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 have responded to their comments.

Please note that all **changes in the BGD version of the manuscript** (including the abstract) are highlighted in red text, to facilitate the editing process. Additional blue text is intended to address the last page of comments by the second reviewer as we mistakenly didn't address these in our initial set of responses. We apologize for any confusion that this may have caused.

Before progressing to the comments of the reviewers, we would like to first mention two changes that we have made 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 have also changed subsection title 3.5 to "Time of emergence in individual multiple ecosystem drivers", 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 included a small subsection of the Discussion section where we address the implications for coral reef regions (see below in the response to Reviewer #2).

More broadly, we have now divided the Discussion section into subsections, and added a new Conclusions section.

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 be 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. 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. Following their suggestion, we have created a separate paragraph. 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.

Minor Comments

Page 18194, Line 7: snapshots

This has been fixed.

Page 18194, Line 11: perturbations

This has been fixed.

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

These have been changed.

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

We have restructured the entire sentence for the sake of clarity.

Anonymous Referee # 2

Major Comments

The paper is dedicated to the statistical analysis of ensemble simulations of an ESM. The focus is on a topic of emergence of climate stressors (or drivers) of ocean ecosystems. The approach and results are well written, with some interesting results, but 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. The same can be said about multiple driver emergences. Why are they low in the tropics? What drives the difference? Such a narrow focus on statistical characteristics is especially surprising when the authors aim to make their results useful for the optimization of the observational strategies.

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 have added three three paragraphs 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 is 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** identified with a 30 yr window is largely reflecting the small amplitude 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 in the northern North Atlantic) relative to the Southern Ocean (**Fig. A1c**). The lack of SST warming reflects large-scale interhemispheric asymmetries in the mean ocean circulation. The strong upwelling in the Southern Ocean nearly anchors sea surface temperature at pre-industrial level (Stouffer et al. 1989; Marshall and Speer, 2012; Frölicher et al., 2015).

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 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 O_2 from the surface into the ocean interior (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. This is due to the fact that reduced biological production and export of organic matter in the overlying near-surface waters are then associated with reduced O_2 demand in the ocean interior (Gnanadesikan et al., 2012; Steinacher et al., 2010). 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, as the GFDL ESM2M has biases in its representation of today's observed O_2 distribution (Gnanadesikan et al., 2012), a feature common to the current generation of Earth System models (Bopp et al., 2013).

Further in the Discussion and conclusion section the authors state that "the temporal and special characteristics of emergence should be model-dependent". If this is true, trying to explain or at least discuss emergent statistical properties in a view of the features of underlying dynamics is especially important. Otherwise what is the value of conclusions, which would change, in the next model? I suggest a thorough reworking and expanding the discussion section in respect to the criticism above.

Again, we agree fully. The new subsection of the Discussion section (Section 4.2) addresses the concerns of the reviewer. Note that at the end of this section we refer to the inter-model comparison study of Bopp et al. (2013), which considers differences between the CMIP5 models.

"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 have prepared 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 of the system. Earlier OSSEs (Christian et al., 2008; Park et al., 2010; Plancherel et al., 2013; Majkut et al., 2014; Cassar 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 Earth System through selective sub-sampling of the model output. The target is to test the available 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 on all timescales of the system of interest, but where the target is to identify the secular trend. 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; Christian, 2014). The question is then as follows: Given an observing system with perfect skill that allows one to perfectly monitor the evolution of the system, how many years of continuous measurements are needed to identify the secular trend above the noise of background variability? Our main result with this question 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 mult-decadal observing system for ocean biogeochemistry and ecosystem drivers. For the four drivers considered here, the confidence intervals found with a 30 yr window for calculating trends (**Fig. 2**) are significantly higher than those found with a 10 yr window (**Fig. 5**). For the case with a 10 yr window, even W_{arag} reveals broad expanses of non-emergence over the 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 continuous measurements in order to identify the rate of acidification associated with the secular trend in these regions.

More generally, our analysis of conficence intervals for emergence for two versus four drivers (**Fig. 4c** for the period 2005-2014) largely highlight the combined effects of W_{arag} and SST in the tropics. This implies that even with high resolution of temporal and spatial scales, a sustained multi-decadal (30 yr) observing system of the type considered by Ishii et al. (2009) in western Equatorial Pacific surface waters is needed to detect the rate of the secular trend in acidification against the background noise of natural variability with confidence.

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 have removed from the abstract the phrase pertaining to risk assessment and mitigation strategies.

Page 18191, Line 5: Remove possibly

This has been fixed.

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

We have added an additional sentence to clarify this point.

Page 18192, Line 19: What is "comprehensive"?

We have removed this word, both from the Abstract and from the location cited by the reviewer.

Page 18193, Lines 1-2: I'm afraid I don't understand this sentence

The confusing sentence has been removed.

Page 18194: Statement that ENSO is the most pronounced driver of decadal variability. Reference?

A reference to the study of Wittenberg (2009) is now included in the text.

Page 18194, line 26: Just to clarify that "trend" refers to linear trend?

Yes, the word "linear" has been inserted into the text.

Page 18195, line 3: 30 years choice is getting even more confusing here.

We have cleaned up the source of confusion here by removing the confusing sentence discussed immediately above (Page 18193, Lines 1-2). Our interest in 30 years is now clearly stated as being motivated by the approximate length of relatively continuous elements of the global observing system.

Page 18200: I find that the authors here are getting into more and more details which progressively lose their importance because of the lack of proper interpretation and wider outlook as explained at the beginning of the review.

We agree with the Reviewer that the scope of the original presentation in the text of Figure 4 was insufficient. In the revised version of the manuscript, this has been addressed in two ways. First, we have scaled back the number of panels from six to four. The middle row (with three drivers) had not received attention in the original text, and we removed this. Second, we have dedicated significantly detailed new text to our interpretation of the final row (now Fig. 4c and Fig. 4d) emphasizing the case of two drivers. This has also led to a change of emphasis in Section 4.3 (in the Discussion). The interpretive text in Section 4.3 text reads:

"We have previously defined perceptible changes in drivers of ocean ecosystems as secular changes that are above the noise level of natural variability. For the case where two of the four ecosystem drivers are used (**Fig. 4c** and **Fig. 4d**), our analysis has revealed that Ω_{arag} and SST are the dominant drivers with early emergence in the tropics. In fact, it should be emphasized that this is a result of our two-driver analysis, rather than an assumption or an imposed constraint. In particular, the two-driver analysis presented in **Fig. 4c** for the recent past indicates that tropical coral reef habitats may be the primary regions currently experiencing perceptible changes for an observer relative to the background natural variability (Pelejero et al., 2005). Thus our results are consistent with previous studies that argue that coral reefs are the marine ecosystems that are threated most by environmental changes (see the cross-chapter box on coral reefs in IPCC AR5, Gatusso et al, 2014a; also see Gattuso et al., 2014b), if we equate drivers and stressors for the case of Ω_{arag} and SST for tropical coral reef habitats. Although the results of the two-driver analysis are seen to hold through the tropics, the results may warrant particular attention in the Coral Triangle biodiversity hotspot region that is thought to account for more than 75% of the world's coral species (Veron et al., 2009), spanning Indonesia, the Philippines, Malaysia, Papua New Guinea, and the Solomon Islands (Allen, 2008)."

Figure 6 (same as for FigA5): A very poor choice for the color scheme

The color scheme for both Figure 6 and Figure A5 have been modified to satisfy the suggestions of the reviewer.

P. 18203. I would disagree with the statement. The trend will become evident to the observer, not to the ecosystem. To state that it would become evident to the ecosystem needs demonstration that the ecosystem is sensitive to the driver. For example most of the organisms are not sensitive to the oxygen concentration until it falls below a certain threshold no matter what the variability and the trend are. Please also define "perceptible".

We have made a number of changes to the text to satisfy this important point, both in the Introduction and in the Discussion. First and foremost, we distinguish between "perceptible" to an observer and an ecosystem.

In the Introduction, our text now reads:

"Identifying and understanding when and where the secular trends in ocean drivers emerge above the noise is important for two reasons. The first reason is that emergence characterizes when the secular trend becomes evident or perceptible locally relative to the background variability for an observer. This is intended as a predecessor step to considering perceptible changes for the ecosystems themselves, as emergence or perceptible change to an observer will be a necessary but not sufficient condition for perceptible change being experienced by ecosystems. This distinction between perceptible change for an observer and an ecosystem is important for example for the case of O_2. Most organisms are not very sensitive to O_2 levels as long as the O_2 concentrations are sufficient. However below a critical threshold (e.g. hypoxia) many organisms start to suffer from several physiological stresses (e.g. Vaquer-Sunyer and Duarte, 2008). The value of Earth System Models is that they can characterize perceptible change from the vantage point of an observer. The second reason is that understanding of the emergence of multiple drivers will be important for optimizing the design of the ocean observing system. Inferring trends in drivers from Repeat Hydrography is complicated by natural variability in the ocean (Rodgers et al., 2009), and natural variability can also complicate trend detection using time series data (Henson et al., 2014)."

Additionally, the first paragraph of the Discussion now reads:

"We set out to evaluate the emergence characteristics of four ecosystem drivers (surface Ω_{arag}, SST, subsurface O_2, and NPP) of marine ecosystems, with two questions driving this investigation. The first motivation stems from an interest in identifying when the secular trend in drivers becomes evident or perceptible for an observer relative to the natural background decadal variability. This is precisely what Large Initial-Condition Ensemble Simulations with Earth System Models can provide, and it is our intention that this will add value to more ambitious efforts to identify when the secular trend becomes a perceptible change that can stress ecosystems. Identifying perceptible change for an *observer* is a necessary but not sufficient condition for identifying perceptible change for an *ecosystem*, as is evident for example in the fact that many organisms are not sensitive to O_2 concentrations until they fall below a particular threshold (e.g. Vaquer-Sunyer and Duarte, 2008)."

P. 18203, Line 9: Again, reference to the observing system and now also the community of researchers evaluating the network design. The authors really need to explain this properly.

We have now included a separate section of the Discussion (Section 4.4) to clarify our interpretation of the implications of our study for observing system design. The new text reads as follows:

“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 of the system. Earlier OSSEs (Christian et al., 2008; Park et al., 2010; Plancherel et al., 2013; Majkut et al., 2014; Cassar 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 Earth System through selective sub-sampling of the model output. The target is to test the available 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 on all timescales of the system of interest, but where the target is to identify the secular trend. 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; Christian, 2014). The question is then as follows: Given an observing system with perfect skill that allows one to perfectly monitor the evolution of the system, how many years of continuous measurements are needed to identify the secular trend above the noise of background variability? Our main result with this question 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 mult-decadal observing system for ocean biogeochemistry and ecosystem drivers. For the four drivers considered here, the confidence intervals found with a 30 yr window for calculating trends (**Fig. 2**) are significantly higher than those found with a 10 yr window (**Fig. 5**). For the case with a 10 yr window, even Ω_{arag} reveals broad expanses of non-emergence over the 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 continuous measurements in order to identify the rate of acidification associated with the secular trend in these regions.

“More generally, our analysis of conficence intervals for emergence for two versus four drivers (**Fig. 4c** for the period 2005-2014) largely highlight 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 (30 yr) observing system of the type considered by Ishii et al. (2009) in western Equatorial Pacific surface waters is needed to detect the rate of the secular trend in acidification against the background noise of natural variability with confidence.”