AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Oceans

RESEARCH ARTICLE

10.1002/2017JC013023

Key Points:

- Geostatistical techniques isolate mesoscale variability independent of satellite platform
- Unresolved variability scales proportional to surface chlorophyll except low-chlorophyll regions
- Physical stirring of large-scale gradient supports observed mesoscale variability in surface chlorophyll

Correspondence to:
D. M. Glover,
dglover@whoi.edu

Citation:
Glover, D. M., Doney, S. C., Oestreich, W. K., & Tullo, A. W. (2018). Geostatistical analysis of mesoscale spatial variability and error in SeaWiFS and MODIS/Aqua global ocean color data. *Journal of Geophysical Research: Oceans*, *123*. https://doi.org/10.1002/2017JC013023

Received 24 APR 2017
Accepted 4 DEC 2017
Accepted article online 11 DEC 2017

© 2017. The Authors.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Geostatistical Analysis of Mesoscale Spatial Variability and Error in SeaWiFS and MODIS/Aqua Global Ocean Color Data

David M. Glover[1], Scott C. Doney[1,2], William K. Oestreich[1,3], and Alisdair W. Tullo[1,4]

[1]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA, [2]Now at Department of Environmental Sciences, University of Virginia, Charlottesville, VA, USA, [3]Now at Department of Biology, Hopkins Marine Station, Stanford University, Pacific Grove, CA, USA, [4]Now at School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK

Abstract Mesoscale (10–300 km, weeks to months) physical variability strongly modulates the structure and dynamics of planktonic marine ecosystems via both turbulent advection and environmental impacts upon biological rates. Using structure function analysis (geostatistics), we quantify the mesoscale biological signals within global 13 year SeaWiFS (1998–2010) and 8 year MODIS/Aqua (2003–2010) chlorophyll *a* ocean color data (Level-3, 9 km resolution). We present geographical distributions, seasonality, and interannual variability of key geostatistical parameters: unresolved variability or noise, resolved variability, and spatial range. Resolved variability is nearly identical for both instruments, indicating that geostatistical techniques isolate a robust measure of biophysical mesoscale variability largely independent of measurement platform. In contrast, unresolved variability in MODIS/Aqua is substantially lower than in SeaWiFS, especially in oligotrophic waters where previous analysis identified a problem for the SeaWiFS instrument likely due to sensor noise characteristics. Both records exhibit a statistically significant relationship between resolved mesoscale variability and the low-pass filtered chlorophyll field horizontal gradient magnitude, consistent with physical stirring acting on large-scale gradient as an important factor supporting observed mesoscale variability. Comparable horizontal length scales for variability are found from tracer-based scaling arguments and geostatistical decorrelation. Regional variations between these length scales may reflect scale dependence of biological mechanisms that also create variability directly at the mesoscale, for example, enhanced net phytoplankton growth in coastal and frontal upwelling and convective mixing regions. Global estimates of mesoscale biophysical variability provide an improved basis for evaluating higher resolution, coupled ecosystem-ocean general circulation models, and data assimilation.

1. Introduction

Before remote sensing satellites provided regular, near-synoptic views of ocean basin scale surface distributions of phytoplankton, conferences (e.g., Steele, 1978) were being held to discuss and debate the character of, biophysical mechanisms responsible for, and means of measuring plankton patchiness. This interest was driven by a desire to predict both the size of these plankton patches and understand their underlying dynamics. This knowledge is sought to improve our ability to use satellite data to guide models (data assimilation) to assess the impact of marine ecosystems (e.g., fisheries) upon the ocean global role in planetary climate modulation. Initially, much of the characterization and measurement was carried out through 1-D spectral analysis of chlorophyll *a* distributions from field measurements along ship-tracks (Platt, 1972), theory (Steele, 1974), or theory and measurements (Denman & Platt, 1976). With time, the mesoscale variability patterns displayed by plankton in the surface layer of the ocean (of which patches are just one) came to be associated with the interaction of turbulent mixing and horizontal gradients of tracer distributions (Abraham, 1998; Bennett & Denman, 1985; Franks, 2005; Garrett, 1989, 2006; Lévy, 2003; Lévy & Martin, 2013; Mackas et al., 1985; Powell & Okubo, 1994, to name just a few). A few, more adventurous, researchers attempted to answer some of these questions with structure function analysis (Denman & Freeland, 1985; Yoder et al., 1987) once satellite images became digitally available. It has been over 40 years since the early 1970s and characterizing the nature and causes of patchiness continues to be a challenge (van Gennip et al., 2016). Into this milieu we offer our structure function analysis of the remotely sensed, mesoscale resolution distributions of surface chlorophyll *a* derived from global, multiyear data sets from NASA's Sea-viewing Wide Field-of-view Sensor (SeaWiFS) and Moderate resolution Imaging Spectroradiometer onboard the EOS-AM satellite (MODIS/A).

The idea of using remotely sensed data to illuminate features of scalar variables near the sea surface has been around for a long time. In the past, spatial scale analysis has been accomplished largely with spectral methods requiring various data smoothing preprocessing to avoid problems introduced by data losses due to masking clouds and optically thick aerosols, or sensor performance issues (Denman & Abbott, 1988, 1994; Denman & Platt, 1976; Gower et al., 1980). Some of these drawbacks can be ameliorated with the use of structure function analysis (Doney et al., 2003; Denman & Freeland, 1985; Yoder et al., 1987). Denman and Freeland (1985) present one of the earliest applications of spatial structure functions (variograms) to oceanographic measurements of chlorophyll *a* taken shipboard during a 3 year period off the west coast of Vancouver Island, Canada. Their efforts to estimate the correlation functions and unresolved noise levels for the purpose of objective mapping of four oceanographic variables, including log-transformed phytoplankton chlorophyll *a* values had no obscuration by clouds. Yoder et al. (1987) followed with an application of structure function analysis to Coastal Zone Color Scanner (CZCS) data from the southeast U.S. continental shelf. Their results demonstrate that satellite data (with cloud induced data dropouts) could be used as input to structure function analysis in the mesoscale range of oceanographic variability. Doney et al. (2003) present monthly, globally distributed structure function results of spatial variability for 1 year (1998) of daily, Level-3 (L-3) standard mapped image (SMI) SeaWiFS data. In this paper, we expand upon their work and present an analysis of both the geographic patterns of mesoscale spatial variability and their corresponding time series for the 13 year (1998–2010) and 8 year (2003–2010) periods of L-3 SMI (nominal 9 km resolution) data from SeaWiFS and MODIS/A sensors.

In this study, we use structure function analysis, also known as variogram analysis (Clark, 1979; Journel & Huijbregts, 1978), to explore the spatial distribution of mesoscale variance in ocean color imagery. By analyzing the distribution of variance spatially we can separate significant resolved mesoscale biological signals from smaller-scale unresolved geophysical variability and systematic noise. Additionally, structure function determination is a necessary prerequisite for objectively mapping ocean color imagery (Bretherton et al., 1976). Previously, spectral methods have been the mathematics of choice to explain patchiness in variability of phytoplankton chlorophyll *a* (Bennett & Denman, 1985; Denman et al., 1977; Franks, 2005; Garrett, 1989, 2006; Lévy & Martin, 2013; Mackas et al., 1985; Martin, 2003; Powell & Okubo, 1994) in theoretical studies and (Denman & Abbott, 1988, 1994; Denman & Platt, 1976; Gower et al., 1980; Platt, 1972; Shi et al., 2015; Smith et al., 1988; van Gennip et al., 2016) in remotely sensed and shipboard data studies. Many of these previous studies either averaged several images together or carefully chose only clear images to avoid gaps in the data too large to ignore. Our choice of structure function analysis (Yaglom, 1957) was guided by our desire to probe multiple length scales of mesoscale satellite ocean color data comparable to spectral analysis but without the operational problems presented by missing data.

In this paper we present results of our study that are an expanded analysis of the geographic, mesoscale patterns of variability begun in Doney et al. (2003). Estimates of resolved and unresolved variance (uncertainty) as well as decorrelation scale lengths are found for the global, daily 13 year and 8 year time series of SeaWiFS and MODIS/A. We also present a mathematical framework to link the decorrelation scales results from the structure function analysis to mixing scales (L_{tracer}) that act upon horizontal tracer gradients (Garrett, 1989, 2006) producing patchiness. This paper starts with a mathematical derivation of tracer mixing scales related to the root mean square chlorophyll anomaly fluctuations and the large-scale chlorophyll mean gradient followed by a brief discussion of structure function usage and ocean color data sources and processing. In Section 3, we present geographic and temporal ocean color variability at both local and global scales separating resolved, mesoscale phytoplankton variance from apparent smaller-scale geophysical or systematic noise or their combination. Similarities and differences between results derived from SeaWiFS and MODIS/A are also provided. The paper concludes with the global, time varying patterns of decorrelation scales and their relationship with an eddy scale tracer mixing length demonstrating possible geographic distribution of passive and nonpassive tracer behavior of phytoplankton in the upper ocean. We then summarize our results and make a few remarks about the apparent lack of any long-term temporal trend in our results.

2. Data and Methods

2.1. Mathematical Framework

The primary variables analyzed in this investigation are the near-surface chlorophyll *a* (Chl *a*) concentration estimates from the ocean color algorithms OC4 and OC3M applied to data from the SeaWiFS and MODIS/

Aqua satellite instruments, respectively (O'Reilly et al., 2000). These chlorophyll *a* estimates are log transformed (Campbell, 1995) to render them into approximately normal distributed data $C = \log_{10}(\mathrm{Chl}a)$ (see Appendix A on log-normal variables). For more details about the satellite data handling and processing see section 2.3.

Our mathematical framework begins with a standard Reynolds decomposition applied to the log transformed, satellite Chl *a* estimates (refer to any introductory text on turbulence, e.g., Tennekes & Lumley, 1972).

$$C = \overline{C} + C' \tag{1}$$

where C represents the base-10 log transform of the satellite chlorophyll estimate, $\overline{C}$ represents the long-term, large-scale mean of the base-10 log transform of chlorophyll, and C' represents the base-10 log transform anomalies from the long-term, large-scale mean.

We apply an analysis taken from Klocker and Abernathey (2014) where we introduce the concept of a mean free path (L_{mix}), as in thermodynamics, defined as the length over which a parcel of water is moved by eddy rotation without changing its properties (i.e., without mixing). Following Klocker and Abernathey (2014), we create a parameter that is a mixing length dependent upon a tracer (L_{tracer}) and make an empirical function of the tracer fluctuation anomaly as in:

$$L_{\mathrm{tracer}} = \frac{\sqrt{\overline{C'^2}}}{\partial \overline{C}/\partial x} \tag{2}$$

where $\sqrt{\overline{C'^2}}$ is the RMS of the tracer fluctuation (an estimate of anomaly variance) and $\partial \overline{C}/\partial x$ is the large-scale gradient of the tracer mean. Equation (2) is derived from the simple idea that tracer anomalies arise from turbulent mixing across a mean gradient (Garrett, 1989) with two assumptions (Klocker & Abernathey, 2014): (1) tracer fluctuations are generated by *local* stirring of larger-scale tracer gradients (with no advection of tracer variance from upstream regions) and (2) the gradient of the tracer varies slowly over the distance L_{tracer}. As such, equation (2) is a theoretical expression for an eddy scale relevant for mixing this tracer (in our case, chlorophyll *a*). A true estimate of tracer fluctuation variance (uncontaminated by processes unrelated to turbulent mesoscale mixing of the upper ocean) when plotted against the local larger-scale gradient should yield a linear relationship with a slope of L_{tracer} if the tracer behaves in a passive fashion.

2.2. Structure Functions

To measure the local spatial variability apparent in ocean color imagery, we use structure function, or variogram, analysis that estimates the change in spatial variance as a function of the vector distance ($\mathbf{h}$) between data points (Journel & Huijbregts, 1978). Structure function analysis is covered in many places (Cressie, 1993; Glover et al., 2011; Journel & Huijbregts, 1978). Briefly, frequent interruption of satellite image data (clouds, etc.) requires that semivariances that make up the structure function be computed empirically as in this one-dimensional example:

$$\gamma^*(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{k}^{N(\mathbf{h})} \left[C'(\mathbf{x}_k) - C'(\mathbf{x}_k + \mathbf{h})\right]^2 \tag{3}$$

where γ^* is the empirical semivariance, $\mathbf{h}$ is the vector of distances between data pairs, $N(\mathbf{h})$ is the number of data pairs at each distance h, C' as above and $\mathbf{x}_k$ is a spatial location within the data set.

Variograms are rarely analyzed in their empirical form and are rather reduced to parameters of a model function obtained from a nonlinear regression of the empirical semivariances. In this work, we use a spherical model given as:

$$\gamma(\mathbf{h}) = \begin{cases} c_0 + (c_\infty - c_0)\left[\frac{3}{2}\frac{\mathbf{h}}{d} - \frac{1}{2}\frac{\mathbf{h}^3}{d^3}\right] & \text{for} \quad h \le d \\ c_\infty & \text{for} \quad h > d; \end{cases} \tag{4}$$

where c_0 represents the unresolved variance (or *nugget* in geostatistical terms), c_∞ represents the total variability (*sill*), d represents the decorrelation scale length (*range*), and the difference $c_\infty - c_0$ is known as the

resolved variability (*relative sill*). This function is a curve that tends to zero as h tends to zero and rises in an arc fashion to a plateau and then remains constant. The intercept (c_0) represents unresolved variance in the data, variability that cannot be resolved at the spatial resolution of the data. The plateau value (c_∞) is closely related to the total data spatial variance and the difference between the two ($c_\infty - c_0$) represents the variance that can be resolved by the data set under analysis. The decorrelation scale length (d) is an estimate of how far apart any two data points must be to exhibit statistically independent behavior.

2.3. Ocean Color Data

Data from both SeaWiFS and MODIS/Aqua (MODIS/A) satellite instruments provide near-global coverage every 2 days in 8 or 36 visible and near-infrared spectral bands, respectively (NASA, 2014a, 2014b). Spectral bands common to both SeaWiFS and MODIS/A are combined into ratios of water leaving radiances to make an estimate of the upper-water column chlorophyll *a* concentration following the OC4 or OC3M algorithms, respectively (O'Reilly et al., 2000). These data are then processed into various levels of resolution and provided via the NASA Ocean Color web site (http://oceancolor.gsfc.nasa.gov/cms/) by the Ocean Biology Processing Group (OBPG) at NASA's Goddard Space Flight Center (GSFC) as either standard mapped image (SMI) or binned (BIN) data files. Level-3 SMI data are created from L-3 BIN data stored on a nearly equal-area, sinusoidal grid. Level-3 BIN data are, in turn, created from lower processing level (Level-2) data aggregated onto this grid over specified time periods (daily, 8 day weeks, monthly, etc.). The L-3 SMI data are stored on an equirectangular grid introducing some distortion in the east–west directions toward the pole. The distortion follows a simple relationship between the pixels and their geographic location and is compensated for during our analysis.

There are arguably significant differences between the two sensors SeaWiFS and MODIS/A. SeaWiFS data quantization is 10 bit versus 12 bit quantization for MODIS/A data. Further SeaWiFS signal-to-noise ratio (SNR) is on the order of three times lower than MODIS/A. This is of particular importance in the near-infrared bands that are used for atmospheric correction because greater noise in these bands can lead to greater noise in Chl *a* concentrations, particularly in waters with low Chl *a* concentrations such as oligotrophic gyres (Hu et al., 2012), where the water leaving signal at the longest wavelength employed in the Chl *a* algorithm (547 and 555 nm, respectively) is minimal. Additionally, SeaWiFS L-3 SMI data products are produced from the Global Area Coverage (GAC) observational sampling strategy employed by the SeaStar satellite platform, i.e., nominal 1 km observations are subsampled onboard the satellite as one sample saved every fourth pixel every forth scan line. Consequently each SeaWiFS 9 km pixel has 1/16 as many samples in its 9 km bin as does MODIS/A (McClain et al., 2004). Fewer samples translate to larger variance in the mean. Finally, the subsampling scheme has another effect on SeaWiFS data noise, namely that complete information about the light field is not available when SeaWiFS data is postprocessed on the ground. Hence, stray light identification and correction of SeaWiFS data is incomplete (stray light being defined as light in the optical system that was not intended in the design) (Hu et al., 2000; Uz & Yoder, 2004). For these reasons, SeaWiFS data should be expected to have more noise than MODIS/A, and our work helps to quantify its magnitude.

The preparation of ocean color data for analysis in this paper closely follows the procedures in Doney et al. (2003). Briefly, daily L-3 SMI SeaWiFS and MODIS/A data are obtained from the OBPG OceanColor web site. We use reprocessing R2010.0 data for both SeaWiFS and MODIS/A satellite instruments. These roughly 9 km resolution images of chlorophyll *a* (Chl *a*) are first log transformed to render them approximately normally distributed (Campbell, 1995) following the "law of proportionate effect" (Aitchison & Brown, 1966). The daily log-transformed images are subsequently spatially low-pass filtered with a 31 day centered averaged (Hamming windowed), spatially moving two-dimensional Gaussian sparse-data, low-pass filter with a half-height width set to one half the filter width (i.e., 100 km and 200 km respectively). Each daily low-pass filter is then subtracted from the unfiltered daily image to produce a daily anomaly field (all missing data are set to a default value). For comparison, Doney et al. (2003) used calendar monthly, block averaged means as the subtrahend in the formation of the daily anomalies.

2.4. Analysis

The procedures for creating and analyzing the structure functions (variograms) derived from ocean color data have both similarities to and differences from those applied in Doney et al. (2003). Here, as before, the global, daily anomaly fields (C') are divided into nonoverlapping $5^\circ \times 5^\circ$ subsets and the empirical

variograms are computed from those subsets daily. For each 5° cell, the nominal lag distance of 9 km is set by the resolution of underlying L3 SMI data, corrected for the latitude and direction analyzed. In Doney et al. (2003), the structure functions were computed from all-possible data pairs in the spatial domain for the North-South and East-West directions only. However, this approach was found to be computationally inefficient. In this paper, fully two-dimensional variograms are computed for each 5° daily anomaly field with a Fast Fourier Transform (FFT)-based variogram algorithm (Marcotte, 1996).

The daily empirical variograms are further processed by compositing them on a calendar monthly basis to form monthly averaged, data pair weighted, empirical variograms. From these monthly two-dimensional, empirical variograms the North-South and East-West directional semivariances and number of data pairs are extracted as a function of the lag distance between data pixels. A one-dimensional, spherical model structure function, equation (4), is fit to each extracted, monthly empirical variogram with a Levenberg-Marquardt nonlinear regression routine to yield estimates of the model function parameters and their uncertainties.

2.5. Chlorophyll Gradient

One property of the Chl *a* field that was not calculated by Doney et al. (2003) was an estimate of the gradient field magnitude. We start with an estimate of $\log_{10}(\mathrm{Chl}a)$ values at 9 km resolution, low-pass filtered, Hamming windowed, monthly moving mean centered on the 15th of each month. These values ($\overline{C}$ in equation (1)) are then averaged into 1° blocks and the Matlab gradient function is applied to these 1° blocks to compute a discretized form of the N-S and E-W partial derivatives:

$$\nabla\overline{C}=\left[\frac{\partial\overline{C}}{\partial x},\frac{\partial\overline{C}}{\partial y}\right]. \tag{5}$$

The resultant $\log_{10}(\mathrm{Chl}_a/\mathrm{m}^3)\,\mathrm{degree}^{-1}$ values are converted to $\log_{10}(\mathrm{Chl}_a/\mathrm{m}^3)\,\mathrm{km}^{-1}$ allowing for the curvature of the Earth. The monthly root mean square (RMS) value of the per km results are then averaged onto the same 5° grid that the geostatistical properties were calculated in either the North-South or East-West directions and displayed in our tables and figures. The gradient algorithm uses a straightforward application of the discretized form of partial derivatives.

3. Results

There are three geostatistical parameters of primary interest in this study. First, unresolved variance (c_0) represents variability present in ocean color data that cannot be resolved at that data resolution. It can be variability due to submesoscale geophysical variability, instrument or algorithmic noise, or any combination of the three. Second, resolved variance ($c_\infty - c_0$) represents an estimate of the true mesoscale variability in terms of the variance of the log-10 transformed data anomalies. An advantage of displaying variability in terms of c_0 and ($c_\infty - c_0$) for the log-10 transformed data is that they effectively normalize over the large seasonal and global range in surface chlorophyll. For some applications, however, we also want measures of the absolute variability of surface chlorophyll concentration. As described in Appendix A, both resolved ($c_\infty - c_0$) and unresolved (c_0) variance can be transformed, using equation (A5), into estimates of the corresponding arithmetic variance (and standard deviation) of the original, untransformed data. When plotted as the log of arithmetic standard deviation ($\log_{10}(s_\xi)$) against log of mean Chl *a*, a variable with a constant coefficient of variation would plot as a straight line. We use this to bracket the data with parallel contours of constant relative variability as shown in Figures 2 and 4. Third, decorrelation scale length (*d*) that represents a distance beyond which any two data points will no longer exhibit any regional correlation. The magnitude of this parameter is dependent upon the size of the low-pass filter used to generate the anomalies initially. This parameter tells us how great a distance the chlorophyll *a* concentration retains some level of statistical correlation and, in anisotropic data fields, is also direction dependent.

3.1. Interannual Variability

Figure 1 shows the time series of $\log_{10}(\mathrm{Chl}a)$, L_{tracer}, horizontal gradient magnitude, resolved standard deviation ($\sqrt{c_\infty - c_0}$), and unresolved standard deviation ($\sqrt{c_0}$) for the 5° square with its northwest corner at 45°N and 40°W along the North Atlantic Aerosol and Marine Ecosystem Study (NAAMES) cruise track (http://naames.larc.nasa.gov/). This figure displays aspects of the ocean color data record that were not explored in

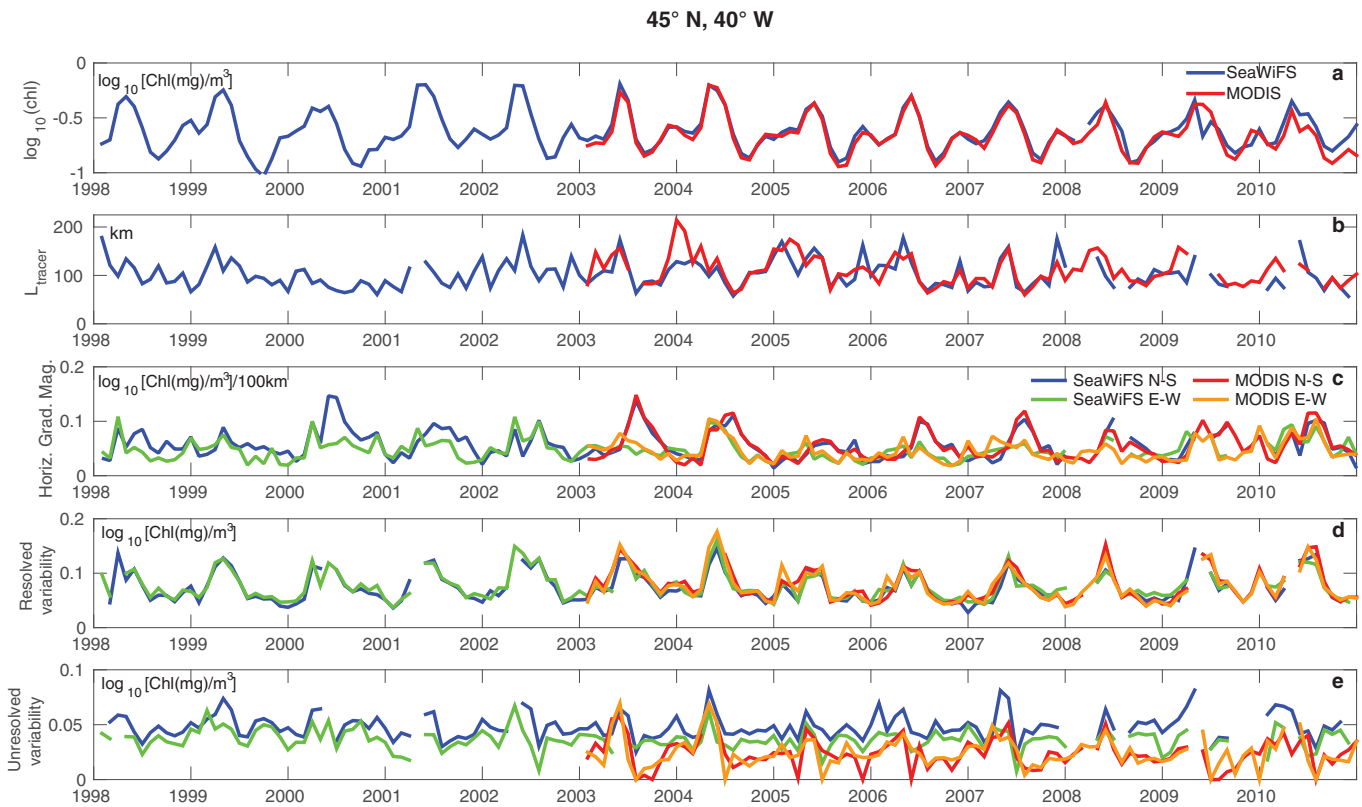

Figure 1. Time series from along the NAAMES program cruise track at 45°N and 40°W showing (a) log-transformed chlorophyll *a*, (b) the tracer-based mixing length (L_{tracer}) defined in equation (2), (c) the horizontal gradient magnitude of Figure 1a, (d) the geostatistical resolved variability of SeaWiFS and MODIS/Aqua chlorophyll *a*, and (e) the geostatistical unresolved chlorophyll *a* variability of the same satellite instruments. Geostatistical parameters (in Figures 1d and 1e) $\sqrt{c_\infty - c_0}$ and $\sqrt{c_\infty}$ are in standard deviation units of the $\log_{10}$ transformed data.

Doney et al. (2003), namely interannual variability, coherent seasonal patterns, and horizontal gradient magnitude. At this site (NAAMES), and in virtually every 5° grid cell, a seasonal cycle in all of the above properties can be clearly seen. In Figure 1a, a maximum surface chlorophyll signal appears in the northern hemisphere late spring-early summer and a lower, secondary, maximum occurs during the late fall, early northern winter. The secondary maximum is more pronounced at some locations than others depending on whether or not the site has nutrient-limited spring bloom characteristics of a mid-latitude site (Behrenfeld et al., 2005; Longhurst, 1995; Siegel et al., 2002).

Although there is agreement between the SeaWiFS and MODIS/A sensor estimates of Chl *a* from year to year, there are noticeable differences and similarities in the other properties displayed in Figure 1. The time series of the tracer-based mixing length L_{tracer} (calculated using $\sqrt{\overline{C'^2}} \approx \sqrt{c_\infty - c_0}$) has a seasonal cycle similar to the $\log_{10}(\mathrm{Chl}a)$ and appears to be in phase with the log-transformed chlorophyll *a*. The seasonal cycle of the horizontal gradient has a different magnitude depending whether one considers the north-south or the east-west direction for both SeaWiFS and MODIS/A (north-south generally being larger than east-west for both instruments). Further, the gradient seasonal cycle is out of phase with the $\log_{10}(\mathrm{Chl}_a)$ signal, with the gradient lagging the Chl *a* peak by a month or two. Estimates of resolved standard deviation ($\sqrt{c_\infty - c_0}$) from SeaWiFS and MODIS/A agree well at this location, with seasonality peaking in agreement with surface chlorophyll. This demonstrates the strength of the geostatistical technique in isolating the mesoscale geophysical signal. The unresolved variability (Figure 1e) estimates from MODIS/A are lower than SeaWiFS estimates following the aforementioned greater expected noise in SeaWiFS data (section 2.3), assuming the geophysical noise is the same for both instruments. Table 1 shows the time series average summary of the MODIS/A key variables for select time series sites from the northern and southern hemisphere including Ocean Observatory Initiative (OOI) sites (http://oceanobservatories.org/).

Table 1
MODIS/Aqua Time Series Summary for BATS, HOT, North Atlantic Aerosols and Marine Ecosystems Study (NAAMES), and Ocean Observatory Initiative (OOI) Sites

	BATS (31.7°N, 64.2°W)	HOT (22.8°N, 158°W)	NAAMES (45°N, 40°W)	Station PAPA (50°N, 145°W)	Irminger Sea (60°N, 39°W)	Argentine Basin (42°S, 42°W)	Southern Ocean (55°S, 90°W)
$\log_{10}$(Chl)[a]	−1.07	−1.23	−0.66	−0.55	−0.46	−0.39	−0.99
n	96	96	96	96	88	96	88
Seasonal amplitude	0.55	0.18	0.52	0.09	0.62	0.37	0.21
Phasing (month)	3	1	5	9	7	2	1
N-S unresolved variability[b]	0.061	0.063	0.051	0.056	0.059	0.062	0.049
n	94	96	93	92	69	95	71
Seasonal amplitude	0.03	0.03	0.06	0.05	0.03	0.04	0.04
Phasing (month)	7	5	4	1	6	11	9
E-W unresolved variability[b]	0.067	0.075	0.049	0.051	0.055	0.058	0.032
n	95	94	94	94	71	92	72
Seasonal amplitude	0.04	0.03	0.05	0.05	0.03	0.05	0.05
Phasing (month)	7	4	5	1	2	12	9
N-S resolved variability[b]	0.16	0.16	0.13	0.12	0.13	0.18	0.09
n	91	94	91	89	69	94	67
Seasonal amplitude	0.12	0.06	0.16	0.10	0.12	0.22	0.10
Phasing (month)	4	7	5	10	8	12	1
E-W resolved variability[b]	0.15	0.16	0.13	0.12	0.14	0.16	0.074
n	93	93	94	92	69	92	69
Seasonal amplitude	0.12	0.06	0.14	0.08	0.08	0.25	0.08
Phasing (month)	4	5	5	10	7	12	1
N-S range[c]	82.4	63.3	74.0	66.8	60.8	73.5	55.0
n	93	93	95	91	71	94	70
Seasonal amplitude	33.4	21.8	27.6	39.9	14.5	42.1	46.8
Phasing (month)	9	10	1	2	3	3	9
E-W range[c]	81.8	82.4	65.6	58.6	51.4	72.4	52.4
n	93	94	93	94	69	94	72
Seasonal amplitude	98.2	57.1	27.8	25.0	23.1	50.6	30.5
Phasing (month)	9	2	8	1	2	3	11
L_{tracer}[c]	113.0	206.0	110.5	94.6	99.5	136.3	91.1
n	94	95	91	87	71	94	70
Seasonal amplitude	69.3	81.6	59.0	65.5	36.1	109.1	51.3
Phasing (month)	8	6	5	10	9	12	9
Gradient magnitude[d]	0.074	0.038	0.074	0.077	0.082	0.089	0.060
n	96	96	96	95	72	96	73
Seasonal amplitude	0.082	0.014	0.064	0.063	0.054	0.110	0.066
Phasing (month)	4	9	6	1	8	2	1

[a]In units of $\log_{10}[\mathrm{Chl(mg)/m^3}]$.
[b]Unresolved and resolved variability are in units of coefficient of variation (equation (A6)).
[c]N-S and E-W ranges and L_{tracer} are in units of kilometers.
[d]Gradient magnitude is in units of $\log_{10}[\mathrm{Chl(mg)/m^3}]/100$ km.

3.2. Local Temporal Relationships

The relationships between geostatistical variability as arithmetic standard deviation (s_ξ) for the NAAMES site and mean chlorophyll *a* become apparent when the time series in Figure 1 are collapsed into scatter plots (Figure 2). The monthly climatologies (solid dots) are also shown to emphasize the seasonal behavior of the relationships versus the additional scatter introduced by interannual variability. Lines of constant coefficient of variation (CV) are plotted by combining equation (A6) with equation (A5) (making allowance for conversion between base-10 and natural logarithms). The monthly resolved variability data points (open circles) and seasonal monthly climatologies for both instruments lie in a single group. At Chl *a* concentrations below ~0.32 mg(Chl)/m^3 they group around a line of constant CV, but jump to a line of higher constant CV

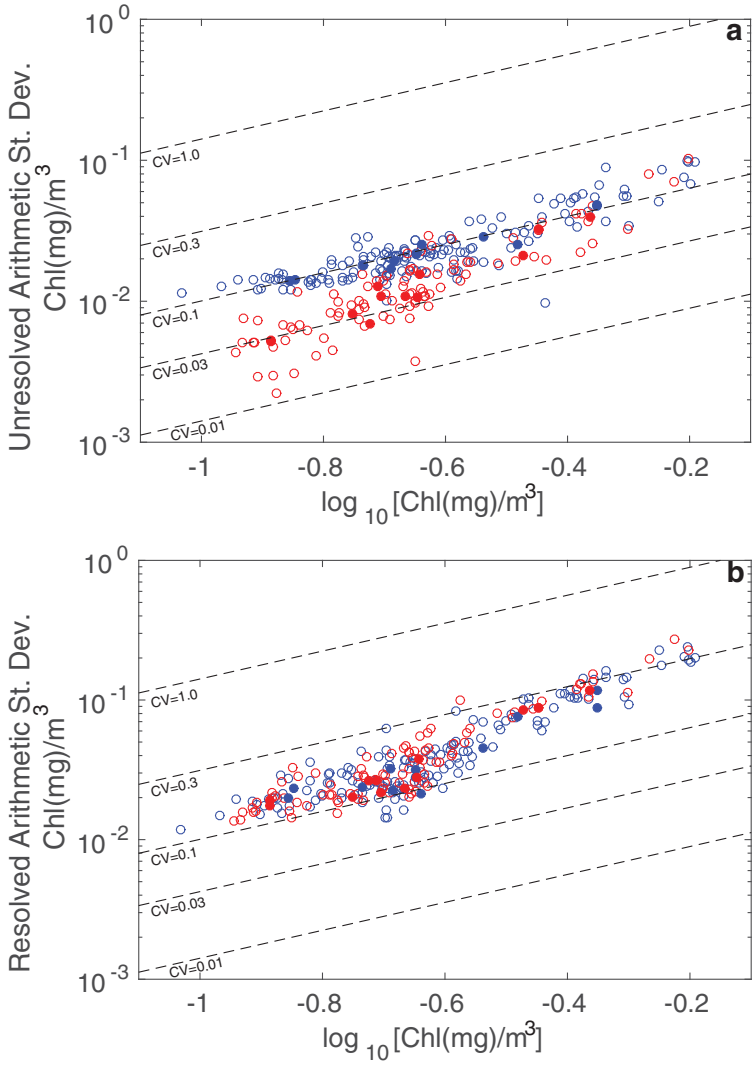

Figure 2. Log-log plots of geostatistical variability as arithmetic standard deviation (Chl(mg)/m^3) plotted against logarithm of chlorophyll *a* ($\log_{10}[\mathrm{Chl(mg)/m^3}]$) for the 5° cell time series along the NAAMES program cruise track (45°N and 40°W). On these axes, lines of constant coefficient of variation plot as straight lines (black dashed lines) and bracket the data in terms of relative variability. Open symbols for MODIS/A (red, N = 92) and SeaWiFS (blue, N = 145) are monthly values and solid symbols (N = 12) are monthly climatologies for same satellites. (a) Unresolved variability as the arithmetic standard deviation (${}^{\mathrm{unres}}s_{\xi}$) defined as the square root of equation (A5) (mg Chl/m^3) versus $\log_{10}(\mathrm{Chl}a)$. Figure 2b same as Figure 2a for resolved variability (${}^{\mathrm{res}}s_{\xi}$) versus $\log_{10}(\mathrm{Chl}a)$.

at higher Chl *a* concentrations. The unresolved variability ranges with a CV from about 3% to 10% (Figure 2a) and the resolved variability with a CV from about 10% to 30% (Figure 2b). There is a divergence in behavior between SeaWiFS and MODIS/A in Figure 2a, which we will return to in our discussion of global relationships (section 3.3).

3.3. Global Patterns

In an attempt to deconvolve the underlying causes of the seasonal and interannual variability observed in Figures 1 and 2, global spatial patterns of geostatistical parameters are examined. Figure 3 displays geographic patterns, in 5° cells, of variance weighted, long-term means of unresolved variability ($\sqrt{c_0}$), and resolved variability ($\sqrt{c_\infty - c_0}$) of the monthly time series (SeaWiFS 1998–2010 and MODIS/A 2003–2010). The patterns shown here are very similar to the patterns found in Doney et al. (2003) even though the N-S and E-W results have been averaged together in Figure 3 for each instrument. Immediately apparent are

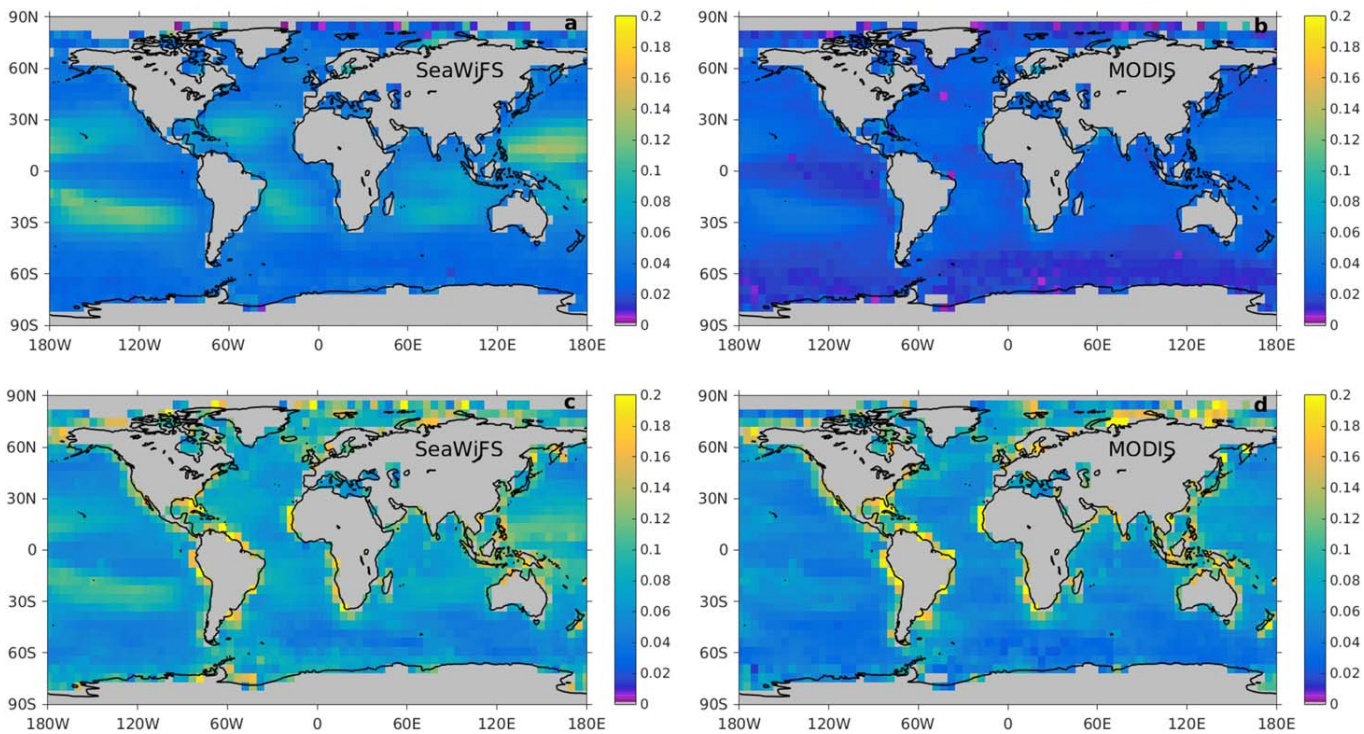

Figure 3. Global distribution maps of resolved and unresolved variability for both SeaWiFS and MODIS/Aqua. The values displayed here are the average of the north-south and east-west results (see text for further discussion). (a) The unresolved variability derived from SeaWiFS data (as $\sqrt{c_0}$). (b) The unresolved variability from MODIS/A data (as $\sqrt{c_0}$). (c) The resolved variability derived from SeaWiFS data (as $\sqrt{c_\infty - c_0}$). (d) The resolved variability from MODIS/A data (as $\sqrt{c_\infty - c_0}$). Note that the unresolved variability from MODIS/A data (b) is much lower than SeaWiFS yet the resolved variability is still slightly elevated in the oligotrophic gyres in the same pattern as SeaWiFS (c).

the much lower unresolved variability values for MODIS/A (Figure 3b) versus SeaWiFS (Figure 3a). This difference is especially noticeable in the oligotrophic gyres where the general background chlorophyll *a* concentrations are low. We interpret this difference as due, in part, to a greater number of bits recorded by the MODIS/A detector than SeaWiFS (12 versus 10 bit quantization) allowing MODIS/A to distinguish more effectively between noise and signal at low Chl *a* concentrations (see section 2.3). Regardless, the pattern of elevated unresolved variability in the center of the basin scale gyres (areas of low Chl *a*) is clearly visible for both instruments in both the unresolved (Figures 3a and 3b) and, to a lesser degree, as a bleed-over into the resolved variability (Figures 3c and 3d). This pattern implies an inverse relationship between the underlying chlorophyll concentrations and the relative level of uncertainty with respect to the unresolved variability at low-chlorophyll levels in SeaWiFS.

Perhaps nowhere else are the differences between SeaWiFS and MODIS/A and the attributes of our analysis better shown than in the comparisons displayed in Figure 4. Similar to Figure 2, dashed lines of constant CV are drawn on the plot to show relative variation, however, in Figure 4 all of the 5° cell, long-term means are plotted as points with those poleward of 60° latitude as open circles. We assume there is a true geophysical variability signal (variance) that is scale dependent spanning from submesoscale through mesoscale. Choosing a sampling resolution of 9 km (SMI grid) partitions that true variability signal into two components, resolved (>9 km or mesoscale) and unresolved ($\leq$ 9 km or submesoscale). Figure 4a shows that increase in the measurement value adds additional noise (more variance) to the unresolved component with a magnitude dependent on characteristics of the satellite sensor data product noise. As in Figure 2a, the scatter of data points for unresolved variability tends to range with a CV from ~3% to ~10% for both instruments. However, at low Chl *a* concentrations (below ~0.18 mg Chl/m^3) the behavior of the two instruments diverge. While MODIS/A relative variability of unresolved error stays at or below a CV of ~10%, SeaWiFS unresolved variability rises to a CV of ~30%. The relationship between unresolved variability in terms of arithmetic standard deviation, ${}^{\text{unres}}s_{\xi}$, and $\log_{10}(\text{Chl}a)$ (Figure 4a) shows SeaWiFS data have a larger unresolved variance at low chlorophyll. Assuming that the true geophysical submesoscale signal is the same for

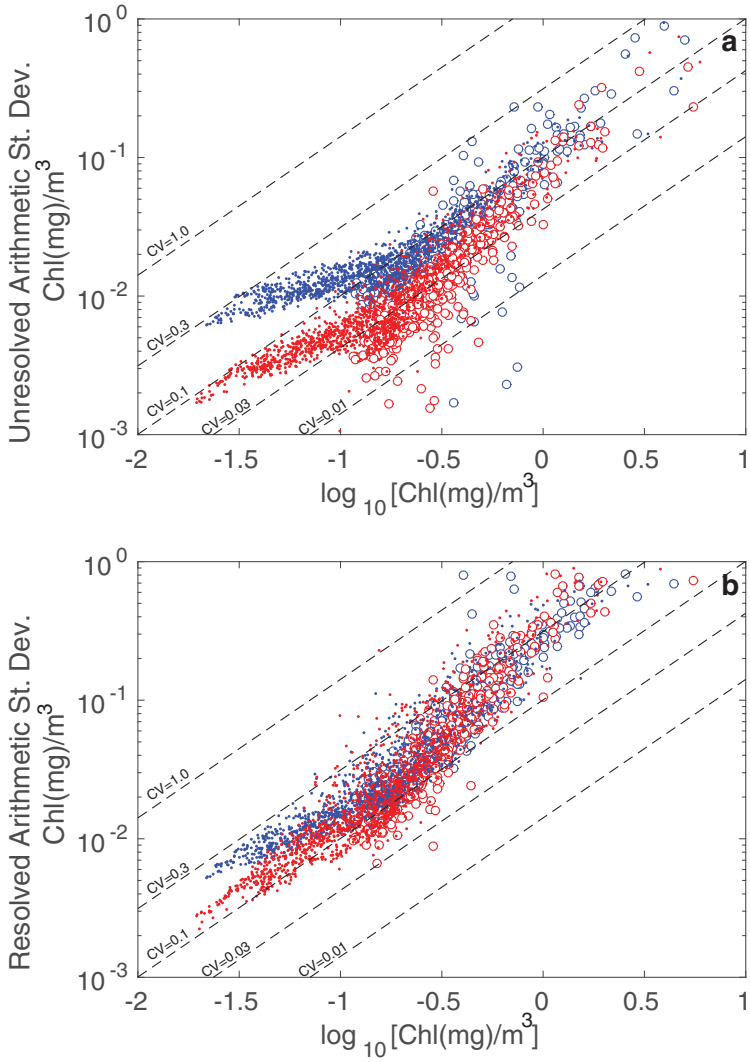

Figure 4. Log-log plots of geostatistical variability as arithmetic standard deviation plotted against logarithm of chlorophyll *a* ($\log_{10}[\mathrm{Chl(mg)/m^3}]$) for the annual mean of the global collection of 5° cells. As in Figure 2 lines of constant coefficient of variation plot as straight lines (black dashed lines) and bracket the data in terms of relative variability. All points poleward of 60° are plotted as open circles for SeaWiFS (blue) and MODIS/A (red). (a) Unresolved variability as the arithmetic standard deviation (${}^{\mathrm{unres}}s_{\xi}$) defined as the square root of equation (A5) (mg Chl/m^3) versus $\log_{10}(\mathrm{Chl}a)$. Figure 4b same as Figure 4a for resolved variability (${}^{\mathrm{res}}s_{\xi}$) versus $\log_{10}(\mathrm{Chl}a)$.

SeaWiFS and MODIS this implies a larger contribution of instrument-specific noise for the SeaWiFS data (section 2.3).

Resolved variability in terms of arithmetic standard deviation, ${}^{\mathrm{res}}s_{\xi}$, plotted versus $\log_{10}(\mathrm{Chl}a)$ is shown in Figure 4b and demonstrates a similarity between SeaWiFS and MODIS/A. Most resolved variability plots between the 10% and 30% coefficient of variation contours and only a small amount of divergence between the two instruments at low Chl *a* is apparent. Figure 4b implies that the geostatistical technique is isolating a robust measure of biophysical, mesoscale variability that is largely independent of measurement platform. Nevertheless, SeaWiFS does exhibit larger resolved variability than MODIS in the same low-chlorophyll regions as unresolved variability (compare Figure 4a with Figure 4b). In an ideal situation, both instruments should capture the same magnitude of resolved variability reflecting only the true geophysical mesoscale signal. This would hold if our geostatistical method partitioned all sources of instrument-specific noise into the unresolved component, but clearly this is not the case. The larger SeaWiFS variances in both

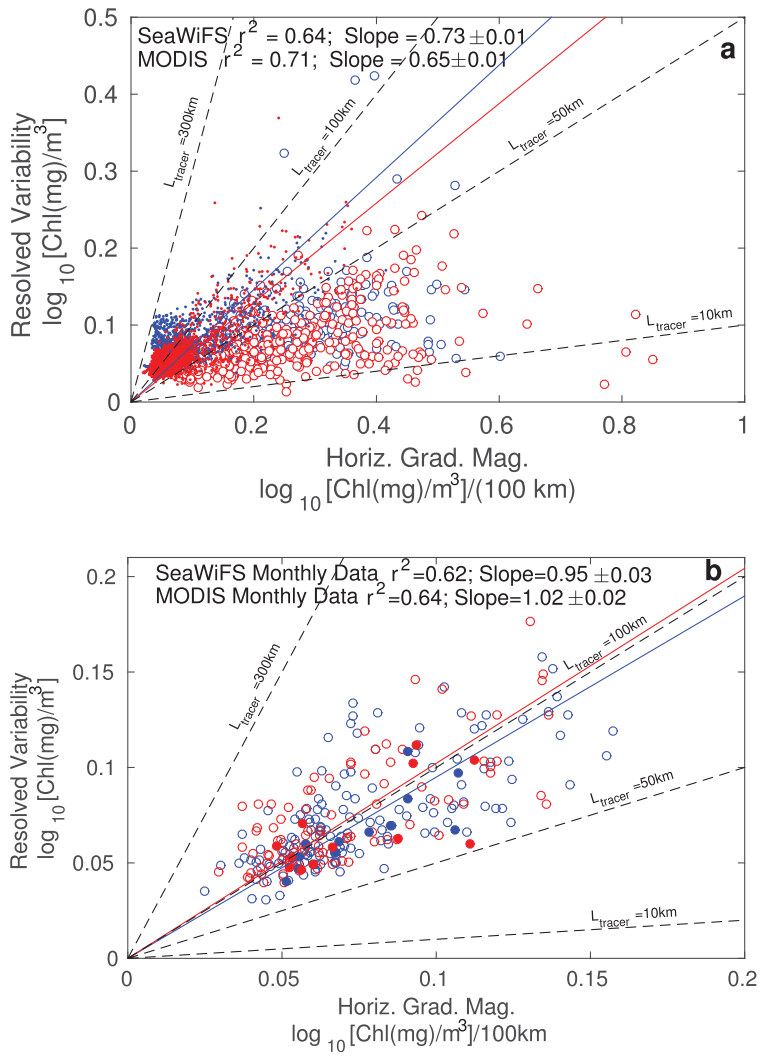

Figure 5. The relationship between chlorophyll anomaly RMS and chlorophyll gradient displayed as the square root of the resolved variability versus the mean chlorophyll gradient magnitude. Here we estimate $\sqrt{\overline{C'^2}}$ with $\sqrt{c_\infty - c_0}$ (see section 2.2) and $\partial\overline{C}/\partial x$ by equation (5). (a) The same global set of data points as in Figure 4 with two sets of linear lines, lines of constant L_{tracer} (black dashed lines) and the linear regression of SeaWiFS (blue) and MODIS/A (red) data. All points poleward of 60° have open symbols. (b) The same plot as in Figure 5a, but restricted to the NAAMES data displayed in Figure 2, open symbols are monthly and filled symbols monthly seasonal climatologies.

the resolved and unresolved components suggest that our geostatistical method is not fully partitioning the larger noise levels in SeaWiFS between the unresolved and resolved components, resulting in an apparent leakage of noise from the unresolved into the resolved component. This could reflect deficiencies in low-pass filtering, the variogram approach, or that some of the noise in SeaWiFS is spatially correlated (memory effects of optical sensors, cloud ringing, stray light artifacts, etc.) as previously mentioned in section 2.3 and Doney et al. (2003), or any combination of the three. Our results apply to the variance components of SeaWiFS and MODIS/A and do not directly address possible biases in the low-frequency time or space means. Any consideration of the reasonableness of either instruments Chl *a* measurements, in any region, is a question best decided by the investigator using the data.

Our analysis yields insight into the underlying mechanism of patchiness creation. Building from the definition of L_{tracer} in equation (2), Figure 5a displays the square root of the resolved variability $\sqrt{c_\infty - c_0}$, an

estimate for $\sqrt{\overline{C'^2}}$, for both instruments as a function of the local horizontal gradient magnitude. Figure 5a has lines of constant, operationally defined empirical mixing scale, L_{tracer} (equation (2)) plotted as dashed lines. Inspection reveals that the data poleward of 60° (open circles) plot between 10 and 50 km, while the solid points (±60°) plot between 50 and 300 km, reflecting a connection of surface chlorophyll mesoscale length scales to the Rossby radius of deformation noted earlier by Doney et al. (2003). Additionally, linear regressions of $\sqrt{c_\infty - c_0}$ (in units of $\log_{10}(\text{Chl}a)$) as a function of the local horizontal gradient magnitude (also in units of $\log_{10}(\text{Chl}a)$) were done and the r^2 and slope are reported. These regressions are both statistically significant at the 95% confidence level (p-value <0.0001). Since the x axis is in units of gradient per 100 km, 100 times the slope yields an estimate of the global, average L_{tracer} (73 km for SeaWiFS and 65 km for MODIS/A). The SeaWiFS values of L_{tracer} are likely larger because of the positive bias, noted above, in resolved variability in low-chlorophyll subtropical regions that also have relatively low horizontal gradient magnitude. Returning, briefly, to the NAAMES site, Figure 5b shows the monthly (open circles) and monthly seasonal climatologies (filled circles) on the same axes as Figure 5a. Note that at 45°N the data points all plot between ~50 and ~200 km, the linear regressions (also significant) imply mean L_{tracer} distances of 95 and 102 km for SeaWiFS and MODIS/A, respectively. Both plots demonstrate that the resolved mesoscale variability can be expressed as a function of the absolute magnitude of the local, horizontal gradient of the low-pass filtered chlorophyll field consistent with the argument that physical stirring acting on the large-scale gradient is an important factor supporting observed mesoscale variability.

4. Discussion

4.1. Error

NASA has a primary interest in quantifying the error of the measurements they make from orbit and the distribution of that error in both space and time. The sources of error in Chl *a* remote measurements are manifold, but can be grouped into one of two classes. Either the error is due to true random processes, generally referred to as noise, or to processes that scale with the size of the variable (Aitchison & Brown, 1966; Campbell, 1995). Mathematically, the actual measurement can be expressed as a sum:

$$\xi_{\text{obs}} = \xi_{\text{true}} + \eta\overline{\xi} + \epsilon \tag{6}$$

where ξ_{obs} are the actual observations (before the low-pass filter has been subtracted) and ξ_{true} are the true values. Here η represents errors that are proportional to the size of the mean ($\overline{\xi}$), in this case a spatial mean, and ϵ represents a constant background noise; both are random and independent. Then we can write for the unresolved variability component (other components are similar) the arithmetic variance as:

$$s_\xi^2 = s_\eta^2\overline{\xi}^2 + s_\epsilon^2 \tag{7}$$

For coefficient of variation defined as $\text{CV} = s_\xi/\overline{\xi}$, then

$$\text{CV} = \sqrt{s_\eta^2 + \frac{s_\epsilon^2}{\overline{\xi}^2}} \tag{8}$$

Figure 4 demonstrates different aspects of this relationship for large and small Chl *a* concentrations. We see, in Figure 4a and equation (8), as $\overline{\xi}$ becomes large ($\log_{10}(\text{Chl}) > -0.8$), both MODIS unresolved error and SeaWiFS unresolved error scale approximately linearly with Chl *a*, i.e., CV approaches a constant (s_η). This implies that unresolved error can be roughly modeled as the first RHS term of equation (7), $s_\eta\overline{\xi}$. At low Chl *a*, this is not true for SeaWiFS where coefficient of variation increases substantially as Chl *a* decreases, implying that an additional constant error term (s_ϵ) is also present, i.e., as $\overline{\xi}$ becomes small, CV increases with the background noise $s_\epsilon/\overline{\xi}$ (equation (8)).

4.2. Length Scales

There are two key length scales in this work, the operationally defined, tracer-based mixing scale (L_{tracer}) and the geostatistically derived decorrelation scale length (d). Both L_{tracer} and d lengths are data derived, in fact from different aspects of the same data set. They are different approaches but are not fully independent. The tracer-based mixing scale (L_{tracer}) is computed from estimates of resolved property anomalies and large-scale lateral gradient and is meant to represent a distance a mesoscale eddy

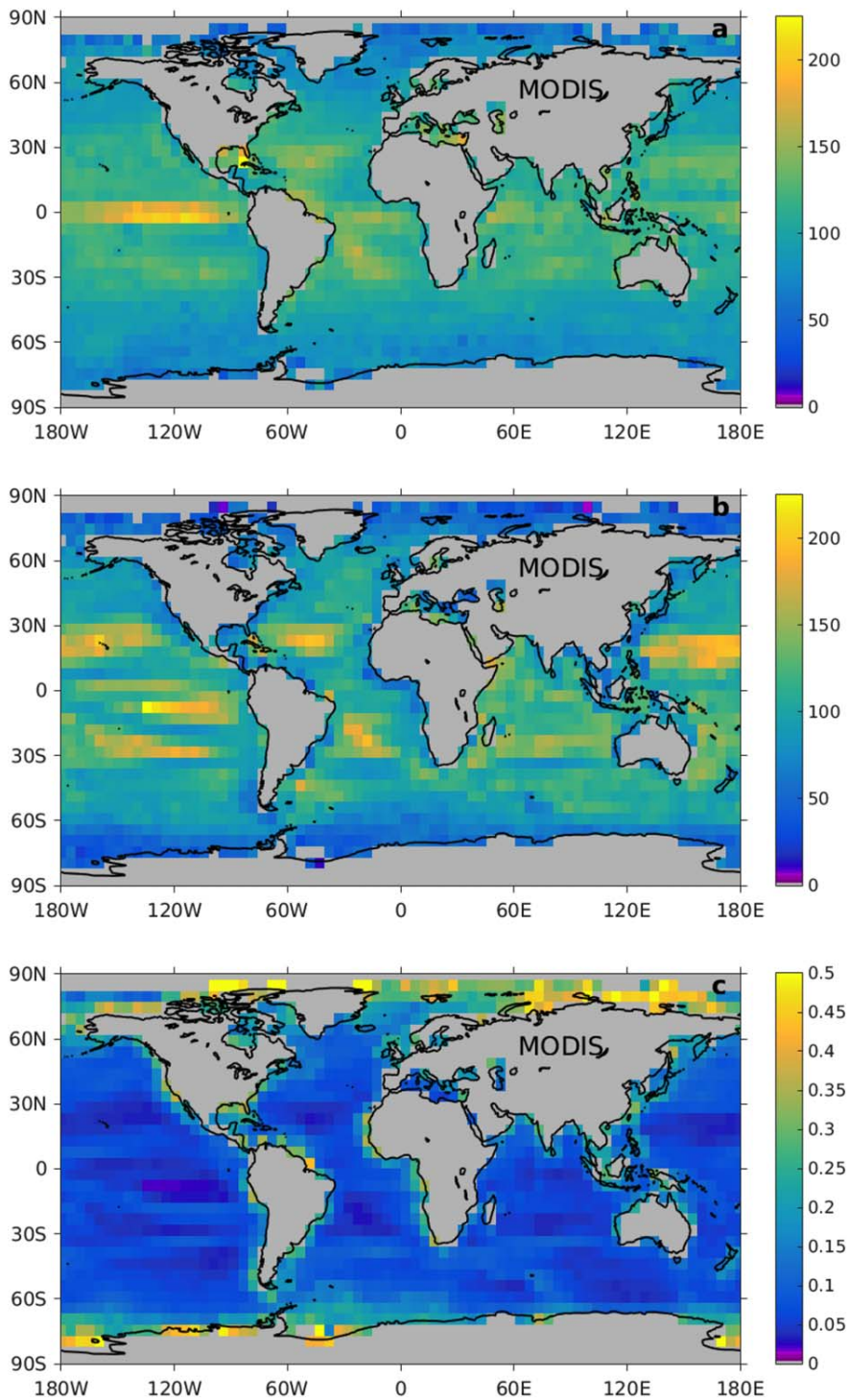

Figure 6. Maps of the length parameters derived from MODIS/A (see text for a discussion of SeaWiFS data). In each image, the annual mean value for each 5° cell is plotted for the entire time series (2003–2010). (a) The range (d) in km. (b) L_{tracer} following equation (2) in km. Here we estimate $\sqrt{\overline{C'^2}}$ with $\sqrt{\overline{c_\infty - c_0}}$ (see section 2.2) and $\partial\overline{C}/\partial x$ by equation (5). (c) The local gradient magnitude of the mean chlorophyll concentration ($\partial\overline{C}/\partial x$) in units of $\log_{10}[\text{Chl}(\text{mg})\,\text{m}^{-3}]/100$ km.

would, or could, stir a water parcel containing, in this case, chlorophyll a anomalies. The decorrelation scale length (d) represents a data-derived estimate of the actual distance any two regional mean anomalies must be separated to ensure that they are statistically independent. The interpretation of

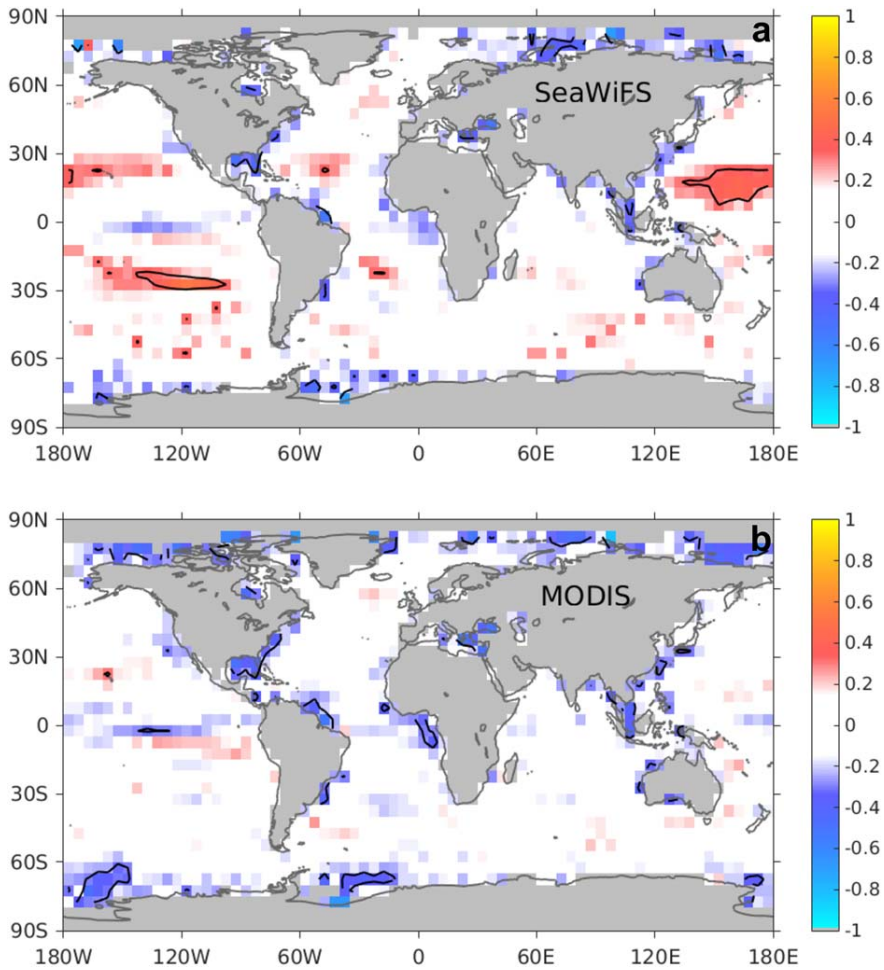

Figure 7. Ratio (R) of observed resolved variability ($\sqrt{c_\infty - c_0}$) to predicted variability $(\sqrt{\overline{C'^2}})_{stir}$ as explained in the text and equation (10). The white space, centered on zero, encompasses ± 1 standard deviation of R (SeaWiFS: ± 0.1662 and MODIS/A: ± 0.1461). A single contour at the $\pm 95\%$ confidence level encloses 5° grid cells outside 95% of all R values. (a) SeaWiFS R values, note the large positive areas that are beyond the 95% confidence limits may be due, in part, to aliasing of unresolved variability into the resolved variability estimate. (b) MODIS/A R values, note the areas of negative R in common with SeaWiFS. See text for further discussion.

differences between L_{tracer} and d has many potential challenges because of the differential impacts of physical effects and mesoscale biological variations on the two length scales. After all, L_{tracer} should, in theory, represent physical effects while d (range) should also capture mesoscale biological variations. In principle, it may be possible to separate physical stirring from biological effects, however, this will have to wait for future improvements in the methodology that are not fully resolved at this time. Even if L_{tracer} and d were exactly proportional, there is no reason the ratio should be exactly one (1). For example, choice of a different variogram model would result in a different d and hence a different proportionality. Further, this framework does not account for large regional variations in mesoscale eddy kinetic energy, the effects of which are unknown. Therefore, we caution readers to be careful not to over-interpret perceived differences between L_{tracer} and d; the following discussion should be considered as a scaling argument not an exact relationship.

With these caveats in mind, we can explore effective differences in L_{tracer} and d in terms of the implied chlorophyll anomalies caused by physical stirring. Scaling arguments suggest that physical stirring with a tracer length scale (L_{tracer}) replaced by the observed decorrelation length scale (d) acting on a large-scale gradient ($\partial\overline{C}/\partial x$) should generate tracer anomalies with a root mean square (RMS) of:

$$\left(\sqrt{\overline{C'^2}}\right)_{\text{stir}} = d\frac{\partial \overline{C}}{\partial x} \tag{9}$$

where we can consider $\left(\sqrt{\overline{C'^2}}\right)_{\text{stir}}$ an estimate of the anomaly variability due to physical stirring. The magnitude of these tracer anomalies can be compared against the observed resolved variability $\sqrt{c_\infty - c_0}$ derived from geostatistical methods. The two variability estimates can be combined in a ratio (R):

$$R = \frac{\left(\sqrt{c_\infty - c_0}\right)_{\text{obs}} - \left(\sqrt{\overline{C'^2}}\right)_{\text{stir}}}{\left(\sqrt{c_\infty - c_0}\right)_{\text{obs}} + \left(\sqrt{\overline{C'^2}}\right)_{\text{stir}}} \tag{10}$$

and are displayed in Figure 7 for both SeaWiFS and MODIS/A. These maps have a color table chosen so that values within plus or minus one standard deviation of zero are colored white. Further, 5° grid cells with values outside the ±95% confidence level around zero are enclosed with a black contour line.

The ratio R presents the spatial distribution of two different estimates of physical stirring derived either from structure function parameters or mean Chl *a* gradients. In addition to stirring, the resolved variability ($\sqrt{c_\infty - c_0}$) includes contributions from mesoscale biological sources and sinks, as well as leakage (aliasing) of unresolved variability into the resolved variability component, and geophysical noise. If the ratio equals zero, within statistical uncertainties, then it is likely that stirring is a major factor in the observed resolved variability. Positive values ($R \gg 0$) indicate areas, such as in the oligotrophic ocean, where it is possible that there is aliasing of unresolved variability into the resolved variability estimate (Figures 3c and 3d). It could be noise; it could also be real mesoscale or submesoscale variability. Positive values are also seen in the temperate North Atlantic, Southern Ocean, and Equatorial Pacific where there may be real biological factors enhancing the observed variability. Negative values ($R \ll 0$), such as in some of the western boundary currents (North and South America and Asia), indicate regions where the gradients (Figure 6c) are quite large leading to large estimates of physical stirring. It may be that the simple scaling arguments do not hold in these complex coastal regions where transport is tied to topography and is not isotropic because of the boundary (Gruber et al., 2006, 2011).

To within the uncertainties noted above, the following can be stated about the mixing length scale observed in our results. Length scales computed from the two different approaches give roughly comparable results with values similar to that expected for mesoscale features (Figures 6a and 6b). Similar to the findings of Doney et al. (2003) both length scales decrease from the tropics to the poles, comparable to the Rossby radius of deformation. Regional differences occur (Figure 7) but it is unclear from where these biases arise; are they due to methodological issues or real geophysical signals?

5. Summary

Application of geostatistical analysis techniques allows the variability observed in ocean color imagery to be partitioned into resolved and unresolved variability and gives an estimate of how variable pixels are from each other as a function of the distance between them. This variability can be expressed in terms of a coefficient of variation that allows relative magnitudes of variability to be assessed. Also, the scale of decorrelation clearly relates to the underlying mean chlorophyll *a* gradient and can be expressed in terms of an idealized mesoscale tracer mixing length.

Patchiness is, by its very nature, a spatial phenomenon. Our estimates of resolved and unresolved variability as well as decorrelation scale lengths are spatial estimates resolved into monthly time series of individual 5° grid cells (or more precisely the ocean beneath that cell) of geostatistical parameters. In this paper, we have demonstrated that the mesoscale-isolated geostatistical parameters exhibit a temporal (seasonal) behavior and vary geographically. We have found that SeaWiFS and MODIS/A resolve approximately equal variance. SeaWiFS data have an unresolved variability component that is larger than MODIS/A at low Chl *a* concentrations, most likely due to sensor noise characteristics. SeaWiFS and MODIS/A have resolved variability in the range of 10–30% globally, when expressed as coefficient of variation. Globally SeaWiFS and MODIS/A imply a L_{tracer} between 50 and 300 km for sites ±60° latitude and smaller poleward. When comparing satellite derived ocean color data to model output and in developing data assimilation methods, the scale of the comparison (basin, mesoscale, submesoscale, etc.), both magnitude and

nature of data source variability (resolved versus unresolved), and the correlation scale lengths should be carefully considered.

The reprocessing version of the ocean color Chl *a* data used in this study (R2010.0) is an older version of the data than is currently available (for both SeaWiFS and MODIS/A). Improvements in the current reprocessing (R2014.0) might significantly reduce the difference in unresolved variability between the two sensors, a hypothesis that should be explored in future work using the geostatistical techniques outlined here.

On a closing note, it is interesting that no time series of any geostatistical or any other of our derived parameters show any sign of a temporal trend for either instrument. Perhaps this is not surprising because the detection of any trend in ocean color based data is obscured by the shortness of the time series and the large natural interannual and decadal variability of the primary variable (Chl *a*). Both Henson et al. (2010) and Yoder et al. (2010) found that the SeaWiFS time series record (the longest as of 2010) was insufficient in length to differentiate a global trend from natural variability found in the time series. And although the MODIS/A time series is, as of this writing, 6 years longer, it is only 1 year longer than the SeaWiFS time series analyzed for this study. Henson et al. (2010), in particular, estimate a time series longer than 40 years will be necessary to accomplish this task. For approximately 10^{+} years, the patterns have been rather stable and apparently unchanging, displayed either as time series or as global maps. If we were to have similar information from a time series twice as long (Callander & Mitchell, 1996; Clark et al., 2016; Santer et al., 1995), we might be able to make inferences about how these spatial patterns change or do not change with time. One thing that we can do is monitor the spatial distribution of chlorophyll in the current ocean into the future and observe any changes that may indicate a response to changing forcing factors (Boyd et al., 2015).

Appendix

A1. Log-Normal Variables

There are a number of references where one can read about log-normal distributions (Aitchison & Brown, 1966; Baker & Gibson, 1987; Campbell, 1995; Limpert et al., 2001). In this short appendix, we hope to provide a discussion of log-normally distributed variables from the data analyst point-of-view.

Consider data, ξ, that you know from analysis or previous work to be log-normally distributed (for example, chlorophyll *a* concentrations). It can be transformed into a normal (or near-normal) distributed variable with $C=\ln(\xi)$. From here on we will be working in a natural logarithm base system, but the following is true in any base (to within a base-conversion constant, e.g., for base-10, 2.3026). The arithmetic mean and variance of C are calculated in a straightforward manner

$$\overline{C}=\frac{1}{N}\sum_i C_i \tag{A1}$$

$$\sigma^2=\frac{\sum_i (C_i-\overline{C})^2}{N-1} \tag{A2}$$

where $\overline{C}$ is the arithmetic mean of the log-transformed data (C), σ^2 is its variance and N represents the number of data points. From here the simple inverse transformation:

$$\overline{\xi}_G=\exp(\overline{C}) \tag{A3}$$

yields the *geometric* mean of ξ, equal to the median of ξ, for log-normally distributed data. We use $\exp(a)=e^a$ to make clear the base and the exponent. The arithmetic mean ($\overline{\xi}$) and variance (s_ξ^2) of ξ are given by:

$$\overline{\xi}=\exp(\overline{C}+\sigma^2/2) \tag{A4}$$

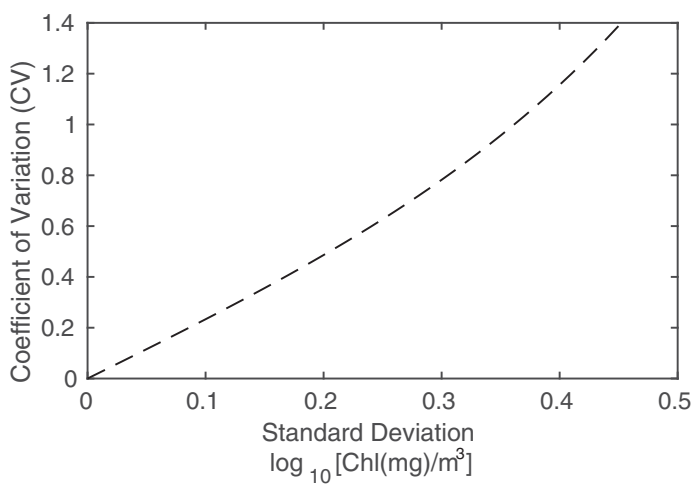

Figure A1. Coefficient of variation (unitless) as a function of the arithmetic standard deviation of the log base-10 transformed data ($\log_{10}(\mathrm{Chl(mg)m^{-3}})$) as per equations (A5) and (A6), assuming original data ξ is log-normally distributed.

$$s_{\xi}^{2}=\exp\left(2\overline{C}+\sigma^{2}\right)\left(\exp\left(\sigma^{2}\right)-1\right) \tag{A5}$$

and represent the maximum likelihood estimators (MLE) of $\overline{\xi}$ and s_{ξ}^{2}. Not to put too fine a point on it, $\overline{\xi}_{G} < \overline{\xi}$ and the coefficient of variation (CV) for ξ is given by:

$$\mathrm{CV}=s_{\xi}/\overline{\xi}=\sqrt{\exp\left(\sigma^{2}\right)-1} \tag{A6}$$

being equivalent to the standard deviation divided by the mean. The coefficient of variation of a log-normal variable (equation (A6)) is not dependent upon the mean value of ξ because it is, by definition, effectively divided by (normalized) its mean (Glover et al., 2011). Combining equation (A6) with equation (A5) yields an expression of CV in terms of $\overline{C}$ and s_{ξ}. Using this relationship, lines of constant CV are plotted on Figures 2a and 2b and 4a and 4b. A plot of CV as a function of the standard deviation of the log, base-10, transformed data (σ) is shown in Figure A1.

Acknowledgments
This work is the product of many years research generously supported by NASA's Ocean Biology and Biogeochemistry research program under grants NNG05GG30G, NNG05GR34G, NNX14AM36G, NNX15AH13G, and NNX15AE65G to D.M.G. and S.C.D. The authors are grateful to one anonymous reviewer, Carlos E. Del Castillo, and Bryan Franz for improving the manuscript. D.M.G. wishes to especially acknowledge receipt of the FFT-based variogram algorithm from Denis Marcotte, Polytechnique Montréal. All data are available from https://oceancolor.gsfc.nasa.gov/NASA's OceanColor web site supported by the Ocean Biology Processing Group (OBPG) at NASA's Goddard Space Flight Center.

References

Abraham, E. (1998). The generation of plankton patchiness by turbulent stirring. *Nature*, *391*, 577–580.
Aitchison, J., & Brown, J. (1966). *The lognormal distribution with special reference to its uses in economics*. New York, NY: Cambridge University Press.
Baker, M. A., & Gibson, C. H. (1987). Sampling turbulence in the stratified ocean: Statistical consequences of strong intermittency. *Journal of Physical Oceanography*, *17*(10), 1817–1836.
Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*, *19*, GB1006. https://doi.org/10.1029/2004GB002299
Bennett, A. F., & Denman, K. L. (1985). Phytoplankton patchiness: Inferences from particle statistics. *Journal of Marine Research*, *43*(2), 307–335.
Boyd, P. W., Lennartz, S. T., Glover, D. M., & Doney, S. C. (2015). Biological ramifications of climate-change mediated oceanic multi-stressors. *Nature Climate Change*, *5*, 71–79. https://doi.org/10.1038/NCLIMATE2441
Bretherton, F. P., Davis, R. E., & Fandry, C. (1976). A technique for objective analysis and design of oceanographic experiments applied to mode-73. *Deep Sea Research and Oceanographic Abstract*, *23*, 559–582.
Callander, B., & Mitchell, J. (1996). Can we detect climate change? *Physics World*, *9*(7), 37–41.
Campbell, J. W. (1995). The lognormal distribution as a model for bio-optical variability in the sea. *Journal of Geophysical Research*, *100*(C7), 13237–13254.
Clark, I. (1979). *Practical geostatistics* (141 p.). New York, NY: Elsevier Applied Science.
Clark, P. U., Shakun, J. D., Marcott, S. A., Mix, A. C., Eby, M., Kulp, S.,. . . Plattner, G.-K. (2016). Consequences of twenty-first-century policy for multi-millennial climate and sea-level change. *Nature Climate Change*, *6*(4), 360–369. https://doi.org/10.1038/NCLIMATE2923
Cressie, N. (1993). *Statistics for spatial data* (928 p.). New York, NY: Wiley.
Denman, K. L., & Abbott, M. R. (1988). Time evolution of surface chlorophyll patterns from cross-spectrum analysis of satellite color images. *Journal of Geophysical Research*, *93*(C6), 6789–6798.
Denman, K. L., & Abbott, M. R. (1994). Time scales of pattern evolution from cross-spectrum analysis of advanced very high resolution radiometer and coastal zone color scanner imagery. *Journal of Geophysical Research*, *99*(C4), 7433–7442.
Denman, K. L., & Freeland, H. J. (1985). Correlation scales, objective mapping and a statistical test of geostrophy over the continental shelf. *Journal of Marine Research*, *43*(3), 517–539.
Denman, K. L., Okubo, A., & Platt, T. (1977). The chlorophyll fluctuation spectrum in the sea. *Limnology and Oceanography*, *22*(6), 1033–1038.
Denman, K. L., & Platt, T. (1976). The variance spectrum of phytoplankton in a turbulent ocean. *Journal of Marine Research*, *34*, 593–601.
Doney, S. C., Glover, D. M., McCue, S. J., & Fuentes, M. (2003). Mesoscale variability of sea-viewing wide field-of-view sensor (SeaWIFS) satellite ocean color: Global patterns and spatial scales. *Journal of Geophysical Research*, *108*(C2), 3024. https://doi.org/10.1029/2001JC000843
Franks, P. J. (2005). Plankton patchiness, turbulent transport and spatial spectra. *Marine Ecology Progress Series*, *294*, 295–309.
Garrett, C. (1989). A mixing length interpretation of fluctuations in passive scalar concentration in homogeneous turbulence. *Journal of Geophysical Research*, *94*(C7), 9710–9712.
Garrett, C. (2006). Turbulent dispersion in the ocean. *Progress in Oceanography*, *70*, 113–125. https://doi.org/10.1016/j.pocean.2005.07.005
Glover, D., Jenkins, W., & Doney, S. (2011). *Modeling methods for marine science* (571 p.). New York, NY: Cambridge University Press.
Gower, J., Denman, K., & Holyer, R. (1980). Phytoplankton patchiness indicates the fluctuation spectrum of mesoscale oceanic structure. *Nature*, *288*, 157–159.
Gruber, N., Frenzel, H., Doney, S. C., Marchesiello, P., McWilliams, J. C., Moisan, J. R.,. . . Stolzenbach, K. D. (2006). Eddy-resolving simulation of plankton ecosystem dynamics in the California current system. *Deep Sea Research, Part I*, *53*, 1483–1516. https://doi.org/10.1016/j.dsr.2006.06.005
Gruber, N., Lachkar, Z., Frenzel, H., Marchesiello, P., Műnnich, M., McWilliams, J. C.,. . . Plattner, G.-K. (2011). Eddy-induced reduction of biological production in eastern boundary upwelling systems. *Nature Geoscience*, *4*, 787–792. https://doi.org/10.1038/ngeo1273
Henson, S., Sarmiento, J., Dunne, J., Bopp, L., Lima, I., Doney, S.,. . . Beaulieu, C. (2010). Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity. *Biogeosciences*, *7*, 621–640. https://doi.org/10.5194/bg-7-621-2010
Hu, C., Carder, K., & Muller-Karger, F. (2000). How precise are SeaWIFS ocean color estimates? Implications of digitization-noise errors. *Remote Sensing of Environment*, *76*, 239–249. https://doi.org/10.1016/S0034-4257(00)00206-6
Hu, C., Feng, L., Lee, Z., Davis, C., Mannino, A., McClain, C., & Franz, B. (2012). Dynamic range and sensitivity requirements of satellite ocean color sensors: Learning from the past. *Applied Optics*, *51*(25), 6045–6062. https://doi.org/10.1364/AO.51.006045
Journel, A., & Huijbregts, C. (1978). *Mining geostatistics* (600 p.). San Diego, CA: Academic Press.
Klocker, A., & Abernathey, R. (2014). Global patterns of mesoscale eddy properties and diffusivities. *Journal of Physical Oceanography*, *44*, 1030–1046. https://doi.org/10.1175/JPO-D-13-0159.1

Lévy, M. (2003). Mesoscale variability of phytoplankton and of new production: Impact of the large-scale nutrient distribution. *Journal of Geophysical Research*, *108*(C11), 3358. https://doi.org/10.1029/2002JC001577

Lévy, M., & Martin, A. P. (2013). The influence of mesoscale and submesoscale heterogeneity on ocean biogeochemical reactions. *Global Biogeochemical Cycles*, *27*, 1139–1150. https://doi.org/10.1002/2012GB004518

Limpert, E., Stahel, W., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *BioScience*, *51*(5), 341–352.

Longhurst, A. R. (1995). Seasonal cycles of pelagic production and consumption. *Progress in Oceanography*, *36*, 77–167.

Mackas, D. L., Denman, K. L., & Abbott, M. R. (1985). Plankton patchiness: Biology in the physical vernacular. *Bulletin of Marine Science*, *37*(2), 652–674.

Marcotte, D. (1996). Fast variogram computation with FFT. *Computers and Geosciences*, *22*(10), 1175–1186.

Martin, A. (2003). Phytoplankton patchiness: The role of lateral stirring and mixing. *Progress in Oceanography*, *57*, 125–174. https://doi.org/10.1016/S0079-6611(03)00085-5

McClain, C. R., Feldman, G. C., & Hooker, S. B. (2004). An overview of the SeaWIFS project and strategies for producing a climate research quality global ocean bio-optical time series. *Deep Sea Research, Part II*, *51*, 5–42. https://doi.org/10.1016/j.dsr2.2003.11.001

NASA (2014a). *SeaWIFS ocean color data*. Greenbelt, MD: NASA OB.DAAC. Retrieved from https://doi.org/10.5067/ORBVIEW-2/SEAWIFS_OC.2014.0, accessed 27 September 2010

NASA (2014b). *MODIS/Aqua ocean color data*. Greenbelt, MD: NASA OB.DAAC. Retrieved from https://doi.org/10.5067/AQUA/MODIS_OC.2014.0, accessed 27 September 2010

O'Reilly, J., Mueller, J. L., Mitchell, B. G., Kahru, M., Chavez, F. P., Strutton, P.,. . . Culver, M. (2000). *SeaWIFS postlaunch calibration and validation analyses, Part 3* (NASA Tech. Memo. 2000–206892, Vol. 11, 49 p.). Greenbelt, MD: NASA Goddard Space Flight Center.

Platt, T. (1972). Local phytoplankton abundance and turbulence. *Deep Sea Research and Oceanographic Abstract*, *19*, 183–187.

Powell, T. M., & Okubo, A. (1994). Turbulence, diffusion and patchiness in the sea. *Philosophical Transactions of the Royal Society London B*, *343*, 11–18. https://doi.org/10.1098/rstb.1994.0002

Santer, B. D., Taylor, K. E., Wigley, T. M. L., Penner, J. E., Jones, P. D., & Cubasch, U. (1995). Towards the detection and attribution of an anthropogenic effect on climate. *Climate Dynamics*, *12*, 77–100.

Shi, Y., Zhou, X., Yang, X., Shi, L., & Ma, S. (2015). Merging satellite ocean color data with Bayesian maximum entropy method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(7), 3294–3304.

Siegel, D., Doney, S., & Yoder, J. (2002). The North Atlantic spring phytoplankton bloom and Sverdrup's critical depth hypothesis. *Science*, *296*, 730–733. https://doi.org/10.1126/science.1069174

Smith, R. C., Zhang, X., & Michaelsen, J. (1988). Variability of pigment biomass in the California current system as determined by satellite imagery 1. Spatial variability. *Journal of Geophysical Research*, *93*(D9), 10863–10882.

Steele, J. H. (1974). *The structure of marine ecosystems* (128 p.). Cambridge, MA: Harvard University Press.

Steele, J. H. (Ed.) (1978). *Spatial pattern in plankton communities* (470 p.). New York, NY: Plenum Press.

Tennekes, H., & Lumley, J. L. (1972). *A first course in turbulence* (300 p.). Cambridge, MA: MIT Press.

Uz, B. M., & Yoder, J. A. (2004). High frequency and mesoscale variability in SeaWIFS chlorophyll imagery and its relation to other remotely sensed oceanographic variables. *Deep Sea Research, Part II*, *51*, 1001–1017. https://doi.org/10.1016/j.dsr2.2004.03.003

van Gennip, S., Martin, A. P., Srokosz, M. A., Allen, J. T., Pidcock, R., Painter, S. C., & Stinchcombe, M. C. (2016). Plankton patchiness investigated using simultaneous nitrate and chlorophyll observations. *Journal of Geophysical Research: Oceans*, *121*, 4149–4156. https://doi.org/10.1002/2016JC011789

Yaglom, A. (1957). Some classes of random fields in n-dimensional space, related to stationary random processes. *Theory of Probability and its Applications*, *2*, (3), 273–320 (Translated by R.A. Silverman).

Yoder, J. A., Kennelly, M. A., Doney, S. C., & Lima, I. D. (2010). Are trends in SeaWIFS chlorophyll time-series unusual relative to historic variability. *Acta Oceanologica Sinica*, *29*(2), 1–4. https://doi.org/10.1007/s13131-010-0016-o

Yoder, J. A., McClain, C. R., Blanton, J. O., & Oey, L.-Y. (1987). Spatial scales in CZCS-chlorophyll imagery of the southeastern U.S. Continental Shelf. *Limnology and Oceanography*, *32*(4), 929–941.