Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth System Model

K. B. Rodgers¹, J. Lin² and T. L. Frölicher³

¹[Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, USA]
²[Princeton University, Princeton, USA]
³[Environmental Physics, Institute of Biogeochemistry and for Pollutant Dynamics, ETH Zürich, Switzerland]

Correspondence to: K. B. Rodgers (krodgers@princeton.edu)

Abstract

Marine ecosystems are increasingly stressed by human-induced changes. These ocean drivers – including warming, acidification, deoxygenation and perturbations to biological productivity - can co-occur in space and time, but detecting their trends is complicated by the presence of noise associated with natural variability in the climate system. Here we use Large Initial-Condition Ensemble Simulations with a comprehensive Earth System Model under a historical/RCP8.5 pathway over 1950-2100 to consider emergence characteristics for the four individual and combined drivers. Using a one-standard deviation (67% confidence) threshold of signal-to-noise to define emergence with a 30 yr trend window, we show that ocean acidification emerges much earlier than other drivers, namely during the 20th century over most of the global ocean. For biological productivity, the anthropogenic signal does not emerge from the noise over most of the global ocean before the end of the 21st century. The early emergence pattern for sea surface temperature in low latitudes is reversed from that of subsurface oxygen inventories, where emergence occurs earlier in the Southern Ocean. For the combined multiple-driver field, 41% of the global ocean exhibits emergence for the 2005-2014 period, and 63% for the 2075-2084 period. The combined multiple-driver field reveals emergence patterns by the end of this century that are relatively high over much of the Southern Ocean, North Pacific, and Atlantic, but
relatively low over the tropics and the South Pacific. For the case of two drivers, the tropics including habitats of coral reefs emerges earliest, with this driven by the joint effects of acidification and warming. It is precisely in the regions with pronounced emergence characteristics where marine ecosystems are expected to be pushed outside of their comfort zone determined by the degree of natural background variability to which they are adapted. The results here also have implications for optimization of the ocean observing system.

1 Introduction

An important priority in climate research is to understand the potential vulnerabilities of marine ecosystems in the face of anthropogenic climate change (e.g. Doney et al., 2012). Over the last decade, multiple drivers of marine ecosystems such as ocean warming, ocean acidification, nutrient stress and low oxygen levels have been identified to be among those of greatest concern (e.g. Gruber, 2011; Hall et al., 2013). We have chosen to use “drivers” rather than “stressors” as some drivers (for example temperature) can be beneficial to some organisms or processes. On a global scale, the development of these drivers is largely a consequence of the increase in atmospheric CO₂ and the associated climate change. The oceanic response to these changes, namely the oceanic uptake of excess heat and anthropogenic CO₂ causes ocean warming and ocean acidification, i.e. a decrease in both oceanic pH and in the saturation state of seawater with regard to mineral calcium carbonate (Doney et al. 2009). The warming of the ocean tends to stratify the upper ocean (Sarmiento et al. 1998), leading to a reduced supply of nutrients to the euphotic zone (Bopp et al. 2001, Steinacher et al. 2010), but also to a reduced resupply of oxygen to the ocean’s interior (Frölicher et al., 2009; Keeling et al., 2010), causing a loss of oxygen there. The magnitude of these global drivers will likely continue to grow, given current trends in fossil fuel CO₂ emissions and the strong inertia within the global community with respect to efforts to decarbonize (Friedlingstein et al., 2014).

The detection of secular trends in driver fields on regional- to global-scales is complicated by the presence of natural variability in the climate system, as has been shown for dissolved oxygen by Frölicher et al. (2009). The presence of background natural variability motivates the introduction
of the concept of emergence to identify where the signal (the secular trend) becomes larger than
the noise (the background natural variability). 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 for
the local marine species relative to the background variability to which they have adapted. Here
emergence becomes a measure of perceptible changes for the ecosystems. 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).

Previous studies exploring the concept of emergence have largely focused on physical state
variables of the atmosphere and ocean, such as temperature, precipitation and sea level (e.g.
Diffenbaugh and Scherer, 2011; Hawkins and Sutton, 2012; Mahlstein et al., 2012; Mora et al.,
2013; Lyu et al., 2014). Mora et al. (2013), for example emphasized that the tropics, which hold
the worlds greatest diversity of marine species, will exhibit emergence in ocean warming ten
years earlier than any of the other global ocean regions. Far less attention has been devoted to
date to signal-to-noise ratios in ocean biogeochemistry, a notable exception being the study of
Keller et al. (2014).

Here we introduce a new suite of Large Initial-Condition Ensemble Simulations using a
comprehensive Earth System Model to understand the local emergence characteristics of the
ocean ecosystem drivers sea surface temperature (SST), sea water saturation state with respect to
aragonite ($\Omega_{arag}$), a mineral phase of calcium carbonate, oxygen levels ($O_2$), and net primary
productivity (NPP) over an interdecadal (30 yr) timescale. The ocean state as expressed in SST
exhibits pronounced decadal-to-interdecadal variability (Zhang et al., 1997), and variations on
this timescale are well documented for oxygen (Emerson et al., 2004; Mecking et al., 2008;
Kouketsu et al., 2010; Takatani et al., 2012). This has also been considered for the case of
phytoplankton in the study of Martinez et al. (2009).
In opting to use a suite of Large Initial-Condition Ensemble Simulations, we emphasize in particular the uncertainty in estimates of emergence due to natural variability inherent in the climate system. Deconvolving the signature of the forced response from the background natural variability with one coupled model requires Large Initial-Condition Ensemble Simulations. Only with a sufficiently large number of ensemble members can the effects of natural variability be removed by averaging over the ensemble members. With a single model run of a coupled climate model one is forced to estimate noise either from simulated pre-industrial variability or through high-pass filtering of a scenario run (e.g. Deser et al., 2014). The problem is that the forced response is in general imbedded in a stochastic dynamical system that exhibits variability on all timescales. Additionally, the amplitude of major modes of variability such as El Nino-Southern Oscillation (ENSO) is not stationary in their amplitude over climate change timescales (Timmermann et al., 1999). The ensemble approach to coupled modeling thereby offers an important opportunity when applied to the case of ecosystem drivers (Frölicher et al., 2009).

2 Methods

2.1 Model and simulations

We conducted 30 ensemble simulations over the 1950-2100 period following historical and RCP8.5 concentration pathways (van Vuuren et al., 2011). All 30 ensemble members are run with the same coupled Earth System Model developed at the Geophysical Fluid Dynamics Laboratory (Dunne et al., 2012; Dunne et al., 2013): GFDL’s ESM2M. The physical state model underlying ESM2M is the updated version of the coupled model CM2.1 (Delworth et al., 2006), consisting of the 1-degree version of the MOM4p1 ocean model (Griffies, 2009) coupled to an approximately 2-degree configuration of the AM2 atmospheric model (Anderson et al., 2004). The ocean biogeochemical model is Tracers of Ocean Phytoplankton and Allometric Zooplankton code version 2 including 30 tracers to represent cycles of carbon, oxygen, and the major macronutrients and iron (Dunne et al., 2010).
The initial conditions for the 30 ensemble members for January 1\textsuperscript{st} 1950 differed in the initial state of the atmosphere/land/ocean/sea ice components of the Earth System Model. This was accomplished by using model state snapshots for the ends of days 1-29 in January 1950 as the initial model states for January 1\textsuperscript{st} 1950 for each of the ensemble members 2-30. As has been shown by Wittenberg et al. (2014) using significantly smaller initial perturbations to the ocean component only with nearly the same underlying physical coupled model, our initial condition perturbations lead to a randomization of the ENSO state amongst the individual ensemble members within five years. Given that decadal modulations of ENSO are the most pronounced driver of decadal physical variability in this coupled model, decadal variability will be randomized amongst the individual ensemble members.

The four drivers considered in this study are (i) surface $\Omega_{\text{arag}}$, (ii) SST, (iii) subsurface $O_2$ vertically integrated from 100m to 600m, and (iv) NPP vertically integrated over the top 100m. Our focus on subsurface $O_2$ concentrations is intended to characterize regimes ranging from oxygen minimum zones to the main thermocline of polar and circumpolar regions.

2.2 Confidence intervals and detection of time of emergence

In order to quantify emergence characteristics, it is necessary to specify a timescale over which trends are calculated. We use decadal trends considered over 30 yr intervals on a gridpoint-by-gridpoint basis to quantify signal-to-noise ratios for each of the four drivers. The signal is the trend obtained using the ensemble-mean, and the noise is the standard deviation of the 30 yr trends for the individual ensemble members. The signal-to-noise ratio is thereby calculated as the ratio of these two terms, and thereby associated with this specific timescale. The choice of a 30 yr trend window is motivated by the approximate length of relatively continuous elements of the global observing system and by the timescale of important natural variability events such as the Pacific Decadal Oscillation. Given that our model runs span 1950-2100, and our choice of 30 yr trends, signal-to-noise can effectively be calculated over each year spanning the period 1965-2085.
Central to our analysis is the model-derived quantification of confidence intervals for trend detection. We assume that the 30 yr trends for the 30 individual ensemble members are normally distributed. For a given time-window (30 yr), the signal-to-noise ratio (SNR) is calculated using the relationship between the ensemble mean TREND and the standard deviation ($\sigma_{\text{noise}}$) of the trends of the various ensemble members, which we denote as NOISE:

\[ \text{SNR} = \frac{\text{TREND}}{\text{NOISE}}. \]  

Here we focus on a threshold of one for SNR, representing a confidence interval of 67% for emergence. In other words, the threshold of one is used to characterize when the “signal” of anthropogenic climate change rises above the “noise” of natural background variability. However, the sensitivity to a choice of two (95% confidence) for the threshold will also be considered. We choose two time intervals over which we consider average confidence intervals: the first is for the most recent decade 2005-2014, and the second is the decade 2075-2084 towards the end of the 21st century. The confidence intervals are calculated using 30 yr trends year-by-year over each of the 10 yr intervals before considering 10 yr averages over the respective intervals. An averaging interval of 10 yr was sufficient to remove noise present in analysis for individual years in the confidence intervals, presumably due to the fact that even with 30 ensemble members the modes of variability aren’t sufficiently randomized.

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 characteristic of the underlying coupled (atmosphere/ocean) model, it has been shown earlier studies that SST variability is more pronounced over 10 yr timescales than over 30 yr timescales (Fig. 2 in Wittenberg (2009) and Fig. 7 in Dunne et al. (2012)). 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.

Our analysis also includes diagnostics of Time of Emergence (ToE). In contrast to the confidence interval diagnostic for fixed time intervals, ToE requires that we first specify a fixed threshold (here we choose 67%) for confidence intervals and then calculate the time (year) at which that is satisfied on a gridpoint-by-gridpoint basis for each driver. For consistency, a 30 yr trend window is used here as well in the calculation of ToE.

3 Results

3.1 Temporal hierarchy of global and regional emergence of individual drivers

We start the analysis with the temporal hierarchy of emergence for the globe as well as for various regions, such as the Southern Ocean (90-45°S), the southern subtropics (45-15°S), the equatorial band (15°S-15°N), the northern subtropics (15-45°N), and the subpolar and polar Northern Hemispheres (45-90°N) (Fig. 1). For each driver, the signal-to-noise ratio was calculated gridpoint-by-gridpoint, and then the area-weighted mean of this quantity were considered by region year-by-year (solid lines in Fig. 1). There is a distinct hierarchy in the emergence of the drivers on a global scale (Fig. 1a), with $\Omega_{arag}$ (red line in Fig. 1a) already having risen above the one-standard deviation (67%, black solid vertical axis in Fig. 1a) level by the beginning of the period considered (by 1965). This is followed by SST (green line in Fig. 1a), which globally emerges from the 67% confidence level by the year 2000. $O_2$ inventories rising above the 67% confidence interval by approximately 2060 (blue line in Fig. 1a), whereas NPP remains below the 67% confidence level over the entire time period (purple line in Fig. 1a).

The regional behavior of the four drivers is shown in Fig. 1b-f. Overall, the hierarchy found in the global analysis tends to be reproduced on a regional scale but with a few notable exceptions. First, the Southern Ocean differs from the other regions in that the $O_2$ inventories (blue solid line in Fig. 1b) are more detectable than SST (green solid line in Fig. 1b) over the duration.
Additionally, the subtropical SST tends to emerge from the 67% confidence interval by 1990-2010 (green solid lines in Fig. 1c-e), with this not being the case for the Southern Ocean and the Northern subpolar/Arctic regions (green solid lines in Fig. 1b,f). The commonality is that $\Omega_{\text{arag}}$ is above the 67% confidence interval in all regions for the duration of the analysis period, whereas the NPP signal remains below the 67% confidence level for the duration of the analysis period in all regions, with the exception of the southern subtropics for the period post-2070.

3.2 Local emergence confidence intervals for individual drivers

We next consider the spatial distributions of the confidence level that the signal (ensemble mean) in the four drivers has emerged from the natural variability (standard deviation among the 30 ensemble members) for two different time periods: 2005-2014 (Fig. 2) and 2075-2084 (Fig. 3). For $\Omega_{\text{arag}}$ it is found that 99.8% of the global domain exhibits the signal emerging above the noise with more than 67% certainty (Fig. 2a). This indicates that a global observing platform of 30 yr duration would be able to detect trends nearly everywhere. The reasons for the early emergence is that changes in surface $\Omega_{\text{arag}}$ are mainly dictated by the increase in surface DIC, which closely follows atmospheric pCO$_2$ trends (e.g. Keller et al. 2014). For SST (Fig. 2b), 74.8% of the global surface ocean has emerged with 67% certainty from the noise. Especially the lower latitudes tend to have emerged with high confidence by 2005-2014, whereas the Southern Ocean south of 45°S and the northern subpolar and Arctic regions reveals only few confidence that they have emerged by 2005-2014. For O$_2$ inventories, 22.6% of the global ocean has emerged with 67% certainty from the noise (Fig. 2c). For NPP, only 14.8% of the global ocean has emerged with 67% certainty over 2005-2014 (Fig. 2d). Interestingly, the pattern for the O$_2$ inventories shown in Fig. 2c reveals emergence over important parts of the Pacific and Atlantic sectors of the Southern Ocean (defined as the region to the south of 45°S), in contrast to what is seen in SST. The lower latitudes reveal relatively little emergence by 2005-2014, but there are small areas within the subtropics of the North Pacific with emergence. For the case of NPP (Fig. 2d), there is very little evidence of emergence by 2005-2014 over most of the global domain, although the Equatorial Pacific Cold Tongue regions, parts of the Equatorial Atlantic, and the Agulhas regions show
marginal emergence. This indicates that even with complete 30 yr time series for global NPP, it 
would not be possible in most regions to identify a significant secular trend.

Next we consider in Fig. 3 the confidence intervals calculated over 2075-2084. The patterns 
shown in Fig. 3 largely reinforce what was seen in Fig. 2, although clearly the emergent 
structures have expanded. Not surprisingly, $\Omega_{\text{arag}}$ (Fig. 3a) now stands emerged over 100% of 
the global surface area with 67% certainty at the end of the 21st century, consistent with what 
was shown over 2005-2014. For SST (Fig. 3b), 90.9% of the global ocean has emerged with 
67% certainty, clearly having expanded beyond their limits from 2005-2014. This includes 
expansion into the North Atlantic, and the high latitudes over the Southern Ocean. For O$_2$ 
inventories (Fig. 3c), with 42.3% of the globe emerged with 67% certainty, the Southern Ocean 
increasingly emerges towards the end of the century, and the North Pacific sub tropics have 
emerged with high confidence. For the case of NPP (Fig. 3d), with 23.7% of the globe emerged 
with 67% certainty, there continues to be only weak emergence over most of the globe, the 
exceptions being a coherent structure spanning the southern subtropics and subtropical front 
regions of the South Indian Ocean, the equatorial Pacific upwelling region, and then patchy 
regions over the Southern Ocean the eastern mid-latitude and subpolar North Atlantic.

Taken together, the results here indicate that the four drivers are not advancing in unison with the 
same patterns and rates of their detectability over the globe. $\Omega_{\text{arag}}$ emerges first, consistent with 
what has been found in previous modeling studies (Friedrich et al., 2012), NPP emerges last, and 
then O$_2$ inventories and SST have their own contrasting patterns of emergence. Stated 
differently, random uncertainty associated with natural decadal variability poses significantly 
more important challenges for NPP than for $\Omega_{\text{arag}}$, with O$_2$ inventories and SST falling between 
NPP and $\Omega_{\text{arag}}$. 


3.3 Local emergence confidence intervals for the multi-driver mean

We now consider an average across the four ecosystem drivers of the confidence intervals for the same time intervals 2005-2014 (Fig. 4a) and 2075-2084 (Fig. 4b). To facilitate presentation and interpretation, we calculated averages within 14 biomes as defined by Henson et al. (2010). The biome definition separates the regions where phytoplankton growth is seasonally light limited (for mid to high latitudes), regions where the ocean is gaining heat (equatorial regions) and oligotrophic regions.

For the 2005-2014 interval (Fig. 4a), the multi-driver confidence intervals have already risen above the threshold of one standard deviation (67% confidence interval) in the Equatorial Atlantic, the South Atlantic, and the Arabian Sea. The Indian and Pacific sectors of the Southern Ocean weigh in at 60% confidence intervals, thereby below one standard deviation. With the exception of the South Atlantic, subpolar regions are minimum regions for confidence intervals. For the case of the Equatorial Atlantic, where in fact O\textsubscript{2} inventories are increasing (Gnanadesikan et al., 2012; Cocco et al., 2013) (Fig. A1e), it is important to acknowledge that emergent O\textsubscript{2} inventories are thereby not to be understood as a driver. Interestingly, the hemispheric asymmetry between the northern and southern subtropics of the Pacific is the reverse of the asymmetry found in the Atlantic.

By the later period 2075-2084 (Fig. 4b), the confidence intervals averaged over the four drivers are higher than during the earlier period 2005-2014 except for the eastern equatorial Atlantic and the western equatorial Indian Oceans. There continues to be a hemispheric asymmetry between the subtropics of the northern and southern subtropical Pacific. Interestingly, the fact that subpolar regions have become maxima for the multi-driver mean during the later period represents a reversal of what is found during the earlier period. For both hemispheres this reflects an important contribution from the O\textsubscript{2} inventories, as can be seen in the subpolar time series changes in Figs. 1b and 1f. A relative decrease in variability in SST can be seen over the subpolar regions (Figs. A1d and A2d), contributing to the increased confidence in emergence over these regions (the differences between Fig. 2b and Fig. 3b).
As we did for the individual drivers, we also consider here the fraction of the global area that has emerged for the multi-driver mean over the periods 2005-2014 and 2075-2084. Following the procedure used in Figs. 2 and 3, a 30 yr trend window was used to characterize signal-to-noise over the respective time interval year-by-year. The result was then averaged over each of the 10 yr time intervals gridpoint-by-gridpoint. Subsequently, area-weighted spatial averages were considered. For the earlier period (2005-2014), 40.9% of the global ocean exhibits emergence with 67% confidence. By 2075-2084, 62.5% of the global ocean exhibits emergence with 67% confidence.

Next we consider the emergence characteristics for the three drivers with highest emergence confidence intervals (rather than all four drivers) in the subsequent two panels. This is shown for the time interval 2005-2014 in Fig. 4c, and for the time intervals 2075-2084 in Fig. 4d. For each case, the fields presented in Fig. 2 and Fig. 3 were considered gridpoint-by-gridpoint to find the three maximum drivers, and these were then averaged over the same biome regions. Unsurprisingly, the confidence intervals are higher for this case with three drivers than with four. For the time interval 2005-2014, 87.8% of the global surface area has emerged above the 67% confidence interval (Fig. 4c), and for 2075-2084 98.8% has emerged above 67% (Fig. 4d).

We then consider averages over the two top drivers for the period 2005-2014 (Fig. 4e) and 2075-2084 (Fig. 4f). Clearly the emergence characteristics are significantly higher here. As is revealed in the structures of locally maximum and minimum confidence intervals, the dominant influences over most of the global ocean are $\Omega_{an}g$ and SST in setting the even higher confidence intervals. For the time interval 2005-2014, 97.94% of the global ocean is above the 67% confidence interval. For the time interval 2075-2084, this has risen to 99.92% of the global ocean having risen above the 67% confidence interval.
Taken together, the six panels in Fig. 4 reveal a strong sensitivity to the number of drivers used to characterize multiple-driver emergence. For the case of emergence over 2075-2084, it is interesting to note that the patterns are different. For four drivers, the Southern Ocean has the strongest emergence, whereas with two drivers the tropical band has the strongest emergence.

3.4 Sensitivity of Confidence Intervals to Length of Decadal Trend Window

Next we consider the sensitivity of our confidence interval calculation for the period 2005-2014 (Fig. 2) to the length of the window over which trends are calculated. The sensitivity of the confidence intervals over 2005-2014 to the width of the window is considered in Fig. 5, where a window of 10 yr has been used instead of the 30 yr used in Fig. 2. We show that for all four drivers, the choice of a 10 yr trend window results in important decreases in the confidence intervals over the global domain relative to a 30 yr trend window.

For $\Omega_{\text{arag}}$ (Fig. 5a), the confidence intervals are much lower in the regions directly impacted by El Niño variability with a 10 yr window than they are with a 30 yr window (Fig. 2a). Additionally, smaller confidence intervals are in evidence through the North Atlantic and parts of the Arctic, as well as in a well-defined band across the North Pacific. In fact, along much of both the west and east coasts of North America, confidence intervals are significantly lower than with the 30 yr window. This may have important implications in pointing to the need for sustained (multi-decadal) observing systems for ocean biogeochemistry in these regions.

For SST as well, the pattern obtained with a 10 yr window (Fig. 5b) reveals large differences from the pattern obtained with a 30 yr window (Fig. 2b). The relatively elevated values throughout much of the tropics and the subtropics have now disappeared, revealing relatively weak confidence intervals over global domain. A similar loss of confidence is found in for O$_2$ inventories (Fig. 5c) and for NPP (Fig. 5d). In fact for SST, O$_2$ inventories, and NPP, using a 10 yr window results in very weak confidence nearly everywhere over the global domain.
3.5 Time of emergence in individual ecosystem drivers

Next we consider Time of Emergence (ToE) in Fig. 6. Here, we use a one standard deviation threshold to define ToE, and our intention is to represent another dimension to the emergence question, namely that of time. A 10 yr tolerance window with a robust Loess filter has been applied to the fields in Fig. 6, as described in Appendix B. First we consider $\Omega_{\text{ang}}$ (Fig. 6a), with this revealing that most of the emergence has already occurred by the start of the analysis period (before 1965), consistent with what was seen in the confidence interval analysis. The only exceptions are relatively high dynamical variability zone within the equatorial Pacific, and certain high variability structures associated with western boundary current regions. For SST (Fig. 6b), the ToE is within the 20th century for much of the tropics, with a tendency for this ToE to shift to the 21st century for much of the subtropical and subpolar regions. For much of the high-latitude Southern Ocean region, there is no emergence during the analysis period (before 2085). This is also true for important sectors of the northern North Atlantic. For O$_2$ inventories (Fig. 6c) an early ToE is evident for the Southern Ocean as well as the eastern equatorial Atlantic, largely consistent with what was seen in Fig. 2. Alternating zonal structures of early and late ToE are seen in the Equatorial Pacific, and a patchwork of structures with very different ToE is in evidence over much of the ocean away from the Southern Ocean. In fact, the patchwork-like structures are in even stronger evidence for NPP (Fig. 6d). Consistent with what was seen in Fig. 2d, relatively early ToE is in evidence for both the Agulhas and the Equatorial Pacific Cold Tongue regions. Otherwise there are alternating bands of saturation for both early (pre-1965) and late (post-2085) ToE. The adjacent early and late saturation regions are the consequence of choosing a threshold for ToE. Thus the boundaries between such regions may not in general be reflecting real biome structures in the model domain. This is addressed in more detail in the Appendix (Fig. A3 and Fig. A4).

4 Discussion
4.1 Interpretation of Main Results

We set out to evaluate the emergence characteristics of four ecosystem drivers (surface $\Omega_{\text{arag}}$, 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 local marine ecosystems relative to the natural background decadal variability to which the organisms have adapted. The second pertains to the optimization of the ocean biogeochemical observing system, and the application of models to advance this optimization through iterative communication with the community of researchers evaluating network design. Building on the previous work of Frölicher et al. (2009), this was pursued using a suite of Large Initial-Condition Ensemble Simulations, as it is only with this approach that one can infer the secular trend (ensemble mean) for a model by filtering natural variability through an averaging procedure. The averaging procedure operates on not only patterns of climate modes such as ENSO (Wittenberg et al., 2014), but also on natural variability on smaller scales associated with variations in gyre boundaries that are not correlated to climate modes. Importantly, with this large ensemble approach, one does not need to assume that variability in the system is stationary in time.

Our main result is that there is a temporal hierarchy in the emergence of the four ocean ecosystem drivers above the level of background natural variability. This is strongly evident in Fig. 1. $\Omega_{\text{arag}}$ emerges earliest, NPP emerges latest, and both $O_2$ inventories and SST fall between the two. Additionally, three (SST, $O_2$ inventories, and NPP) of the four drivers considered here exhibit large regions where detection of secular trends is significantly complicated by the presence of natural decadal variability in the climate system. Our results also revealed very pronounced differences in the patterns for the confidence intervals for the emergence of SST and $O_2$ inventories (Fig. 2). For SST, the Southern Ocean emerges relatively late (post-2014) and the tropics emerge rather early (pre-2014), in line with earlier studies (Mora et al., 2013). For $O_2$ inventories, on the other hand, the Pacific and Atlantic sectors of the Southern Ocean exhibit regions of relatively early emergence (pre-2014) and the tropics emerge rather late (post-2014).
The coalescence of the different global drivers in certain regions is already creating a number of hot-spots (Fig. 4), with the Southern Ocean and more generally the high latitudes projected to increase in importance by 2075-2084.

Consistent results regarding the temporal hierarchy of ecosystem driver emergence were found through the analysis of confidence intervals (Fig. 2 and Fig. 3) and time of emergence (ToE) (Fig. 6) as diagnostics. However, we prefer the confidence interval analysis over the ToE analysis for two reasons. First, for the four ecosystem drivers the saturation characteristics of the ToE analysis (emergence before 1965 or after 2085) are widespread, complicating interpretation. Second, and perhaps more importantly, ToE diagnostics require a specification of a threshold of signal-to-noise that is somewhat arbitrary (here we have considered both 1 and 2 standard deviations, but have chosen to emphasize the less conservative value of 1 in Fig. 6).

The quantification of signal-to-noise at the center of our analysis relies on joint use of a suite of Large Initial-Condition Ensemble Simulations using an individual Earth System Model (GFDL’s ESM2M). However, previously published analyses indicate that the collection of Earth System Models developed by different modeling centers exhibit disparate amplitudes for the secular trends in individual drivers (Bopp et al., 2013). For example, different ESMs are likely to differ more in their projected changes in NPP than they are in $\Omega_{\text{arag}}$ or SST, since at least with $\Omega_{\text{arag}}$ or SST the ESMs are largely consistent in the sign of their response (Bopp et al. 2013, Steinacher et al. 2010, Cocco et al. 2013). In high latitudes, ESMs are generally consistent in simulating decreasing ocean O$_2$ inventories under 21st century climate change (Cocco et al. 2013). However, it is worth noting that in GFDL ESM2M global NPP changes by only 2% under the historical/RCP8.5 scenario, whereas global NPP tends to decrease by approximately 10% for a multi-model of CMIP5 ESMs under the same scenario (Laufkötter et al., 2015). Additionally, different Earth System Models exhibit different noise or underlying natural variability characteristics for individual drivers (Keller et al., 2014). Such inter-model differences strongly suggest that the temporal and spatial characteristics of emergence should be model-dependent, as has been shown for the case of surface air temperature by Hawkins and Sutton (2011).
Alternative emissions scenarios may also lead to changes in both the signal and the noise. Investigations of the sensitivity of our results to alternative scenarios for anthropogenic emissions and other model projections may be subject to further studies.

4.2  Mechanistic Interpretation

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 O₂ 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 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 O₂ inventories, the reverse holds. In the Southern Ocean, de-oxygenation is much larger than natural variability due to the stratification-induced reduced supply of O₂ from the surface into the thermocline (Frölicher et al, 2009; Gnanadesikan et al., 2012). In contrast, almost no O₂ changes are projected to occur in the low O₂ regions of the tropical and subtropical thermocline owing to a reduced O₂ demand because of the lower
biological production and export of organic matter in the overlying near-surface waters (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).

4.3 Perceptible changes in ocean drivers

We have previously defined perceptible changes in drivers of ocean ecosystems as anthropogenic changes that are above the noise level of the natural background decadal variability to which organisms are adapted. For the case where two of the four ecosystem drivers are used (Fig. 4e and Fig. 4f), our analysis has revealed that $\Omega_{\text{arag}}$ 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. 4e for the recent past indicates that tropical coral reef habitats may be the primary regions currently experiencing perceptible changes 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 threatened 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). 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, spanning Indonesia, the Philippines, Malaysia, Papua New Guinea, and the Solomon Islands (Allen, 2008).

It is important however to emphasize that the analyses for multiple drivers seen in Fig. 4 consisted of averaging of confidence intervals obtained for the individual drivers. In interpreting these results for SST and $\Omega_{\text{arag}}$ in the Coral Triangle as the confidence intervals for the impacts of multiple drivers, it remains a scientific challenge to determine whether they are in fact acting
additively, synergistically, or antagonistically in their impact (Boldt et al., 2014, and references therein). To date resource management strategies have tended to focus on the impact of individual drivers, with little consideration or attention to potential relationships and feedbacks between the drivers. Addressing these questions is beyond the scope of our study, but it is our hope that the analysis considered here will contribute to motivating future work in this direction. What is clear, however, is that anthropogenic CO₂ emissions to the atmosphere are the common driver of the perturbations considered here, and this underscores the necessary of a single policy response (reduction in emissions). Substantial mitigation efforts are required if ocean ecosystems are to be spared from the “quadruple whammy” of the drivers considered here.

4.4 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; 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 associate 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, 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 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 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 $\Omega_{\text{arag}}$ 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 confidence intervals for emergence for two versus four drivers (Fig. 4e) largely highlight the combined effects of $\Omega_{\text{arag}}$ 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 rate of the secular trend in acidification against the background noise of natural variability with confidence.

5 Conclusions

Here we have considered a suite of Large Initial-Condition Ensemble Simulations with GFDL’s Earth System Model ESM2M to evaluate the emergence characteristics of four drivers of ocean ecosystems under anthropogenic climate change. The drivers chosen were $\Omega_{\text{arag}}$, SST, upper-ocean inventories of $O_2$, and net primary productivity (NPP). There were two questions
underlying the analysis: First, when and where will marine ecosystems experience perceptible changes in the ocean drivers, changes that lie outside of the range of natural variability to which they are adapted or accustomed? Second, what are the implications for optimal design of a global observing system for the four drivers?

The main result of our study is that among our four ecosystem drivers, there is a pronounced temporal and spatial hierarchy to emergence over global scales. Using a one-standard deviation (67%) confidence interval threshold for signal-to-noise, and a 30 yr trend window, we find that the acidification driver ($\Omega_{\text{arag}}$) emerges earliest and NPP emerges latest over global scales. Between these two outliers, SST and O$_2$ inventories have intermediate timescales of emergence, but opposing patterns between the two of them. SST emerges earlier in the low latitudes and later in the high latitudes, while O$_2$ inventories exhibit earlier emergence over high latitude regions than in the tropics. We also considered a multiple-driver analysis where we combined the four individual drivers. There we found that whereas 41% of the global ocean area exhibits emergence over 2005-2014, 63% has exhibited emergence by 2075-2084. This four-driver analysis reveals a more pronounced emergence pattern over the extra-tropics than over the tropics. We also considered the multiple-driver case where we included only two drivers by biome region, with very different results. There the tropics emerge earliest, with the dominant drivers being $\Omega_{\text{arag}}$ and SST. Given that $\Omega_{\text{arag}}$ and SST are the two most important drivers for coral reef ecosystems, this analysis identifies coral reefs as being especially vulnerable under 21st century climate change.

Considered as an Observing System Simulation Experiment (OSSE), our results emphasize the need for a sustained global observing system for multiple decades for the task of identifying anthropogenic trends in ecosystem drivers. This is true even for the case of a global bio-Argo array within a broad multi-platform observing system, as is revealed in the contrast between the confidence intervals for emergence of the secular trend between a 10 yr sustained observing system (Fig. 5) and a 30 yr sustained observing system (Fig. 2).
Appendix A: Secular trends and natural variability underlying signal-to-noise analysis

We characterize here the secular trends (left columns in Figs. A1,A2; referred below as TREND) and standard deviations of the secular trends (right columns in Figs. A1,A2; referred below as NOISE) separately. Recall that the signal-to-noise ratio is defined as the ratio of these two fields. $\Omega_{\text{arag}}$ decreases everywhere over the global domain (Fig. A1a), with minimum relative rates of decrease in the equatorial regions, and a general tendency towards stronger relative rates of decrease at high latitudes. Largest NOISE of $\Omega_{\text{arag}}$ is simulated in a number of dynamically active regions, including the margins of the subtropical gyres and the equatorial Pacific. However, with the exception of a few isolated regions, the TREND is everywhere significantly larger than the NOISE for $\Omega_{\text{arag}}$. For SST, the TREND over 2005-2014 is positive (warming) over most of the globe, with the notable exception of the western subpolar North Atlantic and large expanses of the Southern Ocean (cooling). This stands in contrast to $\Omega_{\text{arag}}$, where the trend had the same sign over the entire domain.

The NOISE for SST finds largest expression in the subpolar regions of the Northern Hemisphere and over parts of the Southern Ocean. In fact it can be seen in the Fig. 1c,d that the extratropical regions of weak or negative TREND in SST are associated with enhanced variability. The tropics, on the other hand, reveal only modest NOISE amplitude relative to the TREND. Taken together, this helps to account for the fact that the confidence interval map in Fig. 2b reveals high confidence in the tropics relative to the subpolar regions. As has been stated in the main text, it may seem somewhat surprising that the signal-to-noise-ratio is relatively elevated in the Equatorial Pacific, given that this is the region of largest natural variability in the climate system. It is important to emphasize here that we are considering trends over a 30 yr interval in our quantification of NOISE rather than considering the standard deviation associated with interannual variability for each of the drivers.

For the case of O$_2$ inventories, a decreasing TREND can be seen in the well-ventilated thermocline of the high latitudes. Within the tropics and subtropics, structures of positive trend
do occur. As with SST but in contrast with $\Omega_{\text{arag}}$, the sign of the ensemble-mean response of $O_2$ inventories is not of the same sign everywhere. However, there are also pronounced structures of larger NOISE in $O_2$ inventories, which over many regions are associated with decadal variations in gyre boundaries and frontal regions. The most prominent extended region where the TREND is larger than the NOISE is over a broad expanse of the Pacific and Atlantic sectors of the Southern Ocean (as seen in Fig. 2c).

For the case of Net Primary Productivity (NPP), both the TREND and the NOISE patterns show relatively narrowly-defined but large-amplitude structures. There is a trend in the zonal gradient in NPP across the equatorial Pacific, as well as a trend towards enhanced NPP along the poleward flanks of the Southern Hemisphere subtropical gyres. For the NOISE, a series of relatively narrow structures of high amplitude are found winding through the tropics. Over most regions of the globe, the NOISE is of sufficient amplitude relative to TREND to give the consistently lowest Confidence Interval distribution of the four drivers considered here (Fig. 2d).

It is worth noting in Fig. A1 that the unforced components (right column) of the four drivers exhibit large-scale spatial coherence rather than small grid-scale noise. However, these structures are distinct for each of the drivers.

The same fields for the time interval 2075-2084 are considered in Fig. A2. The amplitudes in TRENDS have increased in general, but important elements of the TREND structures are similar as for the earlier period. The structures of the NOISE are quite similar to those found for the equivalent drivers during the earlier period. However, the amplitude of the variations is in many cases different for 2075-2084 than for 2005-2014. This indicates that the amplitude of the background natural variability is not stationary. For the case of SST and $O_2$, the standard deviation of natural decadal variations decreases over the Southern Ocean.

Appendix B: Temporal filtering and trend detection
The task of calculating Time of Emergence (ToE) in Fig. 6 is complicated by the fact that the signal-to-noise ratio (SNR) in time series for individual grid points in the ocean model does not in general tend to be monotonically increasing over the period 1950-2100. Rather, the evolution of the SNR can reflect that for this particular diagnostic, the 30 ensemble members may not be sufficient to eliminate noise when averaged.

As an illustration of this problem, we consider in Fig. A3 the evolution of the SNR for sea surface temperature a region in the Pacific sector of the Southern Ocean bounded by 130°W-100°W and 45°S-60°S. The time series of the SNR for the various points in this region are superposed in Fig. A3a. The non-monotonic nature of the increase in SNR through time is evident. With this unfiltered SNR calculated directly from the annual mean model output, there are relatively short-timescale excursions above the two standard deviation threshold that precede by a number of decades the more permanent crossing of the one standard deviation threshold. However, this does not occur for all of the grid points in the domain of interest. As a consequence of these early excursions above the one standard deviation threshold, the spatial pattern of ToE using a strict definition of first crossing (Fig. A3b) reveals a spatial pattern that has a ToE before 2014 (present time) over more than 50% of the region.

It is important to understand the degree to which the ToE structure in Fig. A3b reflects short-term versus longer-term or permanent transitions of the SNR about the one standard deviation threshold, rather than short-term excursions. To evaluate this, we apply to the full suite of time series shown in Fig. A3a (gridpoint-by-gridpoint SNR for SST) a robust Loess filter. We have chosen to use a 10 yr tolerance window with the robust Loess filter, with this serving effectively as a low-pass filter that is obtained using 10 yr local regressions over the entire time period for each individual gridpoint. The result of applying the robust Loess filter is shown in Fig. A4a, where the filtered time series are shown as red lines overlaying the full time series shown in blue. The filtered time series are seen to behave as low-pass filters that effectively remove the higher frequency components. When these smoothed time series are used to define ToE with the same two standard deviation threshold, the resulting pattern (Fig. A4b) reveals important differences
relative to the unfiltered time series. The structures with post-2014 ToE now occupy most of the
domain. This indicates the strong sensitivity of the spatial pattern of ToE to the time-filtering of
the SNR time series.

The sensitivity of the Time of Emergence (ToE) to the width of the window used for calculating
trends is considered in Fig. A5. Here the window is chosen to be 10 yr, and this result is to be
contrasted to what was found for ToE using a 30 yr window as considered in Fig. 6. Clearly the
narrower trend interval of 10-years results in saturation (post-2095 emergence for this case of a
10-year trend window) on nearly global scales for all of the drivers except for Ωarag. This
indicates a strong sensitivity of ToE to the timescale chosen for the analysis, consistent with what
was shown in Fig. 4 for the confidence interval sensitivity analysis. Fig. A5 is valuable in that
the strong saturation characteristics for three of the drivers (post-2085 emergence) are much more
difficult to interpret than the parallel and favored analysis with confidence intervals.

We also consider the sensitivity of the Time of Emergence (ToE) to the width of the tolerance
window used for the robust Loess filter in Fig. A6a. For each case, the sensitivity of the ToE to
the width of the tolerance window (described above) is considered for each of the four drivers.
For each case, the area-weighted global mean sensitivity is considered. The sensitivity is weak
for Ωarag, with the dynamic range of the sensitivity being only a few years as the tolerance
window is modified from 5 yr to 25 yr. The sensitivity is largest for O2 inventories, with a
decrease of the ToE of 80 years as one transitions from a 5 yr to a 25 yr tolerance window. This
strong sensitivity is likely the expression of the red spectrum of modes of variability in
thermocline depth impacting O2 inventories in their temporal variability. The sensitivities of SST
and NPP are quite similar, both being approximately 50% of the amplitude of the sensitivity seen
for O2 inventories. In fact, the sensitivity of NPP should be expected to be larger than that of
SST, given that the sensitivity shown here may be obscured by the saturation characteristics of
NPP.
We also consider in Fig. A6b the sensitivity of the ToE to the choice of a SNR ratio of one in Fig. 6. Here as well, the sensitivity considers the global area-weighted mean. The sensitivity is strongest for SST, and then the second strongest is found for $O_2$ inventories. The sensitivity is weakest or $\Omega_{\text{arg}}$ and for NPP, but as a caveat it needs to be emphasized that both of these fields exhibit saturation behavior in Fig. 6. If the suite of 30 ensemble runs with ESM2M had been runs for a significantly longer time intervals, say from 1860-2300, then one would expect that the sensitivity for these two fields to be more pronounced.

Acknowledgements

The contribution of K.B. Rodgers came through awards NA17RJ2612 and NA08OAR4320752, which includes support through the NOAA Office of Climate Observations (OCO). The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NOAA or the U.S. Department of Commerce. T.L. Frölicher acknowledges financial support from the SNSF (Ambizione grant PZ00P2_142573). J. Lin’s work in a summer internship was supported through the Princeton Environmental Institute (PEI). The authors would like to thank John Dunne at GFDL for his contributions with model development and for access to computing resources. We would also like to thank Brendan Carter, Niki Gruber, Jorge Sarmiento, Richard Slater, Paul Yi, and Sarah Schlunegger for fruitful discussions.
References


Cocco, V., Joos, F., Steinacher, M., Frölicher, T.L., Bopp, L., Dunne, J., Gehlen, M., Heinze, C.,
Orr, J., Oschlies, A., Schneider, B., Setschneider, J., and Tjiputra, J.: Oxygen and indicators of
stress for marine life in multi-model global warming predictions, Biogeosciences, 10, 1849-1868,
2013.

Delworth, T.L., Broccoli, A.J., Rosati, A., Stouffer, R.J., Balaji, V., Beesley, J.A., Cooke, W.F.,
Dixon, K.W., Dunne, J., Dunne, K.A., Durachta, J.W., Findell, K.L., Ginoux, P., Gnanadesikan,
A., Gordon, C.T., Griffies, S.M., Gudgel, R., Harrison, M.J., Held, I.m., Hemler, R.S., Horowitz,
Malysev, S.L., Milly, P.C.D., Ramaswamy, V., Russell, J., Schwarzkopf, M.D., Shevliakova,
and Zhang, R.: GFDL’s CM2 Global Coupled Climate Models. Part I: Formulation and

Climate over the Next 50 Years: Uncertainty due to Internal Variability, Journal of Climate, 27,
2271-2296, 2014.

Diffenbaugh, N. S., and Scherer, M.: Observational and model evidence of global emergence of


Doney, S.C., Ruckelshaus, M., Duffy, J.E., Barry, J.P., Chan, F., English, C.A., Galindo, H.M.,
Grebeimeier, J.M., Hollowed, A.B., Knowlton, N., Polovina, J., Rabalais, N.N. Sydeman, W.J.,
and Talley, L.D.: Climate change impacts on marine ecosystems, Annual Review of Marine

Dunne, J.P., Gnanadesikan, A., Sarmiento, J.L., and Slater, R.D.: Technical description of the
prototype version (v0) of Tracers of Phytoplankton with Allometric Zooplankton (TOPAZ) ocean
biogeochemical model as used in the Princeton IFMIP model, Biogeosciences, 7 (Suppl.), 3593,

Dunne, J.P., John, J.G., Adcroft, A.J., Griffies, S.M., Hallberg, R.W., Shevliakova, E., Stouffer,
R.J., Cooke, W., Dunne, K.A., Harrison, M.J., Krasting, J.P., Halysev, S.L., Milly, P.C.D.,


Gruber, N.: Warming up, turning sour, losing breath: ocean biogeochemistry under global change, Phil. Trans. R. Soc. A, 369, 1980-1996, 2011.


Steinacher, M., Joos, F., Frölicher, T., Bopp, L., Cadule, P., Cocco, V., Doney, S.C., Gehlen, M., Lindsay, K., Moore, J.K., Schneider, B., and Segschneider, J.: Projected 21st century decrease in


Figure 1. Time series of (area weighted) averages of signal-to-noise ratio (SNR) for four marine ecosystem drivers, considered over a number of ocean regions: (a) Global, (b) 90°S-45°S, (c) 45°S-15°S, (d) 15°S-15°N, (e) 15°N-45°N, and (f) 45°N-90°N. The four drivers are $\Omega_{\text{arag}}$, SST, $O_2$ inventories, and NPP. For each driver, trends have been calculated individually on a gridpoint-by-gridpoint basis using a 30 yr trend window, and the SNR was subsequently calculated. The vertical axis has a logarithmic scale (non-dimensional) representing the SNR. The 1-standard deviation threshold is shown as a solid horizontal line, and the 2-standard deviation threshold is shown as a dashed line in each panel.
Figure 2. Confidence interval maps averaged over 2005-2014 for (a) $\Omega_{\text{arag}}$, (b) SST, (c) $O_2$ inventories, and (d) NPP. For the case of $O_2$ inventories in panel (c), shelf regions where the ocean depth is less than 600m deep are not included in the analysis. For each case, a 30-year window has been used to calculated trends gridpoint-by-gridpoint for each year between 2005 and 2014. An average over 10 years was considered to remove shorter timescale fluctuations in the signal-to-noise ratio. Note that the color scheme here is chosen such that saturation occurs (maroon color) above the 67% confidence interval. Warm colors indicate confidence intervals ranging from 67-95% (one to two standard deviations), and cool colors span the range 0%-67% (less than one standard deviation).
Figure 3. Same as Fig. 2, but averaged over the 2075-2084 period.
Figure 4. Confidence intervals for two ten-year intervals, namely 2005-2014 (left column), and 2075-2084 (right column). The confidence intervals for (a) the four drivers (Ω_{arag}, SST, O_{2} inventories, and NPP) over 2005-2014 are taken as the average over the fields shown in Fig. 2, and (b) the confidence intervals shown for the four drivers over 2075-2084 are taken as the average of the fields shown in Fig. 3. This is then considered for the case of the dominant three drivers in (c) for the period 2005-2014, and in (d) for the period 2075-2084. Finally, the confidence intervals averaged for the dominant two drivers are shown in (e) over 2005-2014 and (f) over 2075-2084.
Figure 5. The confidence intervals for emergence of each of the drivers for the time period 2005-2014, using a 10 yr window for calculating trends. The analysis here is otherwise identical to that shown in Fig. 2, except that there a 30 yr window was chosen. The panels show the distributions for (a) $\Omega_{\text{arag}}$, (b) SST, (c) $O_2$ inventories, and (d) NPP. This analysis reveals significantly lower confidence intervals for the 10 yr window than for the 30 yr window for each of the four drivers.
Figure 6. Time of Emergence (ToE), calculated using a threshold of 1 standard deviation (67% confidence) for (a) $\Omega_{\text{arag}}$, (b) SST, (c) $O_2$ inventories, and (d) NPP. The color scheme has been chosen to distinguish between relative to the present (2014) with warm colors indicating a ToE post-2014 and cold colors indicating a ToE pre-2014.
Figure A1. Linear trends (left) and standard deviations of the linear trends (right) for (a-b) $\Omega_{arag}$, (c-d) SST, (e-f) $O_2$ inventories from 100m-600m, and (g-h) net primary production averaged over 2005-2014. All fields are calculated using 30-year trend windows, and the trends are shown in units of %/year.
Figure A2. Same as Fig. A1, but averaged over 2075-2084.
Figure A3. Time of emergence of SST considered for a boxed region of the Pacific sector of the Southern Ocean (130°W-100°W and 45°S-60°S). The SNR calculated individually using annual mean SST for each gridpoint in the domain is shown in panel (a), where a 30-year window is used to calculated TREND and NOISE. Short timescale excursions of less than ten years are in evidence in modulations of maximum SNR. The spatial pattern of ToE for SST over this region is shown in panel (b). A ToE before the present time (pre-2014) is indicated by cool (blue) colors, while a later ToE is indicated with warm (orange) colors. Saturation (ToE post-2085) is represented with maroon color.
Figure A4. The ToE is considered for the same region, but this time with application of a robust Loess filter with a tolerance window of ten years applied to the SNR of SST calculated for the individual surface grid points in the domain. The SNR after application of the robust Loess filter is shown in panel (a) in red, superposed over the same blue time series considered in Fig. A3a. The filtering effect on the short timescale maximum excursions of the SNR is evident in the Fig. The net effect over the full domain of the robust Loess filter is shown in panel (b), revealing a later ToE over significant portions of the region of interest.
The Time of Emergence (ToE) for each of the four drivers has been calculated using a 10-year window for the calculation of trends, considering the time interval 1955-2095. This is the complement to the results with ToE for a 30 yr window, shown in Fig. 6.
Figure A6. Sensitivity analysis for tolerance window width in years (left) -- using a one standard deviation threshold and revolved around a 10 yr window, as well as threshold level (right) -- using a 10 yr tolerance window and revolved around a standard deviation of one. Caveats regarding averaging over fields that experience saturation are discussed in the text of the Supplementary Materials.