UNIVERSITY OF CAPE TOWN

A preliminary investigation on the relationships between upwelling and commercial hake fishery in the Southern Benguela.

Author:
MSc. Candidate.
Nkululeko MEMELA

Supervisor:
Associate Professor.
Marcello Vichi

Co-Supervisor:
S.H. Fleet Executive.
Russell Hall

A mini-dissertation submitted in fulfilment of the partial requirements of the Master of Science Degree In Applied Ocean Science.
The department of:

Ocean and Atmosphere Sciences
&
Marine Research Institute

April 16, 2021

Declaration of Authorship

I, Nkululeko Charles Memela, MSc. candidate, student number [MMLNKU001] declare that this dissertation titled, **A preliminary investigation on the relationships between upwelling and commercial hake fishery in the Southern Benguela.** and the work presented in it is done by myself in candidature for the master of science qualification. I also confirm that:

- This work was done wholly or mainly while in the UNIVERSITY OF CAPE TOWN in partnership with Sea Harvest Corporation.
- Where any part of this project have previously been published by any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this project is entirely my own work.
- I have acknowledged all main sources of help.
- I commit to NOT distributing the findings and the contents of this report to any individual outside the company except parts of the research that will be deemed necessary by the UNIVERSITY OF CAPE TOWN and Sea Harvest Group limited as publishable non-competitive company information as some critical data used in it is property of Sea Harvest Group limited.

Signed: Signed by candidate

Date of first submission: 28 October 2020

UNIVERSITY OF CAPE TOWN

Abstract

OCEAN SCIENCE

Ocean and Atmosphere Sciences

&

Marine Research Institute

Master of Science Degree In Applied Ocean Science.

A preliminary investigation on the relationships between upwelling and commercial hake fishery in the Southern Benguela.

by MSc. Candidate.

Nkululeko MEMELA

Abstract

The hake bottom trawl division is the largest component of the fishing industry in South Africa and it is one of the strong pillars of the food industry and the national economy. It is the main source of livelihood for many people in the West Coast and as such, finding ways of advancing it towards the direction of fourth industrial revolution is at the top of societal interests and a top priority for the major companies that are key players in the industry. Sea Harvest Group Limited is one of these key players and as such, it has undertaken to be a part of the study to improve predictability of fishing by collecting data which will contribute towards the scientific study of the patterns which determine the viability of some fishing locations over the others at different times under various conditions. The studied region is the West Coast grounds located in the southern Benguela at grid 32°S: 34°S, 16°E :19°E. The study is based on the hypothesis that the main driver of the availability of hake is the upwelling, separated into its coastal Ekman transport and curl-driven components. These two components of upwelling are driven by winds and they are known to stimulate primary production and support a larger marine food web. The correlations between these upwelling types and the mean monthly catch per unit effort (CPUE) of this region is assessed. The relationship between chlorophyll abundance and hake CPUE hypothesises lagged association of hake abundance to the underlying biological food chain driven by the upwelling events. A multiple regression model is then produced as a basic step towards quantification. The results suggest that Hake CPUE is lag correlated with upwelling and that some degree of predictability can be derived from the observation of combined upwelling patterns.

Acknowledgements

- I thank my supervisor, the Marine Research institute (*Ma-Re*) director Prof. Marcello Vichi for awakening in me a sleeping giant that can confront complex scientific problems with courage. I thank you a lot for seeing hidden potential in me since undergrad Prof. Vichi, I will be walking tall into the future because you allowed me to stand on your shoulders.
- I thank my co-supervisor, the Sea Harvest fleet Executive Mr Russel Hall for recommending that Sea Harvest must fund this study and for believing in my abilities to carry it through. I also thank you for giving my first opportunity to showcase my talent in the field of work and for making me a part of Sea Harvest scientific problem solving team.
- I thank the University of Cape Town for opening its doors of learning for me to push my academic career this far.
- I thank Sea Harvest Group limited for trusting in my abilities, for sharing their private information for the performance of this project and for significantly funding this project.
- I thank the Sea Harvest strategic services group executive Mr Khumalo for laying the conceptual idea of this project and for constant support with literature material and further ideas.
- I thank the department of Ocean and Atmosphere Sciences for offering me yet another opportunity to be one of their departmental scholars when the number of deserving applicants is always very high and competitive.
- I thank the National Research Foundation (NRF) for putting additional funds on the project and for ensuring that my main preoccupation is with completing the degree other than worrying about any shortage of basic necessities.
- I thank *Ma-Re* for assisting me with urgent funds and for ensuring that my focus is not sidetracked by non-academic preoccupations. I particularly thank Sharon Bosma for being behind it all.
- I thank my lovely woman Sanelisiwe Buthelezi for daily emotional support even though a thousand miles away.
- I thank my big friend Ramontseng Rapolaki for constant moral support and great data handling ideas.
- I thank my big friend Shingirai Gwini for being my number one fan and for always seeing the best in my work ethic and effort. This kept me very energised.
- I thank my huge friends Sizwe and Palesa for making sure that there is warm supper for me after a long 16 hour day behind the desk and for making sure there's something cold to drink after.

- I thank my parents uMama no Baba uMemela. They always supported my hard work and inspired my success in a number of ways. May God keep them long for me.

Contents

List of Figures

List of Tables

List of Abbreviations

AMJ	**A**pril **M**ay **J**une
ANOVA	**AN**alysis **O**f **VA**rience
CPUE	**C**atch **P**er **U**nit **E**ffort
DDL	**D**igital **D**ata **L**oggers
EBC	**E**astern **B**oundary **C**urrent
ENSO	**E**l **N**iño **S**outhern **O**scillation
ERA5	**E**arth **R**e**A**nalysis data **5**
GC	**G**rade **C**atch
HPS	**H**igh **P**ressure **S**ystem
ITCZ	**I**nter **T**ropical **C**onvergance **Z**one
JAS	**J**uly **A**ugust **S**eptember
JFM	**J**anuary **F**ebruary **M**arch
LPS	**L**ow **P**ressure **S**ystem
MLRA	**M**arine **L**iving **R**esources **A**ct
SST	**S**ea **S**urface **T**emperature
OND	**O**ctober **N**ovember **D**ecember
SAHPS	**S**outh **A**tlantic **H**igh **P**ressure **S**ystem
SASH	**S**outh **A**tlantic **S**ubtropical **H**igh
SAWS	**S**outh **A**frican **W**eather **S**ervice
SC	**S**pecies **C**atch
SLP	**S**ea **L**evel **P**ressure
Sv	**S**verdrop
SWC	**S**outh **W**estern **C**ape
VPA	**V**irtual **P**opulation **A**nalysis

Physical Constants

Hake conversion factor	$c = 1.46$ (Headed and gutted)
Sea-level air density	$\rho_1 = 1.225\,(\mathrm{kg\,m^{-3}})$ (at sea level)
Sea water density	$\rho_0 = 1025\,(\mathrm{kg\,m^{-3}})$ (on the boundary layer)
Earth's angular velocity	$\Omega = 7.2921 \times 10^{-5}\,(\mathrm{rev/s})$
Atmospheric pressure	$\mathrm{P} = 101\,325\,(\mathrm{Pa})$ (at sea level)
Rossby radius of deformation	$R_d = 30\,000\,(\mathrm{m})$
Sverdrup	$Sv = 1\,000\,000\,(\mathrm{m^3\,s^{-1}})$
Area (selected region of study)	$A = 8.5834 \times 10^5\,(\mathrm{m^2})$

List of Symbols

A	Area	(m^2)
R_d	Rossby radius of deformation	(m)
W	Wind speed	$(m\,s^{-1})$
u	Zonal wind speed	$(m\,s^{-1})$
v	Meridional wind speed	$(m\,s^{-1})$
t	Time	(s)
f	Coriolis parametre	(s^{-1})
ρ_0	Sea water density	$(kg\,m^{-3})$
ρ_1	Sea-level Air density	$(kg\,m^{-3})$
∂	Rate of change	
z	Vertical distance	(m)
V_f	Volume flow	$(m^3\,s^{-1})$
Chl_a	Chlorophyll concentration	(mg/m^3)
$\hat{k}$	Vertical unit vector	
T_{E_x}	Ekman transport	$(m^2\,s^{-1})$
w_{curl}	Vertical curl driven velocity	$(m\,s^{-1})$
Ω	Earth's angular velocity	(rev/s)
τ	Wind stress	$(N\,m^{-2})$
τ_x	Zonal wind stress	$(N\,m^{-2})$
τ_y	Meridional wind stress	$N\,m^{-2}$
τ_a	Alongshore wind stress	$(N\,m^{-2})$
$Curl_z(\tau)$	Wind stress curl	$N\,m^{-3}$
ω	Angular frequency	(rad)
i, j	Row & column index for a grid cell	
Dx	Width of the grid cell	(m)
Dy	Height of the grid cell	(m)
$CPUE$	Catch per unit effort	(ton/h)
Sv	Sverdrup	$(m^3\,s^{-1})$

Dedicated to my Mom MaNdosi. Lala ngoxolo Magay'aydlele. Khumbuza, Dubandlela ka-Vico!

1 Introduction

The hake fishery in the South and West Coast of South Africa is a major player in the food industry and in the sovereign economy of the republic. As such, improving its efficiencies and scientific understanding is at the pinnacle of sustaining and advancing its economic importance. The hake fishery contributes substantially to the GDP of the country, accounting for about 40% of all landed fish in the country according to SADC-EU (2017). The hake fishery creates 65 jobs for every landed 1000 tones & the Western Cape province accounts for 95% of the deep sea and inshore catches as well as 70% of the industry's income and jobs (SADC-EU, 2017). Improving efficiency in the fishing operations would make the fishing process less costly and decrease the carbon footprint associated with high fishing effort. If the industry produces products at environmentally sustainable and cheaper cost, it can supply affordable products to ensure food security and it can keep more people employed.

A study of the ocean conditions and the behaviour of fish that determine their temporal and spatial distributions is undertaken using the data collected over a 10 year period (2009-2018) by the bottom trawler fleet during the fishing operations performed over this period at Sea Harvest corporation. Sea Harvest corporation is one of the major deep hake trawler companies in South Africa for which I am an employee who has been brought in to scientifically analyse the above mentioned database and determine any existence of correlations between catches and environmental conditions. The company has been collecting this data with intentions of utilising it for studying the existence of fishing patterns in order to make informed fishing decisions. This would help the fishing management to allocate necessary fishing effort (number of vessels) if the catch can be reliably predicted and the anticipated catches can be confidently outlined prior to the fishing trip.

Winds are the main driver of surface ocean currents at sea and water masses moves from one position to another with their properties in accordance to the forcing of the prevailing wind. Winds that push away surface waters from the coast results in upwelling on the coast which leads to coastal biological productivity. Biological productivity is generally indicated by chlorophyll presence (detected as the ocean colour from space) after an upwelling event. The correlation between the upwelling, chlorophyll and catches is then assessed based on the following quotation, "There are many examples in literature that demonstrate correlations between physical oceanographic factors and changes in fish stocks, these correlations hold for a few years and

then break off again within a decade" (Mann, 1993).

This study assesses if there are no observable linkages in the decade of our data collection between the environment and the catches. Mann (1993) further claimed that the breaking of these correlations does not necessarily mean that the correlation was invalid but instead it informs that there are deeper hidden mechanisms and factors that need to be explored and understood. This particular study assesses the correlations in the marine ecosystem wherein the intensity of southerly winds along shore induce a coastal upwelling by laterally displacing the water mass within the mixed layer $\approx$ 30 m deep according to Veitch, Penven, and Shillington (2010) offshore such that the nutrient rich bottom waters would be sucked up to the euphotic zone to replace the offshore displaced water in the upper ocean (Chavez and Messié, 2009). In the presence of sunlight, this upwelled nutrient rich water stimulates a rapid growth in the phytoplankton biomass, which after some time leads to the growth of zooplankton that primarily feed on the phytoplankton. Small pelagic fish such as anchovies and sardines feed on the zooplankton, therefore, they also after some time grow in numbers which mean that there will be an abundance of food for the higher trophic level species such as hake which feed primarily on anchovies and sardines. This means that a larger biomass or congregation of hake will be expected in this environment after some time lag from the booming of the small-pelagic fish population.

This intuitive process is the basis of our hypothesis which states that hake availability should increase after a strong upwelling event. We are looking to statistically analyse this lagged effect to see if it exists. The upwelling is estimated in two ways detailed in Chapter 3 as originally proposed by Rykaczewski and Checkley (2008); the coastal upwelling and wind stress curl induced upwelling.Rykaczewski and Checkley (2008) performed a study of this nature in the West Coast of California and obtained that correlation between the curl driven upwelling, coastal upwelling, the chlorophyll, the halocline and nutricline depth is strong and that there is increased surplus production during periods of strong curl driven and coastal upwelling. Our study is designed along these lines to observe the effects in our fishing region here in southern Africa. Most of the nutrient considerations will be based on literature but the chlorophyll is analysed, together with the winds effect. The study is conducted by using the chlorophyll as an indicator for primary production. There is no database available to us for the primary consumers (zooplankton) and for the secondary consumers (i.e. small pelagic fish), the data attainment and processing would have expanded the scope of the project substantially, therefore, due to time constraints they were avoided even though they are acknowledged to be a critical component that should be included in future studies. The only biomass estimate data available is the predators database i.e. hake that we are tracking. The loose connection will, therefore, be made between primary production and the predator's abundance which mean that there might be a lag between peak(s) of the predictor variable(s) and the response variable, hake.

Goals

- The first aim which is achievable in the context of this thesis is to provide useful statistical relationships that can be used for predicting the next hake fishing hot-spot so that when the fleet executive sends the vessel into the sea, there would be an element of predictability in the decision process to assist the skippers (Vessel captains) and to complement their expertise.
- The long-term goal is to enhance cost saving within the company and in the industry, as well as to promote effective fishing so that the industry gains competitive edge in the trading market.
- The second long-term goal is to improve the lifestyle and balance of well-being of the fishermen such that they are efficient in doing a months worth of fishing during a fishing trip in a shorter amount of time so that they can spend more time at home with their families while maintaining their good income.

Objectives

- The number one objective is to analyse some of the various environmental factors that drives hake availability in the Southern Benguela region.
- The second objective is to familiarize the growing business intelligence of the South African fishing industry with the upwelling patterns and chlorophyll patterns that exhibit the Southern Benguela region in order to lay the grounds for connecting dots in other physical oceanographic researches that depend on this basic understanding.

This thesis contains a proprietary data and they will be treated in a special way. The private data is obtained and handled in the method outlined in Chapter 3, Section 3.2 and 3.3. The rest of the analysis of this data will be performed in the protected chapter, Chapter 5. Sea Harvest catch data has been used for all fish stock estimations unless stated differently.

2 Literature Review

2.1 Climate of Western Cape and South Benguela Region

The extended austral winter (April-September) in the Western Cape shown in Figure 2.1 is characterised by wet conditions associated with cold fronts and north westerly winds according to Favre et al. (2013), Weldon and Reason (2014) and according to Engelbrecht and Landman (2016). These conditions reverse during the extended austral summer (September-April) as the wind reverse direction. The climate of this region is known as a Mediterranean climate because of its characteristics which match those of the countries surrounding the Mediterranean Sea. This match comes about because of the corresponding latitudinal positioning around the equator (subtropical latitudes). The characteristics of this region include hot dry summers, cold wet winters, austral south westerly wind gusts and south-easterly summer winds. The most pertinent of all these to our study is the south easterly winds because it is related to upwelling of colder and nutrient-rich waters. The upwelling is effected by south easterly winds shape of our coastline shown in Figure 2.1. This relationship of the coastline shape and southerly wind direction resulting in upwelling is clarified deeper in Chapter 3.

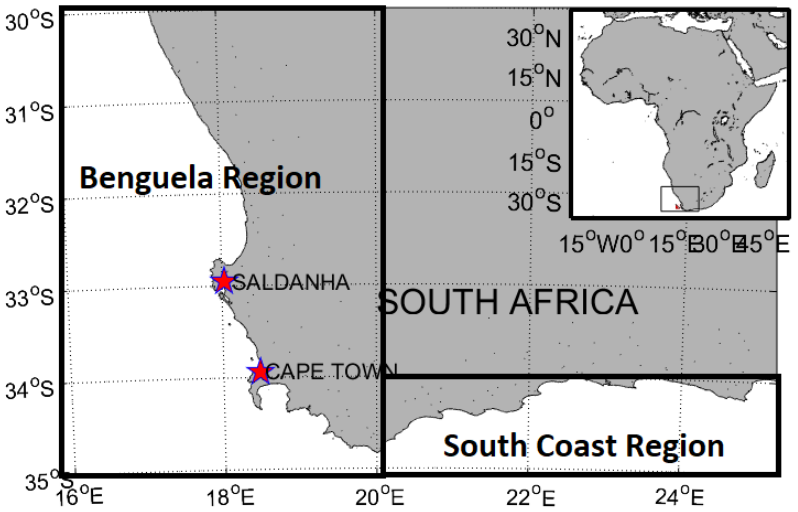

FIGURE 2.1: The Western Cape and Southern Benguela region and the south coast.

Hutchings et al. (2009) suggest that the pulsed low pressure systems suppress upwelling and that these low pressure systems arrive at an interval of 3-10 days in the coast of the Southern Benguela as shown in Figure 2.2 panel (a). Hutchings et al. (2009) further assert that these pulsed coastally trapped low pressure systems in the West Coast drive a pulsed upwelling effect demonstrated in pane (b) of Figure 2.2 where the index signal centred at zero is the meridional wind speed (labeled Average of NS wind speed) in the figure. It can be seen that during a meridional wind, there is a cumulative divergence per upwelling event once a low pressure system arrives as a pulse, this is broken which means that there will be a momentary suppression of upwelling during the coastaly trapped low pressure system passing.

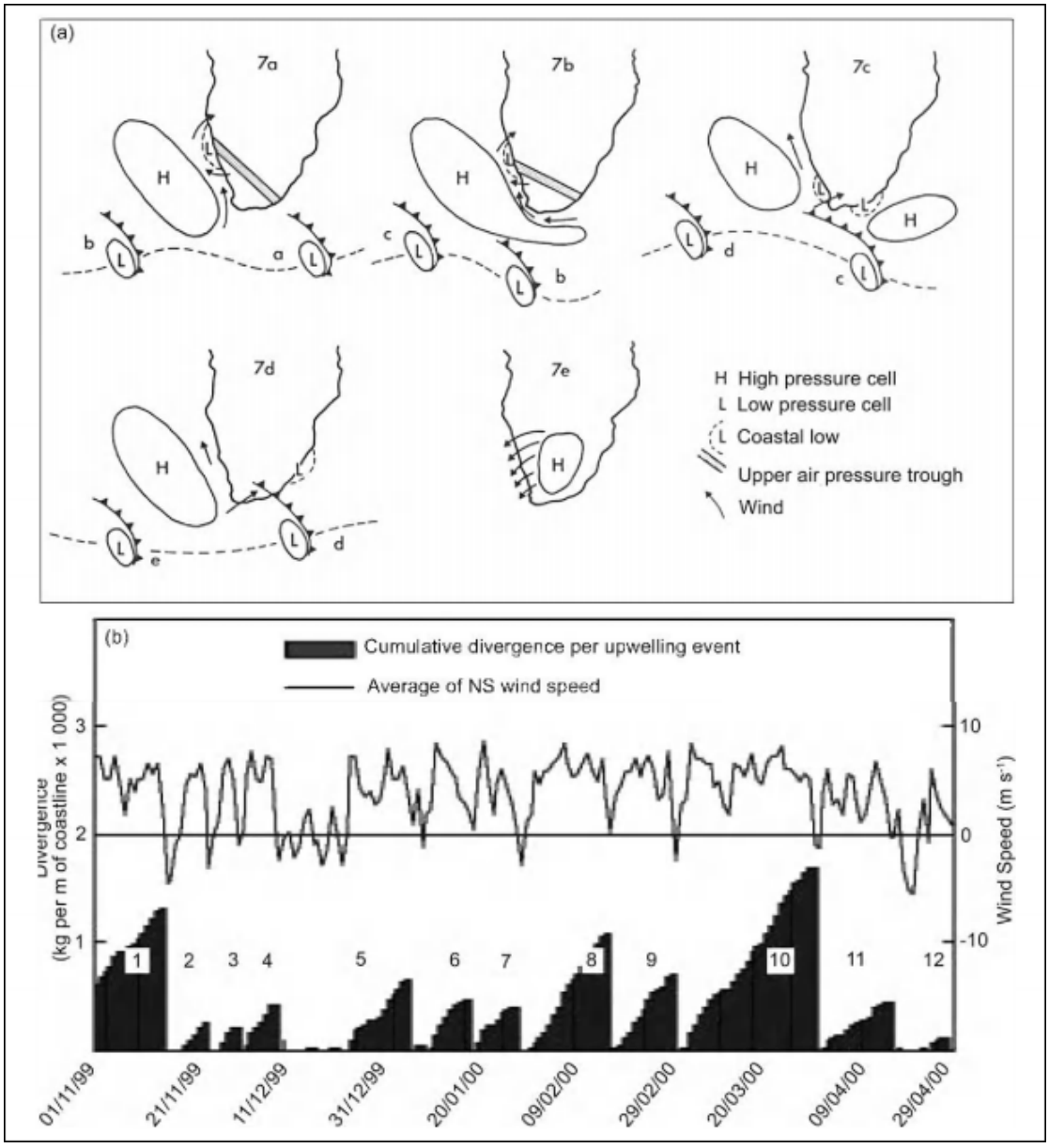

FIGURE 2.2: The impact of high and low pressure cells in the Southern Benguela region (Hutchings et al., 2009).

The seasonality of upwelling patterns is embedded in the seasonality of the winds themselves which are driven by the pressure systems which interact with the coast in the way that has been explained above. The cold fronts are a prevalent phenomenon in the austral winters in this area during this time as depicted in Figure 2.3 which is a snapshot of a low pressure system pattern in winter in this region. The South Atlantic high pressure system is located in its position in the south Atlantic during summer, hence the severely reduced occurrence of cold fronts hitting the coast during this season as shown in Figure 2.4. However, during winter it is pushed up meridionally and replaced by cold fronts that emanate from the meridionally shifted low pressure systems.

The curved bold black-line in Figure 2.3 represents a cold front extending from a cold pressure cell and approaching sub-continent during winter, a phenomenon that does not arise in austral summer. Coastally bound high pressure systems present winds that are oriented South Easeterly or Southerly which are important for inducing upwelling as they are positively meridional and parallel to the coast. The upwelling of sea waters contribute to calm conditions in the atmosphere since the upwelling waters are colder and they cool the surface atmosphere, making it less evaporative. These are conditions typically seen during summer months rather than in winter.

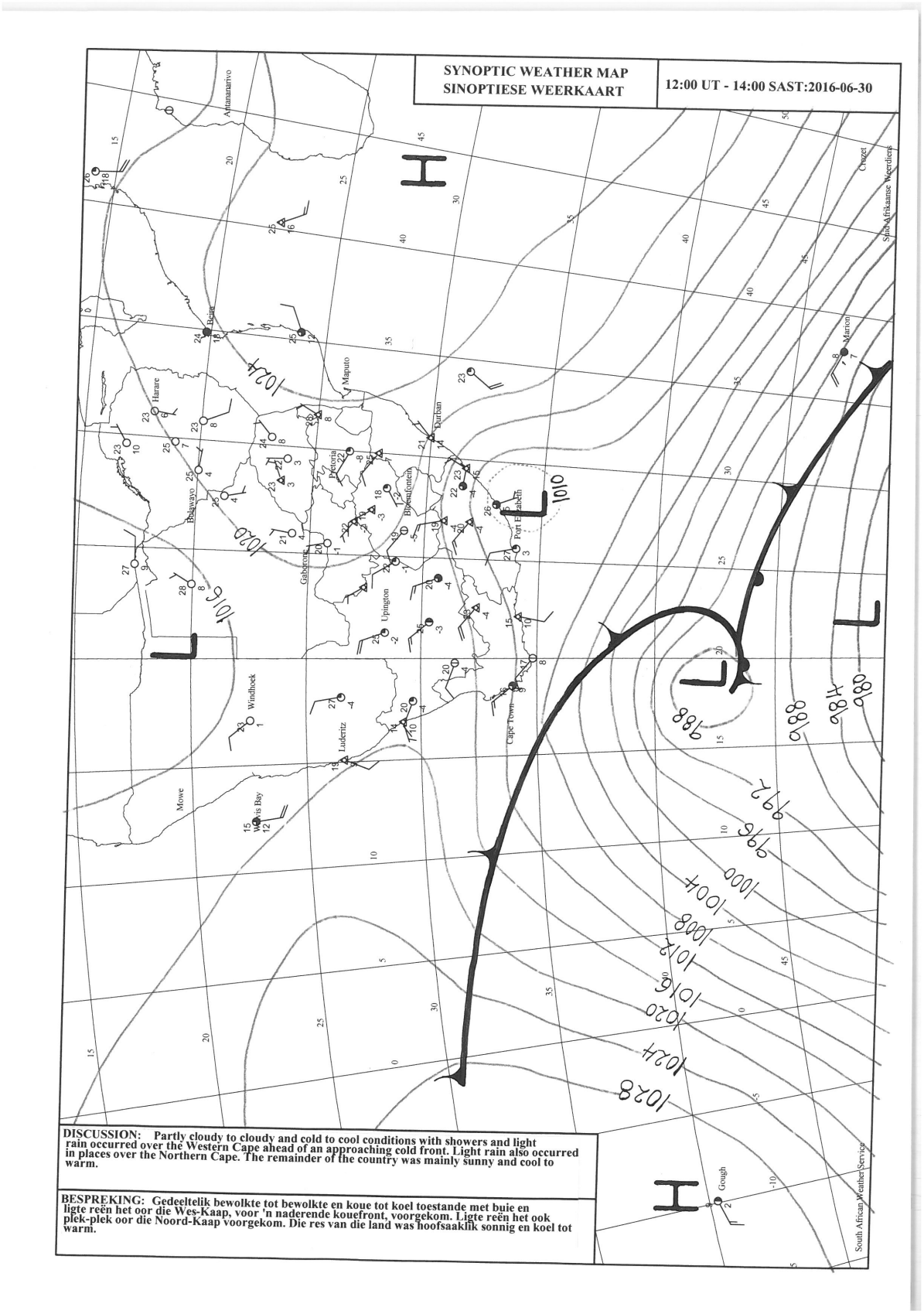

FIGURE 2.3: Synoptic weather map showing a snap shot of a synoptic pressure systems in a typical Winter day. Courtesy of SAWS (*South African Weather Service*)

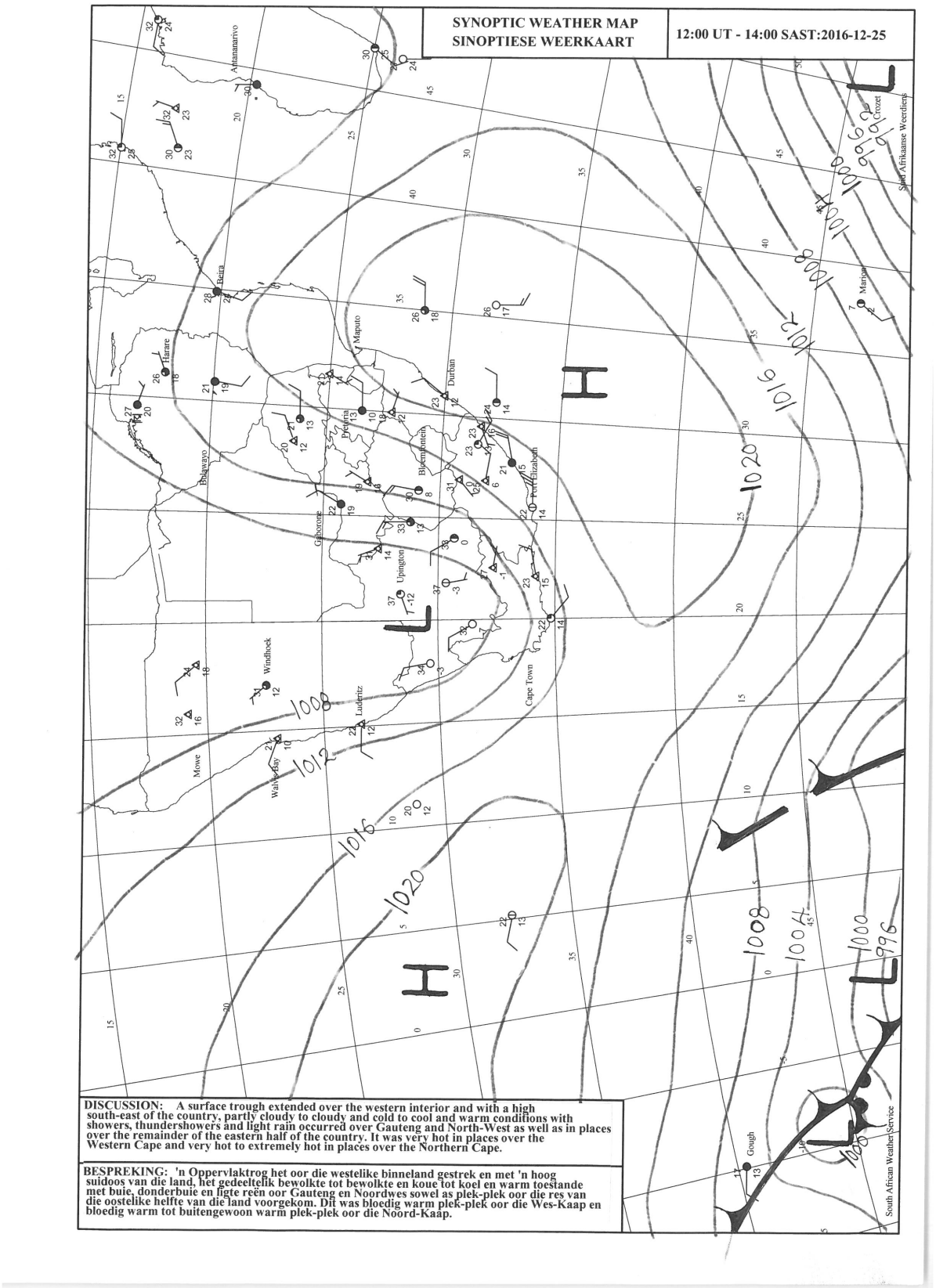

FIGURE 2.4: Synoptic weather map showing a snap shot of a showing synoptic pressure systems in a typical Summer day. Courtesy of SAWS (*South African Weather Service*)

2.2 Oceanography of the Southern Benguela Region

There are four major Eastern Boundary Current systems (EBCs) that are recognised for their unique ocean properties in the world, among which is the Benguela current system. The EBCs are the most productive marine ecosystems in the global ocean basin (Carr, 2001). Their productivity accounts for 5% of the global marine production and 17% of the global fish catches (Carr, 2001), while they are only less than 1% of the world's ocean by fraction (Messié and Chavez, 2015). The four EBCs namely are the California Current, Canary Current (West Africa), Humboldt Current System (Near Peru), and the Benguela Current System located on the South West Coast of Africa. According to Carr (2001), the most productive of all the four EBCs is the Benguela Current System with 0.37 Gt C per year based on the 1997-1999 study of the satellite-based estimates of potential primary production. According to Chavez and Messié (2009), the upwelling favourable winds generally increase offshore to maximum velocities at around 50-200 km away from the coast and then flatten or weaken further into the high seas.

Hutchings et al. (2009) describe the Southern Benguela region, a subregion in which our study area is located as an unusual system as it is surrounded by two stratified warm temperate boundaries on both sides of the wind-driven upwelling system. These boundaries are the Agulhas current leaks (rings) characterised by warm Agulhas current properties and warm Angola current as displayed in Figure 2.5. The currents and features shown in Figure 2.5 in the Benguela upwelling system include upwelling cells and currents. The black lines showing main currents are the surface currents such as the Agulhas Current (AgC), its retroflection on the South coast, the Agulhas rings formed through the breaking of the AgC during retroflection, the Benguela coastal current (BCC), the Benguela oceanic current (BOC) which is part of the larger Sub tropical gyre (STG). The other currents that are part of the large scale circulation around our study area are the South Atlantic current (SAC) and the Antarctic circumpolar current (ACC).

The Equatorial Under Current, South Equatorial Current and counter current (EUC), (SEC) and (SECC) respectively are very far to have a known effect on our region of study even though they are also shown as part of the Benguela current Large Marine Ecosystem (BCLME). Not shown in this figure (Figure 2.5) is the South Atlantic Central Water (SAWC) that flow underneath the sea surface as the pole-ward counter current on the Benguela upwelling system. The SACW is responsible for the nutrient rich cold cells that are in the Southern Benguela upwelling system (Ismail et al., 2015).

The entire West Coast constitutes what can be referred to as an upwelling region. However, there are hot-spots where upwelling takes place in more intense rates and for longer periods in a year, aided by the shape of the bathymetry. These locations are called upwelling cells depicted in grey patches in Figure 2.5. The biggest in the entire Benguela region is the Lüderitz plume (cell) located on the coast of Namibia.

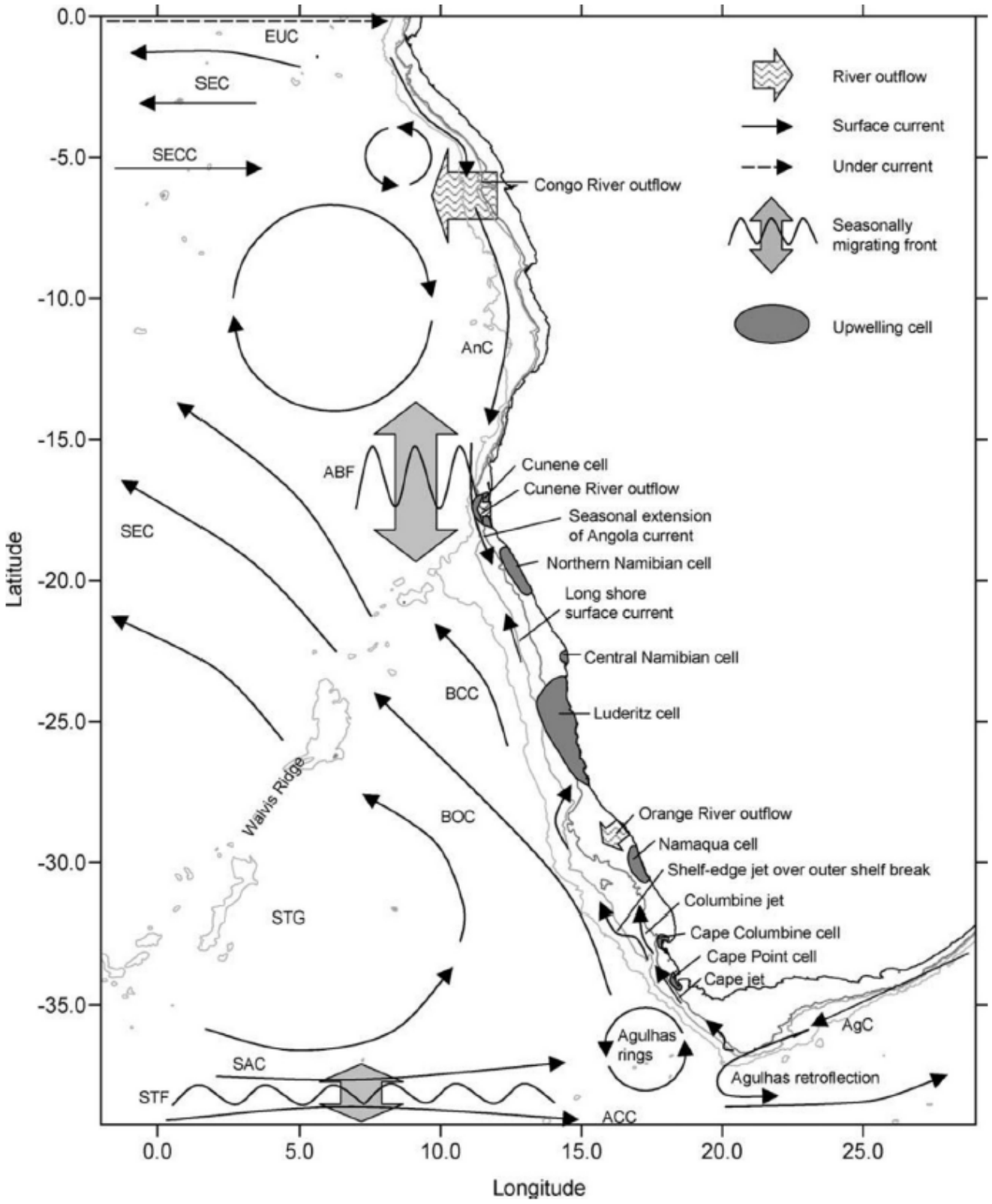

FIGURE 2.5: A conceptual depiction of major aspects of the physical oceanography around the Benguela upwelling system, particularly the southern Africa on the West Coast as shown by Verheye et al. (2016), also demonstrated by Payne and Punt (1995).

The cells that are most relevant to our sub-region (study area) are located below the Orange river and near the coast of Cape Town, namely the Cape Columbine cell and the Cape point cell located 1° North of Cape Town and at the coast of Cape Town respectively as described by Pfaff et al. (2011) where the Cape point cell is also referred to as the Cape Columbine plume by Payne and Punt (1995). Further description of the remote features presented in this figure is can be found in the reference paper mentioned above.

The features of the nutrient distributions have been analysed in the EBC's by various papers. The important nutrients are those that have a biological importance on the organisms and ones that are transported by wind-driven processes namely nitrate, phosphate, silicate and iron (Messié and Chavez, 2015). In the Benguela region along side with the Peru ecosystem, the latitudinal variation of primary production is driven by iron supply from sediments (Messié and Chavez, 2015). Messié and Chavez (2015) further claims that iron regulation is generally a major micro nutrient alongside light supply in other regions as well, therefore, there may be further hidden interactions on the chemical level that need to be understood in the Benguela ecosystem to ascertain the one primary driver that is unique from other systems.

Wind-driven nitrate supply through upwelling of nutrient-rich bottom waters is a main regulator in this region but nitrate may remain underutilised if the iron and light are low as in the Peruvian Austral winters (Messié and Chavez, 2015). Similarly the high abundance of iron is not useful if there is no upwelling (Messié and Chavez, 2015). The source waters for upwelling in the Benguela region 20-35° S come almost exclusively from 26.5 σ_θ of potential density according to Carr and Kearns (2003). Carr and Kearns (2003) state that the characteristics of the upwelled Benguela current water are 12-18 °C of temperature and 34.9-35.2 psu of dissolved salt, although Ismail et al. (2015) claimed a value that is slightly outside this range (34.8 psu) on a more recent study.

The process of upwelling involves vertical transport of colder water masses from the deep ocean to the shallower reaches of the ocean as stated earlier in the chapter. This process supports majority of fishing industries because it brings nutrient rich deep sea water into the euphotic zone (Schumann, Perrins, and Hunter, 1982). The fish then prey on the low trophic level species that prey on the plankton and micro-organisms that rely on the upwelled nutrients. The vertical transport of nutrients is due to two forms of upwelling, the transport driven and the curl driven upwelling. Rykaczewski and Checkley (2008) showed that the two mentioned upwellings have a positive effect on production in two unique ways. They demonstrated that zooplankton are relatively larger in areas of Ekman upwelling as opposed to areas of curl driven upwelling which is characterised by slower vertical transport. Rykaczewski and Checkley (2008) goes further to demonstrate that curl driven upwelling even though slower in vertical velocity, brings up more water volume as it tends to cover larger areas, therefore, both forms must be estimated in order to fully estimate availability of nutrients in each time.

Albert et al. (2010) demonstrate with blue vectors that more surface area of upwelling is possible due to the combined effect of curl-driven and coastal upwelling on the left plot of Figure 2.6. The pole-ward undercurrent is the source water referred to earlier as the SAWC. The into the page red back end arrow vectors shown on the top of the diagrams depict wind stress. It is appearing on the left hand plot that these vectors are diminishing towards coast. This is the phenomenon referred to as wind curl. It is driven by the friction of the coastal land and mountains.

The Ekman upwelling of the right hand-side depicts winds over the surface of the ocean that induce friction on the water column and this frictional effect is altered in direction as it is transferred down the water column due to Coriolis force (not shown). This Coriolis effect drives the net movement of the Ekman layer of water ($\approx$ 150 m deep) offshore, resulting in coastal vacancy which is then filled by deep waters near the coast by surging up to near surface. This is called coastal upwelling.

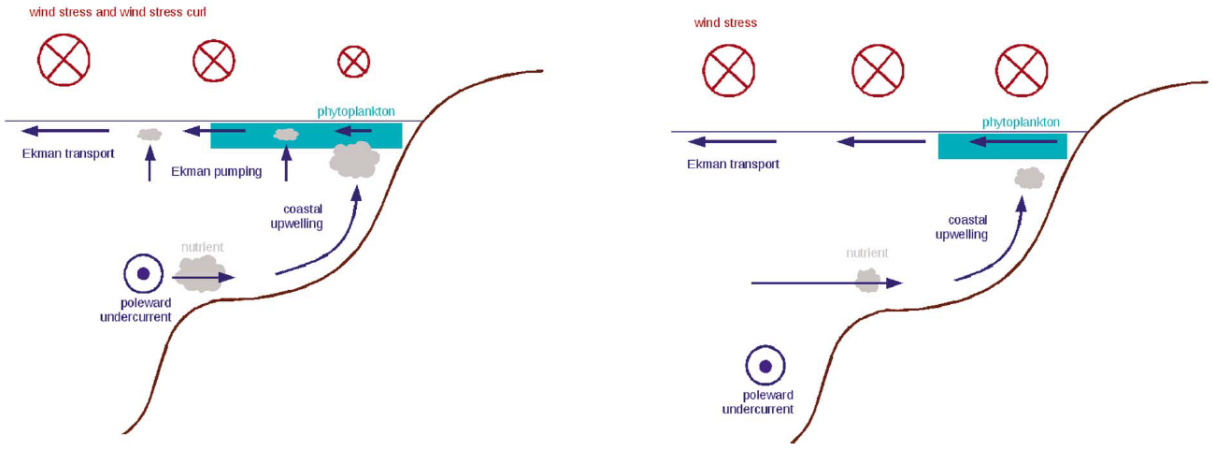

FIGURE 2.6: Schematic scenarios of the curl driven and transport driven upwelling, on the left and right respectively Albert et al. (2010)

Rykaczewski and Checkley (2008) fully demonstrated that the approach of estimating fish production from the analysis of upwelling favourable winds is a working approach in the California coast (Northern hemisphere) by estimating the effects of upwelling on sardine production. They discovered that there were positive correlations between the curl driven upwelling of the Northern hemisphere summer upwelling and the surplus production of sardine. The link between physical oceanography, food chains and food stocks is a very curious and intuitive way of doing scientific analysis through correlations of recruitment with climatic variables.

2.3 Hake Distribution in the Southern Benguela Region

The hake fishery is the most economically prominent sector within the fishing industry in south Africa, it is also the largest exporter of the perishable frozen products and has an international reputation of quality and reliability (Kleinschmidt, Sauer, and Britz, 2003). This adds to all the reasons why there is a need for improved efforts to fish effectively. Since the early 20th century, the Cape hake fishery have formed the basis of a substantial local trawl fishery (Botha, 1985). In the late 1950's, the South African trawler fleet has been split into two parts; the inshore segment (vessels operating in water depths of range 30-120 m) comprising of the hand-line and longline and the deep sea fleet that operates in depths that are deeper than 110 m (Japp, 1990), the latter being the focus of this paper.

Marine biologists consider the species of hake in the fishing grounds of South Africa to be two distinct species, namely the *Merluccius capensis* and the *Melrluccius paradoxus* (Garavelli et al., 2012). It is important to know about their distinction when considering conservation and species distribution analysis, however there is no empirical separation during fishing activity since the two are considered a single stock in the fishing industry as stated by Gordoa, Masó, and Voges (2000), Botha (1985), and Fairweather et al. (2006).

Hake is the target species for the deep trawl fisheries in the south and West Coast of South Africa. However, it is important to note that a whole host of other well known fish species get harvested as an incidental bi-catch. These fish include the *Genypterus capensis* commonly known as the Kingklip, *Zeus faber* commonly known as John Dory, *Thyrsites atun* known as Snoek, *Lophius vomerinus* known as Monkfish and many more. Table 2.1 displays the distribution of catches over the last few years and gives an idea of the proportion of by-catch as compared to the targeted hake that is landed by the deep hake fishery in a period of a year. One of the aims of this study is to also investigate the most suitable fishing months for hake in terms of volume (ton) where there is a good size mix of hake (large-medium-small in approx. equal proportions) and the bi-catch is minimal in the harvest uptake. The hake sizes (grades) are given as small $\leq$ 15 cm, medium $\leq$ 16-34 cm and large being $\geq$ 35 cm from head to tail (Mbatha et al., 2019). This is a grading standard also widely accepted across the hake fisheries.

TABLE 2.1: Hake and by-catch catches in 2016 and 2017 according to the fishing industry handbook 2018.

Landed nominal mass (kg) Deep-sea trawl catches		
YEAR	**2016**	**2017**
Hake	125 781 176	117 170 272
Cape Horse Mackerel	2 679 680	2 518 065
Monk	7 342 960	7 877 236
Kingklip	2 600 967	2 202 815
Snoek	2 689 530	4 147 047
Ribbon-fish	3 439 281	4 686 955
Jacopovers	1 064 476	1 001 112
John Dory	1 260 311	1 267 778
Atlantic Pompfret	4 378 397	3 292 173
Chokka	349 126	248 108
Other	2 864 522	1 478 692
Total	**152 770 746**	**145 890 253**

Note: Source: (*Fishing Industry Handbook*, 2018)

Fish recruitment depends largely on the environmental conditions, particularly those experienced by early ichthyoplanktonic life stages of the fish (Garavelli et al., 2012). The spawning of hake takes place virtually all year round and it is also possible that female hake may spawn more than once a year according to Punt and Leslie (1991). This annually scatted bi-modal distribution of the reproductive cycle makes it possible for the West Coast hake to have an all year long availability when linking availability to the species natality. It is important to highlight that we are not running a production model in this study but a simple correlation analysis model. Most methodologies which can be used in the management of this renewable resource are mathematically and statistically complex and computationally intensive and they are mired with high level of variance since their conclusions are based on assumptions that are sometimes hard to verify (Punt, 1992). The variety of conditions that exist during a fishing activity makes the ability of the fishing gear to have a strong bias when assessing the fish stocks in different environments (Pillar and Barange, 1997). This adds to the level of variance that arises due to the environmental conditions.

Most studies have concluded that hake generally feed at night in mid-waters as opposed to the bottom that they inhibit during the day (Pillar and Barange, 1997). They are also not daily feeders, juveniles rise to the feeding locations once in 2 days while adults can dwell at the bottom for up to six days. This sheds light on the reason why low catches are generally observed at night. The large fish are believed to be cannibalising on the small fish. This is considered as

another possible reason for less movement into the mid waters by the large hake (Payne, Rose, and Leslie, 1987). According to Botha (1985) and Booth (2000), the sub-species of hake that inhabit the mid waters is the *Merluccius capensis* which is situated at regions of around 400-500 m, the deeper dwelling sub-species is the generally bigger *Melrluccius paradoxes*. (Botha, 1985) suggests that most fished marine ecosystems studied for over a decade show structural shifts in fish assemblages. This could imply that the observations made in the earlier stages of our decade long study may be less important in the conclusions and deductions of patterns from the study. Shallower assemblages are more variant on a seasonal basis compared to the deeper assemblages through-out the year (Atkinson et al., 2011). This signifies the importance of the role played by water temperature since the shallower waters will be more affected by solar heating than waters in the deep.

Atkinson et al. (2011) also assert that fish assemblages in the deeper shelves could be more affected by conditions such as oxygen concentration in those regions more than the temperature. However, the strong parallel relationship between water temperature and dissolved oxygen concentration in water cannot be ignored as it could actually be the driving mechanism even in the shallower waters since cold water is usually low in oxygen. This could become a seasonally induced oxygenation and deoxygenation of the environment. Even though bottom fish are not at risk of having their habitat lacking in oxygen since deep waters are seldom deoxyginated, they are most probably less tolerant of high temperature.

Mbatha et al. (2019) found that juveniles (small hake) are the most affected hake size by dissolved oxygen compared to variables like bottom temperature and inter-annual variation in the 2003-2013 decade of study. They also found that small hake specifically gets negatively affected by both the extreme lows and extreme highs of dissolved oxygen while the medium and large hakes only gets badly impacted by very low oxygen concentrations only. This study will be comparing the portion of its decade's results with results from Mbatha et al. (2019) during the period of overlap of the two studies 2009-2013.

3 Methods

3.1 Study Area

The vessels operated in the displayed region in Figure 3.1 and collectively collected data for the time period of 12 December 2009 till present (but the study runs till 30 September 2018). The data is all located within the depth contours depicted in Figure 3.1 which is where all of all the catch data we have was obtained for this study.

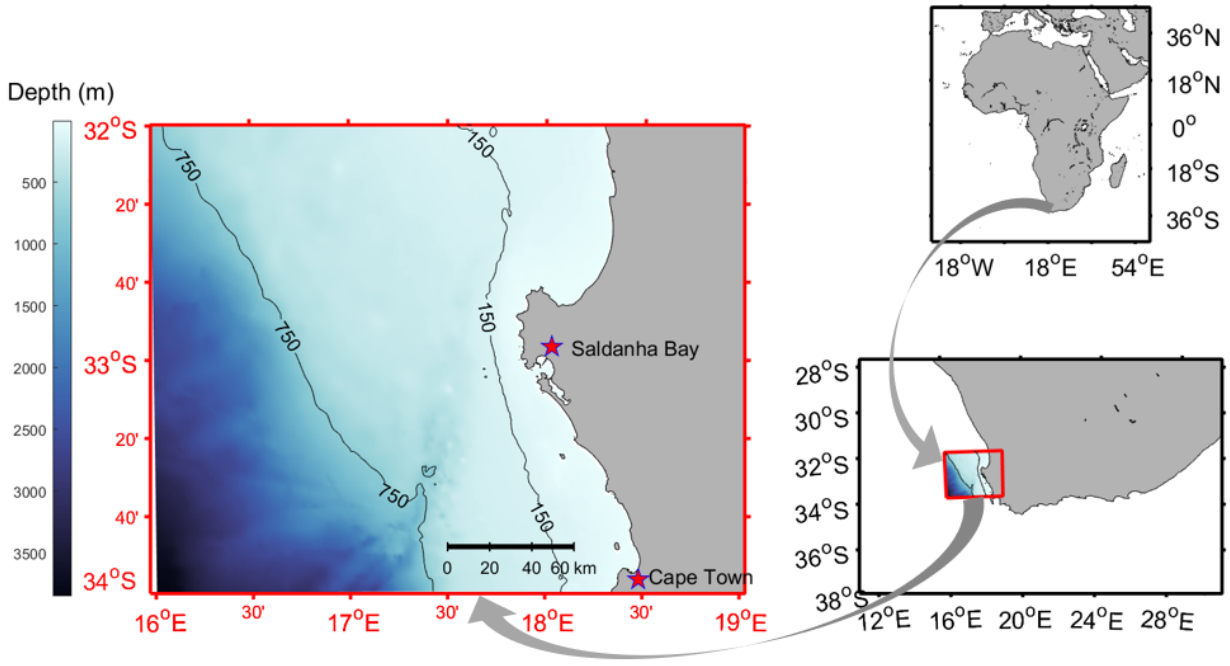

FIGURE 3.1: The study area is named the West Coast fishing grounds and it is the southern region of the Benguela upwelling system.

The West Coast fishing grounds in the red-boxed region are located at the southern most part of the Benguela upwelling system which spans from the coast of Angola down to the Cape (the southern most edge of Africa) . The analysis is performed within the 150-750 m isobaths because the deep hake fishery is permitted to fish in waters deeper than 120 m as already outlined in the previous chapter. The selection of the minimum depth as 150 m as opposed to 120 m is based on the Sea Harvest catch data since the company deliberately targets 150 m as its minimum. The

750 m isobath is based on the gear strength limitations. This region is also bounded by the 32°-34°S latitude which this is where the major focus of the research will be located. The justification for these latitudes is that this region is hypothesized to have different patterns from the region on latitudes above and below it in terms of seasonality of catches according to fishermen at Sea Harvest. It was found to be logical to treat it as a stand alone study area so that it may be compared with the outcomes of the other regions in future to test the said hypothesis. This whole region is referred to as the West Coast fishing grounds for the purposes of this study. The West Coast grounds are at the door step of the Saldanha Bay harbour, the busiest of all the hake fishing grounds in the coast of South Africa according to Sea Harvest database.

3.2 Fishing Method Process & Associated Error Sources

3.2.1 Fishing method

The fishing data used in this study have been obtained and recorded through digital data loggers (DDLs) installed in four Sea Harvest fresh fish vessels (V1,V2,V3 & V4) and a single freezer vessel (F1). The trawl catch data is collected during normal fishing activities by means of deploying a large net that is 50-60 m from the foot-rope to the codend see Figure 3.2, down to the ocean where the sea floor is relatively flat and then the vessel steams at 3-4 knots in the trawling direction. Trawl doors get stretched and opened by the force of the water on them and can stretch as wide as 100-140 m apart from each other allowing fish to be directed into the opening of the net. The warp on both doors is the cable that connects the vessel to the oval doors for pulling the net. The fishing trawl process, commonly referred to as a *drag* in the fishing industry usually takes 1-3 hours. There is lack of interest for longer drags as they dramatically lower the quality of catch. This is the time that is factored in as the *effort* during the CPUE computation as detailed in Section 3.3.2.

The sensors are located at the top of the net on two locations labelled 1 and 2 as shown on Figure 3.2. The sensor in location 1 is called the trawl sounder and it records the activity on the net opening, the speed and bottom temperature. Please note that temperature will not form part of this study. The calibration of the sensors is not always guaranteed and the small uncertain variations in bottom temperature are less likely to give indications on upwelling conditions, Therefore have been excluded from the study at this point. The sensor in the second location is called the catch sensor, it gives an indication of tension on the net mesh which signifies the tonnage of fish that are already caught in the bag. These sensors are general equipment that serves to notify the fishermen about what is happening in the net during trawling and updates them when should they pull the net back up to the vessel.

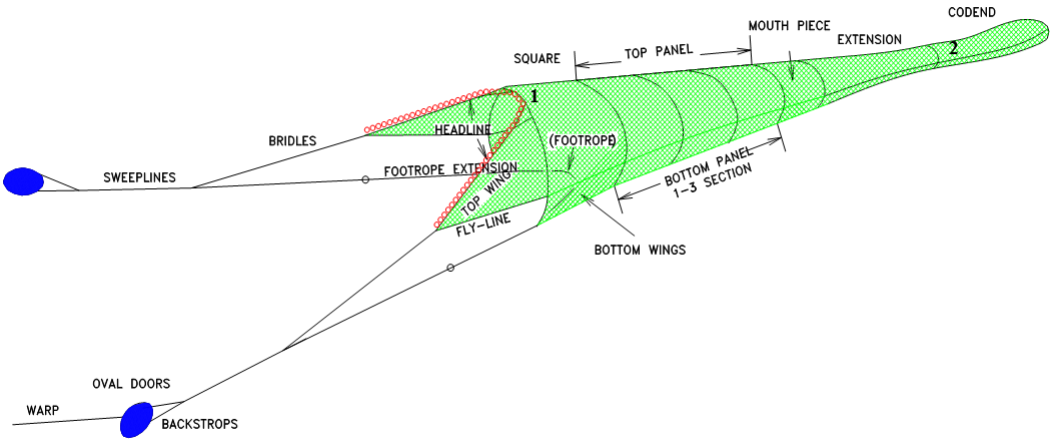

FIGURE 3.2: Trawler net diagram demonstrating a bottom trawl fishing gear

3.2.2 Sources of error

There is a significant amount of error that is associated with the catch equipment. The Sea Harvest database does not have a well documented history of misfires in the fishing launch. The misfire is a failed proper deployment of a fishing net, which is a seldom scenario that occurs when the trawling doors (oval blue parts on the net diagram) falls flat and can not open the net. The record of such an event would indicate a low catch or a zero ton catch on the DDL because there is no fish caught when in actual fact this entry must be completely discarded. During the data cleaning process, all extreme outliers were removed in order to eliminate errors of this nature and other forms. However, if a misfire occur during the mid-stage of a successful trawl, then the time spent in a trawl with the nets closed would negatively affect the CPUE estimation. This is worth pointing out as a possible source of error which will be attributed to the white noise on the data signal alongside other sources of error.

The mesh-grid of the net ranges between 140-160 mm for vessels V1,V2,V3 & V4, 160 mm being mostly for vessel F1. Vessel V2 and V3 have the same trawling gear which indicates that they have the same opening sizes during the fishing process. The Vessel V1 and V4 have the same gear but V4 was converted into a freezer and soon retired from the fleet. F1 has a unique gear equipment as it is a freezer vessel. The mention of all these variations is to acknowledge and indicate further minor sources of bias in the catch estimations. However, it is important to note that there is no intention of taking the vessel to vessel biases on this study as it would expand the scope of the study. Further sources of error are environmental. The most obvious among them being the underwater topography as well as the current speeds and directions. For an example, fish may be influenced by current directions and they may prefer rocky environments which are not trawl-able. These factors among many others makes our study seem narrow as

they will all not be factored in. However, the general patterns should be detectable as the factors we are analysing are known to be very important.

3.3 In situ Data Handling Procedure and Analysis

3.3.1 Hake data description

The Sea-harvest data is a proprietary database that has been made available for the purposes of our study but not for complete open access.The detailed analysis of this data is therefore performed on the protected chapter, Chapter 5. This chapter will be accessible to members of the scientific community and general public upon request if a non-disclosure agreement is signed by that individual. The summery of these data is detailed as follows:

- The data structure is in a tabular form. In it among many variables, the daily catch quantities for each species are given in each drag. The computation of these catch quantities is given in the next subsection, Section 3.2.2
- The starting times and the end times of each drag are part of the tabulated structure, therefore the duration of the drag is determined from their difference. This represents what is referred to as the $effort$ in Section 3.2.2.
- The CPUE is given as a column in this tabular structure after having been determined by making use of the above variables, catch and effort. This is detailed in the computation section as well. Section 3.2.2.
- The CPUE is then averaged to give a daily average CPUE. The daily average CPUE is again averaged to give a monthly average CPUE. The rationale behind using a monthly CPUE is that a daily or weekly data would have too many gaps resulting from weeks of no fishing recorded possibly because of holiday breaks towards the end of the year where most vessels are on leave or accidental skipping of the data recording process on the DDLs during fishing. Too many breaks are minimised by using monthly averages.

3.3.2 Hake data computation.

The first operation performed on the structure was to insert an additional column of conversion factors for each and every species fished, where conversion factor (c) is 1.46 for headed and gutted hake. The rest of the species that may be caught are listed in Table 2.1, they each have a different conversion factor but all only but hake will be considered in this work. These conversion factors are gazetted and shared by government to the fishing companies. These conversion factors are considered as accurate means of measuring the initial weight of an unprocessed fish after it has been processed. The conversion factor for an unprocessed round fish would be 1 to

show that it is whole. The conversion factors are important in the calculation of a *Catch* because their inverse ($\frac{1}{c}$) signifies a proportion of hake that remains after the head and guts have been removed through the process of heading and gutting. The number of packing units n, refers to the number of bins or tubs used to stock the fish after heading and gutting and the packing unit capacity p is their volume in kilograms. The product of conversion factor c, number of packing units n and packing unit capacity p gives the computation of the catch as shown below.

$$Catch_g = cn_g p_g \tag{3.1}$$

where g=1,2,3...n is the number of drags. *Catch* is used to compute the CPUE. Fairweather et al. (2006) gives further background information on the computation of a CPUE.

$$CPUE_g = Catch_g(\frac{60}{effort_g}) \tag{3.2}$$

The CPUE is a strong indicator of the catchability of fish because it takes into account the time spent on achieving a single catch. The unit of time used for the $effort$ is minutes but because of the numerator of 60 factored on the CPUE equation over the $effort$, the output of the CPUE will be in ton/hour. It is a general practice in statistics to take ecological data sets that have a skewed distribution through a logarithm transformation in order to make it fit a normal distribution. This is the case in our study, therefore, a log of CPUE was taken before finding the standard deviation above the mean for the hake species.

$$\mu = \frac{1}{N}\sum_{g=1}^{N}\log(CPUE_g) \tag{3.3}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{g=1}^{N}(\log(CPUE_g) - \mu)^2} \tag{3.4}$$

The Standardised Anomaly Index (I) is a component that is very useful in determining the level of catch on a comparative level in a time-line. In other words a high catch is a catch that is more than 1 standard deviation (σ) above the mean (μ) of all fishing drags performed in the timeline. This will play an important role in Chapter 5.

$$I_g = \frac{\log(CPUE_g) - \mu}{\sigma} \tag{3.5}$$

3.4 Publicly Available Scientific Wind Data

3.4.1 Reanalysis wind data.

The wind field data utilised in this study is obtained from the ERA interim database found in the European Centre for Medium-Range Weather Forecasts (ECMWF[1]) repository at a grid resolution of 0.75° × 0.75° and the ERA 5 database product from the Copernicus Climate Data Store[2] repository. The ERA interim product available from 1 January 1979 to 31 August 2019 is now fully replaced by the ERA 5 product which means that a replication of this study in future would be fully based in ERA 5 in the same way that the improvements and modelling advances in ERA-Interim also necessitated that it replace ERA-40 reanalysis in 2002 (Dee et al., 2011). The ERA 5 data product entails a high resolution zonal and meridional wind components at 0.25° × 0.25° degree resolution. The 10 m above sea surface wind velocity components are used as representative of surface winds.

The retrieval of the reanalysis data is performed on python programming language by connecting to the ECMWF database server through their connecting Application Programming Interface (API) and by making use of the automated python script for a recursive data download process. The data download script is freely available on the ECMWF repository for this purpose. It only requires a few adjustments for the specific requirements of the user. The ERA 5 data download process is a very similar process on the Copernicus Climate Data Store except that you don't need to run a script outside the repository platform to start retrieving data since the data can be queried there on the platform and then the Network Common Data Form (netCDF) data file(s) can be retrieved. The netCDF is the data format for both wind products retrieved.

The ERA 5 and ERA interim wind data was processed in a Linux shell terminal. The Era 5 wind data structure is a 4 dimensional matrix array (X,Y,W,t), where X,Y,W and t are the zonal component, meridional component, wind magnitude and time (3 hourly) respectively. For the processing of the 3 hourly (00:00, 03:00, 06:00... , ...18:00 & 21:00) reanalysis data into daily & monthly mean files, I used the Common Data Operator software (CDO). The ERA interim wind data set also follows the same data structure, but it is available as 6 hourly winds instead. Hourly winds were avoided to save computer space, this is why the data was taken through CDO in order to make it monthly instead of 3 or 6 hourly. The final analysis of these data was performed on MATLAB for data visualization, time series analysis and lagged correlation computations. All the equations given in this section represents MATLAB programming operations performed on various computations. The R programming language was employed in the computation of the statistical comparisons and on the modelling section. All the statistical section

[1] https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim
[2] https://cds.climate.copernicus.eu/!/home

outputs were computed in R.

3.4.2 Ekman transport upwelling estimation

Ekman transport driven upwelling (Ekman upwelling) occurs on the coast at a location of Ekman transport, in a region that is within the Rossby radius of deformation. Ekman transport is the flux of the offshore displaced water that moves due to the forcing of the wind stress (τ) component parallel to the coast. The wind stress (Nm^{-2}) is computed by taking the product of the density of air, the drag coefficient of the steady sea surface and the wind magnitude acting on that location as shown below.

$$\tau = \rho_1 C_D W^2 \tag{3.6}$$

The Coriolis force on the sphere is expressed by means of the Coriolis parameter (f) is an inertial force associated with a rotating object (earth) that impacts large scale features and induce a deviation in a perpendicular direction to the prevailing motion. Its effect in the upper portion of the ocean when acted upon by an alongshore wind stress is to drift the water transport in a direction that is perpendicular to the wind stress in an anticlockwise drift (leftward) in the southern hemisphere. The omega (Ω) is the angular velocity of the earth and the latitudinal position is given as θ.

$$f = 2\Omega \sin(\theta) \tag{3.7}$$

The rate of upwelling can be estimated from the Ekman transport in magnitude because the Ekman upwelling replaces the volume of space that would be left vacant by water that is subjected to Ekman transport ($m^2 s^{-1}$) on the upper ocean when it is drifted by the wind. This is to ensure the conservation of mass as the volume of the ocean stays fixed. The meridional wind stress component (τ_y) is approximately equal to the alongshore wind stress ($\tau_y \approx \tau_a$) since the shape of the West Coast grounds and North Coast area coastline is mostly oriented North-South. The Ekman transport is therefore given by the following expression.

$$T_{E_x} = \frac{\tau_a}{\rho_0 f} \tag{3.8}$$

The Rossby radius of deformation (R_d) is an important parameter to measure the spatial scale of balance between rotational and gravity effects. R_d diminishes with latitude from the equator and it is estimated to be around 20 km at 30°S according to (Bachèlery, Illig, and Dadou, 2016).

The associated vertical volume flow is computed by considering the Rossby radius of deformation and the area (A) of the study. The area between 32° and 34°S and within the 150-750 m

underwater contours where all the permitted fishing takes place is multiplied by the estimated vertical upwelling velocity. Marshall and Plumb (2007) shows a full derivation for both the Ekman and curl driven upwelling. Volume flow (V_f) of the Ekman upwelling in this case is given as:

$$V_f = A \times \frac{T_{E_x}}{R_d} \tag{3.9}$$

3.4.3 Wind stress curl upwelling estimation

Wind stress curl is the torque acting on the surface of the ocean as a result of induced shear by the unequal distribution of surface wind stress. Wind stress curl is denoted as:

$$Curl_z(\tau) = \frac{\partial \tau_y}{\partial x} - \frac{\partial \tau_x}{\partial y} \tag{3.10}$$

A central difference method is applied in order to determine the curl field. The expression for the numerically estimated curl is:

$$Curl_z(\tau) = \frac{\tau_y(i, j+1) - \tau_y(i, j-1)}{2Dx} - \frac{\tau_x(i+1, j) - \tau_x(i-1, j)}{2Dy} \tag{3.11}$$

- i and j are the row and column index of the current grid cell
- Dx and Dy are the widths and heights of the current grid cell in the $0.25° \times 0.25°$ resolution data. Dy is constant ($\frac{111000}{4}$) m while Dx changes with latitude. This is because the Dx gets narrower the further you are from the equator.
- The denominator 2 in the equation indicates the number of indices (i and j) that a grid has from its centre going either side of the grid pixel as shown in Figure 3.3.

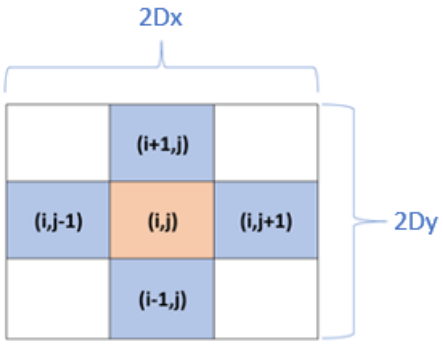

FIGURE 3.3: Central difference grid model schematic of the spatial wind data.

The expression above is computed through the MATLAB function courtesy of Ramkrushn Patel[3] which had to be edited, tailored and corrected for parts that did not work in it such as the

[3]https://www.mathworks.com/matlabcentral/fileexchange/53392-wind-stress-curl-from-wind-data

structure of the dimensions and the computation of the backward difference of the function. The Ekman vertical velocity estimation is approximated by the wind stress curl. The computation is performed by determining w_{curl} which is the upward velocity of the water column.

$$w_{curl} = \frac{1}{\rho_0}\left(\frac{\partial}{\partial x}\frac{\tau_y}{f} - \frac{\partial}{\partial y}\frac{\tau_x}{f}\right) \tag{3.12}$$

The associated vertical volume flow can be computed by considering the area of study and the upward velocity. Volume flow (V_{f2}) in Sverdrup (Sv) is the final output. It is determined as follows:

$$V_{f2} = A \times w_{curl} \tag{3.13}$$

3.5 Chlorophyll Data

The chlorophyll data is obtainable from the NASA Ocean Colour[4] data repository. The resolution of this product is obtained as a 4 km monthly mean in order to be consistent with the wind and hake data frequency. The MODIS (Moderate-resolution Imaging Spectro-radiometer) terra level-3 product from satellite orbiting the earth and is capable of collecting ocean colour data imagery through the use of microwave radiation sensing. Radiation sensing at these wavelengths poses a set of problems for daily analysis as there is often a band of clouds in the area which makes it impossible for the transmission of data from the sea surface to the satellite. This is part of the reason why the monthly mean chlorophyll data has been chosen as opposed to the daily data. 8 day chlorophyll composites are also avoided as they do not align well with the catch data which has been aggregated to monthly means. The log base 10 of the chlorophyll concentration has been taken in order easily visualise the concentrations. Otherwise chlorophyll concentrations are very close to zero for the most part and a few are quite high, therefore, a log base 10 transformation is necessary and is usually taken to flatten an otherwise skewed logarithmic data. The chlorophyll concentration is in (mg/m^3). The chlorophyll monthly mean computations are performed by the computation of the total number of pixles (P) in the target region and then the estimation of the mean chlorophyll in area A defined the previous section. Monthly values are used in a time series to determine the biological product (Chl_a) average over the area. The following equation estimates Chl_a concentrations averaged over the entire region of study in a month:

$$(\overline{Chl_a})^t = \frac{1}{P}\sum_{n=1}^{P} \log_{10}(Chl_a)_n^t \tag{3.14}$$

[4]https://oce andata.sci.gsfc.nasa.gov/

Where t=1,2,3...T, the number of months within the span of 2009-2018 of fish data collection and n =1,2,3...P, is the total number of 4×4 km grid pixels, and $\overline{Chl_a}$ is the mean monthly chlorophyll at month t.

3.6 Other Data Sources and Tools

For the bathymetry, the GEBCO[5] (The General Bathymetric Chart of the Oceans) database repository was consulted for the high resolution 1 minute ocean floor data. GEBCO is a combination of bathymetric data obtained through direct acoustic soundings and satellite altimetry data. The two methods are then integrated to determine an interpolated depth over the grid. The bathymetric data has been important in defining the boundaries of the study area and in the interpolation of the wind effects within the studied region for quantification. All analysis in this project have been performed by the use of Linux shell terminal, MATLAB programming software (Piretzidis and Sideris, 2016), (MATLAB, 2018), Python (Van Rossum and Drake, 2009) and R-Studio (R Core Team, 2019) for statistical analysis, visualisation and modelling.

3.7 Statistical Methods :Regression

The statistical method selected in the study is the multiple regression modelling where all predictor variables are investigated for their influence over the response variable. This is performed in detail in Chapter 6.3. The predictor variables are the monthly Ekman upwelling, curl driven upwelling and chlorophyll concentration while the response variable is monthly Hake CPUE. The objective is to come up with a regression that models a continuous variable Y as a mathematical function of the response X variables (Rencher and Schaalje, 2008). The Y represents hake CPUE whereas X represents each of the predictor variables already mentioned. The generalized mathematical expression of the multivariate regression is as follows:

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \epsilon \tag{3.15}$$

Where β_0 is the intercept and β_n is the coefficient of the predictor variable. ϵ is the error term, the part of Y the regression model is unable to explain.

3.7.1 Regression Model selection

Model selection is based on the next parametres which are each computed separately for each model and eventually compared in a tabular format (shown in Chapter 6) in order to ascertain

[5]https://www.gebco.net/

the most probable model to fit our data. The first one is the r^2 is given as

$$R^2 = 1 - \frac{SS_E}{SS_T} \tag{3.16}$$

The r^2 is a quantity that represents the amount of variation in the response variable that can be explained by the predictor variable(s). The SS_E is the sum of squares errors and the SS_T is the total sum of squares computed as $SS_E = \sum_i^n (y_i - \hat{y}_i)^2$ and $SS_T = \sum_i^n (y_i - \bar{y}_i)^2$ respectively. The $\hat{y}_i$ is the predicted value of the observation i whereas $\bar{y}$ is the mean of the continuous predicted variable. The adjusted r^2 on the other hand is good for comparing models that has multiple predictor models because each additional variable adds also adds to the variation therefore, an adjusted R squared penalizes every additional variable added in order to optimise r^2 of the model. For this reason adjusted r^2 is more valuable than r^2 in a multivariate regression.

$$R^2_{adj} = 1 - \frac{MS_E}{MS_T} \tag{3.17}$$

Where, $MS_E = \frac{SS_E}{(n-q)}$ and $MS_T = \frac{SS_T}{(n-1)}$ are the mean squared error and mean squared total respectively where n is the number of observations and q is the number of coefficients in a model which represents the number of predictor variables.

The standard error measures the goodness of fit where q is the same as above.

$$Std.Error = \sqrt{MS_E} = \sqrt{\frac{SS_E}{n-q}} \tag{3.18}$$

The t-statistic is determined by making use of the model coefficients and standard error when they are known. This serves to give the p-values as well as is performed by the following computation:

$$t_{Statistic} = \frac{\beta - coefficient}{Std.Error} \tag{3.19}$$

The p-value or $\Pr(>|t|)$ is the probability that a t-value is as high or higher than the observed test statistic value when the null hypothesis is true i.e when there is no detectable relationship. A high t-value indicates that it is less likely that the null hypothesis is true, i.e. that the coefficient is zero. A low p-value signifies that the coefficients are significant. Our study is performed at a 95% confidence, therefore the significance level of 0.05 becomes the threshold where p-value must be below 0.05 in order to conclude that our model is indeed significant.

Similarly, $\Pr(>|F|)$ is also a p-value and it is the probability that a F-value is as high or higher than the observed test statistic value when the null hypothesis is true. A null hypothesis is the assumption that there is no relationship between coefficients as already differently stated above.

The F-statistic, similar to the standard error measures the goodness of fit where MS_R is the mean square regression given as $MS_R = \frac{\sum_i^n (\hat{y}_i - \bar{y})}{q-1} = \frac{SS_T - SS_E}{q-1}$

$$F_{statistic} = \frac{MS_R}{MS_E} \tag{3.20}$$

The Akaike Information Criterion (AIC) proposed by Akaike (1974) is useful in ensuring a balance between the under-fitting of the model and over-fitting. It is important to ensure that the model is not under-fitted because in this case it will be too simple to detect the underlying trend and it would lead to a bias, however if it is over-fitted it will also pose a problem of over-complicating the pattern. The complex model will model the variance whereas a simple/under-fitted model can't predict reliably. AIC is employed to give balance in this regard and it is given in the expression below where $log(L(\hat{\theta}|Data))$ is the log-likelihood of a particular model and K is the number of estimated parametres.

$$AIC = -2\log(L(\hat{\theta}|Data)) + 2K \tag{3.21}$$

In this expression $-2\log(L(\hat{\theta}|Data))$ is the measure of how close the model fits the observed data (hake CPUE). The larger the $-2\log(L(\hat{\theta}|Data))$ is, the worse the model and the converse is true. The $2K$ part of the equation can be viewed as the penalty for model complexity. This is to ensure that over-fitting is prevented in the best model because the more the parameters, the more complex the model is. This will be used in the model selection in Chapter 6.3.

TABLE 3.1: Summary of the statistics and criteria that are used in the selection of a regression model.

Statistic used	Criterion applied
r^2	Higher value is preferable
r^2_{adj}	Higher value is preferable
F-Statistic	Higher value is preferable
Std. Error	Closer to zero the better
t-statistic	Should be greater 1.96 for p-value to be less than 0.05
AIC	Lower value is preferable
MS_E (Mean squared error)	Lower value is preferable
SS (Sum of squares)	Lower value is preferable
MS (Mean squares)	Lower value is preferable

3.7.2 Model validation

Model validation and diagnostics of the multiple linear regression methods are presented in accordance to chapter 9 of the Linear Models in Statistics book by Rencher and Schaalje (2008) and Linear Models with R by Faraway (2014). Model validation is performed in Section 6.3.3 where the model residual analysis, normality test, homoscedasticity test and leverage points assessment are performed. The four respective robust model validation steps mentioned on the previous line are detailed as follows:

Residuals plotted against fitted values. "If the model and attendant assumptions are correct, then a plot of the residuals versus predicted values should show no systematic pattern" (Rencher and Schaalje, 2008, pg.230). This quote clarifies that the model must show random distribution of residuals when plotted against the predicted values to show that the model is well fitting of the data it models. The underlying computations and derivations are all shown in the referenced books mentioned above. The model normality is generally displayed with a **Q-Q plot** or a histogram and it assists in eyeballing the normality of the data. If the data is not normal it must be normalized through a mathematical transformation. The **standardised residuals vs fitted values**. Tests homogeneity of variance of the residuals (homoscedasticity). The standardized residual is the residual divided by its standard deviation. Variance must be homogeneous as this is an important underlying assumption for the regression models as it has an impact on the residuals and the least square method is based on constant variance.

According to Selva Prabhakaran[6], the **Cook's distance** in a multivariate model approach computes the influence exerted by each data point on the predicted outcome. In simpler terms, this means that it locates the most influential points on the multivariate model. The most influential point is a point that is above the cook's distance as such points would generally be above 4 times greater than the mean of the observations. A point that is found to be above cook's distance is influential and a decision to remove it in order to improve the model can be taken if it can be justifiable to take it down. Otherwise it must be left on but noted as an exception. Cook's distance is computed with the following expression.

$$D_i = \frac{\sum_{j=1}^{n} \left(\hat{Y}_j - \hat{Y}_{j(i)} \right)^2}{\beta_t \times MS_E} \tag{3.22}$$

Where, $\hat{Y}_j$ is the value of j_{th} modelled response when all the observations are considered. $\hat{Y}_{j(i)}$ is the value of j_{th} modelled response, where the fit does not include observation i, where i is any single observation. MSE is the mean squared error as already defined. The β_t is the total number of coefficients in the regression model.

[6] http://r-statistics.co/Outlier-Treatment-With-R.html

R statistical packages are used to perform the above computations and manual computations are avoided for all steps. The R packages used are tidyxl for excel data importing and analysis , ggplot2 for graphical visualization (Wickham, 2016), MASS package for statistical tools (Venables and Ripley, 2002), dplR package for cross correlation computations (Bunn, 2010) other packages under the umbrella of the tidyverse package (Wickham et al., 2019) which is a collection of R packages dedicated to data science.

4 Results: Wind, Upwelling and Chlorophyll Analysis

4.1 Wind Analysis

4.1.1 Wind product visual comparison

The two wind products used in this project illustrated in Chapter 3.4 are shown in Figure 4.1. This clearly indicates the importance of the higher resolution wind products to capture the wind features, such as the wind curl and the near-coast circulation.

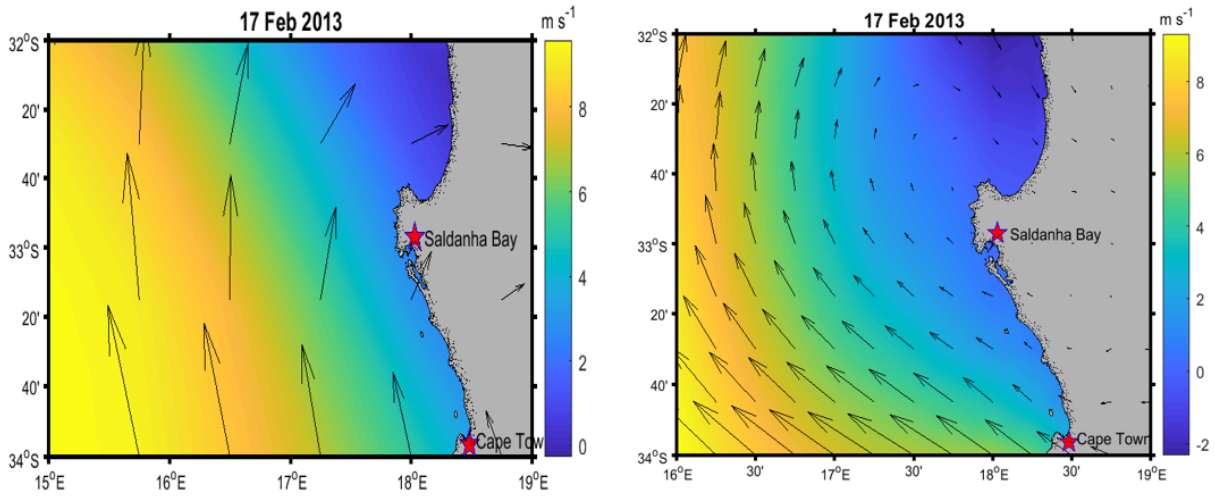

FIGURE 4.1: Mean daily wind vectors from a typical summer day (Left) Era interim 0.75°×0.75° winds show a poor resolution of the wind spin. (Right) ERA 5 0.25°×0.25° winds demonstrate the wind curl in this region.

4.1.2 Examples of wind conditions in different seasons

The SAWS website was consulted for the record of historical synoptic weather maps on which the ones that showed meridional wind pattern along the West Coast were arbitrarily selected on the basis of being in the middle of summer or in the middle of winter. These were picked in order to observe what are pressure regimes that leads to winds that support/suppress upwelling in different seasons. The winds that support upwelling in our study region are positive meridional winds, whereas the ones that suppress upwelling are negative meridional winds. These winds were scouted by paying attention to Cape Town weather station and other coastal stations all the way down to the southern most tip of Africa at Agulhas on the synoptic map. The selected days have already been displayed in synoptic maps in Figure 2.2 and 2.3 in Chapter 2 and the winds with meridional components can be seen in them. Figures 4.2 and 4.4 sits in the middle of winter and summer seasons respectively and displays the associated wind stress patterns associated with the pressure systems that exists during the two seasons.

The wind stress acting over the ocean transfers momentum which results in two forms of upwelling, the wind-stress driven upwelling and the wind-stress curl driven upwelling as detailed in Chapter 2. The cyclonic winds that are observable along the South African coast, in Figure 4.2 are the result of the low pressure system shown in the synoptic map in Figure 2.3. These cyclonic winds on the West Coast of southern Africa are north westerly (southwards towards east). The blue shading depicts winds that are southward component of the wind and the red depicts winds that are the northward component. This wind pattern does not generate upwelling but rather favours downwelling in the West Coast of South Africa.

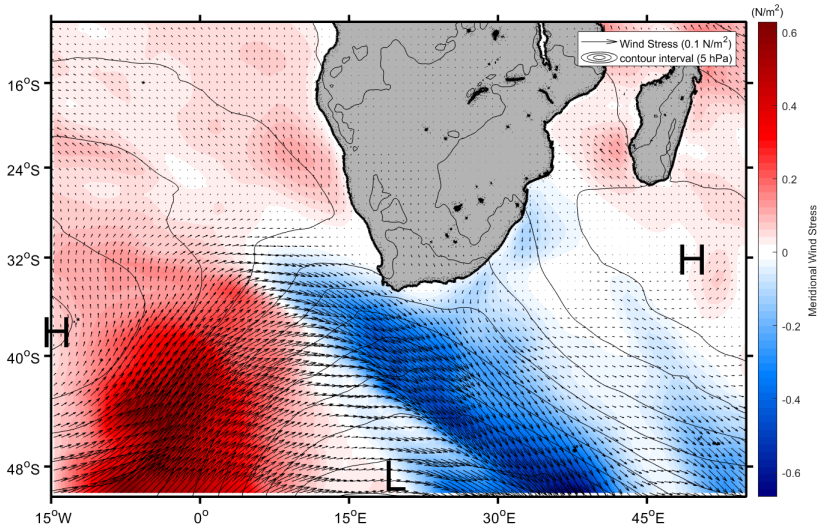

FIGURE 4.2: Wind stress (N/m^2) based in Era-interim wind data is given as vectors and wind stress meridional direction given as coloured field over the Southern Africa under under northerly winds in the presence of a low pressure system midday snapshot of 30^{th} June 2016.

The wind stress curl in the winter season depicted below during a low pressure system exhibits the form of pattern displayed in Figure 4.3. A south westerly wind stress through interaction with the coastal topography decreases in strength towards the coast, as can be seen with diminishing vectors towards the coastline in Figure 4.3. The take away message is that the coast induces a positive curl on the southward wind stress creating a downwelling favourable phenomenon near the coast. The negative wind stress curl further from the coast shown in blue indicates that the tendency of the wind spin in that area is clockwise as opposed to the anti-clockwise tendency near coast resulting in positive curl shown in red.

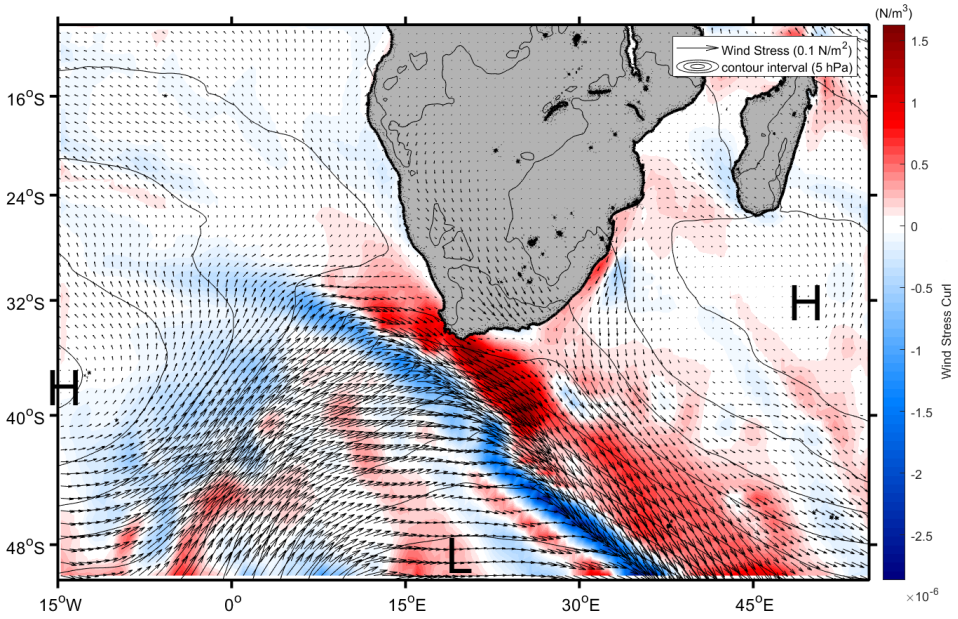

FIGURE 4.3: Wind stress (N/m^2) based in Era-interim wind data is given as vectors and wind stress curl (N/m^3) as coloured field over the Southern Africa under northerly winds in the presence of a low pressure system (30 June 2016) pattern.

The orientation of pressure systems in the presence of southerly winds is depicted in Figure 4.4. Pressure systems do move around and different HPS locations would come about as a result of a daily variation but they would not veer very far away from this pattern as this pattern is what brings about the positive meridional component of coastal winds in the West Coast. These are conducive conditions for Ekman upwelling as the north-ward meridional wind stress parallel to the coast generates an offshore drift of the Ekman layer of the ocean, resulting in upwelling.

The sea surface pressure patterns shown in the West Coast in Figure 4.4 & 4.5, are indicative of the South Atlantic high pressure system which drives the south-easterly winds that are required for Ekman driven upwelling. The positive shading indicates positive meridional wind stress on the West Coast which is primary for Ekman upwelling regimes. The random positive meridional winds day selected in the middle of summer displays the wind stress and curl conditions at midday.

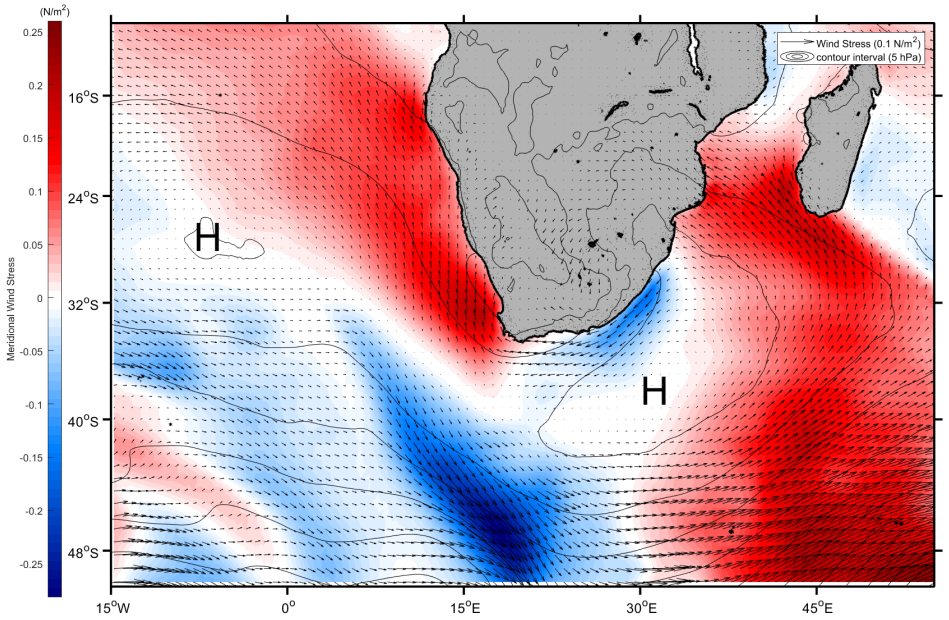

FIGURE 4.4: Wind stress (N/m^2) based in Era-interim wind data is given as vectors and meridional wind stress direction and intensity given as coloured field over the Southern Africa under summer conditions in the presence of high pressure systems snapshot (25 December 2016) pattern.

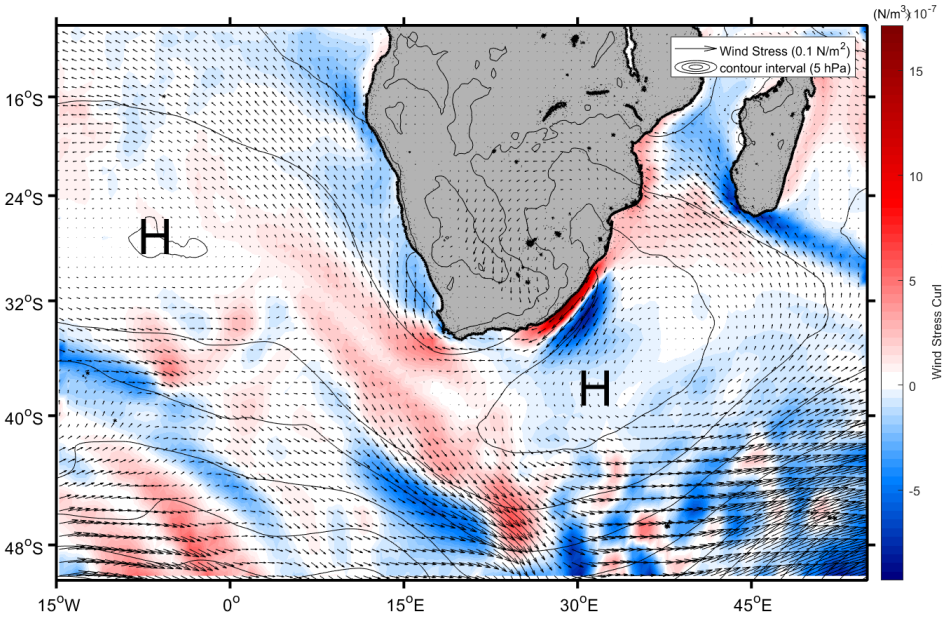

FIGURE 4.5: wind stress (N/m^2) based in Era-interim wind data is given as vectors and wind stress curl (N/m^3) as coloured field over the Southern Africa under summer conditions in the presence of high pressure systems (25 December 2016) pattern.

4.2 Upwelling and chlorophyll seasonality

Figure 4.6 displays the seasonal variation of the meridional (northward/southward) wind stress where throughout the summer months northward meridional wind stresses are higher, peaking in December and January. The winter months have no northward meridional wind stress or have very little of it as it can be shown in the month of June and on other winter months. The observation that strikes from this figure is that the strongest gradient is observable in the December month giving rise to a strong wind shear.

The Ekman upwelling velocity computed according to Section 3.4.1 is shown in Figure 4.7 for year 2010's. The selection of this year is on the basis that the number of fishing observations done per month in this year were the highest (over 30 drags per month shown in the next chapter in Figure 5.2, which allowed for a proper comparison of hake catches and physical attributes that may have led to them. The extended austral summer months (October, November, December, January February and March) of 2010 appear to have a lot more upwelling velocities as depicted in reddish field compared to extended austral winter months displayed in blue field of that year. These estimations are performed within the study area in which West Coast fishing grounds are located. The strong Ekman upwelling that appears in December is due to a strong meridional wind that is tilted nearly perfectly parallel to the coast.

The curl driven upwelling computed in Section 3.4.2 is shown on Figure 4.8. A seasonal variation of the curl driven upwelling is mostly visible in mid summer; December and January in 2010. March, April and November also presents stronger rates of this kind of upwelling compared to the winter months. The curl driven vertical transport is strongest in December where the strongest wind stress shear is also observed as pointed up in Figure 4.6.

The concentration of chlorophyll used in this study is described in Section 3.5. The chlorophyll monthly concentrations for year 2010 are shown on Figure 4.9. This year is selected as a case study for the reasons already outlined above. The chlorophyll depicts a seasonal variation when observed on a month to month basis. The late summer months (January February and March) of 2010 appear to have higher chlorophyll concentration including April. The blue field which signifies low concentrations of chlorophyll are more prevalent in winter months and they begin to increase in the late winter (August), although it is always high and above 0.5 mg/m^3 along the coast. In August and September during the late winter period.

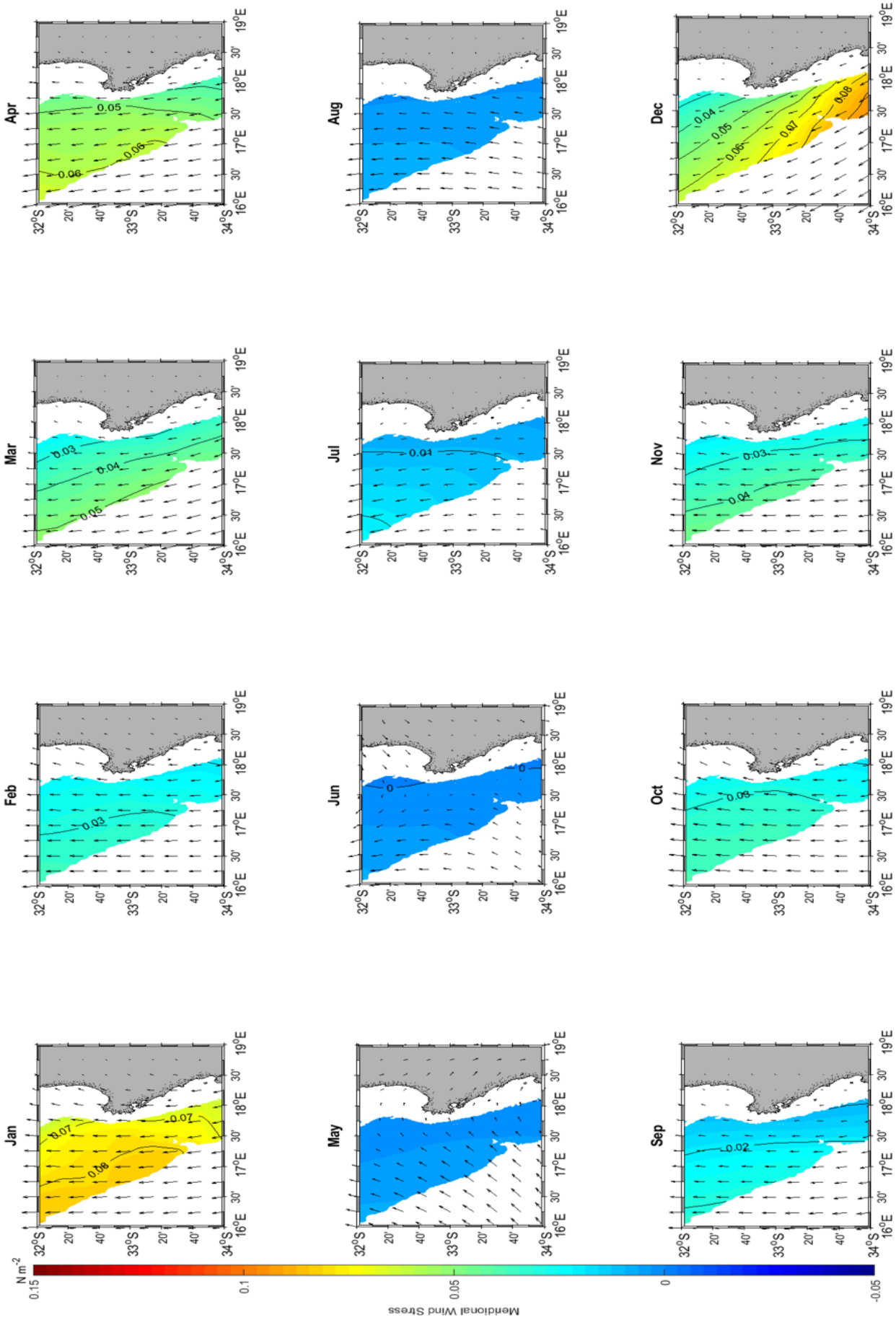

FIGURE 4.6: 2010 monthly mean meridional wind stress based in ERA5 wind data over the Southern Benguela region.

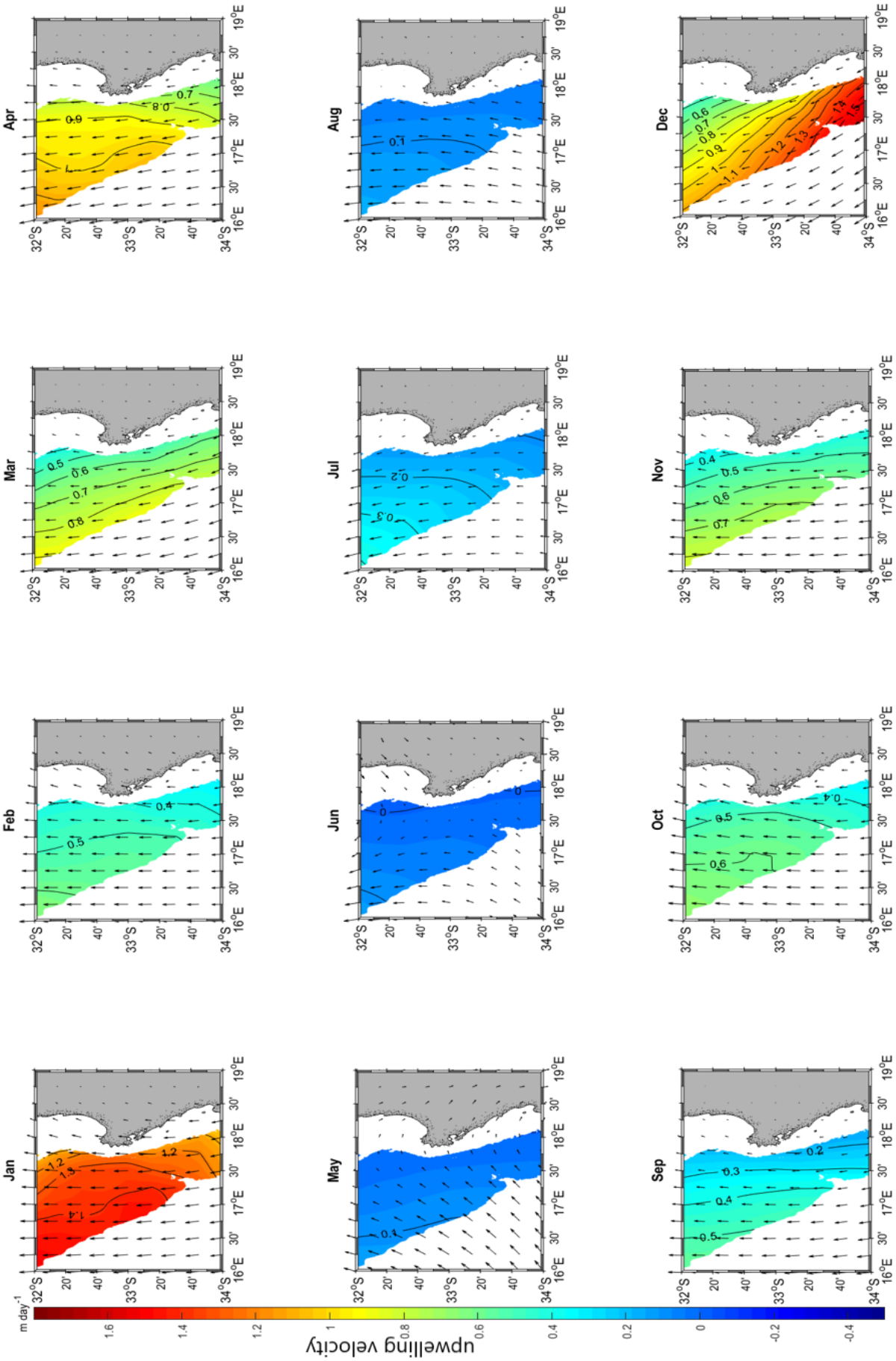

FIGURE 4.7: 2010 Monthly mean Ekman driven upwelling velocities in m/day based in ERA5 wind data estimated over the Southern Benguela region.

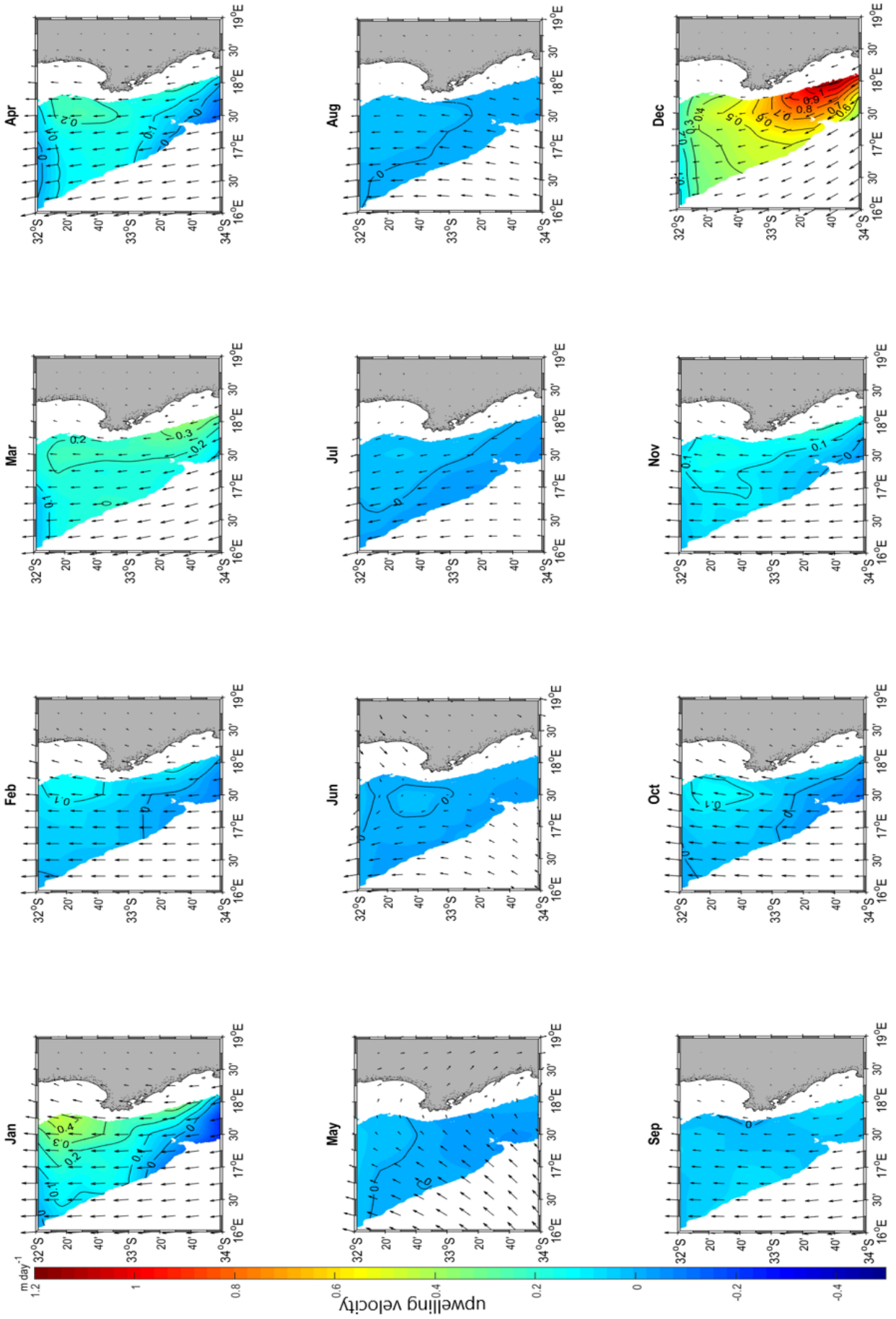

FIGURE 4.8: 2010 monthly mean Curl driven upwelling velocities based in ERA5 wind data over the Southern Benguela region.

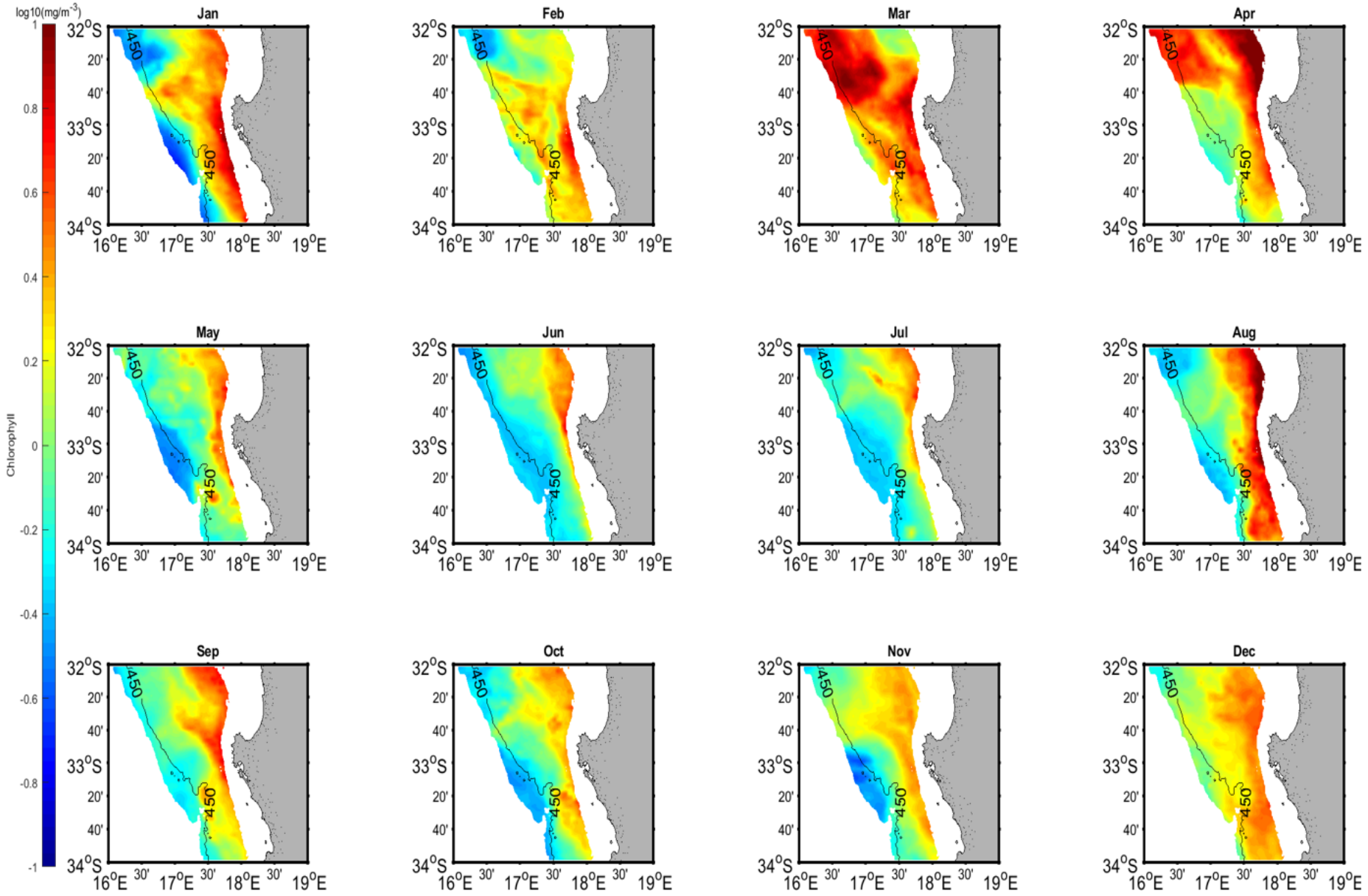

FIGURE 4.9: 2010 monthly mean chlorophyll $\log_{10}(mg/m^3)$ variation within the 150-750 m depth region in the studied area. The 450 m contour depicts the rapid (high hake catch volumes) fishing region

4.3 Time Series Analysis

The time-series patterns of Ekman upwelling, curl driven upwelling and chlorophyll concentrations computed as described in section 3.4 are comparatively displayed in Figure 4.10. The timelines in Figure 4.10 are established based on Equations 3.7, 3.11 & 3.12 for Ekman upwelling, curl driven upwelling and chlorophyll respectively. Chlorophyll peaks and troughs appears to be slightly delayed from the upwelling volume peaks and troughs. This is expected to be the case because the upwelling brings about nutrient rich water into the euphotic zone which supports primary production as already described in Chapter 2.

This delay is the lagged response of chlorophyll, the indicator of primary production to upwelling. The Ekman upwelling volumes are much bigger than the curl driven upwelling totals in each month in this region throughout the 2009-2018 decade. The Ekman upwelling depicts a clear seasonal trend while curl driven upwelling does not show a very strong seasonal trend even though summers still appear relatively more intense than winters. Chlorophyll appears to also show a seasonal pattern that has slightly delayed peaks compared to the Ekman upwelling as shown in Figure 4.10. The Ekman upwelling shows a distinctly more variable pattern when compared to the curl driven upwelling. Curl driven upwelling on the other hand, in addition to being low in upwelled volume, tends to fluctuate above and below the zero line (upwelling and downwelling).

The spatial patterns displayed in Figure 4.9 show a peak in chlorophyll in March. The March chlorophyll peak appear to be in response to the January peaks in Ekman upwelling estimated volumes and this relationship is visible every year as seen on the time series although the months of these peaks varies. It is also notable that some years have a double peak in chlorophyll such as 2013. Fish data is not shown in this timeline as it will be illustrated in Chapter 5.

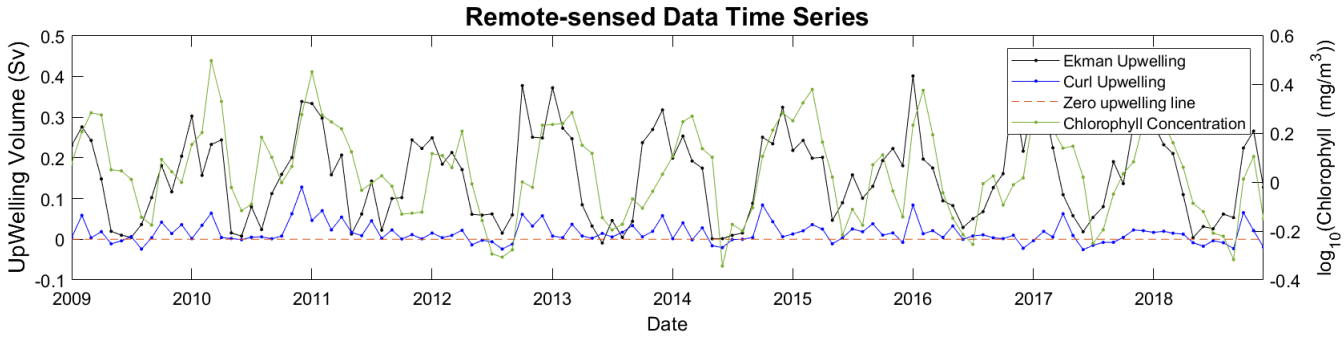

FIGURE 4.10: The Ekman driven upwelling, curl driven upwelling & chlorophyll time series between 2009 and 2018, a timeline that coinsides with the hake fishing data of Sea Harvest corporation.

The lagged correlation is therefore computed and presented in Figure 4.11 to determine the

average lag period between the upwelling and chlorophyll. The curl driven upwelling is in phase with Ekman upwelling as shown in Figure 4.11 (i). Therefore chlorophyll share the same mean lag of one month from the upwelling estimations as shown in Figure 4.11 (ii) & (iii).

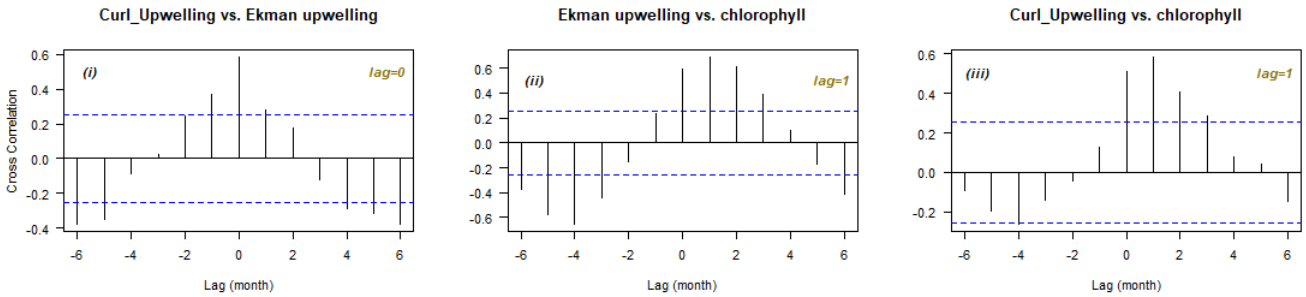

FIGURE 4.11: Cross-correlogram showing the lag between the physical oceanographic variables Ekman upwelling, curl driven upwelling and chlorophyll.

The lagged correlation, sometimes refereed to as "cross-correlation" is a Pearson correlation performed at a lag. A lag is computed by shifting the analysed trends about each other and superimposing data peaks and troughs of the two trends such that the alignment results in the highest correlation between the two data trends (upwelling & chlorophyll). The number of single step shifts that became necessary to achieve the highest correlation is called the lag. Since the data at hand is a monthly data, then a lag referrers to a single month. Figure 4.11 shows that on average chlorophyll peaks a month after the peak in upwelling. The blue dotted line denotes a 95% confidence interval. The same lagged correlation operation has been performed in hake data in Chapter 5 to determine the lagged correlation of hake from the upwelling and chlorophyll.

The blue line represents correlation significance where correlation values that falls within the blue lines are statistically insignificant. Correlation coefficients for the most significantly correlated lag are tabulated in Table 6.1 for all the above three plots and the hake CPUE plots shown in the next chapter.

5 Results: In-situ Hake Catch Data Analysis (Protected Chapter)

5.1 Monthly Average Catches

The hake database was established in December 2009 and during this first month of data collection, only 12 drags were recorded in this area. The observations (drags) were made in every each month between December 2009 until October 2014. There was no observations in November 2014 but there were 28 drags performed and recorded in this region in December that year. This is shown in Figure 5.1 where the data collection was continued until September 2018. The monthly data in 2010 had more than 30 samples each month which prop up the validity of each month as being a good enough estimation of a month's average. The catches are high CPUE during the winter months when compared to the summer months in 2010. In subsequent years this similar trend appears to continue even though it becomes less pronounced.

Figure 5.2 shows a rather disappointing picture when it comes to data collection as it is clear that the data collection was now becoming rare. In 2015 only five months managed to show more than 30 observations and four months have zero observations. In 2016 all but one month had observations. However, the number of drags is small (<30) in every month. 2017 and 2018 continued on the trajectory of sparsity as it can be seen. For this reason, the majority of the statistical analysis had to be performed on the initial 5 years since most time series analysis requires consistent data with no gaps.

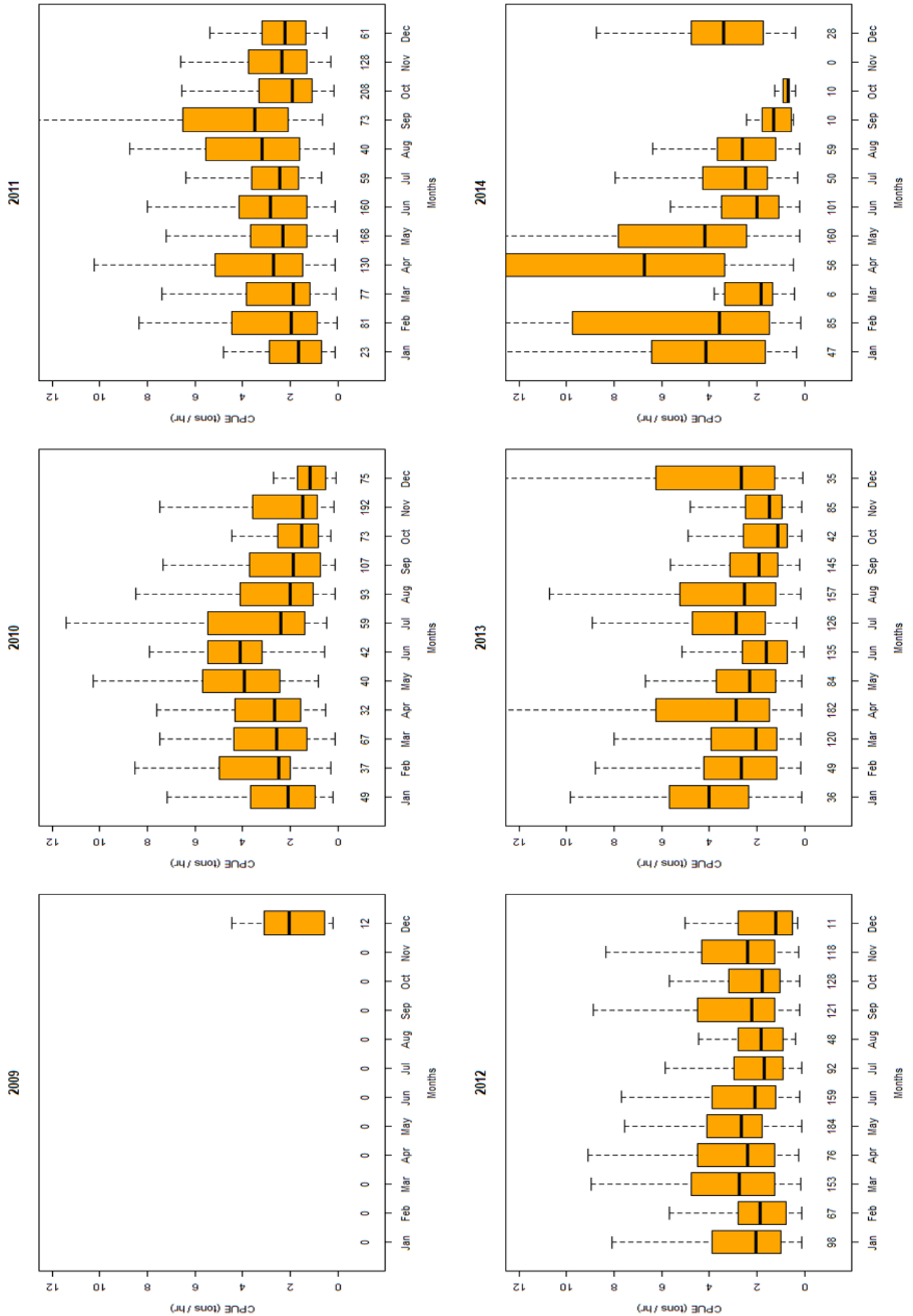

FIGURE 5.1: Annual monthly hake CPUE box and whisker plots 2009-2014

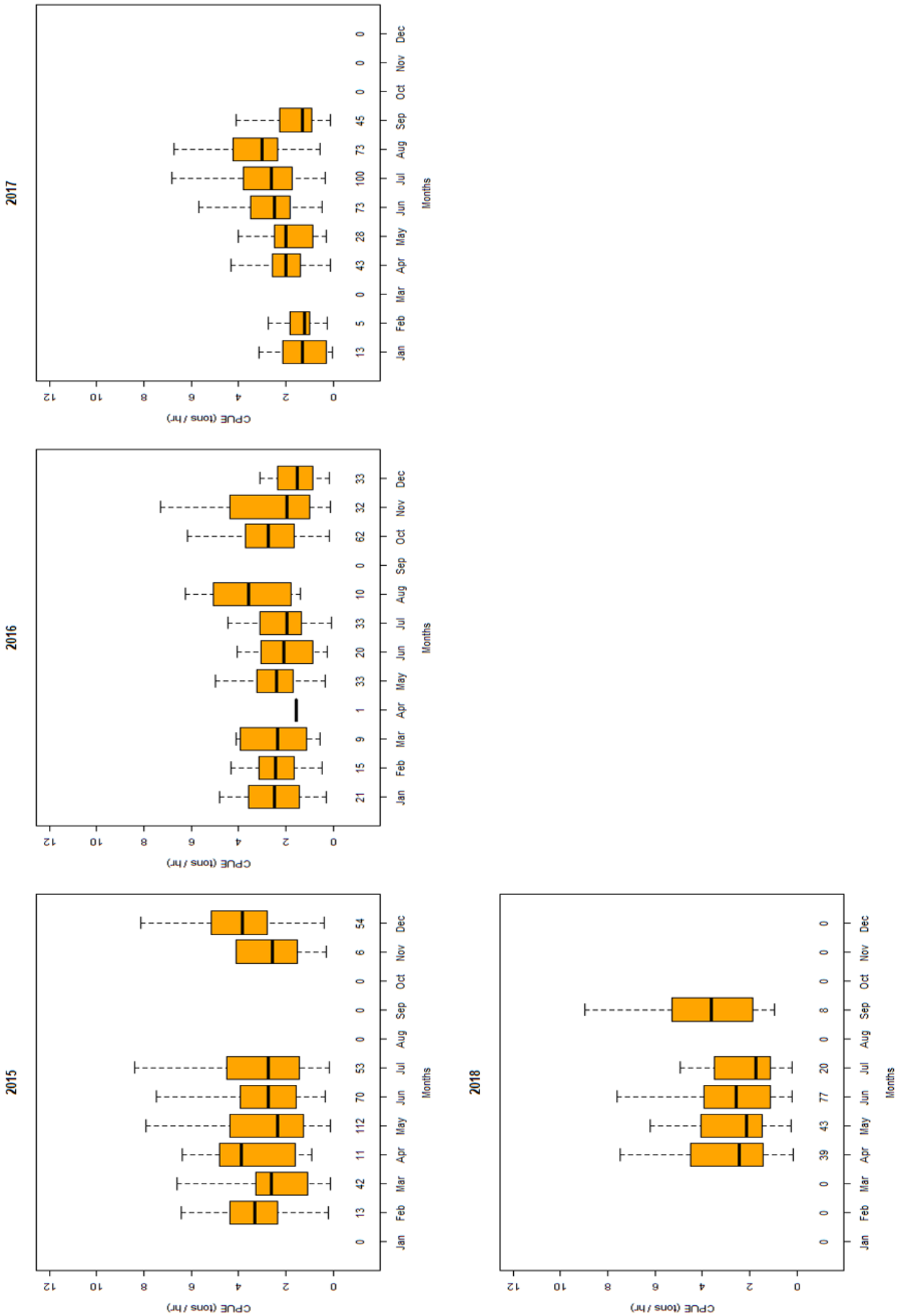

FIGURE 5.2: Annual monthly hake CPUE box and whisker plots 2015-2018

5.2 Upwelling, chlorophyll and CPUE monthly comparison

The pattern of observations between the wind-driven Ekman upwelling, wind stress curl driven upwelling, chlorophyll and hake CPUE is visualised in Figure 5.3. The upwelling volume is in Sverdrups (Sv), where "1 Sv = 1 million cubic metres of water per second" (Rykaczewski and Checkley, 2008) as displayed in Figure 5.3 panel A which has already been presented and discussed in Section 4.3. It is only reported here for comparison with the CPUE. The Ekman upwelling volumes are much bigger than the curl driven upwelling totals in each month in this region throughout the 2009-2018 decade. The CPUE time series shows a fragmented pattern due to a lack of data availability in later years of the decade. The pattern is not very distinct in terms of seasonality although it has a slight tendency that can be seen in some years. 2010 and 2014 stand out as they appear to have high catches (CPUE averages) compared to other years as shown in Figure 5.1 and 5.2. There is also an apparent lack of relationship between hake CPUE and chlorophyll as it appears in Figure 5.3 panel B, but this will be further examined in the next chapter.

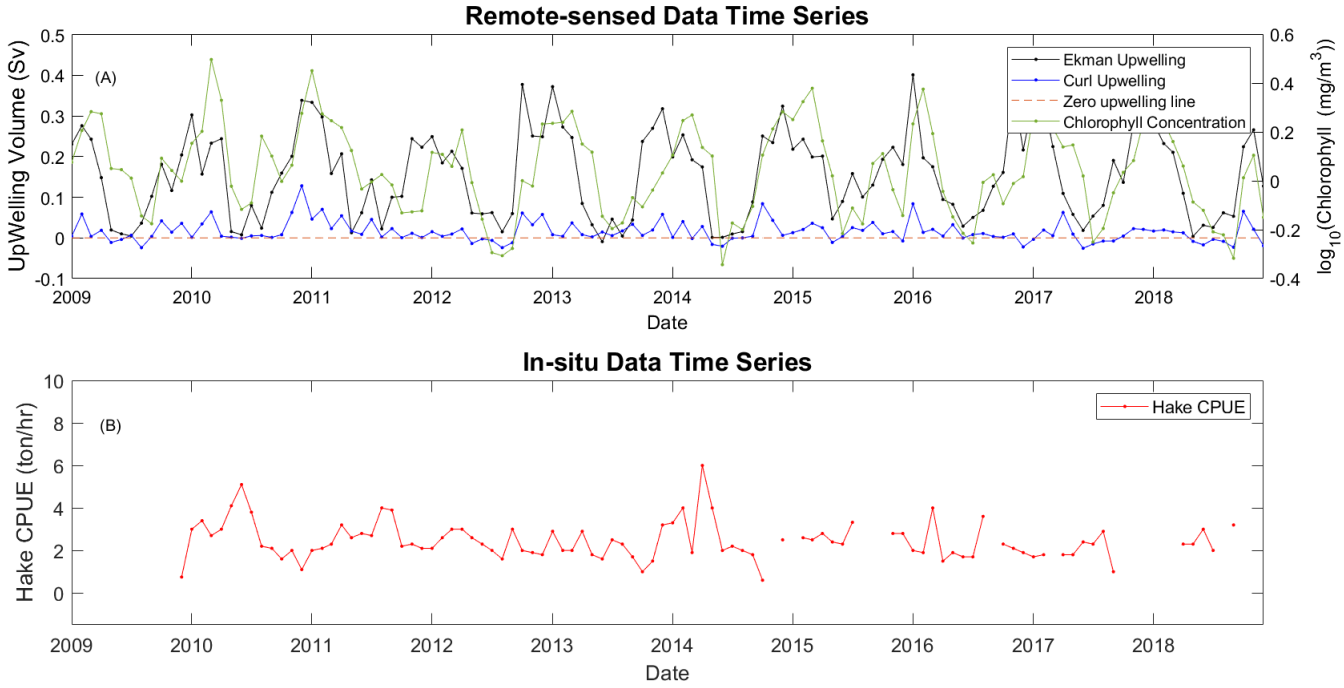

FIGURE 5.3: The Ekman driven upwelling, curl driven upwelling, chlorophyll and CPUE time series are correlated at different lags.

Figure 5.4 shows the different lags that exist between the different predictor variables and the predicted variable, hake CPUE. In Figure 5.4 (i), the correlation between the hake CPUE and Ekman upwelling is observed to be strongest and significant when the lag is 4 months. It is also appearing that the 2^{nd} month and 3^{rd} month have significantly correlated lags as well even

though they are not as strong as the 4^{th} month. The lag between hake CPUE and curl driven upwelling is 4 months and the hake CPUE to chlorophyll lag is 3 month. This means that the delay after an intense upwelling month for a hake CPUE increase is 4 months as also shown in Figure 5.4 (i) and (ii). Figure 5.4 is a continuation of the cross-correlation analysis started in Section 4.5 where the predictor variables were assessed independently before taking into account the response variable, Hake CPUE assessed in this chapter. Correlation coefficients for the most significantly correlated lag are tabulated in Table 6.1 for all the above three plots. The blue dotted line again denotes a 95% confidence interval

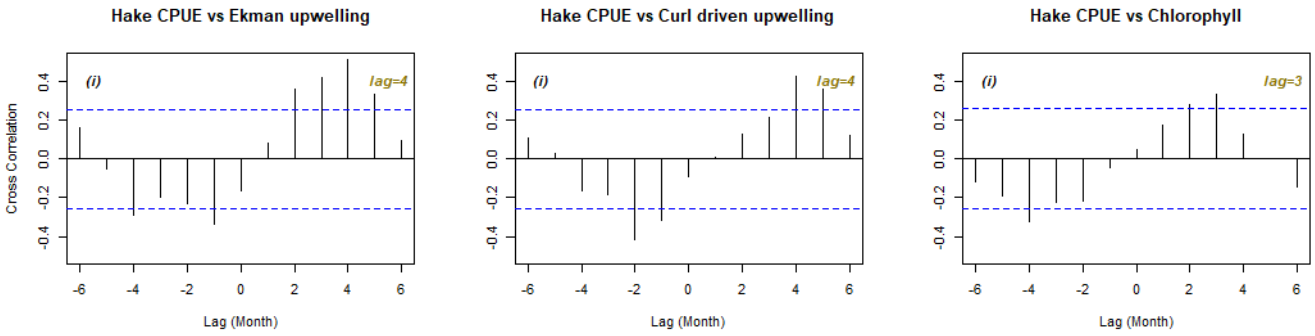

FIGURE 5.4: Cross-correlogram showing the lag between the predictor variables and the response variable (Hake CPUE).

5.3 Reconstructions of Fishing Events in Days of the Month

The Anomaly index (I_g) computed in Section 3.3.2 in the methods chapter has been tabulated and displayed in Table 5.1. This table displays magnitudes of the hake catches (ton/hour) that could be classified as big, medium and small in the form of high CPUE, mid-range CPUE and low CPUE respectively. Note that the hake CPUE data in the table below and the following sections are not monthly but are on a per drag level.

TABLE 5.1: The anomaly index of (year 2009-2018) shows hake CPUE in-terms of st andard deviations above and below the mean.

Anomaly index (I_g) Std-dev σ from the mean	**Hake CPUE in (tons/hour)**	**CPUE Level**
less than -1σ	≤ 0.80	Low CPUE
Between -1σ - 1σ	0.80 - 5.77	Mid-range CPUE
greater than 1σ	≥ 5.77	High CPUE

Note: Mean Hake CPUE = 2.15 tons/hour.

Table 5.2 shows daily upwelling averages and anomalies for the selected months of December 2010 and April 2011. These months are selected to illustrate what is detailed in section 5.3.1 and

5.3.2. The standardised anomalies for total upwelling are computed over the 2009-2018 decade. Total upwelling refers to the sum of Ekman upwelling and curl driven upwelling estimations. Anomalies that are below 0 represent upwelling velocities that are below average and are denoted as low. Anomalies that are between 0 and 1 represents upwelling velocities that are within a standard deviation above mean and are denoted as medium upwelling velocities. The high anomalies are showing daily upwelling velocities that are above a standard deviation from the mean of the said decade. The color coding used in the table is intended to emphasize the above upwelling velocity categorization and are intended to make eyeballing easier.

TABLE 5.2: Estimated daily upwelling velocities (m day^{-1}) during a mid-summer month (Dec 2010) and after a four month lag (Apr 2011). Anomaly computed over the decade 2009-2018 where μ=3.1 m day^{-1}, σ=4.4. m day^{-1}

Day	**December**	Anomaly		**April**	Anomaly	
1	2.5	-0.1	low	0.4	-0.6	low
2	3.0	0.0	low	2.4	-0.2	low
3	3.4	0.1	medium	3.9	0.2	medium
4	2.0	-0.2	low	9.2	1.4	high
5	10.0	1.6	high	3.9	0.2	medium
6	13.8	2.4	high	6.4	0.8	medium
7	6.5	0.8	medium	15.3	2.8	high
8	1.5	-0.4	low	3.2	0.0	low
9	1.9	-0.3	low	6.1	0.7	medium
10	11.2	1.8	high	2.6	-0.1	low
11	20.4	3.9	high	-0.1	-0.7	low
12	8.3	1.2	high	0.9	-0.5	low
13	12.1	2.1	high	5.7	0.6	medium
14	16.4	3.0	high	11.6	1.9	high
15	3.0	0.0	low	12.8	2.2	high
16	1.5	-0.4	low	13.0	2.3	high
17	0.5	-0.6	low	3.2	0.0	low
18	0.1	-0.7	low	5.5	0.6	medium
19	2.4	-0.2	low	2.4	-0.1	low
20	4.5	0.3	medium	1.1	-0.5	low
21	0.7	-0.5	low	2.8	-0.1	low
22	-1.4	-1.0	low	4.7	0.4	medium
23	0.8	-0.5	low	0.8	-0.5	low
24	3.7	0.1	medium	-0.3	-0.8	low
25	8.1	1.1	high	-0.4	-0.8	low
26	11.2	1.8	high	2.4	-0.2	low
27	16.8	3.1	high	2.8	-0.1	low
28	16.6	3.1	high	2.0	-0.2	low
29	16.1	2.9	high	4.7	0.4	medium
30	16.2	3.0	high	5.1	0.5	medium
31	10.0	1.6	high			

Note: Low $< \mu$, medium is between $\mu - 1\sigma$ and high $>1\sigma$.

5.3.1 Comparing an intense upwelling month with its catches

Daily upwelling velocities during a high upwelling month

The reconstructions of fishing days presented below in this section are performed with the intention to visualize the relationship between high intensity total upwelling and hake CPUE on a more refined spatio-temporal landscape, the days of the month. The days are shown in the mid-summer month of December 2010 where the upwelling velocities are observed to be the highest in that year as seen in figure 5.3. The reason of focusing on the upwelling month of December 2010 is that it coincides with the highest number of fishing observations (75 drags) performed in any December month (mid-summer) through out the data collection decade as shown in Figure 5.1 and the number of observations performed after the 4 month lag from this December month are also very high (130 drags). The selection of months with very high number of fishing observations (drags) makes the comparison more powerful. December 2010 and January 2011 are the only two consecutive summer months that have upwelling volumes that are greater than 1 Sv in the whole decade as shown in the time series figure, Figure 5.3. This shows that there was continuous high upwelling for an extended period of time in this summer better than in any other summer which also justifies using it for assessing hake response as a case study.

The strong upwelling days shown in intense red are very predominant in December 2010 as it can be seen in Figure. 5.5, 5.6 & 5.7 which displays the daily upwelling velocities of this month as a case study. The intense upwelling days are the 5^{th}, 6^{th}, 10^{th}, 11^{th}, 12^{th}, 13^{th}, 14^{th}, 25^{th}, 26^{th}, 27^{th}, 28^{th}, 29^{th}, 30^{th} & 31^{st} of December 2010 as they have mean rates above 8 m/-day as also seen in Table 5.2. The many other days in the yellowish sect of the spectrum (3-8 m/day) are moderate upwelling days while the blue end of the spectrum (<3 m/day) shows days that are having nearly zero upwelling velocities such as the 2^{nd}, 8^{th} & 22^{nd} of December 2010 as also clarified in Table 5.2. It is observable that a high upwelling month is made up of predominantly high upwelling days although it also has moderate and low upwelling days within itself. It is observable that the strongest upwelling incidences occur when the winds are perfectly southerly and high wind stress magnitudes are present as these are alongshore winds necessary for upwelling as explained in the earlier chapters. Northerly winds typically produce zero or less upwelling velocities. The strongest upwelling day with the highest upwelling velocities was the 11^{th} of December which showed a record of over 32 m/day, where the wind stress magnitudes appeared to be in the order of +0.2 Nm^{-2} in the southerly direction which is perfectly parallel to our coastline to create the stronger Ekman upwelling rates. The noted record of 32 m/day upwelling velocity is in the range of the values observed by Andrews and Hutchings (1980), which was the highest they recorded in the region in 1980.

Fishing activities during a high upwelling month

The spatial reconstructions of fishing activities (drags) are created for both the month of December 2010 and January 2011 to observe the impact of these intense upwelling months on fishing as discussed on the above section. The aim is to observe the daily fishing patterns during a mid-summer high upwelling before looking at the patterns after the four months lag which will be done in the next section. Figure 5.8 and 5.9 shows fishing activities that took place during an upwelling peak season. The red line in these figures indicates a drag that has high CPUE, the black lines shows medium-sized CPUE (Mid-range CPUE) and the yellow lines shows a low CPUE as outlined in Table 5.1. The grey lines that form the background of each figure are historical drags that show the fishing traffic that has occurred in this region between 2009-2018. The reference cities shown in red stars are for easy orientation within the region and are also locations of Sea Harvest docking ports.

The fishing activities are not very intense or consistent in December of 2010 during an intense upwelling month apart from the notable 4^{th} of December 2010 where 16 drags were achieved. All 75 fishing observations in this month took place within 12 days as shown in Figure 5.8. Some observations are not easy to spot as they were performed in the same location as another one. It is also important to note that there are no large catches (High CPUE drags) recorded in December 2010. Figure 5.1 is a complementary figure that shows the total number of observations (drags) achieved in each month.

On the second month of the intense upwelling period, January 2011, the fishing activities are still not intense at all as shown Figure 5.9. This is shown by the low occurrence of recorded fishing days (11) in the same figure and a low number of drags (23) as also shown in Figure 5.1. Figure 5.9 visualise the reconstructions of the fishing activities of the month of January 2011 and they emphasize what was observed in December 2010 in Figure 5.8, that the low CPUE drags abundance is indeed indicative of low catchability of fish through out this upwelling period. It is important to re-mention that other drags are not easy to visualise as they occur exactly over the same area the previous drag may have occurred. This would happen when the skipper was very happy about the catch he got in that area and he would assume there is still more fish stock available within the transect he took.

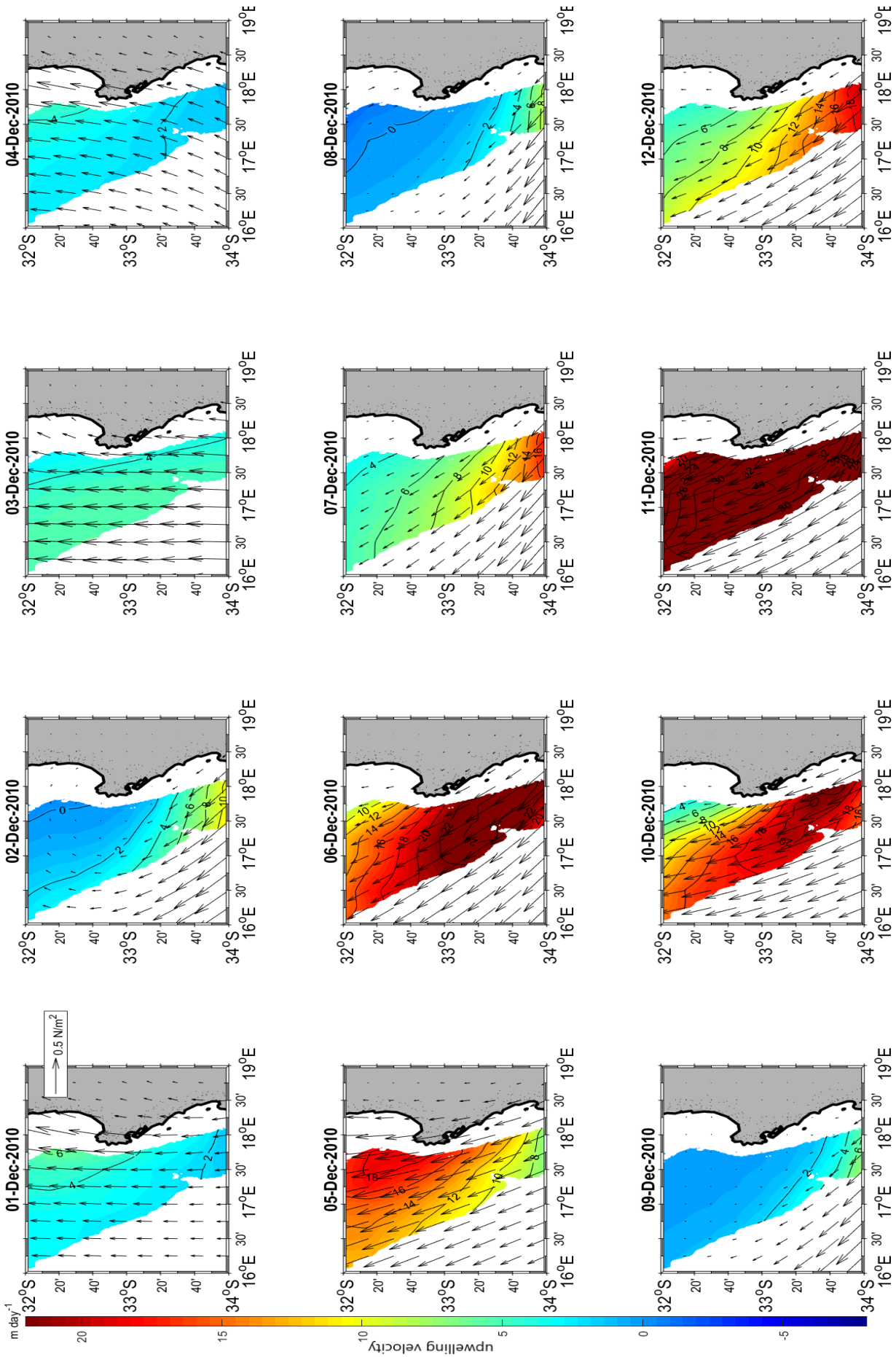

FIGURE 5.5: The daily upwelling velocities in the Southern Benguela region during a peak upwelling month: December 2010

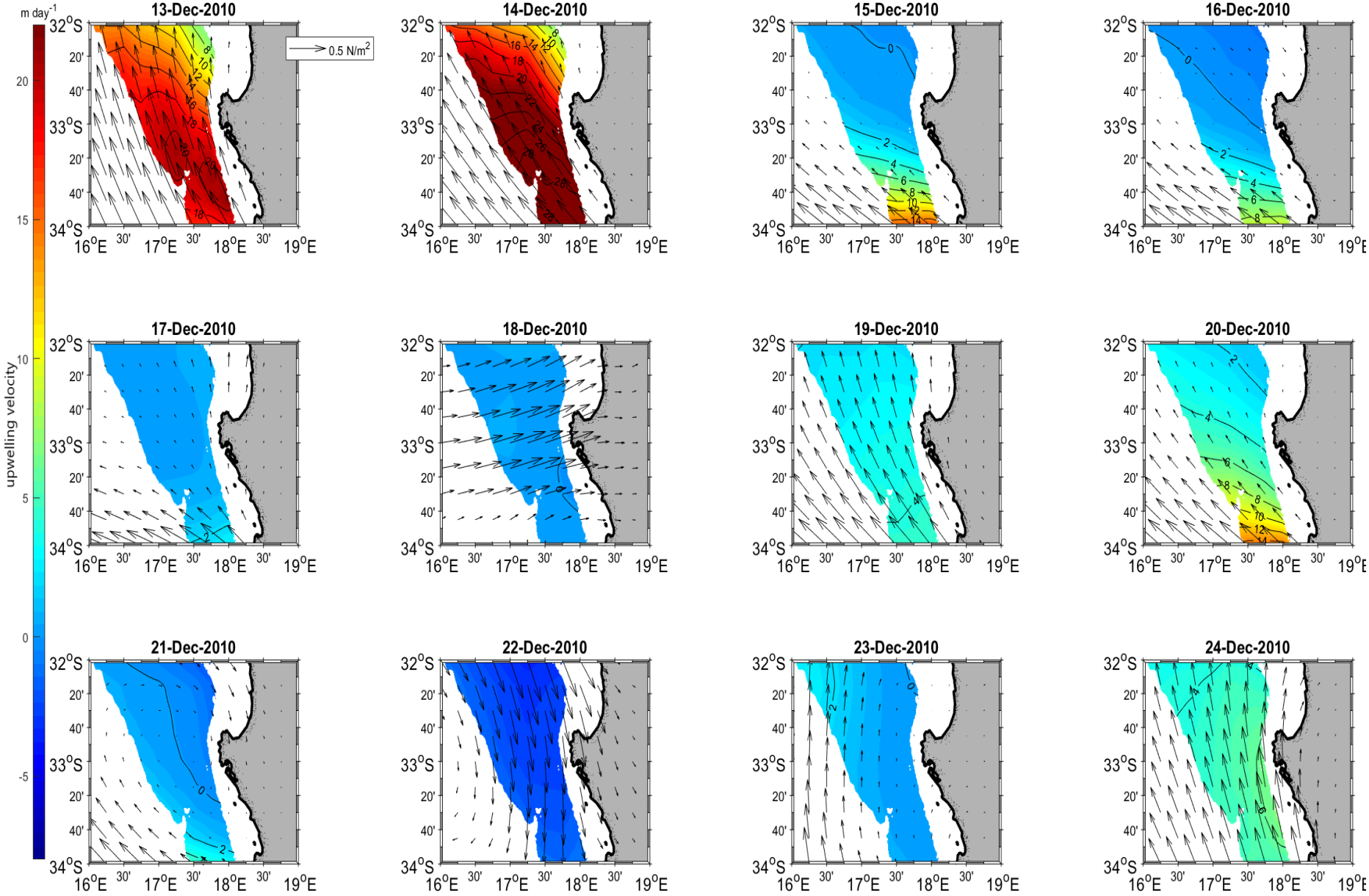

FIGURE 5.6: Continued from Figure 5.5

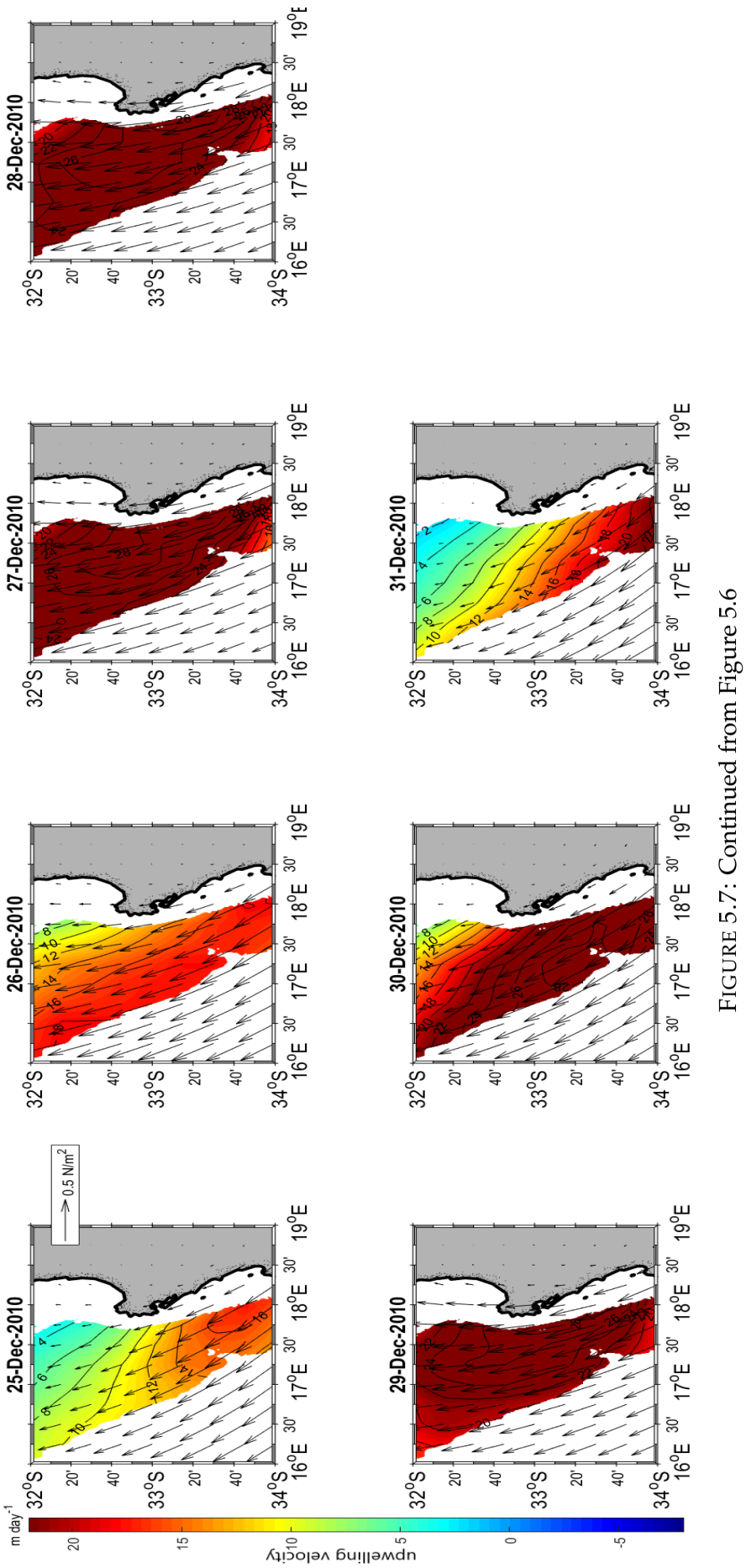

FIGURE 5.7: Continued from Figure 5.6

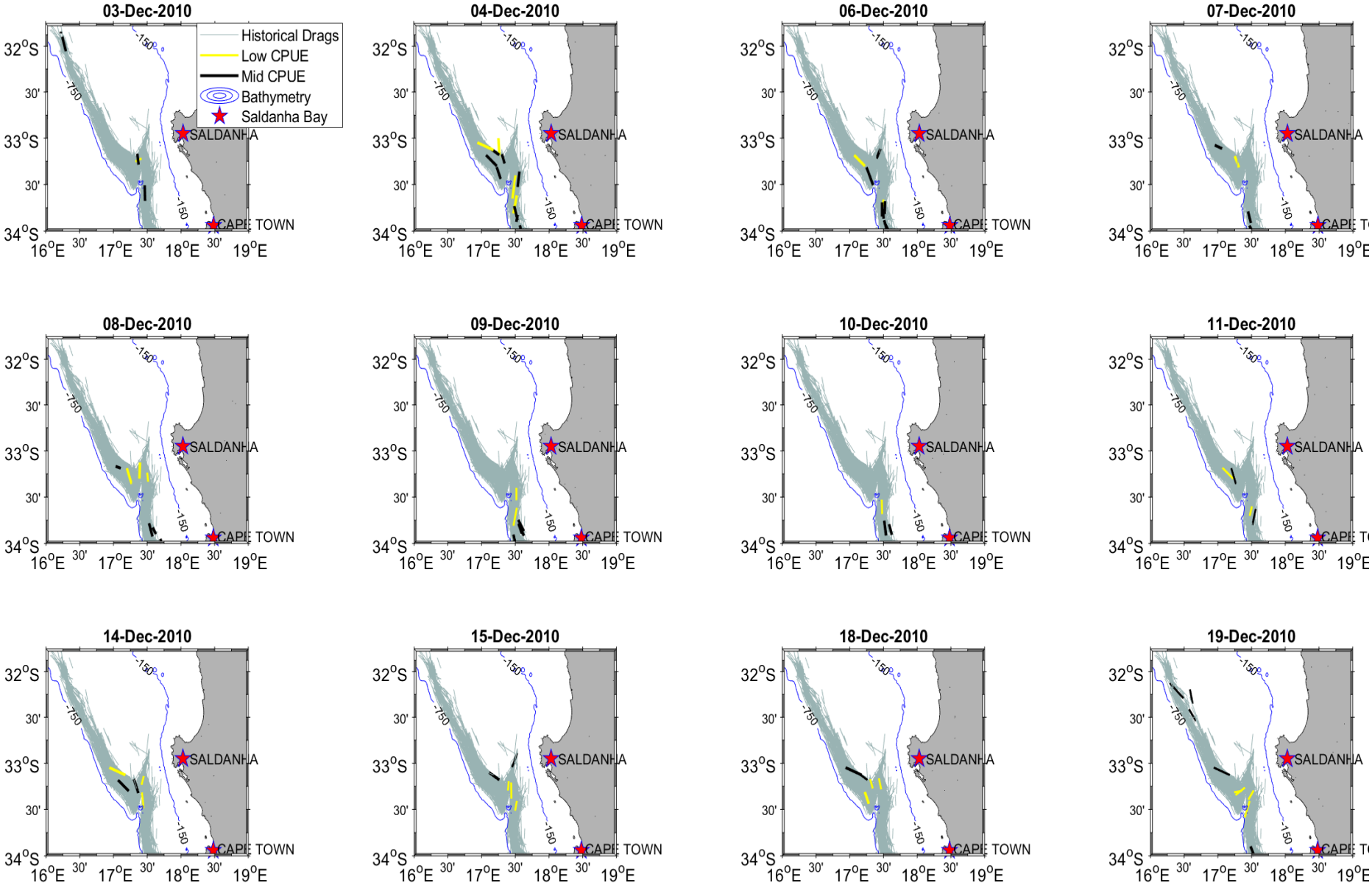

FIGURE 5.8: The reconstruction of fishing drags performed in each day during a peak upwelling month December 2010

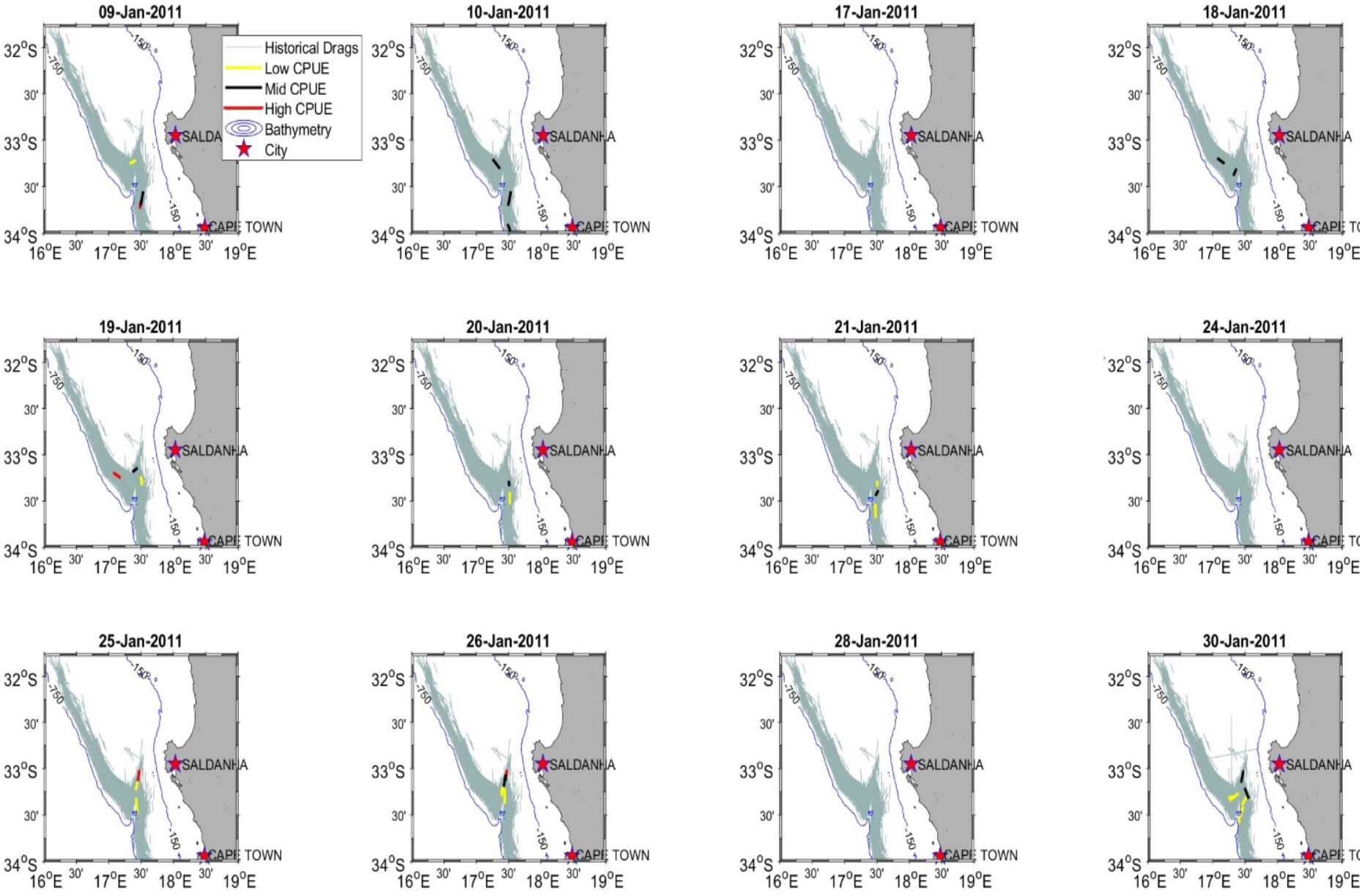

FIGURE 5.9: The reconstructions of fishing drags performed in each day in a month following a peak upwelling month

5.3.2 Comparing daily upwelling velocities with hake CPUE after a 4 month lag

Daily upwelling velocities after a 4 month lag

The strong upwelling days shown in intense red (upwelling velocities higher than 15 m/day) are much fewer in April than they were in December as it can be seen in Figure 5.10, 5.11 & 5.12 which displays the daily upwelling velocities for the month of April 2011. This is shown as a continuation of the case study presented in the previous section. The intense upwelling days are the 4^{th}, 7^{th}, 14^{th}, 15^{th} & 16^{th} of April 2011. The low upwelling days are the 1^{st}, 11^{th}, 24^{th}, & 25^{th} of April 2011. It is visible at this stage that the number of intense upwelling days in April were much fewer (5) compared to the 14 counted in December 2010. The intensities are also lower as it can be seen that the highest upwelling velocity in this month, observed on the 7^{th} of April had up to 24 m/day compared to the peak of 32m/day observed on the 11 th of December 2010. The daily wind directions and magnitudes vary continually but the occurrence of the best upwelling winds is less common in April compared to what happens in the mid-summer month as displayed in the December figure on the previous section.

Fishing activities after a 4 month lag

The fishing activities are very intense during the month of April 2011. This is four months after the recent peak upwelling month, December 2010. This is shown by the high occurrence of fishing days (24) shown in Figure 5.13 & 5.14 and a high number of drags (130) as also shown in Figure 5.1. These figures visualise the reconstructions of the fishing activities of the month of April 2011. The recorded drags appear to shows most fishing activity to be taking place in close proximity to each other within the region instead of being randomly spread through-out the whole study region. This is generally because vessels would communicate and assist each other in identifying productive areas. There is fewer low CPUE drags, a lot of mid-range CPUE drags (black lines) and quite a few of red lines representing high CPUE drags in the month of April even though the upwelling productivity of April is relatively low compared to December. This indicates the that upwelling in December and January may indeed have had the delayed influence on hake catch quantities four-months later as per the computed lagged correlation in Section 5.2.

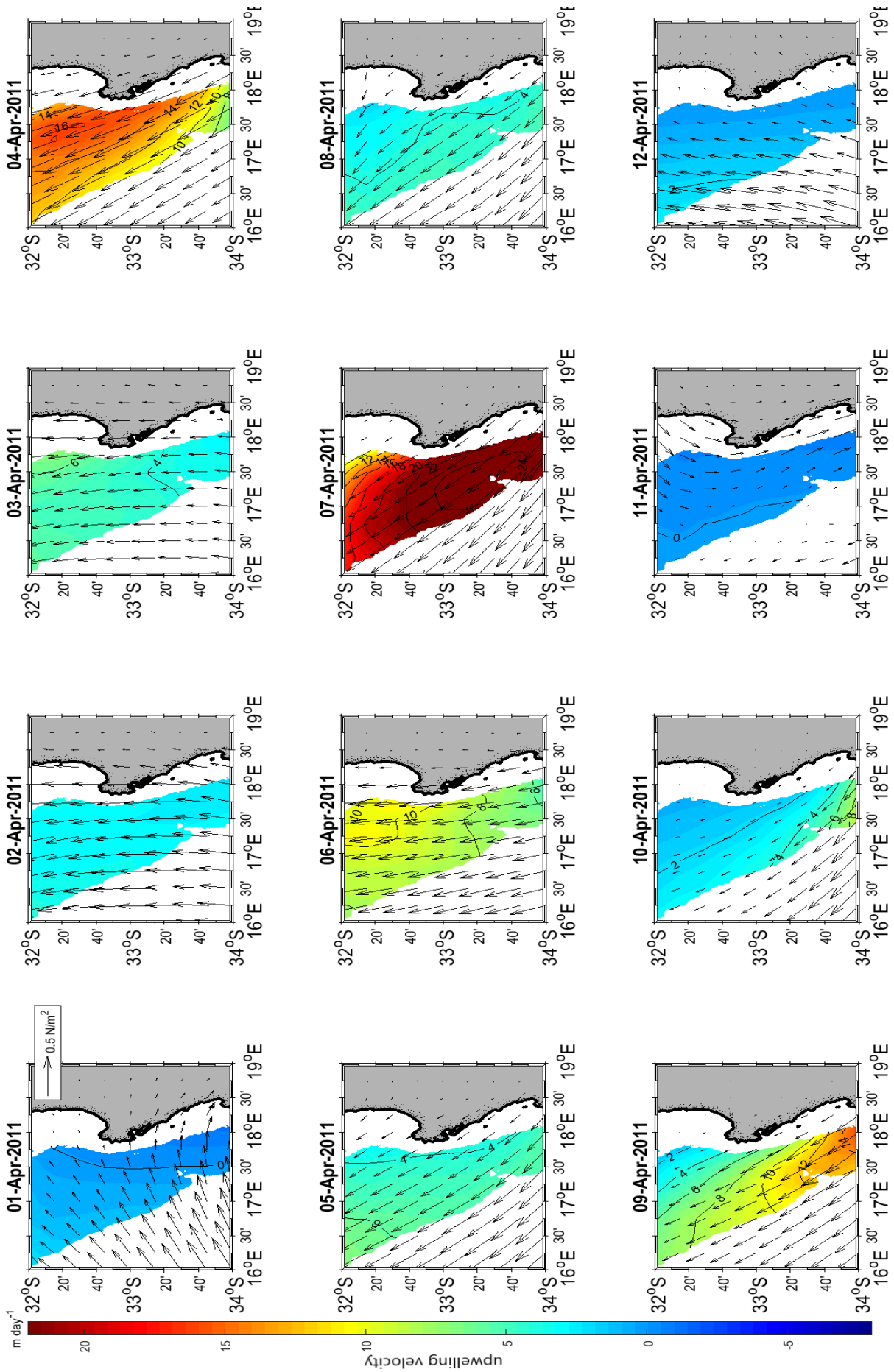

FIGURE 5.10: The daily upwelling velocities in the Southern Benguela region during a peak fishing month: April 2011

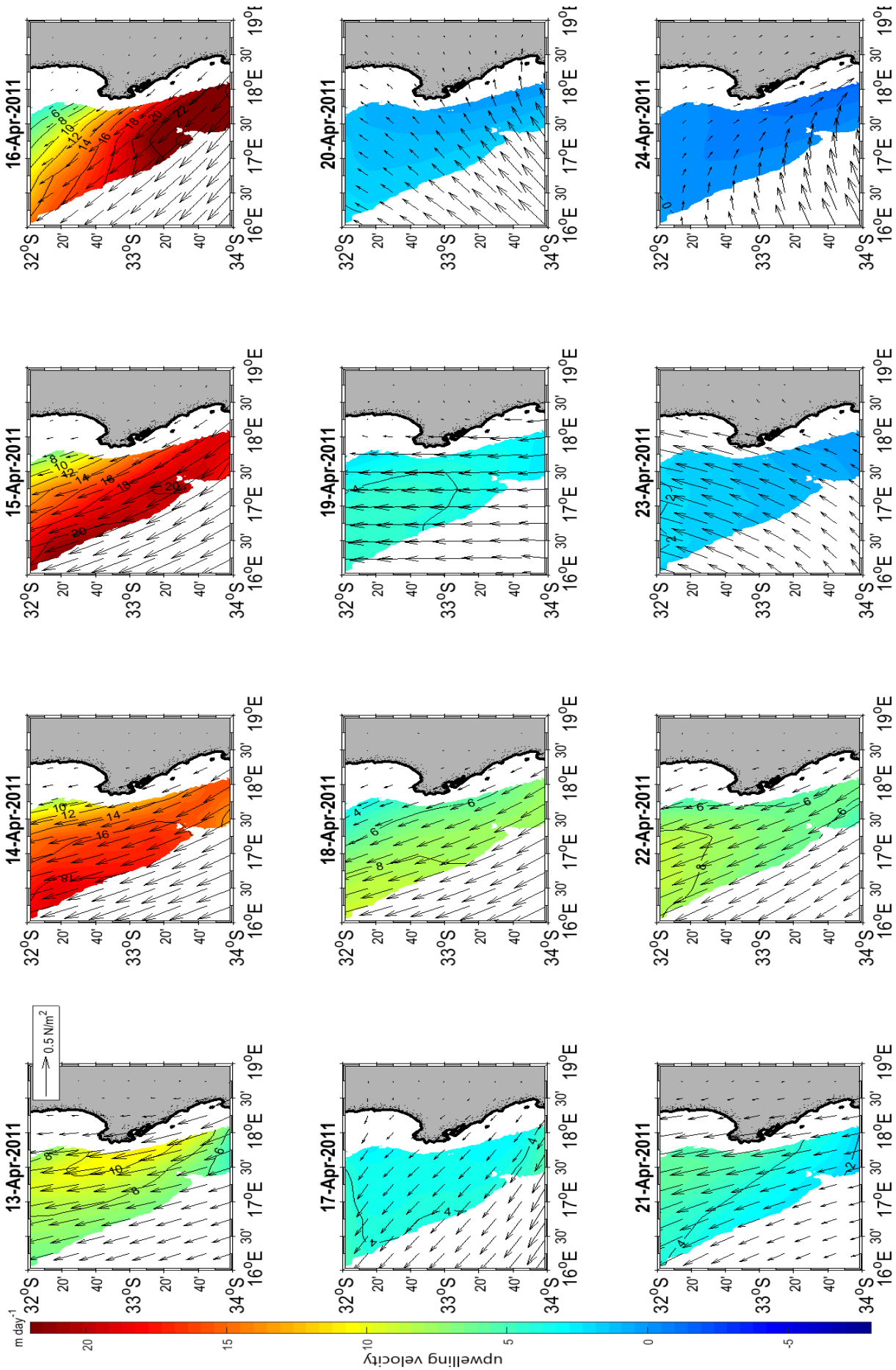

FIGURE 5.11: Continued from Figure 5.10

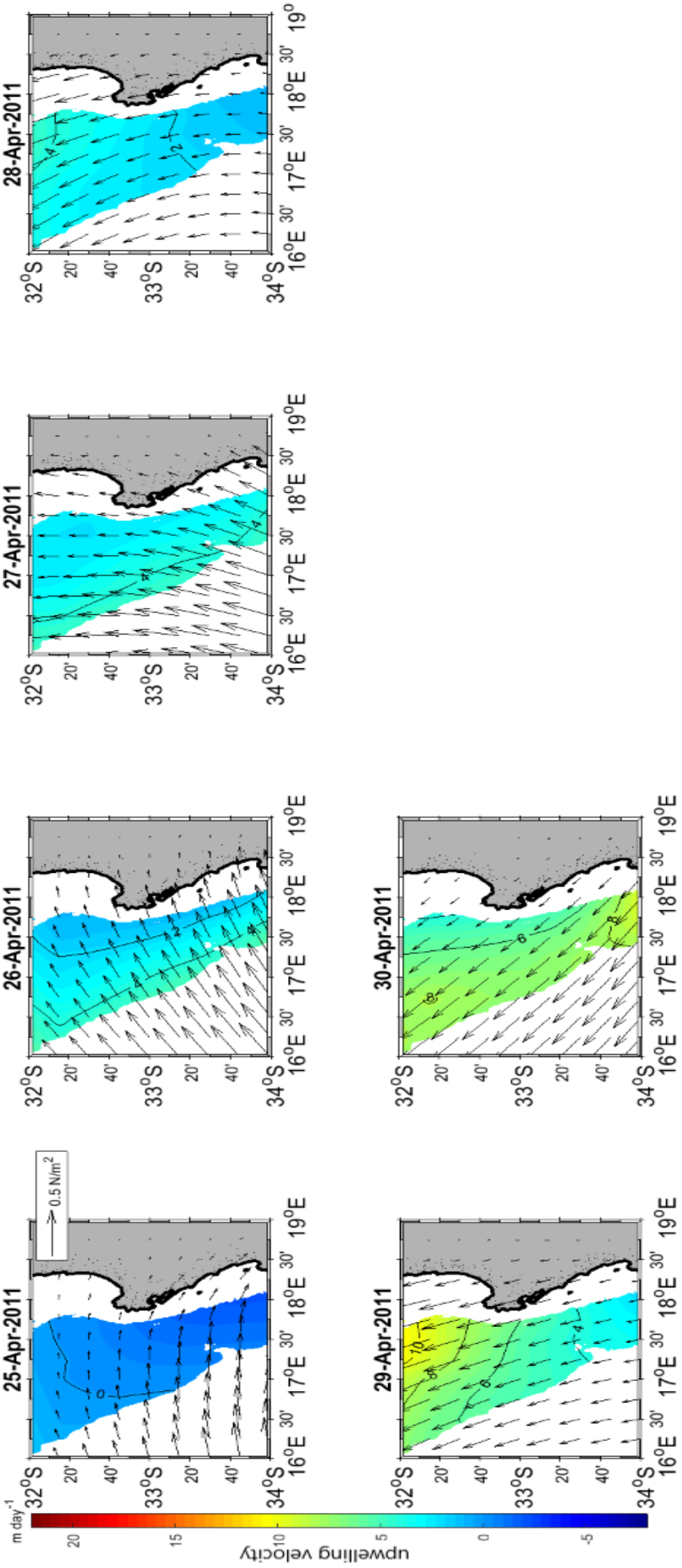

FIGURE 5.12: Continued from Figure 5.11

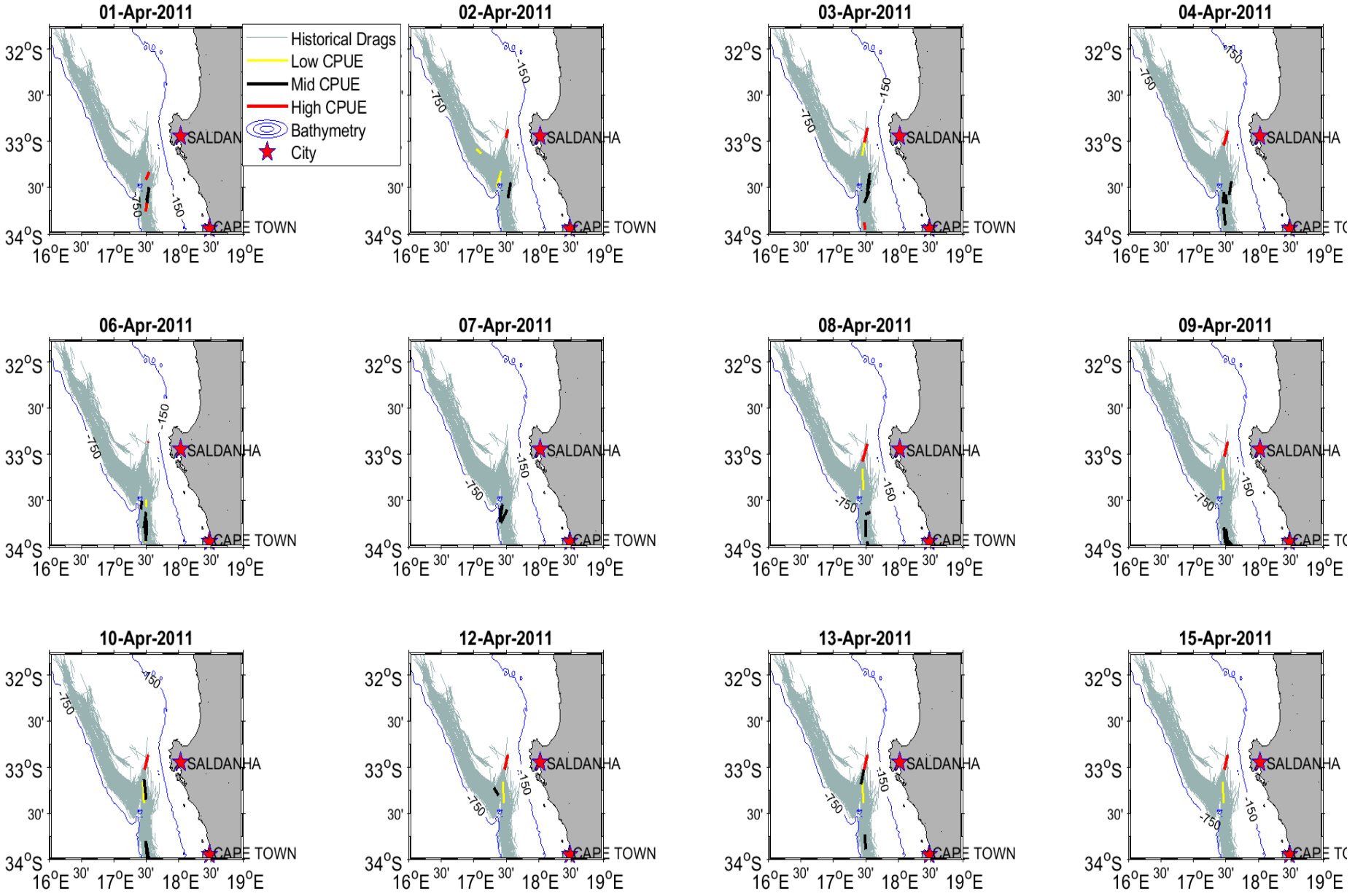

FIGURE 5.13: The reconstructions of fishing drags performed in each day during a peak fishing month following a peak upwelling period observed in figure [5.5]

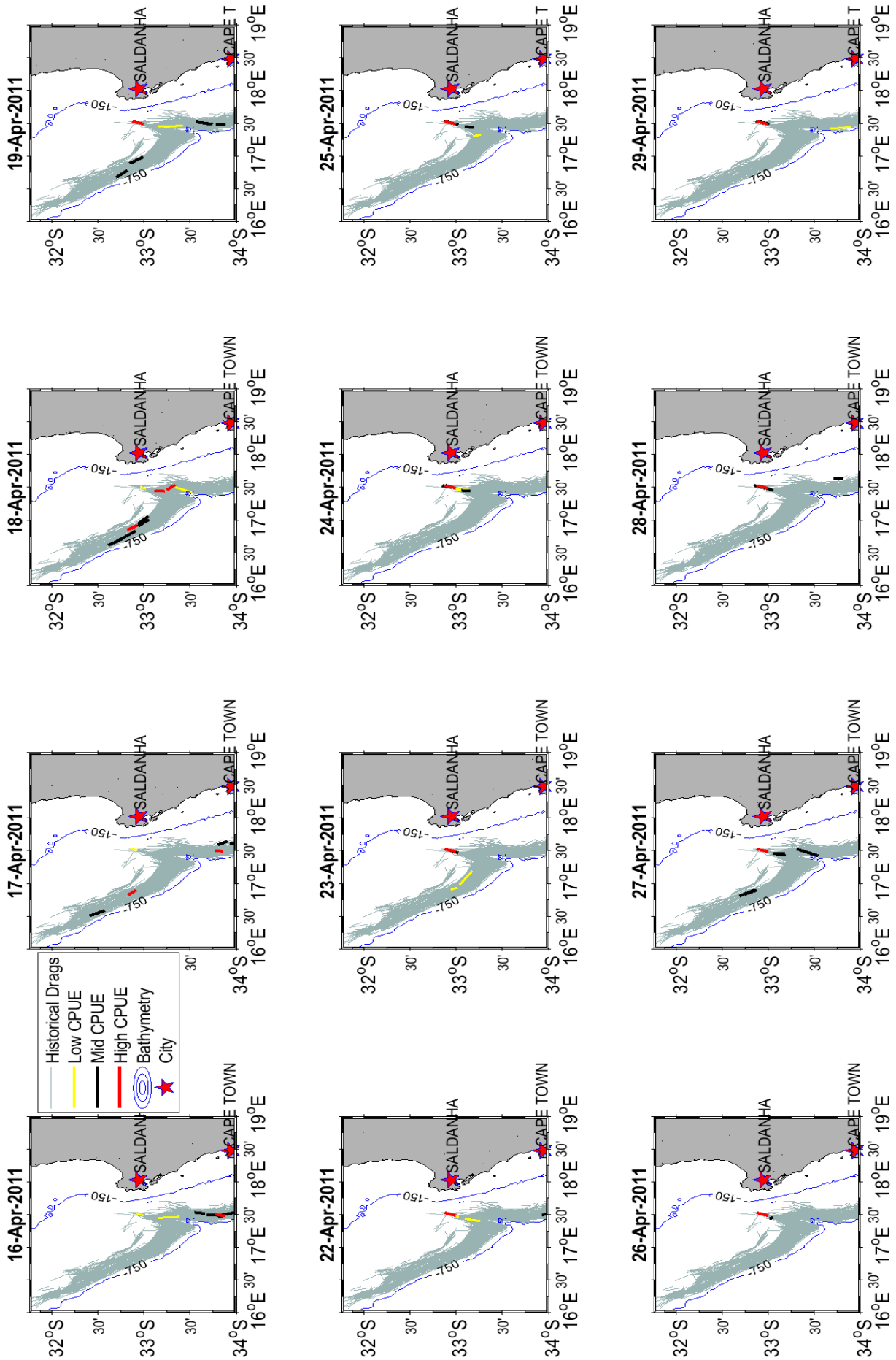

FIGURE 5.14: Continued from Figure 5.13

6 Results: Statistical analysis and regression modelling

6.1 Data Transformation

The hake data, Ekman upwelling, curl driven upwelling and chlorophyll data all produce leftward skewness when observed at a raw data level as depicted in Figure 6.1. The skewed data do not yield good results when analysed through parametric statistical methods (methods that depend on the normality of the data). The selection of the number of bins is on the basis of the square-root approach that is widely used as one of the selection methods. This is $\sqrt{n}$ = number of bins, where n is the number of observations.

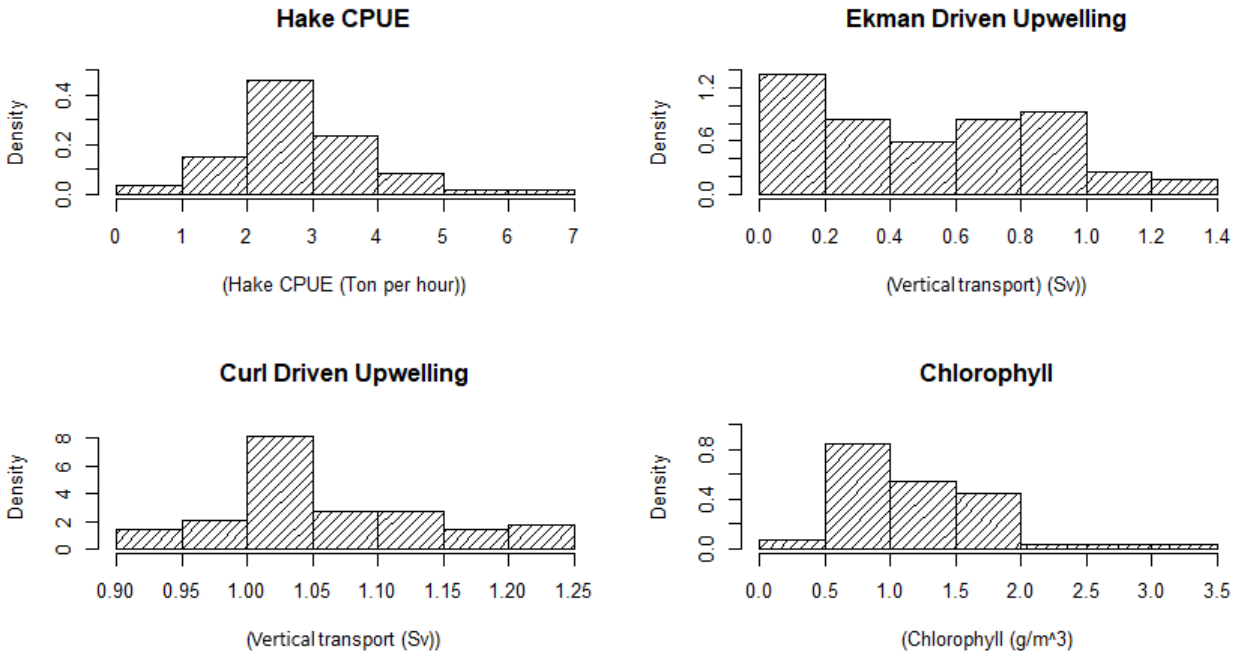

FIGURE 6.1: Raw data density distribution histograms of the hake CPUE in (ton/hour), Ekman upwelling volume (Sv), curl driven upwelling volume (Sv) and chlorophyll.

As shown in Figure 6.2, the hake CPUE only needed to be logarithmically transformed to become relatively normal. The blue line fitted over the curve is a normality line. The normality

transformation is also successfully performed on the Ekman driven upwelling data. Curl driven upwelling data is also log-transformed and the Chlorophyll data is taken through log base 10 because oceanographic reasons which namely are to make the data visually intuitive as it is otherwise generally too skewed for eyeballing and log base 10 is a widely accepted transformation for chlorophyll data.

The normality of the Ekman upwelling and curl driven upwelling is not very much resembling normal distribution as they somewhat look right skewed and left skewed respectively in Figure 6.2. The Shapiro Wilk-test is therefore employed to further statistically assess if the distributions of these two data is significantly different from a normal distribution. The Ekman upwelling Shapiro Wilk-test output (W = 0.97606, p-value = 0.3388) implies that there is no basis to assume that the distribution of the Ekman Driven upwelling data is significantly different from normal distribution. Similerly for the curl driven upwelling, Shapiro Wilk-test (W = 0.93567, p-value = 0.0561) reveals that there is no basis to assume that the data distribution is significantly different from the normal distribution. These conclusions are reached based on the p-value (Pr>0.05) remaining insignificant, therefore providing no grounds for rejecting the null hypothesis.

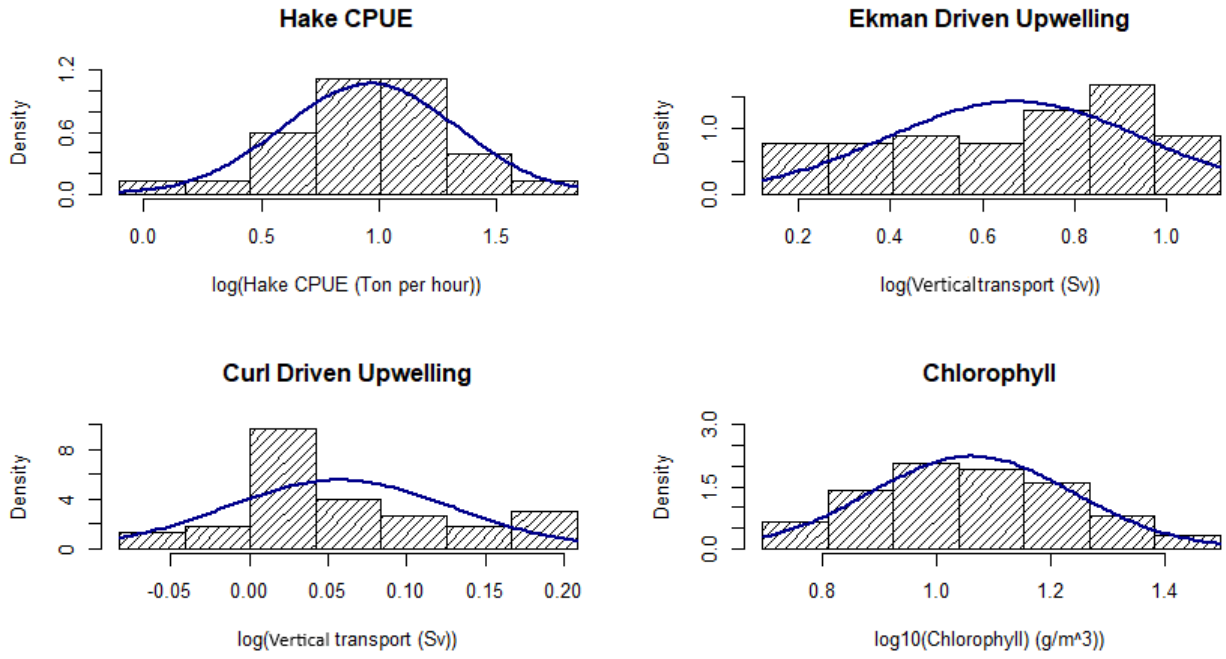

FIGURE 6.2: Normalised data density distribution histograms of the hake CPUE in (ton/hour), Ekman upwelling volume (Sv), curl driven upwelling volume (Sv) and chlorophyll.

The kurtosis and skewness tests have both been avoided as they are known to heavily depend on the size of the sample. They are heavily in favour very large samples as opposed to small and

modest sided samples. The normality test is thus performed solely based on the Shapiro-wilk test and eye inspection of the data histograms as shown above.

6.2 Correlation Analysis

This section is a detailed expansion of the correlation analysis shown in Chapter 4.3 and Chapter 5.2. This is done to show how the data transformations, significance of the correlation and the correlation itself is established as it is only stated and shown in the cross-correlogram figures, Figure 4.11 & 5.4 which are generated through the R software in the previous chapters. This section performs the lagged Pearson correlation manually in order to show the underlying principles that brought up the information contained in the cross-correlogram figures mentioned above. The case example used in this detailed hypothesis testing approach is the hake CPUE and Ekman upwelling as it is shown below in Section 6.2.1 and 6.2.2. All other parameters in addition to the hake CPUE vs Ekman upwelling mentioned above are summarized and tabulated in Table 6.1 as they are arrived at through the same process and this corresponds with what is shown in the cross-correlogram mentioned above.

6.2.1 Pearson Correlation

There appears to no relationship at all between hake CPUE and Ekman upwelling when the relationship is assessed on each month between the two variables when they are not transformed as shown in the left hand side plot in Figure 6.3. Once the data is taken through the transformations outlined in Figure 6.2, there appears to be a negative and really weak relationship between hake CPUE and Ekman upwelling as shown on the right hand side in Figure 6.3. There is still no relationship even after both variables are transformed as shown in the third pane. This shows that what ever transformations you apply will not create a significant correlation between these variables as they are not significantly correlated on the same month.

Hypotheses testing:

H_0: $\rho = 0$ (There's no significant relationship between Ekman upwelling and CPUE)
H_1: $\rho \neq 0$ (There is a significant relationship between Ekman upwelling and hake CPUE)
Significance level = $\alpha = 0.05$
The results show that there is no significant relationship between Ekman upwelling and hake CPUE (r = -0.19, t = 3.522, df = 57, p = 0.145), therefore, we would not be justified to perform a regression analysis. The Pearson's test produced a p-value = 0.145 which is greater than 0.05, therefore, we fail to reject the null hypothesis which states that the is no significant relationship between Ekman upwelling and hake CPUE.

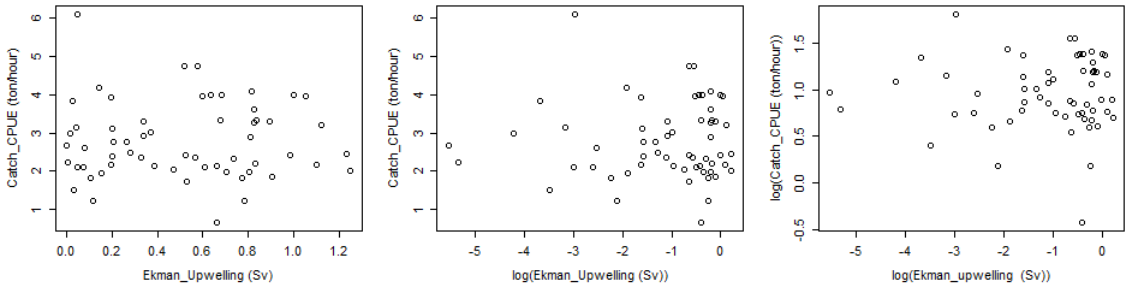

FIGURE 6.3: Scatter plots on raw data demonstrates poor and somewhat negative relationship between wind-driven Ekman upwelling and Hake CPUE.

6.2.2 Pearson Lagged Correlation (cross-correlation)

Initially, it is apparent that even after a the necessary lag has been considered on the data, there appears to be no significant relationship between hake CPUE and Ekman upwelling when the data is not transformed as shown in the left hand side plot in Figure 6.4. Once the data is taken through the transformations displayed in Figure 6.2, there appears to be a positive relationship between hake CPUE and Ekman upwelling as shown on the right hand side panel of Figure 6.4. The middle panel plot in this figure serves to show that the transformation is necessary on both axis as the scaling of the data otherwise it becomes skewed to the right and it shows that further log transformation is necessary on hake CPUE as well as shown on right hand side panel of Figure 6.4.

Hypotheses testing:

H_0: $\rho = 0$ (There's no significant relationship between Ekman upwelling and hake CPUE)
H_1: $\rho \neq 0$ (There is a significant relationship between Ekman upwelling and hake CPUE)
Significance level = $\alpha = 0.05$
The Pearson's test produced a p-value < 0.001 which is less than 0.05, therefore, we reject the null hypothesis which states that the is no significant relationship between Ekman upwelling and hake CPUE. The results show that there is a significant relationship between Ekman upwelling and hake CPUE (r = 0.48, t = 4.172, df = 57, p <0.001), therefore, we would be perfectly justified to perform a regression analysis on these two variables.

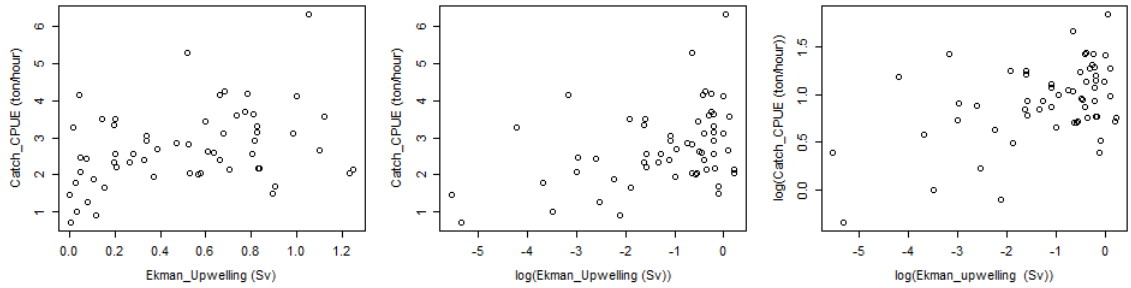

FIGURE 6.4: Pearson correlation on a four month lag data demonstrates a positive relationship between wind-driven Ekman upwelling and Hake CPUE. The positive relationship is also significant, therefore, justifies a regression analysis (r = 0.48, t = 4.172, df = 57, p <0.001)

For a linear regression model between Ekman upwelling and hake CPUE, the data passes the model validation as shown in Figure 6.5. Plot (**a**) shows the residuals and predicted values, the distribution of which forms no particular data pattern and there is randomness on the distribution of the residuals above and below the zero line. The spread of the residuals is similar above and below the line indicating that the variances are similar. Plot (**b**) shows a pattern of most data falling along the quartile-quartile line which suggests that the data is somewhat normally distributed as can also be seen in Plot (**c**) on the histogram.

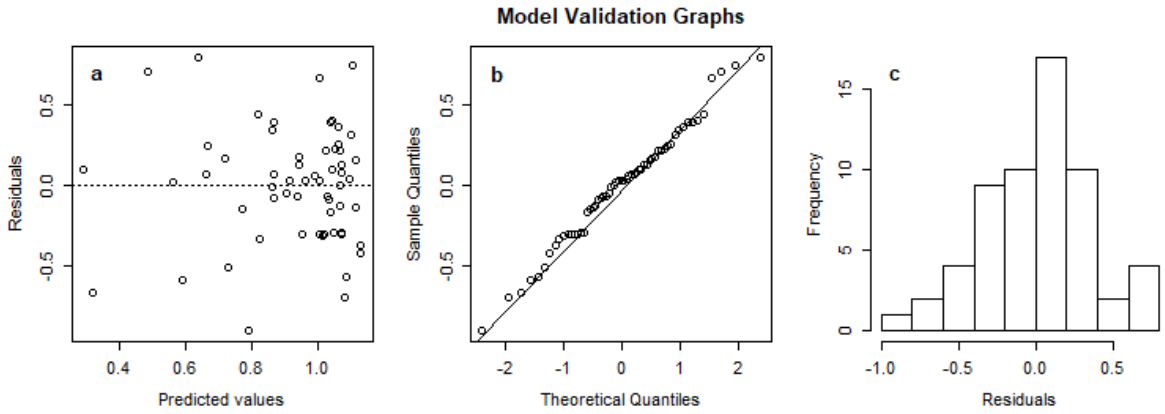

FIGURE 6.5: Model validation for transformed Ekman driven upwelling and Hake CPUE. Plot (**a**) on the left shows shows randomly distributed residuals which indicates equal variance, the Q-Q plot in Plot (**b**) shows the normal distribution of the residuals as well as Plot (**c**) the histogram show more data centred around mean.

All the other correlation analyses has been performed through the above process and summarised into the table below, Table 6.1.

TABLE 6.1: Results of all correlation tests performed against Hake CPUE. Ekman upwelling, Curl driven upwelling and Chlorophyll are correlated with Hake CPUE during a high upwelling and after a lag for comparison.

Correlated variable	**Pearson Cor.(r)**	**Significance** $\alpha = 0.05$	**Cross-Cor.(r)**	**Significance** $\alpha = 0.05$
Ekman - Curl upwelling	**0.59**	**<0.001**		
Ekman upwelling - Chlorophyll			**0.62**	**<0.001**
Curl upwelling - Chlorophyll			**0.58**	**<0.001**
Ekman upwelling - Hake CPUE			**0.48**	**<0.001**
Curl upwelling - Hake CPUE			**0.42**	**<0.001**
Chlorophyll - Hake CPUE			**0.37**	**0.004**

Note: significantly correlated variables shown in bold.

6.3 Multiple Regression Modelling

6.3.1 Model hypotheses

The regression model is constructed based on the above significant correlations. The various hypotheses listed below represent a possible way in which the catch can be explained by the studied predictor variables.

1. Model 1: No relationship can be modelled based on the predictor variables. The variables are not sufficient to model an underlying relationship.
2. Model 2: The hake CPUE is best described linearly by all three predictor variables, Ekman upwelling, curl upwelling and chlorophyll such that an in increase in any one of them results in the increase in hake CPUE.
3. Model 3: The hake CPUE is best linearly described by the strongest correlated predictor variable, the Ekman upwelling. The other variables are not impacting the hake CPUE in a way that enables modelling.
4. Model 4: An interaction of two of the predictor variables has the greatest linear impact on the hake CPUE. The Ekman upwelling and curl driven upwelling interaction results in the best hake CPUE.

One hypothesis of the above four will be adopted at the end of the analysis and the other three will be dismissed.

6.3.2 Model assessment(s)

The R^2 and adjusted R^2 of the four models are displayed in Table 6.2. The MSE's are also shown for comparison of the models. The outputs show how each model describes the predicted variable (hake CPUE). A high MSE means that there high variance of the residuals from the model which means that the model is bad. Low MSE is desired for the best model possible.

TABLE 6.2: Comparison of four hake CPUE predictive models by R^2, adjusted R^2 and residual mean square (MSE).

Models	R^2	Adjusted R^2	MSE
m1	0	0	0.16
m2	0.28	0.24	0.12
m3	0.23	0.22	0.12
m4	0.33	0.29	0.11

A full description of the model output tables, Table 6.3 - 6.8 is given on the statistical methods section in Chapter 3.**Df** refers to the degrees of freedom.

Model 1:

Model 1 (m1) is our null hypothesis model and it is rejected forthright as it is clearly the least likely model as shown in Table 6.2 where R^2 is zero and its MSE is the highest. This leaves us with the 3 alternative hypothesis models m2, m3 m4. which are explored below in detail.

Model 2:

The outputs displayed in Table 6.3 are the results of model 2 (m2). The only significant predictor variable is the Ekman upwelling at an α=0.05. Curl upwelling appears to be significant at an α=0.1, a significance level we are not interested in conducting this study on as it is less powerful than a significance level of α=0.05. The intercept is also very significant.

The analysis of variance (ANOVA) compares all predictor variables at once to the responses variable log(Hake CPUE) as displayed in Table 6.4. The Ekman upwelling is still the most significant even on this approach as shown with the Pr(>F) <0.05. The F value is also the highest in this predictor variable 17.75 $\sim F_{1,57}$. This means that taking noise into account through the F test, Ekman upwelling is the only significant contributor to the prediction of the response term.

TABLE 6.3: summary of results of a linear regression model m2 with all the factors taken into consideration.

Coefficients:	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**Significance**
Intercept	0.97898	0.09272	10.558	<0.0001	***
log(Ekman_Upwelling)	0.10311	0.04445	2.32	0.0241	*
log(Curl_Upwelling)	1.29621	0.76306	1.699	0.0950	.
$\log_{10}(Chlorophyll)$	0.06841	0.29855	0.229	0.8196	

Note: The log transform is taken to normalise data, significant terms shown in (*)

TABLE 6.4: Analysis of variance (ANOVA) results of a linear regression model m2 with all the factors taken into consideration.

Response: log(Hake_CPUE)	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
log(Ekman_Upwelling)	1	2.2395	2.23951	17.7586	<0.0001
log(Curl_Upwelling)	1	0.3911	0.39114	3.1017	0.0838
$\log_{10}(Chlorophyll)$	1	0.0066	0.00662	0.0525	0.8196
Residuals	55	6.9359	0.1		

Model 3:

The outputs displayed in Table 6.5 are the results of model 3 (m3). The only variable used in it is the Ekman upwelling and it is a significant predictor at α=0.05 as shown by its p-value and the p-value of its intercept.
The analysis of variance compares all predictor variables at once to the responses variable log(Hake CPUE) as shown in Table 6.6. The Ekman upwelling is the only variable available in this case However, because ANOVA uses the F-test, for testing the impact of noise, ANOVA was still employed as shown. The Ekman upwelling is significant Pr(>F) <0.05 even in this approach as expected with a high F statistic of 17.64 $\sim F_{3,55}$. A high F-value indicates a model that best fits the data.

TABLE 6.5: summary of results of a linear regression model m3 with the strongest correlated factor only taken into consideration.

Coefficients:	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**Significance**
(Intercept)	1.09963	0.06234	17.639	< 0.0001	***
log(Ekman_Upwelling)	0.14609	0.03502	4.172	0.0001	***

TABLE 6.6: ANOVA results of a linear regression model m3 with the strongest correlated factor only taken into consideration.

Response: log(Hake_CPUE)	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
log(Ekman_Upwelling)	1	2.2395	2.23951	17.406	0.0001
Residuals	57	7.3337	0.12866		

Model 4:

The summary output displayed in Table 6.7 are the results of model 4 (m4). The predictor variables employed are Ekman upwelling, curl driven upwelling and their interaction. It appears that curl driven upwelling on its own is not a significant contributing predictor variable in predicting the response variable at α=0.05 as shown by its p-value. The Ekman upwelling and the interaction of Ekman upwelling and curl driven upwelling are however significantly contributing as indicated by the stared p-values that are below α=0.05.
The analysis of variance compares all predictor variables at once to the responses variable log(Hake CPUE) as shown in Table 6.8. The Ekman upwelling is the strongest contributing variable followed by the interaction of Ekman and curl upwelling. As ANOVA uses the F-test, for testing the impact of noise, the Ekman upwelling showed the highest significant F-value of 19.13 $\sim F_{3,55}$. A high F-value indicates a model that best fits the data.

TABLE 6.7: summary of results of a linear regression model m4 with an interaction of the most significant predictor variables taken into consideration.

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	Significance
(Intercept)	1.00325	0.08732	11.49	<0.0001	***
log(Ekman)	0.10247	0.03973	2.579	0.0126	*
log(Curl)	0.66091	0.79235	0.834	0.4078	
log(Ekman) x log(Curl)	-1.10685	0.53286	-2.077	0.0425	*

TABLE 6.8: ANOVA results of a linear regression model m4 with an interaction of the most significant predictor variables taken into consideration.

Response: log(Hake_CPUE)	Df	Sum Sq	Mean Sq	F value	Pr(>F)
log(Ekman)	1	2.2395	2.23951	19.1336	<0.0001
log(Curl)	1	0.3911	0.39114	3.3418	0.07297
log(Ekman) x log(Curl)	1	0.505	0.50503	4.3148	0.04247
Residuals	55	6.4375	0.11705		

6.3.3 Model selection & validation

Model selection

Based on the AIC selection approach introduced in Chapter 3.7, Table 6.9 displays that the best model of the bunch is **m4**. This is the best possible model of those tested, however it only describes 29% of the variation but for assessing the linear relationship it is better than all the other three considered models. This is a model that has eliminated some parameter (chlorophyll) which was adding complexity without significantly improving the model such as m1. A model with the smallest AIC value in the set is the best model as can be observed in the table.
I will now investigate whether **m3**, the second-best model seen in Table 6.9 doesn't differ by much from **m4** or it is really second-best by a long distance. For this assessment, ΔAIC becomes important as it has a quantity that is proportional to the likelihood.

$$L(g_i|x) \propto exp(-\frac{1}{2}\Delta_i) \tag{6.1}$$

The likelihood $L(g_i|x)$ shown above represents a relative strength of the case for each model. While the $exp(-\frac{1}{2}\Delta_i)$ is the quantity derived from ΔAIC values. They have been scaled to sum up to one in order to make them easily interpret-able. This is shown in Table 6.9 so that we can see the proportion of the individual contribution of each model through the expression known as Akaike weights w_i.

$$w_i = \frac{exp(-\frac{1}{2}\Delta_i)}{\sum_{r=1}^{R} exp(-\frac{1}{2}\Delta_r)} \tag{6.2}$$

w_i is the measure of how likely is each model in percentage proportions. The table shows that **m4** is by far the best model at 79% likelihood compared to the 12% associated with **m3**.

TABLE 6.9: Model selection table for the hake CPUE prediction.

Models	K	AIC	ΔAIC	w_i
m1	2	64.14	17.41	0.00
m2	5	51.13	4.40	0.09
m3	5	50.42	3.69	0.12
m4	5	46.73	0.00	0.79

To assess the odds of the data being better described by **m4** over **m3**, a comparison or the two w_i's of the models shows: $\frac{w_4}{w_3}$ = 6.6 which simply means that **m4** is nearly 7 times more likely than **m3**, its closest competitor.

Model validation

The model validation for **m4** linear regression is displaying valid model outcomes as shown in Figure 6.6. In Plot (**a**) the spread of the residuals is similar above and below the line indicating that the variances are similar. Plot (**b**) shows a pattern of most data falling along the quartile-quartile line which suggests that the data is somewhat normally distributed as can also be seen in Plot (**c**) on the histogram where most data is centred around the mean. Therefore, a regression model based on parametric methods is right to fit the transformed data.

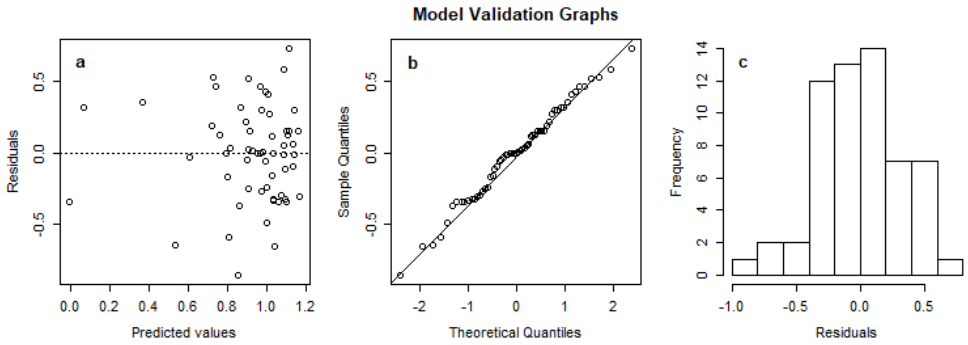

FIGURE 6.6: Model validation for log-transformed Ekman driven upwelling and Hake CPUE linear regression shows a good model validation output. Plot (**a**) shows randomly distributed residuals which indicates equal variance, the Q-Q plot in plot (**b**) shows the normal distribution of the residuals as well as plot (**c**) the histogram show more data centred around mean.

The diagnostic plot in Figure (6.7) below depicts residuals plotted in four different ways which assist in robustly assessing the model fitness.

- **Residuals plotted against fitted** values on the top left plot indicates that the model is linear since the fitted line shown in red is a straight line.
- The **Q-Q plot** of the top right depicts normality as most of the standardised residuals fall within the quantile-quantile line. This justifies the use of parametric methods which are powerful in detecting relationships.
- The bottom left plot depicts **scale-location** which is the standardised residuals vs fitted values. This test homogeneity of variance of the residuals (homoscedasticity). This does not appear to be perfectly the case in our model, so there is an apparent heteroscedasticity problem. The deviation from the homoscedasticity does not appear very strong, therefore, can be neglected since there is a random spread of residuals around the horizontal line.
- **Residuals vs leverage** is the plot on the bottom right. This plot depicts the spread of points and demonstrates the most influential points. The most influential point is a point that is above the cook's distance. In our case it appears that there is non that lies beyond cooks distance, therefore, there is no point (outlier) that can be removed in order to improve the model.

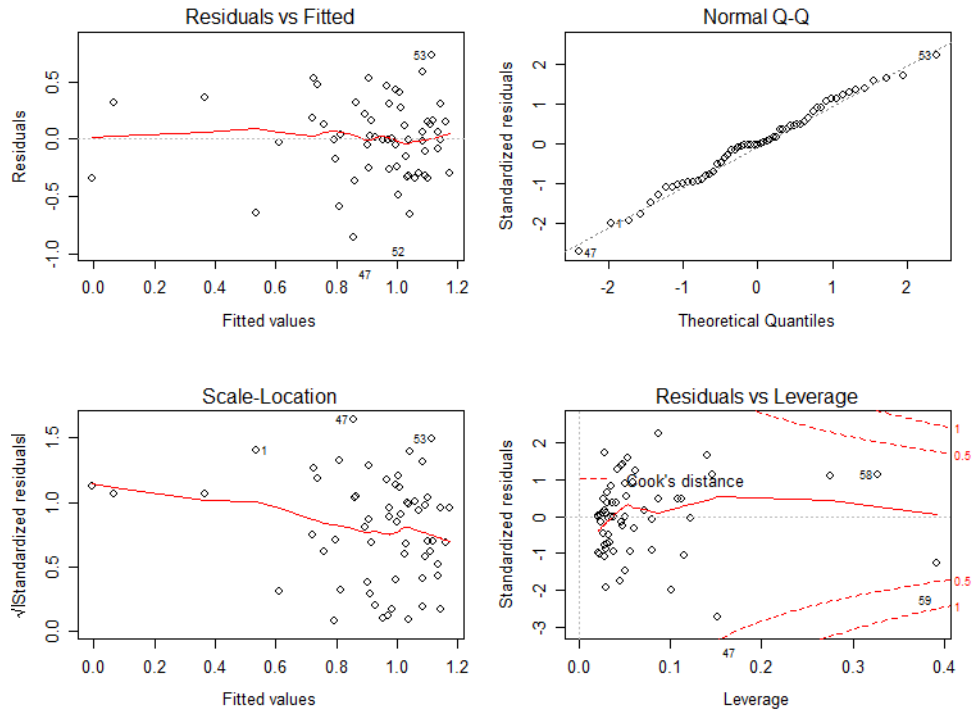

FIGURE 6.7: Model diagnostics plots fitted by residuals plotted in four different ways.

The regression model is therefore developed on the basis of the lagged correlation as it has been shown that lagged correlation on the transformed data yields a positive and significant correlation which allows for regression modelling. The regression model that best describes the pattern between hake CPUE and upwelling in the Southern Benguela region is therefore given as:

$$log(Hake\ CPUE) = \beta_0 + \beta_1 \log(Ekman) \times \log(curl) + \beta_2 \log(curl) + \beta_3 \log(Ekman) + \epsilon_i \qquad [log(Ton/hr)] \quad (6.3)$$

Where the estimated values for β and 95% confidence interval are:

$\beta_0 = 1.003$ (0.83, 1.18)
$\beta_1 = -1.107$ (-2.17,-0.04)
$\beta_2 = 0.661$ (-0.93, 2.25)
$\beta_3 = 0.102$ (0.83, 1.18)

The error ϵ_i are normally distributed errors following N(0, σ^2):
The regression model is showing the relationship that existed between December 2009 and October 2014 which is a period of consistent monthly data collection. The regression line is shown in red and the 95% confidence interval above regression is (0.259-0.658) log(Hake CPUE) is shown in blue curves around the regression line as demonstrated in Figure 6.8. This regression models a relationship between Hake CPUE and the significant predictor variables, Ekman and curl upwelling with a 0.48 and 0.42 significant correlations. The regression model has an adjusted $r^2 = 0.29$, (3,55) degrees of freedom and an F statistic of 8.93$\sim F_{3,55}$ and it is significant at p<0.001 (F(3,55) = 8.93, p<0.001).

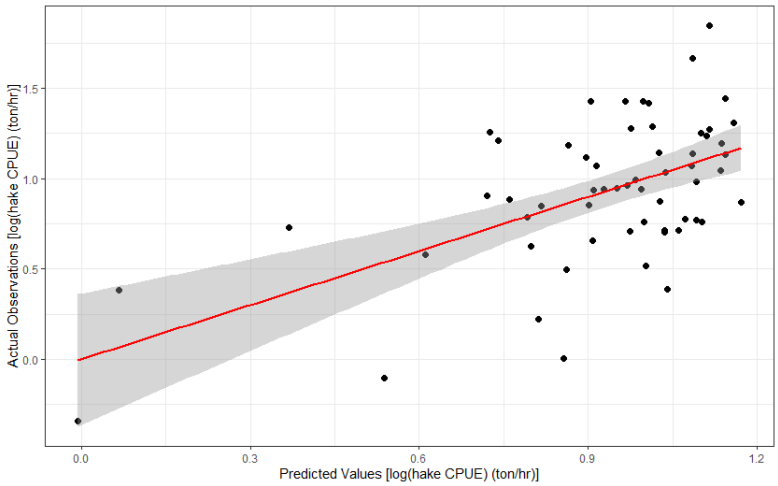

FIGURE 6.8: The red line is the regression line while the shade of grey is the 95% confidence interval. A multiple linear regression model was calculated to predict a Hake Catch based on Upwelling. A significant regression model was found ($F\sim(3,55)=8.93$, $p<0.001$), with adjusted $r^2 = 0.29$.

7 Discussion

7.1 How pressure systems create upwelling winds

The patterns shown in Chapter 4 in Figure 4.2 display conditions that do not favour Ekman upwelling. The negative meridional wind stress indicates winds with a downwelling effect (which is a strong southward wind component parallel to the coast in the south Benguela region) shown in the blue field. This signifies the possible conditions that prevail during the approach of a low-pressure system which are more prevalent in winter in this region. The opposite is true during summer. The presence of the South Atlantic high-pressure system (SAHPS) creates anticlockwise winds which manifest as meridional winds along the Benguela coastline and in the West Coast of South Africa. Figure 4.4 demonstrates this with intense positive wind stress which suggests upwelling favourable wind conditions over the West Coast of South Africa.

The winter and summer wind stress curl figures show low upwelling favourable and high upwelling favourable conditions respectively as seen in Figure 4.3 and 4.5 in the Southern Benguela region. The winter wind stress curl is characterised by negative curl near the coast and positive curl further from the coast. This is because due to coastal friction the winds acquire an anticlockwise spin which induces downwelling momentum on the ocean shores of the Southern hemisphere when northerly winds generated from the cyclonic winds are present due to the prevailing LPS. Figure 4.5 on the other hand shows that the converse is true during summer as it appears that the lower the SAHPS the better formed are the South Easterly winds that favour upwelling in the our study area.

It is important to note that LPSs propagate eastward in the Southern Ocean as part of the larger Rossby wave, therefore they would arrive and pass over this region during winter at a frequency bringing about low upwelling conditions. This could well describe the pattern shown in Figure 5.4 where the upwelling and downwelling fluctuations are clearly seasonal. The winter months generally avoid going intensely negative such that there is full scale downwelling during the decade of our study. The zero line in Figure 5.3 (A) helps in demonstrating this point. This indicates that the Benguela region can not really be classified as a downwelling region even on a seasonal basis since downwelling seldom occurs and and when it does take place the rates are much less than upwelling velocities that characterize it during upwelling. The resulting upwelling velocities from the winds of 2010 depicted in Figure 4.6 are shown in Figure 4.7 and

4.8 where Figure 4.7 is the Ekman driven upwelling and it shows that most rapid Ekman upwelling happens during the summer months as expected and that near zero happens in winter. The meridional wind direction is important to upwelling but is greatly enhanced by high wind stress intensities for producing the highest Ekman upwelling velocities possible otherwise they upwelling velocities are low. The same story holds true for the curl driven upwelling as seen in Figure 4.8. Months with the highest meridional wind stress intensity and with the most parallel to the coast winds which were December and January in 2010 had the highest curl driven upwelling velocities.

7.2 How chlorophyll responds to upwelling

Chlorophyll patterns exhibit a one month lag after the upwelling as shown in chapter 5, Figure 5.4. The monthly chlorophyll concentrations of 2010 are displayed in Figure 4.9 where it appears that most chlorophyll concentration is nearer to the coast in each month except in March 2010 where the concentrations were so high that they intensely covered the entire region of study. The 450 m contour represents the most active trawling region where the highest hake catches are also obtained according to the database used in this study. It is appearing that chlorophyll abundance does not take place directly above this contour region and that the food chain energy transfer from the chlorophyll all the way down to hake is not vertical in the ocean profile but it is transported in a lateral fashion offshore by the marine food chain trophic levels that exist between hake and primary production. It has been demonstrated in the statistical section that chlorophyll is not central to the determination of hake CPUE in this region. Chlorophyll has a low yet positive and significant correlation ($r=0.37$, $df=57$, $p=0.004$) to hake CPUE.

The correlation between Ekman upwelling and chlorophyll is ($r=0.62$, $df=57$, $p<0.001$) within this 10 year period as it can be seen with strong seasonal oscillations in Figure 5.3 (A) & (B). This indicates that the principal determinant of chlorophyll in this region is Ekman upwelling however there is another 38 % which is not explained by curl driven or Ekman upwelling which appears to be associated to undetermined variables. This correlation exists when the chlorophyll lag to upwelling is taken into consideration. The lag between peak upwelling and peak chlorophyll as found to be 1 month as shown earlier in Chapter 4, Section 4.3.

7.3 Time series analysis and hake response

The time series analysis performed on Chapter 4 and Chapter 5 are extensions of each other on the same analysis where chapter 5 focuses on hake CPUE while Chapter 4 scrutinizes the field data (predictor variables), upwelling and chlorophyll. Figure 4.10 displays a monthly variation of the predictor variables where it is apparent that curl driven upwelling and Ekman upwelling are in phase more often than not. The reason for this is that coastal parallel meridional winds that leads to Ekman upwelling often results on a wind stress curl as well when their interaction

with the coastal land mass friction is such that wind magnitude increases offshore. This can be seen in spacial figures such as Figure 4.8 on the December plot where wind vectors are much more intense offshore and less so near the coast.

The chlorophyll timeseries is not in phase with either the Ekman or curl driven upwelling as also seen in Figure 4.10 even though it also appears to have a seasonal pattern like the upwellings. This has justified a cross-correlation analysis where the search is on to find the lag that exists between the predictor variables. It is shown in Figure 4.11 (i) that there is no time lag (months) between Curl upwelling and Ekman upwelling. Figure 4.11 (ii) and (iii) however shows that there is a lag of one month between Ekman upwelling and chlorophyll as well as between curl upwelling and chlorophyll respectively. Figure 5.4 (i), (ii) and (iii) shows the lags that exist between hake CPUE and Ekman upwelling, curl upwelling and chlorophyll. The exact lags are given in chapter 5 as all the hake data quantification is performed and kept in that chapter.

7.4 Reconstructions of fishing days as a case study with time lag considerations

The reconstructions of fishing days performed in a month of intense upwelling, is depicted in Figure 5.5, 5.6 and 5.7 shows that when compared with the trawling intensity of that same month, there was fewer catches as shown in Figure 5.8 as there was fewer drags and a high prevalence of low CPUE drag occurrences. This signifies that the upwelling rates of the current month are not fundamental in the determination of hake CPUE for that same month. The intense upwelling month of December 2010 showed 12 days of very high upwelling days but produced fewer high CPUE drags and a lot of low CPUE drags compared to a month observed after a lag period.

After a time-lag determined in chapter 5, the period it takes for hake CPUE to peak after an upwelling peak. It was found that after the intense upwelling peak, hake CPUE increases dramatically as shown in Figure 5.13 and 5.14 where hake CPUEs are higher than in December 2010. It is also shown in Figure 5.10, 5.11 and 5.12 that this increase in hake CPUE can not be associated with the upwelling of the lag-month because the upwelling velocities in the lag month are actually lower than those observed in December 2010. This leads to the conclusion that the high hake CPUE drags after the lag period are associated to the upwelling that happened before the lag.

7.5 Statistical analysis and regression modelling

The lag analysed above is statistically tested by performing a Pearson correlation analysis on the observed predictor variables against each other and against the response variable. The Pearson correlation between the Ekman upwelling and hake CPUE is negative and insignificant as

shown in Table 6.1. The same Pearson correlation performed after a time lag (cross-correlation) showed an existing positive and significant relationship (r=0.48, df=57, p<0.05). The curl driven upwelling exhibit the same relationship with hake CPUE where the relationship is insignificant when no lag is considered but the correlation is positive and significant at (r=0.43, df=57, p<0.05) once the lag is considered as depicted in Table 6.1. The chlorophyll-hake CPUE correlation also became positive and significant as shown in the same table. A multivariate linear regression model was established on the basis of the interaction effect between Ekman upwelling and curl upwelling in determining hake CPUE. A significant regression model was found (F(3,55)=8.93, p<0.05), with R^2 = 0.33 and adjusted R^2 =0.29. Ekman upwelling and curl upwelling in [Sv] predicted the below hake model equation where hake is calculated in tons per hour.

$$Hake\ CPUE = e^{\beta_0+\beta_1 \log(Ekman)\times\log(curl)+\beta_2 \log(curl)+\beta_3 \log(Ekman)+\epsilon_i} \qquad [Ton/hr]$$

The R^2 = 0.29 signifies the amount of variability that the predictor variable (upwelling) can explain on the response variable (hake CPUE). The R^2 of 29% looks small. However, it explains a third of the variation. This means it is a very important outcome although it is still possible to improve the R^2 of the of the hake statistical model in the Southern Benguela region. One way is through taking into account further variables such as bottom temperature, bottom depth, dissolved oxygen etc.

8 Conclusions.

8.1 Summary

A study of wind-driven upwelling effect and hake catches has revealed that there is a link between hake CPUE and the two forms of upwelling (Ekman driven and curl driven upwelling). The study is conducted at a significance level of $\alpha = 0.05$. The obtained adjusted R^2 of 29% shows that with the adopted multiple regression model, a third of the variation in the of hake CPUE can be explained by the interaction of the Ekman and curl driven upwelling where Ekman upwelling was a bigger contributor. A significant correlation of 0.57 exists between the modelled values and the observed hake CPUE data. This indicates that the coexistence (interaction) of the Ekman and curl driven upwelling positively correlates with hake CPUE at more than half the time (r=0.57, df=(57),p<0.05) and that you can predict a hake CPUE magnitude with regression model Equation 6.3. It is important to be mindful of the fact that a regression model only describes the relationship during the time range of the analysed data as decadal variations may interfere with this relationship. The above regression model is based on the monthly period December 2009 to October 2014. The reason for the relationship to be only modelled within this 5 year period instead of the the entire 10 year period of data collection is the lack of consistency in data collection in the the latter 5 years. There is no causality association that can be firmly deduced from the upwelling and hake CPUE relationship because the R^2 is relatively low and it only describes the third of the variation.

A lag has been found to exist between the upwelling, chlorophyll and hake CPUE as demonstrated in the results, Chapter 4 and Chapter 5. Lagged correlation can only be determined if the time series data-set is unbroken and an unbroken time series exists within the above mentioned period for hake CPUE which explains why modelling was only performed within the said time range (December 2009 to October 2014). The positive relationship and significant correlation between the predictor variables, (Ekman upwelling, curl driven upwelling and chlorophyll) and the response variable (hake CPUE) is, therefore, found to be only existing when the lag is taken into consideration otherwise there is no significant correlation which answers the research question of the project which questions if there is any predictability of hake that can be achieved on the basis of understanding the impact of winds. Chlorophyll however does not significantly contribute to the model that predicts hake CPUE in this region, therefore is excluded.

The comparison of the findings in the Southern Benguela region and the findings made in the coast of California by Rykaczewski and Checkley (2008) shows that in the south Benguela region, the Ekman upwelling is more important than the curl driven upwelling which was a major contributor in the California current ecosystem. The correlation between curl driven upwelling and sardine biomass was higher at (r=0.62, df=22, $\alpha < 0.05$) than our correlation numbers. However, it is important to take into account that hake is a top predator when compared to sardine. This means that the further dissipation of correlation strength can be attributed to a lot more white noise that is bound to exist when the lag is bigger as hake has a more delayed response to upwelling than sardine through the food chain. The Californian study was also performed at a lower temporal resolution than this one as its comparisons were done on an annual level. This would have not been possible in this study nor would it have been desired because we have a shorter time series in an annual scale and we are interested in prediction at an operational level (timely).

The lack of data collection consistency during the latter 5 years of the of the data collection program is due to two suspected major reasons;

1. The vessel skippers (Captains) may have fallen out of favour with the routine of collecting data. This may have been the case because of the lack of feedback in the form of prediction based on the collected data at that time.
2. The skippers may have deduced a pattern within the first five years and then arbitrarily decided to act on it by choosing to fish in this region in seasons they thought were productive and else where on seasons they doubted.

The first reason suggests a situation where there was no data collection taking place at all and the second possible reason suggests that they may have moved their vessels to fish elsewhere during certain seasons which mean there were no vessels located in this region for intense data collection in some months as compared to others. The chlorophyll yielded a surprisingly lower correlation outputs compared to upwelling when correlated with hake CPUE. This signifies that other variables are involved in the marine food chain of the Southern Benguela region instead of chlorophyll being the driver of the whole process after an upwelling event.

Based on the analysis of all the variables and based on the regional and international literature cited throughout this study, it is concluded that hake in the Southern Benguela region is part of the food chain that is driven by upwelling, predominantly Ekman upwelling even though the chlorophyll may not strongly support this staggered process with its minimal correlation.

8.2 Recommendations

- It is apparent that incorporating more variables into the model might yield improved results by elevating the R^2. Therefore, one way of taking this study forward is through incorporating further predictor variables in the study such as the bottom temperature, bottom depth, oxygen concentration and other possible predictor variables for hake CPUE.
- It is recommended that new data gets collected and be assessed in comparison with the correlations obtained over the assessed period to check if they still hold considering that there may be a decadal shift among correlations in some instances as suggested by *Mann (1993)*.
- It is recommended that other fishing regions be studied in the similar fashion in order to ascertain if there is existence of habitat shifts under different seasonal conditions as is the case in this region. This would also give information on the upwelling regimes in different parts of the South African coast as the wind directions will impact hake fishing in different ways depending on the orientation of the fishing ground along the coast.

References

Akaike, Hirotugu (1974). "A new look at the statistical model identification". In: *IEEE transactions on automatic control* 19.6, pp. 716–723.

Albert, A. et al. (2010). "Impact of nearshore wind stress curl on coastal circulation and primary productivity in the Peru upwelling system". English. In: *Journal of Geophysical Research* 115.C12, p. C12033. DOI: 10.1029/2010JC006569. URL: https://hal.archives-ouvertes.fr/hal-00691549.

Andrews, W.R.H. and L. Hutchings (1980). "Upwelling in the Southern Benguela Current". In: *Progress in Oceanography* 9.1, pp. 1 –81. ISSN: 0079-6611. DOI: https://doi.org/10.1016/0079-6611(80)90015-4. URL: http://www.sciencedirect.com/science/article/pii/0079661180900154.

Atkinson, LJ et al. (2011). "Changes in demersal fish assemblages on the west coast of South Africa, 1986-2009". English. In: *African Journal of Marine Science* 33.1, pp. 157–170. DOI: 10.2989/1814232X.2011.572378. URL: http://www.tandfonline.com/doi/abs/10.2989/1814232X.2011.572378.

Bachèlery, Marie-Lou, Serena Illig, and Isabelle Dadou (2016). "Interannual variability in the South-East Atlantic Ocean, focusing on the Benguela Upwelling System: Remote versus local forcing". In: *Journal of Geophysical Research: Oceans* 121.1, pp. 284–310. DOI: 10.1002/2015JC011168. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2015JC011168. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JC011168.

Booth, A. J. (2000). "Incorporating the spatial component of fisheries data into stock assessment models". English. In: *ICES Journal of Marine Science* 57.4, pp. 858–865. DOI: 10.1006/jmsc.2000.0816. URL: http://www.ingentaconnect.com/content/ap/jm/2000/00000057/00000004/art00816.

Botha, L. (1985). "Occurrence and distribution of Cape hakes Merluccius capensis Cast. and M. paradoxus Franca in the Cape of Good Hope area". English. In: *South African Journal of Marine Science* 3.1, pp. 179–190. DOI: 10.2989/025776185784461207. URL: http://www.ingentaconnect.com/content/nisc/sajms/1985/00000003/00000001/art00017.

Bunn, Andrew G. (2010). "Statistical and visual crossdating in R using the dplR library". In: *Dendrochronologia* 28.4, pp. 251–258. ISSN: 1125-7865. DOI: 10.1016/j.dendro.2009.12.001.

Carr, Mary-Elena (2001). "Estimation of potential productivity in Eastern Boundary Currents using remote sensing". English. In: *Deep-Sea Research Part II* 49.1, pp. 59–80. DOI: 10.1016/S0967-0645(01)00094-7. URL: https://www.sciencedirect.com/science/article/pii/S0967064501000947.

Carr, Mary-Elena and Edward J. Kearns (2003). "Production regimes in four Eastern Boundary Current systems". English. In: *Deep-Sea Research Part II* 50.22, pp. 3199–3221. DOI: 10.1016/j.dsr2.2003.07.015. URL: https://www.sciencedirect.com/science/article/pii/S0967064503001863.

Chavez, Francisco P. and Monique Messié (2009). "A comparison of Eastern Boundary Upwelling Ecosystems". English. In: *Progress in Oceanography* 83.1, pp. 80–96. DOI: 10.1016/j.pocean.2009.07.032. URL: https://www.sciencedirect.com/science/article/pii/S0079661109000998.

Dee, Dick P et al. (2011). "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system". In: *Quarterly Journal of the royal meteorological society* 137.656, pp. 553–597.

Engelbrecht, Christien and Willem Landman (2016). "Interannual variability of seasonal rainfall over the Cape south coast of South Africa and synoptic type association". English. In: *Climate Dynamics* 47.1, pp. 295–313. DOI: 10.1007/s00382-015-2836-2. URL: https://search.proquest.com/docview/1799581293.

Fairweather, TP et al. (2006). "Spatial description of hake-directed fishing activity off the west coast of South Africa". English. In: *African Journal of Marine Science* 28.1, pp. 13–24. DOI: 10.2989/18142320609504129. URL: http://www.tandfonline.com/doi/abs/10.2989/18142320609504129.

Faraway, Julian J (2014). *Linear models with R*. CRC press.

Favre, Alice et al. (2013). "Cut-off Lows in the South Africa region and their contribution to precipitation". English. In: *Climate Dynamics* 41.9, pp. 2331–2351. DOI: 10.1007/s00382-012-1579-6. URL: https://search.proquest.com/docview/1446215025.

Fishing Industry Handbook, (2018). 46th edition. South Africa: George Warman PUBLICATIONS.

Garavelli, L. et al. (2012). "Modeling the dispersal of Cape hake ichthyoplankton". English. In: *Journal of Plankton Research* 34.8, pp. 655–669. DOI: 10.1093/plankt/fbs039. URL: 10.1093/plankt/fbs039.

Gordoa, Ana, Mercedes Masó, and Lizette Voges (2000). "Monthly variability in the catchability of Namibian hake and its relationship with environmental seasonality". English. In: *Fisheries Research* 48.2, pp. 185–195. DOI: 10.1016/S0165-7836(00)00160-0. URL: https://www.sciencedirect.com/science/article/pii/S0165783600001600.

Hutchings, L et al. (2009). "The Benguela Current: An ecosystem of four components". In: *Progress in Oceanography* 83.1-4, pp. 15–32.

Ismail, Hassan Ebrahiem et al. (2015). "Relation between upwelling intensity and the variability of physical and chemical parameters in the southern Benguela upwelling system". In: *International Journal of Oceanography* 2015.

Japp, D. W. (1990). "A new study on age and growth of kingklip Genypterus capensis off the south and west coasts of South Africa, with comments on its use for stock identification". English. In: *South African Journal of Marine Science* 9.1, pp. 223–237. DOI: 10.2989/02577619078437875 4. URL: http://www.ingentaconnect.com/content/nisc/sajms/1990/00000009/00000001/art00019.

Kleinschmidt, H., W. H. H. Sauer, and P. Britz (2003). "Commercial Fishing Rights Allocation in Post-apartheid South Africa: Reconciling Equity and Stability". English. In: *African Journal of Marine Science* 25.1, pp. 25–35. DOI: 10.2989/18142320309503998. URL: http://www.tandfonline.com/doi/abs/10.2989/18142320309503998.

Mann, K. (1993). "Physical oceanography, food chains, and fish stocks: a review". English. In: *ICES Journal of Marine Science* 50.2, pp. 105–119. DOI: 10.1006/jmsc.1993.1013.

Marshall, John and R. Alan Plumb (2007). *Atmosphere, Ocean and Climate Dynamics*, First edition. South Africa: Academic Press.

MATLAB (2018). *version 9.4.0.813654 (R2018a)*. Natick, Massachusetts: The MathWorks Inc.

Mbatha, FL et al. (2019). "Oxygen and temperature influence the distribution of deepwater Cape hake Merluccius paradoxus in the southern Benguela: a GAM analysis of a 10-year time-series". In: *African Journal of Marine Science* 41.4, pp. 413–427.

Messié, Monique and Francisco P. Chavez (2015). "Seasonal regulation of primary production in eastern boundary upwelling systems". English. In: *Progress in Oceanography* 134, pp. 1–18. DOI: 10.1016/j.pocean.2014.10.011. URL: https://www.sciencedirect.com/science/article/pii/S0079661114001773.

Payne, A. I. L., B. Rose, and R. W. Leslie (1987). "Feeding of hake and a first attempt at determining their trophic role in the South African west coast marine environment". English. In: *South African Journal of Marine Science* 5.1, pp. 471–501. DOI: 10.2989/025776187845226 67. URL: http://www.ingentaconnect.com/content/nisc/sajms/1987/00000005/00000001/art00037.

Payne, Andrew IL and André E Punt (1995). "Biology and fisheries of South African Cape hakes (M. capensis and M. paradoxus)". In: *Hake*. Springer, pp. 15–47.

Pfaff, Maya C et al. (2011). "Upwelling intensity and wave exposure determine recruitment of intertidal mussels and barnacles in the southern Benguela upwelling region". In: *Marine Ecology Progress Series* 425, pp. 141–152.

Pillar, S. C. and M. Barange (1997). "Diel variability in bottom trawl catches and feeding activity of the Cape hakes off the west coast of South Africa". English. In: *ICES Journal of Marine Science* 54.3, pp. 485–499. DOI: 10.1006/jmsc.1996.0169. URL: http://www.ingentaconnect.com/content/ap/jm/1997/00000054/00000003/art00169.

Piretzidis, Dimitrios and Michael G. Sideris (2016). *MAP-LAB: A MATLAB Graphical User Interface for generating maps for geodetic and oceanographic applications*. DOI: 10.13140/RG.2.2.16099.76323.

Punt, A. E. (1992). "Selecting management methodologies for marine resources, with an illustration for southern African hake". English. In: *South African Journal of Marine Science* 12.1, pp. 943–958. DOI: 10.2989/02577619209504754. URL: http://www.ingentaconnect.com/content/nisc/sajms/1992/00000012/00000001/art00069.

Punt, A. E. and R. W. Leslie (1991). "Estimates of some biological parameters for the Cape hakes off the South African west coast". English. In: *South African Journal of Marine Science* 10.1, pp. 271–284. DOI: 10.2989/02577619109504637. URL: http://www.ingentaconnect.com/content/nisc/sajms/1991/00000010/00000001/art00024.

R Core Team (2019). *R: A Language and Environment for Statistical Computing, version 3.6.2 (2019-12-12)*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rencher, Alvin C and G Bruce Schaalje (2008). *Linear models in statistics*. John Wiley & Sons.

Rykaczewski, Ryan R. and David M. Checkley (2008). "Influence of Ocean Winds on the Pelagic Ecosystem in Upwelling Regions". English. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.6, pp. 1965–1970. DOI: 10.1073/pnas.0711777105. URL: https://www.jstor.org/stable/25451397.

SADC-EU (2017). *SOUTH AFRICAN FISHERIES AND THE SACD-EU ECONOMIC PARTNERSHIP AGREEMENT*. URL: Https://Sadc-Epa-Outreach.Com/Images/Files/Sadc-Eu-Epa-Fisheries-July-2017.Pdf..

SAWS. *South African Weather Service*. https://www.weathersa.co.za/Home/HistoricalSyno

Schumann, Eckadth, L. Perrins, and I. Hunter (1982). "Upwelling along the South Coast of the Cape Province, South Africa." In: *South African Journal of Science* 78, pp. 238–242.

Van Rossum, Guido and Fred L. Drake (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace. ISBN: 1441412697.

Veitch, Jennifer, Pierrick Penven, and Frank Shillington (2010). "Modeling equilibrium dynamics of the Benguela Current System". In: *Journal of Physical Oceanography* 40.9, pp. 1942–1964.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. Fourth. ISBN 0-387-95457-0. New York: Springer. URL: http://www.stats.ox.ac.uk/pub/MASS4/.

Verheye, Hans M et al. (2016). "Plankton productivity of the Benguela current large marine ecosystem (BCLME)". In: *Environmental Development* 17, pp. 75–92.

Weldon, D. and C. Reason (2014). "Variability of rainfall characteristics over the South Coast region of South Africa". English. In: *Theoretical and Applied Climatology* 115.1, pp. 177–185. DOI: 10.1007/s00704-013-0882-4. URL: https://search.proquest.com/docview/1473940283.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

Wickham, Hadley et al. (2019). "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43, p. 1686. DOI: 10.21105/joss.01686.