NATURAL VARIABILITY IN A CHANGING OCEAN:
EMERGENCE AND IMPACTS

SARAH SCHLUNEGER

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF GEOSCIENCE
PROGRAM IN ATMOSPHERIC AND OCEANIC SCIENCE
ADVISER: PROFESSOR JORGE SARMIENTO

JUNE 2019
© Copyright by Sarah Schlunegger, 2019.
All rights reserved.
Abstract

Anthropogenically-forced changes in the ocean are underway and critical for the ocean’s role as a carbon sink and marine habitat. Detecting such changes will require quantification of not only the magnitude of the change (anthropogenic signal) but also the natural variability inherent to the climate system (noise). This work uses Earth System Models (ESMs) to (1) evaluate timescales over which anthropogenic signals in the contemporary ocean emerge from natural climate variability and (2) interpret observed variability in the Pacific Basin.

We apply time-of-emergence (ToE) diagnostics to a Large Ensemble experiment of an ESM, providing both a conceptual framework for interpreting the detectability of anthropogenic impacts on the ocean carbon cycle and observational sampling strategies required to achieve detection. We find ToEs for different components of the ocean carbon cycle range from under a decade to over a century, a consequence of the time-lag between chemical and radiative impacts of rising atmospheric CO$_2$ on the ocean. Processes sensitive to chemical changes emerge rapidly, such as impacts of acidification on the calcium-carbonate pump (10-20 years), and the invasion of anthropogenic CO$_2$ into the ocean (20-30 years). Processes sensitive to the ocean’s physical state, such as the soft-tissue pump, which depends on nutrients supplied through circulation, emerge decades later (20-80+ years).

Next, we evaluate the model- and scenario-sensitivity of ToEs through comparing Large Ensembles from four ESMs. We find ToEs are robust across models for variables that are tied directly with the rise in atmospheric CO$_2$, namely rising sea surface temperature and the invasion of anthropogenic CO$_2$ into the ocean (ToE 20-30 years). For the soft-tissue pump, ocean color, and sea surface salinity, ToEs are longer (50+ years), less robust across the ESMs, and more sensitive to the forcing scenario considered.
Finally, we investigate potential mechanisms responsible for recent variability in the Pacific Ocean. We conduct wind-substitution simulations with GFDL-ESM2M in which decadal trends in the trade-winds are nudged toward observed values. These simulations provide better agreement between simulated and observed variability in ocean temperature, circulation, sea-level and air-sea CO$_2$ exchange, indicating this variability could be attributed to strengthening trade winds.
Acknowledgements

This dissertation was finically supported by NASA award NNX17AI75G, Princeton University, Princeton University’s Carbon Mitigation Initiative (CMI) and the Cooperative Institute for Modeling the Earth System (CIMES, formerly CICS). This dissertation was scientifically supported by the primary advisement of Professor Jorge Sarmiento and Keith Rodgers, and secondary advisement of John Dunne, Thomas Frölicher and Tatiana Ilyina and additional co-authors who are listed within the chapters (3-5) which are or will be stand-alone publications. This dissertation was technologically supported through the computational resources of NOAA/GFDL and assistance of Rick Slater. Lastly, this dissertation was inter-personally supported by my loving parents, partner and friends.
# Contents

Abstract ................................................................. iii
Acknowledgements ....................................................... v
List of Figures ............................................................ vii

1 Introduction ............................................................ 1
   1.1 Overview .......................................................... 1
   1.2 The ocean carbon cycle: components and drivers .............. 3
   1.3 Uncertainties in projections and observations of the Earth System . 9

I Emergence of Anthropogenic Trends in the Ocean .......... 14

2 Large Ensemble and Emergence Methods ......................... 15
   2.1 The GFDL-ESM2M Initial Condition Large Ensemble ............ 15
   2.2 Signal, Noise and Emergence ...................................... 16
   2.3 Partitioning Uncertainty Through Intercomparison of Large Ensembles 21

3 Emergence of Anthropogenic Signals in the Ocean Carbon Cycle 24
   3.1 Introduction ....................................................... 25
   3.2 Overview of Emergence Chronology ............................... 27
   3.3 The CaCO₃ Pump .................................................. 31
   3.4 The Invasion Flux ................................................ 33
   3.5 The Soft-Tissue Pump ............................................. 35
3.6 Emergence as a Lower Bound on Detection

3.7 Conclusions

3.8 Author Contributions

4 Time of Emergence & Large Ensemble Intercomparison For Secular Trends in Ocean Biogeochemistry

4.1 Introduction

4.2 Results

4.2.1 Signals and Time of Emergence

4.2.2 Partitioning Uncertainty in Projections

4.2.3 Signal-to-Noise for trends over the observational era

4.3 Discussion and Conclusions

4.4 Author Contributions

II Natural Trends in the Pacific Ocean

5 Strengthening Trade Winds & an Enhanced Equatorial Pacific Carbon Source

5.1 Introduction

5.2 Methods

5.2.1 Data-based estimates

5.2.2 Models

5.3 Results

5.3.1 Regional trends in Air-Sea CO₂ fluxes

5.3.2 Trends in SST and SSH

5.3.3 Trends in subsurface circulation, temperature and carbon

5.4 Discussion

5.5 Author Contributions
6 Conclusions

6.1 Observational Constraints on Natural Variability and ToE . . . . . . 75
6.2 Concluding remarks on the 'Ensemble Advantage' . . . . . . . . . 79

A Supplementary Material for Chapter 3 . . . . . . . . . . . . 80
A.1 Anthropogenic Changes in the Ocean Carbon Pumps . . . . . . 80
A.2 Time of Emergence for O$_2$, Nutrients and other ocean properties . . . 82
A.3 Alternative methods for ToE computations . . . . . . . . . . . . . 85

B Supplementary material for Chapter 4 . . . . . . . . . . . . 92

C Supplementary material for Chapter 5 . . . . . . . . . . . . 103

Bibliography . . . . . . . . . . . . . . . . . . . . . . 106
# List of Figures

1.1 Schematic of the ocean carbon cycles processes and pumps . . . . . 4
1.2 Venn Diagram schematic of sources of uncertainty in simulating . . 9

2.1 The GFDL Large Ensemble Experimental Design . . . . . . . . . . 15
2.2 Maps of 30-year trends, Signals and Noise in SST . . . . . . . . . . 17
2.3 Distribution of trends in global mean SST . . . . . . . . . . . . . 19
2.4 Time-series of Signal, Noise and SNR for trends in global mean SST . 20
2.5 The GFDL, CESM, CanESM and MPI Large Ensembles . . . . . . . 22

3.1 Fractional Area of Emergence . . . . . . . . . . . . . . . . . . . . . 28
3.2 Global and regional year of emergence . . . . . . . . . . . . . . . . 29
3.3 Time of Emergence and Signal Maps for the three carbon pumps. . . . 32
3.4 Seasonality of \( \Delta pCO_2 \) ToE and Signals . . . . . . . . . . . 34
3.5 ToE and Signal Maps for surface vs. depth-integrated chlorophyll . . 36

4.1 Global Changes in the 4 Large Ensembles, RCP8.5 Forcing . . . . . 45
4.2 Global and Regional Time of Emergence . . . . . . . . . . . . . . 47
4.3 Pace of Local Emergence . . . . . . . . . . . . . . . . . . . . . . . 49
4.4 Maps of Multi-model mean Time of Emergence . . . . . . . . . . . 49
4.5 Maps of Multi-model mean Signal . . . . . . . . . . . . . . . . . . 50
4.6 Partitioning uncertainty . . . . . . . . . . . . . . . . . . . . . . . . 51
4.7 Maps of SNR for the 4 Large Ensembles, Observational Era . . . . . 54
4.8 Maps of Signal-to-Noise Ratio for the 4 Large Ensembles, 1990-2019

5.1 Regional Trends for Pacific Air-Sea CO₂ Flux 1990-2009

5.2 Trends for Air-Sea CO₂ Flux 1990-2009

5.3 Trends for SST 1990-2009

5.4 Trends for SSH 1993-2012

5.5 Wind-induced Trends for Subsurface Transport, Temperature and DIC

6.1 Noise and Emergence for HadiSST and the GFDL-LE

6.2 Time of Emergence using ESM Signals and HadiSST Noise

A.1 Changes in ocean carbon pumps for the LE and sensitivity experiments

A.2 Change in surface ocean pCO₂ due to biological export of carbon

A.3 Time of Emergence for selected variables

A.4 Signal at Time of Emergence

A.5 Decadal variability and mean state for CaCO₃ export

A.6 Depth differences in trends for iron limitation

A.7 Error for ToE calculation with Pre-Industrial noise

B.1 Global Changes in the 4 Large Ensembles, RCP4.5 Forcing

B.2 Regional time series of changes, RCP8.5 Forcing

B.3 Regional time series of changes, RCP4.5 Forcing

B.4 SST Emergence and Signals

B.5 Air-Sea CO₂ Flux Emergence and Signals

B.6 POC Export Emergence and Signals

B.7 Cholorophyll Emergence and Signals

B.8 SSS Emergence and Signals

B.9 Difference between RCP8.5 vs. RCP4.5 ToE

B.10 Partitioning Regional Uncertainty
Chapter 1

Introduction

1.1 Overview

Human activity over the previous centuries has resulted in rising concentrations of atmospheric CO$_2$ and concomitant warming. Much of the anthropogenic carbon (30-40%) and heat (>90%) have been sequestered by the global ocean, slowing the pace of global warming [1] [2]. However, this service that the ocean performs comes at a cost – namely ocean acidification and ocean warming, which poses threat to marine ecosystems, and to the ocean’s capacity for continued sequestration of anthropogenic carbon and heat.

Seminal work on climate and ocean sensitivity to anthropogenic forcings focused on understanding the globally averaged or globally integrated response. Later, with increasing model complexity and resolution, the global response could be understood as the sum of heterogeneous regional responses. Recently, the increasing duration of climate observations has revealed significant temporal heterogeneity in the oceans sequestration of carbon and heat with this temporal heterogeneity being largely a consequence natural climate variability [3] [4]. This thesis uses Earth System Models (ESMs) to contribute to the growing understanding of how natural climate variability interacts with climate change on interannual to decadal time scales.
This thesis contends with three simple statistical concepts: signal, noise, and the signal-to-noise Ratio (SNR). Any observed or simulated climate time-series is a superposition of the impacts of climate change (the signal) and natural background variability (the noise). In the climate record, the signal and the noise are difficult (or impossible) to untangle, however with a climate model and a specific experimental design (namely, a Large Ensemble), the signal and the noise can be separated, and their ratio computed. The SNR provides a means to statistically evaluate the likely contribution of climate change to a trend in the climate system. When the SNR is sufficiently large, then the signal of anthropogenic climate change can be robustly distinguished from background noise. This is referred to as the time of emergence (ToE). Given that ESMs simulate SNRs equitable to those in the observed world, this ToE provides a baseline for how long an observing platform would need to monitor the climate system in order to theoretically detect an anthropogenic trend. Additionally, ToE informs impact timescales, since systems or organisms are likely to feel anthropogenic change once it exceeds the envelope of natural variability to which they are accustomed or adapted.

This thesis consists of two primary parts, looking separately at ToEs and Noise. The first part (Chapters 2, 3 and 4) evaluates timescales over which the impact of climate change on the ocean carbon cycle emerge from natural climate variability. Chapter 2 presents the methodology, Chapter 3 presents results from a Large Ensemble experiment of a single ESM (GFDL-ESM2M) and Chapter 4 presents results from Large Ensemble experiments from four ESMs (GFDL-ESM2M, CESM1-BGC, CanESM2, MPI-ESM1-LR) and different warming scenarios (RCP85 and RCP45). The second part (Chapter 5) evaluates the representation of natural variability (noise) in GFDL-ESM2M and presents new experiments in which GFDL-ESM2M is nudged toward reproducing recently observed decadal variability in the Pacific Ocean.
Chapters 3, 4 and 5 are derived from stand-alone publications. Chapter 3 is derived from a manuscript which is accepted at Nature Climate Change. Chapter 4 and 5 are derived from manuscripts that will be submitted imminently. These three publication-based-chapters (3, 4 and 5) begin with inclusion of all co-authors and a chapter summary or abstract, and end with an explanation of author contributions to the work.

The following sections of Chapter 1 introduces and motivates the study of the ocean carbon cycle, and the sources uncertainty involved in modeling and observing the climate system.

1.2 The ocean carbon cycle: components and drivers

On both geologic and contemporary time scales, the ocean plays a critical role in determining the atmospheric concentration of carbon and thus the transient and equilibrium response of the Earth system to anthropogenic forcing [5] [6]. The ocean’s contribution to the global carbon cycle can be thought to start at the air-sea interface, where atmospheric carbon will enter the ocean if the partial pressure of CO$_2$ (pCO$_2$,atm) of the atmosphere is higher than that of the ocean (pCO$_2$,sea), and vice-versa. In the atmosphere, pCO$_2$ is dominantly set by the mixing ratio, or partial pressure of gaseous CO$_2$. In the ocean, pCO$_2$ is dominantly set by the concentration of dissolved inorganic carbon (DIC), alkalinity (or buffering capacity), and temperature, with CO$_2$ being inversely soluble with temperature (e.g. increasing temperatures result in lowered solubility via increased pCO$_2$,sea). The magnitude of the exchange or flux of carbon (Θ) between the ocean and atmosphere is driven by (a) the magnitude of the pCO$_2$ gradient (ΔpCO$_2$) between the ocean and atmosphere, and (b) the transfer or piston velocity (k), which varies non-linearly with magnitude of the wind speed (U):
Due to the inverse solubility of CO$_2$, carbon is preferentially taken up in cooler waters, high-latitude waters. These waters are denser and therefore more likely to sink to depth, leading to the accumulation of DIC at depth (See Figure 1). This is referred to as the solubility pump, as it promotes the vertical gradient in DIC observed in the ocean, for which DIC concentrations increase with depth, and the resulting drawdown of atmospheric carbon by up to 100 ppm $[5]$. 

Figure 1.1: Schematic of the ocean carbon cycles processes and pumps

\[
\Delta pCO_2 = pCO_2,_{atm} - pCO_2,_{sea} \quad \text{(1.1)}
\]

\[
\Theta = k(U) \cdot \Delta pCO_2 \quad \text{(1.2)}
\]
The distribution of DIC in the ocean is also impacted by biological processes, namely photosynthesis at the ocean’s surface and remineralization throughout the water column. Photosynthesis involves the fixation of DIC into organic carbon by phytoplankton, and remineralization involves the bacterial processing of organic matter back into DIC. On net, more photosynthesis occurs in surface waters than does remineralization, leading to a net export of carbon to depth. This is referred to as the biological pump, and is responsible the drawdown of atmospheric carbon by 200 ppm [5].

The biological pump has two components: a soft-tissue pump which exports organic carbon and a hard tissue or calcium carbonate (CaCO$_3$) counter pump which exports both organic carbon and carbonate ion (CO$_3^{2-}$), an alkaline species of inorganic carbon. The loss of inorganic carbon from the surface reduces pCO$_2$, however the loss of alkalinity via the hard-tissue pump increases surface pCO$_2$. On net, the hard-tissue pump promotes the exhaustion of CO$_2$ from the ocean, hence the term ‘counter’ pump, as it impacts surface pCO$_2$ and air-sea CO$_2$ exchange in the opposite direction as the other pumps.

The conceptual framework of the ocean carbon ‘pumps’ is useful for understanding anthropogenically-forced changes in the contemporary ocean — both changes in the pumps themselves and changes in the important ocean properties and processes to which the pumps are coupled. The impacts of anthropogenic emissions on the ocean carbon pumps can be separated into two categories: those that arise directly or chemically from increasing concentration of atmospheric CO$_2$, and those that arise indirectly as a consequence of the Earths radiative imbalance, i.e. due to climate change. These two distinct pathways by which anthropogenic forcings induce changes in the carbon cycle are termed the concentration-carbon and climate-carbon feedbacks [7][8].
The pumps are impacted differentially by the direct vs. indirect (concentration vs. climate) components of anthropogenic emissions, and have different contributions to the evolving strength of the ocean carbon sink. Previous studies have found that the invasion of anthropogenic CO$_2$ into the ocean is predominantly driven by the increase in atmospheric carbon concentrations (and ΔpCO$_2$) and thereby can be categorized as a subcomponent of the solubility pump, called the invasion flux [9][8]. Climate-induced changes in the invasion flux are also important and may reduce uptake by 10% by end of the 21st century [8].

Anthropogenic changes in the biological pumps are generally smaller than those associated with the solubility pump (or invasion flux) and less important to the magnitude integrated uptake of anthropogenic carbon [10][11]. Changes in the soft-tissue pump are likely to occur due to changes in climate and their impact on ocean circulation, and changes in the hard-tissue pump are likely to occur due to both changes in climate and changes in ocean chemistry that result from the invasion of anthropogenic CO$_2$. Although less important than the solubility pump for the evolving strength of the ocean carbon sink, changes in the biological pumps have implications for other important ocean properties, such as oxygen concentrations, and processes, such as the transfer of energy between marine trophic levels.

The ocean carbon pumps transport not only carbon but are also coupled to the distribution of marine ecosystem tracers and stressors such as acidification, oxygen and nutrients. Once gaseous CO$_2$ invades the ocean, it rapidly reacts with water to form carbonic acid (H$_2$CO$_3$*), which dissociates to form hydrogen ions (H$^+$), bicarbonate (HCO$_3^-$), and carbonate (CO$_3^{2-}$):
On net, the reactions can be summarized as:

\[
H_2CO_3^* + CO_3^{2-} \rightleftharpoons 2HCO_3^-
\]  

(1.6)

Therefore, as a consequence of increasing \( \text{CO}_2_{\text{gas}} \), the system is pushed towards production of \( H^+ \) (resulting in decreasing pH) and the associated consumption of (or buffering by) the \( \text{CO}_3^{2-} \) ion, a critical building block for calcifying organisms. The invasion of anthropogenic \( \text{CO}_2 \) reduces the abundance of \( \text{CO}_3^{2-} \), which could act to diminish the production of \( \text{CaCO}_3 \) and consequently the function of the hard-tissue pump, constituting a concentration-carbon feedback on the climate system. Additionally, the decrease in \( \text{CO}_3^{2-} \) abundance will impact other ecologically-important calcifying organisms like corals and pteropods.

Oxygen and nutrient distributions are also coupled to the biological carbon pumps. The supply of nutrients from the ocean’s interior to the ocean’s euphotic zone promotes photosynthetic activity and the conversion of DIC into oxygen and organic matter. The export of organic matter from the surface to depth provides food for mesopelagic species, and once remineralized, returns nutrients and consumes oxygen from the ocean’s interior.

Modeling studies indicate that in the contemporary ocean, the invasion of anthropogenic heat, changes in atmospheric circulation and changes in fresh-water fluxes at the ocean’s surface impact ocean circulation, mixing and stratification which consequently alter (generally reduce) the supply of nutrients to the euphotic zone [12].
Reduced nutrient concentrations favor smaller phytoplankton which are anatomically more efficient at nutrient uptake but are less efficient at sinking to depth after mortality. Both effects are due to the increased surface area to volume ratio small phytoplankton have relative to larger phytoplankton. Due to this reduction in phytoplankton biomass, reductions in export of organic carbon is projected to occur. One impact of reduced export is increased oxygen concentrations at depth, potentially offsetting the solubility-induced reductions in global ocean oxygen concentrations.

Through these pathways the ocean carbon pumps are both impacted by and impact (i.e. are coupled to) anthropogenic climate change. However, at any given point in time, the impact of anthropogenic forcings on the ocean carbon cycle and carbon pumps can be significantly smaller than the variation induced through natural causes. The carbon cycle and carbon pumps undergo substantial variability on seasonal to interannual to decadal timescales due to the natural variability of the climate system and its impacts on temperature, circulation, light, chemicals, etcetera in the ocean. For example, with a phase-switch of the El Nino South Oscillation (ENSO), a mode of variability that describes the atmospheric and oceanic circulation patterns in the Pacific Basin, dramatic changes include (1) concentration of carbon and nutrients in the surface ocean, (2) surface ocean temperatures (SST), (3) the cloud-cover, mixed layer depths and subsequently light availability for photosynthesis and (4) surface wind speeds and subsequently gas transfer velocities. To put the magnitude of such natural variability into perspective, the changes over the eastern Equatorial Pacific in SST (3°-4°C) and concentrations of DIC (100+ µg kg⁻¹) which can occur over the course of a year due to natural variability are equivalent to those climate change is projected to impart on the ocean’s mean state over the coming decades to century.

A key scientific challenge is therefore distinguishing between and understanding the interaction between natural and anthropogenic drivers of change in the ocean
carbon cycle. Doing so requires quantification of the uncertainty associated with each.

1.3 Uncertainties in projections and observations of the Earth System

Both projections and observations of the coupled carbon-climate system are subject to uncertainties (Fig. 1.2). Earth System Models (ESMs), which are global climate models that include representation of an interactive carbon cycle, are the current means through which projections of biogeochemical changes are made. For projection of carbon and climate change trends with ESMs, uncertainty stems from the three primary sources: (1) model response uncertainty (2) scenario or forcing uncertainty and (3) internal variability uncertainty [13].

Figure 1.2: Venn Diagram schematic of sources of uncertainty in simulating

Venn Diagram schematic of sources of uncertainty in simulating (using Earth-System Modeling approach) and observing changes in the Earth system. For emergence, detection or attribution of an observed or simulated signal to occur, the signal must overcome the sources of uncertainty in their respective brackets.

Firstly, model or structural uncertainty in projection arises from scientifically incomplete knowledge and model representation of the Earth system. The approxi-
mately 20 state-of-the art ESMs from different modeling institutions internationally have differing model constructions (i.e. different physical and biogeochemical representation and parameterizations) and as a consequence project different deterministic responses to anthropogenic forcing. For example, end of century global mean annual temperature change ranges from 3.2 to 5.4 °C [14]. These differences between the ESMs responses to anthropogenic forcing give a lower bound on the model or structural uncertainty inherent to projection.

Secondly, scenario uncertainty arises due to uncertainty in the pathway of future emissions of greenhouse gasses (GHG) and other climactically-important constituents. To represent this uncertainty, standardized representative concentration pathways (RCPs), which prescribe the evolution of atmospheric GHGs, aerosols, and land-use change, amongst other factors, have been developed for use by the climate modeling community [15]. The RCPs are constructed to provide a specific and consistent radiative imbalance across various the ESMs throughout the century, with the high-emissions or business-as-usual scenario, RCP8.5, producing 8.5 Watts m$^{-2}$ imbalance at year 2100. Two medium emissions scenarios, RCP6.0 and RCP4.5, and a low emission scenario, RCP2.6 complete the suite of 4 RCPs. Examining differences between projections using various emissions scenarios provides an estimate of scenario uncertainty.

Finally, internal variability uncertainty arises as a consequence of the chaotic nature of the atmosphere and coupled ocean-atmosphere system. The exact trajectory of the climate on annual to decadal timescales is extremely sensitive to the initial conditions or initial state of the simulation. Internal variability can be characterized by either (1) examining the variability of an unforced pre-industrial control simulation, in which changes in the simulated climate over time are due only to internal variability and can there be isolated quantified or (2) by use of an Initial Condition Large Ensemble (LE). The central idea with LEs is that the initial conditions only
need modest perturbations (e.g. $10^{-14}$ °C perturbations to a single grid cell; [16]) such that the climate modes and variability structures quickly become randomized between the ensemble members for any particular time slice. Differences between ensemble members is due only to natural variability, and as such it can be isolated and quantified. The advantage with the LE method for examining variability is that contemporary (anthropogenic) forcing or boundary conditions can be used. This allows for (1) the consideration of how anthropogenic forcing and natural variability coexist and coevolve and (2) the explicit removal of natural variability from the projection, such that the deterministic response of the ESM to anthropogenic forcing can be revealed [17].

Now turning to sources of uncertainty in observing anthropogenic change in the Earth system, we characterize three primary sources: (1) measurement error, (2) sampling error and (3) internal variability uncertainty [11].

The first two sources of uncertainty arise from the limitations on the capability of the observing system. Measurement error and/or imprecision confer uncertainty upon trends in the observational record. Sampling error arises from the fact that the measurements of the Earth are not sampled perfectly in space or time, but rather measurements are generally sparse in either space or time or both. This necessitates the use of gap-filling or map-filling interpolation or reanalysis methods to reconstruct missing temporal or spatial information. These methods induce additional uncertainty in the resulting data or data-based product.

The third source of uncertainty, internal variability uncertainty, arises from the aforementioned chaotic nature of the Earth System, and is a shared source of uncertainty between observational and modeling efforts to quantify climate change and its impacts. The climate-change literature invokes specific terminology to categorize simulated and observed changes or trends that overcome various sources of uncertainty. In climate simulations, a forced trend is *emergent* if it is statistically unlikely (e.g.
< 5% chance) that it was produced through internal variability alone. In the observational record, a trend is detected if the magnitude of the trend exceeds the uncertainty imposed by internal variability and measurement capability of the observing system. A detected trend can then be attributed to a given cause or forcing if the observed trend is consistent with model simulations which impose the hypothesized forcing, but inconsistent with alternative forcings and/or internal variability.

The timescales over which emergence and detection of forced trends is possible are of interest for numerous reasons. Emergence informs the timescales over which impacts of climate change might occur, as this is when the anthropogenic signal exceeds the natural variability to which systems and organisms are adapted. Detection timescales inform observing system design, determining timescales over which monitoring platforms must be sustained in order to identify forced trends. Detection timescales are also of potential socio-political importance, for example treaty verification and/or monitoring the impacts of future mitigation efforts.

The Time of Emergence (ToE) is the timescale over which or the point in time at which an anthropogenic signal is statistically distinguishable from internal climate variability (See Chapter 2 for documentation on methodology). For a perfectly resolved observing system, emergence and detection times would be equivalent. However, for an imperfectly resolved observing system, detection times exceed emergence times, due to additional uncertainty of sampling and/or measurement error which must be overcome for detection to be achieved. Therefore, ToE sets a lower-bound or minimum for monitoring timescales required for detection and attribution of anthropogenic trends. Chapter 3 evaluates the ToE for changes in the ocean carbon cycle using a LE from a single ESM.

ToEs are estimated from the projections of ESMs. As such, ToE estimates are subject to the sources of uncertainty inherent to projecting change in the Earth system. ToE is sensitive to the model-specific forced response, the model-specific internal
variability and the magnitude of forcing (which partially determines the magnitude of the forced response). Chapters 4 examines the model- and scenario- sensitivity of ToEs for anthropogenic changes in ocean biogeochemistry for a suite of ESMs and different warming scenarios.

ESM representation of natural variability uncertainty is considered in Chapter 5. For a handful of ocean properties, globally resolved observational records or data-based estimates now span timescales of order decades. This length of time-series, combined with the multiple realizations provided by Large Ensembles experiments, affords the opportunity to evaluate the ESMs on their simulated internal variability on interannual to multi-decadal timescales. Accurate ESM representation of the climate system’s ‘noise’ is important for accurate interpretation of the observational record, as ESMs are a primary tool within the detection-attribution process. If ESMs under-simulate decadal noise, this would result in early-biased ToE estimates and elevate the rate of type I errors (false positive) in detecting anthropogenic trends in the observational record. Previous studies have identified exactly this bias for the ESM-simulated variability of the physical state of the Pacific Ocean over recent decades [18] [19] [3] [20]. We evaluate the associated variability of the ocean carbon cycle over the same time period, finding the amplitude of GFDL-ESM2M’s decadal variability in air-sea carbon fluxes over the Equatorial and Subtropical Pacific regions to be smaller than data-based estimates. To address this model deficiency and explore its implications, a new, small ensemble with GFDL-ESM2M is presented, which imposes stronger trends in the trade-winds over the observational era. The wind ensemble produces trends in SST, SSH and air-sea carbon fluxes that are in agreement with data-based estimates. The new simulations allow for improved understanding of recent biogeochemical variability in the Pacific Ocean, and indicate that the recently observed Equatorial Pacific carbon source might in part be attributed to strengthening trade-winds.
Part I

Emergence of Anthropogenic Trends in the Ocean
Chapter 2

Large Ensemble and Emergence Methods

2.1 The GFDL-ESM2M Initial Condition Large Ensemble

The Large Ensemble experiment is conducted with the coupled Earth system model GFDL-ESM2M developed at the Geophysical Fluid Dynamics Laboratory [21][22] for which fidelity of the biogeochemical model (TOPAZ) has been documented for preindustrial [22], historical [2] and future [12] boundary conditions.

Figure 2.1: The GFDL Large Ensemble Experimental Design

Change in SST relative to the mean-state of the preindustrial control for ensemble member number 1 from years 1860-2100 (dark green), and for the envelope of possible climate trajectories produced by ensemble members 2-10 (1950-2100, light green), and the ensemble mean (dark blue, 1950-210).
We use a 30-member ensemble simulation over the period 1950-2100 (Figure 2.1), first presented in [23]. Each simulation follows a historical (1950-2005) and RCP8.5 (2006-2100) concentration pathway boundary condition with strongly restored atmospheric CO$_2$ concentrations. The initial conditions of ensemble members 2-30 are modestly perturbed through using the climate state (ocean, atmosphere, land, sea-ice) from January 2$^{nd}$-30$^{th}$ from year 1950 from the first ensemble member as the January 1$^{st}$ 1950 condition for ensemble members 2-30, respectively. Initial condition perturbation results in a randomization of internal modes of variability across ensemble members. By design, differences between ensemble members are solely due to internal variability, and similarities in the evolution of ensemble members over time are due to anthropogenic forcing.

2.2 Signal, Noise and Emergence

Changes in the contemporary climate are due to a superposition of anthropogenic forcing (the signal) and internal variability (the noise). Anthropogenic forcing generally produces monotonically increasing (or decreasing) change, i.e. the slow, upward progression of global temperatures. In contrast, internal variability produces oscillatory trends at various timescales, i.e. warming during positive ENSO and cooling during negative ENSO. As such, taking trends in climate data over long time horizons diminishes the impacts of internal variability (since negative and positive anomalies will average toward zero) and reveals the impact of anthropogenic climate change. Analogously, averaging trends across ensemble members also acts to mute the internal variability and reveal the anthropogenic trend. Once the anthropogenic trend is determined, it can be removed from the time-series, and the residual yields an estimate of internal variability. The essential advantage of the LE is that it provides the ability to separate the natural variability from anthropogenic trend on much shorter
timescales and finer spatial scales than would be possible with a single realization of the climate.

Figure 2.2: Maps of 30-year trends, Signals and Noise in SST

The linear 30-year trend in SST from the period 1990-2019 for (a) each of the 30 ensemble members (b) the signal [mean of the 30 members trends] (b) the noise [standard deviation of the 30 members trend] and (d) the signal-to-noise ratio [the magnitude of the signal divided by the noise].

A central question in the scientific communities that model, observe and plan for climate change is ‘When will the impacts of climate change be realized?’ Large Ensembles offer the statistical or probabilistic climate projections necessary to address this question. The common metric used is Time of Emergence (ToE). Time of Emergence (ToE) denotes the time at which a signal of interest is statistically distinguishable from background noise. The signal is the anthropogenic or forced trends, and the noise is trends that arise from natural or internal variability. The signal
is computed by averaging across the 30 trends (linear, least-squares) given by each LE. The noise is computed by taking the standard deviation of these 30 trends. The forced signal represents the common trend amongst the ensemble members. The noise represents the disagreement amongst ensemble members, which by design can only be due to natural variability. The signal-to-noise ratio (SNR), which is the magnitude of the signal divided by the noise, gives a measure of the significance of the anthropogenic trend relative to natural variability. As shown in Figure 2.2, 30-year trends in SST can differ significantly between ensemble members. For example, ensemble member 1 has sustained, strong warming in the equatorial and subtropical Pacific (5°C/century), whereas ensemble member 7 has moderate cooling. Upon averaging the trends produced by the 30 members, the anthropogenic signal is revealed (Figure 2.2b). Upon taking the standard deviation of the trends produced by the 30 members, an estimate of the noise is produced (Figure 2.2c).

Decadal trends in a stationary (unforced) climate system are for most purposes expected to be normally distributed about zero, however in a forced climate, the mean will be shifted to non-zero values. This is illustrated in Figure 2.3, which shows the 30-member time-series of global mean annual SST anomalies over the period 1990 to 2020, and the distribution (histogram) of trends that occur for 5-, 10- and 15-year trends. The signal is the mean of each distribution, and the noise is the standard deviation. The signals are indeed non-zero and positive, as expected in a warming world, and the trends are indeed normally distributed. Therefore, a standard 2-sided Student’s t-test can be used to test whether a given signal could be explained by natural variability alone. The null-hypothesis (that the signal could be due to natural variability) is rejected with >95% confidence when the magnitude of the signal is twice the magnitude of the noise, i.e. when the SNR equals or exceeds two. The ToE is the first year at which SNR ≥ 2.
Returning to Figure 2.3, the anthropogenic signal is not emergent after 5 or 10 years, but emerges sometime between 10 and 15 years. To pin-point the exact year, the SNR is computed for each year in time (Figure 2.4). Both the signal and the noise are larger at shorter trend lengths, as temporal smoothing has yet to reduce the noise, and the signal can be spuriously large due to randomly aligned natural variability between ensemble members. At year 12, however, the signal emerges above twice the value of the noise. The 2xNoise value represents a 95% confidence interval for how large decadal trends can be, i.e. 95% of the natural, decadal trends the ensemble produces will fall below this threshold. As such, we can be 95% confident that the trend in the ensemble mean is produced by anthropogenic forcing. After the ToE is met, at year 12, the signal stabilizes at a value of 1.5°C/century, whereas the noise continues to decline as longer-term trends converge across ensemble members (Figure 2.4a). As a consequence, the SNR increases throughout the period, and the signal becomes increasingly emergent and perhaps even detectable and attributable (Figure 2.4b).
Returning to Figure 2.2, looking at local or grid-cell, 30-year trends in SST, the anthropogenic signal is only emergent (SNR > 2) over the tropics and subtropics. The Southern Ocean has a weak trend combined with strong natural variability, making the region particularly non-emergent. The North Atlantic has a strong section of cooling, however the natural, multi-decadal variability is also strong, resulting in non-emergence over the 30-year period. On this 30-year time scale, the Equatorial Pacific, a region of pronounced interannual variability (due to ENSO), has unremarkable 30-year natural variability. In contrast, the areas of deep convection (North Atlantic and Weddell gyre) or of strong boundary currents (the Kuroshio and it’s extension region) exhibit the largest multi-decadal variability.

In the chapters that follow, all trend calculations are performed on annual means (unless otherwise stated) and started in year 1990, as this is the approximate beginning of the data-rich ocean carbon observing era. ToE calculations are performed at the grid-cell level (1° x 1°), regionally, and globally. At each grid-cell, there are 30 individual time-series. At each domain (either global or regional), first a single time-
series of the domain-averaged or integrated quantity is taken, providing 30 individual
time-series. From these individual time-series, the trends, signal, noise and ToE are
computed.

The regional bounds from the Regional Carbon Cycle Assessment Project (RECCAP)
protocol (http://www.globalcarbonproject.org/reccap/protocol.htm), are used. The Southern Ocean is defined as lying south of 45°S. The Arctic is defined
as the region north of 65°N. For the Pacific and Atlantic basins, North is defined as
18°N-65°N, Equatorial is defined as 18°N-18°S and South is defined as 18°S to 44°S.
For the Indian basin, North is defined as lying north of 0°N, and South is defined as
44°S-0°S.

2.3 Partitioning Uncertainty Through Intercomparison of Large Ensembles

Large ensemble simulations have been conducted separately with four Earth system
models (Figure 2.5): (i) GFDL-ESM2M, (ii) NCARs CESM1-GBC, (iii) MPI-ESM,
and (d) CanESM2. ESM2M is described by [21][22]. The large ensemble suite with
CESM1 is described by [16][25], with the marine biogeochemistry model described by
[26][27][28]. The MPI-ESM is described by [29], with the marine biogeochemistry
model described by [30], and the Large Ensemble first described [31]. For CanESM2
the model is described in [32][33][34].

The models have a number of important aspects that are similar, including similar non-eddy-permitting ocean resolution, surface ocean carbonate chemistry broadly
following protocols (OCMIP2; [35]), and similar formulations for primary production.
However, there are also a few important differences, including the types of phytoplankton represented, whether or not chlorophyll is explicitly and interactively
computed, and the particle aggregation, remineralization and sinking schemes.
For each of these models, at least 30 ensemble members have been run over the historical period spanning 1950-2005, and extended though 2100 with RCP8.5 forcing. For each of these models, an RCP4.5 extension was also performed, with either a large (> 30 members, GFDL and MPI), medium (15 members, CESM) or small (5 members, CanESM2) ensemble.

Figure 2.5: The GFDL, CESM, CanESM and MPI Large Ensembles

Time series of change in SST for (a) the GFDL Large Ensemble, (b) the CESM Large Ensemble, (b) the CanESM2 Large Ensemble and (d) 30 members of the MPI Grand Ensemble. For each ensemble, historical forcing is used over the 20th century, and starting in year 2006, separate ensembles are generated for RCP8.5 (darker shading) and RCP4.5 (lighter shading) forcing scenarios.

We use this suite of ESMs from multiple scenarios to partition projection uncertainty into contributions from model, scenario and internal variability uncertainty. By virtue of working with LEs, we are able to improve on or refine the methods considered in the original the definition of model uncertainty (e.g. Hawkins and Sutton 2009). We define model uncertainty \( U_M \) as simply the spread between the LE means \( \overline{LE} \), or the difference between the minimum and maximum LE mean for the historical-to-RCP8.5 forcing scenario. The ensemble mean from a Large Ensemble gives the forced signal of the given model, eliminating the need to fit a polynomial or assume a distribution, as was necessary methodology in pre-large-ensemble studies.

\[
U_M(t) = Max\{\overline{LE}_{m,85}(t, m = 1 : 4)\} - Min\{\overline{LE}_{m,85}(t, m = 1 : 4)\} \quad (2.1)
\]
Where $t$ is the years between 1990 and 2100 and $m$ denotes the 4 ESMs. We use the RCP8.5 LEs for two reasons: (1) this is the scenario with the most ensemble members available, at least 30 members for each ESM and (2) the larger forcing that persists through the century will reveal model differences more effectively than will the moderate and declining forcing scenario.

With a Large Ensemble, our definition of internal variability uncertainty differs from [13]. Internal variability uncertainty ($U_{IV}$) is given as the spread between the minimum and maximum ensemble member at a given year. This allows for changes in natural variability over time.

\[
U_{IV}(m, t) = \text{Max}\{LE_{m,85}(t, e = 1:30)\} - \text{Min}\{LE_{m,85}(t, e = 1:30)\} \quad (2.2)
\]

Where $t$ is the years between 1990 and 2100 and $m$ denotes the 4 ESMs, and $e$ denotes the ensemble members for each LE. We use the RCP8.5 LEs as this is the scenario with the most ensemble members available for each of the LEs.

Scenario uncertainty ($U_S$) is defined as the differences between the RCP8.5 and RCP4.5 multi-LE mean (the mean of the 4 LE means) for each scenario.

\[
U_S(t) = \text{Mean}\{\bar{LE}_{85}(t, m = 1:4)\} - \text{Mean}\{\bar{LE}_{45}(t, m = 1:4)\} \quad (2.3)
\]

Where $t$ is the years between 2006 and 2100 and $m$ denotes the 4 ESMs. Our choice of using RCP4.5 rather than RCP2.6 is motivated by the limitation of available RCP2.6 ensembles (namely that CESM only has a single simulation with RCP2.6 forcing).
Chapter 3

Emergence of Anthropogenic Signals in the Ocean Carbon Cycle

S. Schlunegger¹, K. B. Rodgers²,³, J. L. Sarmiento¹, J. P. Dunne⁴, T. L. Frölicher⁵,⁶, M. Ishii⁷ & R. Slater¹

¹Program in Atmospheric and Oceanic Sciences, Princeton University, USA
²Center for Climate Physics, Institute for Basic Science, Busan, South Korea
³Pusan National University, Busan, South Korea
⁴NOAA Geophysical Fluid Dynamics Laboratory, Princeton, USA
⁵Climate and Environmental Physics, Physics Institute, University of Bern, Switzerland
⁶Oeschger Centre for Climate Change Research, University of Bern, Switzerland
⁷Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan
Chapter Summary

Attribution of anthropogenically-forced trends in the climate system requires understanding when and how such signals will emerge from natural variability. We apply time-of-emergence diagnostics to a Large Ensemble of an Earth System Model, providing both a conceptual framework for interpreting the detectability of anthropogenic impacts in the ocean carbon cycle and observational sampling strategies required to achieve detection. We find emergence timescales ranging from under a decade to over a century, a consequence of the time-lag between chemical and radiative impacts of rising atmospheric CO$_2$ on the ocean. Processes sensitive to chemical changes emerge rapidly, such as impacts of acidification on the calcium-carbonate pump (10 years for the globally-integrated signal, 9-18 years regionally-integrated), and the invasion flux of anthropogenic CO$_2$ into the ocean (14 globally, 13-26 regionally). Processes sensitive to the ocean’s physical state, such as the soft-tissue pump, which depends on nutrients supplied through circulation, emerge decades later (23 globally, 27-85 regionally).

3.1 Introduction

The invasion of anthropogenic carbon and heat into the global ocean occur through a cascade of biogeochemical and physical change, which are coupled to the ocean’s carbon cycle [30]. The ocean carbon cycle, and in particular the ocean carbon pumps, redistribute not only carbon, but also nutrients, oxygen and organic matter between the ocean’s surface and the ocean’s interior, playing an important role in determining the mixing ratio of atmospheric CO$_2$ and the functioning of marine ecosystems [5][10]. As such, anthropogenic changes in the ocean carbon pumps and the ecosystem parameters and processes to which the pumps are coupled are critical for both future climate sensitivity and marine health [37][36][38][39][40].
The characterization of ocean carbon pumps arises from a conceptual framework originally proposed by Volk and Hoffert ([5]) which distinguishes three pumps: a solubility pump, operating via the increased solubility of CO$_2$ in cold (deep) waters, and two biologically operated pumps which export organic carbon (the soft-tissue pump) and inorganic carbon (the calcium carbonate pump) from the surface to depth. In the contemporary ocean, the sequestration or invasion of anthropogenic carbon is predominantly driven by rising atmospheric CO$_2$ concentrations (e.g. [8]), constituting a sub-set of the solubility pump that can be denoted the invasion flux [9].

Given the importance of the ocean carbon pumps for marine life, and of the invasion flux for the strength of the ocean carbon sink and climate, there have been significant efforts to quantify and detect changes in global carbon budgets ([11] [12] [13] [14] [15] [16] [17] and project changes in the pumps ([10] [18] [19] [20]). However, both model projections and the observational record of the coupled carbon-climate system are subject to uncertainties (Fig. 1.2). Natural variability uncertainty, the uncertainty stemming from natural variability inherent in the climate system, is a shared uncertainty amongst modeling and observational efforts and has been shown to be a significant source of uncertainty for assessing anthropogenic changes in the ocean carbon sink ([21] [25] [31]). Emergence denotes when a simulated anthropogenic trend 'emerges', in a statistical sense, above either natural variability (this work, [21] [22] [23] [24] [25] [26] [27] [28] [29]) and/or model uncertainty ([2] (Fig. 1.2). Detection denotes when an observed trend exceeds the uncertainty posed by both natural variability and measurement capability of the observing system ([30]. For a perfectly resolved observing system, emergence and detection times are equivalent. However, for an imperfectly resolved observing system, detection times exceed emergence times, due to additional uncertainty of mapping and/or measurement error which must be overcome for detection to be achieved. Therefore emergence timescales set a lower-bound or minimum for timescales required for detection and attribution.
In this study, we use an ESM as an observing system simulation experiment (OSSE) to quantify minimum detection time scales (i.e. time of emergence, ToE) for changes in the ocean carbon cycle. To estimate emergence timescales, we unravel natural variability from anthropogenic trends in the ocean carbon cycle using a 30-member Large Ensemble (LE) experiment conducted with GFDLs ESM2M (21, 22; see Chapter 2 for Methods). We reference all ToE calculations to year 1990, as this is the approximate beginning of the ocean-observing era 24 and therefore the start of the reference period from which contemporary anthropogenic trends can emerge.

The analysis centers around the three ocean carbon pumps, with complimentary tracers and processes to which the pumps are coupled, such as acidification, warming, oxygen, nutrients and ocean color, provided for mechanistic insight and connection to the observational record and observing system optimization. Additional model experiments which separate the rapid chemical vs. slow climate impacts of rising atmospheric CO$_2$ provide attributive insight into the anthropogenic drivers of changes in the ocean (Fig. A1). The resulting chronology of emergence provides a roadmap for when and why underway and imminent changes in important ocean biogeochemical processes and tracers might be detectable.

### 3.2 Overview of Emergence Chronology

Variables which reflect the accumulated or integrated invasion of anthropogenic CO$_2$ into the ocean, such as surface pH and pCO$_2$, emerge most rapidly, with 100% of the ocean area emerging within 10 years (Fig. 3.1), and the majority of regional signals emerging in under a decade (Fig. 3.2). The impact of acidification on calcium carbonate (CaCO$_3$) export follow closely behind, with timescales of local emergence between 20-30 years (Fig. 3.1), regional emergence between 9 years (the Southern Ocean) and 18 years (the Arctic).
A time series of the percent of the global ocean area with locally-emergent anthropogenic trends illustrates the disparity of emergence timescales for anthropogenic changes in the ocean carbon cycle. Omega applies to the saturation of both aragonite and calcite forms of calcium carbonate, for which emergence times are approximately equivalent. Exports are taken at 100 meters depth. Heat inventory is taken from 0-700 meters depth. Oxygen inventories are taken between 200 and 600 meters depth. Chlorophyll inventory taken for 0-500 meters depth. NPP is integrated from 0-100 meters depth. All other variables taken at surface.

Next to emerge, with local ToEs between 30-50 years and regional ToEs between 10 and 30 years, is sea surface temperature (SST), upper ocean heat content (integration between 0 and 700 meters depth), and carbon variables that are sensitive to both the physical and chemical state of the ocean: $\Delta pCO_2$ and air-sea CO$_2$ fluxes (Fig. 3.1 & Fig. 3.2). Upper ocean heat content emerges rapidly (4 years) on a global scale, consistent with a detection-attribution study for which data-based estimates of global upper ocean temperature increases emerged within a decade [57]. Southern Ocean SST presents an outlier, with ToEs extending beyond year 2100 (24). Non-emergence results from weak anthropogenic trends in SST (Fig. A4) attributed to the dynamical effects of surface freshening stabilizing the water column and decreasing convective heating from warmer subsurface waters, thereby offsetting the surface invasion of anthropogenic heat [58].
For $\Delta$pCO$_2$ and air-sea CO$_2$ fluxes, global signals emerge within 17 and 14 years respectively, and regional emergence is more homogeneous than for SST, with all regions emerging within the range of 13 – 29 years (Fig. 3.2). These emergence estimates are consistent with a detection attribution study which found reconstructed, global and regional air-sea CO$_2$ fluxes are emergent sometime within the 46 year period considered [59]. We interpret these to reflect emergence timescales for changes in the invasion flux as the impact of changes in the biological pumps on surface ocean pCO$_2$ is small (0.5 uatm, global average) relative to that of rising atmospheric and surface ocean pCO$_2$ (550 uatm, global average) between 1990 and 2100 (Fig. A1 & A2).

Figure 3.2: Global and regional year of emergence

![Global and regional year of emergence](image)

Global and regional year of emergence (after 1990) for globally- and regionally-integrated anthropogenic signals and 50% of local anthropogenic signals for the given biogeochemical variables. For each domain a single domain-averaged or domain-integrated timeseries for each variable is used to compute ToE. Color of each cell corresponds to the emergence year. Dashed boxes around the 3 ocean carbon pumps. Variables same as defined in caption of Figure 3.1

Last to emerge are variables tied indirectly to changes in the three-dimensional physical state of the ocean (i.e. circulation, ventilation, stratification), with these changes reflected also in biological processes (Fig. 3.1). For these variables, eumer-
gence of globally-integrated signals are considerably shorter than local emergence, with the soft-tissue pump (given by the export of particulate organic carbon - POC) emerging globally in 23 years, and with most regions under 50 years (Fig. 3.2). Emergence of surface chlorophyll, the primary observable currently used to monitor biological productivity and export, follows closely behind, with the global signal emerging after 25 years, and regional signals taking up to 8 decades for emergence, exceeding previously published, biome-scale ToE estimates by up to 4 decades [55]. If inventories over the upper 500 meters, rather than surface concentrations of chlorophyll are considered, emergence times decrease by 10 years for the global signal and by multiple decades for many regions (Fig. 3.2).

Emergence of O₂ inventories decrease to 17 years upon global integration, however the Arctic, the Indian Ocean and regions of the Pacific Ocean require more than 100 years to emerge (Fig. 3.2). The North Pacific stands out with only 20 years for emergence of O₂ inventories. The relatively early emergence of global and North Pacific oxygen inventories is consistent with finding of fingerprinting methods, for which observations of global and Pacific O₂ inventories over 20 years were found to be anthropogenically forced, whereas inventory changes in all other regions were indistinguishable from natural variability [60]. The thermally-driven components of oxygen trends (O₂,SAT) emerge more rapidly than the full O₂ signal, locally, globally and for most regions, whereas the non-thermal component of O₂ (AOU, apparent oxygen utilization) has equivocal ToEs (Fig. 3.1 & Fig. 3.2), with the exception of non-emergence of globally integrated changes in AOU, a consequence of compensating regional trends (Fig. A4).

Local emergence of over half the ocean surface area occurs this century (Fig. 3.1 & final column of Fig. 3.2), even for slowly emerging variables such as net primary production (NPP) and mixed layer depth (MLD). This highlights the value of time series observations in climate change monitoring efforts, particularly for the fields
that are not directly remotely observed (e.g. nutrients, O\textsubscript{2}). Broadly, integration over space (moving from the local to regional to global scale) reduces emergence timescales as noise decreases when averaging over compensating features of natural variability (e.g. East and West Pacific SST anomalies during ENSO events). For nearly all variables considered, the order of emergence is global, then regional, and finally local demonstrating the utility of observing networks with large spatial footprints for early trend detection.

3.3 The CaCO\textsubscript{3} Pump

The contribution of changes in CaCO\textsubscript{3} export to the sequestration of anthropogenic carbon is small (Fig A2), however changes in the CaCO\textsubscript{3} pump still represent a change in the ocean carbon cycle and a biological impact of climate change. We present emergence timescales for the export of CaCO\textsubscript{3} rather than the indirect chemical (buffering/alkalinity) contributions that impact surface pCO\textsubscript{2} and air-sea CO\textsubscript{2} gas exchange, as these are not explicit diagnostics in ESM2M.

Decreases in the export of CaCO\textsubscript{3} emerge rapidly, with about 50\% of the ocean being emergent within 30 years (Fig. 3.2, Fig. 3.3a & 3.4d), lagging approximately a decade behind its principal driver of declining $\Omega$, the saturation of CaCO\textsubscript{3}, which is critical for biological calcification. In ESM2M calcification rates are directly proportional to the degree of supersaturation of CaCO\textsubscript{3}, calcification immediately transforms into detritus, and dissolution does not occur in the upper 100 meters where waters are super saturated with respect to CaCO\textsubscript{3}. Therefore, it is changes in the production (and not dissolution) which are responsible for changes in export \[^{22}\]. The lag between declines in $\Omega$ and corresponding declines in the CaCO\textsubscript{3} export is due to contributions from the noisier co-drivers temperature and nutrients upon CaCO\textsubscript{3} formation and ultimately export.
Decreases in export of CaCO$_3$ are due entirely to the invasion of anthropogenic CO$_2$, and not changes in the physical climate (Fig. A1), consistent with the emergence times mirroring changes in carbonate chemistry rather than physics. Emergence of CaCO$_3$ export occurs sooner at the high latitudes (20 years) than the mid- and low-latitudes (30-40 years) (Fig. 3.3a), despite the decline in export being strongest at low-latitudes (Fig. 3.3d). This occurs because the spatial pattern of ToE for CaCO$_3$ export is strongly determined by the magnitude of decadal variability, which varies strongly by latitude (Fig. A4a). In ESM2M, decadal variability of CaCO$_3$ export scales with the magnitude of export (i.e., lowest at the high latitudes and greatest in the tropics, Fig. A5). As a consequence, the strongest signals and earliest emergence are anti-correlated (i.e. areas with the most pronounced trend emerge the slowest).

Figure 3.3: Time of Emergence and Signal Maps for the three carbon pumps.

Changes in CaCO$_3$ cycling result in changes in the buffering capacity of seawater and salinity-normalized alkalinity (nALK, Fig. A3 & A4). Changes in nALK represent an accumulated or integrated response to changes in export, and therefore emerge
prior to changes in CaCO$_3$ export. In a previous study, observational and interannual uncertainty for salinity and organic-matter-cycling normalized alkalinity were combined to consider ToE and detection of anthropogenic signals, showing detection times of 20-30 years for the low-to-mid latitudes, and in excess of 40 years at the high latitudes (56, and consistent with Fig. A3). Longer nALK ToEs in the high-latitudes results from the reduced exposure of upwelled waters to biogenic CaCO$_3$ cycling and is consistent with the relatively weaker high-latitude anthropogenic trends in CaCO$_3$ export shown here in Fig. 3.3d. Estimates of detection times are only modestly longer than our estimates of emergence times as contributions to uncertainty in trend detection from measurement error is of the same order magnitude as contributions from natural variability on decadal timescales.

### 3.4 The Invasion Flux

Emergence of air-sea CO$_2$ fluxes occurs at the high and tropical latitudes within 20-30 years, however the subtropics remain non-emergent throughout the twenty-first century (Fig. 3.3b). This is consistent with the emergence times and spatial patterns of LE simulations with NCARs CESM1-BGC [25]. Non-emergence of annual trends in air-sea CO$_2$ fluxes occurs over regions with weakly positive or even negative trends in exchange (Fig. 3.3e). The same pattern of emergence unsurprisingly holds for $\Delta$pCO$_2$, the primary driver of changes in solubility and air-sea carbon fluxes (Fig. 3.4a). In the subtropics, the annual mean trends in air-sea fluxes and sea-air $\Delta$pCO$_2$ are weak (Fig. 3.4b). However, this obscures significant increases in the amplitude of the seasonal cycle (Fig. 3.4d, 3.4f). The seasonal cycle of surface ocean pCO$_2$ is driven primarily through the seasonal cycles of dissolved inorganic carbon (DIC) and SST [61], but it is not changes in the seasonality of the drivers which result in the seasonal amplification of pCO$_2$ (i.e. for the subtropics, DIC seasonality decreases by 7% and SST seasonality increases by 4% over the 21st century). Rather, amplification of pCO$_2$ seasonality is
largely sustained through the cumulative effect of invading anthropogenic CO₂ upon the carbon dioxide buffering capacity of seawater, the Revelle factor. The impacts of reduced buffering capacity on the seasonal cycle of air-sea CO₂ fluxes and ΔpCO₂ finds maximum expression in summer over the subtropics (Fig. 3.4d), where the seasonal cycle is thermally dominated [62][63]. This is illustrated in SST-DIC phase-space (Fig. 3.4g), for which the seasonal cycle of pCO₂ at the subtropical location (35N) exhibits more amplification during the 21st century (its trajectory crosses more pCO₂ contours during the seasonal cycle) than the equatorial (0N) or subpolar (55N) locations.

Figure 3.4: Seasonality of ΔpCO₂ ToE and Signals

Time of Emergence for (5a.) annual (5c.) local summer and (5e.) local winter trends in sea-air ΔpCO₂ and the corresponding trend (signals) in same order (5b, d, f). July-September and January-March define summer and winter. Signals are the linear trend between 1990 and the ToE for each grid-cell. Panel 5g. shows the ensemble mean seasonal cycle of ΔpCO₂ at 3 locations along 160W (marked in 4f.) for year 1990 (solid) and year 2100 (dashed). The amplification of the seasonal pCO₂ cycle is shown on the maps and on the diagram of the season cycle at the 3 locations.

As a result of diverging seasonal trends, ToE in the subtropical convergence regions is significantly earlier for ΔpCO₂ and fluxes considered separately for winter and summer (Fig 3.4c & 3.4e) than for the annual mean (Fig. 3.4a). Observation-based
products of ocean surface pCO$_2$ demonstrate enhanced seasonality over the recent decades [64], which is consistent with the 30-year emergence timescales of seasonal trends in ΔpCO$_2$ (Fig. 3.4c & 3.4e).

3.5 The Soft-Tissue Pump

As with the CaCO$_3$ pump, the contribution of changes in the export of organic carbon to the sequestration of anthropogenic carbon is small (Fig A2). For most of the global ocean, reductions in the soft-tissue pump (POC export) emerge by the mid-to-end of century (Fig. 3.3c & 4f). These reductions are ultimately a consequence of the reduced supply of nutrients to the surface ocean resulting from slowly-emerging changes in ocean circulation and stratification ([50]; Fig. A3 & A4). POC export and surface chlorophyll broadly agree on both the pattern and timing of emergence (Fig. 3.3 & 3.5). For the two fields, ToEs agree (within 20% of each other, i.e. $|\text{ToE}_\text{POC}-\text{ToE}_\text{chlorophyll}|/\text{mean}(\text{ToE}_\text{POC},\text{ToE}_\text{chlorophyll}) < 0.2$) for 66% of the ocean area, and the underlying signal direction agrees for 87% of ocean area (i.e. decreased chlorophyll corresponding to decreased export). The agreement in timing and direction of changes in surface chlorophyll and soft tissue export supports the underlying assumption of field campaigns such as NASAs EXPORTS [46] that anthropogenic signals in ocean color do correspond to the strength of biological pump on decadal to centennial timescales, at least in ESM2M.

An exception to coupling between surface chlorophyll and soft tissue export occurs over the Southern Ocean, where export decreases despite increasing surface chlorophyll. During austral summer, iron limitation increases with depth over the 21st century, producing the divergent surface-subsurface trends in chlorophyll and productivity (Fig. A6). Chlorophyll integrated over the upper 500 meters, however, decreases in the Southern Ocean (Fig. 3.5c), consistent with reduced net primary production (NPP; Fig. A3, A4) and biological export (Fig. 3.3c & 3.4f). Disagree—
Figure 3.5: ToE and Signal Maps for surface vs. depth-integrated chlorophyll

ToE (a-b) and Signal Maps (d-e) for surface vs. depth-integrated chlorophyll. Surface trends are two orders of magnitude greater than depth integrated (0-500m) trends, however emergence of integrated chlorophyll is generally earlier due to the noise reduction that occurs with depth integration. The maximum on the color scale for chlorophyll inventory signal is $[4 \times 10^{-5} \text{ mmol m}^{-3} \text{ yr}^{-1}]$ and for surface chlorophyll is $[2.3 \times 10^{-3} \text{ mmol m}^{-3} \text{ yr}^{-1}]$. $[\text{NO}_3] = 0.5 \mu \text{mol kg}^{-1}$ contours imposed on panel d.

3.6 Emergence as a Lower Bound on Detection

Our emergence timescales provide a lower bound for detection timescales, for the following five reasons. Firstly, overcoming measurement and sampling error extends the duration of observational time-series needed for detection. Secondly, uncertainty in the methodology of emergence calculation, and the indication that alternative methods can produce longer ToEs in the case of surface chlorophyll and ToEs differing by $>20\%$ with the use pre-industrial rather than contemporary noise for a variety of ocean variables (Fig. A7).
Thirdly, we consider a high-emissions scenario, which would act to shorten ToEs relative to a lower-emission scenario, but only for slowly emerging variables like the soft-tissue pump. In contrast, ToEs for rapidly emerging variables like pH, CaCO$_3$ export and SST are scenario-insensitive, as these variables generally emerge prior to the separation between future scenarios (year 2006 formally, but an additional 2-3 decades for the impact of differential emissions to be evident in upper-ocean temperature [67]), indicating change induced from committed surface warming and acidification is sufficient for emergence.

Fourthly, uncertainty in model representation of natural variability could extend detection times. Model inter-comparison indicates ESMs show significant differences in natural variability estimates [68] and model-observation comparison indicate models such as ESM2M under-predict natural variability in the oceans physical state relative to the natural variability estimated from observational products [19][3]. Such insufficiencies in simulated variability have been shown to arise on interannual timescales and shorter timescales due to insufficient resolution to permit eddies [69] and decadal timescales due to inadequately strong atmospheric variability to which the ocean is coupled [3]. Suppressed natural variability (noise) implies emergence calculations would be biased early.

Fifthly, uncertainty in model response to anthropogenic forcing poses an additional scientific uncertainty for estimating ToE (i.e. other ESMs have different forced responses, internal variability and potentially emergence timescales). For variables like O$_2$ and NPP, inclusion of model uncertainty in the framework for calculating emergence timescales extends emergence estimates by decades [68] in comparison to the values presented here. Thus for the carbon pumps and drivers, its inclusion could significantly alter the ToEs presented here.
3.7 Conclusions

The three ocean carbon pumps considered have distinct spatial patterns of emergence, including rapid emergence of CaCO$_3$ export at high-latitudes and non-emergence of the annual-mean invasion flux in the sub-tropics. The three pumps have disparate ToEs, ranging from under a decade to over a century. This disparity reflects slower emergence for physical upper-ocean properties which determine emergence timescales for the soft-tissue pump, and more rapid emergence for the invasion of anthropogenic CO$_2$ and its biological impacts on calcification. The primary observables tied to each flux can emerge before (alkalinity preceding CaCO$_3$ pump), in tandem ($\Delta$CO$_2$ and the invasion pump), or after (ocean color lagging soft-tissue pump), further widening the gap between detection timescales for changes in the ocean carbon pumps.

Our results highlight the considerable observing system requirements for trend detection including high temporal and spatial resolution and multidecadal length sampling. For example, the analysis presented in this work shows that full seasonal resolution of surface pCO$_2$ and depth resolution of ocean color is critical to optimal observing system design. This LE OSSE is best understood as complementary to parallel OSSE efforts that consider constraints of observing platforms and address optimal spatiotemporal sampling strategy [70][71].

Another important challenge will be to apply the results derived here to better constrain mechanistic controls on marine feedbacks to the climate system, an important source of uncertainty in climate projection over the coming decades to centuries [72]. We present the emergence times for changes in the ocean carbon cycle induced by the summation of direct anthropogenic forcings and climate-carbon feedbacks. Distinguishing between the two, within an emergence framework, could provide timescales over which the magnitude of the ocean’s climate-carbon feedback could be observationally constrained, and contribute to the mechanistically-based framework for interpreting emergence timescales presented here.
3.8 Author Contributions

S.S. performed all analysis and writing, with continuous guidance from K.B.R, J.L.S, T.L.F, J.P.D, M.I. and R.S. The Large Ensemble simulations were set up by K.B.R and T.L.F and performed and post-processed by K.B.R. The sensitivity experiments and control runs were performed by R.S.
Chapter 4

Time of Emergence & Large Ensemble Intercomparison For Secular Trends in Ocean Biogeochemistry

S. Schlunegger\textsuperscript{1}, K. B. Rodgers\textsuperscript{2,3}, J. L. Sarmiento\textsuperscript{1}, T. Ilyina\textsuperscript{4}, J. P. Dunne\textsuperscript{5}, J. R. Christian\textsuperscript{6}, M. Long\textsuperscript{7}, Y. Takano\textsuperscript{4}, T. L. Frölicher\textsuperscript{8,9}, R. Slater\textsuperscript{1} \& F. Lehner\textsuperscript{7}

\textsuperscript{1}Program in Atmospheric and Oceanic Sciences, Princeton University, USA
\textsuperscript{2}Center for Climate Physics, Institute for Basic Science, Busan, South Korea
\textsuperscript{3}Pusan National University, Busan, South Korea
\textsuperscript{4}Max Plank Institute for Meteorology, Hamburg Germany
\textsuperscript{5}NOAA Geophysical Fluid Dynamics Laboratory, Princeton, USA
\textsuperscript{6}Canadian Center for Climate Modeling and Analysis, Victoria, British Columbia, Canada
\textsuperscript{7}National Center for Atmospheric Research, Boulder, USA
\textsuperscript{8}Climate and Environmental Physics, Physics Institute, University of Bern, Switzerland
\textsuperscript{9}Oeschger Centre for Climate Change Research, University of Bern, Switzerland
Chapter Summary

Anthropogenically-forced changes in the ocean biogeochemistry are underway and critical for the ocean’s role as a carbon sink and marine habitat. Detecting such changes in ocean biogeochemistry will require quantification of not only the magnitude of the change (anthropogenic signal) but also the natural background variability inherent to the climate system (noise). Here we use Large Ensemble (LE) experiments from four Earth System Models (ESMs) to estimate the Time of Emergence (ToE) and partition projection uncertainty for anthropogenic signals in five biogeochemically-important upper-ocean variables. We find ToEs are robust across models of different climate sensitivities for variables that are tied directly with the rise in atmospheric CO$_2$, namely rising sea surface temperature and the invasion of anthropogenic carbon into the ocean which emerge within 20-30 years. For the export of organic carbon, ocean color, and sea surface salinity, emergence timescales are longer (50+ years), less robust across the ESMs, and more sensitive to the forcing scenario considered. Our findings provide important lower bounds for the maintenance of a sustained multi-platform marine carbon observing system.

4.1 Introduction

The physical and biogeochemical state of the ocean determine the ocean’s habitability and capacity for sequestration of anthropogenic carbon. Rising temperatures, changing circulation and invading carbon threaten marine ecosystems and alter the cycling of carbon within the ocean – however due to uncertainty in the projections of future climate change, the timing and magnitude of these potential impacts is uncertain [10] [12] [39] [68]. Projections of future changes in climate and biogeochemistry are predominantly made through use of Earth System Models (ESMs), which are global climate models that include representation of an interactive carbon cycle. Uncer-
tainties in these projections stem from the following three sources: (1) the structural uncertainty associated with the different models used to make projections, (2) the societal uncertainty associated with different future emission pathways and (3) the internal variability uncertainty associated with natural fluctuations of the coupled climate system [13].

Time of Emergence (ToE), or the timescale over which an anthropogenic or forced trend in the climate system emerges above the uncertainty induced by internal variability. ToE is a multi-purpose metric which stands to (1) inform observing system design by providing a baseline for monitoring time scales required for trend detection (2) inform impact timescales, since systems or organisms are likely to ‘feel’ anthropogenic change once it exceeds the envelope of natural variability to which they are adapted and (3) intelligently normalize anthropogenic responses across disparate variables, across different ESMs, and across a spectrum of forcing scenarios — providing a framework for model, scenario and impact intercomparison.

ToE can be estimated with projections made from Initial Condition Large Ensemble (LE) experiments of ESMs [23][25][53][11]. The central idea with LEs is that the initial conditions only need modest perturbations such that the climate modes and variability structures quickly randomize between the ensemble members for any particular time slice. Differences between projections of ensemble members are solely due to natural variability, and therefore natural variability can be identified and separated from the ESM’s forced response [17]. Over the duration of the projection, the magnitude of the forced response (signal) may become statistically distinguishable from internal variability (noise). This point in time defines the ToE.

Previous studies have shown that anthropogenic changes in different ocean properties exhibit vastly different timescales of emergence [54][73][68][11]. Schlunegger [11] provides mechanistic insight as to why anthropogenic trends for disparate components of the ocean carbon cycle emerge over such different timescales, illustrating with an
LE of a single ESM (GFDL-ESM2M) the relationship between ToEs for the physical and biological pumping of carbon within the ocean and to a broader suite of related variables. Variables which represent the integrated effect of invading anthropogenic carbon into the global ocean, like acidification, emerge most rapidly, with ToEs of only a few years. A few decades later, changes in sea surface temperature (SST) and in air-sea CO$_2$ fluxes emerge. The delay between acidification and warming/invasion is due to the time-lag between the chemical and radiative consequences of rising atmospheric CO$_2$ and the susceptibility of SST and surface pCO$_2$ to natural variability. Changes in the physical state of the upper ocean, including upper ocean mixing, and associated changes in biological processes, like the export of organic matter which depends on nutrients supplied through mixing, only emerge after many (5+) decades.

Does this sequencing of emergence timescales characterized by a single ESM hold true across LEs of different ESMs or different forcing scenarios? Here, we use LEs from four ESMs with two forcing scenarios to explore the model and scenario sensitivity of ToE as well as test the hypothesis that the chronology or sequencing of emergence times for changes in ocean biogeochemistry is robust across a representative suite of four ESMs. We then utilize the multiple LEs and multiple scenarios to partition contributions from sources of uncertainty — scenario, model, internal — in projecting biogeochemical change in the ocean. Previous studies which evaluated uncertainty in projections of ocean biogeochemical variables have a number of inconsistencies across the studies in terms of the definition of noise (e.g. preindustrial vs. contemporary, single-model vs. CMIP5 mean) and filtering to retrieve the forced signal (e.g. temporal smoothing techniques vs. 4th order polynomial fitting), making comparisons across studies difficult. A central objective here is to take advantage of the opportunity offered by the LE approach to provide a consistent and unifying framework for evaluating trend detection and projection uncertainty.
We focus the analysis on five observable or observationally-constrained biogeochemical variables that impact the cycling of carbon within the global ocean: sea surface temperature (SST), air-sea CO$_2$ flux, export of organic carbon from the surface ocean, sea surface chlorophyll, and sea surface salinity (SSS). SST, observable by satellite, impacts carbon through setting the solubility of CO$_2$, through contributions to stratification, density and overturning, and through the temperature dependence of biological activity and the export of biological carbon to depth. The flux of CO$_2$ at the air-sea interface allows for the invasion of anthropogenic carbon into the global ocean and is observationally constrained through the use of shipboard surface pCO$_2$ measurements and interpolation in space and time through known relationships with more densely observed oceanographic properties like SST, sea surface height (SSH) and ocean color [47] or through ocean or atmospheric inversion models [74]. The biological export of carbon to depth and its primary associated observable, ocean color, provide a means to monitor the impact of climate change on the ecologically important transfer of energy from the base of the marine food web and the associated climatologically important transfer of carbon from the upper ocean to depth [46]. Finally, SSS, monitored via remote sensing (since year 2010; [75]), the Argo program (since year 2000; [76]), and ship-board measurements (reliably since the 1970s; [77]), provides a means to monitor the impact climate change has on freshwater fluxes and ocean circulation, important drivers of the cycling of carbon in the ocean.

Numerous observational programs with the intended goal of monitoring changes in the ocean’s cycling of carbon are currently underway, and have observational records which extend 20-30 years. We note in particular (1) the Regional Carbon Cycle Assessment Project (RECCAP; [74]) for which the last phase of analysis focused on Air-Sea CO$_2$ fluxes over the 20-year period 1990-2009 and forthcoming analysis will focus on the 30-year period 1990-2019, and (2) NASAs EXport Processes in the Ocean from Remote Sensing (EXPORTS; [46]) for which the duration of globally
resolved observational constrains now exceeds 20 years. In the final section of this work, as a means to directly facilitate interpretation of the observational record, we utilize the multiple Large Ensembles to provide confidence intervals for emergence of anthropogenic signals over the observational period.

All methods for this chapter are contained within Chapter 2.

4.2 Results

4.2.1 Signals and Time of Emergence

Mean State Changes

For SST (Figure 4.1a), the four LEs span the spread of CMIP5 given in [12] climate sensitivity, with the between 2°C (GFDL) and 3.5°C (CanESM2) warming by end of century, relative to 1990 temperatures. For globally-integrated air-sea CO$_2$ flux (Figure 4.1b), the four LEs spread the range of RCP8.5 CMIP5 uptake given in [19], for which the ocean takes up an additional 2.5 PgC/yr (CanESM2) to 4 PgC/yr (GFDL) by end of century. We note that the inverse relationship between SST and air-sea CO$_2$ fluxes is consistent with what is identified with the transient climate sensitivity of [6].

Figure 4.1: Global Changes in the 4 Large Ensembles, RCP8.5 Forcing

Global annual changes relative to year 1990 for a. global mean sea surface temperature (SST), b. globally integrated Air-Sea CO$_2$ Flux, c. POC Export, d. surface chlorophyll and e. sea surface salinity (SSS). Values given in the left-hand column are global mean in year 1990 and in the right hand column are the 95% confidence intervals for magnitude of global natural decadal trends for each LE in units of a. [°C/decade] and b.-c. [PgC/decade] d. [mg Chl/m$^3$/decade] and e. [practical salinity units/decade]. Larger values indicate stronger decadal variability over global domain.
Global POC export declines for all models, however the magnitudes of decline is model-dependent (Figure 4.1c). For POC export, declines range between 0.5 and nearly 2.0 PgC/yr by end of century. A subset of the CMIP5 models is used in [50] to project and attribute changes in POC flux (Figure 4.1c), including GFDL and CESM shown here. For POC export we note the pronounced variability of CanESM2, which is partially attributable to the larger baseline export (∼11 PgC/yr relative to ∼8 PgC/yr for the other ESMs, but also differences in model formulation.

Globally averaged surface chlorophyll concentrations decline for all models except CanESM2, for which global chlorophyll concentrations increase modestly, as a residual of regionally heterogeneous trends (Figure B2 and B7g). Pronounced variability and decline of surface chlorophyll in the MPI LE is related to the mean state bias (4 times higher chlorophyll concentrations relative to the other 3 models), which arises from chlorophyll in MPI-ESM1.1LR being diagnostic, rather than prognostic. For the case of CanESM2, the rise in globally averaged surface chlorophyll concentrations occurs despite the decrease in biological export – complicating the use of ocean color observations to infer changes in biological export.

The four LEs freshen over the 21st century, which is qualitatively consistent with CMIP5 projections, in which increased precipitation (freshening) over the Pacific basin overwhelm increased evaporation (salinification) of Atlantic basin [78]. The magnitude of global freshening scales with the magnitude of global warming for the four LEs considered here.

Global and Regional ToE

When do global and regional changes emerge above natural variability, for each of the LEs? The ToEs referenced to the year 1990 for each LE, for SST, Air-Sea CO₂ Flux, POC Export, and surface chlorophyll concentrations and SSS are given in Figure 4.2. Globally, and for most regions, SST emerges between 10 and 20 years, then air-sea
CO$_2$ flux between 20 and 30 years. The Southern Ocean is the only region with model disagreement exceeding a decade. The GFDL model projects a weak cooling trend over this region, which does not emerge during this century, and thus disagrees with the ~20-30 year emergence times for significant warming estimated by the other LEs (Fig. B4).

Figure 4.2: Global and Regional Time of Emergence

Air-sea CO$_2$ exchange is the variable with highest agreement between the LEs, for which all LEs are within approximately a decade of each other, even for hotspots of variability like the Southern Ocean, North Atlantic and Equatorial Pacific. For Air-sea CO$_2$ fluxes and SST, the global and regional emergence times for the RCP4.5 and RCP8.5 scenarios are equivalent. This is due to the fact that emergence for these variables occurs prior to the separation between the impacts of differential forcing scenarios on the ocean [67].
Global changes in POC export emerge between 25 and 40 years, however regional changes have longer timescales of emergence (40-110+ years) and larger disagreement between LEs (20 to 60+ year disagreement depending on region). Differences in emergence times for POC export arise from a combination of different warming rates, natural background variability and sensitivity of biological processes (including primary production, grazing, particle aggregation, sinking and remineralization) to a changing ocean. Surface chlorophyll concentrations also take many decades to emerge and the model spread is even greater than for POC export. For most regions, surface chlorophyll emergence times lag export emergence by a few years to a decade.

Global freshening emerges within 20-30 years for CESM, CanESM2 and MPI and within 50 years for GFDL. Regional emergence times, even for areas of stronger trends, like in the Atlantic, still require at least a few decades, and the models diverge by more than 50 years, despite overall model agreement on the sign and magnitude of the trend (Figure B2).

For biological export, chlorophyll and SSS, emergence times for an RCP4.5 scenario can extend decades beyond those of an RCP8.5 scenario, depending on the region. This indicates that in some regions, mitigation delays the impacts of climate change, and the benefit stands to be observationally verifiable.

**Local ToE**

Now we consider emergence and signals at the local scale (Figures 4.3-4.5). Upon global aggregation of local ToEs (Figure 4.3) the LEs agree within approximately a decade upon the pace or rate of local emergence, with the exception of slowed emergence at end of century for GFDL SST (Figure 4.3a) and for chlorophyll and SSS for GFDL and MPI (Figure 4.3d & de). Slowed emergence at end of century for GFDL SST is a consequence of a non-emerging Southern Ocean, due to weak cooling in this region ([58]; Figure B4).
Figure 4.3: Pace of Local Emergence

Pace of local emergence (percent of global ocean area emerged at each year) for RCP8.5 (solid lines) and RCP4.5 (dashed lines) scenario. CanESM and CESM1-BGC excluded from RCP4.5 local scale emergence presentation of air-sea CO$_2$ flux, POC flux and chlorophyll due to insufficient ensemble size.

Air-Sea CO$_2$ Fluxes have strong agreement between the LEs not just upon global and regional integration (Figure 4.2), but also locally (Figure 4.3b) and spatially (Figure 4.4b). All LEs share the common feature of non-emergence in the Ekman convergence regions of the subtropical gyres (Figure 4.5b), as previously shown for CESM [25] and GFDL [11]. Non-emergence of annual trends arises from the superposition of opposing seasonal trends: summertime outgassing and wintertime ingassing [11].

Figure 4.4: Maps of Multi-model mean Time of Emergence

The mean emergence time of the four LEs. White hatching indicates pixels where the spread (standard deviation) of the 4 LEs ToE is more than half the mean ToE, and red stippling indicates pixels where at-least 2 of the 4 LEs are non-emergent at end of century. For averaging purposes, year 2100 was used when emergence was not achieved for a given LE.

POC Fluxes, despite showing agreement across the LEs in pace of local emergence (Figure 4.3c), do not agree on which localities are emerging (Figure 4c, Figure B6), or the magnitude or direction of the signal that is emerging (Figure 4.5c, Figure B6). For example, in the Pacific cold tongue, GFDL has a mixture of weak and non-
emergent negative and positive trends, CESM has positive trends, and CanESM2 and MPI have negative trends (Figure B7). Similarly, the Southern Ocean has divergent trends, with declining POC export for GFDL, increasing POC export for CanESM2 and non-emergent local trends for MPI and CESM. Chlorophyll, like POC export, exhibits long timescales of local emergence, and significant model disagreement upon the timing of emergence, which localities emerge, and with what signal magnitude and direction (Figure 4.4, 4.5, B7).

For SSS, the models agree on the underlying features of the signal – namely salinification of the tropical and subtropical Atlantic and southern subtropical gyre of the Pacific Ocean, a freshening of the Arctic and North Atlantic, and weak freshening of the Equatorial and North Pacific, Indian and Southern Ocean (Figure 4.5e, B8). Despite agreement on the signal’s spatial pattern, the LEs do not agree on local ToE (Figure 4.3e, 4.4e), as a consequence of the different signal magnitudes (larger for CanESM2 and MPI) and different noise (Figures B4-B8).

4.2.2 Partitioning Uncertainty in Projections

We now turn to partition the sources of uncertainty evident in the LE projections (Figure 4.6 & Figure B10). Partitioning uncertainty in projected change is of interest
since the different sources of uncertainty convey different consequences or implications. For example, if structural uncertainty is relatively large then scientific advances in the understanding and modeling of the Earth system would provide improved projection skill. Alternatively if scenario uncertainty is large, this indicates that societal decisions are important for the given outcome. If internal variability uncertainty is large, this indicates the change may not be discernable from background noise, and neither improved models nor differing societal decisions will significantly alter the projection uncertainty. Large internal variability uncertainty also implies that detecting the given response of the Earth system would require sustained observational platforms, and potentially organisms and systems have increased tolerance or resilience to change as the envelope of variability to which they are adapted is large relative to the differential impact of the considered forcing scenarios.

Figure 4.6: Partitioning uncertainty

Partitioning uncertainty for a. SST, b. Air-Sea CO$_2$ Flux, c. POC Export, d. Chlorophyll and SSS, for scenario uncertainty (red, RCP4.5 vs RCP8.5), model uncertainty (green shading) and internal variability (yellow shading). The contribution of natural variability from each ensemble is given by the colored lines, same model-color relation as previous figures. The interface between yellow and green is determined by maximum contribution from internal variability, i.e. the model with the largest internal variability at that point in time. Ten year box filter was applied to the estimates of natural variability.

Consistent with previous work of [68] and [49], at global scales, scenario uncertainty dominates SST and Air-Sea CO$_2$ fluxes (i.e. model and internal variability are relatively small, Figure 4.6). This is also reflected in the generally robust ToE estimates given for SST and Air-Sea CO$_2$ fluxes in Figures 4.2-4.6. For the biologically mediated carbon flux, however, model uncertainty dominates at the global scale, and
depending on region, either internal variability or model uncertainty dominates at the regional scales, with scenario uncertainty only contributing an average of 19% at end of century (Figures B10). For SSS, model uncertainty dominates at both the global and regional scale (55% globally and 47% on average over the regions, Figure 4.6 & Figure B10).

With the use of multiple LEs, we have multiple estimates of the contribution of internal variability uncertainty to total uncertainty, which is given by the individual colored lines within the yellow 'internal' sections of each panel. Previous methods used either the noise from a pre-industrial control run of a single climate model or ESM, or an average across single ensemble members from multiple ESMs, or the noise from a LE form a single ESM. However here, we can provide noise estimates from contemporary LEs from multiple ESMs. This provides estimate of the uncertainty of internal variability uncertainty, which is represented by the spread of the four LEs’ individual lines.

For SST the contribution of internal variability to prediction uncertainty is equivalent across the four LEs – about 15% by end of century. For Air-Sea CO₂ fluxes, the contribution of internal variability ranges from 75% (CanESMs) to 45% (CESM) at the beginning of the century. For the biological carbon flux, contributions range from 30% (GFDL and CESM) to 90% (CanESM2) at the beginning of the century and 25% (GFDL and CESM) to 50% (CanESM2) at mid-century. For both Air-Sea CO₂ fluxes and POC fluxes, as model and scenario uncertainty increase throughout the century, the contribution of internal variability uncertainty decreases, as does the importance of the differences between individual ESMs estimates of internal variability, i.e. convergence of the 4 colored lines at end of century is a consequence of the decline in the contribution of internal variability uncertainty rather than the convergence of the noise estimates given by the four LEs.
At the regional scale, however, differences in the internal variability uncertainty provided by the various models remains significant throughout the century for many regions, for Air-Sea CO$_2$ fluxes, POC flux and chlorophyll (Figure B10). For the Southern Ocean, in particular, the contribution of internal variability uncertainty at year 2100 for POC export ranges from $<5\%$ (GFDL) to $60\%$ (MPI) and for chlorophyll from $<5\%$ (GFDL) to $51\%$ (Figure B10), illustrating the importance of the estimation of natural variability for the partitioning of uncertainty in projection of future ecosystem function.

### 4.2.3 Signal-to-Noise for trends over the observational era

Considering ToE from the same start date (1990) across variables (Figures 4.2 to 4.5) is necessary for scientific and mechanistic interpretation of intrinsic detectability and impact timescales across the spectrum of variables. However, in order to directly aid interpretation of the observational record, we include Signal-to-Noise Ratios (SNR) for local, regional and global trends over the period for which globally resolved observations or data-based estimates are available for each of the 5 variables (Figure 4.7) and for the general period of 1990-2019 as a means to equitably compare SNRs across variables (Figure 4.8).

The first full year of satellite estimates of globally resolved SST was 1979, and therefore we provide the SNR for SST the 41 year period 1979 to 2019. Most of the available data-based products for air-sea CO$_2$ fluxes [47] start on or before year 1990, and so for air-sea CO$_2$ fluxes we provide SNRs for the 30 period 1990-2019. The first full year of satellite estimates of globally resolved surface ocean chlorophyll measurements was 1998, and therefore we provide the SNR for both surface chlorophyll and POC export for the 22 year period 1998-2019. For SSS, the year 2000 marks the beginning of the ARGO program, the first continuous, near-global salinity observing system [79] and so we consider SNR for SSS over the 20 year period 2000-2019.

53
Maps of mean Signal-to-Noise Ratio for the 4 Large Ensembles, for local and regionally-integrated signals. White hatching over locations of where LEs disagree [mean of SNR of 4 LE is less than the standard deviation of SNRs across models] Number of years that SNR ratio is taken given in parenthesis. The end-year of the trend for each variable is 2019. For SST, 41-year trends for 1979-2019, for air-sea CO$_2$ flux, 30-year trends for 1990-2019, for POC export and surface chlorophyll, 22-year trends for 1998-2019 and for SSS, 20-year trends from 2000-2019. The LE mean Global SNR ratio the standard deviation across the LEs given below the maps for each variable.

SST is emergent for the globally-integrated signal, for all regionally-integrated signals, and for over half of the global area for local (1°x1°) signals, consistent with the scientific consensus that global and regional warming of the upper-ocean from mid-century to present is attributable to anthropogenic forcing [80]. Air-sea CO$_2$ fluxes, although only emergent for 19% of global area when local signals are considered, are emergent for all regionally and globally-integrated signals. The subtropics stand-out with particularly low (<1) SNR ratios. Export, chlorophyll and SSS have no local, regional, or global emergence over the observational period, motivating their sustained monitoring as the current observational time-series is insufficient to robustly capture anthropogenic change. However, we note the ESMs with high-sensitivity, CESM and CanESM, have emergent POC and/or chlorophyll trends in the Atlantic ocean (Figure B11 & B12).

As a complement to Figure 4.7, we show SNR for a consistent time period, 1990-2019 (Figure 4.8). The 11-year reduction in the trend duration for SST decreases local emergence from 66% to 29% of the global ocean area. The duration of the trend and resulting SNRs are identical for Air-Sea CO$_2$ fluxes in Figure 4.7 and Figure 4.8. For
Figure 4.8: Maps of Signal-to-Noise Ratio for the 4 Large Ensembles, 1990-2019

Maps of mean Signal-to-Noise Ratio for the 4 Large Ensembles, for local and regionally-integrated signals. White hatching over locations of where LEs disagree [mean of SNR of 4 LE is less than the standard deviation of SNRs across models]. Number of years that SNR ratio is taken given in parenthesis. For each variable, SNR is given for the 30-year period 1990-2019. The LE mean Global SNR ratio the standard deviation across the LEs given below the maps for each variable. The percent of ocean area with SNR > 2 shown on upper right corner of each map.

POC export, Chlorophyll and SSS the increase in observational period by ~10 years results in the regional emergence of the North and Equatorial Atlantic for export and chlorophyll and for the South Atlantic for SSS. This indicates that approximately a decade more of observations of these fields could result in the detection of regionally emergent anthropogenic trends. Furthermore, the LE-mean global signals for these three fields become emergent during the 30-year observational window.

Finally we comment further on Air-Sea CO$_2$ fluxes, and the difference between emergence between the initial 20-year time period considered in RECCAP (1990-2009), and the 30-year time period which will soon be available. 30-year trends enable the emergence (at the 99% confidence level) of all the large regions considered (Figure 4.7), however if only a 20-year trend is considered, less than half of the regions are emergent, and local emergence reduces from 20% of the global area to 0% (Figure B15).
4.3 Discussion and Conclusions

For fields like SST and POC Flux the time of emergence is more robust across models than the magnitude or direction of the trend – meaning that the models considered here indicate that the timescales over which anthropogenic signals emerge has more certainty than the characteristic of the underlying signal. Despite uncertainty in what signal will emerge, there is certainty in how long, at minimum, we must monitor in order to detect such a signal. The models also agree on which variables do and do not exhibit different emergence timescales with mitigation. For variables that emerge rapidly, like SST and Air-Sea CO$_2$ fluxes, historical emissions and committed warming is sufficient to produce an emergent signal prior to impacts of climate mitigation efforts. However, for variables which emerge more slowly, like POC flux, chlorophyll and SSS, ToEs are scenario-sensitive, and can be delayed by multiple decades with moderate climate mitigation efforts in line with the Paris Agreement of a 1.5°C target.

Our results highlight an apparent paradox, in that the scenario-sensitivity of ToEs and scenario-uncertainty in projection are inversely related, e.g. that variables whose ToEs are insensitive to mitigation (SST, Air-Sea CO$_2$ flux) have the largest, future scenario uncertainty. However, these findings are indeed consistent, as variables which are sufficiently sensitive to emissions emerge early, prior to the impact of differential scenarios, such that the scenario considered does not alter ToE but do strongly influence the evolution of the signal over the 21st century.

We find the chronology of emergence to be consistent amongst the four LE, i.e. SST first, followed by Air-Sea CO$_2$ flux, biological export, surface chlorophyll concentrations and finally SSS. We interpret this chronology to reflect the time-lag between the underlying drivers of the change. Rising CO$_2$ warms the atmosphere, the surface ocean, and chemically-induces a net positive air-sea CO$_2$ flux trend for most regions. Changes in ocean circulation subsequently and slowly result from the invasion of anthropogenic heat and changes in freshwater fluxes – with changes in circulation,
stratification and/or transport altering biological activity and export through model-
dependent pathways [50].

Finally, the analysis indicates that the current observational record is long enough
to identify global and regional anthropogenic trends in SST and Air-Sea CO₂ fluxes,
the properties associated directly with rising CO₂ concentrations and atmospheric
temperature. For the attribution of local trends, however, the duration of observa-
tions is insufficient, particularly given the additional uncertainties associated with
observations (like measurement error and gap-filling) which we do not consider here.
For Air-Sea CO₂ fluxes, the LEs agree that within the last decade, the duration
of observational record has surpassed a critical threshold for regional emergence of
anthropogenic trends.

For the POC export, surface chlorophyll and SSS, properties indirectly associated
with rising atmospheric CO₂ concentrations, the observational record is likely insuf-
fficient for anthropogenic trends to be identified. However, in the coming decade, the
Large Ensembles agree that regional trends in biological activity, export and salinity
could begin to emerge.

4.4 Author Contributions

S.S. performed all analysis and writing, with guidance from K.B.R, J.L.S, T.I., T.L.F,
J.P.D, J.C., M.L., and R.S. The Large Ensemble simulations for GFDL RCP8.5 were
set up by K.B.R and T.L.F and performed and post-processed by K.B.R. The Large
Ensemble simulations for GFDL RCP4.5 were set up by S.S and R.S, performed by
S. S. and post-processed by K.B.R. Y.T. performed and post-processed the MPI-GE
RCP8.5 simulations.
Part II

Natural Trends in the Pacific Ocean
Chapter 5

Strengthening Trade Winds & an Enhanced Equatorial Pacific Carbon Source

S. Schlunegger\textsuperscript{1}, T. L. Frölicher\textsuperscript{2,3}, K. B. Rodgers\textsuperscript{4,5}, J. L. Sarmiento\textsuperscript{1}, J. P. Dunne\textsuperscript{6}, R. Zhang\textsuperscript{6}, T. Delworth\textsuperscript{6}, & R. Slater\textsuperscript{1}

\textsuperscript{1}Program in Atmospheric and Oceanic Sciences, Princeton University, USA
\textsuperscript{2}Climate and Environmental Physics, Physics Institute, University of Bern, Switzerland
\textsuperscript{3}Oeschger Centre for Climate Change Research, University of Bern, Switzerland
\textsuperscript{4}Center for Climate Physics, Institute for Basic Science, Busan, South Korea
\textsuperscript{5}Pusan National University, Busan, South Korea
\textsuperscript{6}NOAA Geophysical Fluid Dynamics Laboratory, Princeton, USA
Chapter Summary

We find that data-based products for ocean-atmosphere CO$_2$ fluxes over the observationally-rich period 1990-2010 show a pronounced trend towards stronger outgassing of CO$_2$ over the Tropical Pacific. This is consistent with the dynamics of a persistent La Nina-like state over the second half of this period, over which enhanced upwelling in the East and Equatorial Pacific brings carbon rich waters to the surface, increasing pCO$_2$ and thus Sea-to-Air carbon fluxes. We hypothesize that enhanced CO$_2$ -source behavior results from the dynamical impacts of strengthening trade winds on subtropical overturning, as identified previously in the climate dynamic literature (e.g.,[18] [19] [3] [81]).

To attribute the strengthening source of carbon to the strengthening trade winds we have conducted a suite of wind-substitution (‘data-override’) experiments with GFDL’s ESM2M model over the period 1979-2017. For the wind-override experiments, wind stress anomalies from the ERA-Interim reanalysis product have been imposed over the model’s Equatorial Pacific domain. The new simulations (1) provide better agreement between simulated and observed trends in the Pacific basin for not only air-sea CO$_2$ exchange, but also ocean temperatures, sea level, and the subtropical shallow overturning circulation (2) allow for improved understanding of recent biogeochemical variability in the Pacific Ocean, and (3) indicate that the recently observed Equatorial Pacific carbon source might in part be attributed to strengthening trade-winds.

5.1 Introduction

Since the industrial revolution, the ocean has sequestered $\sim$30-40\% of the carbon humans have emitted to the atmosphere [82] [11], thus mitigating the rate of increasing atmospheric CO$_2$ and anthropogenic climate change. Understanding how the ocean
carbon cycle has changed over the observation-rich period 1990-2009 is an important part of the broader task of quantifying uncertainty in ocean carbon uptake. Substantial efforts have been made to monitor the ocean carbon cycle and develop synthesis products to estimate trends in air-sea CO₂ exchange, as reflected in the Regional Carbon Cycle Assessment and Processes (RECCAP) [83] synthesis project as a part of the Global Carbon Project (GCP) [84] and the Surface Ocean pCO₂ Mapping intercomparison (SOCOM) [47].

During this relatively observation-rich interval 1990-2009 for carbon measurements, the physical climate system was observed to undergo a regime shift of the Pacific Decadal Oscillation (PDO) resulting in a substantial decadal trend towards La Nina-like condition in the Equatorial Pacific [85]. This manifested itself in enhanced overturning of the Pacific subtropical cells [85], decreasing sea surface temperatures (SSTs), and increasing wind stress over the region, with these thought to reflect natural variability rather than a secular trend [18] [19] [20]. Scientific inquiry in the climate modeling community arose when the amplitude of these observed trends was outside the range of trends projected by the full suite of state-of-the-art coupled climate models and Earth System Models (CMIP5; [86]) [19] [3], with McGregor [3] suggesting that this bias may be due to missing processes in the planetary boundary layer of the atmosphere.

These biases towards an under-representation of physical variability on decadal timescales should then be expected to have implications for their representation of decadal variability in the ocean carbon cycle and more generally biogeochemistry and ecosystems. The Equatorial and Extra-tropical Pacific represents the largest oceanic source of carbon to the atmosphere annually, e.g. [87], and is home to numerous biodiversity hot spots [88], motivating the examination of ESM representation of decadal trends in carbon-cycle dynamics of this climatically and ecologically important region.
Initial Condition Large Ensembles, like those conducted with GFDL-ESM2M \[23\] and CESM1-BGC \[16\] and multi-model ensembles like the CMIP5 provide a test-bed for evaluating natural variability and model uncertainty in the representation of carbon-cycle dynamics. In this paper, we find that no individual realization in the GFDL Large Ensemble, the CESM Large Ensemble or the CMIP5 multi-model ensemble captures the magnitude of data-based trends toward Equatorial Pacific outgassing. We hypothesize this is due to anemic decadal variability of the wind-stress trends over the Equatorial Pacific. We then test this hypothesis by nudging GFDL-ESM2M with winds-trends from the ERA-Interim reanalysis product \[89\] over the Equatorial Pacific.

5.2 Methods

5.2.1 Data-based estimates

Air-sea CO\(_2\) fluxes

Four observational data-based products (OBS) spanning the period 1990-2009 of ocean carbon fluxes from the Surface Ocean pCO\(_2\) Mapping intercomparison (SO-COM) initiative are considered. Spatially- and temporally-sparse carbon observations were assimilated into a spatially and temporally continuous carbon flux products through four different methods, generating the suite of four data-based observational products as documented in \[17\]. The products include, the neural-network based interpolation and mapping (SOM-FFN) \[90\], statistical interpolation (Jena MLS) \[91\], linear regression techniques (JMA) \[92\] and model-based regression involving scaling GFDL-ESM2M to optimally match pCO\(_2\) data (OTTM) \[23\]. The trends of the four products over the observational period can be found in Figure C1.
Wind Stress / Momentum Fluxes

The European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis product involves assimilation of observed meteorological and surface conditions into a forward, forecast model to reproduce the trajectory of the climate from 1979 to present [89]. The product fills gaps in space and time for observed variables, such as sea surface temperature (SST) as well as produces estimates of the spatial and temporal patterns for variables that are not directly observable but represented in the model, such as wind velocities. We use the wind stress from the ERA-Interim reanalysis product to evaluate the statistics of wind stress in the GFDL-ESM2M Large Ensemble as well as to force trends in equatorial Pacific wind stress in new simulations with GFDL-ESM2M.

Sea Surface Temperature & Sea Surface Height

In addition to data-based estimates of air-sea CO$_2$ fluxes, we use satellite estimates of sea surface temperature (SST) and sea surface height (SSH) over the Pacific Basin to evaluate the results of our wind substitution experiments. For SST, we consider the HadiSST2 product [94]. For SSH, we consider the altimeter products of Ssalto/Duacs and distributed by the Archiving, Validation, and Interpretation of Satellite Oceanographic (AVISO), with support from CNES (http://www.aviso.altimetry.fr/duacs/).

5.2.2 Models

Unperturbed or free-running ESM simulations

We use Large Ensembles (LEs) from two Earth System Models, GFDL-ESM2M [21] [22] [23] and CESM1-BGC [16] [25]. For each of the LEs, the simulations are concentration driven and follow historical and RCP8.5 boundary conditions. Each
LE has at least 30 members which were initialized in the early- to mid-twentieth century. Modest perturbation to initial conditions of each ensemble member results in future randomization of the modes and structures of variability.

Decadal trends in Air-Sea carbon fluxes for 10 additional ESMs of CMIP5 are considered. For each of the models, the simulations are concentration driven and follow historical and RCP8.5 boundary conditions. The spread of the multi-model ensemble results from both structural differences between the different model physics and different (asynchronous) paths of internal variability. Single realizations of the following ESMs are considered: BCC-CSM.1, CanESM2, HadGEM2-CC and HadGEM2-ES, IPSL-CM5B-LR, MIROC-ESM-CHEM and MIROC-ESM, For MPI-ESM-LR & MPI-ESM-MR and NorESM1.

Wind-substitution simulations with GFDL-ESM2M

With GFDL-ESM2M it is possible to impose or prescribe wind-stress induced momentum fluxes over specified domains of the ocean model. By design, this imposition or substitution does not directly impact “bulk” quantities tied to wind speed such as the piston velocity of the gas exchange routine or rate of sensible and latent heat exchange between the ocean and atmosphere i.e. all other surface transfers or exchanges use the prognostic (model-calculated) value of the wind while only the momentum exchange is prescribed.

We use the method developed and documented in Delworth for creation and application of the imposed wind-stress fields in the override experiments. This method involves using the high-frequency (daily and seasonal) variations from the model but the low-frequency variations (i.e interannual to decadal) from the ERA-Interim reanalysis product. The control simulations for these override experiments must be done in two steps. The first step is to run unperturbed simulations, with historical-to-RCP8.5 forcing and saving daily wind-stress momentum fluxes from the atmosphere.
to the ocean. This is done for 5 ensemble members, for which the initial conditions of the GFDL-ESM2M Large Ensemble members 1-5 were used to initialize the new control (‘free-control’) runs.

The second step is to run a ‘forced-control’ for each ‘free-control’ ensemble member. For the ‘forced-control’ the daily momentum fluxes from the ‘free-control’ are processed with the same treatment as we process the ERA winds (described in detail below) and then imposed on the ‘forced control’ in the manner that they are for the ERA experiments (‘forced-ERA’). The use of the ‘forced-control’ is necessary as the free-control and the ‘forced-ERA’ runs actually have two differences: first is the wind stress trends, and second the suppression of ocean-atmosphere coupling over the domain. On the other hand the ‘forced-control’ and the ‘forced-ERA’ only differ in the magnitude of trend prescribed.

For both the forced-control and forced-ERA experiments, the imposed momentum flux timeseries consists of (1) high-frequency or daily fluctuations from the control simulation, (2) a repeating seasonal cycle derived from the climatology of the control simulation, and (3) monthly anomalies from ERA-Interim (or, for the forced-control, from the free-control run). The resulting time-series has daily and seasonal fluctuations, as well as the time-mean, consistent with the original free-running control, but the longer-term variability (interannual to decadal) of ERA-Interim. The momentum flux \(MF\) for each of the 5 ensemble members \((e)\) at each point in space \((x, t)\) and time \((t)\) can be represented for the forced-control and for the simulations with ERA wind-stress trends as:

\[
MF_{\text{forced-control}}(e, x, y, t) = MF_{\text{gfdl,HF}}(e, x, y, t) + MF_{\text{gfdl,season}}(e, x, y, t) + ... + MF_{\text{GFDL,anom}}(e, x, y, t)
\]

(5.1)
The seasonal component ($MF_{gfdl,season}$) is simply the mean seasonal climatology over the simulation period. The imposed perturbation is either the monthly anomalies from the control run ($MF_{GFDL,anom}$) or from ERA-Interim ($MF_{ERA,anom}$). The monthly anomalies are computed by removing a climatology of the seasonal cycle – this also produces a time series with a mean of zero over the period – ensuring the forced simulations do not drift due to non-zero integrated momentum fluxes over the period. The high-frequency component ($MF_{gfdl,HF}$) is computed by first filtering the low-frequency variability from the full timeseries and then removing it, yielding the residual as the high-frequency component:

$$MF_{gfdl,HF}(e, x, y, t) = MF_{gfdl}(e, x, y, t) - MF_{gfdl,LF}(e, x, y, t)$$  \hspace{1cm} (5.3)$$

Where the low frequency ($MF_{gfdl,LF}$) component, greater than 30-days, is identified through a standard finite impulse response filter \[103\].

The domain of full-imposition of the wind-substitution extends from 15S to 15N and from the eastern boundary of Pacific Ocean (i.e. the western coast of South America) and 150E. A 10° buffer for which the imposed forcing linearly tapers from the full value at the edge of the box to zero at buffer’s edge, is used around the non-continental boundaries of the box, i.e. from 25S to 15S, from 15N to 25N and from 150E to 140E. Previous modeling studies found that the Subtropical Cell overturning strength is determined by the strength of the trade winds at the boundary of the equatorial and subtropical ocean \[81\] \[104\], thus motivating the use of this 15N-15S wind-override footprint.
5.3 Results

We present the data-based estimates alongside the Large Ensembles, the CMIP5 multimodel ensemble and the results from the wind-substitution experiments. First we discuss the regional features of decadal trends in Air-Sea CO$_2$ fluxes for the free-running ESMs, the data-based estimates, and the wind-substitution experiments. Then we consider SST, SSH and subsurface trends in circulation, temperature, and dissolved inorganic carbon (DIC) concentrations.

5.3.1 Regional trends in Air-Sea CO$_2$ fluxes

The data-based estimates of Equatorial Pacific Air-Sea CO$_2$ flux all demonstrate a trend toward enhanced outgassing over the 20-year period 1990-2009, with a mean value of 10.6 TgC year$^{-1}$ year$^{-1}$ outgassing trend, or $\sim$200 TgC or 1/5 of a PgC cumulative over the 20 year period (Figure 5.1a). The outgassing in the data-based Air-Sea CO$_2$ flux products are consistent with previous findings [105] which used an empirical relationship between shipboard pCO$_2$ and SST to calculate carbon fluxes from satellite SST, resulting in an apparent increase (+27\%) in outgassing of carbon from the Equatorial Pacific coincident with the 1997–1998 PDO regime shift, and long-term trends toward increasing surface ocean $\Delta$ pCO$_2$ as derived from moored time-series data from the TOA/TRITON array which spans the Equatorial Pacific [106].

The trends given by the Air-Sea CO$_2$ products are outside the range of the GFDL LE, the CESM LE and the CMIP5 ESMs. The majority of the distribution of trends in the free-running simulations is greater than zero, indicating a trend toward decreased outgassing or enhanced uptake, which is to be expected given the increasing concentrations prescribed for atmospheric CO$_2$. The distributions of each LE is comprised of contributions from natural variability (produces the spread) and the forced or anthropogenic trend (approximately the mean of the distribution). For the GFDL-
Figure 5.1: Regional Trends for Pacific Air-Sea CO₂ Flux 1990-2009

Trends in regionally integrated air-sea CO₂ flux for a. the Equatorial Pacific and b. the North Pacific over the 20-year period 1990-2009. Positive values indicate a trend toward ocean carbon uptake or an increasing air-sea CO₂ flux, negative values indicate a trend toward ocean carbon outgassing or a decreasing air-sea CO₂ flux. Distributions for the CMIP5, GFDL-LE and CESM-LE are shown. Individual estimates from the data-based estimates are given (orange markers). Results from the wind-substitution experiments, ‘Forced ERA’, are shown with purple markers. The green triangle indicates the forced or anthropogenic trend for the GFDL-LE. In parenthesis, the mean value of the data-based estimates, the ‘Forced ERA’ simulations and the GFDL-LE area given.

LE, the forced trend is toward enhanced uptake (or rather reduced outgassing) of 4.6 TgC year⁻¹ year⁻¹. Given the LEs provide a forced trend equitable with that observed, then we can interpret the distance between the mean of the LE distribution (given by the green triangle) and any individual member’s trend or those of the data-based estimates as being due to natural variability. Therefore, the largest, natural, negative trend rendered by the ESMs is approximately equivalent to the smallest magnitude natural trend of the data-based estimates. A consistent, but inverse relation is true of the North Pacific region, for which the data-based estimates of decadal trends in Air-Sea CO₂ fluxes are also at the upper-end or fully outside the range of those produced by the LEs or CMIP5 (Figure 5.1b). For this region, the data-based products estimate larger trends toward uptake than do the ESMs.

The wind-substitution simulations close the gap between the trends in data-based estimates and the ESMs. For the Equatorial Pacific, the trends in Air-Sea CO₂
fluxes for the wind substitution experiments fall fully outside the distribution of the GFDL-LE trends, and cluster around the central value of the four data-based estimates (Figure 5.1a). The impact of imposing the ERA-intrium wind-stress trends on ESM2M is therefore $15.4 \pm 1.25$ TgC year$^{-1}$ year$^{-1}$ ($\pm$ standard deviation of the 5 'forced-ERA' ensemble members) or $\sim 300$ gC or $1/3$ a PgC over the 20 year period. For the North Pacific, two of the five wind ensembles fall outside the range of the GFDL-LE, and are in better agreement with the data-based estimates (Figure 5.1a). We note for ESM2M, the additional uptake of carbon over the North Pacific ($3.8 \pm 2.7$ TgC year$^{-1}$ year$^{-1}$) due to increasing trade winds does not compensate for the outgassing over the Equatorial Pacific.

Figure 5.2: Trends for Air-Sea CO$_2$ Flux 1990-2009

Turning to the maps of Air-Sea CO$_2$ flux (Figure 5.2), the GFDL-ERA wind-substitution experiments produce similar patterns and magnitude of change as the data-based estimates. Both are characterized by a band of strong outgassing trends over the Equatorial upwelling region and trends toward uptake over the subtropics, and pounced uptake trends in the Kuroshio extension region.
5.3.2 Trends in SST and SSH

The wind-substitution experiments produce equatorial cooling and extra-tropical warming (Figure 5.3), consistent with HadiSST and previous wind-substitution experiments [19][20]. The wind-substitution experiments also produce stronger trends in the gradient of SSH across with Equatorial Pacific, consistent with those of altimetry estimates (Figure 5.4), and of previous wind-substitution experiments [19]. For both the wind-substitution experiments and the data-based estimates of SSH, the Western pacific rises by 6 cm year\(^{-1}\) or over a meter over the course of the 20 years, and the Eastern Pacific suppresses by 4 cm year\(^{-1}\) or almost a meter by the end of the 20 years.

Altimetry measurements are considered of the most precise and accurate remotely sensed quantities available [107]. As such they provide an important, independent constraint on (1) the interpretation of air-sea CO\(_2\) fluxes and (2) the validation that our over-ride experiments are reproducing SST and air-sea CO\(_2\) flux trends for the right (dynamical) reasons.

Figure 5.3: Trends for SST 1990-2009

![Figure 5.3: Trends for SST 1990-2009](image)

Trends in SST for a. the mean of the 4 data-based estimates, b. the wind-substitution experiments 'Forced-ERA' and c. the forced-control experiment over the 20-year period 1990-2009. Units of C/decade. The pronounced cooling in the Equatorial Pacific region is evident in the data-based estimates and the wind-substitution experiments, but not the control simulations.
Figure 5.4: Trends for SSH 1993-2012

Trends in SSH taken over the domain 10N-10S for the AVISO satellite altimetry estimates, the wind-substitution experiments ‘Forced-ERA’ and the control experiments and the GFDL-LE. The trends are taken over the 20-year period 1993-2012 for the simulations and AVISO product. Trends for 1994-2013 and 1995-2014 are also shown for the AVISO product. Units of cm/year. The pronounced increase in the East-West SSH gradient across the Equatorial Pacific Basin is evident in the altimetry product and the wind-substitution experiments, but not the control simulations.

5.3.3 Trends in subsurface circulation, temperature and carbon

Coupled to trends in SSH are trends in sub-surface circulation. As expected due to the wind-induced accumulation of water in the western Pacific, there is a sub-surface trend toward increased eastward transport of waters in the Equatorial under current (EUC), in both the wind-substitution experiments (Figure 5.5a) and in reanalysis products [85] [108]. The strengthening EUC brings with it cooler temperatures and higher concentrations of DIC to the surface and subsurface of the Eastern and Central Pacific (Figure 5.5).

5.4 Discussion

The reason for the disagreement between the magnitude of observed and ESM trends could have one of the following explanations: (1) if we assume the distribution of natural variability produced by the LEs is representative of the true (real climate) value, then the realized magnitude of climate variability over the recent decades represents an extreme event (i.e. outside the 95% confidence bounds) or (2) since it is unlikely (by definition < 5% chance) that the given period is an extrema, the greater possibility is that distribution of the natural variability produced by ESMs has insufficient magnitude or (3) the reanalysis and/or other data-based trends are erroneously large.

At present, the observational record is insufficiently long to characterize the distribution of natural variability inherent to the climate system, and therefore it is difficult to attribute the disagreement between data-based and ESM trends as belonging to either or neither the first or second explanation.

Regarding the second possibility, even with modes of variability for which the amplitude is underestimated, ESMs provide a means to evaluate the mechanistic relationship between the natural carbon cycle variability and physical-state variability.
Documenting the repercussions of anemic modes of physical variability for the carbon cycle can serve to motivate the future improvement of models. For the time being, applying methods which correct for the impact on the ocean of anemic variability in the atmosphere provides a means through which to mechanistically interpret and begin to attribute recent trends in the observational record of the ocean carbon cycle.

Regarding the third possibility, alternative estimates of wind trends based on the TOA array indicate that the ERA interim (and other data-based products) trends are overstated [109]. If this is the case, then attribution of missed-hiatus in the CMIP5 should be revisited. With regard to Air-Sea carbon fluxes, Chiodi [109] finds recent outgassing is also overstated, however, their experimental design does not consider the indirect impacts of trends in wind stress on ocean circulation and consequently surface ocean pCO₂. As we show, the dynamically-driven mechanism is also more consistent with satellite altimetry and satellite-derived SST, as well as reanalysis trends in the EUC.

We note that in our experiments, the impact of winds on the piston velocity freely evolve in response to the dynamics induced through the prescribed momentum flux. As a consequence, we cannot fully partition contributions of trends in air-sea CO₂ exchange induced through the impact of momentum fluxes on ocean dynamics from those induced through changes in the gas exchange that result from prognostic winds adjusting to “forced-ERA” dynamics. However, the advantage of our experimental design, is that it allows for consideration of the overall change in the Walker circulation and the trade winds within the coupled system that drives the CO₂ outgassing trend, rather than isolating either the dynamics or gas-exchange drivers in an un-realistic in inconsistent configuration.

If the Chiodi-hypothesis [109] is correct, and it is only the trends in the wind-products which are overstated, and yet it takes imposing overly-strong trends on climate models in order to reproduce observed SST, SSH and air-sea CO₂ flux trends,
this would indicate the ESMs are insufficiently sensitive to equatorial wind stress and/or the real mechanism(s) responsible for the recent hiatus behavior have yet to be identified. A longer duration of observations, better quantification of observational uncertainty, and further experimentation with mechanisms of natural and anthropogenic change in ESMs is required to assess the magnitude and mechanisms of recent variability in the Equatorial Pacific.

5.5 Author Contributions

S.S. performed all analysis and writing, with guidance from T.L.F, K.B.R, J.L.S, J.P.D, R. Z. and T. D. The wind-override experiments were performed by S.S. with the guidance of R. S.
Chapter 6

Conclusions

6.1 Observational Constraints on Natural Variability and ToE

Emergence timescales represent the union of the magnitude of anthropogenic signals and the distribution of natural climate variability. The projected model uncertainty in anthropogenic signals (e.g. 3°-5°C warming at year 2100) cannot be cleanly reduced or constrained through use of the relatively short observational record. However, some timescales of internal variability could potentially be observationally constrained at present (e.g. the recently observed strong decadal variability of Air-Sea CO₂ fluxes presented in Chapter 5). Therefore, for ToE estimates, which rely on the signal-to-noise ratio, the numerator will be subject to model-uncertainty (until the true anthropogenic trend can be observationally obtained), however for the denominator, observational constraints can inform the fidelity of model-generated internal variability.

In Chapter 4 we show that the 4 ESMs demonstrate different internal variability, and in Chapter 5 we show that for some regions, the model-generated internal variability is insufficiently large in comparison to recent data-based estimates.
For each region, the anthropogenic signal given by the GFDL-LE for the period starting with year 1990 (purple), the 64% (darker blue) and 95% (lighter blue) confidence interval for the magnitude of natural variability of the GFDL-LE, and the largest magnitude trend in HadiSST for each trend-length (orange). In the upper left corner of each panel is the additional time required until emergence if HadiSST noise is used, rather than the GFDL-LE noise. We note that the large interannual to decadal variability given by the HadiSST comes from the later-half of the 20th century, and therefore is not an artifact of data-sparsity inherent to the first half of the century.
Given the model-uncertainty as well as the apparent model inaccuracy in representation of internal variability, a question that arises is “How do emergence estimates change if we use observed variability rather than model-generated variability?”

The challenge in addressing this question arises from the necessary substitution of the large quantity of statistics generated from the Large Ensembles (each LE having 30+ realizations of the climate) with the single, auto-correlated realization which the observational record affords. The computation of ToE requires noise at decadal to multi-decadal time-scales, however the observational record is not yet long enough to procure robust statistics on multi-decadal variability, with the exception of reconstructed products, like HadISST which extend back in time long enough to capture noise at timescales needed. Even from ~100 years of HadISST product, it is not possible to retrieve a probability distribution function (PDF) of natural variability (as it is possible for the LEs, e.g. Figure 2.3b) however what can be computed for HadISST are maximum trend values for various time-scales. These maximum observed trends can then be compared to and substituted for the 95% Confidence Interval (CI) for internal variability estimates given by the LEs, and used to define 'noise' in the ToE estimation. The results of this procedure, applied to HadISST, and compared to the GFDL-LE, are shown in Figure 6.1. For each region, we detrend the last 110 years (1908-2017) of the HadISST reanalysis product and compute the largest trend for each trend length between 2 years and 110 years. Therefore there are 109 2-year trends, 59 50-year trends but only 1 110-year trend. This maximum magnitude trend for each trend-length is then used in lieu of the LEs noise for computing ToE.

For all regions, the maximum natural variability given by HadISST is equal to or exceeds the 95% CI given by GFDL-LE. As a consequence, the resulting HadISST-ToE (given in upper right corner of each panel) is always equivalent to or greater than that given by the LEs. We note that if the LEs produced a realistic PDF of noise, then we would expect that HadISST would fall within the LEs 95% confidence
interval 95% of the time (given HadiSST is a representative sample in both time and space). However, the maximum trends for HadiSST are outside the 95% confidence interval for nearly half the regions considered. For each of the 4 LEs considered in Chapter 4, this same procedure is followed, and the resulting multi-model mean ToE estimates are shown in Figure 6.2. ToE estimates increase by a decade or more for the high-latitudes and for global emergence, when observational (reanalysis) noise, rather than model noise is used.

Figure 6.2: Time of Emergence using ESM Signals and HadiSST Noise

<table>
<thead>
<tr>
<th></th>
<th>Global</th>
<th>Arctic</th>
<th>North Pacific</th>
<th>WestEq Pacific</th>
<th>EastEq Pacific</th>
<th>South Pacific</th>
<th>North Atlantic</th>
<th>Eq Atlantic</th>
<th>South Atlantic</th>
<th>North Indian</th>
<th>South Indian</th>
<th>Southern Ocean *</th>
</tr>
</thead>
<tbody>
<tr>
<td>ToE_LE (4-model-mean)</td>
<td>12</td>
<td>39</td>
<td>16</td>
<td>14</td>
<td>18</td>
<td>15</td>
<td>25</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>ToE_LE-obs (4-model-mean) uses HadiSST noise</td>
<td>26</td>
<td>66</td>
<td>26</td>
<td>14</td>
<td>34</td>
<td>22</td>
<td>28</td>
<td>15</td>
<td>17</td>
<td>17</td>
<td>14</td>
<td>48</td>
</tr>
<tr>
<td>difference ToE_LE-obs minus ToE_LE</td>
<td>14</td>
<td>27</td>
<td>10</td>
<td>0</td>
<td>16</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>25</td>
</tr>
</tbody>
</table>

Global and regional emergence times for the mean of the 4 LEs using method presented in Chapter 2 (row 1), for the case in which HadiSST noise is used for each LE in place of model-generated noise (row 2) and the difference between the model-only vs. model-obs ToE estimates (row 3). For each LE, ToE is first computed separately before the multi-model mean is taken.

For SST, this result is somewhat inconsequential, as detection of anthropogenic signals in SST has already been achieved globally and for most regions [14]. However, if the same bias holds true for other open ocean properties for which detection of anthropogenic signals has yet to be achieved, such as net primary productivity (NPP) or O₂ inventories, then the ToEs given in Chapters 3 and 4 might also under-estimate realized emergence by decades. On the other hand, the convergence of the HadiSST noise and the LE noise on longer-timescales (Figure 6.1) is indicative that either (1) the LEs produces realistic variability at the time-scales relevant to slowly-emerging ocean properties like NPP, and therefore these ToEs remain unchallenged or (2) the smaller-sample size of observed 50-year trends is insufficient to reveal model bias on this time horizon. Regardless, continued observation of the Earth system is required to
detect these important trends as well as diagnose model bias and ultimately improve ESM capability.

6.2 Concluding remarks on the ‘Ensemble Advantage’

Large Ensembles provide statistical context with which to interpret the recent observational record as well as inform the design of future observational platforms which seek to detect anthropogenic change. The Large Ensembles indicate that anthropogenic changes in important ocean properties are underway, however for some regions, it may take many more decades of observations in order to distinguish these changes from natural climate variability.

The use of multiple Large Ensembles provides opportunity for the clean separation of otherwise confounding sources of uncertainty inherent to climate projections – model uncertainty and internal variability uncertainty. We find the magnitude of natural variability, and its potential contribution to future ocean states varies between the Large Ensembles and in some cases does not agree with observational estimates of natural variability.

The way forward, on narrowing uncertainty in detection and projection, is sustained observations and continued model development.
Appendix A

Supplementary Material for Chapter 3

We provide additional analysis through (a) discussion and presentation of the global evolution of ocean carbon fluxes in the Large Ensemble and sensitivity experiments, (b) expanded discussion and presentation of emergence timescales for additional variables related to the ocean carbon pumps that are not covered fully in main text and (c) discussion and presentation of an alternative method for computing time of emergence using a pre-industrial simulation.

A.1 Anthropogenic Changes in the Ocean Carbon Pumps

The evolution of globally integrated air-sea CO$_2$ fluxes and biological pumps for the LE, two sensitivity experiments, and the pre-industrial control runs are shown in Fig. A1. Pre-industrially, the air-sea CO$_2$ flux is zero PgC/yr, and the soft tissue and calcium carbonate pumps (quantified by export of particulate organic carbon (POC) and CaCO$_3$ at 100 meters depth) are 8.0 and 0.7 PgC/yr respectively. Note that the quasi zero net air-sea CO$_2$ pre-industrial flux implies that the export of carbon in the soft tissue and calcium carbonate pumps across the 100 m depth contour is resupplied by upward transport of dissolved inorganic carbon (DIC). The air-sea CO$_2$ flux increases by 6 PgC/yr under anthropogenic forcing by the end of the 21st century, whereas the biological pump decreases by $\sim$1 PgC/yr. Until year $\sim$2050,
Figure A.1: Changes in ocean carbon pumps for the LE and sensitivity experiments

Panel a. shows the physical pump (air-sea CO$_2$ fluxes) and panel b the biological pumps, where POC is particulate organic carbon. Values from the sensitivity experiments are shown in different colors. All forcings indicates a fully coupled, transient run in which the forcing follows a historical to RCP8.5 pathway. In 1950, ensembles 2-29 branch from ensemble member 1, with the ensemble mean shown in darker purple. The Rad-only run includes only the radiative impacts of anthropogenic forcings, and air-sea gas exchange is not affected by rising atmospheric CO$_2$. The BGC-only run has only the chemical impacts of anthropogenic warming (i.e. increased CO$_2$ gas exchange, but no warming). PreInd indicates the pre-industrial control run. End of blue shading and dashed line at 1990 indicates the beginning of the ocean observation era, and the reference state from which emergence is defined for this paper.

Anthropogenic changes in the global air-sea CO$_2$ flux are due entirely to the impact of rising atmospheric CO$_2$ concentrations on air-sea gas exchange, not due to warming, with small contributions (<1 PgC/yr) from changes in the biological pumps. For this reason, we interpret changes in the air-sea flux as representing changes in the solubility or physical carbon pump [110]. Additional confirmation that changes in biological export are not impacting the emergence characteristics of air-sea CO$_2$ fluxes can be found in Fig. A2, which demonstrates that the impact of changing export of surface pCO$_2$ is much smaller than the total change over the 21st century (0.5 uatm vs. 550 uatm, global averages). After year 2050, warming decreases uptake globally by \( \sim 1 \) PgC/yr or 15%, consistent with CMIP5 estimates of climate-carbon feedback magnitude [8]. Anthropogenic changes in the soft tissue pump are driven by warm-
ing (concurrence of All-forcings and Rad-only experiments), whereas anthropogenic changes in the CaCO$_3$ pump are driven by changes in ocean chemistry (concurrence of All-forcings and BGC-only experiments).

Figure A.2: Change in surface ocean pCO$_2$ due to biological export of carbon

The export of soft tissue and calcium carbonate impacts surface ocean pCO$_2$ through changes in the concentration of alkalinity (ALK) and dissolved inorganic carbon (DIC). The impacts are not explicitly diagnosed in ESM2M, however we do offline calculations with CO2SYS (van Heuven et al., 2011) to estimate the impact of changing export on surface pCO$_2$. This is done by first computing pCO$_2$ using monthly output from the year 2100, then adjusting the concentration of DIC and ALK to account for changes in the pumps relative to 1990, and recalculating surface pCO$_2$ with the export-adjusted DIC and ALK concentrations. The difference between the pCO$_2$ at 2100 and the pCO$_2$ at 2100 if the pumps were stationary since 1990, is shown below. The annual mean is taken after the computations of pCO$_2$ are done monthly. For reference, the total pCO$_2$ changes by $\sim$550uatm over this same time period, due to rising atmospheric pCO$_2$ and a warming surface ocean.

A.2 Time of Emergence for O$_2$, Nutrients and other ocean properties

Changes in the ocean carbon pumps do not occur in isolation, but rather they are coupled to many other changing ocean properties and processes. Some, but not all, of these related ocean fields are discussed fully in the main text. Notably, sea surface temperatures (SST), net primary productivity (NPP), controls on the soft-tissue pump, and oxygen inventories have limited discussion, and so we include a more detailed discussion here.
SST couples to the ocean carbon pumps through driving modulations of the CO$_2$ solubility [5] and biological productivity [111]. Satellite-derived SST is a primary input for estimating air-sea CO$_2$ fluxes [47] and biological export [46]. Therefore, the timescales of emergence for anthropogenic changes in SST are of interest to projection, detection and attribution of changes in the ocean carbon pumps. We find SST emerges in 20-30 years for localities in the tropics and extra-tropics (Fig. A3e), where the signal is strongest and positive (Fig. A4e). SST emergence timescales for GFDL-ESM2M LE were first shown in Rodgers [23] and is consistent with multi-model emergence times and patterns of previous multi-model studies [68] [112] [73], with the exception of the high latitudes. Weak warming, and even cooling in the high latitudes results in delayed emergence, and even non-emergence for localities in the North Atlantic and Southern Ocean. We note this is model-dependent, as cooling trends in the Southern Ocean in GFDL-ESM2M are stronger than the majority of CMIP5 models [13].

Changes in the soft-tissue pump occur through a complex sequence of mechanisms. Changing ocean circulation (e.g. stratification, reduced ventilation and mixed layer depths, Fig. A4u) reduces nutrient availability in the upper ocean (Fig. A4i & A4q), shifting phytoplankton species/size distribution towards smaller phytoplankton (i.e. decreasing phytoplankton biomass, Fig. A4n), which are less efficient than large phytoplankton at exporting carbon from the surface ocean (Fig. A4o) [50].

In GFDL-ESM2M NPP has limited attributive value for anthropogenic trends in biological export. However, we discuss it because it is an important ecosystem quantity, and because other ESMs indicate stronger dependence of anthropogenic changes in export on anthropogenic changes of NPP [12]. Emergent trends in NPP (Fig. A3t) follow the sign of trends in nitrate concentration (Fig. A3q). Exceptions occur over (a) the Equatorial Pacific cold tongue, where nitrate concentrations are adequate ($> 0.5$ mol kg$^{-1}$) and warming enhances phytoplankton growth rates (b);
the Pacific sector of the Southern Ocean, where NPP decreases due to decreasing iron (Fig. A4j); and (c) in the North West Pacific where NPP increases due to alleviation of light limitation associated with decreased mixed layer depths (MLD, Fig. A4u). For the North West Pacific, the biological impacts (a reduction in NPP) of a changing physical climate (shallowing MLD) emerge, in a statistical sense, before the physical driver (Fig. A3t vs. Fig. A3u). We note that increases in nitrate (and NPP) in the band 45°S-30°S are due to poleward expansion of the subtropical gyre resulting in an injection of higher-nitrate at the gyre edge.

Finally, we discuss the emergence of O\textsubscript{2} inventories over 200-600m as it is of critical importance to marine life and impacted by the biological export and subsequent remineralization of organic carbon. The trend in O\textsubscript{2} inventories is determined by the relative contributions of changing solubility (reflected by O\textsubscript{2,SAT}), and circulation and biology (primary productivity, biological export and respiration, reflected in Apparent Oxygen Utilization (AOU)). Decreasing O\textsubscript{2,SAT} emerges for most of the globe at mid-century (Fig. A3k & A4k), with the exceptions of the subtropical convergence zones/fronts, which have counter-trends, and remain non-emergent through the century. Consistent with CMIP5 projections presented in Cabre [113], AOU has positive trends in the high latitudes and negative trends in the mid-to-low latitudes, with these trends emerging mid-to-late century (Fig. A3r & A4r). The reason for these changes is slowdown in the ocean overturning circulation results in increased water ages at intermediate depths in the high latitudes (positive AOU) and decreased water age in intermediate depths in the mid-to-low latitudes (negative AOU), as older, deeper, oxygen-depleted waters have decreased interaction with waters at intermediate depths [114][115]. The full O\textsubscript{2} signal emerges most rapidly (~2030-2060) in the high latitudes where O\textsubscript{2,SAT} and AOU have consistent sign of trends (i.e. where AOU is increasing) (Fig. A3s & A4s). Regions with opposing O\textsubscript{2,SAT} and AOU trends emerge slowly (West Equatorial Pacific and Indian), if at all (non-emergence
over the subtropical gyres of the South Pacific and Atlantic). The North Atlantic exhibits limited emergence due to the dipole spatial pattern of trends and strong natural decadal variability. Decreases in POC export (Fig. A3o & A4o) contributes to declining AOU, which combats the effects of declining O$_2$ solubility on O$_2$ inventories, thus extending emergence times for these regions. Contributions from POC export to anthropogenic changes in AOU are not (and cannot be) quantified here, but are generally considered to be important but nevertheless second order to the impact of changes in ocean physics in modeling studies [114][115][35]. However, it is observed to be of first order importance for some regions on seasonal to interannual timescales [116].

A.3 Alternative methods for ToE computations

As a complement to the large ensemble calculations, we also consider output from a 1600-year long pre-industrial run with the same ESM2M model, with this providing alternative means with which to estimate the noise associated with natural variability in determining ToE. For the purposes of ToE estimates, the LE advantage is twofold: 1) providing a purely forced signal at the local or pixel scale, without requiring a smoothing technique and 2) providing robust noise estimates for a perturbed (rather than pre-industrial) climate state. To illustrate both of these advantages, we recompute ToE using sub-sets (between 1 and 30 members) of the LE to estimate the forced signal and using 30, randomly selected 111-year segments from a 1600-year pre-industrial (PI) control run to estimate noise. In practice, if a 30-member ensemble were available, then the PI run would not be needed to estimate the noise, however to isolate the importance of noise, we compare the ToE with the same signal (from the 30 member LE) but different noise (30 member LE noise vs. PI noise). Fig. A7 compares the standard $ToE_{LE}$ (both signal and noise from 30 LE members) to the re-computed $ToE_{PI}$ (signal from 1-30 ensemble members, but noise from the
PI control run). Error in air-sea CO₂ flux ToE estimates using PI noise and 3, 10,
and 30 ensemble members to generate the signal are shown in Fig. A7a, b and c,
respectively. Using PI noise and only 3 ensemble members (Fig. A7a) can result
in errors of 50+ years relative to estimates generated from the full LE. This error
reduces as the number of ensembles used to estimate the signal increases (Fig. A7b,
c), however even when all 30 ensembles members are used to estimate the signal (Fig.
A7c), error remains, particularly at frontal boundaries, where the underlying noise
structures have spatially shifted between the pre-industrial and contemporary climate
state (e.g. poleward expansion of the subtropical gyres, Fig. A3m).

Fig. A7d shows the fractional ocean area with error in the $ToE_{PI}$ estimates
greater than 20% of the $ToE_{LE}$ estimates (i.e. $|ToE_{LE} - ToE_{PI}| / ToE_{LE} > 0.2$). Error
estimates are generated through subsampling the combination space of 30 choose
n, e.g. there are exactly 30 combinations of 30 choose 1, however there are 435
combinations of 30 choose 2 (from which 30 are sampled for these calculations). The
'maximum error' is the maximum error that occurs, between $ToE_{LE}$ and $ToE_{PI}$, at
a particular grid cell for a given sub-sample of LE members. For example, when n
= 3, error is maximized for a particular pixel in the Southern Ocean when ensemble
members 2, 14 and 27 are used to estimate the anthropogenic signal. The 'expected
error' is mean of 30 different combinations of 3 ensemble members. It is the error
one would expect if only a 3 member ensemble were used, however the error could
also be as large (or larger) than the 'maximum error', depending on the random
combination of ensemble members generated or used – and so we show/calculate
both. The 'expected' and 'maximum' error converge at $X = 30$ by construct, as there
is only 1 combination of 30 choose 30.

For all fields, the extent of ocean area with $ToE_{PI}$ error declines when an increasing
number of ensembles are used to estimate the signal, however persistent errors
remain (when n = 30) over 15-30% of the global ocean due to differences between
pre-industrial and contemporary noise, highlighting the necessity of LEs to minimize potential error in local ToE estimates.

There are three main messages from Fig. A7d: (1) When using the pre-industrial control run to generate noise statistics, persistent error remains, even when the full 30 ensemble members are used to generate the anthropogenic signal, and (2) the mean or expected error between the full ensemble and a subset of the ensemble approaches zero (leveling off of the dashed curves) occurs at about 5 ensembles (the persistent 15-30% error is due to the use of PI noise, rather than a smaller ensemble size) and (3), Maximum error (solid lines) between the full and partial ensemble does not decay exponentially (as does expected error), but rather linearly, highlighting the necessity of LEs to minimize potential error in local ToE estimates.

We note that the exact method employed here is not the only way to calculate emergence times. For robust estimates, the 'signal' used must be free of natural variability. To achieve this, without a LE, one must take the trend over a long period of time (e.g. 21st century). This assumes a stationary trend throughout the 21st century, which is appropriate for some fields, however is not appropriate for others, like air-sea CO$_2$ fluxes, which are convex at the end of the century as the oceans buffering capacity is depleted (e.g. Fig. A1a). Assuming stationary noise, and stationary noise structures (e.g. frontal boundaries) proves erroneous (the persistent error over 15-30% of ocean area in Fig. A7d), at least when the difference in ocean state is contemporary vs. pre-industrial. As shown in Frölicher [68] their Fig. 9, differences in noise between PI and RCP8.5 is model and field dependent, so the benefits of the LE method of ToE presented in this paper are also model dependent. One could also argue that a historical simulation with natural forcing (i.e. solar, volcanoes) should be used to estimate noise, instead of the pre-industrial control runs. Future work should explore the model and method dependence of noise estimates on time of emergence estimates.
CaCO₃ and particulate organic carbon (POC) export are evaluated as a flux at 100m. NPP is integrated over 0-100 meters. Oxygen variables are integrated over 200-600m. Heat is integrated from 0-700m. Chlorophyll is integrated over upper 500m. All other variables taken at surface. Asterisks indicate the 3 carbon fluxes (pumps) which are repeated from Fig. 3.3 of the main text. Chlorophyll maps are repeated from Fig. 3.5 of the main text.
Figure A.4: Signal at Time of Emergence or year 2100, whichever comes first. This figure is meant to illustrate the direction and relative magnitude of trends. Blues are negative trends, Reds are positive trend, with increasing color intensity scaling with magnitude of trend. White hatch over non-emergent regions. \([\text{NO}_3] = 0.5 \ \mu\text{mol kg}^{-1}\) contours (1990 ensemble annual mean values) imposed on some fields for explanatory purposes. \(\Delta p\text{CO}_2 = p\text{CO}_2\text{ocn} - p\text{CO}_2\text{atm}\). Asterisks indicate the 3 carbon fluxes (pumps) which are repeated from Fig. 3.3 of the main text. The value of darkest red and corresponding units are as follows: \(p\text{CO}_2 \ [3 \ \text{uatm yr}^{-1}]\), \(\Omega \ [0.01 \ \text{yr}^{-1}]\), nALK \(\ [0.3 \ \text{meq m}^{-3} \ \text{yr}^{-1}]\), CaCO\(_3\) export at 100m \([0.02 \ \text{gC m}^{-2} \ \text{yr}^{-1}]\), SST \([0.05 \ \degree\text{C} \ \text{yr}^{-1}]\), \(dp\text{CO}_2 \ [0.5 \ \text{uatm yr}^{-1}]\), fgCO\(_2\) \([0.3 \ \text{gC m}^{-2} \ \text{yr}^{-1}]\), heat inventory \([8.37 \times 10^7 \ J \ \text{m}^{-2} \ \text{yr}^{-1}]\), chlorophyll inventory \([0.02 \ \text{mmol m}^{-2} \ \text{yr}^{-1}]\), Iron \([1.5 \ \text{nmol m}^{-3} \ \text{yr}^{-1}]\), O\(_2\)sat inventory 200-600m \([0.22 \ \text{mmol m}^{-3} \ \text{yr}^{-1}]\), SSS \([0.008 \ \text{psu yr}^{-1}]\), SSH \([0.3 \ \text{cm yr}^{-1}]\), percent large phytoplankton \([1.4 \ \mu\text{mol m}^{-3} \ \text{yr}^{-1}]\), POC export at 100m \([0.22 \ \text{gC m}^{-2} \ \text{yr}^{-1}]\), surface chlorophyll \([0.0023 \ \text{mmol m}^{-3} \ \text{yr}^{-1}]\), NO\(_3\) \([2.0 \ \mu\text{mol m}^{-3} \ \text{yr}^{-1}]\), AOU inventory 200-600m \([0.22 \ \text{mmol m}^{-3} \ \text{yr}^{-1}]\), NPP \([0.075 \ \text{gC m}^{-2} \ \text{yr}^{-1}]\), and MLD \([0.3 \ \text{cm yr}^{-1}]\).
Figure A.5: Decadal variability and mean state for CaCO₃ export.

(a) Decadal variability in CaCO₃ export, defined by the standard deviation of the 30 ensemble trends over the 20-year period 1990 to 2009, and (b) ensemble mean CaCO₃ export at year 1990. Units are in gC m² per 20 yr⁻¹ and corresponding ranges for the two panels are given below the color bar. Note that the color bar range of the (a) decadal variability is one order of magnitude less than (b) the mean state.

Figure A.6: Depth differences in trends for iron limitation

Depth differences in trends for iron limitation. For (a) small and (b) large phytoplankton, the difference between surface trends and 50 m trends (surface minus 50 meters) in iron limitation during austral summer (January-March). The trend is taken from the ensemble mean over the period 1990 to 2050. Units given in mol kg⁻¹ yr⁻¹. For both large and small phytoplankton, positive values in the Southern Ocean indicate iron limitation trends increases with depth.
Error for ToE calculation when using Pre-Industrial noise instead of contemporary noise derived from the LE. (a), (b), and (c) show the maximum difference between emergence times calculated for air-sea CO$_2$ fluxes between using the full ensemble for both signal and noise estimates ($T_{OE_{LE}}$) and using pre-industrial noise ($T_{OE_{PI}}$) and 3, 10 and all 30 ensemble members to estimate the anthropogenic signal. (d) shows the fractional ocean area with a difference between $T_{OE_{LE}}$ and $T_{OE_{PI}}$ greater than 20%. Increasing the number of ensemble members reduces error, however persistent error remains (~15-30% of the globe), due to the inaccuracy of pre-industrial noise. Dashed lines are the maximum error in ToE estimates, solid lines are mean (expected) error, for the given number of ensemble used to estimate the anthropogenic signal.
Appendix B

Supplementary material for Chapter 4
Global annual changes relative to year 1990 for a. global mean sea surface temperature (SST), b. globally integrated Air-Sea CO$_2$ Flux, c. POC Export, d. surface chlorophyll and e. sea surface salinity (SSS).
Regional time series of changes relative to year 1990 for RCP8.5 scenario. Legend same as Figure 5.1 & B1 (GFDL green, CESM blue, CanESM purple, MPI pink). Variable indicated at top of each column.
Regional time series of changes relative to year 1990 for RCP4.5 scenario. Legend same as Figure 5.1 & B1 (GFDL green, CESM blue, CanESM purple, MPI pink). Variable indicated at top of each column.
Maps of SST Emergence (a-d) and Trends at time ToE (e-h) and the magnitude of 20-year variability (i-l). Units of a-d are years. Units of e-l are °C year$^{-1}$. The trend is given at the Time of Emergence. White hatching on pixels with non-emergent trends are year 2100.

Maps of Air-Sea CO$_2$ Flux Emergence (a-d) and Trends at time ToE (e-h) and the magnitude of 20-year variability (i-l). Units of a-d are years. Units of e-l are gC m$^{-2}$ year$^{-1}$ year$^{-1}$. The trend is given at the Time of Emergence. White hatching on pixels with non-emergent trends are year 2100.
Figure B.6: POC Export Emergence and Signals

Maps of POC Export Emergence (a-d) and Trends at time ToE (e-h) and the magnitude of 20-year variability (i-l). Units of a-d are years. Units of e-l are gC m$^{-2}$ year$^{-1}$ year$^{-1}$. The trend is given at the Time of Emergence. White hatching on pixels with non-emergent trends are year 2100.

Figure B.7: Cholorophyll Emergence and Signals

Maps of Cholorophyll Emergence (a-d) and Trends at time ToE (e-h) and the magnitude of 20-year variability (i-l). Units of a-d are years. Units of e-l are mg m$^{-3}$ year$^{-1}$. The trend is given at the Time of Emergence. White hatching on pixels with non-emergent trends are year 2100.
Figure B.8: SSS Emergence and Signals

Maps of SSS Emergence (a-d) and Trends at time ToE (e-h) and the magnitude of 20-year variability (i-l). Units of a-d are years. Units of e-l are practical salinity units (psu) year$^{-1}$. The trend is given at the Time of Emergence. White hatching on pixels with non-emergent trends are year 2100.

Figure B.9: Difference between RCP8.5 vs. RCP4.5 ToE

Maps of difference in years between RCP4.5 and RCP8.5 Multi-model mean Time of Emergence. White hatching over locations of where mean difference (RCP4.5-RCP8.5) is less than 10% of the mean RCP8.5 TOE. For averaging purposes, year 2100 was used when emergence was not achieved for a given LE. For presentation of RCP4.5 local scale ToEs, CanESM excluded for all variables and CESM1-BGC excluded from air-sea CO$_2$ flux, POC flux and chlorophyll due to insufficient ensemble size.
Partitioning regional uncertainty for a. SST, b. Air-Sea CO$_2$ Flux, c. POC Export, d. Chlorophyll and SSS, for scenario uncertainty (red, RCP4.5 vs RCP8.5), model uncertainty (green shading) and internal variability (yellow shading). The contribution of natural variability from each ensemble is given by the colored lines, same model-color relation as previous figures. The interface between yellow and green is determined by maximum contribution from internal variability, i.e. the model with the largest internal variability at that point in time. Ten year box filter was applied to the estimates of natural variability.

Figure B.13: Maps of 30-year local Signal-to-Noise Ratio for the 4 Large Ensembles

Maps of Signal-to-Noise Ratio for the 4 Large Ensembles, 1990-2019. For each variable, SNR is given for the 30-year period 1990-2019. The percent of ocean area with SNR > 2 shown on upper right corner of each map.

Figure B.14: Maps of 30-year Regional Signal-to-Noise Ratio for the 4 Large Ensembles

Maps of regional Signal-to-Noise Ratio for the 4 Large Ensembles. For each variable, SNR is given for the 30-year period 1990-2019. Each LEs Global SNR ratio given on map for each variable.
Maps of local and regional Signal-to-Noise Ratio for the 4 Large Ensembles for Air-Sea CO$_2$ fluxes over the 20-year period 1990-2009. White hatching over locations of where LEs disagree [mean of SNR of 4 LE is less than the standard deviation of SNRs across models]. The LE mean Global SNR ratio ± the standard deviation across the LEs given below the maps for each variable. The percent of ocean area with SNR > 2 shown on upper right corner of each map.
Appendix C

Supplementary material for Chapter 5

Figure C.1: Trends in air-sea CO₂ flux for the 4 data-based methods

Trends in air-sea CO₂ flux for a. SOM-FFN, b. JMA, c. OTTM and d. Jena-MLS over the 20-year period 1990-2009. Units of gC m² year⁻¹, with positive indicating a trends toward increased uptake. The pronounced outgassing in the Equatorial Pacific region is evident in each of the data-based estimates.
Figure C.2: Wind-induced trends in wind-stress, SST and air-sea CO$_2$ flux

The difference between the Forced-ERA simulations and the Forced Controls for a. wind-stress ($\tau_x$), b. SST and c. air-sea CO$_2$ flux over the 20-year period 1990-2009, with positive indicating a trends toward increased uptake. White hatching over regions where the difference between the forced-control and the wind-substitution experiment is not significant. By design, $\tau_x$ has a significantly negative trend (increasing westward trade-winds) over the Equatorial Pacific. White hatching at bounds of over-ride domain to be expected since the winds there are prescribed to the control values. Cooling and outgassing over the Equatorial Pacific also result from the imposition of stronger trends in the trade winds.
Figure C.3: Regional Trends for Pacific Air-Sea CO₂ Flux 1990-2009

Trends in regionally integrated air-sea CO₂ flux for the given regions. Distributions for the CMIP5, GFDL-LE and CESM-LE are shown. Individual estimates from the data-based estimates are given (orange markers). Results from the wind-substitution experiments, ‘Forced ERA’, are shown with purple markers. Panels e. and b. are repeated from Figure 5.1 of main text.
Bibliography


[8] Vivek K Arora, George J Boer, Pierre Friedlingstein, Michael Eby, Chris D Jones, James R Christian, Gordon Bonan, Laurent Bopp, Victor Brovkin, Patricia Cadule, Tomohiro Hajima, Tatiana Ilyina, Keith Lindsay, Jerry F Tjiputra,


[27] J Keith Moore, Keith Lindsay, Matthew C Long, Scott C Doney, and Kazuhiro Misumi. Marine Ecosystem Dynamics and Biogeochemical Cycling in the Community Earth System Model [CESM1(BGC)]: Comparison of the 1990s with the 2090s under the RCP4.5 and RCP8.5 Scenarios. Journal of Climate, 26(23):9291–9312, November 2013.


W M Berelson, W M Balch, R Najjar Global, and 2007. Relating estimates of CaCO3 production, export, and dissolution in the water column to measure-
ments of CaCO3 rain into sediment traps and dissolution on the .... Wiley Online Library.


[87] Taro Takahashi, Stewart C Sutherland, Rik Wanninkhof, Colm Sweeney, Richard A Feely, David W Chipman, Burke Hales, Gernot Friederich, Francisco Chavez, Christopher Sabine, Andrew Watson, Dorothee C E Bakker, Ute Schuster, Nicolas Metzl, Hisayuki Yoshikawa-Inoue, Masao Ishii, Takashi Midorikawa, Yukihiro Nojiri, Arne Körtzinger, Tobias Steinhoff, Mario Hoppema, Jon Olafsson, Thorarin S Arnason, Bronte Tilbrook, Truls Johannessen, Are Olsen, Richard Bellerby, C S Wong, Bruno Delille, N R Bates, and Hein J W de Baar. Climatological mean and decadal change in surface ocean pCO2, and


[95] Tongwen Wu, Weiping Li, Jinjun Ji, Xiaoge Xin, Laurent Li, Zaizhi Wang, Yanwu Zhang, Jianglong Li, Fang Zhang, Min Wei, Xueli Shi, Fanghua Wu, Li Zhang, Min Chu, Weihua Jie, Yiming Liu, Fang Wang, Xiangwen Liu, Qiaoping Li, Min Dong, Xiaoyun Liang, Yang Gao, and Jie Zhang. Global carbon budgets simulated by the Beijing Climate Center Climate System Model for the last century. Journal of Geophysical Research: Atmospheres, 118(10):4326–4347, May 2013.


J F Tjiputra, C Roelandt, M Bentsen, D M Lawrence, T Lorentzen, J Schwinger, Ø Seland, and C Heinze. Evaluation of the carbon cycle components in the


