*Journal of Marine Research, 63,* 683–704, 2005

# Statistical inversion of South Atlantic circulation in an abyssal neutral density layer

**by Ian W. McKeague[1], Geoff Nicholls[2], Kevin Speer[3] and Radu Herbei[4]**

ABSTRACT

This paper introduces a Bayesian inversion approach to estimating steady state ocean circulation and tracer fields. It is based on a quasi-horizontal flow model and a PDE solver for the forward problem of computing solutions to the tracer field advection-diffusion equations. A typical feature of existing ocean circulation inverse methods is a preprocessing stage in which the tracer data are interpolated over a regular grid and the interpolation error is ignored in the subsequent inversion. Our approach only uses interpolated data at those grid points that have neighboring hydrographic stations. By exploiting physically-based models in an integrated fashion, the method provides a statistically unified inversion and tracer field reconstruction with minimal data smoothing. Solving the problem consists of finding information about the circulation and tracer fields in the presence of a number of assumptions (prior information); the resulting posterior probability distribution summarizes what we can know about these fields. We develop a Markov chain Monte Carlo simulation procedure to extract information from the (analytically intractable) posterior distribution of all the parameters in the model; uncertainty about the "solution" is represented by variation in the output of the simulation runs. Our approach is aimed at finding the time-averaged quasi-horizontal flow and tracer fields for an abyssal neutral density layer in the South Atlantic.

## 1. Introduction

The collection of oceanographic data during the past century has formed the basis for our understanding of the distribution of properties and the circulation in the ocean. However, a description of the steady flow and associated property fields in the oceanic interior that is dynamically consistent with those collective observations and that accounts for the inherent uncertainty associated with measurements, physical variability and model parameterizations, for example, is still lacking. It may be that the steady-state assumption is a poor approximation to the equations of motion, but this hypothesis deserves close attention as a test of our understanding of the basic physics of the ocean circulation. In this paper we introduce a physically-based statistical inversion approach to estimate the

1. Department of Biostatistics, Columbia University, 722 W. 168th Street, New York, New York, 10032, U.S.A. ***email: im2131@columbia.edu***
2. Department of Statistics, Oxford University, Oxford OX1 3TG, United Kingdom.
3. Department of Oceanography, Florida State University, Tallahassee, Florida, 32306, U.S.A.
4. Department of Statistics, Florida State University, Tallahassee, Florida, 32306, U.S.A.

large-scale steady state circulation from tracer observations, and apply the approach to an abyssal neutral density layer of the South Atlantic.

Hydrographic data (hydrostatic pressure, temperature, salinity, and other properties such as silica and oxygen) in each of the ocean basins are strongly nonuniform in their spatial and temporal distribution. Such data have been used to construct climatological maps of ocean properties (Levitus, 1994; Lozier *et al.,* 1995; Curry, 1996) with large-scale structure reflecting the large-scale circulation, whose elements include western boundary currents, wind-driven gyres, and abyssal interior flow driven by some combination of mechanical (wind, tides) and thermodynamic forcing. Due to the scarcity of data in any particular year and the spatial irregularity of sampling stations, statistical smoothing methods are used typically to provide estimates at space-time locations for which there are no measurements (Stammer *et al.,* 2002). Using an approach similar in form to ours, Lavine and Lozier (1999) recently considered a Bayesian smoothing technique to the estimation of temperature changes (not circulation) in the subtropics of the North Atlantic, with prior temperature variations expressed in terms of Gaussian Markov random fields (GMRFs).

For the steady problem, hydrographic data collected at different times are assumed to be representative of a mean state. The degree to which this is true is often impossible to determine for lack of data; however, it is thought that the large-scale distributions of dynamically-passive tracer represent an integrated effect of circulation at many space and time scales, hence provide more insight to a mean flow. For this notion to be meaningful, significant mean flow ought to be associated with tracer distributions, in the face of diffusion by eddies. [The instantaneous flow, measured by the density field (the thermal wind; Pedlosky, 1996), is in contrast unlikely to be close to the mean flow, and smoothing is often used to try to approach the mean.] Passive tracer fields change with the instantaneous flow (e.g., if there is an eddy or other small-scale structure), but they are often sufficiently stable on a fixed neutral density layer to justify combining data sets collected at different times. For abyssal property distributions, we also need corrections for measurement bias. Our analysis begins with corrected tracer observations from recent (1980s and 1990s) oceanographic field work in the South Atlantic Ocean (Thurnherr and Speer, 2004; Fig. 1).

A widely used method of estimating flow is the box inverse method (Wunsch, 1996), which starts from *in situ* density measurements, available from salinity, temperature and hydrostatic pressure data along ship tracks. Integrating the thermal wind equations reduces the problem to the estimation of a reference level velocity field. A system of linear equations expressing conservation of mass (and other constraints) within a collection of connected boxes is then inverted to estimate the reference level velocity via Gauss–Markov or SVD/tapered/ridge regression estimators. A "local" way of approaching the reference level velocity problem, known as the β-spiral method (Stommel and Schott, 1977), has also been widely used. More elaborate general circulation inverse modeling based on nonlinear optimization with weak constraints, allowing errors in the data and in the

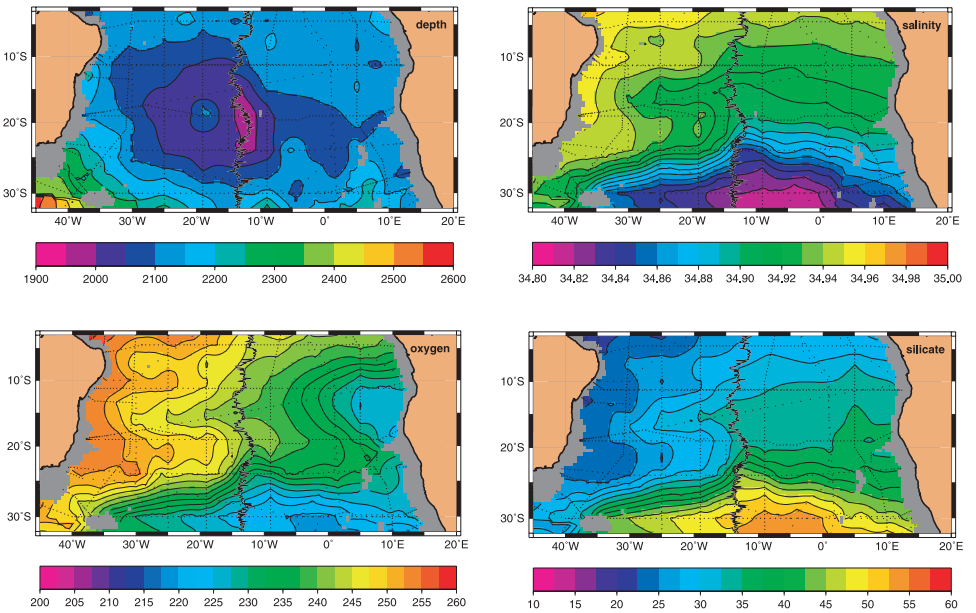


Figure 1. Maps based on smoothed and interpolated data for the 28.0 isoneutral surface (Thurnherr and Speer, 2004); depth in meters, salinity in ppt, oxygen and silica in μmol/kg. The hydrographic stations are indicated by dots, and the mid-Atlantic Ridge is the ragged solid line. Data interpolation done at a 1 degree resolution with a harmonic surface fitting (Smith and Wessel, 1990).

dynamics, have been explored by Bennett (1992), Zhang and Hogg (1993), Wunsch (1994) and Paillet and Mercier (1997), among others.

Our approach is based on a Bayesian inversion method and the use of a PDE solver for the forward problem. Bayesian estimation of parameters in PDEs (in the context of image recovery problems) has been surveyed by Fox and Nicholls (2002). The forward problem in our case is to compute solutions to the tracer field advection-diffusion equations. A typical feature of existing grid-based ocean circulation inverse methods is a preprocessing stage in which the tracer data are interpolated over a regular grid (e.g., Fig. 1) and the interpolation error is ignored in the subsequent inversion. In contrast, while our proposed approach does use a regular grid, we only assume that we have tracer observations at those grid points with nearby hydrographic data. This provides an inversion for flow and tracer field reconstruction consistent with the physics.

The Bayesian method provides a formula (Bayes formula) which allows us to write, up to an unknown normalizing constant, the probability density function of the desired fields, conditional on the data. This high-dimensional joint density defines what is called the posterior distribution. The posterior distribution provides statistically consistent information about the fields (means, variances, etc.) and uncertainty in the inversion in terms of all the parameters involved. However, the normalizing constant is analytically intractable and

the posterior distribution needs to be studied by simulation techniques, in our case Markov Chain Monte Carlo (MCMC). Simulation runs can be designed to produce samples from the posterior distribution and these represent the results; they might be described as "snapshots" of the underlying "solution" that are consistent with the data and the physical model. We make the obvious choice to present the posterior mean values of the fields as the "solutions," and the posterior standard deviation values as the "uncertainty," though many other choices could be made (e.g., one might wish to examine medians and modes). In principle, any queries about the posterior distribution, for example, the dependence of the "solution" on prior information or observations, can be answered in this way.

The paper is organized as follows. In Section 2 we introduce the physical and observation models, the proposed Bayesian inversion technique, and the MCMC procedure used numerically to reconstruct the advection and the tracer fields. The inversion results are presented in Section 3, and some concluding remarks in Section 4. The Appendix contains details about the MCMC implementation.

## 2. Inversion for quasi-horizontal flow

We consider the problem of estimating the large-scale climatological velocity within an abyssal neutral density layer. The analysis is based on Bayesian inversion of tracer data using geostrophic balance and 2D advection-diffusion equations. We restrict attention to a 2D inversion in order to bring out the essential features of our approach; the full 3D inversion problem, in which the vertical velocity also needs to be estimated, will be considered in subsequent papers.

We are interested in the flow along a thin abyssal neutral density layer of the form

$$\{(x, y, z)\colon \gamma_a \le \gamma(x, y, z) \le \gamma_b \text{ for } (x, y) \in S\},$$

where $S$ is the inversion region; the neutral density is $\gamma(x, y, z)$ (McDougall, 1987), and $\gamma_a$, $\gamma_b$ are the lower/upper levels of the layer. In what follows we study the 27.98–28.01 neutral density layer, roughly corresponding to depths within one or two hundred meters of 2000 m, and the boundaries of the region $S$ are 34W, 11E, 32S and 3S. The tracers are salinity, oxygen, silica ($SiO_2$), large-scale potential vorticity (McDougall, 1988), and potential temperature. We use the following standard notation: Coriolis parameter $f = 2\Omega \sin\theta$, where $\theta$ is latitude and $\Omega$ is the earth's angular velocity, $\beta = \partial f/\partial y$, horizontal velocity $\mathbf{u}_H = (u, v)$, *in situ* density $\rho$, and pressure $p$. The horizontal velocity is assumed not to vary vertically throughout the layer.

### *a. Physical model*

The model dynamics are based on exact geostrophic balance:

$$\rho f u = -\frac{\partial p}{\partial y}, \qquad \rho f v = \frac{\partial p}{\partial x}. \tag{1}$$

The pressure field is eliminated by differentiating and adding these equations. Using the Boussinesq approximation to neglect density variations,

$$\nabla \cdot f\mathbf{u}_H = f\frac{\partial u}{\partial x} + \beta v + f\frac{\partial v}{\partial y} = 0 \tag{2}$$

in terms of the horizontal divergence operator (see, e.g., Pedlosky, 1996, p. 7). Assume that the zonal flow vanishes along the eastern boundary $x = x_e$, as it would if the eastern boundary is a N–S wall. Dividing (2) by $f$ and integrating from the eastern boundary gives

$$u(x, y) = \int_x^{x_e} \left(\frac{\partial v}{\partial y} + \frac{\beta v}{f}\right) dx'. \tag{3}$$

The zonal velocity field is thus driven by the meridional velocity to the east of $x$, an implicit acknowledgment of higher order (dissipative) dynamics missing from geostrophic balance (1). Higher order dynamics would put strong boundary currents at the western boundary with zonal flow connecting them to the interior. By expressing $u$ as a known linear operator acting on $v$, we have cut by half the number of advection parameters that need to be estimated. The second term in the integrand above is expected to be one to two orders of magnitude smaller than the first, except near the equator.

To complete the structure of the horizontal advection, we need a suitable representation for $v$. The Bayesian approach makes its first appearance at this stage, because we take the point of view that it suffices to specify a *prior distribution* for $v$. [Note that the posterior flow will have structure derived from the observed tracer fields.] The approach we adopt here is to assume a zero mean spatially stationary GMRF (defined below) for the prior on $v$. GMRF priors are popular in image analysis (see Winkler, 1995) where prior information is similarly unstructured. Later work will relax this assumption.

The remaining part of the physical model is the link between the tracers and the advection. The horizontal advection-diffusion equation for a given climatological tracer concentration $C = C^{(j)}$, where $j$ indexes one of $n_C$ tracers, is written

$$\mathbf{u}_H \cdot \nabla C = \nabla \cdot (K\nabla C) + Q_C, \qquad (x, y) \in S \tag{4}$$

where the (2D) diffusivity matrix $K$ is diagonal with spatially constant components $K^{(x)}$, $K^{(y)}$ (the same for each tracer, and assumed to be positive), and the term $Q_C = -\lambda C$ is only present in the case of oxygen. Here $\lambda$ represents the rate of oxygen consumption (with units 1/s), also assumed to be spatially constant. We use a Dirichlet boundary condition $C^{(j)} = C^{(j)}_{\partial S}$, for $(x, y) \in \partial S$, with spatially varying boundary values $C^{(j)}_{\partial S}$ treated as a 1D GMRF (defined below).

### *b. Observation model*

Bias-corrected modern hydrographic data (Thurnherr and Speer, 2004) are used to find the mean tracer concentrations within a fixed neutral density layer at each station. These

concentrations are filtered to discard outliers on the density surface (more than three standard deviations from the overall mean), and averaged over nearest neighbors only onto the grid $S_G$ used in the discretization of the advection-diffusion equation. Thus, when data are farther than one grid point away from a given location there are no tracer observations specified at that location.

An important aspect of our reconstruction is the use of tracer data located in a fixed neutral density layer, rather than a layer of fixed depth. The strong temporal dependence of the data is illustrated in Figure 2; nearby stations from two separate cruises have very different neutral density and oxygen profiles (first and second panels) when plotted against depth. However, when the oxygen concentration is plotted against neutral density (third panel), the profiles match closely over most of their range and, in particular, in the range 27.98–28.01. The pooling of the data (from all cruises) within this neutral density layer is a standard way to remove unwanted noise.

The resulting concentration $C^{(j)}_{D,i}$ for tracer $j$ at grid points in $S_G$ indexed by $i$ is assumed to satisfy the observation model

$$C^{(j)}_{D,i} = C^{(j)}_i + \epsilon^{(j)}_i, \tag{5}$$

$i = 1, \ldots, n_D$, $j = 1, \ldots, n_C$, where $C^{(j)}_i$ is the underlying (steady state) concentration of the $j$th tracer, and the $\epsilon^{(j)}_i$ are independent and normally distributed measurement errors having zero mean and constant variance $\sigma^2_{(j)}$. The assumption of a constant (and prespecified) measurement error variance $\sigma^2_{(j)}$ was found to be reasonable based on inspection of residuals from nonparametrically smoothed estimates of the tracer concentrations. The actual values of $\sigma^2_{(j)}$ are given in Table 1 below.

The solutions $C^{(j)}(x, y) = C^{(j)}(x, y|\Phi)$ to the advection-diffusion equation may be treated as uniquely determined by the collection of parameters $\Phi = (u, v, K^{(x)}, K^{(y)}, \lambda, C_{\partial S})$, where $C_{\partial S} = (C^{(j)}_{\partial S}, j = 1, \ldots, n_C)$ are the boundary conditions. The zonal velocity $u$ is driven by $v$, even though it appears as a separate variable. Approximate solutions to these PDEs which work well in our setting will be based on multigrid iteration, as implemented in MUDPACK, a collection of Fortran programs for linear PDEs developed by J. C. Adams.

The inverse problem of reconstructing $\Phi$ from the tracer fields (even with complete tracer information) is ill-posed in general: the solutions $C^{(j)}(x, y|\Phi)$ do not determine $\Phi$ without some extra assumptions. This is due to invariance of the advection-diffusion equations under certain transformations of $\Phi$. For example, multiplying $(u, v, K^{(x)}, K^{(y)}, \lambda)$ by any positive constant gives the same advection-diffusion equations after cancellation of the constant, so the solutions $C^{(j)}$ are the same. Nevertheless, if $\lambda$ is fixed (providing a fixed clock rate), this problem does not arise and $\Phi$ can be determined; see Wunsch (1996, p. 282) for a discussion of this issue.

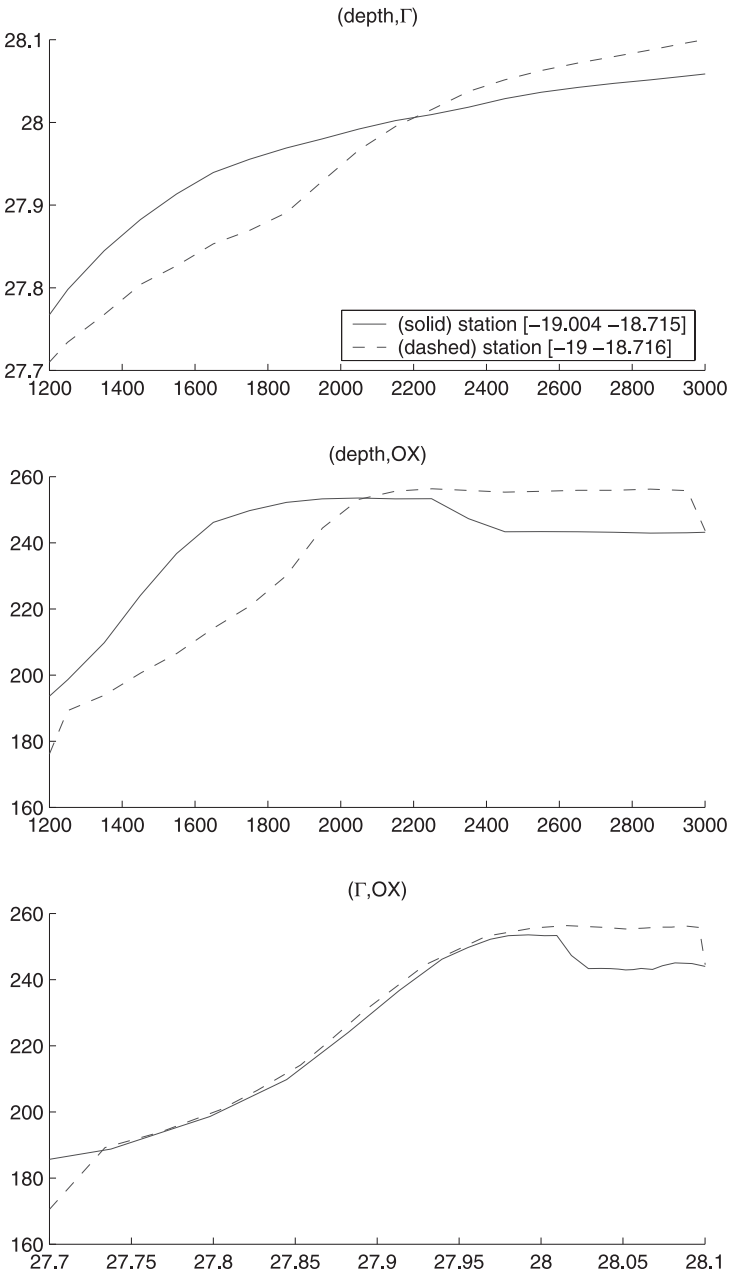


Figure 2. Comparison of oxygen data at two nearby stations from separate cruises: neutral density vs. depth (top), oxygen concentration vs. depth (middle), oxygen concentration vs. neutral density (bottom). Note the close agreement between the curves in the bottom panel.

Table 1. Values of $\delta_1$ in the priors and values of $\sigma^2_{(j)}$ in the observation model.

| | Salinity | Oxygen | Silica | Potential vorticity | Potential temperature | $v$ |
|---|---|---|---|---|---|---|
| $\delta_1$ | $1.75 \times 10^4$ | 0.23 | 0.38 | $10^2$ | $6.5 \times 10^2$ | $2 \times 10^4$ |
| $\sigma^2_{(j)}$ | 0.00757 | 2.4617 | 1.8701 | 0.10446 | 0.04128 | — |

*c. Bayesian inversion*

Conditional on the tracer data

$$C_D = (C^{(j)}_{D,i},\ i = 1, \ldots, n_D,\ j = 1, \ldots, n_C), \tag{6}$$

Bayes formula gives the posterior density of the parameters $\Phi$ in terms of prior densities on the various components:

$$\pi(\Phi|C_D) \propto \left[ \prod_{j=1}^{n_C} \prod_{i=1}^{n_D} \exp\left\{ -\frac{1}{2\sigma^2_{(j)}} (C^{(j)}_{D,i} - C^{(j)}_i)^2 \right\} \right] \pi(v)\pi(K^{(x)})\pi(K^{(y)})\pi(\lambda) \prod_{j=1}^{n_C} \pi(C^{(j)}_{\partial S}). \tag{7}$$

The term in square brackets corresponds to the likelihood and the remaining terms to (independent) priors on the parameters; their product is an unnormalized posterior density for $\Phi$ which is proper (i.e., integrable) provided the prior densities are proper (note that the likelihood component is uniformly bounded as a function of $\Phi$).

The prior for the meridional velocity is specified using a 2D GMRF over the grid $S_G$ and having density

$$\pi(v) \propto \exp\left( -\delta_1 \sum_{s \sim s'} (v_s - v_{s'})^2 - \delta_2 \sum_{s \in S_G} v_s^2 \right), \tag{8}$$

where the first summation runs over pairs of adjacent grid points, and $\delta_1$, $\delta_2$ are positive tuning parameters that control nearest neighbor interaction and variance.

For background on the use of GMRF models in spatial statistics see Besag (1974, 1986) and Winkler (1995). Each of the boundary conditions $C^{(j)}_{\partial S}$ is represented by a 1D GMRF of the same form. Wikle *et al.* (2003) have used a similar prior specification of Dirichlet boundary conditions for an application involving atmospheric streamfunction fields. We take $\delta_2 = 0$ to provide a 'non-informative' prior, even though the prior is then improper; this did not result in an improper posterior judged by the output from the MCMC. The values of the interaction parameter $\delta_1$ used to specify these various GMRF priors are provided in Table 1. The priors on $K^{(x)}$, $K^{(y)}$ are taken to be exponentially distributed with mean 1000 $m^2/s$. The rate of oxygen consumption $\lambda$ (units of 1/s) can be specified using a narrow Gaussian prior, but we found that an exactly specified value ($\lambda = 10^{-11}$) was adequate.

*d. Extracting information from the posterior*

The essence of our approach is that, rather than using unconstrained tracer fields $C$ (as used to estimate density in the box inverse method), we numerically solve the PDE in conjunction with exploration of the posterior advection and tracer fields via MCMC. The normalizing constant needed to make the unnormalized posterior density into a density is not analytically tractable, hence the need for MCMC. For general background on MCMC methods, see, e.g., Liu (2001).

The ill-posedness mentioned earlier does not carry through to our Bayesian formulation because the proper prior on $\Phi$ implies a proper posterior (as we already noted). Nevertheless, to design an MCMC scheme that adequately samples from the posterior, it is important to include state space moves that explore regions where the likelihood is constant; i.e., moves that leave the solutions to the advection-diffusion equations invariant. The data have nothing to say in these regions, so it is all the more important to explore them for assessing posterior uncertainty in the advection. One possible move of this type is to multiply all components of $\Phi$, apart from the boundary conditions, by a random factor $r$, with $r$ uniformly distributed on the interval $(1/c, c)$, for some $c > 1$. However this move can have low acceptance probability (zero if $\lambda$ is specified exactly in the prior), so it is necessary to keep $\lambda$ unchanged. Thus we use the following *ridge* move:

$$\Phi \rightarrow \Phi' = (ru, rv, rK^{(x)}, rK^{(y)}, \lambda, C_{\partial S}).$$

The acceptance probability for the ridge move is given in the Appendix. The advection-diffusion equations are invariant under the ridge move, apart from the equation for oxygen which needs to be re-solved at each iteration.

We also utilize standard random walk Metropolis moves for each component $\phi$ of $\Phi$, with the location for updating the components of $v$ and $C_{\partial S}^{(j)}$ randomly selected (a random scan). The acceptance probability for this type of move is given in the Appendix. The move is $\phi \rightarrow \phi + r$, where $r$ is uniformly distributed on the interval $(-d, d)$, for some $d > 0$; the other components remain the same. In the case of the diffusivities, which are positive, the update is $\phi \rightarrow |\phi + r|$. We also use a *column* move in which all the components of $v$ along a N–S line are simultaneously updated by $v(x, \cdot) \rightarrow v(x, \cdot) + r$, and an adjacent N–S line is updated by $v(x, \cdot) \rightarrow v(x, \cdot) - r$ (with the same $r$ for all the components); such a move respects conservation of mass and updates $v$ with negligible corresponding change to $u$.

The MCMC procedure cycles through the moves described above, with a complete cycle consisting of: (1) update $K^{(x)}$ or $K^{(y)}$ (one chosen at random); (2) update one randomly selected component of $v$; (3) update the tracer concentration at one randomly chosen site on the boundary for one randomly chosen tracer; (4) column move for $v$ along two randomly chosen adjacent N–S lines; (5) ridge move.

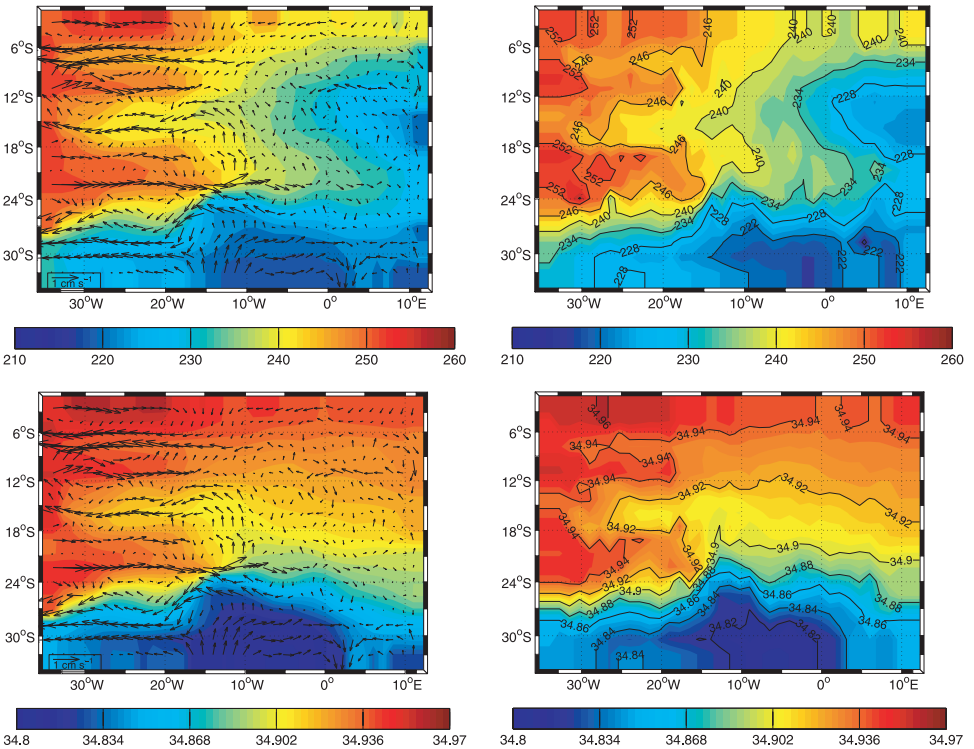


Figure 3. Posterior means of the advection and oxygen fields (top left) and interpolated oxygen data (top right); Posterior mean salinity (bottom left) and interpolated salinity data (bottom right).

## 3. Inversion results

Figure 3 shows the mean circulation, along with the reconstruction of the oxygen and salinity concentrations (left-hand plots) based on their posterior means, as compared with mapped or interpolated data (right-hand plots). The estimated mean tracer concentration is less noisy than the interpolated data, which is to be expected because they are averaged solutions to the advection-diffusion equation and these solutions are not constrained to go through the data at each sampling station. The lack of smoothness in the interpolated data is an artifact of the interpolation method itself: there is a spike at around 7W and 15S, and several other locations, that may be inconsistent with the surrounding region, and there are numerous sharp bends in the contour lines. These could be artificially smoothed away but the result would not be consistent with the physical model.

Note that observations for the tracers are sparse, but the observation model (normally distributed data) forces the result (posterior distribution) to respect those observations as controlled by the measurement error variance, which is relatively small. On the other hand, the posterior mean "solution" is only roughly comparable with the oxygen data itself (or the interpolated data in Fig. 1) because this Bayesian solution is derived from all the tracer

data (not just oxygen) along with the physical model for the advection, and its connection with the underlying tracer fields. Indeed, it would be surprising to find a very close agreement with the mapped oxygen data, which is obtained without any physical model in mind. A sense of the uncertainty involved in the reconstruction can be obtained from Figure 4, which shows the posterior standard deviation for oxygen and salinity (bottom panels) mapped over the whole inversion region, along with the advection, oxygen and salinity concentrations in two draws (upper panels) from the posterior distribution. The posterior standard deviations become inflated and less smooth around the boundary of the inversion region, but overall they are relatively small on the scale of the posterior means for the tracer concentrations, which suggests that our tracer reconstruction has a high degree of accuracy. Generally speaking, the MCMC draws should only be viewed as consistent with the data in the sense that they are sampled from the conditional distribution given the data. With so many variables in the posterior, it is highly likely that some draws will be very different from the data. Moreover, some draws will be extreme in the sense that the posterior probability distribution is relatively small for those draws compared to other draws. The posterior is a balance between information in the data and information in the prior; a low probability in the observational part of the model (the likelihood) may be balanced by a high probability in the prior, and vice-versa. If the MCMC scheme is working well, then one should expect it to explore all such situations. It would be misleading to interpret the two draws in Figure 4 as typical in any sense; only the long run averages of a large number of draws can have such an interpretation. Nevertheless, the tracer concentrations in the two draws agree remarkably well with each other and with their posterior means in Figure 3, as we expect from the relatively small posterior variance. However, the advection in the two draws only roughly reflects the posterior mean, and they differ considerably from each other, especially in the southwest corner, which points to relatively high posterior variance for the advection in that region.

At many locations, the components of the reconstructed advection may not be significantly different from zero (the prior mean), in the sense that the posterior mean is within one or two standard deviations of zero. In Figure 5 we have displayed only the significant values of the zonal and meridional components. This clearly shows a predominant zonal flow pattern in the western half of the domain, with zonal jets reversing with latitude. There is some strong meridional flow over the crest of the Mid-Atlantic Ridge near the middle of the basin, on the eastern boundary, and over the Walvis Ridge (jutting across the southeast corner). Reversals to the direction of the zonal flow in the western half of the region appear to be associated with tongues in the tracer fields, especially oxygen and salinity. Figure 6 displays the posterior standard deviations of the zonal and meridional velocities. The solution for $u$ is seen to have particularly high uncertainty in the southwest corner of the inversion region, and also in the northwest corner, probably because of poor resolution of smaller-scale mean flows in these regions [the latter is close to the equator and the deep western boundary current supplying the deep layer, while the former contains a large topographic feature, the Rio Grande Rise].

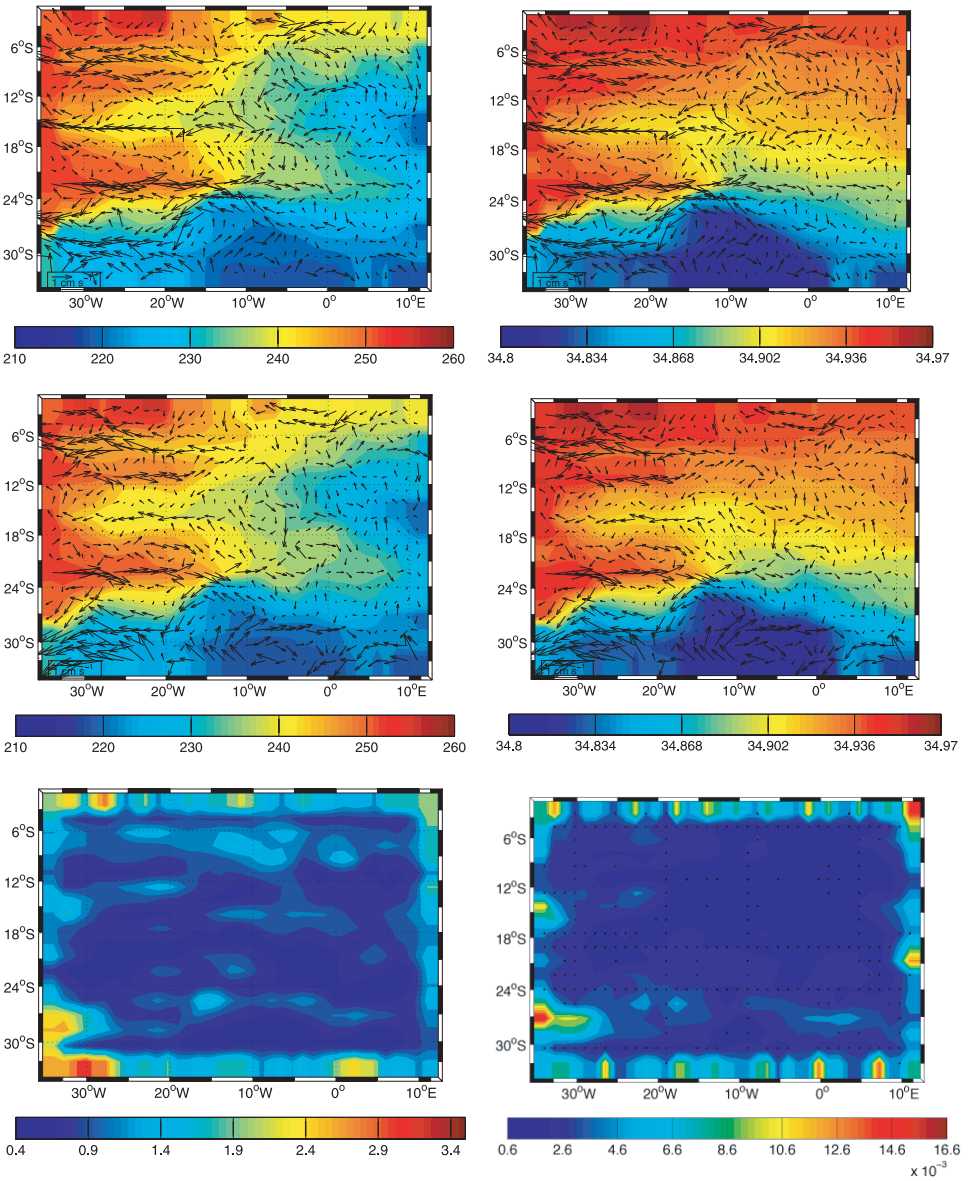


Figure 4. The advection and concentration fields for oxygen (left) and salinity (right) in two draws from the posterior distribution (top and middle); Posterior standard deviations for oxygen (bottom left) and salinity (bottom right).

The analogue of an error bar (or confidence interval) in the Bayesian framework is a credible interval, which has a prespecified (say 95%) posterior probability of containing a given parameter, and these are readily available based on the MCMC output. It should be

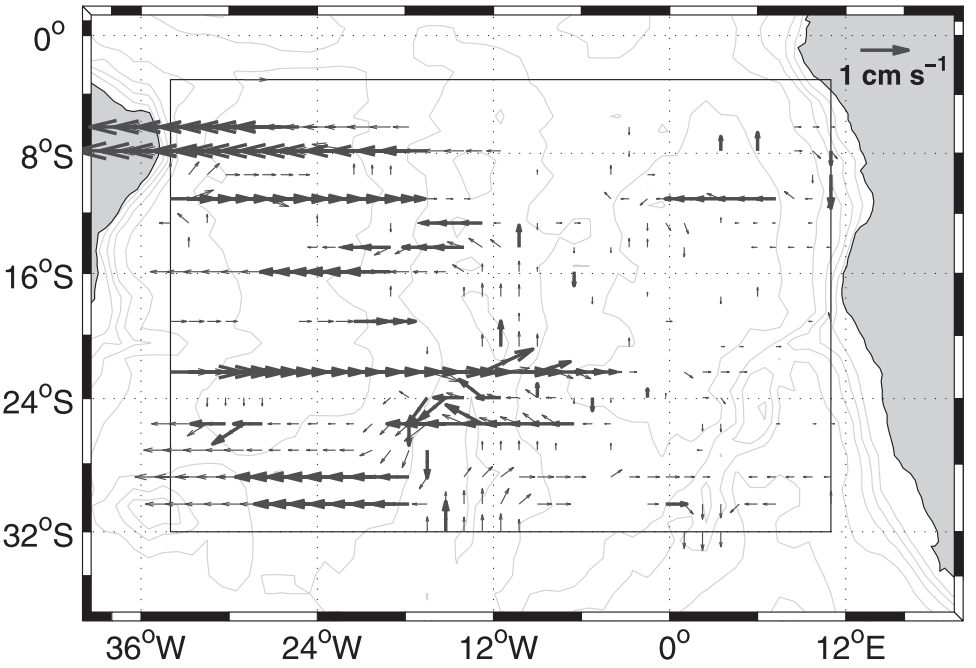


Figure 5. Posterior mean advection that differs significantly from zero [i.e., only the components that differ from zero by at least one (thin arrows) and two (thick arrows) posterior standard deviations]; contours of the bottom topography (interval 1000 m).

noted, however, that caution is needed in interpreting such intervals; the coverage does not hold simultaneously and, with such a large number of parameters in our case (there are about 1400 advection parameters alone), it is almost certain that at least one of the credible

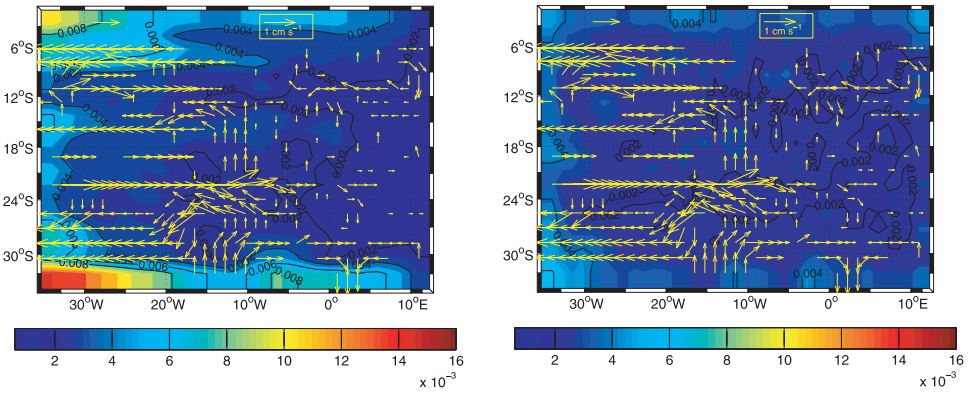


Figure 6. Posterior standard deviation of $u$ (left) and $v$ (right) in units of m/s, and the significant advection (at least one posterior standard deviation from zero).

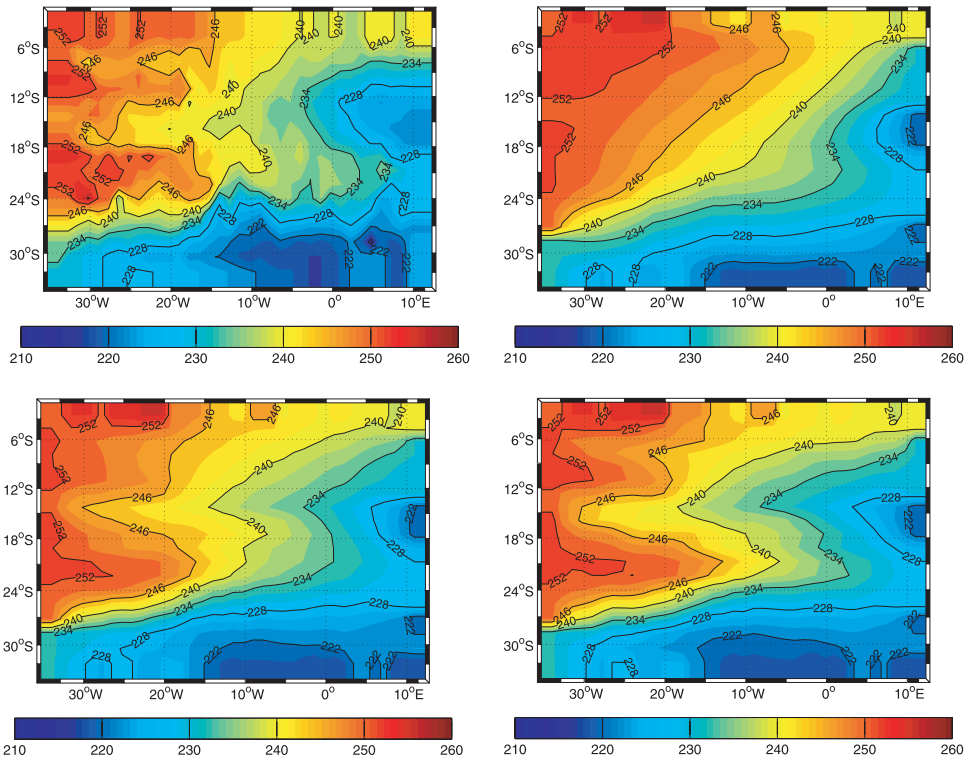


Figure 7. Interpolated oxygen data (top left), and three solutions to the advection-diffusion equation (having the same boundary conditions) with the parameters drawn from the prior. Comparison with the draws from the posterior oxygen concentration distribution in Figure 4 shows that the data are swamping the prior.

intervals does not contain its target parameter. Indeed, we would only expect 95% to do so, and we have no way of telling which 95%. The results in Figure 4 need to be interpreted with this in mind.

Figure 7 compares the interpolated oxygen data with three draws or examples from the prior distribution, using the same boundary conditions (interpolated from the data) in each case. A comparison with the prior is helpful to establish the basic relevance of the assumptions and physical model to the objective of tracer reconstruction. These draws display similar tongues to those from the posterior distribution, and indicate that the prior captures important qualitative features of the oxygen concentration. The only influence of the tracer data in such draws (from the prior) is on the boundary, and they are not solutions to the inversion problem. Their qualitative similarity with the interpolated oxygen data throughout the inversion region implies that the prior information (along with the way we link $v$ to $u$, and to the tracers via the advection-diffusion equation) is adequate in the sense that it is broadly consistent with the data.

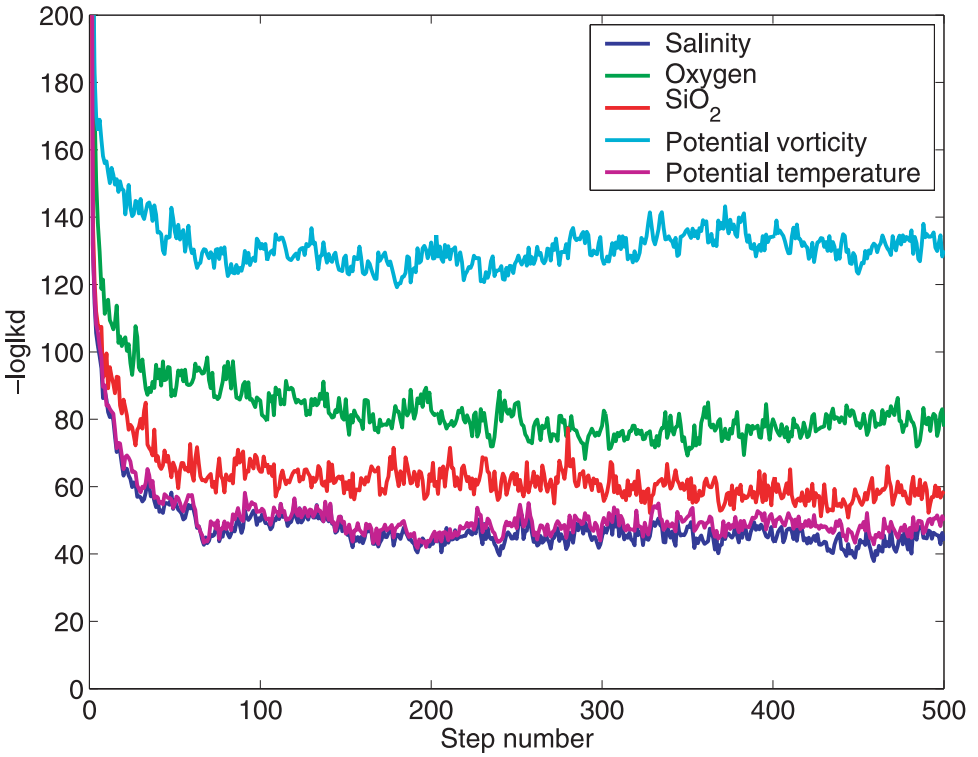


Figure 8. MCMC trace plot showing a time series plot of the simulation output for the parts of minus the log-likelihood (i.e., the error sum of squares normalized by twice the measurement error variance) corresponding to each tracer. Each step on the time axis covers 100 cycles of the moves described at the end of Section 2. It takes about 50 of these steps for the sampler to converge to stationarity; only the subsequent output is used to obtain the inversion results.

The MCMC trace plots in Figure 8 give time series plots of the error sum of squares of the tracer observations, normalized by the measurement error variance, at each 100 cycles of the simulation. These plots show that the sampler settles down rapidly and there are no apparent convergence problems. When the MCMC is slow to explore the range of likely states [the support of the posterior density], averages over the output may be a long way from the true posterior mean, but we see no evidence that this is occurring in our simulations. The extent of disagreement between the tracer data and the posterior can also be assessed through these plots. Here temperature and salinity show the greatest agreement; of course this is to be expected because temperature and salinity are approximately linearly related on a constant density surface.

Histograms of the posterior distribution of the diffusivities are displayed in Figure 9 and show that zonal diffusivity is about seven times the magnitude of meridional diffusivity. Resolution might account for some of this difference, by creating stronger zonal tracer

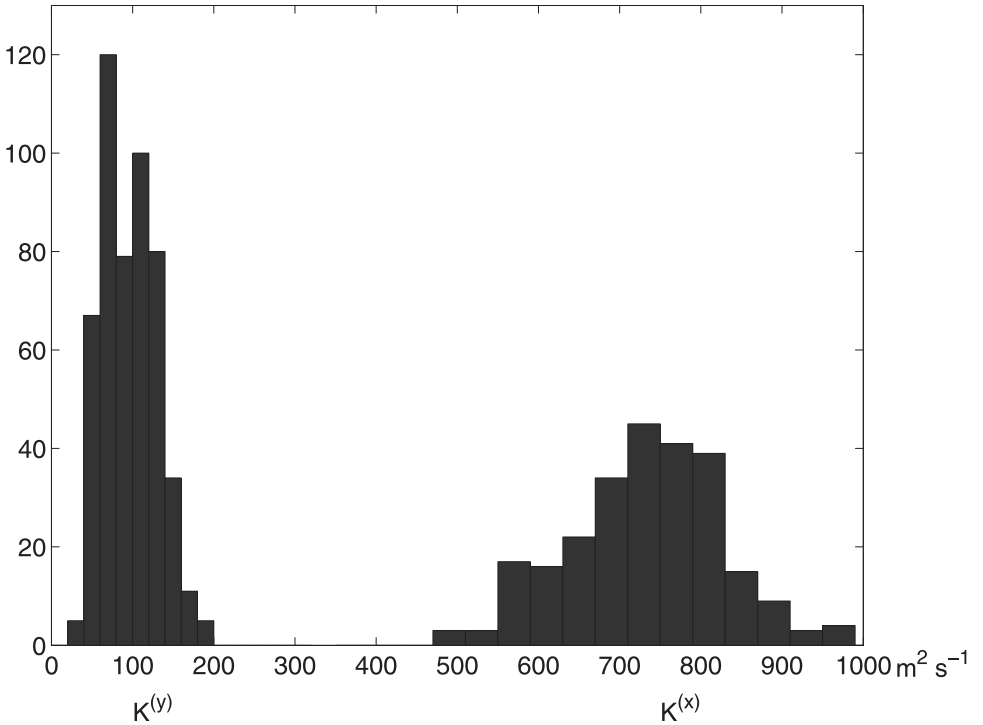


Figure 9. Histograms for the posterior diffusivities $K^{(x)}$ (right), $K^{(y)}$ (left). Units are m$^2$/s.

gradients, but observed tracer concentrations do evolve substantially along the core of the zonal jets, and the same flow and diffusion has to apply to all the tracers. Presumably the dynamics involved in generating the zonal flows also plays a role in enhanced zonal diffusion.

We are fortunate to have direct observations of the abyssal circulation in the western portion of the domain, from neutrally buoyant buoys or floats designed to drift near 2500 m depth (Hogg and Owens, 1999). These floats drifted up to three years and constitute the best direct evidence for the nature of the deep circulation in the region (Treguier *et al.,* 2003). As our domain does not extend to the western boundary a comparison is made over an interior box (Fig. 10). Meridional flow was found to be barely significant in the float observations outside the western boundary and away from the crest of the Mid-Atlantic Ridge, so the comparison is limited to the zonal flow. Within error, the float observations and tracer-based inversion velocities agree. At the northern and southern limits of the domain the float data become very sparse, so we only display the tracer-based zonal velocity.

This comparison can be carried one step further in the case of the zonal flow near 22S, which has received some attention because of its possible role connecting to deep flow around the southern tip of Africa. The posterior mean solution shows a circulation of water

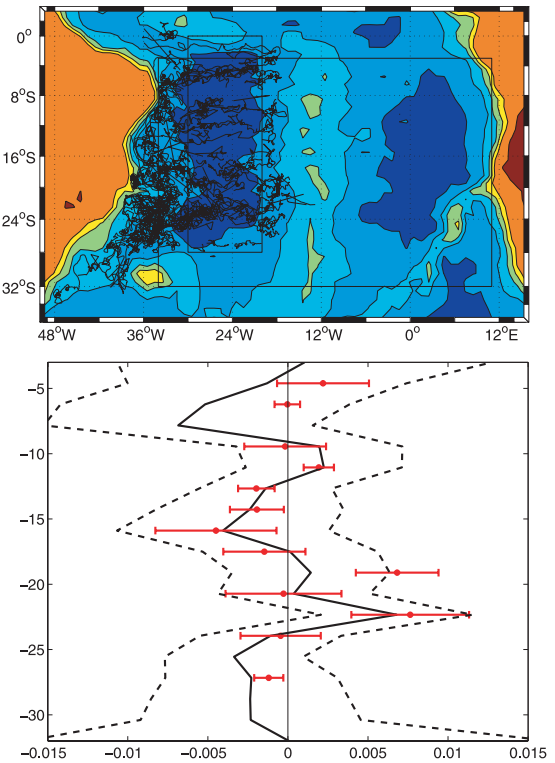


Figure 10. Comparison between posterior mean zonal velocity and mean zonal velocity for float data, units of m/s. Top: geographic distribution of float trajectories illustrating the directly observed flow with floats drifting near 2500 m depth over a period of several years (Hogg and Owens, 1999); the comparison is carried out in the box indicated. Bottom: velocity estimated from float data ± two standard errors (error bars), and tracer-based posterior mean (thin solid line) ± two posterior standard deviations (dashed lines) as a function of latitude.

to the east in a narrow current near 22S, but also substantial flow to the west at higher latitudes. A similar narrow eastward flow in the ocean has been described by Speer *et al.* (1995), who estimated a net transport of 2–3 Sverdrups (Sv, $10^6$ $m^3/s$). To compare our solution to this result we must convert our velocity estimate to a net transport. This is done assuming that the (2D) velocity applies to a layer of finite thickness, in this case 1 km thick. Using this value a calculation of transport in a solution found here shows a value of 2.5 Sv across 22–24S, similar to that found in the ocean. Evidence for the continuation of the flow across the Atlantic is present in both the posterior mean and the tracer fields. Thermal wind calculations over deep layers are consistent with this eastward extension (Speer *et al.,* 1995; Hogg and Thurnherr, 2005).

Near 30S, the flow is dominated by a westward jet away from the crest of the Mid-Atlantic Ridge (Figs. 3 and 5). This flow is mainly supplied by water moving north

along the ridge, entering the domain at the southern boundary. Farther east, the solution near this latitude shows southward flow across a portion of the Walvis Ridge, supplied by convergent flow from the west and east. Arhan *et al.* (2003) inferred branches of deep water flow east of the Mid-Atlantic Ridge crest in the eastern South Atlantic roughly consistent with this pattern. One difference is that their schematic shows a major direct flow from the Mid-Atlantic Ridge crest near 24S to the Walvis Ridge near 30S. Ours shows an indirect connection in a system of zonal currents, with meridional flow significant along the ridges including the Walvis Ridge. A direct connection is inconsistent with the tracer balances enforced in our solutions. Thus, the southward flow across the Walvis Ridge (and into the Cape Basin, the basin east of the Walvis Ridge) is supplied mainly by zonal flow from the region of the Mid-Atlantic Ridge crest, as happens near 24S, but in a less well-defined current. It is possible, however, that the solution in the southern part of the domain is affected by the proximity of the southern boundary of the domain, and more experiments are needed with the boundary set farther south.

## 4. Discussion

In traditional inverse methods in oceanography (and in data assimilation more generally), the physical model (e.g., the thermal wind equation) is connected to the data through minimization of an objective function which represents a trade-off between the data and the model, with a smoothing parameter quantifying the tradeoff. For example, suppose that the parameters are denoted by the vector $x$, and the physical model used to constrain $x$ is $Ax = q$, where $q$ is measured or assumed known. When the statistical model is linear, $y = Bx +$ error, the objective function is $J = (Bx - y)^T(Bx - y) + s(Ax - q)^T(Ax - q)$, where $s > 0$ is the smoothing parameter. Typically neither $A$ nor $B$ is an invertible matrix, but both are known. The sensitivity of the objective function to model error (or perturbations in $q$) is then assessed using the partial derivative of $J$ with respect to $q$. A second partial derivative measures sensitivity to the data $y$, see, e.g., Wunsch (1996, p. 176). These quantities are then interpreted as the extent to which the solution has been constrained by the various components of $y$ and $q$. The Bayesian analogue of this analysis replaces the assumption of a deterministic $x$ by the assumption that $x$ is random, say with a $N(m, I/s)$ prior distribution, where $m$ is the prior mean, $I$ is the identity matrix, and $s > 0$. The constraint $Ax = q$ is replaced by $q = Ax +$ error, and then $q$ is absorbed into $y$, so we have $y = Cx +$ error, where the error is taken as $N(0, I/t)$ say, with $t > 0$. By Bayes theorem the posterior density $\pi(x|y)$ of $x$ given $y$ is proportional to $\pi(y|x)\pi(x)$, or $\exp(-J/2)$, where $J = t(Cx - y)^T(Cx - y) + s(x - m)^T(x - m)$, which is normally distributed. The minimizer of $J$ is the Bayesian analogue of the solution above (the MAP estimate). The extent to which this solution has been constrained by the various components of $y$ can be assessed in the same way.

In our setting (and in many other settings in which Bayesian inversion is used), however, the relationship between $y$ and $x$ takes a more complicated form: $y = C(x) +$ error, where $C$ is a nonlinear function that is not explicitly determined (it is evaluated using a PDE

solver). So the posterior distribution is no longer normal, and its density can only be evaluated up to an unknown constant (because it is impossible to integrate $\exp(-J/2)$ with respect to $x$). Hence it is necessary to extract information about $\pi(x|y)$ using MCMC, and finding the extent to which the solution is determined by the data $y$ (the tracer concentrations) now becomes more challenging because a separate time-consuming MCMC run would be required to find the effect of perturbing any given component of $y$ (and there are thousands of components). Moreover, the effect on the posterior mean or the MAP estimate of $x$ would be difficult to assess because there are thousands of components of $x$. In the general Bayesian inversion setting, it can become computationally prohibitive to assess the effect of perturbations in the data on the solution, even though in principle it is possible to do so.

A referee asked whether the solution we obtain is sensitive to the choice of prior for $v$, or whether the observations are such a strong constraint that the choice of prior is irrelevant. Due to the computational difficulties just mentioned, it is difficult to answer this question precisely. Nevertheless, we expect the solution to be insensitive to the choice of prior for $v$ within the class of Gaussian Markov random fields of the type used, because its only effect is to constrain $v$ to be smooth and it has no influence on the magnitude of $v$. We have experimented in a limited way with various choices of the parameters that specify this prior, and found only minor changes in the results. Moreover, as discussed further below, we have found that the posterior for $v$ is significantly different from its prior distribution, suggesting that the data have had a much stronger effect on the solution than the prior.

A closely related question is whether it is possible to find the source of the information used to determine the solution (e.g., priors vs. observations vs. dynamics). Recall that we have used dynamical information about geostrophic balance as a strict constraint on $u$ in terms of $v$, and the full advection $(u, v)$ necessarily satisfies that constraint. The solution is then constructed using a deterministic link between the true tracer concentrations and $(u, v)$ through the advection-diffusion equation, and a statistical model connecting the true tracer concentrations to the data. All these aspects play a crucial role in the solution, and it is not possible to say that any one of them represents "the source" of information. Nevertheless, in Bayesian statistics it is common practice to compare the posterior distribution of a parameter with its prior distribution, and if the two are markedly different, then one can conclude that the data have swamped the prior in terms of information. In particular, we find that this is the case for the advection, because the posterior mean is significantly different from the prior mean (zero) at many locations (see Fig. 5), for the tracer concentrations (compare the prior and posterior draws in Figs. 4 and 7), and for the diffusivities, because their posteriors appear to be normally distributed rather than exponentially distributed (see Fig. 9).

Although we have begun to construct a plausible circulation and physically consistent tracer distribution, our analysis is lacking in several respects: (1) as a purely 2D solution, it does not allow for vertical advection and diffusion, yet these are traditionally thought to be part of the dominant balance of tracer transport; (2) the dynamics is rudimentary; (3) by

including only minimal dynamics, the statistical inversion is more easily able to fit the data in a nonphysical way. Nevertheless, the simple zonal integration of a random meridional velocity field provides a statistical background for more complete dynamics. More complete dynamics could be readily introduced, for example, by including vertical advection, or by basing the prior on the output of a more complete circulation model.

We conclude by outlining a simple extension of our approach that allows for vertical (cross-isopycnal) advection in a single layer. The idea is to use the 3D advection-diffusion equation in a box $S$ consisting of 3 layers, with the middle layer being the inversion layer for the advection, along with a representation of the vertical velocity $w$ in terms of the vertical diffusivity. The neutral density satisfies an advection-diffusion equation:

$$w \frac{\partial \gamma}{\partial z} = K^{(z)} \frac{\partial^2 \gamma}{\partial z^2},$$

cf. Eq. (12) of McDougall (1991) without the nonlinear terms. Here $K^{(z)} = K^{(z)}(x, y)$ is the same vertical diffusivity as in the 3D advection-diffusion equation linking the tracers to $(u, v, w)$, and $u$ is linked to $v$ in the same way as before. At least five layers of data are needed to evaluate the neutral density gradients, but then we can determine $w$ in terms of $K^{(z)}$. As $K^{(z)}$ reflects sub-grid-scale physics, prior information about its structure is limited, so the absolute value of a 2D GMRF of the type we have been using is appropriate to specify its prior distribution. To specify a prior on the boundary conditions for the tracers, $C_{\partial S}$, it suffices to use two 2D GMRFs (for the top and bottom layers), and a 1D GMRF (along the edge of the middle layer). This extension of our approach will be developed in a forthcoming paper (Herbei *et al.,* in prep.).

*Acknowledgments.* This research was supported in part by NSF Grants ATM-0222244, OCE-0336697, DMS-0207139 and OCE-0117618. The authors thank Ralph Milliff, Andreas Thurnherr and Chris Wikle for helpful discussions.

## APPENDIX

### MCMC acceptance probabilities

The moves of the MCMC sampler described in Section 2 can be specified as reversible jumps (Green, 1995). That is, each move has the form $\Phi \rightarrow \Phi' = F(\Phi, r)$, where $F$ is a non-random operator, $r$ has a given density $q$, and there is a unique $r'$ such that $F(\Phi', r') = \Phi$. The proposal $\Phi'$ is accepted with probability $\alpha(\Phi'|\Phi)$, otherwise the chain stays in its current state $\Phi$. Detailed balance, ensuring that the posterior distribution is the stationary distribution of the Markov chain, is satisfied when

$$\alpha(\Phi'|\Phi) = \min\left(1, \frac{\pi(\Phi'|C_D)q(r')}{\pi(\Phi|C_D)q(r)} \left|\frac{\partial(\Phi', r')}{\partial(\Phi, r)}\right|\right). \tag{A.1}$$

In the case of the ridge type moves, $r' = 1/r$, the Jacobian is $1/r$, and the acceptance probability reduces to

$$\alpha(\Phi'|\Phi) = \min\left(1, \frac{\pi(\Phi'|C_D)}{\pi(\Phi|C_D)r}\right). \tag{A.2}$$

For the random walk Metropolis moves on the components of $\Phi$, we have $r' = -r$, the Jacobean is 1 and the acceptance probability takes the (standard) form

$$\alpha(\Phi'|\Phi) = \min\left(1, \frac{\pi(\Phi'|C_D)}{\pi(\Phi|C_D)}\right). \tag{A.3}$$

## REFERENCES


Arhan, M., H. Mercier and Y.-H. Park. 2003. On the deep water circulation of the eastern South Atlantic Ocean. Deep-Sea Res., *50,* 889–916.

Bennett, A. F. 1992. Inverse Methods in Physical Oceanography, *in* Cambridge Monographs on Mechanics and Applied Mathematics, Cambridge University Press, Cambridge, 364 pp.

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. J. Roy. Stat. Soc. (Series B), *36,* 192–225.

—— 1986. On the statistical analysis of dirty pictures. J. Roy. Stat. Soc. (Series B), *48,* 259–302.

Curry, R. 1996. Hydrobase: A database of hydrographic stations and tools for climatologic analysis. Woods Hole Oceanographic Institution Tech. Report 96-01.

Edwards, C. A. and J. Pedlosky. 1995. The influence of distributed sources and upwelling on the baroclinic structure of the abyssal circulation. J. Phys. Oceanogr., *25,* 2259–2284.

Fox, C. and G. K. Nicholls. 2002. Statistical estimation of the parameters of a PDE. Can. Appl. Math. Quart., *10,* 277–306.

Green, P. J. 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. Biometrika, *82,* 711–732.

Hogg, N. G. and W. B. Owens. 1999. Direct measurement of the deep circulation within the Brazil Basin. Deep-Sea Res., *46,* 335–353.

Hogg, N. G. and A. M. Thurnherr. 2005. A zonal pathway for NADW in the South Atlantic. J. Oceanogr., *61,* 493–507.

Lavine, M. and M. S. Lozier. 1999. A Markov random field spatio-temporal analysis of ocean temperature. Env. Ecol. Stat., *6,* 249–273.

Levitus, S. 1994. World Ocean Atlas 1994 CD-ROM sets. Informal Report 13, National Oceanographic Data Center.

Liu, J. S. 2001. Monte Carlo Strategies in Scientific Computing, Springer, NY, 360 pp.

Lozier, M. S., W. B. Owens and R. G. Curry. 1995. The climatology of the North Atlantic. Prog. Oceanogr., *36,* 1–44.

McDougall, T. J. 1987. Neutral surfaces. J. Phys. Oceanogr., *17,* 1950–1964.

—— 1988. Neutral surface potential vorticity. Prog. Oceanogr., *20,* 185–221.

—— 1991. Parameterizing mixing in inverse models, *in* Dynamics of Oceanic Internal Gravity Waves: Proceedings, Aha Hulikoa Hawaiian Winter Workshop, University of Hawaii at Manoa, January 15–18, 1991, P. Müller, D. Henderson, eds., School of Ocean and Earth Science and Technology, Hawaii.

Paillet, J. and H. Mercier. 1997. An inverse model of the eastern North Atlantic general circulation and thermocline ventilation. Deep-Sea Res., *44,* 1293–1328.

Pedlosky, J. 1996. Ocean Circulation Theory, Springer-Verlag, NY, 453 pp.

Smith, W. H. F. and P. Wessel. 1990. Gridding with continuous curvature splines in tension. Geophys., *55,* 293–305.

Speer, K. G., G. Siedler and L. Talley. 1995. The Namib Col Current. Deep-Sea Res., *42,* 1933–1950.

Stammer, D., C. Wunsch, R. Giering, C. Eckert, P. Heimbach, J. Marotzke, A. Adcroft, C. N. Hill, and J. Marshall. 2002. Global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model. J. Geophys. Res., *107 (C9),* 3118.

Stommel, H. and F. Schott. 1977. The beta spiral and the determination of the absolute vorticity field from hydrographic station data. Deep-Sea Res., *24,* 325–329.

Thurnherr, A. M. and K. G. Speer. 2004. Representativeness of meridional hydrographic sections in the western South Atlantic. J. Mar. Res., *62,* 37–65. [See also http://www.ocean.fsu.edu/sac]

Treguier, A.-M., N. G. Hogg, M. Maltrud, K. G. Speer and V. Thierry. 2003. The origin of deep zonal flows in the Brazil Basin. J. Phys. Oceanogr., *33,* 580–599.

Wikle, C. K., L. M. Berliner and R. F. Milliff. 2003. Hierarchical Bayesian approach to boundary value problems with stochastic boundary conditions. Mon. Weather Rev., *131,* 1051–1062.

Winkler, G. 1995. Image Analysis, Random Fields and Dynamic Monte Carlo Methods, Springer-Verlag, Berlin.

Wunsch, C. 1994. Dynamically consistent hydrography and absolute velocity in the eastern North Atlantic Ocean. J. Geophys. Res., *99,* 14071–14090.

—— 1996. The Ocean Circulation Inverse Problem, Cambridge University Press, 458 pp.

Zhang, H.-M. and N. C. Hogg. 1992. Circulation and water mass balance in the Brazil Basin. J. Mar. Res., *50,* 385–420.

Received: *9 December, 2003;* revised: *8 April, 2005.*