1 **Using tracer observations to reduce the uncertainty of ocean diapycnal mixing and climate**
2 **- carbon cycle projections**
3
4
5 Andreas Schmittner*, College of Oceanic and Atmospheric Sciences, Oregon State University,
6 Corvallis, OR, U.S.A.
7
8 Nathan M. Urban, Department of Geosciences, The Pennsylvania State University, University
9 Park, PA, U.S.A.
10
11 Klaus Keller, Department of Geosciences, The Pennsylvania State University, University Park,
12 PA, U.S.A.
13
14 Damon Matthews, Department of Geography, Planning and Environment, Concordia University,
15 Montreal, Quebec, Canada
16
17 * Corresponding author. 104 COAS Admin Bldg, Corvallis, Oregon 97331-5503, phone 541 737
18 9952, fax 541 737 2064, aschmitt@coas.oregonstate.edu
19
20 Index terms: 3245, 1615, 3275, 0428
21
24

Abstract

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-gridscale diapycnal (vertical) mixing in physical circulation models. We show that model versions with weak mixing lead to higher projections of atmospheric CO_2 and larger global warming than models with vigorous mixing. Slower heat uptake as well as 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 K_v value given the observational constraints. One-dimensional (horizontally averaged) data result in sharper probability density functions compared with the full 3D fields. This unexpected result opens 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.

1. Introduction

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; *Joos et al.*, 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[4]MIP) [*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.

The reasons for the aforementioned model differences are poorly understood. Although the C[4]MIP models showed larger differences in land uptake (-6 to +10 GtC/yr), there were also considerable differences in ocean uptake (+4 to +10 GtC/yr) by the year 2100 [*Friedlingstein et al.*, 2006]. *Matthews et al.* [2005a] shows that differences in the parameterizations of the dependency of terrestrial vegetation growth rates on ambient temperatures have a large effect on 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/yr in the UK Hadley Center model versus 8 GtC/yr in the French IPSL model at

78 700 ppmv atmospheric CO_2) due to increasing CO_2 alone (without climate change) to differences

79 in Southern Ocean circulation [*Friedlingstein et al.*, 2003].

80

81 The models included in the C^4MIP exercise are very heterogeneous and range from box models

82 via zonally averaged and slab mixed layer ocean models to fully three-dimensional coupled

83 ocean-atmosphere general circulation models (GCMs). However, these studies are mostly silent

84 on the question of how probable the different model structures are given the available

85 observational constraints. Since no systematic comparison with observations has been performed

86 for the C^4MIP models, it remains unclear whether or not the uncertainty range of the C^4MIP

87 models is realistic. It is very well possible that the model spread overestimates the true

88 uncertainty if some models are inconsistent with observations, as is the case for the Ocean

89 Carbon-cycle Model Intercomparison Project (OCMIP) models [*Doney et al.*, 2004; *Matsumoto*

90 *et al.*, 2004].

91

92 It is also possible that the C^4MIP models underestimate the true uncertainty, e.g. if the models do

93 not represent a large enough range of unconstrained parameters. Recent Monte Carlo simulations

94 with an atmosphere model suggest that model parameter uncertainties can increase the range of

95 future climate projections considerably [*Murphy et al.*, 2004; *Stainforth et al.*, 2005]. A key

96 uncertain parameter in ocean circulation models is the diapycnal (vertical) diffusivity K_v. The

97 strong sensitivity of the global deep overturning circulation to K_v has been known since the

98 pioneering study by *Bryan* [1987]. Here we investigate the uncertainty in ocean vertical mixing

99 and its effect on future projections of climate and CO_2.

100

101 Earlier studies show that tracer distributions in ocean models are sensitive to changes in ocean

102 circulation and ventilation [*Doney et al.*, 2004; *England and Maier-Reimer*, 2001; *Gnanadesikan*

103 *et al.*, 2004; *Matsumoto et al.*, 2004], but no attempt has been undertaken to quantify the

104 probability of different model structures and parameters given spatially resolved observations of

105 ocean tracer distributions. *Tomassini et al.* [2007] estimates probability density functions for

106 multiple parameters, including K_v, of a simple climate model using global mean surface air

107 temperature and global ocean heat content changes as observational constraints. That study finds

108 the probability distribution for K_v is multimodal, and concludes that the globally averaged

109   metrics do not provide strong limits on the value of K_v. Here we show that multiple physical,

110   geochemical and biogeochemical observations with spatial resolution can provide much stronger

111   constraints on the diapycnal ocean diffusivity. The main goal of this paper, however, is to

112   develop and demonstrate a Bayesian data-model fusion approach for spatially distributed tracer

113   observations that can be used to assess and reduce the uncertainty of future climate projections.

114

2. Methods

2.1. Model

117   The UVic Earth System Climate Model [*Weaver et al.*, 2001] of intermediate complexity,

118   includes a coarse resolution (1.8x3.6º, 19 vertical layers) three-dimensional general circulation

119   model of the ocean. It has state-of-the-art physical parameterizations such as diffusive mixing

120   along and across isopycnals, eddy induced tracer advection [*Gent and McWilliams*, 1990] and a

121   scheme for the computation of tidally induced diapycnal mixing over rough topography

122   [*Simmons et al.*, 2004]. In order to account for other sources of mixing, a globally constant

123   background diffusivity K_{bg} is added to the tidally induced diffusivity $K_v = K_{tidal} + K_{bg}$. It is this

124   background diffusivity K_{bg} that will be varied in our sensitivity study, from 0.01 cm^2/s to 0.5

125   cm^2/s. The tidally induced diffusivity rapidly decays in the water column above the sea floor

126   with an exponential depth scale of 500 m. This results in the background diffusivity determining

127   the value of diapycnal mixing in most parts of the pelagic pycnocline. Observations from the

128   Southern Ocean show that diapycnal mixing is much larger than in other oceans [*Naveira*

129   *Garabato et al.*, 2004]. We account for these observations by limiting K_v to $\geq$ 1 cm^2/s south of

130   40°S. Thus, the variations in K_{bg} affect mixing only in the open ocean north of 40°S.

131

132   A simple one-layer atmospheric energy-moisture balance model (EMBM) interactively

133   calculates heat and water fluxes to ocean, land and sea ice, while wind velocities are prescribed

134   from the NCAR/NCEP monthly climatology in the momentum transfer to the ocean and to a

135   dynamic-thermodynamic sea ice model. The model does not use flux corrections. The model of

136   the terrestrial vegetation and carbon cycle [*Meissner et al.*, 2003] is based on the Hadley Center

137   model TRIFFID. The ocean biogeochemical model is based on the NPZD (nutrient,

138   phytoplankton, zooplankton, detritus) ecosystem model of *Schmittner et al.* [2005b], and

139   includes a parameterization of fast nutrient recycling due to microbial activity after *Schartau and*

140   *Oschlies* [2003]. It solves prognostic equations for two phytoplankton classes (nitrogen fixers

141   and other phytoplankton) as well as for nitrate, phosphate, oxygen, dissolved inorganic carbon,

142   alkalinity, radiocarbon and chlorofluorocarbons as tracers. The biogeochemical/carbon cycle

143   model is described in detail in *Schmittner et al.* [2008]. Biological uptake and release occurs in

144   fixed elemental ratios of carbon, phosphate, nitrate and oxygen. Calcium carbonate production is

145   parameterized as a fixed ratio of the production of particulate organic matter in the water

146   column. Remineralization of calcium carbonate is determined by instantaneous sinking with an

147   e-folding depth of 3500 m.

148

149   The ensemble consists of 8 models with K_v=(0.01,0.05,0.1,0.15,0.2,0.3,0.4,0.5). (In the

150   following, for brevity, we omit the units of K_v, which are in cm^2/s.) Each model version is

151   restarted from an 8000 year control integration with K_v=0.15, and spun up for an additional

152   3000-4000 years (longer for smaller K_v) using constant pre-industrial forcing until climate and

153   carbon cycle are in quasi-equilibrium. Initially atmospheric CO_2 is fixed at 280 ppmv, but for the

154   last ~1000 years of the spin up it is calculated interactively. Equilibrium is determined if changes

155   in atmospheric CO_2 are less than 5 ppmv per 1000 years, so that at the end of the spin up

156   atmospheric CO_2 is within ±5 ppmv of ice core measurements of its pre-industrial value of 280

157   ppmv (Figure 1) for all model versions. Subsequent estimates [*Crowley*, 2000] of historical

158   forcing from year 1800 to 1998 AD are applied, considering changes in solar insolation, volcanic

159   and anthropogenic aerosol and greenhouse gases, followed by CO_2 emission scenario SRES A2

160   until 2100 and a linear decrease of emissions to zero from year 2100 to 2300 (Figure 1). Solar,

161   aerosol and non-CO_2 greenhouse gas forcings have been held constant at 1988-1998 levels for

162   the future simulations.

163

164   **2.2. Observations**

165   We calculate probability densities for nine three-dimensional tracer distributions from two

166   databases. Temperature (*T*) [*Locarnini et al.*, 2006], salinity (*S*) [*Antonov et al.*, 2006], phosphate

167   (PO_4) [*Garcia et al.*, 2006b]*,* apparent oxygen utilization (*AOU*) [*Garcia et al.*, 2006a] and

168   preformed phosphate (*P*=PO_4-AOU/170*) are taken from the World Ocean Atlas 2005 (WOA05,

169   data downloaded from ftp.nodc.noaa.gov/pub/data.nodc/woa/WOA05nc) and radiocarbon

170   ($\Delta^{14}C$), chlorofluorocarbon 11 (*CFC11*), dissolved inorganic carbon (*DIC*), and alkalinity (*ALK*)

171   come from the Global Ocean Data Analysis Project (GLODAP) [*Key et al.*, 2004]. Both

172   databases provide data on a $1\times1°$ grid with 33 vertical levels. The observations are averaged

173   onto the $1.8\times3.6°$ model grid with 19 vertical levels. GLODAP data represent the 1990s and are

174   compared with the decadal model mean from 1990-2000, whereas WOA05 data represent the

175   1950-2000 and are compared to the model mean during this period.

176

177   **2.3. Observation Error Estimates**

178   To quantitatively compare observations to model projections, it is necessary to have an estimate

179   of the observation errors. The error size determines how far from the data a model can be and

180   remain consistent with the observations. Spatially variable error estimates for the observations

181   (σ_{Oi}) are available for $\Delta^{14}C$, *CFC, DIC and ALK* from the GLODAP data set representing errors

182   resulting from the objective analysis (mapping) procedure used to interpolate and extrapolate

183   observations to a global grid. Due to the sparse observations the GLODAP error estimates are

184   horizontally correlated with a correlation length scale of 10-20°. Thus the GLODAP errors are

185   simply averaged onto the model grid.

186

187   The WOA05 provides the standard error for each unanalyzed variable, which is the standard

188   deviation of the mean divided by the square root of the number of observations in each grid cell.

189   Following the recommendation in the WOA05 documentation (available at

190   http://www.nodc.noaa.gov/OC5/WOA05/pr_woa05.html) the error due to the objective analysis

191   is estimated as the difference between the value of the analyzed field and the mean at each grid

192   cell containing observations. For *T* and *S,* which have observations at almost all grid points, this

193   error is horizontally uncorrelated. The global horizontal root mean square is calculated at each

194   depth level, representing the (horizontally uniform but vertically varying) mapping error. This

195   mapping error is added to the standard error to yield the spatially variable total error estimate.

196

197   For PO_4 and *AOU* the data density is too sparse to calculate an error estimate due to the mapping

198   procedure, because the calculation can only be performed for points that include data. For this

199   reason we do not use the analyzed fields but rather we use the unanalyzed mean (the average of

200   the raw observations in any given $1x1°$ data grid box). This limits the number of grid cells to

201   those containing observations. The observations are averaged onto the model grid, and model

202   grid cells without observations are discarded in the analysis. In this case the total error of the

203   observations is only the standard error of the mean (no mapping error). For all WOA05 variables

204   the total errors are horizontally uncorrelated and hence they are averaged onto the model grid

205   and divided by $2.55 = \sqrt{3.6 \times 1.8}$ in order to account for 6.48 independent data grid boxes in one

206   model grid cell.

207

208   2.4. Statistical Analysis

209   We assess the compatibility of different diapycnal diffusivities with observed tracer

210   measurements using Bayesian inference to compute the relative probability of each of the eight

211   diffusivities in our ensemble implied by each of the nine tracer fields. Two different methods are

212   used in the model assessment. The first computes the root mean squared (RMS) error (E) for

213   each model, including the full three-dimensional (3D) spatial fields of observations. This method

214   neglects the correlation of the errors and requires the size of the errors to be specified. The

215   second method considers the correlation of the errors and determines the error magnitude and

216   bias endogeneously from the data-model residuals. However, due to computational constraints it

217   uses only 1-dimensional data (globally horizontally averaged depth profiles). Both methods, as

218   well as the relations between them, are described in detail below.

219

220   2.4.1.   The 3D Method

221   Models that greatly differ from the observations are judged less probable than models whose

222   deviations from the data are small. To quantify this intuition, it is necessary to mathematically

223   specify what "small" means. We introduce an error estimate σ to set the scale against which

224   data-model discrepancies are evaluated. These deviations are deemed large or small relative to

225   the magnitude of σ. Observations can differ from model predictions for two reasons: model

226   structural error, and observational/measurement error. The quality of data-model agreement

227   depends on how large we judge these errors to be (see Sec. 3.7). However, errors can be difficult

228   to estimate *a priori* (before seeing the observational data), especially when model structural

229   errors are substantial. Observational errors usually can be estimated from known properties of

230   the measurement system (Sec. 2.3), but the size of the model error typically cannot be

231   determined without comparing the model output to observations.

232

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

234

235
$$\overline{E_i^{\,2}} = \overline{\left(\frac{O_i - M_i}{\sigma_i}\right)^2}.$$
(1)

236

237   The overbar denotes the global, volume-weighted average. Deviations of each modeled 3D tracer

238   field $M_i = M_i(x,y,z) = \overline{M_i} + M_i'$ from the observations $O_i = O_i(x,y,z) = \overline{O_i} + O_i'$ are weighted by a

239   combined error estimate ($\sigma_i^2 = \sigma_{Oi}^2 + \sigma_{Mi}^2$) for the observations σ_{Oi} and the model σ_{Mi}. (The

240   prime denotes the deviation from the global mean.) Our methods for estimating the observation

241   and model errors are discussed in Section 3.4.1.

242

243   The models often show bias relative to the observations, so that their mean prediction differs

244   from the mean of the observations. To distinguish between the amount of error introduced by

245   model bias and the amount of error unrelated to bias, we also consider the bias-corrected RMS

246   error. This error is calculated by subtracting the global mean bias $b_i = \overline{O_i} - \overline{M_i}$, so that the bias

247   corrected residuals $O_i - M_i - b_i$ have zero mean. The bias-corrected RMS error is then

248    $\overline{E_i'^{\,2}} = \overline{\left(\frac{O_i - M_i - b_i}{\sigma_i}\right)^2} = \overline{\left(\frac{O_i' - M_i'}{\sigma_i}\right)^2}$. The error $\overline{E_i'^{\,2}}$ excludes information about the global mean

249   data-model misfit. Assuming the errors are independent and identically-distributed random

250   variables, the probability density

251

252
$$L(O_i \,|\, K_v) \propto \exp\left(-\frac{1}{2} E_i^{\,2}\right)$$
(2)

253

254   is the likelihood that the observations O_i could arise from the model with parameter K_v. Above,

255    $E_i^{\,2} = \overline{E_i^{\,2}} \times N$ is the (volume-weighted, error-weighted) sum of squared errors, equal to the mean

256   squared error $\overline{E_i^{\,2}}$ times the number of data points. More precisely, assuming a known error σ,

257 the probability in equation (2) is a normal likelihood function: the observations are assumed to

258 be drawn from a normal distribution with mean centered on the model output

259 $(O \sim N(\mu = M, \sigma^2))$. Bayes' theorem states that the posterior probability density function (PDF)

260 for K_v is proportional to the product of the likelihood of the observations with the prior PDF of

261 K_v, p(K_v):

262

263
$$p(K_v|O_i) \propto L(O_i|K_v) \times p(K_v) \ . \tag{3}$$

264

265 We adopt a uniform prior PDF for K_v. If the errors in different tracers are independent of each

266 other – which is generally *not* the case, as discussed below – likelihoods for individual tracers

267 can be multiplied to yield the combined likelihood of all tracers, $L(O|K_v) = \prod_i L(O_i|K_v)$.

268 Probability-weighted projections for a climate variable T are obtained by averaging over the

269 possible values of K_v,

270
$$\bar{T} = \int_{Kvmin}^{Kvmax} T(K_v) \cdot p(K_v|O) \, dK_v \ , \tag{4}$$

271

272 if the PDF is defined on the interval $[K_{vmin}, K_{vmax}]$.

273

274 **2.4.2. The 1D Method**

275

276 The above 3D method ignores spatial autocorrelation of the data-model residuals, $R_i = O_i - M_i$,

277 which is known to lead to overconfident parameter estimates [*Zellner and Tiao*, 1964]. In

278 addition, the above formulation presumes that the residual error σ is known, but as discussed in

279 the previous section, it can be difficult to estimate *a priori.* Here we develop a relatively simple

280 and computationally efficient method to estimate the combined effects of observation errors and

281 model structural errors endogenously from the overall data-model misfit. This method is more

282 computationally expensive than the 3D method, so we apply it to small 1D aggregated data sets

283 instead of to the full 3D spatial fields.

284

285 When the errors are uncorrelated, only their magnitudes σ_i need to be specified. If the errors are

286 correlated, the correlation between errors must be specified in addition to their magnitudes. We

287 generalize from the error variances σ_i^2 to an error covariance matrix Σ, which includes the error

288 variances and the spatial correlations between points. In the 3D method we use the weighted sum

289 of squared errors, $\sum_i (O_i - M_i)^2 / \sigma_i^2$, to quantify model skill. This error measure is not

290 appropriate when the errors are correlated. Correlated errors effectively provide fewer

291 independent data points than uncorrelated errors. An appropriate measure should penalize

292 models less harshly when correlation is present, since fewer independent data are assimilated.

293 To include correlation the sum of squared errors generalizes to a quantity involving the error

294 covariance matrix, known as the Mahalanobis distance [*Mahalanobis*, 1936], which appears in

295 the multivariate normal distribution:

296

297
$$E_i^2 = (O_i - M_i)^T \Sigma_i^{-1} (O_i - M_i). \tag{5}$$

298

299 This expression reduces to the sum of squared errors when the covariance matrix is diagonal

300 with entries σ_i^2, i.e., when it contains only variances but no off-diagonal correlations. (In the

301 remainder of this section we omit the subscript i when referring to each tracer.)

302

303 Only small covariance matrices are used here because matrix inversion is computationally

304 expensive. To reduce the covariance matrix to a size which is feasible to invert, we consider only

305 a 1D globally averaged spatial field of tracer data $O_i(z)$ and $M_i(z)$ as a function of depth z. Each

306 field is reduced to 18 depths data points. We assume the covariance matrix Σ is given by a

307 stationary squared-exponential covariance function between depths z_j and z_k,

308 $\Sigma_{jk} = \sigma^2 \exp(-|z_j - z_k|^2 / \lambda^2)$, where σ^2 is the residual variance and λ is a range or correlation

309 length parameter. Including the possibility of a constant model bias, b, the observations are

310 assumed to be drawn from a multivariate normal likelihood centered on the bias-corrected model

311 output $(O \sim MVN(\mu = M + b, \Sigma))$.

312

313 In the 3D method, the residual error σ, the correlation length λ, and the model bias b are assumed

314 known constants (with $\lambda=0$, and b=0 or set to the difference in observational and model means).

315   These constants may differ between tracers. In the 1D method we relax these assumptions by

316   treating the three constants as unknown statistical parameters. The full Bayesian approach, which

317   we approximate, is to calculate a joint posterior PDF for all the uncertain parameters, including

318   the model parameter K_v and the three statistical parameters. By Bayes' theorem, this posterior

319   probability is proportional to the product of the likelihood of the observations with the prior

320   probability of the parameters,

321

322
$$p(K_v,\sigma,\lambda,b|O) \propto L(O|K_v,\sigma,\lambda,b) \times p(K_v,\sigma,\lambda,b). \tag{6}$$

323

324   We are most interested in the probabilities of the different model diffusivities, not of the

325   statistical parameters. We can obtain the posterior PDF $p(K_v|O)$ for K_v alone by integrating the

326   joint posterior $p(K_v,\sigma,\lambda,b|O)$, Eq. 6, with respect to the three statistical parameters:

327

328
$$p(K_v|O) = \iiint p(K_v,\sigma,\lambda,b|O)\,d\sigma\,d\lambda\,db. \tag{7}$$

329

330   However, for computational simplicity, we avoid performing this integral by fixing the statistical

331   parameters at their best-fit values σ*, λ*, β*. This gives an approximate proportionality

332

333
$$p(K_v|O) \approx p(K_v|O,\sigma^*,\lambda^*,b^*) \propto L(O|K_v,\sigma^*,\lambda^*,b^*) \times p(K_v,\sigma^*,\lambda^*,b^*). \tag{8}$$

334

335   Fixing the statistical parameters ignores their uncertainty but still accounts for the presence of

336   model error, bias, and correlation. These quantities are estimated from the data-model misfit

337   instead of assumed from expert prior judgement. The best estimate for σ*, λ*, β* is obtained by

338   numerically maximizing the posterior probability (Eq. 6) using a global optimization method

339   [*Storn and Price*, 1997] to account for potential multimodality. Posterior maximization is

340   analogous to maximum likelihood estimation [*Lehmann and Casella*, 2003], except that the

341   likelihood is modified by prior constraints on the parameters. The statistical parameters are

342   separately optimized for each tracer, allowing the estimated residual structure to vary between

343   tracers. For every tracer, the parameters are also re-optimized for each member of the ensemble.

344   In other words, the statistical parameters are allowed to depend on K_v. The logic behind this

345   assumption is that the model error depends on the model parameters, since poorly fitting models

346   should have larger model error and bias. We linearly interpolate the posterior probability onto a

347   regular grid of K_v and normalize the integral to unity to arrive at a proper probability density

348   function.

349

350   We choose a uniform prior for the model parameter K_v. The correlation length prior is

351   p(λ)=*Lognormal*(5.5,0.5^2). That is, ln(λ) is normally distributed with a mean 5.5 and standard

352   deviation 0.5. This prior puts most of its probability mass between 0 and 600 meters, excluding

353   unphysically large correlations between the ocean surface and bottom in a globally averaged

354   depth profile. We use a joint prior for the residual variance and bias, $p(b/\sigma) = N(0,0.5^2)$. This

355   prior is selected so the model bias for the best K_v value is assumed to be likely smaller than the

356   residual error (i.e., b/σ is near zero). This gives low prior weight to models with large biases,

357   where "large" is quantified relative to the size of the bias-corrected error, σ. Exploratory

358   analysis indicates that an improper, unbounded uniform prior for the range or bias parameters

359   can lead to ill-conditioned covariance matrices and non-robust results for the K_v posterior

360   distribution.

361

362   **3. Results**

363   **3.1. Global Metrics**

364   Observed atmospheric CO_2 concentrations and global mean surface air temperatures are

365   simulated roughly equally well in all model versions, irrespective of the value of K_v (Figure 1).

366   This is also true for the ocean heat content changes, which are very similar in all simulations

367   (Figure 2). As already concluded in *Tomassini et al.* [2007], these globally aggregated

368   observations provide relatively poor constraints on K_v. The model suggests, however, that this

369   situation might change in the future, because the simulations for different K_v values diverge

370   notably during the 21st century. For example, at year 2100 differences in CO_2 concentrations are

371   about 70 ppmv (Figure 1). This suggests that variations in diapycnal diffusivity alone can

372   account for about 25% of the range in the C^{4}MIP models. At year 2300 differences in CO_2

373   concentrations are more than 200 ppmv. Differences in projected global average surface air

374   temperatures are 0.8°C in model year 2100 and 1°C in year 2300.

375

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

376

377 Larger diapycnal mixing leads to faster oceanic uptake of heat and CO_2 in the model. Both

378 effects tend to delay and reduce atmospheric warming. Faster CO_2 uptake leads to lower

379 atmospheric CO_2 concentrations and thus reduced radiative forcing, whereas faster heat uptake

380 leads to slower warming of surface waters and therefore delayed warming of surface air

381 temperatures. We separate these two effects by comparing a simulation with weak mixing ($K_v =$

382 0.1) forced with interactive CO_2 to one forced with a prescribed CO_2 evolution (and thus

383 radiative forcing) taken from a run with vigorous mixing ($K_v = 0.5$). The difference in surface air

384 temperature evolution between these two simulations is due only to the effect of slower ocean

385 heat uptake. The effect of different ocean carbon uptakes is quantified by comparing the

386 simulation with prescribed CO_2 to the fully coupled run with $K_v = 0.1$ (Figure 3). The global

387 surface air temperature increase in the run with prescribed CO_2 evolution is about half way

388 between the experiments with high and low K_v. About 55% (0.5 K) of the reduced warming of

389 air temperatures in the high versus the low K_v simulation is explained by differences in ocean

390 heat uptake alone, and 45% is caused by faster CO_2 uptake. This demonstrates that both effects,

391 slower heat uptake and slower carbon uptake, contribute about equally to the reduced warming in

392 the high mixing model projections.

393

394 We analyze the sensitivity of land (ΔC_L) and ocean (ΔC_O) carbon uptake until year 2100 with

395 respect to changes in atmospheric CO_2 ($\beta_L = \Delta C^u_L/\Delta C^u_A$; $\beta_O = \Delta C^u_O/\Delta C^u_A$) and climate ($\gamma_L =$

396 $(\Delta C^c_L - \beta_L \Delta C^c_A)/\Delta T^c$; $\gamma_O = (\Delta C^c_O - \beta_O \Delta C^c_A)/\Delta T^c$) following *Friedlingstein et al.* [2006], where *c* and

397 *u* superscripts denote the coupled and uncoupled (constant climate) runs respectively, ΔT is the

398 global mean surface air temperature change and ΔC_A is the atmospheric CO_2 anomaly. As

399 expected the land sensitivities ($\beta_L(K_v = 0.1) = \beta_L(K_v = 0.5) = 1.4$ GtC/ppm; $\gamma_L(K_v = 0.1) = -114$

400 GtC/K; $\gamma_L(K_v = 0.5) = -116$ GtC/K) are very similar between the different K_v simulations. (The

401 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.)

402

403 However, ocean carbon uptake due to changes in atmospheric CO_2 alone is 30% smaller in the

404 low mixing model ($\beta_O(K_v = 0.1) = 1$ CtC/ppm) compared to the high mixing model ($\beta_O(K_v = 0.5)$

405   = 1.4 CtC/ppm). This suggests that differences in ocean diapycnal mixing alone can explain half

406   the range of β_O in the C^{4}MIP models (0.8-1.6 GtC/ppm) and re-emphasizes the important role of

407   diapycnal mixing on anthropogenic carbon uptake by the ocean. There are, of course, other

408   processes that additionally determine ocean carbon uptake (under fixed climate), such as the

409   strength of the overturning circulation and convection, mixed layer depths, and air-sea gas

410   exchange (driven by factors such as sea ice and wind velocities).

411

412   Ocean carbon uptake decreases in the model simulations as climate warms due to increasing

413   stratification of the upper ocean. A greater weakening of the ocean carbon sink corresponds to

414   more negative values of γ_O. In the high mixing models this decrease is larger ($\gamma_O(K_v = 0.5) = -45$

415   GtC/K) than in the low mixing models ($\gamma_O(K_v = 0.1) = -31$ GtC/K). The C^{4}MIP models range

416   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

417   box model; the next largest C^{4}MIP model value of γ_O is -46 GtC/K). At year 2100 the ocean

418   takes up 4.8 GtC/yr in the low mixing model versus 6.2 GtC/yr in the high mixing model. Most

419   (8 out of 11) C^{4}MIP models lay within that range of ocean carbon uptake.

420

421   The strength of positive climate-carbon cycle feedbacks can be quantified by the feedback gain

422   ($g = 1-\Delta C^u_A/\Delta C^c_A = -\alpha(\gamma_L+\gamma_O)/(1+\beta_L+\beta_O)$), where $\alpha = \Delta T^c/\Delta C^c_A$ is the transient climate

423   sensitivity) [*Friedlingstein et al.*, 2006]. The effects of higher β_O and larger (negative) γ_O almost

424   completely compensate each other, but due to the larger transient climate sensitivity ($\alpha (K_v = 0.1)$

425    $= 0.0060$K/ppm versus $\alpha (K_v = 0.5) = 0.0055$K/ppm) there is a modest (10%) increase in gain in

426   the low mixing model ($g(K_v = 0.1) = 0.2$) compared to the high mixing model ($g(K_v = 0.5) =$

427   0.18). The range of g in the C^{4}MIP models is 0.04-0.31, which includes differences in both

428   terrestrial and oceanic carbon cycle contributions to the total climate-carbon cycle feedback, in

429   addition to different values of transient climate sensitivity. According to our analysis, while

430   different K_v values can explain a substantial portion of the range of ocean carbon uptake between

431   models, K_v differences can explain only a relatively small proportion of the inter-model range in

432   net climate-carbon cycle feedback strength.

433

434   **3.3. Influence of Isopycnal Mixing on Climate-Carbon Cycle Projections**

435   Additional experiments with changes in along isopycnal diffusivity from 800 to 2000 m^2/s show

436   no differences in simulated atmospheric CO_2 or climate demonstrating that uncertainties in this

437   parameter do not affect projections.

438

439   **3.4. Model Assessment Using Spatially Resolved Ocean Tracer Observations**

440   **3.4.1. Model and Observation Error Estimates**

441   Assessing model skill requires an estimate of the discrepancy between observations and model

442   predictions. The 3D method's likelihood function, Equation (2), assumes that the standard

443   deviation of the data-model residuals (σ) is known. The 1D method estimates this error from

444   the residuals by an optimization procedure (Sec. 2.4.2). For the 3D method we choose to

445   determine the residual error by more informal means. By definition, the residual error

446    $\sigma_i = \sqrt{\sigma_{Oi}^2 + \sigma_{Mi}^2}$ should be similar to the standard deviation of the residuals, $\sigma_i \approx SD(O_i - M_i)$.

447   For the 3D method we choose the model error σ_{Mi} such that this is the case for one of the best

448   fitting models ($K_v = 0.15$). (See Table 1 for values.) This model error is then applied to all

449   ensemble members.

450

451   The model error estimates σ_{Mi} can also be interpreted as measures of model quality; they can be

452   used for different models and are suitable for model intercomparisons. For example, for

453   temperature and salinity, the values in the second row of Table 1 (3D data and $\sigma_O = 0$)

454   correspond to the global RMS error. They can be compared to those reported for the OCMIP

455   models [*Doney et al.*, 2004, Table 2] and a subset of the Intergovernmental Panel on Climate

456   Change Fourth Assessment Report (IPCC AR4) models [*Schmittner et al.*, 2005a]. The OCMIP

457   range for 3D models without internal restoring is 0.84-2.18 K for temperature and 0.15-0.31 for

458   salinity; for the IPCC AR4 fully coupled ocean atmosphere models it is 0.86-2.97 K for

459   temperature and 0.20-0.38 for salinity.

460

461   We use the observational errors reviewed in Section 2.3 for the 3D method. For the 1D method

462   we assume that the observation error is negligible compared with the model error, since the

463   global averaging leads to very small observational errors (decreasing with $\sim N_{eff}^{-1/2}$, where N_{eff} is

464   the effective number of observations). This is consistent with the 1D data-model residuals, which

465 show a smoothly varying structure more indicative of systematic model error than random

466 observation noise.

467

468 With these error estimates, we evaluate the skill of each of the eight models in the ensemble

469 using three metrics. We use the root mean squared (RMS) error introduced in Section 2.4.1, as

470 well as the bias-corrected RMS error. We also compute the correlation between the observations

471 and each model. A higher correlation indicates greater similarity between the model and the

472 observations. We conduct sensitivity studies to explore how model skill varies with K_v as

473 determined by each of the three skill measures.

474

475 In the following discussion we distinguish between tracers which are influenced by physical

476 processes only such as T, S, $\Delta^{14}C$, and $CFC11$, and those tracers strongly affected by biological

477 processes such as PO_4, AOU, P^*, DIC, and ALK, since the latter also depend on the choice of

478 uncertain biological model parameters. Biological effects on the radiocarbon distribution in the

479 ocean are about 2 orders of magnitude smaller than the physical effects of decay and air-sea gas

480 exchange.

481

482 The RMS and bias-corrected RMS errors, E and E', are plotted in Figure 4 together with the

483 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

484 the full 3-dimensional data. The different tracers show different sensitivities to K_v depending on

485 the global metric considered. When measured by the RMS error E, the model skill for the $\Delta^{14}C$,

486 AOU, P^* and DIC tracers show the largest sensitivity to changes in K_v. Much of this sensitivity,

487 however, is due to the model bias, as revealed by the difference between E and the bias corrected

488 error E'. For S, ALK and PO_4 the bias is zero ($E=E'$) because neither of these tracers exchanges

489 with other climate system components in the model and hence their ocean inventories are fixed.

490 When measured by the correlation coefficient, the model skill for the AOU and S tracers are most

491 sensitive to variations in K_v.

492

493 Most tracers are in better agreement with the observations for small values of K_v, both for E and

494 r as metrics. Correlation coefficients between model output and observations peak between 0.05

495 and 0.15 for all tracers except DIC and T, which are rather insensitive. AOU, DIC, $\Delta^{14}C$ and P^*

17

496 show very large biases for large values of K_v. The deep ocean is much too young ($\Delta^{14}C$ too high),

497 too vigorously ventilated (AOU too low), too poor in inorganic carbon and too high in preformed

498 nutrients. Even if the bias is removed, the bias-corrected RMS error E' in AOU is still much

499 larger for the high K_v models. $CFC11$ and S are both moderately sensitive and show better

500 agreement with the observations for low K_v, irrespective of the metric considered. PO_4 and ALK

501 are also moderately sensitive and show minima in RMS errors and maxima in correlation around

502 $K_v = 0.15$.

503

504 **3.5. Probabilities From the 3D Method**

505 PDFs from the 3D method suggest that $\Delta^{14}C$ is the most sensitive of the physical variables to

506 changes in K_v, followed by $CFC11$, T and S (Figure 5). $\Delta^{14}C$, S, and $CFC11$ show the maximum

507 probability for small values of K_v. For $\Delta^{14}C$ the probability for small K_v is about three times as

508 high as that for high K_v. T shows a broad maximum for $0.2 \leq K_v \leq 0.4$ and smallest probabilities

509 for very high and very low values of K_v. The biological tracers (lower panel in Fig. 5) are all

510 sensitive to variations in K_v, in particular AOU, DIC, and P^* which are 5-10 times more likely

511 for low then high K_v. ALK and PO_4 show maxima for K_v around 0.15-0.2.

512

513 **3.6. Probabilities From the 1D Method**

514 Figure 6 shows PDFs for the same variables but using the 1D method. The most obvious

515 difference is that the 1D PDFs are much sharper than those obtained with the 3D method. This

516 might be counterintuitive, since information was lost by aggregating the data from 3D to 1D (we

517 discuss this effect further below). The 1D method yields maxima for all tracers for $0.01 \leq K_v \leq$

518 0.2. Probabilities for $K_v > 0.4$ are very small for all tracers. Thus the two statistical methods

519 agree that high K_v models are less consistent with the observations than low K_v models. Both

520 methods also exhibit similar shapes for most tracers. E.g. $\Delta^{14}C$, S, AOU, DIC and P^* all have

521 maxima for $K_v < 0.2$, $CFC11$, ALK and PO_4 show maxima for $0.1 \leq K_v \leq 0.2$, and T shows a

522 broad maximum for $0.2 \leq K_v \leq 0.3$.

523

524 **3.7. Sensitivity Tests**

525 We conduct four simple sensitivity analyses of the 3D method to gain some insights into the

526 factors that influence the differences in the posterior PDFs between the 3D and 1D methods

527 (Figure 7). First, we test the assumption of neglecting the error of the observations by setting

528 $\sigma_O=0$ and re-estimating the total error σ (Table 1). Comparing the resulting PDFs (blue lines)

529 with the original PDFs (black lines) shows that this effect is negligible for most tracers. Only

530 *CFC11, $\Delta^{14}C$,* and *DIC* show small differences.

531

532 Second, we calculated the PDFs for 1D (horizontally averaged) data but using the 3D method as

533 described above (red solid lines in Fig. 7). The re-estimated errors (Table 1) are much smaller

534 than in the 3D case for all tracers, indicating that the model has considerably more skill in

535 reproducing the horizontally averaged observations than the full 3D distributions. The resulting

536 PDFs are therefore much sharper than in the 3D case and for most tracers they are similar to the

537 PDFs from the 1D method (Figure 6).

538

539 Third, we evaluate the effects of correcting for spatial autocorrelation. Following *Ricciuto et al.*

540 [2008] we remove the lag-1 autocorrelation (*a*) from the 1D residuals according to:

541

$$\overline{E_i^2} = \overline{\left(R_i(z_k) - a R_i(z_{k-1}) \right)^2}. \tag{9}$$

543

544 As expected from earlier studies [*Ricciuto et al.*, 2008; *Zellner and Tiao*, 1964] this approach to

545 account for the autocorrelation (green lines in Fig. 7) leads to broader PDFs (compared to the red

546 solid lines). Neglecting spatial autocorrelation typically results in overconfident parameter

547 estimates. The fact that the PDFs are quite different emphasizes the importance of properly

548 considering spatial autocorrelation.

549

550 Fourth, the PDFs are re-calculated for the 1D data (without subtracting autocorrelation) but using

551 the error estimate from the 3D method with $\sigma_O=0$ (red dashed lines in Fig. 7). Thus the

552 difference between the red solid lines and the red dashed lines in Fig. 7 isolates the effect of

553 different estimated σ. The difference between the red dashed lines and the blue lines isolates the

554 effect of the reduced information content in the 1D versus the 3D residuals. For most tracers the

555 PDFs are broader than those in the high σ cases (red solid lines) and more similar to the 3D case

556 (blue lines). This indicates that the most important reason for the difference between the 1D and

557 3D methods (and the explanation for the sharper PDFs in the 1D method) is the differently

558 estimated σ. It also suggests that spatial aggregation, despite a loss of information, can help to
559 improve the model skill, and as a consequence lead to sharper PDFs. For PO_4, $CFC11$, and ALK
560 the red dashed lines deviate substantially from the blue solid lines. This indicates an important
561 loss of information due to the averaging. These tracers might not be suitable for the 1D method.
562
563 **3.8. Probabilities for Multiple Tracers**
564
565 Each of the tracers examined above contains different information and leads to a different PDF
566 for K_v. Our goal, however, is to produce a single PDF combining the information from all tracers
567 as outlined in sub-section 2.4.1.
568
569 The distribution of each tracer is influenced not only by diapycnal mixing and the large-scale
570 ocean circulation, but also by other processes. Some tracers, such as T, S, $CFC11$, $\Delta^{14}C$, and
571 DIC, are also influenced by air-sea exchange. Thus, the model errors, and hence the PDF, for T
572 e.g., might be influenced by model biases in ocean-atmosphere heat fluxes, which are controlled
573 by radiative fluxes as well as sensible and latent heat fluxes. The PDF for S, on the other hand, is
574 influenced by surface ocean water fluxes, which are determined by evaporation, precipitation and
575 river runoff, and thus by the atmospheric hydrological cycle. Because different physical
576 processes control heat and water fluxes it is highly unlikely that model errors in heat fluxes are
577 correlated with errors in water fluxes. Similarly, the air-sea fluxes of carbon, radiocarbon and
578 CFCs are presumably independent from heat and water fluxes. Thus, considering multiple tracers
579 can possibly average out model errors in individual air-sea fluxes. If the errors in the tracer
580 residuals are independent between tracers, a combined likelihood for all tracers can be calculated
581 by multiplying the likelihoods of the individual tracers as described at the end of section 2.4.1.
582
583 On the other hand, if tracers are not independent multiplication of the likelihoods would lead to
584 overconfident and possibly biased PDFs. Sinking of particulate organic matter (the soft-tissue
585 biological pump), for instance, influences PO_4, AOU, and DIC and thus errors in those tracers
586 cannot be expected to be independent. An objective way to determine independence between
587 different tracers is to examine correlations between the errors of the residuals. As shown in Table
588 2, the different tracers are generally not independent. PO_4, AOU, and DIC are clearly related for

20

589   the reasons mentioned above, but other tracer residuals (such as T and S) are also correlated, for

590   less obvious reasons.

591

592   At this point no method that we are aware of has accounted for the cross-tracer correlation. It is

593   highly desirable to develop such a method in the future. For the time being we calculate PDFs for

594   different combinations of uncorrelated tracers (Figure 8). All combined PDFs show low

595   probability for models with high mixing rates ($K_v > 0.3$). The different tracer combinations do

596   not agree well for the probability of low mixing models. Some show a distinct maximum around

597   0.1-0.2 and considerably lower probabilities for lower K_v, whereas others show high probabilities

598   for the lowest diffusivities. We conclude that the observations put a firm upper limit on the

599   diffusivities, whereas no unequivocal lower limit can be determined.

600

601   **4. Discussion**

602   One issue that has not been addressed here is parameter interactions. Generally model tracer

603   distributions are influenced by more than one parameter, each of which is uncertain. Thus, the

604   results obtained by varying one parameter depend on the values of many other parameters. This

605   is also true in our case, and hence the probabilities for different K_v presented here are tentative

606   and should be regarded as a test of the methodology rather than a definitive result.

607

608   Parameter interactions might be most obvious for tracers affected by biological processes such as

609   PO_4, AOU, P^*, DIC and ALK, which are sensitive to ill-constrained biological model parameters.

610   Surface nutrient concentrations and deep ocean AOU, P^* and DIC, for instance, all depend

611   strongly on the maximum growth rate of phytoplankton (γ) which determines the efficiency of

612   the biological pump. The vertical alkalinity gradient is controlled by the fixed ratio of calcium

613   carbonate versus particulate organic carbon production ($R_{CaCO3/POC}$). These biological model

614   parameters were tuned for a model version with $K_v = 0.15$ ($\gamma = 0.13$ d^{-1}, $R_{CaCO3/POC} = 0.03$). Thus

615   larger errors for those tracers in models with different K_v can be expected because the biological

616   parameters are unadjusted. Models with $K_v > 0.15$ therefore overestimate surface nutrient

617   concentrations because of more intense advective and diffusive transport of nutrient rich deep

618   waters to the surface. Similarly, models with $K_v > 0.15$ underestimate the efficiency of the

619   biological pump and hence the deep ocean AOU and DIC, and overestimate P^*. Thus γ should be

620   increased together with K_v. Faster rates of nutrient input into the euphotic zone, in the strong

621   mixing models, also lead to increased primary and export production [*Schmittner et al.*, 2005b]

622   and higher production of $CaCO_3$, resulting in overestimated vertical alkalinity gradients. Thus,

623    $R_{CaCO3/POC}$ should be decreased as K_v is increased.

624

625   Due to computational constraints we are currently not able to retune the biological parameters for

626   each model version with different K_v. A simple optimization of biological parameters for the

627   model version with $K_v = 0.5$ ($\gamma = 0.2$ d^{-1}, $R_{CaCO3/POC} = 0.02$) results in a decrease of the errors with

628   respect to the untuned values shown in Figure 3, but the errors are still significantly larger than

629   those of the low K_v models. Thus, the true likelihoods for the biological tracers would

630   presumably increase for model versions with high K_v. It is highly desirable to include these

631   known cross-parameter dependencies in a larger model ensemble in the future. Of course, tracer

632   distributions not affected by biological parameters, such as $\Delta^{14}C$ (radiocarbon in our model is not

633   influenced by biological parameters) and CFCs, do not suffer from this complication. Therefore

634   our conclusion that models with $K_v > 0.3$ cm^2/s are increasingly inconsistent with observations

635   holds true based on these tracers alone.

636

637   An outstanding question remains whether the C^4MIP model results represent the true uncertainty

638   of our understanding of the climate-carbon cycle system. This question can be addressed only by

639   a systematic comparison with observations. We have shown here that low values of K_v are most

640   consistent with ocean tracer observations, and that most of the C^4MIP models fall within the

641   range of ocean carbon uptake simulated by varying K_v values in this study. If the values for K_v

642   were known for the different C^4MIP models, it would be possible to reject projections from

643   models with high K_v values, or judge them as less reliable than those from models with low K_v.

644   However, we are not aware of a published documentation of the values of K_v used by the C^4MIP

645   models (effective diapycnal diffusivity can also contain a difficult-to-evaluate numerical

646   component). There is an additional complication arising from different structural types of ocean

647   models represented in C^4MIP (box models, versus 2D models, versus GCMs). In practice,

648   therefore, it remains difficult to assign the likelihoods we have derived here directly to the

649   C^4MIP model projections. However, we think that the methodology developed here can be used

650 for multi-model assessments in the future, given that spatially resolved tracer model data output

651 is provided.

652

5. Conclusions

654 We have shown that uncertainties in the value of diapycnal mixing in the pelagic ocean

655 contribute to the spread in future model projections of CO_2 and climate in response to

656 anthropogenic carbon emissions. Models with low mixing lead to slower uptake of carbon and

657 heat by the ocean, therefore contributing to higher atmospheric CO_2 and warmer air

658 temperatures. These results suggest that models with large ocean vertical mixing (high K_v)

659 systematically underestimate future warming and CO_2 concentrations, thereby possibly

660 significantly contributing to the large range in transient climate sensitivity and climate-carbon

661 cycle feedbacks diagnosed in earlier model intercomparisons.

662

663 Globally averaged metrics such as historic changes in globally averaged surface air temperature

664 or ocean heat content do not provide strong constraints on K_v [*Tomassini et al.*, 2007]. We show

665 that spatially resolved physical, geochemical and biogeochemical tracer observations in the

666 ocean can be used to reduce the uncertainty of this parameter (and, by extension, that of future

667 climate projections). These observations provide a firm upper limit on the value of K_v, whereas

668 the lower limit is less well constrained. Our best estimate for the background diapycnal

669 diffusivity in the pelagic ocean is 0.05-0.2 cm^2/s, in agreement with independent estimates based

670 on dye dispersion experiments and microstructure turbulence measurements [*Ledwell et al.*,

671 1993; *Toole et al.*, 1994].

672

673 We have developed a Bayesian model-data fusion method that can be used to quantify and

674 reduce the uncertainty in future climate-carbon cycle projections. Remaining issues that need

675 more work are cross-tracer correlations, parameter interactions and the optimal degree of spatial

676 aggregation. Resolution of the second issue is simply one of computational resources, while the

677 first needs further development and refinement of the existing statistical methodology and

678 theory. To resolve the third issue, the optimal degree of aggregation can presumably be

679 determined in a sensitivity study with successively larger spatial scales of averaging. None of

680 those issues seems insurmountable. The prospects of robust likelihood-based model assessment,

681 using multiple observations considering spatial and temporal autocorrelation as well as cross-

682 tracer correlations has the potential to lead towards truly probabilistic projections of climate and

683 the carbon cycle.

684

694

695

References

695

696

697 Antonov, J., R. A. Locarnini, T. Boyer, A. V. Mishonov, and H. E. Garcia (2006), World Ocean

698 Atlas 2005. Volume 2: Salinity, 182 pp, U.S. Government Printing Office, Washington, D.C.

699 Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates

700 in regional and global observed temperature changes: A new data set from 1850, *J. Geophys.*

701 *Res., 111*(D12), D12106, doi: 10.1029/2005JD006548.

702 Bryan, F. (1987), Parameter Sensitivity of Primitive Equation Ocean General Circulation

703 Models, *J. Phys. Oceanogr., 17*(7), 970-985.

704 Canadell, J. G., et al. (2007), Contributions to accelerating atmospheric CO2 growth from

705 economic activity, carbon intensity, and efficiency of natural sinks, *Proceedings of the National*

706 *Academy of Sciences of the United States of America*, *104*(47), 18866-18870.

707 Canadell, J. G., P. Ciais, T. J. Conway, C. B. Field, C. Le Quere, R. A. Houghton, G. Marland,

708 and M. R. Raupach (2008), Carbon budget and trends 2007, Global Carbon Project,

709 www.globalcarbonproject.org, 26 September 2008.

710 Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell (2000), Acceleration of

711 global warming due to carbon-cycle feedbacks in a coupled climate model, *Nature*, *408*(6809),

712 184-187.

713 Crowley, T. J. (2000), Causes of climate change over the past 1000 years, *Science*, *289*(5477),

714 270-277.

715 Domingues, C. M., J. A. Church, N. J. White, P. J. Gleckler, S. E. Wijffels, P. M. Barker, and J.

716 R. Dunn (2008), Improved estimates of upper-ocean warming and multi-decadal sea-level rise,

717 *Nature*, *453*(7198), 1090-1093.

718 Doney, S. C., et al. (2004), Evaluating global ocean carbon models: The importance of realistic

719 physics, *Glob. Biogeochem. Cycles, 18*(3), GB3017, doi: 10.1029/2003GB002150.

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

723   England, M. H., and E. Maier-Reimer (2001), Using chemical tracers to assess ocean models,
724   *Reviews in Geophysics*, *39*, 29-70.

725   Friedlingstein, P., J. L. Dufresne, P. M. Cox, and P. Rayner (2003), How positive is the feedback
726   between climate change and the carbon cycle?, *Tellus B*, *55*(2), 692-700.

727   Friedlingstein, P., et al. (2006), Climate-carbon cycle feedback analysis: Results from the C4MIP
728   model intercomparison, *J. Clim.*, *19*(14), 3337-3353.

729   Garcia, H. E., R. A. Locarnini, T. P. Boyer, and J. Antonov (2006a), World Ocean Atlas 2005.
730   Volume 4: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation, 342 pp,
731   U.S. Government Printing Office, Washington, D.C.

732   Garcia, H. E., R. A. Locarnini, T. P. Boyer, and J. Antonov (2006b), World Ocean Atlas 2005.
733   Volume 4: Nutrients (phosphate, nitrate, silicate), 396 pp, U.S. Government Printing Office,
734   Washington, D.C.

735   Gent, P. R., and J. C. McWilliams (1990), Isopycnal mixing in ocean circulation models, *J. Phys.*
736   *Oceanogr.*, *20*, 150-155.

737   Gnanadesikan, A., J. P. Dunne, R. M. Key, K. Matsumoto, J. L. Sarmiento, R. D. Slater, and P.
738   S. Swathi (2004), Oceanic ventilation and biogeochemical cycling: Understanding the physical
739   mechanisms that produce realistic distributions of tracers and productivity, *Glob. Biogeochem.*
740   *Cycles*, *18*(4), Gb4010.

741   Govindasamy, B., S. Thompson, A. Mirin, M. Wickett, K. Caldeira, and C. Delire (2005),
742   Increase of carbon cycle feedback with climate sensitivity: results from a coupled climate and
743   carbon cycle model, *Tellus B*, *57*(2), 153-163.

744 Jones, C. D., P. M. Cox, R. L. H. Essery, D. L. Roberts, and M. J. Woodage (2003), Strong

745 carbon cycle feedbacks in a climate model with interactive CO2 and sulphate aerosols, *Geophys.*

746 *Res. Lett., 30*(9), 1479, doi: 10.1029/2003GL016867.

747 Joos, F., G. K. Plattner, T. F. Stocker, O. Marchal, and A. Schmittner (1999), Global warming

748 and marine carbon cycle feedbacks on future atmospheric CO2, *Science*, *284*(5413), 464-467.

749 Joos, F., I. C. Prentice, S. Sitch, R. Meyer, G. Hooss, G. K. Plattner, S. Gerber, and K.

750 Hasselmann (2001), Global warming feedbacks on terrestrial carbon uptake under the

751 Intergovernmental Panel on Climate Change (IPCC) emission scenarios, *Glob. Biogeochem.*

752 *Cycles*, *15*(4), 891-907.

753 Keeling, C. D., and T. P. Whorf (2005), Atmospheric CO2 records from sites in the SIO air

754 sampling network., Carbon Dioxide Information Analysis Center, Oak Ridge National

755 Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.

756 Key, R. M., et al. (2004), A global ocean carbon climatology: Results from Global Data Analysis

757 Project (GLODAP), *Glob. Biogeochem. Cycles, 18*(4), GB4031, doi: 10.1029/2004GB002247.

758 Le Quere, C., et al. (2007), Saturation of the Southern Ocean CO2 sink due to recent climate

759 change, *Science*, *316*(5832), 1735-1738.

760 Ledwell, J. R., A. J. Watson, and C. S. Law (1993), Evidence for Slow Mixing across the

761 Pycnocline from an Open-Ocean Tracer-Release Experiment, *Nature*, *364*(6439), 701-703.

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

763 Locarnini, R. A., A. V. Mishonov, J. Antonov, T. Boyer, and H. E. Garcia (2006), World Ocean

764 Atlas 2005. Volume 1: Temperature, 182 pp, U.S. Government Printing Office, Washington,

765 D.C.

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

767 *2*, 49-55.

768   Matear, R. J., and A. C. Hirst (1999), Climate change feedback on the future oceanic CO2

769   uptake, *Tellus B*, *51*(3), 722-733.

770   Matsumoto, K., et al. (2004), Evaluation of ocean carbon cycle models with data-based metrics,

771   *Geophys. Res. Lett., 31*(7), L07303, doi: 10.1029/2003GL018970.

772   Matthews, H. D., M. Eby, A. J. Weaver, and B. J. Hawkins (2005a), Primary productivity control

773   of simulated carbon cycle-climate feedbacks, *Geophys. Res. Lett., 32*(14), L14708, doi:

774   10.1029/2005GL022941.

775   Matthews, H. D., A. J. Weaver, and K. J. Meissner (2005b), Terrestrial carbon cycle dynamics

776   under recent and future climate change, *J. Clim.*, *18*(10), 1609-1628.

777   Meissner, K. J., A. J. Weaver, H. D. Matthews, and P. M. Cox (2003), The role of land surface

778   dynamics in glacial inception: a study with the UVic Earth System Model, *Clim. Dyn.*, *21*(7-8),

779   515-537.

780   Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, and M. Collins (2004),

781   Quantification of modelling uncertainties in a large ensemble of climate change simulations,

782   *Nature*, *430*(7001), 768-772.

783   Naveira Garabato, A. C., K. L. Polzin, B. A. King, K. J. Heywood, and M. Visbeck (2004),

784   Widespread intense turbulent mixing in the Southern Ocean, *Science*, *303*, 210-213.

785   Neftel, A., H. Friedli, E. Moor, H. Lötscher, H. Oeschger, U. Siegenthaler, and B. Stauffer

786   (1994), Historical CO2 record from the Siple Station ice core, Carbon Dioxide Information

787   Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn.,

788   U.S.A..

789   Ricciuto, D. M., K. J. Davis, and K. Keller (2008), A Bayesian calibration of a simple carbon

790   cycle model: The role of observations in estimating and reducing uncertainty, *Glob. Biogeochem.*

791   *Cycles, 22*(2), GB2030, doi: 10.1029/2006GB002908.

792   Sarmiento, J. L., T. M. C. Hughes, R. J. Stouffer, and S. Manabe (1998), Simulated response of

793   the ocean carbon cycle to anthropogenic climate warming, *Nature*, *393*(6682), 245-249.

794   Schartau, M., and A. Oschlies (2003), Simultaneous data-based optimization of a 1D-ecosystem
795   model at three locations in the North Atlantic: Part I - Method and parameter estimates, *Journal*
796   *of Marine Research*, *61*(6), 765-793.

797   Schmittner, A., M. Latif, and B. Schneider (2005a), Model projections of the North Atlantic
798   thermohaline circulation for the 21st century assessed by observations, *Geophys. Res. Lett.,*
799   *32*(23), L23710, doi: 10.1029/2005GL024368.

800   Schmittner, A., A. Oschlies, X. Giraud, M. Eby, and H. L. Simmons (2005b), A global model of
801   the marine ecosystem for long-term simulations: Sensitivity to ocean mixing, buoyancy forcing,
802   particle sinking, and dissolved organic matter cycling, *Glob. Biogeochem. Cycles, 19*(3),
803   GB3004, doi: 10.1029/2004GB002283.

804   Schmittner, A., A. Oschlies, H. D. Matthews, and E. D. Galbraith (2008), Future changes in
805   climate, ocean circulation, ecosystems and biogeochemical cycling simulated for a business-as-
806   usual CO2 emission scenario until year 4000 AD, *Glob. Biogeochem. Cycles, 22*, GB1013,
807   doi:10.1029/2007GB002953.

808   Simmons, H. L., S. R. Jayne, L. C. St Laurent, and A. J. Weaver (2004), Tidally driven mixing in
809   a numerical model of the ocean general circulation, *Ocean Modelling*, *6*(3-4), 245-263.

810   Stainforth, D. A., et al. (2005), Uncertainty in predictions of the climate response to rising levels
811   of greenhouse gases, *Nature*, *433*(7024), 403-406.

812   Storn, R., and K. Price (1997), Differential evolution - a simple and efficient heuristic for global
813   optimization over continuous spaces, *Journal of Global Optimization*, *11*, 341-359.

814   Tomassini, L., P. Reichert, R. Knutti, T. F. Stocker, and M. E. Borsuk (2007), Robust Bayesian
815   uncertainty analysis of climate system properties using Markov chain Monte Carlo methods, *J.*
816   *Clim.*, *20*(7), 1239-1254.

817   Toole, J. M., K. L. Polzin, and R. W. Schmitt (1994), Estimates of Diapycnal Mixing in the
818   Abyssal Ocean, *Science*, *264*(5162), 1120-1123.

819   Weaver, A. J., et al. (2001), The UVic Earth System Climate Model: Model description,

820   climatology, and applications to past, present and future climates, *Atmos.-Ocean*, *39*(4), 361-428.

821   Zellner, A., and G. C. Tiao (1964), Bayesian Analysis of the Regression Model With

822   Autocorrelated Errors, *J. Am. Stat. Ass.*, *59*(307), 763-778.

823   Zeng, N., H. F. Qian, E. Munoz, and R. Iacono (2004), How strong is carbon cycle-climate

824   feedback under global warming?, *Geophys. Res. Lett., 31*(20), L20203, doi:

825   10.1029/2004GL020904.

826

827
828

Figure captions

829

830 **Figure 1.** Hindcasts and projections of atmospheric CO_2 concentration (top) and near surface air

831 temperature (SAT) anomalies from the 1960-1990 levels (bottom) for model versions with

832 different values of K_v. The emission scenario (SRES A2 until year 2100 and linear decrease until

833 year 2300 afterwards) is shown as the heavy dotted line in the top panel with the scale in the

834 bottom right corner ranging from 0-30 Gt C / yr. For reference: current (2007) levels are about

835 8.5 Gt C / yr [*Canadell et al.*, 2008]. The insets in the upper left region of each panel show a

836 zoom into the hindcast period (1800-2007) including CO_2 observations from Mauna Loa

837 [Keeling and Whorf, 2005] and ice cores [Neftel et al., 1994] (circles) and temperature

838 observations from the HadCRUT3 [Brohan et al., 2006] dataset (black wiggly line).

839

840 **Figure 2.** Upper ocean (0-700 m) heat content changes (from year 1961) as simulated by the

841 different model versions (lines) compared to observations (grey shading) from *Domingues et al.*

842 [2008]. The dark grey shading denotes nine year running mean values for comparison with

843 decadal averages plotted for the model simulations. The light gray shading shows three-year

844 averages.

845

846 **Figure 3.** Effects of reduced ocean heat and carbon uptake on projected warming resulting from

847 smaller vertical mixing. Solid line shows the difference in global mean surface air temperature

848 anomaly for the run with low vertical mixing (K_v=0.1) minus the run with high vertical mixing

849 (K_v=0.5) including the effects of both reduced heat and carbon uptake. Dashed line corresponds

850 to a run with K_v=0.1 but with atmospheric CO_2 prescribed from the simulation with K_v=0.5, thus

851 isolating the effect of reduced heat uptake alone. The effect of reduced carbon uptake is the

852 difference between the dashed and solid line.

853

854 **Figure 4.** Normalized RMS errors E (left), E' (center) and correlation coefficients r (right) for

855 3D distributions of different physical (top) and biogeochemical (bottom) tracers as a function of

856 the diapycnal background diffusivity K_v.

857

858   **Figure 5.** Posterior PDFs using the 3D method (eq. 3) for different physical (top) and

859   biogeochemical (bottom) tracers as a function of the diapycnal background diffusivity K_v.

860

861   **Figure 6.** Posterior PDFs using the 1D method (eq. 7) for different physical (top) and

862   biogeochemical (bottom) tracers as a function of the diapycnal background diffusivity K_v.

863

864   **Figure 7.** Sensitivity tests. Posterior PDFs as a function of the diapycnal background diffusivity

865   K_v for different tracers using the 3D method, but different assumptions in the statistical analysis

866   as described in the text. Note that for many tracers the black lines are indistinguishable from and

867   covered by the blue lines.

868

869   **Figure 8.** Posterior PDFs as a function of the diapycnal background diffusivity K_v for different

870   combinations of uncorrelated tracer distributions using the 3D method.

871

872

872

873

Table 1. Estimated model error σ_M for different assumptions and tracers.

		T	*S*	*Δ^{14}C*	*CFC11*	*PO$_4$*	*AOU*	*DIC*	*ALK*	*P**
		(K)		(permil)	(pM)	(μM)	(mM)	(μM)	(μM)	(μM)
3D	$\sigma_O{\neq}0$	0.90	0.184	0	0.240	0.202	24.8	19	13.0	0.16
	$\sigma_O{=}0$	0.92	0.190	20.2	0.353	0.203	24.8	25	15.2	0.16
1D	$\sigma_O{=}0$	0.24	0.079	5.85	0.05	0.064	3.7	9	6.6	0.06

875

876

Table 2. Cross-tracer error correlation for the 3D method in the model with $K_v{=}0.15$ cm^2/s. Absolute values larger than 0.3 are shown in bold.

	T	*S*	*Δ^{14}C*	*CFC11*	*PO$_4$*	*AOU*	*DIC*	*ALK*
T								
S	**0.48**							
Δ^{14}C	0.26	0.03						
CFC11	0.01	-0.04	**0.39**					
PO$_4$	**-0.42**	-0.22	**-0.41**	-0.23				
AOU	-0.09	0.12	**-0.36**	**-0.40**	**0.65**			
DIC	-0.20	**0.31**	**0.33**	-0.13	**0.52**	**0.76**		
ALK	0.04	**0.54**	0.23	-0.07	0.12	**0.42**	**0.72**	
P*	**-0.44**	**-0.42**	-0.19	0.08	**0.69**	-0.08	-0.05	-0.23

879

880

Figures

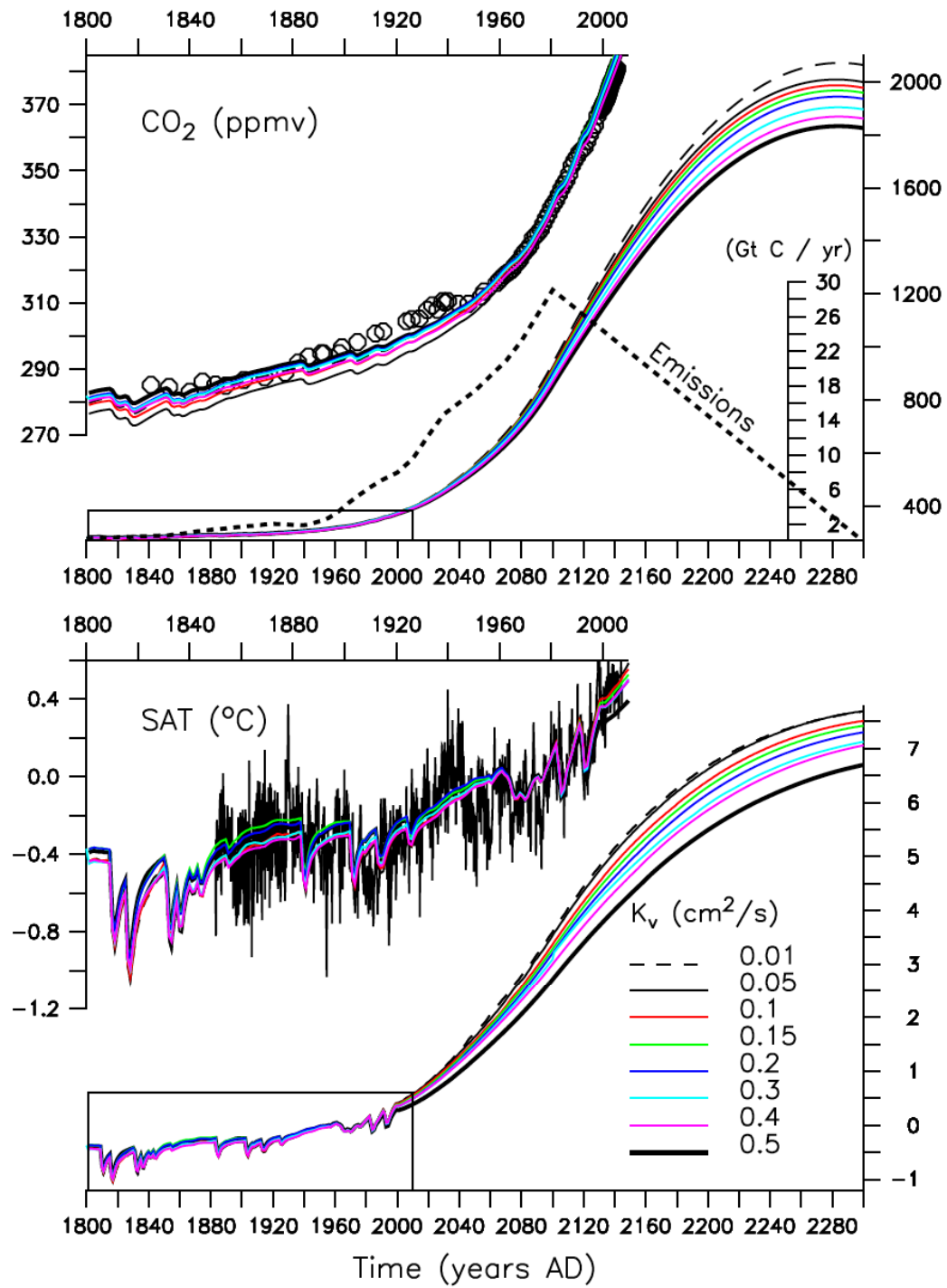**Figure 1**

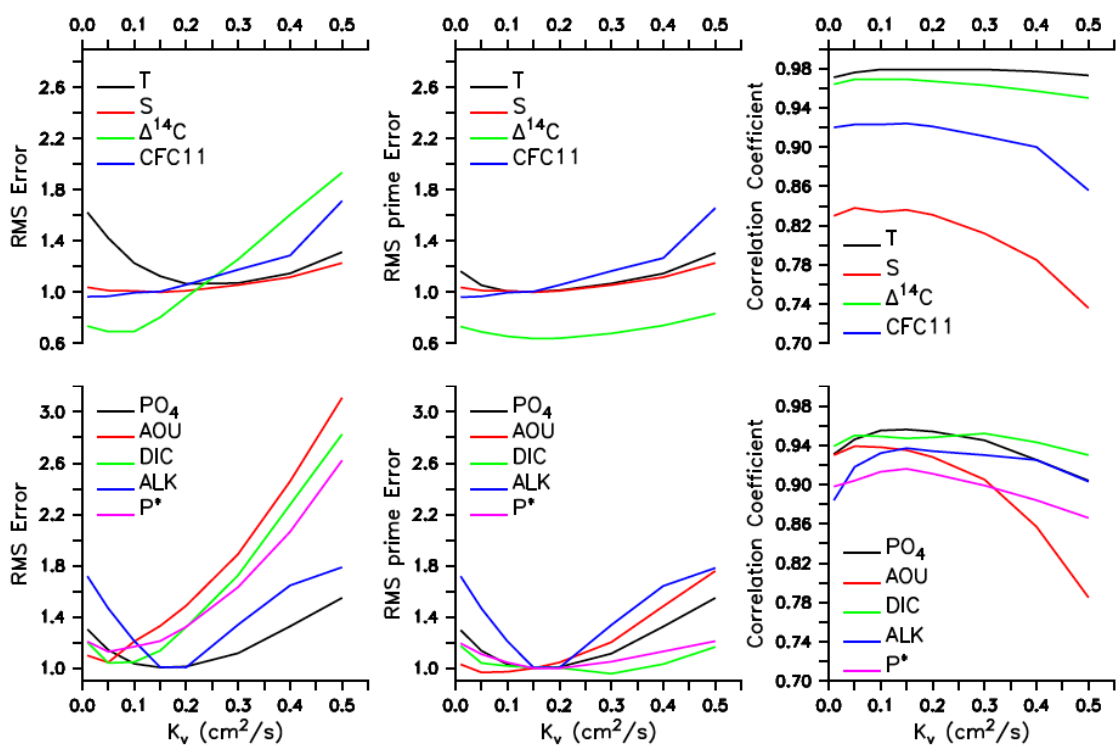Ocean heat content changes, 0−700 m (10^{22} J)
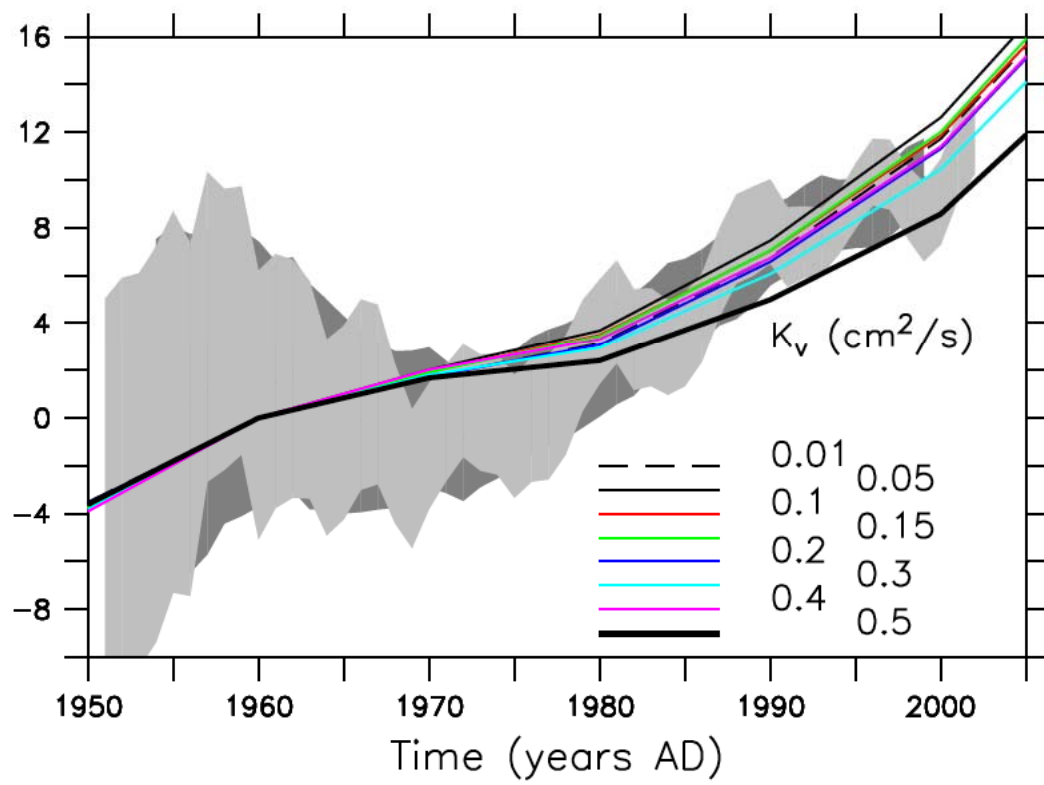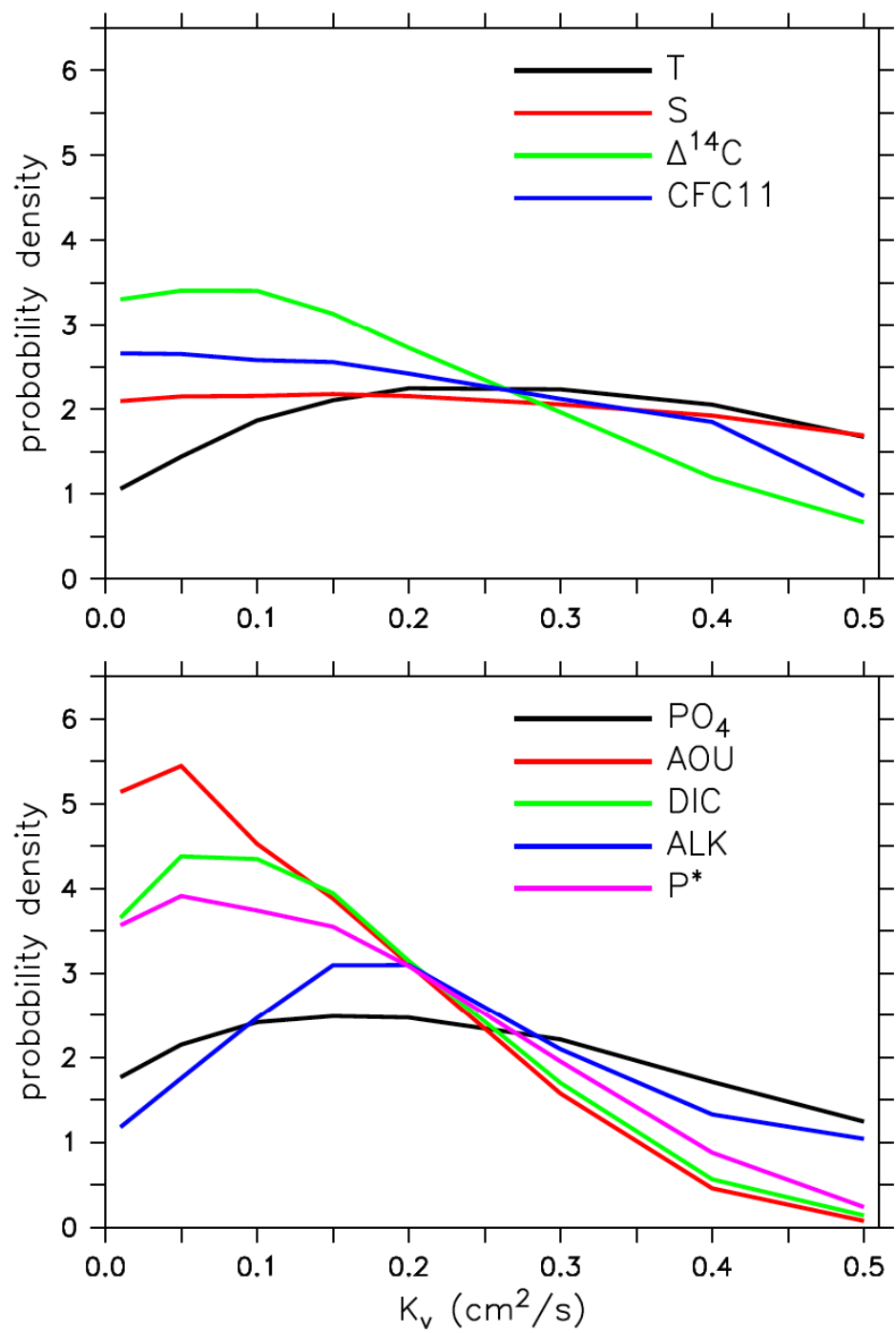
Figure 2
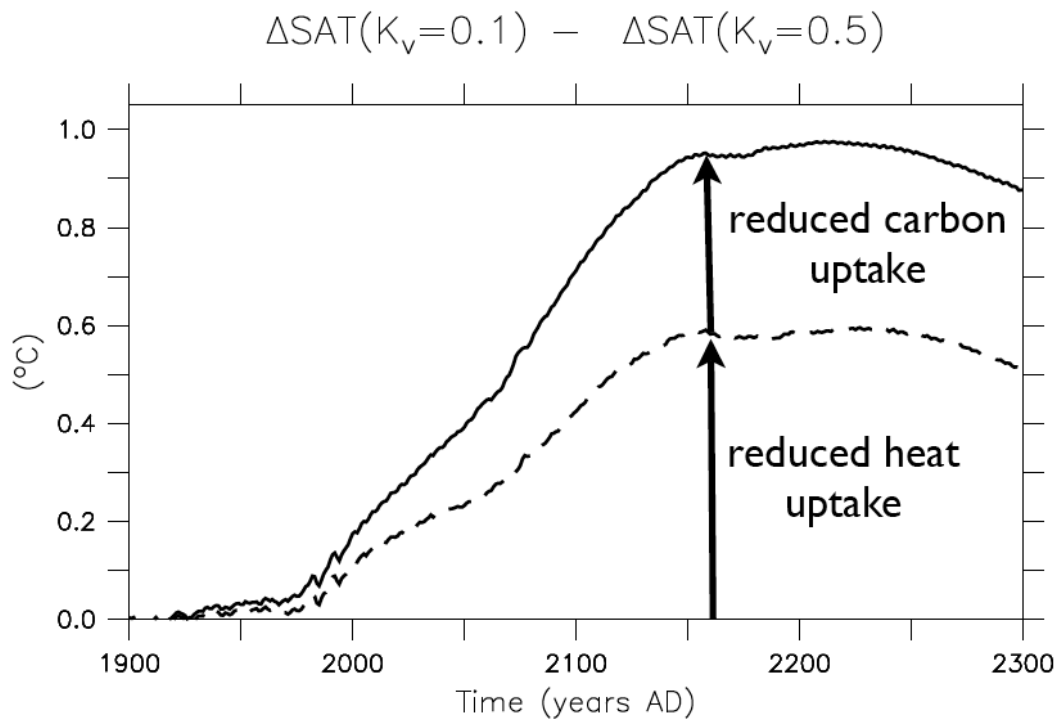
Figure 3

Figure 4

Figure 5
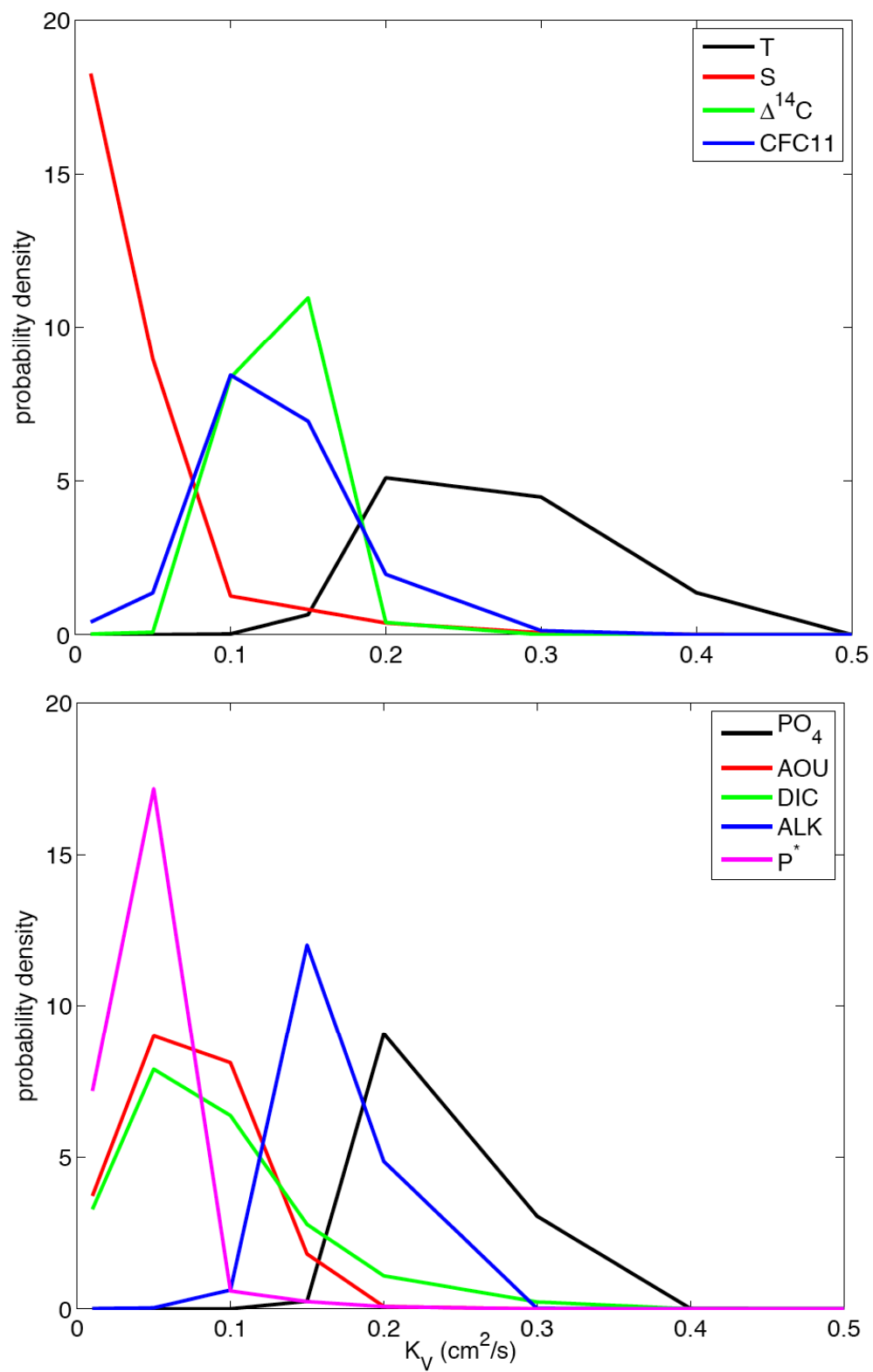
Figure 6

Figure 7

Figure 8