We want to thank the reviewers for their constructive comments and questions pertaining to the manuscript. We believe that addressing their comments and questions can serve to strengthen the manuscript, and we would very much like to resubmit a version of the manuscript that incorporates these suggestions. Detailed below are the ways in which we plan to respond to their comments.

Before progressing to the comments of the reviewers, we would like to first mention two changes that we propose to include pertaining to related publications. First, we will reference the newly published Biogeosciences Discussions study by Laufkötter et al. (2015), as in our previous study this was merely references as "personal communications" with the lead author. We also propose to change subsection title 3.5 to "Time of emergence in individual multiple ecosystem drives", as we chose not to consider combined Time of Emergence diagnostics in this study.

Additionally, following the original comments of the editor before the manuscript was circulated for review, we have proposed a small subsection of the Discussion section where we address the implications for coral reef regions (see below in the response to Reviewer #2).

Anonymous Referee #1

Major Comments

I was confused when I read the introduction and methods as to why the authors decided to use a 30-year timescale for their trend emergence analysis. Do the cited observational studies on page 18192 show that the time series needs to be at least 30 years long? Likely not… Based on Chla data, Henson et al. 2010 (Biogeosciences) suggest that a time-series needs to be at least 40 years long in order to e able to detect secular trends in biological production. In the results presented in this manuscript, the authors test whether 30 years are enough to detect a secular trend in each driver individually. They also show that 10-year windows are not enough to detect a secular trend, especially in certain regions that are strongly driven by large interannual/multidecadal atmospheric patterns. To avoid confusion, it would be good to make this clearer on page 18195 – the last sentence does mention the 10-yr window test, but I only realized that after rereading the methods several times. I also think that the results of these tests are important findings for oceanographers and should be mentioned in the abstract.

We appreciate the comments of the reviewer. Following their suggestion, we have created a separate paragraph to clarify this point. The text for the paragraph would be the following:

Additionally, we also consider the sensitivity of the confidence intervals to the choice of the width of the trend window. To that end, confidence intervals are also considered for the case of a 10 yr window. From the spectral SST characteristics of the underlying coupled (atmosphere/ocean) model, it has been shown in Fig. 2 of Wittenberg (2009) and Fig. 7 of Dunne et al. (2012; Journal of Climate) that SST variability is more pronounced over 10 yr timescales than over 30yr timescales. Thus one may well expect that the signal-to-noise ratio characteristics for the ecosystem drivers reflect these underlying dynamical drivers of variability, at least in the equatorial Pacific. Thus our sensitivity analysis is intended to offer insight into both of our primary interests described in the Introduction, namely identification of perceptible changes for ecosystems and optimization of the observing system. We've followed the reviewer's suggestion in improving the clarity of presentation on this point in the main text, but we are of the opinion that including mention of this sensitivity analysis in the abstract would detract from the main points of the study.

Minor Comments

Page 18194, Line 7: snapshots

This will be fixed.

Page 18194, Line 11: perturbations

This will be fixed.

Page 18199, Line 3: Delete "a" and change "drive" to "drivers"

These will be changed.

Page 18216: ..has been used to calculate trends (erase calculated)

We will restructure the entire sentence for the sake of clarity.

Anonymous Referee # 2

Major Comments

I find the paper to be too narrow in the discussion of its conclusions. The authors describe statistical properties without any attempt to link them to the mechanisms underlying the variability of ocean ecosystems. For instance, there is an interesting conclusion about the early emergence of SST in the tropics as opposed to oxygen which first manifests itself in the Southern Ocean. There is no attempt to explain or link this conclusion to the main features of ocean dynamics in these regions. Why are they low in the tropics? What drives the difference?

The reviewer makes a very good suggestion, in proposing that a mechanistic account should be provided to facilitate interpretation of the main results of this study. To this end, we propose that the following three paragraphs be added to the Discussion section:

Although our analysis has been focused on statistical questions (namely confidence intervals and time of emergence diagnostics), it is also important to consider the mechanisms that control emergence timescales. The most important contrast seen in our results was that in evidence in Fig. 2, namely the early (late) emergence of SST in the tropics (Southern Ocean), and the late (early) emergence of O_2 inventories in the tropics (Southern Ocean) with the 30 yr window. For SST, the contrast between the tropics and the Southern Ocean in Fig. 2 with a 30 yr window is largely reflecting the weakness of the SST trend over the Southern Ocean relative to the tropics. In fact, the contrast between the tropics and the Southern Ocean is more generally representative of most of the rest of the global surface ocean (except the northern North Atlantic) relative to the Southern Ocean (Fig. A1c). The lack of sea surface temperature warming reflects large-scale interhemispheric asymmetries in the mean ocean circulation. The strong upwelling in the Southern Ocean nearly anchors sea surface temperature at preindustrial levels (Manabe et al. 1991; JC, Marshall and Speer 2012; Nature Geosciences, Frölicher et al. 2015; Journal of Climate).

For the tropics, the secular trend is sufficiently large over a 30 yr window to be more important than the natural decadal variability, but consistent with the spectral characteristics of ENSO for the underlying physical model (Wittenberg, 2009) this is no longer true for the case of a 10 yr window in the tropics. For the case of the O_2 inventories, the reverse holds. In the Southern Ocean, deoxygenation is much larger than natural variability due to the stratification-induced reduced supply of oxygen from the surface into the thermocline (Frölicher et al. 2009, Gnanadesikan et al. 2012). In contrast, almost no O_2 changes are projected to occur in the low O_2 regions of the tropical and subtropical thermocline owing to a reduced O_2 demand because of lower biological production and export of organic matter in the overlying near-surface waters (Ganandesikan et al. 2012; Steinacher et al. 2010; Biogeosciences). These biological drivers are expected to be modulated by perturbations to the rates of ocean interior and thermocline ventilation. However, the confidence of the O_2 projections in the low latitudes is low, because the GFDL ESM2M has serious limitations in its ability to correctly simulate today's observed O_2 distribution (Gnandesikan et al. 2012; Biogeosciences), a common feature among the current Earth System Models (Bopp et al. 2013; Biogeosciences).

"Observing system" first mentioned in the abstract and further throughout the text: it looks like an afterthought dropped into the text at a later stage. It might be an important goal, but it is not explained properly. What is this observational strategy/system supposed to observe/achieve/demonstrate? Globally? Regionally? Selectively in some hotspots? I can guess it should relate to the emergence, but how and why is left to the reader to deduce. I suggest either removing all references to it or explaining properly and the dedicating some discussion to more clear recommendations for such a system following conclusions of the study.

We agree with the reviewer that the implications of our study for "observing system design" were not sufficiently explained. For the Discussion section of our manuscript, this concerns page 18205, lines 8-21. In order to better clarify this point, we propose splitting into two separate subsections as follows:

Discussion Subsection: Implications for Observing System Design:

It is also important to consider the implications of our study for optimization of the global ocean observing system. With this goal in mind, our study can be considered as an Observing System Simulation Experiment (OSSE). With an OSSE, one considers a model to be an analog for the real ocean, for which one has the fully resolved state evolution to round-off error. Earlier OSSEs (Christian et al., 2008; Park et al., 2010; Plancherel et al., 2013; Majkut et al., 2014) have tended to focus on one realization of the evolution of the Earth system, and focused on the skill with which different observing strategies can reproduce variability in the system through selective sub-sampling of the model output. The target is to test skill in reproducing the real-world trends and variability with an incomplete observing system, without any claim to separating the signal associated with the secular trend and natural variability.

For our experimental configuration considered as an OSSE, we address a different but complementary question. We consider the case where the observing system has perfect skill in reproducing the trends and variability of the system of interest (the Earth System), but where the target is to identify the secular trend. With observations alone it is not possible to deconvolve the secular trend from natural variability that is present in the system (Deser et al., 2014). It is precisely this deconvolution that we address with the Large Initial-Condition Simulations with the Earth System Model, thereby building on the previous analyses considered with fewer ensemble members (Frölicher et al., 2009; Christan, 2014). The question we address is then as follows: Given an observing system with perfect skill that allows us to perfectly monitor the evolution of the system, how many years of measurements do we need to distinguish between the secular trend and natural variability? Our main result is that sustained decadal measurements will be needed even for the idealized case of a perfect observing system.

Viewed in this way, our main results point to the importance of maintaining a sustained multi-decadal observing system for ocean biogeochemistry and ecosystem drivers. For the four drivers considered here, the confidence intervals found with a 30yr window for calculating trends (Fig. 2) are significantly higher than those found with a 10yr window (Fig. 5). For the case with a 10yr window, even Ω_{arag} reveals expanses of non-emergence over the recent decade (2005-2014). This is in evidence, for example, over important parts of the Coral Triangle biodiversity hot-spot spanning the Indo-Pacific warm pool region, as well as for the North Atlantic. This underscores the potential importance of sustained multi-decadal measurements in order to identify the rate of acidification associated with the secular trend in these regions.

More generally, our analysis of confidence intervals for emergence for two versus four drivers (Fig. 4e) largely highlights the combined effects of Ω_{arag} and SST in the tropics. This implies that even with high resolution of temporal and spatial scales, a sustained multi-decadal (30yr) observing system of the type considered by Ishii et al. (2009) through the western Equatorial Pacific is needed to detect the anthropogenic trend against the background noise of natural variability.

Discussion Subsection: Implications for perceptible changes in ocean drivers:

It is worth noting as well here that given that Ω_{arag} and SST are chosen by the method to be the dominant drivers in the tropics, and that in the two driver analysis this is the region of earliest emergence, that coral reef habitats may be expected to be the primary regions currently experiencing perceptible changes relative to the background variability to which they are adapted. This could be of particular importance for "The Coral Triangle" in the Pacific Warm Pool region, a region of maximum biodiversity that is considered to include at least 75% of the world's coral species (Ministry of Marine Affairs and Fisheries, 2008).

It may well be that the mechanisms driving warming and acidification over the warm pool are distinct. The degree of tropical warming has been argued to be constrained by coupled interactions between the ocean and the atmosphere (Ramanathan and Collins, 1991; Pierrhumbert, 1995; Clement et al., 1996). For the rate of acidification, a number of insights were offered in the study of Ishii et al. (2009). There was demonstrated that the changes in dissolved inorganic carbon (DIC) are not driven by local air-sea fluxes, but must result from water being advected in from elsewhere. This difference between the principal climate dynamical drivers of acidification (DIC increases) and warming reflects differences in the air-sea equilibration timescales for carbon (of order one year) and temperature (of order one month).

Similarly to the first reviewer I was not satisfied by the discussion of the choice of 30 years.

See our response above to the first reviewer on this point. There we explain the new paragraph that has been added to the text to clarify this point.

Minor Comments

Abstract (last sentence): Risk assessment of what? Mitigation strategies of what?

We will remove from the abstract the phrase pertaining to risk assessment and mitigation strategies.

Page 18191, Line 5: Remove possibly

This will be fixed.

Page 18191, Line 17: Loss of oxygen is also caused by other factors

We will add an additional sentence to clarify this point.