Using tracer observations to reduce the uncertainty of ocean diapycnal mixing and climate–carbon cycle projections

Andreas Schmittner,[1] Nathan M. Urban,[2] Klaus Keller,[2] and Damon Matthews[3]

Received 7 November 2008; revised 25 May 2009; accepted 7 July 2009; published XX Month 2009.

[1] What is the uncertainty of climate–carbon cycle projections in response to anthropogenic greenhouse gas emissions, and how can we reduce this uncertainty? We address this question by quantifying the ability of available ocean tracer observations to constrain the values of diapycnal diffusivity in the pelagic ocean (K_v), a key uncertain parameter representing sub-grid-scale diapycnal (vertical) mixing in physical circulation models. We show that model versions with weak mixing (i.e., low K_v) lead to higher projections of atmospheric CO_2 and larger global warming than do models with vigorous mixing. Slower heat uptake and slower carbon uptake by the oceans contribute about equally to the accelerated warming in the low-mixing models. A Bayesian data-model fusion method is developed to quantify the likelihood of different structural and parametric model choices given an array of observed 20th century ocean tracer distributions. These spatially resolved observations provide strong limits on the upper value of K_v, whereas global metrics used in previous studies, such as the historical evolution of global average surface air temperature, global ocean heat uptake, or atmospheric CO_2 concentration, provide only poor constraints. We compare different methods to quantify the probability of a particular diffusivity value given the observational constraints. One-dimensional, globally horizontally averaged data result in sharper probability density functions compared with the full 3-D fields. This perhaps unexpected result opens up an avenue to objectively determine the optimal degree of aggregation at which model predictions have skill, and at which observations are most helpful in constraining model parameters. Our best estimate for K_v in the pelagic pycnocline is around 0.05–0.2 cm^2/s, in agreement with earlier independent estimates based on tracer dispersion experiments and turbulence microstructure measurements.

Citation: Schmittner, A., N. M. Urban, K. Keller, and D. Matthews (2009), Using tracer observations to reduce the uncertainty of ocean diapycnal mixing and climate–carbon cycle projections, *Global Biogeochem. Cycles*, *23*, XXXXXX, doi:10.1029/2008GB003421.

1. Introduction

[2] Atmospheric CO_2 concentrations are rising faster than ever since continuous monitoring began in 1959 [*Canadell et al.*, 2007]. Increasing anthropogenic carbon emissions is the main cause of this accelerating growth, but reduced uptake of atmospheric CO_2 by ocean [*Le Quere et al.*, 2007] and land are also hypothesized to play a role [*Canadell et al.*, 2007]. These observations are consistent with previous coupled climate–carbon cycle model simulations that predict decreases in terrestrial and oceanic carbon uptake in the future due to changes in climate [*Cox et al.*, 2000; *Dufresne et al.*, 2002; *Friedlingstein et al.*, 2006; *Govindasamy et al.*, 2005; *Jones et al.*, 2003; *Joos et al.*, 1999, 2001; *Matear and Hirst*, 1999; *Matthews et al.*, 2005b; *Sarmiento et al.*, 1998; *Zeng et al.*, 2004]. However, the Coupled Climate–Carbon Cycle Model Intercomparison Project (C^4MIP) [*Friedlingstein et al.*, 2006] shows a large range in the projected magnitude of this feedback between different models. Projected atmospheric CO_2 levels for emission scenario SRES A2 at year 2100 range from ~700 ppmv to ~1000 ppmv, and up to 200 ppmv of this difference can be attributed to differences in the climate–carbon cycle feedback [*Friedlingstein et al.*, 2006]. Thus, the unknown magnitude and uncertainty of the future climate–carbon cycle feedback presents a major hindrance in the assessment of the impacts of carbon emission scenarios.

[3] The reasons for the aforementioned model differences are poorly understood. Although the C^4MIP models showed larger differences in land uptake (−6 to +10 GtC/a), there were also considerable differences in ocean uptake (+4 to +10 GtC/a) by the year 2100 [*Friedlingstein et al.*, 2006]. *Matthews et al.* [2005a] show that differences in the parameterizations of the dependency of terrestrial vegetation growth rates on ambient temperatures have a large effect on

[1]College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA.
[2]Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania, USA.
[3]Department of Geography, Planning and Environment, Concordia University, Montreal, Quebec, Canada.

0886-6236/09/2008GB003421$12.00

carbon uptake on land in future warming experiments, suggesting that this might be a major contributor to the uncertainty range observed in the C^{4}MIP models. Even less is known about reasons for the differences in ocean uptake, although more simplified models (either in terms of physics or biology) apparently show a larger sensitivity of carbon uptake with respect to temperature changes than more complex models [*Friedlingstein et al.*, 2006]. A more detailed comparison between two specific models attributes a two-fold difference in oceanic carbon uptake (4 GtC/a in the UK Hadley Center model versus 8 GtC/a in the French IPSL model at 700 ppmv atmospheric CO_2) due to increasing CO_2 alone (without climate change) to differences in Southern Ocean circulation [*Friedlingstein et al.*, 2003].

[4] The models included in the C^{4}MIP exercise are very heterogeneous and range from box models to zonally averaged and slab mixed layer ocean models to fully three-dimensional coupled atmosphere-ocean general circulation models (AOGCMs). These studies are mostly silent on the question of how probable the different model structures are given the available observational constraints. Without a systematic and probabilistic comparison between observations and the C^{4}MIP models, it remains unclear how to interpret the range covered by the C^{4}MIP models. A careful probabilistic analysis of whether the models are consistent with observations can provide important insights into this question [*Doney et al.*, 2004; *Matsumoto et al.*, 2004].

[5] It is also likely that the C^{4}MIP model simulations do not cover the full scope of uncertainty in possible future climate-carbon interactions, because of (for example) an incomplete representation of the range of unconstrained parameters. Recent Monte Carlo simulations with an atmosphere model suggest that model parameter uncertainties can increase the range of future climate projections considerably [*Murphy et al.*, 2004; *Stainforth et al.*, 2005]. A key uncertain parameter in ocean circulation models is the diapycnal (vertical) diffusivity K_v. The strong sensitivity of the global deep overturning circulation to K_v has been known since the pioneering study by *Bryan* [1987]. Here we investigate the uncertainty in ocean vertical mixing and its effect on future projections of climate and CO_2.

[6] Earlier studies show that tracer distributions in ocean models are sensitive to changes in ocean circulation and ventilation [*Doney et al.*, 2004; *England and Maier-Reimer*, 2001; *Gnanadesikan et al.*, 2004; *Matsumoto et al.*, 2004], but no attempt has been undertaken to quantify the probability of different model structures and parameters given spatially resolved observations of ocean tracer distributions. Probabilistic approaches to climate projections have only been developed in recent years. These pioneering studies were designed to estimate the probability density function (PDF) of the climate sensitivity and used simple model structures constrained only by globally aggregated observations such as the global mean surface air temperature evolution since 1850 [*Andronova and Schlesinger*, 2001], global mean ocean heat content changes [*Forest et al.*, 2002; *Knutti et al.*, 2003; *Tomassini et al.*, 2007], atmospheric CO_2 [*Ricciuto et al.*, 2008], global carbon emissions [*Jones et al.*, 2006], or paleoclimate data [*Annan et al.*, 2005; *Schneider von Deimling et al.*, 2006]. *Tomassini et al.* [2007] found a multimodal probability distribution for K_v, and concluded that these globally averaged metrics do not provide strong limits on the value of K_v. Here we show that multiple physical, geochemical and biogeochemical observations with spatial resolution can provide much stronger constraints on the diapycnal ocean diffusivity. The main goal of this paper, however, is to develop and demonstrate a Bayesian data-model fusion approach for spatially distributed tracer observations that can be used to assess and reduce the uncertainty of future climate projections.

2. Methods

2.1. Model

[7] The UVic Earth System Climate Model [*Weaver et al.*, 2001] of intermediate complexity, includes a coarse resolution (1.8 × 3.6°, 19 vertical layers) three-dimensional general circulation model of the ocean. It has state-of-the-art physical parameterizations such as diffusive mixing along and across isopycnals, eddy induced tracer advection [*Gent and McWilliams*, 1990] and a scheme for the computation of tidally induced diapycnal mixing over rough topography [*Simmons et al.*, 2004]. In order to account for other sources of mixing, a globally constant background diffusivity K_{bg} is added to the tidally induced diffusivity $K_v = K_{tidal} + K_{bg}$. It is unlikely that breaking of internal waves and other unconsidered sources of mixing are spatially constant, but lacking process based parameterizations, K_{bg} is assumed constant within the current model context. It is this background diffusivity K_{bg} that we vary in our sensitivity study, from 0.01 cm^2/s to 0.5 cm^2/s. The tidally induced diffusivity rapidly decays in the water column above the seafloor with an exponential depth scale of 500 m. This results in the background diffusivity determining the value of diapycnal mixing in most parts of the pelagic pycnocline. Observations from the Southern Ocean show that diapycnal mixing is much larger than in other oceans [*Naveira Garabato et al.*, 2004]. We account for these observations by limiting K_v to $\geq$1 cm^2/s south of 40°S. Thus, the variations in K_{bg} affect mixing only in the open ocean north of 40°S.

[8] A simple one-layer atmospheric energy-moisture balance model (EMBM) interactively calculates heat and water fluxes to ocean, land and sea ice, while wind velocities are prescribed from the NCAR/NCEP monthly climatology in the momentum transfer to the ocean and to a dynamic-thermodynamic sea ice model. The model does not use flux corrections. The model of the terrestrial vegetation and carbon cycle [*Meissner et al.*, 2003] is based on the Hadley Center model TRIFFID. The ocean biogeochemical model is based on the nutrient, phytoplankton, zooplankton, detritus (NPZD) ecosystem model of *Schmittner et al.* [2005b] and includes a parameterization of fast nutrient recycling due to microbial activity after *Schartau and Oschlies* [2003]. It solves prognostic equations for two phytoplankton classes (nitrogen fixers and other phytoplankton) as well as for nitrate, phosphate, oxygen, dissolved inorganic carbon, alkalinity, radiocarbon and chlorofluorocarbons as tracers. The biogeochemical/carbon cycle model is described in detail by *Schmittner et al.* [2008]. Biological uptake and

release occurs in fixed elemental ratios of carbon, phosphate, nitrate and oxygen. Calcium carbonate production is parameterized as a fixed ratio of the production of particulate organic matter in the water column. Remineralization of calcium carbonate is determined by instantaneous sinking with an e-folding depth of 3500 m.

[9] The ensemble consists of eight models with K_{bg} = (0.01, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, and 0.5). (In the following, for brevity, we omit the units of K_{bg}, which are in cm^2/s.) Each model version is restarted from an 8000-year control integration with K_{bg} = 0.15, and spun up for an additional 3000–4000 years (longer for smaller K_{bg}) using constant preindustrial forcing until climate and carbon cycle are in quasi-equilibrium. Initially atmospheric CO_2 is fixed at 280 ppmv, but for the last ~1000 years of the spin up it is calculated interactively. Equilibrium is determined if changes in atmospheric CO_2 are less than 5 ppmv per 1000 years, so that at the end of the spin up atmospheric CO_2 is within ±5 ppmv of ice core measurements of its preindustrial value of 280 ppmv (Figure 1) for all model versions. Subsequent estimates [*Crowley*, 2000] of historical forcing from year 1800 to 1998 AD are applied, considering changes in solar insolation, volcanic and anthropogenic aerosol and greenhouse gases, followed by CO_2 emission scenario SRES A2 until 2100 and a linear decrease of emissions to zero from year 2100 to 2300 (Figure 1). Solar, aerosol and non-CO_2 greenhouse gas forcings have been held constant at 1988–1998 levels for the future simulations.

2.2. Observations

[10] We calculate probability densities for nine three-dimensional tracer distributions from two databases. Temperature (T) [*Locarnini et al.*, 2006], salinity (S) [*Antonov et al.*, 2006], phosphate (PO_4) [*Garcia et al.*, 2006a], apparent oxygen utilization (AOU) [*Garcia et al.*, 2006b], and preformed phosphate ($P^* = PO_4 - AOU/170$) are taken from the World Ocean Atlas 2005 (WOA05, data downloaded from ftp.nodc.noaa.gov/pub/data.nodc/woa/WOA05nc) and radiocarbon ($\Delta^{14}C$), chlorofluorocarbon 11 (CFC11), dissolved inorganic carbon (DIC), and alkalinity (ALK) are adopted from the Global Ocean Data Analysis Project (GLODAP) [*Key et al.*, 2004], Both databases provide data on a $1 \times 1^\circ$ grid with 33 vertical levels. The observations are averaged onto the $1.8 \times 3.6^\circ$ model grid with 19 vertical levels. GLODAP data represent the 1990s and are compared with the decadal model mean from 1990 to 2000, whereas WOA05 data represent the 1950–2000 and are compared to the model mean during this period.

2.3. Observation Error Estimates

[11] To quantitatively compare observations to model projections requires an estimate of the observation errors. The error size determines how far from the data a model can be and still remain consistent with the observations. Spatially variable error estimates for the observations (σ_{Oi}) are available for $\Delta^{14}C$, CFC, DIC and ALK from the GLODAP data set representing errors resulting from the objective analysis (mapping) procedure used to interpolate and extrapolate observations to a global grid. Because of the sparse observations the GLODAP error estimates are horizontally correlated with a correlation length scale of 10–20°. Thus the GLODAP errors are simply averaged onto the model grid.

[12] The WOA05 provides the standard error for each unanalyzed variable, which is the standard deviation of the mean divided by the square root of the number of observations in each grid cell. Following the recommendation in the WOA05 documentation (available at http://www.nodc.noaa.gov/OC5/WOA05/pr_woa05.html) the error due to the objective analysis is estimated as the difference between the value of the analyzed field and the mean at each grid cell containing observations. For T and S, which have observations at almost all grid points, this error is horizontally uncorrelated. The global horizontal root mean square is calculated at each depth level, representing the (horizontally uniform but vertically varying) mapping error. This mapping error is added to the standard error to yield the spatially variable total error estimate.

[13] For PO_4, preformed PO_4, and AOU the data density is too sparse to calculate an error estimate due to the mapping procedure, because the calculation outlined above can only be performed for points that include data. For this reason we do not use the analyzed fields but rather we use the unanalyzed mean (the average of the raw observations in any given $1 \times 1^\circ$ data grid box). This limits the number of grid cells to those containing observations. The observations are averaged onto the model grid, and model grid cells without observations are discarded in the analysis. In this case the total error of the observations is only the standard error of the mean (no mapping error). For all WOA05 variables the total errors are assumed to be horizontally uncorrelated and are hence averaged onto the model grid and divided by $2.55 = \sqrt{3.6 \times 1.8}$ in order to account for 6.48 independent data grid boxes in one model grid cell.

2.4. Statistical Analysis

[14] We assess the compatibility of different diapycnal diffusivities with observed tracer measurements using Bayesian inference to compute the relative probability of each of the eight diffusivities in our ensemble implied by each of the nine tracer fields. Two different methods are used in the model assessment. The first computes the root mean squared (RMS) error (E) for each model, including the full three-dimensional (3-D) spatial fields of observations. This method neglects the correlation of the errors and requires the size of the errors to be specified. The second method considers the correlation of the errors and determines the error magnitude and bias endogeneously from the data-model residuals. However, because of computational constraints it uses only one-dimensional data (globally horizontally averaged depth profiles). Both methods, as well as the relations between them, are described in detail in sections 2.4.1 and 2.4.2.

2.4.1. Three-Dimensional Method

[15] Models that greatly differ from the observations are judged less probable than models whose deviations from the data are small. To quantify this intuition, it is necessary to mathematically specify what "small" means. We introduce an error estimate σ to set the scale against which data-model discrepancies are evaluated. These deviations are deemed large or small relative to the magnitude of σ. Observations can differ from model predictions for two reasons: model

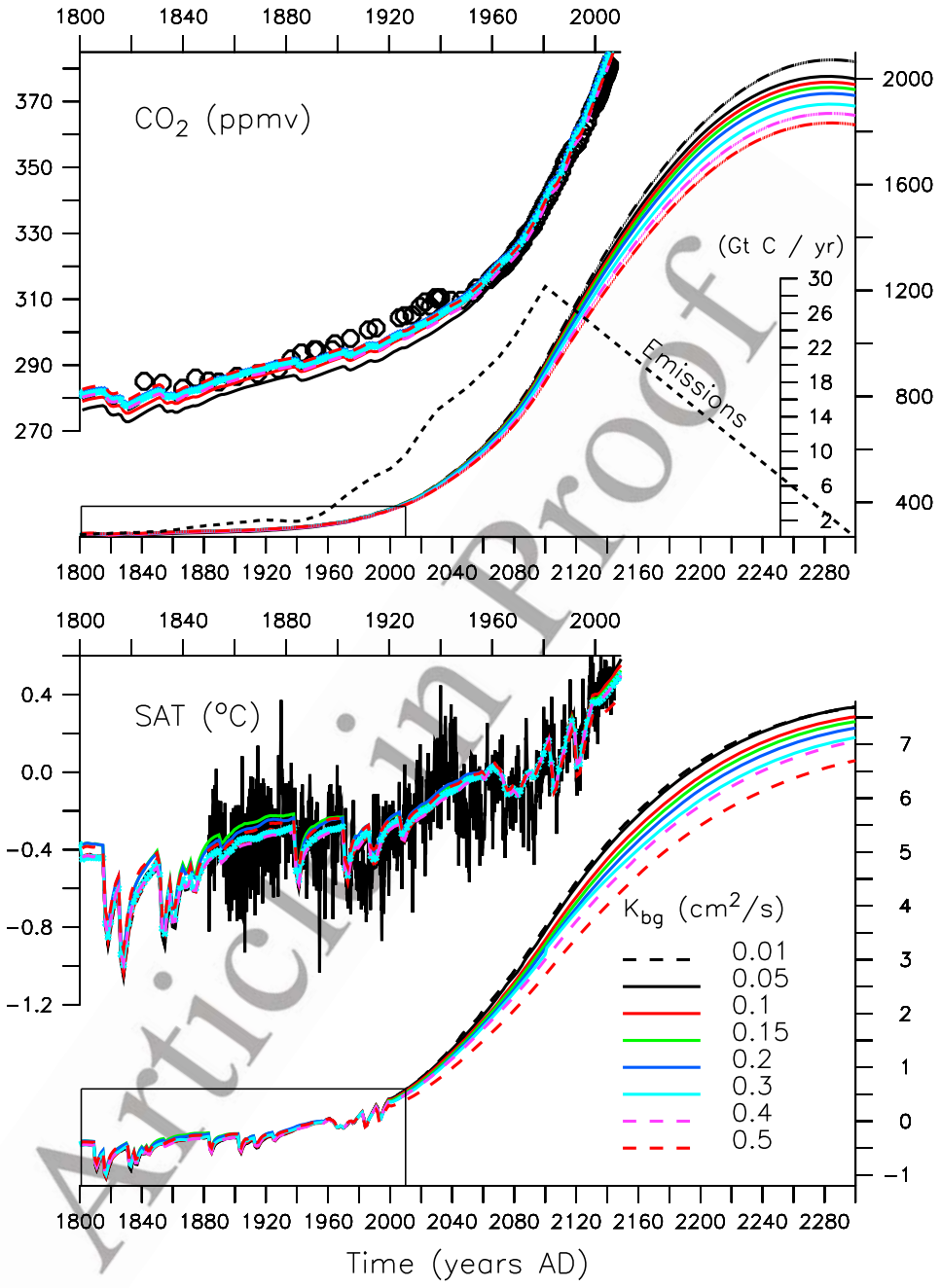

Figure 1. Hindcasts and projections of (top) atmospheric CO_2 concentration and (bottom) near-surface air temperature (SAT) anomalies from the 1960–1990 levels for model versions with different values of K_{bg}. The emission scenario (SRES A2 until year 2100 and linear decrease until year 2300 afterward) is shown as the heavy dotted line in Figure 1, top, with the scale in the bottom right corner ranging from 0 to 30 Gt C/a. For reference: current 2007 levels are about 8.5 Gt C/a (J. G. Canadell et al., Carbon budget and trends 2007, available from the Global Carbon Project, http://www.globalcarbonproject.org, 26 September 2008). The insets in the upper left region of each plot show a zoom into the hindcast period (1800–2007) including CO_2 observations from Mauna Loa [*Keeling and Whorf*, 2005] and ice cores [*Neftel et al.*, 1994] (circles) and temperature observations from the HadCRUT3 [*Brohan et al.*, 2006] data set (black noisy line).

structural error, and observational/measurement error. The quality of data-model agreement depends on how large we judge these errors to be (see section 3.7). However, errors can be difficult to estimate a priori (before seeing the observational data), especially when model structural errors are substantial. Observational errors usually can be estimated from known properties of the measurement system (section 2.3), but the size of the model error typically

cannot be determined without comparing the model output to observations.

[16] To assess model skill for each tracer i, we calculate the error-weighted mean squared error

$$\overline{E_i^2} = \overline{\left(\frac{O_i - M_i}{\sigma_i}\right)^2}. \quad (1)$$

The overbar denotes the global, volume-weighted average. Deviations of each modeled 3-D tracer field $M_i = \overline{M_i}(x, y, z) = \overline{M_i} + M_i'$ from the observations $O_i = O_i(x, y, z) = \overline{O_i} + O_i'$ are weighted by a combined error estimate ($\sigma_i^2 = \sigma_{Oi}^2 + \sigma_{Mi}^2$) for the observations σ_{Oi} and the model σ_{Mi}. (The prime denotes the deviation from the global mean.) Our methods for estimating the observation and model errors are discussed in section 3.4.1.

[17] The models often show bias relative to the observations, so that their mean prediction differs from the mean of the observations. To distinguish between the amount of error introduced by model bias and the amount of error unrelated to bias, we also consider the bias-corrected RMS error. This error is calculated by subtracting the global mean bias $b_i = \overline{O_i} - \overline{M_i}$, so that the bias-corrected residuals $O_i - M_i - b_i$ have zero mean. The bias-corrected RMS error is then $\overline{E_i'^2} = \overline{\left(\frac{O_i - M_i - b_i}{\sigma_i}\right)^2} = \overline{\left(\frac{O_i' - M_i'}{\sigma_i}\right)^2}$. The error $\overline{E_i'^2}$ excludes information about the global mean data-model misfit.

[18] The probabilistic model assessment, however, includes information on the global mean data-model misfit using equation (1). Assuming the errors are independent and identically distributed random variables, the probability density

$$L(O_i|K_{bg}) \propto \exp\left(-\frac{1}{2}E_i^2\right) \quad (2)$$

is the likelihood that the observations O_i could arise from the model with parameter K_{bg}. Above, $E_i^2 = \overline{E_i^2} \times N$ is the (volume-weighted, error-weighted) sum of squared errors, equal to the mean squared error $\overline{E_i^2}$ times the number of data points (N). More precisely, assuming a known error σ, the probability in equation (2) is a normal likelihood function: the observations are assumed to be drawn from a normal distribution with mean centered on the model output ($O \sim N(\mu = M, \sigma^2)$). Bayes' theorem states that the posterior probability density function (PDF) for K_{bg} is proportional to the product of the likelihood of the observations with the prior PDF of K_{bg}, p(K_{bg}):

$$p(K_{bg}|O_i) \propto L(O_i|K_{bg}) \times p(K_{bg}). \quad (3)$$

The prior PDF quantifies expert judgment about the value of K_{bg} before having assimilated the observational data. We adopt a uniform prior PDF for K_{bg}, giving equal prior probability to each model run. The posterior probability of a particular model run is the product of how likely the data are given the model output, weighted by how probable the run is judged to be a priori. See *Gelman et al.* [2004] for a basic reference text on Bayesian methods.

[19] If the errors in different tracers are independent of each other, which is generally not the case, as discussed in section 3.3.5, likelihoods for individual tracers can be multiplied to yield the combined likelihood of all tracers, $L(O|K_{bg}) = \prod_i L(O_i|K_{bg})$. Probability-weighted projections for a climate variable T are obtained by averaging over the possible values of K_{bg},

$$\overline{T} = \int_{Kbgmin}^{Kbgmax} T(K_{bg}) \cdot p(K_{bg}|O)\, dK_{bg}, \quad (4)$$

if the PDF is defined on the interval [K_{bgmin}, K_{bgmax}].

2.4.2. One-Dimensional Method

[20] The above 3-D method ignores spatial autocorrelation of the data-model residuals, $R_i = O_i - M_i$, which is known to lead to overconfident parameter estimates [*Zellner and Tiao*, 1964]. In addition, the above formulation presumes that the residual error σ is known, but as discussed in section 2.4.1, it can be difficult to estimate a priori. Here we develop a relatively simple and computationally efficient method to estimate the combined effects of observation errors and model structural errors endogenously from the overall data-model misfit. This method is more computationally expensive than the 3-D method, so we apply it to small 1-D aggregated data sets instead of to the full 3-D spatial fields.

[21] When the errors are uncorrelated, only their magnitudes σ_i need to be specified. If the errors are correlated, the correlation between errors must be specified in addition to their magnitudes. We generalize from the error variances σ_i^2 to an error covariance matrix Σ, which includes the error variances and the spatial correlations between points. In the 3-D method we use the weighted sum of squared errors, $\sum_i (O_i - M_i)^2/\sigma_i^2$, to quantify model skill. This error measure is not appropriate when the errors are correlated. Correlated errors effectively provide fewer independent data points than uncorrelated errors. An appropriate measure should penalize models less harshly when correlation is present, since fewer independent data are assimilated. To include correlation the sum of squared errors generalizes to a quantity involving the error covariance matrix, known as the Mahalanobis distance [*Mahalanobis*, 1936], which appears in the multivariate normal distribution:

$$E_i^2 = (O_i - M_i)^T \Sigma_i^{-1} (O_i - M_i). \quad (5)$$

This expression reduces to the sum of squared errors when the covariance matrix is diagonal with entries σ_i^2, i.e., when it contains only variances but no off-diagonal correlations. (In the remainder of this section we omit the subscript i when referring to each tracer.)

[22] Only small covariance matrices are used here because matrix inversion is computationally expensive, growing with the cube of the number of data points. To reduce the size of the covariance matrix to a computationally feasible magnitude, we consider only a 1-D globally averaged spatial field of tracer data $O(z)$ and $M(z)$ as a function of depth z. Each field is reduced to 18 data points (depths),

allowing the assimilation to run for all tracers within a few minutes on a single workstation. A small 2-D latitude-depth grid may also be computationally feasible to assimilate in this manner, but this exercise is beyond the scope of this proof-of-concept study. We assume the covariance matrix Σ is given by a stationary squared-exponential covariance function between depths z_j and z_k, $\Sigma_{jk} = \sigma^2 \exp(-|z_j - z_k|^2/\lambda^2)$, where σ^2 is the residual variance and λ is a range or correlation length parameter. A squared-exponential covariance function implies a smooth (infinitely differentiable) spatial process and is chosen because prior judgment, as well as inspection of the residuals, suggests that the globally averaged model structural error varies smoothly with depth. Including the possibility of a constant model bias, b, the observations are assumed to be drawn from a multivariate normal likelihood centered on the bias-corrected model output ($O \sim MVN(\mu = M + b, \Sigma)$).

[23] In the previously discussed 3-D method, the residual error σ, the correlation length λ, and the model bias b are assumed known constants (with $\lambda = 0$, and b = 0 or set to the difference in observational and model means). These constants may differ between tracers. In the 1-D method applied here, we relax these assumptions by treating the three constants as unknown statistical parameters. The full Bayesian approach, which we approximate, is to calculate a joint posterior PDF for all the uncertain parameters, including the model parameter K_{bg} and the three statistical parameters. By Bayes' theorem, this posterior probability is proportional to the product of the likelihood of the observations with the prior probability of the parameters,

$$p(K_{bg}, \sigma, \lambda, b|O) \propto L(O|K_{bg}, \sigma, \lambda, b) \times p(K_{bg}, \sigma, \lambda, b). \quad (6)$$

We are most interested in the probabilities of the different model diffusivities, not of the statistical parameters. We can obtain the posterior PDF $p(K_{bg}|O)$ for K_{bg} alone by integrating the joint posterior $p(K_{bg}, \sigma, \lambda, b|O)$, equation (6), with respect to the three statistical parameters:

$$p(K_{bg}|O) = \iiint p(K_{bg}, \sigma, \lambda, b|O) d\sigma d\lambda db. \quad (7)$$

However, for computational simplicity, we avoid performing this integral by fixing the statistical parameters at their best fit values σ^*, λ^*, β^*. This gives an approximate proportionality

$$p(K_{bg}|O) \approx p\left(K_{bg}|O, \sigma^*, \lambda^*, b^*\right) \propto L\left(O|K_{bg}, \sigma^*, \lambda^*, b^*\right) \times p\left(K_{bg}, \sigma^*, \lambda^*, b^*\right). \quad (8)$$

Fixing the statistical parameters ignores their uncertainty but still accounts for the presence of model error, bias, and correlation. These quantities are estimated from the data-model misfit instead of assumed from expert prior judgment. The best estimate for σ^*, λ^*, β^* is obtained by numerically maximizing the posterior probability (equation (6)) using a global optimization method [*Storn and Price*, 1997] to account for potential multimodality. Posterior maximization is analogous to maximum likelihood estimation [*Lehmann and Casella*, 2003], except that the likelihood is modified by prior constraints on the parameters. The statistical parameters are separately optimized for each tracer, allowing the estimated residual structure to vary between tracers. For every tracer, the parameters are also reoptimized for each member of the ensemble. In other words, the statistical parameters are allowed to depend on K_{bg}. The logic behind this assumption is that the model error depends on the model parameters, since poorly fitting models should have larger model error and bias. We linearly interpolate the posterior probability onto a regular grid of K_{bg} and normalize the integral to unity to arrive at a proper probability density function.

[24] We choose a uniform prior for the model parameter K_{bg}. The correlation length prior is $p(\lambda)$ = lognormal(5.5, 0.5^2). That is, ln(λ) is normally distributed with a mean 5.5 and standard deviation 0.5, which puts most of its probability mass between 0 and 600 m and practically excludes larger correlation lengths. Large correlation lengths imply strong communication between the surface and the deep ocean, which is contrary to the layered nature and highly stratified vertical structure of the ocean. We use a joint prior for the residual variance and bias, $p(b/\sigma) = N(0, 0.5^2)$. This prior is selected so the model bias for the best K_{bg} value is assumed to be likely smaller than the residual error (i.e., b/σ is near zero). This gives low prior weight to models with large biases, where "large" is quantified relative to the size of the bias-corrected error, σ. Exploratory analysis indicates that an improper, unbounded uniform prior for the range or bias parameters can lead to ill-conditioned covariance matrices and nonrobust results for the K_{bg} posterior distribution.

3. Results

3.1. Global Metrics

[25] Observed atmospheric CO_2 concentrations and global mean surface air temperatures are simulated roughly equally well in all model versions, irrespective of the value of K_{bg} (Figure 1). This is also true for the ocean heat content changes, which are very similar in all simulations (Figure 2). As already concluded by *Tomassini et al.* [2007], these globally aggregated observations provide relatively poor constraints on K_{bg}. The model suggests, however, that this situation might change in the future, because the simulations for different K_{bg} values diverge notably during the 21st century. For example, at year 2100 differences in CO_2 concentrations are about 70 ppmv (Figure 1). This suggests also that variations in diapycnal diffusivity alone can account for about 25% of the range in the C^{4}MIP models. At year 2300 differences in CO_2 concentrations are more than 200 ppmv. Differences in projected global average surface air temperatures are 0.8°C in model year 2100 and 1°C in year 2300.

[26] A 1°C variance with respect to a 7°C global warming might not seem significant compared to the much larger variance in the C^{4}MIP or IPCC AR4 model projections. However, it is important to remember that the multimodel spread is caused by numerous differences in model structures and parameter values, whereas here we have only varied a single parameter.

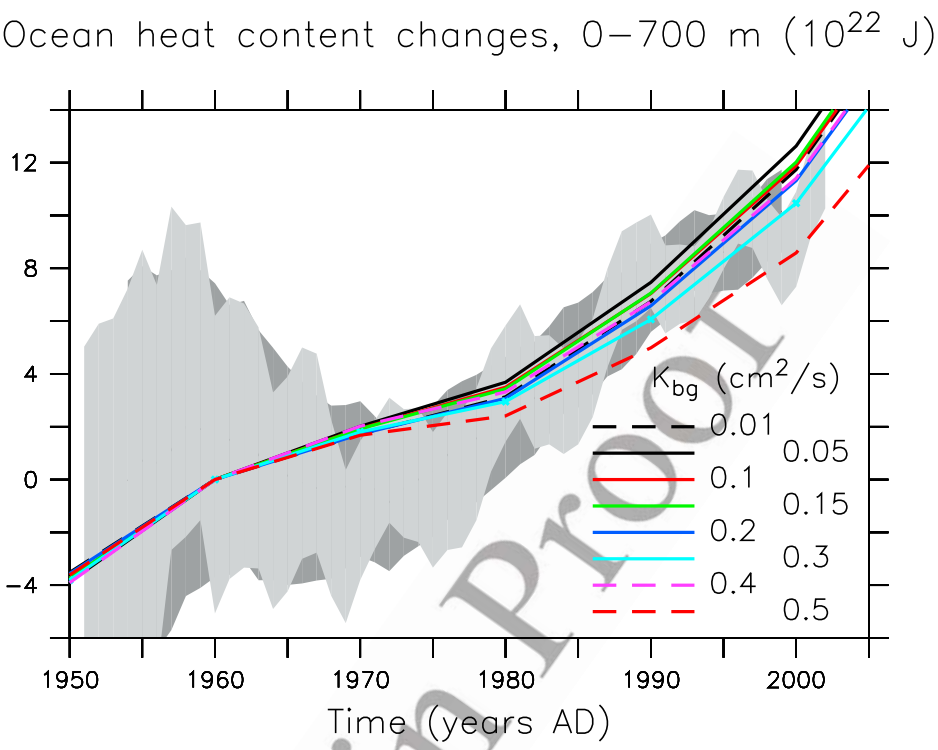

Figure 2. Upper ocean (0−700 m) heat content changes (from year 1961) as simulated by the different model versions (lines) compared to observations (gray shading) from *Domingues et al.* [2008]. The dark gray shading denotes 9-year running mean values for comparison with decadal averages plotted for the model simulations. The light gray shading shows 3-year averages.

3.2. Influence of Diapycnal Mixing on Climate−Carbon Cycle Projections

[27] Larger diapycnal mixing leads to faster oceanic uptake of heat and CO_2 in the model. Both effects tend to delay and reduce atmospheric warming. Faster CO_2 uptake leads to lower atmospheric CO_2 concentrations and thus reduced radiative forcing, whereas faster heat uptake leads to slower warming of surface waters and therefore delayed warming of surface air temperatures. We separate these two effects by comparing a simulation with weak mixing (K_{bg} = 0.1) forced with interactive CO_2 to one forced with a prescribed CO_2 evolution (and thus radiative forcing) taken from a run with vigorous mixing (K_{bg} = 0.5). The difference in surface air temperature evolution between these two simulations is due only to the effect of slower ocean heat uptake. The effect of different ocean carbon uptakes is quantified by comparing the simulation with prescribed CO_2 to the fully coupled run with K_{bg} = 0.1 (Figure 3). The global surface air temperature increase in the run with prescribed CO_2 evolution is about halfway between the experiments with high and low K_{bg}. About 55% (0.5 K) of the reduced warming of air temperatures in the high-K_{bg} versus the low-K_{bg} simulation is explained by differences in ocean heat uptake alone, and 45% is caused by faster CO_2 uptake. This demonstrates that both effects, slower heat uptake and slower carbon uptake, provide similar contributions to the reduced warming in the high-mixing model projections.

[28] We analyze the sensitivity of land (ΔC_L) and ocean (ΔC_O) carbon uptake until year 2100 with respect to changes in atmospheric CO_2 ($\beta_L = \Delta C_L^u/\Delta C_A^u$; $\beta_O = \Delta C_O^u/\Delta C_A^u$) and climate ($\gamma_L = (\Delta C_L^c - \beta_L \Delta C_A^c)/\Delta T^c$; $\gamma_O = (\Delta C_O^c - \beta_O \Delta C_A^c)/\Delta T^c$) following *Friedlingstein et al.* [2006], where c and u superscripts denote the coupled and uncoupled (constant climate) runs, respectively, ΔT is the global mean surface air temperature change and ΔC_A is the atmospheric CO_2 anomaly. As expected, the land sensitivities ($\beta_L(K_{bg} = 0.1) = \beta_L(K_{bg} = 0.5) = 1.4$ GtC/ppm; $\gamma_L(K_{bg} = 0.1) = -114$ GtC/K; $\gamma_L(K_{bg} = 0.5) = -116$ GtC/K) are very similar between the different K_{bg} simulations. (The C^{4}MIP range for β_L is 0.2 to 2.8 GtC/ppm and for γ_L it is −20 to −177 GtC/K.)

[29] However, ocean carbon uptake due to changes in atmospheric CO_2 alone is 30% smaller in the low-mixing model ($\beta_O(K_{bg} = 0.1) = 1$ CtC/ppm) compared to the high-mixing model ($\beta_O(K_{bg} = 0.5) = 1.4$ CtC/ppm). This suggests that differences in ocean diapycnal mixing alone can explain half the range of β_O in the C^{4}MIP models (0.8–1.6 GtC/ppm) and reemphasizes the important role of diapycnal mixing on anthropogenic carbon uptake by the ocean. There are, of course, other processes that additionally determine ocean carbon uptake (under fixed climate), such as the strength of the overturning circulation and convection, mixed layer depths, and air-sea gas exchange (driven by factors such as sea ice and wind velocities).

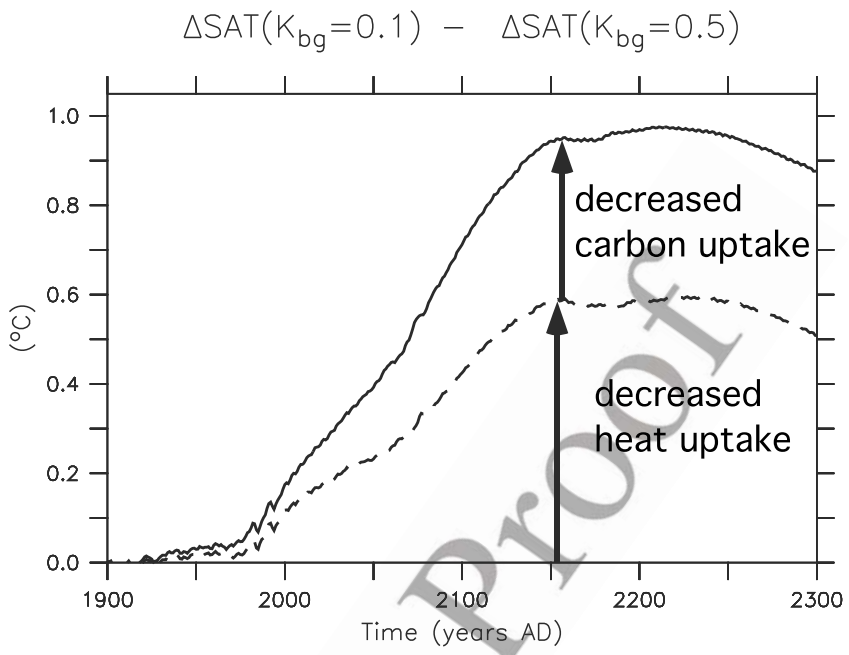

Figure 3. Effects of reduced ocean heat and carbon uptake on projected warming resulting from smaller vertical mixing. The solid line shows the global mean surface air temperature anomaly for a run with low vertical mixing (K_{bg} = 0.1) minus that from a run with high vertical mixing (K_{bg} = 0.5), including both effects, reduced heat and reduced carbon uptake. The dashed line shows the effect of reduced heat uptake alone from a sensitivity experiment with K_{bg} = 0.1 in which atmospheric CO_2 evolution is prescribed to be identical to that from the K_{bg} = 0.5 simulation. The effect of reduced carbon uptake shows as the difference between the dashed and solid lines.

[30] Ocean carbon uptake decreases in the model simulations as climate warms because of increasing stratification of the upper ocean. A greater weakening of the ocean carbon sink corresponds to more negative values of γ_O. In the high-mixing models this decrease is larger ($\gamma_O(K_{bg}$ = 0.5) = −45 GtC/K) than in the low-mixing models ($\gamma_O(K_{bg}$ = 0.1) = −31 GtC/K). The C^{4}MIP models range from −14 to −67 GtC/K (though it is worth noting that the γ_O value of −67 GtC/K is the result of a box model; the next largest C^{4}MIP model value of γ_O is −46 GtC/K). At year 2100 the ocean takes up 4.8 GtC/a in the low-mixing model versus 6.2 GtC/a in the high-mixing model. Most (8 out of 11) C^{4}MIP models lay within that range of ocean carbon uptake.

[31] The strength of positive climate–carbon cycle feedbacks can be quantified by the feedback gain ($g = 1 - \Delta C_A^u/\Delta C_A^c = -\alpha(\gamma_L + \gamma_O)/(1 + \beta_L + \beta_O)$, where $\alpha = \Delta T^c/\Delta C_A^c$ is the transient climate sensitivity) [*Friedlingstein et al.*, 2006]. The effects of higher β_O and larger (negative) γ_O almost completely compensate each other, but because of the larger transient climate sensitivity (α (K_{bg} = 0.1) = 0.0060 K/ppm versus α (K_{bg} = 0.5) = 0.0055 K/ppm), there is a modest (10%) increase in gain in the low-mixing model ($g(K_{bg}$ = 0.1) = 0.2) compared to the high-mixing model ($g(K_{bg}$ = 0.5) = 0.18). The range of g in the C^{4}MIP models is 0.04–0.31, which includes differences in both terrestrial and oceanic carbon cycle contributions to the total climate–carbon cycle feedback, in addition to different values of transient climate sensitivity. According to our analysis, while different K_{bg} values can explain a substantial portion of the range of ocean carbon uptake between models, K_{bg} differences can explain only a relatively small proportion of the intermodel range in net climate–carbon cycle feedback strength.

3.3. Model Assessment Using Spatially Resolved Ocean Tracer Observations

3.3.1. Model and Observation Error Estimates

[32] Assessing model skill requires an estimate of the discrepancy between observations and model predictions. The 3-D method's likelihood function, equation (2), assumes that the standard deviation of the data-model residuals (σ) is known. The 1-D method estimates this error from the residuals by an optimization procedure (section 2.4.2). For the 3-D method we choose to determine the residual error by more informal means. By definition, the residual error $\sigma_i = \sqrt{\sigma_{Oi}^2 + \sigma_{Mi}^2}$ should be similar to the standard deviation of the residuals, $\sigma_i \approx SD(O_i - M_i)$. For the 3-D method we choose the model error σ_{Mi} such that this is the case for one of the best fitting models (K_{bg} = 0.15). (See Table 1 for values.) This model error is then applied to all ensemble members.

[33] The model error estimates σ_{Mi} can also be interpreted as measures of model quality; they can be used for different models and are suitable for model intercomparisons. For example, for temperature and salinity, the values in the second row of Table 1 (3-D data and σ_O = 0) correspond to the global RMS error. They can be compared to those

t1.1 **Table 1.** Estimated Model Error σ_M for Different Assumptions and Tracers as Calculated From the $K_{bg} = 0.15$ Case[a]

t1.2			T (K)	S	Δ^{14}C (per mil)	CFC11 (pM)	PO_4 (μM)	AOU (mM)	DIC (μM)	ALK (μM)	P^* (μM)
t1.3	3-D	$\sigma_O \neq 0$	0.90	0.18	0.0	0.24	0.20	25	19	13	0.16
t1.4		$\sigma_O = 0$	0.92	0.19	20	0.35	0.20	25	25	15	0.16
t1.5	1-D	$\sigma_O = 0$	0.24	0.079	5.9	0.053	0.064	3.7	9.0	6.6	0.062

t1.6 [a]AOU, apparent oxygen utilization; DIC, dissolved inorganic carbon; ALK, alkalinity.

reported for the OCMIP models [*Doney et al.*, 2004, Table 2] and a subset of the Intergovernmental Panel on Climate Change Fourth Assessment Report (IPCC AR4) models [*Schmittner et al.*, 2005a]. The OCMIP range for 3-D models without internal restoring is 0.84–2.18 K for temperature and 0.15–0.31 for salinity; for the IPCC AR4 fully coupled ocean atmosphere models it is 0.86–2.97 K for temperature and 0.20–0.38 for salinity.

[34] We use the observational errors reviewed in section 2.3 for the 3-D method. For the 1-D method we assume that the observation error is negligible compared with the model error, since the global averaging leads to very small observational errors (decreasing with $\sim N_{eff}^{-1/2}$, where N_{eff} is the effective number of observations). This is consistent with the 1-D data-model residuals, which show a smoothly varying structure more indicative of systematic model error than random observation noise.

[35] With these error estimates, we evaluate the skill of each of the eight models in the ensemble using three metrics. We use the root mean squared (RMS) error introduced in section 2.4.1, as well as the bias-corrected RMS error. We also compute the correlation between the observations and each model. A higher correlation indicates greater similarity between the model and the observations. We conduct sensitivity studies to explore how model skill varies with K_v as determined by each of the three skill measures.

[36] In the following discussion we distinguish between tracers which are influenced by physical processes only such as T, S, Δ^{14}C, and CFC11, and those tracers strongly affected by biological processes such as PO_4, AOU, P^*, DIC, and ALK, since the latter also depend on the choice of uncertain biological model parameters. Biological effects on the radiocarbon distribution in the ocean are about 2 orders

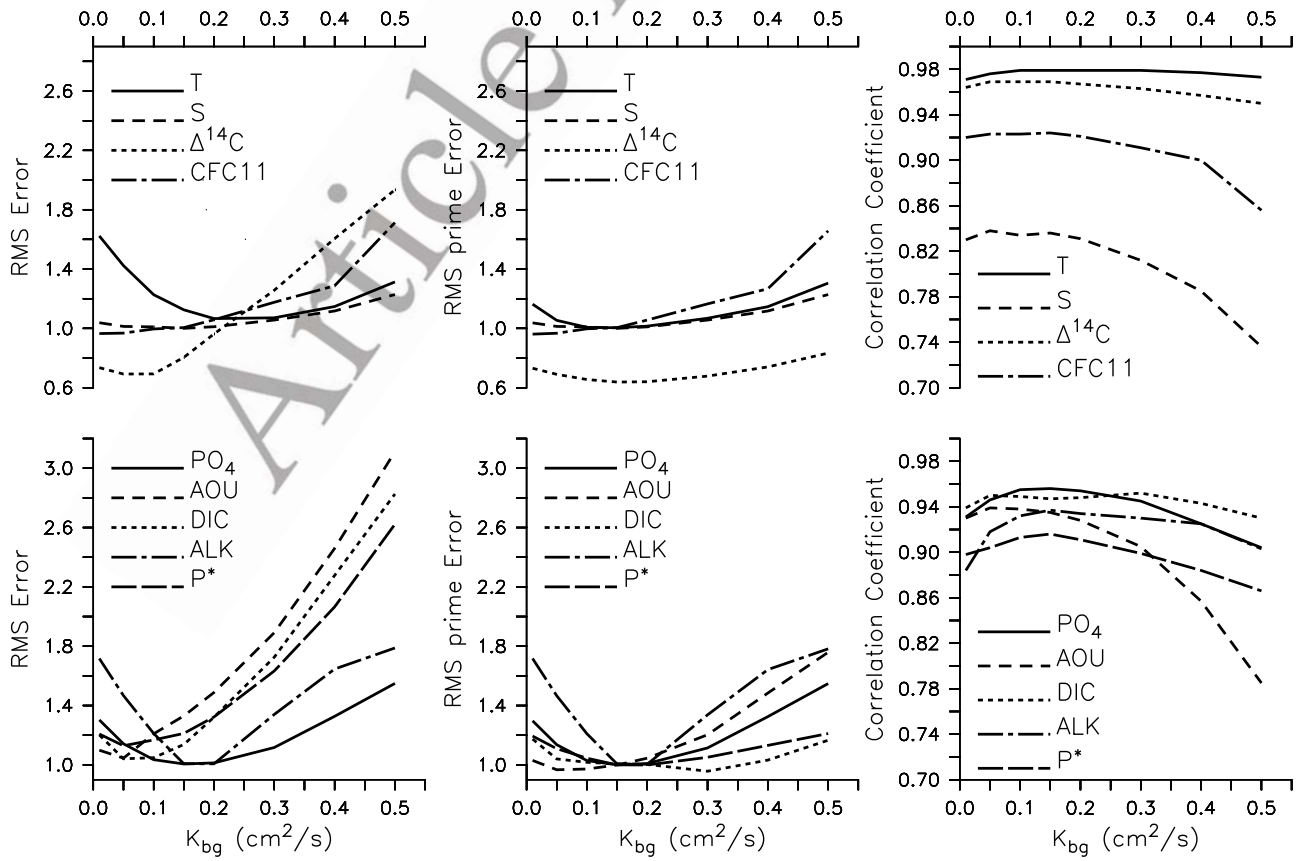

Figure 4. Normalized RMS errors (left) E and (center) E' and (right) correlation coefficients r for 3-D distributions of different (top) physical and (bottom) biogeochemical tracers as a function of the diapycnal background diffusivity K_{bg}.

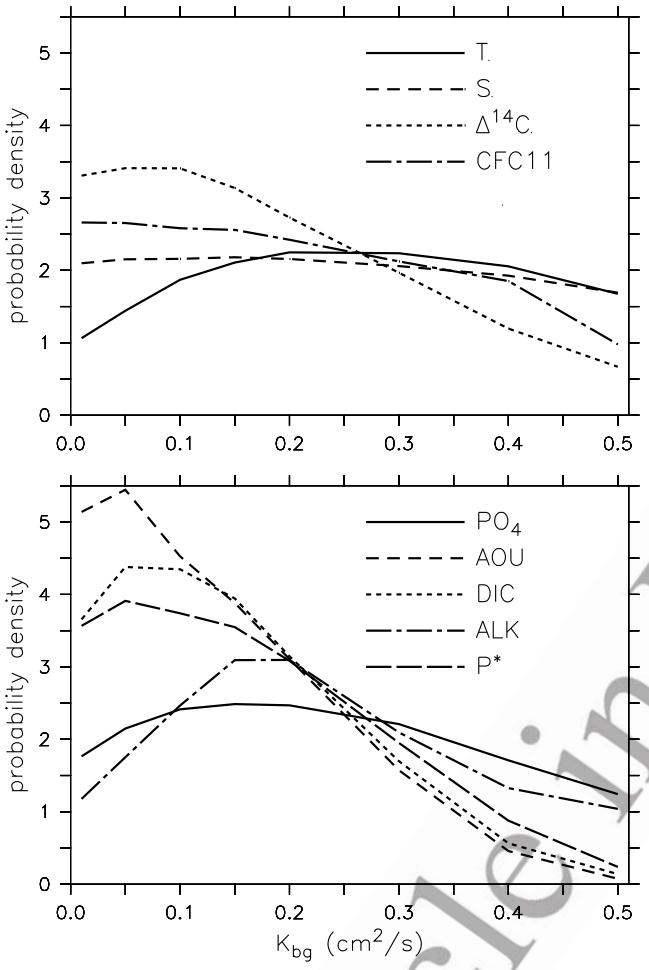

Figure 5. Posterior PDFs using the 3-D method (equation (3)) for different (top) physical and (bottom) biogeochemical tracers as a function of the diapycnal background diffusivity K_{bg}.

of magnitude smaller than the physical effects of decay and air-sea gas exchange.

[37] The RMS and bias-corrected RMS errors, E and E', are plotted in Figure 4 together with the correlation coefficients $r_i = \overline{M_i O_i}/\sqrt{\mathrm{var}(M_i)\cdot\mathrm{var}(O_i)}$, with the variance $\mathrm{var}(x) = \overline{x^2} - \overline{x}^2$, using the full three-dimensional data. The different tracers show different sensitivities to K_{bg} depending on the global metric considered. When measured by the RMS error E, the model skill for the Δ^{14}C, AOU, P^* and DIC tracers show the largest sensitivity to changes in K_v. Much of this sensitivity, however, is due to the model bias, as revealed by the difference between E and the bias-corrected error E'. For S, ALK and PO_4 the bias is zero ($E = E'$) because neither of these tracers exchanges with other climate system components in the model and hence their ocean inventories are fixed. When measured by the correlation coefficient, the model skill for the AOU and S tracers are most sensitive to variations in K_{bg}.

[38] Most tracers are in better agreement with the observations for small values of K_{bg}, both for E and r as metrics. Correlation coefficients between model output and observations peak between 0.05 and 0.15 for all tracers except DIC and T, which are rather insensitive. AOU, DIC, Δ^{14}C and P^* show very large biases for large values of K_{bg}. The deep ocean is much too young (Δ^{14}C too high), too vigorously ventilated (AOU too low), too poor in inorganic carbon and too high in preformed nutrients. Even if the bias is removed, the bias-corrected RMS error E' in AOU is still much larger for the high-K_{bg} models. CFC11 and S are both moderately sensitive and show better agreement with the observations for low K_v, irrespective of the metric considered. PO_4 and ALK are also moderately sensitive and show minima in RMS errors and maxima in correlation around $K_{bg} = 0.15$.

3.3.2. Probabilities From the 3-D Method

[39] PDFs from the 3-D method suggest that Δ^{14}C is the most sensitive of the physical variables to changes in K_v, followed by CFC11, T and S (Figure 5). Δ^{14}C, S, and CFC11 show the maximum probability for small values of K_{bg}. For Δ^{14}C the probability for small K_{bg} is about three times as high as that for high K_{bg}. T shows a broad maximum for $0.2 \le K_{bg} \le 0.4$ and smallest probabilities for very high and very low values of K_{bg}. The biological tracers (Figure 5, bottom) are all sensitive to variations in K_{bg}, in particular AOU, DIC, and P^* which are 5–10 times more likely for low than high K_{bg}. ALK and PO_4 show maxima for K_{bg} around 0.15–0.2.

3.3.3. Probabilities From the 1-D Method

[40] Figure 6 shows PDFs for the same variables but using the 1-D method. The most obvious difference is that the 1-D PDFs are much sharper than those obtained with the 3-D method. This might be counterintuitive, since information was lost by aggregating the data from 3-D to 1-D (we discuss this effect further in section 3.3.4). The 1-D method yields maxima for all tracers for $K_{bg} \le 0.2$. Probabilities for $K_{bg} > 0.4$ are very small for all tracers. Thus the two statistical methods agree that high-K_{bg} models are less consistent with the observations than low-K_{bg} models. Both methods also exhibit similar shapes for most tracers. For example, Δ^{14}C, S, AOU, DIC, and P^* all have maxima for $K_{bg} < 0.2$, CFC11, ALK and PO_4 show maxima for $0.1 \le K_{bg} \le 0.2$, and T shows a broad maximum for $0.2 \le K_{bg} \le 0.3$.

3.3.4. Sensitivity Tests

[41] We conduct four simple sensitivity analyses of the 3-D method to gain some insights into the factors that influence the differences in the posterior PDFs between the 3-D and 1-D methods (Figure 7). First, we test the assumption of neglecting the error of the observations by setting $\sigma_O = 0$ and reestimating the total error σ (Table 1). Comparing the resulting PDFs (blue lines) with the original PDFs (black lines) shows that this effect is negligible for most tracers. Only CFC11, Δ^{14}C, and DIC show small differences.

[42] Second, we calculated the PDFs for 1-D (horizontally averaged) data but using the 3-D method as described in section 2.4.1 (red solid lines in Figure 7). The reestimated errors (Table 1) are much smaller than in the 3-D case for all tracers, indicating that the model has considerably more skill in reproducing the horizontally averaged observations

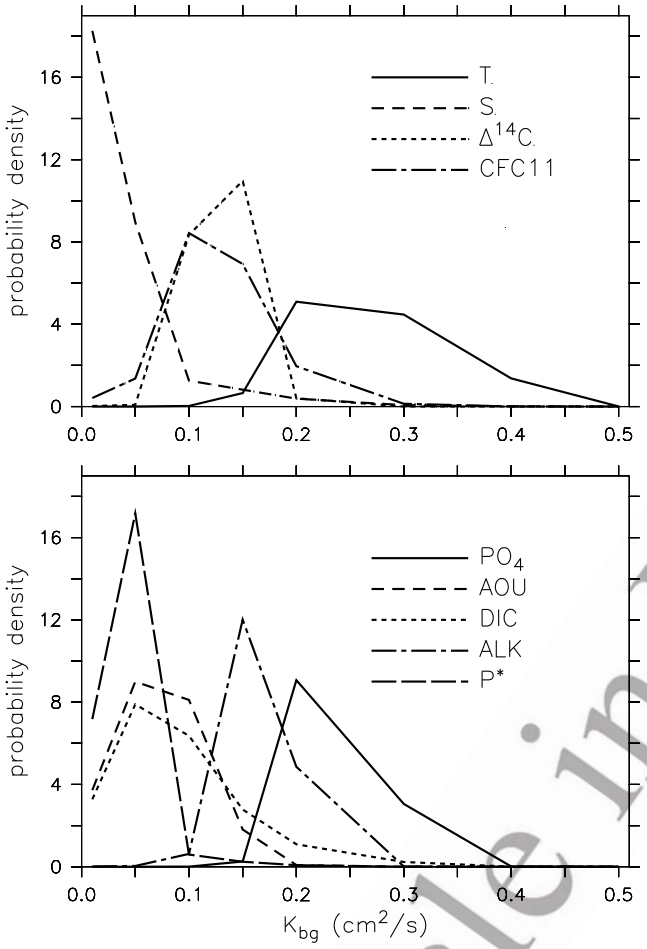

Figure 6. Posterior PDFs using the 1-D method (equation (7)) for different (top) physical and (bottom) biogeochemical tracers as a function of the diapycnal background diffusivity K_{bg}.

than the full 3-D distributions. Intuitively it makes sense that because of the limited resolution a model's skill improves with increasing spatial scale. Comparison of the solid black and solid red lines in Figure 7 shows that smaller σ results in sharper PDFs, which for most tracers are now similar to the PDFs from the 1-D method (Figure 6). This suggests that the main reason for the sharper PDFs in the 1-D method (Figure 6) compared with the 3-D method (Figure 5) is the smaller estimated σ.

[43] Third, we evaluate the effects of correcting for spatial autocorrelation. Following *Ricciuto et al.* [2008] we remove the lag-1 autocorrelation (a) from the 1-D residuals, R, according to:

$$\overline{E_i^2} = \overline{(R_i(z_k) - aR_i(z_{k-1}))^2}. \quad (9)$$

As expected from earlier studies [*Ricciuto et al.*, 2008; *Zellner and Tiao*, 1964] this approach to account for the autocorrelation (green lines in Figure 7) leads to broader PDFs (compared to the red solid lines). Neglecting spatial autocorrelation typically results in overconfident parameter estimates. The fact that the PDFs are quite different emphasizes the importance of properly considering spatial autocorrelation.

[44] Fourth, the PDFs are recalculated for the 1-D data (without subtracting autocorrelation) but using the error estimate from the 3-D method with $\sigma_O = 0$ (red dashed lines in Figure 7). Thus the difference between the red solid lines and the red dashed lines in Figure 7 isolates the effect of different estimated σ. The difference between the red dashed lines and the blue lines isolates the effect of the reduced information content in the 1-D versus the 3-D residuals. For most tracers the PDFs are broader than those in the high-σ cases (red solid lines) and more similar to the 3-D case (blue lines). This indicates that the most important reason for the difference between the 1-D and 3-D methods (and the explanation for the sharper PDFs in the 1-D method) is the differently estimated σ. It also suggests that spatial aggregation, despite a loss of information, can help to improve the model skill, and as a consequence lead to sharper PDFs. For PO_4, CFC11, and ALK the red dashed lines deviate substantially from the blue solid lines. This indicates an important loss of information due to the averaging. These tracers might not be suitable for the 1-D method.

3.3.5. Probabilities for Multiple Tracers

[45] Each of the tracers examined above contains different information and leads to a different PDF for K_{bg}. Our goal, however, is to produce a single PDF combining the information from all tracers as outlined in section 2.4.1.

[46] The distribution of each tracer is influenced not only by diapycnal mixing and the large-scale ocean circulation, but also by other processes. Some tracers, such as T, S, CFC11, $\Delta^{14}C$, and DIC, are also influenced by air-sea exchange. Thus, the model errors, and hence the PDF, for T, e.g., might be influenced by model biases in ocean-atmosphere heat fluxes, which are controlled by radiative

Figure 7. Sensitivity tests. Posterior PDFs as a function of the diapycnal background diffusivity K_{bg} for different tracers using the 3-D method, but different assumptions in the statistical analysis. Black lines show the full 3-D method and are identical to the PDFs shown in Figure 5. Blue lines neglect the error in the observations ($\sigma_O = 0$). Note that for many tracers the black lines are indistinguishable from and covered by the blue lines. Solid red lines use the 3-D method as described in section 2.4.1 but use horizontally averaged 1-D data. Note that the error estimate (σ) is strongly reduced (see also Table 1) and that the PDFs are much sharper compared to the unaveraged 3-D data (black lines). Red dashed lines also use the 3-D method and horizontally averaged 1-D data, but instead of estimating the error (as done for the solid red lines) the error estimate from the 3-D data (black lines) is used. Thus the difference between the red solid and red dashed lines is only due to different σ. Green lines compared to the red solid lines illustrate the broadening effect on the PDFs from removing the spatial autocorrelation using equation (9).

fluxes as well as sensible and latent heat fluxes. The PDF for S, on the other hand, is influenced by surface ocean water fluxes, which are determined by evaporation, precipitation and river runoff, and thus by the atmospheric hydrological cycle. Because different physical processes control heat and water fluxes (except for evaporation which influences both) it is unlikely that model errors in heat fluxes are strongly correlated with errors in water fluxes. Similarly, the air-sea fluxes of carbon, radiocarbon and CFCs are presumably rather independent from heat and

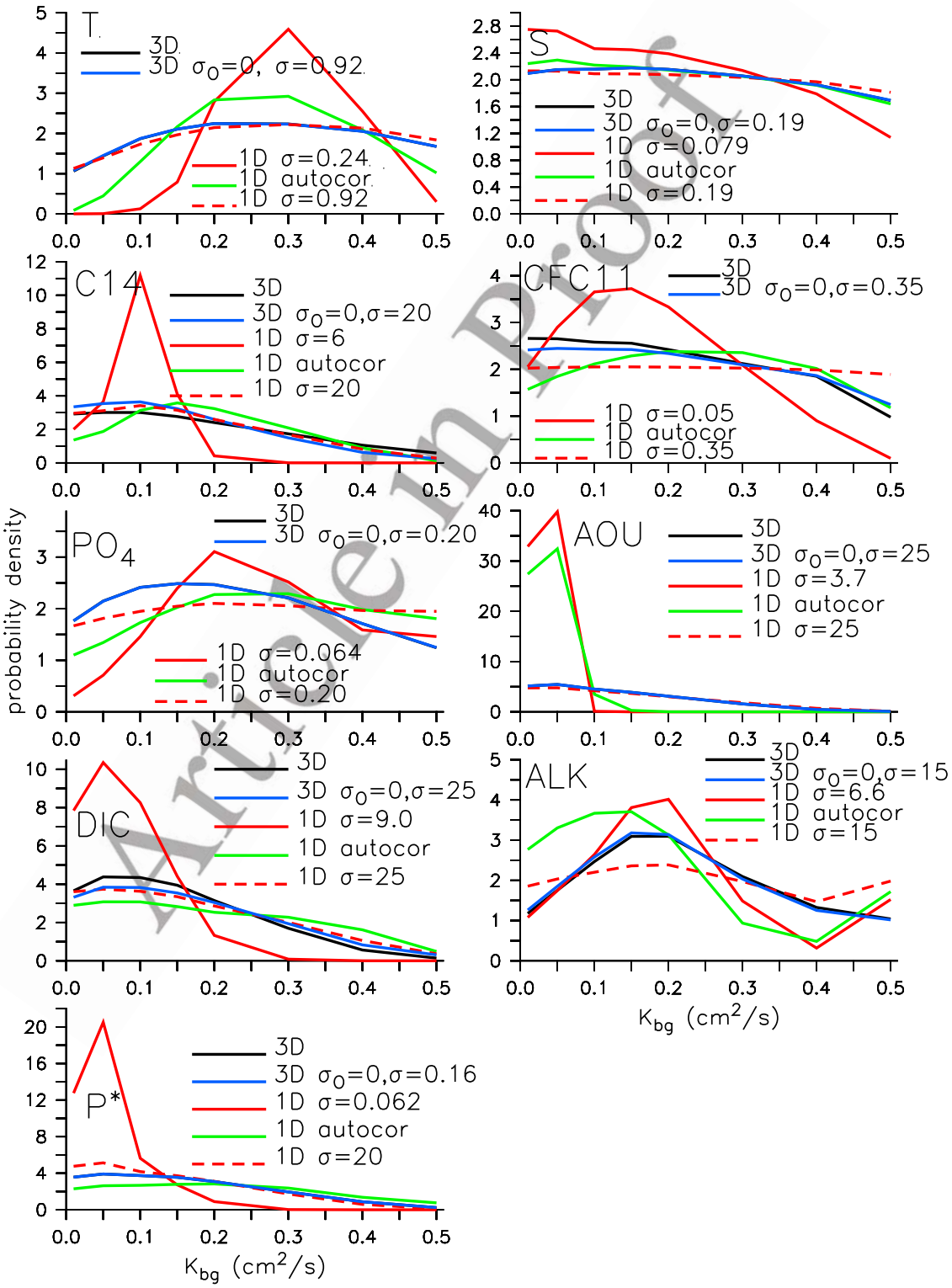

Figure 7

t2.1 **Table 2.** Cross-Tracer Error Correlation for the 3-D Method in the Model With K_{bg} = 0.15 cm^2/s[a]

t2.2		T	S	Δ^{14}C	CFC11	PO_4	AOU	DIC	ALK
t2.3	T								
t2.4	S	**0.48**							
t2.5	Δ^{14}C	0.26	0.03						
t2.6	CFC11	0.01	−0.04	**0.39**					
t2.7	PO_4	**−0.42**	−0.22	**−0.41**	−0.23				
t2.8	AOU	−0.09	0.12	**−0.36**	**−0.40**	**0.65**			
t2.9	DIC	−0.20	**0.31**	**0.33**	−0.13	**0.52**	**0.76**		
t2.10	ALK	0.04	**0.54**	0.23	−0.07	0.12	**0.42**	**0.72**	
t2.11	P^*	**−0.44**	**−0.42**	−0.19	0.08	**0.69**	−0.08	−0.05	−0.23

t2.12 [a]Absolute values larger than 0.3 are shown in bold.

water fluxes. Thus, considering multiple tracers can possibly average out model errors in individual air-sea fluxes. If the errors in the tracer residuals are independent between tracers, a combined likelihood for all tracers can be calculated by multiplying the likelihoods of the individual tracers as described at the end of section 2.4.1.

[47] On the other hand, if tracers are not independent, multiplication of the likelihoods would lead to overconfident and possibly biased PDFs. Sinking of particulate organic matter (the soft-tissue biological pump), for instance, influences PO_4, AOU, and DIC and thus errors in those tracers cannot be expected to be independent. An objective way to determine independence between different tracers is to examine correlations between the errors of the residuals. As shown in Table 2, the different tracers are generally not independent. PO_4, AOU, and DIC are clearly related for the reasons mentioned above, but other tracer residuals (such as T and S) are also correlated, for less obvious reasons.

[48] At this point no method that we are aware of has accounted for the cross-tracer correlation. It is highly desirable to develop such a method in the future. For the time being we calculate PDFs for different combinations of uncorrelated tracers (Figure 8). All combined PDFs show low probability for models with high mixing rates (K_{bg} > 0.3). The different tracer combinations do not agree well for the probability of low-mixing models. Some show a distinct maximum around 0.1–0.2 and considerably lower probabilities for lower K_{bg}, whereas others show high probabilities for the lowest diffusivities. We conclude that the observations put a firm upper limit on the diffusivities, whereas no unequivocal lower limit can be determined on the basis of the information we have presented here.

4. Discussion

[49] One issue that has not been addressed here is parameter interactions. Generally model tracer distributions are influenced by more than one parameter, each of which is uncertain. Thus, the results obtained by varying one parameter depend on the values of many other parameters. This is also true in our case, and hence the probabilities for different K_{bg} presented here are tentative and should be regarded as a test of the methodology rather than a definitive result.

[50] Parameter interactions might be most obvious for tracers affected by biological processes such as PO_4, AOU, P^*, DIC and ALK, which are sensitive to ill-constrained biological model parameters. Surface nutrient concentrations and deep ocean AOU, P^* and DIC, for instance, all depend strongly on the maximum growth rate of phytoplankton (γ) which determines the efficiency of the biological pump. The vertical alkalinity gradient is controlled by the fixed ratio of calcium carbonate versus particulate organic carbon production ($R_{CaCO3/POC}$). These biological model parameters were tuned for a model version with K_{bg} = 0.15 (γ = 0.13 d^{-1}, $R_{CaCO3/POC}$ = 0.03). Thus larger errors for those tracers in models with different K_{bg} can be expected because the biological parameters are unadjusted. Interestingly, though, 3 out of 5 biological tracers prefer K_{bg} = 0.05. Models with K_{bg} > 0.15 therefore overestimate surface nutrient concentrations because of more intense advective and diffusive transport of nutrient rich deep waters to the surface. Similarly, models with K_{bg} > 0.15

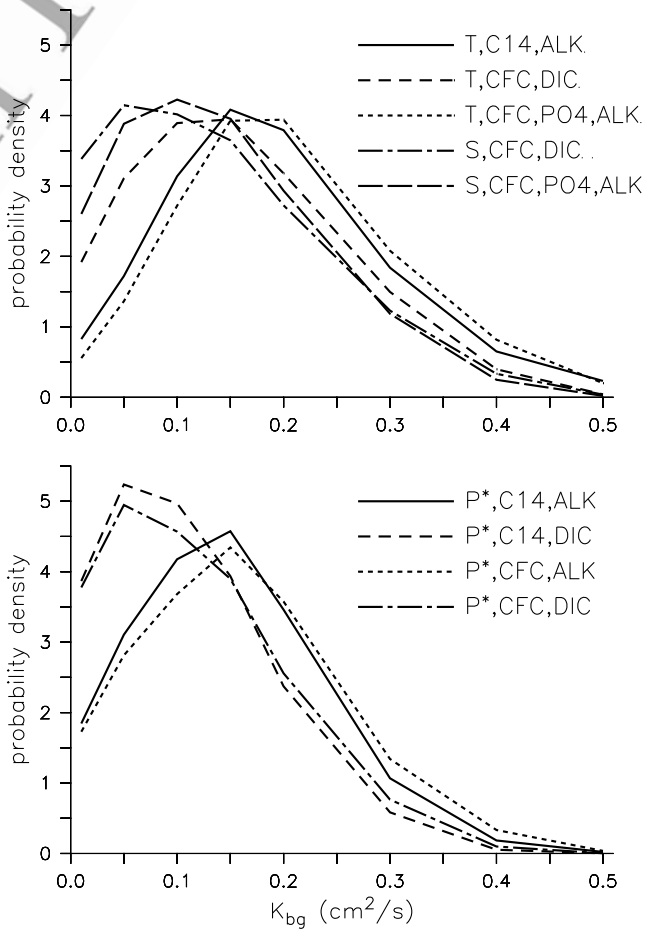

Figure 8. Posterior PDFs as a function of the diapycnal background diffusivity K_{bg} for different combinations of uncorrelated (see Table 2) tracer distributions using the 3-D method. Compared with the PDFs of the individual tracers as shown in Figure 5, the combined PDFs are much sharper, illustrating the power of using multiple tracers to constrain model parameters.

underestimate the efficiency of the biological pump and hence the deep ocean AOU and DIC, and overestimate P^*. Thus γ should be increased together with K_{bg}. Faster rates of nutrient input into the euphotic zone, in the strong mixing models, also lead to increased primary and export production [*Schmittner et al.*, 2005b] and higher production of $CaCO_3$, resulting in overestimated vertical alkalinity gradients. Thus, $R_{CaCO3/POC}$ should be decreased as K_{bg} is increased.

[51] Because of computational constraints we are currently not able to retune the biological parameters for each model version with different K_{bg}. A simple optimization of biological parameters for the model version with K_{bg} = 0.5 (γ = 0.2 d^{-1}, $R_{CaCO3/POC}$ = 0.02) results in a decrease of the errors with respect to the untuned values shown in Figure 3, but the errors are still significantly larger than those of the low-K_{bg} models. Thus, the true likelihoods for the biological tracers would presumably increase for model versions with high K_{bg}. It is highly desirable to include these known cross-parameter dependencies in a larger model ensemble in the future. Of course, tracer distributions not affected by biological parameters, such as $\Delta^{14}C$ (radiocarbon in our model is not influenced by biological parameters) and CFCs, do not suffer from this complication. Therefore our conclusion that models with K_{bg} > 0.3 cm^2/s are increasingly inconsistent with observations holds true on the basis of these tracers alone.

[52] An intriguing result is that horizontally averaged data (1-D method) lead to sharper PDFs than the full 3-D data distribution. We have shown that this is likely due to the improved skill of the model in simulating horizontally averaged observations (smaller σ). This seems to be an advantage of the 1-D method. However, horizontal averaging has the obvious disadvantage that major model problems in the horizontal tracer distribution are undetectable. Consider, for example, a model with deep water formation in the North Pacific instead of the North Atlantic. Such a model might still reproduce the horizontally averaged tracer distributions reasonably well, despite the fact that it is obviously wrong. Nevertheless, our results suggest that an optimal degree of spatial aggregation might exist, at which high model skill and the resulting sharp PDFs could be combined with 3-D spatial information.

[53] *Griffies et al.* [2000] showed that z-level models, such as the one used here, can exhibit spurious diapycnal mixing due to numerical errors. For a model with the same numerical scheme as that used here (the second-order accurate, flux-corrected transport scheme, FCT) and a resolution of 2.4° × 2.4° they found large spurious mixing on the order of 0.3 cm^2/s, whereas for a model with 1.2° × 1.2° the spurious mixing was negligible because of the improved resolution of the western boundary currents. Our zonal grid resolution (which is more important than the meridional resolution for simulating western boundary currents) of 1.8° is in between those reported by *Griffies et al.* [2000]. Thus, we cannot exclude the possibility that our model exhibits spurious mixing, particularly for the low-K_{bg} cases. However, we can exclude the possibility that the model is dominated by numerical diffusion, because in this case changing the explicit diffusivity would not alter the solution. By contrast, in our experiments, the circulation is significantly different between all runs, including those with low diffusivity. The maximum overturning at 25°N in the Atlantic in the unperturbed preindustrial model spinup, for example, is 10.8 Sv, for K_{bg} = 0.01, 12.2 Sv in the K_{bg} = 0.05 case, and 13.8 Sv for K_{bg} = 0.1.

[54] An outstanding question remains as to how to interpret the range spanned by the C^4MIP model results. This question can be addressed only by a systematic and probabilistic comparison with observations that sample the relevant parametric and structural uncertainties. Our study represents a step toward this goal, though here we have sampled only a small fraction of the full range of parametric uncertainty. We have shown that low values of K_v are most consistent with ocean tracer observations, and that most of the C^4MIP models fall within the range of ocean carbon uptake simulated by varying K_{bg} values in this study. If the values for K_v were known for the different C^4MIP models, it would be possible to reject projections from models with high K_v values, or to judge them as less reliable than those from models with low K_v. However, we are not aware of a published documentation of the values of K_v used by the C^4MIP models (effective diapycnal diffusivity can also contain a difficult-to-evaluate numerical component). There is an additional complication arising from different structural types of ocean models represented in C^4MIP (box models, versus 2-D models, versus GCMs). In practice, therefore, it remains difficult to assign the likelihoods we have derived here directly to the C^4MIP model projections. However, we think that the methodology developed here can be used for multimodel assessments in the future, given that spatially resolved tracer model data output is provided.

5. Conclusions

[55] We have shown that uncertainties in the value of diapycnal mixing in the pelagic ocean contribute to the spread in future model projections of CO_2 and climate in response to anthropogenic carbon emissions. Models with low mixing lead to slower uptake of carbon and heat by the ocean, therefore contributing to higher atmospheric CO_2 and warmer air temperatures. These results suggest that models with large ocean vertical mixing (high K_v) systematically underestimate future warming and CO_2 concentrations, and that the range in vertical mixing between models is a contributing factor to the large ranges in transient climate sensitivity and climate–carbon cycle feedbacks that have been diagnosed in earlier model intercomparisons.

[56] Globally averaged metrics such as historic changes in globally averaged surface air temperature or ocean heat content do not provide strong constraints on the vertical diffusivity [*Tomassini et al.*, 2007]. We show that spatially resolved physical, geochemical and biogeochemical tracer observations in the ocean can be used to reduce the uncertainty of this parameter (and, by extension, that of future climate projections). These observations provide a firm upper limit on the value of K_{bg}, whereas the lower limit is less well constrained. Our best estimate for the background diapycnal diffusivity in the pelagic ocean is 0.05–0.2 cm^2/s, in agreement with independent estimates based

on dye dispersion experiments and microstructure turbulence measurements [*Ledwell et al.*, 1993; *Toole et al.*, 1994].

[57] We have developed a Bayesian model-data fusion method that can be used to quantify and reduce the uncertainty in future climate–carbon cycle projections. Remaining issues left for future work include (1) cross-tracer correlations, (2) parameter interactions, and (3) the optimal degree of spatial aggregation. Resolution of the second issue is simply one of computational resources, while the first needs further development and refinement of the existing statistical methodology and theory. To resolve the third issue, the optimal degree of aggregation can presumably be determined in a sensitivity study with successively larger spatial scales of averaging. None of those issues seem insurmountable. The prospect of robust likelihood-based model assessment, using multiple observations considering spatial and temporal autocorrelation as well as cross-tracer correlations has the potential to lead toward truly probabilistic climate–carbon cycle projections.

[58] **Acknowledgments.** We thank Roman Tonkonojenkov, Bob Key, Katrin Meissner, Murali Haran, Sham Bhat, Marlos Goes, Josh Dorin, and Brian Tuttle for helpful discussions. Any potential errors and omissions are, of course, ours. We are grateful for the support of Michael Eby and Andrew Weaver from the University of Victoria model development team. Financial support from the National Science Foundation, the Canadian Foundation for Climate and Atmospheric Sciences, and the Penn State Center for Climate Risk Management is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding entity.

References

Andronova, N. G., and M. E. Schlesinger (2001), Objective estimation of the probability density function for climate sensitivity, *J. Geophys. Res.*, *106*(D19), 22,605–22,611, doi:10.1029/2000JD000259.

Annan, J. D., J. C. Hargreaves, R. Ohgaito, A. Abe-Ouchi, and S. Emori (2005), Efficiently constraining climate sensitivity with ensembles of paleoclimate simulations, *SOLA*, *1*(0), 181–184.

Antonov, J. I., R. A. Locarnini, T. P. Boyer, A. V. Mishonov, and H. E. Garcia (2006), *World Ocean Atlas 2005*, vol. 2, *Salinity, NOAA Atlas NESDIS 62*, edited by S. Levitus, 182 pp., U.S. Govt. Print. Off., Washington, D. C.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.

Bryan, F. (1987), Parameter sensitivity of primitive equation ocean general circulation models, *J. Phys. Oceanogr.*, *17*(7), 970–985, doi:10.1175/1520-0485(1987)017<0970:PSOPEO>2.0.CO;2.

Canadell, J. G., et al. (2007), Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks, *Proc. Natl. Acad. Sci. U. S. A.*, *104*(47), 18,866–18,870, doi:10.1073/pnas.0702737104.

Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell (2000), Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, *Nature*, *408*(6809), 184–187, doi:10.1038/35041539.

Crowley, T. J. (2000), Causes of climate change over the past 1000 years, *Science*, *289*(5477), 270–277, doi:10.1126/science.289.5477.270.

Domingues, C. M., J. A. Church, N. J. White, P. J. Gleckler, S. E. Wijffels, P. M. Barker, and J. R. Dunn (2008), Improved estimates of upper-ocean warming and multi-decadal sea-level rise, *Nature*, *453*(7198), 1090–1093, doi:10.1038/nature07080.

Doney, S. C., et al. (2004), Evaluating global ocean carbon models: The importance of realistic physics, *Global Biogeochem. Cycles*, *18*, GB3017, doi:10.1029/2003GB002150.

Dufresne, J. L., P. Friedlingstein, M. Berthelot, L. Bopp, P. Ciais, L. Fairhead, H. Le Treut, and P. Monfray (2002), On the magnitude of positive feedback between future climate change and the carbon cycle, *Geophys. Res. Lett.*, *29*(10), 1405, doi:10.1029/2001GL013777.

England, M. H., and E. Maier-Reimer (2001), Using chemical tracers to assess ocean models, *Rev. Geophys.*, *39*, 29–70, doi:10.1029/1998RG000043.

Forest, C. E., P. H. Stone, A. P. Sokolov, M. R. Allen, and M. D. Webster (2002), Quantifying uncertainties in climate system properties with the use of recent climate observations, *Science*, *295*(5552), 113–117, doi:10.1126/science.1064419.

Friedlingstein, P., J. L. Dufresne, P. M. Cox, and P. Rayner (2003), How positive is the feedback between climate change and the carbon cycle?, *Tellus, Ser. B*, *55*(2), 692–700, doi:10.1034/j.1600-0889.2003.01461.x.

Friedlingstein, P., et al. (2006), Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison, *J. Clim.*, *19*(14), 3337–3353, doi:10.1175/JCLI3800.1.

Garcia, H. E., R. A. Locarnini, T. P. Boyer, and J. Antonov (2006a), *World Ocean Atlas 2005*, vol. 4, *Nutrients (Phosphate, Nitrate, Silicate), NOAA Atlas NESDIS 64*, 396 pp., U.S. Govt. Print. Off., Washington, D. C.

Garcia, H. E., R. A. Locarnini, T. P. Boyer, and J. Antonov (2006b), *World Ocean Atlas 2005, Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation*, vol. 4, *NOAA Atlas NESDIS 63*, edited by S. Levitus, 342 pp., U.S. Govt. Print. Off., Washington, D. C.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, CRC, Boca Baton, Fla.

Gent, P. R., and J. C. McWilliams (1990), Isopycnal mixing in ocean circulation models, *J. Phys. Oceanogr.*, *20*, 150–155, doi:10.1175/1520-0485(1990)020<0150:IMIOCM>2.0.CO;2.

Gnanadesikan, A., J. P. Dunne, R. M. Key, K. Matsumoto, J. L. Sarmiento, R. D. Slater, and P. S. Swathi (2004), Oceanic ventilation and biogeochemical cycling: Understanding the physical mechanisms that produce realistic distributions of tracers and productivity, *Global Biogeochem. Cycles*, *18*, GB4010, doi:10.1029/2003GB002097.

Govindasamy, B., S. Thompson, A. Mirin, M. Wickett, K. Caldeira, and C. Delire (2005), Increase of carbon cycle feedback with climate sensitivity: Results from a coupled climate and carbon cycle model, *Tellus, Ser. B*, *57*(2), 153–163, doi:10.1111/j.1600-0889.2005.00135.x.

Griffies, S. M., R. C. Pacanowski, and R. W. Hallberg (2000), Spurious diapycnal mixing associated with advection in a z-coordinate ocean model, *Mon. Weather Rev.*, *128*(3), 538–564, doi:10.1175/1520-0493(2000)128<0538:SDMAWA>2.0.CO;2.

Jones, C. D., P. M. Cox, R. L. H. Essery, D. L. Roberts, and M. J. Woodage (2003), Strong carbon cycle feedbacks in a climate model with interactive CO_2 and sulphate aerosols, *Geophys. Res. Lett.*, *30*(9), 1479, doi:10.1029/2003GL016867.

Jones, C. D., P. M. Cox, and C. Huntingford (2006), Climate-carbon cycle feedbacks under stabilization: Uncertainty and observational constraints, *Tellus, Ser. B*, *58*(5), 603–613, doi:10.1111/j.1600-0889.2006.00215.x.

Joos, F., G. K. Plattner, T. F. Stocker, O. Marchal, and A. Schmittner (1999), Global warming and marine carbon cycle feedbacks on future atmospheric CO_2, *Science*, *284*(5413), 464–467, doi:10.1126/science.284.5413.464.

Joos, F., I. C. Prentice, S. Sitch, R. Meyer, G. Hooss, G. K. Plattner, S. Gerber, and K. Hasselmann (2001), Global warming feedbacks on terrestrial carbon uptake under the Intergovernmental Panel on Climate Change (IPCC) emission scenarios, *Global Biogeochem. Cycles*, *15*(4), 891–907, doi:10.1029/2000GB001375.

Keeling, C. D., and T. P. Whorf (2005), Atmospheric CO_2 records from sites in the SIO air sampling network, Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., U.S. Dept. of Energy, Oak Ridge, Tenn.

Key, R. M., et al. (2004), A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP), *Global Biogeochem. Cycles*, *18*, GB4031, doi:10.1029/2004GB002247.

Knutti, R., T. F. Stocker, F. Joos, and G. K. Plattner (2003), Probabilistic climate change projections using neural networks, *Clim. Dyn.*, *21*(3–4), 257–272, doi:10.1007/s00382-003-0345-1.

Le Quere, C., et al. (2007), Saturation of the Southern Ocean CO_2 sink due to recent climate change, *Science*, *316*(5832), 1735–1738, doi:10.1126/science.1136188.

Ledwell, J. R., A. J. Watson, and C. S. Law (1993), Evidence for slow mixing across the pycnocline from an open-ocean tracer-release experiment, *Nature*, *364*(6439), 701–703, doi:10.1038/364701a0.

Lehmann, E. L., and G. Casella (2003), *Theory of Point Estimation*, 2nd ed., Springer, New York.

Locarnini, R. A., A. V. Mishonov, J. Antonov, T. Boyer, and H. E. Garcia (2006), *World Ocean Atlas 2005*, vol. 1, *Temperature, NOAA Atlas NESDIS 61*, edited by S. Levitus, 182 pp., U.S. Govt. Print. Off., Washington, D. C.

Mahalanobis, P. C. (1936), On the generalized distance in statistics, *Proc. Natl. Inst. Sci. India*, *2*, 49–55.

Matear, R. J., and A. C. Hirst (1999), Climate change feedback on the future oceanic CO_2 uptake, *Tellus, Ser. B*, *51*(3), 722–733, doi:10.1034/j.1600-0889.1999.t01-1-00012.x.

Matsumoto, K., et al. (2004), Evaluation of ocean carbon cycle models with data-based metrics, *Geophys. Res. Lett.*, *31*, L07303, doi:10.1029/2003GL018970.

Matthews, H. D., M. Eby, A. J. Weaver, and B. J. Hawkins (2005a), Primary productivity control of simulated carbon cycle-climate feedbacks, *Geophys. Res. Lett.*, *32*, L14708, doi:10.1029/2005GL022941.

Matthews, H. D., A. J. Weaver, and K. J. Meissner (2005b), Terrestrial carbon cycle dynamics under recent and future climate change, *J. Clim.*, *18*(10), 1609–1628, doi:10.1175/JCLI3359.1.

Meissner, K. J., A. J. Weaver, H. D. Matthews, and P. M. Cox (2003), The role of land surface dynamics in glacial inception: A study with the UVic Earth System Model, *Clim. Dyn.*, *21*(7–8), 515–537, doi:10.1007/s00382-003-0352-2.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, and M. Collins (2004), Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, *430*(7001), 768–772, doi:10.1038/nature02771.

Naveira Garabato, A. C., K. L. Polzin, B. A. King, K. J. Heywood, and M. Visbeck (2004), Widespread intense turbulent mixing in the Southern Ocean, *Science*, *303*, 210–213, doi:10.1126/science.1090929.

Neftel, A., H. Friedli, E. Moor, H. Lötscher, H. Oeschger, U. Siegenthaler, and B. Stauffer (1994), Historical CO_2 record from the Siple Station ice core, Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., U.S. Dept. of Energy, Oak Ridge, Tenn.

Ricciuto, D. M., K. J. Davis, and K. Keller (2008), A Bayesian calibration of a simple carbon cycle model: The role of observations in estimating and reducing uncertainty, *Global Biogeochem. Cycles*, *22*, GB2030, doi:10.1029/2006GB002908.

Sarmiento, J. L., T. M. C. Hughes, R. J. Stouffer, and S. Manabe (1998), Simulated response of the ocean carbon cycle to anthropogenic climate warming, *Nature*, *393*(6682), 245–249, doi:10.1038/30455.

Schartau, M., and A. Oschlies (2003), Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part I. Method and parameter estimates, *J. Mar. Res.*, *61*(6), 765–793, doi:10.1357/002224003322981147.

Schmittner, A., M. Latif, and B. Schneider (2005a), Model projections of the North Atlantic thermohaline circulation for the 21st century assessed by observations, *Geophys. Res. Lett.*, *32*, L23710, doi:10.1029/2005GL024368.

Schmittner, A., A. Oschlies, X. Giraud, M. Eby, and H. L. Simmons (2005b), A global model of the marine ecosystem for long-term simulations: Sensitivity to ocean mixing, buoyancy forcing, particle sinking, and dissolved organic matter cycling, *Global Biogeochem. Cycles*, *19*, GB3004, doi:10.1029/2004GB002283.

Schmittner, A., A. Oschlies, H. D. Matthews, and E. D. Galbraith (2008), Future changes in climate, ocean circulation, ecosystems and biogeochemical cycling simulated for a business-as-usual CO_2 emission scenario until year 4000 AD, *Global Biogeochem. Cycles*, *22*, GB1013, doi:10.1029/2007GB002953.

Schneider von Deimling, T., H. Held, A. Ganopolski, and S. Rahmstorf (2006), Climate sensitivity estimated from ensemble simulations of glacial climate, *Clim. Dyn.*, *27*(2–3), 149–163, doi:10.1007/s00382-006-0126-8.

Simmons, H. L., S. R. Jayne, L. C. St Laurent, and A. J. Weaver (2004), Tidally driven mixing in a numerical model of the ocean general circulation, *Ocean Modell.*, *6*(3–4), 245–263, doi:10.1016/S1463-5003(03)00011-8.

Stainforth, D. A., et al. (2005), Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature*, *433*(7024), 403–406, doi:10.1038/nature03301.

Storn, R., and K. Price (1997), Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces, *J. Global Optim.*, *11*, 341–359, doi:10.1023/A:1008202821328.

Tomassini, L., P. Reichert, R. Knutti, T. F. Stocker, and M. E. Borsuk (2007), Robust Bayesian uncertainty analysis of climate system properties using Markov chain Monte Carlo methods, *J. Clim.*, *20*(7), 1239–1254, doi:10.1175/JCLI4064.1.

Toole, J. M., K. L. Polzin, and R. W. Schmitt (1994), Estimates of diapycnal mixing in the abyssal ocean, *Science*, *264*(5162), 1120–1123, doi:10.1126/science.264.5162.1120.

Weaver, A. J., et al. (2001), The UVic Earth System Climate Model: Model description, climatology, and applications to past, present and future climates, *Atmos. Ocean*, *39*(4), 361–428.

Zellner, A., and G. C. Tiao (1964), Bayesian analysis of the regression model with autocorrelated errors, *J. Am. Stat. Assoc.*, *59*(307), 763–778, doi:10.2307/2283097.

Zeng, N., H. F. Qian, E. Munoz, and R. Iacono (2004), How strong is carbon cycle-climate feedback under global warming?, *Geophys. Res. Lett.*, *31*, L20203, doi:10.1029/2004GL020904.

K. Keller and N. M. Urban, Department of Geosciences, Pennsylvania State University, 436 Deike Building, University Park, PA 16802, USA.

D. Matthews, Department of Geography, Planning and Environment, Concordia University, Hall Building, Room 1255-21, 1455 de Maisonneuve Boulevard West, Montreal, QC H3G 1M8, Canada.

A. Schmittner, College of Oceanic and Atmospheric Sciences, Oregon State University, 104 COAS Admin Building, Corvallis, OR 97331-5503, USA. (aschmitt@coas.oregonstate.edu)