Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data

Scott C. Doney[a]*, Ivan Lima[a], J. Keith Moore[b], Keith Lindsay[c], M. Behrenfeld[d], Toby K. Westberry[d], Natalie Mahowald[e], David M. Glover[a], and Taro Takahashi[g]

a. Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, 266 Woods Hole Road, Woods Hole, MA 02543, USA (sdoney@whoi.edu ilima@whoi.edu, and dglover@whoi.edu)

b. Department of Earth System Science**,** University of California Irvine, Irvine CA 92697 USA (jkmoore@uci.edu)

c. Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO 80307 USA (klindsay@ucar.edu)

d. Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331 USA (behrenfm@science.oregonstate.edu and toby.westberry@science.oregonstate.edu)

e. Department of Earth and Atmospheric Sciences, Cornell University, Cornell NY 14850 USA (nmm63@cornell.edu)

f. Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964 USA (taka@ldeo.columbia.edu)

Submitted, Nov. 1st, 2007
Revised, March 4th, 2008
Submitted to *Journal of Marine Systems*
Special Issue "*Skill Assessment for Coupled Biological/Physical Models of Marine Systems*"
guest-edited by Dan Lynch, Dennis McGillicuddy and Cisco Werner.

* corresponding author

Abstract

We present a generalized framework for assessing the skill of global upper ocean ecosystem-biogeochemical models against *in-situ* field data and satellite observations. We illustrate the approach utilizing a multi-decade (1979-2004) hindcast experiment conducted with the Community Climate System Model (CCSM-3) ocean carbon model. The CCSM-3 ocean carbon model incorporates a multi-nutrient, multi-phytoplankton functional group ecosystem module coupled with a carbon, oxygen, nitrogen, phosphorus, silicon, and iron biogeochemistry module embedded in a global, three-dimensional ocean general circulation model. The model is forced with physical climate forcing from atmospheric reanalysis and satellite data products and time-varying atmospheric dust deposition. Data-based skill metrics are used to evaluate the simulated time-mean spatial patterns, seasonal cycle amplitude and phase, and subannual to interannual variability. Evaluation data include: sea surface temperature and mixed layer depth; satellite derived surface ocean chlorophyll, primary productivity, phytoplankton growth rate and carbon biomass; large-scale climatologies of surface nutrients, pCO_2, and air-sea CO_2 and O_2 flux; and time-series data from the Joint Global Ocean Flux Study (JGOFS). Where the data is sufficient, we construct quantitative skill metrics using: model-data residuals, time-space correlation, root mean square error, and Taylor diagrams.

Keywords: marine ecology, biogeochemistry, modeling, evaluation, skill

1. Introduction

The last two decades witnessed a dramatic increase in the volume of global ocean biogeochemical and ecological observations due, in part, to coordinated international field programs (e.g., Joint Global Ocean Flux Study JGOFS; World Ocean Circulation Experiment WOCE), satellite ocean color sensors, and emerging and ongoing ocean observing systems (e.g., *Fasham et al*., 2001; *Doney and Hood*, 2002; *McClain et al*., 2004). Data availability combined with increasing computational power stimulated a rapid growth in basin to global upper ocean ecosystem-biogeochemistry models (e.g., *Sarmiento et al*., 1993; *Six and Maier-Reimer*, 1996; *Oschlies and Garcon*, 1998; *Doney*,

1999; *Gregg et al*., 2003; *Aumont et al*., 2003; *Moore et al*., 2004; *Le Quéré et al.,* 2005; *Doney and Ducklow*, 2006). Such models are now widely applied to questions from seasonal and interannual climate variability (e.g., *Le Quéré et al*., 2000; *McKinley et al*., 2004; *Wetzel et al*., 2005; *Lovenduski et al*., 2007; *Le Quéré et al*., 2007) to anthropogenic climate change (e.g., *Bopp et al*., 2001; *Boyd and Doney*, 2002; *Bopp et al. et al.*, 2003). Marine ecosystem models are also growing in sophistication and complexity, incorporating multiple limiting nutrients and multiple planktonic functional groups at the lower trophic levels (*Denman*, 2003; *Hood et al*., 2006; *Follows et al.*, 2007).

Even with the wealth of new ocean data, the evaluation of basin and global-scale marine ecosystem-biogeochemistry models is challenging. Satellite data provide reasonably high space-time resolution but only for the surface layer and for only a handful of biological properties. Ship-based biological and chemical data are invaluable but comparatively sparse, with the construction of global annual-mean climatologies requiring the aggregation of data from multiple years. Routine underway sampling transects on research vessels and volunteer commercial ships are greatly improving data densities but are mostly limited, to this point, to carbon dioxide system variables (e.g., pCO_2). Quantitative evaluation is further confounded by the fact that model variables in most ecosystem models are highly aggregated, for example lumping all microzooplankton or mesozooplankton into two model compartments. Model variables can also lack direct analogues with observed values, as in the case of the crucial higher trophic level mortality closure terms that combine effects of mortality and predation. Furthermore, even the best sampled locations (e.g., JGOFS sites) often do not have enough data to constrain model dynamics fully (e.g., *Friedrichs et al*., 2007). Despite these issues, data-based verification of model skill is fundamental to advancing the science of marine ecological modeling. Quantitative skill assessment is integral to both the model development cycle and defining confidence estimates on model forecasts. As such, most research groups have in place some form of data-based assessment, though as discussed in *Stow et al*. (2007) these assessments are often partial and qualitative.

Skill assessment requires significant up-front investment to compile a wide range of field and remote sensing data into appropriate data products (often, but not always

gridded) prior to the actual comparison with model results. The problem is not one of simply data management, but rather requires a serious level of data interpretation and analysis effort to combine observations from different researchers, measurement techniques or even satellite platforms. Once created, the utility of such data products is clear, as illustrated by the broad use by the modeling community of data compilations such as the global surface pCO_2 data set and air-sea CO_2 flux estimates of *Takahashi et al.* (2002) and *Takahashi et al*. (submitted), the GLODAP data products created from the WOCE/JGOFS (*Key et al*., 2004), and the level 3 gridded satellite ocean color data from SeaWiFS and MODIS. Standard data products and quantitative metrics of model skill, even if imperfect, stimulate critical assessment of model performance and speed model development.

Lessons for developing a systematic model-data skill assessment can be drawn from the experience of similar efforts in related fields. For example, the ocean biogeochemical community has organized model-data skill assessments under the Ocean Carbon Model Intercomparison Project (OCMIP). Within OCMIP researchers compared about a dozen global ocean biogeochemical models against observations of ocean physics (*Doney et al*., 2004), transient tracers including radiocarbon (*Matsumoto et al*., 2004) and chlorofluorocarbons (*Dutay et al*., 2004), and inorganic carbon, nutrients and oxygen (*Orr et al*., 2005; *Najjar et al*., 2007). *Matsumoto et al*. (2004) argue that most the OCMIP simulations do not adequately match (within error bars) the available ocean transient tracer data. Estimates of less well-constrained model variables (e.g., future ocean uptake of anthropogenic CO_2) should therefore include only the subset of "skillful" simulations or should be constructed using a weighting function based on the transient tracer model skill. A second lesson from OCMIP is that model skill assessment is often best done as a partnership between researchers with expertise in modeling and researchers involved in field observations and remote sensing.

Here we argue for a similar generalized model-data framework for characterizing the skill of global ocean ecosystem-biogeochemical models against in-situ field data and satellite observations. While all of the major ocean ecosystem modeling groups currently assess model skill, the approaches are often specific to their particular model, and the community lacks a set of agreed upon, objective evaluation metrics that can be used to

inter-compare skill across models. Our goal here is to stimulate discussion and dialogue by proposing a prototype scheme that will be open to the community. We realize that a fully comprehensive system will only emerge over time with input from different user groups and that even with a generalized set of skill metrics there will still remain a need for unique verification approaches for different model applications.

As an illustration, we present model-data skill results from a multi-decade (1979-2004) hindcast experiment conducted with the Community Climate System Model (CCSM-3) coupled ocean Biogeochemical Elemental Cycling model (BEC). The BEC model consists of upper ocean ecological (*Moore et al*., 2004) and full-depth biogeochemical (*Doney et al*., 2006) modules embedded in a global 3-D Parallel Ocean Program (POP) ocean general circulation model (*Smith and Gent*, 2004; *Collins et al*., 2006). The model is forced with physical climate forcing from atmospheric reanalysis and satellite data products (*Doney et al*., 2003; 2007) and time-varying dust deposition (*Mahowald et al*., 2003). We focus the analysis on three aspects of the simulation: time-mean spatial patterns, the seasonal cycle, and subannual to interannual variability. Evaluation data include satellite derived surface ocean chlorophyll and primary productivity (SeaWiFS and MODIS), large-scale climatologies of surface nutrients and pCO_2, and time-series data from JGOFS and other field programs. Where the data is sufficient, we construct quantitative skill scores (time-space correlation and model and data rms variability) (*Lima and Doney*, 2004).

2. Hindcast Global Ocean Ecosystem-Biogeochemistry Simulation

2.1 Ecosystem-Biogeochemistry Module

The CCSM-3 BEC model is cast as a set of three-dimensional, time-varying advection diffusion equations for a suite of tracers C:

$$\frac{\partial C}{\partial t} + \nabla \cdot (\vec{u} C) - \nabla \cdot (K \nabla C) = RHS_{bio}^{C} \qquad (1)$$

The physical transport is partitioned into resolved advection and parameterized eddy mixing terms; all of the ecological-biogeochemical source/sink terms and surface and sediment fluxes are grouped into the right hand side term RHS_{bio}. The marine ecosystem module (Figure 1) builds on traditional phytoplankton-zooplankton-detritus-nutrient food

web models (e.g., *Fasham et al.*, 2000; *Doney et al.,* 1996). The module incorporates multi-nutrient limitation (N, P, Si, and Fe) on phytoplankton growth and specific phytoplankton functional groups (*Moore et al*., 2002a; 2004).

There are fourteen main model compartments: small pico/nano-plankton, diatoms, and diazotrophs; zooplankton; suspended and sinking particulate detritus; and dissolved nitrate, ammonia, phosphorus, iron, silicate, oxygen, dissolved inorganic carbon, and alkalinity. The pico/nanoplankton size class is designed to replicate the rapid and highly efficent nutrient recycling found in many subtropical, oligotrophic (low nutrient) environments. Diatoms model a larger, bloom-forming size class. Phytoplankton growth rates are determined by available light and nutrients using a modified form of the *Geider et al*. (1998) dynamic growth model. Photoadaptation is parameterized with dynamically adaptive chl/C ratios. The diazotrophs fix all required nitrogen from N_2 gas, and calcification is parameterized as a fraction of the pico/nanoplankton production as a function of temperature and nutrients adapted for coccolithophores. Size-structure effects are included by varying key zooplankton (e.g., partitioning of fecal pellets between suspended and sinking detritus) depending on the food source (*Lima and Doney*, 2004). Many of the biotic and detrital compartments contain multiple elemental pools, in addition to carbon, to track flows through the ecosystem. The model has one adaptive zooplankton class that grazes on phytoplankton and large detritus.

The biogeochemistry module (*Doney et al*., 2006) is based on an expanded version of the Ocean Carbon Model Intercomparison Project (OCMIP) biotic model (*Najjar et al*., 2007). The module includes full carbonate system thermodynamics and air-sea CO_2 and O_2 fluxes. Gas transfer velocities are computed from the 6-hourly NCEP winds (http://www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml) using the quadratic wind speed relationship of *Wanninkhof* (1992). A dynamic iron cycle is incorporated with atmospheric dust deposition, water-column scavenging and a continental sediment source (*Moore et al*., 2006; see also *Moore and Braucher, submitted* a for discussion on refinements of continental sediment source). Denitrification is simulated in oxygen minimum zones following *Moore and Doney* (2007), and subsurface particle remineralization is parameterized incorporating the mineral ballast arguments of *Armstrong et al.* (2002). The model equations are identical to those reported for the 3-D

implementation of *Moore et al.* (2004) with two important modifications as documented in more detail in *Moore et al.* (2006). First, water column denitrification has been added to the model in order to close the global nitrogen cycle. Second, a number of the parameters associated with the model iron dynamics and scavenging have been adjusted to improve the simulated dissolved iron fields (see Table 1 of *Moore et al*., 2006) (*Moore and Braucher, submitted* b).

2.2 Atmospheric Dust Deposition

Time-varying mineral aerosol deposition to the ocean is simulated using a 3-D atmospheric chemical transport model (*Mahowald et al*., 2003; *Luo et al.*, 2003) based on National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis (*Kistler et al*., 2001). The dust source and deposition scheme is based on the Dust Entrainment and Deposition (DEAD) scheme (*Zender et al*., 2003), and the chemical transport model is the Model of Atmospheric Transport and Chemistry (MATCH) (*Rasch et al*., 1997), which has been developed specifically to be used with reanalysis winds (*Mahowald et al*., 1997). The dust source areas are defined as dry, poorly vegetated regions which have easily erodible sources, using topographic lows as preferential source areas (*Ginoux et al*., 2001). Dust is removed by wet deposition during precipitation events, and by dry deposition from gravitational settling and turbulent processes.

The ability of the dust model to correctly simulate the annual mean, and seasonal cycle of dust has been compared against in situ and satellite observations elsewhere (*Luo et al*., 2003; *Mahowald et al*., 2003). For the in situ concentration data at 10 stations where multiple years are available, the model gets statistically significant correlations for the seasonal cycle, and interannual variability at 8 of 10 stations, with the most difficulty and lowest correlations at stations in the southern hemisphere (Table 1, *Mahowald et al*., 2003). Comparisons of daily averaged concentrations obtain correlation coefficients of 0.31 to 0.84 for the 7 stations with daily averaged data. Correlation coefficients with available satellite data (TOMS AAI and AVHRR optical depth) in regions where dust is adequately sampled are above 0.60 (Figure 2, *Mahowald et al*., 2003). Much of the interannual variability in dust concentrations downwind of the source regions is driven by

atmospheric transport (or transport/source correlations) and not by source interannual variability (*Tegen and Miller*, 1998; *Mahowald et al.*, 2003).

2.3 Atmospheric Forcing and Ocean Physical Hindcasts

The POP is a z-level, hydrostatic, primitive equation model integrated here with a resolution of 3.6° in longitude, 0.8° to 1.8° in latitude, and 25 vertical levels (*Yeager* et al., 2006). Effects of mesoscale eddy transport are parameterized according to *Gent and McWilliams* (1990). The *Large et al.* (1994) K-Profile Parameterization is implemented in the vertical to capture surface boundary-layer dynamics and interior diapycnal mixing. The historical simulation (1979-2004) is integrated with air-sea heat, freshwater, and momentum fluxes derived from a bulk flux forcing method that combines 6-hourly atmospheric surface fields (temperature, humidity, winds) from the NCEP reanalysis (*Kistler et al.,* 2001) with satellite and in-situ derived clouds, precipitation, runoff and sea-ice fraction (*Large and Yeager*, 2004). *Doney et al.* (2007) present a quantitative skill assessment of the ocean physical solutions in terms of interannual variability of temperature, sea surface height, and circulation.

Initial conditions for the nutrient and inorganic carbon variables are prescribed from data based climatologies (e.g., *Key et al*., 2004). The ecological-biogeochemical simulation is spun-up for several hundred years, prior to initiating the interannual varying forcing, using a repeat annual cycle of physical forcing, dust deposition, and fixed preindustrial atmospheric CO_2 mole fraction (280 ppm). The full interannual variability in physics and dust forcing is initiated in model year 815 (equivalent to calander year 1979). In the pre-industrial simulation atmospheric CO_2 mole fraction remains fixed at 280 ppm over the hindcast (1979-2004). In a companion anthropogenic CO_2 simulation, atmospheric CO_2 starts to evolve over time mid-way through the spin-up following ice-core and historical CO_2 observations from the 1700s forward to 1979; in that simulation, atmospheric CO_2 tracks observed global mean temporal trends over the hindcast (1979-2004).

The model ecosystem components converge to a repeat annual cycle within a few years of spin-up. There remains a slow drift in the subsurface nutrient and inorganic

carbonate fields in the pre-industrial simulation. The global net air-sea CO_2 uptake flux is 0.150 PgC y^{-1} (mean areal flux 0.025 mol C m^{-2} y^{-1}), but the change in the drift over the 26 year integration (1979-2004) is only -0.010 PgC y (mean -0.002 mol C m^{-2} y^{-1}) and much smaller than the simulated interannual variability. In a companion anthropogenic CO_2 simulation, atmospheric CO_2 evolves over time during the latter part of the model spin-up following historical observations.

3. Evaluation Data Sets

Table 1 presents details on the specific field and remote sensing data sets used for model evaluation in this study. Information in the table includes each specific variable, its units, spatial and temporal resolution of the underlying data set, data source, and reference(s). Most of the data sets are global in extent, which limits us primarily to data climatologies (annual mean and seasonal cycle) and satellite data products (annual mean, seasonal cycle, and subannual to interannual variability). This is not to argue that other data sets are not of value, a fact that we illustrate using an example 1-D water column time-series. Our emphasis here is mostly on surface water properties.

The data sets are chosen to highlight key aspects of the coupled BEC simulation with regards to physics, chemistry and biology. Physical fields include sea surface temperature (SST), which is important for biological growth and respiration rates as well as air-sea gas exchange, and mixed layer depth (MLD), which influences nutrient entrainment and the average light field observed by the phytoplankton. Biogeochemical fields include surface water dissolved inorganic macronutrients (nitrate (NO_3), silicate (SiO_3), and phosphate (PO_4)) and reflect a balance between physical nutrient supply and net biological nutrient drawdown. We also examine the simulated fields of dissolved gases oxygen (O_2) and the carbon dioxide partial pressure (pCO_2), as well as air-sea O_2 and CO_2 fluxes, which reflect physical transport, solubility variations, net community production, and ocean-atmosphere exchange. The global biological fields are derived from satellite ocean color data and include chlorophyll (Chl) and vertically integrated primary production ($\int PP$), which are measures of phytoplankton pigment standing stock and the photosynthesis that fuels the upper ocean food web. We also examine a pair

of relatively new remote sensing products, phytoplankton specific growth rate (μ) and phytoplankton carbon concentration ($P_{C,}$), as described in more detail below.

The list of evaluation datasets is heavy on bottom-up physical-chemical forcing data, biogeochemical tracers, and phytoplankton responses, but light on many of the higher trophic-level dynamics and loss processes that are also integral to the BEC solutions (Figure 1). For example, we have not yet incorporated measures of zooplankton biomass and grazing rates because of a lack of comprehensive global data. While some macrozooplankton biomass climatologies exist, there is no similar treatment for microzooplankton that are essential to verifying the behavior of our single aggregated zooplankton compartment. For somewhat different reasons, we do not include an explicit measure of export production. While globally gridded export flux maps are available, their construction from satellite data (e.g., *Laws et al*., 2000) involves a considerable level of model or empirical assumptions; they are essentially derived products from derived products, in this case primary production. Unlike most satellite ocean color algorithms there is no direct link to radiances nor extensive in situ validation, and it is unclear whether they serve as independent observational assessments or more of a model-model comparison (*Najjar et al.*, 2007). Recent compilations of deep-sea sediment trap data offer another approach for a point by point assessment of simulated export production, but because only a small fraction of surface export reaches the deep ocean such analysis also folds in the skill of the model subsurface remineralization parameterization, which can have large uncertainties.

Some assessment variables require merging multiple observational data-sets, which can add potential biases to model-data assessments if particular care is not taken. For example, air-sea CO_2 flux maps (e.g., *Takahashi et al*., 2002) are commonly created by joining air-sea ΔpCO_2 data with wind-speed dependent gas transfer velocities. Observational flux estimates thus scale directly, if non-linearly, with wind speed, and the use of different wind speed products will result in different observational flux estimates even for the same underlying ΔpCO_2 data. A good argument can be made for adjusting the global mean transfer velocity to correct for differences in the global mean wind speed from different wind products (or the global mean wind speed squared in the case of a

quadratic wind-speed formulation), but there will still be seasonal and regional spatial flux differences introduced by the different wind products.

For models the issues are somewhat more complicated because wind speed products are used to force both physical circulation and biogeochemistry. The model air-sea CO_2 flux and ΔpCO_2 are dynamically coupled in the model solutions in that a change in flux will alter surface water DIC and thus ΔpCO_2. The use of a different wind speed product between the observational estimate and as forcing for the model could therefore introduce model-data differences in both variables. Despite their interdependence, we include here model skill metrics for both air-sea CO_2 flux and ΔpCO_2 because they are commonly presented observational fields and reflect somewhat different weighting of different regions and seasons within the model. Note, however, that the uncertainties in more derived "observational" products, such as air-sea CO_2 flux, are larger than directly measured fields; in the case of air-sea CO_2 flux, this includes the incorporating error in wind speed based gas transfer relationships (*Wanninkhof*, 1992).

Traditionally, surface chlorophyll concentration has been the primary satellite-derived ecosystem variable related to biological ocean carbon cycling. From chlorophyll concentration, a variety of models have been described for estimating water-column net primary production (see reviews by *Campbell et al*. 2002 and *Carr et al*., 2006). In most cases the estimated primary production is proportional to the satellite derived chlorophyll multiplied by spatially and temporally varying factors that may depend upon estimates of surface light, nutrients, temperature, mixed layer, etc. Similar to the arguments for including both air-sea CO_2 flux and ΔpCO_2, we also include both satellite chlorophyll and net primary production as evaluation data sets because of their common usage, their different weighting of the critical process of marine photosynthesis, and the considerable independent effort going into the validation of satellite productivity using ^{14}C-based field primary productivity data sets (*Carr et al*., 2006 and *Friedrichs et al*., submitted).

The satellite primary productivity algorithms depend on empirical descriptions (generally temperature-dependent) of phytoplankton assimilation efficiencies that may be somewhat unreliable for detecting regional temporal variability, particularly in response to factors such as aeolian iron deposition. Recently, however, an alternative approach to analyzing satellite ocean color data has been developed that yields not only estimates of

phytoplankton mixed layer pigment concentrations but also new and independent information on particulate scattering coefficients (*Garver and Siegel*, 1997; *Maritorena et al.,* 2002; *Siegel et al*., 2002).

With this additional information, it is now possible to directly derive phytoplankton carbon biomass, chlorophyll-to-carbon ratios, and phytoplankton growth rates from space (*Behrenfeld and Boss*, 2003, *Behrenfeld et al.,* 2005), and thus more reliably detect and distinguish physiological- and biomass-dependent responses to changing environmental conditions. Importantly, these three phytoplankton characteristics are directly comparable to ocean model variables. For the current study, we employ a spectrally-resolved version of the Carbon-based Production Model (CbPM) (*Westberry and Behrenfeld*, 2006) that yields improved descriptions of vertical variability in phytoplankton carbon, Chl:C and growth rates compared to the original CbPM (*Behrenfeld et al*., 2005). Due to the tight coupling between phytoplankton growth rates and zooplankton grazing, physical/chemical perturbations to mixed layer growth conditions can regularly occur without a detectable signature in phytoplankton biomass. Environmental changes are, however, invariably imprinted in physiological characteristics of the phytoplankton assemblage.

4. Model-Data Skill Metrics

The skill metric suite includes model-data comparisons of fields of tracers (standing stocks) and biological flows or rates. Monthly averages χ are computed from the model output to match the common temporal resolution of observational climatologies. The observations are interpolated to horizontal model grid for spatial fields and, where applicable, averaged to monthly resolution. For each observed χ_O and model predicted χ_P variable, we compute for each grid point a long-term mean $\langle\chi\rangle$, an annual mean $\overline{\chi}$, and a mean annual cycle χ^a (e.g., average January, average February, etc.) for the period of analysis 1979-2004. We define a series of anomalies (*Doney et al*., 2007):

$$\chi' = \overline{\chi} - \langle\chi\rangle$$
$$\chi'' = \chi - \langle\chi\rangle$$
$$\chi^* = \chi - \chi^a \qquad (2)$$

where χ' are the annual mean anomalies, χ'' are the monthly anomalies, and χ^* are the monthly deseasonalized anomalies. Standard deviations (σ) are computed for each variable and for the various anomalies, as needed.

We apply a standard suite of univariate model-data skill metrics (e.g., *Evans*, 2003; *Stow et al*., submitted) including the model-data correlation coefficient *r*, the root mean square error ε_{rms}, and the average error or bias ε_{bias}. The metrics are applied to different temporal and spatial domains, depending upon the availability of observations and the question of interest. For example, metrics for a time-series at a single grid point for the full monthly data over the full analysis period of the hindcast simulations (1979-2004) would be:

$$r(\chi) = \frac{\sum(\chi_O - \langle\chi_O\rangle)(\chi_P - \langle\chi_P\rangle)}{\sqrt{\sum(\chi_O - \langle\chi_O\rangle)^2 \sum(\chi_P - \langle\chi_P\rangle)^2}} = \frac{\sum(\chi''_O)(\chi''_P)}{\sqrt{\sum(\chi''_O)^2 \sum(\chi''_P)^2}} \tag{3}$$

$$\varepsilon_{rms}(\chi) = \sqrt{\frac{1}{N}\sum(\chi_P - \chi_O)^2} \tag{4}$$

$$\varepsilon_{bias}(\chi) = \frac{1}{N}\sum(\chi_P - \chi_O) = \langle\chi_P\rangle - \langle\chi_O\rangle \tag{5}$$

where the summation Σ is over N=312 months (26 years x 12 months). For variables where only a seasonal climatology is available, we define the corresponding metrics at the grid-point scale using the mean annual mean cycle $r(\chi^a)$, $\varepsilon_{rms}(\chi^a)$, and $\varepsilon_{bias}(\chi^a)$ and n=12. Where appropriate, similar statistics are computed on larger-spatial scales including zonal averages across ocean basins (e.g., Atlantic), global zonal averages, and global averages. Also, the simulated model fields are sub-sampled in time to match the data sampling when observations exist for only a subset of the model hindcast.

For some variables, that have large dynamic ranges, we may choose to analyze the log-transform of the data:

$$\mathrm{X} = \log(\chi) \tag{6}$$

The log-transform tends to give more equal weight to all of the data and not skew the statistics towards the largest data values. The mean of the log-transformed variable $\langle\mathrm{X}\rangle$, can be related to the geometric mean of the untransformed variable $\langle\chi\rangle_G$:

$$\langle \chi \rangle_G = \sqrt[N]{\prod_i \chi_i} = \exp\left(\langle \mathrm{X} \rangle\right) \tag{7}$$

The geometric bias:

$$\varepsilon^G_{bias}(\chi) = \exp\left(\langle \mathrm{X}_P \rangle - \langle \mathrm{X}_O \rangle\right) \tag{8}$$

gives a measure of the typical bias normalized by the value of the variable, $(\chi_P - \chi_O)/\langle \chi_O \rangle$; $\varepsilon^G_{bias}(\chi) < 1$ occurs when the model tends on average to underestimate the observations and $\varepsilon^G_{bias}(\chi) > 1$ when the model tends on average to overestimate. The corresponding geometric root mean square error given by:

$$\varepsilon^G_{rms}(\chi) = \sqrt{\exp\left(\frac{1}{N}\sum\left(\mathrm{X}_P - \mathrm{X}_O\right)^2\right)} \tag{9}$$

reflects the size of the typical model-data error normalized relative to the typical data value. A perfect model with no error would give a ε^G_{rms} value of 1.0; a value of ε^G_{rms} of 2.0 would reflect a case where the typical error at any point is comparable in size to the observed value. Note that for geometric averaging, the errors are not symmetric about the mean in the geophysical space. Geometric averaging is used for several of the bio-optical data sets (chlorophyll, primary production, phytoplankton biomass) that have approximate log-normal distributions (*Campbell*, 1995).

We use Taylor diagrams (*Taylor*, 2001) to display simultaneously information on model-data skill for a suite of variables from the ocean biological model (*Lima and Doney*, 2004). The Taylor diagram combines global r and ε_{rms} normalized by the standard deviation with the ratio of the predicted to observed standard deviation σ_P/σ_O into a single point in a two-dimensional plot. The ratio of the standard deviations indicates the relative amplitude of the simulated and observed variations, while the correlation coefficient indicates whether the fields have similar patterns of variation, regardless of amplitude. The normalized ε_{rms} reflects differences in the overall pattern of variations. In the diagram, the radial distances from the origin are proportional to the ratio of the standard deviations and the azimuthal positions give the correlation between the two fields. The point representing the observational reference field is plotted along the abscissa and has coordinate $\sigma_P/\sigma_O = 1$ and $r = 1$. The distance between the test and reference point is proportional to the normalized ε_{rms} between the two fields. A perfect

match between model output and observations would plot at the x=1.0 point on the x-axis; a point representing no relationship between model and observations whatsoever would plot on the y-axis.

Joliff et al. (submitted) introduce a related diagnostic plot, termed a "target diagram", which displays each variable as a point as a function of the bias ε_{bias} (y-axis) and the unbiased rms error $\varepsilon_{rms}^{unbiased}(\chi)$ (y-axis):

$$\varepsilon_{rms}^{unbiased}(\chi) = \sqrt{\frac{1}{N}\sum(\chi_P - \chi_O - \varepsilon_{bias}(\chi))^2} \qquad (10)$$

Both ε_{bias} and $\varepsilon_{rms}^{unbiased}(\chi)$ are normalized by dividing by the standard deviation of the observations σ_O. Additionally, for the target diagram the values of the normally positive definite $\varepsilon_{rms}^{unbiased}(\chi)$ are treated as positive if $\sigma_P > \sigma_O$ and negative for the opposite. The distance from the origin to a point on the target diagram is the total normalized ε_{rms}.

5. Model-Data Skill Assessment

5.1 Example Diagnostic Plots to Ocean Chlorophyll

Given that we are comparing 3-D time-varying model results against observations for more than a dozen variables, the number of potential diagnostic plots quickly expands beyond the scope of a single journal paper. The solution used here is to illustrate a standard set of diagnostic plots for a specific example variable, in this case surface ocean chlorophyll, and to make available the entire set of figures on a CCSM-BEC diagnostic webpage.

Figures 2-5 display the results of a comparison of model predicted surface ocean chlorophyll (mg Chl m^{-3}) against results from the SeaWiFS ocean color sensor (Sept. 1997- Dec. 2004). The spatial map of the bias in the long-term mean $\varepsilon_{bias}(\langle Chl \rangle)$ (Figure 2, bottom panel) exhibits large-scale, coherent error patterns. The model surface chlorophyll tends to be too high in the subtropical oligotrophic gyres ($\varepsilon_{bias} > 0$) and too low in the subpolar gyres ($\varepsilon_{bias} < 0$). This error pattern may reflect problems with the single adaptive zooplankton pool; relative to primary production, grazing is too weak in picoplankton dominated subtropics and too strong in bloom environments. In particular,

multiple zooplankton pools may allow for a seasonal disconnect in grazing, and thus a stronger bloom, in temperate and high latitudes.

Consistent with previous coarse-resolution global model results, the simulated chlorophyll levels are also underestimated in shallow coastal regimes. The reasons for this coastal bias are numerous but are likely dominated by physical errors: vertical upwelling due to off-shore coastal flow is poorly resolved at coarse scale and without careful treatment of the wind stress curl; the model lacks tidal mixing, an important mechanism for vertical mixing and nutrient supply on continental shelves; the mesoscale eddy parameterizations used in the global model are designed for adiabatic ocean interiors and are not adequate for the highly turbulent, and often topograqphically controlled lateral mixing on shelves.

The long-term subtropical/subpolar bias reflects in part the fact that the model does a relatively poor job capturing the magnitude of the peak surface chlorophyll concentrations during summer in the temperate northern hemisphere, as illustrated in a plot of the model and observed zonal average of the seasonal anomalies $\left(Chl^a - \langle Chl \rangle\right)$ versus month (Figure 3). The phasing of the model northern hemisphere spring bloom is approximately correct, but high chlorophyll levels are not sustained over the summer in the simulation. The phasing of the annual cycle in the southern hemisphere mid-latitudes (40-60°S) matches the observations well, but in this case overestimates the amplitude of the annual cycle, with chlorophyll values too high in the southern hemisphere summer and too low in the winter.

The hindcast exhibits substantial subannual to interannual variability $\varepsilon_{rms}(Chl^*)$ beyond the variance introduced from the mean annual cycle $\varepsilon_{rms}\left(Chl^a\right)$. This is demonstrated in the middle panel of Figure 4, a spatial map of the root-mean-squared variability in the deseasonalized anomalies in surface chlorophyll. The model hindcast simulation exhibits considerable interannual variability in the tropics, and temperate to subpolar oceans. The temporal correlations between the model and SeaWiFS chlorophyll data $r(Chl^*)$ (top panel, Figure 4) are high in the western and central Equatorial Pacific and in the subtropics. However, the model-data correlations tend to drop in the mid- to high-latitude regions, where correlations are often not statistically different from zero.

5.2 Basin and Global Aggregated Skill Metrics

A corresponding set of model-data diagnostic plots can be constructed for all variables of interest, but we have also found it convenient to generate multi-variable synthesis plots and a table assessing skill for more aggregated basin zonal means and global averages/integrals. We include the zonal means because the spatial patterns of the model-data bias or residuals is often as or more interesting than the global average/integrated bias.

Figure 5 illustrates this approach and compares the observed and simulated annual mean zonal averages for a range of variables. The bias, $\varepsilon_{bias}(\chi)$, in the zonal means is simply the difference of the model and data curves. The CCSM BEC hindcast exhibits a number of large-scale biases, many of which are coherent across multiple variables. The simulation displays excess surface macro-nutrients in the tropical Pacific, likely the result of a combination of physical circulation errors and too much iron scavenging. Interestingly, however, the model and observed zonal average phytoplankton growth rates for the tropical and subtropical Pacific are similar and model productivity is actually higher that observed, suggesting that errors may also arise from other aspects of the biological cycling (e.g., export flux, subsurface remineralization).

As discussed above the model chlorophyll underestimated the data in mid- to high latitudes. Note that in the zonal average plots, the overall model bias in the northern hemisphere subtropics is negative. The low simulated chlorophyll in the coastal upwelling regions overwhelms the model positive bias in the subtropics. Another region of marked bias in the model simulations is the tropical Atlantic, where simulated phytoplankton biomass, specific growth rates and primary production are low relative to the observations. This region in the model is strongly P-limitation, which may reflect a combination of model errors in circulation, export and subtropical nitrogen fixation.

Global model skill is summarized in the Taylor (Figure 6) and target (Figure 7) diagrams and Table 2. The global mean bias $\varepsilon_{bias}(\chi)$ for most variables is relatively small (Figure 7, top panel). However, in agreement with Figure 5, the spatial variation in the long-term mean $\langle \chi \rangle$ is not captured well in many of the ecosystem variables (e.g., chlorophyll, primary production, phytoplankton growth rate), which exhibit correlation

coefficients $r(\chi) < 0.4$. The strongest correlations are for SST and nutrients. The model-data correlations for the seasonal anomalies $\left(\chi^a - \langle\chi\rangle\right)$ are for the most part between 0.3 and 0.7, and the model tends to overestimate the seasonal amplitude of some variables (e.g., chlorophyll) while underestimating that of others (e.g., phosphate and silicate) (Figure 6 middle panel). On a global basis, the model exhibits little skill in capturing the interannual anomalies χ^a except for SST.

5.3 Comparison against local (Eulerian) time-series data-sets

While powerful evaluation tools, global data sets also often have associated limitations. The compilation of disjoint field data sets from different time period and from different investigators and/or methodologies introduces errors and blurs important natural variability and climate change signals. Uncertainties arise due to spatial/temporal interpolation and extrapolation involved in creating complete global climatologies from sparse data. Because of limitations in sampling the subsurface ocean and dynamical rates from remote sensing and underway sampling, global data sets illuminate only a portion of most ecosystem models. Data-rich, local time-series sites provide a wealth of complementary information on depth profiles, the annual mean cycle, and interannual variability of physical and biogeochemical variables. These observations, although limited in spatial information, may be compared with global models to evaluate the credibility of the simulations. Time-series data have a rich history in marine ecosystem modeling as test-beds for model development and validation (e.g., Doney et al., 1996; Evans, 1999; Moore et al., 2002; Friedrichs et al. 2007).

An example time-series comparison of the CCSM-BEC hindcast simulation is presented in Figure 8 for the Hawaii Ocean Time-series (HOT) site (http://hahana.soest.hawaii.edu/hot/hot_jgofs.html). Monthly average vertical 1-D profiles are sub-sampled from the hindcast versus time at the nearest grid-point to the HOT Aloha station.

At the surface, the simulated phasing of the seasonal cycle in chlorophyll (Figure 8) is approximately correct with a minimum in the summer and a maximum in the winter, but the model tends to overestimate the mean chlorophyll levels as noted previously for the subtropical North Pacific. The seasonal surface chlorophyll cycle in the model

mimics the seasonal cycle of mixed layer depth; during winter simulated mixed layer depth deepens to 60-70m, resulting in the entrainment of nutrients that drives enhanced productivity. The strength of the simulated winter bloom is somewhat stronger than that observed in the data, however, even though the maximum winter observed mixed layer depth in the observations are somewhat deeper (and shifted later in the year).

The simulated magnitude of the deep chlorophyll maximum is slightly smaller and shallower (by about 20m) than in the observations. The observations also exhibit non-zero chlorophyll levels well below 150m in very low light conditions. The deeper deep chlorophyll structure in the observations results in a characteristic positive/negative dipole pattern in the model-data difference plot. The deep chlorophyll maximum and subsurface chlorophyll penetration depth are largely set by the initial slope of phytoplankton productivity-irradiance curve. The model chlorophyll field can be shifted downward by increasing the initial productivity-irradiance slope (effectively reducing light limitation at low light) but at the expense of replicating surface productivity. The solution, explored in other models, involves incorporating distinct high-light and low-light phytoplankton populations in the subtropical gyre (*Y. Spitz, per. comm.*).

The coarse-resolution global model exhibits substantial interannual variability in both mixed layer depth and chlorophyll. Chlorophyll and productivity gradually drift downward from 1988-1998 and then begin to increase following a deep mixing event in the winter of 1989. The observed mixed layer depth experiences more variability than in the model, reflecting in part the aliasing of mesoscale eddies passing by the time-series site (Doney, 1996). There is an indication of stronger winter mixing in the observations in 1998-1999 followed by higher near surface chlorophyll levels (but not for as an extended period as in the model).

Similar patterns of model skill (and misfit) are found for aggregated skill metrics for a suite of variables displayed in Taylor diagram format (Figure 9) and in tabular form (Tables 4 and 5). The model-data correlation values r are greater than 0.95 with a normalized standard deviation near 1.0 for the annual-mean vertical profiles of temperature, oxygen, macro-nutrients, and primary production; the corresponding chlorophyll correlation is smaller (0.56) in large part because of the shallower simulated deep-chlorophyll maximum. The model-data agreement for the seasonal cycle (monthly-

time Taylor diagram) is considerably weaker, with all of the model-data correlations less than 0.3 and substantial underestimates in the simulated strength of the seasonal cycle ($\sigma_P/\sigma_O < 0.5$) for primary production and macronutrients.

The model's interannual variability is much weaker than that in the data, and the hindcast shows essentially no skill on the observed interannual variability at the HOT site. In contrast to much of the subtropics where model skill is relatively high for interannual chlorophyll variability, the region around Hawaii is a region of little skill. A similar analysis for the Bermuda Atlantic Time-Series (BATS) (not shown) results in some what higher but still low model-data correlations, $r(Chl) = 0.131$ and $r(PP) = 0.222$. This highlights that care is need in making local data comparisons with a global model, the skill of which can vary significantly from region to region and which is confounded by regional biases in physics and unresolved high frequency variability in the observations.

One strategy around these difficulties would be to combine the 3-D model assessment with a complementary assessment of 1-D vertical simulations for targeted time-series stations, so-called "regional test-beds". Surface forcing can be adjusted to give 1-D physical simulations that closely fit specific time-series records (e.g., Doney, 1996), and 1-D ecosystem simulations are more amenable to parameter opimization, data assimilation, and cross-model intercomparison studies.

6. Discussion and Future Directions

A systematic and quantitative approach for assessing model-data skill is an essential tool in model development, evaluation, and data assimilation (*Gregg et al., submitted*), and emerging global-scale field and satellite data sets provide invaluable opportunities for testing upper-ocean coupled ecosystem-biogeochemistry-physical models. The example suite of data sets and skill metrics presented here, while certainly still incomplete, illustrates several general points with regards to the CCSM-3 BEC model. First, overall the skill metrics highlight the fact that the model solutions have a number of deficiencies in replicating the observational data sets. Second, the degree of model skill differs sharply among variables. For example, simulated SST and some of the surface nutrient fields exhibit consistently higher skill across most of the metrics than the

simulated ecological fields, chlorophyll, primary production and the new remote sensing products such as phytoplankton growth rate. Surface pCO_2 and CO_2 and O_2 air-sea fluxes are typically intermediate in skill between the physical and ecological variables. Third, regional spatial biases and seasonal cycle errors are often consistent across multiple variables, pointing towards a common dynamical problem within the coupled model. Fourth, model skill is time and space scale dependent; the model solutions of the seasonal cycle are more skillful than interannual variability. And fifth, it is challenging for a global model solution to replicate observations from local time-series because there are many subgridscale processes and representation issues that tend to confound the comparison.

The emphasis here has been primarily on assessing model skill in replicating ecological and biogeochemical metrics. But the upper ocean system is coupled, and the success of the biological and chemical simulations depends critically on a high-quality underlying physical circulation model (*Doney et al*., 2004; 2007). Physical model biases and errors thus should be an integral component of ecological and biogeochemical model-data assessment. Here we focused on two physical properties, SST and mixed layer depth. Other important facets of the circulation from a biological perspective include upwelling rates and near surface water column structure (Doney, 1996). As an example of how physical biases may influence ecosystem behavior, Figure 10 displays the maximum winter mixed layer depth from the model and an observational estimate (*Boyer-Montegut et al.,* 2004). The CCSM-3 BEC has consistently deeper maximum mixed depths in the tropics and subtropics and shallower maximum mixed depths in the northern hemisphere subpolar gyres and Southern Ocean. The model chlorophyll bias patterns (Figure 2) may in part reflect these physical errors. In the subtropics, simulated winter chlorophyll is too high and summer levels are closer to observations. One possible explanation is that the nutrient supply from excessively deep mixing in winter induces higher simulated chlorophyll/C ratios. The too weak simulated mixing in the subpolar gyres and Southern Ocean, in contrast, limits the supply of nutrients to the surface (e.g., note the negative nutrient bias around the Antarctic) and may allow for a large overwintering zooplankton population, which can then limit the magnitude of the spring bloom.

As a pilot study, we have restricted our large-scale analysis here primarily to globally gridded synthesis data sets. But there are a number of key variables for which data coverage is still too poor to create global synthesis products but which one would want to consider in a more comprehensive assessments. Much of the same diagnostic machinery, however, can rather straightforwardly be used on a collection of stations. For example global maps and zonal mean plots can be generated to examine spatial biases and seasonal and interannual variability where data are plotted only for the grid points where there are observations. The statistics for the summary Taylor and target diagrams are independent of whether one is using gridded or point data. Examples of additional, non-gridded data sets that we plan to include in the future are deep sediment traps (e.g., *Gehlen et al*., 2006), surface and sub-surface iron concentrations (*Moore and Braucher, submitted* b), and phytoplankton taxonomy derived from cell counts or pigments (e.g., *Gregg and Casey*, 2007; http://polar.gsfc.nasa.gov/research/oceanbiology/index.php). Particularly for the Southern Ocean, one can also use atmospheric O_2 and CO_2 data sets to assess ocean model behavior, with of course the caveat that the comparison will also incorporate errors in atmospheric transport used to translate surface fluxes to atmospheric fields (e.g., *Naegler et al., 2007; Nevison et al. in press; submitted*).

Other directions to pursue include pattern analysis and multivariate model-data skill metrics that allow for an investigation of whether the model is correctly capturing the spatial or temporal relationships among the data (*Stow et al*., *submitted*). Small temporal and spatial phase shifts between the model simulations and observations can introduce large apparent biases and seasonal to interannual variability errors. Lagged correlation analysis can help identify phase errors. Empirical orthogonal functions can be used to assess the similarity between model and observed time-space variability patterns in the presence of spatial and temporal phase error. Multivariate analyses (e.g., seasonal property-property phase diagrams; factor analysis; binary-discriminatory receiver-operator methods) can be used to identify regions or times when model dynamics diverge from that seen in the observations. The multivariate analyses of model dynamics may be particular useful when the model-data skill assessments are applied to fully coupled ocean-atmosphere climate models (*Doney et al*., 2006; *Schneider et al*., *submitted*). Direct comparisons to observations are more difficult in this case because persistent

physical biases in coupled models propagate into the ecological/biogeochemical mean state and seasonal cycle and because the coupled ocean-atmosphere models generate their own internal climate variability and thus assessment of simulated interannual to decadal variability can only be done statistically, not directly.

Acknowledgements

This work was supported in part by grants from the NSF/ONR National Ocean Partnership Program (N000140210370), the NASA Ocean Biology and Biogeochemistry Program (NNX07AL80G), and the NSF Center for Microbial Oceanography Research and Education (C-MORE).

References

Aumont, O., Maier-Reimer, E.B.S., Monfray, P., 2003: An ecosystem model of the global ocean including Fe, Si, P colimitations, *Global Biogeochem. Cycles*, **17**, 1060.

Behrenfeld, M., P. Falkowski 1997: Photosynthetic rates derived from satellite based chlorophyll concentration, *Limnol. Oceanogr*., **42**, 1–20.

Behrenfeld, M. J., Boss, E., Siegel, D. A., and Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space, *Global Biogeochemical Cycles*, **19**, GB1006.

Bopp, L., P. Monfray, O. Aumont, J.-L. Dufresne, H. Le Treut, G. Madec, L. Terray, and J.C. Orr, 2001: Potential impact of climate change on marine export production, *Global Biogeochem. Cycles*, 15, 81–100.

Bopp, L., O. Aumont, S. Belviso, P. Monfray, 2003: Potential impact of climate change on marine dimethyl sulfide emissions, *Tellus B*, **55**, 11–22.

Boyd, P.W., Doney, S.C., 2002: Modelling regional responses by marine pelagic ecosystems to global climate change, *Geophys. Res. Lett*., **29**, 1806, doi:10.1029/2001GL014130.

Boyer-Montégut, C., Madec, G., Fischer, A. S., Lazar, A., and Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. Journal of Geophysical Research, 109(C12003).

Campbell, J. W., The lognormal distribution as a model for bio-optical variability in the sea., *J. Geophys. Res*., **100**, 13,237– 13,254, 1995.

Carr, M-E., M.A Friedrichs, M. Schmeltz, M.N. Aita, D. Antoine, K.R. Arrigo, I. Asanuma, O. Aumont, R. Barber, M. Behrenfeld, R. Bidigare, E.T. Buitenhuis, J. Campbell, A. Ciotti, H. Dierssen, M. Dowell, J. Dunne, W. Esaias, B. Gentili, W. Gregg, S. Groom, N. Hoepner, J. Ishizaka, T. Kameda, C. Le Quere, S. Lohrenz, J. Marra, F. Melin, K. Moore, A. Morel, T.E. Reddy, J. Ryan, M. Scardi, T. Smyth, K. Turpie , G. Tilstone, K. Waters, Y. Yamanaka. 2006. A comparison of global estimates of marine primary production from ocean color. *Deep Sea Res.*, **53**, 741-770.

Collins, W.D., M. Blackmon, C.M. Bitz, G.B Bonan, C.S. Bretherton, J.A. Carton, P. Chang, S. Doney, J.J. Hack, J.T. Kiehl, T. Henderson, W.G. Large, D. McKenna, and

B.D. Santer, 2006: The Community Climate System Model: CCSM3, *J. Climate,* **19**(11), 2122-2143.

Conkright, M. E., Locarnini, H. E., Garcia, T. D., O'Brien, T. P., Boyer, C., and Stephens, J. (2001). World ocean atlas 2001: Objective analyses, data statistics, and figures, documentation. National Oceanographic Data Center Internal Report 17, National Oceanographic Data Center, Silver Spring, MD. 17 pp.

Denman, K.L., 2003: Modelling planktonic ecosystems: parameterizing complexity, *Prog. Oceanogr*., **57**, 452–492.

Doney, S.C. and H.W. Ducklow, 2006: A decade of synthesis and modeling in the U.S. Joint Global Ocean Flux Study, *Deep-Sea Res. II,* **53**(5-7), 451-458.

Doney, S.C., 1996. A synoptic atmospheric surface forcing data set and physical upper ocean model for the U.S. JGOFS Bermuda Atlantic Time-Series Study (BATS) site. *J. Geophys. Res., Oceans*, 101, 25,615-25,634.

Doney, S.C., D.M. Glover, and R.G. Najjar, 1996. A new coupled, one-dimensional biological--physical model for the upper ocean: applications to the JGOFS Bermuda Atlantic Time Series (BATS) site. *Deep-Sea Res. II*, **43**, 591-624.

Doney, S.C. and M. Hood, 2002: A Global Ocean Carbon Observation System, A Background Report, Global Ocean Observing System Report No. 118, UNESCO Intergovernmental Oceanographic Commission IOC/INF-1173, 55p.

Doney, S.C., K. Lindsay, K. Caldeira, J.-M. Campin, H. Drange, J.-C. Dutay, M. Follows, Y. Gao, A. Gnanadesikan, N. Gruber, A. Ishida, F. Joos, G. Madec, E. Maier-Reimer, J.C. Marshall, R.J. Matear, P. Monfray, A. Mouchet, R. Najjar, J.C. Orr, G.-K. Plattner, J. Sarmiento, R. Schlitzer, R. Slater, I.J. Totterdell, M.-F. Weirig, Y. Yamanaka, A. Yool, 2004: Evaluating global ocean carbon models: the importance of realistic physics, *Global Biogeochem. Cycles*, **18**, GB3017, doi:10.1029/2003GB002150.

Doney, S.C., K. Lindsay, I. Fung and J. John, 2006: Natural variability in a stable 1000 year coupled climate-carbon cycle simulation, *J. Climate*, **19**(13), 3033-3054.

Doney, S.C., S. Yeager, G. Danabasoglu, W.G. Large, and J.C. McWilliams, 2007: Mechanisms governing interannual variability of upper ocean temperature in a global hindcast simulation, *J. Phys. Oceanogr*., **37**, 1918-1938.

Dutay, J.-C., J.L. Bullister, S.C. Doney, J.C. Orr, R. Najjar, K. Caldeira, J.-M. Champin, H. Drange, M. Follows, Y. Gao, N. Gruber, M.W. Hecht, A. Ishida, F. Joos, K. Lindsay, G. Madec, E. Maier-Reimer, J.C. Marshall, R.J. Matear, P. Monfray, G.-K. Plattner, J. Sarmiento, R. Schlitzer, R. Slater, I.J. Totterdell, M.-F. Weirig, Y. Yamanaka, and A. Yool, 2002: Evaluation of ocean model ventilation with CFC-11: comparison of 13 global ocean models. *Ocean Modelling*, **4**, 89-120.

Evans, G.T., 1999: The role of local models and data sets in the Joint Global Ocean Flux study, *Deep Sea Res. I*, 46, 1369-1389.

Evans, G.T., 2003: Defining misfit between biogeochemical models and data sets, *J. Mar. Syst*., 40-41, 49-54.

Fasham, M.J.R., B.M. Balino, M.C. Bowles, R. Anderson, D. Archer, U. Bathmann, P. Boyd, K. Buesseler, P. Burkill, A. Bychkov, C. Carlson, C.T.A. Chen, S. Doney, H. Ducklow, S. Emerson, R. Feely, G. Feldman, V. Garcon, D. Hansell, R. Hanson, P. Harrison, S. Honjo, C. Jeandel, D. Karl, R. Le Borgne, K.K. Liu, K. Lochte, F. Louanchi, R. Lowry, A. Michaels, P. Monfray, J. Murray, A. Oschlies, T. Platt, J. Priddle, R. Quinones, D. Ruiz-Pino, T. Saino, E. Sakshaug, G. Shimmield, S. Smith, W. Smith, T. Takahashi, P. Treguer, D. Wallace, R. Wanninkhof, A. Watson, J. Willebrand, C.S. Wong, 2001: A new vision of ocean biogeochemistry after a decade of the Joint Global Ocean Flux Study (JGOFS), *AMBIO*, *Sp. Iss. 10*, 4-31.

Friedrichs, M.A.M., J.A. Dusenberry, L.A. Anderson, R. Armstrong, F. Chai, J.R. Christian, S.C. Doney, J. Dunne, M. Fujii, R. Hood, D. McGillicuddy, J.K. Moore, M. Schartau, Y.H. Spitz, and J.D. Wiggert, 2007: Assessment of skill and portability in regional marine biogeochemical models: the role of multiple planktonic groups, *J. Geophys. Res. Oceans*, **112**, C08001, doi:10.1029/2006JC003852.

Friedrichs, M.A.M., M.-E. Carr, R.T. Barber, M. Scardi, D. Antoine, R.A. Armstrong, I. Asanuma, M.J. Behrenfeld, E.T. Buitenhuis, F. Chai, J.R. Christian, A.M. Ciotti, S.C. Doney, M. Dowell, J. Dunne, B. Gentili, W. Gregg, N. Hoepffner, J. Ishizaka, T. Kameda, I. Lima, J. Marra, F. Mélin, J.K. Moore, A. Morel, R.T. O'Malley, J. O'Reilly, V.S. Saba, M. Schmeltz, T.J. Smyth, J. Tjiputra, K. Waters, T.K. Westberry, A. Winguth, Assessing the uncertainties of model estimates of primary productivity in the Tropical Pacific Ocean, *J. Mar. Systems*, *submitted*.

Follows, M.J., S. Dutkiewicz, S. Grant, and S.W. Chisholm, 2007: Emergent biogeography of microbial communities in a model ocean, *Science*, **315**, 1843-1846.

Garcia, H. E. and Keeling, R. F. (2001). On the global oxygen anomaly and air-sea flux. Journal of Geophysical Research, 106(C12):31155–31166.

Gehlen, M., L. Bopp, N. Emprin, O. Aumont, C. Hainze, and O. Ragueneau (2006) Reconciling surface ocean productivity, export fluxes and sediment composition in a global biogeochemical ocean model, *Biogeosciences*, **3**, 521-537.

Ginoux, P., M. Chin, I. Tegen, J. M. Prospero, B. N. Holben, O. Dubovik, and S.-J. Lin (2001), Sources and distribution of dust aerosols with the GOCART model, Journal of Geophysical Research, 106, 20255-20273.

Gregg, W.W. and N.W. Casey (2007) Modeling coccolithophores in the global oceans, *Deep-Sea Research II*, **54**, 447-477.

Gregg, W.W., Ginoux, P., Schopf, P.S., Casey, N.W., 2003. Phytoplankton and iron: validation of a global three dimensional ocean biogeochemical model. *Deep-Sea Research II*, **50**, 3143–3169.

Gregg, W.W., M.A.M. Friedrichs, A.R. Robinson, K.A. Rose, R. Schlitzer, K.R. Thompson, and S.C. Doney, Skill assessment in ocean biological data assimilation, *J. Mar. Systems, in press.*

Gruber, N., Gloor, M., Fan, S., and Sarmiento, J. L. (2001). Air-sea flux of oxygen estimated from bulk data: Implications for the marine and atmospheric oxygen cycles. Global Biogeochemical Cycles, 15(4):783–803.

Gruber, N., M. Gloor, S.E. Mikaloff-Fletcher, S.C. Doney, S. Dutkiewicz, M. Follows, M. Gerber, A.R. Jacobson, F. Joos, K. Lindsay, D. Menemenlis, A. Mouchet, J.L. Sarmiento, and T. Takahashi, Oceanic sources and sinks of atmospheric CO_2, *Global Biogeochem. Cycles, submitted*.

Hood, R.R., E.A. Laws, R.A. Armstrong, N.R. Bates, C.W. Brown, C.A. Carlson, F. Chai, S.C. Doney, P.G. Falkowski, R.A. Feely, M.A.M. Friedrichs, M.R. Landry, J.K. Moore, D.M. Nelson, T.L. Richardson, B. Salihoglu, M. Schartau, D.A. Toole, and J.D. Wiggert, 2006: Pelagic functional group modeling: progress, challenges and prospects, *Deep-Sea Res. II,* **53**, 459-512.

Kistler, R., E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H.v.e. Dool, R. Jenne, and M. Fiorino, (2001) The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation, *Bulletin American Meteorological Society*, 82, 247-267.

Large, W.G., and S.G. Yeager, Diurnal to decadal global forcing for ocean and sea-ice models: the data sets and flux climatologies. NCAR Technical Note NCAR/TN-460+STR, 111 pp, 2004.

Le Quéré, C., J. C. Orr, P. Monfray and O. Aumont, 2000: Interannual variability of the oceanic sink of CO_2 from 1979 through 1997. *Global Biogeochem. Cycles*, 14(4), 1247-1265.

Le Quéré, C., S.P. Harrison, I.C. Prentice, E.T. Buitenhuis, O. Aumont, L. Bopp, H. Claustre, L.C. da Cunha, R. Geider, X. Giraud, C. Klaas, K.E. Kohfeld1, L. Legendre, M. Manizza, T. Platt, R.B. Rivkin, S. Sathyendranath, J. Uitz, A.J. Watson, D. Wolf-Gladrow, 2005: Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, Global Change Biology, 11, 2016-2040.

Lima, I. and S.C. Doney, 2004: A three-dimensional, multi-nutrient, size-structured ecosystem model for the North Atlantic, *Global Biogeochem. Cycles,* **18**, GB3019, doi:10.1029/2003GB002146.

Luo, C., N. Mahowald, and J. d. Corral (2003), Sensitivity study of meteorological parameters on mineral aerosol mobilization, transport and idstribution, *Journal of Geophysical Research*, *108*, 10.1029/2003JD0003483.

Lovenduski, N.S., N. Gruber, S.C. Doney, and I.D. Lima, 2007: Enhanced CO_2 outgassing in the Southern Ocean from a positive phase of the Southern Annular Mode, *Global Biogeochem. Cycles,* **21**, GB2026, doi:10.1029/2006GB002900.

Mahowald, N., C. Luo, J. d. Corral, and C. Zender (2003), Interannual variability in atmospheric mineral aerosols from a 22-year model simulation and observational data, *Journal of Geophysical Research*, *108*, 4352.

Matsumoto, K., J.L. Sarmiento, R.M. Key, J.L. Bullister, K. Caldeira, J.-M. Campin, S.C. Doney, H. Drange, J.-C. Dutay, M. Follows, Y. Gao, A. Gnanadesikan, N. Gruber, A. Ishida, F. Joos, K. Lindsay, E. Maier-Reimer, J.C. Marshall, R.J. Matear, P. Monfray, R. Najjar, G.-K. Platter, R. Schlitzer, R. Slater, P.S. Swathi, I.J. Totterdell, M.-F.

Weirig, Y. Yamanaka, A. Yool, and J.C. Orr, 2004: Evaluation of ocean carbon cycle models with data-based metrics, *Geophys. Res. Lett.,* **31**, L07303, doi:10.1029/2003GL018970.

McClain, C.R., Feldman, G.C., Hooker, S.B., 2004. An Overview of the SeaWiFS Project and Strategies for Producing a Climate Research Quality Global Ocean Bio-optical Time Series. Deep-Sea Research II 51 (1–3), 5–42.

McKinley, G. A., M. J. Follows and J. Marshall, 2004: Mechanisms of air-sea CO_2 flux variability in the equatorial Pacific and the North Atlantic. *Global Biogeochem. Cycles*, 18(2), GB2011, doi:10.1029/2003GB002179.

McKinley, G.A., T. Takahashi, E. Butenhuis, F. Chai, J.R. Christian, S.C. Doney, M.-S. Jiang, K. Lindsay, J.K. Moore, C. Le Quere, I. Lima, R. Murtugudde, L. Shi, and P. Wetzel, 2006: North Pacific carbon cycle response to climate variability on seasonal to decadal timescales, *J. Geophys. Res. Oceans,* **111**, C07S06, doi:10.1029/2005JC003173.

Moore, J.K., S.C. Doney, J.A. Kleypas, D.M. Glover, and I.Y. Fung, 2002: An intermediate complexity marine ecosystem model for the global domain. *Deep-Sea Res., II*, **49**, 403-462.

Moore, J.K., S.C. Doney and K. Lindsay, 2004: Upper ocean ecosystem dynamics and iron cycling in a global 3-D model, *Global Biogeochem. Cycles,* **18**, 4, GB4028, 10.1029/2004GB002220.

Moore J.K., and O. Braucher, Sedimentary and mineral dust sources of dissolved iron to the world ocean, *Biogeosciences, submitted* (a)

Moore J.K., and O. Braucher, Observations of dissolved iron concentration in the world ocean: Implications and constraints for ocean biogeochemical models. *Biogeosciences, submitted* (b).

Naegler, T., P. Ciais, J.C. Orr, O. Aumont, and C. Roedenbeck, On evaluating ocean models with atmospheric potential oxygen, *Tellus 59B (1),138-156*, 2007.

Najjar, R.G., X. Jin, F. Louanchi, O. Aumont, K. Caldeira, S.C. Doney, J.-C. Dutay, M. Follows, N. Gruber, F. Joos, K. Lindsay, E. Maier-Reimer, R.J. Matear, K. Matsumoto, P. Monfray, A. Mouchet, J.C. Orr, G.K. Plattner, J.L. Sarmiento, R. Schlitzer, R.D. Slater, M.-F. Weirig, Y. Yamanaka and A. Yool, 2007: Impact of

circulation on export production, dissolved organic matter and dissolved oxygen in the ocean: Results from Phase II of the Ocean Carbon-cycle Model Intercomparison Project (OCMIP-2), *Global Biogeochem. Cycles*, **21**, GB3007, doi:10.1029/2006GB002857.

Nevison, C.D., N.M. Mahowald, S.C. Doney, I.D. Lima, G.R. van der Werf, J.T. Randerson, D.F. Baker, P. Kasibhatla, and G.A. McKinley, 2008: Contribution of ocean, fossil fuel, land biosphere and biomass burning carbon fluxes to seasonal and interannual variability in atmospheric CO_2, *J. Geophys. Res. Biogeosci.*, **113**, G01010, doi:10.1029/2007JG000408.

Nevison, C.D., N.M. Mahowald, S.C. Doney, I.D. Lima, Variability in air-sea O_2 and CO_2 fluxes and its impact on atmospheric potential oxygen (APO) and the partitioning of land and ocean carbon sinks, *Biogeosciences, submitted*.

Obata, A., and Y. Kitamura, 2003: Interannual variability of the sea-air exchange of CO2 from 1961 to 1998 simulated with a global ocean circulation-biogeochemistry model, J. Geophys. Res., 108, 3337, doi:10.1029/2001JC001088.

Oschlies, A. and V. Garcon, 1998: Eddy-induced enhancement of primary production in a model of the North Atlantic Ocean, *Nature*, **394**, 266-269.

Rasch, P. J., N. M. Mahowald, and B. E. Eaton (1997), Representations of transport, convection and the hydrologic cycle in chemical transport models: Implications for the modeling of short-lived and soluble species, *J. Geophys. Res*., **102**, 18127-28138.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D., and Wang, W. (2002). An improved in situ and satellite SST analysis for climate. Journal of Climate, 15:1609-1625.

Sarmiento, J.L., R.D. Slater, and M.J.R. Fasham, 1993: A seasonal three-dimensional ecosystem model of nitrogen cycling in the North Atlantic euphotic zone, Global Biogeochem. Cycles, 7, 417-450.

Six, K.D., and E. Maier-Reimer, 1996: Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model, *Global Biogeochem. Cycles*, **10**, 559-583.

Smith, R. and P. Gent, 2004: Reference Manual for the Parallel Ocean Program (POP). Ocean Component of the Community Climate System Model (CCSM2.0 and 3.0).

LAUR-02-2484. Los Alamos National Laboratory, Los Alamos, New Mexico.

Stow, C.A. J. Jolliff, D.J. McGillicuddy Jr., S.C. Doney, J.I. Allen, M.A.M. Friedrichs, K.A. Rose, and P. Wallhead, Skill assessment for coupled biological/physical models of marine systems, *J. Mar. Systems, in press*.

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C., Olafsson, J., and Nojiri, Y. (2002). Global sea air CO2 flux based on climatological surface ocean pCO2, and seasonal biological and temperature effects. Deep-Sea Research II, 49:1601–1622.

Takahashi, T. S. C. Sutherland, R. Wanninkhof, C. Sweeney, R. A. Feely, D. W. Chipman, B. Hales, G. Friederich, F. Chavez, A.Watson, D. C. E. Bakker, U. Schuster, N. Metzl, H. Yoshikawa-Inoue, M. Ishii, T. Midorikawa, Y. Nojiri, C. Sabine, J. Olafsson, Th. S. Arnarson, B. Tilbrook, T. Johannessen, A. Olsen, R. Bellerby, A. Körtzinger, T. Steinhoff, M. Hoppema, H. J. W. de Baar, C. S. Wong, B. Delille and N. R. Bates (submitted). Climatological mean and decadal changes in surface ocean pCO_2, and net sea-air CO_2 flux over the global oceans. *Deep-Sea Res. II*.

Wanninkhof, R., 1992, Relationship between wind speed and gas exchange over the ocean, *J. Geophys. Res*., 97(C5), 7373-7382.

Wetzel, P., A. Winguth and E. Maier-Reimer, 2005: Sea-to-air CO_2 flux from 1948 to 2003: A model study. *Global Biogeochem. Cycles*, 19(2), GB2005, doi:10.1029/2004GB002339.

Figure Captions

Figure 1: Schematic of the CCSM Biogeochemical-Ecosystem-Circulation (BEC) model showing the major dissolved inorganic, biological, and detrital tracers and the flows among the tracers. Many of the biological and detrital tracers include multi-element sub-elements to separately track carbon, macro-nutrients, and iron.

Figure 2: Comparison of spatial distribution of observed and simulated annual mean, surface chlorophyll (mg Chl m^{-3}) with CCSM-3 BEC model results (top panel), SeaWiFS satellite ocean color data (middle panel), and model minus data residual (bottom panel).

Figure 3: Comparison of seasonal cycle of observed and simulated surface chlorophyll (mg Chl m^{-3}). Zonal averages of the seasonal anomalies to the zonal mean are displayed (latitude versus month) for CCSM-3 BEC model results (top panel), SeaWiFS satellite ocean color data (middle panel), and model minus data residual (bottom panel).

Figure 4: Comparison of the spatial distribution of interannual variability of observed and simulated surface chlorophyll anomalies from the mean seasonal cycle (mg Chl m^{-3}). Top panel displays the temporal correlation coefficient of CCSM-3 BEC model results against the SeaWiFS satellite ocean color data. The middle and bottom panels display the spatial map of the root mean square of the model and observed anomalies, respectively.

Figure 5: Comparison of observed and CCSM-3 BEC model values for suite of surfaces variables. Zonal averages are displayed (latitude versus month) for data (dashed lines) and model (solid line) for three basins Atlantic (blue), Indian (red), and Pacific (green).

Figure 6: Comparison of observed and CCSM-3 BEC model values for suite of global surfaces variables using Taylor diagrams. Taylor diagrams display in polar coordinates the model-data correlation coefficient (angle from x-axis) and model standard deviation

normalized to observational standard deviation (radius). Diagrams are shown for annual mean spatial distributions (top panel), seasonal anomalies (middle panel), and interannual variability of anomalies from the mean seasonal cycle (bottom panel).

Figure 7: Comparison of observed and CCSM-3 BEC model values for suite of global surfaces variables using Target diagrams. Target diagrams simultaneously display information on normalized model-data biases and unbiased rms differences (see text for more details). Diagrams are shown for the seasonal climatology (top panel) and hindcast including seasonal dynamics and interannual variability (bottom panel).

Figure 8: Comparison of observed and simulated vertical profiles of chlorophyll (mg Chl m^{-3}) for a specific time-series station. Left column displays seasonal average climatology (depth versus month) and right column interannual variability for 1988-2004 (depth versus time) for CCSM-3 BEC model results (top panels), Hawaii Ocean Time-Series data (middle panels), and model-data differences (bottom panels). A dashed line gives model and observed mixed layer depth.

Figure 9: Comparison of observed and CCSM-3 BEC model values for suite of upper ocean variables (0-160 m) for the Hawaii Ocean Time-Series (HOT) site using Taylor diagrams. Taylor diagrams display in polar coordinates the model-data correlation coefficient (angle from x-axis) and model standard deviation normalized to observational standard deviation (radius). Diagrams are shown for annual mean climatological spatial distributions (i.e., vertical profile) (top panel), climatological seasonal anomalies (middle panel), and interannual variability of anomalies from the mean seasonal cycle (bottom panel).

Tables

Table 1: Description of the evaluation data sets used in this study.

Data Set	Variable Name	Temporal Coverage	Reference
Sea Surface Temperature	*SST*	Monthly (1979-2004)	NOAA OI.v2 Reynolds et al. (2002)
Mixed Layer Depth	*MLD*	Monthly Climatology	Boyer-Montégut et al. (2004)
Air-Sea ΔpCO_2 (surface water pCO_2)	ΔpCO_2	Monthly Climatology normalized to 2000	Takahashi et al. (submitted)
Air-Sea CO_2 Flux (ΔpCO_2 and winds)	F_{CO2}	" "	Takahashi et al. (submitted)
Air-Sea O_2 Flux (ocean inverse model)	F_{O2}	Annual & Monthly Clim.	Garcia & Keeling (2001) & Gruber et al. (2001)
Surface Oxygen	O_2	Monthly Climatology	Conkright et al. (2001)
Surface nitrate	NO_3	" "	" "
Surface Phosphate	PO_4	" "	" "
Surface Silicic Acid	$Si(OH)_4$	" "	" "
Surface Chlorophyll SeaWiFS	*Chl*	Monthly (1998-present)	McClain et al. 2004
Vertically Integrated Primary Production	$\int PP$	" "	SeaWiFS & Behrenfeld Falkowski (1997)
Phytoplankton Growth Rate (MODIS)	μ	Monthly (1998-present)	Behrenfeld et al. (2005) & Westberry et al. (2007)
Phytoplankton Carbon Biomass (MODIS)	P_C	" "	" "

Table 2: Globally aggregated model-data skill metrics for the mean annual cycle in the CCSM BEC hindcast.

Variable	Obs. mean	Obs. σ_O	Model mean	Model σ_P	Bias ε_{bias}	Bias (%)	rms error ε_{rms}	corr. coeff. r
SST ($^\circ$C)	18.22	10.29	18.48	9.94	+0.25	+1.4	1.54	0.989
MLD (m)	58.0	37.8	62.3	48.7	+4.3	+7.5	39.2	0.616
pCO_2 (μatm)	357.6	24.4	355.1	29.6	-2.48	-0.7	18.6	0.780
F_{CO2} (mol m^{-2} y^{-1})	0.424	1.65	0.396	1.80	-0.03	-6.6	1.53	0.611
F_{O2} (mol m^{-2} y^{-1})	---	8.34	---	8.33	---	---	5.86	0.753
surf. O_2 (mmol m^{-3})	243.9	60.9	247.2	52.2	+3.3	+1.4	36.9	0.797
surf. NO_3 "	5.12	8.16	6.57	8.95	+1.45	+28.4	3.44	0.923
surf. PO_4 "	0.527	0.553	0.618	0.618	+0.09	+17.2	0.29	0.881
surf. $Si(OH)_4$ "	7.19	13.47	11.09	19.43	+3.91	+54.3	12.38	0.775
Chl (mg m^{-3})	0.285	0.656	0.190	0.173	-0.095	-33.3	0.646	0.185
$\int PP$ (gC m^{-2} mon^{-1})	12.69	13.16	12.50	7.39	-0.19	-1.5	12.80	0.328
P_C (mgC m^{-3})	16.00	15.63	16.27	16.28	+0.27	+1.7	20.27	0.194
μ (d^{-1})	0.55	0.33	0.53	0.48	-0.02	-4.2	0.55	0.116
Geometric statistics using log-transformed data; means converted to geophysical units								
Chl (mg m^{-3})	0.160	---	0.149	---	0.933	---	1.54	0.338
$\int PP$ (gC m^{-2} mon^{-1})	9.71	---	10.01	---	1.032	---	1.32	0.510
P_C (mgC m^{-3})	13.28	---	12.83	---	0.966	---	1.27	0.335

Table 3: Globally aggregated model-data skill metrics for the interannual variability of the monthly deseasonalized anomalies in the CCSM BEC hindcast.

Variable	Obs. σ_O	Model σ_P	rms error ε_{rms}	corr. coeff. r
SST ($^\circ$C)	0.563	0.547	0.510	0.578
Chl (mg m^{-3})	0.266	0.099	0.281	0.032
$\int PP$ (gC m^{-2} mon^{-1})	4.03	2.95	4.75	0.102
P_C (mgC m^{-3})	7.31	8.69	11.25	0.018
μ (d^{-1})	0.20	0.37	0.42	0.010
Geometric statistics using log-transformed data				
Chl (mg m^{-3})				0.126
$\int PP$ (gC m^{-2} mon^{-1})				0.161
P_C (mgC m^{-3})				0.033

Table 4: Aggregated model-data skill metrics for upper-ocean variables (0-160 m) at the Hawaii Ocean Time-Series (HOT) site in the CCSM BEC hindcast.

Variable	Obs. mean	Obs. σ_O	Bias ε_{bias}	Bias (%)	rms error ε_{rms}	σ_P/σ_O	corr. coeff. *r*
SST ($^\circ$C)	22.84	1.60	+0.02	+0.1%	0.66	1.25	0.958
Chl (mg m^{-3})	0.126	0.046	-0.035	-27.8%	0.071	1.48	0.266
PP (mgC m^{-3} d^{-1})	2.50	2.22	-0.83	-33.0%	0.93	0.76	0.923
O_2 (mmol m^{-3})	213.1	4.6	-0.70	-0.3%	3.87	0.70	0.558
NO_3 "	0.73	0.72	+0.35	+48.4%	0.46	1.44	0.921
PO_4 "	0.130	0.053	+0.221	+170%	0.021	0.94	0.921
$Si(OH)_4$ "	2.21	0.88	-1.78	-81.5%	0.66	0.46	0.821

Table 5: Aggregated model-data skill metrics for the interannual variability of the monthly deseasonalized anomalies at the Hawaii Ocean Time-Sereis (HOT) site in the CCSM BEC hindcast.

Variable	Obs. σ_O	Model σ_P	σ_P/σ_O	rms error ε_{rms}	corr. coeff. *r*
SST ($^\circ$C)	0.83	0.36	0.43	0.83	0.209
Chl (mg m^{-3})	0.038	0.013	0.34	0.039	0.073
PP (mgC m^{-3} d^{-1})	1.16	0.47	0.41	1.25	0.004
O_2 (mmol m^{-3})	5.20	3.31	0.64	6.03	0.049
NO_3 "	0.68	0.47	0.69	0.90	-0.207
PO_4 "	0.057	0.033	0.58	0.069	-0.097
$Si(OH)_4$ "	1.22	0.12	0.10	1.22	-0.042

Figures

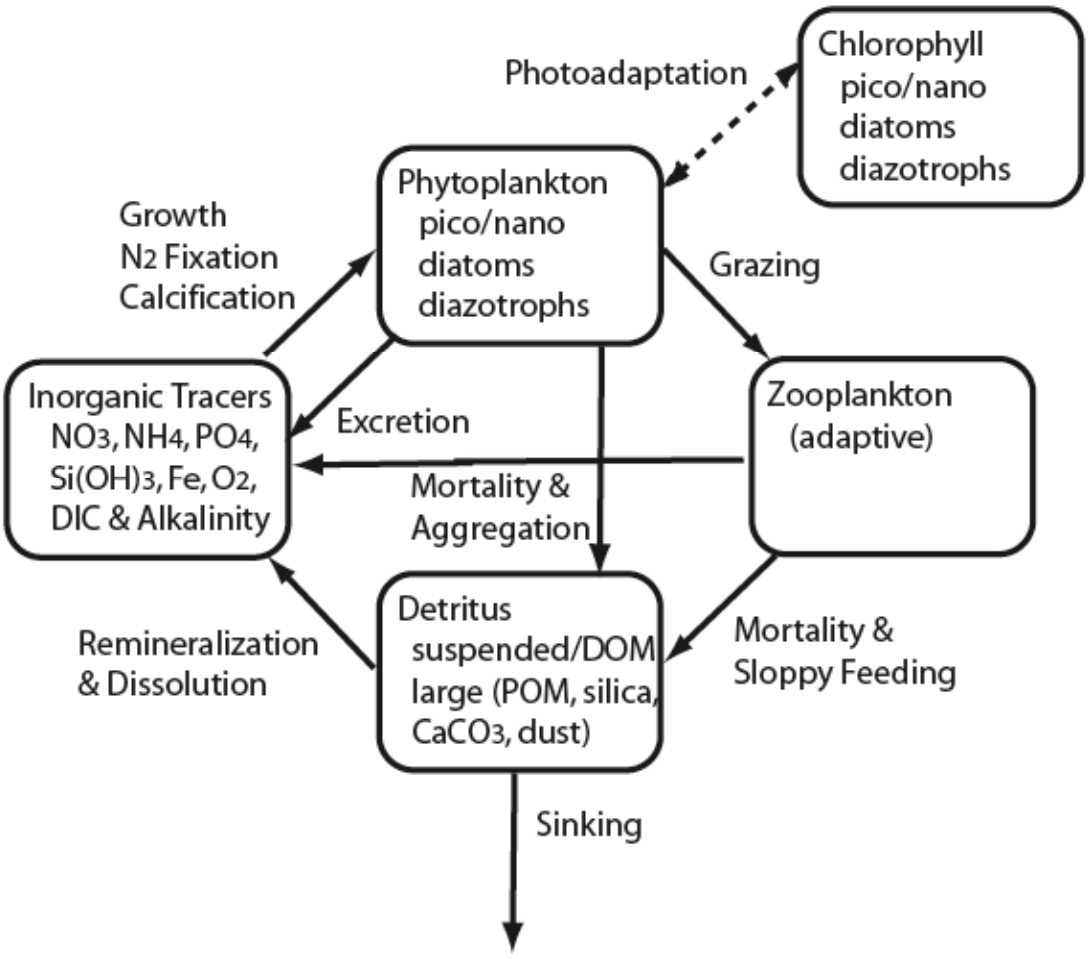


Figure 1: Schematic of the CCSM Biogeochemical-Ecosystem-Circulation (BEC) model showing the major dissolved inorganic, biological, and detrital tracers and the flows among the tracers. Many of the biological and detrital tracers include multi-element sub-elements to separately track carbon, macro-nutrients, and iron.

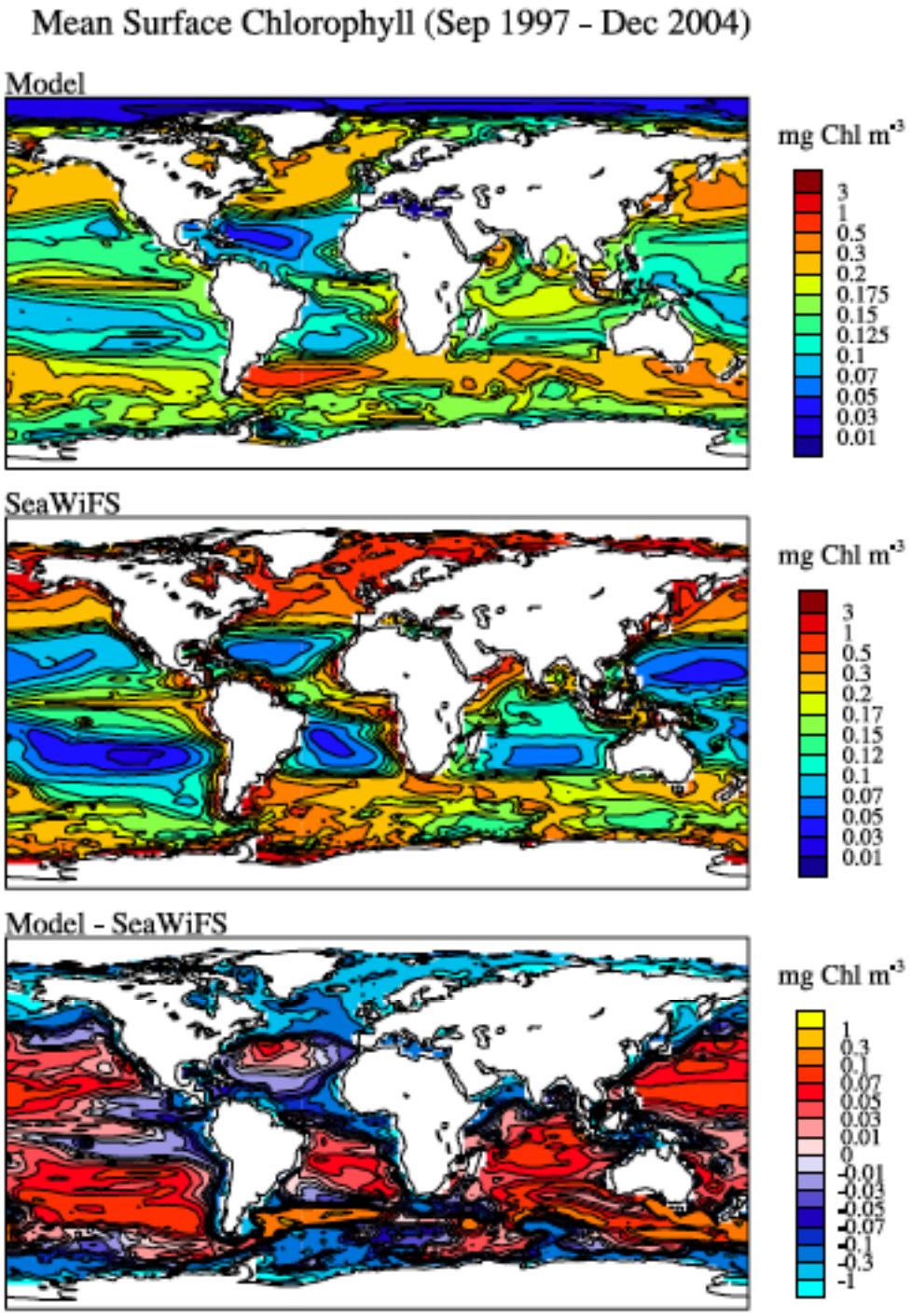


Figure 2: Comparison of spatial distribution of observed and simulated annual mean, surface chlorophyll (mg Chl m^{-3}) with CCSM-3 BEC model results (top panel), SeaWiFS satellite ocean color data (middle panel), and model minus data residual (bottom panel).

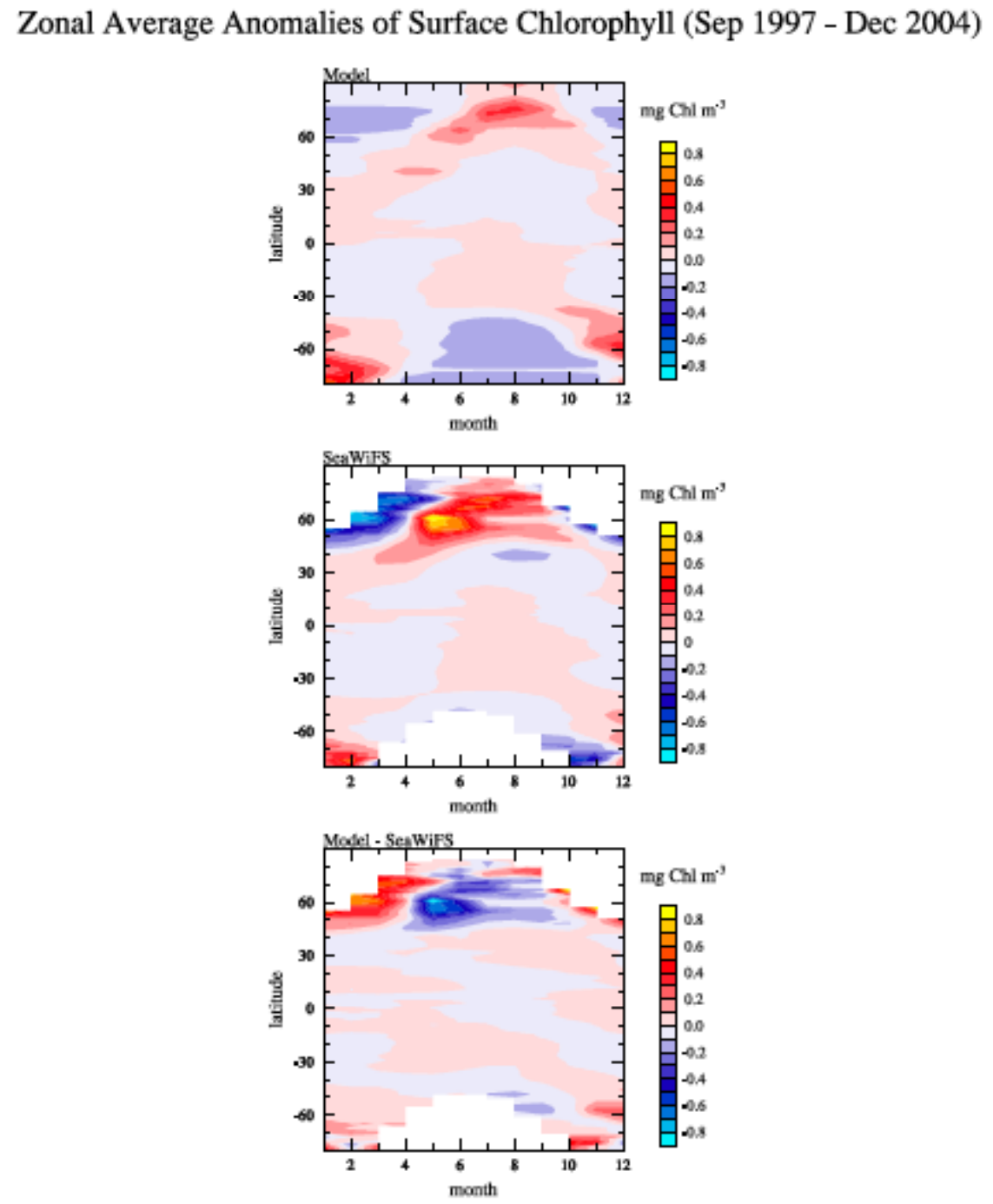


Figure 3: Comparison of seasonal cycle of observed and simulated surface chlorophyll (mg Chl m^{-3}). Zonal averages of the seasonal anomalies to the zonal mean are displayed (latitude versus month) for CCSM-3 BEC model results (top panel), SeaWiFS satellite ocean color data (middle panel), and model minus data residual (bottom panel).

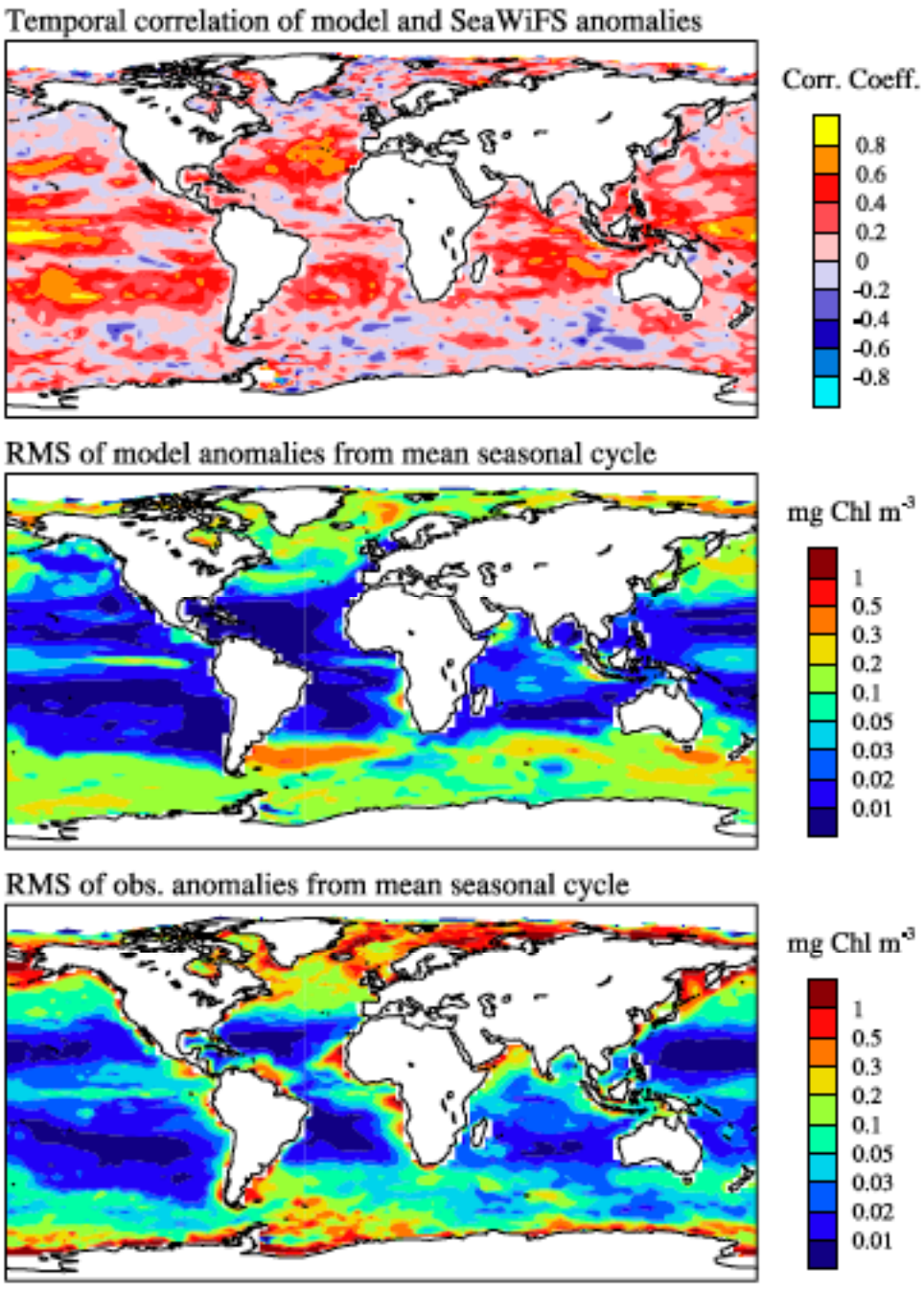


Figure 4: Comparison of the spatial distribution of interannual variability of observed and simulated surface chlorophyll anomalies from the mean seasonal cycle (mg Chl m^{-3}). Top panel displays the temporal correlation coefficient of CCSM-3 BEC model results against the SeaWiFS satellite ocean color data. The middle and bottom panels display the spatial map of the root mean square of the model and observed anomalies, respectively.

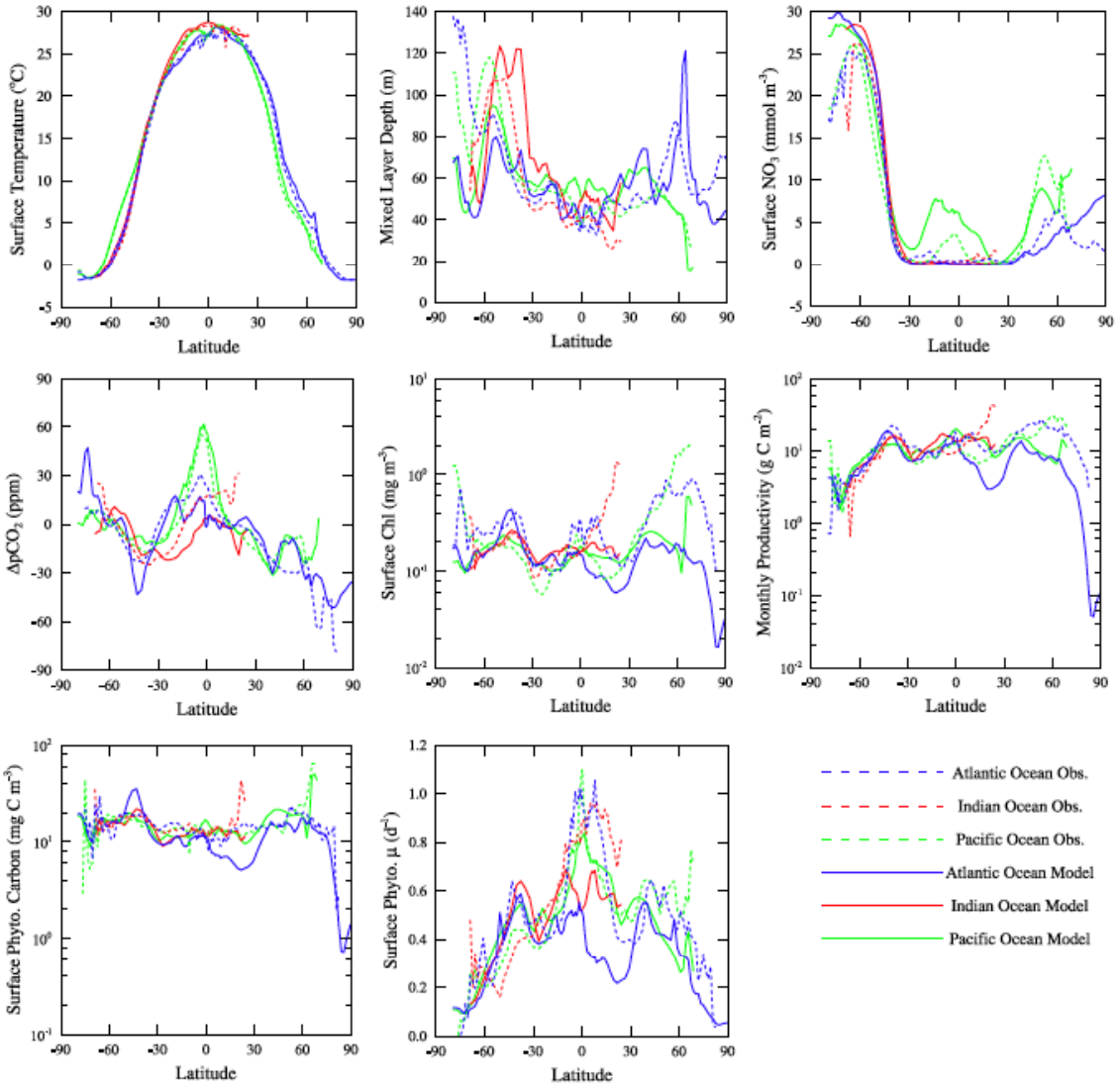
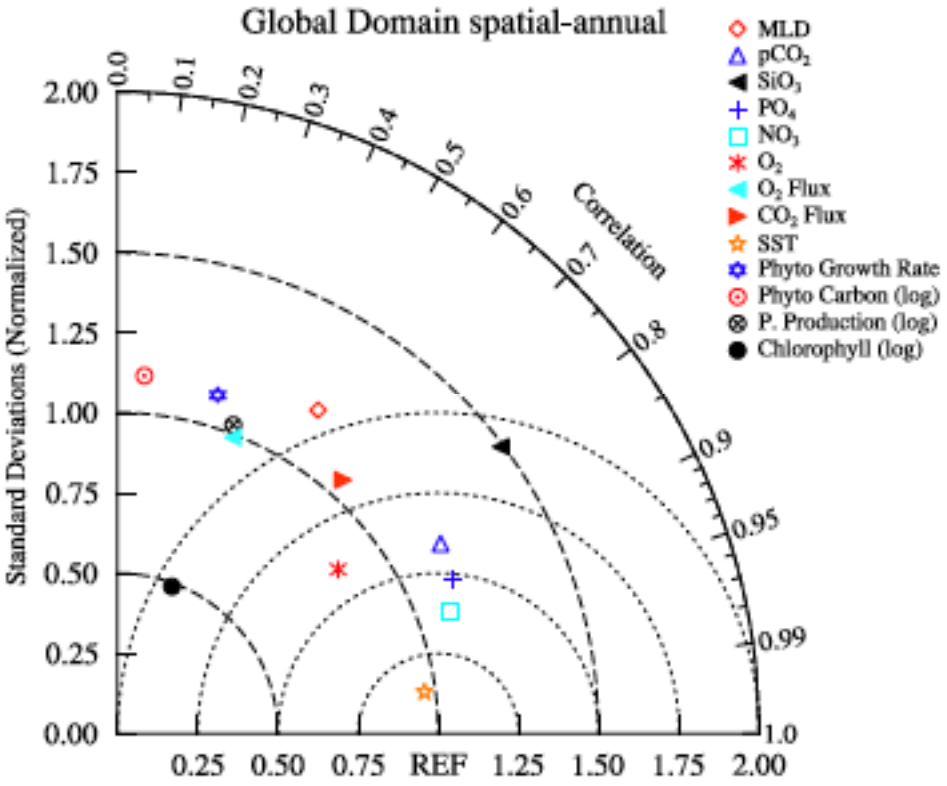
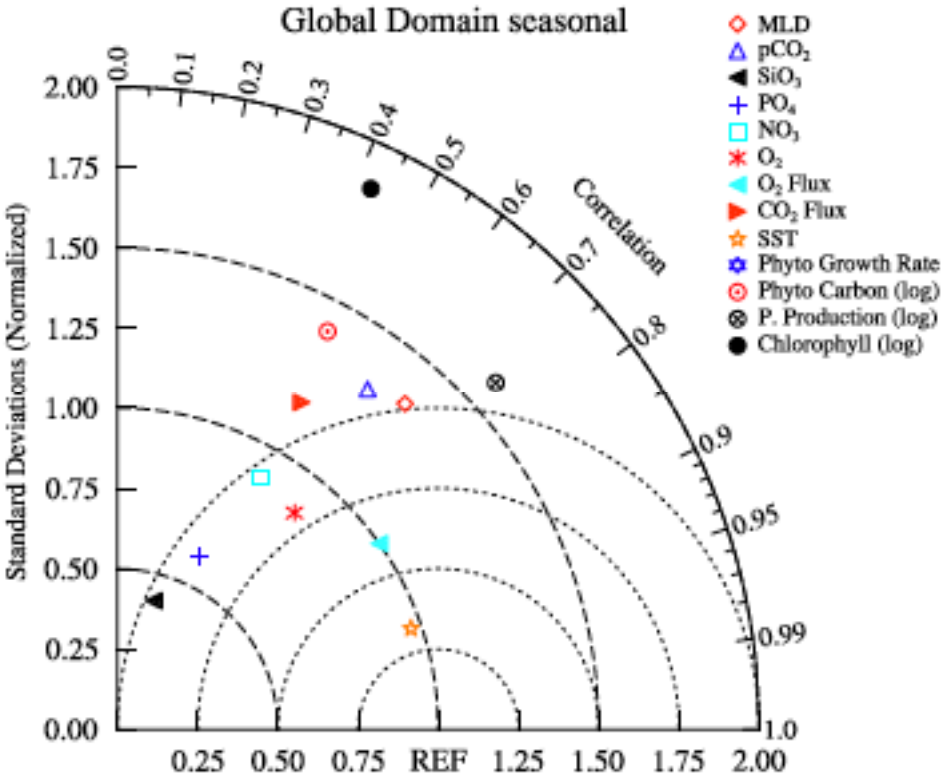


Figure 5: Comparison of observed and CCSM-3 BEC model values for suite of surfaces variables. Zonal averages are displayed (latitude versus month) for data (dashed lines) and model (solid line) for three basins Atlantic (blue), Indian (red), and Pacific (green).


Global Domain spatial-annual
Standard Deviations (Normalized)
Correlation
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.95
0.99
1.0
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
REF
1.25
1.50
1.75
2.00
MLD
pCO2
SiO3
PO4
NO3
O2
O2 Flux
CO2 Flux
SST
Phyto Growth Rate
Phyto Carbon (log)
P. Production (log)
Chlorophyll (log)


Global Domain seasonal
Standard Deviations (Normalized)
Correlation
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.95
0.99
1.0
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
REF
1.25
1.50
1.75
2.00
MLD
pCO2
SiO3
PO4
NO3
O2
O2 Flux
CO2 Flux
SST
Phyto Growth Rate
Phyto Carbon (log)
P. Production (log)
Chlorophyll (log)

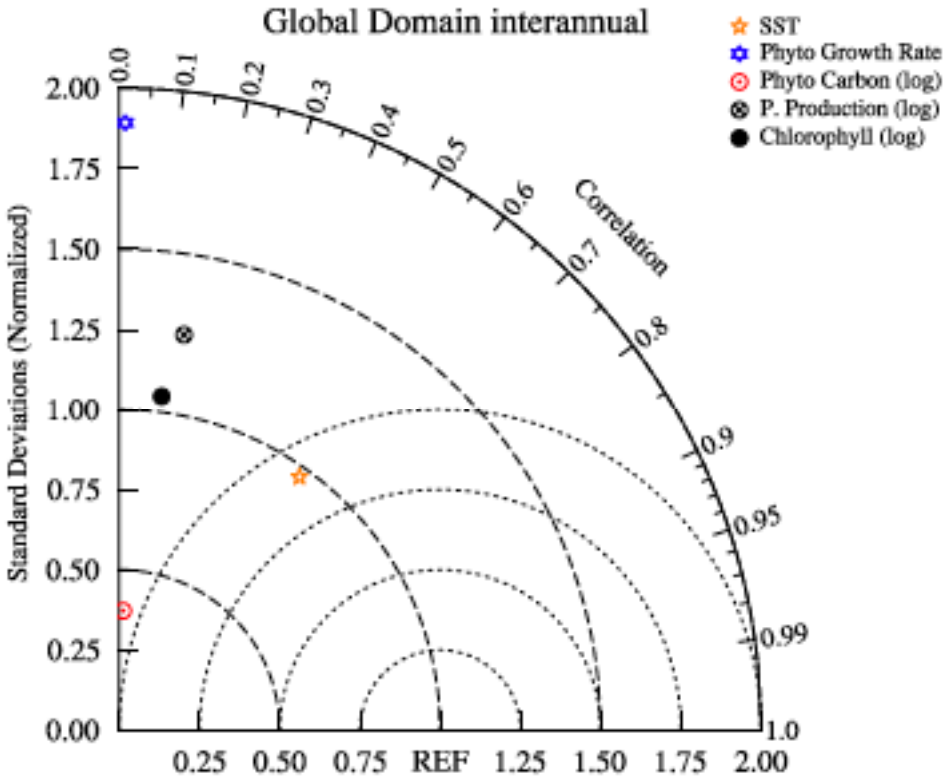
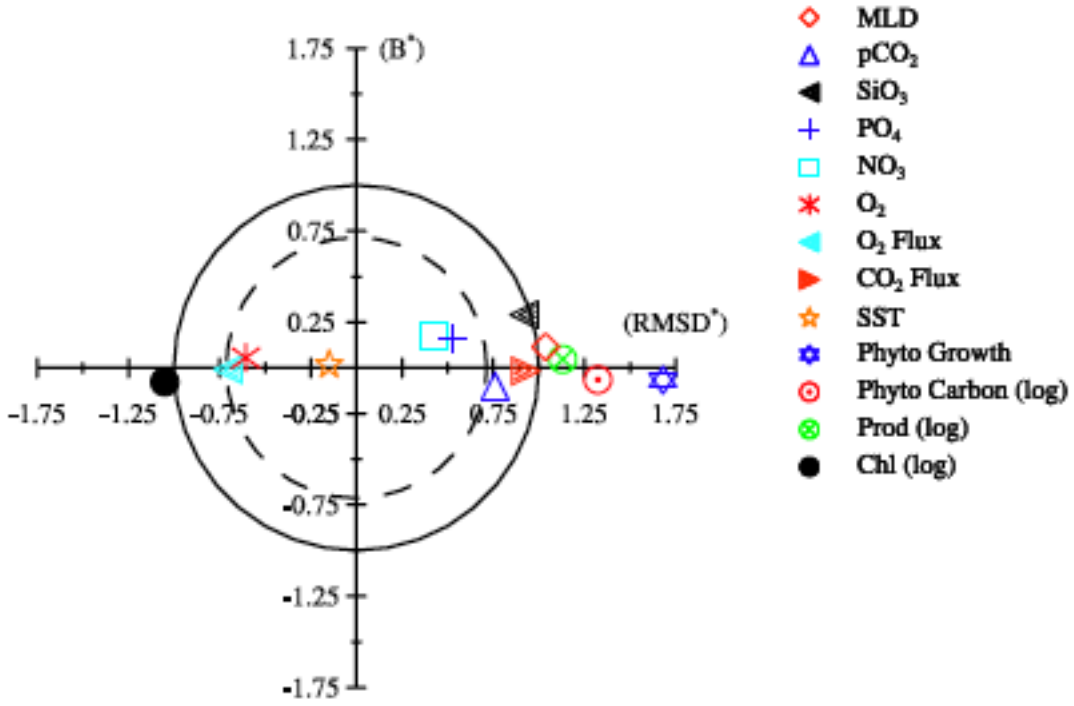


Figure 6: Comparison of observed and CCSM-3 BEC model values for suite of global surfaces variables using Taylor diagrams. Taylor diagrams display in polar coordinates the model-data correlation coefficient (angle from x-axis) and model standard deviation normalized to observational standard deviation (radius). Diagrams are shown for annual mean spatial distributions (top panel), seasonal anomalies (middle panel), and interannual variability of anomalies from the mean seasonal cycle (bottom panel).

Total Space-Time (Seasonal Climatology)



Total Space-Time (Interannual)

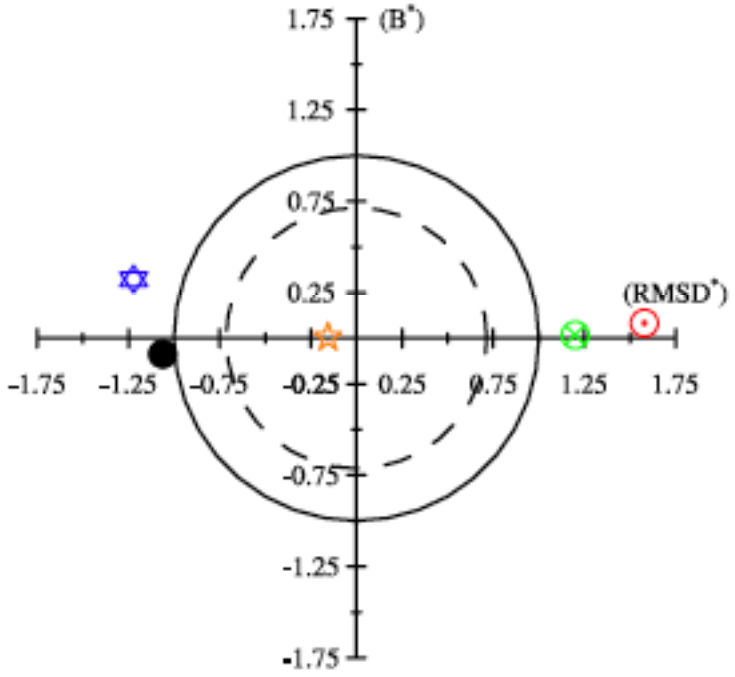


Figure 7: Comparison of observed and CCSM-3 BEC model values for suite of global surfaces variables using Target diagrams. Target diagrams simultaneously display information on normalized model-data biases and unbiased rms differences (see text for more details). Diagrams are shown for the seasonal climatology (top panel) and hindcast including seasonal dynamics and interannual variability (bottom panel).

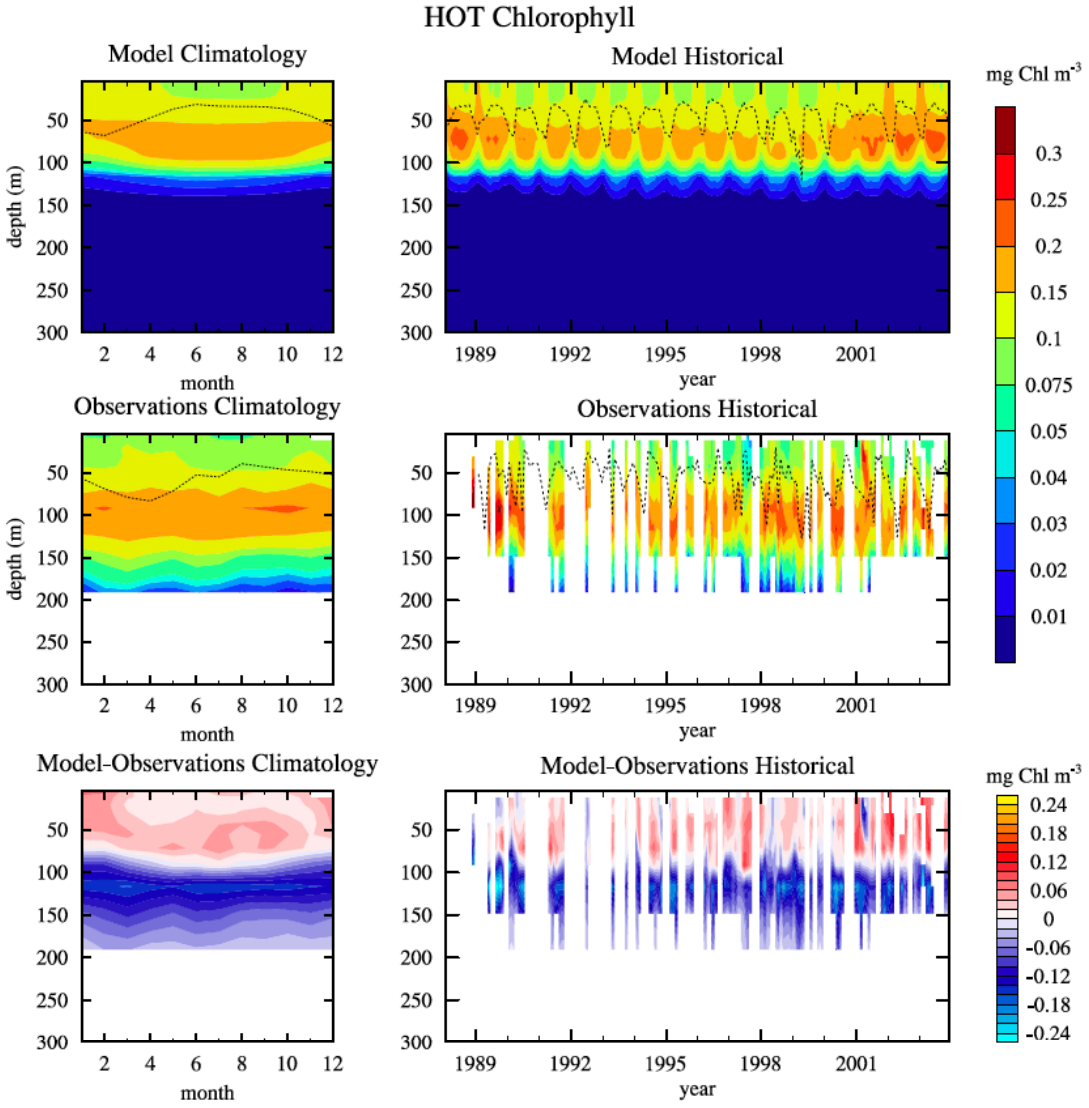
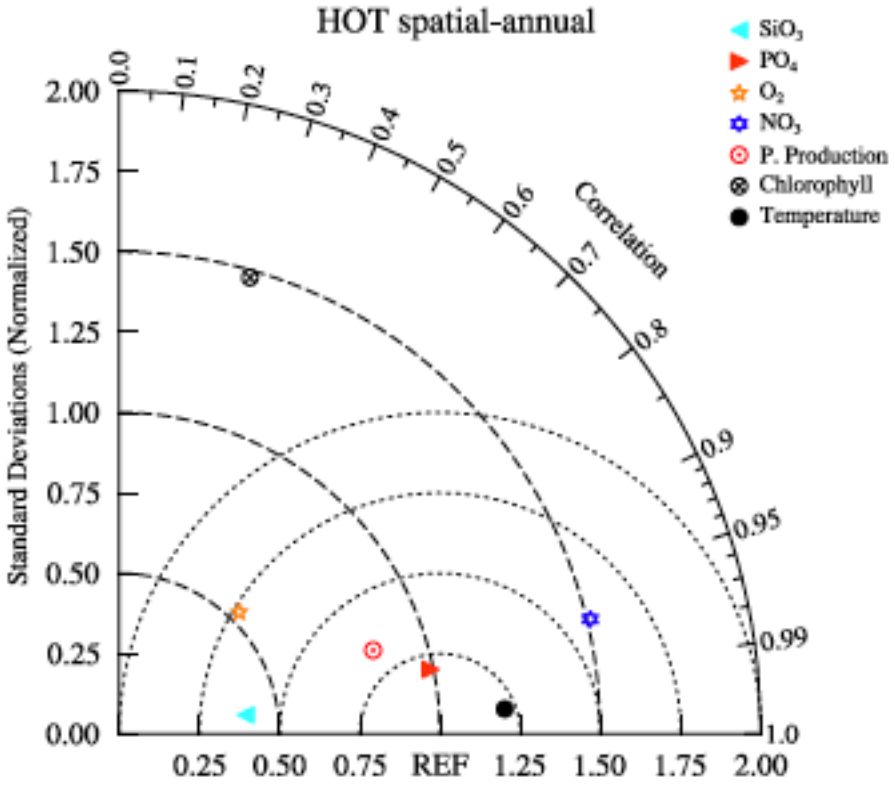
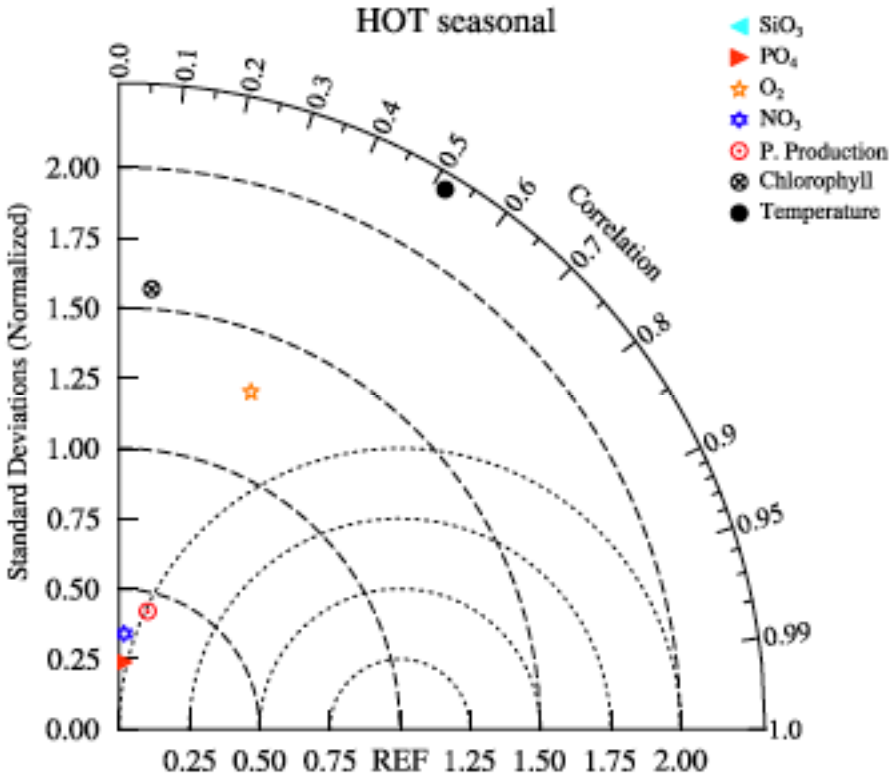


Figure 8: Comparison of observed and simulated vertical profiles of chlorophyll (mg Chl m^{-3}) for a specific time-series station. Left column displays seasonal average climatology (depth versus month) and right column interannual variability for 1988-2004 (depth versus time) for CCSM-3 BEC model results (top panels), Hawaii Ocean Time-Series data (middle panels), and model-data differences (bottom panels). A dashed line gives model and observed mixed layer depth.


HOT spatial-annual
SiO3
PO4
O2
NO3
P. Production
Chlorophyll
Temperature
Correlation
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.95
0.99
1.0
Standard Deviations (Normalized)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
0.25 0.50 0.75 REF 1.25 1.50 1.75 2.00


HOT seasonal
SiO3
PO4
O2
NO3
P. Production
Chlorophyll
Temperature
Correlation
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.95
0.99
1.0
Standard Deviations (Normalized)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
0.25 0.50 0.75 REF 1.25 1.50 1.75 2.00

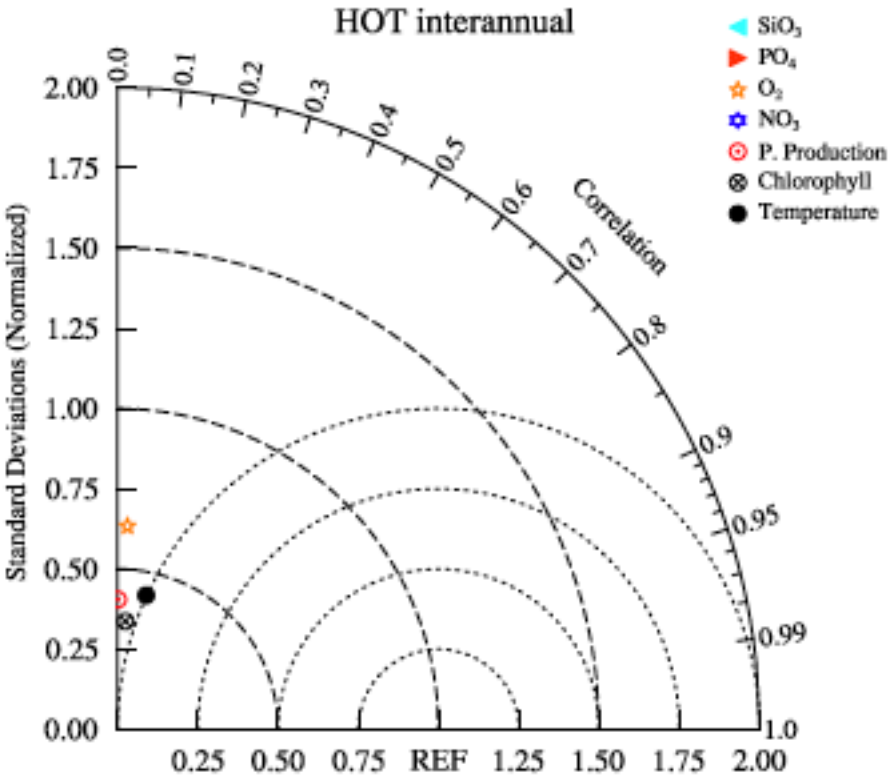


Figure 9: Comparison of observed and CCSM-3 BEC model values for suite of upper ocean variables (0-160 m) for the Hawaii Ocean Time-Series (HOT) site using Taylor diagrams. Taylor diagrams display in polar coordinates the model-data correlation coefficient (angle from x-axis) and model standard deviation normalized to observational standard deviation (radius). Diagrams are shown for annual mean climatological spatial distributions (i.e., vertical profile) (top panel), climatological seasonal anomalies (middle panel), and interannual variability of anomalies from the mean seasonal cycle (bottom panel).

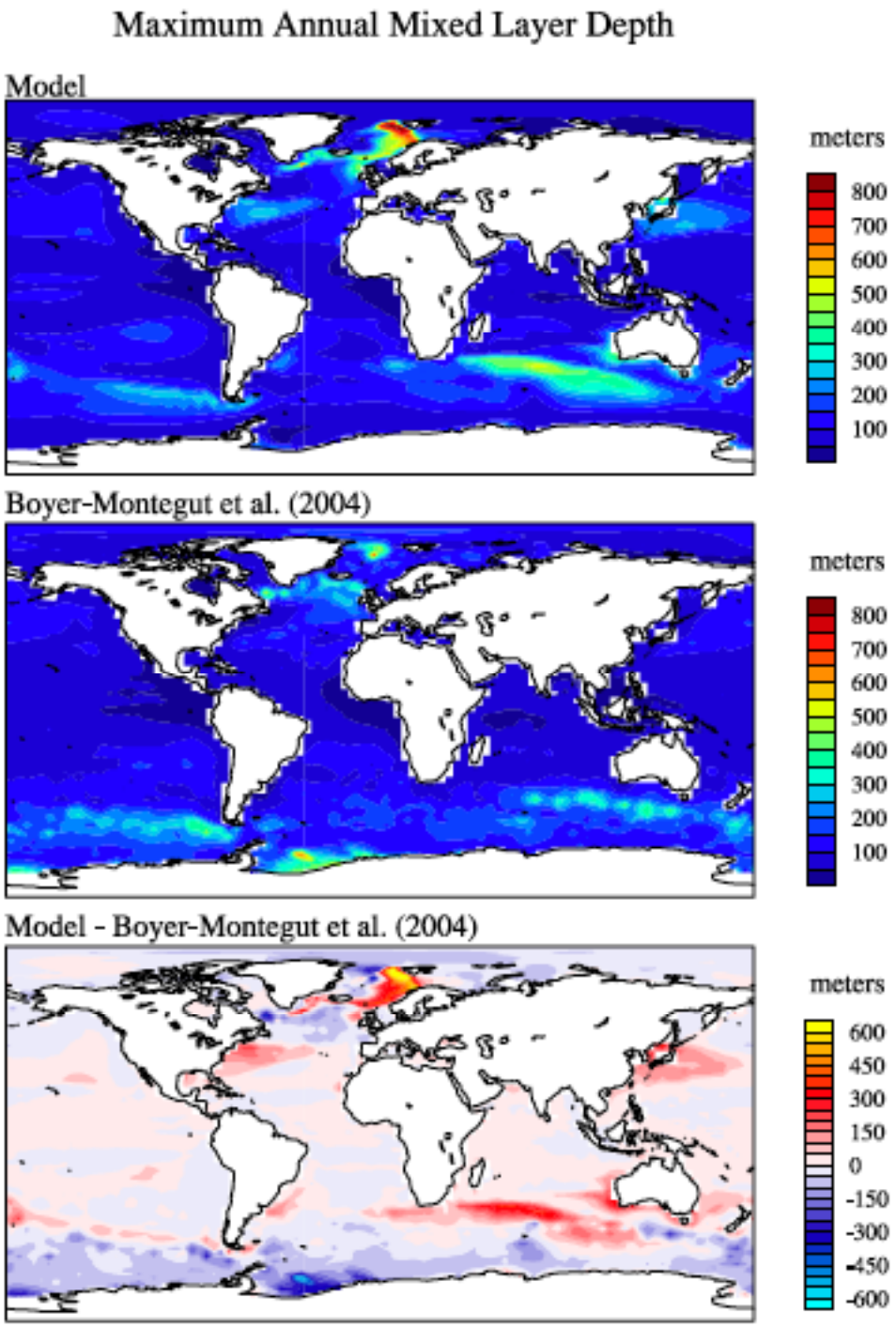


Figure 10: Comparison of spatial distribution of observed and simulated maximum winter mixed layer depth (m) with CCSM-3 BEC model results (top panel), Boyer-Montegut et al. (2004) field data (middle panel), and model minus data residual (bottom panel).