

## ***Interactive comment on “Inconsistent strategies to spin up models in CMIP5: implications for ocean biogeochemical model performance assessment” by R. Séférian et al.***

**F. Joos (Referee)**

joos@climate.unibe.ch

Received and published: 23 November 2015

This is a nice and timely paper that addresses an important issue – model drift. It reflects the authors' broad knowledge in the field of coupled modelling. The authors show that short spin-up simulations initialized with observations lead to a too optimistic error statistic and biases model ranking. The authors also make proposal how model drift may be accounted for in future model assessments. This is an important and original contribution to the field.

I recommend publication after the following comments have been addressed

C3026

1) I am concerned about the way the drift model is presented and introduced and that the drift model may be used inappropriately in future work. The authors apply an exponential model with a single relaxation time scale to approximate the evolution of drift. However, the application of a single time scale is most likely not appropriate to determine the drift in whole ocean RMSE or other global error statistics. For example, this is implicitly demonstrated by the results in section 3.5 where the authors apply the drift models for different depth levels individually and show that time scales are different between depth levels.

In my opinion, the following point should be made very clear in this manuscript and in the method, results and discussion/conclusion section:

different relaxation time scales apply for different regions (and variables). This requires that the drift in RMSE and other metrics is to be determined for different regions or even for different grid boxes individually before the drift in RMSE for the whole ocean is to be determined. In this way, multiple time scales would be applied to estimate the evolution of whole ocean RMSE and to correct error statistics for drift.

2) I am not convinced that selecting depth levels as regions is a good approach. For example, drift at 2000 m in the well-ventilated North Atlantic Deep Water may be quite different from drift in the slowly ventilated North Pacific.

It would be illustrative to compute the relaxation time scale,  $\tau$ , individually for each grid cell and plot  $\tau$  along sections in the Atlantic, Pacific and Indian (or similar). A grid-cell based approach is generally also applied when removing model drift from projections by using a control simulation. Computing  $\tau$  for individual grid cells would be comparable with such an approach.

Further comments:

---

1) A sufficiently long spin up over several hundred years is a prerequisite to estimate

C3027

drift in error statistics and other variables. (High-resolution) models that are initialized with observed fields and not spun-up over several centuries very likely suffer from serious drift problems. It may be very difficult to estimate the future evolution of the drift from a short spin-up. This should be mentioned explicitly in the manuscript. (May be this could even be quantitatively illustrated by estimating relaxation time scales from an initial period, e.g., first 50 or 100 yr as compared to time scales from the last 100 year of the simulation as already presented for three different depth levels.)

2) The authors may also note that rate of drifts (e.g. in the surface) may increase when the mode of model operation is changed, e.g. from prescribed atmospheric CO<sub>2</sub> to freely simulated atmospheric CO<sub>2</sub>.

3) The authors do hardly evaluate the validity of their exponential model. It would be nice if this model could be validated, e.g. in the context of a millennium long control simulation or similar?

Sec 3.6: It is not entirely clear whether the same time scale is applied here across all models considered. Please make this clear. It is also not clear whether different time scales are used for different depth levels. Please clarify.

Sec 3.2: I am somewhat confused here about the role of river outgassing. The clarity of the text should be increased. It is not readily clear whether the model should actually achieve a flux of 0 GtC/yr or an outgassing of ~0.4 to 0.6 GtC/yr at equilibrium.

8767, line 17: additional compared to what?

8778 line 24: conclusion: Is it sufficient to report the drift in global RMSE? Perhaps this clause should be deleted or refined.

---

Interactive comment on Geosci. Model Dev. Discuss., 8, 8751, 2015.