Mapping the fugacity of CO_2 in the Nordic Seas and the northern North Atlantic Ocean

Master Thesis in
Chemical Oceanography

Elise Kvæstad

December 2013

UNIVERSITY OF BERGEN
GEOPHYSICAL INSTITUTE

Contents

Acknowledgements

I would like to thank my supervisors Are Olsen, Siv Kari Lauvset and Truls Johannessen for the opportunity to work with such an interesting thesis. In particularly I would like to thank Are for his guidance and patience along the way.

I would extend my gratitude to Dr. Maciek Telszewski who took the time to come to Bergen to share his knowledge of self organizing maps. Without his Matlab scripts, this work would have been impossible.

I want to thank Benjamin Pfeil for compiling the unpublished SOCAT data for me. Thanks to Sergio Signorini for providing the HYCOM data, which I unfortunately never got time to use, and thanks to Torbjørn Taskjelle for his helping hand with LaTeX.

I would also like to thank my family and friends for supporting me and always being there. Finally, a special thank to Frank, my boyfriend, for always believe in me and encourage me in difficult times.

Abstract

Self organizing maps (SOMs) have been used to estimate seawater fugacity of CO_2 (fCO_2) distribution in the Nordic Seas and the northern North Atlantic Ocean for 2005-2007. Four maps were produced for each region using various combinations of the parameters MLD, SST, CHL and SSS to train the SOM. The combination of MLD+SST gave the most realistic fCO_2 map for all the basins, which was surprising as one would expect that more input parameters would generate more realistic maps. SOMs were estimated in the northern North Atlantic Ocean and the Nordic Seas, the northern North Atlantic Ocean, and the Nordic Seas. The SOMs for the Nordic Seas were also merged with the SOMs for the northern North Atlantic Ocean, which generated better results than the SOM for the whole study area. The merged SOMs had minimal issues with discontinuity. All the SOMs generated incorrect fCO_2 values for the deepest mixed layer depths in the Greenland Sea. A theory is that this may be the result of an artifact in the training dataset. This illustrate the importance of having a broad data coverage when using the SOM technique. The most optimal SOM for the Nordic Seas provided realistic estimates of fCO_2 distribution. This SOM had an uncertainty of 12.5 µatm, which is close to the LSCOP target of 10.8 µatm for estimation of net fCO_2 flux for the Northern North Atlantic.

Abbreviations

BMU	Best matching unit
CCD	Carbonate Compensation Depth
CFCs	Chlorofluorocarbons
CF	Cost Function
CH_4	Methane
CHL	Chlorophyll
CO_2	Carbon dioxide
CO_3^{2-}	Carbonate
$CaCO_3$	Calcium carbonate
DIC	Dissolved Inorganic Carbon
D	Euclidean Distance
GLORYS	GLobal Ocean ReanalYsis and Simulation
H^+	Hydrogen cation
H_2CO_3	Carbonic acid
HCO_3^-	Bicarbonate
ME	Modeling Efficiency
MLD	Mixed Layer Depth
MLP	Multilayer Perceptron Neural Network
MLR	Multivariable Linear Regression
NN	Neural network
NOAA	National Oceanic and Atmospheric Administration
PRSOM	Probabilistic Regularized Self Organizing Map
RMSE	Root-Mean-Square Error
SOCAT	Surface Ocean CO_2 ATlas

SOM	Self Organizing Map
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
SeaWiFS	Sea-viewing Wide Field-of-view Sensor
TA	Total Alkalinity
$\boldsymbol{O_i}$	the ith number of n observations
$\boldsymbol{P_i}$	the ith number of n predictions
$\overline{\boldsymbol{O}}$	averaged observations
$\overline{\boldsymbol{P}}$	averaged predictions
$\boldsymbol{f}\mathbf{CO_2}$	fugacity of CO_2
$\boldsymbol{p}\mathbf{CO_2}$	partial pressure of CO_2
$\boldsymbol{w}_i$	weight vector
$\boldsymbol{x}_i$	input vector
n	number
ppm	parts per million
psu	practical salinity unit
qe	quantization error
$\mathbf{r}^2$	the coefficient of determination
te	topographic error

Chapter 1

Introduction

1.1 Greenhouse Effect and Global Climate

Earth's climate and biogeochemistry are closely intertwined, as the gases responsible for the greenhouse effect takes part in biogeochemical cycles.The greenhouse effect is the result of absorption and re-radiation of energy by atmospheric greenhouse gases and particles, a process that gives a downward flux of radiation from the atmosphere to the surface. 30% of the incoming solar radiation is reflected back to space, giving the Earth an average heat flux of 235 Wm^{-2} and making it a habitable planet. Without the atmosphere and the greenhouse effect the equilibrium temperature on Earth would be 255 K (−18 °C), whereas the actual average temperature is remarkable 33 K warmer, 288 K (15 °C) (Sarmiento and Gruber, 2006). Fig. 1.1 shows the global energy fluxes through the atmosphere.

In order to expand our knowledge of the biogeochemical cycling and how it affects the global climate, it is essential to investigate the biogeochemical interactions between the different reservoirs, namely atmosphere, ocean and land.

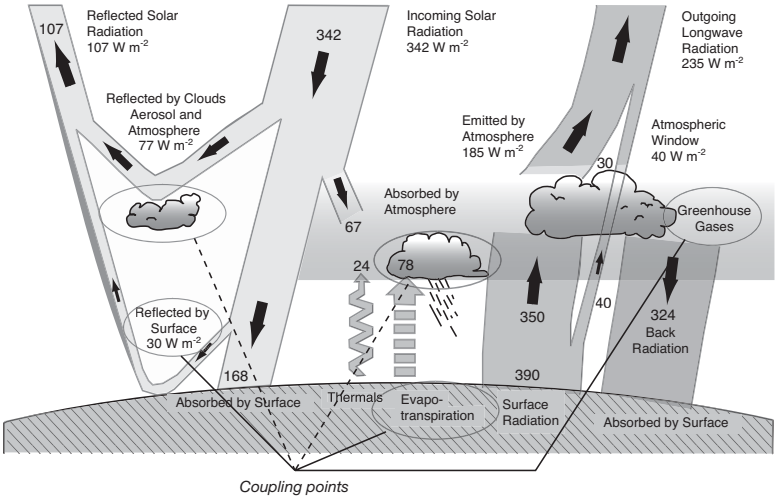


Figure 1.1: Flow of energy through the atmosphere. Numbers are in Wm^{-2}, and some are uncertain by as much as 20%. The coupling points between biogeochemical cycles and the physical climate system are also shown. Figure from Sarmiento and Gruber (2006).

1.1.1 Mechanism of Greenhouse Effect

The physical mechanism of the greenhouse effect is the blocking of outgoing long-wave radiation in our atmosphere, which affects the amount of heat trapped within it. The basic mechanism, on a macro

level, is as follows: The Earth receives energy from the sun in the form of visible light; short-wave radiation from the sun reaches the uppermost atmosphere and travels towards the Earth's surface. Along the way, more than half of the incoming solar radiation is either backscattered to space by clouds, dust and atmospheric gas molecules, or absorbed. When the solar energy reaches the Earth, some of it is absorbed by the surface and heats up the planet, while the remaining is reflected back as a function of the surface's albedo. Albedo is a measure of the reflecting power of a surface. The greater the albedo, the greater the reflection. The ocean surface has a low albedo of around 10%, except for the Arctic and Antarctic where ice fields replace open water (Holden, 2012). Various surfaces and their corresponding albedo values are presented in Tab. 1.1.

Table 1.1: Values of albedo for various surfaces

Surface	Albedo (%)
Fresh, dry snow	80-95
Sea ice	30-40
Dry light sandy soil	35-45
Meadows	15-25
Dry steppe	20-30
Coniferous forest	10-15
Deciduous forest	15-20

Source: (Holden, 2012)

1.1.2 Greenhouse Gases

The atmosphere consists of numerous gases that vary in concentration. Nitrogen (N_2) is the most abundant component, and constitutes 78% of the atmosphere, followed by 21% oxygen (O_2) and 0.9% argon (Ar). The primary greenhouse gases are water vapor (H_2O), carbon dioxide (CO_2) and methane (CH_4). Nitrous oxide (N_2O), ozone (O_3) and chlorofluorocarbons (CFCs) are other gases that contribute to the greenhouse effect, but to a lesser extent. These will not be discussed.

Although water vapor is the most abundant greenhouse gas, the atmospheric concentration is more or less controlled by natural processes linked to the water cycle. The single largest contributor to the increase in the greenhouse effect leading to climate change is CO_2, which now represents 0.04% of the atmosphere, and its levels have been increasing for the past 250 years due to combustion of fossil fuels such as coal, oil and gas, in combination with deforestation. (Sarmiento and Gruber, 2006).

The third most abundant greenhouse gas is methane, which is produced by cattle, land farming and decomposing trash. Methane is considered to be the most powerful greenhouse gas on a molecular level, and its atmospheric levels have increased by 250% since the industrial revolution in 1750 (Trujillo and Thurman, 2010). However, as it accounts for only 0.00017% of the atmosphere its net radiative forcing effect is smaller than for CO_2.

Despite the negative focus on greenhouse gases, their existence are essential for the habitation on Earth. Yet, continued emissions of greenhouse gases will lead to increased temperatures, which in turn will cause changes in climate. Therefore, it is important to increase our knowledge of the chemistry of the atmosphere and the greenhouse gases, with CO_2 being the topic for this thesis.

1.2 The Anthropogenic Perturbation

Over the past 250 years the atmospheric carbon dioxide concentration has increased by nearly 40%, from a pre-industrial level of approximately 280 parts per million (ppm) (Doney et al., 2009) to 393 ppm in 2013 (Tans, n.d.). The atmospheric CO_2 increase is attenuated by oceanic uptake, which is

estimated to account for approximately 1/3 of the total anthropogenic carbon added to the atmosphere (Sabine et al., 2004). Without this carbon drawdown from the atmosphere to the ocean, the levels of atmospheric CO_2 would be approximately 450 ppm today (Doney et al., 2009). Although the CO_2 moves from the atmosphere to the ocean, this does not solve the problem with excessive CO_2 emissions. Ocean CO_2 uptake leads to ocean acidification, as the uptake results in reduction in pH. Acidification is often referred to as the "other CO_2 problem" (Doney et al., 2009), as it prevents formation of calcareous structures by organisms such as phytoplankton. This in turn may disrupt the whole marine food chain.

1.2.1 Trends in Atmospheric Carbon Dioxide

In 1958 the American scientist Charles David Keeling began atmospheric CO_2 measurements at the Mauna Loa Observatory in Hawaii. The observatory is situated near the summit of the volcano Mauna Loa, at an altitude of 3400 m. The goal was to create records of CO_2 in air masses that would be representative for large parts of the northern hemisphere, and with a bit of luck, the globe. The main principle was to select a location that eliminated the influence of large spikes of CO_2 absorbed or emitted, either by plants and soils or by human activities (Tans and Thoning, 2008). His choice of place has been questioned due to the fact that volcanoes outgas CO_2. But looking at the actual levels recorded (Fig. 1.2), it is evident that no volcanic emissions have affected the measurements. Generally the curve has a nice and steady trend with seasonal variations. The annual cycle shows a peak in CO_2 concentration in May, while the lowest concentrations of CO_2 occur during fall. Random dips or spikes indicating volcanic activity are absent in the cycle.

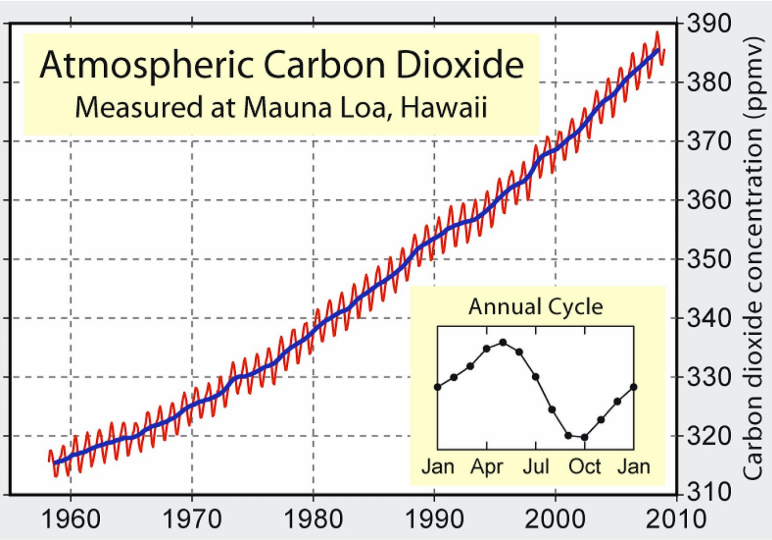


Figure 1.2: Records of atmospheric CO_2 at Mauna Loa Observatory in Hawaii from 1958 to 2009. The red curve shows monthly measurements of CO_2, while the blue curve represent the annual average. Figure from Rohde (2011).

To examine the historical development of anthropogenic CO_2, scientists needed CO_2 records with longer timescales. Law Dome in Antarctica became the solution. By using ice cores obtained at Law Dome from 1987 to 1993, one was able to obtain CO_2 records going back as far as 1006 A.D. The Law Dome site fulfilled many of the criteria required to make it the place of choice to reconstruct historical atmospheric CO_2 records (Etheridge et al., 1998). Fig. 1.3 shows the CO_2 concentration from pre-industrial time to present. Values from 1960 and before are from the the Law Dome ice core in Antarctica, while values from the last 50 years are from the Mauna Loa Observatory.

The remediation of the historical trends in atmospheric CO_2 concentration are of tremendous importance to understand the changes in the Earth's climate. No other indicator is as complete,

updated and direct as the atmospheric CO_2. Needless to say, the oceanic uptake of CO_2 from the atmosphere is an important mechanism to reduce the atmospheric CO_2 levels. In the following section, oceanic CO_2 uptake will be introduced.

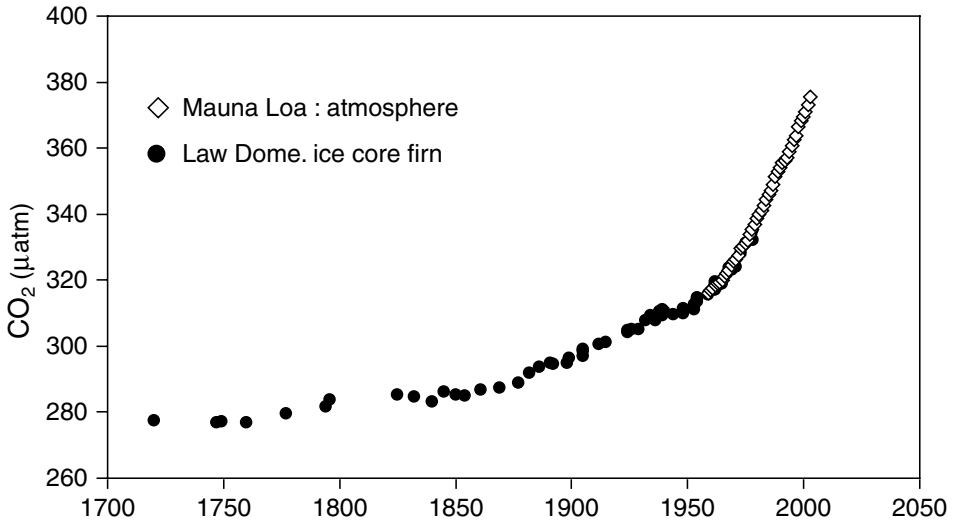


Figure 1.3: The atmospheric CO_2 records at Mauna Loa Observatory in Hawaii from 1958 to present, together with preindustrial CO_2 records from the Law Dome Ice Core in Antarctica. Figure from Emerson and Hedges (2008).

1.3 The Carbon Cycle in the Ocean

1.3.1 Inorganic Carbon Chemistry

When atmospheric carbon dioxide dissolves in seawater, it forms aqueous CO_2 ($CO_2{}^{aq}$). $CO_2{}^{aq}$ reacts with water to form carbonic acid (H_2CO_3), which in two steps dissociate to form bicarbonate ($HCO_3{}^-$) and carbonate ($CO_3{}^{2-}$).

$$\begin{aligned} CO_2^{g} &\rightleftharpoons CO_2^{aq} \\ CO_2^{aq} + H_2O &\rightleftharpoons H_2CO_3 \\ H_2CO_3 &\rightleftharpoons H^+ + HCO_3^- \\ HCO_3^- &\rightleftharpoons H^+ + CO_3^{2-} \end{aligned} \tag{1.1}$$

Simplifying the above reactions yields

$$CO_2 + H_2O \overset{K_1}{\rightleftharpoons} HCO_3^- + H^+ \overset{K_2}{\rightleftharpoons} CO_3^{2-} + 2\,H^+ \tag{1.2}$$

where K_1 and K_2 are equilibrium constants. These are used to describe the relationship between the concentration of the species

$$K_1 = \frac{[HCO_3^-][H^+]}{[CO_2]} \tag{1.3}$$

$$K_2 = \frac{[CO_3^{2-}][H^+]}{[HCO_3^-]} \tag{1.4}$$

and are affected by temperature, salinity and pressure (Zeebe and Wolf Gladrow, 2001).

The carbonate species in the ocean are not measured directly, but calculated from measurements of two of the four measurable parameters: pH, TA, DIC and fCO_2. The total concentration of dissolved

inorganic carbon species in the ocean is called dissolved inorganic carbon, commonly abbreviated DIC. DIC consists of three carbonate species; H_2CO_3, HCO_3^- and CO_3^{2-} (Williams and Follows, 2011). In typical seawater, bicarbonate and carbonate are the two major carbon species (Sarmiento and Gruber, 2006).

$$\mathrm{DIC} \approx [\mathrm{HCO_3^-}] + [\mathrm{CO_3^{2-}}] \tag{1.5}$$

Another important carbon system variable is total alkalinity (TA). TA can be defined as a measure of the capacity of water to neutralize acids, or rather, the sum of all titratable bases. An acid (proton donor) donates hydrogen cations (H^+) when dissolved in water, in contrast to a base (proton acceptor) which accepts H^+. Alkalinity as defined by Dickson et al. (2007), is determined as:

$$\begin{aligned}\mathrm{TA} = {} & [\mathrm{HCO_3^-}] + 2\,[\mathrm{CO_3^{2-}}] + [\mathrm{B(OH)_4^-}] + [\mathrm{OH^-}] + [\mathrm{HPO_4^{2-}}] + 2\,[\mathrm{PO_4^{3-}}] \\ & + [\mathrm{SiO(OH)_3^-}] + [\mathrm{NH_3}] + [\mathrm{HS^-}] + \ldots - [\mathrm{H^+}]_\mathrm{F} - [\mathrm{HSO_4^-}] - [\mathrm{HF}] - [\mathrm{H_3PO_4}] - \ldots\end{aligned} \tag{1.6}$$

Bicarbonate and carbonate are by far the most important terms (Sarmiento and Gruber, 2006). This yields

$$\mathrm{TA} \approx [\mathrm{HCO_3^-}] + 2\,[\mathrm{CO_3^{2-}}] \tag{1.7}$$

pH is a measure of the hydrogen ion concentration in an aqueous solution: $\mathrm{pH} = -\log[\mathrm{H^+}]$. Values on the pH scale range from 0 (strongly acidic) to 14 (strongly basic or alkaline), 7 being the neutral solution. The pH scale is logarithmic, each one-unit change corresponds to a ten-fold change in the hydrogen ion concentration (Trujillo and Thurman, 2010).
The fourth and last carbon system parameter is the fugacity of CO_2 (fCO_2). fCO_2 is the same as partial pressure of CO_2, pCO_2, except that the non-ideality of CO_2 is taken into account, and it is approximately 0.3% lower in value. The distribution of fCO_2 is affected by temperature, salinity, DIC and alkalinity (Sarmiento and Gruber, 2006). The warm equatorial regions are typically supersaturated with respect to CO_2 ($fCO_2^{sea} > fCO_2^{atm}$), which leads to CO_2 outgasing to the atmosphere. Undersaturated waters ($fCO_2^{sea} < fCO_2^{atm}$) are found at higher latitudes and can absorb carbon dioxide from the atmosphere (Heinze et al., 1991).

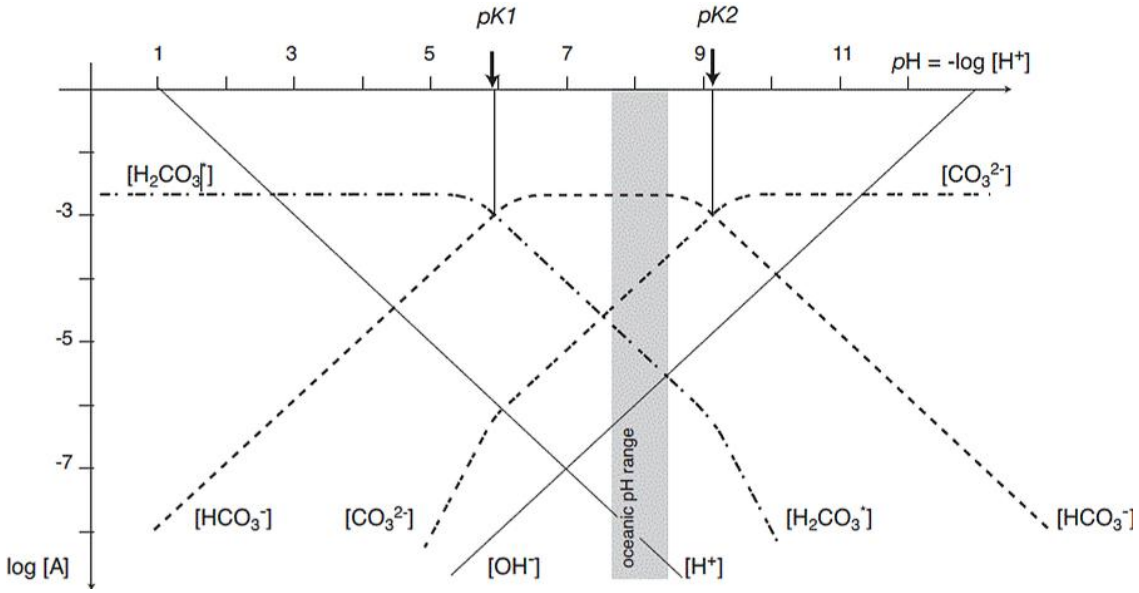


Figure 1.4: Plot of the concentrations of H_2CO_3, HCO_3^-, CO_3^{2-} as functions of pH. Figure from Sarmiento and Gruber (2006).

The Bjerrum plot (Fig. 1.4) shows how the concentrations of the three carbonate species varies as a function of pH. At $\mathrm{pH} = \mathrm{pK_1}$ the concentrations of H_2CO_3 and HCO_3^- are equal, whereas the

concentrations of HCO_3^- and CO_3^{2-} are equal at pH=pK_2. Within the ocean's pH range, marked in Fig. 1.4, the dominant carbonate specie is HCO_3^- (Emerson and Hedges, 2008).

1.3.2 The Solubility Pump

The solubility pump, often referred to as the gas exchange pump, is outlined to the left in Fig. 1.5, and involves the physical processes of transporting carbon from the ocean's surface to its interior and visa versa. Warm surface water circulates from equator towards higher latitudes as part of the large-scale ocean circulation, gradually losing heat to the colder environment. Since the solubility of CO_2 increases with decreasing temperature, the cold waters takes up CO_2 from the atmosphere. When the CO_2-enriched water reaches high latitudes, it sinks to greater depths, and transports large amounts of carbon to the ocean's interior (Trujillo and Thurman, 2010), where it is stored until the water reaches upwelling zones. Subsequent heating of the water decreases the solubility of CO_2, which is then released to the atmosphere.

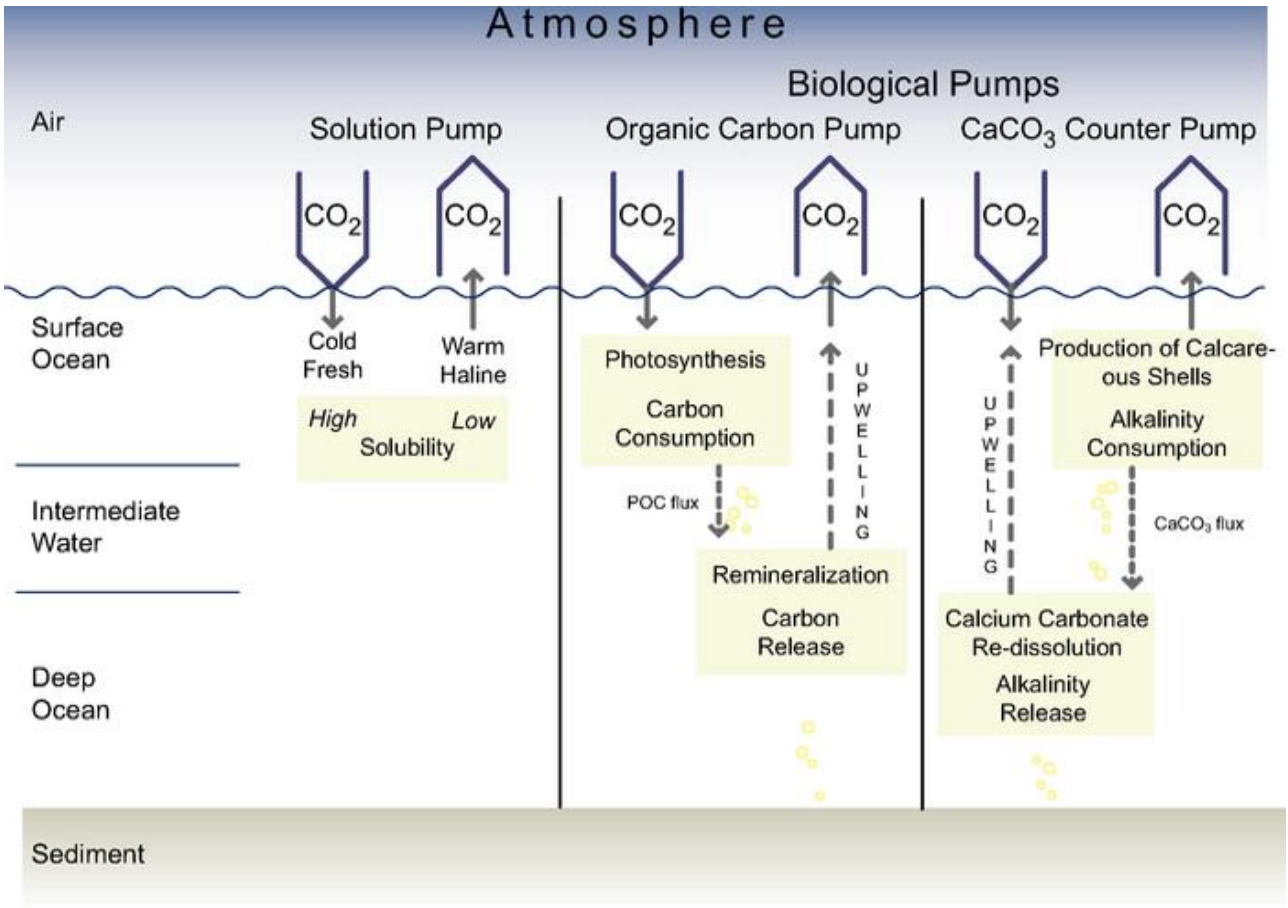


Figure 1.5: The marine carbon pumps; the solubility pump, the organic carbon pump, and the calcium carbonate counter pump. From Heinze et al. (1991) revised by Denman et al. (2007).

1.3.3 The Organic Carbon Pump

The middle panel of Fig. 1.5 shows the organic carbon pump, a component of the biological pump also known as the "soft tissue pump", where photosynthesis and remineralization control the pattern of carbon dioxide. The biological processes in the ocean influencing DIC are the conversion of CO_2 to organic matter through photosynthesis, and the inverse process of respiration and remineralization:

$$106CO_2 + 16HNO_2 + H_3PO_4 + 78H_2O + 18H^+ \underset{\text{Remineralization}}{\overset{\text{Photosynthesis}}{\rightleftharpoons}} C_{106}H_{175}O_{42}N_{16}P + 150O_2 \quad (1.8)$$

As can be seen in reaction (1.8), the concentration of DIC as well as the concentration of free protons (H^+) decrease when organic matter is formed. For organic matter to be produced from inorganic nutrients, it is essential to have access to light, so this process takes place in the surface ocean. Carbon and nutrients are thus consumed in the upper layers by phytoplankton to produce organic matter. A fraction of the organic matter is exported to the abyss and contributes to a net drawdown of CO_2 in the surface (Heinze et al., 1991). During the subsequent remineralization of the organic matter in the water column, the CO_2 is returned to the seawater. Inorganic carbon is thus transported from the surface to the deeper layer. During upwelling of this carbon rich water the CO_2 is released back to the atmosphere (Sarmiento and Gruber, 2006).

1.3.4 The Calcium Carbonate Counter Pump

The biological production of $CaCO_3$ in the surface ocean releases CO_2 to the atmosphere. This process, in addition to the dissolution of $CaCO_3$ in the water column and in the sediments, are the main processes that control the calcium carbonate counter pump, which is shown in Fig. 1.5. This mechanism works in the opposite direction of the organic carbon pump, hence the name calcium carbonate counter pump.

Formation and dissolution of $CaCO_3$, either calcite or aragonite occurs through:

$$2\,HCO_3^- + Ca^{2+} \underset{\text{Dissolution}}{\overset{\text{Formation}}{\rightleftharpoons}} CO_2 + CaCO_3 + H_2O \quad (1.9)$$

As can be seen from equation (1.9), stoichiometric amounts of CO_2 is formed, which can be released to the atmosphere during the formation of $CaCO_3$.

Planktonic species such as coccolithophorids, foraminifera and pteropods form shell structures of calcareous ($CaCO_3$) material (Heinze et al., 1991). Once the calcifying organisms die, their shells sink and dissolve in the water column. The rate of dissolution is mainly determined by two factors: the lysocline and the carbonate compensation depth (CCD). The lysocline is the upper limit where sediments are exposed to corrosive waters, whereas the CCD is the depth where the rates of $CaCO_3$ dissolution and accumulation are equal. Dissolution liberates bicarbonate, which raises the alkalinity. Only a small fraction of the sinking carbonate reaches the ocean seafloor and is buried. The overall process results in a downward transport of DIC and alkalinity from the surface into the deep ocean.

1.4 Thesis Motivation and Aims

1.4.1 Air-sea Gas Exchange of Carbon dioxide

The concentration of some of the atmospheric gases are partially controlled by the mechanisms of air-water exchange (Millero, 2013), with CO_2 being one of them. Understanding how the oceans take up CO_2 from the atmosphere is critical for the prediction of climate change. Air-sea gas exchange of CO_2 is expressed as:

$$F = k_v \alpha (fCO_2^{sea} - fCO_2^{air}) \quad (1.10)$$

where k_v is the gas transfer velocity expressing the rate with which a gas molecule can pass from a gaseous to a liquid phase and vice versa (Watson and Orr, 2003), α is the solubility of CO_2 affected primarily by temperature and fCO_2^{sea} - fCO_2^{air} is the fugacity of CO_2 in surface ocean and air respectively, often expressed as ΔfCO_2 (Phillips, 1994).

The two most important factors controlling air-sea gas exchange are the gas transfer velocity and the CO_2 concentration difference, ΔfCO_2. The transfer velocity is often assumed to be a function

of wind speed (Millero, 2013). Generally, the rate of the gas exchange increases with the wind speed (Phillips, 1994), but there is a high degree of uncertainty concerning this topic.

The oceanic fCO_2 is relatively complex compared to the more uniform atmospheric fCO_2 (Sarmiento and Gruber, 2006). Previously, marine fCO_2 measurements were performed by research vessels that could only cover a fraction of the oceans. Recent technical and financial improvements have ensured that marine fCO_2 data are more effectively collected, with assistance from the shipping industry (Telszewski, 2009). This has led to a large and unique database of fCO_2 measurements. There are still may areas to cover, especially in the Southern Hemisphere. However, installations of equipment on merchant vessels have allowed for new investigations. Ocean uptake of CO_2 can now be studied on daily, weekly and monthly timescales. This has resulted in a greater understanding of the diverse mechanisms controlling ocean uptake of CO_2, such as changes in ocean circulation, biological activity and today's hot topic, climate change.

1.4.2 Previous Work

There are numerous ways to estimate the magnitude and variability of air-sea CO_2 fluxes. Some are based on measurements of the oceanic fCO_2, done at vessels and ships, while others use atmospheric CO_2 mixing ratio measurements (Rödenbeck et al., 2003). Takahashi et al. (2009) are known for their effort to estimate global sea-air CO_2 fluxes based on climatological surface ocean fCO_2. There have been several approaches towards regional estimates, most of them exploiting the relationship between fCO_2 and oceanic state variables. The most common of these is the multivariable linear regression (MLR) (Olsen et al., 2008; Lauvset et al., 2013; Chierici et al., 2012), whereas self organizing maps (Telszewski et al., 2009) have been used a lot less frequently.

1.4.3 Aim of the Thesis

Previous mapping of the Nordic Seas has been done by using MLR (Olsen et al., 2003). However, this method can usually only be applied to one basin at a time since the MLRs, expressing the net effect of the dominating processes affecting fCO_2, is usually quite regionally specific. In this thesis I will explore whether self organizing maps can be used to produce seamless mapping of surface fCO_2 and air-sea exchange of CO_2 over the Nordic Seas and the northern North Atlantic Ocean. The SOM technique, which has previously been adopted by Telszewski et al. (2009) in the North Atlantic Ocean, utilizes the relationship between measured fCO_2 and variables such as temperature, salinity, chlorophyll and mixed layer depth. For chlorophyll I will use data obtained from the SeaWiFS satellite. The fCO_2 data will be obtained from the SOCAT database (Pfeil et al., 2013), which is the most complete collection of surface seawater fCO_2 data from the global oceans. Mixed layer depth, temperature and salinity data will be extracted from the Mercator Ocean database.

Chapter 2

Method and Data

2.1 Artificial Neural Network

The artificial neural network (NN) is a flexible mathematical structure inspired by the biological nervous system. The nervous system's task is to transfer signals between various parts of the body and coordinate movements of the organism. This complex task is performed by a series of interconnected units (neurons), which have the ability to communicate and process signals. In the same manner, an artificial neural network transforms non-linearly correlated input signals, triggering numerous neurons simultaneously, which in turns leads to a final output (Telszewski, 2009). By imitating the human nervous system, an artificial neural network acts as an outstanding non-linear modeling tool to extract patterns from input variables.

For the artificial neural network to operate in the right manner, the network has to be configured. This configuration enables the set of inputs to produce the desired set of outputs, and is done by training the neural network. Training is the procedure where the network is fed teaching patterns. For each input provided there are matching output patterns, which are necessary to discover the optimal operating point. The neural network must undergo this configuration process in order to learn to recognize patterns (Dreyfus, 2005).

The NN is based on a "black-box" model with input-output training data and is able to predict an output pattern when it recognizes a given input pattern (Fig. 2.1). Since the neural network can be applied to approximate any underlying functional relationship, it is easy to understand why the NN

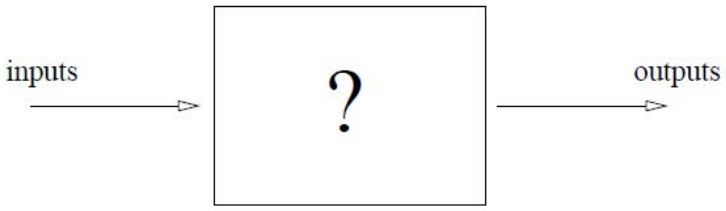


Figure 2.1: Black-box model. Figure from Hollmén (1996).

is favored over other techniques. In linear models for instance, the parameter independence is rarely observed. This is not an issue for the NN technique, as the network is trained to learn an approximation of the relationship by constantly adapting its parameters using observed data (Günther and Fritsch, 2010). The artificial neural network is a valuable statistical tool that has proved to be particularly useful in studies regarding cycles and trends. Over recent decades, this technique has become increasingly popular among the geosciences, in both oceanography (Liu, Wesiberg and He, 2006; Mihanović et al., 2011) and atmospheric science (Niang et al., 2006; Polo et al., 2011) and the evidence does not oppose the future application of the artificial neural network, rather the contrary.

2.1.1 Self Organizing Map

Self organizing map (SOM), often referred to as Kohonen neural network, is one of many types of artificial neural networks available, and appears to be the best suited technique to study empirical relationships in geoscience (Telszewski, 2009). The self organizing map belongs to the category of competitive learning networks and is based on unsupervised learning. The SOM technique can be seen as an advanced look-up table, where the fCO_2 can be determined by the corresponding parameters. This thesis will use the SOM approach in order to create basin-wide maps of sea surface fCO_2 in the Nordic Seas and the northern North Atlantic Ocean.

The SOM mapping consists of three main steps: (1) the training, (2) the labeling, and (3) the mapping. The training takes place without fCO_2 data, and is the process where a set of data, e.g. SST, MLD, SSS and CHL, produces a set of weight vectors that represents this data (Telszewski, 2009). During the training, points that are close to each other in the input are mapped to neighboring neurons, adjusting the distance between the points. All the neurons in the competitive learning network receive the same input. The internal competition between the neurons leads to a "winner" or a best matching unit (BMU, Fig. 2.2), based on the one with most activity (Hollmén, 1996). This can also be described as the weight vector (w_i) with the smallest Euclidean distance D(x_i, w_i) to the input vector (x_i). For a network with MLD, SST, CHL and SSS the Euclidean distance would be:

$$D(x_i, w_i) = \sqrt{(x_{iMLD} - w_{iMLD})^2 + (x_{iSST} - w_{iSST})^2 + (x_{iCHL} - w_{iCHL})^2 + (x_{iSSS} - w_{iSSS})^2} \tag{2.1}$$

where x_{iXXX} and w_{iXXX} are elements of the input vector and the weight vector for each parameter, respectively. The second step is the labeling process wheres fCO_2 data is assigned to each of the trained neurons. The third and final step is the completion of the geographical mapping using the preconditioned SOM as defined in the two previous steps (Telszewski, 2009). Here the SST, MLD, CHL and SSS are used to assign fCO_2 values to all space coordinates in the geographical map.

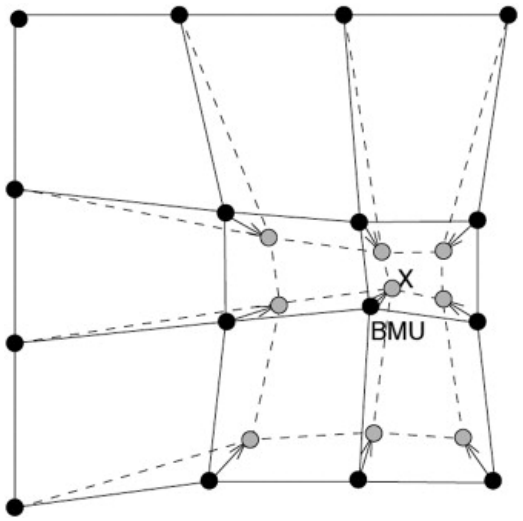


Figure 2.2: An input sample (X) updates the best matching unit (BMU) and its neighboring units. Black dots and solid lines correspond to the situation prior updating, whereas gray dots and dashed lines correspond to the situation after. Figure from Vesanto et al. (2000).

Various factors make the use of Self Organizing Maps favorable compared to other methods. One advantage is their ability to organize large, complex datasets, based on observations. The system has the ability to learn from an example, and does not need to address complex equations regarding physical, chemical and biological processes like other methods; no relationships between input parameters are necessary. Another advantage is the possibility to compare an estimated SOM with a realistic map with measured data. Since no complete fCO_2 map with measured data exists, this advantage

will be shown by comparing a temperature estimated SOM with a map of measured temperatures (section 3.5).

One disadvantage of using the SOM for mapping is related to the input data. In order to produce reliable maps, input values for each parameter is essential. This is sometimes impossible, as discussed in section 2.2.1 regarding chlorophyll data. This can result in limited data and a misrepresentation of the maps. Another disadvantage can be found in the training. Specifically, if the training process occurs over an extremely short or long period of time, it can result in undertraining or overtraining respectively. Undertraining means that the map has not been able to recognize the pattern before the training is ended. Overtraining involves the learning beyond a certain point, where the relevant information is lost due to too much extraction (Hollmén, 1996). Both cases leads to a distortion of the real situations. Further drawbacks are that nearby points must have similar behavior and that mapping can result in clustering.

2.1.2 Other Types of Neural Networks

Several types of NN are used in geoscience, and in addition to SOM, the two most commonly used networks are the Probabilistic Regularized Self Organizing Map (PRSOM) and the Multilayer Perceptron Neural Network (MLP). Due to the previous success of using SOM as an approach to map the sea surface fCO_2 (Lefévre et al., 2005; Telszewski et al., 2009), these two other networks are only briefly discussed.

The probabilistic model, PRSOM, is an extension of SOM for visualizing and clustering high-dimensional datasets. The first step, the training phase, is identical to the SOM technique. An unsupervised classification is built from a dataset by clustering similar input vectors into a certain number of neurons. The second and last step concerns the labeling of the vectors and groups being clustered into classes corresponding to physical characteristics provided by an expert (Awa Niang et al., 2003).

The MLP method is classified as a "feedforward neural network", and consists of an input layer, a hidden layer and an output layer, shown in Fig. 2.3. As the name implies, the diagram must be feed-forward (Bishop, 1995), which results in a layer of neurons being fully connected to the next layer. Normally the MLPs are trained with the back propagation learning algorithm. This involves letting some of the information flow in the backward direction. Based on the amount of errors calculated during the training, the MLP adjusts the weight of neurons to minimize the error. This is repeated until the desired result is reached (Telszewski, 2009).

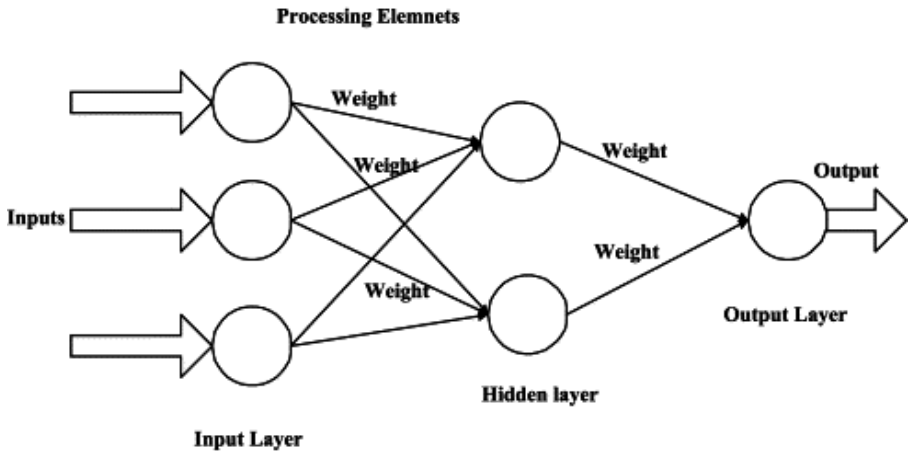


Figure 2.3: Schematic illustration of the Multilayer Perceptron neural network. Circles indicate individual neurons. Figure from Heiat (2002).

2.2 Dataset

Based on previous studies and the successful adoption of SOM in the North Atlantic Ocean by Telszewski et al. (2009), the SOM technique will be used to map fCO_2 in the Nordic Seas and the northern North Atlantic Ocean. The first fCO_2 map will be estimated with the following two parameters:

$$fCO_2 = f_{SOM}(\Theta_{surf}, z_{mix}) \tag{2.2}$$

where Θ_{surf} is sea surface temperature (SST) and z_{mix} is mixed layer depth (MLD). These are selected as default parameters and will be included in all SOMs. Several extensions of Eq. 2.2 will be tested to see whether the use of additional input parameters will optimize the fCO_2 maps. The chlorophyll-a concentration (CHL), defined as c_{chl}, is introduced as the third parameter:

$$fCO_2 = f_{SOM}(\Theta_{surf}, z_{mix}, c_{chl}) \tag{2.3}$$

A fourth parameter, sea surface salinity (SSS), represented as S_{surf} in Eq. 2.4, has proved to work well for estimating sea surface fCO_2 in the North Pacific Ocean (Nakaoka et al., 2013) and in the Barents Sea (Lauvset et al., 2013), and will therefore also be tested:

$$fCO_2 = f_{SOM}(\Theta_{surf}, z_{mix}, S_{surf}) \tag{2.4}$$

Additionally, all four parameters will be tested together:

$$fCO_2 = f_{SOM}(\Theta_{surf}, z_{mix}, c_{chl}, S_{surf}) \tag{2.5}$$

The various combinations of the four parameters are presented in Tab. 2.1.

The chosen parameters affect fCO_2 either directly or indirectly. SST is regulated by physical processes such as solar radiation and mixed layer depth, and has a direct influence on fCO_2. During winter, when the biological productivity is low, fCO_2 is mainly regulated by the thermodynamic effect. Generally, the fCO_2 increases with increasing SST. However, during summer, the warming water contributes to an increased biological production, that causes fCO_2 to decrease with increasing SST. SST has a greater influence on fCO_2 than SSS (Sarmiento and Gruber, 2006). SSS influences fCO_2 indirectly through alkalinity and mixing, in addition to having a minor direct effect on regulation of CO_2 solubility. Mixed layer depth influences fCO_2 indirectly (Olsen et al., 2008). During fall, the sea surface waters cool, and this in combination with intense wind mixing causes deep mixing to occur. The deep mixing transports waters rich in remineralized CO_2 up from the depths and causes the sea surface fCO_2 to increase. Another indirect influence MLD has on fCO_2 is through its influence on biological production. As shallow mixed layers form in response to heating in spring, biological production increases, hence low MLDs tends to be associated with low fCO_2 values (Olsen et al., 2008). Biological production has a direct influence on fCO_2, and is measured by CHL. During summer CHL levels are high, and are typically associated with low fCO_2 values.

Table 2.1: Various input parameters used for the SOMs

Map name	SST	MLD	CHL	SSS
SOM A	√	√		
SOM B	√	√	√	
SOM C	√	√		√
SOM D	√	√	√	√

Self organizing maps will be estimated for three regions: (i) the northern North Atlantic north of 44°N including the Nordic Seas, (ii) the northern North Atlantic (44-63°N) excluding the Nordic Seas, and (iii) the Nordic Seas only (63-85°N). This is in order to see whether mapping two independent basins works better than one large. Additionally, the SOMs for (ii) and (iii) will be merged to see whether this produces more realistic results than the SOM for (i). The dataset for each parameter will be presented in the next two subsections. Tab. 2.2 presents a list over the training and labeling datasets.

Table 2.2: Dataset used and their corresponding source

Parameter	SOCAT	Mercator	SeaWiFS
The training dataset:			
SST	*	√	
MLD		√	
CHL			√
SSS	*	√	
The labeling dataset:			
fCO_2	√		

Note: Dataset for validation (∗) and for estimating SOM (√)

2.2.1 The Training Dataset

Subsets consisting of the parameters SST, MLD, CHL and SSS represent the training dataset. Both the SST and the SSS data exist as two independent datasets. The first temperature and salinity dataset comes from GLobal Ocean ReanalYsis and Simulation (GLORYS2V1) of Mercator Ocean, together with an MLD dataset. The second dataset is the *in-situ* values of SST and SSS observed along with fCO_2 data and is obtained from the Surface Ocean CO_2 Atlas (SOCAT v2.0, www.socat.info). The first dataset of SST and SSS will be a part of the SOM parameters, whereas the second dataset only will be used for validation. As this subsection only addresses the training datasets, fCO_2 will be discussed in the section concerning the labeling dataset.

The training datasets from Mercator Ocean are derived from an assimilation model designed to describe the ocean on a global and regional scale. Both satellite and in-situ measurements are used in order to make the models as reliable as possible. I will use data from the GLORYS reanalysis that includes the period from 2005 to 2007 with a resolution of 1/4°, covering the areas from 0-89°N and 107°W-72°E. This GLORYS reanalysis constitutes a complete global dataset of temperature (°C), mixed layer depth (m) and salinity (psu) henceforth just called Mercator. The Mercator project is a contribution to MyOcean reanalyses, a European network project aiming to describe the ocean in 3 dimensions and in real-time (www.myocean.eu).

For the CHL data (mg m^{-3}) the Level 3 standard mapped image (8-daily 1/9°) derived from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) is used (oceancolor.gsfc.nasa.gov). CHL is here represented by chlorophyll-a, a specific form of chlorophyll used by phytoplankton in photosynthesis. The concentration of CHL represents the amount of photosynthetic organisms in the surface layer. A disadvantage with SeaWiFS is that the data collection is limited by poor light conditions due to shorter days in the winter, and cloudiness. Consequently the chlorophyll data can only be obtained for summer (April-September).

The range of years included in this study was restricted by the limitation of the SOCAT and SeaWiFS databases. The SOCAT database has few measurements from prior to year 2005, including these years would therefore be detrimental to the analysis. A drawback with SeaWiFS is that no data has been collected since December 2010, and a distinct reduction in the number of annual data collected occurred after 2008. Due to this drawback and the fact that the majority of the data from this period is from the Pacific Ocean, the 2008-2010 data was eliminated. This study will therefore focus on the period from 2005 to 2007.

All datasets were re-gridded onto an 8-daily time scale to fit the CHL data, as it is easier to convert the SST, MLD and SSS datasets to an 8-daily dataset, than vice versa. The datasets were then re-gridded to a 1° latitude x 1° longitude resolution, in order to make them more manageable. Coastal waters (bottom depth < 350m) have been removed from the datasets, as they are influenced by coastal activities and nutrient pollution from coastal rivers. This was done by using the global model of ocean bathymetry ETOPO2v2g from National Oceanic and Atmospheric Administration (NOAA),

available at www.ngdc.noaa.gov/mgg/global/relief/ETOPO2/. Additionally, SST values below -1.85°C were assumed to indicate sea-ice cover and removed. This temperature was found by trial and error and proved to produce the most realistic ice cover around Greenland compared to satellite maps. The calendar year was divided into 4 quarters; winter (December, January, February), spring (March, April, May), summer (June, July, August) and fall (September, October, November) in order to distinguish seasonal variations.

2.2.2 The Labeling Dataset

After the SOM neurons have been trained, it is necessary to label them with fCO_2 values (µatm) in order to use them to estimate the fCO_2 maps. This is done by the labeling dataset which consists of fCO_2 measurements with corresponding Mercator and SeaWiFS SST, MLD, CHL and SSS values. To match the re-gridded training dataset, the fCO_2 data from between 2005 and 2007 were bin-averaged in an 8-daily frequency, with a 1°latitude x 1°longitude resolution. This thesis uses the dataset SOCATv2.0_NorthAtlantic (Bakker et al., 2012), which includes data from 1968-2011. This dataset has a particularly large contribution of carbon data from the cargo ship R/V Nuka Arctica, which runs monthly between Denmark and West Greenland (Olsen et al., 2008).

Fig. 2.4 depicts the number of fCO_2 data in the labeling dataset for the individual years. The majority of the data for the first six months was collected in 2007, whereas most of the data for the remaining six months was collected in 2006. The distribution of the fCO_2 measurements is illustrated in Fig. 2.5. Year 2005 has few measurements in both the northern North Atlantic Ocean and the Nordic Seas, whilst the number of measurements seems to improve spatially the following two years.

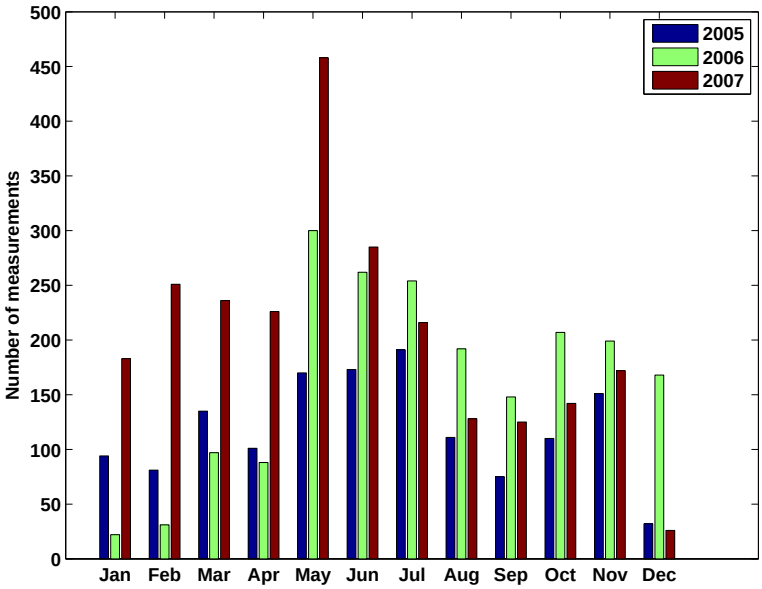
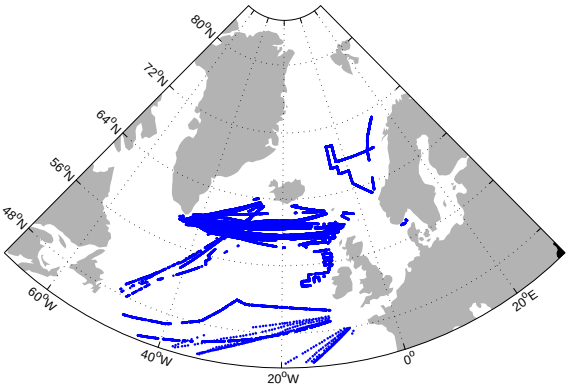
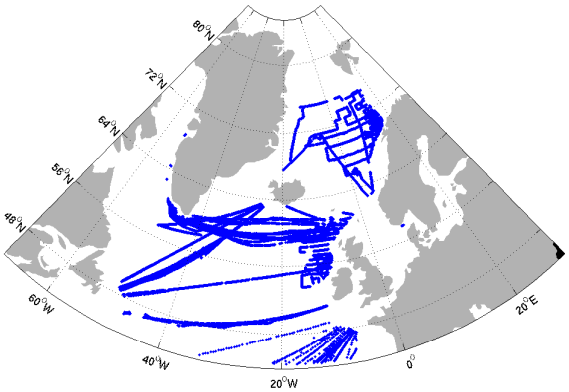


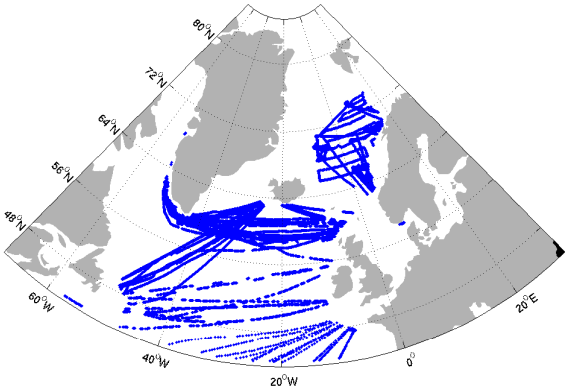
Figure 2.4: Number of fCO_2 measurements in the re-gridded labeling dataset (1°latitude x 1°longitude, 8-daily).



(a)



(b)



(c)

Figure 2.5: Distribution of fCO_2 measurements from SOCAT database in year 2005 (a), 2006 (b) and 2007 (c).

2.3 Validation of Input Parameters

In order to determine the SOM's reliability, a validation of each input parameter is performed. Fig. 2.6 shows the monthly input parameters for the years 2005-2007, averaged over the entire study area. Unlike the values for fCO_2 (Fig. 2.6a) and MLD (Fig. 2.6b), the graphs for SST, SSS and CHL (Fig. 2.6c-e) show very minor interannual differences. The MLD data have some small deviations in January and March. Most of the monthly values for fCO_2 show interannual variations, but nonetheless the graph shows a distinctive seasonal cycle, which implies that the data are realistic. The relatively high fCO_2 values in June and July 2005 are due to the lower abundance of phytoplanktonic species, represented by CHL.

Additionally, the input re-gridded parameters were checked against the original data (Fig. 2.7). This was to ensure that the re-gridded data was able to represent the original data in a realistic manner. The re-gridded SST dataset (Fig. 2.7a) is missing some values around Greenland due to the definition of sea ice cover invoked in the thesis. As the re-gridded CHL dataset has lower resolution than the original, the amount of data present appears to be larger in the re-gridded dataset, which is not the case.

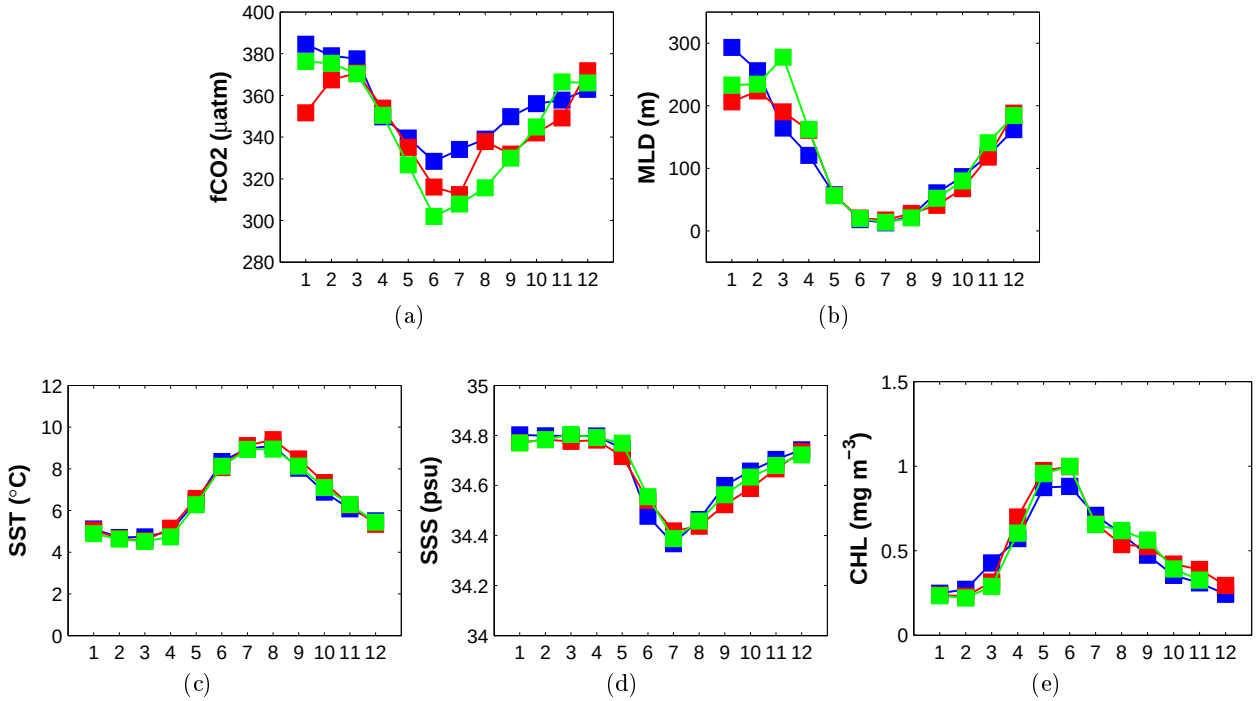


Figure 2.6: Mean monthly values of $fCO2$ (a), MLD (b), SST (c), SSS (d) and CHL (e) input data over the entire study area for year 2005 (blue), 2006 (red) and 2007 (green). Numbers on x-axis (1-12) represent months.

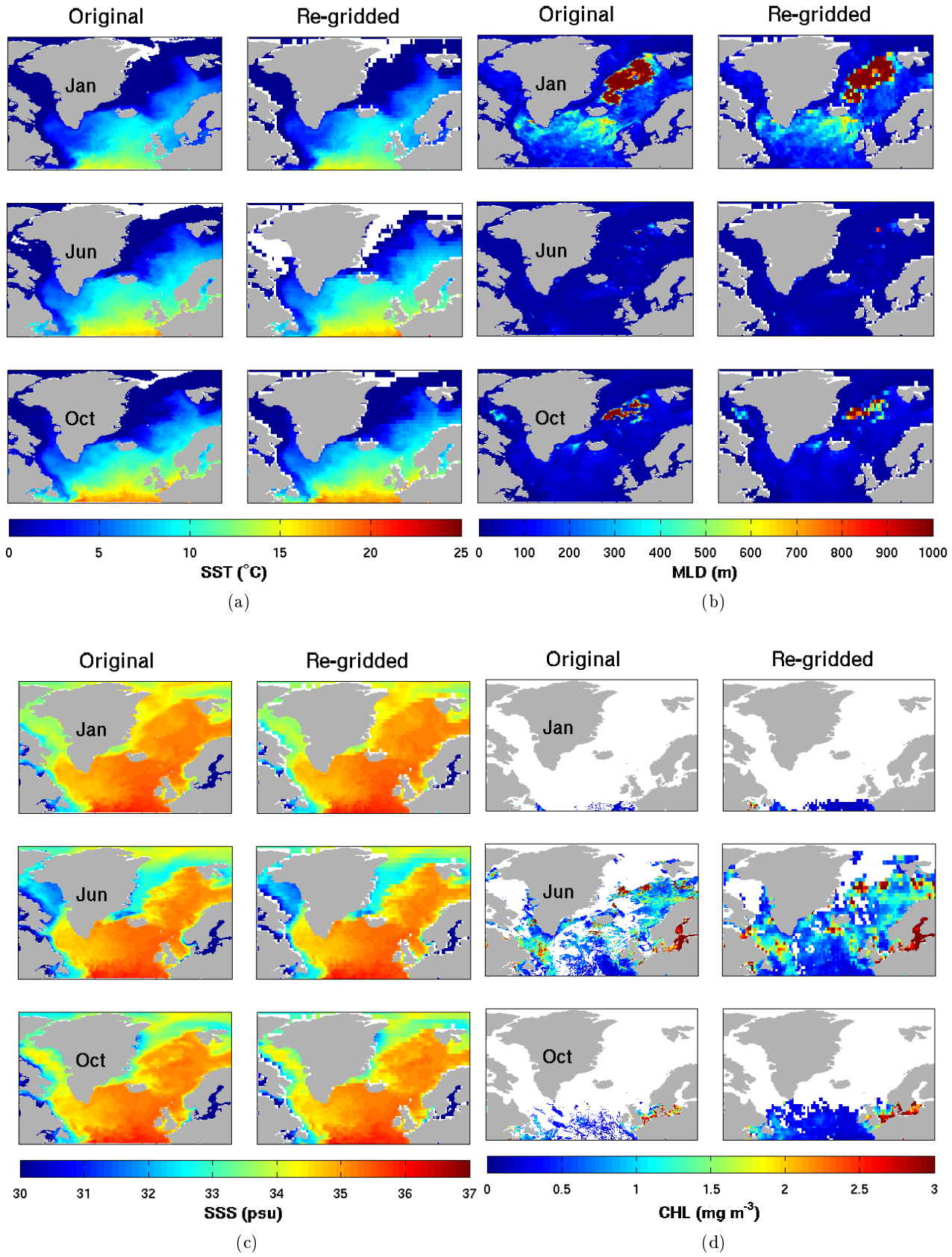
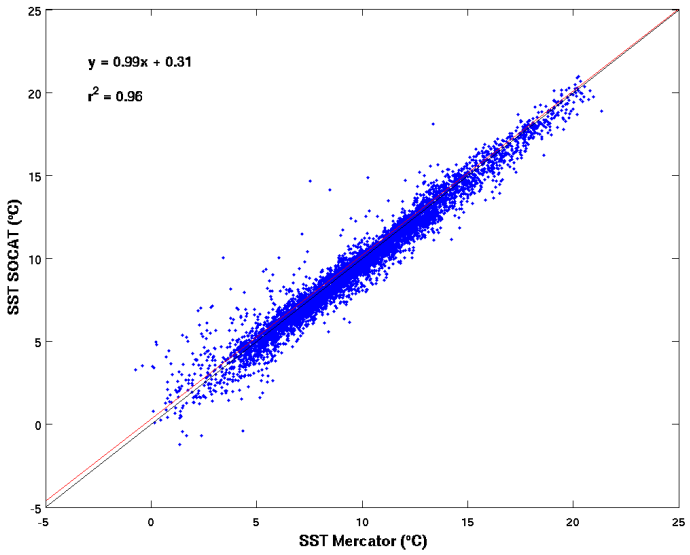
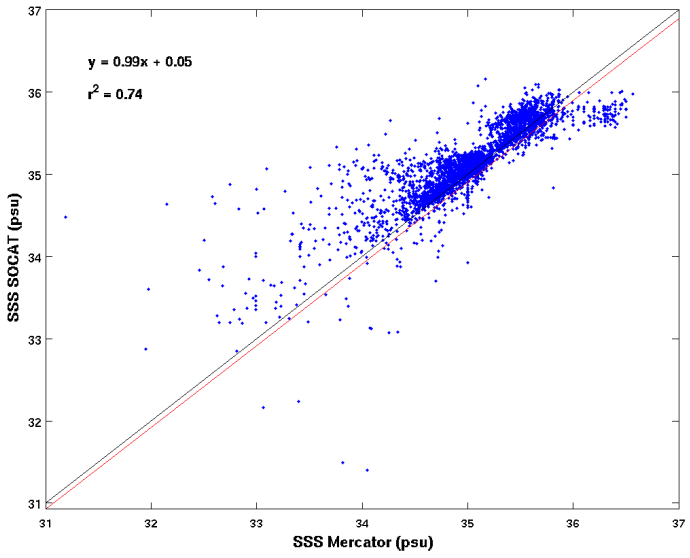


Figure 2.7: Original SST (a), MLD (b), SSS (c) and CHL (d) data plotted against matching re-gridded data for January, June and October.

The SST and SSS data from Mercator consist of simulated values and are plotted for comparison with measured data from SOCAT in Fig. 2.8. Only the data from Mercator will be used to estimate fCO_2 maps, whilst the SOCAT SST and SSS data is used for validation. As can be seen in Fig. 2.8a the correlation between temperature from Mercator and SOCAT is nearly perfect, with an r^2 of 0.96. The correlation for salinity between the two datasets are not quite as satisfactory, but lies within an acceptable range, with an r^2 of 0.74.



(a)



(b)

Figure 2.8: Comparison of Mercator input data with measured data from SOCAT: (a) temperature, (b) salinity. The function represents the regression line (red), while the identity line (black) represents x=y.

2.4 Seasonal Cycles

This section deals with the distribution of each individual parameter as illustrated by monthly maps (Fig. 2.9-2.12). The illustrations are produced from data for year 2005, but are representative for 2006 and 2007 as well.

2.4.1 Sea Surface Temperature

Seasonal variations in solar radiation dominates the distribution of SST. Since the temperature is affected by insolation, the distribution have a tendency to have a zonal pattern. Polar regions receive lower amounts of heat compared to equatorial regions. Warm surface waters are seen in the eastern basin, in sharp contrast to the cooler waters in the west. This pattern is due to a combination of isolation and ocean circulation. The Gulf Stream is transported from equatorial latitudes and brings warm surface waters along the western coast, whereas the East Greenland current transports cold water masses from the Arctic Ocean in to the Atlantic Ocean. The lowest temperatures are found in March, while the highest are found in August. This is due to variations in solar radiation. During spring the oceans gradually warm, reaching maximum temperatures at the end of summer. The decrease in temperature during fall and winter is caused by a decrease in solar radiation as well as the occurrence of deep mixing.

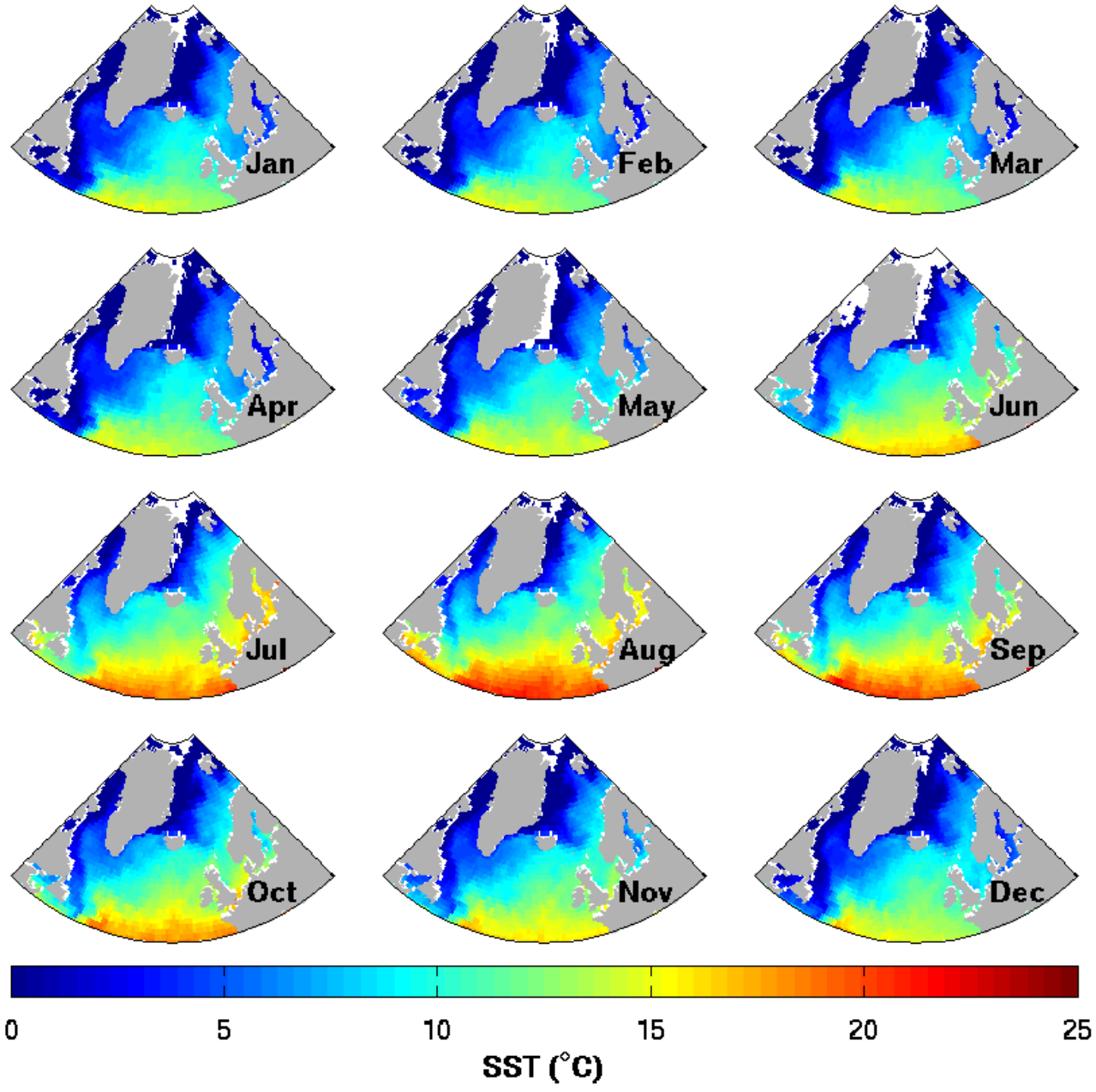


Figure 2.9: Seasonal cycle of SST(°C) represented by the re-gridded dataset for 2005.

2.4.2 Mixed Layer Depth

Most of the ocean stratification is determined by temperature, and this also applies to the mixed layer depth. During winter the MLD may reach hundreds to a thousand meters due to the cold destabilized water masses. When spring arrives, the upper ocean water masses are gradually heated due to increased solar radiation, leading to a distinctive temperature gradient in the upper water column. This prevents the water masses in the interior to mix with the sea surface, resulting in a shallow MLD.

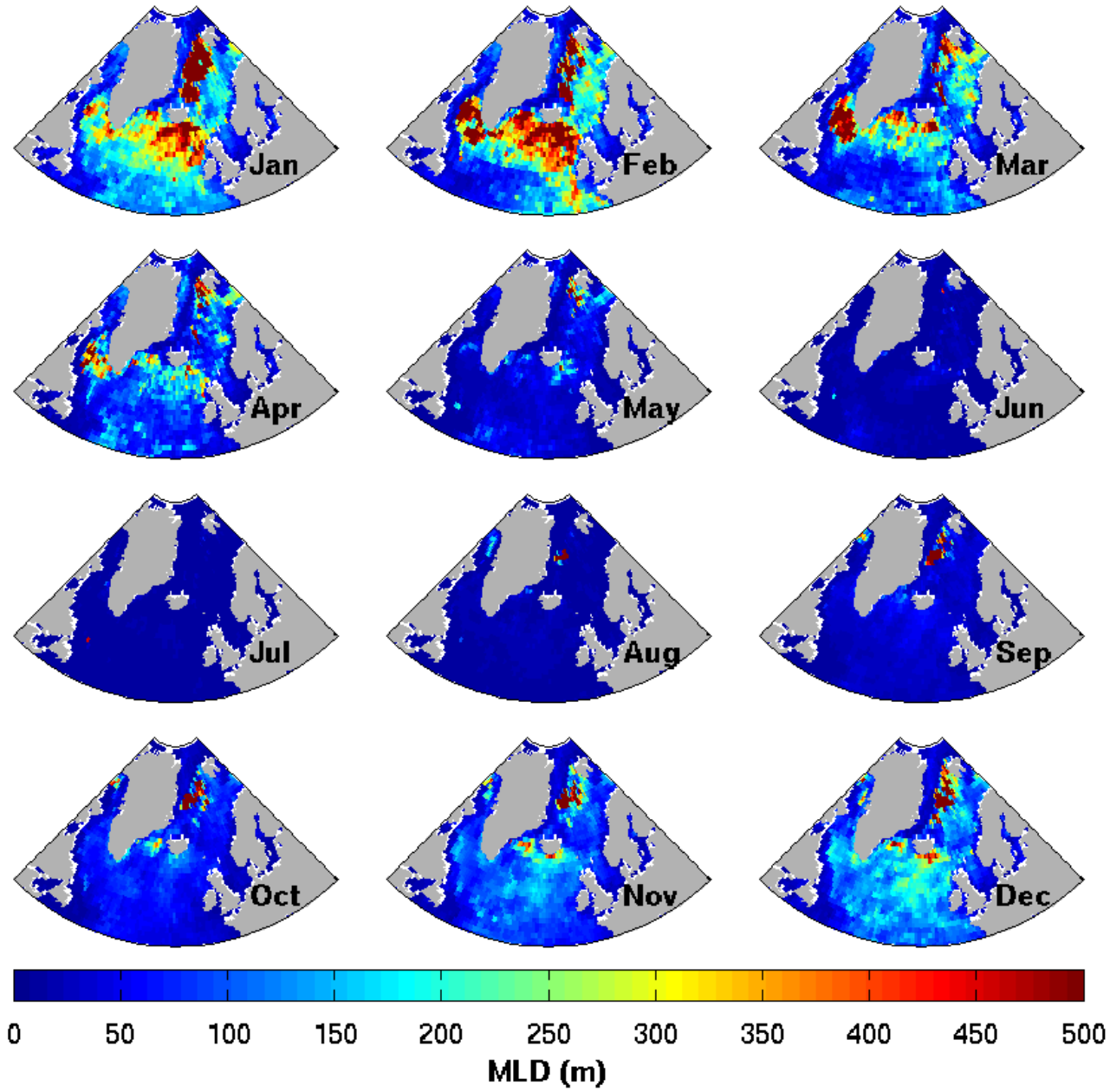


Figure 2.10: Seasonal cycle of MLD (m) represented by the re-gridded dataset for 2005.

2.4.3 Sea Surface Salinity

A tendency towards zonal arrangement can be seen in the salinity distribution. Seasonal variations occur to a lesser extent and are mainly driven by ocean circulation and the addition and removal of freshwater, which occurs through river input and formation or melting of sea ice. The saltier water masses along the eastern side of the basin are a result of the salt ocean currents transported from the equator, whilst the fresher areas on the western side are due to the fresher water masses from the Arctic Ocean. A noticeable change can be seen in the summer months around Greenland, where freshwater is released due to ice melting. These areas become saltier during winters, when ice formation takes place.

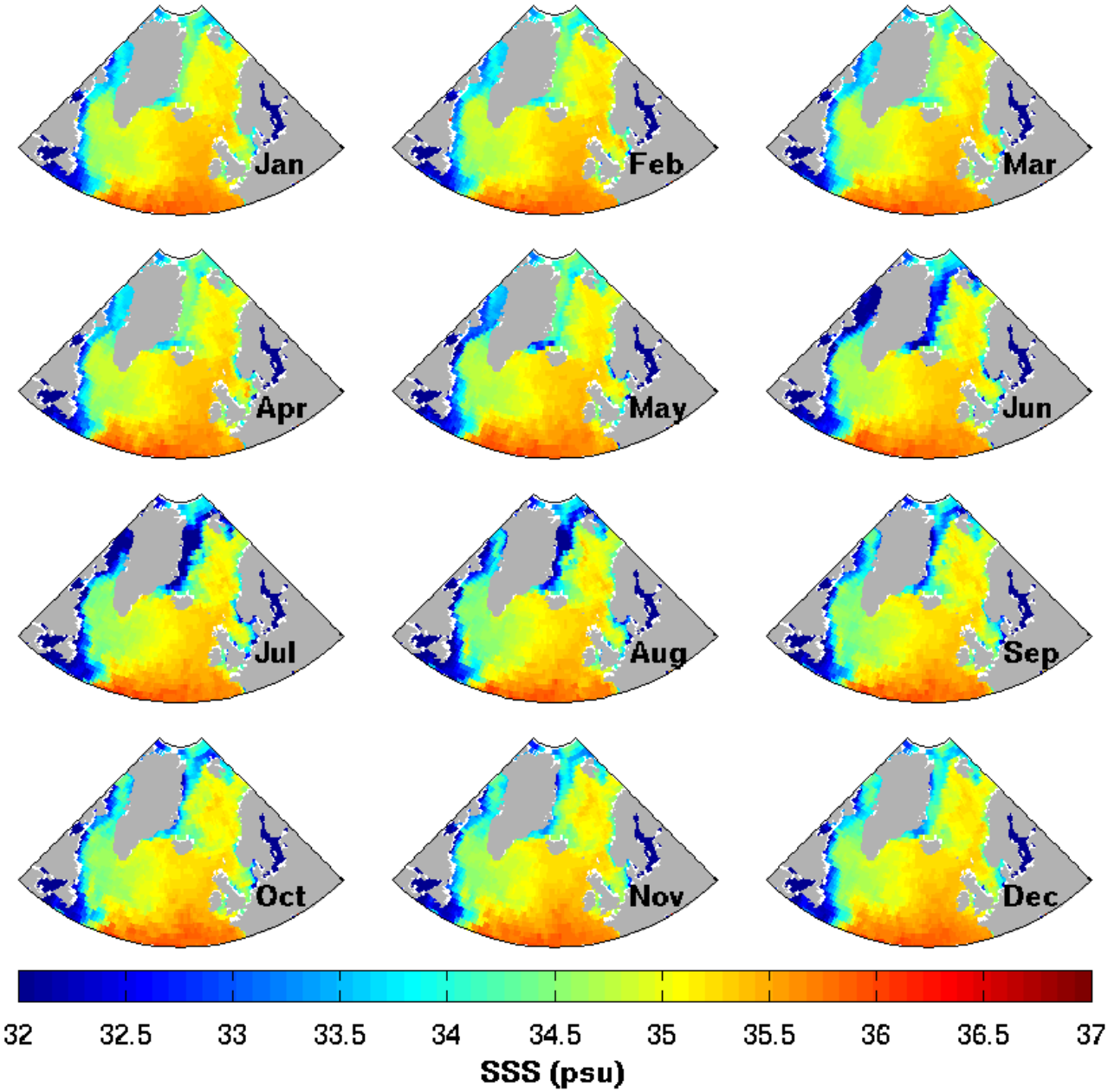


Figure 2.11: Seasonal cycle of SSS (psu) represented by the re-gridded dataset for 2005.

2.4.4 Chlorophyll

Due to shorter days and increased cloudiness during winter months, the CHL cycle lacks data at higher latitudes from October to March. Since phytoplankton use chlorophyll to absorb energy from light to carry out photosynthesis, CHL data are influenced by seasonal variations. When both nutrients and sun light are available at the surface and the phytoplankton is trapped in a shallow mixed layer, the spring bloom begins, which leads to a strong increase in phytoplankton abundance. The highest chlorophyll concentrations are often found in cool waters. This is not due to the temperature itself, but a result of the cold, nutrient-rich waters brought to surface during deep winter mixing. In April the spring bloom begins in the North Sea due to stratification by coastal currents. In the Nordic Seas stratification is induced by the increase in surface temperature, here the spring bloom peaks about a month later, in May and June.

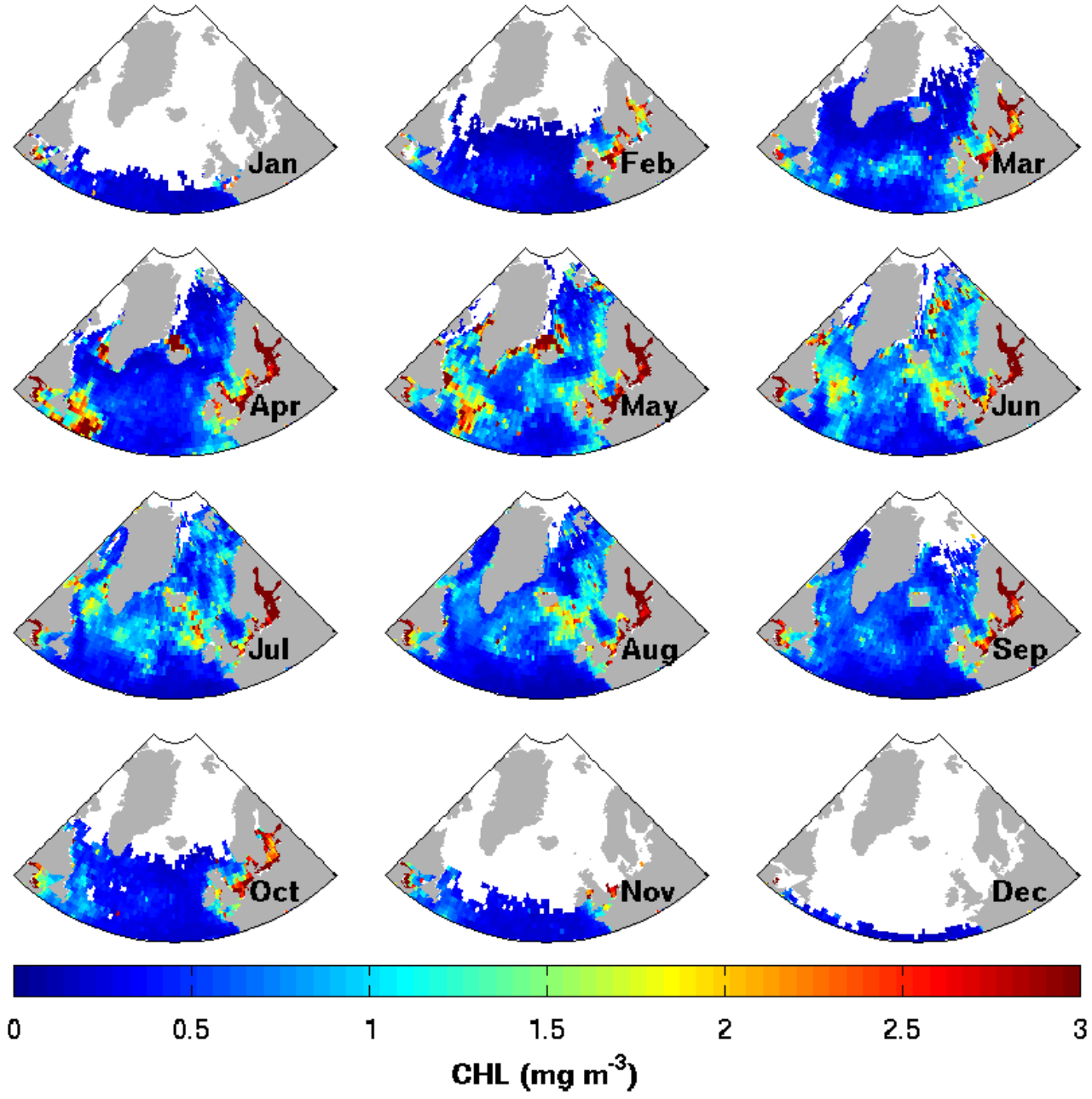


Figure 2.12: Seasonal cycle of CHL (mg m^{-3}) represented by the re-gridded dataset for 2005.

2.5 SOM Parameters and Settings

This section addresses the SOM parameters and the settings used to estimate my fCO_2 maps. The SOM Toolbox version 2.0 for Matlab 5 (Vesanto et al., 2000) adopted in order to produce the maps is freely downloadable at www.cis.hut.fi/projects/somtoolbox. Matlab was used to create the SOMs and as a data visualization tool.

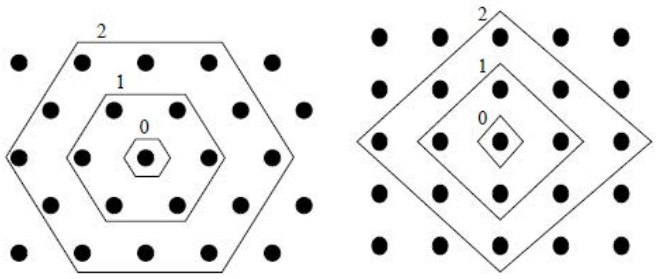
In the process of producing the most optimal map, both the selection of the input parameters and the SOM settings influences the outcome. The initialization and training part is the first step and also the most crucial. The function "som_make" carries out this step and requires a set of arguments, with the most commonly used settings for each given in Tab. 2.3.

Table 2.3: Changeable arguments in the 'som_make' function

Arguments	Options			
msize	<map grid size>			
lattice	'hexa'	'rect'		
shape	'sheet'	'cyl'	'toroid'	
neigh	'bubble'	'gaussian'	'cutgauss'	'ep'
mask	<input columns>			
training	<rough/fine>			

The map size 'msize' is not the resolution of the fCO_2 map, but an argument that determines the number of neurons used in the training process. Too many neurons causes the maps to be overrepresented, which makes it difficult for the SOM to recognize characteristics patterns. Too few neurons leads to an underrepresentation of the data that produces too smooth patterns. The map size for the different SOMs was determined by trial and error.

While choosing the most desirable map size, the mean quantization (qe) and topographic (te) errors are used as indicators to determine the SOM's quality. The qe describes the average distance between each data vector and its BMU, while the te measures whether the first and second BMU are neighboring neurons. The lower the qe and te values, the better the representation of the training data. Generally, the qe and te decreases with increasing neurons.



(a) (b)

Figure 2.13: Hexagonal (a) and rectangular (b) SOM lattices. The values 0, 1 and 2 correspond to neighborhood neurons. Figure from Vesanto et al. (2000).

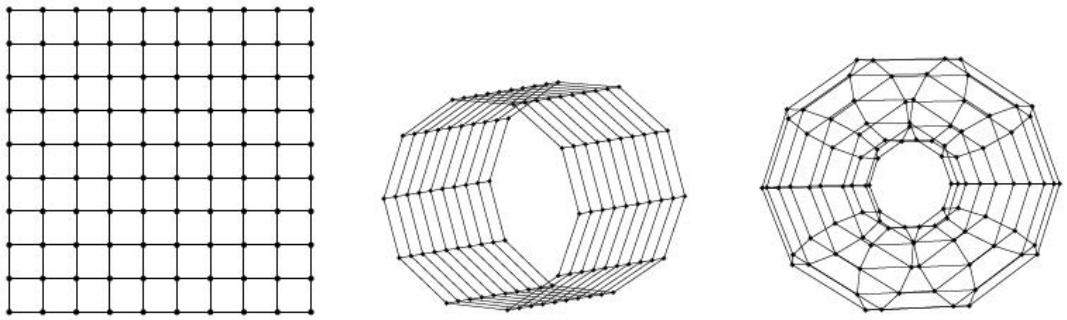
In addition to map size determination, the lattice and shape must be chosen. The neurons are placed in either a hexagonal and rectangular lattice (Fig. 2.13). The difference between the two lies in the map topology and the distance between each neuron. Different varieties of lattice shapes can be chosen, such as sheet, cylinder or toroid (Fig. 2.14). Based on the findings of Liu, Weisberg and Mooers (2006), only the hexagonal lattice and the flat sheet shape were used to estimate the maps here.

Four neighborhood functions are selectable in the SOM Toolbox: 'bubble', 'gaussian', 'cutgauss' and 'ep' (epanechicov). Based on previous experiments with the neighborhood function (Telszewski, 2009), and the good reproduction of the original data, the 'gaussian' function was chosen for the argument 'neigh'.

The extent of training of the SOM is specified by the argument 'training' and consists of two phases: rough training and fine-tuning.The rough training takes place first, followed by final adjustments in the fine-tuning phase. The amount of training is crucial for the outcome, but can unfortunately only be determined by trial and error. If the map is excessively trained it can lead to overtraining, whilst inadequate training leads to undertraining. Both cases may result in improper visualization of the data, which is why the maps should be trained with moderation. This is achieved by avoiding too low or too high values for the arguments for rough and fine-tuning.

Both the training and the labeling dataset were linearly normalized in order to minimize redundancy. This is particularly important since the SOM algorithm uses Euclidean distance and because the large range of the input values (e.g. 10-3000 m for MLD and -1.8-25°C for SST), greatly affects the calculation.

The fCO_2 maps are presented in the next chapter, where the following questions are investigated: Are the self organizing maps able to reproduce the fCO_2 distribution in the Nordic Seas and the northern North Atlantic Ocean? If so, which parameters are required in order to estimate the most optimal fCO_2 maps?



(a) (b) (c)

Figure 2.14: Different map shapes. The default sheet shape (a), and two shapes where the map topology accommodates circular data: cylinder (b) and toroid (c). Figure from Vesanto et al. (2000).

Chapter 3

Results

In order to determine whether the SOM method is useful in estimating fCO_2 maps in the Nordic Seas and the northern North Atlantic Ocean, skill assessment statistics that quantify model performance are used to determine the maps' reliability. The statistics adopted are based on metrics used by Stow et al. (2009) and are the square of the correlation coefficient (r^2, Eq. 3.1), the root-mean-square-error (RMSE, Eq. 3.2) and the Nash Sutcliffe model efficiency (ME, Eq. 3.3), in addition to the Cost Function (CF, Eq. 3.4) used by Holt et al. (2005):

$$r^2 = \frac{\sum_{i=1}^{n}(O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2 \sum_{i=1}^{n}(P_i - \overline{P})^2}} \tag{3.1}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \tag{3.2}$$

$$\mathrm{ME} = \frac{\left(\sum_{i=1}^{n}(O_i - \overline{O})^2 - \sum_{i=1}^{n}(P_i - O_i)^2\right)}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \tag{3.3}$$

$$\mathrm{CF} = \frac{1}{n\sigma^2}\sum_{i=1}^{n}(P_i - O_i)^2 \tag{3.4}$$

where n is the number of observations, O_i is the ith number of n observations, P_i is the ith number of n predictions, $\overline{O}$ and $\overline{P}$ are averaged observations and predictions respectively and σ^2 is the standard deviation of the observations.

The r^2 value ranges from 0 to 1 and measures how well the predicted data fits the observed data. A correlation of 1 equals a perfect match. The RMSE is a measure of the discrepancy between observed and predicted values. The lower the value, the better the match (Stow et al., 2009). The ME measures how well a model simulates the observed values (Nash and Sutcliffe, 1970; Stow et al., 2009). Based on Allen et al. (2007), the efficiency is rated as: ME>0.65 excellent, 0.65-0.5 very good, 0.5-0.2 good, and ME<0.2 poor. The CF is weighted with σ^2, which means that it measures the accuracy of the predictions, in addition to how accurate they are compared to the variation in the data. CF is therefore more informative than RMSE (Holt et al., 2005).

Maps were determined for year 2005, 2006 and 2007. However, only the maps for year 2005 is presented in this chapter. The maps for 2006 and 2007 are presented in Appendix B. Year 2005 was selected in order to compare my maps with the SOM fields prepared by Telszewski et al. (2009) for the North Atlantic Ocean.

As mentioned in section 2.2, bottom depths shallower than 350 m were excluded (Fig. 3.1). During visualization of the fCO_2 maps in Matlab, the function "m_contourf" had difficulties with visualization of the defined bottom depths. The excluded depths therefore appear in some of the estimated fCO_2 maps as dark blue areas, especially around Iceland. This is incorrect and the reader should be aware of this.

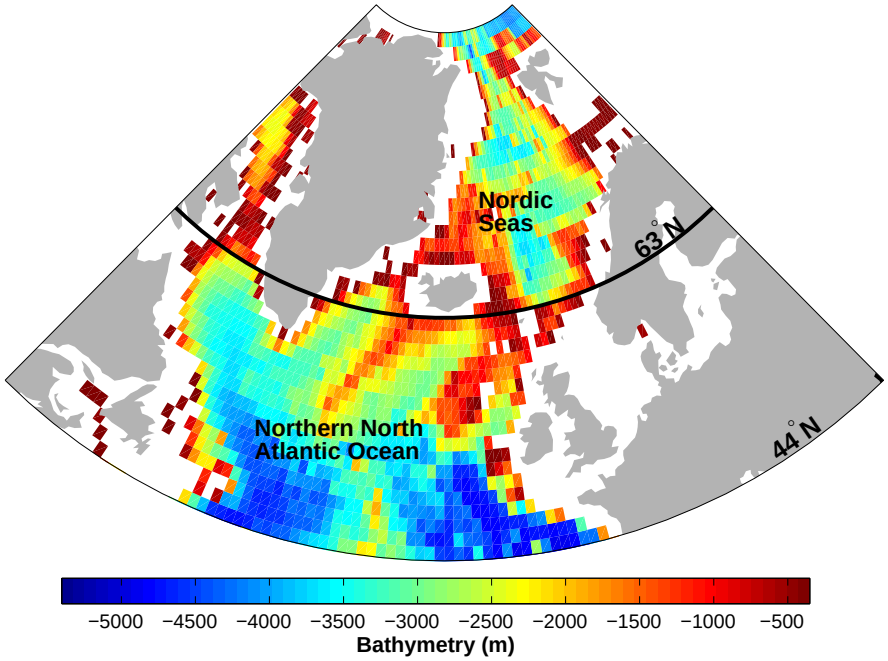


Figure 3.1: Map showing the separation of the two seas at 63°N. The northern North Atlantic Ocean is defined as 44-63°N, while the Nordic Seas is defined as 44-85°N. Colors indicate bottom depths greater than 350m, whereas white spaces indicate bottom depths shallower than 350 m.

As SOM B and D, which contains the CHL parameter, have difficulties with mapping fCO_2 at higher latitudes during winter, these maps will be created with two different sets of parameters, one for winter and one for summer. CHL is only part of the summer parameters, as no data is available for the winter months. This should work well, since the CHL concentration and the biological productivity are low during winter. The winter months are defined as September to March, and the summer months as April to August. The monthly breakdown was based on biological production (Fig. 2.12). Months with no distinctive biological activity were categorized as winter months, while months with distinctive biological production were categorized as summer months. Since SOM A and C do not contain the CHL parameter, their winter and summer parameters are identical. The different sets of parameters for all SOMs are listed in Tab. 3.1.

Table 3.1: Different sets of parameters used for SOM A-D

Map name	Winter parameters	Summer parameters
SOM A	MLD/SST	MLD/SST
SOM B	MLD/SST	MLD/SST/CHL
SOM C	MLD/SST/SSS	MLD/SST/SSS
SOM D	MLD/SST/SSS	MLD/SST/SSS/CHL

The optimal combinations of map size and training arguments were selected by performing skill assessment statistics. The arguments 'msize' and 'training' for each combination of predictive variables were optimized iteratively by running each SOM with different settings. The SOMs with the best results are shown in Tab. 3.2. The optimalization had good effect on the parameters, e.g. r^2 typically improved by 10%. The SOM parameters for each map are also specified under each figure. However, this is not done for the merged SOMs, but they can be found in Tab. 3.2 by combining the SOM parameters for the northern North Atlantic Ocean and the Nordic Seas for each SOM. Most of the SOMs were trained with an 'msize' of [100 65] and a 'training' of [25 20]. The various combinations are discussed in section 4. Quality parameters for the SOMs were calculated and are presented in Tab. A.1 in Appendix A. These parameters can function as an indicator on how well a SOM will perform.

Table 3.2: Parameters used for the optimized SOMs

Map name & Parameters	The northern North Atlantic Ocean and the Nordic Seas		The northern North Atlantic Ocean		The Nordic Seas	
	Winter	**Summer**	**Winter**	**Summer**	**Winter**	**Summer**
SOM A: 'msize'	[90 50]		[85 50]		[100 65]	
SOM A: 'training'	[20 15]		[15 10]		[25 20]	
SOM B: 'msize'	[90 50]	[90 50]	[85 50]	[100 65]	[100 65]	[100 65]
SOM B: 'training'	[20 15]	[25 20]	[15 10]	[25 20]	[25 20]	[25 20]
SOM C: 'msize'	[100 65]		[100 65]		[100 65]	
SOM C: 'training'	[20 15]		[25 20]		[25 20]	
SOM D: 'msize'	[100 65]	[60 40]	[100 65]	[100 65]	[100 65]	[100 65]
SOM D: 'training'	[20 15]	[15 10]	[25 20]	[15 10]	[25 20]	[15 10]

3.1 The northern North Atlantic Ocean and the Nordic Seas [44-85°N]

3.1.1 Underlying Functional Relationships

To see how suitable the different input parameters are as predictive variables, the MLD, SST, SSS and CHL were plotted against fCO_2. Thus it is possible to investigate the coupling between them (Fig. 3.2).

Fig. 3.2a-d shows all available data, whereas Fig. 3.2e-h and Fig. 3.2i-l show only data obtained for winter and summer months, respectively. All of these figures were prepared using data from areas with bottom depths >350 m.

The relationship between fCO_2 and MLD for 2005-2007 (Fig. 3.2a) has a nearly logarithmic shape and deep mixed layer depths are associated with high fCO_2 values, whereas waters with shallow MLD has fCO_2 values within the whole range. There is a clear transition in MLD from winter (Fig. 3.2e) to summer (Fig. 3.2i). During winter the fCO_2 values are high regardless of MLD. In summer on the other hand, the MLD is shallow with fCO_2 values ranging from low to high.

An almost logarithmic shape can also be seen for the CHL data (Fig. 3.2d). Typically the fCO_2 values decrease with increasing amount of CHL. This is expected as CHL measures the presence of phytoplankton, which consumes CO_2 in photosynthesis. The amounts of CHL during winter (Fig. 3.2h) and summer (Fig. 3.2l), reveal that the CHL data in Fig. 3.2d are dominated by data from summer.

For the SST (Fig. 3.2b) and SSS (Fig. 3.2c) there are no clear patterns. There is a warming trend with decreasing fCO_2 values from 5-10°C (green trend line in Fig. 3.2b), which is in accordance with the higher biological fCO_2 consumption during summer. Another slight trend can be seen between 10-20°C (red trend line in Fig. 3.2b), where fCO_2 values increases with increasing temperature, i.e. the thermodynamic effect. This is typical for oligotrophic areas, such as the subtropical eastern North

Atlantic Ocean where the biological production is low. Some values with both low temperature and fCO_2 can be seen in Fig. 3.2b, and are most likely due to coastal areas. During winter the SST values are relatively stable, with high fCO_2 values (Fig. 3.2f). Heating of the water masses during summer leads to an increase in fCO_2.

The fCO_2 value is nearly constant relatively to SSS during winter, but has a wider range in the summer. Fig. 3.2k shows two groups with low fCO_2 during summer. The first group, with values between 34 and 35 psu, are most likely associated with the East Greenland Current. These water masses are fresh and cold and originates from the Arctic Ocean flowing south along Greenland. The second group, with intermediate fCO_2 and SSS between 32 and 34, are presumably elements of coastal waters at the western side of the basin.

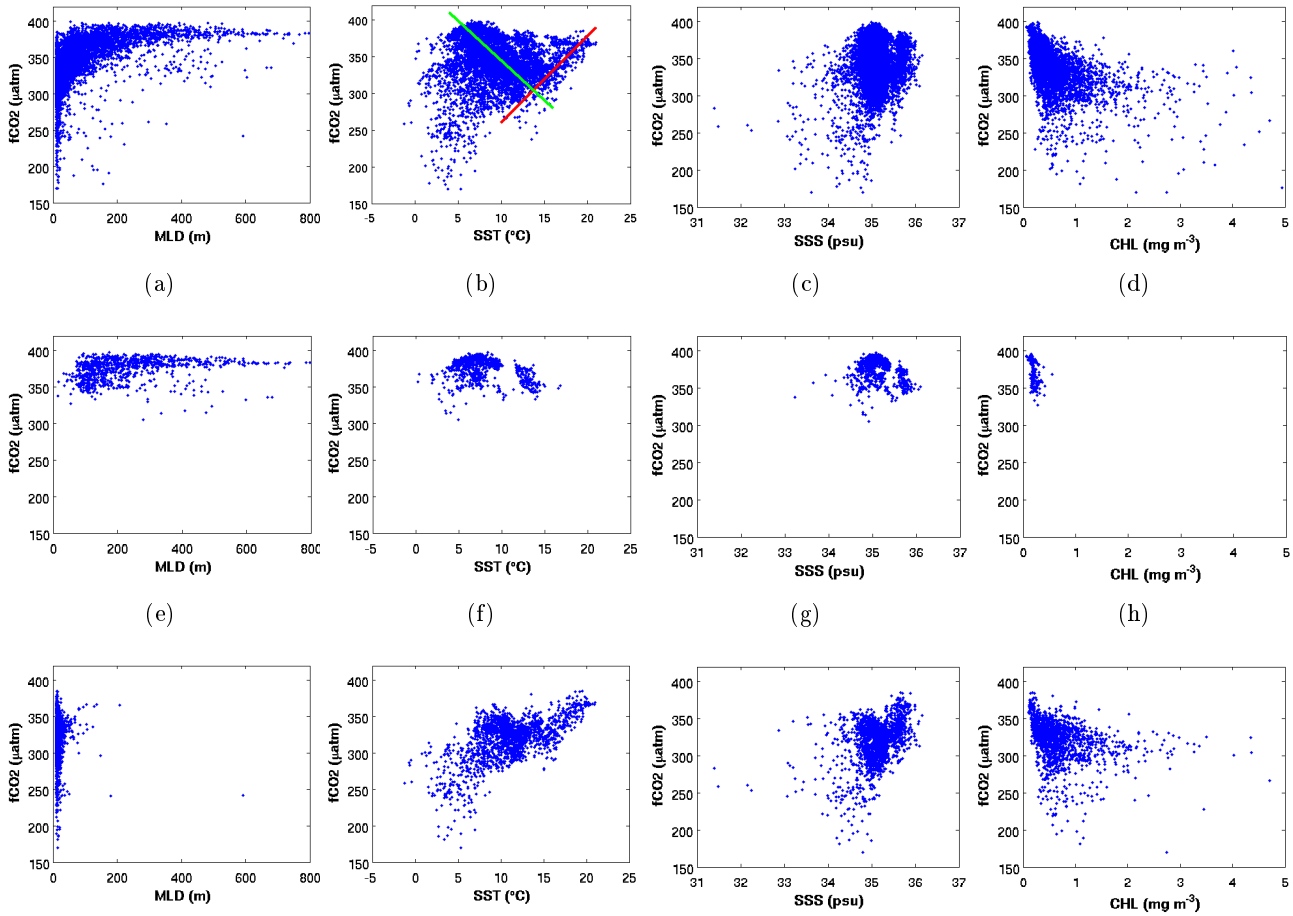


Figure 3.2: Plots of MLD, SST, SSS and CHL (from left to right) versus fCO_2 from data for the northern North Atlantic Ocean and the Nordic Seas: (a)-(d) total distribution; (e)-(h) relationship in winter; (i)-(l) relationship in summer. The red and green lines in figure (b) are drawn to illustrate trends in SST.

3.1.2 Self Organizing Maps

For each SOM analysis, A-D, the estimated and the measured *in-situ* fCO_2 is compared (Fig. 3.3). The results of the statistical analyses are presented in Tab. 3.3. The r^2 for the self organizing maps varies from 0.48 in SOM D to 0.63 in SOM A. The smallest RMSE, 22.6 µatm, is found in SOM A, while the largest RMSE, 27.1 µatm, is found in SOM C and D. Based on the modeling efficiency, SOM A and B are rated as 'very good', whereas SOM C and D are rated as 'good'. The cost function also

indicates that the most optimal map is SOM A, followed by B, C and D with the values 0.39, 0.43, 0.56 and 0.58 respectively.

Table 3.3: Skill assessment statistics for the Nordic Seas and the northern North Atlantic Ocean

Map name	r^2	RMSE (µatm)	ME	CF
SOM A	0.63	22.6	0.61	0.39
SOM B	0.59	23.3	0.57	0.43
SOM C	0.50	27.1	0.44	0.56
SOM D	0.48	27.1	0.42	0.58

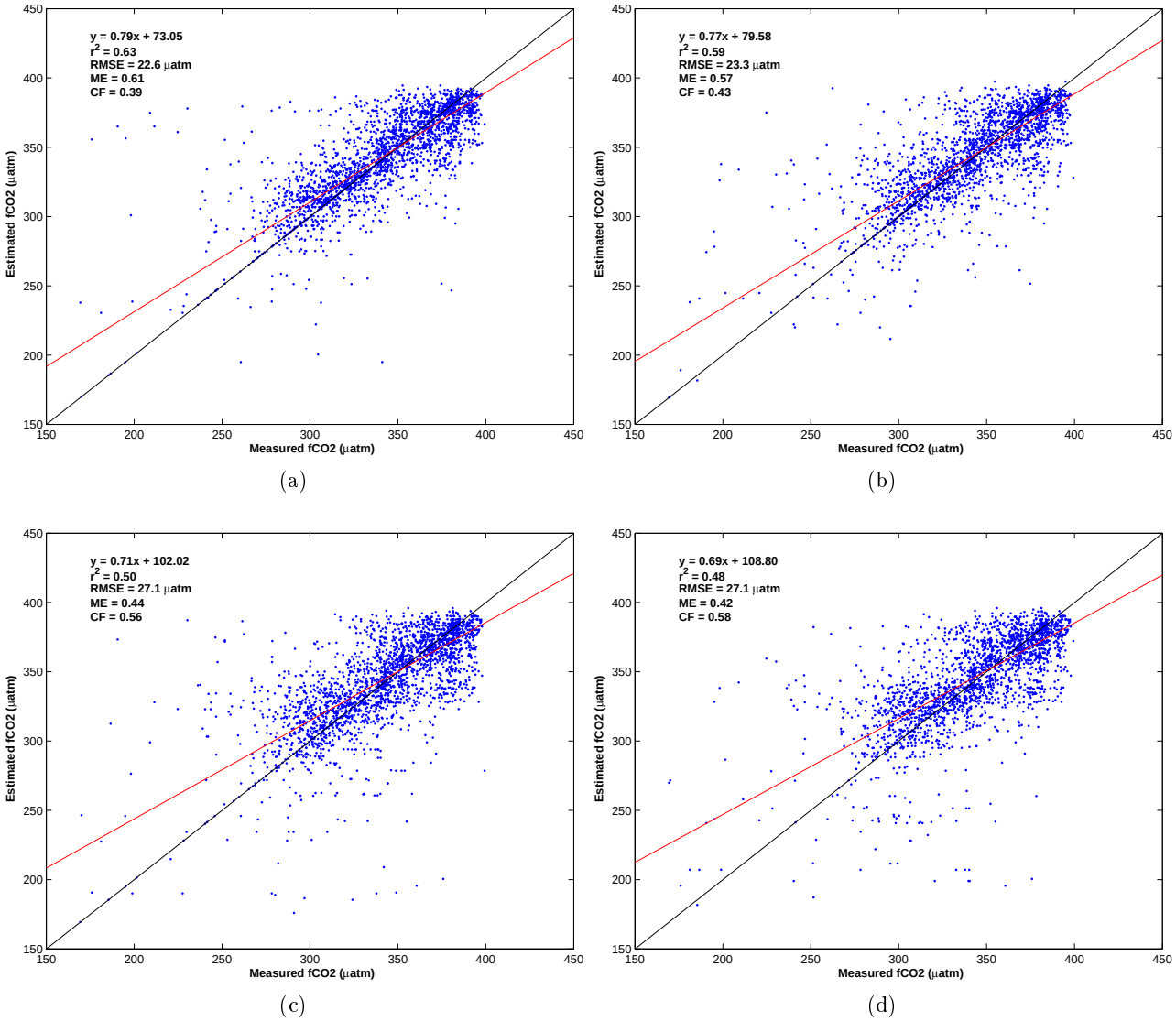


Figure 3.3: Comparison of fCO_2 estimates with measured values from the northern North Atlantic Ocean and the Nordic Seas for the different types of SOM: SOM A (a), SOM B (b), SOM C (c) and SOM D (d). The function represents the regression line (red), while the identity line (black) represents x=y.

Monthly fCO_2 maps for 2005 are presented in Fig. 3.4-3.7. Since the estimated maps have more

or less the same patterns, a description of the seasonal cycle of fCO_2 will only be given for the first SOM.

The seasonal fCO_2 cycle (Fig. 3.4) is characterized by a winter maximum and a summer minimum. This is caused by the two factors that control fCO_2 during the season; temperature and deep mixing in the winter season, and biological production in the summer season.

The Nordic Seas and the northern North Atlantic Ocean have high estimated fCO_2 values from January through February. In March a gradual depletion begins, due to biological production (Fig. 2.12). This occurs as the water masses warm (Fig. 2.9) and stratifies (Fig. 2.10). The biological impact is maintained throughout April and May, and the lowest fCO_2 values are reached in June. The low values persists in the majority of the ocean in July and August. A change in fCO_2 can bee seen at around 44°N-48°N at the southern reaches of the study area, in July, August and September, with values noticeable higher than for the rest of the basin. This is most likely due to the nutrient-poor water masses in the subtropical gyre (∼20°N-45°N), an area where biological production is absent. These warm waters cause the fCO_2 increase. The mixed layer deepens during fall and winter, which leads to an enhanced uptake of fCO_2 from the atmosphere to the ocean.

SOM A: MLD and SST

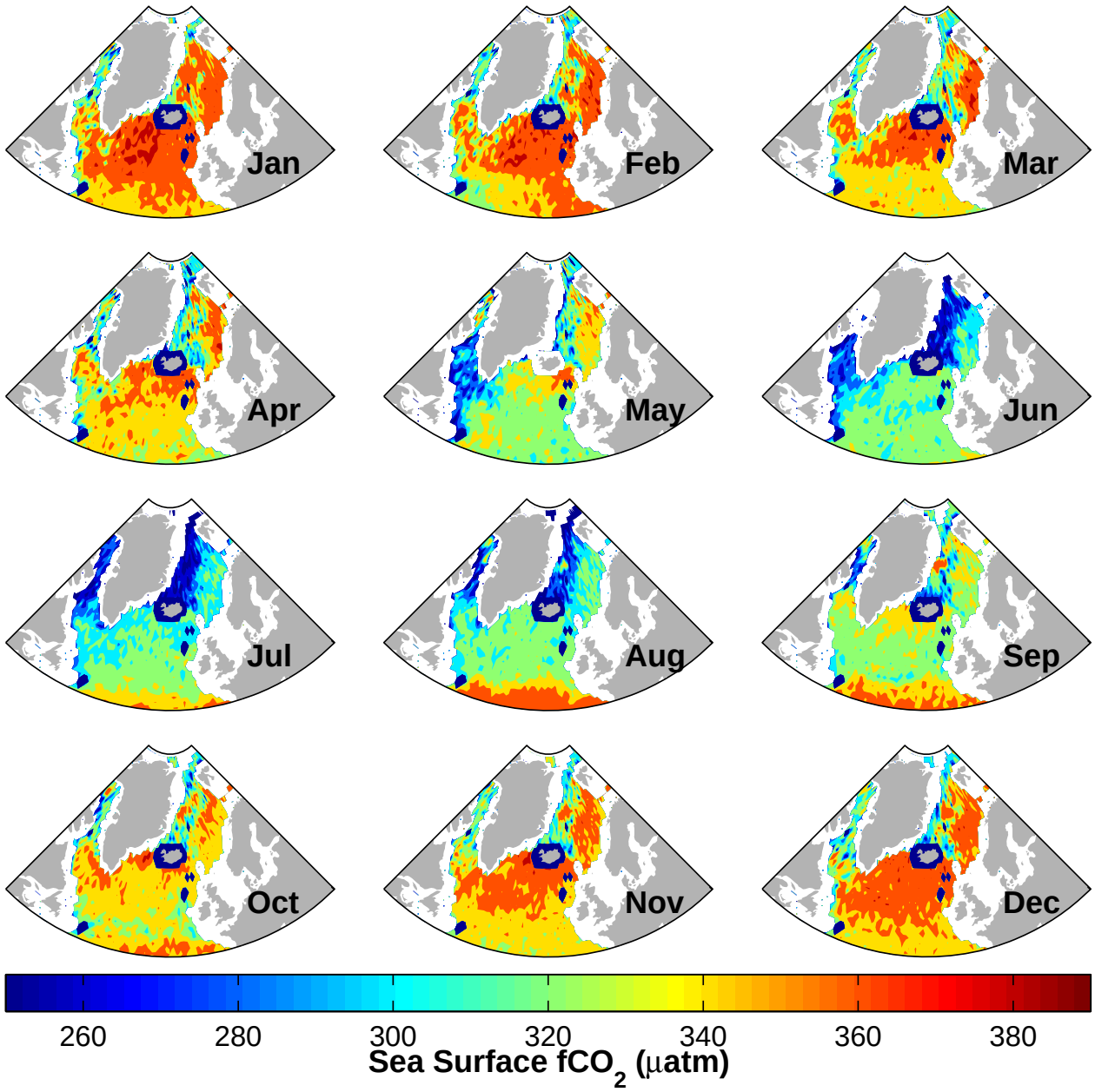


Figure 3.4: SOM A: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean and the Nordic Seas for 2005. SOM A was prepared using the following settings: 'msize': [90 50], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': [20 15].

SOM B: MLD, SST and CHL (summer only)

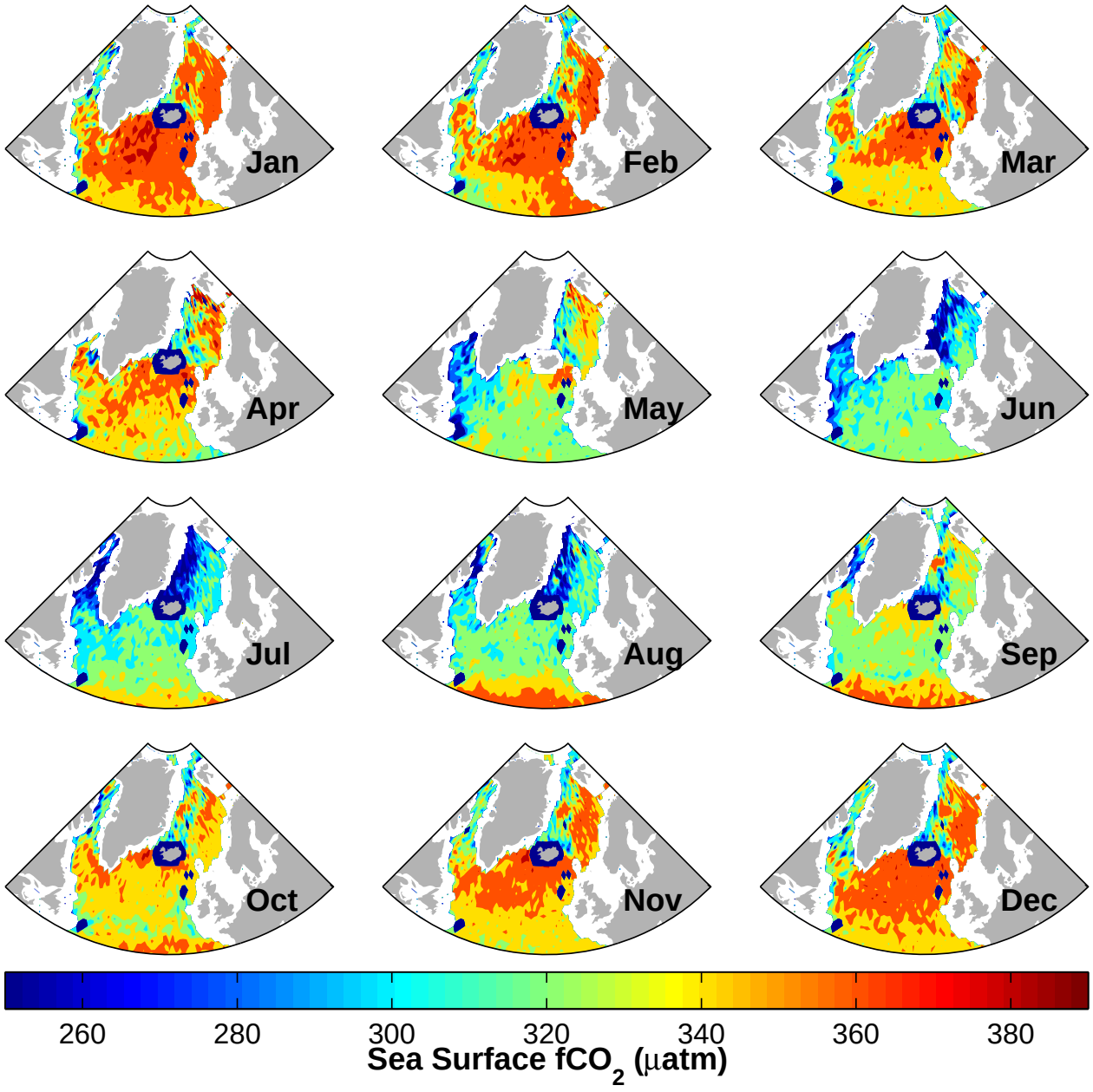


Figure 3.5: SOM B: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean and the Nordic Seas for 2005. SOM B was prepared using the following settings: 'msize': Summer/Winter: [90 50], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': Summer: [25 20], Winter: [20 15].

SOM C: MLD, SST and SSS

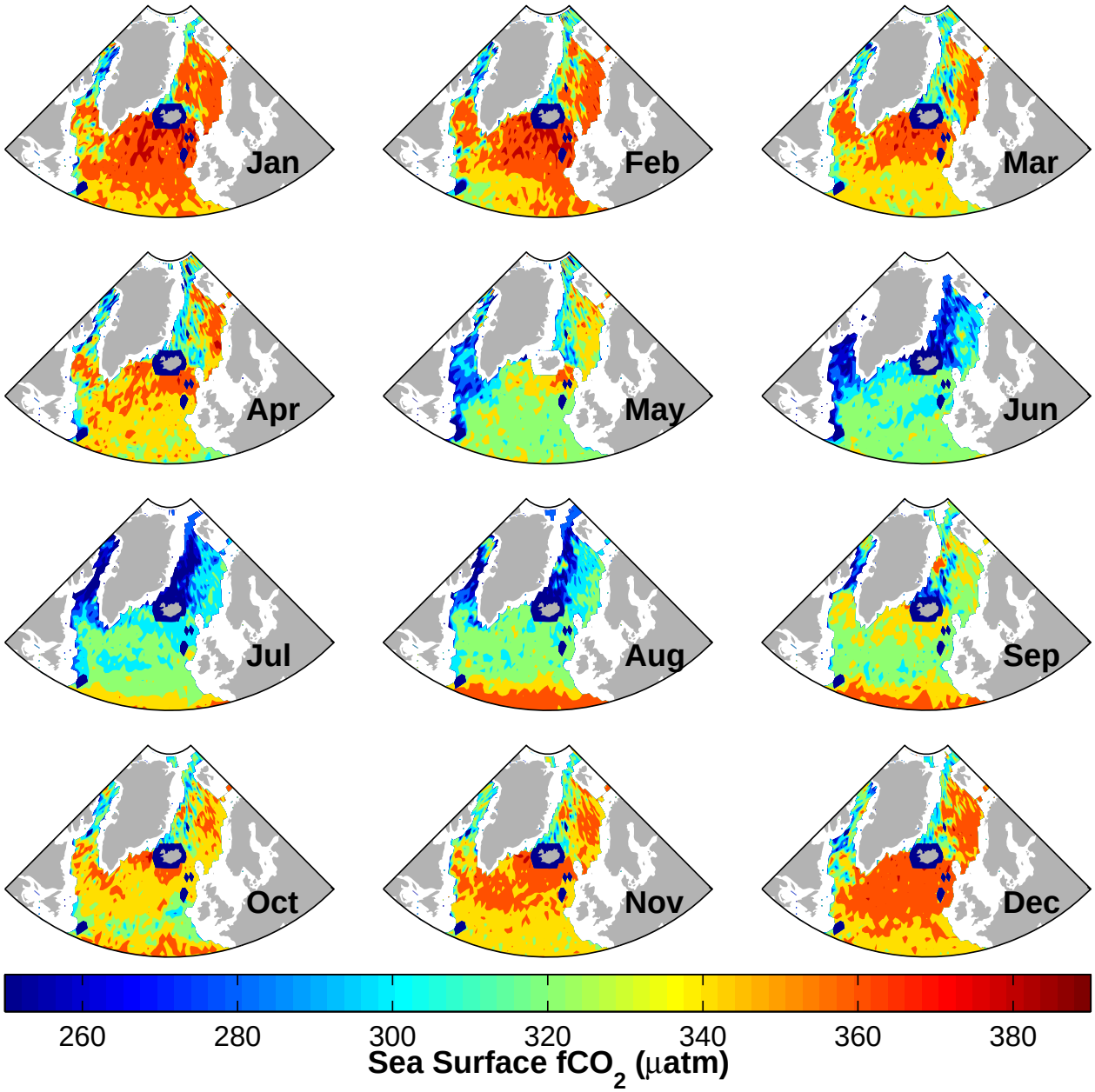


Figure 3.6: SOM C: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean and the Nordic Seas for 2005. SOM C was prepared using the following settings: 'msize': [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': [20 15].

SOM D: MLD, SST, SSS and CHL (summer only)

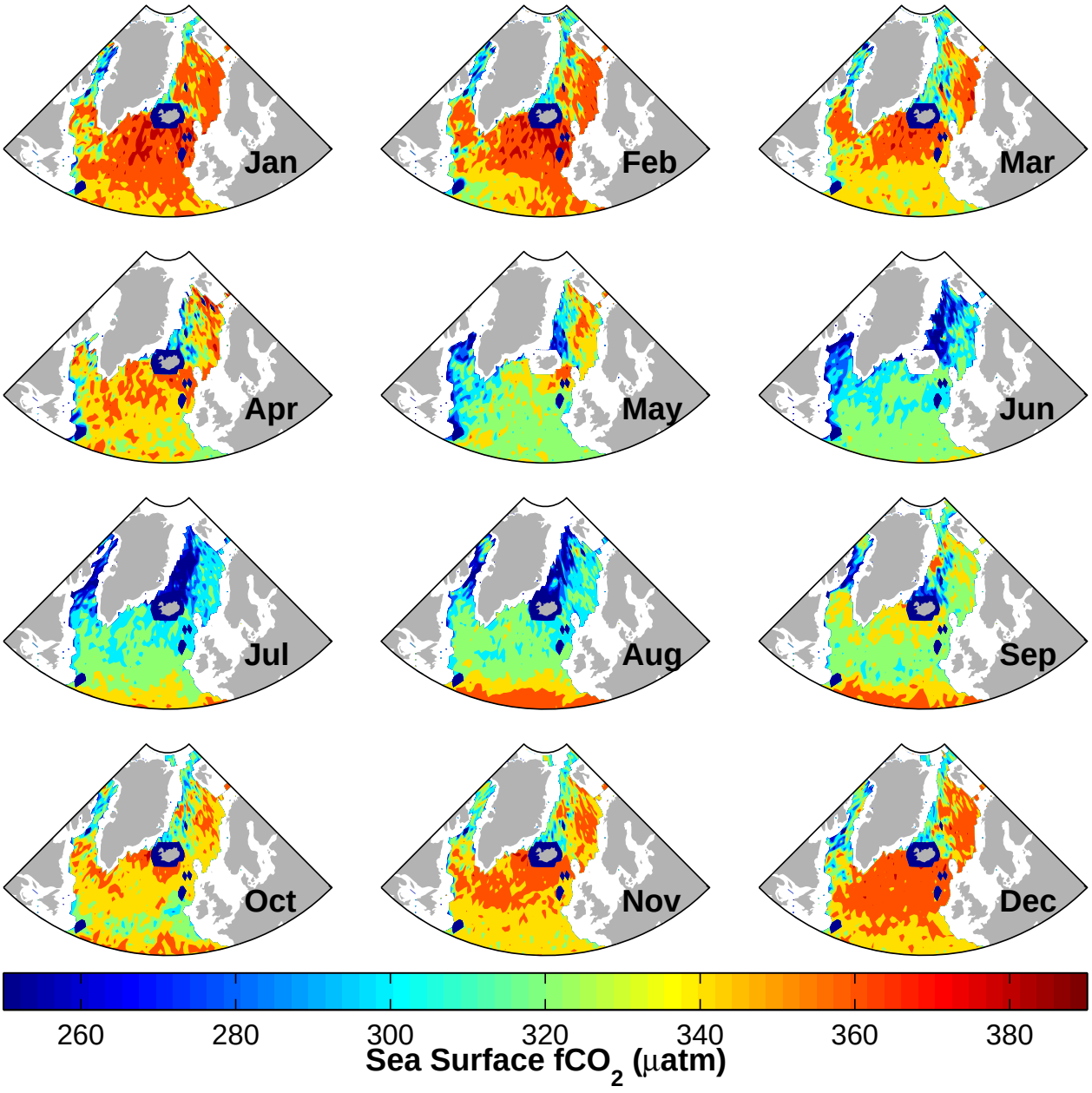


Figure 3.7: SOM D: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean and the Nordic Seas for 2005. SOM D was prepared using the following settings: 'msize': Summer: [60 40], Winter: [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': Summer: [15 10], Winter: [20 15].

3.2 The northern North Atlantic Ocean [44-63°N]

3.2.1 Underlying Functional Relationships

Comparison of the plots of MLD, SST, SSS and CHL against fCO_2 for the northern North Atlantic Ocean and the Nordic Seas (Fig. 3.2) with the same plots for only the northern North Atlantic Ocean (Fig. 3.8), reveals more or less similar patterns. Except for some outliers with low SST and SSS values that occurred in the northern North Atlantic Ocean and the Nordic Seas, Fig. 3.8 is essentially identical to Fig. 3.2.

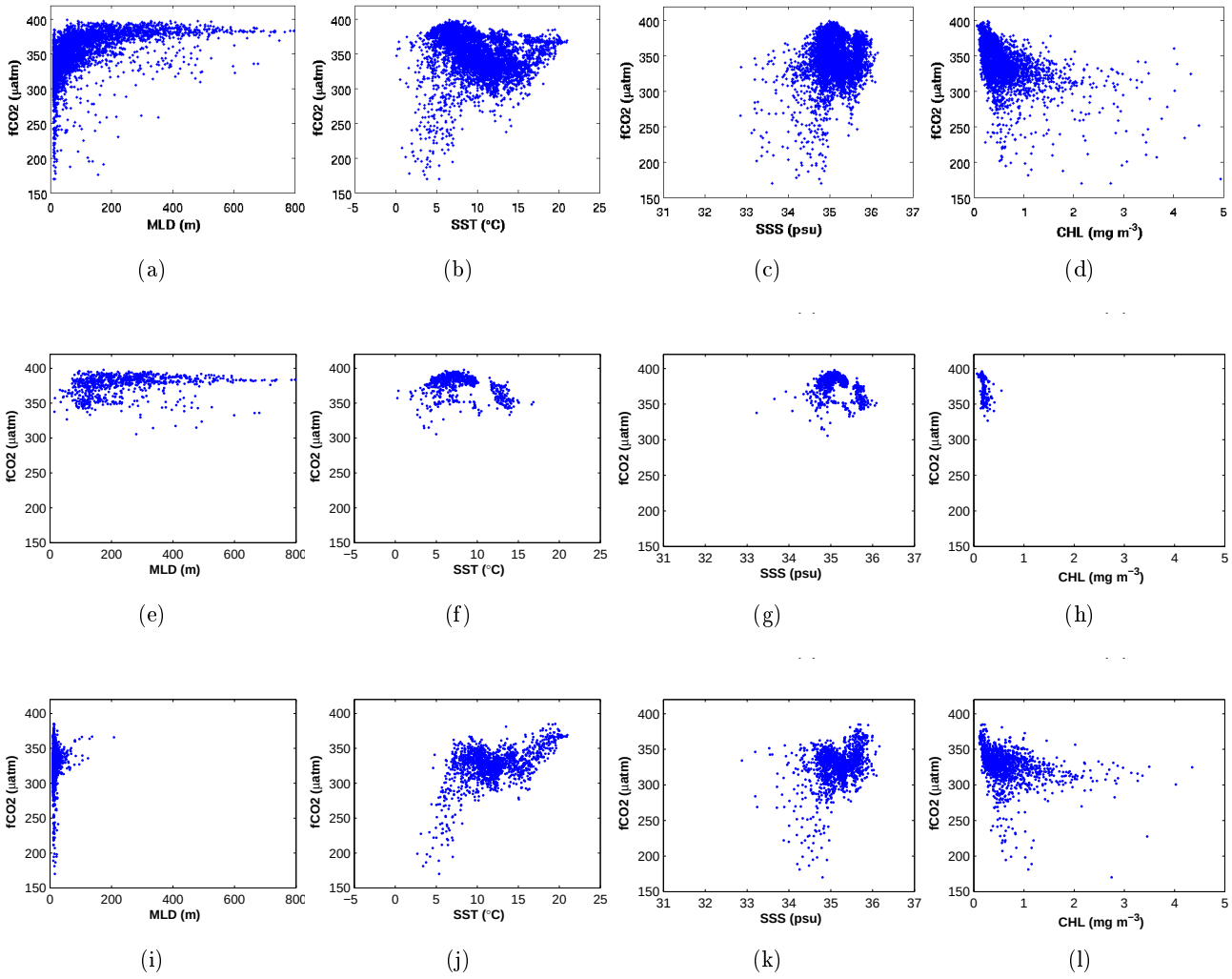


Figure 3.8: Plots of MLD, SST, SSS and CHL (from left to right) versus fCO_2 from data for the northern North Atlantic Ocean: (a)-(d) total distribution; (e)-(h) relationship in winter; (i)-(l) relationship in summer.

3.2.2 Self Organizing Maps

Plots of estimated versus measured fCO_2 values are shown in Fig. 3.9, while the skill assessment metrics are presented in Tab. 3.4. The r^2 values range from 0.51 to 0.61 for SOM D and SOM A, respectively. The RMSEs are more or less equal to those for the northern North Atlantic Ocean and the Nordic Seas, ranging from 23.1-26.1 µatm. The ME's are the same as for the previous estimations, where SOM A and B are rated as 'very good' and the two others are rated as 'good'. CF is calculated to be between 0.41 and 0.53 for SOM A and SOM D, respectively. Monthly fCO_2 maps for 2005 are

presented in Fig. 3.10-3.13, and will be further discussed in chapter 4.

Table 3.4: Skill assessment statistics for the northern North Atlantic Ocean

Map name	r^2	RMSE (µatm)	ME	CF
SOM A	0.61	23.1	0.59	0 41
SOM B	0.57	24.2	0.55	0.45
SOM C	0.52	25.8	0.49	0.51
SOM D	0.51	26.1	0.47	0.53

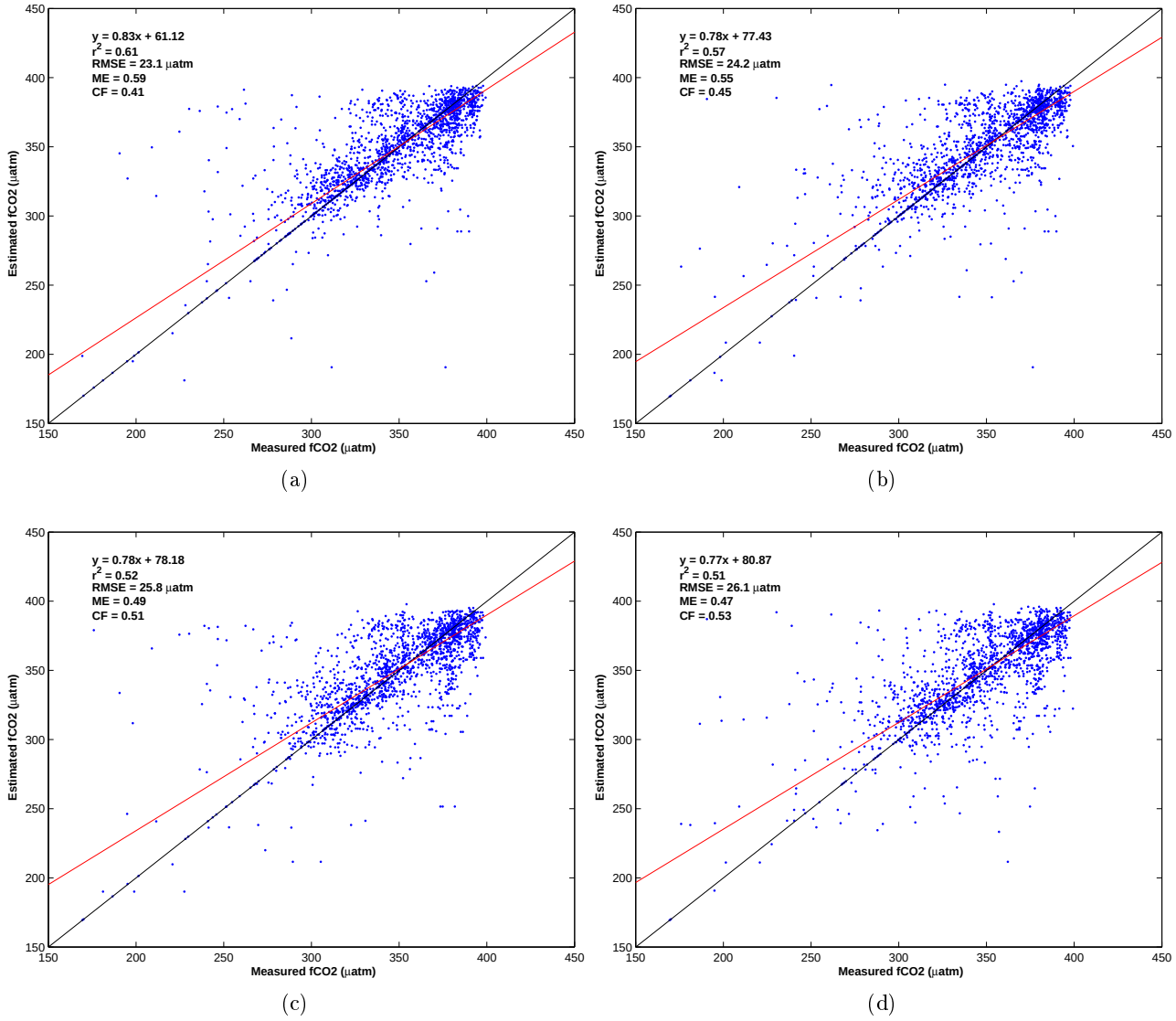


Figure 3.9: Comparison of fCO_2 estimates with measured values from the northern North Atlantic Ocean for the different types of SOM: SOM A (a), SOM B (b), SOM C (c) and SOM D (d). The function represents the regression line (red), while the identity line (black) represents x=y.

SOM A: MLD and SST

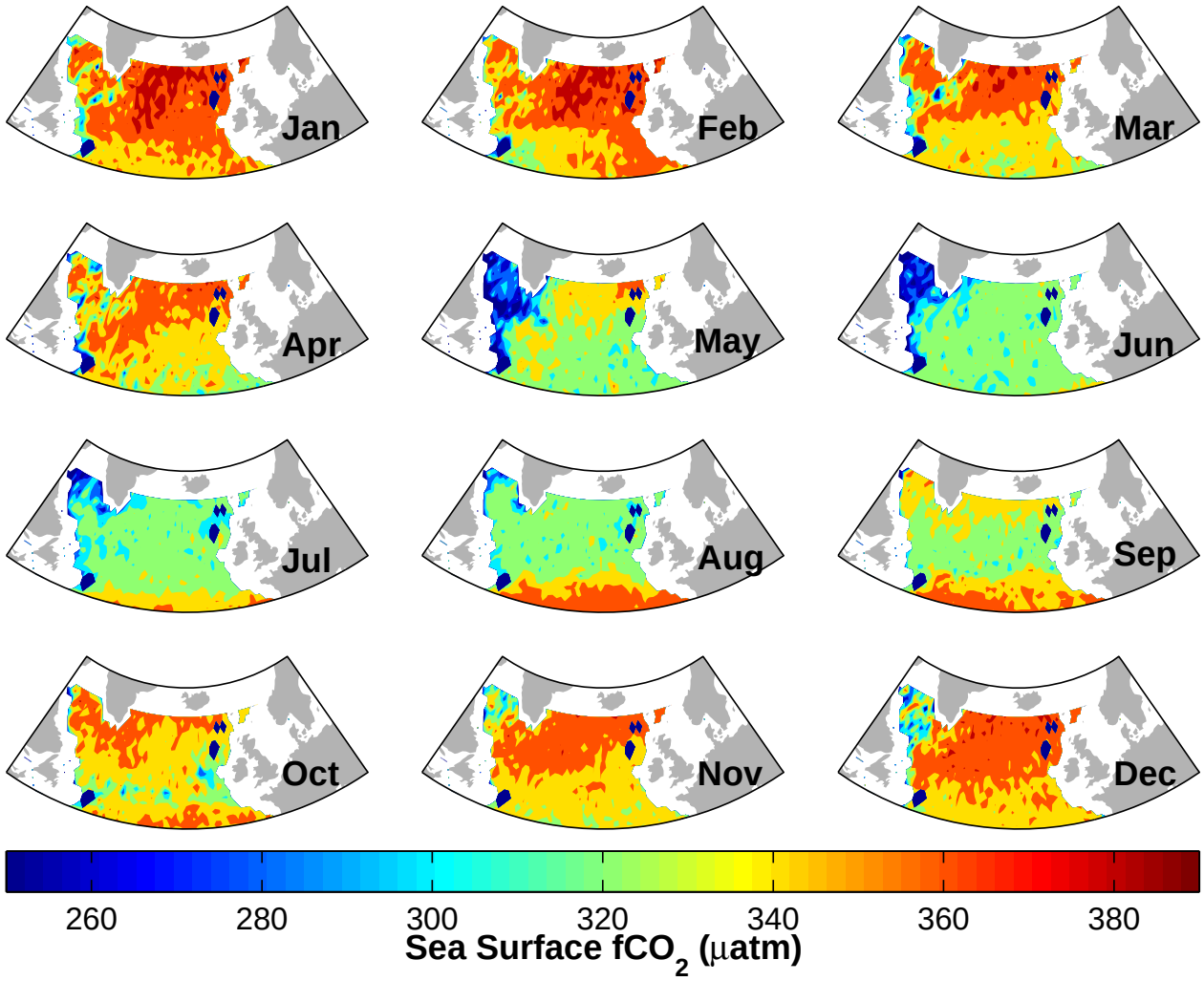


Figure 3.10: SOM A: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean for 2005. SOM A was prepared using the following settings: 'msize': [85 50], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': [15 10].

SOM B: MLD, SST and CHL (summer only)

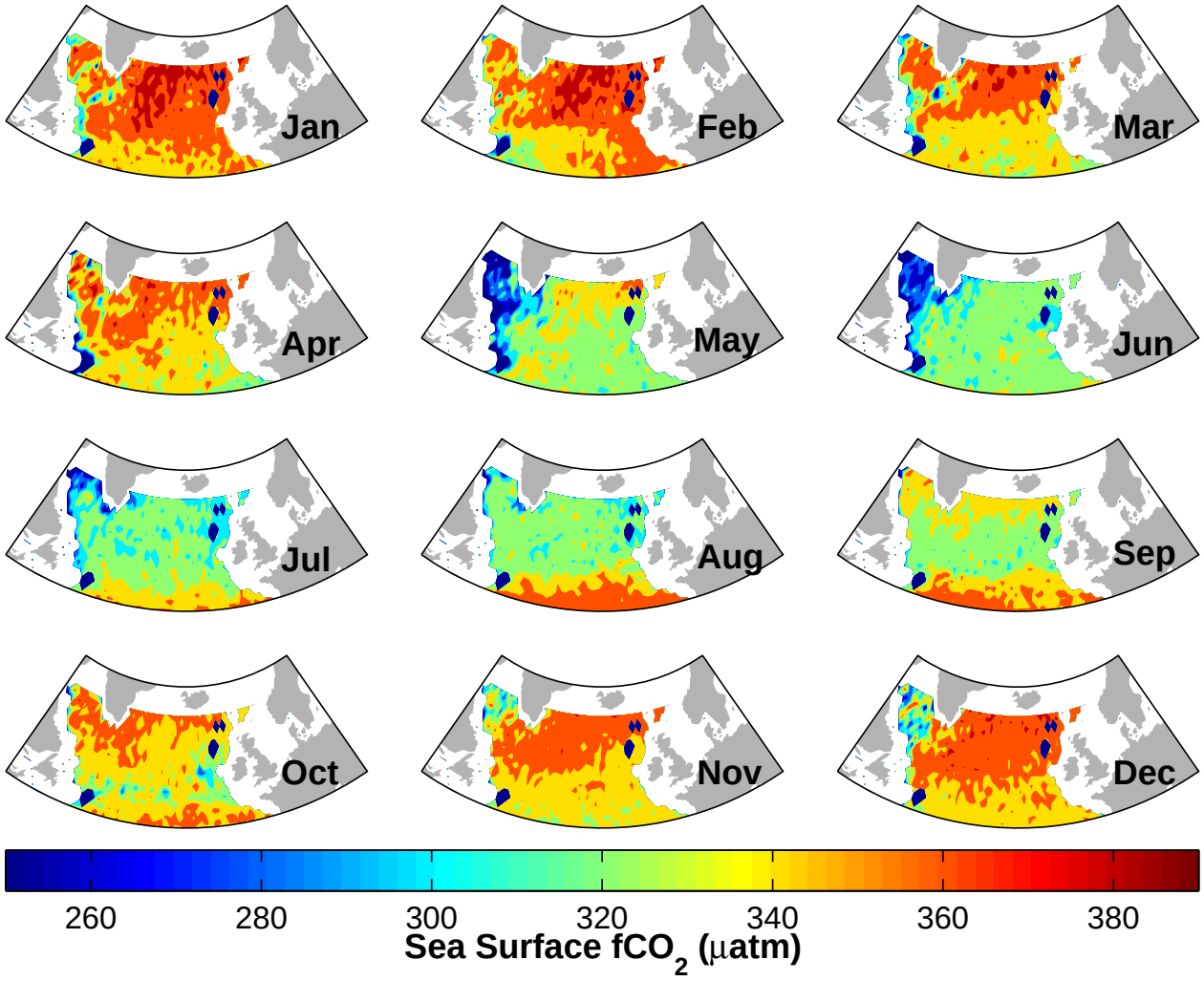


Figure 3.11: SOM B: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean for 2005. SOM B contains the parameters: 'msize': Summer: [100 65], Winter: [85 50], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': Summer: [25 20], Winter: [15 10].

SOM C: MLD, SST and SSS

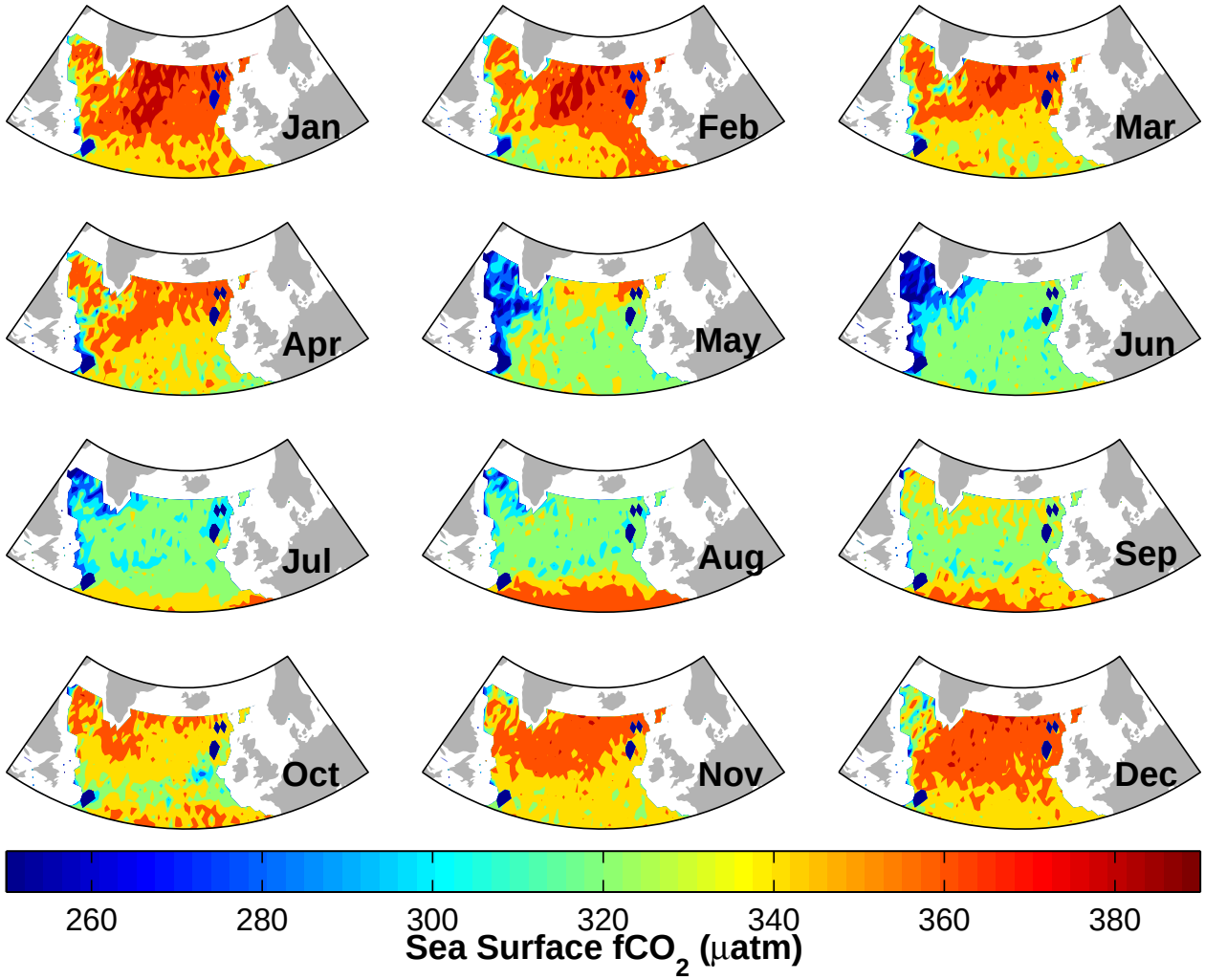


Figure 3.12: SOM C: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean for 2005. SOM C was prepared using the following settings: 'msize': [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': [25 20].

SOM D: MLD, SST, SSS and CHL (summer only)

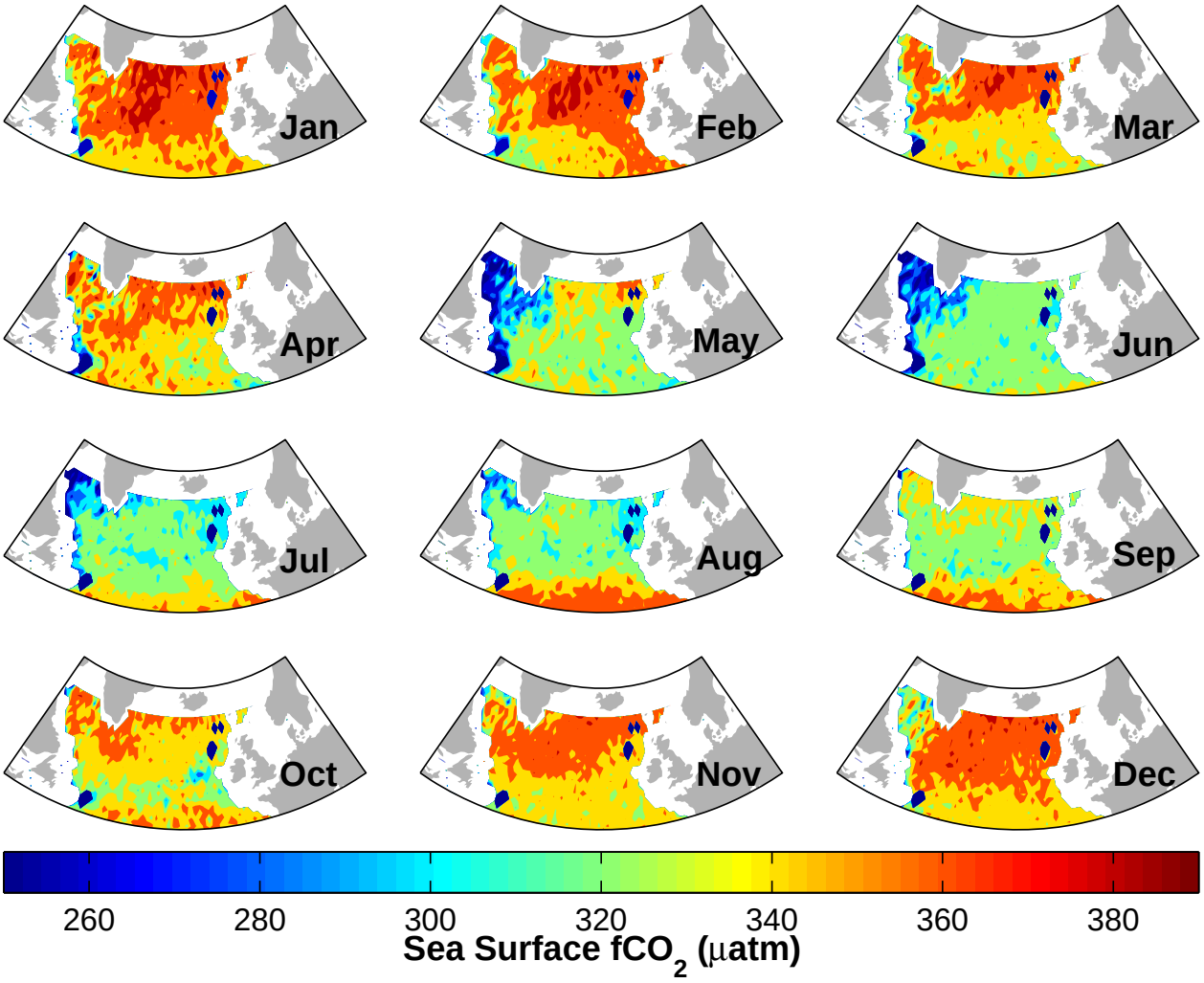


Figure 3.13: SOM D: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean for 2005. SOM D was prepared using the following settings: 'msize': Summer/Winter: [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': Summer: [15 10], Winter: [25 20].

3.3 The Nordic Seas [63-85°N]

3.3.1 Underlying Functional Relationships

When comparing the property-fCO_2 plot for the northern North Atlantic Ocean and the Nordic Seas (Fig. 3.3) with the plot for the Nordic Seas (Fig. 3.15), the most prominent difference is the lower amount of data. A logarithmic shape can still be seen for the MLD-fCO_2 relationship (Fig. 3.14a), although the amount of data is significantly reduced. Despite the reduced patterns, the MLD transition from winter (Fig. 3.14e) to summer (Fig. 3.14i) can still be seen. Fig. 3.14h is blank, as expected, since no CHL data is available during winter. The low SST and fCO_2 values seen in Fig. 3.2j, seems to come from the Nordic Seas.

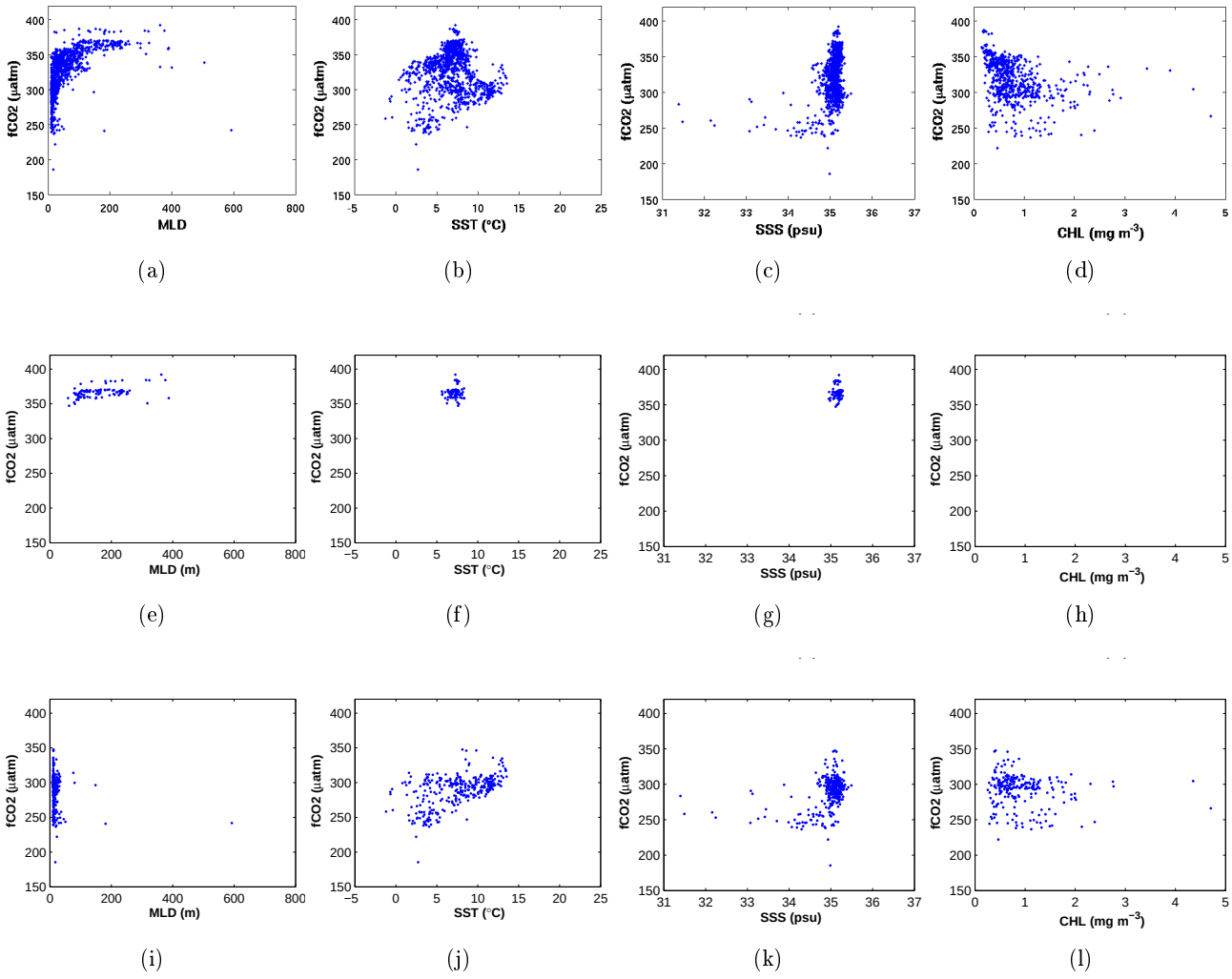


Figure 3.14: Plots of MLD, SST, SSS and CHL (from left to right) versus fCO_2 from data for the Nordic Seas: (a)-(d) total distribution; (e)-(h) relationship in winter; (i)-(l) relationship in summer.

3.3.2 Self Organizing Maps

The skill assessment statistics, presented in Tab. 3.5 based on the results in Fig. 3.15, are better for the Nordic Seas than the other two basins. The r^2 reaches 0.85 in SOM A, whereas the other SOMs range from 0.70 to 0.78. There is a drastic improvement in the RMSE for all SOMs, and the lowest and highest errors are 12.5 and 17.9 µatm, respectively. An enhanced modeling efficiency is also seen for all the SOMs, where SOM A-D are rated to have an 'excellent' ME. The highest ME belongs to

SOM A, with an ME of 0.84. CF values are also noticeably smaller than for the others basins, with following results: 0.16, SOM A; 0.24, SOM B; 0.32, SOM C; 0.30, SOM D. Monthly fCO_2 maps for 2005 are presented in Fig. 3.16-3.19, and will be further discussed in chapter 4.

Table 3.5: Skill assessment statistics for the Nordic Seas

Map name	r^2	RMSE (µatm)	ME	CF
SOM A	0.85	12.5	0.84	0.16
SOM B	0.78	14.6	0.76	0.24
SOM C	0.70	17.9	0.68	0.32
SOM D	0.72	16.4	0.70	0.30

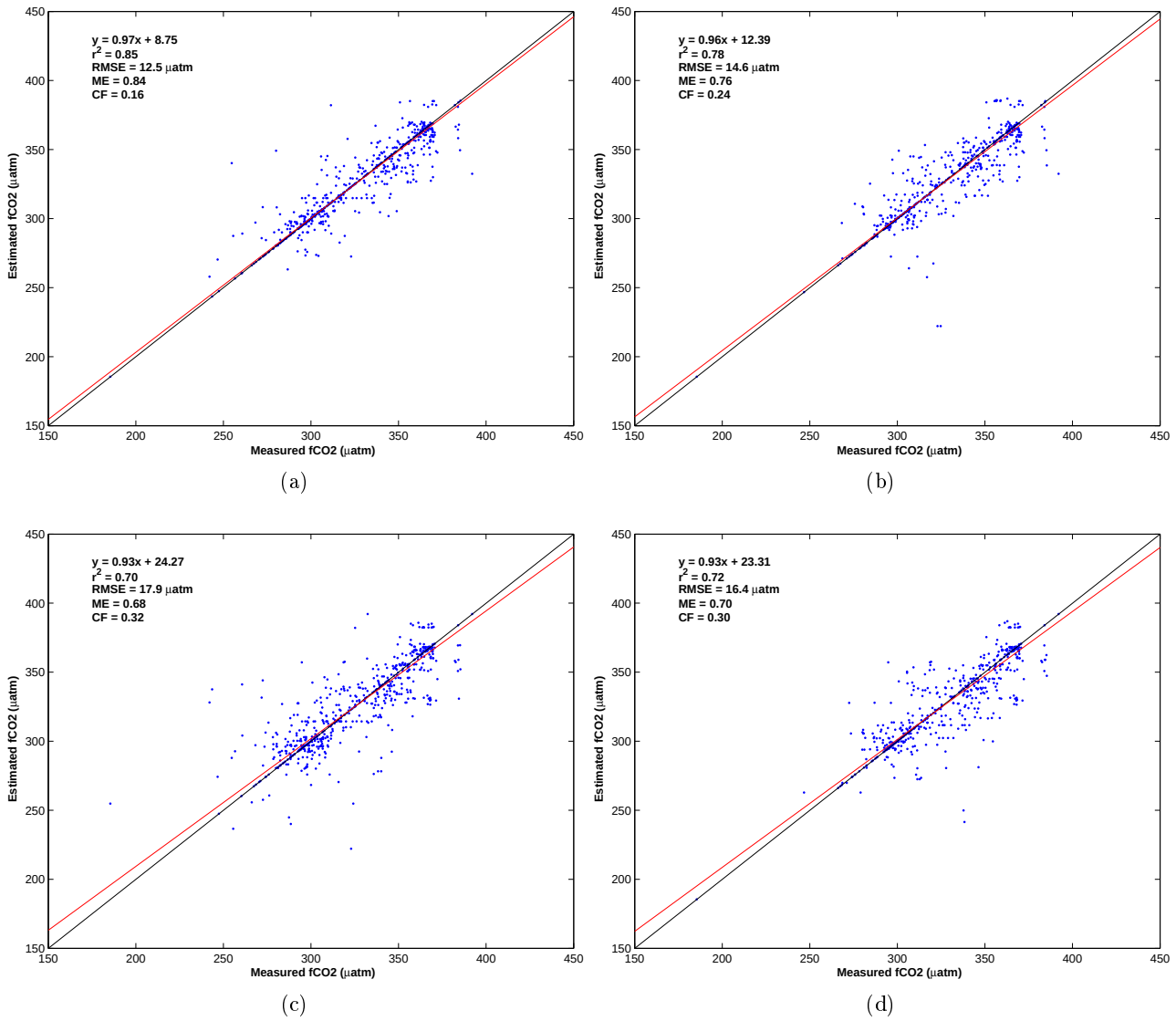


Figure 3.15: Comparison of fCO_2 estimates with measured values from the Nordic Seas for the different types of SOM: SOM A (a), SOM B (b), SOM C (c) and SOM D (d). The function represents the regression line (red), while the identity line (black) represents x=y.

SOM A: MLD and SST

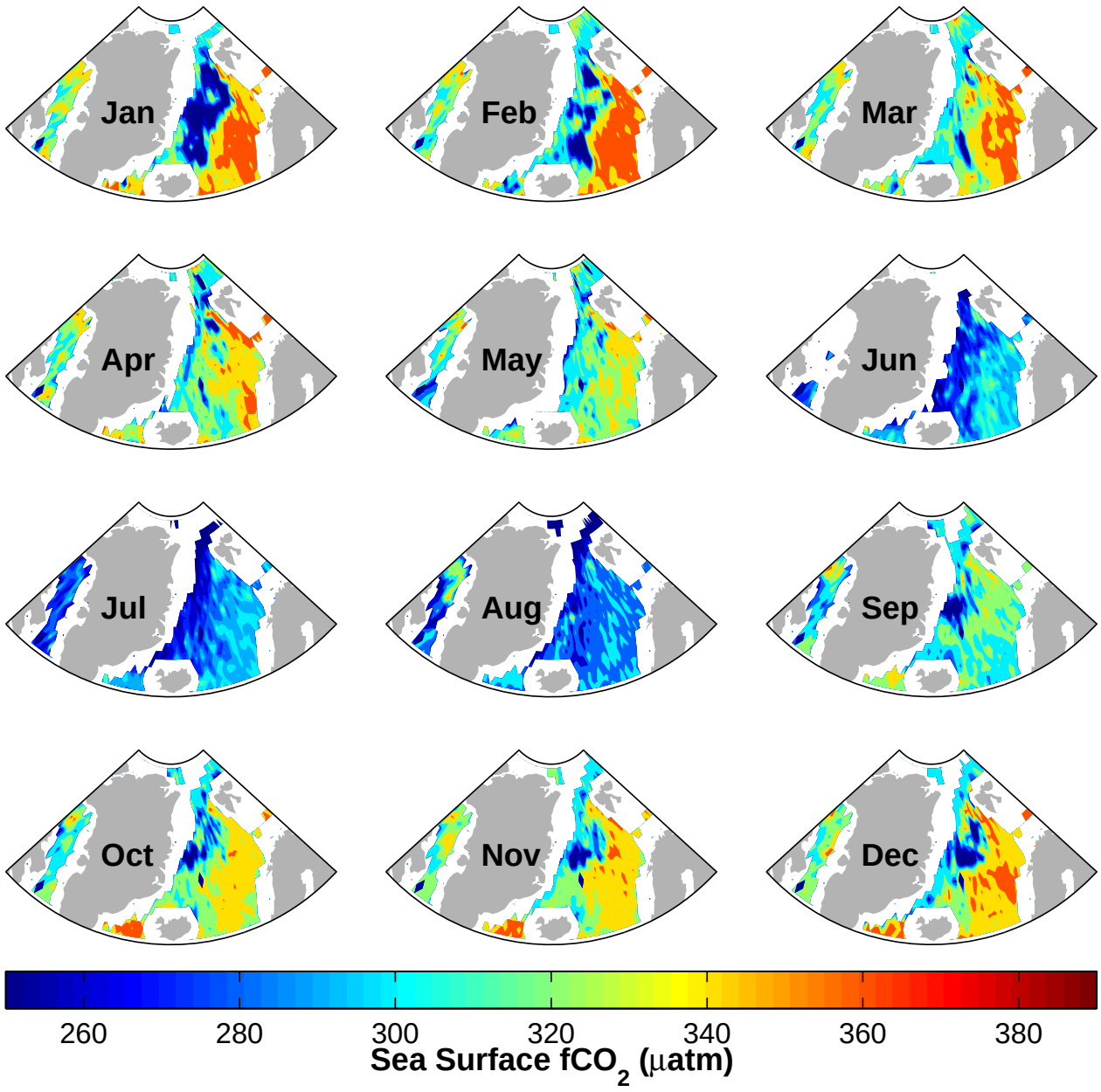


Figure 3.16: SOM A: Monthly distribution of estimated sea surface fCO_2 in the Nordic Seas for 2005. SOM A was prepared using the following settings: 'msize': [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': [25 20].

SOM B: MLD, SST and CHL (summer only)

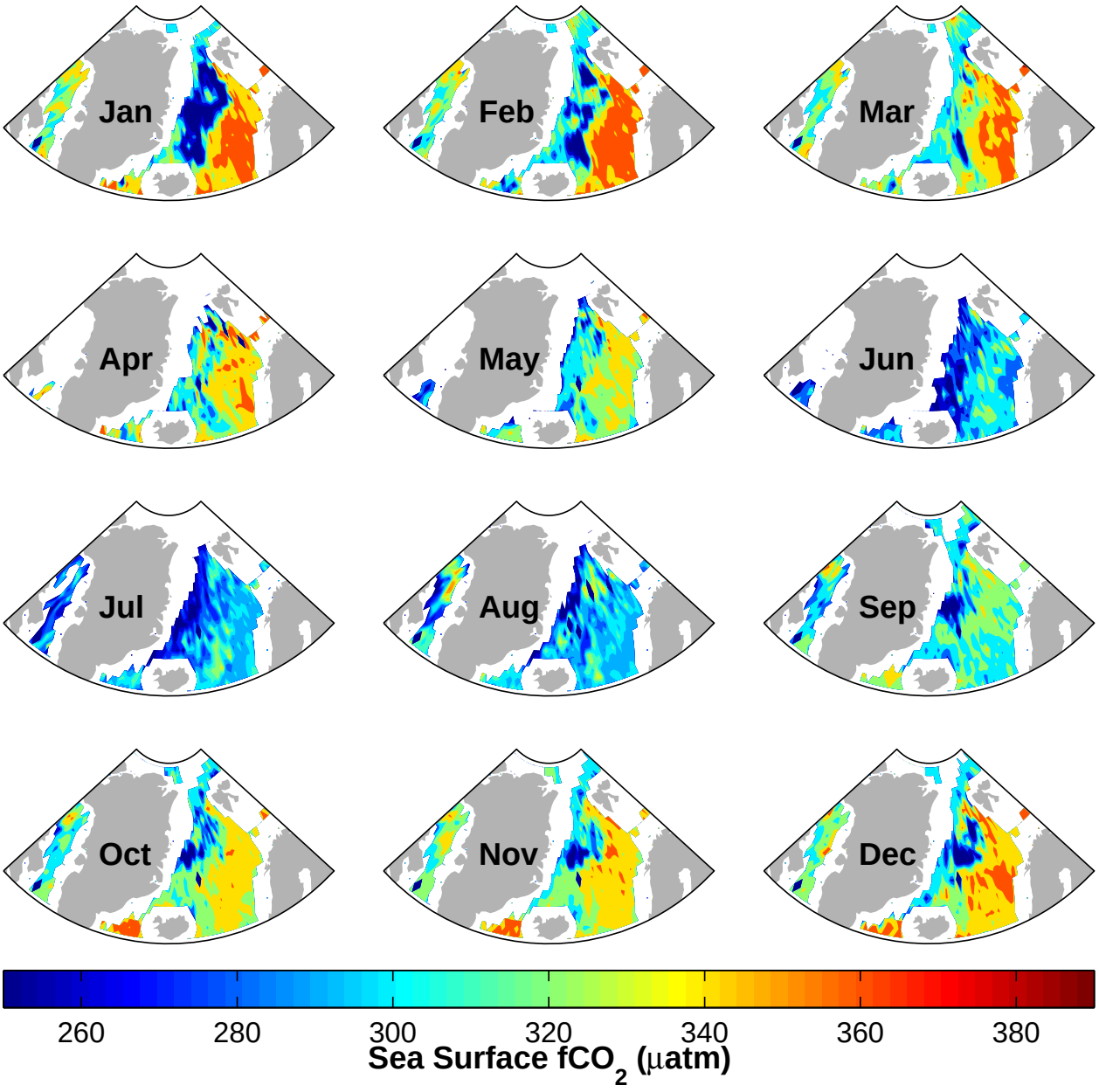


Figure 3.17: SOM B: Monthly distribution of estimated sea surface fCO_2 in the Nordic Seas for 2005. SOM B was prepared using the following settings: 'msize': Summer/Winter: [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': Summer/Winter: [25 20].

SOM C: MLD, SST and SSS

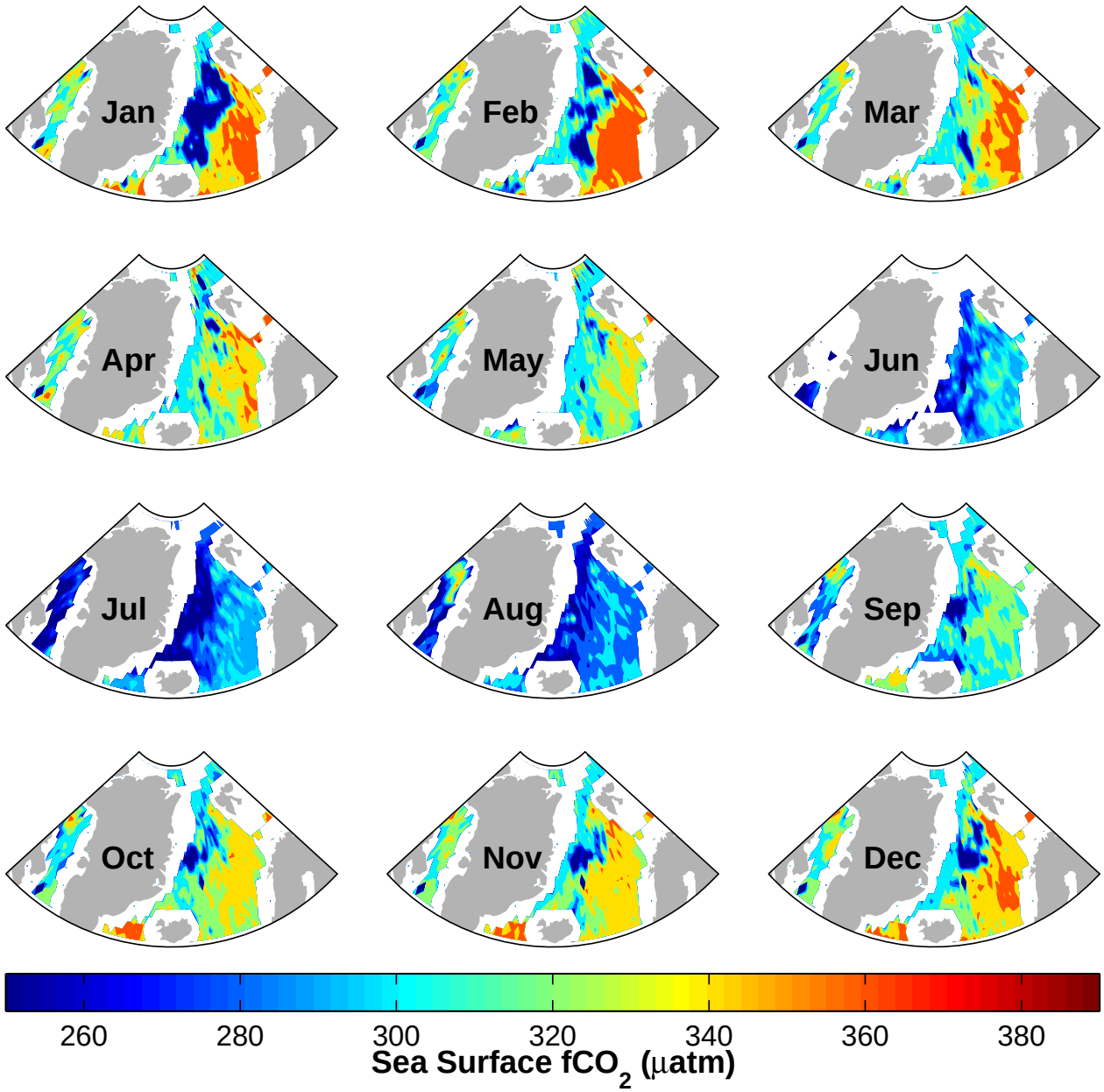


Figure 3.18: SOM C: Monthly distribution of estimated sea surface fCO_2 in the Nordic Seas for 2005. SOM C was prepared using the following settings: 'msize': [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': [25 20].

SOM D: MLD, SST, SSS and CHL (summer only)

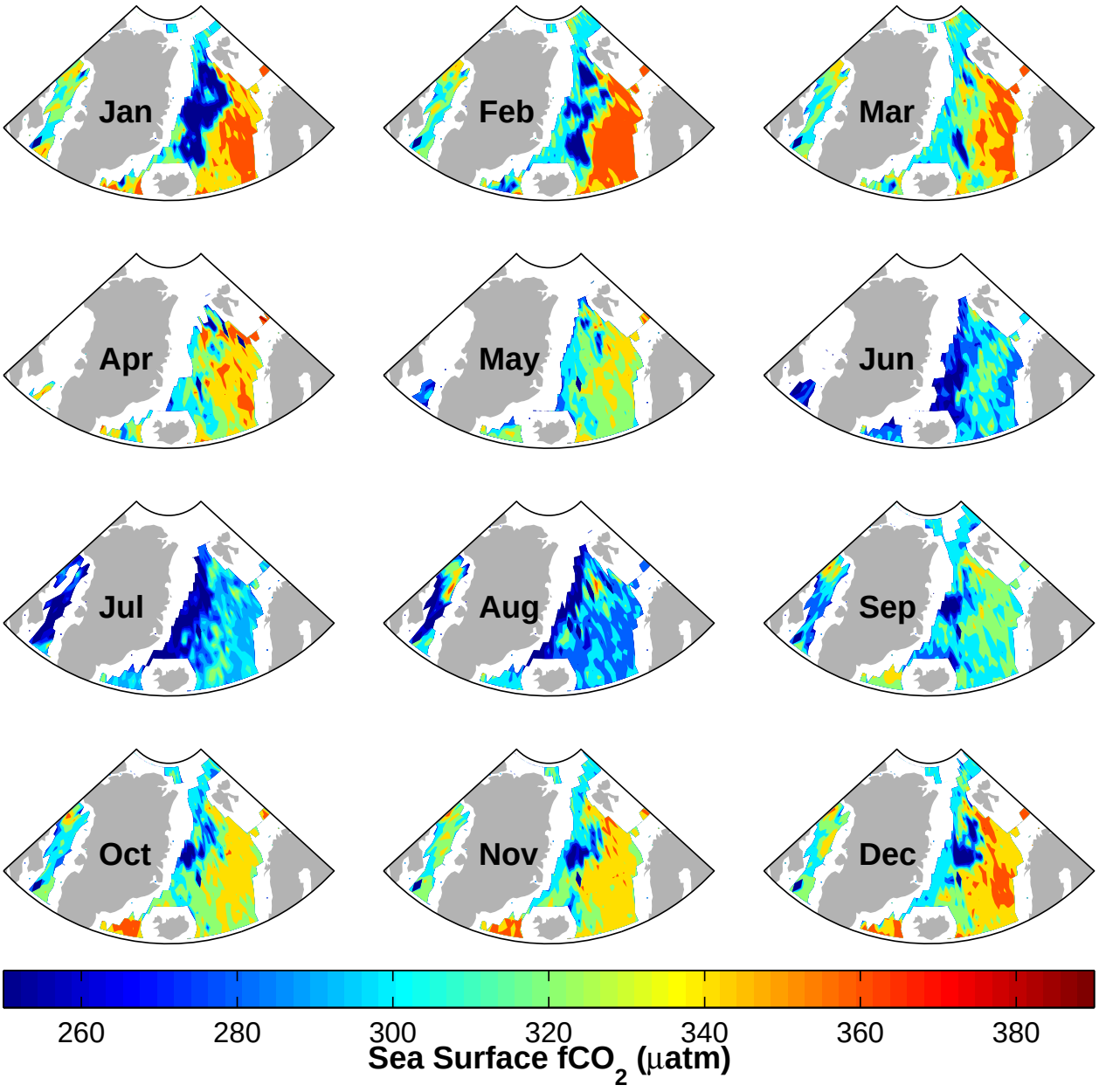


Figure 3.19: SOM D: Monthly distribution of estimated sea surface fCO_2 in the Nordic Seas for 2005. SOM D was prepared using the following settings: 'msize': Summer/Winter: [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': Summer: [15 10], Winter: [25 20].

3.4 Merging Two SOMs [44-85°N]

An attempt was made to merge the SOMs of the northern North Atlantic Ocean (44-63°N) and the Nordic Seas (63-85°N), to create one large basin. This was done in order to compare the merged estimates with the estimates for the entire region (44-85°N). Skill assessment statistic are presented in Tab. 3.6, while the linear regression of the estimated and measured fCO_2 are shown in Fig. B.13. The coefficient of determination ranges from 0.57 to 0.68, where SOM A has the best fCO_2 prediction. The difference between the best and the worst RMSE is quite small, only 3,3 µatm. SOM A is the only map that is rated as 'excellent', whereas the others are rated as 'very good'.

Monthly fCO_2 maps for 2005 are presented in Fig. 3.21-3.24, which will be discussed in chapter 4.

Table 3.6: Skill assessment statistics for merged SOMs

Map name	r^2	RMSE (µatm)	ME	CF
SOM A	0.68	20.9	0.67	0.33
SOM B	0.63	22.3	0.61	0.39
SOM C	0.59	24.1	0.56	0.44
SOM D	0.57	24.2	0.54	0.46

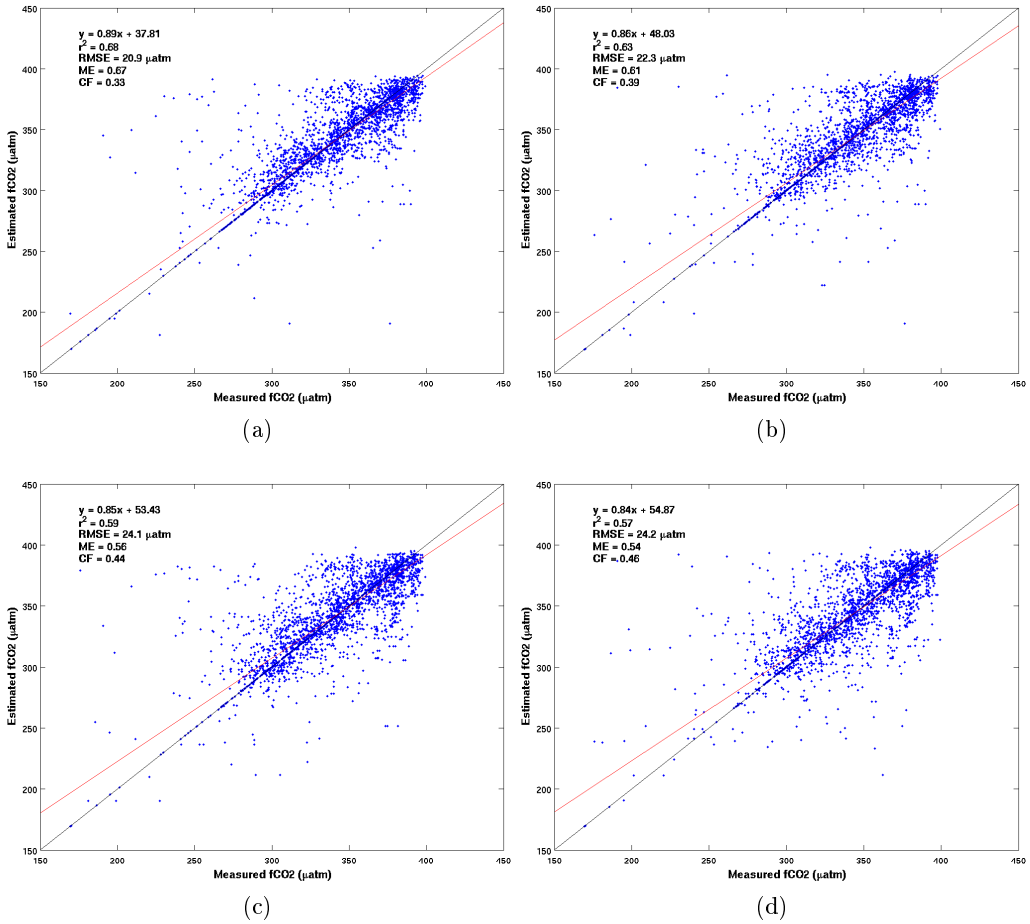


Figure 3.20: Comparison of fCO_2 estimates with measured values from the merged basins. SOM A (a), SOM B (b), SOM C (c) and SOM D (d). The function represents the regression line (red), while the identity line (black) represents x=y.

SOM A: MLD and SST

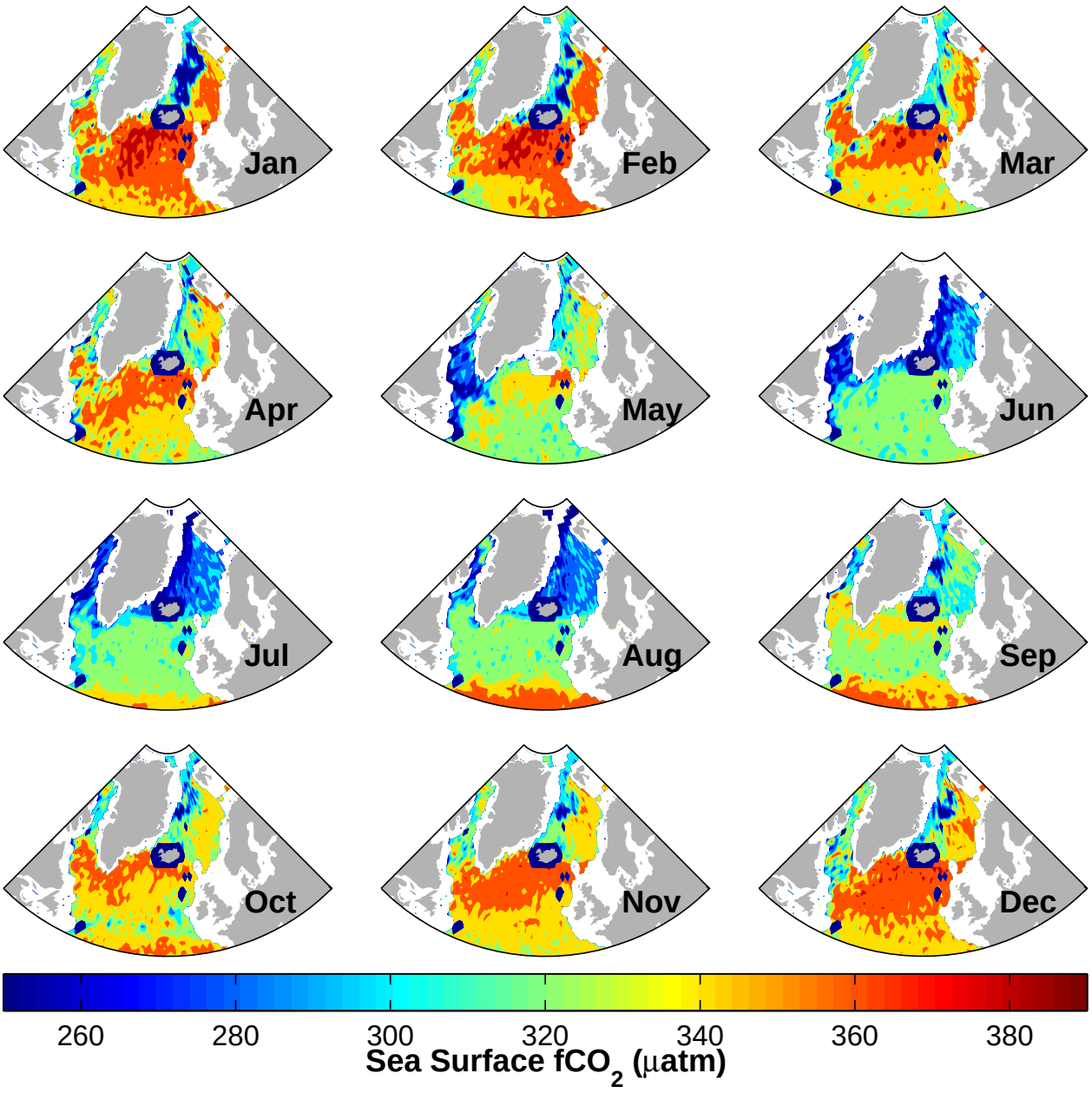


Figure 3.21: SOM A: Monthly distribution of estimated sea surface fCO_2 for the merged SOM in 2005.

SOM B: MLD, SST and CHL (summer only)

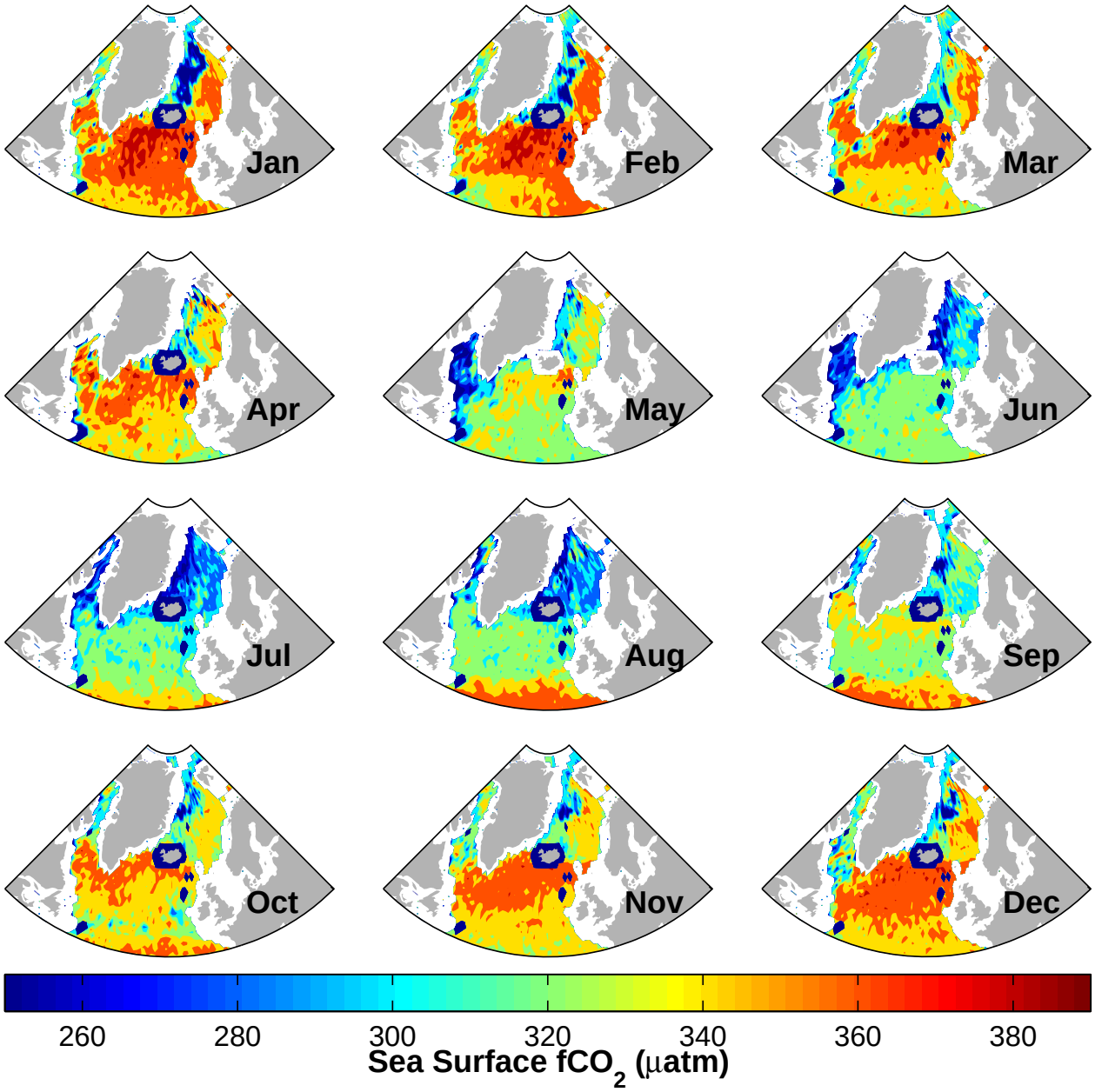


Figure 3.22: SOM B: Monthly distribution of estimated sea surface fCO_2 for the merged SOM in 2005.

SOM C: MLD, SST and SSS

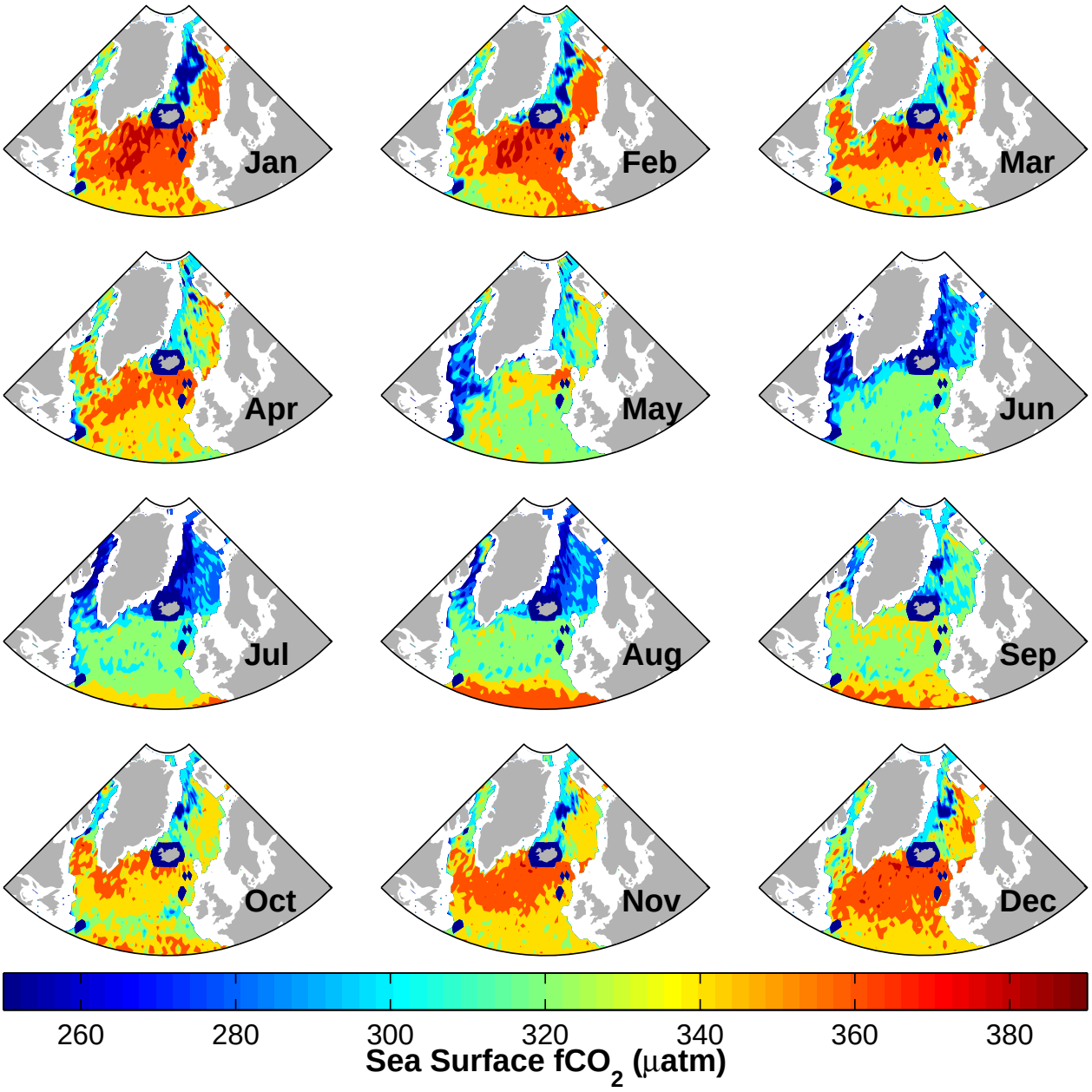


Figure 3.23: SOM C: Monthly distribution of estimated sea surface fCO_2 for the merged SOM in 2005.

SOM D: MLD, SST, SSS and CHL (summer only)

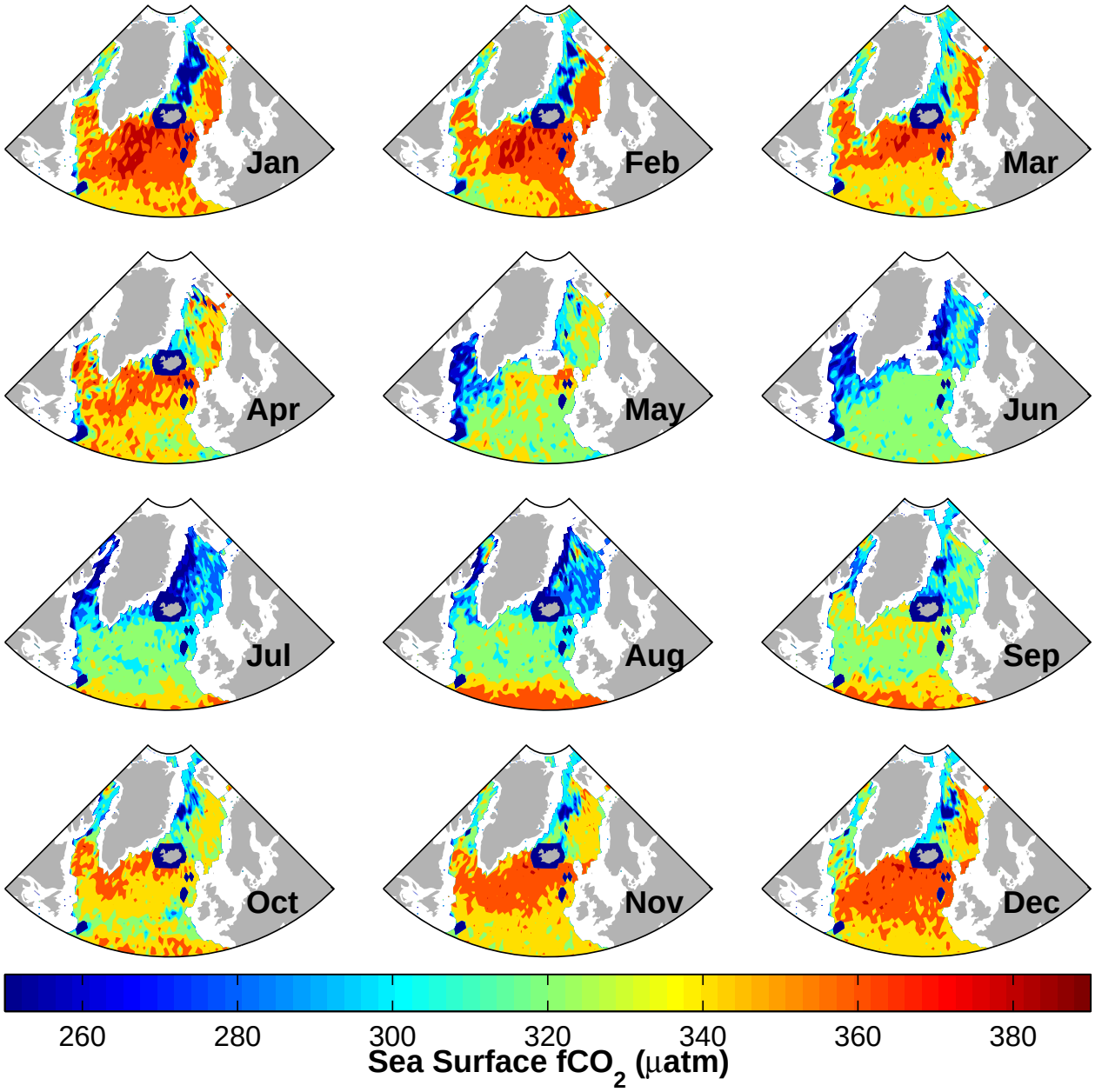


Figure 3.24: SOM D: Monthly distribution of estimated sea surface fCO_2 for the merged SOM in 2005.

3.5 Mapping SST by the use of Self Organizing Map

In order to verify whether the SOM technique is able to estimate realistic maps, a SST map was estimated and compared with the mapped seasonal cycle of temperature (Fig. 2.9) in chapter 4. SOM SST was estimated using the input parameters MLD and SSS, with the training arguments 20 for rough and 15 for fine-tuning, and with map size set to [100 65]. Skill assessment statistics are presented in Tab. 3.7.

Table 3.7: Skill assessment statistics for SOM SST

Map name	r^2	RMSE (µatm)	ME	CF
SOM SST	0.71	3.1	0.70	0.30

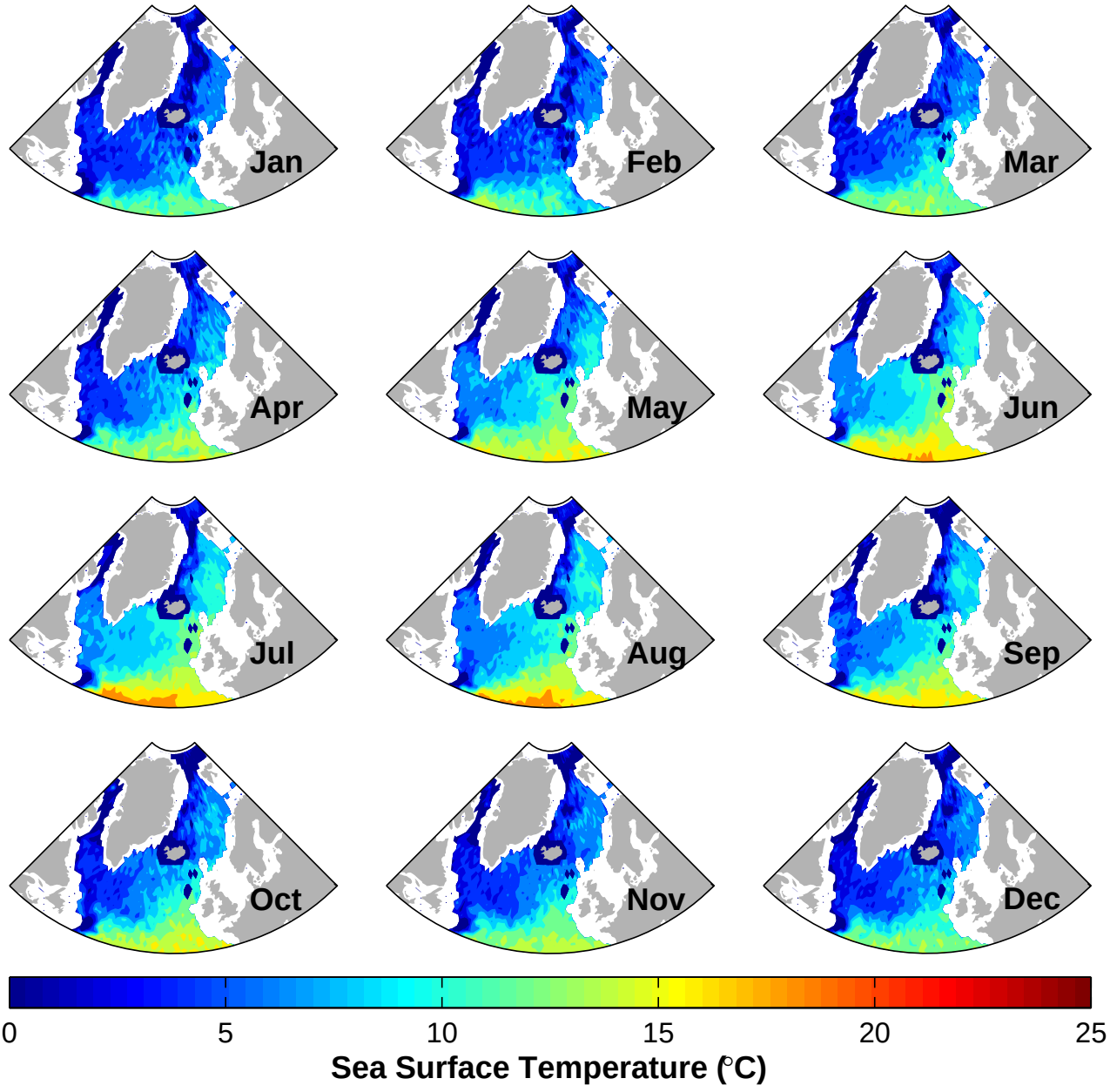


Figure 3.25: SOM SST: Monthly distribution of estimated sea surface temperature in 2005.

Chapter 4

Discussion

4.1 The Most Optimal Self Organizing Map

The skill assessment statistics for the different SOMs are presented in Tab. 4.1. From this table it can clearly be seen that SOM A for the Nordic Seas, with the input parameters MLD and SST, provides the most realistic estimates of fCO_2 distribution. The net fCO_2 flux for the Northern North Atlantic required for 0.1 Pg C yr^{-1} was estimated to 10.8 µatm in the LSCOP report (Sweeney et al., 2002). The uncertainty for SOM A in the Nordic Seas, 12.5 µatm, is close to this target, which proves that the SOM technique works well for the Nordic Seas. However, the technique is inadequate for the northern North Atlantic Ocean as the accuracy of the fCO_2 estimates here are far from the target.

Table 4.1: Skill assessment statistics for all self organizing maps

Basin	Map name	r^2	RMSE (µatm)	ME	CF
The northern	SOM A	0.63	22.6	0.61	0.39
North Atlantic	SOM B	0.59	23.3	0.57	0.43
Ocean and the	SOM C	0.50	27.1	0.44	0.56
Nordic Seas	SOM D	0.48	27.1	0.42	0.58
The northern	SOM A	0.61	23.1	0.59	0 41
North Atlantic	SOM B	0.57	24.2	0.55	0.45
Ocean	SOM C	0.52	25.8	0.49	0.51
	SOM D	0.51	26.1	0.47	0.53
	SOM A	**0.85**	**12.5**	**0.84**	**0.16**
The Nordic	SOM B	0.78	14.6	0.76	0.24
Seas	SOM C	0.70	17.9	0.68	0.32
	SOM D	0.72	16.4	0.70	0.30
	SOM A	0.68	20.9	0.67	0.33
Merged	SOM B	0.63	22.3	0.61	0.39
SOMs	SOM C	0.59	24.1	0.56	0.44
	SOM D	0.57	24.2	0.54	0.46

For all the basins, SOM A always estimates the most realistic fCO_2 maps. The addition of the CHL parameter (in SOM B) gives the second best estimates. SOM C has better skill assessment statistics than SOM D, except for the Nordic Seas. This may be explained by the fact that the biological production plays a key role in the seasonal cycle of fCO_2 in the Nordic Seas, and that an exclusion of the CHL parameter, as done in SOM C, leads to a weaker estimation of fCO_2.

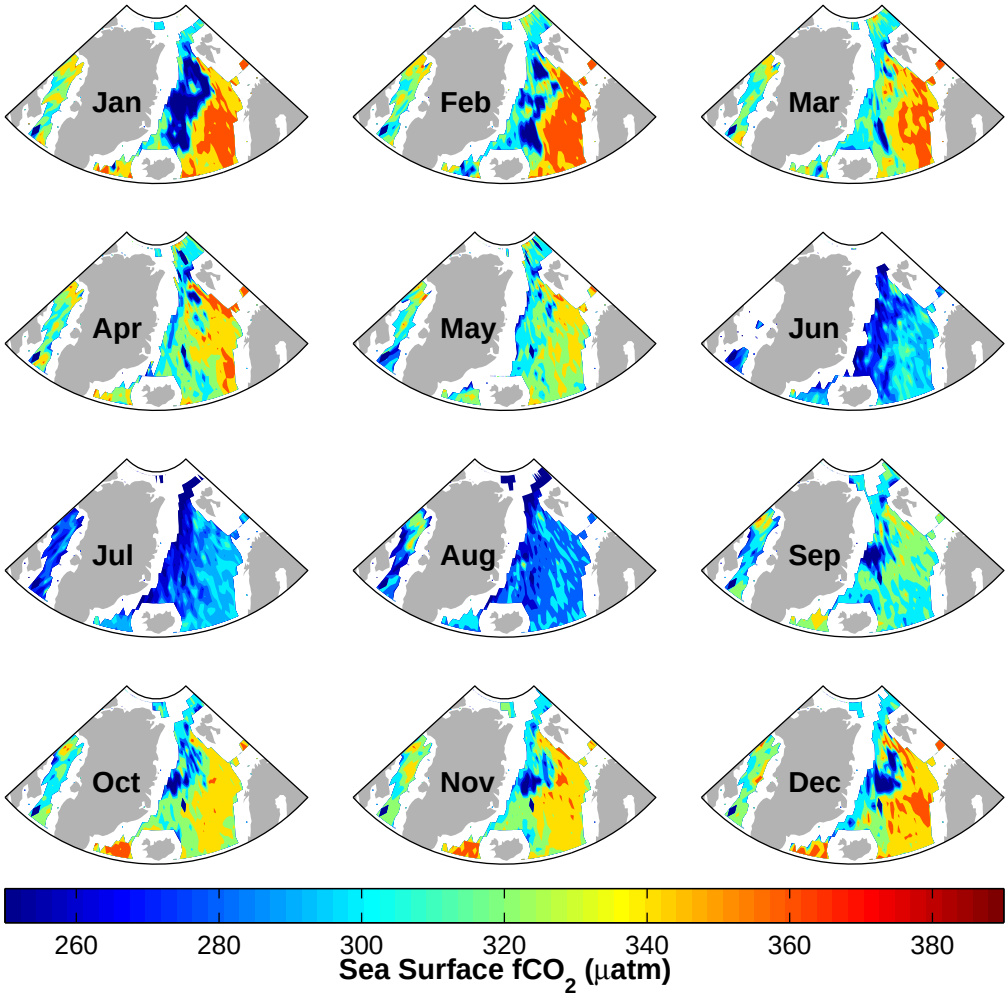


Figure 4.1: Monthly distribution of the most optimal sea surface fCO_2 map for 2005.

Combining the default MLD and SST parameters with additional parameters affected the skill assessment statistics as summarized in Tab. 4.2. Adding the CHL parameter to represent biological production, led to a 4-7% decrease in r^2. Adding the SSS parameter led to a 9-13% decrease, whereas adding both the CHL and SSS led to a 10-15% decrease. This is surprising, as one would expect that more input parameters would yield better statistics. From Tab. 4.2, one can conclude that the most optimal maps are made from the combination of the MLD and SST parameters, and that if any additional parameters is to be included, it should be CHL as this yields better statistics than SSS. This is probably because CHL has a larger impact on fCO_2 than SSS.

The is a distinct difference between winter and summer in the most optimal SOM (Fig. 4.1), an indication that the SOM estimates are able to solve the transition between physical and biological processes. The next section discusses the seasonal and interannual variations in the map.

Table 4.2: The effect of additions of CHL and SSS to the default input parameters SST and MLD, on the performance of the SOMs

Parameter	r^2	RMSE (µatm)	ME	CF
CHL	-(0.04-0.07)	+(0.7-1.4)	-(0.04-0.08)	+(0.04-0.08)
SSS	-(0.09-0.13)	+(2.7-5.4)	-(0.10-0.17)	+(0.10-0.17)
CHL and SSS	-(0.10-0.15)	+(3.0-4.5)	-(0.12-0.19)	+(0.12-0.19)

4.2 Seasonal and Interannual fCO_2 Variations in the Nordic Seas

In this section the seasonal, spacial and interannual fCO_2 variations in the Nordic Seas are analyzed in regard to SST (Fig. 4.4) and MLD (Fig. 4.5), as the most optimal SOM is for this region. The northern North Atlantic Ocean is therefore not discussed. The color scale for June to September in Fig. 4.5 is different from the remaining months, as the MLD values for these months are very low.

4.2.1 Seasonal

The three years in this study have similar seasonal cycles of fCO_2, therefore only year 2005 will be discussed.

There is a clear distinction between east and west in January (Fig. 4.3a), with very low fCO_2 values in the Greenland Sea, and high values in the Norwegian Sea. A comparison with SST (Fig. 4.4a), shows that fCO_2 follows the same pattern as SST, which explains the clear distinction between the Norwegian Sea and the Greenland Sea. Comparing annual fCO_2 with MLD shows that the lowest fCO_2 values occur in areas with deep MLD, while the highest values occur in areas with intermediate MLD (Fig. 4.5a). This is rather unexpected, as fCO_2 generally increases with increasing MLD. This trend is also seen in the rest of the figures, and may be an artifact caused by an anomalous data point in the training dataset. As waters with shallow MLD have intermediate fCO_2 values, and waters with intermediate MLD have high fCO_2 values, it seems illogical that waters with high MLD should have the lowest fCO_2 values. However, in the fCO_2 versus MLD plot (Fig. 3.14a), there is a single data point with MLD of approximately 600 m and fCO_2 of approximately 250 µatm. The temperature for this data point is approximately 3°C (Fig. 3.14b). It is not unlikely that this has defined a single neuron, which has been used to determine the fCO_2 in the Greenland Sea in winter, as the MLD and SST in this area typically has such values (Fig. 4.4 and 4.5).

The distinction between east and west is also present in February, however to a lesser extent. Low fCO_2 values (Fig. 4.3d) in the Greenland Sea are associated with deep MLD (Fig. 4.5d), and higher values in the Norwegian Sea is explained by the increased MLD from January to February.

March shows a gradual decrease in fCO_2 in both the Greenland Sea and the Norwegian Sea, compared to February (Fig. 4.3g). This is explained by the decreased MLD (Fig. 4.5g), which probably has caused some primary production to occur.

As the surface water temperature gradually increases during spring (Fig. 4.4j), the MLD decreases (Fig. 4.5j). The decrease in MLD in April increases primary production, causing fCO_2 values to further decrease (Fig. 4.3j). The Norwegian Sea still has higher fCO_2 values than the Greenland Sea, but the distinction between east and west has gradually decreased.

As the Nordic Seas continue to warm in May (Fig. 4.4m) and the MLD keeps decreasing (Fig. 4.5m), the biological production increases further (Fig. 2.12). Shallow MLD thus results in intermediate levels of fCO_2 (Fig. 4.3m). The distinction in the SST between the Norwegian Sea and the Greenland Sea becomes less clear, which is reflected in fCO_2. The relatively high fCO_2 values northeast of the Greenland Sea are due to intermediate MLD.

In June, July and August the fCO_2 distribution is more or less homogeneous, with low values in the Greenland Sea and slightly higher in the Norwegian Sea (Fig. 4.3p, 4.3s and 4.3v). The low fCO_2 values are caused by shallow MLD and resultant high biological productivity during summer.

The fCO_2 values in September gradually increases after being low during summer (Fig. 4.3y). The increase can be explained by the decrease in biological activity and the slight increase in MLD (Fig. 4.5y). This deep mixing is caused by intense wind mixing and decreased temperature (Fig. 4.4y). The distinction between the Norwegian Sea and the Greenland Sea becomes more clear, and is clearly visible in October.

In November, cooling water masses (Fig. 4.4ae) and deepening MLD (Fig. 4.5ae) causes the fCO_2 values in the Norwegian Sea and the Greenland Sea to increase and decrease, respectively (Fig. 4.3ae). The low fCO_2 values in response to deep MLD in the Greenland Sea are illogical as mentioned above,

and should be investigated further. The high values in the Norwegian Sea are explained by intermediate MLD.

During December the MLD continues to deepen (Fig. 4.5ah), while the SST gradually cools (Fig. 4.4ah). As a result the distinction in fCO_2 values increases further, and begins to resemble the patterns in January (Fig. 4.3ah). Low fCO_2 values are explained by deep MLD, while intermediate MLD are explained by intermediate values.

4.2.2 Interannual

Differences in physical and biological processes between years results in interannual variability of fCO_2. The interannual fCO_2 variability for 2005-2007 (Fig. 4.3) is analyzed and compared with the interannual SST (Fig. 4.4) and MLD variations (Fig. 4.5).

January 2005 has a much clearer distinction in the fCO_2 distribution between the Greenland Sea and the Norwegian Sea than the other years (Fig. 4.3a). This distinction is also visible in January 2006 (Fig. 4.3b), but to a lesser extent. January 2007 (Fig. 4.3c) has a distinction that lies in between the two previous years. Comparing annual fCO_2 with MLD shows that 2005 has much larger areas with deep MLD than 2006 and 2007, and therefore the lowest fCO_2 values (Fig. 4.5a-4.5c). This may be an artifact due to the abnormal data point as described above. The problem occurs in all of the months, and is not further commented. All three years have different extent of areas with intermediate MLD in the Norwegian Sea, which results in different fCO_2 values.

In February 2005 and 2007 the fCO_2 values in the Norwegian Sea are slightly higher than in 2006 (Fig. 4.3d-4.3f). This is explained by the larger MLD in 2005 and 2007 (Fig. 4.5d-4.5f). The low values in the Greenland Sea for all years correspond to the greatest MLD values.

In March, the greatest fCO_2 difference between the years occurs in the Greenland Sea (Fig. 4.3g-4.3i). Comparison of the values shows that 2005 has the highest values and that 2007 has the lowest. The areas with low values in 2006 and 2007 are the result of deep MLD (Fig. 4.5g-4.5i). Intermediate MLD is found in the Norwegian Sea at slightly deeper depths in 2006 and 2007 than in 2005. This yields higher fCO_2 values for 2006 and 2007.

April is a month with strong interannual variations. 2005 shows more or less intermediate fCO_2 values in the Nordic Seas (Fig. 4.3j), while 2006 shows low and intermediate values in the Greenland Sea and the Norwegian Sea, respectively (Fig. 4.3k). 2007 shows very low and high values in the Greenland Sea and the Norwegian Sea, respectively (Fig. 4.3l). 2005 has smaller areas with deep MLD (Fig. 4.5j), but still higher fCO_2 levels this year compared to 2006 and 2007. 2006 has a distinctive area in the Greenland Sea with deep MLD, in contrast to the rest of the basin which has shallow MLD (Fig. 4.5k). For 2007, the deep MLD is found in the same area as 2006 (Fig. 4.5l). The intermediate MLD in the Norwegian Sea corresponds to the high fCO_2 values. By comparing the fCO_2 for 2006 and 2007 with MLD, it is evident that intermediate MLD in the Norwegian Sea gives high fCO_2 values, as April 2006 lacks both intermediate MLD and high fCO_2 values.

May 2005 has a different fCO_2 pattern than 2006 and 2007 (Fig. 4.3m-4.3o). The higher fCO_2 values this year in the Norwegian Sea corresponds to the deeper MLD (Fig. 4.5m-4.5o). The MLD decreases in 2006 and 2007, which leads to lower fCO_2 values. The interannual variation in May is probably the result of variations in primary production, and it is quite intriguing to examine the SeaWiFS CHL maps in May, which shows larger CHL concentrations in 2006 and 2007 compared to 2005 (Fig. 4.2a-4.2c). These data have not been used in this SOM, which indicates strongly that the MLD dataset from Mercator is quite robust in this month.

June has small interannual variations (Fig. 4.3p-4.3r). 2005 has slightly higher fCO_2 values in the Iceland Sea, while 2007 has slightly higher values in the Norwegian Sea. Both are explained by the deeper MLD in these areas these years (Fig. 4.5p-4.5r).

In July the interannual differences are almost negligible (Fig.4.3s-4.3u). However, the biological distribution for this month do not correspond to the fCO_2 distribution (Fig. 4.2d-4.2f). In 2005 and 2006 the biological productivity is larger than in 2007, which is not reflected in the fCO_2 distribution. The Mercator MLD estimates are clearly not always a robust proxy for the biological processes. High

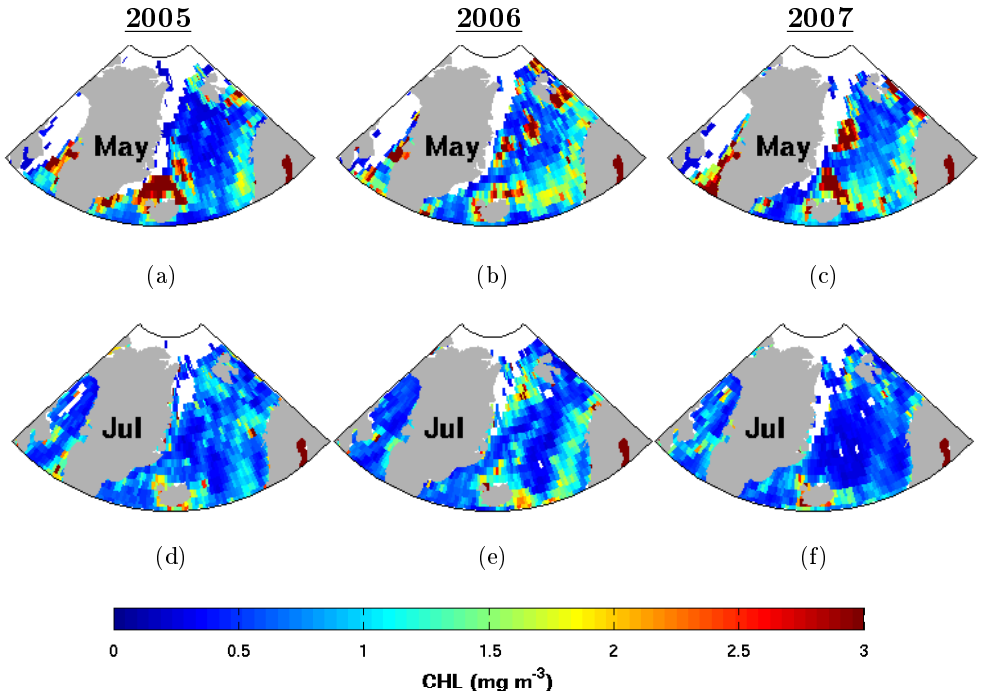


Figure 4.2: Interannual differences in the CHL (mg m^{-3}) distribution in May (a-c) and June (d-f) in 2005-2007.

fCO_2 values are found southwest of Svalbard, which can be explained by deeper MLD (Fig. 4.5s-4.5u). The indistinct fCO_2 distribution between the Greenland Sea and the Norwegian Sea, is caused by the warming waters.

August 2007 has high values of fCO_2 (Fig. 4.3x) in the northern Greenland Sea as a result of deep MLD (Fig. 4.5x), otherwise, there are little interannual variations for this month.

The interannual variation of fCO_2 in September is small, but visible (Fig. 4.3y-4.3aa). The high fCO_2 values in the eastern Greenland Sea for 2006 and 2007, are due to an increase in MLD (Fig. 4.5y-4.5aa). 2005 has more shallow MLD, and therefore lower fCO_2 values than the other years.

There is a small interannual variation in October (Fig. 4.3ab-4.3ad), with the largest difference in 2005 in the Irminger Sea. This area has high fCO_2 values compared to the other years, which is caused by deep MLD (Fig. 4.5ab). A small variation can also be seen east in the Iceland Sea, where 2006 has lower fCO_2 values than 2005 and 2007. The MLD in 2006 for this area is slightly shallower than the rest (Fig. 4.5ac), but not shallow enough to explain the large difference in fCO_2. As there is no CHL data for this time of the year, it is difficult to say whether this is caused by biological production or not.

November 2006 has higher fCO_2 values than 2005 and 2007 in the north Greenland Sea (Fig. 4.3ae-4.3ag), which is an area with deep MLD (Fig.4.5ae-4.5ag). The Irminger Sea has more shallow MLD in 2007, and therefore lower fCO_2 values.

December 2006 has more shallow MLD in the southern Norwegian Sea compared to 2005 and 2007 (Fig. 4.5ah-4.5aj), which is reflected in the lower fCO_2 values for this year (Fig. 4.3ah-4.3aj). The deep MLD in the north Greenland Sea in 2006 explains the high fCO_2 values in this area.

Overall, the results in the Greenland Sea indicate that MLD control interannual fCO_2 variations in winter, where deep MLD leads to low fCO_2. In summer on the other hand, the biological production controls interannual fCO_2 variations. In the Norwegian Sea, MLD controls interannual fCO_2 variations in winter, where deep MLD leads to high fCO_2. This is known as the thermodynamic effect. In summer, the interannual fCO_2 variations in the Norwegian Sea appears to be controlled by biological production.

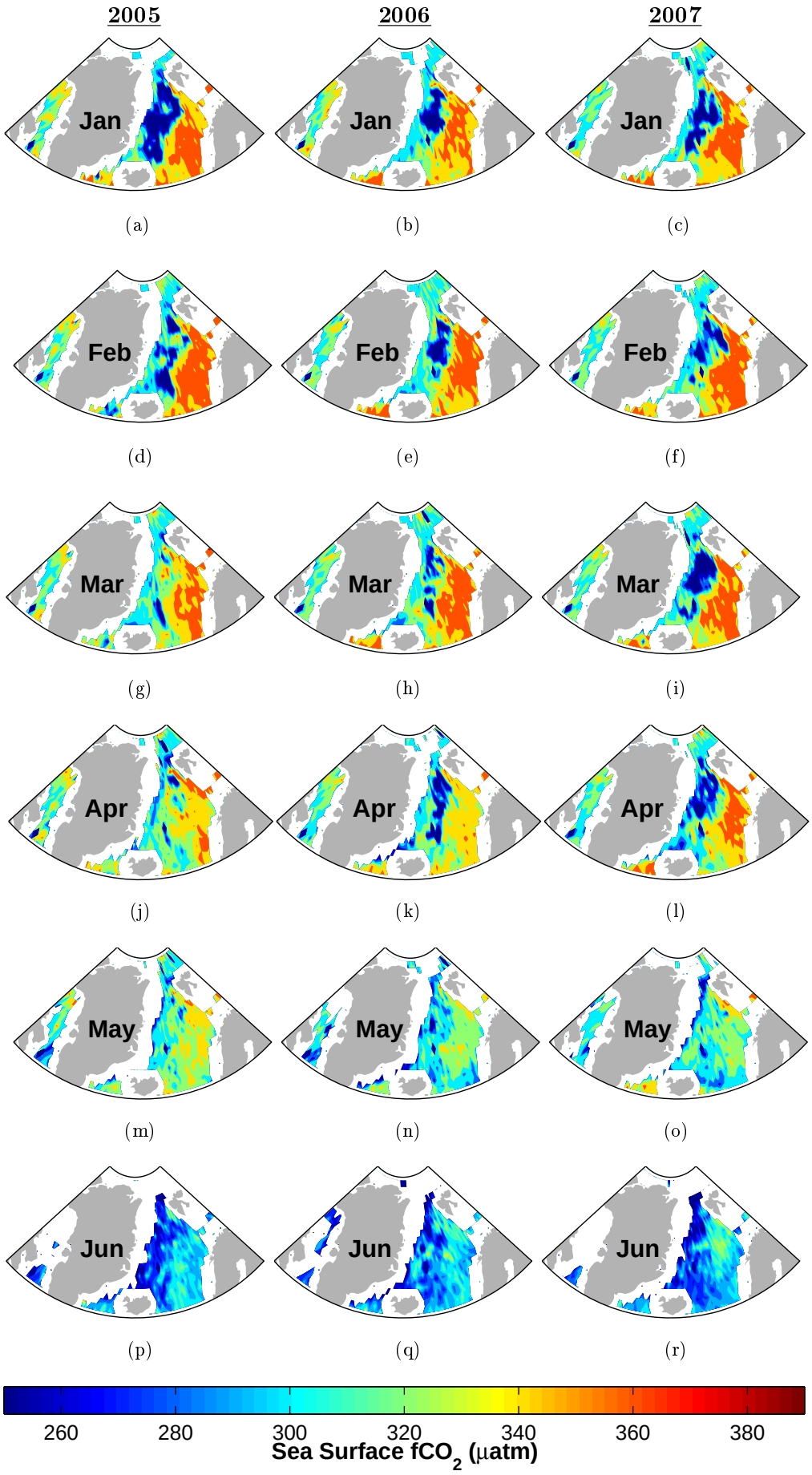


Figure 4.3: Continued on the next page.

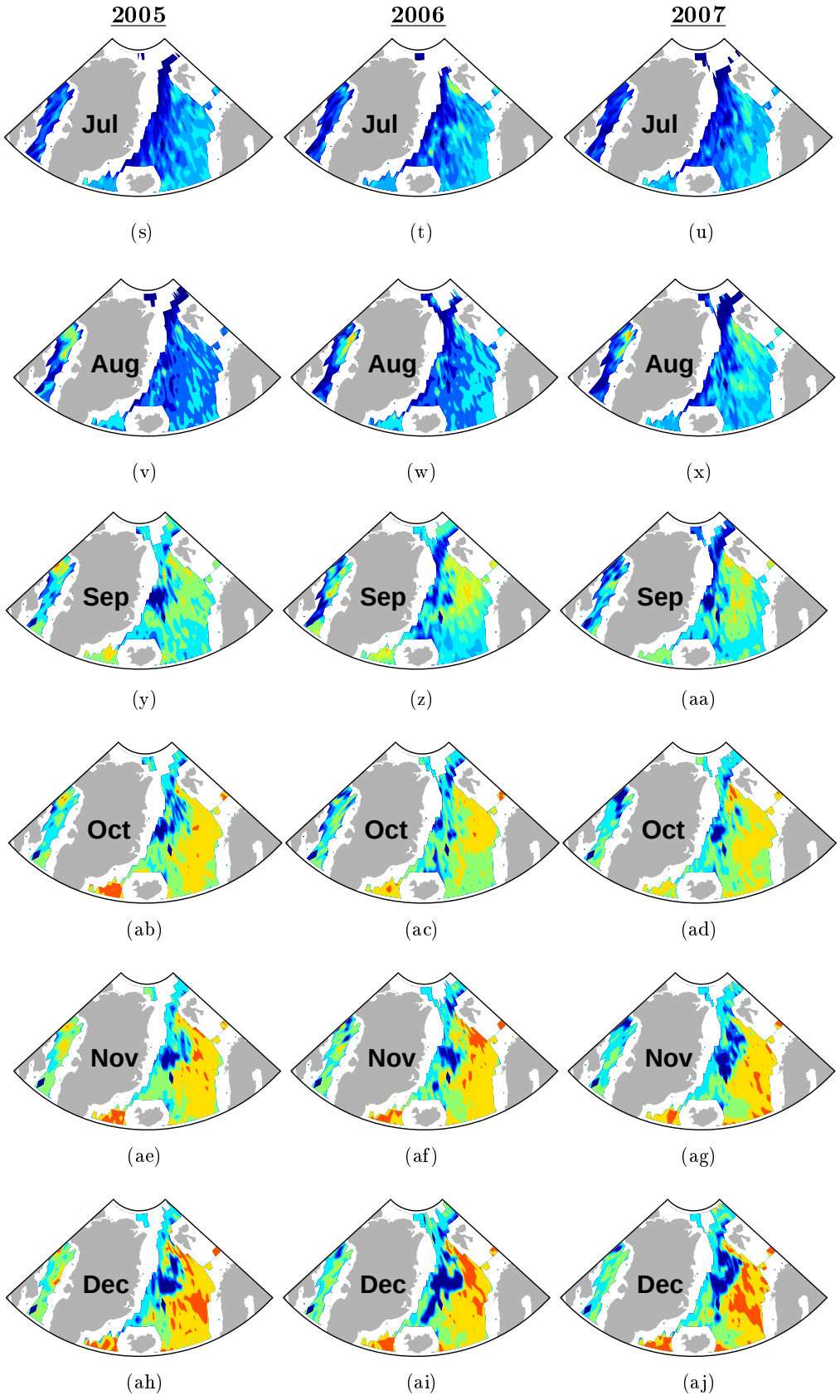


Figure 4.3: Interannual differences in the distribution of sea surface fCO_2 (µatm) from January-December in 2005-2007.

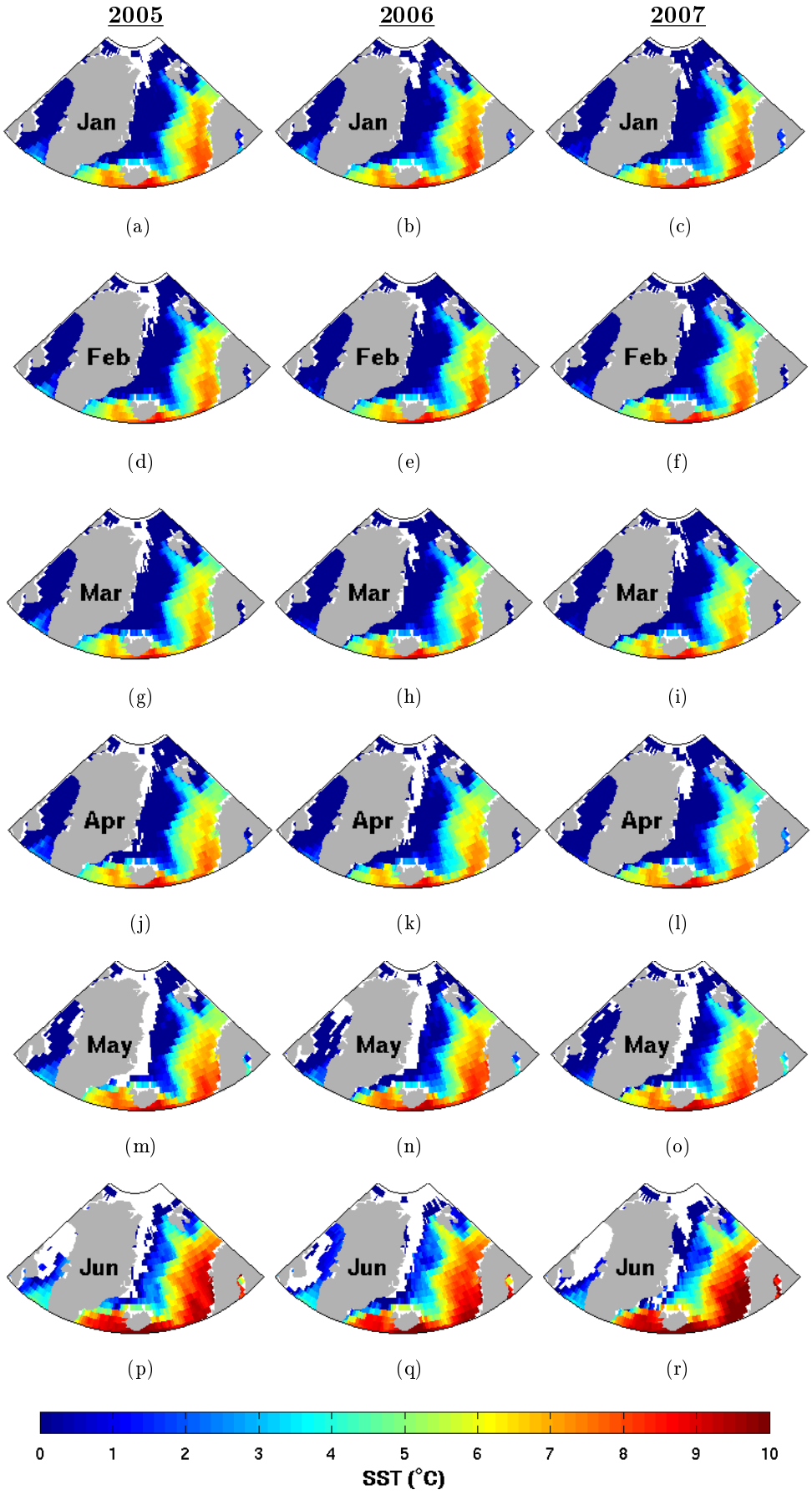


Figure 4.4: Continued on the next page.

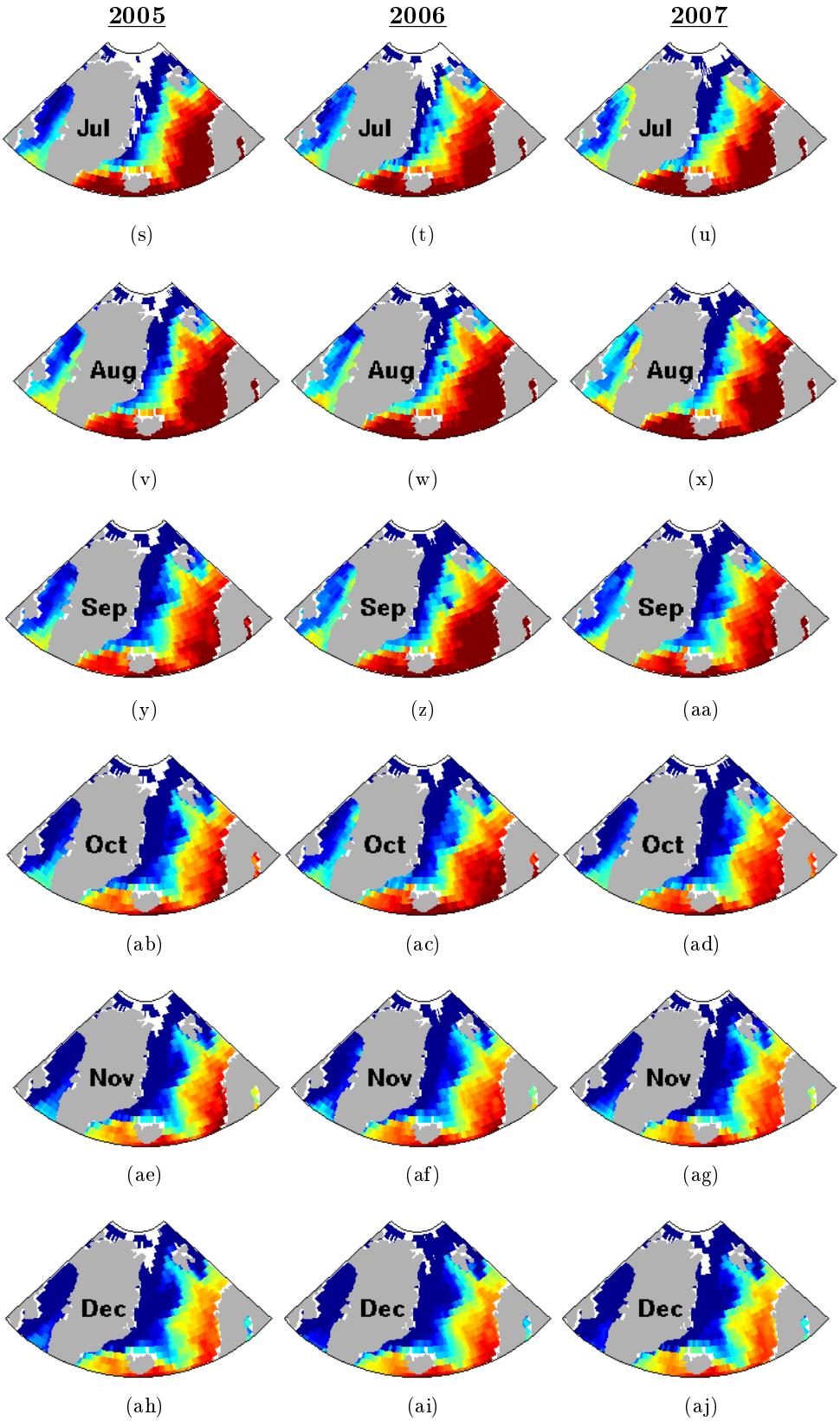


Figure 4.4: Interannual differences in the distribution of SST (°C) from January-December in 2005-2007.

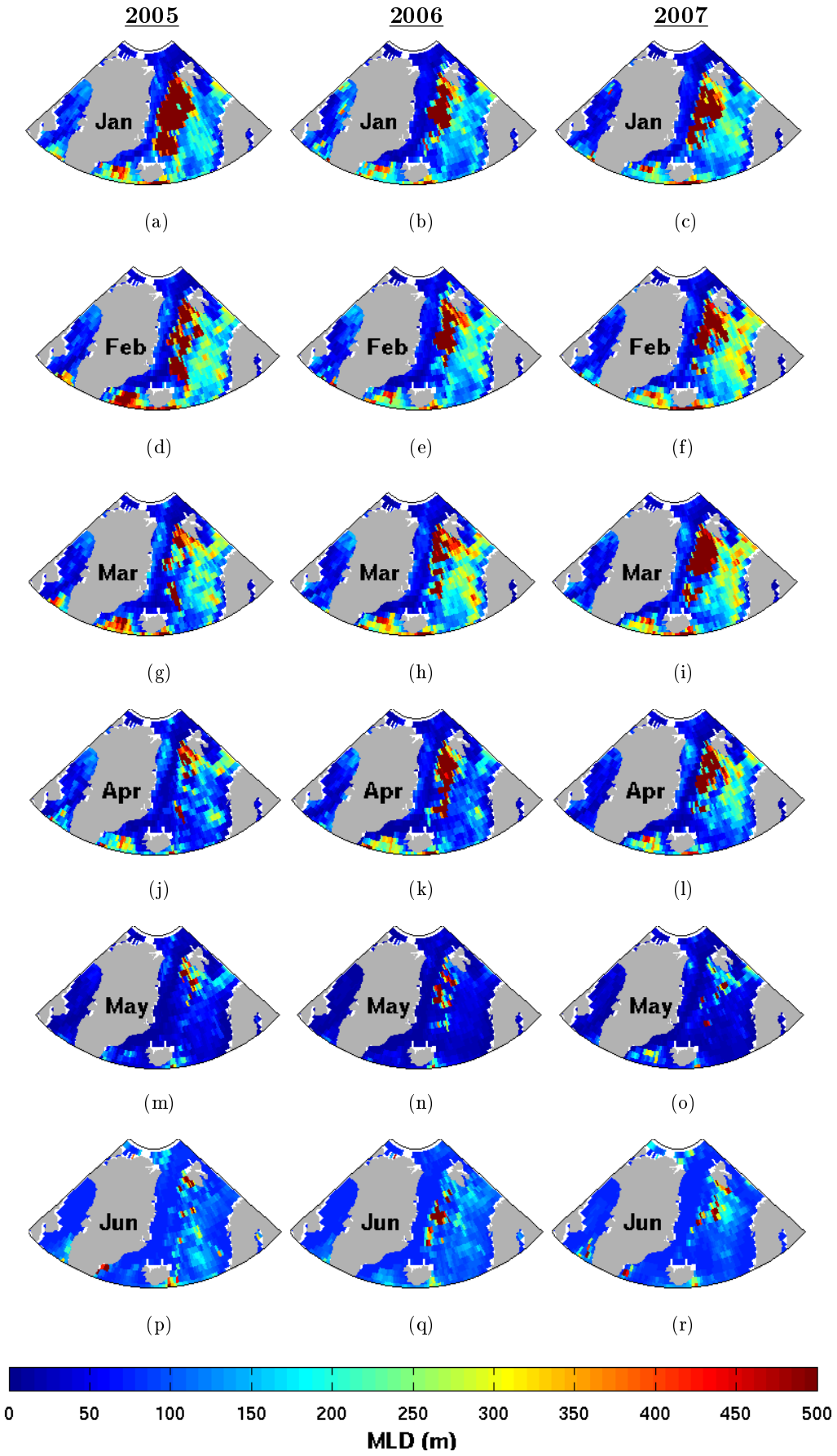


Figure 4.5: Continued on the next page.

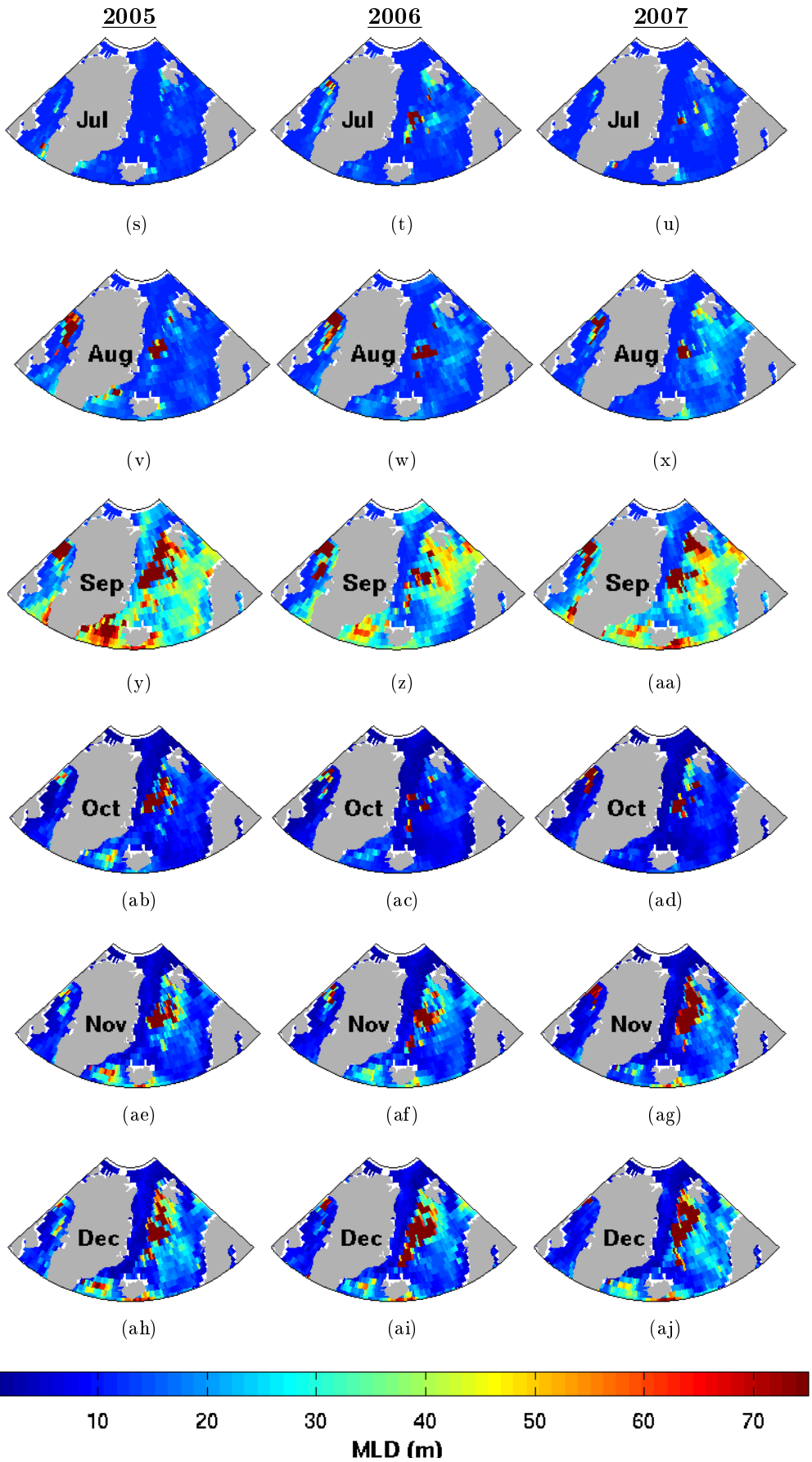


Figure 4.5: Interannual differences in the distribution of MLD (m) from January-December in 2005-2007. The color scale for June to September ranges from 0-75 m, while October to May ranges from 0-500 m.

4.3 Normal Versus Merged SOMs for the northern North Atlantic Ocean and the Nordic Seas

The normal maps were produced by mapping all data for the northern North Atlantic Ocean and the Nordic Seas simultaneously. In the merged maps, each basin was trained and mapped separately, and then merged together. Skill assessment statistics for the two approaches are presented in Tab. 4.3. By comparing the statistics, it is evident that the merged maps are slightly better for all SOMs, with a 4-9% higher correlation. The RMSEs for the merged maps are typically 1-3 µatm lower than for the normal maps, whereas the MEs are 4-12% higher. The cost function is also lower for all merged SOMs.

Table 4.3: Skill assessment statistics for the normal and merged SOMs

Map name	r^2	RMSE (µatm)	ME	CF
Normal SOM A	0.63	22.6	0.61	0.39
Merged SOM A	0.68	20.9	0.67	0.33
Normal SOM B	0.59	23.3	0.57	0.43
Merged SOM B	0.63	22.3	0.61	0.39
Normal SOM C	0.50	27.1	0.44	0.56
Merged SOM C	0.59	24.1	0.56	0.44
Normal SOM D	0.48	27.1	0.42	0.58
Merged SOM D	0.57	24.2	0.54	0.46

Only the self organizing map with the best statistics, SOM A, is presented and discussed (Fig. 4.6). A comparison of the merged SOM A with the normal SOM A, showed that they have more or less similar patterns. fCO_2 values typically peaks in January and gradually decrease through spring. The lowest values are found in July, and they start to increase during fall. There is a slight overestimation of fCO_2 in the Nordic Seas, and a slight underestimation in the northern North Atlantic Ocean for the normal SOM, compared to the merged SOM. This can clearly be seen in January, where the normal SOM (Fig. 4.6a) has much higher fCO_2 values along the Greenland coast than the merged SOM (Fig. 4.6b). This over- and underestimation may be explained by the fact that the northern North Atlantic Ocean contains more data than the Nordic Seas, therefore the physical and biological processes for the northern North Atlantic Ocean dominate the normal SOM. Another explanation may be pattern smoothing of the regional characteristic during the training process. The normal SOM is trained with rough and fine-tuning parameters of 20 and 15, respectively, whereas the merged SOM is trained with rough and fine-tuning parameters of 25 and 20 for the Nordic Seas, and with 15 and 10 for the northern North Atlantic Ocean. This may have led to an overtraining of the characteristics pattern in the Greenland Sea, but was done nonetheless as these settings reflected the most realistic fCO_2 values. The same assumptions can be made for the difference in the Nordic Seas at the eastern side of the basin in July (Fig. 4.6f), where the fCO_2 values for the merged map are slightly lower than for the normal map (Fig. 4.6e). From Tab. 4.3 and Fig. 4.6 one can conclude that the merged SOM produces better fCO_2 maps than the normal SOM.

A problem associated with merging two basins, is discontinuity along the border where they meet (at 63°N). This occurs when the two separate estimates do not match perfectly, and a small discontinuity can be seen west of Iceland in March (Fig. 4.6d) and July (Fig. 4.6h). For the rest of the months, discontinuity is more or less absent, proving that merged SOMs can successfully be applied.

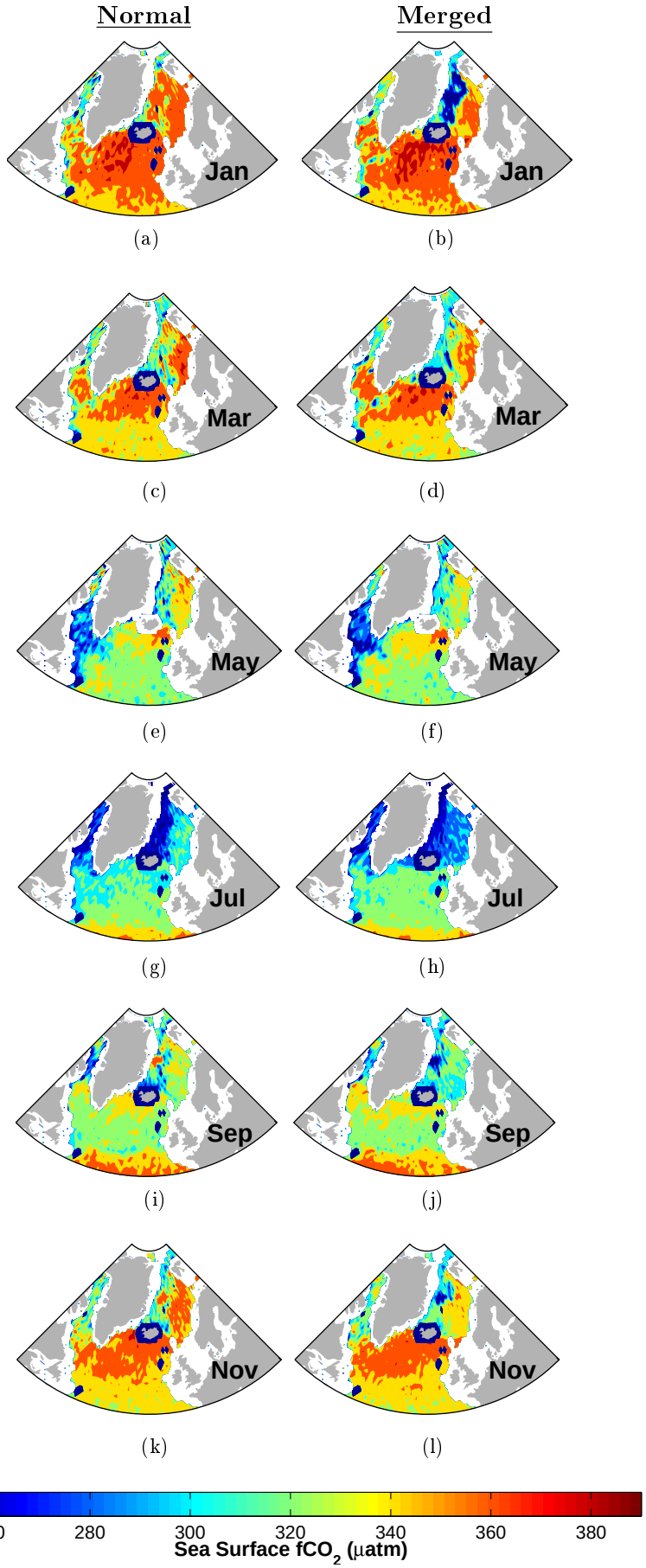


Figure 4.6: Comparison of fCO_2 estimates in the Northern North Atlantic and the Nordic Seas for SOM A 'normal' (left column) and SOM A 'merged' (right column).

4.4 Comparison with Earlier Self Organizing Maps

In this section the most optimal pCO_2 map from Telszewski et al. (2009)'s study of the North Atlantic Ocean between 10.5°N to 75.5°N, SOM Main, is compared with the most optimal fCO_2 map for the Nordic Seas from this study, SOM A (Fig. 4.7). Telszewski's SOM Main was trained with the input parameters MLD, SST and CHL, as they were found to be the most optimal, whereas in this study MLD and SST proved to be the most optimal for the region he used. Tab. 4.4 presents the skill assessment statistics for SOM A and SOM Main. The r^2 for SOM Main was 0.93 compared to 0.85 for SOM A, whereas the RMSEs were 11.6 µatm and 12.5 µatm, respectively. Telszewski's better statistical results may be explained by the different study areas. His study incorporated larger areas of the North Atlantic Subtropical gyre, where fCO_2 tends to be more homogeneous than in the northern areas that I focus on. In fact this oligotrophic area is mainly driven by the thermodynamic effect and differs from the Nordic Seas where biological production has a major impact on fCO_2 during summer (Takahashi et al., 2002).

Table 4.4: Skill assessment statistics for SOM A and SOM Main

Map name	Input Parameters	r^2	RMSE (µatm)
SOM A	MLD/SST	0.85	12.5
SOM Main	MLD/SST/CHL	0.93	11.6

Fig 4.7 compares, on a month by month basis, mine and Teszewski's fCO_2 maps for 2005. Since Telszewski's estimates are done more southward than this study, this discussion is limited to the region north of 63°N. The color scale for SOM A was changed in order to match the scale used for SOM Main, which is 280-440 µatm. The CHL parameter in SOM Main makes a comparison with SOM A impossible for January, February, October, November and December, as these months in SOM Main do not contain any data at higher latitudes. SOM Main has higher fCO_2 values in March than SOM A. There is a clear distinction between the eastern and western side of the basin in SOM A, with high and low fCO_2 values respectively. This feature is rather indistinct in SOM Main. For April, the SOMs are slightly more similar, and SOM Main is now able to display the distinction between eastern and western side of the Nordic Seas. May seems to be the best match between the SOMs, with similar fCO_2 values and patterns. Low values are found along the Greenland coast, whereas slightly higher values are found on the western side of the basin. June, July and August show more or less similar patterns, with fCO_2 values in SOM Main being slightly higher than in SOM A. The comparison of September reveals large differences between the SOMs. SOM Main has a clear distinction between east and west, with moderate levels of fCO_2 in east and low levels in west. SOM A has low fCO_2 levels with an approximately homogeneous distribution, in contrast to SOM Main.

From Fig. 4.7, it seems like SOM A is slightly underestimated with respect to fCO_2, compared to SOM Main. This may have been caused by the defined color, or the fact that visualization of the maps is done with two different Matlab functions: "m_contourf' for SOM A and "m_pcolor" for SOM Main. Given that SOM Main has been trained over a larger area, the overestimated fCO_2 values in the Nordic Seas may have been slightly influenced during the training. From Fig. 4.7, it is evident that SOM A works better than SOM Main in the Nordic Seas.

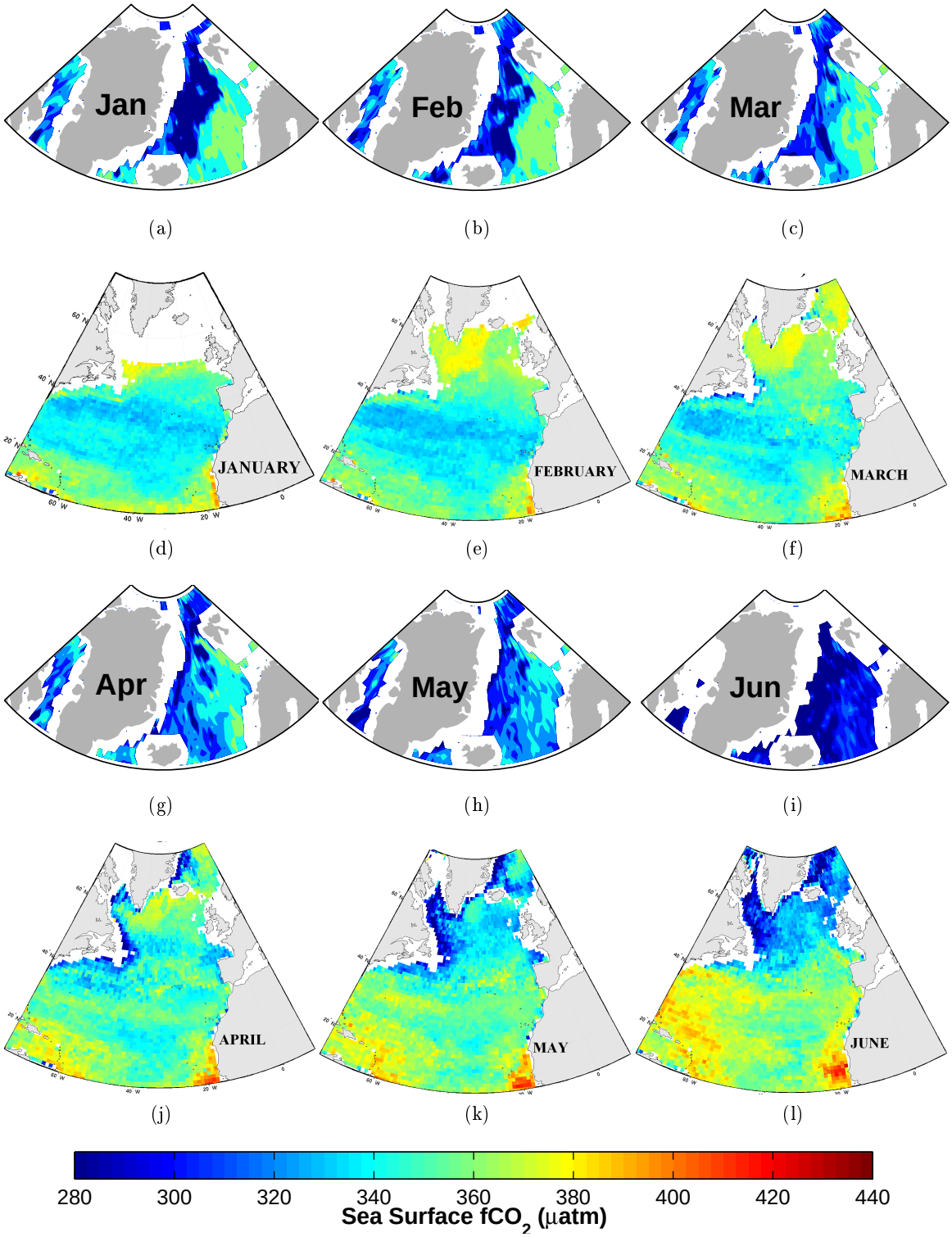


Figure 4.7: Continued on the next page.

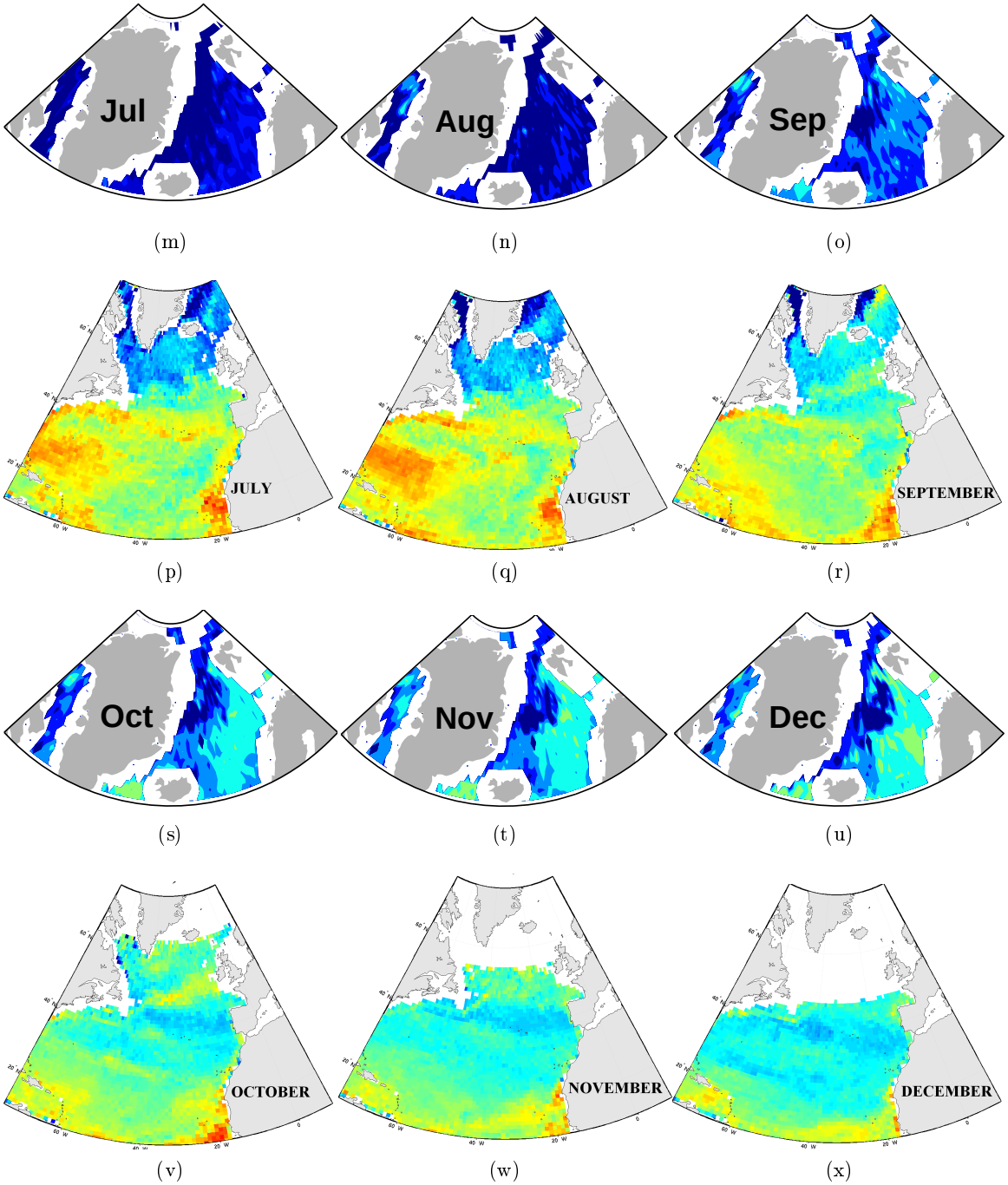


Figure 4.7: Comparison of SOM A for the Nordic Seas with Telszewski's SOM Main. Colors in SOM A have been changed to match the colors in SOM Main. Figures (d)-(f), (j)-(l), (q)-(s) and (w)-(y) from Telszewski (2009)

Chapter 5

Summary and Conclusion

The main objectives of this thesis were to see whether self organizing maps were able to produce seamless mapping of surface fCO_2 in the Nordic Seas and the northern North Atlantic Ocean, and to investigate which parameters were required in order to estimate the most optimal fCO_2 maps.

The SOMs were estimated for the northern North Atlantic Ocean including the Nordic Seas (44-85°N), the northern North Atlantic (44-63°N) excluding the Nordic Seas, and the Nordic Seas only (63-85°N). This was done in order to see whether mapping two independent basins worked better than one large.

Four maps were produced for each region, SOM A-D, using different parameters. SOM A was estimated with MLD and SST, SOM B with MLD, SST and CHL, SOM C with MLD, SST and SSS, and SOM D with MLD, SST, CHL and SSS.

Surprisingly, SOM A (MLD and SST) always estimated the most realistic fCO_2 map for all the basins. Adding the CHL parameter (in SOM B) provided the second best estimate. Of the two poorest performing SOMs, SOM C was usually better than SOM D, except for the Nordic Seas. The fact that adding additional parameters to the default MLD and SST, actually lead to poorer skill assessment statistics was unexpected, as one would expect that more input parameters would improve the statistics.

SOM A for the Nordic Seas provided the most realistic estimates of fCO_2 distribution. The uncertainty for this SOM, 12.5 µatm, is close to the LSCOP target of 10.8 µatm for estimation of net fCO_2 flux for the Northern North Atlantic (Sweeney et al., 2002).

The fCO_2 maps that were produced showed a realistic seasonal cycle. The Norwegian Sea has high fCO_2 values of typically 360 µatm during winter. The values gradually decrease throughout spring, until the lowest values, typically 290 µatm, are reached in summer. This is caused by primary production. During fall, primary production decreases and deep mixing causes fCO_2 to increase, and the $fCO2$ values continues to increase during the winter. The Greenland Sea has low $fCO2$ values during winter, typically 250 µatm. As spring approaches the values gradually increase to 320 µatm, before they decrease again in summer, typically to 280 µatm. In fall, the fCO_2 values decrease and continues to decrease throughout winter. Winter in the northern North Atlantic Ocean has high fCO_2 values of typically 360 µatm. These values gradually decrease as the biological production increases throughout the spring, and are at their lowest, approximately 320 µatm, in summer. A decrease in biological activity and an increase in MLD causes the fCO_2 values to increase for the remainder of the year. The Irminger Sea follows the same trend as the Norwegian Sea and the northern North Atlantic Ocean, with high fCO_2 values in winter and low values in summer.

The SOM A fCO_2 maps for the Nordic Seas were scrutinized for interannual variability. The most substantial interannual variations occurred in the central areas during winter and were tied to the MLD. In May the interannual variations in fCO_2 were described by the MLD, and this was confirmed by the CHL distribution. However, in July the homogeneous fCO_2 distribution was not in accordance with the interannual differences in CHL. Here the MLD did not seem to accurately represent the biological

production.

In all the SOMs, comparison of fCO_2 with MLD in the Greenland Sea showed that the lowest fCO_2 values occurred at deep MLD, while the highest values occurred at intermediate MLD. This was rather unexpected, as fCO_2 generally increases with increasing MLD. As waters with shallow MLD have intermediate fCO_2 values, and waters with intermediate MLD have high fCO_2 values, it seems illogical that waters with high MLD should have the lowest fCO_2 values. A theory is that a neuron has been overexpressed during training, resulting in an artifact in the SOMs. This seems to be a weakness with SOM, if one do not have typical values for all regions in the study, one might end up with unrealistic results.

The SOMs for the Nordic Seas were also merged with the SOMs for the northern North Atlantic Ocean, to see whether this produced more realistic results than the SOMs for the northern North Atlantic Ocean and the Nordic Seas (called normal SOMs). The merged SOMs showed more or less the same patterns as the normal SOMs. For the normal SOM, there was a slight overestimation of fCO_2 in the Nordic Seas, and a slight underestimation in the northern North Atlantic Ocean, compared to the merged SOM. The over- and underestimation may be explained by the fact that the northern North Atlantic Ocean contains more data than the Nordic Seas, therefore the physical and biological processes for the northern North Atlantic Ocean dominate the normal SOM. Discontinuity along the border is a problem that can occur when merging the fCO_2 maps from two basins. Here only a small discontinuity was seen west of Iceland in March and July. For the rest of the months, discontinuity was more or less absent, proving that merged SOMs can successfully be applied.

The most optimal fCO_2 map for the Nordic Seas, SOM A, was compared with the most optimal pCO_2 map from Telszewski et al. (2009)'s study. SOM A had better results for the Nordic Seas, because Telszewski's SOM was trained with the CHL parameter. As there is no CHL data available during winter in the Nordic Seas, his most optimal SOM was unable to reproduce the fCO_2 distribution at higher latitudes. Telszewski's statistical results were better than the statistics for the most optimal SOM from this study, which may be explained by the fact that his study incorporated larger areas of the North Atlantic Subtropical gyre, where fCO_2 tends to be more homogeneous than in the Nordic Seas.

The most optimal SOM had good statistics and performed well in the Nordic Seas, it was able to reproduce seasonal variations and a realistic fCO_2 distribution, but appears unable to reproduce correct fCO_2 values for the deepest mixed layer depths in the Greenland Sea. SOM performed well in the Nordic Seas, but not as well as in the northern North Atlantic Ocean, it is therefore recommended to use merged SOM when mapping both regions together. There is a need for more data from the winter season in the Nordic Seas, especially in central areas, as SOM seems to produce unrealistic fCO_2 values here.

Chapter 6

Further Work

The SOMs for the Nordic Seas reproduced deep mixed layer depths with low fCO_2 values in the Greenland Sea, which is counterintuitive. The training dataset should therefore be scrutinized for possible outliers, and more data should be obtained from the Greenland Sea in winter, in order to resolve the fCO_2 distribution and its controlling mechanism in this area during this season. The training dataset should also be trained with MLD from alternative ocean reanalysis products such as the Hycom (www.hycom.org/reanalysis) and TOPAZ (topaz.nersc.no), to see whether this yields better results than MLD from Mercator. To determine whether sea surface fCO_2 is increasing with the increasing atmospheric CO_2, the SOMs should be trained yearly, instead of simultaneously. Adding additional parameters to the default parameters gave poorer results than just using the default parameters, which was unexpected. The causes for this behavior should also be investigated further.

Appendix A

Self Organizing Map Quality Parameters

Table A.1: Quality parameters for the SOMs: quantization error (qe) and topographic error (te)

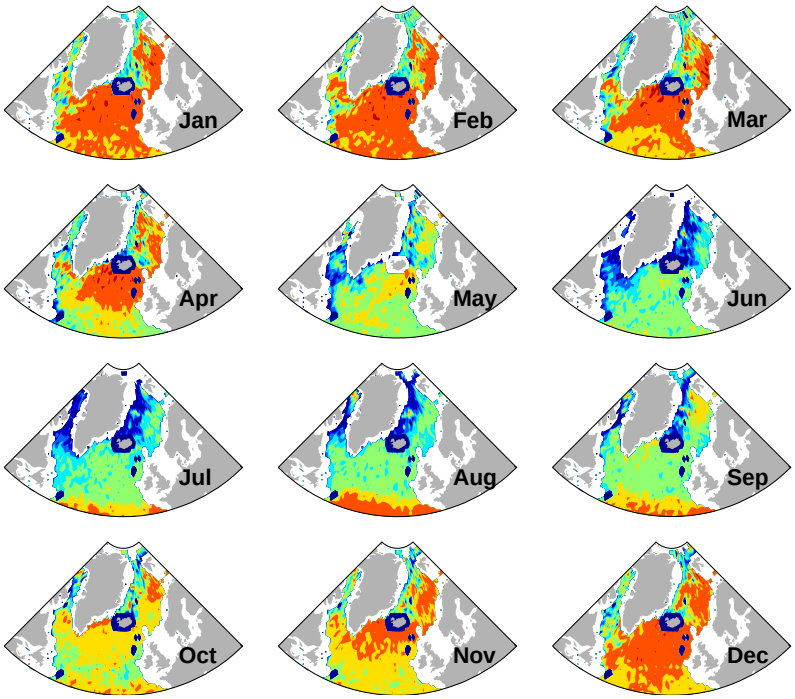
Map name & Parameters	The northern North Atlantic Ocean and the Nordic Seas		The northern North Atlantic Ocean		The Nordic Seas	
	Winter	**Summer**	**Winter**	**Summer**	**Winter**	**Summer**
SOM A: 'qe'	0.019		0.024		0.013	
SOM A: 'te'	0.058		0.015		0.089	
SOM B: 'qe'	0.019	0.132	0.024	0.017	0.013	0.109
SOM B: 'te'	0.058	0.043	0.015	0.044	0.089	0.036
SOM C: 'qe'	0.087		0.095		0.065	
SOM C: 'te'	0.046		0.041		0.039	
SOM D: 'qe'	0.087	0.284	0.095	0.222	0.065	0.195
SOM D: 'te'	0.046	0.055	0.041	0.050	0.039	0.048

Appendix B

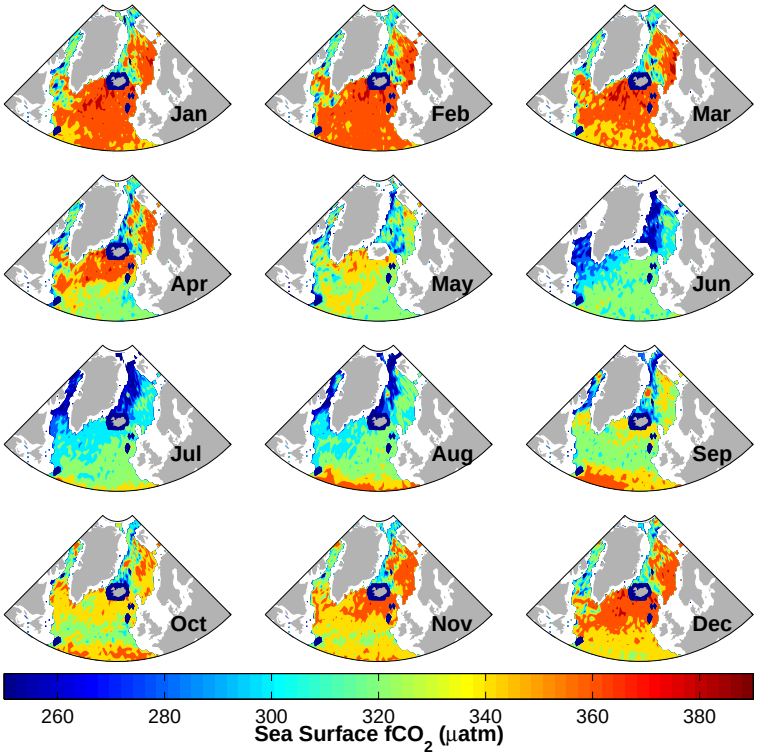
Supplementary Self Organizing Maps

The northern North Atlantic Ocean and the Nordic Seas [44-85°N]

SOM A: MLD and SST



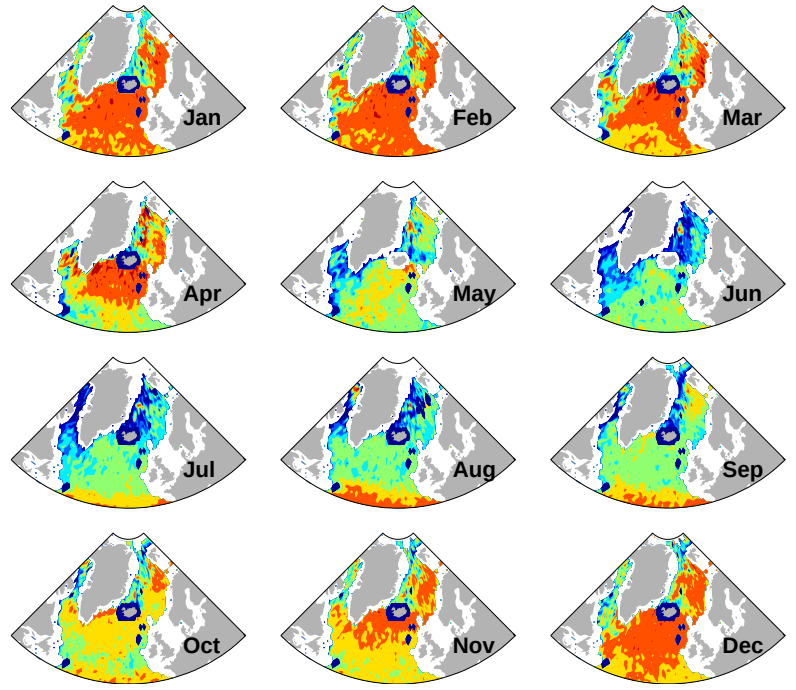
(a)



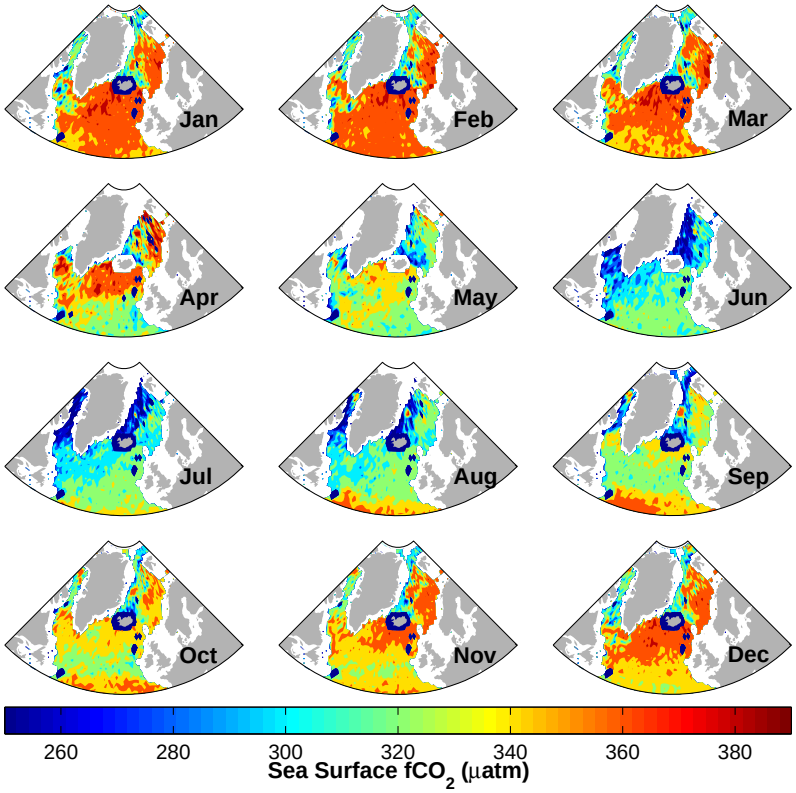
(b)

Figure B.1: SOM A: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean and the Nordic Seas for year 2006 (a) and 2007 (b). SOM A was prepared using the following settings: 'msize': [90 50], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': [20 15].

SOM B: MLD, SST and CHL (summer only)



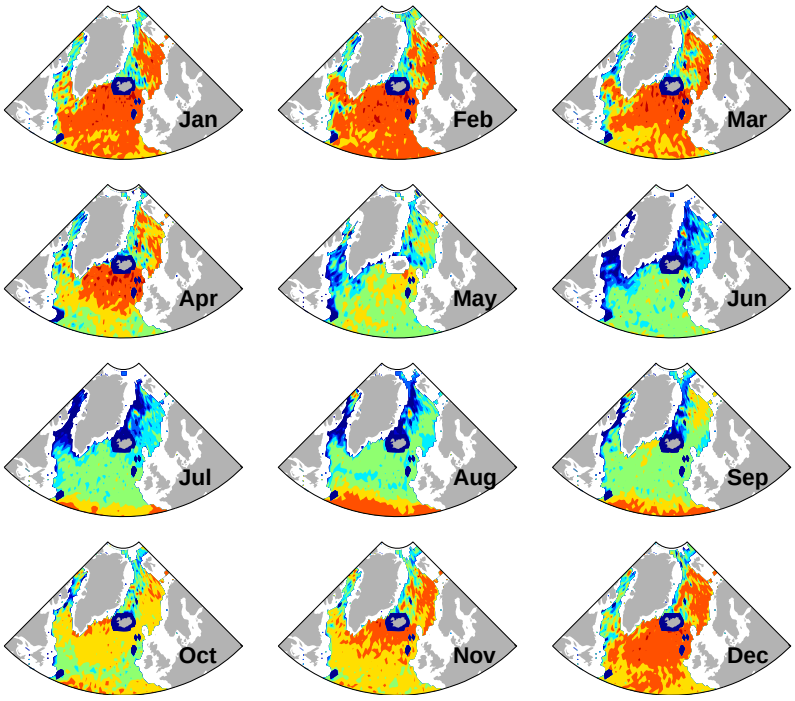
(a)



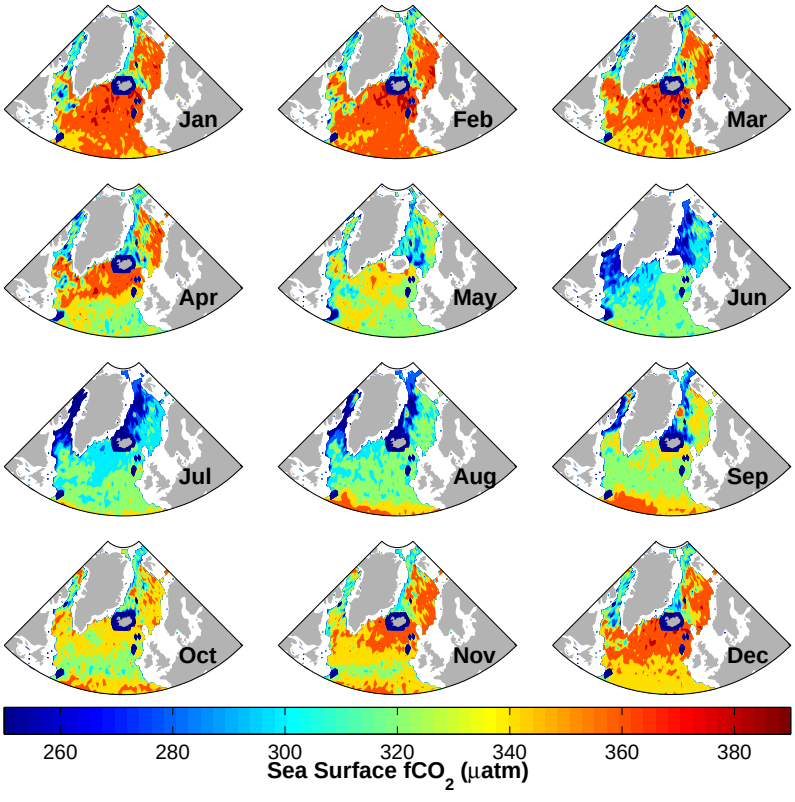
(b)

Figure B.2: SOM B: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean and the Nordic Seas for 2006 (a) and 2007 (b). SOM B was prepared using the following settings: 'msize': Summer/Winter: [90 50], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': Summer: [25 20], Winter: [20 15].

SOM C: MLD, SST and SSS



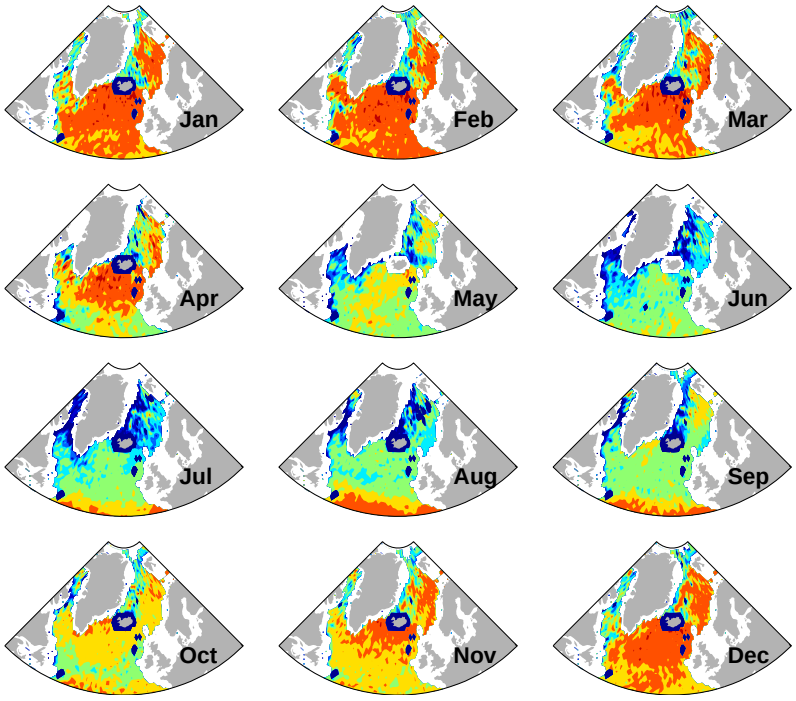
(a)



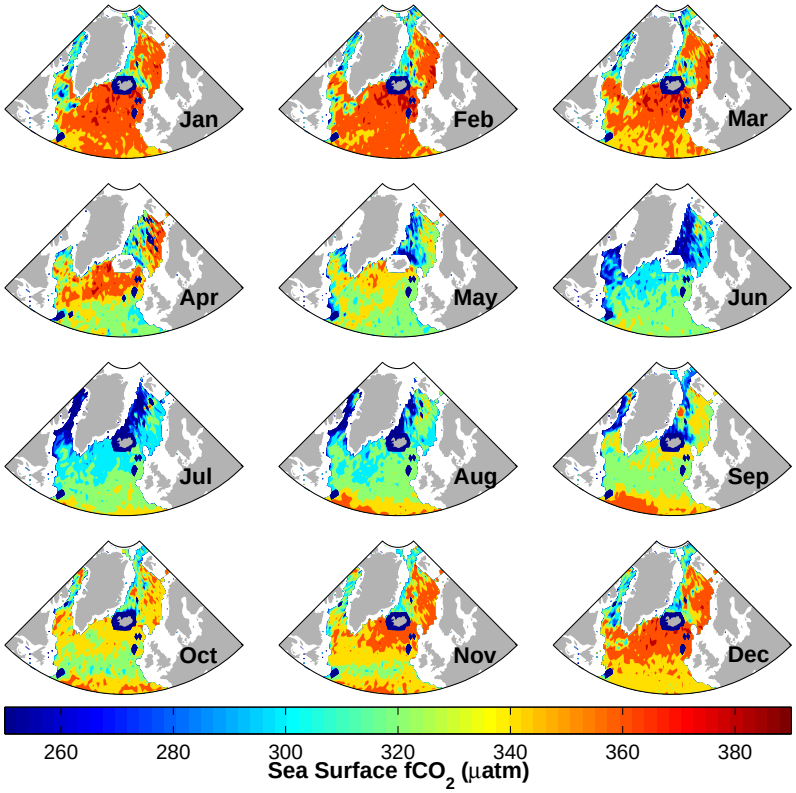
(b)

Figure B.3: SOM C: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean and the Nordic Seas for year 2006 (a) and 2007 (b). SOM C was prepared using the following settings: 'msize': [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': [20 15].

SOM D: MLD, SST, SSS and CHL (summer only)



(a)

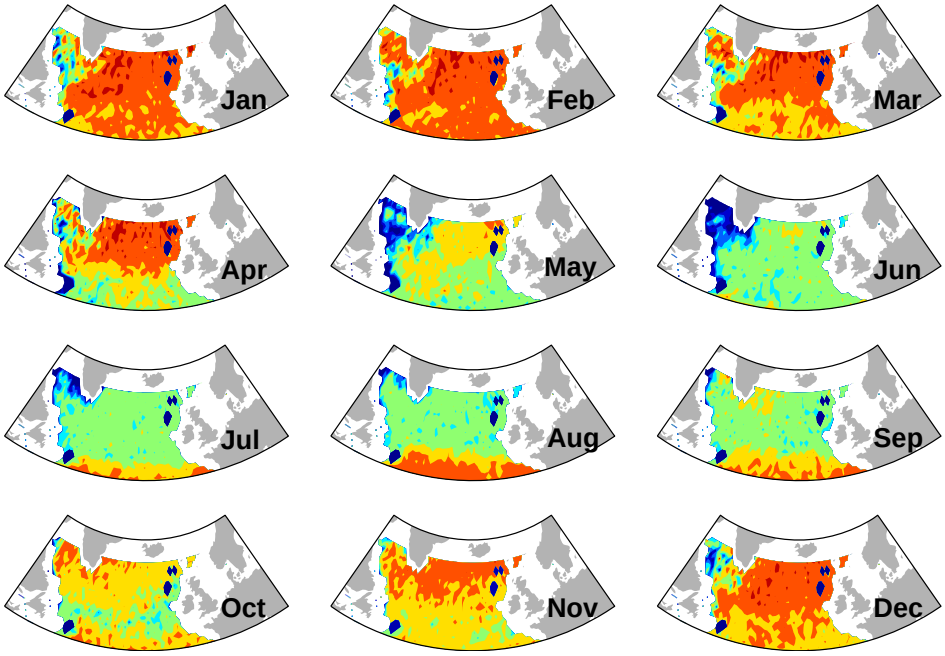


(b)

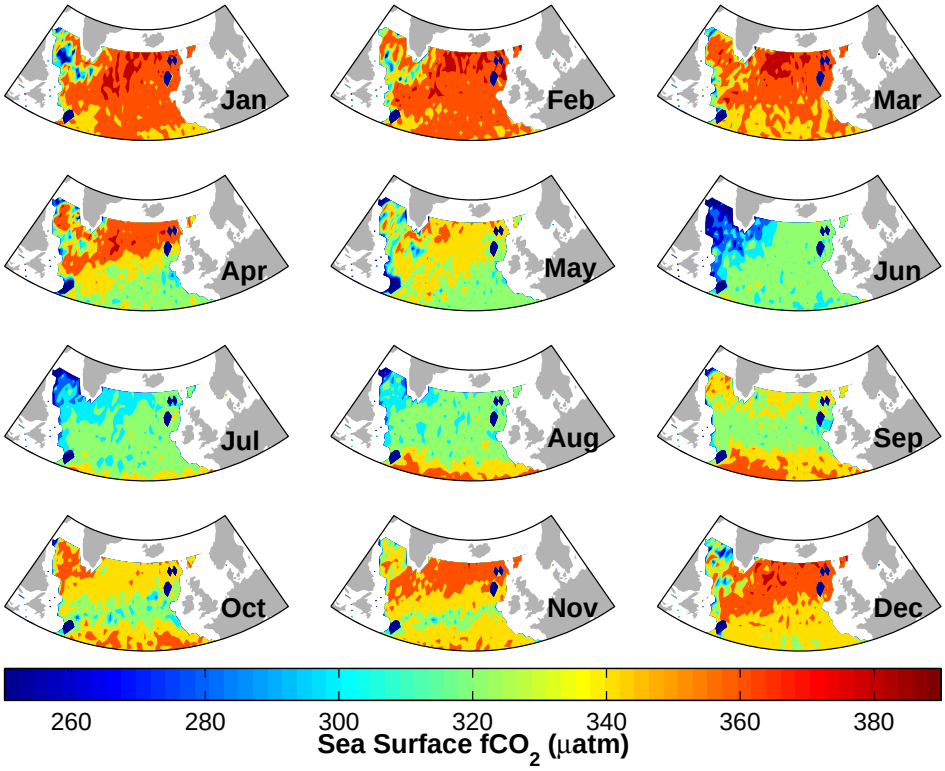
Figure B.4: SOM D: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean and the Nordic Seas for year 2006 (a) and 2007 (b). SOM D was prepared using the following settings: 'msize': Summer: [60 40], Winter: [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': Summer: [15 10], Winter: [20 15].

The northern North Atlantic Ocean [44-63°N]

SOM A: MLD and SST



(a)



(b)

Figure B.5: SOM A: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean for 2006 (a) and 2007 (b). SOM A was prepared using the following settings: 'msize': [85 50], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': [15 10].

SOM B: MLD, SST and CHL (summer only)

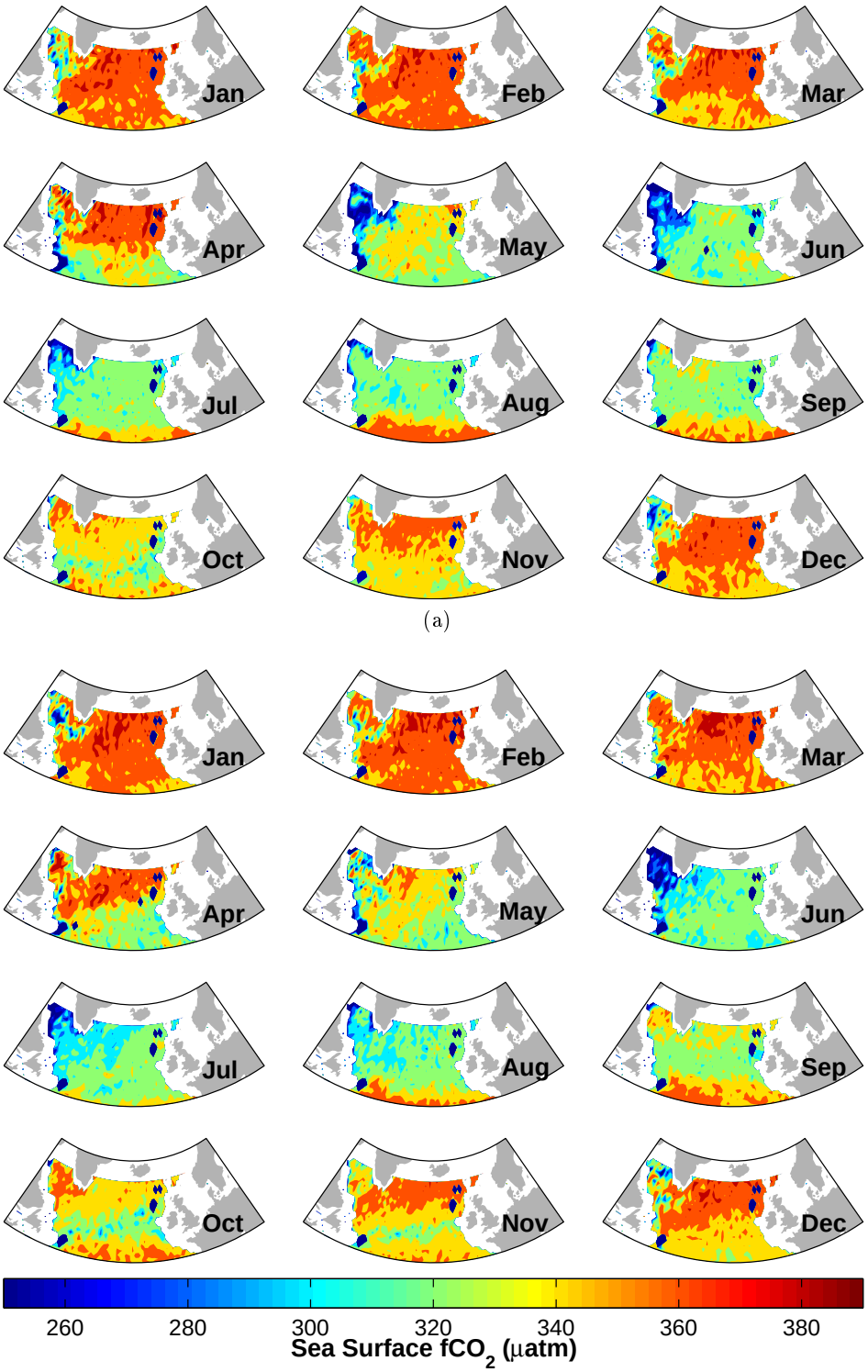
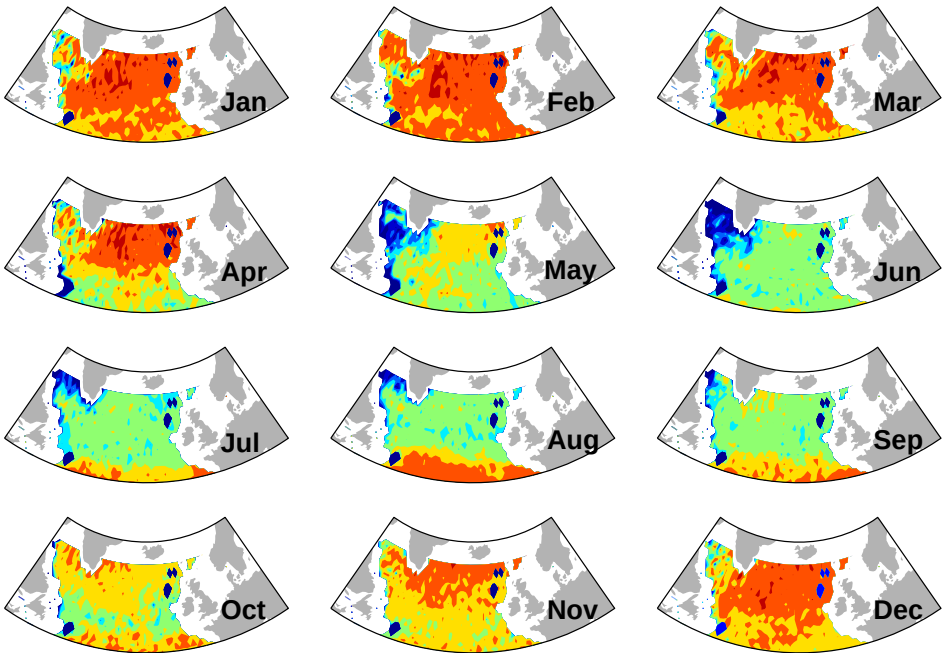
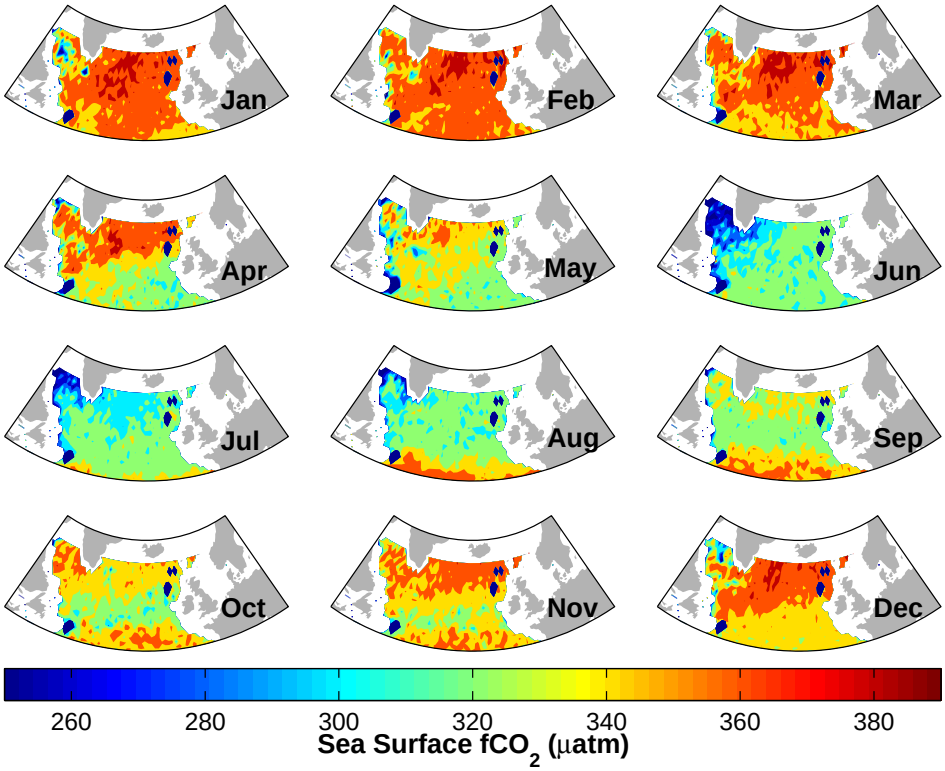


Figure B.6: SOM B: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean for 2006 (a) and 2007 (b). SOM B contains the parameters: 'msize': Summer: [100 65], Winter: [85 50], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': Summer: [25 20], Winter: [15 10].

SOM C: MLD, SST and SSS



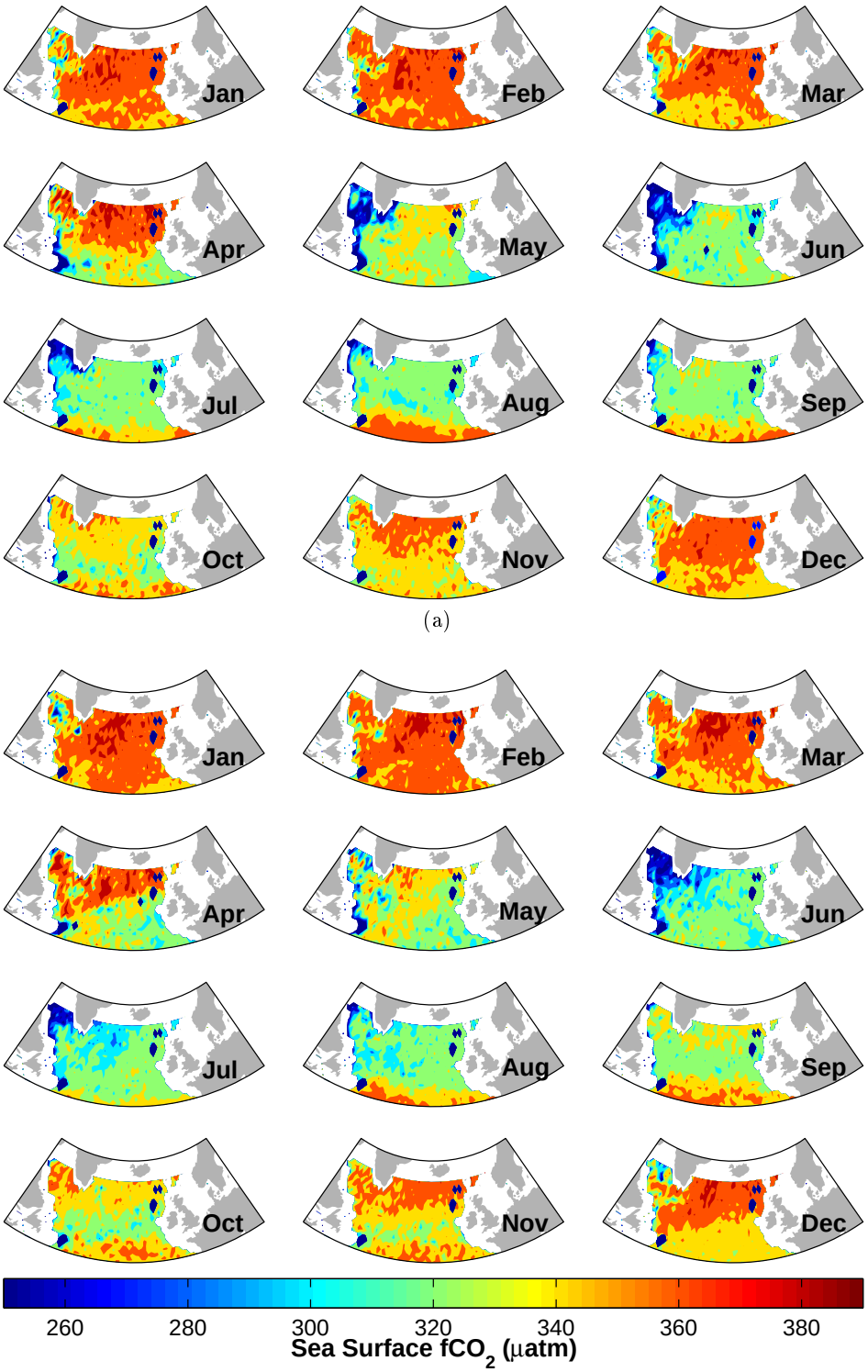
(a)



(b)

Figure B.7: SOM C: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean for 2006 (a) and 2007 (b). SOM C was prepared using the following settings: 'msize': [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': [25 20].

SOM D: MLD, SST, SSS and CHL (summer only)



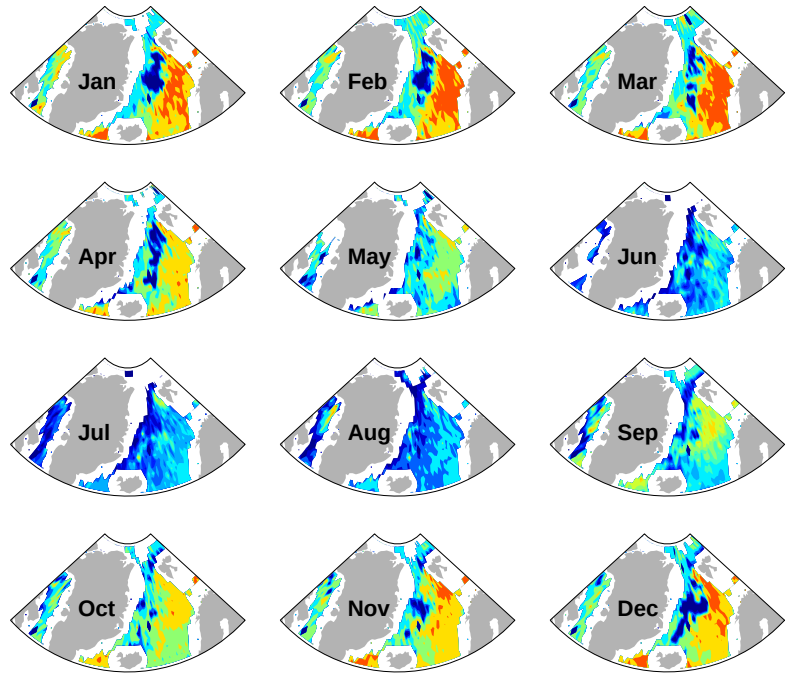
(a)

(b)

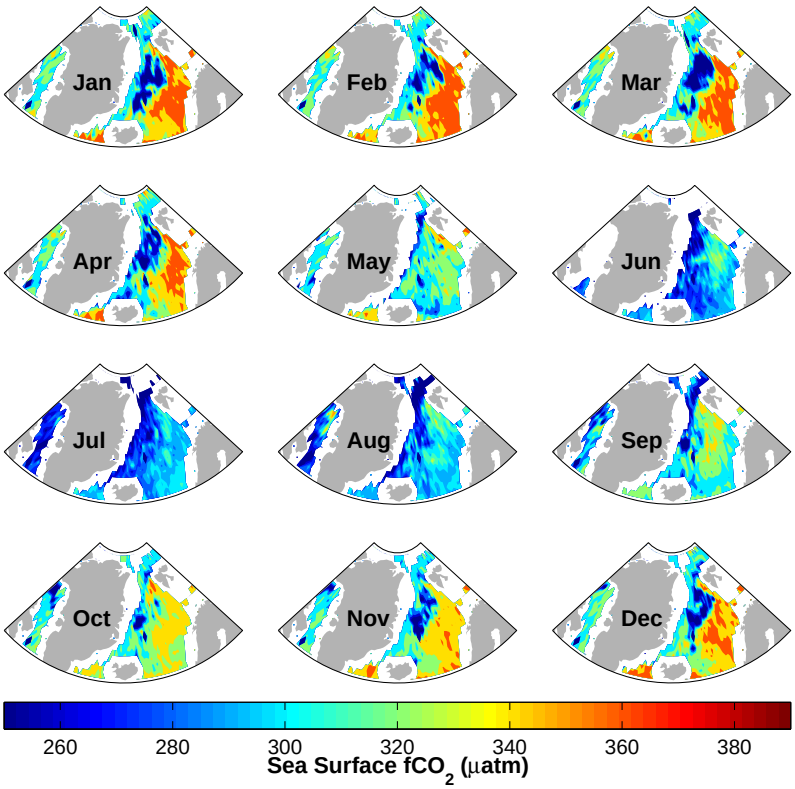
Figure B.8: SOM D: Monthly distribution of estimated sea surface fCO_2 in the northern North Atlantic Ocean for 2006 (a) and 2007 (b). SOM D was prepared using the following settings: 'msize': Summer/Winter: [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': Summer: [15 10], Winter: [25 20].

The Nordic Seas [63-85°N]

SOM A: MLD and SST



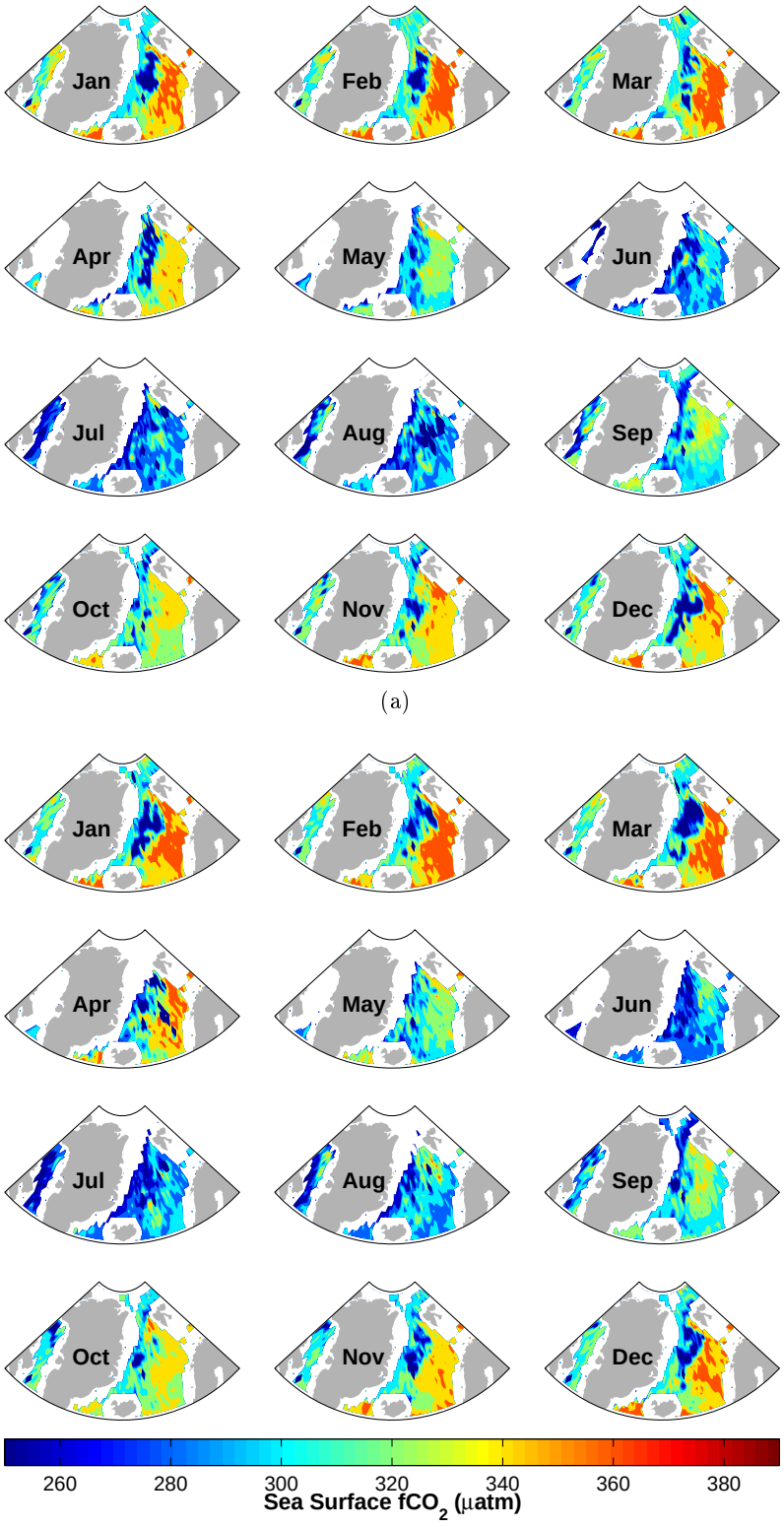
(a)



(b)

Figure B.9: SOM A: Monthly distribution of estimated sea surface fCO_2 in the Nordic Seas for 2006 (a) and 2007 (b). SOM A was prepared using the following settings: 'msize': [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': [25 20].

SOM B: MLD, SST and CHL (summer only)

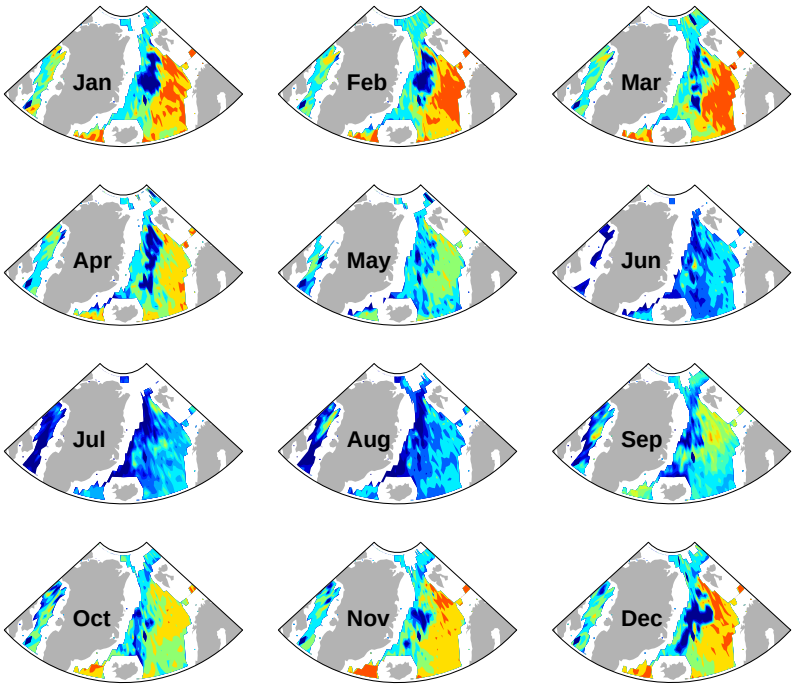


(a)

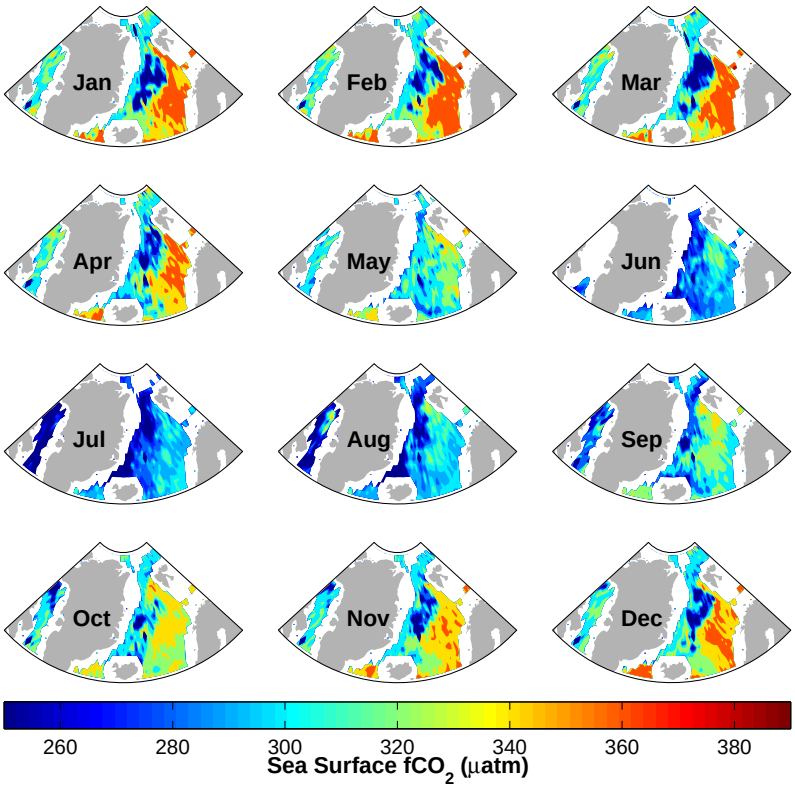
(b)

Figure B.10: SOM B: Monthly distribution of estimated sea surface fCO_2 in the Nordic Seas for 2006 (a) and 2007 (b). SOM B was prepared using the following settings: 'msize': Summer/Winter: [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': Summer/Winter: [25 20].

SOM C: MLD, SST and SSS



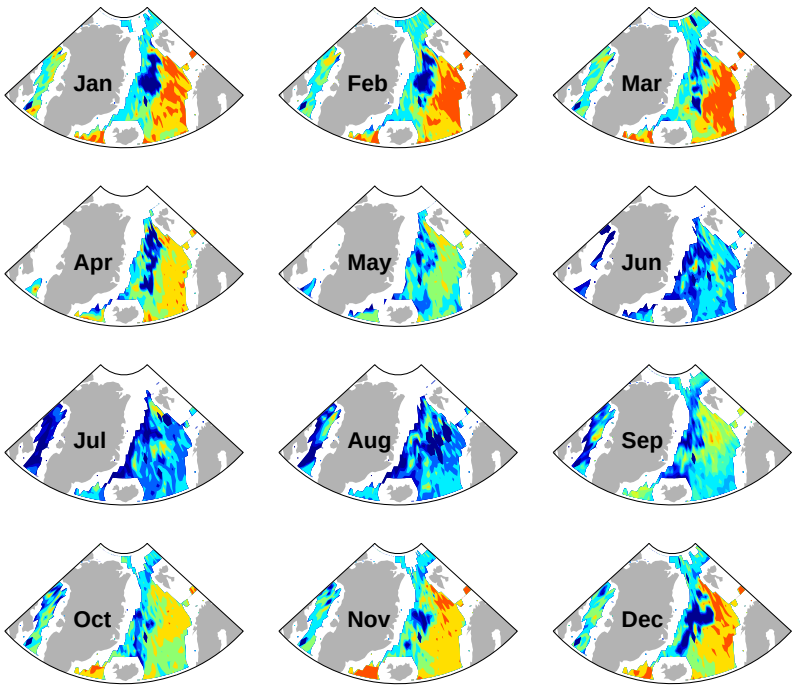
(a)



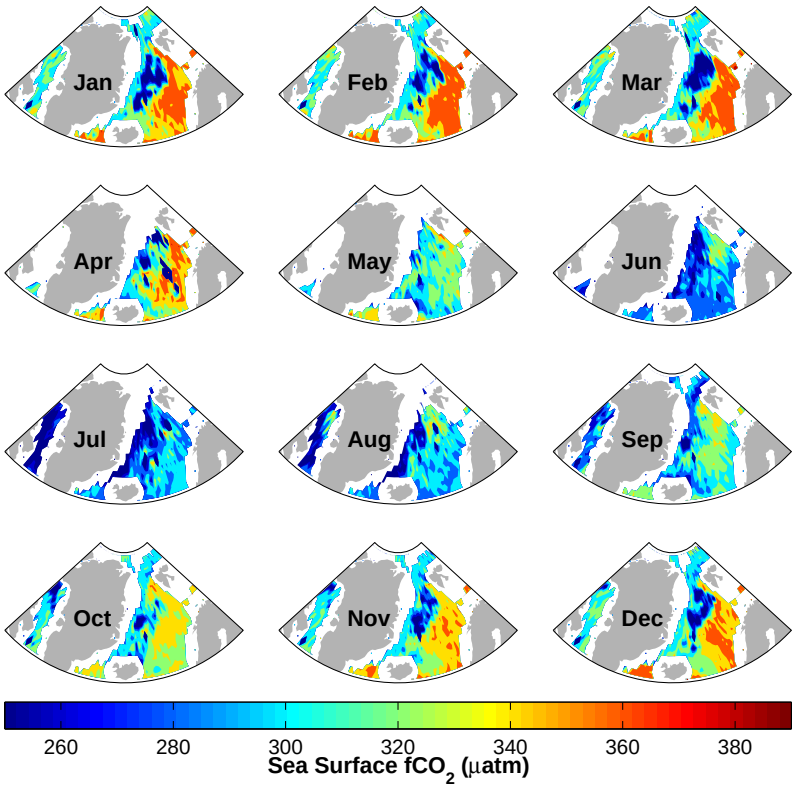
(b)

Figure B.11: SOM C: Monthly distribution of estimated sea surface fCO_2 in the Nordic Seas for 2006 (a) and 2007 (b). SOM C was prepared using the following settings: 'msize': [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': [25 20].

SOM D: MLD, SST, SSS and CHL (summer only)



(a)

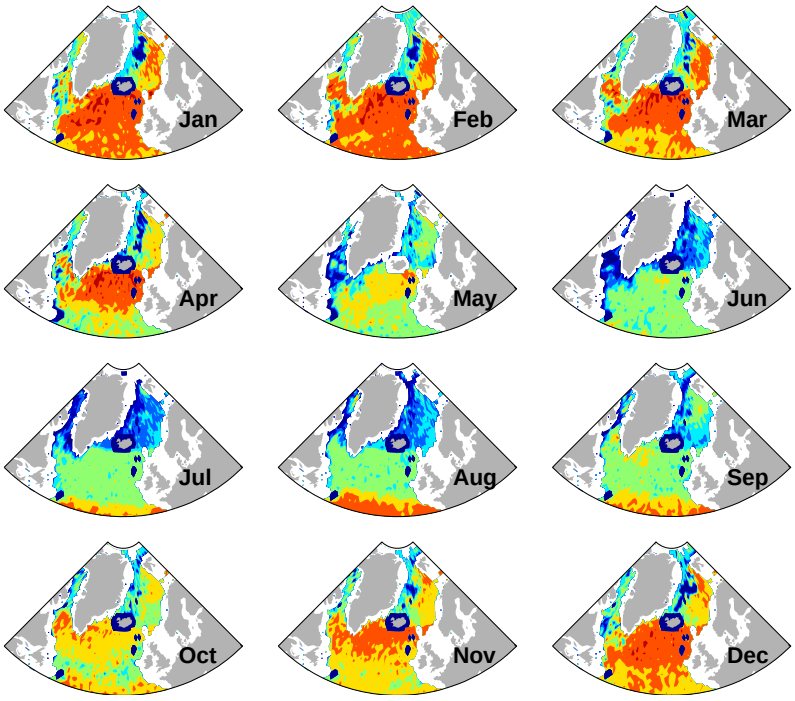


(b)

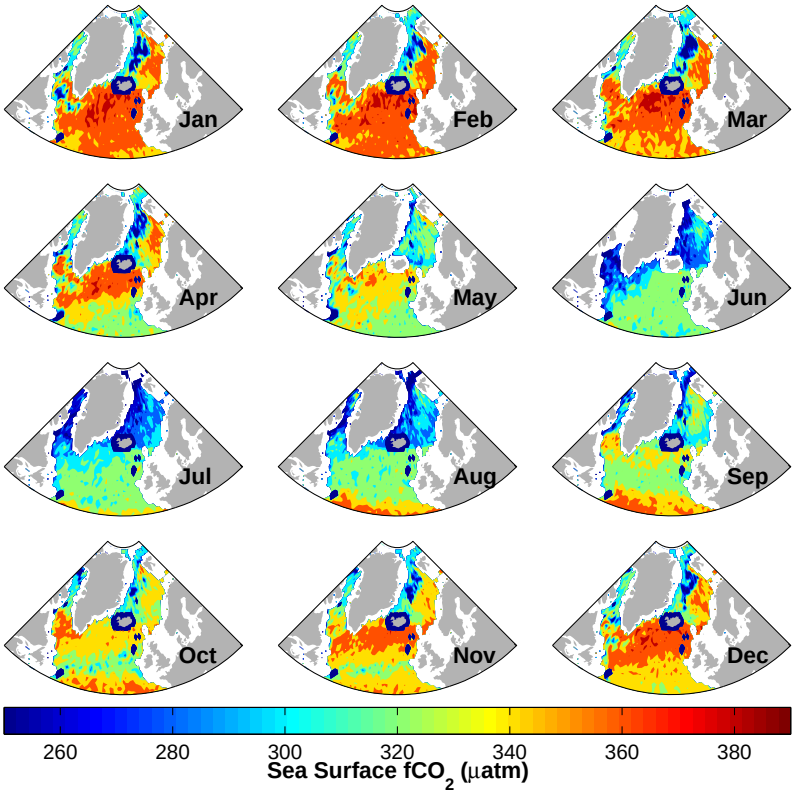
Figure B.12: SOM D: Monthly distribution of estimated sea surface fCO_2 in the Nordic Seas for 2006 (a) and 2007 (b). SOM D was prepared using the following settings: 'msize': Summer/Winter: [100 65], 'lattice': 'hexa', 'shape': 'sheet', 'neigh': 'gaussian' and 'training': Summer: [15 10], Winter: [25 20].

Merging Two SOMs [44-85°N]

SOM A: MLD and SST



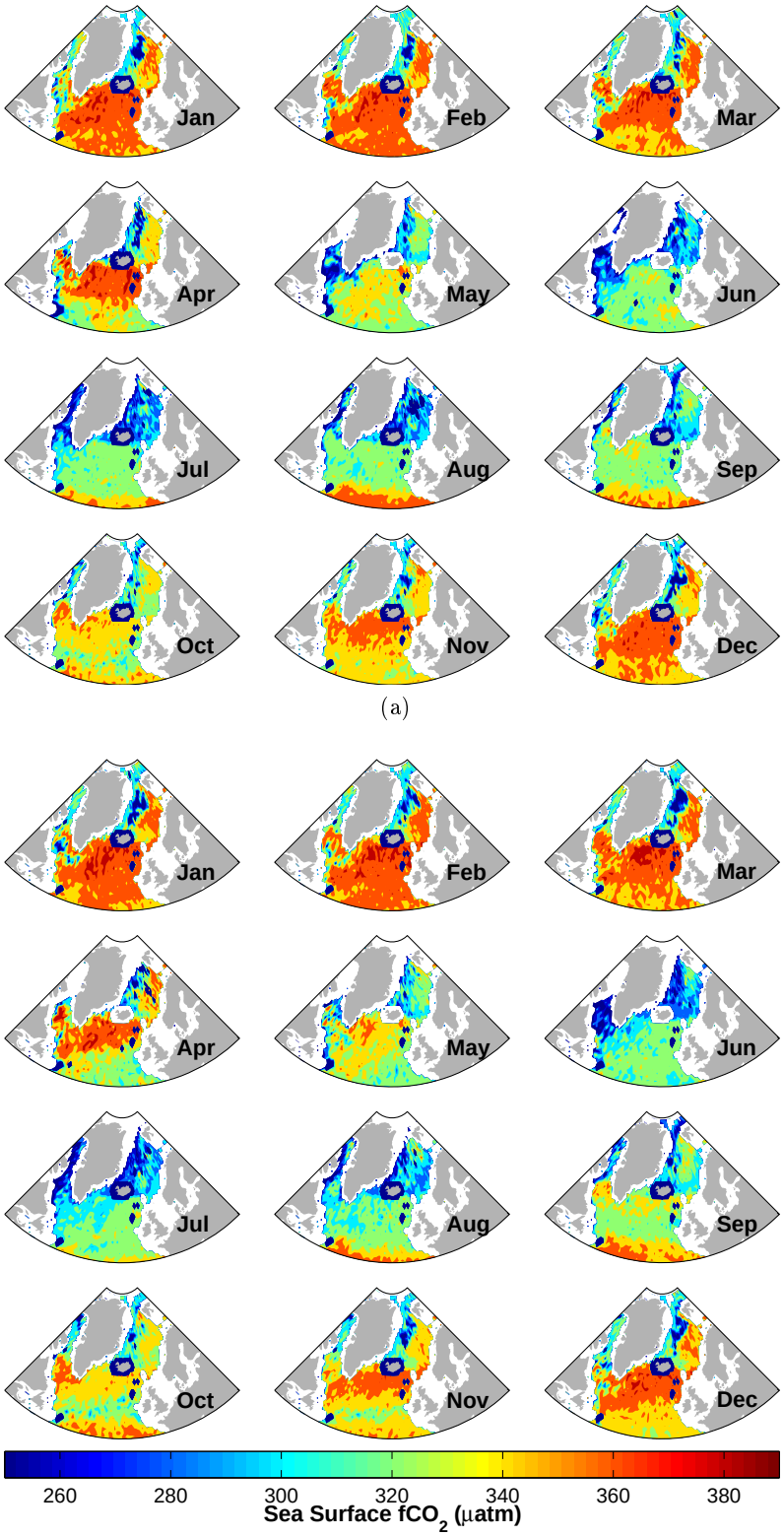
(a)



(b)

Figure B.13: SOM A: Monthly distribution of estimated sea surface fCO_2 for the merged SOM in 2006 (a) and 2007 (b).

SOM B: MLD, SST and CHL (summer only)

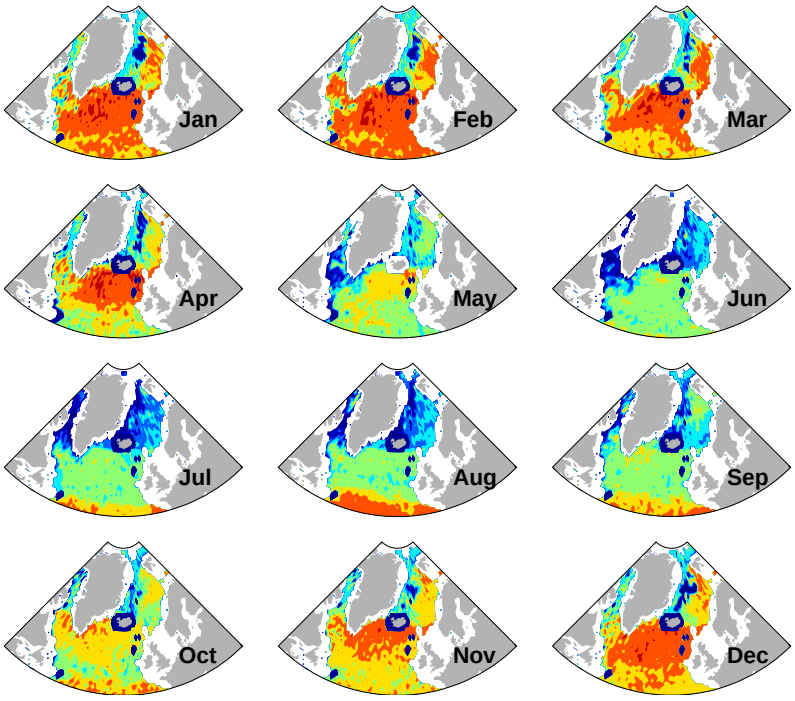


(a)

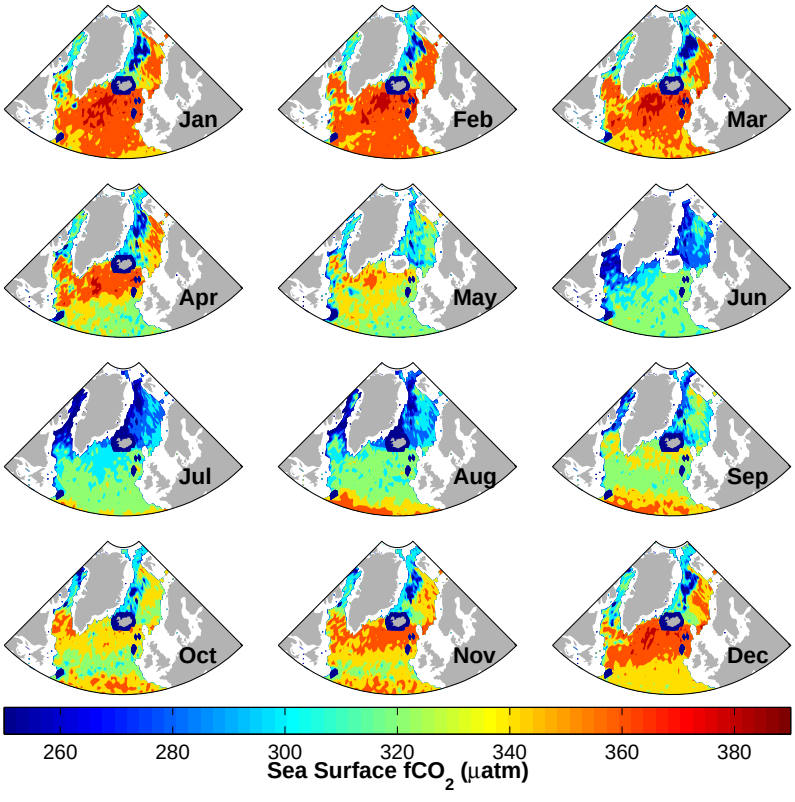
(b)

Figure B.14: SOM B: Monthly distribution of estimated sea surface fCO_2 for the merged SOM in 2006 (a) and 2007 (b).

SOM C: MLD, SST and SSS



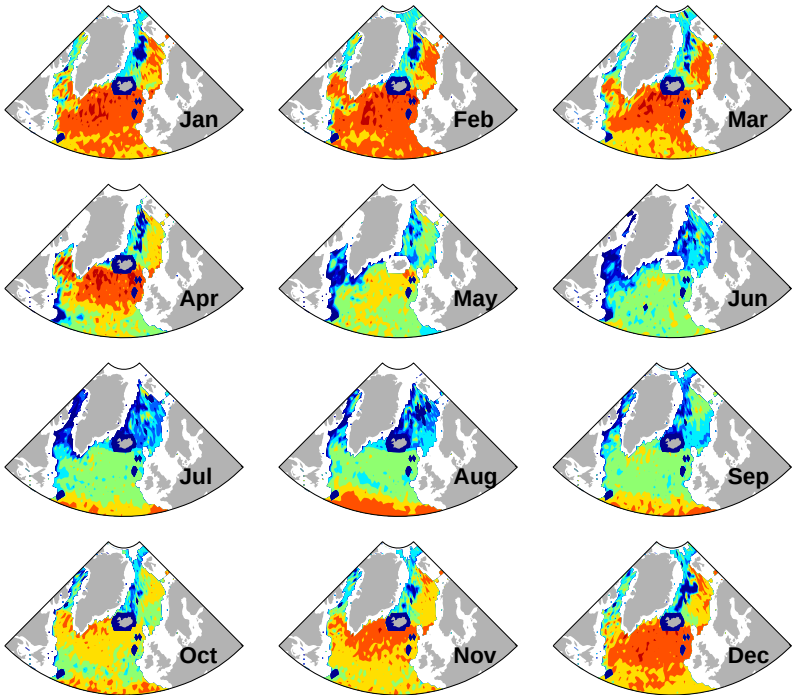
(a)



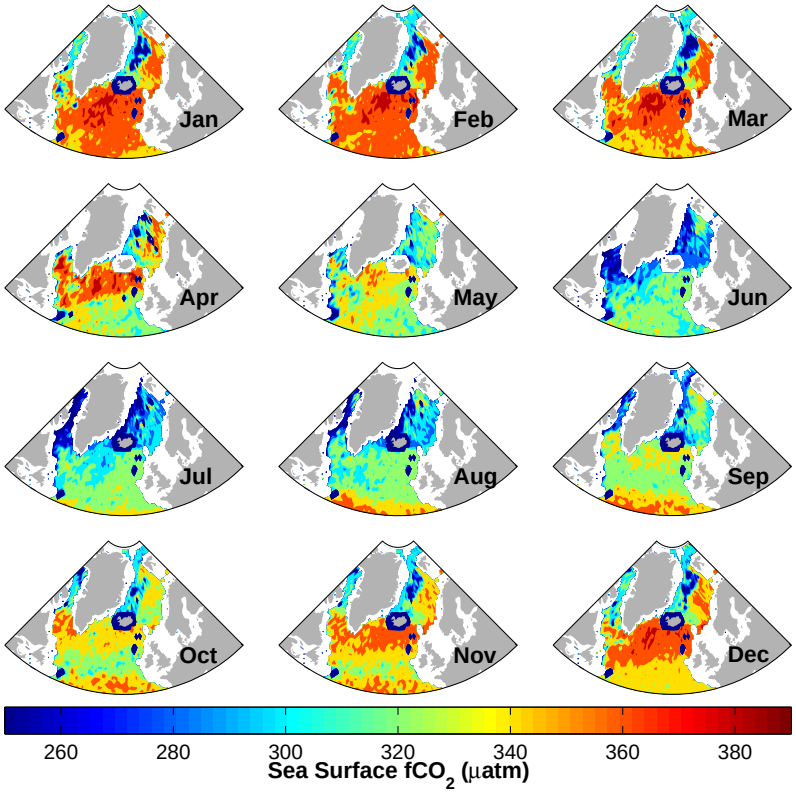
(b)

Figure B.15: SOM C: Monthly distribution of estimated sea surface fCO_2 for the merged SOM in 2006 (a) and 2007 (b).

SOM D: MLD, SST, SSS and CHL (summer only)



(a)



(b)

Figure B.16: SOM D: Monthly distribution of estimated sea surface fCO_2 for the merged SOM in 2006 (a) and 2007 (b).

Bibliography

Allen, J., Holt, J. T., Blackford, J. and Proctor, R. (2007), 'Error quantification of a high-resolution coupled hydrodynamic-ecosystem coastal-ocean model: Part 2. Chlorophyll-a, nutrients and SPM', *Journal of Marine Systems* **68**(3), 381–404.

Awa Niang, Lidwine Gross, Sylvie Thiria, Fouad Badran and Cyril Moulin (2003), 'Automatic neural classification of ocean colour reflectance spectra at the top of the atmosphere with introduction of expert knowledge', *Remote Sensing of Environment* **86**(2), 257 – 271.
URL: *http://www.sciencedirect.com/science/article/pii/S0034425703001135*

Bakker, D., Pfeil, B., Olsen, A., Metzl, N., Hankin, S., Hoppema, M. et al. (2012), 'Introducing the Surface Ocean CO2 Atlas (SOCAT) for the polar regions'.

Bishop, C. M. (1995), *Neural networks for pattern recognition*, Oxford university press.

Chierici, M., Signorini, S. R., Fransson, A. and Olsen, A. (2012), 'Surface water CO2 algorithms for the high-latitude Pacific sector of the Southern Ocean', *Remote Sensing of Environment* **119**, 184–196.

Denman, K., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P., Dickinson, R., Hauglustaine, D., Heinze, C., Holland, E., Jacob, D., Lohmann, U., Ramachandran, S., da Silva Dias, P., Wofsy, S. and Zhang, X. (2007), 'Couplings Between Changes in the Climate System and Biogeochemistry', *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* pp. 499–587.

Dickson, A. G., Sabine, C. L. and Christian, J. R. (2007), 'Guide to best practices for ocean CO2 measurements'.

Doney, S., Fabry, V., Feely, R. and Kleypas, J. (2009), 'Ocean acidification: the other CO2 problem', *Annual Review in Marine Science* **1**.

Dreyfus, G. (2005), *Neural networks: methodology and applications*, Springer Berlin.

Emerson, S. and Hedges, J. (2008), *Chemical oceanography and the marine carbon cycle*, Cambridge University Press.

Etheridge, D., Steele, L., Langenfelds, R., Francey, R., Barnola, J. and Morgan, V. (1998), 'Historical CO2 records from the Law Dome DE08, DE08-2, and DSS ice cores', *Trends: a compendium of data on global change* pp. 351–364.

Günther, F. and Fritsch, S. (2010), 'neuralnet: Training of neural networks', *The R Journal* **2**(1), 30–38.

Heiat, A. (2002), 'Comparison of artificial neural network and regression models for estimating software development effort', *Information and Software Technology* **44**(15), 911–922.

Heinze, C., Maier Reimer, E. and Winn, K. (1991), 'Glacial pCO2 reduction by the world ocean: Experiments with the Hamburg carbon cycle model', *Paleoceanography* **6**(4), 395–430.

Holden, J. (2012), *An Introduction to Physical Geography and the Environment, Third edition*, Pearson Education, Limited.

Hollmén, J. (1996), 'Process modeling using the self-organizing map'.

Holt, J. T., Allen, J., Proctor, R. and Gilbert, F. (2005), 'Error quantification of a high-resolution coupled hydrodynamic–ecosystem coastal–ocean model: part 1 model overview and assessment of the hydrodynamics', *Journal of Marine Systems* **57**(1), 167–188.

Lauvset, S., Chierici, M., Counillon, F., Omar, A., Nondal, G., Johannessen, T. and Olsen, A. (2013), 'Annual and seasonal fCO2 and air–sea CO2 fluxes in the Barents Sea', *Journal of Marine Systems* **113-114**, 62–74.

Lefévre, N., Watson, A. and Watson, A. (2005), 'A comparison of multiple regression and neural network techniques for mapping in situ pCO2 data', *Tellus B* **57**(5), 375–384.

Liu, Y., Weisberg, R. H. and Mooers, C. N. (2006), 'Performance evaluation of the self-organizing map for feature extraction', *Journal of Geophysical Research: Oceans (1978–2012)* **111**(C5).

Liu, Y., Wesiberg, R. H. and He, R. (2006), 'Sea surface temperature patterns on the West Florida Shelf using growing hierarchical self-organizing maps', *Journal of Atmospheric and Oceanic Technology* **23**, 325–338.

Mihanović, H., Cosoli, S., Vilibić, I., Ivanković, D., Dadić, V. and Gačić, M. (2011), 'Surface current patterns in the northern Adriatic extracted from high-frequency radar data using self-organizing map analysis', *Journal of Geophysical Research* **116**(C8), C08033.

Millero, F. J. (2013), *Chemical oceanography*, CRC PressI Llc.

Nakaoka, S., Telszewski, M., Nojiri, Y., Yasunaka, S., Miyazaki, C., Mukai, H. and Usui, N. (2013), 'Estimating temporal and spatial variation of ocean surface pCO2 in the North Pacific using a Self Organizing Map neural network technique', *Biogeosciences Discussions* **10**, 4575–4610.

Nash, J. and Sutcliffe, J. (1970), 'River flow forecasting through conceptual models part I—A discussion of principles', *Journal of hydrology* **10**(3), 282–290.

Niang, A., Badran, F., Moulin, C., Crepon, M. and Thiria, S. (2006), 'Retrieval of aerosol type and optical thickness over the Mediterranean from SeaWiFS images using an automatic neural classification method', *Remote sensing of Environment* **100**(1), 82–94.

Olsen, A., Bellerby, R. G., Johannessen, T., Omar, A. M. and Skjelvan, I. (2003), 'Interannual variability in the wintertime air–sea flux of carbon dioxide in the northern North Atlantic, 1981–2001', *Deep Sea Research Part I: Oceanographic Research Papers* **50**(10), 1323–1338.

Olsen, A., Brown, K., Chierici, M., Johannessen, T., Neill, C. et al. (2008), 'Sea-surface CO2 fugacity in the subpolar North Atlantic', *Biogeosciences* **5**(2), 535–547.

Pfeil, B., Olsen, A., Bakker, D., Hankin, S., Koyuk, H., Kozyr, A., Malczyk, J., Manke, A., Metzl, N., Sabine, C. et al. (2013), 'A uniform, quality controlled Surface Ocean CO2 Atlas (SOCAT)', *Earth* .

Phillips, L. F. (1994), 'The theory of air-sea gas exchange', *Accounts of chemical research* **27**(8), 217–223.

Polo, I., Ullmann, A., Roucou, P. and Fontaine, B. (2011), 'Weather regimes in the Euro-Atlantic and Mediterranean sector, and relationship with West African rainfall over the 1989-2008 period from a self-organizing maps approach', *Journal of Climate* **24**(13), 3423–3432.

Rödenbeck, C., Houweling, S., Gloor, M. and Heimann, M. (2003), 'CO 2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport', *Atmospheric Chemistry and Physics* **3**(6), 1919–1964.

Rohde, R. (2011), 'Global Warming Art'.
URL: *http://www.globalwarmingart.com/*

Sabine, C., Feely, R., Gruber, N., Key, R., Lee, K., Bullister, J., Wanninkhof, R., Wong, C., Wallace, D., Tilbrook, B. et al. (2004), 'The oceanic sink for anthropogenic CO2', *Science* **305**(5682), 367.

Sarmiento, J. and Gruber, N. (2006), *Ocean biogeochemical dynamics*, Vol. 503, Cambridge University Press.

Stow, C. A., Jolliff, J., McGillicuddy Jr, D. J., Doney, S. C., Allen, J., Friedrichs, M. A., Rose, K. A. and Wallhead, P. (2009), 'Skill assessment for coupled biological/physical models of marine systems', *Journal of Marine Systems* **76**(1), 4–15.

Sweeney, C., Takahashi, T., Gnanadesikan, A., Wanninkhof, R., Feely, R. A., Friedrich, G., Chavez, F., Bates, N., Olafsson, J. and Sarmiento, J. (2002), 'Spatial and temporal variability of surface water pCO2 and sampling strategies', *A large-scale CO2 observing plan: In situ oceans and atmosphere, edited by: Bender, M., Doney, S., Feely, RA, et al., National Technical Information Service, Springfield, Virginia, USA* pp. 155–175.

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C. et al. (2002), 'Global sea–air CO2 flux based on climatological surface ocean pCO2, and seasonal biological and temperature effects', *Deep Sea Research Part II: Topical Studies in Oceanography* **49**(9), 1601–1622.

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C. et al. (2009), 'Climatological mean and decadal change in surface ocean pCO2, and net sea–air CO2 flux over the global oceans', *Deep Sea Research Part II: Topical Studies in Oceanography* **56**(8), 554–577.

Tans, P. (n.d.), 'Trends in Atmospheric Carbon Dioxide', NOAA/ESRL.
URL: *http://www.esrl.noaa.gov/gmd/ccgg/trends/*

Tans, P. and Thoning, K. (2008), 'How we measure background CO2 levels on Mauna Loa'.
URL: *http://www.esrl.noaa.gov/gmd/ccgg/about/co2_measurements.html*

Telszewski, M. (2009), An investigation of sea surface carbon dioxide and its distribution in the North Atlantic using a neural network technique, PhD thesis, University of East Anglia.

Telszewski, M., Chazottes, A., Schuster, U., Watson, A., Moulin, C., Bakker, D., Gonzalez Davila, M., Johannessen, T., Kortzinger, A., Luger, H. et al. (2009), 'Estimating the monthly pCO2 distribution in the North Atlantic using a self-organizing neural network'.

Trujillo, A. and Thurman, H. (2010), *Essentials of Oceanography*, Pearson Education.

Vesanto, J., Himberg, J., Alhoniemi, E. and Parhankangas, J. (2000), *SOM toolbox for Matlab 5*, Citeseer.

Watson, A. J. and Orr, J. C. (2003), Carbon dioxide fluxes in the global ocean, *in* 'Ocean Biogeochemistry', Springer, pp. 123–143.

Williams, R. and Follows, M. (2011), *Ocean Dynamics and the Carbon Cycle: Principles and Mechanisms*, Cambridge University Press.

Zeebe, R. and Wolf Gladrow, D. (2001), *CO2 in seawater: equilibrium, kinetics, isotopes*, Vol. 65, Elsevier Science.