Marine Science Faculty Publications   College of Marine Science

2019

Time of Detection as a Metric for Prioritizing Between Climate Observation Quality, Frequency, and Duration

B. R. Carter
Joint Institute for the Study of the Atmosphere and Ocean

N. L. Williams
Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, Seattle, nancywilliams@usf.edu

W. Evans
Hakai Institute

A. J. Fassbender
Monterey Bay Aquarium Research Institute

L. Barbero
Cooperative Institute for Marine and Atmospheric Studies

See next page for additional authors

Follow this and additional works at: https://scholarcommons.usf.edu/msc_facpub

Part of the Life Sciences Commons

Authors

B. R. Carter, N. L. Williams, W. Evans, A. J. Fassbender, L. Barbero, C. Hauri, R. A. Feely, and A. J. Sutton

Time of Detection as a Metric for Prioritizing Between Climate Observation Quality, Frequency, and Duration

B. R. Carter[1,2] (ID), N. L. Williams[2] (ID), W. Evans[3] (ID), A. J. Fassbender[4] (ID), L. Barbero[5,6] (ID), C. Hauri[7], R. A. Feely[2], and A. J. Sutton[2] (ID)

[1]Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, WA, USA, [2]Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, Seattle, WA, USA, [3]Hakai Institute, Heriot Bay, British Columbia, Canada, [4]Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA, [5]Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL, USA, [6]Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, Miami, FL, USA, [7]International Arctic Research Center, University of Alaska Fairbanks, Fairbanks, AK, USA

Abstract We advance a simple framework based on "time of detection" for estimating the observational needs of studies assessing climate changes amidst natural variability and apply it to several examples related to ocean acidification. This approach aims to connect the Global Ocean Acidification Observing Network "weather" and "climate" data quality thresholds with a single dynamic threshold appropriate for a range of potential ocean signals and environments. A key implication of the framework is that measurement frequency can be as important as measurement accuracy, particularly in highly variable environments. Pragmatic cost-benefit analyses based on this framework can be performed to quantitatively determine which observing strategy will accomplish a given detection goal soonest and resolve a signal with the greatest confidence and to assess how the trade-offs between measurement frequency and accuracy vary regionally.

1. Introduction

A growing body of scientific literature focuses on assessing the "time of emergence" of various climate signals or the time at which a climate signal grows to exceed a multiple of the noise in an observed or modeled climate record. The idea has been used both to characterize how rapidly modeled climate signals become meaningful relative to natural variability (e.g., Christian, 2014; Hawkins & Sutton, 2012; Henson et al., 2017; Keller et al., 2014; Rodgers et al., 2015) and to determine the length of a record required for detection of a climate signal despite natural variability and/or observational uncertainty (e.g., Carter et al., 2016; Ilyina et al., 2009; McKinley et al., 2016; Weatherhead et al., 1998). The latter application is referred to here as "time of detection," and is our focus. There is also currently scientific discussion regarding what level of measurement uncertainty is allowable for climate research and how requirements might differ when researching shorter-timescale natural phenomena (e.g., ocean weather; Newton et al., 2015). We contend that these topics are connected by a simple, practical question: "What is needed to confidently assess whether the Earth system is changing over a given length of time in a given region?" The answer to this question is, in some cases, certainly "better measurements," but other possible answers include "a longer measurement time series," "more frequent measurements of the system," or even "knowledge of how much natural variation we should expect." In this study, we present a simple approach that combines the time-of-detection concept with measurement uncertainties into a framework that can be used to test the effectiveness of monitoring efforts within the context of the potential answers to this question. The approach accounts for differing timescales of variability and allows for strategies that might be used to reduce the impacts of natural variability in the measurement record.

We first present the framework and then show how it can be applied using a handful of examples that are related to ocean acidification (OA). While we focus on OA applications, we note that this approach could be applied to any climate signal (e.g., warming, eutrophication, and deoxygenation). With a large number of emerging climate signals to observe—and a large and growing number of sensor and measurement

platform technologies being used to observe them—our vision is that this approach will provide means to optimize the effectiveness of future observation strategies.

The framework is built around the minimum requirements for signal detection, but we caution that "good enough for detection" is only good enough provided some common but seldom tested simplifications hold (e.g., normally distributed uncertainties and variabilities). It is therefore advisable that one aims higher than the minimum benchmarks provided by this framework, and we note that the more one exceeds the minimum requirements by, the sooner a signal can be detected and the better the magnitude of the signal will be constrained. Nevertheless, the minimum signal detection criterion is a convenient quantitative measure with which different observation strategies can be compared.

2. The Framework

The framework is a restatement of the time of detection "detection at 95% confidence" criterion, with 1.96 standard deviations rounded to 2 to not overstate the precision of this approach. For detection of a difference, the signal of interest (S, expressed as a difference from an initially or distantly measured value) must then equal or exceed twice the noise (N, expressed as a standard deviation (e.g., Keller et al., 2014)

$$|S| \geq 2N \tag{1}$$

The units in these and all subsequent equations are flexible provided they are consistent for all terms in the equations, and the units reflect approximately linear responses (e.g., avoid considering very large changes in log-space quantities such as pH). The nuance of the method is then in the estimation and interpretation of the terms S and N and their components.

The S and N terms are estimated from a consideration of the signals of interest, the variability of the associated systems, and the uncertainties inherent in the measurements and simplifying assumptions. For marine systems, the common component terms for S and N are measurement uncertainty (U_M), daily variability (V_D), subseasonal variability (V_{SS}), seasonal variability (V_S), interannual and longer timescale natural variability (V_I), long-term change (Δ) or trend multiplied by time ($T \times dt$), and uncertainty in any assumptions (U_A) made during data reduction. Depending on the region and question of interest, one must decide whether each of these components is part of the signal or the noise. Once the components are partitioned between S and N, components of the noise can be combined as the square root of the sum of their squares. A common example in climate science is detecting the presence of a long-term trend in a raw signal such as seawater alkalinity. In this case the signal would equal Δ while noise would be the combination of all other terms (e.g., Carter et al., 2016; Ilyina et al., 2009) added in quadrature

$$N = \sqrt{U_M{}^2 + U_A{}^2 + V_D{}^2 + V_{SS}{}^2 + V_S{}^2 + V_I{}^2} \tag{2}$$

Like measurement uncertainty, natural variability is treated here as a component of the noise because it acts to prevent detection of long-term climate signals (Frölicher et al., 2009). However, it is sometimes practical to make either diverse or frequent measurements frequently enough to resolve the impacts of modes of natural variability. For example, monthly measurements from each year of a monthly time series might be averaged together and subtracted from each monthly value prior to analyzing interannual variability, thereby reducing the impact of V_S and allowing it to be neglected. Alternately, a researcher interested in V_S (e.g., Fassbender, Alin, et al., 2018) could "detrend" monthly measurements by subtracting a linear fit to a multiyear time series, allowing the impacts of T and the longer-term portion of V_I to be neglected (these two terms can be challenging to distinguish from one another in short time series; see work by Weatherhead et al. (1998) for more on this topic). However, in both cases errors in the assumption that these approaches completely removed the impacts of the variabilities/trends must be included (as U_A) in the noise estimate. For example, U_A might equal the uncertainty on the annual mean value in the first example above and the uncertainty on the subtracted trend line in the second. The possibility of reducing the variability by averaging portions of a highly temporally resolved time series gives rise to an important insight for observation network design: An observation network providing a large number of moderately accurate measurements may be preferable to one that provides infrequent yet highly accurate measurements, particularly when natural variability is large compared to the measurement uncertainty. The proposed framework

allows for quantitative comparison of the merits of the full spectrum of such possibilities using, for example, the S to N ratio.

Two additional complications that could be (but are not here) included in the framework are evolving variability terms (e.g., a growing V_S; Fassbender, Rodgers, et al., 2018) and spatial variability. Spatial variability is an additional mode of variability that should be considered for observing platforms that can cross natural seawater property gradients (e.g., floats, drifters, and ship measurements). For this reason, these platforms are often implemented as arrays that partially cancel these biases using a large number of independent measurements within a region. These complications are topics for the future development of this method.

The framework can be solved for many terms, each of which can be of scientific interest. For example, a researcher curious about the maximum amount of measurement uncertainty that would still permit the detection of a specified, long-term change Δ (S in this case) could solve for this U_M (Figure 1a) as

$$U_M \leq \sqrt{\left(\frac{S}{2}\right)^2 - V_T{}^2 - V_D{}^2 - V_{SS}{}^2 - V_S{}^2 - V_I{}^2 - U_A{}^2} \qquad (3)$$

Alternately, one could solve for the smallest trend (T) that could be detected for a time series of length dt (here the S is expanded to $T \times$ dt):

$$|T| \geq \frac{2N}{dt} \qquad (4)$$

Equations (1)–(4) are differentiable allowing them to be solved for many related relationships, for example, the incremental impact on dt expected for a given increase in U_M (supporting information Figure S1).

When rearranged to solve for dt (Figure 1b), equation (4) is one formulation for "time of emergence/detection" (e.g., Hawkins & Sutton, 2012; Keller et al., 2014). This framework approach to time-of-emergence estimation differs from the formulation of Weatherhead et al. (1998)—which is also used in climate science (Beaulieu et al., 2013; Henson et al., 2010; Henson et al., 2018; Sutton et al., 2018)—in that it considers timescales of variability separately and explicitly rather than inferring the net impact of all variability collectively from deseasonalized monthly time series data. Our approach has the advantage of being applicable to a broader range of climate records but the disadvantage that it requires independent information about the impacts of all timescales of variability. For our purposes therefore, even data sets or model simulations that show null results of "no emergent trends" are often useful for characterizing regional variability on different timescales.

3. Example Applications

We consider example applications related to underway ship (section 3.1), mooring (4.2), and repeat hydrographic cruise data (summarized in Table 1 and explained in subsequent sections).

3.1. pCO$_2$ Trend Near Unimak Pass, Alaska

Our first example tests whether frequently repeated surface seawater pCO$_2$ measurements (81 sets of measurements in ~21 years) through Unimak Pass, a 19-km wide and 55-m deep waterway in the Alaskan Aleutian Islands (Figure 2a), are sufficient to establish whether a long-term change has emerged from the envelope of natural variability and how large of a role measurement uncertainty has played in the ability to detect a change there. The signal in this example is the long-term trend T, combined with any portion of interannual variability V_I that acts on timescales comparable to or longer than the ~20-year data record available in version 6 of the Surface Ocean CO$_2$ Atlas data set (see Bakker et al., 2016 for a version 3 description). The noise is all other modes of variability combined with a comparatively small (~2 to 5 µatm) U_M. The constraints of Unimak Pass minimize spatial variability from ship tracks, but this is nevertheless an unresolved additional potential contribution to the signal and noise (Figure 2a).

We use several approaches to reduce the impacts of variability on this record. First, we average all pCO$_2$ values from each transect of Unimak Pass, keeping only data from 77 transects with at least 10 independent pCO$_2$ values. This averaging reduces our statistical degrees of freedom to reflect the fact that multiple
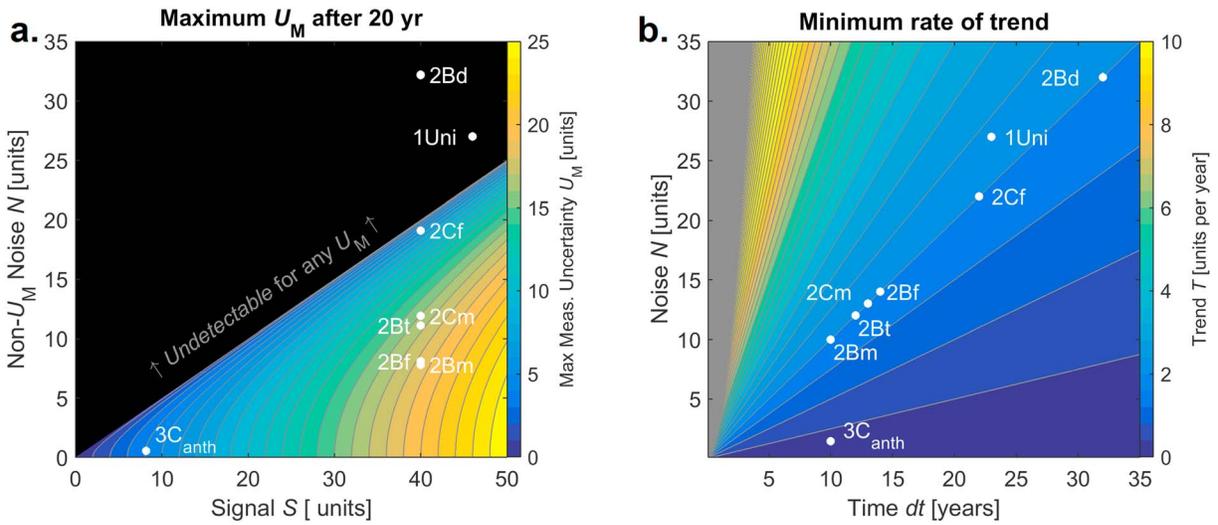
Figure 1. A schematic representation of framework equations (3) and (4) solved for (a) the maximum measurement uncertainty (U_M) that would still permit detection of various 20-year signals (S, x axis) and (b) the minimum trend T that would still be detected for a given unresolved noise N (y axis) after a length of time dt (x axis). Many example calculations in section 4 are plotted as white dots labeled as indicated by the "code" in Table 1. The black areas in Figure 1a indicate undetectable signals, and the gray area on the left of Figure 1b results from closely spaced contours (i.e., only large trends are detectable).

Table 1
Terms in All of the Various Calculations Discussed in the Examples

Code	Description	Units	U_M	V_D	V_{SS}	V_S	V_I	U_A	N	T (year^{-1})	dt (year)	S
Figure 1a, calculating maximum measurement uncertainty for given signal and variability												
1Uni	Unimak	µatm	x	-	-	-	~	-	27	2.3	<u>20</u>	46
2Bd	BATS decadal	µatm	x	2	8	31	4	-	20	2	<u>20</u>	40
2Bt	BATS monthly time series	µatm	**17**	2	8	~	4	7	20	2	<u>20</u>	40
2Bf	BATS float array	µatm	**18**	2	~	~	4	7	20	2	<u>20</u>	40
2Bm	BATS mooring	µatm	**18**	~	~	~	4	7	20	2	<u>20</u>	40
2Cd	*CCE2 decadal*	*µatm*	*x*	*15*	*36*	*33*	*9*	*-*	*20*	*2*	<u>*20*</u>	*40*
2Ct	*CCE2 monthly time series*	*µatm*	*x*	*15*	*36*	*~*	*9*	*7*	*20*	*2*	<u>*20*</u>	*40*
2Cf	CCE2 float array	µatm	**6**	15	~	~	9	8	20	2	<u>20</u>	40
2Cm	CCE2 mooring	µatm	**16**	~	~	~	9	8	20	2	<u>20</u>	40
3C$_{anth}$	Decadal C$_{anth}$	µmol kg^{-1}	**4.1**	~	~	~	~	0.55[a]	4.1[a]	0.41	<u>20</u>	8.2
Figure 1b, calculating minimum trend rate or length of trend with a given rate												
1Uni	Unimak	µatm	-	-	-	-	~	-	27	2.3	**23**	54
2Bd	BATS decadal	µatm	2	2	8	31	4	-	32	<u>2</u>	**32**	64
2Bt	BATS monthly time series	µatm	2	2	8	~	4	9	12	<u>2</u>	**12**	18
2Bf	BATS float array	µatm	11	2	~	~	4	8	14	<u>2</u>	**14**	24
2Bm	BATS mooring	µatm	2	~	~	~	4	10	10	<u>2</u>	**10**	8
2Cd	*CCE2 decadal*	*µatm*	*2*	*15*	*36*	*33*	*9*	*-*	*52*	<u>*2*</u>	**52**	*104*
2Ct	*CCE2 monthly time series*	*µatm*	*2*	*15*	*36*	*~*	*9*	*5*	*40*	<u>*2*</u>	**40**	*80*
2Cf	CCE2 float array	µatm	11	15	~	~	9	7	22	<u>2</u>	**22**	41
2Cm	CCE2 mooring	µatm	2	~	~	~	9	10	13	<u>2</u>	**13**	18
3C$_{anth}$	Decadal C$_{anth}$	µmol kg^{-1}	-	~	~	~	~	-	1.45[a]	**0.15**	<u>10</u>	-
Additional calculations discussed in text												
	Unimak float array	µatm	**11**	~	~	~	-	-	29	2.3	<u>25</u>	58
	Peru float array	µatm	11	15	~	~	66	4	66	<u>2</u>	**66**	131
	Peru mooring	µatm	2	~	~	~	63	4	63	<u>2</u>	**63**	126

Note. Each row is one calculation. Assumed quantities are underlined. Calculated quantities are in bold. The "code" refers to dots in Figures 1a and 1b with the number indicating the example set. Italicized calculations fall beyond Figure 1 axis limits. "x" = signal does not exceed variability alone, so no additional U_M noise can be calculated; "-" = not needed for calculation; see text for details; "~" = neglected, assumed small, or dealt with using simplifying assumptions.
[a]From Carter et al. (2017, Appendix B).
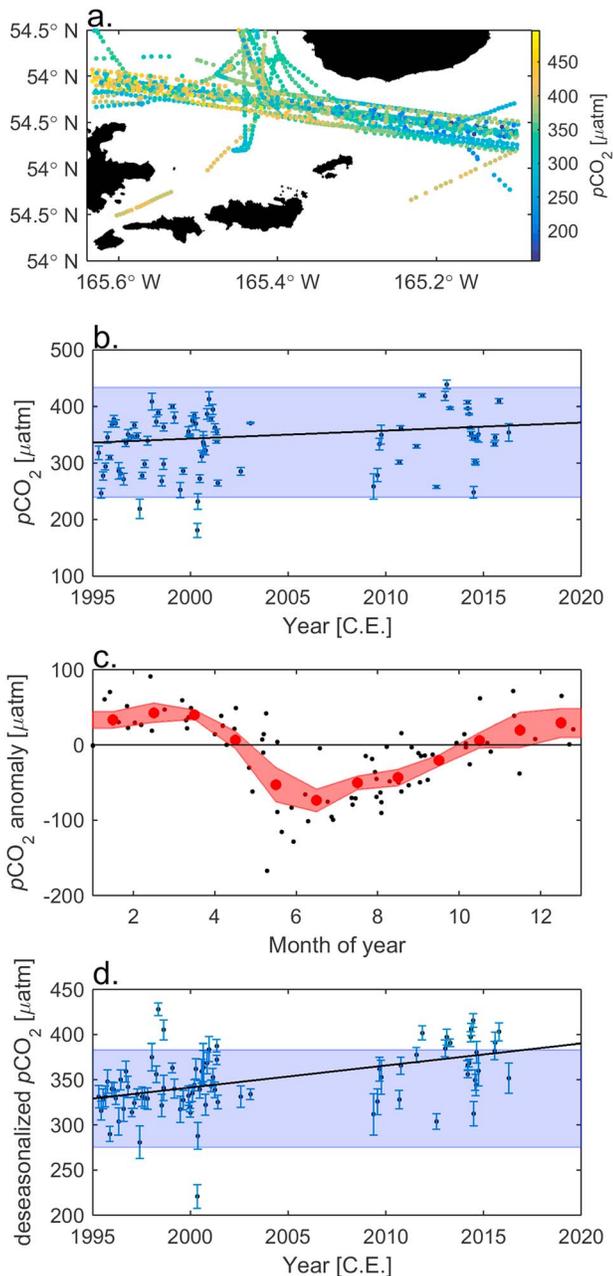
Figure 2. (a) A map of Unimak Pass with pCO_2 measured on transects plotted in color. (b) Average transect pCO_2 values plotted with error bars reflecting standard uncertainties for the mean values. A linear fit of these data reveals an insignificant trend (black line) that will not exceed twice the noise (shaded blue area with width equaling $\pm 2N$ or twice the trend line root-mean-square error of 49 µatm) in the near future. (c) These data, plotted by month of year (black dots), can be binned and averaged for each month and shown as a monthly anomaly from the long-term average pCO_2 (red dots, with the red band indicating standard uncertainties on the monthly mean values). When these bin average anomalies are interpolated to the measurement months of each transect and then subtracted from the original transect data (in b), the result is (d) deseasonalized data. Here error bars reflect combined errors from both the transect averaging and the deseasonalization. Panel (d) shows a significant trend (black line) that exceeds twice the noise (shaded blue area with a width of twice the trend line root-mean-square error of 27 µatm) before 2020.

measurements along brief transects do not provide independent realizations of V_{SS}, V_S, or V_I or arguably of U_M or V_D (with an only 2- to 3-hr transit through the region at ~20 knots). A U_A term is necessary to account for uncertainty regarding our assumption that the averaged values represent the true mean value expected along each transect. This U_A equals the standard deviations of the individual pCO_2 measurements propagated through the averaging for each transect (Figure 2b, error bars). There remains enough combined variability that the long-term trend T of the transectaveraged data (estimated using an uncertainty weighted regression) is not significant over the ~20-year observational period ($p = 0.068$, trend root-mean-square error [RMSE[= 49 µatm). However, these data still have an unknown influence from V_D, V_{SS}, and V_S. Next, an adjustment is applied to "d-seasonalize" the data (see Bates, 2001; Sutton et al., 2018; Takahashi et al., 2009): the average monthly pCO_2 anomalies relative to the long-term mean measured pCO_2 (Figure 2c) are interpolated to the average month of each transect and subtracted from the transect-mean pCO_2. The observed trend after deseasonalization is 2.3 µatm/year ($p = 5 \times 10^{-6}$; Figure 2d). The 27-µatm RMSE of this fit is due to the combination of unresolved V_D, V_{SS}, the portion of V_I acting on <20-year timescales, U_M, and a U_A term reflecting both errors in the transect averaging and errors in the deseasonalization adjustment (the latter of which averages ~13 µatm, estimated from the standard uncertainty on the average monthly anomalies used for the adjustment).

Our approach allows us to calculate that it would be ~23 years from 1995 before the observed rate of change in Unimak Pass would exceed the background of deseasonalized variability (Figure 2d), so the signal should be detectable within the next few annual releases of the Surface Ocean CO_2 Atlas. This analysis demonstrates that a significant trend can be identified even before the trend can be shown to have fully "emerged" from the envelope of natural variability provided some of the influences of natural variability can be averaged out with frequent measurements. However, since we did not separately estimate V_I acting on >20-year timescales in this example, we cannot definitively say whether such a trend owes to very long timescale interannual variability.

We can also use equation (3) to solve for the maximum U_M that would result in trend detection within 25 years (i.e., by 2020). The framework suggests that the signal could still be detected within 25 years had U_M been as large as ± 11 µatm (Table 1) or the approximate uncertainty of surface pCO_2 estimates made from profiling floats equipped with pH sensors (Williams et al., 2017). Climate quality data are therefore not required for pCO_2 trend detection within 25 years in this highly variable Aleutian pass when given a time series long and dense enough to constrain the seasonal cycle.

3.2. Measurement Timescales and Platforms

In the next example we use highly temporally resolved data and model output to estimate all timescales of variability in the framework at two disparate locations, thereby allowing an exploration of the interplay between measurement frequency, measurement uncertainty, natural variability, and signal detection in different regions. We consider surface pCO_2 variability at two locations, both of which are equipped with moored autonomous pCO_2 measurement systems (Sutton et al., 2014) sampling at intervals of 3 hr: the Bermuda Atlantic Time Series (BATS,

64°W, 31.5°N; Joyce & Robbins, 1996) in the oligotrophic North Atlantic subtropical gyre and a mooring in the California Current Ecosystem (CCE2, 121.8°W, 34.324°N; plotted in supporting information S2). The CCE2 location experiences intense variability from upwelling and biological productivity on seasonal and subseasonal timescales, whereas the BATS location is an open ocean site with substantially less high frequency variability. V_D, V_{SS}, and V_S are estimated as the average moving window standard deviations of sets of eight consecutive three-hourly pCO_2 values, 30 consecutive daily average pCO_2 values, and 12 consecutive monthly average pCO_2 values, respectively. Measurement uncertainty is contributing to the observed daily variability, so V_D has the ± 2 μatm U_M (Sutton et al., 2014) subtracted from this average moving window standard deviation in quadrature. We supplement these data records with fully coupled Earth system model output from the HadGEM preindustrial control run (r1i1p1) with fixed atmospheric pCO_2 (Collins et al., 2008; supporting information S3). This affords V_I estimates at these locations as the standard deviations of annually averaged values (see Table 1).

A plausible surface ocean pCO_2 increase of ~2 μatm/year (Sutton et al., 2018) would be expected to exceed natural variability and a small U_M (± 2 μatm) after 26 years at BATS and after 42 years at CCE2 using decadal measurements that resolve no timescales of variability. If we could eliminate all impacts of V_S only, V_S and V_{SS} only, and all three of V_S, V_{SS}, and V_D, then these detection times would decrease from 26 to 9, 5, and 4 years at BATS and from 42 to 30, 18, and 9 years at CCE2, respectively (not shown in Table 1). Example platforms that are capable of resolving these modes of variability are monthly time series measurements (Joyce & Robbins, 1996), 5- to 10-day cycling profiling biogeochemical Argo floats (Johnson & Claustre, 2016), and continuously measuring autonomous moorings (Sutton et al., 2014), respectively. However, realistically these platforms only gradually improve their resolution of the modes of variability as moreindependent realizations of the variability are recorded. Therefore, for the calculations in Table 1 we assess a U_A equaling the eliminated modes of variability divided by the square root of the number of realizations of each mode of variability after dt years and added in quadrature. For example, for moorings, U_A equals

$$U_A = \sqrt{\frac{V_D{}^2}{365\mathrm{d}t} + \frac{V_{SS}{}^2}{12\mathrm{d}t} + \frac{V_S{}^2}{\mathrm{d}t}} \qquad (5)$$

We then solve for dt iteratively (an exact solution is achievable but many termed). These calculations suggest that monthly time series data are nearly as effective as moorings (12 years to emerge with a monthly timeseries compared to 10 years for a mooring) at BATS but would be significantly less effective at CCE2 (40 and 13 years, respectively) due to high daily and subseasonal pCO_2 variability. The moored autonomous systems are effective in both locations due to their high sampling frequency and measurement accuracy.

With an ± 11-μatm pCO_2 uncertainty (at 400 μatm; Williams et al., 2017), the framework suggests that profiling float observations would require 14 and 22 years to detect a pCO_2 trend at BATS and CCE2, respectively. This is longer than is required for moorings (10 and 13 years, respectively) due primarily to the larger float-based pCO_2 uncertainty at BATS and the inability of floats to resolve daily variability at CCE2. Floats are also less effective than time series measurements at BATS, but the reverse is true at CCE2 where floats can detect a trend sooner by resolving subseasonal variability. While the dt required for a profiling float at CCE2 is nearly double the requirement for a mooring, the float dt would only be 5% greater at CCE2 had that region exhibited the much larger (± 63 μatm) interannual variability seen in the upwelling region offshore Peru in the same HadGEM simulation (Table 1, supporting information S3). The ideal measurement platform therefore depends significantly on the expected magnitudes and timescales of variability. For example, moorings are particularly effective in regions where high-frequency variability dominates, whereas a more cost-effective platform that samples less frequently could be preferable in regions with larger low-frequency variability.

A nearly identical (~13 years) time of detection value is obtained from Weatherhead et al.'s (1998) formulation when applied to deseasonalized monthly mooring data at CCE2 (Sutton et al., 2018, omitting the conservative 40% dt increase that they assign to allow for potential additional unresolved interannual variability).

3.3. Anthropogenic Carbon Accumulation

Examples so far primarily demonstrate the importance of measurement frequency, but meeting the stringent "climate quality" data requirements for carbonate system measurements (Newton et al., 2015) is critical for some OA monitoring applications. One example of this comes from inferring decadal interior ocean anthropogenic carbon (C_{anth}) storage from repeat hydrography (e.g., Carter et al., 2017; Kouketsu et al., 2013; Pardo et al., 2014; Waters et al., 2011; Williams et al., 2015; Woosley et al., 2016). With only one synoptic "snapshot" of ocean chemistry each decade, this observation strategy does not allow any modes of natural variability to be directly resolved through repeated observations. However, researchers have developed means of accounting for the impacts of natural variability on dissolved inorganic carbon (DIC) to resolve the signal of interest. One example approach is the extended multiple linear regression (eMLR) technique (Friis et al., 2005) which uses empirical regressions relating DIC to other measured properties affected by natural variability to estimate and remove the impacts of natural variability on DIC (i.e., using diverse measurements to resolve natural variability instead of frequent measurements). The eMLR method has been shown to be partially effective (Carter et al., 2017; Clement & Gruber, 2018; Plancherel et al., 2013), and the purely methodological error was recently estimated from simulations of model output with known C_{anth} to average ±0.55-μmol/kg C_{anth} (U_A) over the 200- to 1,500-m depth range where eMLR is most often applied (Carter et al., 2017). However, this uncertainty increases to 1.45-μmol/kg C_{anth} when simulated measurement errors common for "climate quality" repeat hydrographic observations are also included (i.e., U_A and U_M combined). These two uncertainties are the only remaining noise terms in the framework since the variability terms are assumed to be accounted for by eMLR.

An example C_{anth} signal is the average C_{anth} storage over the 200- to 1,500-m depth range of the Pacific Ocean along ~180°E (equaling 4.1-μmol/kg C_{anth} from 1995 to 2005 in the model simulation considered by Carter et al., 2017, Appendix B). Measurement uncertainties increase noise from 13% (U_A) to 35% (U_A and U_M) of that signal and increase the time of detection from 2.7 to 7.1 years. The eMLR method is therefore capable of detecting the signal with each decadal repeat, and the simulated DIC measurement uncertainty is the dominant source of uncertainty for the decadal C_{anth} estimates. The significant noise contribution from measurement uncertainties both prevents detection in (e.g., deeper) water masses with lower C_{anth} accumulation rates and increases the estimate uncertainty (i.e., lowers the S/N ratio) in water masses with higher accumulation rates. The N term could therefore be directly reduced by more accurate measurements or—as measurement uncertainties are mostly statistically independent between synoptic surveys—by more frequent or numerous hydrographic line reoccupations to reduce the DIC trend uncertainty. At the current rate of sampling, even the most stringent climate quality observation uncertainties limit the ability to infer interior ocean decadal C_{anth} distribution changes in broad swaths of ocean. Therefore, the highly accurate "climate quality" data produced by repeated hydrographic surveys are indeed needed.

4. Conclusions

The framework is intended to quantify the trade-offs between different observing approaches with respect to the impacts of measurement uncertainty, frequency, and duration and to assess how the importance of these factors depends on the potential signal magnitude and regional natural variability. However, a full consideration of the best measurement approach for a climate observing challenge must involve practical concerns not covered by the framework. Example additional considerations include measurement costs, platform suitability for the region(s) of interest, sensor availability and viability for the signals of interest, the availability of personnel with appropriate expertise, and potential synergies with existing, planned, and past observing efforts.

The examples we consider reveal lessons for an observing system planning exercise: (1) highly accurate measurements are needed to resolve some, but not all, critical climate signals, and the impacts of improving measurement confidence are quantifiable within the time-of-detection context; (2) emergent climate trends can generally be identified sooner with more frequent measurements, particularly for systems that are highly variable on shorter timescales; (3) records that do not identify emergent climate signals still provide important constraints for future oceanographic research; and (4) the comparative merits of different observing approaches (e.g., repeat hydrography, time series, float arrays, and moorings) depend significantly on the expected magnitudes of variabilty over a range timescales, on the uncertainties of the measurements

produced, and on the uncertainties of the data reduction strategies applied during data workup. Broadly, repeat hydrography can be ideal when data transformations can limit the impacts of natural variability on climate signals, monthly time series are effective at detecing climate signals in areas with low daily and subseasonal variability, float arrays are effective for a range of environments where broad regional trends and patterns are of interest, and moorings can be ideal when resources can be directed at a single question or region, especially when the region experiences high-frequency natural variability.

References

Bakker, D., Pfeil, B., Landa, C., Metzl, N., & O'Brien, K. (2016). A multi-decade record of high-quality fCO$_2$ data in version 3 of the Surface Ocean CO$_2$ Atlas (SOCAT). *Earth System Science Data*, *8*(2), 383–413. https://doi.org/10.5194/essd-8-383-2016

Bates, N. R. (2001). Interannual variability of oceanic CO$_2$ and biogeochemical properties in the Western North Atlantic subtropical gyre. *Deep Sea Research Part II: Topical Studies in Oceanography*, *48*(8-9), 1507–1528. https://doi.org/10.1016/S0967-0645(00)00151-X (Accessed September 6, 2018).

Beaulieu, C., Henson, S. A., Sarmiento, J. L., Dunne, J. P., Doney, S. C., Rykaczewski, R. R., & Bopp, L. (2013). Factors challenging our ability to detect long-term trends in ocean chlorophyll. *Biogeosciences*, *10*(4), 2711–2724. https://doi.org/10.5194/bg-10-2711-2013 (Accessed September 7, 2018).

Carter, B. R., Feely, R. A., Mecking, S., Cross, J. N., Macdonald, A. M., Siedlecki, S. A., et al. (2017). Two decades of Pacific anthropogenic carbon storage and ocean acidification along Global Ocean Ship-based Hydrographic Investigations Program sections P16 and P02. *Global Biogeochemical Cycles*, *31*, 306–327. https://doi.org/10.1002/2016GB005485 (Accessed April 17, 2017).

Carter, B. R., Frölicher, T. L., Dunne, J. P., Rodgers, K. B., Slater, R. D., & Sarmiento, J. L. (2016). When can ocean acidification impacts be detected from decadal alkalinity measurements? *Global Biogeochemical Cycles*, *30*, 595–612. https://doi.org/10.1002/2015GB005308 (Accessed October 10, 2018).

Christian, J. R. (2014). Timing of the departure of ocean biogeochemical cycles from the preindustrial state. *PLoS ONE*, *9*(11), e109820. https://doi.org/10.1371/journal.pone.0109820 (Accessed December 6, 2018).

Clement, D., & Gruber, N. (2018). The eMLR(C*) method to determine decadal changes in the global ocean storage of anthropogenic c. *Global Biogeochemical Cycles*, *32*, 654–679. https://doi.org/10.1002/2017GB005819 (Accessed April 5, 2018).

Collins, W., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Hinton, T., Jones, C. D., et al. (2008). Evaluation of the HadGEM2 model. Retrieved from http://www.inscc.utah.edu/~reichler/publications/papers/Collins_08_MetOffice_74.pdf (Accessed July 31, 2018).

Fassbender, A. J., Alin, S. R., Feely, R. A., Sutton, A. J., Newton, J. A., Krembs, C., et al. (2018). Seasonal carbonate chemistry variability in marine surface waters of the US Pacific northwest. *Earth System Science Data*, *10*(3), 1367–1401. https://doi.org/10.5194/essd-10-1367-2018 (Accessed October 25, 2018).

Fassbender, A. J., Rodgers, K. B., Palevsky, H. I., & Sabine, C. L. (2018). Seasonal asymmetry in the evolution of Surface Ocean pCO$_2$ and pH thermodynamic drivers and the influence on sea-air CO$_2$ flux. *Global Biogeochemical Cycles*, *32*, 1476–1497. https://doi.org/10.1029/2017GB005855 (Accessed December 6, 2018).

Friis, K., Körtzinger, A., Pätsch, J., & Wallace, D. W. R. (2005). On the temporal increase of anthropogenic CO$_2$ in the subpolar North Atlantic. *Deep Sea Research Part I: Oceanographic Research Papers*, *52*(5), 681–698. https://doi.org/10.1016/j.dsr.2004.11.017, http://www.sciencedirect.com/science/article/pii/S0967063705000087 (Accessed November 17, 2015).

Frölicher, T. L., Joos, F., Plattner, G.-K., Steinacher, M., & Doney, S. C. (2009). Natural variability and anthropogenic trends in oceanic oxygen in a coupled carbon cycle-climate model ensemble. *Global Biogeochemical Cycles*, *23*, GB1003. https://doi.org/10.1029/2008GB003316 (Accessed September 24, 2015).

Hawkins, E., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, *39*, L01702. https://doi.org/10.1029/2011GL050087 (Accessed June 20, 2018).

Henson, S. A., Beaulieu, C., Ilyina, T., John, J. G., Long, M., Séférian, R., et al. (2017). Rapid emergence of climate change in environmental drivers of marine ecosystems. *Nature Communications*, *8*. https://doi.org/10.1038/ncomms14682

Henson, S. A., Cole, H. S., Hopkins, J., Martin, A. P., & Yool, A. (2018). Detection of climate change-driven trends in phytoplankton phenology. *Global Change Biology*, *24*, e101–e111. https://doi.org/10.1111/gcb.13886 (Accessed September 6, 2018).

Henson, S. A., Sarmiento, J. L., Dunne, J. P., Bopp, L., Lima, I., Doney, S. C., et al. (2010). Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity (pp. 621-640). Retrieved from www.biogeosciences.net/7/621/2010/ (Accessed September 6, 2018).

Ilyina, T., Zeebe, R. E., Maier-Reimer, E., & Heinze, C. (2009). Early detection of ocean acidification effects on marine calcification. *Global Biogeochemical Cycles*, *23*, GB1008. https://doi.org/10.1029/2008GB003278 Accessed September 24, 2015

Johnson, K., & Claustre, H. (2016). Bringing biogeochemistry into the Argo age. *Eos, Transactions of the American Geophysical Union*, *97*. https://doi.org/10.1029/2016EO062427

Joyce, T. M., & Robbins, P. (1996). The Long-term hydrographic record at Bermuda. *Journal of Climate*, *9*(12), 3121–3131. https://doi.org/10.1175/1520-0442(1996)009<3121:TLTHRA>2.0.CO;2 (Accessed September 24, 2015).

Keller, K. M., Joos, F., & Raible, C. C. (2014). Time of emergence of trends in ocean biogeochemistry. *Biogeosciences*, *11*(13), 3647–3659. https://doi.org/10.5194/bg-11-3647-2014 (Accessed September 24, 2015).

Kouketsu, S., Murata, A., & Doi, T. (2013). Decadal changes in dissolved inorganic carbon in the Pacific Ocean. *Global Biogeochemical Cycles*, *27*, 65–76. https://doi.org/10.1029/2012GB004413 (Accessed January 12, 2017).

McKinley, G. A., Pilcher, D. J., Fay, A. R., Lindsay, K., Long, M. C., & Lovenduski, N. S. (2016). Timescales for detection of trends in the ocean carbon sink. *Nature*, *530*(7591), 469–472. https://doi.org/10.1038/nature16958 (Accessed June 20, 2018).

Newton, J. A., Feely, R. A., Jewett, E. B., Williamson, P., & Mathis, J. (2015). Global Ocean Acidification Observing Network: Requirements and Governance Plan Second Edition GOA-ON Global Ocean Acidification Observing Network. Retrieved from www.iaea.org/ocean-acidification (Accessed July 30, 2018).

Pardo, P. C., Pérez, F. F., Khatiwala, S., & Ríos, A. F. (2014). Anthropogenic CO$_2$ estimates in the Southern Ocean: Storage partitioning in the different water masses. *Progress in Oceanography*, *120*, 230–242. https://doi.org/10.1016/J.POCEAN.2013.09.005, https://www.sciencedirect.com/science/article/pii/S0079661113001833 (Accessed April 5, 2018).

Plancherel, Y., Rodgers, K., Key, R., Jacobson, A., & Sarmiento, J. (2013). Role of regression model selection and station distribution on the estimation of oceanic anthropogenic carbon change by eMLR. *Biogeosciences*, *10*(7), 4801–4831. https://doi.org/10.5194/bg-10-4801-2013, http://biogeosciences-discuss.net/9/C6421/2012/bgd-9-C6421-2012-supplement.pdf (Accessed November 17, 2015).

Rodgers, K., Lin, J., & Frölicher, T. (2015). Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model. *Biogeosciences*, *12*(11), 3301–3320. https://doi.org/10.5194/bg-12-3301-2015, http://www.researchgate.net/profile/Thomas_Froelicher/publication/278859338_Emergence_of_multiple_ocean_ecosystem_drivers_in_a_large_ensemble_suite_with_an_Earth_system_model/links/5587a49908aeb0cdade0c9aa.pdf (Accessed September 24, 2015).

Sutton, A. J., Feely, R. A., Maenner-Jones, S., Musielwicz, S., Osborne, J., Monacci, N., et al. (2018). Autonomous seawater pCO$_2$ and pH time series from 40 surface buoys and the emergence of anthropogenic trends. *Earth System Science Data Discussions*, 1–26. https://doi.org/10.5194/essd-2018-114 (Accessed October 4, 2018).

Sutton, A. J., Sabine, C. L., Maenner-Jones, S., Lawrence-Slavas, N., Meinig, C., Feely, R. A., et al. (2014). A high-frequency atmospheric and seawater pCO$_2$ data set from 14 open-ocean sites using a moored autonomous system. *Earth System Science Data*, *6*(2), 353–366. https://doi.org/10.5194/essd-6-353-20142014

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., et al. (2009). Climatological mean and decadal change in surface ocean pCO$_2$, and net sea–air CO$_2$ flux over the global oceans. *Deep Sea Research Part II: Topical Studies in Oceanography*, *56*(8-10), 554–577. https://doi.org/10.1016/J.DSR2.2008.12.009, https://www.sciencedirect.com/science/article/pii/S0967064508004311 (Accessed September 6, 2018).

Waters, J. F., Millero, F. J., & Sabine, C. L. (2011). Changes in South Pacific anthropogenic carbon. *Global Biogeochemical Cycles*, *25*, GB4011. https://doi.org/10.1029/2010GB003988 (Accessed April 5, 2018).

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X. L., Choi, D., Cheang, W. K., et al. (1998). Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *Journal of Geophysical Research*, *103*(D14), 17,149–17,161. https://doi.org/10.1029/98JD00995 (Accessed September 6, 2018).

Williams, N. L., Feely, R. A., Sabine, C. L., Dickson, A. G., Swift, J. H., Talley, L. D., & Russell, J. L. (2015). Quantifying anthropogenic carbon inventory changes in the Pacific sector of the Southern Ocean. *Marine Chemistry*, *174*, 147–160. https://doi.org/10.1016/j.marchem.2015.06.015, http://www.sciencedirect.com/science/article/pii/S0304420315001309 (Accessed October 21, 2015).

Williams, N. L., Juranek, L. W., Feely, R. A., Johnson, K. S., Sarmiento, J. L., Talley, L. D., et al. (2017). Calculating surface ocean pCO$_2$ from biogeochemical Argo floats equipped with pH: An uncertainty analysis. *Global Biogeochemical Cycles*, *31*, 591–604. https://doi.org/10.1002/2016GB005541 (Accessed June 2, 2017).

Woosley, R. J., Millero, F. J., & Wanninkhof, R. (2016). Rapid anthropogenic changes in CO$_2$ and pH in the Atlantic Ocean: 2003–2014. *Global Biogeochemical Cycles*, *30*, 70–90. https://doi.org/10.1002/2015GB005248 (Accessed January 23, 2016)