track could be taken and train a model (e.g. RF) on each one. This means that, the number of the models that will be produced will be equal to the number of relative orbits. In the case that the prediction performance is high, then any point on a grid would have as many SSHA predictions as the number of models. The final SSHA prediction of that point would be some inverse distance weighted mean of every model's prediction. The distance would represent the distance between the query point and the relative orbit
- The autocorrelation graphs of the SSHA seemed to be promising for coastal altimetry applications. The reason for this is that most of altimetry tracks presented moderate to strong autocorrelations which in fact revealed spatially periodic SSHA signal. Taking into account the well-known problem of high errors of SSH close to the coasts, it could be alleviated by an SSHA along-track prediction using some sort of auto-regressive model.
- Since we are dealing with a two-dimensional problem, then the way to tackle it might be to use 2D techniques. An idea is that after having gap-filled data, 2D CNNs could be used. In this was spatial relationship in all directions would be taken into account, thus more realistically approaching the problem. This type of work would be directly comparable to the average weekly SSHA products of AVISO.

8. References

Akima, H. 1970. A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the Assosiation for Computing Machinery*, *17,* 4. pp. 589–602.

Aschbacher, J., Milagro-Perez, M., P. 2012. The European Earth monitoring (GMES) programme: Status and perspectives. *Remote Sensing of Environment,* 120. pp. 3-8.

AVISO. 2019. Mesoscale Eddy Trajectory Atlas Product Handbook, SALP-MU-P-EA-23126-CLS, issue 3.0

Barron, J., T. 2019. A general and Adaptive robust loss function. *Computer Vision and Pattern Recognition.* pp. 4331-4339.

Beckers, J., M., Barth, A., Alvera-Azcarate, A. 2006. DINEOF reconstruction of clouded images including error maps. Application to the Sea Surface Temperature around Corsican Island. *Ocean Science* 2, 2. pp. 183-199.

Behzad, S. 2014. Predicting the Trend of Land Use Changes Using Artificial Neural Network and Markov Chain Model (Case Study: Kermanshah City). *Research Journal of Environmental and Earth Sciences*, 6. pp. 215-226.

Booth, E., Mount, J., Viers, J., H. 2006. Hydrologic variability of the Cosumnes River Floodplain. *San Fransisco Estuary and Watershed Science,* 4, 2.

Bouffard, J., Vignudelli, S., Cipollini, P., Menard, Y. 2008. Exploiting the potential of an improved multimission altimetric data set over the coastal ocean. *Geophysical Research Letters*, 35.

Boyce, D., G., Lewis, M., R., Worm, B. 2010. Global phytoplankton decline over the past century. *Nature*, 466. pp. 591–596.

Brasseur, P., Bahurel, P., Bertino, L., Birol, F., Brankart, J., M., Ferry, N., Losa, S., Remy, E., Schroter, J., Skachko, S., Testut, C., E., Tranchant, B., van Leeuwen, P., J., Verron, J. 2005. Data assimilation for marine monitoring and prediction: The Mercator operational assimilation systems and the MERSEA developments. *Quarterly Journal of the Royal Meteorological Society,* 131. pp. 3561–3582.

Breiman, L. 2001. Random forests. *Machine Learning,* 45. pp. 5–32.

Chao, Y., Li, Z., Farrara, J., D., Huang, P. 2009. Blended sea surface temperatures from multiple satellites and in-situ observations for coastal oceans. *J. Atmos. Ocean. Technol.*, 26. pp. 1415-1426.

Chelton, D., B., Schlax, M., G., Samelson, R., M. 2011. Global observations of nonlinear mesoscale eddies. *Progr. Oceanogr*., 91. pp. 167–216.

Chou, Y. 1975. *Statistical Analysis.* Holt International. ISBN 0-03-089422-0

Cipollini, P., Calafat, M., F., Jevrejeva, S., Melet, A., Prandi, P. Monitoring Sea Level in the Coastal Zone with Satellite Altimetry and Tide Gauges. *Surv. Geophys.* 38. pp. 33-57

Conrad, C., P. 2013. The solid Earth's influence on sea level. *Bull. Geol. Soc. Am.*, 125. pp. 1027-1052.

Cummings, J., A. 2005. Operational multivariate ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society,* 131. pp. 3583–3604.

Davis, R., E., 1991: OBSERVING THE GENERAL-CIRCULATION WITH FLOATS. *Deep Sea Research Part I: Oceanographic Research Papers*, 38**.** pp. 531-571.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M., H., Femenias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., Nieke, J., Rebhan, H., Seitz, B., Stroede, J., Sciarra, R. 2012. The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. *Remote Sensing of Environment* 120. pp. 37-57.

Ducet, N.*,* Le Traon, P., Y.*,* Reverdin, G. 2000. Global high resolution mapping of ocean circulation from the combination of TOPEX/POSEIDON and ERS-1/2*. J. Geophys. Res.,* 105, C8. pp. 19477–19498.

Eleveld, M., Pasterkamp, R., van der Woerd, H., Pietrzak, J., D. 2008. Remotely sensed seasonality in the spatial distribution of sea-surface suspended particulate matter in the southern North Sea. *Estuarine, Coastal and Shelf Science,* 80. pp. 103-113.

Elsberry, R., L., Garwood, R., W., Jr. 1980. Numerical Ocean Prediction Models-Goals for the 1980s. *Bulletin American Meteorological Society,* 61, 12. pp. 1556-1566.

EUMETSAT. 2017. Sentinel-3 SRAL Marine User Handbook
EUMETSAT. 2018a. Sentinel-3 OLCI Marine User Handbook
EUMETSAT. 2018b. Sentinel-3 SLSTR Marine User Handbook
European Space Agency (ESA). 2017. Sentinel-3 User Handbook. Available at: https://earth.esa.int/documents/247904/685236/Sentinel-3_User_Handbook iss1_v1_20170113/960ff616-87f5-43cc-b23e-3e96030bd13a. Retrieved on: 26/09/2018
Evensen, G. 2006. *Data Assimilation: The Ensemble Kalman Filter*. Springer. pp.280.
Frederikse, T., Riva, R., Kleinherenbrink, M., Wada, Y., van den Broeke, M., Marzeion, B. 2016. Closing the sea level budget on a regional scale: Trends and variability on the Northwestern European continental shelf. *Geophys. Res. Lett.* 43. pp. 10864-10872.
Frederikse, T. 2018. Sea-level changes on multiple spatial scales: estimates and contributing processes. PhD Dissertation, Delft University of Technology, The Netherlands.
Freund, Y., Schapire, R., E. 1999. Large margin classification using the perceptron algorithm. *Machine Learning*, 37, 3. pp. 277-296.
Fu, L., L., Alsdorf, D., Morrow, R., Rodriguez, E. 2012. SWOT: The Surface Water and Ocean Topography Mission. Wide-Swath altimetric Measurement of Water Elevation on Earth. Retrieved from: https://swot.jpl.nasa.gov/documents.htm. Retrieved at: 06-2019
Fu, L., L., Cazenave, A. 2001. *Satellite Altimetry and Earth Sciences. A Handbook of Techniques and Applications*, 69. Academic Press, San Diego, California. ISBN 0-12-269545-3.
Fujii, Y., Kamachi, M. 2003. Three-dimensional analysis of temperature and salinity in the equatorial Pacific using a variational method with vertical coupled temperature-salinity EOF modes. *Journal of Geophysical Research,* 108, C9. pp. 3297.
Fukumori, I. 2002. A partitioned Kalman filter and smoother. *Monthly Weather Review,* 130. pp. 1370–1383.
GHRSST Science Team. 2010. The Recommended GHRSST Data Specification (GDS) 2.0, document revision 4. pp. 123.
Gubner, J., A. 2006. *Probability and Random Processes for Electrical and Computer Engineers.* Cambridge University Press. ISBN 978-0-521-86470-1.
Hahnloser, R., Sarpeshkar, R., Mahowald, M., A., Douglas, R., J., Seung, H., S. 2000. Digital Selection and Analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405. pp. 947-951.
Halliwell, G., .R, Bleck, R., Chassignet, E., P., Smith, L., T. 2000. Mixed layer model validation in Atlantic Ocean simulations using the Hybrid Coordinate Ocean Model (HYCOM). *EOS*, 80, OS304.
Halliwell, G., R., Jr., 1998. Simulation of North Atlantic decadal/multi-decadal winter SST anomalies driven by basin-scale atmospheric circulation anomalies. *J. Phys. Oceanogr.*, 28. pp. 5-21.
Hastie, T., TIbshirani, R., Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer, New York, NY. ISBN 978-0-387-84858-7.
Hausmann, U., Czaja, A. 2012. The observed signature of mesoscale eddies in sea surface temperature and the associated heat transport. *Deep Sea Res. Part I, Oceanogr. Res. Papers*, 70. pp. 60–72.
Hawkings, D., M. 1980. Identification of Outliers. *Chapman and Hall,* London-New York. ISBN 978-94-015-3994-4.
Held, I., M., Pierrehumbert, R., T., Garner, S., T., Swanson, K., L.. 1995. Surface quasi-geostrophic dynamics. *J. Fluid Mech.*, 282. pp. 1–20.
Hoge, F.,E., Vodacek, A., Swift, R., N., Yungel, J., K., Blough, N., V., 1995. Inherent optical properties of the ocean: retrieval of the absorption coefficient of chromophoric dissolved organic matter from airborne laser spectral fluorescence measurements. Applied Optics. 34, 30. pp. 7032–7038.
Huber, P., J. 1964. Robust Estimation of a Location Parameter. *Annals of Statistics,* 53, 1. pp. 73-101.
Isern-Fontanet, J., Lapeyre, G., Klein, P., Chapron, B., Hecht, M., W. 2008. Three-dimensional reconstruction of oceanic mesoscale currents from surface information. *J. Geophys. Res.*, 113, C9. pp. 1-17.

Jelesnianski, C., P., Chen, J., Shaffer, W., A. 1992. SLOSH - Sea, Lake, and Overland Surges from Hurricanes. U.S. Department of Commerce, NOAA Tech. Report NWS 48. pp. 71.

JPL OurOcean Project. 2010. GHRSST Level 4 G1SST Global Foundation Sea Surface Temperature Analysis. Ver. 1. PO.DAAC, CA, USA. Dataset accessed [2019-07-17]

Ketkar, N. 2017. Deep Learning with Python: A hands-on introduction. *Apress,* India. ISBN 978-1-4842-2766-4.

Kingma, D., P., Ba, J. 2014. Adam: A Method for Stochastic Optimization. *ICLR*

Kun Il Park. 2018. *Fundamentals of Probability and Stochastic Processes with Applications to Communications.* Springer. ISBN 978-3-319-68074-3.

Kuragano, T., Kamachi, M. 2000. Global statistical space-time scales of oceanic variability estimated from the TOPEX/POSEIDON altimeter data. *J. Geophys. Res.* 105. pp. 955-974.

Le Roy, Y., Deschaux-Beaume, M., Mavrocordatos, C., Aguirre, M., Heliere, F. 2007. SRAL SAR Radar Altimeter for Sentinel-3 mission. *IEEE International Geoscience and Remote Sensing Symposium*. 23-28 July, Barcelona, Spain.

Le Roy, Y., Deschaux-Beaume, M., Mavrocordatos, C., Borde, F. 2010. SRAL, A radar altimeter designed to measure a wide range of surface types. *Geoscience and Remote Sensing Symposium*, 2009 IEEE International, IGARSS 2009.

LeCun, Y., Bengio, Y., Hinton, G. 2015. Deep Learning. *Nature*, 521, 7553. pp. 436-444.

Lguensat, R., Tandeo, P., Ailliot, P., Pulido, M., Fablet, R. 2017. The Analog Data Assimilation. *American Meteorological Society.* pp. 4093-4107.

Liu, Y., Racah, E., Prabhat, Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M., Collins, W. 2016. Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets. *Int'l Conf. on Advances in Big Data Analytics.* pp. 81-88.

Maas, A., L., Hannun, A., Y., Ng, A., Y. 2013. Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proceedings of the 30th International Conference on Machine Learning,* Atlanta, Georgia, USA. JMLR:W&PC, 28.

Martin, M., J., Hines, A., Bell, M., J. 2007. Data assimilation in the FOAM operational short-range ocean forecasting system: A description of the scheme and its impact. *Quarterly Journal of the Royal Meteorological Society,* 133. pp. 981–995.

Martin, S. 2014. An Introduction to Ocean Remote Sensing, 2nd ed. *Cambridge University Press,* UK. ISBN 978-1-107-01938-6.

McCandliss, R., R., Jones, S., E., Hearn, M., Latter, R., Jago, C., F., 2002. Dynamics of suspended particles in coastal waters (southern North Sea) during a spring bloom. *Journal of Sea Research* 47. pp. 285–302.

McDowell, S., E., Rossby, H., T. 1978. Mediterranean water: An intense mesoscale eddy off the Bahamas. *Science*, 202. pp. 1085–1087.

Merchant, C., J., Harris, A., R., Murray, M., J., Závody, A., M. 1999. Toward the Elimination of Bias in Satellite Retrieval of Sea Surface Temperature 1. Theory, Modelling and Interalgorithm Comparison. *Journal of Geophysical Research*, 104. pp. 23565-23578.

Milliman, J., D., Meade, R., H., 1983. World-wide delivery of river sediment to the oceans. *Journal of Geology,* 91, 1. pp. 1–21.

Minnett, P., Kaiser-Weis, A. 2012. Near-surface oceanic temperature gradients. *Group for High Resolution Sea Surface Temperature (GHRSST)*. Discussion document.

Mobley, C. 1994. Light and Water: Radiative Transfer in Natural Waters. Academic Press.

Nair, V., Hinton, G. 2010. Rectified Linear Units Improve restricted Boltzmann Machines. *Proceedings of 27th International Conference on Machine Learning,* Haifa, Israel.

NASA. 2008. Handbook for Limiting Orbital Debris

NASA. 2016. Orbital Debris. *Quarterly News,* 20, 3.

Niedzielski, T., Kosek, W. 2009. Forecasting sea level anomalies from TOPEX/Poseidon and Jason-1 satellite altimetry. *J Geod*, 83. pp. 469-476.

Oke, P.R., G.B. Brassington, D.A. Griffin, and A. Schiller. 2008. The Bluelink Ocean Data Assimilation System (BODAS). *Ocean Modelling* 21:46–70.

Ott, L., R., Longnecker, M. 2010. *An introduction to statistical Methods and Data Analysis.* 6th ed. Brooks/Cole, USA.

Patten, B., C. 1975. Systems Analysis and Simulation in Ecology, Academic Press. ISBN 0-12-547203-X.

Pietrzak, J., D., de Boer, G., J., Eleveld, A., M. 2011. Mechanisms controlling the intra-annual mesoscale variability of SST and SPM in the southern North Sea. *Continental Shelf Research* 31, 6. pp. 594-610.

Premier, V. 2017. Analysis of Satellite Altimetry vs GLOSSIS. *Internship Report.*

Pujol, M., I., Faugere, Y., Taburet, G., Dupuy, S., Pelloquin, C., Ablain, M., Picot, N. 2016. DUACS DT2014: the new multi-mission altimeter data set reprocessed over 20 years. *Ocean Sci.*, 12. pp. 1067-1090.

Rixen, M., Beckers, J., M., Alvarez, A., Tintore, J. 2001. Results on SSH neural network forecasting in the Mediterranean Sea. *Remote Sensing of the Ocean and Sea Ice, Proceedings of SPIE* 4544. pp. 24-33.

Robinson, I., S. 2010. Discovering the Ocean from Space. The unique applications of satellite oceanography. Springer. Verlag Berlin Heidelberg. ISBN 978-3-540-24430-1. e-ISBN 978-3-540-68322-3. DOI 10.1007/978-3-540-68322-3.

Shao, Q., Zhao, J. 2015. Comparing the steric height in the nordic seas with satellite altimeter sea surface height. *Acta Oceanologica Sinica*, 34, 7. pp. 32-37.

Shepard, D. 1968. A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 ACM National Conference.* pp. 517-524.

Shutler, J., D., Quartly, G., D., Donlon, C., J., Sathyendranath, S., Platt, T., Chapron, B., Johannessen, J., A., Girard-Ardhuin, F., Nightingale, P., D., Woolf, D., K., Hoyer, J., L. 2016. Progress in satellite remote sensing for studying physical processes at the ocean surface and its borders with the atmosphere and sea ice. *Progress in Physical Geography* 40, 2. pp. 215-246.

Siegel, D., A., Maritorena, S., Nelson, N., B., Behrenfeld, M., J., McClain, C., R. 2005. Coloured dissolved organic matter and its influence on the satellite-based characterization of the ocean biosphere. *Geophysical Research Letters*, 32, L20605.

Srivastana, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15. pp. 1929-1958.

Stammer, D., Ray R., D., Andersen, O., B., Arbic, B., K., Bosch, W., Carrere, L., Cheng, Y., Chinn, D., S., Dushaw, B., D, Egbert, G., D., Erofeeva, S., Y., Fok, H., S., Green, J., A., M., Griffiths, S., King, M., A., Lapin, V., Lemoine, F., G., Luthcke, S., B., Lyard, F., Morison, J., Muller, M., Padman, L., Richman, J., G., FShriver, J., F., Shum, C., K., Taguchi, E., Yi, Y. 2014. Accuracy assessment of global barotropic ocean tide models. *Reviews in Geophysics*, 52. pp. 243–282.

Sterlini, P., Le Bars, D., de Vries, H., Ridder, N. 2017. Understanding the spatial variation of sea level rise in the North Sea using satellite altimetry. *J. Geophys. Res: Oceans,* 122. pp. 6498-6511

Tandeo, P., Ailliot, P., Chapron, B., Lguensat, R., Fablet, R. The Analog data assimilation: Application to 20 years of altimetric data. *5th International Workshop on Climate Informatics,* September 24-25

Tompson, J., Goroshin, R., Jain, A., LeCun, Y., Bregler, C. 2015. Efficient Object Localization Using Convolutional Networks. *CVPR2015*, *Computer Vision Foundation*. pp. 648-656.

Tukey, J., W. 1977. Exploratory Data Analysis. *Addison-Wesely.*

Valente, A., Sathyendranath, S., Brotas, V., Groom, S., Grant, M., Taberner, M., Antoine, D., Arnone, R., Balch, W., M., Kathryn, B., Barlow, R., Belanger, S., Berthon, J., F., Besiktepe, S., Brando, V., Canuti, E., Chavez, F., Claustre, H., Crout, R.,, Frouin, R., Garcia-Soto, C., Gibb, S., W., Gould, R., Hooker, S., Kahru, M., Klein, H., Kratzer, S., Loisel, G., McKee, D., Mitchell, B., G., Moisan, Tiffany, Muller-Karger, F., O'Dowd, L., Ondrusek, M., Poulton, A., J., Repecaud, M., Smyth, T., Sosik, H., M., Twardowski, M., Voss, K., Werdell, J., Wernand, M., Zibordi, G. 2016 *Earth Syst. Sci. Data,* 8. pp. 235-252.

Weaver, A., T., Deltel, C., Machu, E., Ricci, S., Daget, N. 2005. A multivariate balance operator for variational ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society,* 131. pp. 3605–3625.

Wingham, D., J., Francis, C., R., Baker, S., Bouzinac, C., Brockley, D., Cullen, R., de Chateau-Thierry, P., Laxon, S., W., Mallow, U., Mavrocordatos, C., Phalippou, L., Ratier, G., Rey, L., Rostan, F., Viau, P., Wallis, D., W. 2006: CryoSat: A mission to determine the fluctuations in Earth's land and marine ice fields. *Adv. Space Res.*, 37. pp. 841–871.

Woppelman, G., Marcos, M. 2016. Vertical land motion as a key to understanding sea level change and variability. *Rev. Geophys.,* 54. pp. 64-92.

Yamashita, R., Nishio, M., Gian Do, R., K., Togashi, K. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9. pp. 611-629.

Yelekci, O. 2017. Submesoscale dynamics in the Bay of Biscay continental shelf. *Oceanography*, University Pierre et Marie Curie, Paris VI.

Závody, A., M., Mutlow, C., T., Llewellyn-Jones, D., T., 1995. A Radiative Transfer Model for Sea Surface Temperature Retrieval from the Along-Track Scanning Radiometer. *Journal of Geophysical research*, 100. pp. 937-952.

9. Appendix

9.1. Low-pass filter

In addition to the moving average filter, two more low-pass filtering techniques were used; namely, the median and the Butterworth. They were both abandoned because they produced not realistic results. On the one hand, the median filter in spite of being robust to outliers, hence it would not create peaks and valleys at positions were the did not originally exit, it produced flats and the smoothing was not significant, which are not realistic (Fig. 54).

On the other hand, the butterworth filter was not a heuristic one since it needed to be designed first. It was not choses because it produced an extremely smooth outcome, as well as it probably produced some periodic effect that might be artificial (Fig. 55). For these reason, it was also abandoned.

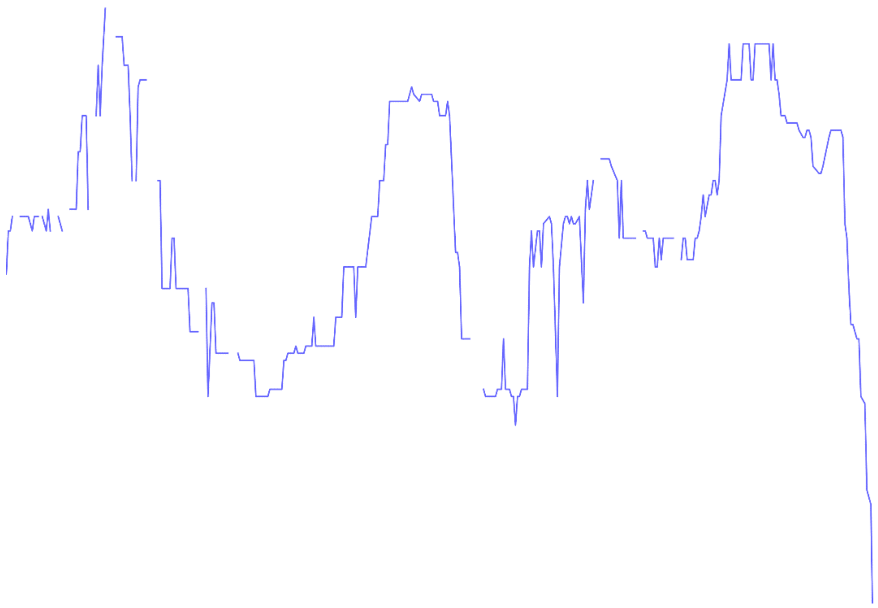


Figure 54: Close-up of along-track SSHA which shows the flat effect produced by median filter

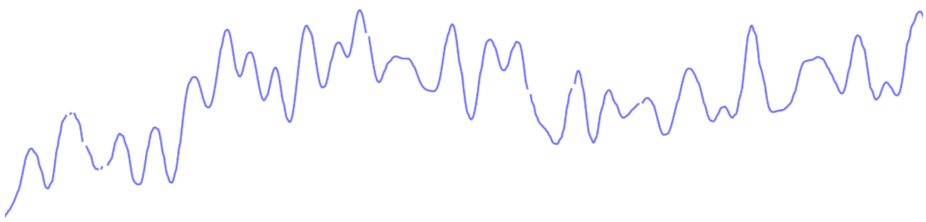


Figure 55: Close-up of along-track SSHA which shows the Butterworth filter with periodic-like effect and high smoothness

9.2. Temporally Grouped SST data

Another approach that was conducted was the temporal grouping of the tracks. Every 10 days the tracks were grouped together and the training approach as described in the methodology chapter (complete track archive approach) was applied, although with different hyperparameters. In reality, the number of tracks that were included in ever 10-day group was varying. The reason for this is that there were several days that they were full of clouds, therefore they could not be used. Also, Sentinel-3 B data delivery (after the Calibration/Validation phase) begun on March 2019, which means that after that the dataset was more dense, hence each group had more tracks. This means that the RF training was based on a heterogeneous distribution of data in the groups, thus the RF result of each one of them could not be trusted therefore the approach was soon abandoned after seeing the not promising first results.

9.3. Learning rate decay on CNN

Instead of a constant learning rate, a learning rate with a decay was tried at the SSHA/SST L4 datasets, without luck. The formula used was:

$$lr = \frac{lr_i}{1 + lr_{decay} \cdot iteration}$$

$$where\ lr, the\ updated\ learning\ rate$$

$$lr_i, the\ initialized\ learning\ rate$$

$$lr_{decay} = \frac{lr_i}{N}, N\ is\ number\ of\ epochs$$

$$iteration, the\ current\ update\ number\ of\ each\ iteration$$

9.4. Scatter plots of SSHA and SST

In addition to the direct scatter plots and correlations between SSHA and SST, time-lagged scatter plots were also examined as far as the North Sea is concerned. The lag interval used was 1 day while the maximum lag was 90 days. Some indicative plots can be seen below. The outstanding majority of them showed pure randomness (Fig. 56, Fig, 57, Fig. 61) while very few showed clear positive correlations (Fig, 59,Fig. 60) or negative correlations (Fig. 58).

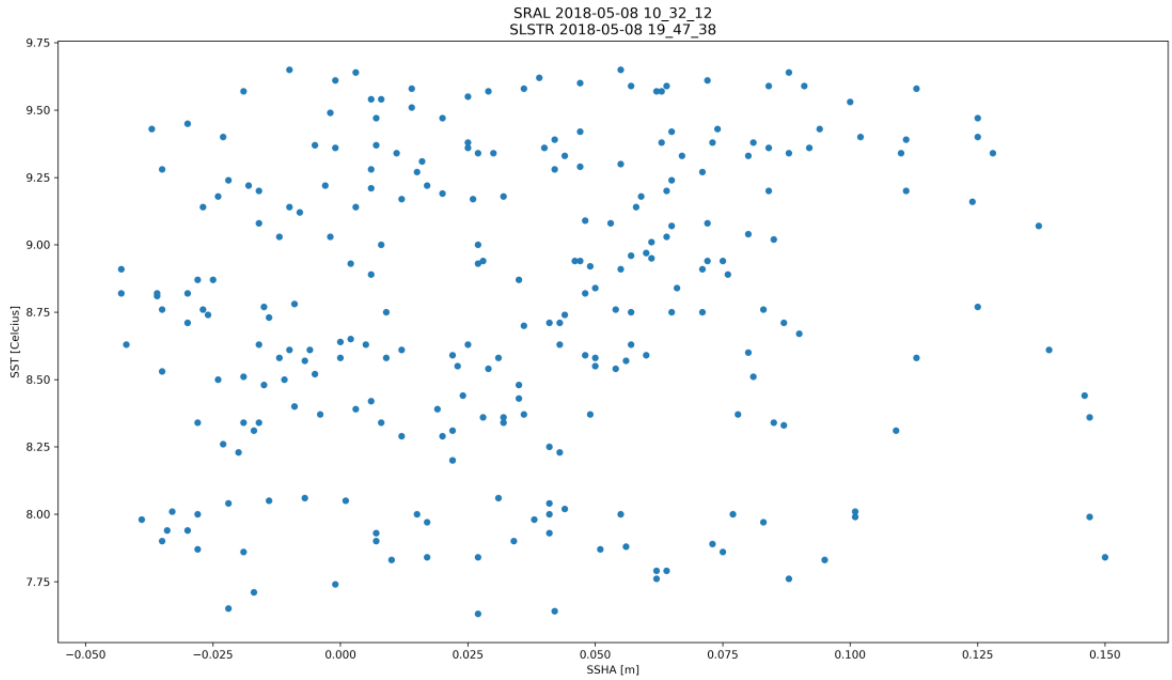


Figure 56: No lag between SSHA and SST without filter. Complete randomness is observed.

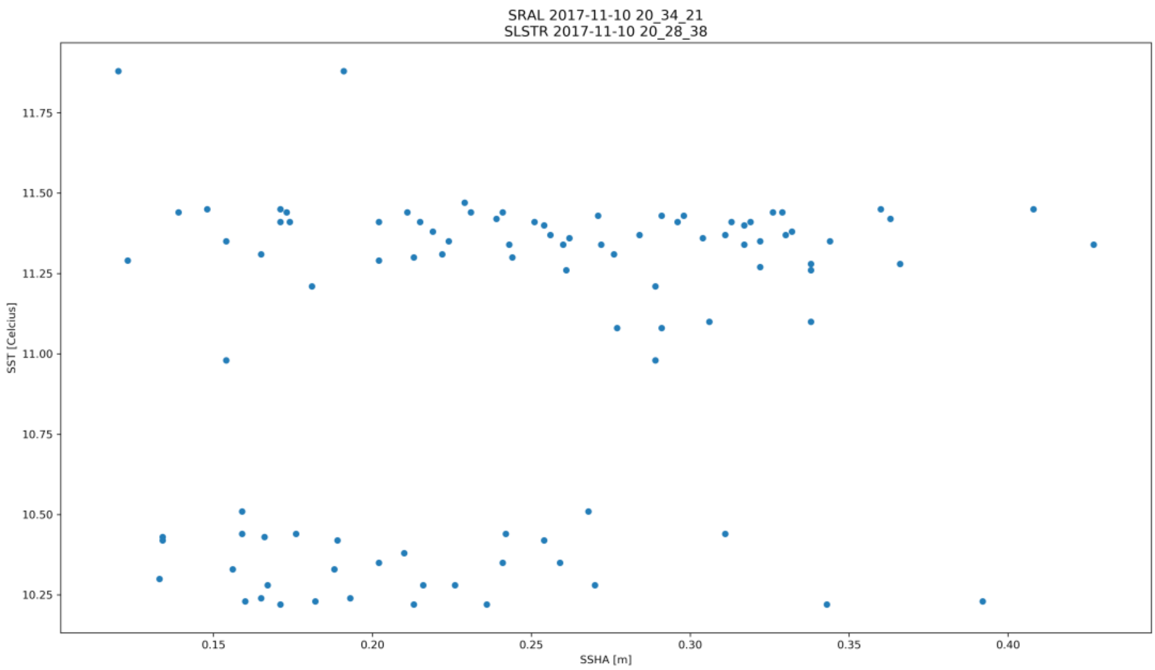


Figure 57: No lag between SSHA and SST without filter. No correlation is observed.

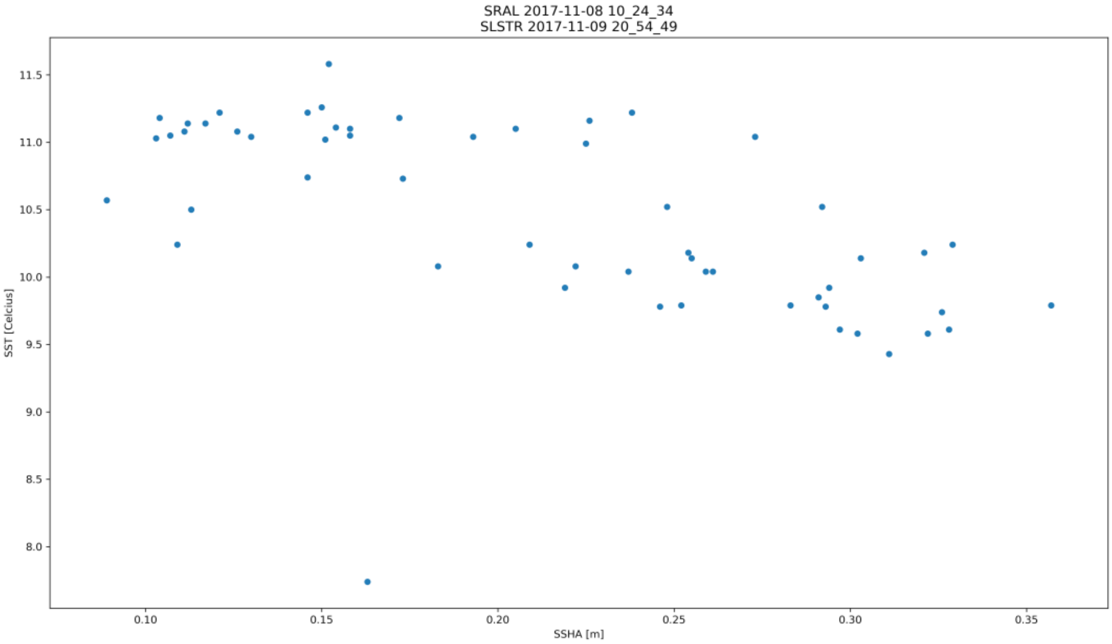


Figure 58: 20-day lag between SSHA and SST. Very slight negative correlation can be distinguished.

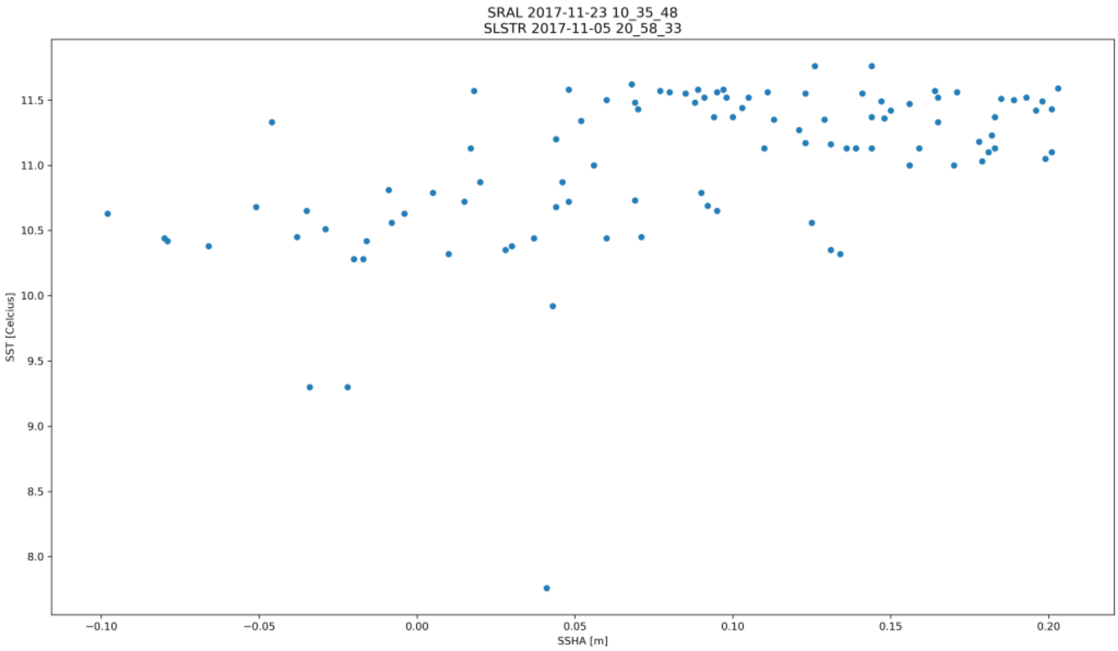


Figure 59: 18-day lag between SSHA and SST. Very slight positive correlation can be observed.

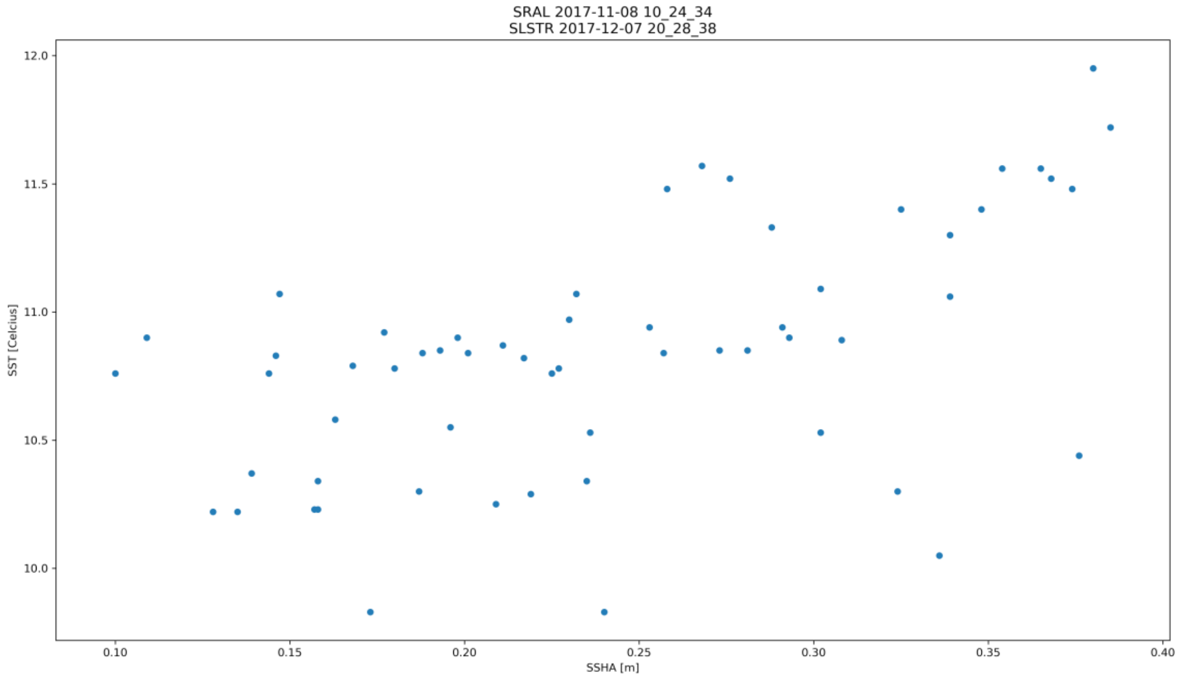


Figure 60: 30-day lag between SSHA and SST without filter. A very slight positive correlation can be distinguished

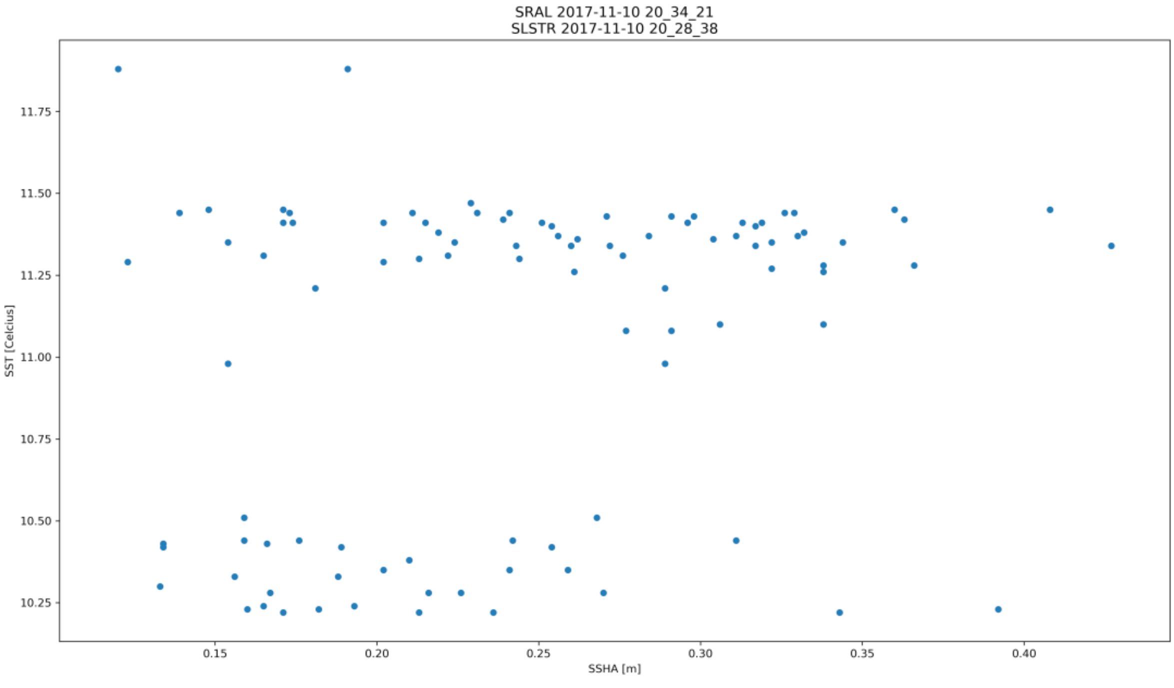


Figure 61: No lag between SSHA and SST without filter. No correlation occurs.

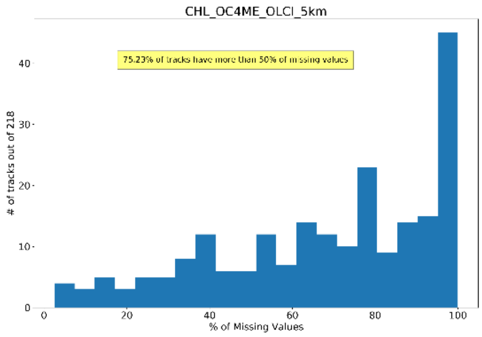
Table 5: Spectral bands that the SLSTR Sentinel-3 instrument uses (from EUMETSAT, 2018b).

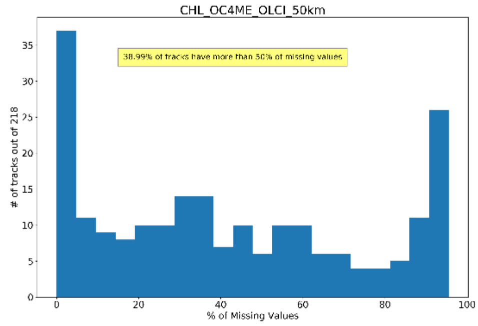
Band	λ centre (μm)	Width (μm)	Function	Comments		Res. (m)
S1	0.555	0.02	Cloud screening, vegetation monitoring, aerosol	**Visible Near IR**	**Solar reflectance bands**	**500**
S2	0.659	0.02	NDVI, vegetation monitoring, aerosol			
S3	0.865	0.02	NDVI, cloud flagging, Pixel co-registration			
S4	1.375	0.015	Cirrus detection over land	**Short-Wave IR**		
S5	1.61	0.06	Cloud clearing, ice, snow, vegetation monitoring			
S6	2.25	0.05	Vegetation state and cloud clearing			
S7	3.74	0.38	SST, LST, Active fire	**Thermal infra-red Ambient bands (200 K - 320 K)**		**1000**
S8	10.85	0.9	SST, LST, Active fire			
S9	12	1	SST, LST			
F1	3.74	0.38	Active fire	**Thermal infra-red fire emission bands**		
F2	10.85	0.9	Active fire			

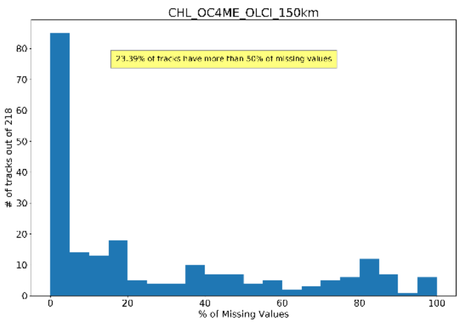
Band	λ centre (nm)	Width (nm)	Function
Oa1	400	15	Aerosol correction, improved water constituent retrieval
Oa2	412.5	10	Yellow substance and detrital pigments (turbidity)
Oa3	442.5	10	Chl absorption max., biogeochemistry, vegetation
Oa4	490	10	High Chl, other pigments
Oa5	510	10	Chl, sediment, turbidity, red tide
Oa6	560	10	Chlorophyll reference (Chl minimum)
Oa7	620	10	Sediment loading
Oa8	665	10	Chl (2nd Chl abs. max.), sediment, yellow substance/vegetation
Oa9	673.75	7.5	For improved fluorescence retrieval and to better account for smile together with the bands 665 and 680 nm
Oa10	681.25	7.5	Chl fluorescence peak, red edge
Oa11	708.75	10	Chl fluorescence baseline, red edge transition
Oa12	753.75	7.5	O2 absorption/clouds, vegetation
Oa13	761.25	2.5	O2 absorption band/aerosol corr.
Oa14	764.375	3.75	Atmospheric correction
Oa15	767.5	2.5	O2A used for cloud top pressure, fluorescence over land
Oa16	778.75	15	Atmos. corr./aerosol corr.
Oa17	865	20	Atmos. corr./aerosol corr., clouds, pixel co-registration
Oa18	885	10	Water vapour absorption reference band. Common reference band with SLSTR instrument. Vegetation monitoring
Oa19	900	10	Water vapour absorption/vegetation monitoring (max. reflectance)
Oa20	940	20	Water vapour absorption, atmos./aerosol corr.
Oa21	1 020	40	Atmos./aerosol corr.

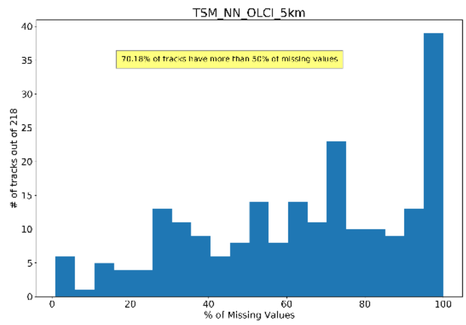
Table 6: Spectral bands the OLCI instrument uses (from EUMETSAT, 2018).

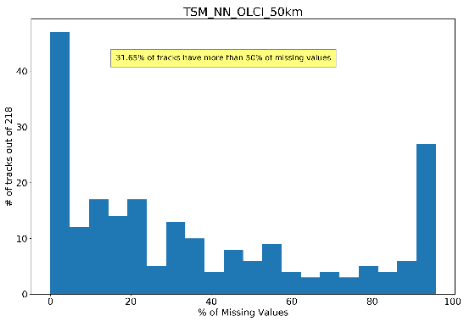
9.5. Missing Values OC variables

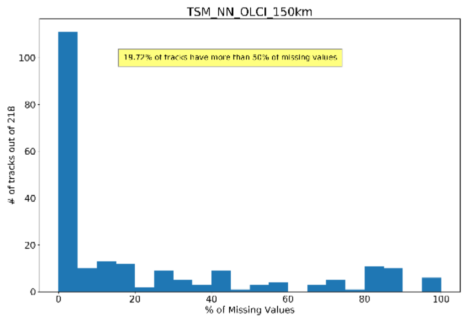

CHL_OC4ME_OLCI_5km
75.23% of tracks have more than 50% of missing values
# of tracks out of 218
% of Missing Values

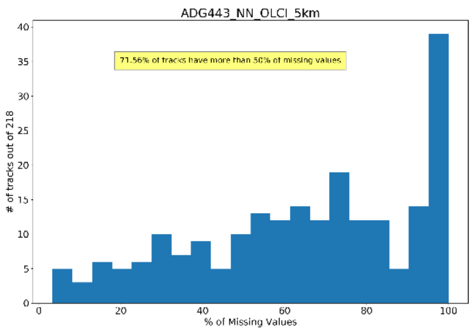

CHL_OC4ME_OLCI_50km
38.99% of tracks have more than 50% of missing values
# of tracks out of 218
% of Missing Values

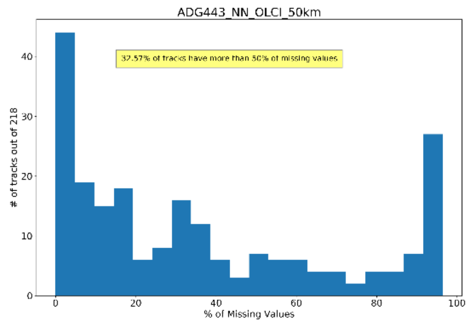

CHL_OC4ME_OLCI_150km
23.39% of tracks have more than 50% of missing values
# of tracks out of 218
% of Missing Values


TSM_NN_OLCI_5km
70.18% of tracks have more than 50% of missing values
# of tracks out of 218
% of Missing Values


TSM_NN_OLCI_50km
31.65% of tracks have more than 50% of missing values
# of tracks out of 218
% of Missing Values


TSM_NN_OLCI_150km
19.72% of tracks have more than 50% of missing values
# of tracks out of 218
% of Missing Values


ADG443_NN_OLCI_5km
71.56% of tracks have more than 50% of missing values
# of tracks out of 218
% of Missing Values


ADG443_NN_OLCI_50km
32.57% of tracks have more than 50% of missing values
# of tracks out of 218
% of Missing Values

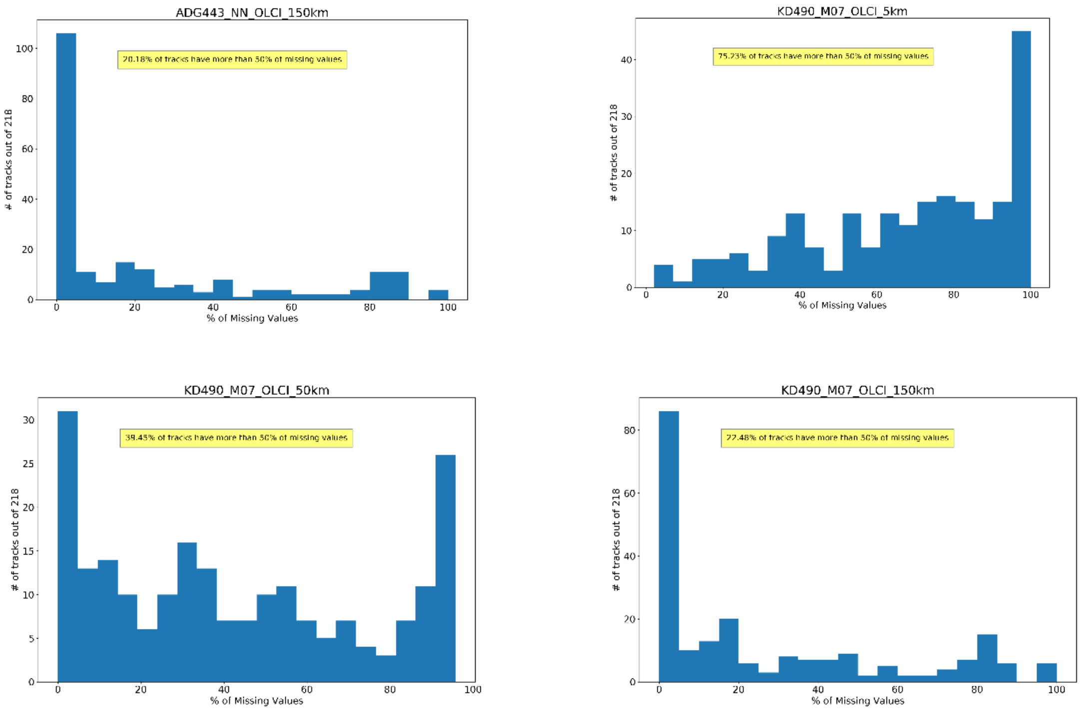


Figure 62: The above graphs show the percentage of missing values for different OC variables of various spatial scales. In larger spatial scales less missing values occurs since they are inter/extra-polated, while in smaller scales the values are very high (over 50%)