

## Abstract

In this dissertation, the aim is to investigate the empirical relationship between the partial pressure of CO<sub>2</sub> (pCO<sub>2</sub>) and other ocean variables in the Southern Ocean, by using a small percentage of the available data.

CO<sub>2</sub> is one of the main greenhouse gases that contributes to global warming and climate change. The concentration of anthropogenic CO<sub>2</sub> in the atmosphere, however, would have been much higher if some of it was not absorbed by oceanic and terrestrial sinks. The oceans absorb and release CO<sub>2</sub> from and to the atmosphere. Large regions in the Southern Ocean are expected to be a CO<sub>2</sub> sink. However, the measurements of CO<sub>2</sub> concentrations in the ocean are sparse in the Southern Ocean, and accurate values for the sinks and sources cannot be determined. In addition, it is difficult to develop accurate oceanic and ocean-atmosphere models of the Southern Ocean with the sparse observations of CO<sub>2</sub> concentrations in this part of the ocean.

In this dissertation classical techniques are investigated to determine the empirical relationship between pCO<sub>2</sub> and other oceanic variables using *in situ* measurements. Additionally, sampling techniques are investigated in order to make a judicious selection of a small percentage of the total available data points in order to develop an accurate empirical relationship.

Data from the SANAE49 cruise stretching between Antarctica and Cape Town are used in this dissertation. The complete data set contains 6103 data points. The maximum pCO<sub>2</sub> value in this stretch is 436.0  $\mu\text{atm}$ , the minimum is 251.2  $\mu\text{atm}$  and the mean is 360.2  $\mu\text{atm}$ . An empirical relationship is investigated between pCO<sub>2</sub> and the variables Temperature (T), chlorophyll-a concentration (Chl), Mixed Layer Depth (MLD) and latitude (Lat). The methods are repeated with latitude included and excluded as variable respectively. D-optimal sampling is used to select a small percentage of the available data for determining the empirical relationship. Least squares optimization is used as one method to determine the empirical relationship. For 200 D-optimally sampled points, the pCO<sub>2</sub> prediction with the fourth order equation yields a Root Mean Square (RMS) error of 15.39  $\mu\text{atm}$  (on the estimation of pCO<sub>2</sub>) with latitude excluded as variable and a RMS error of 8.797  $\mu\text{atm}$  with latitude included as variable. Radial basis function (RBF) interpolation is another method that is used to determine the empirical relationship between the variables. The RBF interpolation with 200 D-optimally sampled points yields a RMS error of 9.617  $\mu\text{atm}$  with latitude excluded as variable and a RMS error of 6.716  $\mu\text{atm}$  with latitude included as variable. Optimal scaling is applied to the variables in the RBF interpolation, yielding a RMS error of 9.012  $\mu\text{atm}$  with latitude excluded as variable and a RMS error of 4.065  $\mu\text{atm}$  with latitude included as variable for 200 D-optimally

# Investigating the empirical relationship between oceanic properties observable by satellite and the oceanic pCO<sub>2</sub>

M.VAN DER WALT

20238703

Dissertation submitted in fulfilment of the requirements for the degree  
*Master of Science in Applied Mathematics* at the Potchefstroom Campus of  
the North-West University

Supervisor: Prof G. Drevin

Co-supervisor: Dr S. Kok

Assistant supervisor: Prof J. Spoelstra

November 2011

sampled points.

## **Keywords**

Southern Ocean; Global warming; Genetic algorithm; Radial basis function interpolation; Curve fitting

## Opsomming

Die doel van hierdie verhandeling is om die empiriese verwantskap tussen partiële druk van CO<sub>2</sub> (pCO<sub>2</sub>) en ander oseaneienskappe te ondersoek in die Suidelike Oseaan deur slegs 'n klein persentasie van die beskikbare data te gebruik.

CO<sub>2</sub> is een van die primêre kweekhuisgasse wat bydra tot aardverwarming en klimaatsveranderinge. Die konsentrasie van antropogeniese CO<sub>2</sub> in die atmosfeer sou egter baie hoër gewees het as 'n gedeelte daarvan nie deur die oseane en landelike gebiede opgeneem was nie. Die oseane neem CO<sub>2</sub> op vanaf die atmosfeer en stel ook CO<sub>2</sub> vry in die atmosfeer. Daar word beweer dat groot areas in die Suidelike Oseaan 'n CO<sub>2</sub> sink is, menende dat dit meer CO<sub>2</sub> opneem vanuit die atmosfeer as wat dit vrylaat. Die lesings van CO<sub>2</sub> in die Suidelike Oseaan is egter yl, en dus kan die mate waartoe die oseaan 'n CO<sub>2</sub> sink of bron is, nie akkuraat bepaal word nie. Benewens is dit moeilik om akkurate oseaan en oseaan-atmosfeer modelle te ontwikkel vir die Suidelike Oseaan met die yl gesaaide metings van CO<sub>2</sub>.

In hierdie verhandeling word klassieke tegnieke ondersoek, om die empiriese verwantskap tussen pCO<sub>2</sub> en ander oseaneienskappe te bepaal deur *in situ* metings te gebruik. Verder word die seleksie van 'n klein persentasie punte (van die volledige data stel) ondersoek. Die doel is om oordeelkundig 'n seleksie van punte te maak, sodanig dat die fout op die benadering van pCO<sub>2</sub> vanuit die empiriese verwantskap so klein as moontlik is.

Die data wat gebruik word in hierdie verhandeling, is verkry deur lesings wat geneem is gedurende die SANAE49 skeepsvaart tussen Antarktika en Kaapstad. Die volledige datastel bevat 6103 punte. Die maksimum pCO<sub>2</sub> waarde gedurende die vaart is 436.0  $\mu\text{atm}$ , die minimum is 251.2  $\mu\text{atm}$  en die gemiddeld is 360.2  $\mu\text{atm}$ . 'n Empiriese verwantskap word ondersoek tussen pCO<sub>2</sub> en Temperatuur (T), chlorofil-a konsentrasie (Chl), Gemengde Laag Diepte (MLD) en breedtegraad (Lat). Die metodes word herhaal met en sonder breedtegraad onderskeidelik. D-optimale seleksie word gebruik om 'n klein persentasie punte van die volledige datastel te selekteer waarmee die empiriese verwantskap bepaal word. Kleinste-kwadrante-optimalisering is een metode wat gebruik word om die empiriese verwantskap te ondersoek. Vir 200 punte wat met D-optimale seleksie gekies is, lewer die 4de orde vergelyking 'n wortel-gemiddelde-kwadraat-fout (RMS) van 15.39  $\mu\text{atm}$  (op die benadering van pCO<sub>2</sub>) sonder breedtegraad ingesluit as veranderlike en 'n RMS fout van 8.797  $\mu\text{atm}$  met breedtegraad ingesluit as veranderlike. Radiale basisfunksie-interpolasie (RBF) is nog 'n metode wat gebruik is om die empiriese verwantskap tussen die veranderlikes te bepaal. Die RBF interpolasie, wat gedoen is met 200 punte gekies deur D-optimale seleksie, lewer 'n RMS fout van 9.617  $\mu\text{atm}$  sonder breedtegraad as veranderlike en 'n RMS fout van 6.716  $\mu\text{atm}$  met breedtegraad ingesluit as veran-

derlike. Wanneer die veranderlikes optimaal skaleer word met die RBF interpolasie, met 200 punte wat met D-optimale seleksie gekies is, lewer dit 'n RMS fout van  $9.012 \mu\text{atm}$  sonder breedtegraad as veranderlike en 'n RMS fout van  $4.064 \mu\text{atm}$  met breedtegraad ingesluit as veranderlike.

## Sleutelwoorde

Suidelike Oseaan; Aardverwarming; Genetiese algoritme; Radiale basisfunksie-interpolasie; Kromme-passing

# Contents

<b>Definitions</b>	<b>13</b>
<b>Acronyms</b>	<b>17</b>
<b>Nomenclature</b>	<b>19</b>
<b>List of Figures</b>	<b>22</b>
<b>List of Tables</b>	<b>26</b>
<b>1 Introduction</b>	<b>29</b>
1.1 Background . . . . .	29
1.2 Previous work . . . . .	30
1.3 Scope of dissertation . . . . .	31
1.4 Layout of dissertation . . . . .	32
<b>2 Literature Review: CO<sub>2</sub> and the oceans</b>	<b>35</b>
2.1 SRP Project . . . . .	35
2.2 Carbon dioxide and the oceans . . . . .	36
2.2.1 Background . . . . .	36
2.2.2 Data availability . . . . .	38

2.2.3	Factors affecting the CO <sub>2</sub> fluxes . . . . .	38
2.2.4	Ocean-atmosphere CO <sub>2</sub> transfers . . . . .	41
2.3	Conclusion . . . . .	42
<b>3</b>	<b>Literature Review: Methods</b>	<b>43</b>
3.1	Radial Basis Function interpolation . . . . .	43
3.2	Sampling techniques . . . . .	47
3.2.1	Random Sampling . . . . .	47
3.2.2	D-optimal Sampling . . . . .	47
3.3	Genetic algorithms . . . . .	50
3.3.1	Background . . . . .	50
3.3.2	Genetic algorithms as optimization method . . . . .	51
3.3.3	Steps in a genetic algorithm . . . . .	52
3.3.4	Motivation for using genetic algorithms . . . . .	58
3.4	Conclusion . . . . .	60
<b>4</b>	<b>Data</b>	<b>61</b>
4.1	Introduction . . . . .	61
4.2	Data Collection . . . . .	61
4.2.1	Introduction . . . . .	61
4.2.2	<i>In situ</i> measurements . . . . .	62
4.2.3	GlobColour dataset . . . . .	75
4.2.4	MLD data . . . . .	79
4.2.5	Satellite data . . . . .	83
4.2.6	Argo Floats . . . . .	85

4.2.7	Conclusion . . . . .	86
4.3	Data processing . . . . .	86
4.3.1	<i>In situ</i> data . . . . .	86
<b>5</b>	<b>Methods and Results</b>	<b>93</b>
5.1	Introduction . . . . .	93
5.2	Sampling . . . . .	94
5.2.1	Random sampling . . . . .	95
5.2.2	D-optimal sampling . . . . .	95
5.2.3	Genetic algorithm application in D-optimal sampling . . . . .	97
5.2.4	Selection . . . . .	98
5.2.5	Crossover . . . . .	99
5.2.6	Mutation . . . . .	99
5.2.7	Elitism . . . . .	99
5.2.8	Implementation . . . . .	100
5.3	Least Squares curve fitting . . . . .	104
5.3.1	Excluding latitude as a variable . . . . .	105
	Linear curve fit . . . . .	106
	Quadratic curve fit . . . . .	108
	Cubic curve fit . . . . .	110
	Fourth order curve fit . . . . .	113
	Removing terms . . . . .	119
5.3.2	Including latitude as variable . . . . .	122
	Linear curve fit with latitude included as variable . . . . .	122
	Quadratic curve fit with latitude included as variable . . . . .	124



Cubic curve fit with latitude included as variable . . . . .	126
Fourth order curve fit with latitude included as variable . . . . .	129
Removing terms . . . . .	136
5.4 Radial basis function interpolation . . . . .	141
5.4.1 Introduction . . . . .	141
5.4.2 D-optimal sampled points . . . . .	143
5.4.3 RBF interpolation with variables T, MLD and Chl . . . . .	144
5.4.4 RBF interpolation with variables T, MLD, Chl and Lat . . . . .	153
5.5 Conclusion . . . . .	162
<b>6 Discussions and future work</b>	<b>163</b>
6.1 Discussion . . . . .	163
6.2 Future work . . . . .	166
<b>7 Bibliography</b>	<b>169</b>
<b>Appendices</b>	<b>179</b>
<b>A Linear curve fit results with latitude excluded as variable</b>	<b>179</b>
<b>B Quadratic curve fit results with latitude excluded as variable</b>	<b>189</b>
<b>C Cubic curve fit results with latitude excluded as variable</b>	<b>199</b>
<b>D Fourth order curve fit with latitude excluded as variable</b>	<b>211</b>
<b>E Linear curve fit results with latitude included as variable</b>	<b>221</b>
<b>F Quadratic curve fit results with latitude included as variable</b>	<b>231</b>

G Cubic curve fit results with latitude included as variable	243
H Fourth order curve fit results with latitude included as variable	255
I Radial basis function interpolation with latitude excluded as variable	269
J Radial basis function interpolation with latitude excluded as variable and optimal scaling of variables	275
K Radial basis function interpolation with latitude included as variable	281
L Radial basis function interpolation with latitude included as variabel and optimal scaling of variables	287



# Definitions

*Advection:* The transfer of a property of the atmosphere, such as heat, cold or humidity, by the horizontal movement of an air mass. The rate of change of an atmospheric property caused by the horizontal movement of air [18].

*Anthropogenic:* Anthropogenic refers to something that is caused by humans or human activity [18].

*Buffer capacity:* Buffer capacity is referred to as the relative ability of a buffer solution to resist pH change when an acid or base is added to the solution [17].

*Carbon budget:* An amount of carbon dioxide that a country, company or organization has agreed upon to be the largest amount that it will produce in a particular period of time [3].

*Carbon cycle:* The combined processes of photosynthesis, decomposition and respiration by which carbon compounds are transferred between the major carbon reservoirs, including the atmosphere, ocean and living organisms [18].

*Carbon dioxide flux:* (Or CO<sub>2</sub> flux) Refers to the CO<sub>2</sub> transfer between the ocean and the atmosphere. It is the flow of CO<sub>2</sub> particles through a given surface [18].

*Carbon sink:* A carbon sink is a terrestrial or oceanic area that absorbs carbon dioxide released by the burning of fossil fuels [16].

*Entrainment:* To carry (suspended particles, for example) along in a current [18].

*Fugacity:* The thermodynamic property of a gas that is related to its partial pressure. It is an indication of the tendency of the gas to escape or expand [18].

*Kyoto Protocol:* The Kyoto Protocol is an international agreement linked to the United Nations Framework Convention on Climate Change. In short, the Kyoto protocol sets targets for 37 industrialized countries and the European community for reducing greenhouse gas emissions. This protocol

commits the countries to reduce the emissions by 2012. It is generally seen as an important first step towards a fully global emission reduction regime that will contribute to stabilizing greenhouse gas emissions [77].

*Nautical mile:* The nautical mile is a unit of length used in marine navigation [8]. One nautical mile = 1.852 km [8].

*Ocean deep layer:* The deep layer of the ocean is the bottom part of the ocean, and makes up most of the ocean's volume (90% of the ocean) [60], [14]. The density of the water increases as the depth of the ocean increases in the deep layer [60]. The colder, denser waters sink down to the deep waters of the ocean where the deep ocean layer acts as a "store" for these waters [14].

*Ocean surface mixed layer:* The ocean surface mixed layer is assumed to be approximately the upper 500m of the ocean [58]. This is the layer that is considered when studying climate, biological activity and pollution of the ocean [11]. Satellites, ships and aeroplanes can easily monitor the mixed layer dynamics, thus we know the most about this layer of the ocean [60].

*Pack ice:* Ice in the polar regions of the ocean, that consists of floating ice that joined together to cover the sea surface such that there is little or no open water in these parts [18].

*Polar Frontal Zone:* The Polar Frontal Zone in the Southern Hemisphere is the area where the salinity of the ocean is low, between the Antarctic Polar Front and the Subantarctic Front [15]. This zone is shown in Figure 1.

*Radial:* Moving or directed along a radius, or similarly developing symmetrically about a central point for a specified radius [18]. Spreading out or developing uniformly on all sides of a central point [16].

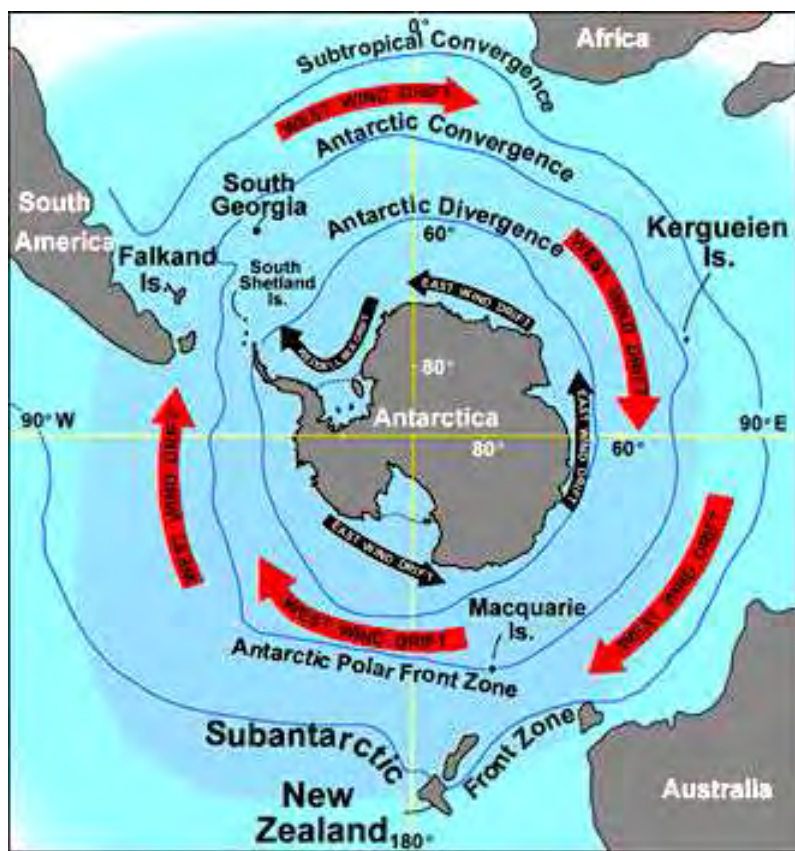


Figure 1: Position of the Polar Frontal Zone [21].



# Acronyms

AATSR	Advanced Along Track Scanning Radiometer
ACC	Antarctic Circumpolar Current
AVHRR	Advanced Very High Resolution Radiometer
CDIAC	Carbon Dioxide Information Analysis Center
CFSR	Climate Forecast System Reanalysis
CLIVAR	Climate Variability
CSIR	Center for Scientific and Industrial Research
DDS	Diagnostic Data Sets
DIC	Dissolved Inorganic Carbon
DoE	Department of Energy
EEZ	Exclusive Economic Zone
GA	Genetic Algorithm
GEOSECS	Geophysical Sections Experiment
GIS	Geographic Information System
GMES	Global Monitoring for Environment and Security
GODAS	Global Ocean Data Assimilation System
HIRS	High-resolution Infra-red Radiation Sender
HPLC	High Performance Liquid Chromatography
IOCCG	International Ocean Colour Coordinating Group
IOCCP	International Ocean Carbon Coordination Project
ISIN	Integerised Sinusoidal projection
JGOFS	Joint Global Ocean Flux Study
LDEO	Latmont Doherty Earth Observatory
MERIS	Medium Resolution Imaging Spectrometer



MLD	Mixed Layer Depth
MODIS	Moderate Resolution Imaging Spectrometer
NODC	National Oceanographic Data Centre
NCEP	National Centres for Environmental Prediction
OC-TAC	Ocean Colour Thematic Assembly Centre
OISST	Optimum Interpolation Sea Surface Temperature
pCO <sub>2</sub>	Partial pressure of Carbon Dioxide
PiRATA	Prediction and Research Moored Array in the Tropical Atlantic
POC	Particulate Organic Carbon
POOZ	Permanent Open Ocean Zone
PSU	Practical Salinity Unit
RAMA	Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction
RBF	Radial Basis Function
RMS	Root Mean Square
SEM	Scanning Electron Microscopy
SIZ	Seasonal Ice Zone
SOCAT	Surface Ocean CO <sub>2</sub> Atlas
SRP	Strategic Research Project
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
SSU	Stratospheric Sounding Unit
THC	Thermohaline Circulation
TOVS	TIROS-N Operational Vertical Sounder
UCTD	Underway Conductivity Temperature and Depth
VOS	Volunteer Observing Ships
WOA05	World Ocean Atlas 2005
WOCE	World Ocean Circulation Experiment
XBT	Expendible Bathythermograph

# Nomenclature

Symbol	Description
$\alpha$	Vector of coefficients for the RBF
$a_i$	The $i^{th}$ scaling coefficient of the variables
$\mathbf{A}$	The $m \times n$ matrix where $m$ is the number of rows in the entire data set and $n$ is the number of terms in the equation that is considered
$\alpha_j$	Coefficient of the $j^{th}$ term in the radial basis function
$\mathbf{B}$	The $k \times n$ matrix obtained by extracting $k$ rows of matrix $\mathbf{A}$
$\beta$	Vector of coefficients for the polynomial part of RBF interpolation
Chl	Chlorophyll-a concentration
Chl <sub><math>j</math></sub>	$j^{th}$ chlorophyll-a concentration value
Chl <sub><math>max</math></sub>	Maximum chlorophyll-a concentration value
Chl <sub><math>min</math></sub>	Minimum chlorophyll-a concentration value
$e$	Measurement error
$\mathbf{E}$	Error on the estimation
$f$	Fitness
fCO <sub>2</sub>	Fugacity of CO <sub>2</sub>
$\gamma$	Vector of coefficients of the RBF interpolation $\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$
$\mathbf{I}$	Identity matrix
$k$	Number of points sampled from the data set for processing
Lat	Latitude
Lat <sub><math>j</math></sub>	$j^{th}$ latitude value
Lat <sub><math>max</math></sub>	Maximum latitude value
Lat <sub><math>min</math></sub>	Minimum latitude value

$m$	Number of points in the complete data set
$\mathbf{M}$	Fisher information matrix
$\mathbf{M}_{b,b}$	Matrix for the RBF part of the RBF interpolation
MLD	Mixed layer depth
$MLD_A$	Mixed layer depth at point $A$
$MLD_B$	Mixed layer depth at point $B$
$MLD_C$	Mixed layer depth at point $C$
$MLD_j$	$j^{th}$ MLD value
$MLD_{max}$	Maximum MLD value
$MLD_{min}$	Minimum MLD value
$n_b$	Number of points selected for RBF interpolation
$OF$	Objective function
$\mathbf{P}$	Variance covariance matrix
$pCO_2$	Partial pressure of $CO_2$
$\mathbf{P}_b$	Matrix for the polynomial part of the RBF interpolation
$P_{equ}$	Pressure in the equilibrator
$pH_2O$	Water vapour pressure of $CO_2$
$\phi$	Radial basis function
$P_{water}$	Equilibrium of vapour pressure at the temperature of equilibration
$r_j$	Radius defined as $\  \mathbf{x} - \mathbf{x}_j \ $
$\mathbf{R}$	Matrix for RBF interpolation
$\mathbf{S}$	Sensitivity matrix
$SS$	Specific sampling schedule
SST	Sea surface temperature
$\sigma$	Standard deviation
$\sigma^2$	Variance
$\sigma_{random}$	Standard deviation of randomly sampled points
$\sigma_{D-optimal}$	Standard deviation of D-optimally sampled points

$T$	Sea surface temperature
$T_{eq}$	Temperature of equilibration
$T_j$	$j^{th}$ Temperature value
$T_{max}$	Maximum temperature value
$T_{min}$	Minimum temperature value
$u$	Input signal
$x_i$	$i^{th}$ $x$ value
$x_{max}$	Maximum $x$ value
$x_{min}$	Minimum $x$ value
$x_{scaled}$	Scaled $x$ values
$XCO_2$	Mole fraction of $CO_2$
$\xi$	The ratio of the $x$ value divided by the radius $r$
$y_{avg}$	Average modelled $pCO_2$ value
$y_i$	Modelled $pCO_2$ values for the $i^{th}$ point in space
$\mathbf{y}$	Actual $pCO_2$ values
$\tilde{\mathbf{y}}$	Modelled $pCO_2$ values



# List of Figures

1	Position of the Polar Frontal Zone. . . . .	15
3.1	An example illustrating Roulette wheel selection. . . . .	54
3.2	An example illustrating single point crossover. . . . .	55
3.3	An example illustrating two point crossover. . . . .	56
4.1	The ACC and other currents in the Southern Ocean. . . . .	63
4.2	Map of the SANAE cruise. . . . .	65
4.3	SOCAT database areas covered in the ocean. . . . .	69
4.4	LDEO database. . . . .	72
4.5	Chlorophyll-a concentration distribution by GlobColour. . . . .	78
4.6	ARGO floats distribution in the ocean. . . . .	85
4.7	The pCO <sub>2</sub> data in the SANAE49L6 data set. . . . .	89
4.8	The temperature data in the SANAE49L6 data set. . . . .	89
4.9	The MLD data in the SANAE49L6 data set. . . . .	90
4.10	The chlorophyll-a concentration in the SANAE49L6 data set. . . . .	90
4.11	Multivariate plot of the SANAE49L6 data set. . . . .	91
5.1	Increasing fitness of genetic algorithm. . . . .	101
5.2	Histogram of selected points of GA. . . . .	101

5.3	Distribution of points - T-MLD. . . . .	102
5.4	Distribution of points - T-Lat. . . . .	102
5.5	Distribution of points - T-Chl. . . . .	102
5.6	Distribution of points - Lat-Chl. . . . .	103
5.7	Distribution of points - MLD-Chl. . . . .	103
5.8	Distribution of points - MLD-Lat. . . . .	103
5.9	Fourth order: results with 400 randomly selected points. . . . .	117
5.10	Fourth order: indication of 10% error with 400 randomly selected points. . . . .	117
5.11	Fourth order: histogram of errors with 400 randomly selected points. . . . .	117
5.12	Fourth order: results with 400 D-optimally selected points. . . . .	118
5.13	Fourth order: indication of 10% error with 400 D-optimally selected points. . . . .	118
5.14	Fourth order: histogram of errors with 400 D-optimally selected points. . . . .	118
5.15	Error vs. terms removed. . . . .	121
5.16	Fourth order: results with 400 randomly sampled points (Lat included). . . . .	134
5.17	Fourth order indication of 10% error with 400 randomly sampled points (Lat included).134	
5.18	Fourth order: histogram of errors with 400 randomly sampled points (Lat included). . 134	
5.19	Fourth order: results with 400 D-optimally sampled points (Lat included). . . . .	135
5.20	Fourth order indication of 10% error with 400 D-optimally sampled points (Lat included).135	
5.21	Fourth order: histogram of errors with 400 D-optimally sampled points (Lat included). 135	
5.22	Error vs. Terms removed (Lat included). . . . .	137
5.23	RBF interpolation: results with 200 randomly sampled points. . . . .	146
5.24	RBF interpolation: indication of 10% error with 200 randomly sampled points. . . . .	146
5.25	RBF interpolation: histogram of errors with 200 randomly samples points. . . . .	146
5.26	RBF interpolation: results with 200 D-optimally sampled points. . . . .	147

5.27 RBF interpolation: indication of 10% error with 200 D-optimally sampled points. . . .	147
5.28 RBF interpolation: histogram of errors with 200 D-optimally samples points. . . . .	147
5.29 RBF interpolation: results with 200 randomly sampled points and optimal scaling of variables. . . . .	151
5.30 RBF interpolation: indication of 10% error with 200 randomly sampled points and optimal scaling of variables. . . . .	151
5.31 RBF interpolation: histogram of errors with 200 randomly sampled points and optimal scaling of variables. . . . .	151
5.32 RBF interpolation: results with 200 D-optimally sampled points and optimal scaling of variables. . . . .	152
5.33 RBF interpolation: indication of 10% error with 200 D-optimally sampled points and optimal scaling of variables. . . . .	152
5.34 RBF interpolation: histogram of errors with 200 D-optimally sampled points and optimal scaling of variables. . . . .	152
5.35 RBF interpolation: results with 200 randomly sampled points (Lat included). . . . .	155
5.36 RBF interpolation: indication of 10% error with 200 randomly sampled points (Lat included). . . . .	155
5.37 RBF interpolation: histogram of errors with 200 randomly sampled points (Lat included). . . . .	155
5.38 RBF interpolation: results with 200 D-optimally sampled points (Lat included). . . .	156
5.39 RBF interpolation: indication of 10% error with 200 D-optimally sampled points (Lat included). . . . .	156
5.40 RBF interpolation: histogram of errors with 200 D-optimally sampled points (Lat included). . . . .	156
5.41 RBF interpolation: results with 200 randomly sampled points and optimal scaling of variables (Lat included). . . . .	160
5.42 RBF interpolation: indication of 10% error with 200 randomly sampled points and optimal scaling of variables (Lat included). . . . .	160



5.43 RBF interpolation: histogram of errors with 200 randomly sampled points and optimal scaling of variables (Lat included). . . . .	160
5.44 RBF interpolation: results with 200 D-optimally sampled points and optimal scaling of variables (Lat included). . . . .	161
5.45 RBF interpolation: indication of 10% error with 200 D-optimally sampled points and optimal scaling of variables (Lat included). . . . .	161
5.46 RBF interpolation: histogram of errors with 200 D-optimally sampled points and optimal scaling of variables (Lat included). . . . .	161

# List of Tables

3.1	Radial basis functions with global support. . . . .	45
3.2	Radial basis functions with local support. . . . .	46
3.3	Wu's compactly supported RBFs. . . . .	46
4.1	Summary of statistics of variables in the final SANAE49L6 data set. . . . .	88
5.1	Results: Linear curve fitting. . . . .	107
5.2	Ratio of standard deviation of coefficients of linear curve fitting for 200 sampled points. . . . .	107
5.3	Results: Quadratic curve fitting. . . . .	109
5.4	Ratio of standard deviation of coefficients of quadratic curve fitting for 200 sampled points. . . . .	109
5.5	Results: Cubic curve fitting. . . . .	111
5.6	Ratio of standard deviation of coefficients of cubic curve fitting for 200 sampled points. . . . .	112
5.7	Results: Fourth Order curve fitting. . . . .	114
5.8	Ratio of standard deviation of coefficients of fourth order curve fitting for 200 sampled points. . . . .	115
5.9	Table of RMSE when systematically removing terms for the fourth order equation. . . . .	119
5.10	Results: Linear curve fitting. . . . .	123
5.11	The ratio of the standard deviation of coefficients of linear curve fitting with latitude included as variable for 200 sampled points. . . . .	123
5.12	Results: Quadratic curve fitting. . . . .	125

5.13	The ratio of the standard deviation of coefficients of quadratic curve fitting with latitude included as variable for 200 sampled points. . . . .	125
5.14	Results: Cubic curve fitting. . . . .	127
5.15	The ratio of the standard deviation of coefficients of cubic curve fitting with latitude included as variable for 200 sampled points. . . . .	127
5.16	Results: Fourth Order curve fitting. . . . .	130
5.17	The ratio of the standard deviation of coefficients of fourth order curve fitting with latitude included as variable for 200 sampled points. . . . .	131
5.18	Table of RMSE when systematically removing terms from the fourth order equation with latitude included. . . . .	138
5.19	Results of RBF interpolation with variables T,MLD and Chl with the 0-2 scaled data set. . .	145
5.20	Optimal scaling found for different points selected by D-optimal sampling. . . . .	149
5.21	Range of values for the variables. . . . .	149
5.22	Results of RBF interpolation for the variables T,MLD and Chl with optimally scaled variables.	150
5.23	Results of RBF interpolation with variables T,MLD, Chl and Lat with the 0-2 scaled data set.	154
5.24	Optimal scaling for consecutive iterations of the optimization. . . . .	158
5.25	Range of values for the variables. . . . .	158
5.26	Results of RBF interpolation for variables T, MLD, Chl and Lat with optimally scaled variables.	159