AALBORG UNIVERSITY
COPENHAGEN

Salinity Fluctuations in the Labrador Sea 2012 - 2016

James Fethers, Sarah Lafaire, Nijole Makovskaja, Georgina Scholes

Surveying, Planning and Land Management / Geoinformatics 2nd Semester,

Summer 2017

The data used in this project were collected and made freely available by the International Argo Program and the national programs that contribute to it. (http://www.Argo.ucsd.edu, http://Argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System.

Table of Contents

Acronyms

- AIC - Argo Information Centre
- CRS - Coordinate Reference System
- .CSV - Comma Separated Values
- CTD - Conductivity Temperature Data
- DAC - Data Assembly Centres
- ESRI - Environmental Systems Research Institute
- GDAC - Global Data Assembly Centre
- GIS - Geographical Information Systems
- GOCB - Global Ocean Conveyor Belt
- GPS - Global Positioning System
- GTS - Global Telecommunications System
- GWR - Geographically Weighted Regression
- NASA - National Aeronautics and Space Administration
- NetCDF - Network Common Data Form
- NOAA - National Centres for Environmental Information
- OLS - Ordinary Least Squares
- PMEL - Pacific Marine Environmental Laboratory
- PPT - Parts Per Thousand
- PSU - Practical Salinity Unit
- QGIS - Quantum Geographic Information Systems
- UTM - Universal Transverse Mercator

Table of Codes

Table of Figures

1. Introduction

1.1 Background

The ‘Great Salinity Anomaly’ was first observed in the North Atlantic Ocean during the 1960’s, where a cool, relatively fresh water mass was observed as a result of above average freshwater runoff from North America. This was followed by greater than average formation of sea ice in the Arctic and suppression of deep convection, then a reduction of cyclonic activity and precipitation over Northern Canada in the following years. Sea ice was observed drifting from the Arctic to the waters around Greenland and melting during the summer, contributing to the cool, fresh layer. This process suggests a self-sustained feedback loop with a period of approximately 15-20 years, which is supported by observation of a similar process in the late 1980’s (Mysak and Power, 1990).

This example highlights the large-scale factors influencing ocean temperature and salinity, which are key components in deep convection that drives thermohaline circulation and the meridional heat transport that regulates the earth’s climate (Klemas, 2001). It demonstrates how salinity and temperature are key variables to understand these processes, because they provide information about evaporation and precipitation at the ocean surface that are important to understand the hydrological cycle, and also water density which drives deep ocean currents through the sinking of water masses.

The purpose of this project is to collect salinity and temperature data, and explore, analyse and visualise it using statistical analysis and geographical information systems (GIS) to determine any trends over a 5-year period.

It is hypothesised that a reduction in salinity will be observed over the study period due to increased ice melt and reduction of sea ice formation as a consequence of climate change, and that temperature with a combination of other factors will be an explanatory variable.

1.2 Research Questions

What are the factors influencing surface salinity in the Labrador Sea?

1. Does surface salinity change over a 5-year period in the Labrador Sea?
2. Where is the surface (at what depth does salinity change)?
3. Does a correlation exist between salinity and ocean temperature?
4. Do any other factors influence salinity?

1.3 Report Structure

This report consists of seven Chapters, beginning with an introduction in which the background ideas are outlined as well as the purpose of the project and research questions. Chapter 2 contains a discussion of theory relating to ocean salinity and temperature, the reasons why the Labrador Sea was chosen as the study area, and information about the Argo program and other potential sources of ocean salinity data. The processes and tools used in the development of the project are outlined in Chapter 3. These include project management techniques, software, geographical data analysis tools and geoprocessing and visualisation methods. In Chapter 4 the data used in the project is described with a focus on the complexities of the data, data handling and processing that was necessary for analysis. The scripts written in Python and R are also discussed and explained in this section. The results of the data analysis are presented in Chapter 5, which includes images and descriptions of the exploratory plots created to see trends and anomalies in the data, statistical analysis of salinity levels and any explanatory variables that were found to impact them. The interpolations of the point data are also included in Chapter 5 to provide visualisation of the pattern of salinity in the Labrador Sea. Chapter 6 contains a discussion and analysis of the project findings and how they can be used to respond to the research questions described in Chapter 1. Finally, the project conclusions are presented in Chapter 7.

2. Theory

2.1 Climate and ocean science

2.1.1 Ocean salinity and why it fluctuates

Ocean salinity is a measure of the amount of salt per 1000g of water (PPT). It fluctuates due to the balance between salinity raising and decreasing factors. Salinity raising factors include evaporation and mineral input from the weathering of rocks, and decreasing factors include melting of ice, freshwater input from rivers, and precipitation of rain and snow (NASA, 2017).

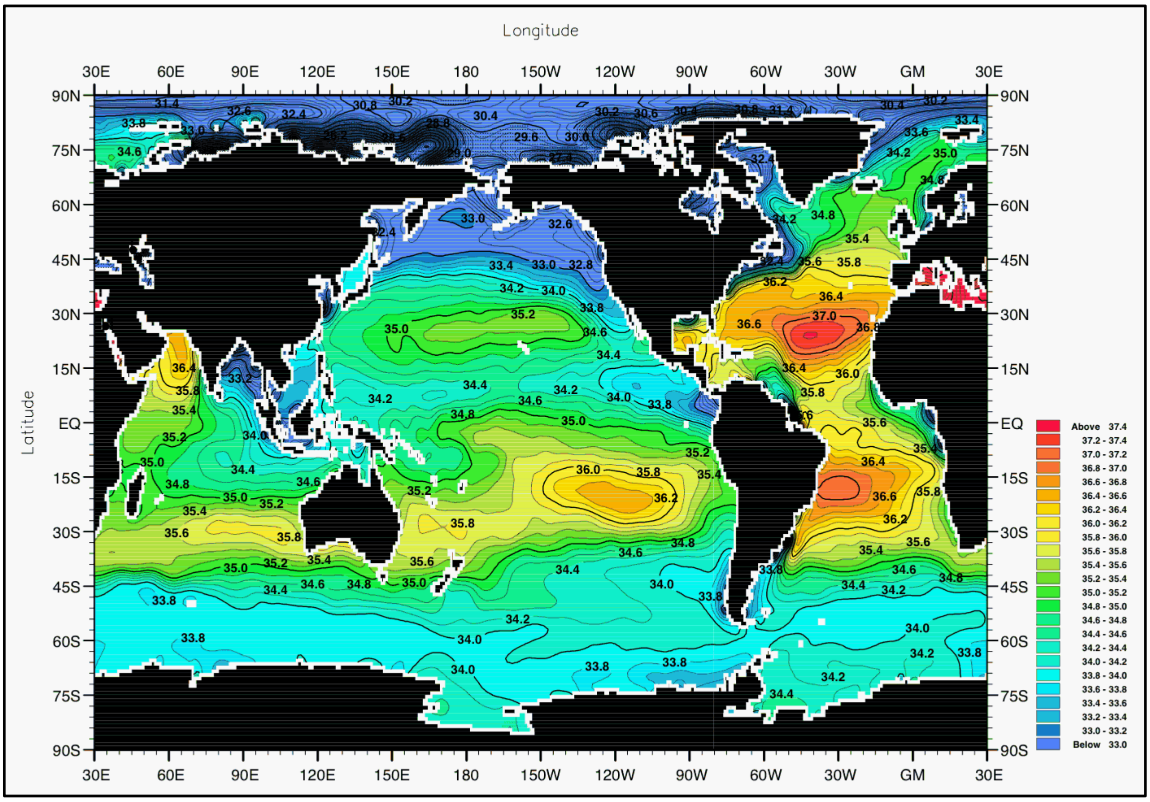


Figure 1: Global salinity concentrations from 2001(Source: NOAA - National Centres for Environmental Information, 2001)

Salinity levels are a balance of these factors, with waters around the tropics more saline due to evaporation exceeding freshwater input, and concentrations reducing towards the poles as fresh water sources exceed evaporation, as illustrated at Figure 1.

2.1.2 Why ocean salinity is important

Monitoring ocean salinity is important for several reasons. Firstly, because salinity is a product of evaporation and precipitation, it is a key indicator of variations in the water cycle such as land runoff, sea ice freezing and melting, and evaporation and precipitation over the oceans (NASA, 2017). Fluctuation in salinity provides evidence for imbalances in these processes, and therefore provides an indicator for phenomena such as climate change.

Secondly, ocean salinity, combined with temperature, determines density; which is a key driver for deep ocean currents that transport heat and maintain the Earth's climate (NASA, 2017). Density increases with increased salinity and decreased temperature, therefore a combination of these two factors result in denser water that sinks and drives deep ocean currents. Studies indicate seawater is becoming fresher in high latitudes and tropical areas dominated by rain, while in sub-tropical high evaporation regions, waters are getting saltier, which could impact on ocean circulation and global climate (ibid).

Knowledge about temperature is also useful for information about heat content for measuring climate change and predicting el Niño cycles.

2.1.3 Labrador Sea

The Labrador Sea south of Greenland, as depicted at Figure 2, is subject to seasonal variations in salinity and temperature due to variations in ocean currents and meteorological conditions (Buch, 1984).

Factors that influence ocean salinity and temperature in this region are a complex combination of precipitation, meteorological factors, and ocean currents (Straneo, 2005; Buch, 1984). In Spring and Summer (March → August) cool freshwater is released from ice melt, which lowers salinity and forms a stratified layer on the surface due to the relatively high water density of the water below (Buch, 1984; Straneo, 2005). In this instance the relatively low salinity has a greater influence on water density than temperature. During Autumn and Winter (September → February), salinity increases due to the formation of sea ice, which in turn increases the density of the surface layer and results in convective mixing that homogenises the sea layers (Straneo, 2005). This results in a broad trend of a layer of low salinity on the ocean surface for 6 months, and higher salinity with high water mixing and no layers for the other 6 months.

Figure 2: The Labrador Sea south of Greenland

Ocean currents in the region also play an important role, contributing a relatively consistent influx of cold and warm water during certain times of the year. For example, the East Greenland Current offsets warmer surface temperature during Spring with an influx of cool water and drift ice, however the temperature of this current can vary based on its proximity to the coastline (Buch, 1984). This highlights the complexities involved with predicting large-scale phenomenon such as ocean salinity and temperature, and therefore caution must be exercised when determining cause and effect relationships.

Observing salinity from 2001 in the Labrador Sea, a trend of salinity increasing moving west to east can be observed (Figure 3). This trend will be further assessed later in the project.

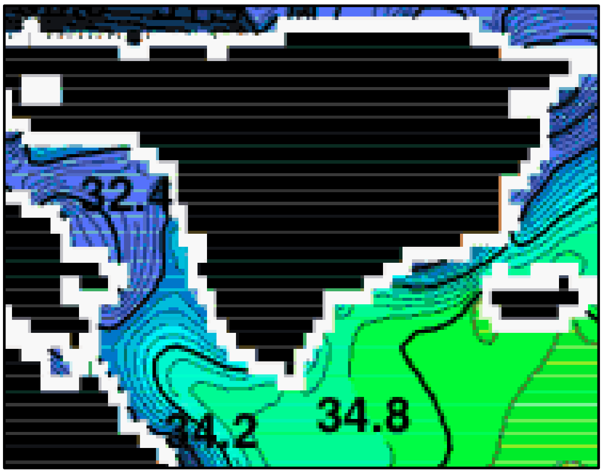


Figure 3: Salinity increases from west to east (Source: NOAA - National Centre for Environmental Information, 2001)

The Labrador Sea is an important region for several reasons, however the primary reason is because it is the region where the sinking of dense seawater drives deep ocean currents that form a component of the Global Ocean Conveyor Belt (GOCB), which is characterized by a northward flow of warm, salty water in the upper layers of the Atlantic, and a southward flow of colder water in the deep Atlantic (Figure 4).

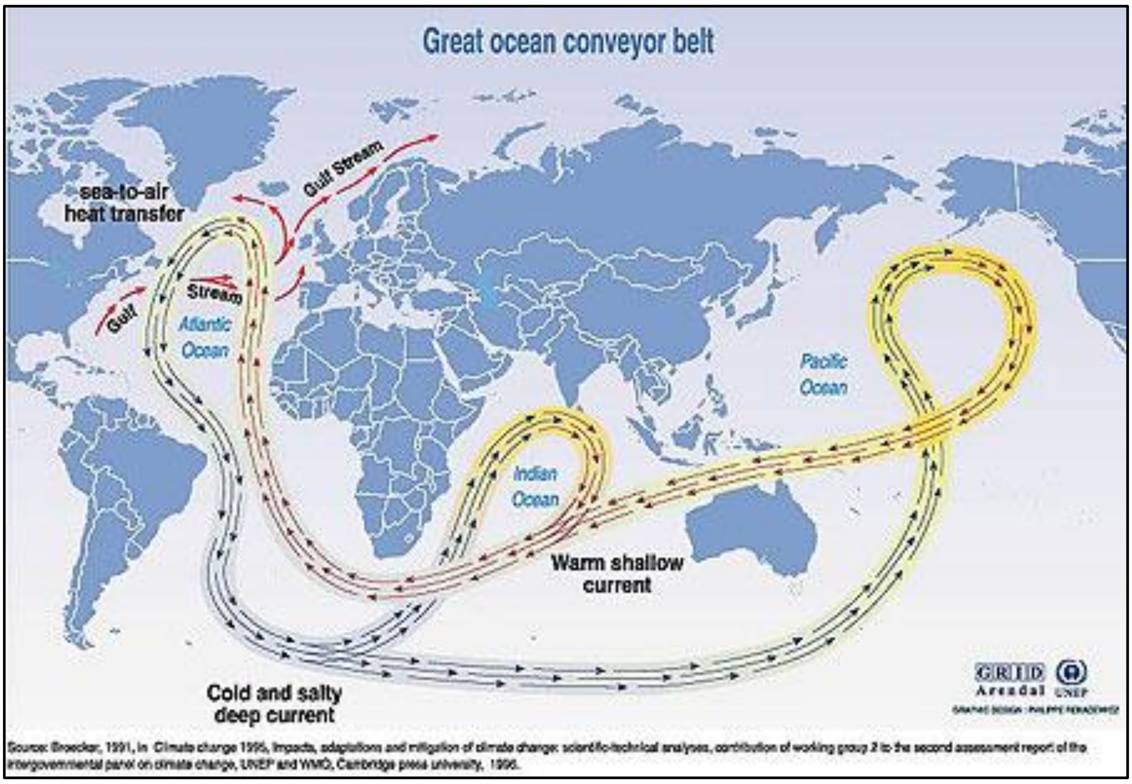


Figure 4: Global water flow

The GOCB is important because it transports a substantial amount of heat from the Tropics and Southern Hemisphere toward the North Atlantic, where the heat is transferred to the atmosphere. Changes to this circulation system could have profound impacts on the global climate system, for example, changes in African and Indian monsoon rainfall, atmospheric circulation of relevance to hurricanes, and climate over North America and Western Europe (Buch, 1984).

The Labrador Sea is also significant because it is an important breeding area for several fish species. Temperature conditions of the upper water layer in June and early July depend greatly on the intensity of inflow of polar water, which in turn greatly influences biological production and the survival of fish larvae in the region (Buch, 1984).

2.1.4 Sea surface

There are multiple definitions of the sea surface, which are mostly dependent on the variable being measured, for example light, salinity, temperature, density etc. With regard to sunlight, the maximum depth at which any significant light reaches is 200m (NOAA 2015). This ocean level is called the euphotic or sunlight zone.

Sea surface in terms of temperature monitoring and measurement is less well defined with some programs using the near-surface layer of the ocean, being the top 2-3m (Soloviev & Lukas 2014), and others using up to the first 20m of depth (HaMAARAG 2017).

Remote sensing to measure salinity can generally return data for the top couple of centimetres of the ocean (for example the Aquarius satellite) (NASA Official 2017), which therefore restricts the definition of the surface to a much smaller scale than for other types of measurements, including Argo.

For the purposes of this project, several surface depth measurements were used: down to 200m, 50m and 20m, as well as the shallowest value for each float. These various surface depths were based either on the literature, or on analysis of the Argo data, and enabled nuances in the data to be observed.

2.1.5 Salinity and temperature

Sea surface temperatures are driven by a complex combination of solar radiation, precipitation, convective mixing, fresh water runoff, and ocean currents. As these factors can also influence salinity, understanding this relationship can provide insights into the balance between these factors.

In Arctic waters low salinity and low temperatures are expected because of low evaporation and high freshwater input from rivers and glacial melt. Any anomalies could indicate an imbalance of these factors, and an indication of long-term trends. For example higher surface temperatures could indicate increased solar radiation, which in the long-term results in increased salinity and could have implications for deep ocean currents.

Factors influencing ocean salinity and temperature are extremely complex; therefore any study that attempts to determine cause should be conducted over a large area and broad timescale. As this study is conducted over a relatively short timeframe and only examines a small subset of the factors that influence salinity and temperature, it is intended to only speculate about possible causal factors.

2.2 Argo

The Argo program is a global array of 3,800 free-drifting robotic profiling floats that measure the temperature and salinity in the upper 2000 m of ocean. It is a pioneering international collaboration that began in the year 2000, and is the sole source of global subsurface ocean datasets (Argo, 2017a). A total of 30 countries contribute to the program, with the United States responsible for around half of the global array. Despite the number of floats, the Argo array is considered sparse as they are deployed at approximately 3°x 3° spacing.

The purpose of is to improve the quantity and quality of upper ocean data, and document the seasonal, annual and decadal changes to temperature, current and salinity. This data is and will be used to formulate models to analyse and predict climate and ocean change and variability. The Argo program allows for continuous monitoring of the upper ocean, with all data being relayed and made publicly available within hours after collection.

The uses of Argo data are varied and include education, operations and research. Argo data is used for monitoring environmental conditions, which aid in making long term weather forecasts based on subsurface ocean temperatures, keeping track of fish stocks and biological productivity, and providing early warning signs of significant anomalies in ocean temperature currents and salinity (Argo 2017d).

Perhaps the most integral purpose of the Argo program is observation, monitoring and analysis of data concerning climate change. Although the program has not been operational long enough to observe long term global climate change reflected in the data, regional, seasonal and intra-annual variability can be observed in ocean temperatures, salinity, freshwater content, and sea-level heights. Given that Argo is a novel and pioneering program, part of the difficulty in making judgements about long-term climate change is the lack of historical data to compare.

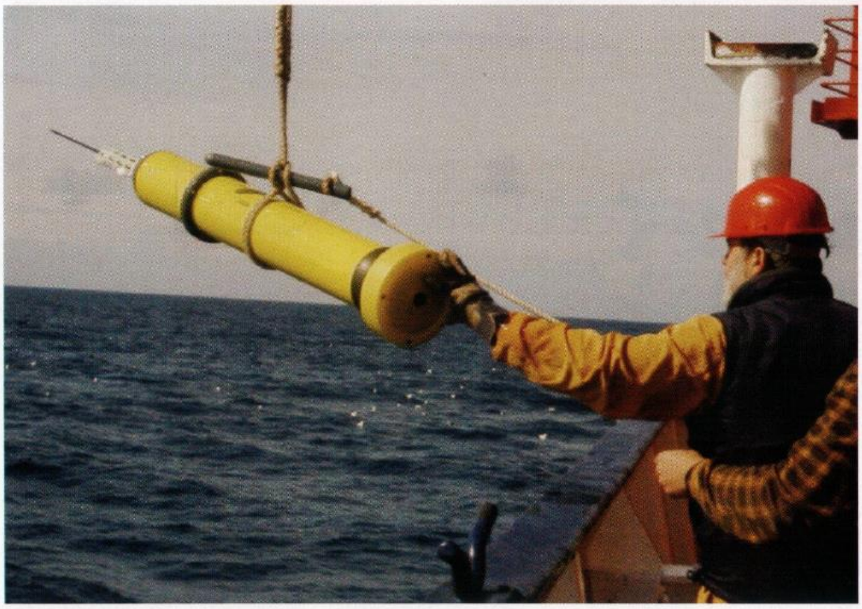


Fig. 1. An Argo float is launched from a research vessel. (Photo courtesy of Institut für Meereskunde/ GEOMAR, Kiel, Germany.)

Figure 5: Argo float being launched

2.2.1 Argo data collection

As demonstrated in Figure 6, each Argo float descends to a depth of 1000m and drifts in the ocean current for approximately 9 days, measuring mid-depth current, before descending further to 2000m. The float then pumps oil into an internal bladder and ascends to the surface, whilst collecting salinity and temperature data at various depths. Once at the surface the position of the float is determined by satellite or GPS, and the process repeats. Each float has a life of approximately 4 years due to the expiration of their lithium batteries, and will conduct approximately 140 cycles (Argo FAQ, 2015).

The data transmitted to the data centres consists of about 200 pressure, temperature and salinity measurements. Some models of Argo floats are capable of measuring and transmitting a greater number of measurements more frequently, resulting in several hundred values per profile. To ensure error free transmission of data, each float using the Argo system must spend 6-12 hours at the surface. Newer floats using GPS and Iridium satellites have a shorter surface time requirement and have two-way communication enabled. (Argo 2017b)

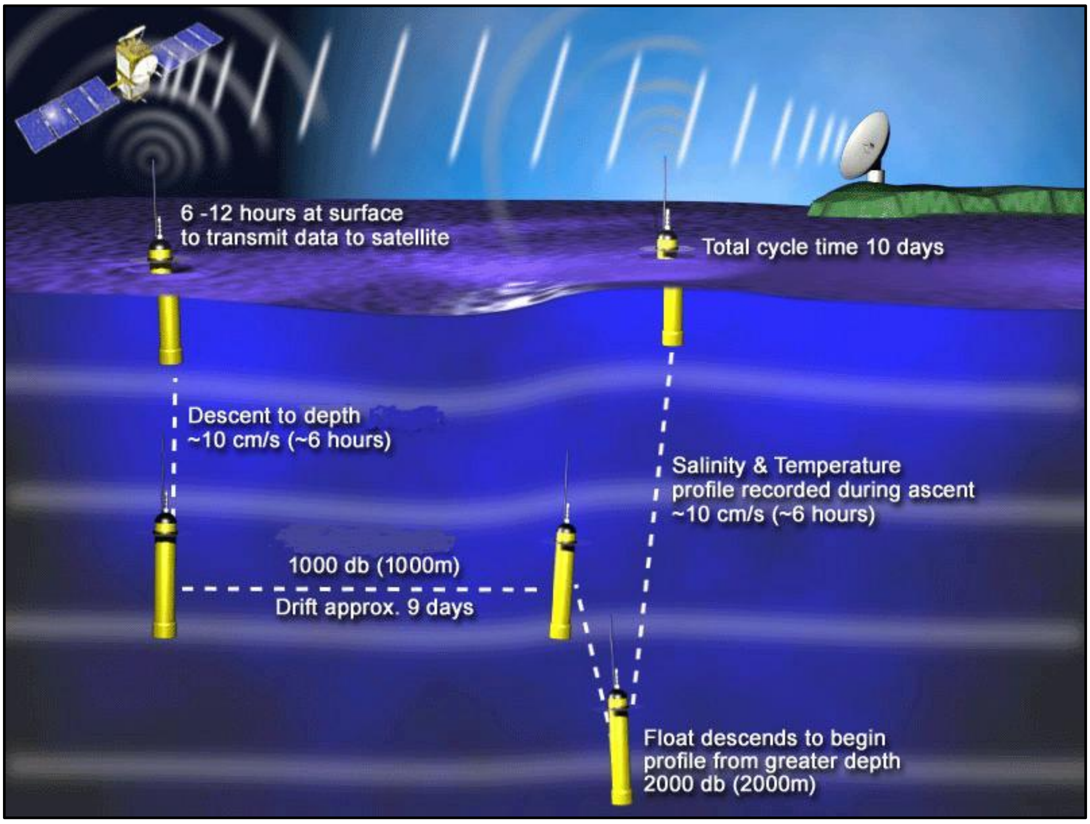


Figure 6: Argo float descends

Usually floats are released from ships on an opportunistic basis: sometimes they are launched from merchant vessels along their normal routes. In regions that are hardly visited by ships, there is a need for dedicated vessels or aircraft. Besides real-time data streaming, Argo offers data calculations which shows salinity profiles that line up with ship based data accuracy. Overall, it is not possible to do calibration checks on floats once they have left the laboratory, but it is possible to regulate and check salinity by stable, deep temperature/salinity climatology (Wong et al., 2003), or link and compare similar floats.

2.2.2 GDAC

The data collected via satellite or GPS from newly surfaced floats is sent to the Argo Information Centre (AIC) in France, before quality control is conducted at national Data Assembly Centres (DAC). Quality control consists of a real time system that conducts a set of agreed automatic checks, where errors are flagged and/or corrected (Argo 2017c). The second step in data management is a semi-automated recalibration of the salinity sensor carried out by the Pacific Marine Environmental Laboratory (PMEL). The comparison is made by sufficient temperature/salinity climatology and float data (Wong et al, 2003).

Data collected by the Argo floats is usually publicly available within 24 hours via the two Global Data Assembly Centres (GDAC) in Brest, France and Monterrey in the US, which synchronize their holdings to ensure consistency between the datasets.

2.2.3 Access to Argo data

Argo data for use by operations centres is accessed via the Global Telecommunications System (GTS), which transmits meteorological data from satellites and weather stations. For research and general purposes, the data is made available via the two GDACs discussed above. All Argo data is available in NetCDF or Ascii Comma Separated Values (.csv) format. The user’s location determines which GDAC is most appropriate, with European-based searches generally conducted through the Coriolis GDAC interface. Specific subsets of the data can be selected based on global coordinates, time period, DAC and quality (ibid).
As the Argo datasets are large and complex - four dimensions including latitude, longitude, depth and time - gridded fields are also available to aid in the analysis of certain water properties. These are produced by users (research institutes and organisations) and provide interpolated Argo data at various pressure levels and for various geographical regions (Argo 2017e).

2.3 Remote sensing

This project focuses on Argo data, as this method was already known and easy to access. However, further research was done which other methods and data set are available.
Besides the data collection via Argo floats, there is also salinity data provided through remote sensing. The measurement of ocean salinity with remote sensing uses radiometer antenna to determine the microwave emissivity and also the surface temperature. Salt dissolves in water which creates charged ions that increase the conductivity and therefore decrease the emissivity of the water (Klemas, 2001).

Radiometers can be mounted to satellites or aircraft. Satellites were first used for measuring salinity from around 2000 (Cracknell and Hayes, 2007; Lagerloef, 2000). Satellite remote sensing is a more convenient method of accumulating spatial and temporal salinity data than ocean sampling. The measurements can be done more quickly, on a denser time scale, and places which are difficult to reach even with aircraft can be observed (e.g. very northerly locations during wintertime) (Lagerloef, 2001; Le Vine et al., 2004).

2.4 Why Argo data was selected

The Argo program was selected as the data source over the remote sensing techniques for several reasons. Firstly, because Argo data is significantly more accurate than satellite and aircraft mounted radiometers. The accuracies of the data was estimated as 0.005 °C for temperature, 5 decibars for pressure, and 0.01 for salinity (Wong et al., 2003). Accuracy was important because of the relatively short time frame for the project, which emphasised the importance of accurate data to detect small changes.

Secondly, Argo data provides a complete profile of salinity and temperature over a 2000m depth profile, which provides data from the deep ocean and therefore information about the contribution from ocean currents. Data from remote sensing techniques is limited to the extreme upper surface (~2cm), and was considered outside the scope of this project because of the intention to plot the salinity and temperature profile to determine the depth at which changes occur.

Finally, Argo was selected because one of the group members was the point of contact with Argo representatives for deploying Argo floats, and therefore familiar with the project. The group member was responsible for deploying approximately 12 floats in the Arabian Sea during a naval deployment, as Argo relies on merchant mariners and naval vessels to deploy floats on an opportunistic basis.

3. Processes and Tools

3.1 Project management

This project was undertaken by a group of 4 members working together and remotely over a period of 2 months. Regular group meetings were conducted between group members and the university project supervisor to structure the report, coordinate tasks, and problem solve the analysis.

Remote group meetings were conducted via Skype to enable face to face conversations, which was beneficial compared with voice alone. File sharing for data and resources was done on Google Drive as opposed to Dropbox or Microsoft SharePoint, primarily because other content, such as the presentation and the report, could also be shared. The report was drafted on Google Docs, which enabled simultaneous editing by all group members. The report was migrated to Microsoft Word for final editing and a PDF created for final submission.

The project was managed with Agile methodology, where small goals were identified and achieved over a short timeframe. This resulted in incremental progress to review against project goals, which could be revised against the project timeframe.

A project timeline was developed (Figure 7) to provide a broad oversight of tasks, and to nominate a project manager each week who was responsible for achieving goals, communication with supervisor, and coordinating group meetings.

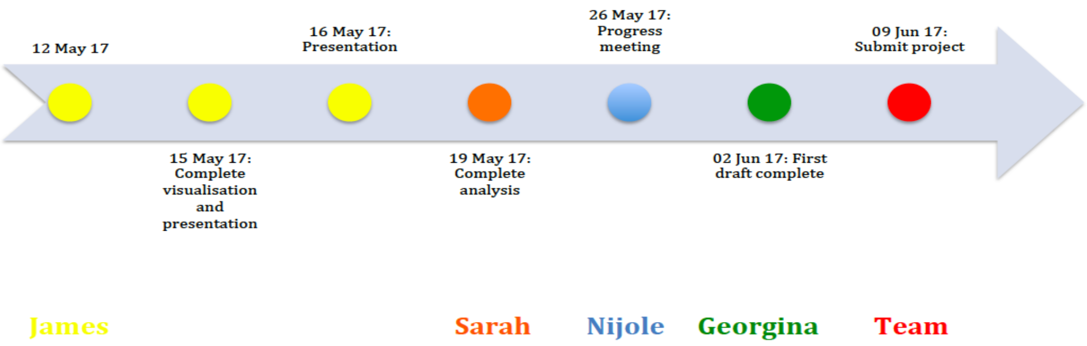


Figure 7: project management arrow

To decide how to best display the data, all group members conducted brainstorming of ideas which are illustrated at Figure 8.

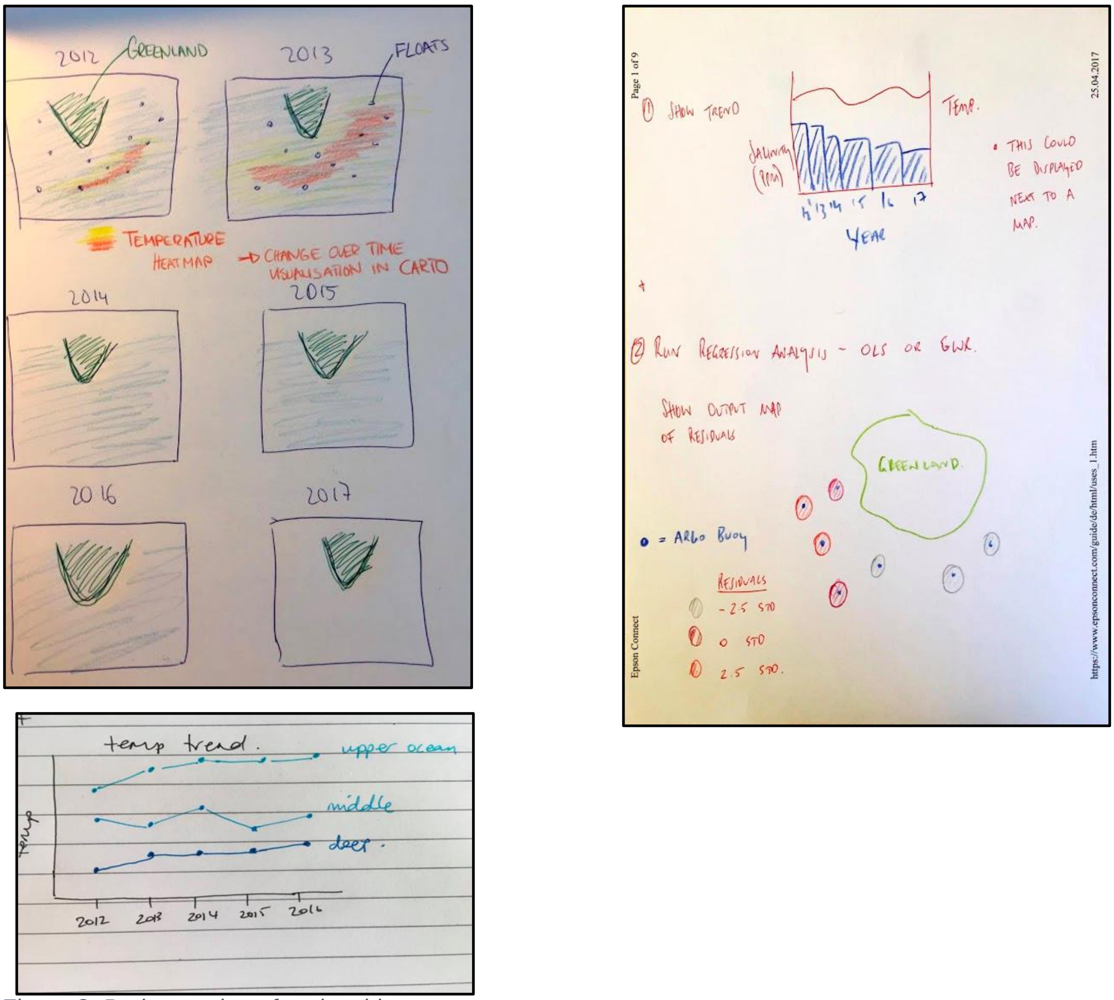


Figure 8: Brainstorming of project ideas

Different concepts were hypothesised by each group member, which was beneficial to ensure all options were considered, such as a trend graph combining salinity and temperature, temperature line graphs at different ocean depths, and interpolation of temperature data over the study period.

This process helped the project group to discuss different options and ensure that all group members shared the same vision for the project outcome.

3.2 Software

3.2.1 R

R is a free and open source programming language and environment developed for statistical analysis and modelling (The R Foundation 2017). It is a true programming language based on the statistical analysis language S, and allows users to define their own functions to extend the functionality. The core functionalities of R are: data handling and storage; calculations on arrays of data; statistical and data analysis tools; graphical display of data analysis, as well as input and output functionality.

The usefulness of R for statistical analysis of all types is extended by the availability of packages specific to the requirements of the users. There are thousands of R packages, which are simply accessed through the command line. Some examples of R packages that are used in this project are: **xlsx** (for reading, writing and formatting data stored in Microsoft Excel format); **sp** (classes and methods for spatial data); **plyr** (Tools for Splitting, Applying and Combining Data); **lubridate** (handling the data as dates) and **rgdal** (bindings for the geospatial data abstraction library) (Comprehensive R Archive Network 2017).

RStudio is a free and open source integrated development environment for the R programming language. It includes a graphical user interface that allows the user to enter command line prompts, define variables and functions and preview data visualisations (RStudio 2016). The R programming for this project was done in RStudio.

3.2.2 Python

Python is an object oriented computer programming language. In this instance Python was used for the initial handling of the Argo datasets, which were downloaded in .csv format. The script for cleaning up the data (Code 1) is written in Python to efficiently deal with missing data, missing attribute names and additional unnecessary attributes, as well as to save the large number of separate files as a single .csv.

3.2.3 ArcGIS

A Geographic Information System (GIS) is designed to store, retrieve, manage, display, and analyse all types of geographic and spatial data. Spatial data is stored as points, lines, or polygons; or as raster images. In the context of a city map, buildings could be stored as points, roads stored as lines, and boundaries stored as polygons; while aerial photos or scanned maps could be stored as raster images (Chang, 2008).

The two most popular and capable GIS software applications are ArcGIS and Quantum GIS (QGIS). Both applications are similar in terms of functionality - each can intuitively join tables, process a wide variety of data types, supports a wide variety of coordinate reference systems (CRS), and offer extensive plugins to expand capability (GIS Geography, 2016).

The major difference between the applications is that QGIS is an open source application and is free to use, whereas ArcGIS is commercial software from ESRI and requires an annual licence fee. Depending on the subscription however, ESRI provides an extensive suite of GIS services in addition to ArcGIS, such as ArcCatalog and ArcGIS for Server. Another difference is that QGIS is available on all 3 major operating systems (windows, Linux, mac), whereas ArcGIS is only available on one operating system - Windows (GIS Geography, 2016).

3.2.4 PostGIS and PostgreSQL

PostgreSQL is an open source object-relational database management system that supports Structured Query Language (SQL). It is essentially a database server, that allows users / other software applications to store, manipulate and retrieve data using queries in SQL. The data is stored in tables which are in turn stored in databases. (postgresql.org, 2016).

For the spatial databases the extender PostGIS is used. It adds spatial functions (distance, area, union, intersection, and other spatial geometry data types) to the database and enables location based calculations in SQL.

As a PostgreSQL development platform we used pgAdmin III. It is running on different platforms like macOS, Linux and Windows and it is freely available. PgAdmin can be used for simple SQL but also to create complex databases (pgadmin.org, 2016).

3.3 Processes

3.3.1 Interpolation

Interpolation is the process of using points with known values to estimate values at other points (Chang, 2008). It is typically conducted using a raster, and is therefore a means of creating surface data from sample points to conduct analysis and modelling. Creating a continuous surface enables trends in a data set to be observed, and prepares the data for analysis, such as trend analysis (ArcGIS for Desktop, 2017). It requires 2 basic inputs – known points and an interpolation method.

Known points, or control points, are important because the number and distribution can greatly influence the accuracy of interpolation. A general assumption is that values closer to the point to be estimated is more influential than values further away. Control points should be widely and evenly distributed as possible to maximise the accuracy of estimation.

Global		*Local*	
Deterministic	**Stochastic**	**Deterministic**	**Stochastic**
Trend surface (inexact)*	Regression (inexact)	Thiessen (exact)	Kriging (exact)
		Density estimation (inexact)	
		Inverse distance weighted (exact)	
		Splines (exact)	

Figure 9: Interpolation methods

As demonstrated at Figure 9, Interpolation methods can be categorised in several ways. Firstly and most broadly, they can be classified as global or local. Global interpolation methods use every point available to estimate an unknown value, whereas local methods only use a sample of known points. Global methods should be used to capture large scale trends and local methods to determine local or small range variation.

Secondly, interpolation methods can be categorised as exact and inexact. Exact interpolation generates a surface that passes through control points and allocates the exact value of the surface that passes through that control point. Inexact on the other hand predicts a value at a point location that differs from its known value. The exact method produces a surface that is more rigid and displays small scale variation, whereas the inexact produces a smoother surface. The decision about which to use depends on size of data set and viewing requirements.

Lastly, interpolation methods may be deterministic or stochastic. Deterministic methods assign values to unknown points using the surrounding values and specific formulas that determine the smoothness of the resulting surface but provide no assessment of errors with predicted values, whereas stochastic methods incorporate autocorrelation which provide an assessment of prediction errors and estimated variances (ArcGIS for Desktop, 2017; Chang, 2008). The decision about which to use therefore depends on whether analysis, and what type of analysis, is required. Additionally, a condition of stochastic models is it assumes a random process to generate data points.

3.3.2 Regression analysis

Regression analysis enables spatial relationships to be modelled and explored to better understand the factors behind observed patterns (ArcGIS for Desktop, 2017). Several analysis methods are available, depending on whether the data is influenced by global or local factors. The most commonly used are Ordinary Least Squares (OLS) and

Geographically Weighted Regression (GWR), which are only suitable for data with linear relationships.

OLS is a common regression technique that should be the starting point for all spatial regression analyses. It creates predictions for a global model or process by creating a single regression equation to represent all data points (ArcGIS for Desktop, 2017), and therefore poorly accounts for local variation processes. GWR however, uses the information for each known point to fit a regression equation to every feature in the dataset and create a local model (ArcGIS for Desktop, 2017).

A linear regression involves comparing one dependent variable and one explanatory variable, whereas multiple regression analyses more than one explanatory variable. Causation cannot be inferred through a regression analysis (de Smith et al 2015), but the validity of the model can be checked using the in-built statistical outputs in ArcGIS. These six checks are:

1. Are the explanatory variables statistically significant?
2. Are the relationships as expected?
3. Are any explanatory variables redundant?
4. Is the model biased - are the residuals normally distributed?
5. Is there spatial autocorrelation in the model residuals?
6. Is the R^2 value sufficient?

3.3.3 Visualisation

Advances in computer technology, scientific theory and available data are enabling the use of 3D visualisation in new and exciting ways (Wood et al, 2005). However, with these advancements come challenges to ensure 3D visualisations remain usable by ensuring visual and navigational realism are used effectively.

When compared with 2D visualisations, 3D visualisations create more dynamic visualisations which enable users to observe phenomena more closely to real life and thereby potentially gain a better understanding of the visualisation. A 3D object can be interrogated in a variety of ways, such as rotating and zooming a model to observe various trends and phenomenon from different perspectives, whereas a 2D model can only be observed from top down which limits the amount of information.

3D models are generally more appealing visually than 2D models, which can be an advantage to draw attention to a display, such as in a public setting.

The visualisation process almost always involves a simplification and abstraction of information (Wood et al, 2005). While this practice is sufficiently widespread and understood by users, 3D visualisation in a 2D display can help reduce this simplification by adding a cartographic ‘degree of freedom’ in the z dimension using perspective, occlusion and parallax motion (Wood et al, 2005). An example of this is adding time to a map display.
With the emergence of 3D printing, 3D visualisations enable projection of information away from a display to an actual object that can be handled and examined by users.

Effective 3D visualisation is dependent on the developments in 3D technologies such as hardware, software and rendering (Wood et al, 2005), which somewhat limits the progress 3D can be developed and is made usable.

Comparing 2D and 3D, usability is the key issue. While advancements in graphic displays and computer processing improve 3D accessibility, people are more familiar with traditional 2D maps and are more comfortable to draw information from them. 3D visualisations require sophisticated software and graphical displays for viewing, which are not accessible for all people. 3D visualisations can be printed on a 2D surface, however the ability to rotate and interrogate the model is lost and therefore loses some information. Finally Software and hardware to create 3D visualisations are more expensive than for creating 2D.

4. Data

4.1 Acquisition

The Argo data used in this project was downloaded from Coriolis.eu.org, the data distribution website for the Argo data in Europe. As discussed in Chapter 1, this report focuses on the North Atlantic ocean and Labrador Sea to the south of Greenland. This area is selected with a bounding box to ensure the same data frame for each dataset.

The data encompassing five years of temperature and salinity fluctuations is too much to be delivered as one data set, so it had to be downloaded in smaller parts. Figure 10 shows a screenshot of the data selection process for the first data set.

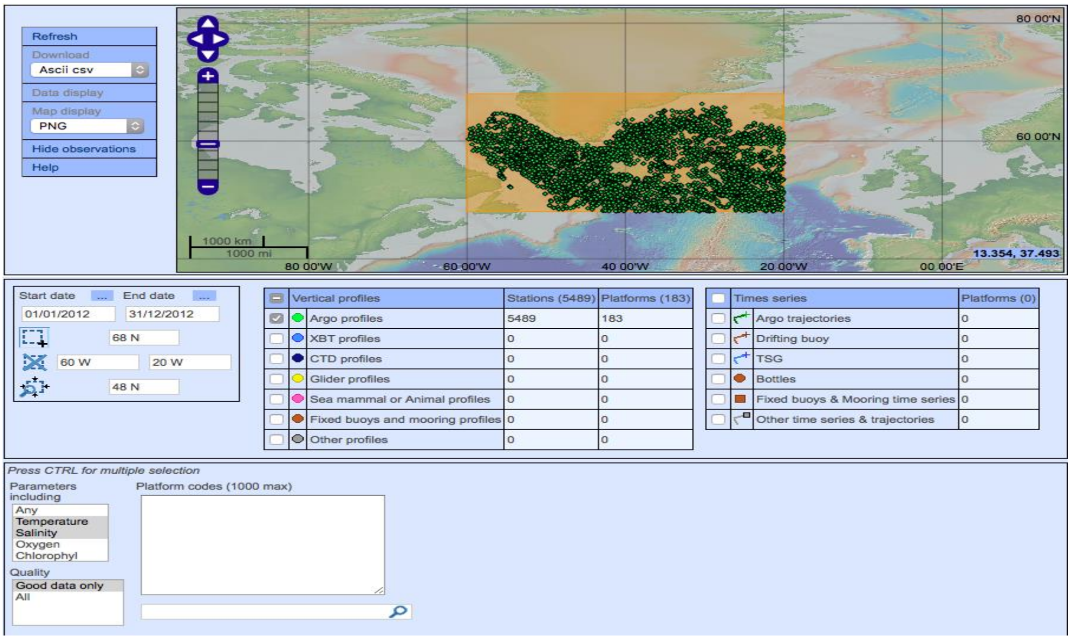


Figure 10: Screenshot Coriolis.eu.org

The data is delivered in separate .csv files for each platform which includes the data for the selected time period and the measurements of temperature and salinity for all the depth levels until a depth of around 2000 meters.

As the bounding box was set, but the platforms are untethered, some platforms drifted out of the selected frame or came into it at a later point of the selected years. This led to a varying number of platforms and measurements over the study period. However, this is not considered to be a problem considering the relatively high density of floats. To analyse the data, one data sheet is needed, which combines all platforms over all years.

4.2 Issues with the data

In selecting data on the Coriolis data centre website, the option to only include ‘good’ data’, meaning data that has been quality checked to some extent, rather than real-time data, is possible. Despite this type of data being used in this project, there were a number of anomalies in the datasets we received from the data centre. Argo floats are unmonitored after launch, because they are free-drifting and communicate data via satellite or GPS (as outlined in Chapter 2). Floats that might have problems with one or more of their sensors are ‘grey listed’, and their data is not considered reliable or included in the packages sent to the GTS for operational use in various countries. It is possible that some unreliable data slips through and is included in the ‘good’ data when selected.

The issues that were encountered with the data include: missing column headings - or put another way, more data than columns; missing data - cells with no value; and certain platforms that had missing columns, which meant that the following data was misaligned with the rest of the dataset. These issues were not only of concern when performing data analysis and interpretation, but also caused problems in the initial data processing and handling stage.

4.3 Clean up

The Argo data is provided in .csv files organised and separated by platforms. These files contain temperature and salinity measurements for every few meters below sea level, reported approximately every ten days. The download of float data for each year of our study period resulted in 766 .csv files. Along with attribute measurements for salinity, depth and temperature, additional information is also provided by Argo relating to how the data was collected and how it was processed afterwards. Within the context of this project, the data was taken as it is and no further examination of the processing behind the data was done. Therefore, the columns with post-processing data were deemed unnecessary and were dropped in the data cleaning process.

Due to the type of data anomalies found in the datasets downloaded from Coriolis, reading the .csv files into R-studio resulted in various errors. We therefore used Python to do the initial reading of files, selecting of relevant columns and saving to a new combined .csv file.

This code is shown in Code 1. The Python script first defines which variables are required for analysis (line 4). For simplicity, the first eight columns were chosen, despite including some irrelevant information, for example ‘Argos_ID’. The first eight columns contain data for: platform, Argo ID, date, latitude, longitude, pressure, temperature and salinity.

Each file in the project folder containing the Argo .csv datasets was read (line 14-16), header names were ignored (lines 24-25) and only the required columns were copied to the new combined .csv (lines 30-35).

```
1   import .csv
2   import glob
3
4   required_cols = ['PLATFORM', 'ARGOS_ID', 'DATE (YYYY-MM-DDTHH:MI:SSZ)', 'LATITUDE (degree_north)', 'LONGITUDE
5   (degree_east)', 'PRES (decibar)', 'TEMP (degree_Celsius)', 'PSAL (psu)']
6
7   out_file_name = 'out_file..csv'
8
9   # First open a file for writing the results as .csv
10  with open(out_file_name, 'a') as outfile:
11          out_.csv = .csv.DictWriter(outfile, fieldnames=required_cols)
12          out_.csv.writeheader()
13
14          # Go through each file in the current directory ending on ..csv
15          for file in glob.glob("*..csv"):
16          with open(file, 'rb') as .csvfile:
17
18             # Open the next file and only get the columns with names 'required_cols'
19          .csv_file = .csv.DictReader(.csvfile, fieldnames=required_cols)
20
21          # See what file we're up to
22          print(file)
23
24          # Remove the header of current file
25          .csv_file.next()
26
27          # Go through each row in the current file
28          for item in .csv_file:
29
30          # Remove the items in the row not in 'required_cols'
31          item.pop(None)
32
33          # Now let's write the row of values to the new .csv file
34          try:
35          out_.csv.writerow(item)
36          except Exception as ex:
37          # If something goes wrong print the error
38          print(ex)
39
```

Code 1: Python

The programming language R was used for the majority of the data analysis and processing for the project. The secondary clean up of the data was also done in R and can be seen in Code 2.

Firstly, the unnecessary ‘Argo ID’ column was dropped because the platform number was deemed sufficient for identification purposes. Then the column headers were renamed for ease of use and readability. After examining the data in ArcGIS a few false values were discovered. The identified nine platforms with missing columns and unusable data were deleted with the subset function. A new subset is created where the platform ID does not match one of the outliers (line 3-5). After deleting all rows with these platform numbers, there were still some missing values in the data. These are detected with *complete.cases* and then omitted with *na.omit* (lines 7-9). The date column was converted to *date* format and the month and year information was added to a new column to aid in temporal subsetting of the data.

```
R code

Argodata <- read..csv("/Users/sarah/Documents/project2/out_file..csv")

# deleting ARGO_ID column
drops <- c("ARGOS_ID")
floats <- Argodata[ , !(names(Argodata) %in% drops)]

#rename columns
library(plyr)
AD <- rename(Argodata, c("PLATFORM"="ID", "DATE..YYYY.MM.DDTHH.MI.SSZ."="DATE",
        "LATITUDE..degree_north."="LAT", "LONGITUDE..degree_east."="LONG",
        "PRES..decibar."="PRES", "TEMP..degree_Celsius."="TEMP", "PSAL..psu."="PSAL"))


# deleting outlier platforms
sub1 <- subset(AD, ID != 6901762)
sub2 <- subset(sub1, ID != 6901760)
sub3 <- subset(sub2, ID != 6900614)
sub4 <- subset(sub3, ID != 6900567)
sub5 <- subset(sub4, ID != 4900778)
sub6 <- subset(sub5, ID != 4901162)
sub7 <- subset(sub6, ID != 6901627)
sub8 <- subset(sub7, ID != 6902661)
sub9 <- subset(sub8, ID != 6900563)

# deleting missing values
CleanD <- na.omit(sub9)

#  Convert to date
CleanD$DATE <- as.Date(CleanD$DATE)

#  Get months
CleanD$Month <- months(CleanD$DATE)

#  Get years
CleanD$Year <- format(CleanD$DATE,format="%y")
```

Code 2: R - data clean up

The final step in the initial data clean up was transforming the dataset into a projected coordinate reference system (CRS). An appropriate projection for the geographic area with meters as the unit of measure was chosen (UTM 3183, 23N), and the data transformed. This transformation was done using the rgdal library and is shown in Code 3. Firstly, the columns with the longitude and latitude were defined and a list with the existing EPSGs was created. The original CRS has to be set in order to transform it into the required projection. So, the latitude and longitude coordinates are set to WGS84 and then transformed into 3183. The right projection format was copied from epsg.io (epsg.io, 2017).

```
library(rgdal)

# setting coordinates
coordinates(final) <- c("LONG", "LAT")

# creating a list containing all EPSG
EPSG <- make_EPSG()

# setting the CRS to 4326 and transforming it to 3183 afterwards
proj4string(final) <- CRS("+proj=longlat +datum=WGS84 +no_defs")
finalCRS <- spTransform(final, CRS("+proj=utm +zone=23 +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +units=m +no_defs"))
```

Code 3: R - projections

4.4 Subsetting

The Argo data is not only very extensive but also complex in itself. In addition to the location of the floats on the x and y axes according to the coordinates there are measurements taken in different depths, which can be shown on a z axis. Furthermore, this report is looking at a change over a period of five years, which means there is also a fourth dimension - time. After the cleaning process there were still 4.259.338 rows in the data frame. To make it possible to process and compare the data, subsets were required.

The processing and plotting of this amount of data is time consuming and this project focuses on the surface rather on the deep ocean. Therefore, the measurements from the topmost 200m were subsetted based on the maximum possible definition of sea surface. Therefore a new data frame is created using only columns which have pressure measurements below 200 meters.

```
# selecting the depth
first200 <- CleanD[which(CleanD$PRES < 200), ]
```

Code 4: R - Depth selection

To reduce it even further salinity and a pressure subsets of 200 meters were plotted for every season to see where change stopped and the salinity data gets more consistent. In order to do so, the 200 meters dataset was again subsetted into different seasons. For an even more accurate way, some months were subsetted and plotted to see the needed depth level. The day value is set as a short form for the date column of the data frame. Afterwards, each season is defined with the operations < and > applied to the start and end day of the each season.

```
# season subsets Summer (1.March-31.August) and Winter (1.September-29.February)
day <- first200CRS$DATE

S12 <- first200CRS[day > "2012-03-01" & day < "2012-08-31", ]
```

```
S13 <- first200CRS[day > "2013-03-01" & day < "2013-08-31", ]
S14 <- first200CRS[day > "2014-03-01" & day < "2014-08-31", ]
S15 <- first200CRS[day > "2015-03-01" & day < "2015-08-31", ]
S16 <- first200CRS[day > "2016-03-01" & day < "2016-08-31", ]

W12 <- first200CRS[day > "2012-01-01" & day < "2012-02-29", ]
W1213 <- first200CRS[day > "2012-09-01" & day < "2013-02-28", ]
W1314 <- first200CRS[day > "2013-09-01" & day < "2014-02-28", ]
W1415 <- first200CRS[day > "2014-09-01" & day < "2015-02-28", ]
W1516 <- first200CRS[day > "2015-09-01" & day < "2016-02-29", ]
W16 <- first200CRS[day > "2016-09-01" & day < "2016-12-31", ]
```

Code 5: R - Season Subset

For the different interpolations and following analyses and plots, different subsets are needed. The process for each subset is similar to the two shown above. One for selecting the depth and the other functions to subset seasons.

The created data frames were written and exported into .csv files in order to display and process them in ArcGIS. This is done, using the xlsx library.

```
write..csv(W12, file = "/Users/sarah/Desktop/W12..csv")
```

Code 6: R - export .csv

5. Data Analysis Results

5.1 Exploration

5.1.1 Trend analysis

The first stage of the project was data exploration where the whole data was plotted in R to determine any broad trends (Figure 11).
This calculation was done in R with the code below and written into a .csv file.

```
# calculating the mean for each month
psalMean <- aggregate( PSAL ~ Month + Year , CleanD , mean )
```

Code 7: R - Salinity mean value

The salinity averages for each month were plotted against time and a linear trend line was fitted.

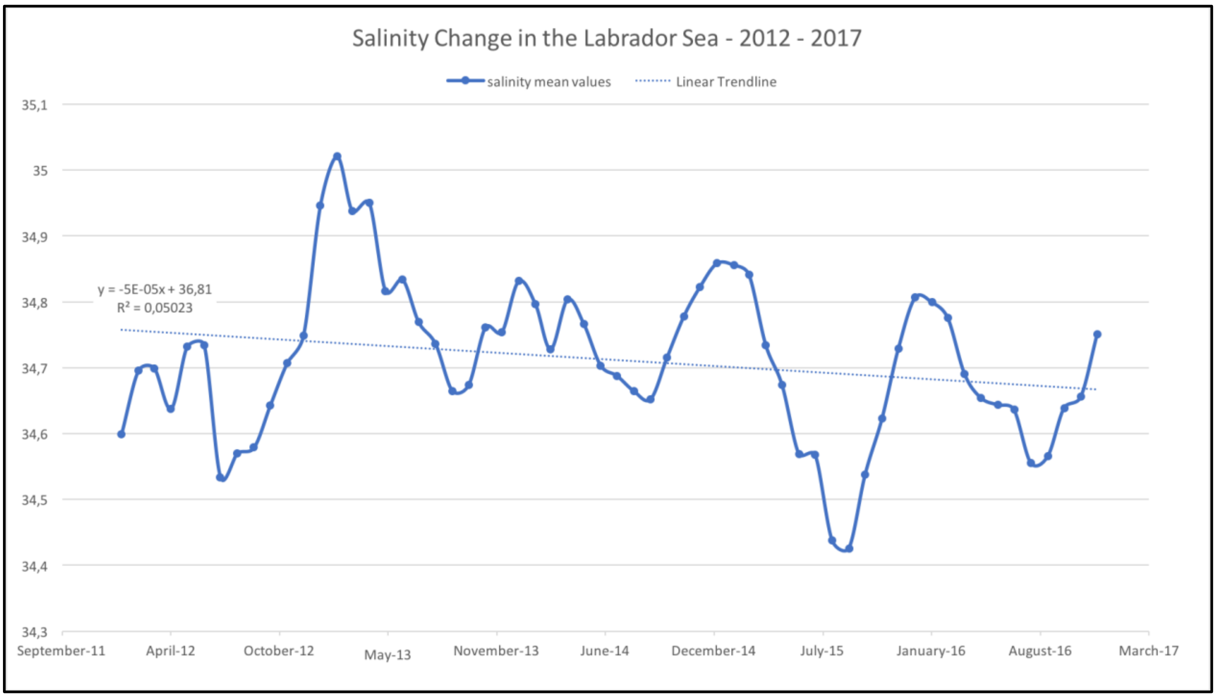


Figure 11: Salinity change in the Labrador Sea, 2012-2017

It can be observed that mean salinity decreased approximately 1 psu between January 2012 to December 2016. R^2 is 0.05, which indicates the trend line is a poor fit to the data. Salinity fluctuated from a minimum of 34.42 psu to 35.02 psu.

After this first plot of salinity over time, another dimension was observed, the change based on location. Using the season subsets for the first 200 meters, the floats were plotted based on their latitude and longitude coordinates. To see the difference between the salinity measurements, a colour scale was used.

At the plots on the next page, it can be seen that the latitude and longitude coordinates are not transformed. New subsets for the season without the transformation had to be done, because the used library, plotly, can not plot spatial points data frames, which would have been the transformed subsets. However, for the aim of these plots it is not necessary to have the other projection, because no metric calculations were done.

To create a facet plot based on the seasons a new column with seasons was added. Therefore, the mutate function was used to create a new column. It contains *if/else* statements with the wanted month for each season. The *%in%* can be translated to "contained with", so the syntax means that if the month column contains one of the given months then it gets the season assigned. An *error* value will be assigned, if a month column does not fall within any of the given criteria (Jones et al., 2017)

```
#adding season column
t20 <- t20 %>%
  mutate(season =
       ifelse(Month %in% c("September", "October", "November", "December", "January", "February"), "Winter",
       ifelse(Month %in% c("March", "April", "May", "June", "July", "August"), "Summer", "Error")))

# facet plot
PSY <- ggplot(t20, aes(LONG, LAT, color=PSAL, fill=PSAL, scale_color_continuous(guide = guide_legend(reverse=TRUE))), size = 0.4) +
  +  geom_point() +
  +  ggtitle("Salinity Change in the Labrador Sea") +
  +  xlab("Longitude") + ylab("Latitude")

PSY + facet_grid(Year ~ season, scales = "free", space = "free")
```

Code 8: R - adding season column and plotting salinity on latitude/longitude

The plots below were generated in R in order to do a basic spatio-temporal analysis of salinity in the study area. The plots show how the salinity measurements in the two predominant seasons over the five year study period changed. The salinity values appear to be lower with proximity to the west, and higher nearer to the east.

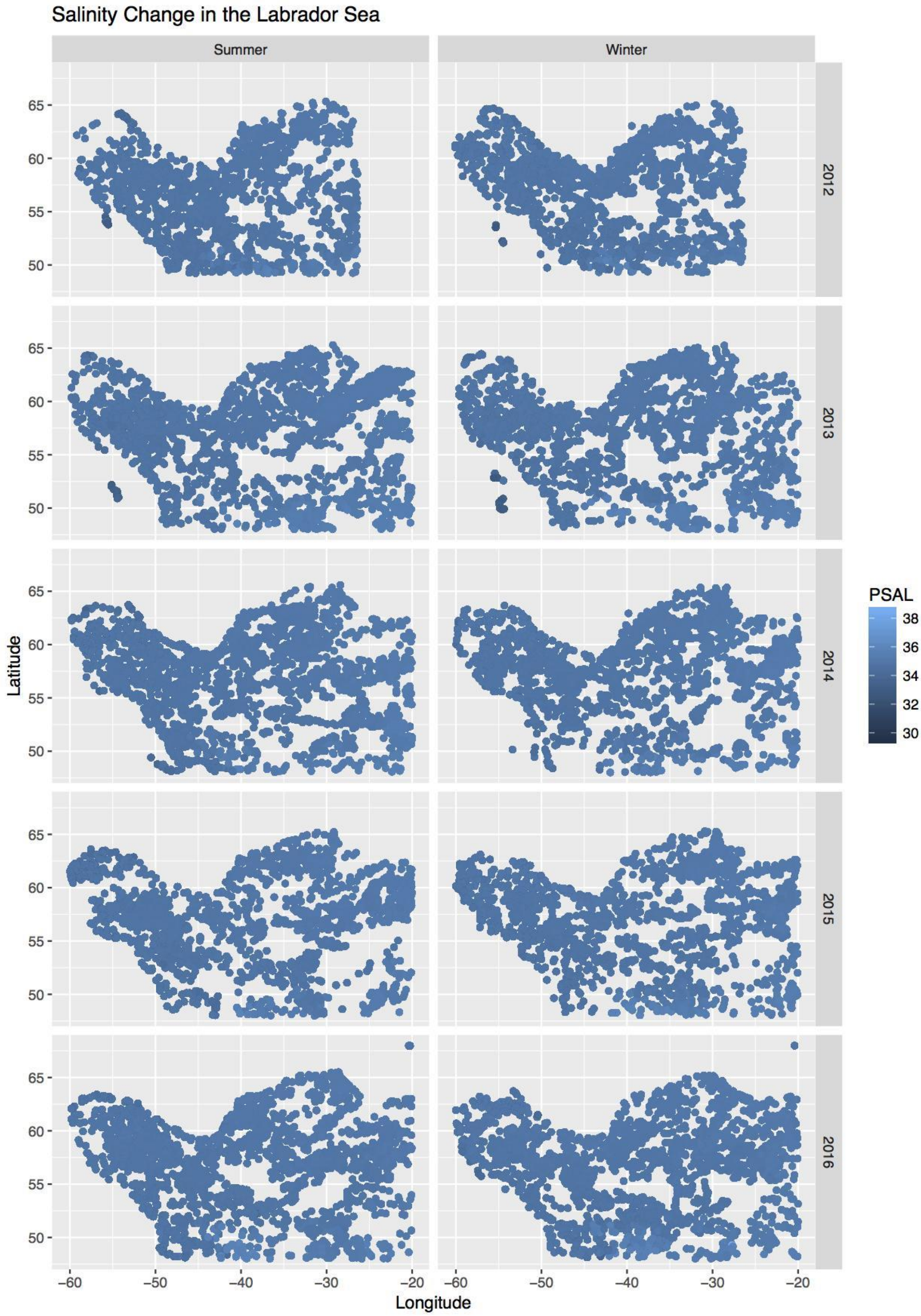


Figure 12: Salinity scatter plots demonstrate variation with longitude

5.1.2 Surface determination

The next step was to determine whether any correlation between salinity and depth was present. Data was grouped into seasons and scatter plots of the raw data were undertaken in R (Figures 13-16). Based on the literature a depth of 200 meters was taken, to reduce the processing.

The season's subsets were used to plot them separately. First the columns of the season dataframe "S12" for the x and y axes are selected. The $ selects the wanted column of the data frame. With "main", "xlab" and "ylab" the main title and labels for the axes are set. With "pch=16" a dot as a symbol is chosen and with "cex" the size is set. The same syntax was used for all seasons and salinity and repeated for temperature.

```
# plotting salinity and depth
plot(S12$PRES, S12$TEMP, main="Summer 2012 - Temperature", xlab="Depth (Meter)", ylab="Temperature (Celsius)", pch=16, cex=0.7)
```

Code 9: R - salinity / depth plots

A broad trend can be observed of greater salinity range at shallow depths, which is expected due to surface freshwater input from rivers, ice melt and precipitation, that decreases sea surface salinity. But it is not possible to determine an obvious cut-off depth, where the salinity stops changing. However, it can be observed that the widest ranges and therefore biggest changes appear in the first 20 to 50 meters. Especially in Winter the range gets more narrow after 20 rather 50 meters. An anomaly in the data can be seen in Winter 2016, as we do not look at outliers due to the limitations of the report, this will be deleted.

Change of Salinity by depth

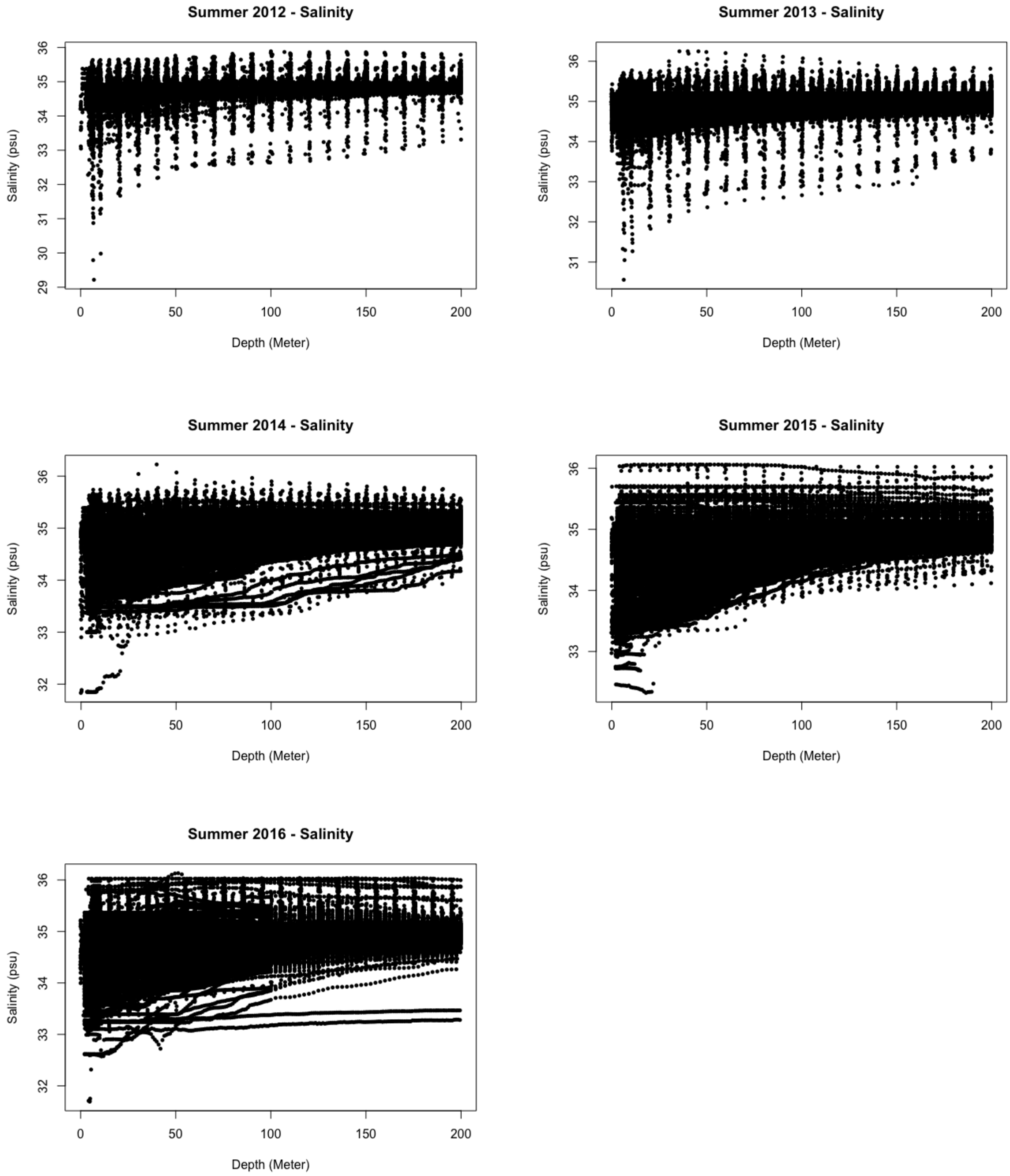


Figure 13: Salinity / Depth, Summer plots

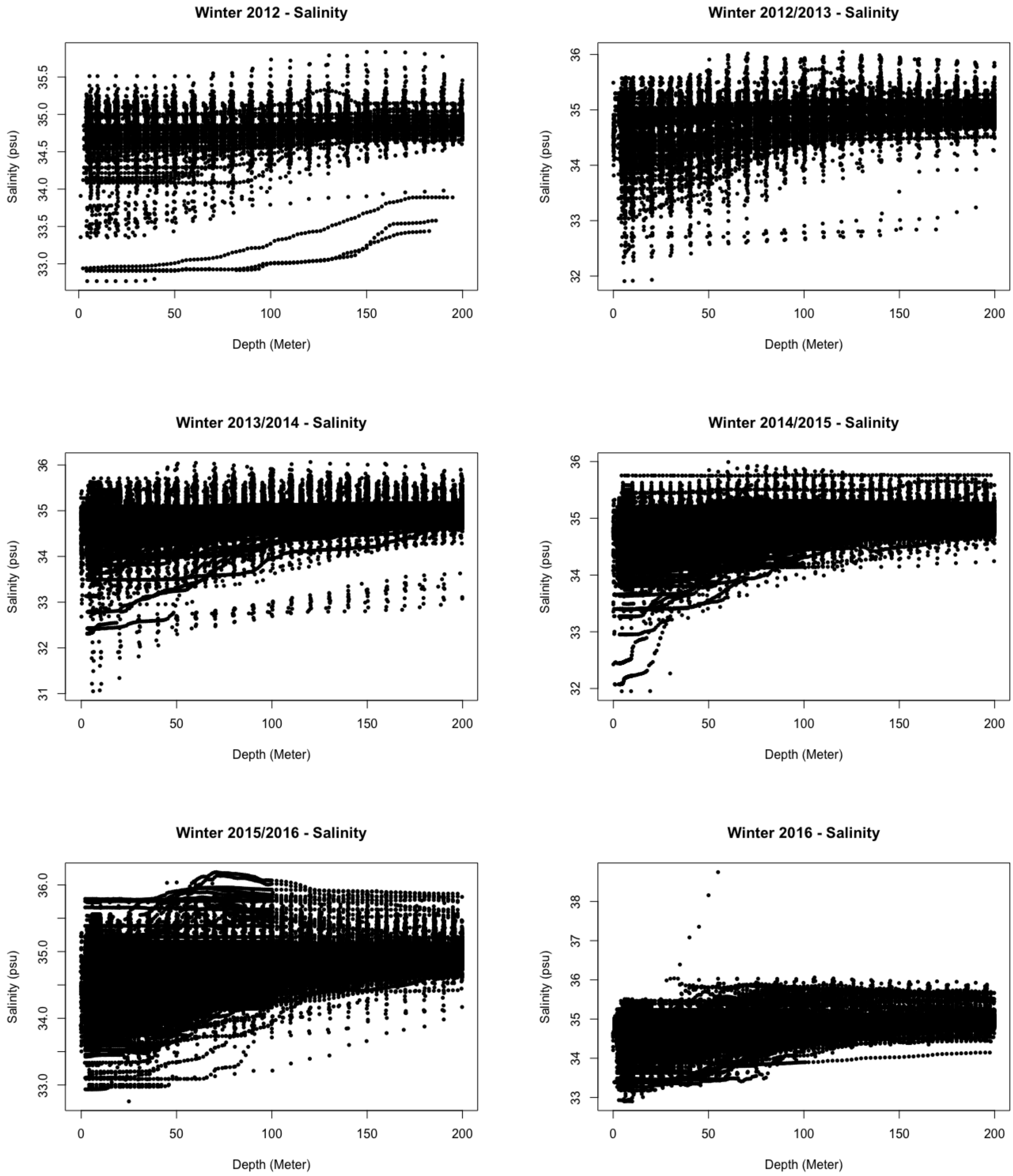


Figure 14: Salinity / Depth, Winter plots

Change of temperature by depth

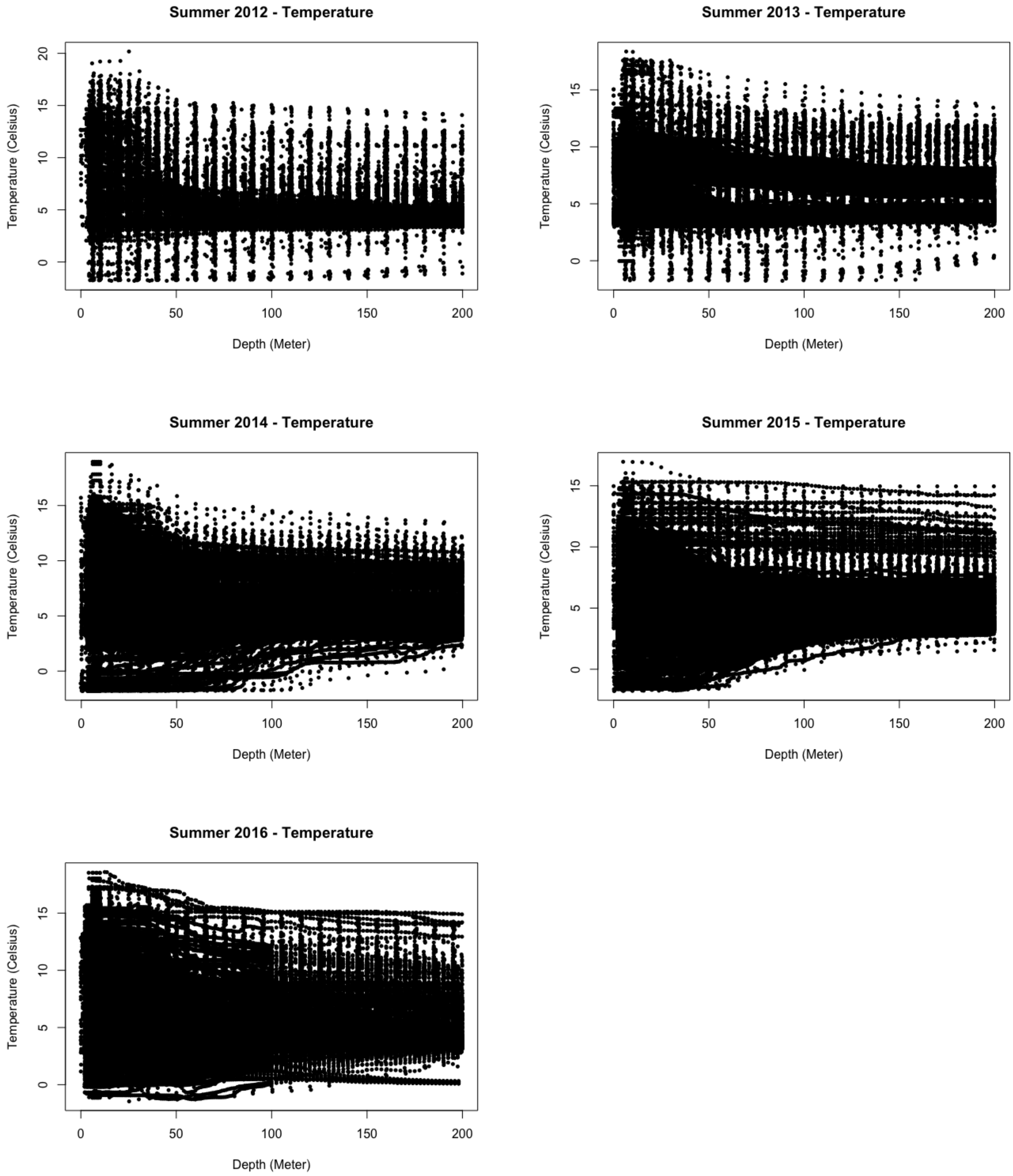


Figure 15: Temperature / Depth, Summer plots

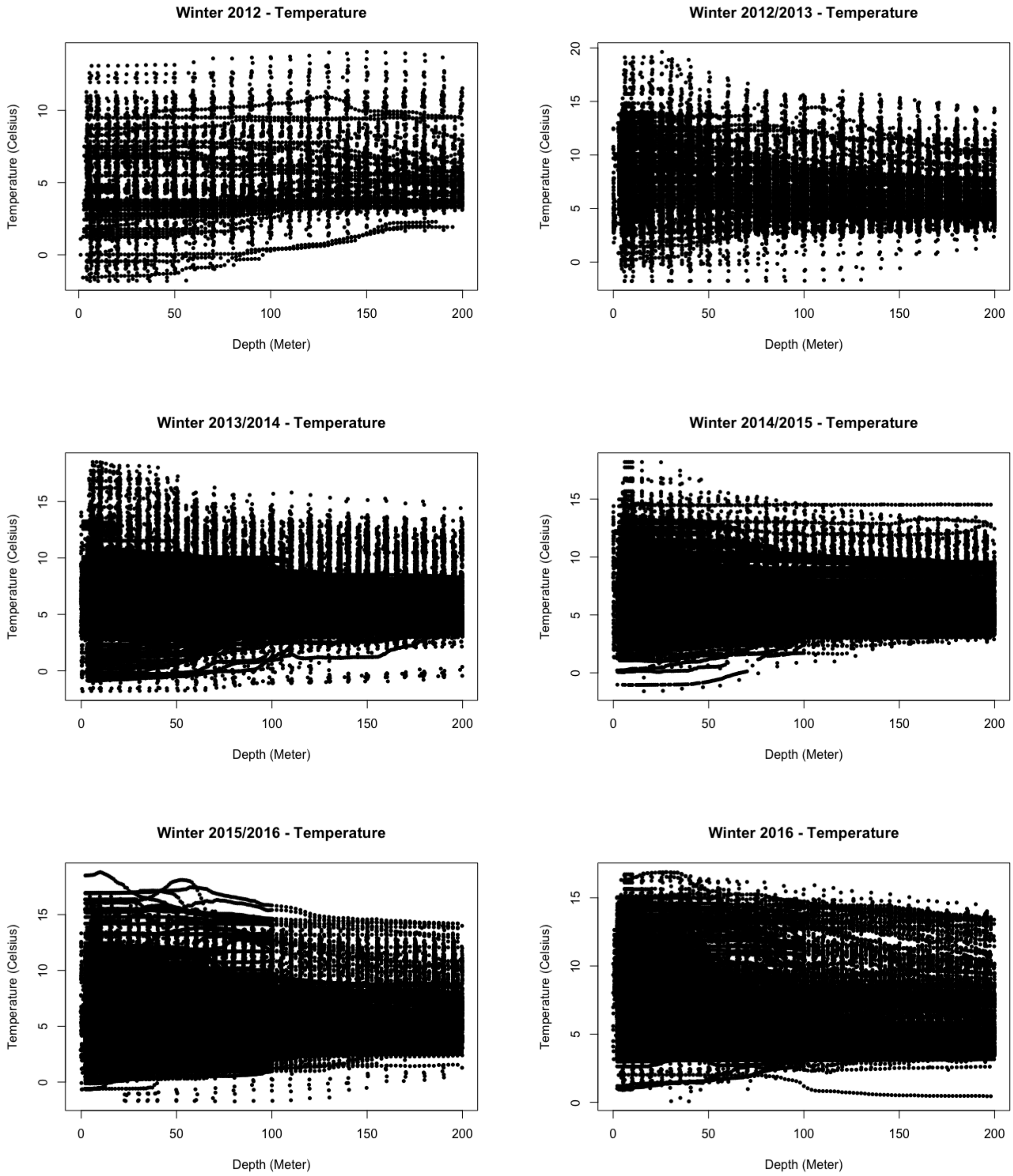


Figure 16: Temperature / Depth, Winter plots

To determine a useful depth a look into the change of temperature by depth was done in the same way as salinity. The scatter plots were again based on the same seasons. Similar observations can be made, although a wider range of temperatures can be observed at slightly greater depths. Overall, it can be seen that both salinity and temperature range decreases after 20 meters. Accordingly, the following analyses are based on a surface subset of 20 meters.

As the seasons have a bigger range due to the longer time period, plots were also conducted for salinity and temperature over a single month. As can be seen below the broad scale trends as explained for the seasons are also occurring over a smaller timeframe.

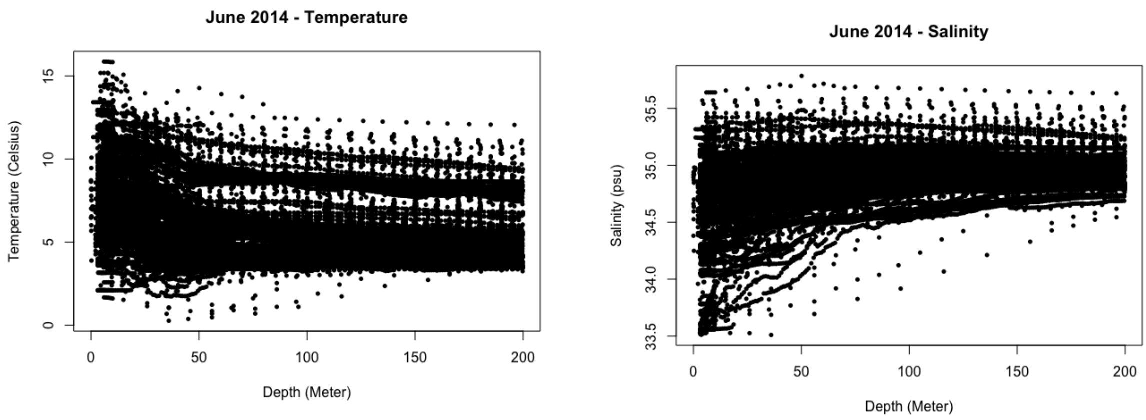


Figure 17: Salinity and Temperature / Depth, June

Based on the scatter plots for salinity, the trend analysis to determine whether salinity changed was modified to include only the upper 20m water profile (Figure 18).

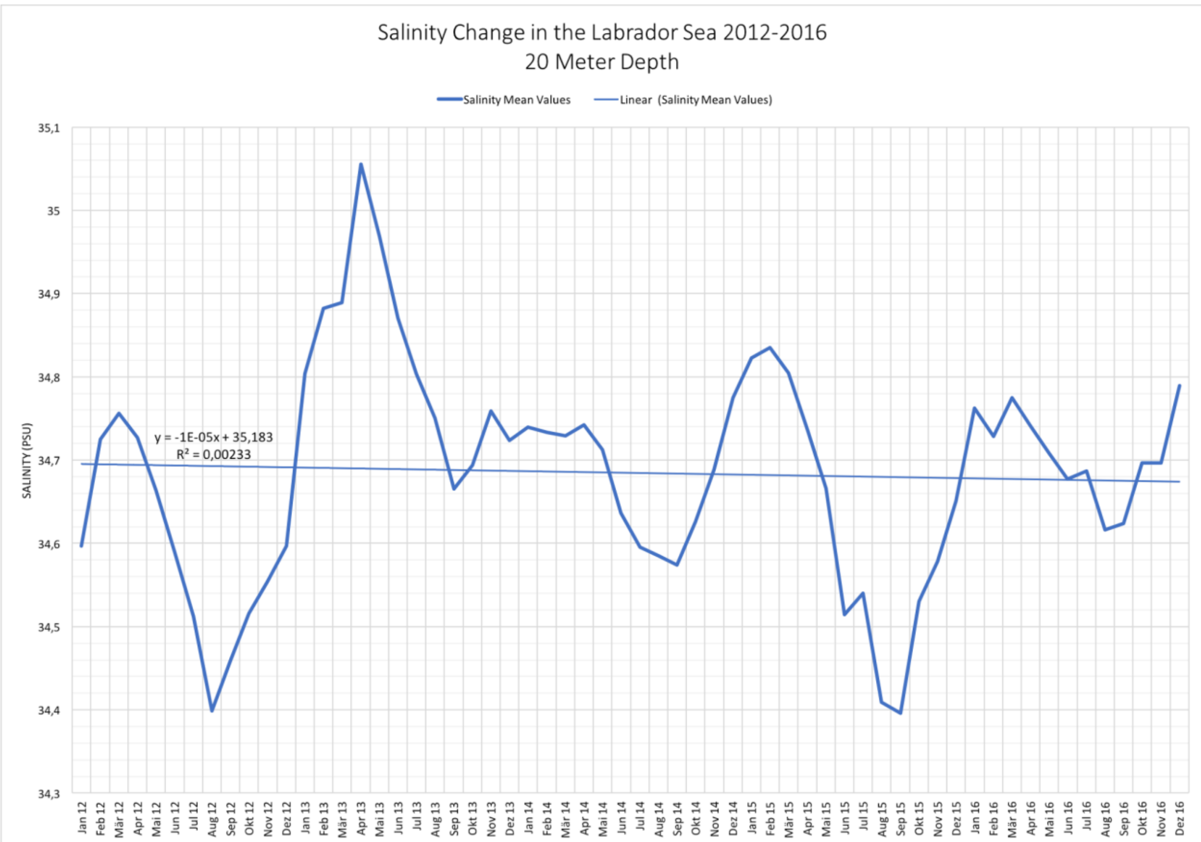


Figure 18: Salinity Change in the Labrador Sea 2012-2016, 20 m

Compared with the graph for the entire water profile, the following can be observed from the upper 20m:

- salinity varies more
- the trend line shows a decrease in salinity, but not as significant a reduction
- R^2 (0.005) has reduced

5.2 Interpolation and visualisations

5.2.1 Interpolation method

Interpolation was conducted to visualise salinity changes over the 5 year study period. The choice of interpolation method was conducted using interpolation hierarchies (see Figure 9). It was determined that a local method of interpolation would be used because ocean salinity and temperature are only influenced by points in the near vicinity. Global interpolation uses every point available, which is unsuitable for interpolating temperature and salinity variation on a small scale.

It was determined that both exact and inexact interpolation methods would be suitable, because although exact interpolation forces the generated surface to pass through the predicted points which can result in sharp peaks and troughs (ArcGIS, 2017a), the scale of the dataset would minimise this effect and was suitable for the intended display purposes. Inexact interpolation methods would also be suitable, and possibly preferred over exact methods, because the model is not forced to pass though predicted points and therefore results in a smoother surface which is suitable for displaying depths.

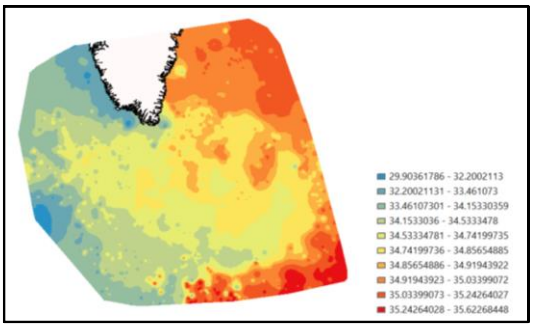
Finally, it was determined that deterministic interpolation methods would be used instead of stochastic, because stochastic methods assume a random process is generating the point data, which in the case of salinity and temperature is not appropriate because both are determined by established processes such as evaporation and precipitation, which are not random processes.

Through this process of elimination, it was determined that Inverse Distance Weighted (IDW) interpolation would be suitable. The major assumption of this method is that things that are close to one another are more alike than those that are farther apart, and it therefore gives greater weights to points closest to the prediction location and the weights diminish as a function of distance, hence the name inverse distance weighted (ArcGIS 10.5 Help). This is consistent with ocean salinity and temperature, as measurements near a point are more likely to be accurate than measurements further away. IDW also assumes that phenomenon being modelled are driven by local variation (ArcGIS 10.5 Help), which is consistent with hypothesise for factors influencing salinity and temperature in the Labrador Sea.

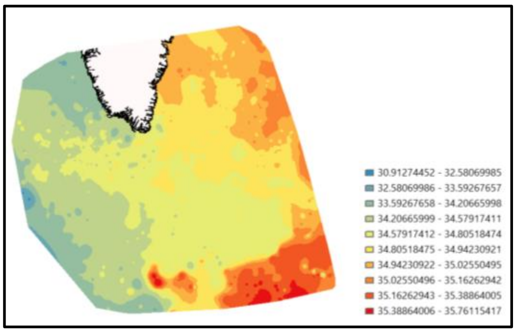
The interpolation methods that are used to generate a surface give the best results if the data is normally distributed (a bell-shaped curve). If data is skewed (lopsided), you might choose to transform the data to make it normal. Thus, it is important to understand the distribution of your data before creating a surface.

5.2.2 Inverse Distance Weighted interpolation

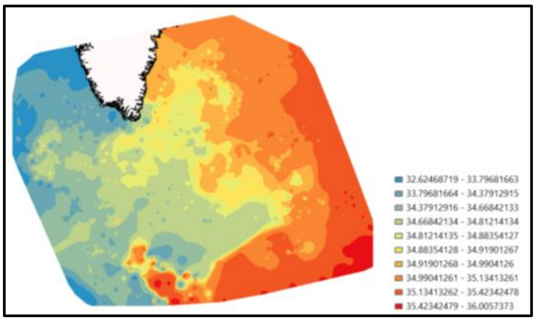
IDW interpolation was used to display seasonal Salinity changes from 2012 Spring/Summer season to 2016 Spring/Summer (Figure 19) and 2012 - 2016 Winter/Autumn season (Figure 20). Interpolation provides visual information about salinity quantity in our research area.



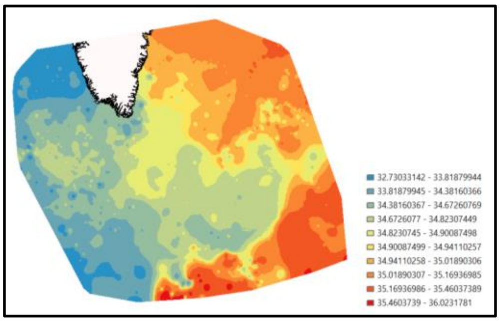
Spring/Summer 2012



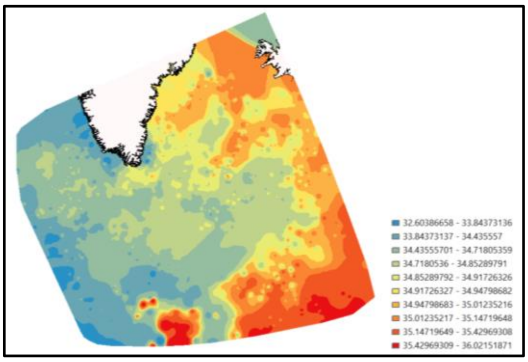
Spring/Summer 2013



Spring/Summer 2014

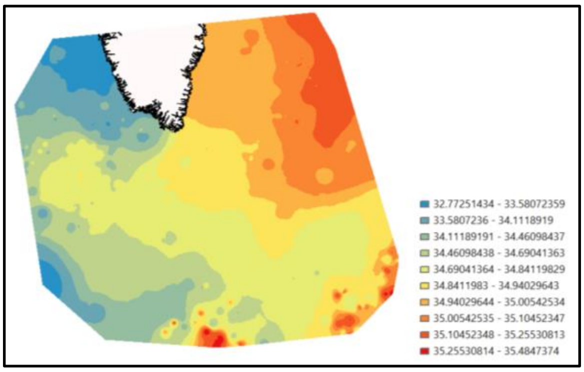


Spring/Summer 2015

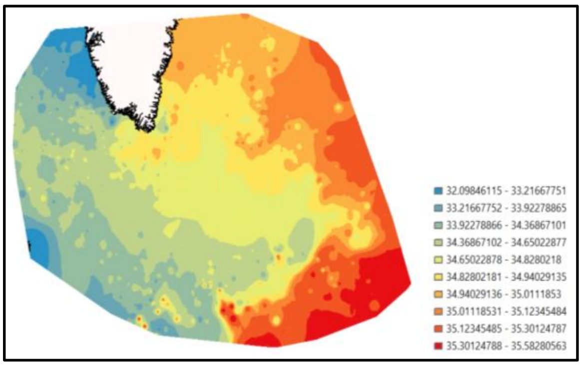


Spring/Summer 2016

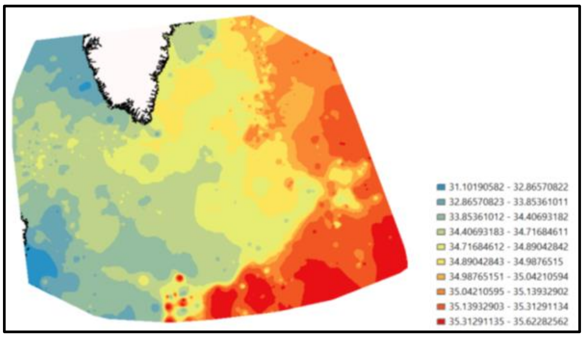
Figure 19: Spring/Summer salinity changes, 2012 – 2017



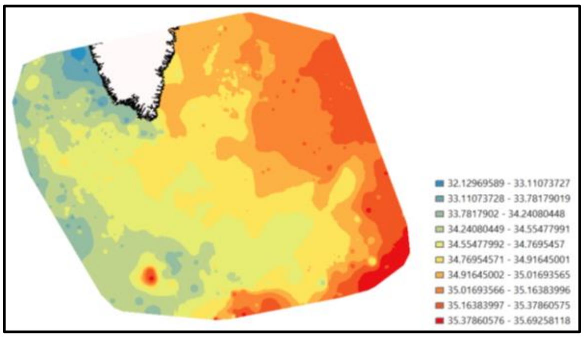
Winter 2012



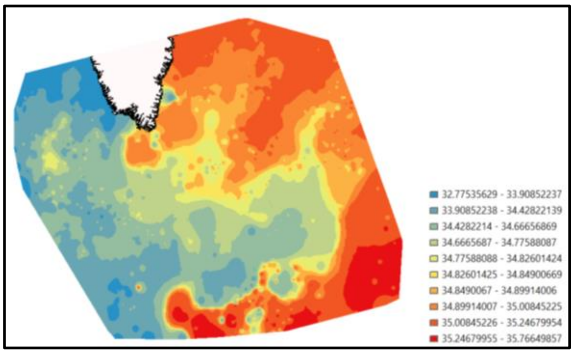
Winter/Autumn 2012-2013



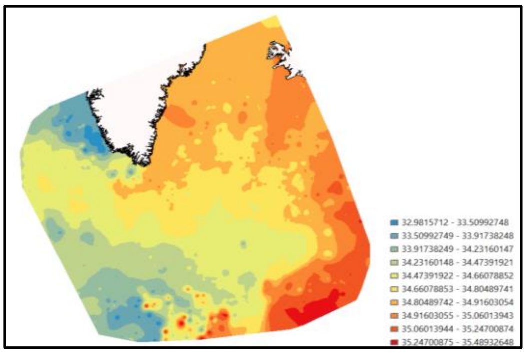
Winter/Autumn 2013-2014



Winter/Autumn 2014-2015



Winter/Autumn 2015-2016



Winter 2016

Figure 20: Winter/Autumn salinity changes, 2012 - 2017

5.2.3 Salinity changes

ArcGIS Minus tool was used to create a new surface, which shows changes in salinity between 2012 and 2016. Minus tool calculates subtraction between two rasters analysing them cell by cell (ArcGIS, 2017c).

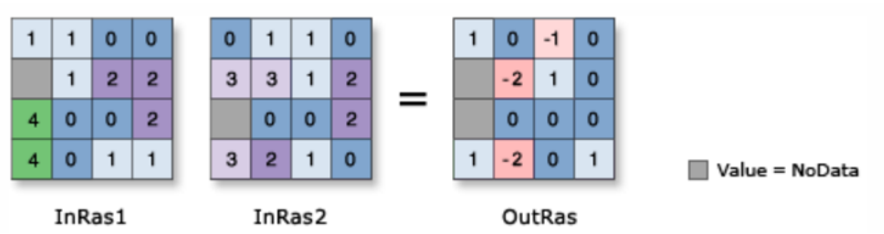


Figure 21: ArcGIS Minus tool (Source: ArcGIS, 2017)

Analysing Spring/Summer shows that high salinity concentrations are clustered in the south and low concentrations are clustered to the north adjacent the Greenland coast (Figure 22).

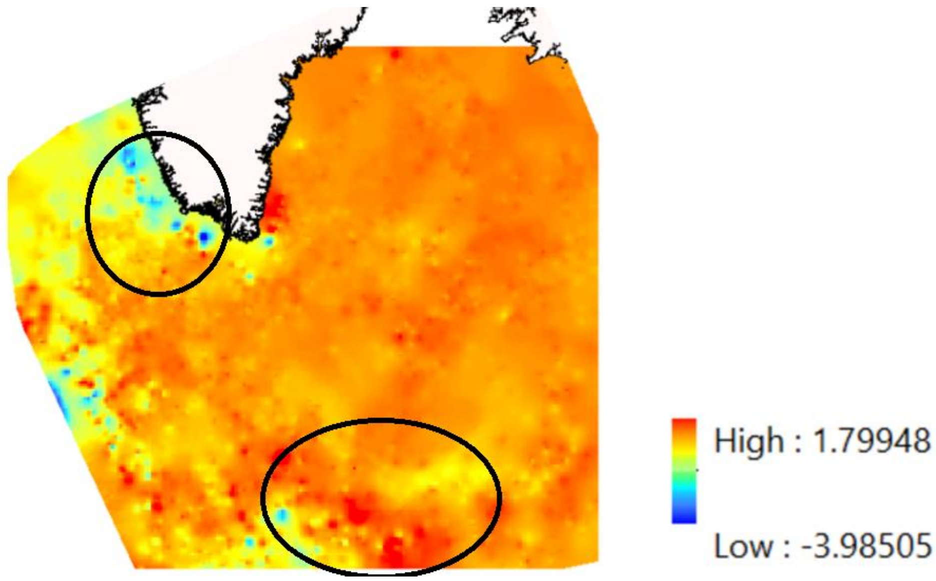


Figure 22: Salinity Difference Spring/Summer 2012 compared to 2016

Interpreting Winter/Autumn Salinity changes, similar trends can be observed as in spring/summer - high salinity concentrations are clustered in the south and low concentrations are clustered to the north, next to the Greenland west coast (Figure 23).

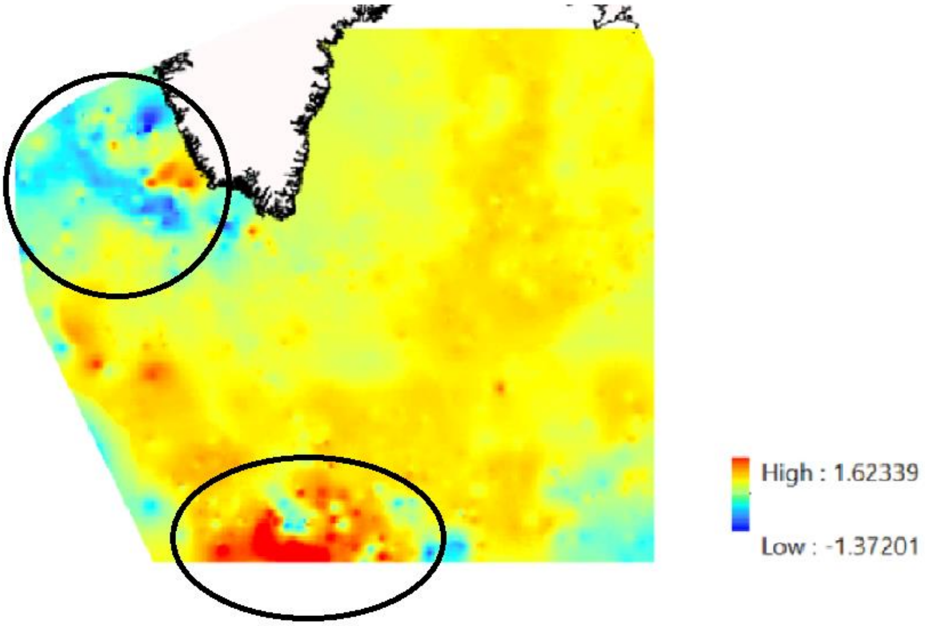


Figure 23: Salinity Difference winter/Autumn 2012 compared to 2016

In the summer it can be observed that the areas that decreased in salinity were mainly in the west of the study area in small localised pockets. The areas that showed an increase in salinity were more dispersed and located for the most part in the south. The maximum decrease is almost 4 psu, while the maximum increase is lower at around 1.8 psu in summer. The majority of the study area is shown in either orange or yellow, which indicates that the salinity change over the five years was relatively stable, with perhaps a slight increase. The pockets of decreasing salinity had thus an outsized impact on the overall change.

In contrast, the change in salinity shown in the winter comparison is much more even, with the maximum increase being around 1.6 psu and the maximum decrease being almost 1.4 psu. The localised pockets of increasing and decreasing salinity levels are approximately in the same areas as for the summer months, but the changes in winter appear to have a smaller impact on the overall change than the summer changes.

5.3 Regression analysis

5.3.1 Temperature and Salinity

An Ordinary Least Squares (OLS) analysis was conducted to determine influence of temperature on salinity as illustrated at Figure 24. Data for depths to 50m across the 5 years was used in the analysis

Input Features:	Export_Output_3	Dependent Variable:	PSAL
Number of Observations:	495989	Akaike's Information Criterion (AICc) [d]:	313669.220158
Multiple R-Squared [d]:	0.157670	Adjusted R-Squared [d]:	0.157668
Joint F-Statistic [e]:	92840.155007	Prob(>F), (1,495987) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	54259.038156	Prob(>chi-squared), (1) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	503.395222	Prob(>chi-squared), (1) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	1145686.892189	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Figure 24: Ordinary Least Squares (OLS) analysis of temperature on salinity

The Adjusted R-Squared value of 0.157670 indicates temperature is only explaining 16% of the story behind salinity variation. Residuals were plotted to determine if any clustering was present (Figure 25).

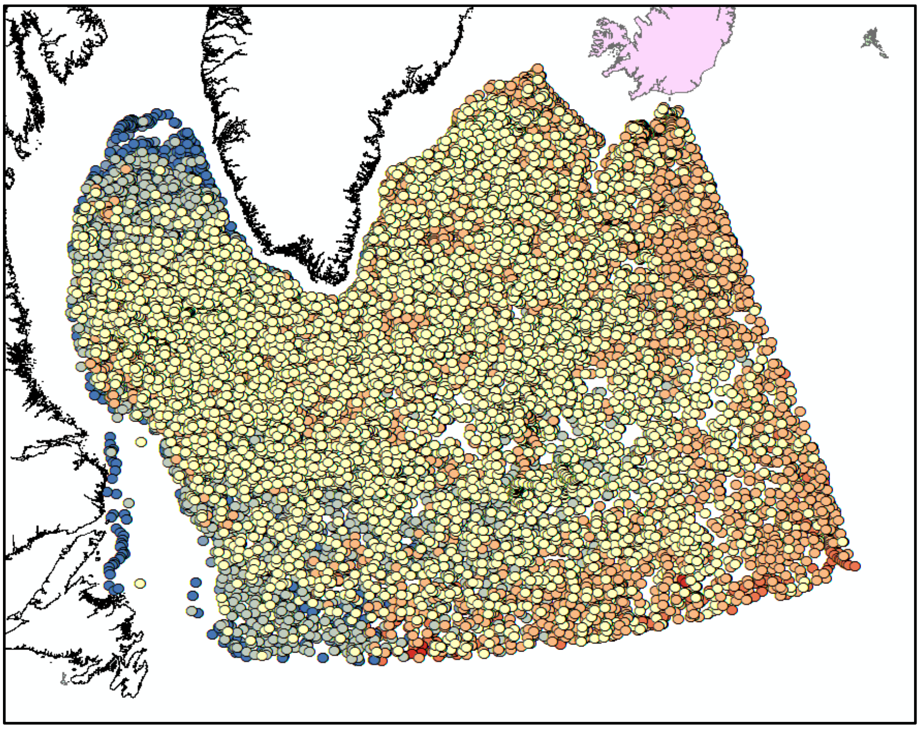


Figure 25: Residual points from OLS analysis

As demonstrated at Figure 25, it was not possible to determine any clustering due to the large number of data points. Interpolation of points was conducted to obtain a smooth surface and observe any trends (Figure 26).

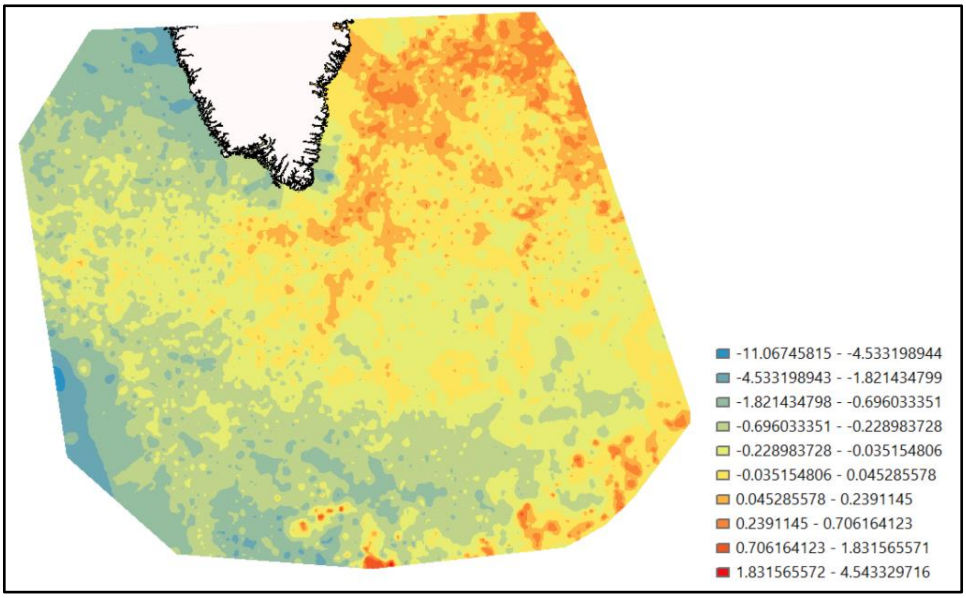


Figure 26: Interpolation to observe trends

Some clustering of residuals can be observed, indicating over-predictions (red) and under-predictions (blue). Large clusters of under predictions can be observed to the west, and large clusters of over predictions to the east. Areas of yellow indicate normally distributed variables, which indicates temperature is explaining salinity variation well. For the model to be valid, residuals must be normally distributed.

Histograms and scatterplots were created for temperature (explanatory) and salinity (dependant), to determine whether a correlation exists as illustrated at Figure 27. The histogram illustrates how each variable is distributed and the scatter plot depicts the relationship between the explanatory variable and the dependent variable (Esri Help, 2013).

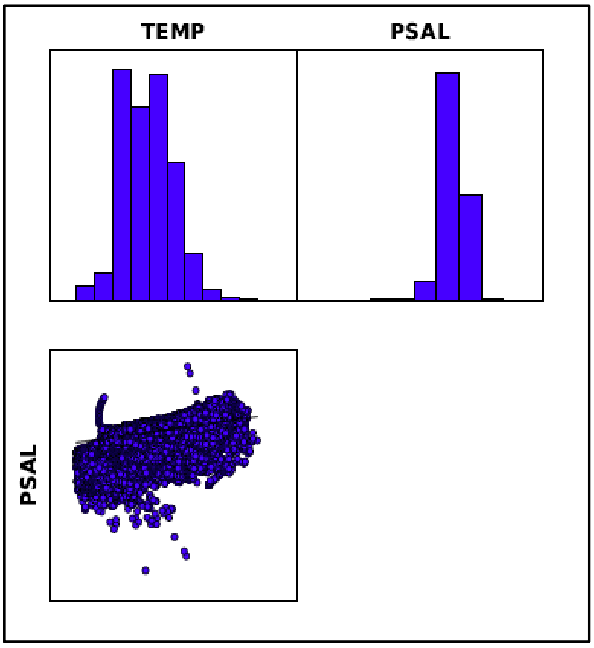


Figure 27: Histograms and scatterplots of salinity and temperature

It can be observed from the scatterplot that a positive linear relationship exists, indicating salinity increases as temperature increases (Figure 27). Temperature and salinity have an inverse relationship according to the literature, but our data shows opposite results, because there are more factors which influence changes in salinity, for example melting of ice, freshwater input from rivers, and precipitation in the form of rain and snow. The histogram indicates distribution of variables is normally distributed, which is not essential for OLS analysis, however can be beneficial for establishing a properly specified model (ArcGIS, 2017b).

The analysis was expanded to include longitude as an explanatory variable, which has been demonstrated to influence salinity (Klemas, 2001) and is consistent with interpolations (see Figure 12).

5.3.2 Salinity, temperature and longitude

Ordinary Least Squares (OLS) analysis was conducted on the upper 50m over the 5 year study period to determine the influence of longitude and temperature on salinity (Figure 28).

Input Features:	final2	Dependent Variable:	PSAL
Number of Observations:	120271	Akaike's Information Criterion (AICc) [d]:	58179.676672
Multiple R-Squared [d]:	0.375424	Adjusted R-Squared [d]:	0.375413
Joint F-Statistic [e]:	36145.659960	Prob(>F), (2,120268) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	88940.180909	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	3300.183029	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	223977.361646	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Figure 28: Ordinary Least Squares (OLS) analysis of temperature and longitude on salinity

The adjusted R-Squared value improved from 0.157670 to 0.375424, which indicates temperature and longitude is explaining 38% of the story behind salinity variation. Residuals were plotted to determine if any clustering was present.

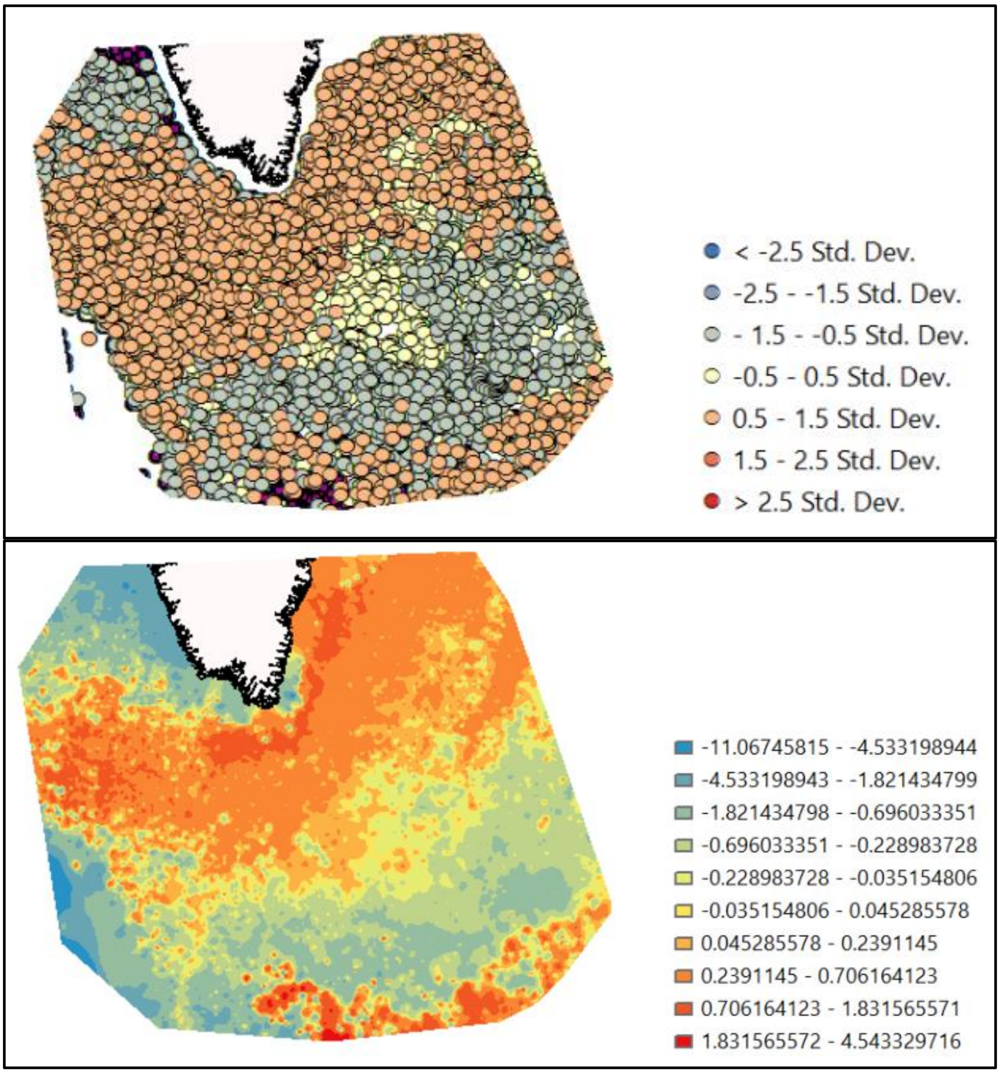


Figure 29: Clustering of residuals from OLS analysis of salinity, temperature and longitude

The following clustering can be observed:

- Over-predictions (red) east of the Greenland coast and southern area of interest; and,
- Under-predictions (blue) off the west coast of Greenland.

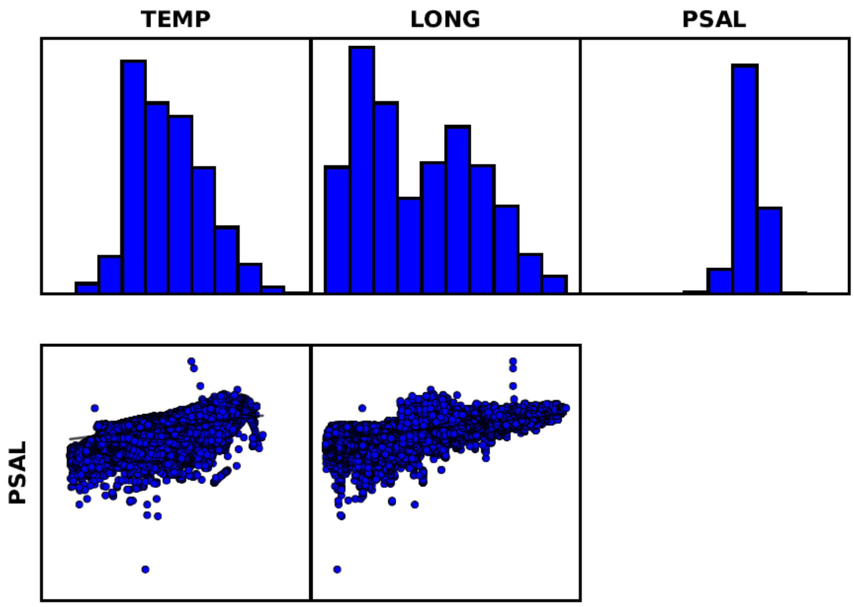


Figure 30: Histograms and scatterplots of temperature and longitude based on salinity

A positive relationship exists between salinity and temperature, and salinity and longitude, indicating both variables increase with salinity. Salinity increases moving east and decreases moving west (Figure 30).

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	34.442775	0.002011	17128.316048	0.000000*	0.002794	12327.362124	0.000000*	--------
TEMP	-0.005596	0.000348	-16.093914	0.000000*	0.000534	-10.488668	0.000000*	1.737124
LONG	0.000000	0.000000	214.116800	0.000000*	0.000000	194.742214	0.000000*	1.737124

Input Features:	final2	Dependent Variable:	PSAL
Number of Observations:	120271	Akaike's Information Criterion (AICc) [d]:	58179.676672
Multiple R-Squared [d]:	0.375424	Adjusted R-Squared [d]:	0.375413
Joint F-Statistic [e]:	36145.659960	Prob(>F), (2,120268) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	88940.180909	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	3300.183029	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	223977.361646	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Figure 31: Summary of OLS Results

6 checks to validate the model (Figure 31):

1. **Coefficients have the expected sign**. TEMP has a negative coefficient, indicating temperature decreases with increased salinity, which contradicts the scatterplot. LONG has a positive coefficient indicating salinity increases to the east, which was expected.

2. **No redundancy among explanatory variables**. VIF is low in both cases (below 7.5), therefore both have equal weighting to explaining the relationship.
3. **Coefficients are statistically significant**. An asterix next to the probability and robust_pr figures indicate significance, which is observed.
4. **Residuals are normally distributed**. The Jarque-Bera Statistic has a high number with a P value <0.01, which indicates residuals are clustered.
5. **Strong R Square value**. The value of 0.375424 indicates there are other factors to explain the distribution.
6. **Residuals are not spatially auto correlated**. It was already established the residuals failed this test.

5.3.3 Temperature, distance and salinity

OLS analysis was then expanded to include distance to the coast (Figure 36), as fresh water runoff from adjacent land was hypothesised to be an influencing factor. Data from the first depth measurement was used, and for this analysis only a single month (January 2012) was used due to the large size of the dataset.

To calculate the distance of each float to the coastline a shapefile for the coastline was downloaded from Open Street Maps Data (2017). The shapefile was reduced to the coastline of Greenland, Iceland and Canada by creating a shapefile from selection in ArcGIS. This new shapefile was loaded into pgAdmin III using the shapefile loader and setting the SRID to 4326. The transformed float data for the upper 20m was also imported into pgAdmin III. To be able to calculate the distance in meters, both shapefiles had to be in the same projection. The float data was already transformed in R, but the coastline had to be transformed. This was done with the st_distance postgreSQL function. Afterwards, a new column named coastdist was added with integer as a format.

```
ALTER TABLE coastline
 ALTER COLUMN geom TYPE geometry(MultiLineString,3183)
  USING ST_Transform(geom,3183);

ALTER TABLE public.surf ADD COLUMN coastdist integer;
```

Code 10: SQL – projection

The initial idea was to calculate all the data points and group them into distances to plot each of them with salinity to see if there is a pattern. The st_distance function returns the minimum

distance between two geographies in meter. This calculation exceeded our data processing capability, so the calculation was done on a very small bit of the data set.
Hence, the surface was limited to only the very first measurements and copied into a new table called surf. The distinct function selects the unique values, which are ordered based on the columns given in the order by function.

```
CREATE TABLE surf AS
SELECT DISTINCT ON (surface.geom) surface.geom, surface.gid, surface.id, surface.date, surface.pres,
        surface.temp, surface.psal, surface.month, surface.year, surface.long, surface.lat
FROM surface
ORDER BY surface.geom, surface.pres ASC;
```

Code 11: SQL - creating new table

Thereafter, the coastdist column is set to the a subquery which calculates the distance and also limits it with the order by function. It puts the outcome into the right row based on the WHERE clause, where the columns match.

```
UPDATE surf
SET coastdist = subquery.coastdist
FROM (SELECT DISTINCT ON (allunique.geom) allunique.geom, ST_Distance(coastline.geom,
allunique.geom) AS coastdist
FROM coastline,
        (SELECT DISTINCT geom
        FROM surf) AS allunique
ORDER BY allunique.geom, ST_Distance(coastline.geom, allunique.geom)) AS subquery
WHERE surf.geom=subquery.geom;
```

Code 12: SQL - distance calculation

This newly created table was then exported to a .csv file to run the OLS analysis on.

Input Features:	dist	Dependent Variable:	PSAL
Number of Observations:	288	Akaike's Information Criterion (AICc) [d]:	132.756716
Multiple R-Squared [d]:	0.577419	Adjusted R-Squared [d]:	0.574454
Joint F-Statistic [e]:	194.713449	Prob(>F), (2,285) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	192.913120	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	7.059313	Prob(>chi-squared), (2) degrees of freedom:	0.029315*
Jarque-Bera Statistic [g]:	340.371873	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Figure 32: Ordinary Least Squares (OLS) analysis of temperature and distance on salinity

As demonstrated at Figure 32, combining temperature and distance to the shore improved R-Squared to 0.577419, which indicates a combination of these factors accounts for 58% of salinity change in January 2012.

Residuals were plotted and interpolated to determine clustering (Figure 33).

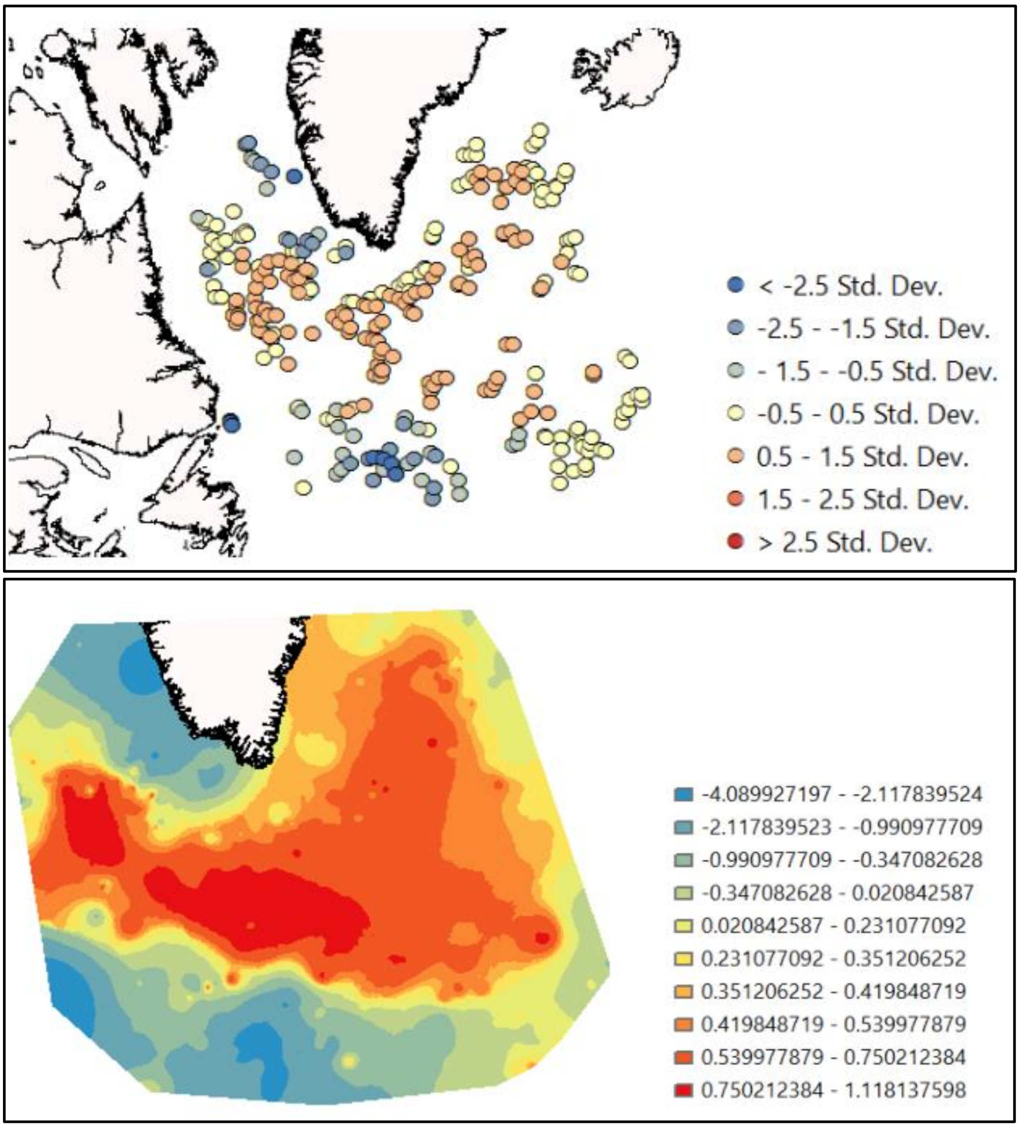


Figure 33: Standard deviation for temperature and distance

The following clustering can be observed:

- over prediction clustering to the south of Greenland (red); and,
- under predictions off the west coast and further south (blue).

Histograms and scatterplots were created for temperature and distance from coast (explanatory) and salinity (dependant) (Figure 34).

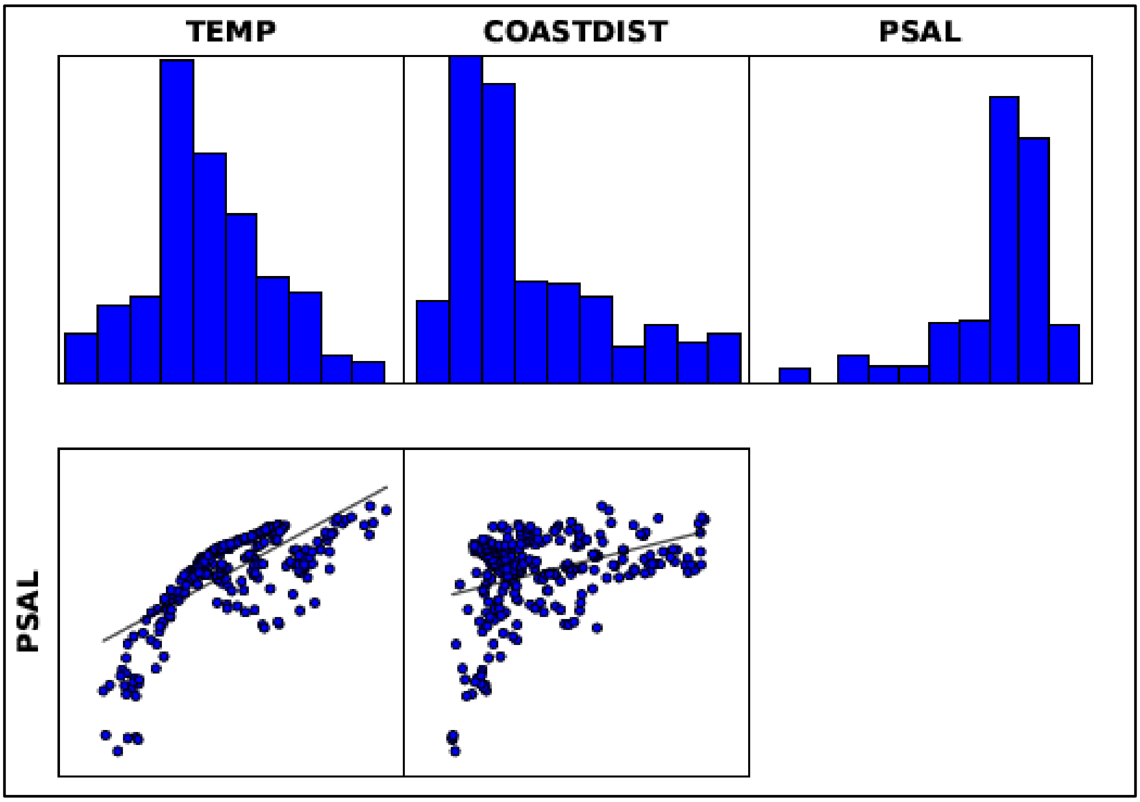


Figure 34: Histograms and scatterplots of temperature and distance based on salinity

It can be observed that a positive relationship exists between salinity and distance to the coast, meaning salinity increases with temperature and distance from shore (Figure 34). The temperature bar chart is normally distributed, whereas distance from shore is positively skewed, indicating the majority of measurements were collected nearer to the shore.

Spatial autocorrelation of regression residuals analysis was then conducted to determine the clustering of residuals (Figure 35).

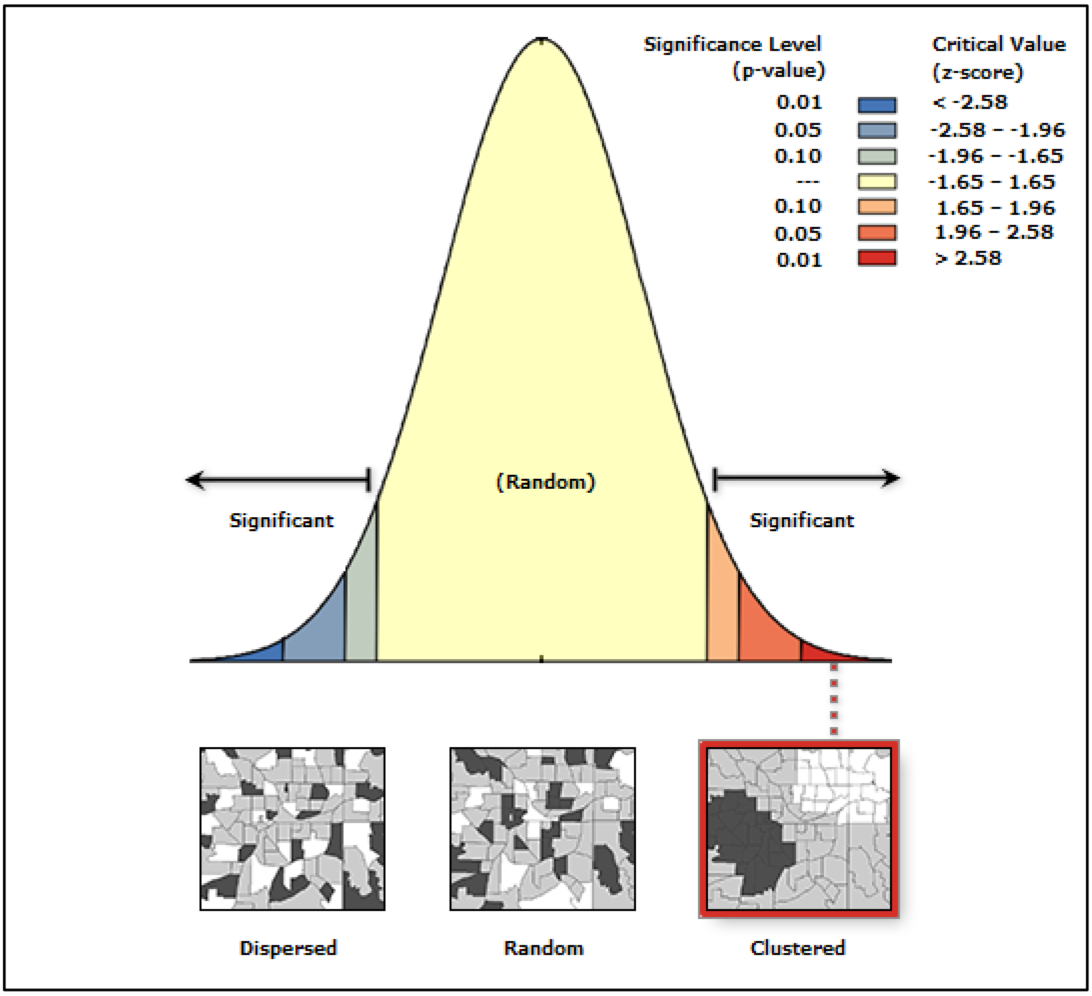


Figure 35: Spatial autocorrelation analysis

As illustrated at Figure 35 the Spatial autocorrelation analysis confirmed the residuals are clustered, which means the model is still missing key explanatory variables and results cannot be trusted at this stage. OLS analysis was then conducted, followed by 6 checks for a properly specified model (ArcGIS for Desktop, 2017).

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	34.192626	0.035496	963.279929	0.000000*	0.049060	696.958280	0.000000*	--------
TEMP	0.176383	0.010084	17.491234	0.000000*	0.012750	13.834229	0.000000*	2.660996
COASTDIST	-0.000001	0.000000	-8.218024	0.000000*	0.000000	-11.752015	0.000000*	2.660996

OLS Diagnostics

Input Features:	dist	Dependent Variable:	PSAL
Number of Observations:	288	Akaike's Information Criterion (AICc) [d]:	132.756716
Multiple R-Squared [d]:	0.577419	Adjusted R-Squared [d]:	0.574454
Joint F-Statistic [e]:	194.713449	Prob(>F), (2,285) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	192.913120	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	7.059313	Prob(>chi-squared), (2) degrees of freedom:	0.029315*
Jarque-Bera Statistic [g]:	340.371873	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Figure 36: OLS Diagnostics for distance to shore and temperature

1. **Coefficients have the expected sign**. TEMP has a positive coefficient, indicating a positive relationship, which is expected because increased surface temperature results in increased salinity. COASTDIST has a negative coefficient, indicating salinity decreases as distance to shore increases. This result is not expected because it contradicts the observations from the salinity/distance scatterplot (Figure 34).
2. **No redundancy among explanatory variables**. VIF is low in both cases (below 7.5), therefore both have equal weighting to explaining the relationship.
3. **Coefficients are statistically significant**. An asterix next to the probability and robust_pr figures indicate significance, which is observed.
4. **Residuals are normally distributed**. The Jarque-Bera Statistic has a high number with a P value <0.01, which indicates residuals are clustered.
5. **Strong R Square value**. The value of 0.577419 indicates there are other factors to explain the distribution.
6. **Residuals are not spatially autocorrelated**. It was already established the residuals failed this test.

6. Discussion

6.1 Does surface salinity change over a 5 year period in the Labrador Sea?

Mean salinity for the entire dataset (down to depths of 2000 meters) decreased approximately 1 psu over the study period, which is consistent with the hypothesis that salinity would decrease due to an increase in inflow of freshwater from ice melt and precipitation associated with climate change. As can be seen in the line graph at Figure 11, salinity concentrations fluctuated greatly during the study period, with troughs occurring in September and peaks in January. The trend line indicates that the small decrease in salinity over the five years is a pattern that will continue in the future. A possible explanation is the input of freshwater from ice melt during summer, and ice formation during winter that removes fresh water, however, due to the complexities associated with determining ocean salinity discussed earlier, regression was conducted to test possible causal factors.

Salinity change was further explored in the plots at Figure 12, which show salinity for the upper 20 meters for the summer and winter season for each study year. A general trend of salinity decline can be observed, as well as higher salinity towards the east. The line graph (Figure 18) showing mean salinity change per month over the study period for the surface 20 meters of water reflects the same pattern as the overall dataset, which indicates this phenomenon is driven by surface factors. Seasonal fluctuations in salinity were evident, with higher surface salinity in the winter and lower in the summer, which was expected because the Labrador sea adjacent to the south west coast of Greenland is sheltered from ocean currents, and as a result freshwater from ice melt during summer accumulates and lowers salinity. In winter convective mixing homogenises the water column, with the net effect of increasing salinity.

The two visualisations at Figures 22 and 23 show two main areas of change in salinity over the study period and are the same for summer and winter. These areas are indicated within ellipse markers. Salinity increased to the south, which is an area of the Atlantic Ocean that is subject to strong currents from the GOCB discussed earlier (Klemas, 2001). A hypothesis for this increase is saline water from the equator mixing with the North Atlantic water mass, producing a net salinity increase.

A decrease in salinity can be observed off the west coast of Greenland, which is sheltered from ocean currents and therefore more influenced by local factors such as precipitation and freshwater run-off (Klemas, 2001). During summer this area can be observed directly adjacent to the west coast of Greenland, which could be a result of freshwater inflow from summer ice melt. In the winter this fresh water mass can be observed further away from the coast, which could be due to offshore winds or ocean currents.

6.2 Where is the surface (at what depth does salinity change)?

The surface was important to determine for several reasons. Firstly due to the large number of data points, a subset was necessary to reduce the number of data points. Secondly, the surface is most sensitive to small scale factors that influence salinity such as climate change and precipitation, which are of interest to this project. Lastly, existing satellite data measuring salinity exists for only the surface, which is useful as a comparison to verify the Argo data.

Salinity was expected to change with depth because the Labrador Sea is subject to high interseasonal stratification of the water column, which creates density layers of varying salinity, followed by mixing which homogenises the layers and evens out salinity (Mysak and Power, 1990; Klemas, 2001). To investigate this phenomenon, plots were created for salinity and depth for each season (Figure 19-20), to define the depth of the surface by determining the depth at which a change occurs in salinity upper or lower limits.

Data between months and seasons was highly variable, and the reliability of the data was questionable because of many outliers. However, it was observed the upper 20m depth had the most consistent salinity range, which is a depth consistent with literature sources (HaMAARAG , 2017). This depth was used as the surface depth for the majority of analysis in the project to ensure consistency. Only the first measurement from each float was used due to the large number of measurements, which resulted in measurements at random between 0m and 20m depth.

In addition to determining the surface depth, several other observations were made. When the graph from the upper 20m was compared with the graph for the entire water profile, a smaller range of salinity values were observed with increased depth, which was expected because the deep ocean is subject to less small scale variation compared with the surface. Salinity in the deep ocean is influenced over a large timeframe and large scale factors such as ocean currents. R^2 reduced, which indicated a reduction in how well the trend line fitted the data, which is consistent with greater variability in the surface.

The plots comparing the seasons indicated greater consistency of salinity in winter compared with summer was observed (less variability), which could be due to convective mixing of water layers during winter that homogenises the salinity content.

6.3 Does a correlation exist between salinity and ocean temperatures?

Temperature and salinity interact to determine water density which drives ocean currents (Marinebio, 2017). A global trend exists of warmer water and saltier water at the equator which reduced towards the poles (Bright hub engineering, 2017), which is driven by evaporation from solar radiation and freshwater influx from precipitation (Marinebio, 2017). This research question was proposed because an understanding of the balance of these factors can provide an insight into the dynamics operating in the region. Although the Labrador Sea is small on the global scale and the observation timeframe is relatively short, a correlation was expected with lower surface temperatures and lower salinities to coincide with cool freshwater from ice melt.

The OLS regression demonstrated a 16% correlation between salinity and temperature over 5 years (Figure 23), with a positive linear relationship between the variables. This is consistent with the hypothesis that cool freshwater influx is lowering salinity, however it indicates other factors are influencing salinity to a greater extent than temperature alone. The bell shape in the histograms indicate normally distributed variables, which enables different types of analysis, such as Geographically Weighted Regression (GWR), to investigate further (Figure 26).

6.4 Do any of the other variables influence salinity?

The regression analysis for salinity and temperature indicated other factors were influencing salinity. Salinity was observed to vary with factors such as longitude and distance to shore in explanatory plots (refer Figures 18, 19 and 20), therefore these factors were added to the analysis.

Challenges were encountered with adding distance to shore to the analysis, because the processing required for the pgAdmin query was too great for the number of data points. Therefore this analysis was undertaken using only one month of data (Jan 2012), and as such could not be combined with the temperature and salinity analysis which used data from the entire study period, which is why it is a stand alone analysis.

Regression analysis of temperature, longitude and salinity indicated a correlation of 37.5%, which was a 21.5% improvement on temperature alone. A positive correlation was observed, which indicates salinity increasing moving east away from North America and Greenland. It is hypothesised this is due to the westerly landmass freezing in winter and melting in the spring/summer, whereas the easterly area is part of or adjacent to the open ocean. This would likely cause a greater decrease in salinity levels from ice melt than in the study area in the North Atlantic Ocean.

Correlation between distance to shore and salinity indicated an improved R Squared value of 0.577419, however as previously described the analysis was conducted over a shorter time period and fewer data points, which reduces the spatial and temporal scale of the analysis. Given that the month chosen for the distance to coastline calculation was January, the mid-winter time frame could be a factor in the improved explanatory value of the distance variable.

The interpolation results (Figures 19 and 20) indicates that distance to shore is not a consistent influence on salinity levels. For example, proximity to the North American coastline appears to correlate with decreased salinity, however proximity to the east coast of Greenland does not. Factors such as precipitation, air temperature, evaporation, wind and current could impact salinity levels, and could be included in future analysis.

7. Conclusions

A reduction in ocean salinity was observed between 2012 and 2016. An exact sea surface definition was not established, however 20 meters was selected based on the consistency of salinity and temperature to that depth.

A correlation with temperature and salinity was demonstrated, however additional factors including longitude and distance to shore are also explanatory variables. This is consistent with the hypothesis that a reduction in salinity will be observed over the study period due to increased ice melt and reduction of sea ice formation as a consequence of climate change. Proximity to the shoreline, particularly in westerly longitudes nearer to North America, was observed to be correlated with salinity; which was a discovery additional to the original project hypothesis.

7.1 Challenges

Argo data is extremely comprehensive with various data types and formats, such as raw and unverified, so-called ‘good’ data, gridded fields, and global or regional data, and the dataset is increasing in size exponentially as more floats are added to the array every year. Although Argo is a valuable resource, it presents challenges for researchers, particularly those with limited experience with data handling, ocean science or hydrology.

The challenges encountered in this project were predominantly related to the size and complexity of the Argo dataset. Initially the dataset included approximately 750 .csv files, which comprised over 4.5 million rows of data. Exploring, subsetting, analysing and visualising a dataset this size proved challenging and time consuming, especially given the relatively low processing power of the group’s personal laptop computers.

The large scope of Argo data can be a great benefit to study the complexity of ocean systems, however for the purposes of this research project, it was necessary to limit the scope of the data to ensure a manageable project. The challenge was to balance data coverage with a large enough geographic area and time frame to make results meaningful, without limiting the scope for the project timeframe and capacity.

Argo data consists of four dimensions: latitude, longitude, depth and time; which proved a challenge to conceptualise the data and determine what type of analysis would be meaningful and logical. The aim of the project was to create a spatio-temporal analysis of ocean salinity in the Labrador Sea, which proved ambitious given the complexity of dealing with 4D data, and the dataset was modified accordingly.

The data used in the project was ‘good’ data from the GDAC, however multiple errors were encountered, such as missing columns, additional nameless columns and missing values. A high level of scrutiny was applied to filtering the dataset, to ensure these errors did not skew the results and affect the analysis.

Software expertise proved a significant challenge. Processing a large and complex dataset required powerful software, and R was selected as the most appropriate. However since no group members were experienced with R, the learning curve was steep to use the full capabilities of the language.

The use of the Argo dataset alone, without additional variables such as precipitation, air temperature, or freshwater influx, likely oversimplify a large and complex system, however this was determined necessary due to the scope and timeframe of the project. Large scale environmental processes, such as climate change and ocean salinity, occur over a long timeframe and the project could be extended to better capture these processes.

7.2 Project Strengths

This project addresses a significant global issue - salinity change in the Labrador Sea, which could have implications for global climate systems due to modification of the GOCB. It utilised a precise and extensive data set from Argo, in conjunction with powerful geoprocessing tools, to demonstrate salinity has decreased in the Labrador Sea.

In addition to highlighting this phenomenon, this project emphasises the importance of the Argo program, which receives minor funding and relies on vessels to deploy the floats on a voluntary and opportunistic basis. This report provides exposure for the Argo program, and presents collaboration opportunities with institutions such as universities.

Salinity change influencing density is not only important for ocean currents, and the Labrador Sea is not the only region in which this phenomenon is significant. Salinity is also important for influencing the distribution of several fish species (Buch, 1984), and variation in density could impact the GOCB at many areas. In theory, this project could be extended to the entire ocean area. Indeed, ocean salinity and temperature monitoring is an important research field due to the serious implications of climate change and associated impact on biodiversity, aquaculture, sea-levels and weather patterns, among others.

Effective use of geoinformatics can be extremely useful for conveying and interpreting complex phenomena such as ocean science. This project combines the disciplines of geoinformation, environmental science, and statistical analysis to provide a powerful assessment of salinity change in the Labrador Sea, as well as quantify influencing factors.

7.3 Future Directions

The project was limited in terms of additional factors and explanatory variables due to the restricted time frame. It could be expanded with additional variables such as currents, rainfall, freshwater flow from rivers, ice melt from remote sensing, and air temperature. These factors will most likely increase the explanation of changes in salinity. Additionally, a larger time frame could expand the trend in salinity change, which is especially important for understanding climate change. The data could also be expanded globally, as Argo provides data from all over the world to account for large scale factors.

A Web Map Service could be created to attract more people's attention to salinity changes. This perfectly presents environmental pollution based on human activities. Visualization of changes and future predictions can be valuable in teaching process. Additionally, an interactive salinity map could be used as a base map for other environmental applications.

Moreover, 3D visualization today is extremely accessible, so this salinity model could be moved to 3D visualization which provides more natural space effect. 3D models also convey information more clearly, and can be rotated, reversed and analysed differently than 2D. This information is more attractive and easier to understand.

A further area of development that was unable to be achieved was the creation of a series of surfaces showing salinity at various depth levels. These are known as gridded fields and are available for download from multiple Argo affiliates, however there are no specific datasets for the study area of this project. It was intended that this would be a possible output of this project, however time restrictions did not allow for this work to be produced.

References

ArcGIS (2017a). How IDW works [Online] Available at: http://desktop.arcgis.com/en/arcmap/10.3/tools/3d-analyst-toolbox/how-idw-works.htm. [Accessed 08 April 2017].

ArcGIS (2017b). Regression analysis basics. [Online] Available at: http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/regression-analysis-basics.htm. [Accessed 08 April 2017].

ArcGIS (2017c). Minus tool [Online] Available at: http://pro.arcgis.com/en/pro-app/tool-reference/3d-analyst/minus.htm. [Accessed 08 June 2017].

ArcGIS for Desktop, (2017). *Understanding Interpolation Analysis.* [online] Available at: http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/understanding-interpolation-analysis.htm . [Accessed 29 May 2017].

Argo (2000). *Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)* [online] Available at: http://doi.org/10.17882/42182 [Data downloaded on 3 May 2017]

Argo (2017a). *About Argo*. [online] Available at: http://www.Argo.ucsd.edu/About_Argo.html [Accessed 25 May 2017].

Argo (2017b). *How Argo Floats Work.* [online] Available at: http://www.Argo.ucsd.edu/How_Argo_floats.html [Accessed 25 May 2017].

Argo (2017c). *Argo Data and How to Get It*. [online] Available at: http://www.Argo.ucsd.edu/Argo_data_and.html [Accessed 25 May 2017].

Argo (2017d). *Uses of Argo Data*. [online] Available at: http://www.Argo.ucsd.edu/Uses_of_Argo_data.html [Accessed on June 4 2017].

Argo (2017e). *Data Products*. [online] Available at: http://www.Argo.ucsd.edu/Gridded_fields.html [Accessed on 5 June 2017]

Argo - part of the integrated global observation strategy, (2015). *Argo FAQ*. [online] Available at: http://www.Argo.ucsd.edu/FAQ.html [Accessed 20 May 2017].

Bright hub engineering (2017). Effect of Temperature on Salinity [Online] Available at: http://www.brighthubengineering.com/naval-architecture/104494-effect-of-temperature-on-salinity/. [Accessed 08 June 2017].

Buch, E. (1984). *Variations in Temperature and Salinity of West Greenland Waters, 1970-82*. NAFO Sci. Coun. Studies, 7: 39-43

Comprehensive R Archive Network (2017). *Available CRAN Packages By Name*. [online] Available at: https://cran.r-project.org/web/packages/available_packages_by_name.html. [Accessed 17 May 2017]

Cracknell, A.P. and Hayes, L., 2007. Introduction to Remote Sensing.
Boca Raton, Florida: CRC Press.

de Smith, M., Goodchild, M., Longley, P. (2015) *Geospatial Analysis - A Comprehensive Guide to Principles, Techniques and Software Tools* 5th Edition [online] Available at: http://www.spatialanalysisonline.com/HTML/index.html. [Accessed on 25 May 2017]

epsg.io (2017). *EPSG:3183*. [online] Available at: https://epsg.io/3183. [Accessed 15 May 2017].

HaMAARAG - Ocean Health Index, Israeli Mediterranean Assessment (2017) *Sea Surface Temperature* [online] Available at: http://www.hamaarag.org.il/en/ocean-health-index/sea-surface-temperature [Accessed 8 June 2017]

Jones, M..; Guarinello, M.; Soderberg, C.; Wasser, L.(2017) *Time Series 06: Create Plots with Multiple Panels, Grouped by Time Using ggplot Facets.* [online] Available at: http://neondataskills.org/R/time-series-plot-facets-ndvi/ [Accessed 4 June 2017].

Lagerloef, G.E., 2000. Recent progress towards satellite measurements of the global sea surface salinity field. In: Halpern, D. (ed.), Satellites, Oceanography and Society. Amsterdam: Elsevier Oceanography Series 63, 367p

Lagerloef, G.S.E., 2001. Satellite measurements of salinity. In: Steele, J.; Thorpe, S., and Turekian, K. (eds.), Encyclopedia of Ocean Sciences. London: Academic Press, pp. 2511–2516.

Le Vine, D.; Koblinsky, C.; Pellerano, F.; Lagerloef, G.; Chao, Y., and Wilson, W., 2004. A sensor to measure salinity in the open ocean from space. International Journal of Remote Sensing, 25, 1313–1318.

MarineBio (2017). The Ocean and Temperature [Online] Available at: http://marinebio.org/oceans/temperature/. [Accessed 08 June 2017].

Mysak, L.A., and Power, S.B. (1990). *Greenland Sea Ice and Salinity Anomalies and Interdecadal Climate Variability*. Centre for Climate and Global Change Research and Department of Meteorology, McGill University

NASA Official (2017) *Aquarius - Overview FAQ* [online] Available at: https://aquarius.umaine.edu/cgi/ov_faqs.htm [Accessed 8 June 2017]

NOAA - National Ocean Service (2015) *How Far Does Light Travel in the Ocean?* [online] Available at: http://oceanservice.noaa.gov/facts/light_travel.html. [Accessed 8 June 2017].

NOAA - National Centres for Environmental Information, (2017) *Ocean Climate Laboratory* [online] Available at: https://www.nodc.noaa.gov/about/oceanclimate.html. [Accessed 02 June 2017].

Open Street Map Data. (2017) [online] Available at: http://openstreetmapdata.com/data/coastlines [Accessed 29 May 2017].

PostgreSQL. n.d. About. [online] Available at: https://www.postgresql.org. [Accessed 07 June 2017].

Pgadmin.org (2016) [online] Available at: https://www.pgadmin.org [Accessed 07 June 2017].

RSL/UPC (Remote Sensing Laboratory, Universitat Politecnica Catalunya), 2011. Sea surface salinity. http://www.tsc.upc.edu/ (accessed March 29, 2011).

RStudio (2016). *Why RStudio?* [online] Available at: https://www.rstudio.com/about/. [Accessed 17 May 2017].

Science.nasa.gov, (2017). *Salinity - Science Mission Directorate*. [online] Available at: https://science.nasa.gov/earth-science/oceanography/physical-ocean/salinity. [Accessed 08 May 2017].

Soloviev, A. & Lukas, R. (2014) *The Near-Surface Layer of the Ocean*, Springer, Atmospheric and Oceanographic Sciences Library, Vol.48

Straneo, F. "Heat and Freshwater Transport through the Central Labrador Sea." Journal of Physical Oceanography 36 (2005)

The Argo Project Global Ocean Observations for Understanding and Prediction of Climate Variability,2003. Scripps Institution of Oceanography,La Jolla,CA,92093-0230, 2-7

The R Foundation (2017). *What is R?*. [online] Available at: https://www.r-project.org/about.html [Accessed 17 May 2017].

Wong,A. PS., G. C. Johnson, and W B.Owens (2003), Delayed-mode calibration of Autonomous CTD profiling float salinity data by Theta-S climatology, J.Atmos. Ocean. Technol., 20(2), 308-318.