Chapter 1

Introduction

1.1 Background

In this dissertation, focus is placed on the CO_2 transfer between the ocean and the atmosphere in the Southern Ocean. The aim of this dissertation is to develop an empirical relationship between pCO_2 in the ocean and the variables that influence the CO_2 transfer between the ocean and the atmosphere.

This dissertation is part of a bigger CSIR Strategic Research Project (SRP), with the title: *Understanding and Predicting the Seasonal Cycle of Carbon in the Southern Ocean: A High Resolution Global Carbon-Climate Model Study and Model Development Platform at the CSIR.* In short, the aim of the bigger project is to understand the seasonal cycle of carbon in the Southern Ocean. The specific topic of interest is the estimation, with a low uncertainty, of seasonally unbiased annual air-sea CO_2 fluxes in the Southern Ocean. The measurements of CO_2 fluxes in the Southern Ocean are sparse. Additionally, the number and quality of the measurements (of the fluxes) are seasonally biased, as it is more difficult to obtain accurate measurements of ocean properties in winter months. It is not feasible to obtain more samples of CO_2 fluxes in winter, since the presence of ice and clouds necessitate more expensive and time consuming methods for taking measurements. Thus, there exists a need to develop empirical pCO_2 relationships using other oceanic properties for which less sparse measurements are readily available in the Southern Ocean.

In recent years, climate change has developed into an increasingly important area of research [22]. Within this area of research, a great deal of emphasis has been placed on the relationship between the ocean and the atmosphere, and the manner in which the climate is affected by this relationship

[22]. According to Bye [22], both the ocean and the atmosphere are CO_2 storage systems that are connected by the surface of the water, where transfers between the two storage systems take place. Note that the ocean absorbs CO_2 — and this CO_2 is converted to other forms of carbon (in the ocean) through chemical processes. The increasing amount of carbon in the oceanic storage system has a positive effect on reducing the amount of CO_2 in the atmosphere. The exchange between the ocean and the atmosphere takes place at varying rates [29], depending on various aspects that will be discussed in this dissertation. The ocean stores significant amounts of carbon. There is approximately 60 times more carbon in the ocean than in the atmosphere. The challenge is to calculate the amount of carbon dioxide taken up and released by the ocean, taking into account the seasonal and regional variations that occur in the ocean [41].

1.2 Previous work

Previous work done on finding the aforementioned empirical relationship shows that various modern techniques have been applied to find this relationship. Metzl *et al.* [55], uses a semi-prognostic biogeochemical model to simulate the annual cycles of pCO_2 in the permanent open ocean zone (POOZ). The annual flux for the circumpolar zone, based on regional biogeochemical characteristics, are then extrapolated from the model results. The model is constrained by monthly sea surface temperature (SST), climatological winds, MLD (the depth at which the density of the water changes by $0.25\,\mathrm{kgm}^{-3}$ from the surface density) and chlorophyll-a concentration. The extrapolation does not yield accurate estimations of the pCO_2 and it is suggested that the model should be region-specific [55]. Chen *et al.* [25], uses the Levenberg-Marquardt algorithm to minimize the sum of the square errors to find an empirical relationship between pCO_2, chlorophyll concentration and SST. The estimation of the CO_2 flux in this case has an uncertainty of 45%.

D'Ortenzio *et al.* [33], uses a 1D coupled physical-biogeochemical model with satellite-derived information to estimate the CO_2 flux in the Mediterranean Sea. Friedrich and Oschilies [39] use a self-organizing neural network that is applied to satellite SST and chlorophyll concentration data to estimate monthly maps of the pCO_2 in the North Atlantic ocean. With this method, a root mean square (RMS) error of $19.0\,\mu$atm is obtained for the estimation of pCO_2. In addition, absolute errors in excess of $50\,\mu$atm are observed in regions where pCO_2 gradients are large. Chierici *et al.*[26], develop algorithms to estimate the fugacity of CO_2 in the surface ocean water from SST, chlorophyll-a concentration and MLD. The Levenberg-Marquardt routine is used to determine the error when adding each variable separately to the estimation equation. Furthermore it is shown that the smallest error for the pCO_2 estimation is obtained by adding all three variables. Telszweski *et* *al.* [76], uses a self organizing neural network to estimate monthly pCO_2 values across the North Atlantic for 2004 to 2006. In their analyses, this part of the ocean is divided into five regions (or biogeochemical provinces) for which different models are developed. Telszweski *et al.* also takes the seasonal variability of the ocean properties into account and obtains a model that reproduces the *in situ* measurements with a mean RMS error of 11.55 μatm.

1.3 Scope of dissertation

The question remains whether there is a compelling reason to use these modern techniques to determine the relationship between pCO_2 and other ocean variables. In this dissertation the aim is to quantify the degree to which classical techniques can be used to determine an empirical relationship. The empirical relationship is investigated using established techniques of regression and interpolation respectively. Both curve fitting and interpolation methods are used to establish a relationship between pCO_2 and other ocean variables. A rigorous assessment of the techniques is performed and the errors on the prediction (using these standard techniques) are quantified.

More than eight million data points are available for the Southern Ocean and the need exists to use a subset of the available points to investigate the empirical relationship. To use all the available data to establish this relationship would be uneconomical and time consuming. One way to use large data sets to construct an equation for the pCO_2 in the Southern Ocean is to sample a subset of data from the complete data set. Then, appropriate techniques including least squares approximation can be used to find a relationship between the variables. Random sampling is one way to sample the data, however, this is not a robust method of sampling. The set of data can be sampled by using D-optimal sampling [46], [47], [70], [50], [37], which minimizes the variance on the parameter estimates. D-optimal sampling is performed in this dissertation using genetic algorithms [40], [72], [44], [23], [19], [73]. A genetic algorithm is an optimization technique based on Darwin's principle of evolution and "survival of the fittest". This sampling procedure selects a set of points to fit the data optimally, resulting in a prediction of the pCO_2 with a smaller error. The algorithm is initiated with a random selection of possible solutions to the problem and is iterated by applying reproductive operators to the possible solutions, in order to find an optimal solution to the particular problem. Genetic algorithms are used widely in a number of different applications and have been proven to be a powerful and robust method in complex and non-linear optimization problems.

Regression can be used to determine the empirical relationship between CO_2 and other ocean variables. Several regression methods exist, including linear regression and non-linear regression methods.

A non-exhaustive list of regression methods include: least squares optimization, least absolute deviations, quantile regression and maximum likelihood estimation. In this dissertation, ordinary least squares optimization is used to investigate the possibility of accurately determining an empirical relationship between the ocean variables, by only using a small percentage of the available data. Emphasis is placed primarily on the ability of D-optimal sampling to yield accurate results for the estimation of pCO_2, from a small number of data points.

Another technique that can be used to determine the empirical relationship is Radial Basis Function (RBF) interpolation [38], [54], [30], [48], [35]. RBF interpolation is a method that is widely used to interpolate multivariate scattered data points. It can be applied to find an interpolation function for the ocean, wherewith the pCO_2 in the ocean can be predicted.

In this dissertation, a series of standard methods and techniques will be applied in order to obtain a relationship between the pCO_2 and other ocean variables. The aim is to obtain the relationship or estimation function that will yield pCO_2 values with the smallest error. Furthermore, in this dissertation, the empirical relationship is determined from only a small percentage of the data points by using the D-optimal sampling technique. The central contribution of this dissertation is to quantify the degree to which classical methods can determine an empirical relationship between pCO_2 and other oceanic properties. Additionally, the applicability of using the D-optimal sampling to select a set of points, from which the empirical relationship is established, is shown in this dissertation. It is hypothesised, in this dissertation, that it is possible to establish an empirical relationship between pCO_2 and other oceanic variables using classical techniques. Furthermore, it is hypothesised that it is possible to judiciously sample a subset of points from the complete data set in order to determine an empirical relationship, from only a small percentage of the available points, that will accurately estimate pCO_2.

1.4 Layout of dissertation

In Chapter 2, background about the project and the CO_2 transfers between the ocean and the atmosphere is given. Chapter 3 contains information about the methods that are used in this dissertation. The data that is available for the ocean and the different methods of obtaining and verifying data as well as the specific data set that is used in this dissertation is discussed in Chapter 4. In Chapter 5, the methods that are used to investigate the empirical relationship between pCO_2 and other oceanic variables, as well as the results obtained, are presented. Finally, the results and future work are discussed in Chapter 6. In Appendix A to D, the figures of the results for the least square curve fitting (with latitude excluded as variable) are given, as well as the tables of the mean coefficients of the fitted equation and the standard deviation of each coefficient in the equation. The figures of the results for the least square curve fitting (with latitude included as variable), along with the tables of the mean coefficients for the fitted equation and the standard deviation of the coefficients of the equation, are given in Appendix E to H. The figures for the RBF results for 50 and 100 sampled points are given in Appendix I to L respectively.