



## Water mass distributions in the Southern Ocean derived from a parametric analysis of mixing water masses

Anouk de Brauwere,<sup>1</sup> Stéphanie H. M. Jacquet,<sup>1</sup> Fjo De Ridder,<sup>1,2</sup> Frank Dehairs,<sup>1</sup> Rik Pintelon,<sup>2</sup> Johan Schoukens,<sup>2</sup> and Willy Baeyens<sup>1</sup>

Received 1 June 2006; revised 13 September 2006; accepted 6 October 2006; published 20 February 2007.

[1] An empirical method to reconstruct distributions of mixing water masses is developed and applied to two Southern Ocean (Australian sector) data sets. The method is a parametric extension of the well-known Optimum Multiparameter (OMP) analysis, therefore called POMP (parametric OMP). The main development consists of parameterizing the mixing fractions as a function of position (e.g. latitude and depth) and estimating the parameters of the resulting function, instead of estimating the mixing fractions for every sampling point individually. Because this enables to reduce the total number of unknowns to be estimated, the proposed adjustments result in more robust estimates of the mixing fractions, as illustrated on the Southern Ocean CLIVAR-SR3 data set (early spring 2001). An additional consequence of the new working scheme is that the mixing distributions are smoothed. It is emphasized that the degree of smoothing should be chosen with care in order not to neglect any significant features. Statistical rules to do this are applied. Application of the POMP method to the second Southern Ocean data set (SAZ'98, summer 1998) revealed the importance of a careful selection of the mixing source water types. This study was limited to the subantarctic region and upper 1000 m, where variations in water mass characteristics and distributions are most important. Indeed, it seems necessary to redefine the source water characteristics for each data set. The direction of these changes suggests that interannual variability or long term changes of water masses overwhelm the seasonal variability.

**Citation:** de Brauwere, A., S. H. M. Jacquet, F. De Ridder, F. Dehairs, R. Pintelon, J. Schoukens, and W. Baeyens (2007), Water mass distributions in the Southern Ocean derived from a parametric analysis of mixing water masses, *J. Geophys. Res.*, 112, C02021, doi:10.1029/2006JC003742.

### 1. Introduction

[2] To understand the role of the Southern Ocean in the global climate functioning, knowledge about ocean circulation and mixing is of particular interest. The subantarctic region merits special attention because it represents a key interface that mediates the influence of the Southern Ocean on the globe, more especially concerning CO<sub>2</sub> uptake and other exchanges with the atmosphere [e.g., Takahashi *et al.*, 2002]. In addition, the Subantarctic Zone is affected by intense circulation changes and corresponds to a critical water mass formation region. Resolving seasonal and interannual variations of the ocean energetics (mainly temperature and salinity budgets) is thus of particular importance to the understanding of Southern Ocean dynamics.

[3] Oceanic circulation and water mass mixing processes can be studied via complex physical models [Siedler *et al.*, 2001]. Such an approach has the advantage to be highly

interpretable since every aspect can be traced back to fundamental physical laws. However, validation of these models using extensive data sets is not straightforward. On the other hand, a conceptually more simple, but completely data-based approach is often used: Optimum Multiparameter (OMP) analysis [Tomczak, 1981; Thompson and Edwards, 1981; Mackas *et al.*, 1987; Tomczak and Large, 1989]. In this paper we consider this type of inverse method to reconstruct distributions of mixing water masses.

[4] Classical OMP analysis usually considers a water parcel sampled on a grid (e.g. latitude versus depth) with a total of  $N$  grid points. At every sampling position of the grid,  $n_v$  physical and/or chemical properties are measured, such as temperature, salinity, dissolved oxygen concentration, etc. These data are considered to be the result of mixing of a number ( $n_s$ ) of source water masses, whose characteristics are assumed to be known. Following the definition given by Tomczak [1999], a set of properties corresponding to such a source water mass is called a source water type (SWT). Defining the SWTs is a delicate job and the final outcome of the analysis highly depends on the choices made here. Once the SWTs are characterized, each measured property can be written as a linear combination of the SWTs' values for that hydrographic property. For

<sup>1</sup>Department of Analytical and Environmental Chemistry, Vrije Universiteit Brussel, Brussels, Belgium.

<sup>2</sup>Department of Electricity and Instrumentation, Vrije Universiteit Brussel, Brussels, Belgium.

example, measured temperature at location  $k$  can be represented by

$$T_m(k) = x_1(k)T_1 + x_2(k)T_2 + \dots + x_{n_S}(k)T_{n_S} + e_T(k), \quad (1)$$

where  $x_i$  is the  $i$ th SWT contribution or fraction and  $T_i$  its temperature ( $i = 1, \dots, n_S$ ), and  $e_T(k)$  the disturbing noise. For each hydrographic property a similar equation can be written. If enough properties are measured, the unknown fractions  $x_1, \dots, x_{n_S}$  at position  $k$  can be estimated by simultaneously optimizing all mixing equations of form (1) and the mass balance equation

$$x_1(k) + \dots + x_{n_S}(k) = 1. \quad (2)$$

Usually a nonnegativity constraint is imposed on the mixing fractions, for obvious reasons of physical relevance. The procedure can be repeated for all grid points, resulting in optimal values for the fractions at each sampling location. These are the basic ideas of the OMP approach.

[5] This method has first been presented more than two decades ago by *Tomczak* [1981]. Several improvements have been proposed since then. The most objective way to assign a weight to each variable in order to reflect differences in reliability of each variable has been discussed [*Thompson and Edwards*, 1981; *Mackas et al.*, 1987; *Tomczak and Large*, 1989]. Later, multivariate methods, like cluster analysis and principal components analysis, have been applied in an attempt to more objectively determine the source characteristics (i.e. the  $T_i$  in equation (1)) [*You and Tomczak*, 1993; *de Boer and van Aken*, 1995; *You*, 1997; *Fogelqvist et al.*, 2003]. Although these methods are efficient in reconstructing multivariate data sets, they lack the physical interpretability that OMP analysis has. In particular, the “SWTs” are mere mathematical constructs representing the directions in multivariate space with maximal variation, but their characteristics generally do not correspond to real source water masses. Therefore, these methods are not further considered here. The OMP methodology has been widely used in oceanographic literature to investigate source water contributions in various regions and systems [e.g., *You*, 1997; *Castro et al.*, 1998; *Coatanoan et al.*, 1999; *Perez et al.*, 2001; *You et al.*, 2003; *Lo Monaco et al.*, 2005].

[6] This is probably due to its main advantage: its simplicity. It is indeed a very intuitive approach, not requiring many assumptions about the physics of the investigated system. The outcome of the computations (i.e. the fraction fields) is directly visualizable and interpretable. This does not mean that the analysis input should not be chosen carefully, but rather that the quality of the results can often be checked even visually. Yet, this OMP method exhibits a number of imperfections, especially from a statistical point of view.

[7] 1. Since every grid point is optimized individually and independently, the position dependent variation of the source contributions over the grid is not taken into account. In other words, it is not considered that if a SWT contribution is high in one point it will probably be quite high in a neighboring point too. A consequence of the current working scheme is that the mixing fractions for a given point only depend on the data associated with this point, and

hence are influenced by errors in that point, i.e. both stochastic and systematic errors.

[8] 2. To estimate all unknown fractions  $x_1, \dots, x_{n_S}$  from the set of equations (mixing equations of form (1) and the mass balance equation (2)), the number of unknown fractions,  $n_S$ , may not exceed the number of measured variables ( $n_v$ ) + 1. Otherwise the problem is degenerated, which means there is an infinite number of solutions for  $x_1, \dots, x_{n_S}$  which are all equally good. This situation can be compared to estimating two unknowns  $a$  and  $b$  from only one equation (e.g.  $a + b = 1$ ): there is an infinite number of solutions for  $a$  and  $b$ . In the extreme situation where  $n_S = n_v + 1$ , there is exactly one mathematically correct solution for the unknown fractions, but it is of little statistical significance because there are no residual degrees of freedom. This situation can be compared to a linear regression (two unknown parameters) through two data points (two equations): a straight line can perfectly be fitted through two points, but the noise contained in the data is entirely modeled as significant information. Consequently, the fitted line will certainly not equal the true line through the true data values. To conclude, from the statistical point of view, the OMP method should be performed in overdetermination, i.e. with  $n_S \leq n_v$ , and preferably even  $n_S \ll n_v$ . However, due to the reality that only a small number of conservative properties are usually available, it is common practice to apply OMP analysis to the extreme situation with  $n_S = n_v + 1$ . The main consequences are: (1) the resulting fraction values are highly influenced by stochastic errors in the data and (2) they are also associated with a higher uncertainty.

[9] 3. In the OMP procedure mass balance (equation (2)) is treated as if it were a mixing equation (see equation (1)). Conceptually this is not entirely correct since the mixing equations are actually modeling a measurement, whereas the mass balance equation only states a physical constraint that should (exactly) be obeyed. As a consequence of this handling, the fractions optimized by OMP analysis only approximately obey mass balance. An additional difficulty with this approach arises when weights have to be chosen for each equation. Indeed, before optimizing the mixing equations, the tracers are weighted. The usual procedure is to attribute a weight to each equation, expressing the reliability of the measurement described in the equation. This highly influences the final fractions and is a healthy thing to do since it is known that some variables are measured with much higher precision than others. However, as mentioned just before, it is not possible to attribute such a weighting factor to the mass balance equation, because there is no measurement involved in this equation. In practice, “the largest of the weights (...) is allocated to the mass conservation equation” [*You and Tomczak*, 1993]. This dilemma was already noted by *Tomczak and Large* [1989]: “More thought has to be given to the most suitable weights. How important in this context is conservation of mass?”

[10] The goal (and organization) of the current article is twofold. First, it aims at curing the abovementioned imperfections while retaining the attractive fundamental ideas and interpretability of OMP analysis. The new procedure is based on the idea of parameterizing the unknown fraction fields using two-dimensional spline functions (section 2).

This approach will primarily make the method more robust to stochastic and systematic errors in the data and enable a consistent weighting scheme. The second aim is to apply the method to two Southern Ocean data sets (WOCE SR3 line, Australian sector) representing two different years and seasons (section 3). This exercise serves to illustrate the new method and allows to investigate the changes in water mass characteristics and distribution between the two data sets in terms of interannual and seasonal variability.

## 2. Method

[11] In this part the parametric method to estimate water mass mixing fractions is described. First, the model equations are derived formally (2.1) and the choice of the parametric model is discussed in more detail in section 2.2. Next, some practical remarks are given about constraints (2.3) and weighting (2.4). All remarks are combined in an algorithm which estimates the optimal values for the sources' mixing coefficients. A schematic outline of the computation algorithm is given in Appendix A.

### 2.1. Parameterization of the Unknown Fractions

[12] If  $n_v$  hydrographic properties are considered in OMP analysis, a mixing equation like equation (1) can be written for each variable at position  $k$ . This set of mixing equations for position  $k$  can formally be combined in one matrix equation

$$\mathbf{y}_k = \mathbf{S} \cdot \mathbf{x}_k + \mathbf{e}_k, \quad (3)$$

with  $\mathbf{y}_k$  the  $(n_v \times 1)$  measurement vector at position  $k$ ,  $\mathbf{S}$  is the  $(n_v \times n_S)$  SWT definition matrix, which is independent of position,  $\mathbf{x}_k$  is the  $(n_S \times 1)$  fractions vector for position  $k$  and  $\mathbf{e}_k$  stands for the error vector at position  $k$ . If this is repeated for all  $(N)$  positions of the grid where the hydrographic properties are measured, one can gather all equations in a similar matrix equation

$$\mathbf{Y} = \mathbf{S} \cdot \mathbf{X} + \mathbf{E}, \quad (4)$$

where  $\mathbf{Y}$  is the  $(n_v \times N)$  measurement matrix,  $\mathbf{S}$  is still the SWT definition matrix,  $\mathbf{X}$  is the  $(n_S \times N)$  fractions matrix and  $\mathbf{E}$  stands for the noise matrix. At this point  $\mathbf{X}$  contains the unknown fractions to be estimated and there is still one set of fractions for every grid point as mentioned in comment (A). To make the model more robust, the discrete fraction field of each SWT (i.e. each row of  $\mathbf{X}$ ) will from now on be described by a continuous function. We consider the situation where the samples lay on a two-dimensional (e.g. latitude-depth) grid, hence two-dimensional functions are used. So, instead of having values for the fractions of a given SWT (say SWT  $s$ ) on a discrete grid of points, we now consider a two-dimensional fraction function, which is function of two position coordinates (e.g. latitude  $l$  and depth  $d$ ). In short, the fractions of SWTs are parameterized by a number  $(n_B)$  of two-dimensional spline basis functions  $B$  [Dierckx, 1995]. Formally, this means that the mixing contribution of SWT  $s$ ,  $x_s$ , at position  $(l, d)$  is modeled by a linear combination of  $n_B$  basis functions evaluated in  $(l, d)$

$$x_s(l, d) = \sum_{b=1}^{n_B} c_{s,b} B_b(l, d). \quad (5)$$

In practice, the two-dimensional functions  $B$  are formed by combinations of two one-dimensional B-spline functions [Vanlanduit, 2001]

$$B_b(l, d) = N_{i,k_i}(l)M_{j,k_d}(d), \text{ with } i \in \{1, \dots, n_l\} \\ \text{and } j \in \{1, \dots, n_d\}. \quad (6)$$

$N_{i,k_i}(l)$  is the  $i^{\text{th}}$  (out of  $n_l$ ) B-spline, with order  $k_i$ , in the latitude direction evaluated in  $l$ . Similarly,  $M_{j,k_d}(d)$  is the  $j^{\text{th}}$  (out of  $n_d$ ) B-spline, with order  $k_d$ , in the depth direction evaluated in  $d$ . These B-splines are defined by knots (fixing their number  $n_l$  and  $n_d$ ) and by an order ( $k_i$  and  $k_d$ ).  $n_l n_d$  is the total number of basis functions used to parameterise the mixing fractions, i.e.  $n_B = n_l n_d$ . We used the basic configuration where the knots defining the transition between B-splines are ordered equidistantly over the study area, but in principle their position can be optimized to suit specific features of the investigated system. More details about the choice of the numbers of splines ( $n_l$  and  $n_d$ ) are given in the next section.

[13] Combined in matrices

$$\mathbf{X} = \mathbf{C} \cdot \mathbf{B}^T, \quad (7)$$

which reformulates equation (4) to

$$\mathbf{Y} = \mathbf{S} \cdot \mathbf{C} \cdot \mathbf{B}^T + \mathbf{E}. \quad (8)$$

In equations (7) and (8),  $\mathbf{C}$  represents the  $(n_S \times n_B)$  spline coefficients matrix,  $\mathbf{B}$  is a  $(N \times n_B)$  basis function matrix and  $T$  stands for the transpose. Now, the unknown parameters to be estimated are the spline coefficients contained in  $\mathbf{C}$ . More details about how to optimize these coefficients are given in sections 2.3–2.4.

[14] The approach essentially consists of a parameterization of the original OMP equations. Therefore we propose to call it Parametric OMP analysis or POMP. A number of remarks can be made.

[15] 1. By parameterizing the source fractions, they are made explicitly function of position (see equation (5)), thus addressing imperfection 1. In other words, truly a continuous fraction field is estimated, and not a discrete fraction grid. However, it is always possible to get the values of the mixing fractions on the sampling grid, using equation (5). Note also that to visualize the discrete results of classical OMP analysis in a contour plot, as is usually done, it is anyway necessary to fit a function through the fraction values in order to interpolate and form the contour lines. POMP performs this parameterization a step earlier.

[16] 2. Generally the number of basis functions  $n_B$  will be much smaller than the number of sampling positions  $N$ . Hence the number of unknowns (spline coefficients  $\mathbf{C}$ ) to be estimated, namely  $n_B n_S$ , is much smaller than the original number of fractions that had to be estimated, that is  $N n_S$ . This implies there will be a certain ‘‘smoothing’’, comparable to what happens in classical linear regression, which reduces the influence of individual noise contributions and makes the resulting mixing fraction estimates more robust. On the other hand, too much smoothing should be avoided, otherwise risking to discard significant features present in the data. Making this trade-off is an inclusive part of the

POMP methodology and it will be further discussed in section 2.2.

[17] 3. POMP possibly creates an opening to objectively optimize the SWT characteristics  $\mathbf{S}$ . Thanks to the reduction in unknowns to be estimated (see consequence 2), enough residual degrees of freedom are left over to extract additional information from the data, possibly about the elements of matrix  $\mathbf{S}$ . This is still an optimization challenge and therefore will not further be discussed in this paper. Some recent work on the characterization of SWTs has been done in the framework of OMP analysis applied to time series [Henry-Edwards and Tomczak, 2006].

[18] 4. The main assumption underlying the POMP method is that the mixing contributions systematically vary with position. This implicitly constrains the possible values of the mixing fractions and therefore reduces their sensitivity to individual measurement errors. Pointwise variation is interpreted as an insignificant random effect which should be removed. So, clearly POMP analysis is not suitable to study small-scale (grid scale) phenomena. Exactly how much “detail” will be provided by the POMP analysis depends on the particular function used to describe the mixing patterns. This issue is addressed in the next section.

## 2.2. Selection of Function Complexity

[19] The quality of the POMP analysis outcome obviously depends on the choices made concerning the functions parameterizing the mixing fractions. Spline functions were chosen to describe the fraction fields, because they are very flexible in shape and stable to compute [Dierckx, 1995]. But an essential choice to be made, regardless of the type of functions used, concerns the complexity of the functions. Indeed, if a too simple function (or model) is chosen, some significant features present in the data will not be captured, so the results will be biased although they are robust. Conversely, if a too complex model is used (like OMP analysis in most occasions), not only the significant underlying processes are modeled, but a part of the stochastic noise as well. As a consequence, such a model will closely fit the data but is highly variable due to the noise’s influence. Summarizing, a model should have an “intermediate” complexity, such that it optimally balances accuracy and precision.

[20] In the case of spline functions, this boils down to carefully selecting the numbers ( $n_l$  and  $n_d$ ) and orders ( $k_l$  and  $k_d$ ) of the B-splines, which is comparable to choosing the best order in polynomial regression, or the number of variables to consider in multivariate regression. All these models are purely empirical (i.e. they do not have a physical interpretation) and the most suitable model complexity only depends on the data they describe. The numbers ( $n_l$  and  $n_d$ ) and orders ( $k_l$  and  $k_d$ ) of spline functions can be chosen by performing the POMP analysis for all relevant combinations ( $n_l$ ,  $n_d$ ,  $k_l$ ,  $k_d$ ) and inspecting the results according to statistical rules designed to select an optimal model complexity [de Brauwere et al., 2005; De Ridder et al., 2005, and references therein]. The only restriction is that  $n_B \leq N$ . The extreme situation when  $n_B = N$  represents the special case that the POMP model reduces to the original OMP solution, where the mixing fractions are estimated for each sampling position separately. Since it is not always obvious whether all assumptions underlying the statistical tests (e.g. normality of noise distributions) are satisfied with this kind

of field data, we suggest to apply a number of these tests and to use them together as a guide, in combination with common sense.

[21] The goal of the parameterization is to make the resulting mixing fractions more robust, by only cutting off that part of the variability which is due to noise. Depending on the sampling density and the noise level, this “filtering” will also cut away a certain portion of some systematic, but small-scale, processes occurring (e.g. eddies). If these processes are of particular interest, they can be further investigated in the residual plots. In this perspective, the aim of a POMP analysis (combined with a statistically based selection of the splines complexity) is to only include those features in the mixing model which are most likely to be significant, while all the remaining information will be in the residuals. With classical OMP, almost everything is packed in the model.

[22] The complexity of the splines functions, or the degree of smoothing, could also completely be interpreted in terms of ocean dynamics. If some processes are associated with specific scales, they could possibly be isolated using a POMP-like technique. The remaining processes could then be visualized by subtracting these results from those obtained with classical OMP analysis. This remark is only meant as a hint for future applications, in this article the selection of the splines complexity is based on the statistical criteria.

[23] To summarize, parameterization of the unknown mixing fractions, combined with an effective choice of the splines complexity improves imperfections 1 and 2 mentioned in the introduction. How to handle remark 3 will be discussed in the next sections.

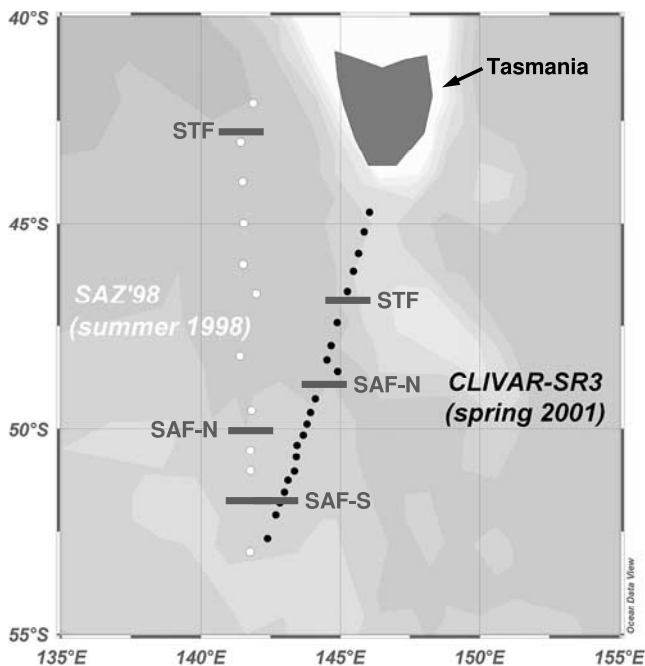
## 2.3. Constraints

[24] If the mass balance equation (2) is not included in the mixing model, this has the advantage that the weighting can be performed following the same logic for all equations (see remark 3 in the introduction). On the other hand, the optimized fractions must satisfy mass balance, otherwise they are not realistic. Both requirements can be assured when the mass balance equation is imposed as an equality constraint during the optimization. Generally in OMP analysis, an optimization procedure is used that constrains the fractions to be positive (e.g. nonnegative least-squares algorithm “lsqnonneg” in MATLAB [OMP2, 2005]). The mass balance imposes an additional constraint that can be inserted in the optimization algorithm (e.g. using the “lsqin” MATLAB function). More details about the algorithm are given in Appendix A. The main consequences of this approach are as follows:

[25] 1. Mass balance will be exactly (i.e. within computation precision) satisfied. In other words, only the mixing equations will have residuals. This is a healthy situation, because the mass balance equation does not represent the modeling of a measurement and is therefore not affected by measurement errors or environmental variability.

[26] 2. The only equations really optimized are the mixing equations, so a consistent weighting scheme can be applied to all equations (see section 2.4).

[27] Apart from these positive changes, the method and results are hardly influenced by treating mass conservation this way. For instance, the condition for mathematical solu-



**Figure 1.** Map of the study area with the positions of the stations considered in this study. The stations from the CLIVAR-SR3 cruise in spring 2001 are shown as black bullets; those from the SAZ'98 campaign in summer 1998 are shown as white bullets. The specific front positions (STF, Subtropical Front; SAF-N and SAF-S, northern and southern branches of the Subantarctic Front) for the two cruises are indicated in gray along the respective cruise tracks. The figure was elaborated using Ocean Data View [Schlitzer, 2004].

tion that  $n_s \leq n_v + 1$  still holds. Also, the resulting fractions will not differ much from those found with the original procedure, especially when the weight attributed to the mass balance equation was very large, as is the custom procedure. This confirms the observation of Tomczak and Large [1989]: “Trial runs with (arbitrary chosen) larger weights for mass conservation show that the OMP analysis converges rapidly to nearly perfect mass conservation, with only minor adjustments in the mixing ratios”. To summarize, imposing the mass conservation as a strict constraint does not influence the results much. It is rather a technical issue allowing to attribute weights to all equations in a more consistent way. This issue is further discussed in the next section.

#### 2.4. Weighting

[28] The final values for the mixing fractions can be greatly influenced by weighting the equations before optimization. For instance, temperature and salinity measurements are usually associated with lower relative uncertainty than the other variables. This knowledge should be incorporated in the method such that these measurements are treated with greater importance. This is achieved by multiplying every equation with a weight symbolizing the reliability of the variable in question. Since this is not possible for a mass conservation equation (uncertainty is zero?), it was suggested in the previous section to treat this equation as a constraint rather than as an equation to be optimized.

[29] In this paper we choose to use a weighting scheme compatible with the statistical Weighted Least Squares (WLS) framework [Pintelon and Schoukens, 2001]. This implies that the weights are the inverses of the uncertainty (in terms of standard deviation) associated with each equation. In other words, optimal values for the unknown spline coefficients  $\mathbf{C}$  (and hence the unknown mixing fractions) can be found by minimizing the squared residuals multiplied by a weighting factor ( $w$ ) which is the inverse of the standard deviation ( $std$ )

$$\sum_{i=1}^N \sum_{j=1}^{n_v} (w_{ij} (\mathbf{S}_j \mathbf{c} \mathbf{B}_i^T - \mathbf{Y}_{ij}))^2 = \sum_{i=1}^N \sum_{j=1}^{n_v} \left( \frac{\mathbf{S}_j \mathbf{c} \mathbf{B}_i^T}{std_{ij}} - \frac{\mathbf{Y}_{ij}}{std_{ij}} \right)^2. \quad (9)$$

Here  $\mathbf{S}_j$  stands for the  $j$ th row of  $\mathbf{S}$ ,  $\mathbf{B}_i$  is the  $i$ th row of  $\mathbf{B}$  and  $w_{ij}$  represents the weighting for the  $i$ th sampling point for the  $j$ th hydrographic property. In equation (9) it can also be seen that the weighting simultaneously serves to non-dimensionalize the variables.

[30] In the most general case, the weighting factor varies with the property considered (e.g. a different weight for temperature than for salinity) and the location (e.g. a different weight for measurements made at the surface than for the deep waters). For simplicity and conform to common OMP practice, one weight per variable will be attributed (no location dependency).

[31] In practice, the difficulty remains to quantify the uncertainties to be used as weights. An accurate estimate of the standard deviation associated with each variable should be known. Ideally, this uncertainty is assessed based on the standard deviation of a number of replicates. However, this information is not always available because it is expensive and time-consuming to repeat experiments. In section 3.3.1 a possible strategy to retrieve uncertainty information is proposed.

### 3. Results and Discussion

[32] In the previous sections a refinement is proposed for the classical OMP analysis. In the next sections, this adapted procedure is applied to two Southern Ocean data sets, sampled along closely sections but during different seasons and years (spring 2001 and summer 1998). The benefit of this exercise is twofold: (1) the POMP method is illustrated and can prove to be trustworthy, and (2) the seasonal variability of the SWTs and their mixing patterns can be investigated. The structure of the next sections is as follows. First the data sets (3.1) and the hydrological setting (3.2) are discussed. Next, the POMP method is applied to the spring 2001 data (3.3) and the summer 1998 data using the same SWT definitions (3.4). The results from this analysis are used in a final step to formulate better SWT characteristics for the summer 1998 situation, enabling some discussion on the variability of the SWTs and the mixing patterns (3.5).

#### 3.1. Data Sets

[33] The POMP method described above is applied to two data sets sampled along the WOCE SR3 line at  $\sim 142^\circ\text{E}$  in the Australian sector of the Southern Ocean during two cruises with the R/V *Aurora Australis*: (1) the CLIVAR-SR3 (October – December 2001) and (2) the SAZ'98 (February – April 1998) cruises. The first data set represents the situation of late winter–early spring 2001 and the second

**Table 1.** Source Water Type (SWT) Characteristics and Total Uncertainties (Expressed as Standard Deviation, std) Used for the Weighting in All Analyses<sup>a</sup>

		Pot. Temperature, °C	Salinity	PO, $\mu\text{M}$	NO, $\mu\text{M}$
AAIW	Spring 2001	4.07	34.355	560	490
	Summer 1998	3.8	34.25	610	519
	Lit. values	3.4 – 5.6 <sup>b</sup> 4 – 5 <sup>c</sup>	< 34.4 <sup>b</sup> 34.35 – 34.44 <sup>c</sup>		
SAMW	Spring 2001	9.17	34.655	434	399
	Summer 1998	8.6	34.58	468	422
	Lit. values	$\pm 9^{\text{d}}$ 8.8 – 9.5 <sup>b</sup> 8.75 – 9.25 <sup>c</sup>	34.58 – 34.73 <sup>b</sup> 34.60 – 34.70 <sup>c</sup>		
STSW	Spring 2001	13.28	35.34	318	297
	Summer 1998	12.94	34.71	388	345
	Lit. values	11 – 16.25 <sup>c</sup>	35 – 35.30 <sup>c</sup>		
CSW	Spring 2001	3.3	33.796	609	562
	Summer 1998	4.6	33.60	602	550
	Lit. values				
UCDW	Spring 2001	2.52	34.56	555	486
	Summer 1998	2.275	34.661	565.6	488.46
	Lit. values	>0.5°C <sup>f</sup> 2.6 <sup>b</sup>	34.40 – 34.70 <sup>f</sup> 34.585 <sup>b</sup>	Nutrient max. <sup>f</sup>	Nutrient max. <sup>f</sup>
Weight = 1/std		1/0.07	1/0.011	1/9	1/8

<sup>a</sup>The SWT characteristics are given for the spring 2001 and summer 1998 situations, and for ease of comparison some values from literature are shown. AAIW, Antarctic Intermediate Water; SAMW, Subantarctic Mode Water; STSW, Subtropical Surface Water; CSW, Circumpolar Surface Water; UCDW, Upper Circumpolar Deep Water. Derivation of the values is explained in the text.

<sup>b</sup>Rintoul and Bullister [1999].

<sup>c</sup>Rintoul and Sokolov [2001].

<sup>d</sup>Rintoul et al. [1997].

<sup>e</sup>Sokolov and Rintoul [2002].

<sup>f</sup>Orsi et al. [1995].

of summer 1998. We focus on the subantarctic region (northern part of the transect), roughly between 42°S and 53°S. Moreover, particular interest is given to the upper 1000 m water column, because investigating seasonality is mostly relevant in this depth range [Rintoul et al., 1997; Coatsanoan et al., 1999]. Figure 1 shows the study area. Salinity, temperature and nutrient data from CTD casts have been provided by Mark Rosenberg (ACE-CRC, Hobart). More details about the hydrological setting of the respective cruises are given below (section 3.2).

### 3.2. Hydrological Setting

[34] In this section an overview of the hydrological setting of the region is given. Definitions and positions of zones, fronts and water masses are mainly based on potential temperature and salinity characteristics. From north to south along the WOCE SR3 line one can find: the Subtropical Zone (STZ), the Subtropical Front (STF), the Subantarctic Zone (SAZ), the Subantarctic Front (SAF, northern and southern branch) and the Polar Front Zone (PFZ). The fronts along the SR3 section are observed in approximately the same location year to year [Rintoul et al., 1997] but seasonal and interannual variabilities are apparent. The STF is generally reported to be located around 45°S [Rintoul and Bullister, 1999; Sokolov and Rintoul, 2002], with specific locations of 47°S for the CLIVAR cruise in spring 2001 [Jacquet et al., 2004] and 43°S for the SAZ section in summer 1998 [Rintoul and Trull, 2001]. The SAF is the major front in this section. Sokolov and Rintoul [2002] consistently find two branches of the SAF along the SR3 line (SAF-N and SAF-S), at mean latitudes of 50.5°S and 52°S respectively. During spring 2001, the northern branch appeared at 49°S [Jacquet et al., 2004]. In Figure 1 the specific front positions during the two cruises are indicated.

Both spring 2001 and summer 1998 situations were characterized by intense SAF cross-frontal exchange [Rintoul and Trull, 2001].

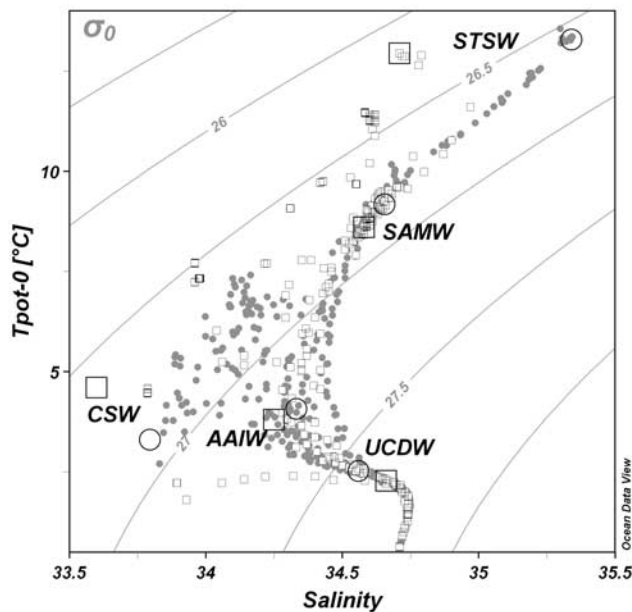
[35] These displacements of front locations are related to changes in water mass formation. Seasonal and interannual changes of water mass properties, alternating cooling or warming, freshening or saltier shifts, have been identified by Aoki et al. [2005a], Chaigneau and Morrow [2002] and Rintoul et al. [1997]. The major water masses in the study area are: the Subtropical Surface Water (STSW), extending north of the STF, the Subantarctic Mode Water (SAMW), located between the STF and the SAF, the Antarctic Intermediate Water (AAIW), subducting between the PFZ and the STF, the Circumpolar Surface Water (CSW) in the PFZ and the Upper Circumpolar Deep Water (UCDW), located deeper in the PFZ.

### 3.3. Spring Data

#### 3.3.1. Input Into POMP

[36] For the mixing analysis, four conservative variables were used: potential temperature ( $\theta$ ), salinity (Sal), PO and NO. PO and NO are tracers based on phosphate and nitrate, but corrected for respiration, according to Broecker [1974]:  $\text{PO} = 170 \text{ PO}_4 + \text{O}_2$  and  $\text{NO} = 9 \text{ NO}_3 + \text{O}_2$ . Redfield ratios are similar to those used in other Southern Ocean studies [You, 2002; You et al., 2003; Lo Monaco et al., 2005].

[37] The source water types relevant for the spring 2001 data are listed in Table 1 and shown in a  $\theta$ -S diagram in Figure 2. They are defined based on the data set itself, usually on the boundaries of the study area. This way, the SWTs represent local water masses, which is justified since the study area is small and we are concerned with local processes. The selected SWTs are in good agreement with values reported in literature, also listed in Table 1.



**Figure 2.** T-S diagram showing all data and the SWT definitions for both data sets. Spring 2001 is represented by circles: small gray dots for the data and large open circles for the sources (see also Table 1). Summer 1998 features are shown as squares: small squares for the data and large open squares for the source characteristics (see also Table 1). For the sake of clarity the data from all depths (i.e. also deeper than 1000 m) are shown. The figure was elaborated using Ocean Data View [Schlitzer, 2004].

[38] To parameterize the fraction fields, two-dimensional splines were used. As mentioned above, it is necessary to determine the optimal complexity of the spline function. For this spring data set, first order splines appeared to deliver the best results (lowest residuals for the same number of optimized parameters). The optimal number of B-splines in latitude and depth directions ( $n_l$  and  $n_d$  in equation (6)) was determined based on the MDL<sub>s</sub> (Minimum Description Length for small samples) criterion [De Ridder et al., 2005]. The optimal complexity appeared to be 10 B-splines in latitude direction and 4 in depth direction.

[39] The weight for each variable is given in Table 1. PO and NO weights are based on analytical precision information obtained from M. Rosenberg et al. (unpublished report). The values for temperature and salinity were taken larger than the analytical precision, because preliminary tests showed that just taking into account analytical precision greatly underestimates the total variability associated with these data. To assess these overall variabilities of temperature and salinity, raw CTD data (downcast and upcast) were used. For several depth horizons the standard deviation for temperature and salinity measurements within a 2 meter water layer was computed and the average of these numbers is given in Table 1. This approach implies that the water within a 2 meter layer is homogeneous, or at least reflects the same (local) water mass. It is also assumed that the variability is similar throughout the water column (because we take one averaged standard deviation), but this only confirms the earlier decision to take one weight per variable and not to take any location dependency into

account (see section 2.4). In this case, the standard deviation of the water characteristics within this layer is a better measure of the total variability than the analytical precision. Chaigneau and Morrow [2002] use a similar procedure but using latitudinal neighboring data to estimate sea surface temperature and salinity precisions (giving resp.  $\sim 0.01^\circ\text{C}$  and  $\sim 0.02$ ).

### 3.3.2. Results From POMP Analysis: Mixing Fractions

[40] The mixing fractions estimated by the POMP method are shown in Figure 3. To facilitate interpretation, sampling depths, density isolines and front positions are indicated as well. The five selected water masses are clearly significant in the investigated spatial window.

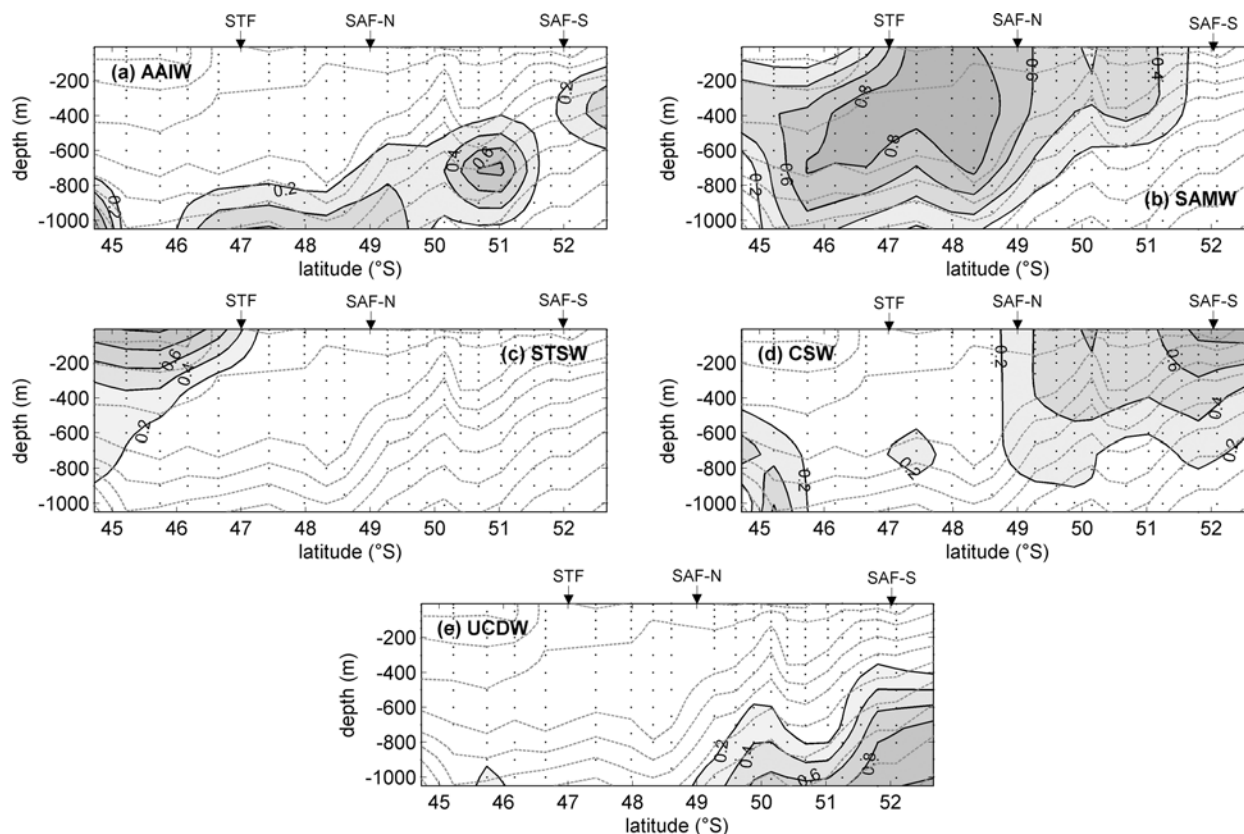
[41] The Antarctic Intermediate Water (AAIW) appears between 200 and 400 m at  $52^\circ\text{S}$  and sinks northward below the SAMW and reaches depths greater than 1000 m north of the STF ( $\sim 47^\circ\text{S}$ ). Its apparent extension at 1000 m in the northern part of the transect ( $\sim 45^\circ\text{E}$ ) could reflect the presence of another variety of AAIW [Rintoul and Bullister, 1999]: a “northern” type of Tasman Sea origin. The fact that the density isolines rise at this position supports this hypothesis. Similarly, the discontinuity of the AAIW branch between  $51$  and  $52^\circ\text{S}$  probably reflects that the water immediately south of the SAF does not have the same origin as the low salinity core north of the SAF. Instead, the AAIW is supplied by a number of water types upstream and by the time they enter the studied area, these waters have developed distinct signatures [Rintoul and Bullister, 1999]. This could explain why one AAIW definition cannot perfectly account for all intermediate water samples in the investigated data set.

[42] As reported by Sokolov and Rintoul [2002], the Subantarctic Mode Water (SAMW) appears as the principal water mass extending north of the northern branch of the Subantarctic Front (SAF-N;  $\sim 49^\circ\text{S}$ ) in the upper 700 m water column, a zone characterized by extremely homogeneous density. Between 700 m and 1000 m, SAMW still has a noticeable contribution, but it mixes with AAIW. The SAMW core between the STF and SAF-N reaches the surface, which illustrates the deep winter mixing in this zone, as noticed by Rintoul and Trull [2001]. North of the STF, SAMW does not reach the surface anymore, its contribution is concentrated between 300 and 800 m. Like the AAIW, the SAMW does not form a single homogeneous water mass [Rintoul and Bullister, 1999]; it results in the large extension of SAMW observed here.

[43] The surface waters north of the STF ( $\sim 47^\circ\text{S}$ ) form a homogeneous water mass, the Subtropical Surface Water (STSW). It corresponds to the Subtropical Lower Water identified by Sokolov and Rintoul [2002]. STSW appears as a compact water mass up to 250 m, lying on top of and slightly mixing with SAMW.

[44] At the surface of the most southern part of the window, a fourth surface water is significant: the Circumpolar Surface Water (CSW). Its core lies at  $\sim 52^\circ\text{S}$  and its contribution extends to 500–600 m in depth and to SAF-N ( $\sim 49^\circ\text{S}$ ) in latitude. As for AAIW an unexpected contribution in the north is observed, which is probably due to the poor resolution in this zone, as mentioned before.

[45] Finally, the southern part of the section undergoes significant contributions of the Upper Circumpolar Deep Water (UCDW) between 800 and 1000 m. This deep water



**Figure 3.** Mixing fractions associated with (a) Antarctic Intermediate Water, (b) Subantarctic Mode Water, (c) Subtropical Surface Water, (d) Circumpolar Surface Water and (e) Upper Circumpolar Deep Water, estimated using the POMP method for the spring 2001 situation. Density isolines are plotted as dotted gray lines and range from 26.6 (upper north corner) to 27.6 (lower south corner) with increments of 0.1. Dots represent the sampling depths.

is normally found between 1500 and 2500 m north of the SAF, and rises to depths of 1000 m just south of the SAF-N and 500 m south of the SAF-S [Sokolov and Rintoul, 2002], coinciding with the reconstruction shown in Figure 3.

### 3.3.3. Uncertainties

[46] To quantitatively interpret the mixing distributions, it is necessary to quantify the variability of the estimated mixing coefficients due to measurement errors (these latter can be found in Table 1). One thousand Monte-Carlo simulations were performed, where in each simulation the original data are perturbed with a random number sampled from the measurement noise distribution [Coatanoan *et al.*, 1999]. The uncertainty of the mixing coefficients is different for each source: 3.5% for AAIW, 1.7% for SAMW, 0.8% for STSW, 1.6% for CSW and 1.9% for UCDW (average values of 1 standard deviation). Note that this uncertainty is much smaller (approximately a factor 10) than the variability if no positivity constraints would be applied. This means that the constraints are active and reduce the space of possible mixing fractions, resulting in a smaller variability.

### 3.3.4. Residuals

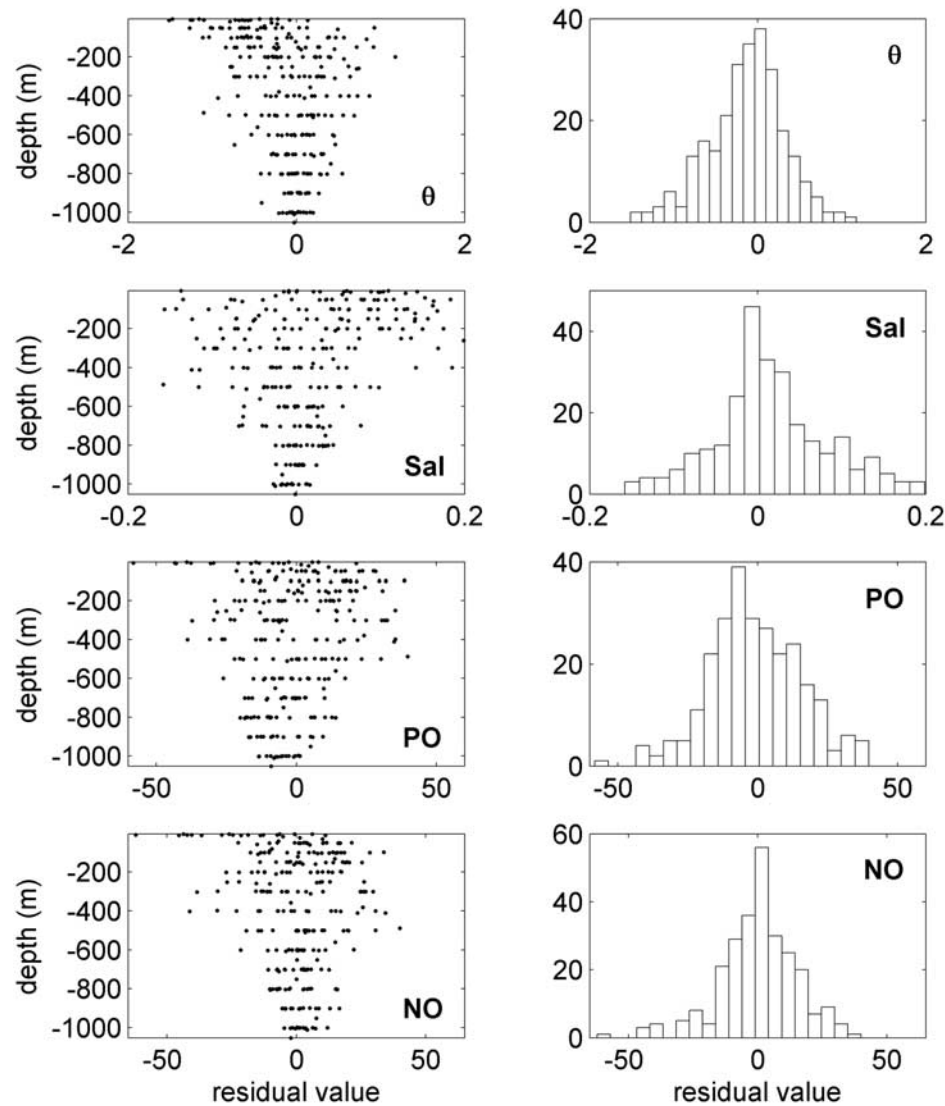
[47] The reliability of the estimated mixing fraction fields can be assessed by examining the residuals. Notice that the residuals are defined as the difference between model reconstruction and measurement (see expression (9)). The

residuals for this spring 2001 analysis are presented in Figure 4. In the figures on the left, showing the residuals as a function of depth, it is clear that the residuals are higher in the upper waters, which is probably an indication that the overall uncertainty associated with a property is also dependent of depth. But the histograms on the right show that on the whole the residuals behave as normally distributed variables: symmetric and centered around zero. This suggests that they only represent random stochastic noise (including features of ocean physics such as eddies and other products of geostrophic turbulence that can be treated as random).

### 3.3.5. Comparison With Classical OMP Analysis

[48] As an illustration of the different performance of the POMP approach, a classical OMP analysis is performed using the same observations and SWT definitions. That is, the fractions are estimated directly from equations (2) and (3), instead of parameterizing them first and estimating the spline coefficients from equation (8). The same weighting scheme and contour visualization is applied as for the POMP results (see equation (9)), so any differences are due to the parameterization.

[49] In Figure 5 the resulting mixing fraction fields are shown. The water masses' average mixing pattern corresponds to the results from the POMP method (see Figure 3). However, the distributions clearly are much more "fitful",

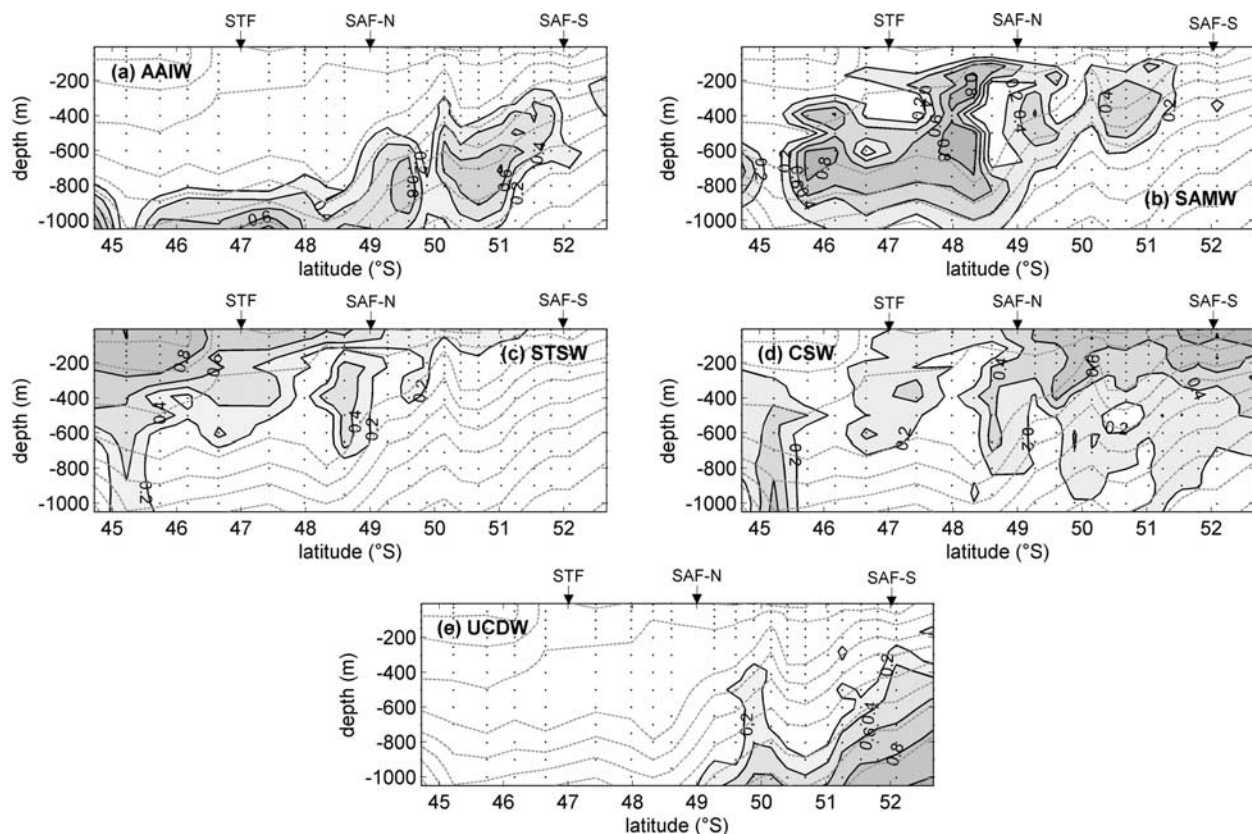


**Figure 4.** Residuals associated with results presented in Figure 3, shown (left) as function of depth and (right) as histograms (right column). Dimension of residuals is  $^{\circ}\text{C}$  for potential temperature and  $\mu\text{M}$  for PO and NO (see Table 1). Dots represent the sampling depths.

due to the fact that the mixing fractions are optimized for every point independently. This approach allows the noise inevitably present in the measurements to fully propagate in the final mixing fractions (see remark 2 in introduction). This can lead to locally unrealistic features such as the “hole” present in the SAMW fractions around  $49^{\circ}\text{S}$ . To provide additional evidence to evaluate the POMP versus OMP results, density isolines were added on the plots in Figures 3 and 5. Most, although not all, features present in the density pattern are recovered in the POMP fractions. Furthermore, the SAMW nicely fills its characteristic pycnostad, whereas from OMP analysis it would seem that CSW and STSW also contribute in this area. This is probably the result of the pointwise strategy allowing discontinuous mixing distributions. Conversely, the mixing distributions provided by the POMP procedure are modeled as a function of position and smoothed in a natural way because the effect of the noise on the individual points is

suppressed by the parameterization. It is important to stress that if the complexity of the spline functions is chosen accurately, all significant features that can be described by linear mixing will be represented in the final mixing fractions. In other words, the fact that the fraction fields calculated by the POMP method are smoother does not mean that significant information has been lost. However, some small-scale features (e.g. eddies) can obviously only be captured if the sampling density is high enough, which is also true for the classical OMP analysis.

[50] The residuals for the classical OMP analysis are not shown, but they are much smaller than those shown in Figure 4. Small residuals are not necessarily the result of an accurate modeling, but could also be due to overmodeling (see linear regression through two data points). To make a real quantitative statement about the residual level which is acceptable, a more profound uncertainty analysis would be required, taking into account all aspects of environmental



**Figure 5.** Mixing fractions estimated using the classical OMP method for the spring 2001 situation. Density isolines are plotted as dotted gray lines and range from 26.6 (upper north corner) to 27.6 (lower south corner) with increments of 0.1. Dots represent the sampling depths.

and SWT variability in addition to mere measurement uncertainties, which is beyond the scope of this paper.

[51] An additional consequence of overmodeling is that the mixing fractions are much more sensitive to stochastic errors. This is illustrated by the uncertainties estimated by Monte-Carlo simulations, as described in section 3.3.3. The uncertainties of the mixing fractions derived using the classical OMP analysis are much higher than with the POMP method, namely: 10% (versus 3.5%, see section 3.3.3) for AAIW, 14% (versus 1.7%) for SAMW, 9% (versus 0.8%) for STSW, 10% (versus 1.6%) for CSW and 6% (versus 1.9%) for UCDW.

[52] This comparison demonstrates the performance of the POMP method. The derived mixing distributions (Figure 3) correspond to what is expected from literature and are more reliable (robust) than those estimated using the classical OMP analysis. The residual analysis confirms that no systematic errors are present. In the next sections, we will further apply the POMP method to the summer 1998 data set, to investigate the seasonal and interannual variability of the SWT characteristics and of their mixing fractions.

### 3.4. Summer Data Modeled With Spring Sources

#### 3.4.1. Input Into POMP

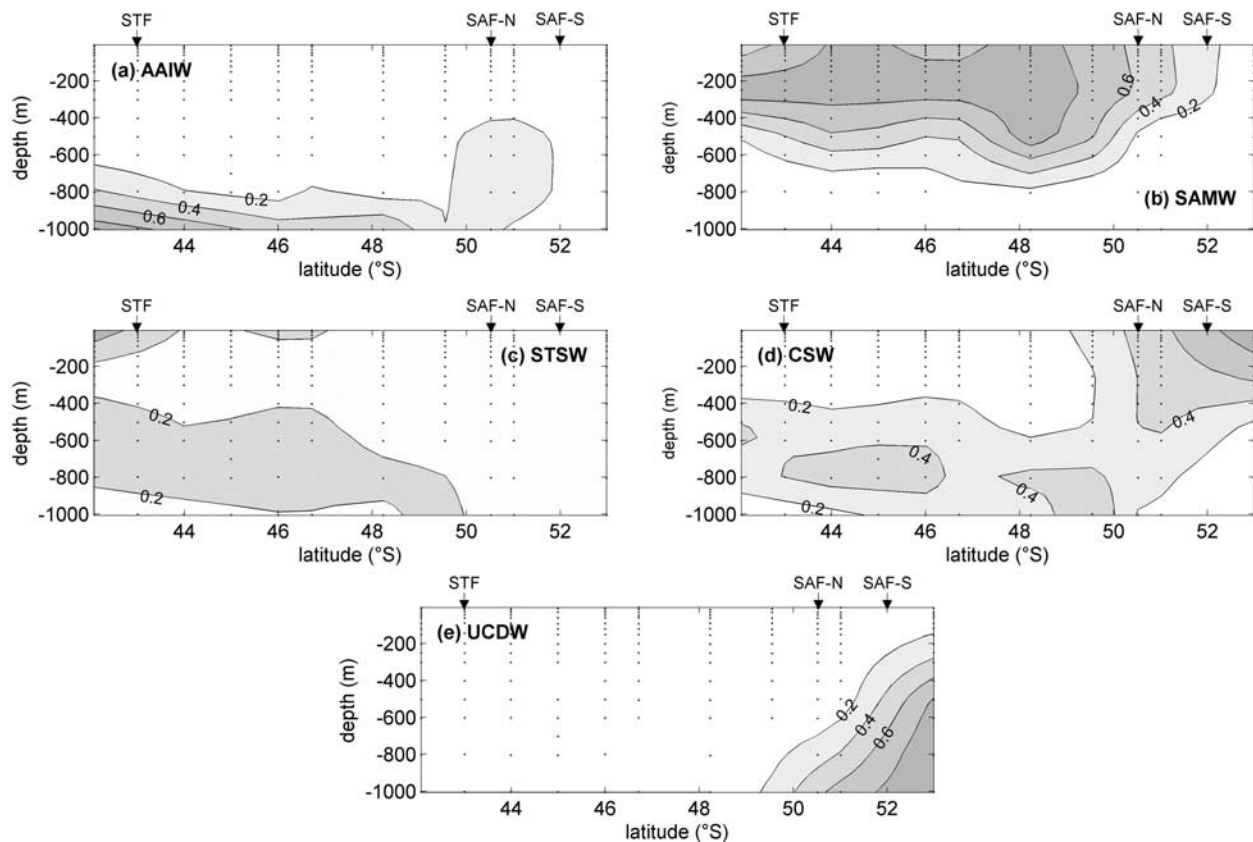
[53] In this analysis, the same variables (potential temperature, salinity, PO and NO) and the same SWTs (Table 1) are used as for the spring situation (section 3.3). Since the summer data are principally sampled along the same tran-

sect, but at a different season (and year), this is an interesting test. A priori it is an open question whether, based on existing hydrographic knowledge, the reconstructed mixing patterns will be realistic. Analysis of the residuals will be an additional tool to assess the quality of the reconstruction.

[54] This summer 1998 data set has a lower resolution in the latitude axis than the spring 2001 data (11 versus 21 stations). Therefore, we cannot use the same spline complexity as for the spring 2001 data. The optimal spline complexity (number of splines) was investigated in the same way as for the spring 2001 data, using the  $MDL_s$  criterion [De Ridder *et al.*, 2005]. Because it seems reasonable to consider that the specific summer 1998 SWTs (see section 3.5) are closer to the true situation, these SWTs were used for the complexity determination. The optimal complexity was 6 splines in the latitude direction and 5 in the depth direction. The same weighting as for the spring data was used (Table 1).

#### 3.4.2. Results From POMP Analysis: Mixing Fractions and Uncertainties

[55] The reconstructed mixing fractions estimated by the POMP method for the summer 1998 data, using the spring 2001 SWTs, are shown in Figure 6. The most striking features are the distributions of the surface waters (STSW and CSW) and of SAMW which extend respectively much deeper and much wider than is reasonably expected. Our interpretation is that these waters are very sensitive to



**Figure 6.** Mixing fractions found by POMP analysis for the summer 1998 situation estimated using the SWTs defined for the spring 2001 data (Table 1). Abbreviations are the same as in Figures 3 and 5. Dots represent the sampling depths.

seasonal variations like temperature and salinity changes; it is particularly true for mode waters [Rintoul and Bullister, 1999]. By using spring (2001) definitions for these SWTs, they are forced to explain deeper zones (AAIW). Except for UCDW, which is expected to be less affected by seasonal variations, all SWTs exhibit an unrealistic mixing pattern. These results emphasize the need to define local (both in space and in time) SWTs in OMP-like analyses, especially when studying the upper water column.

### 3.4.3. Residuals

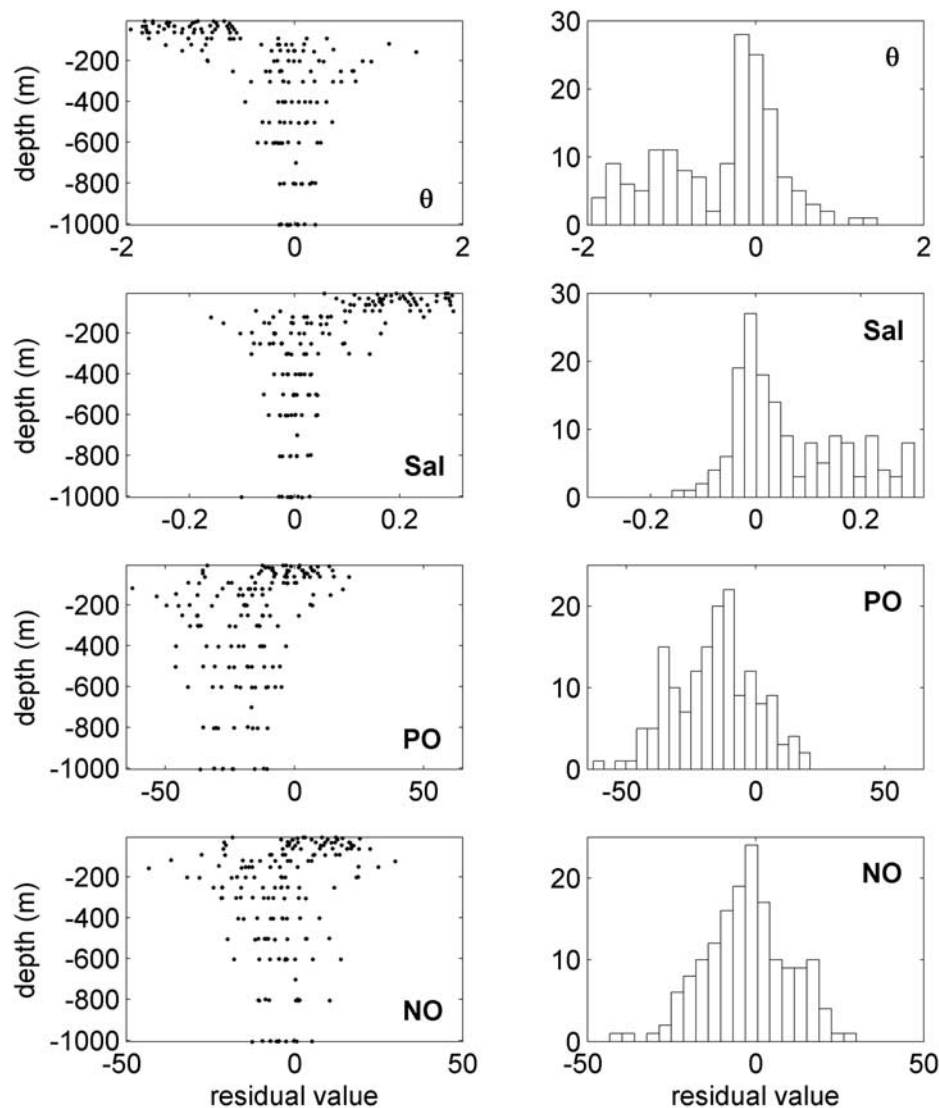
[56] Beside the suspect fraction fields, also the residuals indicate that the model is not able to accurately describe the data. Figure 7 displays the residuals as function of depth and as histograms. It is clear that the residuals are not random. Especially for temperature and salinity there is a considerable shift in the residuals for the surface waters. For temperature the surface residuals are systematically negative, whereas for salinity they are positive. This signifies that the (surface) SWT definition is too cold and too saline, which is what one could reasonably expect when modeling summer data with spring SWTs. The deeper waters seem to be relatively well defined for temperature and salinity, since the residuals in this zone are well centered around zero. However, PO exhibits an opposite pattern: the surface residuals are centered around zero, but the deeper waters are enriched in PO compared to the model's reconstruction. Also, the general distribution of PO residuals (see histogram in Figure 7) is clearly shifted towards more negative values,

while this is not so much the case for the NO residuals (see discussion below, section 3.5.1).

[57] In the next section, the SWTs are redefined in an attempt to improve the data description. This exercise should allow for some more insight into seasonal and interannual changes in water mass characteristics and their distribution.

### 3.5. Summer Data Modeled With Specific Seasonal Sources

[58] In the previous section, we attempted to model the summer 1998 data using spring 2001 SWTs. This approach does not seem to provide satisfactory results (see discussion in 3.4.2). In this section, we will try to improve the estimated mixing patterns by re-defining the SWTs for this specific summer 1998 situation. If a better (i.e. realistic) reconstruction can be obtained, the changes in SWT characteristics can be interpreted as the seasonal and interannual variability of these water masses. Also, the mixing patterns may show some differences compared to the spring 2001 situation. At this point our approach diverges from the one followed by Coatanoan *et al.* [1999]. They performed a seasonal analysis of the water mass distribution in the eastern Indian Ocean, but used the same SWT definitions for both investigated seasons. In the current study, it is clear that this practice does not provide satisfactory results. This is probably linked to the fact that local SWTs were selected, which are subject to seasonal and interannual variations.



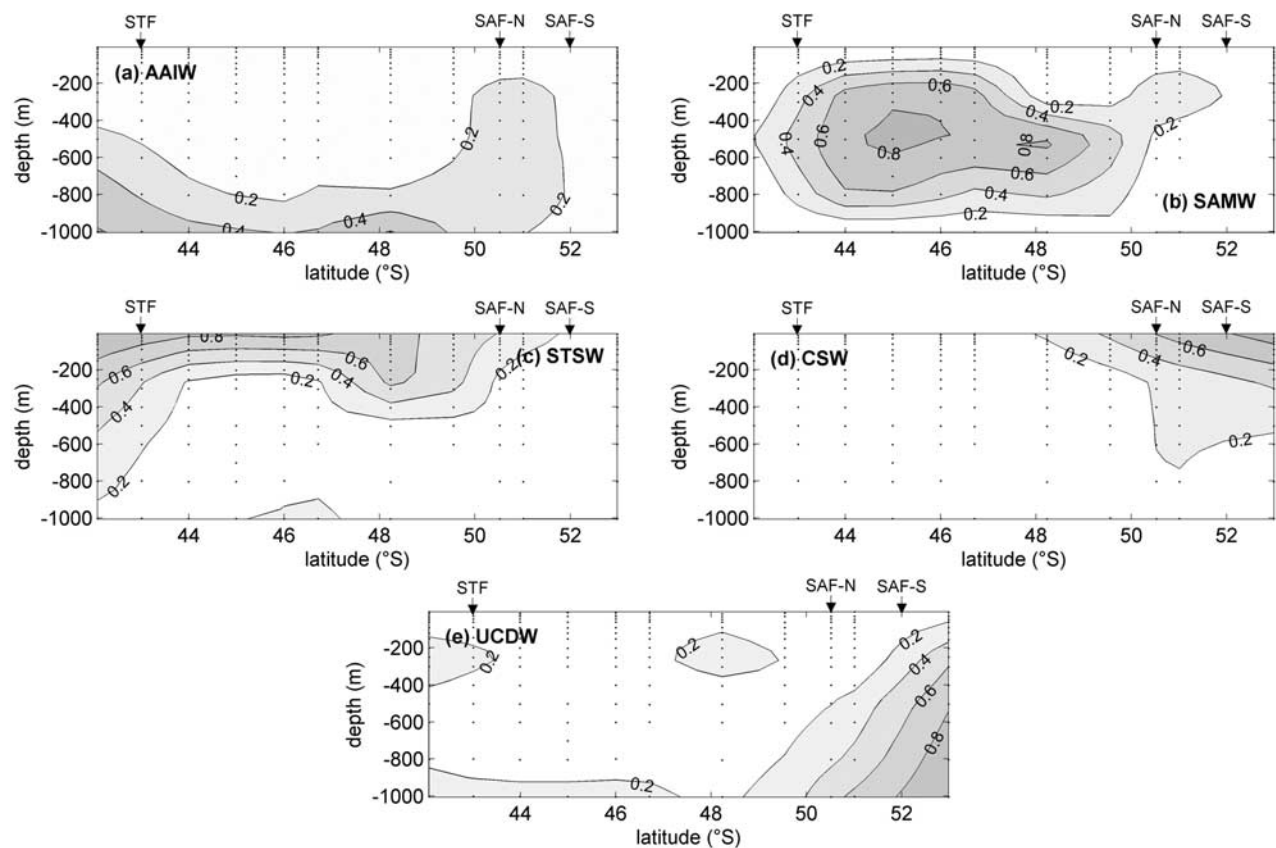
**Figure 7.** Residuals associated with Figure 6, shown (left) as a function of depth and (right) as histograms for all variables. Residuals have same dimensions as the respective properties (see Table 1).

Furthermore, it is expected that seasonal changes are most significant in the surface waters whereas *Coatanoan et al.* [1999] excluded the upper 200 m from their analysis.

### 3.5.1. Input Into POMP: New SWT Definition

[59] The same data as in section 3.4, but different SWTs are used in this analysis. The new SWTs defined for the summer 1998 situation are given in Table 1 and Figure 2. Similarly to the spring 2001 SWTs, summer 1998 SWTs are selected based on the data. Within the acceptable values based on literature, the characteristics of some SWTs were further optimized if this provided a substantially better fit (lower residuals), which means that some SWTs do not exactly correspond to one data point (see Figure 2). The new definitions are again in relatively good agreement with the values reported in literature (Table 1), indicating that this change in characteristics falls within the expected seasonal and/or interannual variability. All SWTs, except UCDW, appear to be fresher in summer 1998 than in spring 2001 (Figure 2). This is expected, especially for the surface

waters, due to the influence of sea ice extent and ice melt [Chaigneau et al., 2004]. In addition, the same authors report that the sea surface salinity in summer 1998 was exceptionally low, “creating a patch of water with a salinity minimum ( $\sim 33.7$ ) around  $59^{\circ}\text{S}$ ”, justifying the low salinity content of the CSW. Also STSW is much fresher than for the spring 2001 situation (Table 1). However, from Figure 2 it is clear that this salinity value is supported by the summer 1998 data. A potential explanation for these fresher STSW waters is the more northern extension of the STZ and SAF during the SAZ’98 cruise (Figure 1). Apart from salinity, the variation of the other parameters might seem less evident in a seasonal perspective, indicating that interannual variability is probably more important. Temperature (except for CSW) is slightly lower than in the spring 2001 situation. The corrected nutrient contents PO and NO are mostly higher than in the spring 2001 situation. Notice that these unexpected “seasonal” changes were already observed in section 3.4.3 when analyzing the residuals. Further inspec-



**Figure 8.** Mixing fractions estimated using the POMP method for the summer 1998 situation, using the SWT specially defined for these data (Table 1). Abbreviations are the same as in Figures 3, 5 and 6. Dots represent the sampling depths.

tion shows that PO and NO are significantly higher in particular for AAIW, SAMW and STSW and that this difference is mainly due to higher nutrient (phosphate and nitrate respectively) concentrations in the summer 1998 data. However, this apparent nutrient increase is relatively small ( $\sim 0.3 \mu\text{M}$  for phosphate and  $\sim 3 \mu\text{M}$  for nitrate) compared to the depth or latitudinal gradients ( $\geq 1 \mu\text{M}$  for phosphate and  $\geq 20 \mu\text{M}$  for nitrate). Furthermore, similar shifts are observed by *Rintoul and Trull* [2001] for the mixed layer nitrate concentration in the subtropical and polar zones. The reason why the changes in temperature, PO and NO contents cannot easily be interpreted as seasonally forced is probably that two seasons from a different year (spring 2001 and summer 1998) and slightly different meridional locations (Figure 1) are examined. So at least for some parameters the interannual variability (e.g., due to fronts displacements [cf. *Rintoul et al.*, 1997; *Sokolov and Rintoul*, 2003]) overwhelms the seasonal cycle. For instance, *Chaigneau and Morrow* [2002] observed for this region an interannual temperature variation ( $3\text{--}5^\circ\text{C}$ ), which is similar or greater than the variation of the total seasonal cycle ( $3\text{--}4^\circ\text{C}$ ). Finally, long term changes in water mass properties have been reported in the same direction (cooling and freshening [*Aoki et al.*, 2005b]). These observations illustrate the variability of the water mass composition and emphasize the importance to carefully characterize the sources in a mixing analysis.

[60] The same spline complexity and weighting as in section 3.4 is used. Please refer to that section for argumentation.

### 3.5.2. Results From POMP Analysis: Mixing Fractions and Uncertainties

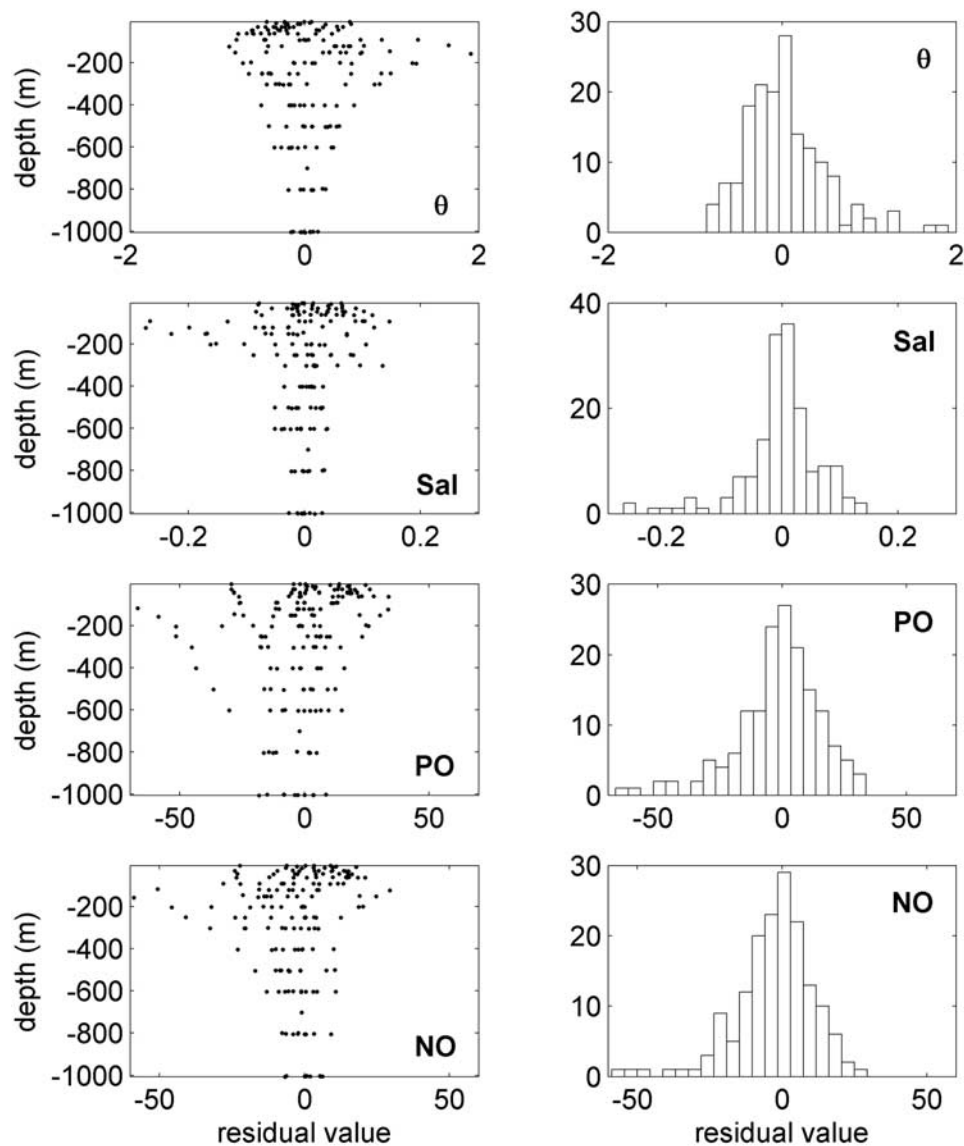
[61] With the new SWT definitions, the mixing fractions are quite different (Figure 8 versus Figure 6) and appear more realistic. The main features are the following:

[62] 1. The SAMW again appears as a homogeneous layer between  $43^\circ\text{S}$  (the STF) and  $50^\circ\text{S}$  (the SAF) but in contrast with the situation in spring, SAMW does not reach the surface anymore. Rather, its core is located between 200 and 700 m, which corresponds to the observations of *Rintoul et al.* [1997] and *Rintoul and Trull* [2001].

[63] 2. The water column above the SAMW is described by a mixing between STSW and CSW, expressing both the summer stratification (seasonal thermocline) and the complexity of the structure of the subantarctic surface waters subject to intense northward cross-front exchanges and transport of water masses [*Rintoul et al.*, 1997].

[64] 3. The mixing fractions for UCDW and AAIW exhibit patterns very similar to the spring 2001 situation. As exposed in section 3.3.2 for the spring 2001 data set, the large extension of the AAIW branch from south to north results from the heterogeneity of AAIW found along the transect [*Rintoul and Bullister*, 1999].

[65] This small study comparing two data sets can obviously not be used to quantify the variability associated with



**Figure 9.** Residuals associated with Figure 8, shown (left) as a function of depth and (right) as histograms for all variables. Residuals have same dimensions as the respective variables (see Table 1).

water masses in this region. *Tomczak and Liefvink* [2005] analyzed the interannual variations of water mass volumes (calculated from OMP mixing contributions) along the same section using five data sets from the period 1991 to 1996. They observed a variability of up to 20% for some of the water masses.

[66] The uncertainty associated with the estimated fractions is assessed in the same way as described in section 3.3.3 (one thousand Monte-Carlo simulations). The average standard deviations are 3.5% for AAIW, 6.3% for SAMW, 3.7% for STSW, 1.1% for CSW and 2.6% for UCDW.

### 3.5.3. Residuals

[67] This reconstruction with specific summer 1998 SWTs is not only more realistic based on the fraction fields, it also describes the data better: the weighted residual sum of squares has decreased from 44765 (spring 2001 SWTs, section 3.4) to 13980 in this last computation. This means

that the new model is able to capture much more of the information contained in the data.

[68] In Figure 9, the residuals are shown. When comparing these residuals to those in Figure 7, obviously the residuals have become more symmetric and well centered around zero, indicating that most of the systematic features have disappeared. Even for the surface waters it appears that the description by the newly selected SWTs is satisfactory.

## 4. Conclusion

[69] The aim of this paper was twofold. First, a parametric extension of the classical OMP method for water mass analysis was presented. The main effect of this parameterization is a reduction of the number of unknowns to be estimated, which results in a natural smoothing of the fraction fields, an explicit position dependence of the source contributions and an increased robustness to measurement

errors. An essential part of the method consists of an objective selection of the model's complexity, which can be performed based on statistical criteria. This part of the method is important since a too simple model will give inaccurate, too smooth, mixing distributions. The main assumption underlying the POMP method is that the mixing fractions vary systematically and relatively slowly with position: any pointwise variation is interpreted as due to noise. Therefore, phenomena varying on the grid scale risk to remain uncaptured by a POMP analysis. However, they will still be apparent in the residual plots.

[70] In the second part, this method was applied to two data sets of approximately the same section but measured during different seasons and years. The results from this application confirm that the parametric OMP method (POMP) is more robust than the original procedure, i.e. the resulting mixing fractions are less influenced by noise in the measurements and consequently are smoother and easier to interpret, at least in terms of larger scale phenomena. On the other hand, this exercise illustrated that in OMP-like analyses SWTs and their characteristics have to be defined carefully, in this case even for each data set individually. This reflects the general variability of water mass characteristics, which in our examples appeared to be not so much seasonally influenced, but seem to exhibit a longer term change. This change can be either steady (i.e. long term warming/cooling etc.) or irregular as related to e.g. front displacements, but the studied data sets are not sufficient to disentangle these factors or to determine their relative importance.

[71] Finally, we would like to suggest an outlook to future applications. The separation of ocean current fields into a slowly evolving mean field and a perturbation field with zero mean to represent geostrophic turbulence is an established practice in oceanography. It appears to us that POMP analysis has the potential to establish an equivalent technique in water mass analysis. By carefully choosing the correct level of spline complexity POMP analysis can extract the slowly evolving fields of the hydrographic parameters from a complex data set. The difference fields (OMP – POMP, or difference fields between POMP runs with different spline complexities) then represent the contribution from geostrophic turbulence to the hydrographic parameter fields.

## Appendix A: POMP Algorithm

[72] In this section, the estimation algorithm is briefly described. All functions were written in MATLAB [MATLAB, 2005]. The goal is to estimate the spline coefficients using equation (8). Once these are known, the source contributions can be calculated with equation (5).

### A1. Input

[73] Data matrix  $\mathbf{Y}$ ; source definition matrix  $\mathbf{S}$ ; weights for all variables; spline parameters (order, number).

### A2. Calculations

[74] Construct spline matrix  $\mathbf{B}$ .

[75] Equation (8) must be vectorized, because the optimization functions are written for this kind of input. Using the  $vec$  operator, all columns of a matrix are stacked

underneath each other, giving one long column vector. For equation (8)

$$\begin{aligned} \mathbf{Y} &= \mathbf{S} \cdot \mathbf{C} \cdot \mathbf{B}^T \Rightarrow vec(\mathbf{Y}) = \mathbf{B} \otimes \mathbf{S} \cdot vec(\mathbf{C}) \Leftrightarrow \mathbf{Y} \\ &= \mathbf{B} \otimes \mathbf{S} \cdot \mathbf{C}, \end{aligned} \quad (\text{A1})$$

where  $vec(\mathbf{Y})$  and  $vec(\mathbf{C})$  stand for the vectorized matrices.  $\otimes$  is the Kronecker product operator which multiplies each element of the first matrix with the second matrix [Brewer, 1978].

[76] Optimal values for  $vec(\mathbf{C})$  can now be found by solving the linear system equation (A1), taking into account the mass balance and nonnegativity constraints. For this optimization, we used the MATLAB function “lsqin” [MATLAB, 2005] for constrained linear least squares optimization.

[77] This function does not have the possibility to input a weighting matrix. Hence, the weighting has to be done before starting the optimization: each element in  $vec(\mathbf{Y})$  and in  $\mathbf{B} \otimes \mathbf{S}$  has to be divided by the according standard deviation (see expression (9)).

### A3. Output

[78] Optimal values for the spline coefficients  $vec(\mathbf{C})$ ; optimal values for the source contributions at the sampling positions can be found by applying equations (5) or (7); residuals can be computed.

### A4. Spline Complexity Selection

[79] To select the optimal spline complexity, the above calculations should be repeated for all relevant number of splines. For each fixed number of splines, the analysis should also be repeated for a number of orders (we tested orders 1 to 3). The best order for a fixed number of splines is that one giving the smallest residuals (residual sum of squares is a good measure), since varying the order does not change the number of model parameters (unknowns). Now the optimal number of splines still has to be selected, which we suggest to do based on model selection criteria described by *de Brauwere et al.* [2005] and *De Ridder et al.* [2005].

[80] **Acknowledgments.** Anouk de Brauwere and Fjo De Ridder are researcher assistant and postdoctoral fellow of the Research Foundation Flanders (FWO - Vlaanderen), respectively. This work was conducted within the framework of projects supported by the Belgian Government (IUAP V/22), the Federal Office for Science Policy program on Global Change, Ecosystems and Biodiversity (BELCANTO Antarctic research network; contracts A4/DD/B11, EV/03/7A, SD/CA/03A), the Flemish Government (GOA22, GOA-IliNoS), the Vrije Universiteit Brussel (HOA9), the Australian Antarctic Science Division (ASAC projects 1343 and 2572), the Australian Greenhouse Office, and the Cooperative Research Centre Program. We thank Mark Rosenberg for making the oceanographic data available and Steve Rintoul for his valuable help evaluating the water mass distributions, as well as Miguel Gilcoto, Matthias Tomczak and an anonymous reviewer for stimulating discussions and useful comments.

## References

- Aoki, S., N. L. Bindoff, and J. A. Church (2005a), Interdecadal water mass changes in the Southern Ocean between 30°E and 160°E, *Geophys. Res. Lett.*, *32*, L07607, doi:10.1029/2004GL022220.
- Aoki, S., S. R. Rintoul, S. Ushio, S. Watanabe, and N. L. Bindoff (2005b), Freshening of the Adélie Land Bottom Water near 14°E, *Geophys. Res. Lett.*, *32*, L23601, doi:10.1029/2005GL024246.
- Brewer, J. W. (1978), Kronecker products and matrix calculus in system theory, *IEEE Trans. Circuits Syst.*, *25*, 772–781.
- Broecker, W. S. (1974), “NO,” a conservative water-mass tracer, *Earth Planet. Sci. Lett.*, *23*, 100–107.

- Castro, C. G., F. F. Perez, S. E. Holley, and A. F. Rios (1998), Chemical characterisation and modelling of water masses in the northeast Atlantic, *Prog. Oceanogr.*, *41*, 249–279.
- Chaigneau, A., and R. Morrow (2002), Surface temperature and salinity variations between Tasmania and Antarctica, 1993–1999, *J. Geophys. Res.*, *107*(C12), 8021, doi:10.1029/2001JC000808.
- Chaigneau, A., R. A. Morrow, and S. R. Rintoul (2004), Seasonal and interannual evolution of the mixed layer in the Antarctic Zone south of Tasmania, *Deep Sea Res., Part I*, *51*, 2047–2072.
- Coatanoan, C., N. Metzl, M. Fieux, and B. Coste (1999), Seasonal water mass distribution in the Indonesian throughflow entering the Indian Ocean, *J. Geophys. Res.*, *104*, 20,801–20,826.
- de Boer, C. J., and H. M. van Aken (1995), A study of objective methods for water mass analysis, applied to the Iceland Basin, *Dtsch. Hydrogr. Z.*, *47*, 5–22.
- de Brauwere, A., F. De Ridder, R. Pintelon, M. Elskens, J. Schoukens, and W. Baeyens (2005), Model selection through a statistical analysis of the minimum of a weighted least squares cost function, *Chemometrics Intell. Lab. Syst.*, *76*, 163–173.
- De Ridder, F., R. Pintelon, J. Schoukens, and D. P. Gillikin (2005), Modified AIC and MDL model selection criteria for short data records, *IEEE Trans. Instrum. Meas.*, *54*, 144–150.
- Dierckx, P. (1995), *Curve and Surface Fitting with Splines*, Oxford Univ. Press, New York.
- Fogelqvist, E., J. Blindheim, T. Tanhua, S. Osterhus, E. Buch, and F. Rey (2003), Greenland-Scotland overflow studied by hydro-chemical multivariate analysis, *Deep Sea Res., Part I*, *50*, 73–102.
- Henry-Edwards, A., and M. Tomczak (2006), Remote detection of water property changes from a time series of oceanographic data, *Ocean Sci.*, *2*, 19–25.
- Jacquet, S. H. M., F. Dehairs, and S. Rintoul (2004), A high resolution transect of dissolved barium in the Southern Ocean, *Geophys. Res. Lett.*, *31*, L14301, doi:10.1029/2004GL020016.
- Lo Monaco, C., N. Metzl, A. Poisson, C. Brunet, and B. Schauer (2005), Anthropogenic CO<sub>2</sub> in the Southern Ocean: Distribution and inventory at the Indian-Atlantic boundary (World Ocean Circulation Experiment line 16), *J. Geophys. Res.*, *110*, C06010, doi:10.1029/2004JC002643.
- Mackas, D. L., K. L. Denman, and A. F. Bennett (1987), Least-squares multiple tracer analysis of water mass composition, *J. Geophys. Res.*, *92*, 2907–2918.
- MATLAB (2005), version 7, The Mathworks Inc., Natick, Mass.
- OMP2 (2005), Water mass analysis package. (Available at [http://www.ldeo.columbia.edu/~jkarsten/omp\\_std/](http://www.ldeo.columbia.edu/~jkarsten/omp_std/))
- Orsi, A. H., T. Whitworth, and W. D. Nowlin (1995), On the meridional extent and fronts of the Antarctic Circumpolar Current, *Deep Sea Res., Part I*, *42*, 641–673.
- Perez, F. F., L. Mintrop, O. Llinas, M. Glez-Davila, C. G. Castro, M. Alvarez, A. Kortzinger, M. Santana-Casiano, M. J. Rueda, and A. F. Rios (2001), Mixing analysis of nutrients, oxygen and inorganic carbon in the Canary Islands region, *J. Mar. Syst.*, *28*, 183–201.
- Pintelon, R., and J. Schoukens (2001), *System Identification: A Frequency Domain Approach*, IEEE Press, New York.
- Rintoul, S. R., and J. L. Bullister (1999), A late winter hydrographic section from Tasmania to Antarctica, *Deep Sea Res., Part I*, *46*, 1417–1454.
- Rintoul, S. R., and S. Sokolov (2001), Baroclinic transport variability of the Antarctic Circumpolar Current south of Australia (WOCE repeat section SR3), *J. Geophys. Res.*, *106*, 2815–2832.
- Rintoul, S. R., and T. W. Trull (2001), Seasonal evolution of the mixed layer in the Subantarctic Zone south of Australia, *J. Geophys. Res.*, *106*, 31,447–31,462.
- Rintoul, S. R., J. R. Donguy, and D. H. Roemmich (1997), Seasonal evolution of upper ocean thermal structure between Tasmania and Antarctica, *Deep Sea Res., Part I*, *44*, 1185–1202.
- Schlitzer, R. (2004), Ocean Data View. (Available at <http://odv.awi-bremerhaven.de/>)
- Siedler, G., J. Church, and J. Gould (2001), *Ocean Circulation and Climate*, Elsevier, New York.
- Sokolov, S., and S. R. Rintoul (2002), Structure of Southern Ocean fronts at 140°E, *J. Mar. Syst.*, *37*, 151–184.
- Sokolov, S., and S. R. Rintoul (2003), Subsurface structure of interannual temperature anomalies in the Australian sector of the Southern Ocean, *J. Geophys. Res.*, *108*(C9), 3285, doi:10.1029/2002JC001494.
- Takahashi, T., et al. (2002), Global sea-air CO<sub>2</sub> flux based on climatological surface ocean pCO<sub>2</sub>, and seasonal biological and temperature effect, *Deep Sea Res., Part II*, *49*, 1601–1622.
- Thompson, R. O. R. Y., and R. J. Edwards (1981), Mixing and water-mass formation in the Australian Sub-Antarctic, *J. Phys. Oceanogr.*, *11*, 1399–1406.
- Tomczak, M. (1981), A multi-parameter extension of temperature/salinity diagram techniques for the analysis of non-isopycnal mixing, *Prog. Oceanogr.*, *10*, 147–171.
- Tomczak, M. (1999), Some historical, theoretical and applied aspects of quantitative water mass analysis, *J. Mar. Res.*, *57*, 275–303.
- Tomczak, M., and D. G. B. Large (1989), Optimum multiparameter analysis of mixing in the thermocline of the eastern Indian Ocean, *J. Geophys. Res.*, *94*, 16,141–16,149.
- Tomczak, M., and S. Liefink (2005), Interannual variations of water mass volumes in the Southern Ocean, *J. Atmos. Ocean Sci.*, *10*, 31–42.
- Vanlanduit, S. (2001), High spatial resolution experimental modal analysis, Ph.D. thesis, 304 pp., Vrije Univ., Brussels.
- You, Y. Z. (1997), Seasonal variations of thermocline circulation and ventilation in the Indian Ocean, *J. Geophys. Res.*, *102*, 10,391–10,422.
- You, Y. (2002), Quantitative estimate of Antarctic Intermediate Water contributions from the Drake Passage and the southwest Indian Ocean to the South Atlantic, *J. Geophys. Res.*, *107*(C4), 3031, doi:10.1029/2001JC000880.
- You, Y., and M. Tomczak (1993), Thermocline circulation and ventilation in the Indian Ocean derived from water mass analysis, *Deep Sea Res., Part I*, *40*, 13–56.
- You, Y., J. R. M. Lutjeharms, O. Boebel, and W. P. M. de Ruijter (2003), Quantification of the interocean exchange of intermediate water masses around southern Africa, *Deep Sea Res., Part II*, *50*, 197–228.

W. Baeyens, A. de Brauwere, F. Dehairs, and S. H. M. Jacquet, Department of Analytical and Environmental Chemistry, Vrije Universiteit Brussel, B-1050 Brussels, Belgium. (adebrauw@vub.ac.be)

F. De Ridder, R. Pintelon, and J. Schoukens, Department of Electricity and Instrumentation, Vrije Universiteit Brussel, B-1050 Brussels, Belgium.