

Evaluation of ocean carbon cycle models with data-based metrics

K. Matsumoto,¹ J. L. Sarmiento,² R. M. Key,² O. Aumont,³ J. L. Bullister,⁴ K. Caldeira,⁵ J.-M. Campin,⁶ S. C. Doney,⁷ H. Drange,⁸ J.-C. Dutay,³ M. Follows,⁹ Y. Gao,⁸ A. Gnanadesikan,¹⁰ N. Gruber,¹¹ A. Ishida,¹² F. Joos,¹³ K. Lindsay,¹⁴ E. Maier-Reimer,¹⁵ J. C. Marshall,⁹ R. J. Matear,¹⁶ P. Monfray,³ A. Mouchet,⁶ R. Najjar,¹⁷ G.-K. Plattner,¹¹ R. Schlitzer,¹⁸ R. Slater,² P. S. Swathi,¹⁹ I. J. Totterdell,²⁰ M.-F. Weirig,¹⁸ Y. Yamanaka,¹² A. Yool,²⁰ and J. C. Orr³

Received 25 December 2003; accepted 5 March 2004; published 2 April 2004.

[1] New radiocarbon and chlorofluorocarbon-11 data from the World Ocean Circulation Experiment are used to assess a suite of 19 ocean carbon cycle models. We use the distributions and inventories of these tracers as quantitative metrics of model skill and find that only about a quarter of the suite is consistent with the new data-based metrics. This should serve as a warning bell to the larger community that not all is well with current generation of ocean carbon cycle models. At the same time, this highlights the danger in simply using the available models to represent the state-of-the-art modeling without considering the credibility of each model. **INDEX TERMS:** 4255 Oceanography: General: Numerical modeling; 4805 Oceanography: Biological and Chemical: Biogeochemical cycles (1615); 4842 Oceanography: Biological and Chemical: Modeling; 4599 Oceanography: Physical: General or miscellaneous. **Citation:** Matsumoto, K., et al. (2004), Evaluation of ocean carbon cycle models with

data-based metrics, *Geophys. Res. Lett.*, 31, L07303, doi:10.1029/2003GL018970.

1. Introduction

[2] Three-dimensional ocean general circulation models (OGCMs) of the carbon cycle are the primary tools to characterize the ocean's response to increasing atmospheric CO₂ concentration. Although individual models are evaluated to various extents by their developers, how well they perform relative to other models has been difficult to assess. A primary reason is the absence of quantitative, data-based metrics of model skill. Further complications arise because models have different grids, numerics, boundary conditions, and parameterizations and because of the lack, until recently, of uniform simulation protocols. This difficulty represents a major stumbling block for the ocean carbon cycle modeling community. Outside the community, this in part causes the Intergovernmental Panel of Climate Change (IPCC), for example, to oversimplify the representation of the community with just the mean and spread of models that happened to be available at the time [Allen, 2003].

[3] We have come together under the Ocean Carbon Cycle Intercomparison Project (OCMIP) and formulated a set of standard protocols and global-scale data sets for model intercomparison and model evaluation of tracers and anthropogenic carbon [Dutay et al., 2002]. In this work, we have developed new quantitative metrics, based on the new CFC-11 and radiocarbon (¹⁴C) measurements from the World Ocean Circulation Experiment [Key et al., 1996, 2002; WOCE Data Products Committee, 2002], to assess the skill of a suite of models that have used the simulation protocols of the second phase of OCMIP (OCMIP-2). Our assessment brings to light large data-model discrepancies in ocean ventilation on decadal and centennial time scales. Our major goal is to provide some guidelines as to what constitutes a reasonable model behavior and to demonstrate that many of the models available today do not behave well.

2. Models and Data

[4] The suite of 19 models evaluated here includes those from the thirteen participating modeling groups of OCMIP-2 and a family of Princeton models (Table 1 and auxiliary material¹). These models have dissimilar physical

¹Geological Survey of Japan, Agency of Industrial Science and Technology, Tsukuba, Japan.

²Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, New Jersey, USA.

³Laboratoire des Sciences du Climat et de l'Environnement, Unite Mixte de Recherche CEA-CNRS, Gif-sur-Yvette, France.

⁴Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, Seattle, Washington, USA.

⁵Lawrence Livermore National Laboratory, Livermore, California, USA.

⁶University of Liege, Liege, Belgium.

⁷Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA.

⁸Nansen Environmental and Remote Sensing Center, Bergen, Norway.

⁹Program in Atmospheres, Oceans and Climate, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

¹⁰NOAA Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA.

¹¹Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California, USA.

¹²Frontier Research System for Global Change, Yokohama, Japan.

¹³Physics Institute, University of Bern, Bern, Switzerland.

¹⁴National Center for Atmospheric Research, Boulder, Colorado, USA.

¹⁵Max Planck Institut für Meteorologie, Hamburg, Germany.

¹⁶CSIRO Division of Marine Research, Hobart, Tasmania, Australia.

¹⁷Department of Meteorology, Pennsylvania State University, University Park, Pennsylvania, USA.

¹⁸Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany.

¹⁹CSIR Centre for Mathematical Modeling and Computer Simulation, Bangalore, India.

²⁰Southampton Oceanography Centre, Southampton, UK.

Table 1. OCMIP-2 and Princeton Sensitivity Models^a

Model Name	Institutions (Country)
<i>OCMIP-2 Models</i>	
1) AWI	Alfred Wegener Inst. Polar Marine Res. (Germany)
2) CSIRO	CSIRO Division of Marine Research (Australia)
3) CCSR	Frontier Research for Global Change (Japan)
4) IPSL	Institute Pierre Simon LaPlace (France)
5) LLNL	Lawrence Livermore National Laboratory (USA)
6) MIT	Massachusetts Institute of Technology (USA)
7) MPIM	Max Planck Inst. Meteorologie-Hamburg (Germany)
8) NCAR	National Center for Atmospheric Research (USA)
9) NERSC	Nansen Env. and Remote Sensing Center (Norway)
10) PIUB	Physics Institute, University of Bern (Switzerland)
11) PRINCE	Princeton University (Low Kv; Low AI) (USA)
12) PRINC2	Princeton University (Intermediate Kv) (USA)
13) SOC	Southampton Oceanography Centre (UK)
14) UL	University of Liege (Belgium)
<i>Princeton Sensitivity Models (USA)</i>	
15) LH	Low Kv; High AI
16) HL	High Kv; Low AI
17) HH	High Kv; High AI
18) HiS	High Kv in Southern Ocean; otherwise PRINCE
19) SW	ECMWF reanalysis winds

^aFor most OCMIP-2 modeling groups the phase-2 models are not their latest. The exceptions are the Princeton models, some of which were developed more recently as part of in-house sensitivity studies [Gnanadesikan *et al.*, 2002]. The “IPSL” model presented here is the degraded version (IPSL.DM1 GM model). Kv and AI represent vertical and along-isopycnal mixing coefficients respectively.

models [Dutay *et al.*, 2002; S. C. Doney *et al.*, preprint, 2003] but use the same OCMIP-2 biogeochemistry formulations of air-sea gas exchange and seawater CO₂ system. Thus differences in the simulated tracer distributions can be attributed to different physical model behavior.

[5] For our data-model comparisons, we focus on two passive tracers: CFC-11, which is measured directly, and natural ¹⁴C, which is relatively straightforward to calculate from data. There is little ambiguity associated with either tracer when interpreting data-model discrepancies. In addition, these two tracers provide information on the ocean circulation that brackets the decadal to centennial timescales relevant for the uptake of anthropogenic CO₂. The need to separate natural ¹⁴C, which is produced only in the atmosphere, from the measured abundance [Broecker *et al.*, 1995; Rubin and Key, 2002] arises because oceanic ¹⁴C has been significantly altered in the upper ocean over the last decades with the introduction of thermonuclear bomb-produced ¹⁴C from the atmosphere. Over most of the ocean, the penetration of the bomb ¹⁴C is shallow and does not extend beyond several hundred meters. Therefore, with the exception of the North Atlantic where North Atlantic Deep Water (NADW) is formed, and locations around Antarctica where Antarctic Bottom Water is formed, the deep ocean can be assumed to be free of bomb ¹⁴C, eliminating most of the uncertainties associated with the separation.

3. Results and Discussion

[6] The deep ocean ventilation is readily assessed by the distribution of natural ¹⁴C, whose abundance is a direct measure of ocean ventilation on centennial time scales. We make data-model comparison of the deep natural ¹⁴C distribution by examining the natural ¹⁴C abundance of NADW, the North Pacific Deep Water (NPDW), and the

Circumpolar Deep Water (CDW). The three water masses are important “end-members” that can characterize the deep ocean. Their typical natural ¹⁴C abundance, expressed with $\Delta^{14}\text{C}$ in per mil (‰), as determined from measurements, are -70‰ , -160‰ , and -230‰ respectively for NADW, CDW, and NPDW (Figure 1; see auxiliary material for water mass definition). The oldest NPDW has the most negative and least $\Delta^{14}\text{C}$, and the most recently ventilated NADW has the highest ¹⁴C abundance. Figure 2 shows that simulated natural $\Delta^{14}\text{C}$ varies substantially [Orr, 2002], particularly in the North Pacific and Southern Ocean. In the North Pacific, $\Delta^{14}\text{C}$ varies between -138‰ and -380‰ from the most ventilated (OCMIP-2 IPSL) model to the most stagnant (Princeton LH), respectively. Many OCMIP-2 models ventilate the deep Pacific excessively. Interestingly, natural $\Delta^{14}\text{C}$ is correlated rather well between NPDW and CDW (diamonds, Figure 2) unlike that between NADW and CDW (triangles, Figure 2), suggesting the important role that CDW plays in these models in ventilating the deep Pacific.

[7] Ocean ventilation on shorter, decadal time scales is better characterized by the presence of CFC-11. Figure 3 shows the combined Indo-Pacific inventories of CFC-11 inside and outside the Southern Ocean relative to long term, deep Southern Ocean ventilation, as indicated by the CDW $\Delta^{14}\text{C}$. This figure allows us to evaluate models in their decadal and centennial time scale ventilation at the same time. We examine the Southern Ocean separately because of the large inter-model variability in CFC uptake [Dutay *et al.*, 2002] and in natural ¹⁴C abundance in this region (Figure 2).

[8] In the Southern Ocean, model-based inventories of CFC-11 are positively correlated with CDW $\Delta^{14}\text{C}$ (Figure 3c), indicating the importance of deep vertical transport (mixing and advection) in these models that carries both CFC-11 and natural ¹⁴C into the deep ocean. Model-simulated inventories encompass the observations but vary roughly by a factor of 4 between the most stratified models (i.e., Princeton LH and LL) and the most vertically well-mixed models (OCMIP-2 AWI, NERSC, UL). The large inter-model variability in the Southern Ocean [Orr *et al.*, 2001] contrasts with a much smaller variability outside the Southern Ocean (Figure 3a), suggesting how sensitive Southern Ocean tracer uptake is to the formulations of surface forcing and physical parameterizations in the current generation of OGCMs.

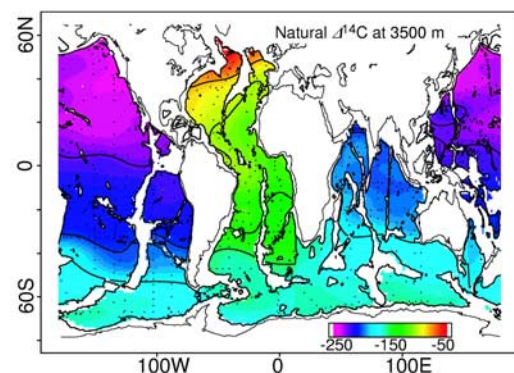


Figure 1. Data-based natural $\Delta^{14}\text{C}$ on 3500 m level [after Matsumoto and Key, 2004].

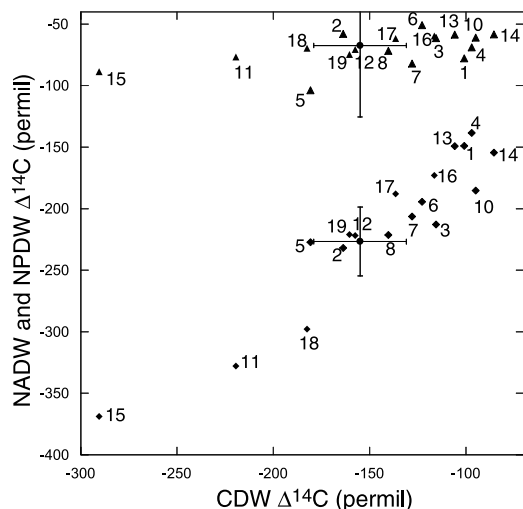


Figure 2. Observed and modeled deep natural $\Delta^{14}\text{C}$. Symbols with error bars (2σ) are observations. See auxiliary material for the definitions of North Atlantic Deep Water (NADW), North Pacific Deep Water (NPDW), and Circumpolar Deep Water (CDW). Diamonds and triangles are model-predicted mean NPDW and NADW $\Delta^{14}\text{C}$ respectively, vs. model predicted CDW. See Table 1 for model indices. The results of model #9 (NERSC) are excluded here, because the model had not been run out to equilibrium.

[9] We can then use the observed natural ^{14}C distribution and CFC-11 inventories (Figures 3a and 3c) as metrics of model skill to identify those models that are most consistent with the observations. For reasonable uncertainties of 2σ for mean CDW $\Delta^{14}\text{C}$ and 15% for CFC-11 inventories (see auxiliary material⁵), there are only 4 out of 19 models that agree with observations. For smaller uncertainties of 1σ and 10%, there is but one model. For larger uncertainties of 3σ and 20%, there are seven models. The majority of the models analyzed in this study are thus not ventilating the ocean on decadal and centennial time scales as indicated by the observed CFC-11 and natural radiocarbon.

[10] Presumably these smaller groupings of the “better” models make more reliable estimates of the global oceanic uptake of anthropogenic carbon than larger groupings that include models that are not as consistent with data. The uptake estimates from the subset of four models, which are essentially the same as those from the subsets of one and seven models, are 1.9 ± 0.2 and 2.2 ± 0.2 Pg-C/yr ($\text{Pg} = 10^{15}\text{g}$; the uncertainty of ± 0.2 represents half the range of the model spread) during the 1980s and 1990s respectively. We made a 3% downward correction [Orr, 2002] to the 1990s estimates, which are from simulations using the IPCC S650 scenario, whose 1990s atmospheric CO_2 concentrations were slightly higher than observed. The new estimates are somewhat lower and, as expected, have smaller spread than those from the full set of OCMIP-2 models for both the 1980s (2.0 ± 0.4 Pg-C/yr) and 1990s (2.4 ± 0.5 Pg-C/yr) [Orr, 2002]. For the 1980s, the model-based estimates are consistent with the most recent IPCC estimate of 1.9 ± 0.6 Pg-C/yr, based on atmospheric CO_2 and O_2 measurements [Houghton et al., 2001], and with 1.6 ± 0.4 Pg-C/yr, based on global CFC distribution [McNeil et al., 2003]. It is

difficult to directly translate the new model-based estimates into reality, because we do not account for a number of important processes, including interannual variability and climate-carbon feedbacks. Therefore, our new estimates should simply be regarded as improved model-based estimates without necessarily implying an improved understanding of the true global carbon budget.

[11] Finally, we draw attention to the comparison of the model-simulated anthropogenic CO_2 inventories with those reconstructed from ocean carbon observations using the ΔC^* method [Gruber et al., 1996]. In Figures 3b and 3d, we show only the Indian [Sabine et al., 1999] and Pacific [Sabine et al., 2002] basins, because the paucity of WOCE data in the far North Atlantic makes it difficult to determine the preformed properties of NADW (i.e., properties of seawater when it was last at the surface before it became part of NADW) and hence anthropogenic carbon. The comparison suggests that the models that successfully simulate the correct CDW $\Delta^{14}\text{C}$ and CFC-11 inventories tend to simulate higher anthropogenic CO_2 inventories than the data-based estimates. This is clearly evident north of 40°S , where simulated anthropogenic carbon inventories are on average 10 Pg-C higher than the data-based inventory of 47 ± 7 Pg-C. Our analysis shows that much of the discrepancy originates from the Pacific inventories. It may result from possible model deficiencies, such as misrepresentation of the oceanic buffering capacity or of the air-sea gas exchange, or possible biases in the data-based estimates.

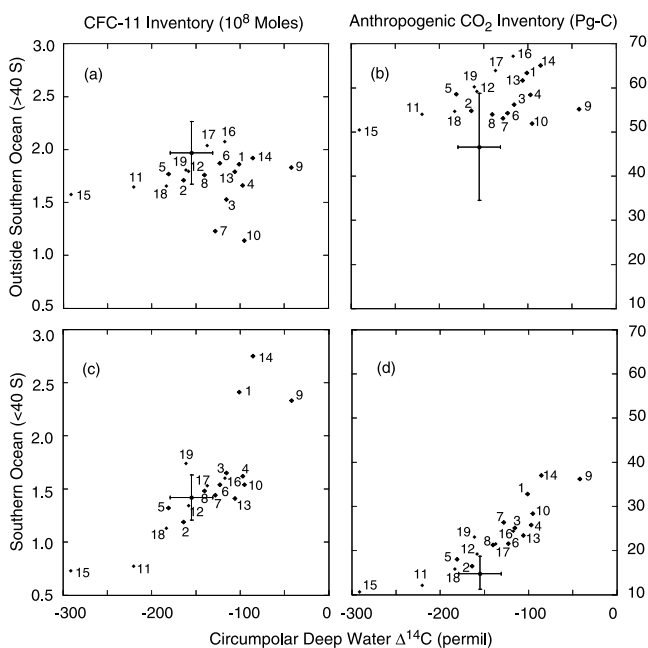


Figure 3. Observed and modeled CFC-11 and anthropogenic carbon inventories in the Indian and Pacific Ocean versus Southern Ocean natural $\Delta^{14}\text{C}$. Observed inventories are based on WOCE CFC-11 measurements [WOCE Data Products Committee, 2002] and anthropogenic carbon reconstructions [Sabine et al., 2002, 1999]. Error bars represent 2σ for natural ^{14}C , 15% uncertainty for CFC-11 inventories, and 25% uncertainty for anthropogenic carbon inventories (see auxiliary material). Southern Ocean is defined here as south of 40°S .

At the same time, our results highlight the importance of better understanding the fidelity of the data-based estimates of anthropogenic carbon, which have recently been suggested as being too low [Hall et al., 2004].

4. Conclusions

[12] We have developed quantitative, data-based metrics of ocean carbon cycle model skill and show that only about a quarter of the 19 models we examined are reasonably consistent with those metrics. While the simple metrics we introduced here are certainly not the best or the only ones, our study illustrates clearly that some models are inadequate. This recognition is timely for the larger community today, when ocean carbon cycle models are becoming more numerous, and relatively easy metrics to assess them are needed to ensure quality control. Identifying what in the models are responsible for this inadequacy will require sensitivity studies, where important model parameterizations and physics are changed one at a time.

[13] **Acknowledgments.** While at Princeton, K. Matsumoto was supported by NSF grants OCE-9819144 and OCE0097316.

References

- Allen, M. R. (2003), Possible or probable?, *Nature*, 425, 242.
- Broecker, W. S., S. Sutherland, W. Smethie, T.-H. Peng, and G. Ostlund (1995), Oceanic radiocarbon: Separation of the natural and bomb components, *Global Biogeochem. Cycles*, 9, 263–288.
- Dutay, J.-C., et al. (2002), Evaluation of ocean model ventilation with CFC-11: Comparison of 13 global ocean models, *Ocean Modell.*, 4, 89–120.
- Gnanadesikan, A., R. D. Slater, N. Gruber, and J. L. Sarmiento (2002), Oceanic vertical exchange and new production: A comparison between models and observations, *Deep Sea Res., Part II*, 49, 363–401.
- Gruber, N., J. L. Sarmiento, and T. F. Stocker (1996), An improved method for detecting anthropogenic CO₂ in the oceans, *Global Biogeochem. Cycles*, 10, 809–837.
- Hall, T. M., D. W. Waugh, T. W. N. Haine, P. E. Robbins, and S. Khatiwala (2004), Estimates of anthropogenic carbon in the Indian Ocean with allowance for mixing and time-varying air-sea CO₂ disequilibrium, *Global Biogeochem. Cycles*, 18, GB1031, doi:10.1029/2003GB002120.
- Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, and D. Xiaosu (2001), *Climate Change 2001—The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, 944 pp., Cambridge Univ. Press, New York.
- Key, R. M., P. Quay, G. A. Jones, R. J. Schneider, A. McNichol, and K. F. V. Reden (1996), WOCE AMS Radiocarbon I: Pacific Ocean results; P6, P16, & P17, *Radiocarbon*, 38, 425–518.
- Key, R. M., P. D. Quay, P. Schlosser, A. McNichol, K. F. von Reden, B. J. Schneider, K. L. Elder, M. Stuiver, and H. G. Ostlund (2002), WOCE Radiocarbon IV: Pacific Ocean results; P10, P13N, P14C, P18, P19, & S4P, *Radiocarbon*, 44, 239–392.
- Matsumoto, K., and R. Key (2004), Natural radiocarbon distribution in the deep ocean, in *Global Environmental Change in the Ocean and on Land*, edited by M. Shiomi and H. Kawahata, TERRAPUB, Tokyo, 45–58.
- McNeil, B. I., R. J. Matear, R. M. Key, J. L. Bullister, and J. L. Sarmiento (2003), Anthropogenic CO₂ uptake by the ocean based on the global chlorofluorocarbon data set, *Science*, 299, 235–239.
- Orr, J. C. (2002), *Global Ocean Storage of Anthropogenic Carbon*, 116 pp., Inst. Pierre Simon Laplace, Gif-sur-Yvette, France.
- Orr, J. C., E. Maier-Reimer, U. Mikolajewicz, P. Monfray, J. L. Sarmiento, J. R. Toggweiler, N. K. Taylor, J. Palmer, N. Gruber, C. L. Sabine, C. Le Quere, R. M. Key, and J. Boutin (2001), Estimates of anthropogenic carbon uptake from four three-dimensional global ocean models, *Global Biogeochem. Cycles*, 15, 43–60.
- Rubin, S., and R. M. Key (2002), Separating natural and bomb-produced radiocarbon in the ocean: The potential alkalinity method, *Global Biogeochem. Cycles*, 16(4), 1105, doi:10.1029/2001GB001432.
- Sabine, C. L., R. M. Key, K. M. Johnson, F. J. Millero, A. Poisson, J. L. Sarmiento, D. W. R. Wallace, and C. D. Winn (1999), Anthropogenic CO₂ inventory of the Indian Ocean, *Global Biogeochem. Cycles*, 13, 179–198.
- Sabine, C. L., R. A. Feeley, R. M. Key, J. L. Bullister, F. J. Millero, K. Lee, T.-H. Peng, B. Tilbrook, T. Ono, and C. S. Wong (2002), Distribution of anthropogenic CO₂ in the Pacific Ocean, *Global Biogeochem. Cycles*, 16(4), 1083, doi:10.1029/2001GB001639.
- WOCE Data Products Committee (2002), *WOCE Global Data, Version 3.0*, WOCE Int. Proj. Off., Southampton, UK.
- O. Aumont, J.-C. Dutay, P. Monfray, and J. C. Orr, Laboratoire des Sciences du Climat et de l'Environnement, F-91191 Gif-sur-Yvette, France.
- J. L. Bullister, PMEL/NOAA, 7600 Sand Point Way, NE, Seattle, WA 98115, USA.
- K. Caldeira, Lawrence Livermore National Laboratory, 7000 East, L-103, Livermore, CA 94550, USA.
- J.-M. Campin and A. Mouchet, Astrophysics and Geophysics Institute, University of Liege, Allée du 6 Août, 17 Bât. B5c, B-4000 Liege, Belgium.
- S. C. Doney, Woods Hole Oceanographic Institution, Bigelow 209b-MS 11, Woods Hole, MA 02543, USA.
- H. Drange and Y. Gao, Nansen Environmental and Remote Sensing Center, N-5059 Bergen, Norway.
- M. Follows and J. C. Marshall, Program in Atmospheres, Oceans and Climate, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.
- A. Gnanadesikan, GFDL/NOAA, P.O. Box 308, Princeton, NJ 08542-0308, USA.
- N. Gruber and G.-K. Plattner, Institute of Geophysics and Planetary Physics, UCLA, 5853 Slichter Hall, Los Angeles, CA 90095-4996, USA.
- A. Ishida and Y. Yamanaka, Frontier Research System for Global Change, 3173-25 Showamachi, Kanazawa-ku, Yokohama 236-0001, Japan.
- F. Joos, Physics Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland.
- R. M. Key, J. L. Sarmiento, and R. Slater, Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ 98544-0710, USA.
- K. Lindsay, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, USA.
- E. Maier-Reimer, Max Planck Institut für Meteorologie, Bundesstr. 55, D-20146 Hamburg, Germany.
- R. J. Matear, CSIRO Division of Marine Research, GPO Box 1538, Hobart, Tasmania 7001, Australia.
- K. Matsumoto, corresponding author, Geological Survey of Japan, AIST Site 7, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8567, Japan. (katsumi@ni.aist.go.jp)
- R. Najjar, Department of Meteorology, Pennsylvania State University, 522 Walker Building, University Park, PA 16802-5013, USA.
- R. Schlitzer and M.-F. Weirig, Alfred Wegener Institute for Polar and Marine Research, D-27568 Bremerhaven, Germany.
- P. S. Swathi, CSIR Centre for Mathematical Modeling and Computer Simulation, Nal Belvr Campus, Bangalore 560 037, India.
- I. J. Totterdell and A. Yool, Southampton Oceanography Centre, Empress Dock, Southampton SO14 3AZ, UK.