Extreme value distributions describe interannual variability in the North Atlantic spring bloom

Gregory Britten[1,1]

[1]Massachusetts Institute of Technology

November 30, 2022

Abstract

The North Atlantic phytoplankton bloom depends on a confluence of environmental factors that drive transient periods of exponential phytoplankton growth and interannual variability in bloom magnitude. I analyze interannual bloom variability in the North Atlantic via extreme value theory where the Generalized Extreme Value Distribution (GEVD) is fitted spatially to annual maxima of satellite-measured surface chlorophyll. I find excellent agreement between the observed distribution of interannual bloom maxima and those predicted from the GEVD. The spatial distribution of fitted GEVD parameters closely follows basin bathymetry where the largest extremes and heaviest distribution tails are found on the continental shelves and slopes. Trend analyses suggest weak evidence for changes in GEVD parameters, despite regional trends in mean chlorophyll levels and sea surface temperature. These results provide a framework to quantify interannual bloom variability and call for further work examining how extreme blooms propagate through food webs and contribute to carbon export.

1 # Extreme value distributions describe interannual variability in the
2 # seasonal North Atlantic phytoplankton bloom
3
4 Gregory L. Britten
5 Program in Atmospheres, Oceans, and Climate
6 Massachusetts Institute of Technology, Cambridge, MA, USA
7 gregleebritten@gmail.com
8
9 **Scientific Significance Statement**: The North Atlantic phytoplankton bloom helps fuel the
10 marine food web, impacts fisheries recruitment, contributes to carbon export, and is predicted to
11 change with climate warming. Understanding interannual bloom variability is thus of central
12 oceanographic importance. This study provides a framework for quantifying interannual bloom
13 variability via statistical extreme value theory. I characterize the spatial distribution of extreme
14 value parameters using satellite chlorophyll observations and test whether the distribution of
15 chlorophyll extremes has changed over time in relation to trends in background chlorophyll
16 levels and sea surface temperature.
17
18 ## Abstract
19
20 The North Atlantic phytoplankton bloom depends on a confluence of environmental factors that
21 drive transient periods of exponential phytoplankton growth and interannual variability in bloom
22 magnitude. I analyze interannual bloom variability in the North Atlantic via extreme value theory
23 where the Generalized Extreme Value Distribution (GEVD) is fitted spatially to annual maxima
24 of satellite-measured surface chlorophyll. I find excellent agreement between the observed
25 distribution of interannual bloom maxima and those predicted from the GEVD. The spatial
26 distribution of fitted GEVD parameters closely follows basin bathymetry where the largest
27 extremes and heaviest distribution tails are found on the continental shelves and slopes. Trend
28 analyses suggest weak evidence for changes in GEVD parameters, despite regional trends in
29 mean chlorophyll levels and sea surface temperature. These results provide a framework to
30 quantify interannual bloom variability and call for further work examining how extreme blooms
31 propagate through food webs and contribute to carbon export.
32
33 ## Introduction
34
35 Phytoplankton form the base of the marine food web (Falkowski et al. 2003) and play a major
36 role in the global carbon cycle (Falkowski et al. 1998; Ito and Follows 2005). At mid and high-
37 latitudes, the seasonality of phytoplankton is characterized by a large seasonal bloom where
38 phytoplankton experience a transient period of exponential growth and biomass concentrations
39 reach an annual maximum (Behrenfeld and Boss 2014). The bloom is important for sustaining
40 higher trophic levels, including species whose annual migrations are timed to coincide with the
41 bloom (Visser et al. 2011). Blooms also contribute to carbon export and are often associated with
42 large export pulses to depth (Briggs et al. 2011).
43
44 Several biophysical mechanisms are thought to control bloom initiation and development
45 (Behrenfeld and Boss 2014, 2018). For example, the critical depth hypothesis predicts that bloom
46 initiation can only begin when mixed layer light levels exceed the threshold of respiration rates,

47 allowing phytoplankton biomass to increase if division rates exceed other losses (Behrenfeld and
48 Boss 2014, 2018). The critical turbulence hypothesis assumes the same threshold as the critical
49 depth hypothesis but focuses on division rates in the actively mixing turbulent layer (Huisman et
50 al. 1999; Taylor and Ferrari 2011). The dilution-recovery hypothesis suggests a more general
51 balance between phytoplankton division and loss rates (often driven by grazing) allowing blooms
52 to develop when division rates are declining so long as loss rates decline below division rates; for
53 example, with dilution via mixed layer deepening (Evans et al. 1985; Behrenfeld 2010). Across
54 proposed mechanisms, the population-level phenomena of blooms arise due to transient positive
55 imbalances between division and loss rates (driven by variations in division, loss, or both) such
56 that the net exponential growth rate $r(t)$ is positive, i.e.

$$\frac{dc}{dt} = (\mu(t) - l(t))c = r(t)c,$$

60 with $\mu(t) - l(t) = r(t) > 0$ prior to the bloom, where $\mu(t)$ is the division rate and $l(t)$ is the
61 total loss rate. Days to weeks post-bloom, losses increase to match and exceed division, most
62 often because of increased grazer abundance supported by the elevated phytoplankton biomass
63 (Behrenfeld and Boss 2014, 2018). The magnitude and duration of the transient exponential
64 growth period determines the magnitude of the bloom biomass maximum. We hereafter refer to a
65 'bloom' as the period of positive net exponential growth and 'bloom magnitude' as the
66 concentration maximum achieved over this period (see Behrenfeld and Boss 2018 for a
67 discussion of bloom definitions).

69 The North Atlantic basin exhibits one of the largest and well-studied seasonal blooms across the
70 global ocean. Deep winter mixing and rapid re-stratification creates ideal conditions for
71 exponential phytoplankton growth, with bloom initiation following a northward progression as
72 re-stratification occurs earliest at low latitude (Dutkiewicz et al. 2001; Siegel et al. 2002).
73 Meteorological variability then modulates the precise timing and magnitude of the bloom
74 according to the impacts on local mixing dynamics (Dutkiewicz et al. 2001; Follows and
75 Dutkiewicz 2002). Importantly, variability in bloom timing and magnitude has been linked to the
76 magnitude of carbon export (Briggs et al. 2011) and to fisheries productivity via the 'match-
77 mismatch hypothesis' (Platt et al. 2003). Climate change is expected to alter North Atlantic
78 bloom dynamics via a range of factors, including changes in seasonal mixing depths, nutrient
79 fluxes, and the metabolic impacts of warmer temperatures (Sommer and Lengfellner 2008).
80 Quantifying the dynamics of the North Atlantic spring bloom is thus of central importance for
81 understanding the relevant oceanographic and ecological processes and will aid in tracking the
82 associated impacts of climate change.

84 Viewing interannual bloom magnitudes as extreme values in observed phytoplankton time series
85 brings to bear the statistical theory of extreme values. Under the Fisher-Tippet-Gnedenko
86 theorem, the maximum of a sequences of random variables converges in distribution to the
87 Generalized Extreme Value Distribution (GEVD), itself a generalization of the Gumbel, Frechet,
88 and Weibull distributions (Coles 2001). The theorem yields a three-parameter probability density
89 function describing the limiting distribution of maximum values generated from samples of a
90 stochastic process, analogous to the central limit theorem for the mean of a distribution (Coles
91 2001; described below). Like the lognormal or chi-square distribution, the GEVD can exhibit a

92   strong right skew; however, the GEVD has the advantage of being derived specifically for
93   stochastic maxima. While the GEVD is increasingly applied to geophysical and climate studies
94   (Easterling et al. 2000; Katz 2010; Aghakouchak et al. 2020) there have been fewer applications
95   to biological time series (Batt et al. 2017; but see interesting exceptions, e.g. Gaines and Denny
96   1993; Benedetti-Cecchi et al. 2015; Butitta et al. 2017). Consistent with the exponential nature of
97   biological growth, Batt et al. (2017) found that biological time series exhibit consistently heavier
98   tails in their extreme value distributions relative to chemical and geophysical time series. These
99   findings suggest North Atlantic bloom magnitudes as an ideal target of extreme value analysis,
100 due to its annually repeating cycle of transient and variable exponential growth.
101
102 Here I analyzed the North Atlantic satellite chlorophyll record to quantify seasonal
103 phytoplankton bloom variability via extreme value analysis. I estimated GEVD parameters at a
104 ¼° latitude-longitude scale. I mapped the fitted parameters spatially and evaluated the GEVD
105 goodness-of-fit to the chlorophyll time series. I correlated the fitted parameters to bathymetric
106 properties of the North Atlantic basin. I further evaluated evidence for non-stationarity (i.e.,
107 time-variability) in GEVD parameters in the context of satellite-observed chlorophyll and
108 temperature trends across the basin. Results of this study will provide a statistical framework to
109 describe interannual bloom variability and allow us to test an important hypothesis with respect
110 to basin-scale environmental change.
111
112 **Methods**
113
114 *Observations*
115
116 I analyzed two sets of basin-scale satellite chlorophyll observations. First, I used chlorophyll
117 estimates from the Moderate-resolution Imaging Spectroradiometer-Aqua (MODIS-Aqua)
118 sensor, as used in standard net primary productivity products, spanning years 2002-2021 from
119 the Oregon State University Ocean Productivity database
120 (https://sites.science.oregonstate.edu/ocean.productivity/). Temporal resolution of the satellite
121 images was eight days per image. Eighteen complete time series years yielded eighteen annual
122 maxima observations for fitting GEVDs. Associated inherent optical properties were estimated
123 using the Garver-Siegel-Maritorena (GSM) algorithm (Maritorena and Siegel 2005). MODIS-
124 Aqua based chlorophyll estimates were gap-filled for missing observations due to clouds
125 according to the algorithm described at http://orca.science.oregonstate.edu/gap_fill.php. While
126 gap-filling can alter the underlying chlorophyll distribution, it is not expected to affect the
127 annually measured maximum value as the method calculates a temporal average over observed
128 time points and would not generally exceed observed magnitudes. Secondly, I used the Ocean
129 Colour Climate Change Initiative (OC-CCI) chlorophyll product (Sathyendranath et al. 2019)
130 covering the same time period as the MODIS-Aqua dataset, accessed from
131 www.oceancolour.org. The same time period was chosen to be consistent between the two
132 analysis and compare results. OC-CCI chlorophyll is a synthetic product generated by combining
133 information from multiple sensors, including SeaWiFS (Sea-viewing Wide-Field-of-view
134 Sensor), MODIS-Aqua, MERIS (Medium spectral Resolution Imaging Spectrometer) and VIIRS
135 (Visible and Infrared Imaging Radiometer Suite). OC-CCI was not gap-filled and therefore
136 contained missing values due to clouds. I gridded both sets of chlorophyll observations to ¼°
137 latitude-longitude resolution. Wintertime satellite chlorophyll observations were not available for

138   higher latitudes due to light limitation and were replaced with a value of zero concentration
139   which did not affect the extreme value analysis. Sea surface temperature data from the MODIS-
140   Aqua sensor were obtained from the Oregon State University Ocean Productivity Database site
141   cited above. Bathymetric depth data were obtained from
142   https://www.gebco.net/data_and_products/gridded_bathymetry_data/. I analyzed chlorophyll
143   observations as a proxy for phytoplankton carbon biomass, noting that chlorophyll and carbon
144   can decouple due to photoacclimation processes, particularly at low light (Behrenfeld et al. 2005;
145   Sathyendranath et al. 2020). Despite this, chlorophyll is generally estimated with lower
146   uncertainty than carbon (Sathyendranath et al. 2020) and plays a central role in rates of primary
147   production. Carbon biomass estimates may also be used in future work with the analytical
148   framework presented here.
149
150   *Extreme Value Analysis*
151
152   Given a time series of chlorophyll observations at an individual location, I estimated the
153   parameters of the GEVD via the block maxima approach (Gilleland and Katz 2016), taking time
154   series blocks as individual years. I define $x = \max(y_1, y_2, \dots, y_n)$ as the maximum chlorophyll
155   measurement in a single year of n measurements. Over m years I have m yearly maxima, thus
156   defining the observed annual maxima time series $x_1, x_2, \dots, x_m$ (the map showing the month
157   when the maxima most frequently occurred is displayed in **Supplementary Figure 1**). For an
158   annual maximum x, the GEVD has a probability density function given by
159

$$p(x) = \frac{1}{\sigma} t(x)^{\xi+1} e^{-t(x)}$$

161   with

$$t(x) = \begin{cases} \left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}, & \xi \neq 0, \\ e^{-\frac{x-\mu}{\sigma}}, & \xi = 0, \end{cases}$$

163
164   where μ, σ, and ξ are the location, scale, and shape parameters, respectively. The location
165   parameter shifts the GEVD along the x axis, the scale parameter controls the spread, and the
166   shape parameter controls the peaked-ness of the mode and heaviness of the distribution tail.
167   Examples of how μ, σ, and ξ modulate the GEVD are given in **Supplementary Figure 2**.
168   Formulas for the expected value, variance, and mode of the GEVD are given in **Appendix A**.
169   For the satellite observations, m=18 for years 2002-2021 and n=46, on average, due to eight day
170   spacing through the year. Goodness of fit was evaluated by quantile-quantile plots where the
171   empirical quantiles of the observations are correlated against the theoretical quantiles predicted
172   from the fitted GEVD distributions.
173
174   In addition to the three-parameter GEVD described above, I also apply a nonstationary extension
175   where the parameters are described as simple linear functions of time (Gilleland and Katz 2016)
176   of the form
177

$$\theta(t) = \theta_0 + \theta_1 t,$$

179

180 where θ is one of the GEVD parameters, θ_0 is the intercept of the linear relationship, and θ_1 is
181 the slope, i.e. the rate of change with respect to time.

182
183 The log-likelihood for the GEVD parameters (Gilleland and Katz 2016) is given by

184

185 $$l(\mu, \sigma, \xi | x_1, x_2, \ldots, x_m) = -m \ln \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{m} \ln \left(1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right) - \sum_{i=1}^{m} \left(1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right)^{1/\xi}.$$

186
187 I maximized the log-likelihood function with respect to the parameters to obtain empirical
188 estimates. I maximized with respect to μ, σ, and ξ in the case of stationary GEVDs, and with
189 respect to the intercept and slope in the case of nonstationary GEVDs. I restricted the analysis to
190 estimating one nonstationary parameter at a time due to data restrictions and weak identifiability
191 when multiple parameters are allowed to vary. I only considered linear functions of time and
192 suggest nonlinear functions for future work. I used numerical optimization routines implemented
193 in the extRemes library within the R programming language (Gilleland and Katz 2016). I
194 compared stationary and nonstationary fits according to the Bayesian Information Criterion
195 (BIC), given by BIC $= -2l(\hat{\mu}, \hat{\sigma}, \hat{\xi}) - k \log n$, where $l(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ is the maximized likelihood at
196 empirical estimates $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\xi}$. k is the number of parameters in the GEVD (three for stationary
197 GEVDs, four for nonstationary GEVDs), and n is the number of yearly maxima used in the fit. A
198 GEVD was fit to chlorophyll time series in each ¼° pixel. The parameters were mapped spatially
199 and correlated with basin bathymetry. Parameter uncertainty was derived by taking the square-
200 root of the inverse Hessian matrix evaluated at the maximum likelihood estimates.

201
202 **Results**

203
204 The distributions of annual chlorophyll maxima showed excellent agreement with those
205 predicted from the GEVD. Across the Atlantic basin, observed distribution quantiles correlated
206 with those from the fitted GEVDs at $r = 0.97$ on the arithmetic scale (**Supplementary Figure**
207 **3a**) and $r = 0.98$ on the log scale (**Supplementary Figure 3b**). The spatial distribution of the
208 quantile-quantile correlation was also consistent across the basin, with no apparent relationship
209 with latitude or distance from the coast (**Supplementary Figure 3c**). The region of largest
210 disagreement occurred off the southwest coast of Europe, yet correlations were still above $r =$
211 0.7 and remained so across the basin.

212
213 When fitted GEVD parameters were mapped spatially I found that parameter magnitude closely
214 followed basin bathymetry (**Figures 1-2**). Location, scale, and shape parameters were
215 consistently elevated on the shelf and adjacent waters (<700m depth; **Figure 1**; **Figure 2a-c**).
216 Location parameters were elevated by over fivefold, scale parameters elevated approximately
217 twofold, and shape parameters elevated over threefold in waters shallower than 700m. The
218 distinction in magnitude between shelf and open ocean water was strongest for location and
219 shape parameters, and weaker for scale parameters, demonstrating a correlation between the
220 mean interannual bloom magnitude and the 'heaviness' of the underlying distribution tail. This
221 pattern appeared on the eastern and western sides of the basin. Deeper slope waters off the coast
222 of Greenland also showed elevated location and shape parameters. The largest parameter
223 magnitudes were found in the shallow Baltic Sea (**Figure 1c,d**; **Figure 2a,c**). The scale
224 parameter showed a different pattern with bathymetry where parameter magnitude was modestly

225   elevated in shelf waters but also showed a step-change decrease in the deepest waters (>4000m;
226   **Figure 1b**; **Figure 2b**). However, the aerial distribution of shelf vs. deep water is markedly
227   different, with deep waters limited to the southern half of the basin around the Mid-Atlantic ridge
228   while shelf seas are widely distributed. Parameter uncertainty weakly correlated with parameter
229   magnitude for location and shape parameters, particularly around the Greenland slope, but had
230   no consistent relationship across the basin (**Supplementary Figure 4**). Scale parameter
231   uncertainty was elevated in deeper water (**Supplementary Figure 4**). The area-weighted average
232   of the parameter coefficients of variation (the standard deviation of the parameter uncertainty
233   divided by the fitted mean parameter) was 0.11, 0.38, 0.24 for location, scale, and shape
234   parameters, meaning the 1σ uncertainty was 11%, 23.5%, and 24% of the mean, respectively
235   (**Supplementary Figure 4**). I repeated the uncertainty analysis by artificially halving and
236   doubling the number of observations to diagnose the impact of sample size on estimation
237   uncertainty. Area-weighted coefficients of variation increased to 14%, 26.3%, and 34.9% when
238   sample size was halved, and decreased to 8%, 16.6%, and 17.0% when sample size was doubled,
239   respectively (**Supplementary Figure 5**).
240
241   I quantified the correlation between fitted GEVDs parameters using linear relationships (**Figure
242   2d-f**). Bivariate relationships between parameters were well described by a linear intercept and
243   positive slope, indicating strong positive scaling relationships. The strongest relationship was
244   found between fitted location and shape parameters, reflecting that GEVD distributions increase
245   in magnitude and heavy tailed-ness with decreasing bathymetric depth. I visually characterized
246   how the GEVD changes from deep to shelf waters using the fitted linear relationships (**Figure
247   2g-h**), noting that this characterization represents the basin-averaged relationships so may not
248   necessarily be representative of individual regions. The empirical pattern underlying positive
249   scaling between location, scale, and shape parameters is reflected in the extreme value time
250   series where the mean and median of the annual extremes are positively related to the extreme
251   variance (**Supplementary Figures 6-7**).
252
253   Using a nonstationary GEVD analysis, I found weak evidence for temporal trends in GEVD
254   parameters, despite significant trends in chlorophyll levels and sea surface temperature across the
255   North Atlantic (**Figure 3**). Nonstationary parameters were favored in 34.3%, 29.9%, and 36.3%
256   of basin area for location, scale, and shape parameters, respectively (**Figures 3a-c**). Where non-
257   stationary parameters were favored, trends in the location parameter positively correlated with
258   background chlorophyll trends (r=0.70; **Figure 3, Supplementary Figure 8**). However, trends in
259   the scale and shape parameters did not correlate with chlorophyll nor sea surface temperature
260   trends (**Supplementary Figure 8**; r<0.1 in all cases). The weakest evidence for nonstationary
261   parameters was found in the area with the strongest sea surface temperature trends, specifically
262   the warming-cooling dipole pattern on the western side of the basin caused in-part by a
263   slowdown of the Atlantic overturning circulation and associated northward heat transport (i.e. the
264   North Atlantic 'warming hole'; Keil et al. 2020). I also examined parameter variability at the
265   level of Longhurst biogeochemical provinces but found no consistent pattern, with high inter-
266   and intra-province variability (**Supplementary Figure 9**).
267
268   I repeated the GEVD parameter estimation using the OC-CCI chlorophyll product and found
269   consistent results to those based on MODIS observations presented above. Quantile-quantile
270   correlations showed a similarly good fit between observed OC-CCI quantiles and those predicted

271   from the fitted GEVDs with a correlation of $r = 0.98$ in arithmetic space and $r = 0.98$ in log
272   space (**Supplementary Figure 10-11**). Spatial patterns in fitted GEVD parameters were
273   consistent across parameters estimated using the two datasets, despite some evidence for slightly
274   reduced scale parameter magnitudes using OC-CCI chlorophyll (**Supplementary Figure 10**),
275   which may be due to reduced variance in the synthetic multi-sensor OC-CCI dataset. Finally, I
276   examined the annually observed extremes for autocorrelation and found weak evidence across
277   the basin (**Supplementary Figure 12**).
278
279   **Discussion**
280
281   Our analysis demonstrates that annual chlorophyll maxima are well-described by the GEVD
282   based on the statistical theory of extreme values. I achieved a high goodness-of-fit and
283   interpretable spatial patterns across the North Atlantic basin. A clear pattern emerged in the
284   correlation between GEVD parameters and bathymetric depth, with the magnitude and tailed-
285   ness of chlorophyll extremes increasing in shelf and slope environments. The mechanism for this
286   pattern is unclear and will require a further understanding of how nutrients, shelf stratification
287   dynamics, grazing, and other factors interact to generate extreme chlorophyll concentrations. The
288   heavy distribution tail on the shelf may be related to variability in bloom timing which has been
289   shown to impact interannual variability in shelf bloom magnitude (Friedland et al. 2015).
290   However, the observed relationships may also be a more general phenomenon where the tailed-
291   ness of the chlorophyll distribution is a function of background concentrations which are
292   coincidentally elevated in shelf environments. I suggest further work examining the potential
293   mechanisms that could explain different extreme value distributions under contrasting
294   oceanographic environments. Classical models of phytoplankton blooms (e.g. Behrenfeld and
295   Boss 2014) may be extended to include stochastic forcing and generate statistical distributions of
296   bloom interannual magnitude. Targeted sensitivity analyses of annual maxima distributions to
297   different underlying forcing and background chlorophyll levels may uncover mechanisms for
298   changes in distribution parameters.
299
300   Results of the nonstationary analysis suggested weak evidence for changes in GEVD parameters
301   over time. The lack of correlation with observed trends in background chlorophyll and sea
302   surface temperature suggests that climate-related drivers are playing a limited role in modulating
303   interannual bloom magnitude, despite the North Atlantic showing significant climate change
304   (Keil et al. 2020). I note, however, that the current nonstationary analysis is limited by time
305   series length, here applied to eighteen years of observed annual maxima. Uncertainty in the
306   chlorophyll satellite record is also introduced via missing values due to clouds which partly
307   obscures the distribution of maxima. Here I found that gap-filling the missing observations had
308   no appreciable effect on the estimated extreme value parameters when comparing the gap-filled
309   and non-gap-filled chlorophyll datasets. Continued studies will be required to monitor changes in
310   bloom magnitude over time. An extended satellite record will provide greater statistical power to
311   detect climate-driven trends. An extended record will also constrain more complicated functions
312   describing the variation of parameters with time and their statistical association with
313   environmental factors. Biogeochemical general circulation models may also be analyzed to gain
314   insight into the distribution and generating mechanisms of annual maxima.
315

316   Ecologically, interesting questions arise about how extreme blooms propagate through food
317   webs, contribute to carbon export, and impact ecological processes more broadly. For example,
318   annual fisheries recruitment is often characterized by heavy-tailed distributions where individual
319   years exhibit extremely large cohorts, often fueling the fishery for years (Saetre et al. 2002).
320   Extreme bloom years may increase the probability of strong cohorts via the match-mismatch
321   mechanism (Platt et al. 2003), perhaps with temporal lags between blooms and recruitment
322   modulated by trophic transfer. With respect to carbon export, I expect extreme blooms to
323   contribute disproportionately to interannual carbon fluxes due to the commonly-observed
324   positive effect of phytoplankton productivity on carbon export efficiency (Britten and Primeau
325   2016) and observed carbon fluxes associated with blooms in the North Atlantic (Briggs et al.
326   2011). The observed positive relationship between location and shape parameters may be of
327   particular interest where the frequency and intensity of extreme blooms scales with the
328   background chlorophyll levels. Databases of carbon flux observations may be used to test these
329   relationships at the basin and global scales (Mouw et al. 2016). Beyond fisheries and carbon
330   export, I envision extreme value distributions to be broadly useful in characterizing the response
331   of ecological processes to environmental extremes. The increase in applications of extreme value
332   theory to environmental processes (Aghakouchak et al. 2020) naturally leads to questions of how
333   environmental extremes impact ecology. The GEVD is one extreme value analysis tool that is
334   theoretically motivated when considering block maxima (which was particularly appropriate here
335   to describe the maxima of a repeating annual cycle) however other statistical descriptions of
336   ecological heavy tailed-ness can also be useful in this context, for example the lognormal
337   distribution (e.g. Anderson et al. 2017).

339   In summary, the GEVD provided a useful statistical description of bloom variability and a
340   general framework to quantify the spatiotemporal statistics of interannual bloom maxima. I hope
341   this study spawns further analysis of marine ecosystem variability using extreme value theory to
342   better understand how environmental conditions give rise to ecological extreme events, how
343   extreme blooms contribute to fisheries productivity and carbon export, and how these processes
344   may change with climate.

346   **Acknowledgements**

352   **Data and Code Availability**: Original data used in this study are publicly available via the
353   sources cited in the text. Processed data and metadata used for analysis will be made available in
354   the Dryad data repository (Britten 2021). Code used to perform the GEVD analysis and visualize
355   the results is publicly available at: https://github.com/gregbritten/chl_extremes_public.

357   **References**

359   Aghakouchak, A., F. Chiang, L. S. Huning, and others. 2020. Climate extremes and compound
360   hazards in a warming world. Annu. Rev. Earth Planet. Sci. **48**: 519–548.
361   Anderson, S. C., T. A. Branch, A. B. Cooper, and N. K. Dulvy. 2017. Black-swan events in

362　　　　animal populations. Proc. Natl. Acad. Sci. **114**: 3252–3257.

363　Batt, R. D., S. R. Carpenter, and A. R. Ives. 2017. Extreme events in lake ecosystem time series.
364　　　　Limnol. Oceanogr. Lett. **2**: 63–69.

365　Behrenfeld, M. J. 2010. Abandoning sverdrup's critical depth hypothesis on phytoplankton
366　　　　blooms. Ecology **91**: 977–989.

367　Behrenfeld, M. J., and E. S. Boss. 2014. Resurrecting the ecological underpinnings of ocean
368　　　　plankton blooms. Ann. Rev. Mar. Sci. **6**: 167–194.

369　Behrenfeld, M. J., and E. S. Boss. 2018. Student's tutorial on bloom hypotheses in the context of
370　　　　phytoplankton annual cycles. Glob. Chang. Biol. **24**: 55–77.

371　Behrenfeld, M. J., E. Boss, D. A. Siegel, and D. M. Shea. 2005. Carbon-based ocean productivity
372　　　　and phytoplankton physiology from space. Global Biogeochem. Cycles **19**: 1–14.

373　Benedetti-Cecchi, L., A. Canepa, V. Fuentes, and others. 2015. Deterministic factors overwhelm
374　　　　stochastic environmental fluctuations as drivers of jellyfish outbreaks. PLoS One **10**: 1–16.

375　Briggs, N., M. J. Perry, I. Cetinić, C. Lee, E. D'Asaro, A. M. Gray, and E. Rehm. 2011. High-
376　　　　resolution observations of aggregate flux during a sub-polar North Atlantic spring bloom.
377　　　　Deep. Res. Part I Oceanogr. Res. Pap. **58**: 1031–1039.

378　Britten, G. L. 2021. Regridded subset of MODIS chlorophyll, OC-CCI chlorophyll, MODIS sea
379　　　　surface temperature; basin bathymetric depth. Dryad Dataset.
380　　　　doi:https://doi.org/10.5061/dryad.k6djh9w6w

381　Britten, G., and F. Primeau. 2016. Biome-specific scaling of ocean productivity, temperature,
382　　　　and carbon export efficiency. Geophys. Res. Lett. **43**: 5210–5216.

383　Butitta, V. L., S. R. Carpenter, L. C. Loken, M. L. Pace, and E. H. Stanley. 2017. Spatial early
384　　　　warning signals in a lake manipulation. Ecosphere **8**.

385　Coles, S. 2001. An introduction to statistical modeling of extreme values, Springer-Verlag.

386　Dutkiewicz, S., M. Follows, J. Marshall, and W. W. Gregg. 2001. Interannual variability of
387　　　　phytoplankton abundances in the North Atlantic. Deep. Res. Part II Top. Stud. Oceanogr.
388　　　　**48**: 2323–2344.

389　Easterling, D. R., G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, and L. O. Mearns.
390　　　　2000. Climate extremes: observations, modeling, and impacts. Science. **289**: 2068–2075.

391　Evans, G. T., J. S. Parslow, G. T. Evans, and J. S. Parslow. 1985. A model of annual plankton
392　　　　cycles. Deep Sea Res. Part B. Oceanogr. Lit. Rev. **32**: 759.

393　Falkowski, P., R. Barber, and V. Smetacek. 1998. Biogeochemical controls and feedbacks on
394　　　　ocean primary production. Science. **281**: 200–206.

395　Falkowski, P., E. Laws, R. Barber, and J. Murray. 2003. Phytoplankton and their role in primary,
396　　　　new and export production, p. 99–121. *In* M.J.R. Fasham [ed.], Ocean Biogeochemistry.
397　　　　Springer.

398　Follows, M., and S. Dutkiewicz. 2002. Meteorological modulation of the North Atlantic spring
399　　　　bloom. Deep Sea Res. Part II Top. Stud. Oceanogr. **49**: 321–344.

400　Friedland, K. D., R. T. Leaf, J. Kane, and others. 2015. Spring bloom dynamics and zooplankton
401　　　　biomass response on the US Northeast Continental Shelf. Cont. Shelf Res. **102**: 47–61.

402　Gaines, S. D., and M. W. Denny. 1993. The largest, smallest, highest, lowest, longest, and
403　　　　shortest: extremes in ecology. Ecology **74**: 1677–1692.

404　Gilleland, E., and R. W. Katz. 2016. extRemes 2.0: An extreme value analysis package in R. J.
405　　　　Stat. Softw. **72**: 1–39.

406　Huisman, J., P. van Oostveen, and F. J. Weissing. 1999. Critical depth and critical turbulence:
407　　　　Two different mechanisms for the development of phytoplankton blooms. Limnol.

408   Oceanogr. **44**: 1781–1787.

409   Ito, T., and M. J. Follows. 2005. Preformed phosphate, soft tissue pump and atmospheric CO2. J.
410   Mar. Res. **63**: 813–839.

411   Katz, R. W. 2010. Statistics of extremes in climate change. Clim. Change **100**: 71–76.

412   Keil, P., T. Mauritsen, J. Jungclaus, C. Hedemann, D. Olonscheck, and R. Ghosh. 2020. Multiple
413   drivers of the North Atlantic warming hole. Nat. Clim. Chang. **10**: 667–671.

414   Maritorena, S., and D. a. Siegel. 2005. Consistent merging of satellite ocean color data sets using
415   a bio-optical model. Remote Sens. Environ. **94**: 429–440.

416   Mouw, C. B., A. Barnett, G. A. Mckinley, L. Gloege, and D. Pilcher. 2016. Global ocean
417   particulate organic carbon flux merged with satellite parameters. Earth Syst. Sci. Data **8**:
418   531–541.

419   Platt, T., C. Fuentes-Yaco, and K. T. Frank. 2003. Spring algal bloom and larval fish survival.
420   Nature **423**: 398–399.

421   Saetre, R., R. Toresen, and T. Anker-Nilssen. 2002. Factors affecting the recruitment variability
422   of the Norwegian spring-spawning herring (Clupea harengus L.). ICES J. Mar. Sci. **59**:
423   725–736.

424   Sathyendranath, S., R. J. W. Brewin, C. Brockmann, and others. 2019. An ocean-colour time
425   series for use in climate studies: The experience of the ocean-colour climate change
426   initiative (OC-CCI). Sensors **19**: 1–31.

427   Sathyendranath, S., T. Platt, Ž. Kovač, and others. 2020. Reconciling models of primary
428   production and photoacclimation. Appl. Opt. **59**: 100–114.

429   Siegel, D. A., S. C. Doney, and J. A. Yoder. 2002. The North Atlantic spring phytoplankton
430   bloom and Sverdrup's critical depth hypothesis. Science. **296**: 730–733.

431   Sommer, U., and K. Lengfellner. 2008. Climate change and the timing, magnitude, and
432   composition of the phytoplankton spring bloom. Glob. Chang. Biol. **14**: 1199–1208.

433   Taylor, J. R., and R. Ferrari. 2011. Ocean fronts trigger high latitude phytoplankton blooms.
434   Geophys. Res. Lett. **38**: 1–5.

435   Visser, F., K. L. Hartman, G. J. Pierce, V. D. Valavanis, and J. Huisman. 2011. Timing of
436   migratory baleen whales at the azores in relation to the north atlantic spring bloom. Mar.
437   Ecol. Prog. Ser. **440**: 267–279.

438

To be submitted to *Limnology & Oceanography Letters*
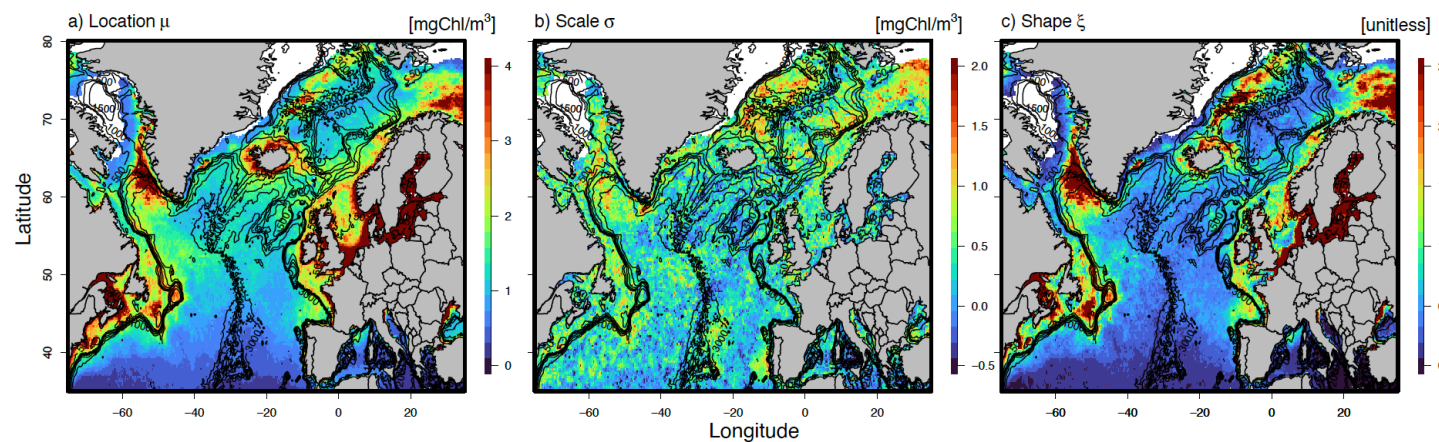
439 **Figures**
440
441

442
443 Figure 1. Spatial distribution of fitted GEVD parameters and basin bathymetry. Panels **a**, **b**, **c**
444 give the estimated location, scale, and shape parameters, respectively. Bathymetry contours are
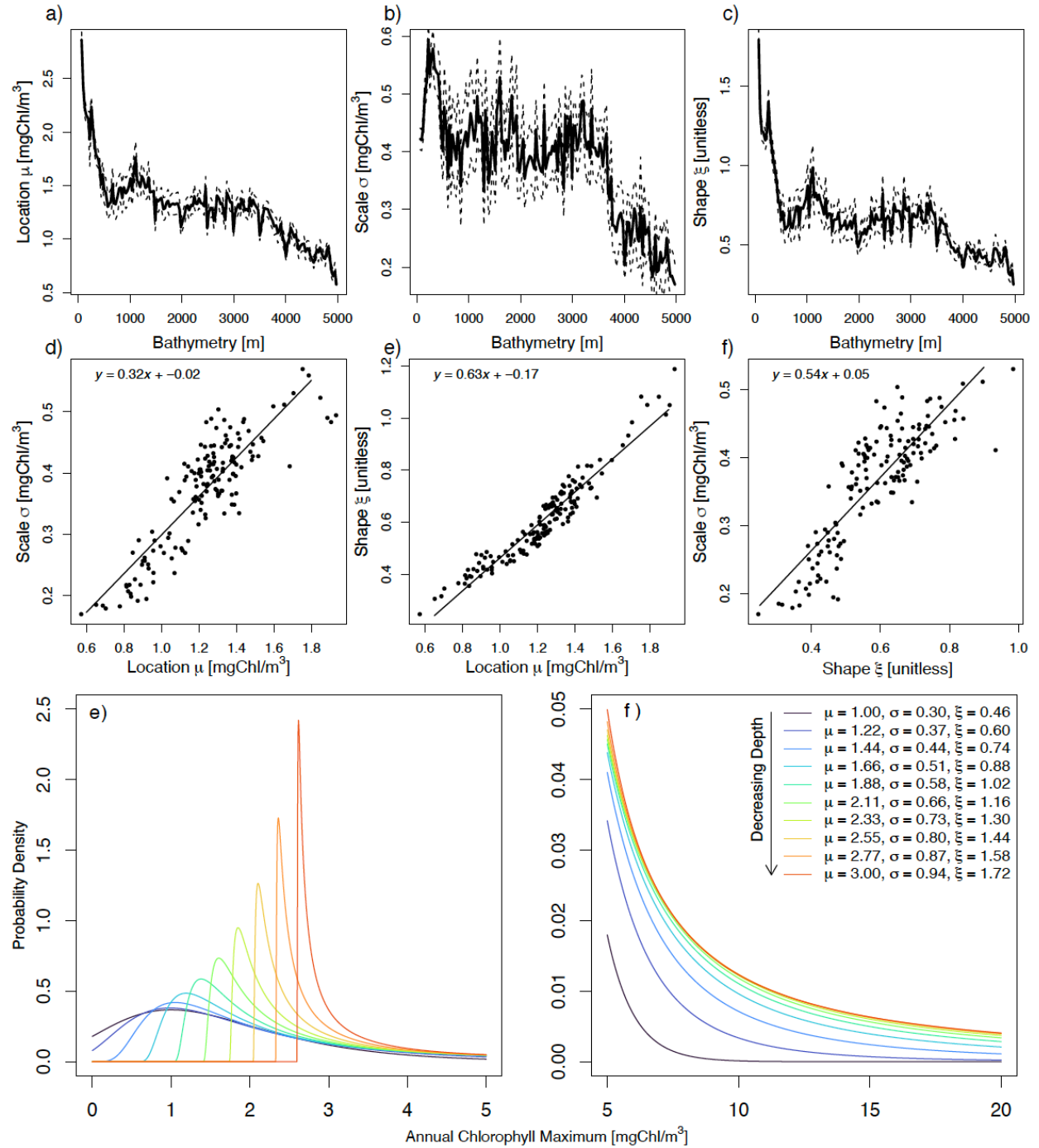445 overlaid with a contour interval of 500 meters.
446

447
448
449 Figure 2. Relationships between fitted GEVD parameter and bathymetric depth. Panels **a-c** give
450 the basin-averaged relationship between the fitted parameters and bathymetric depth with 95%
451 intervals. Panels **d-f** give the inter-relationships among fitted parameters with the least squares
452 regression line and equation. Panels **g** and **h** give the basin-averaged change of the GEVD with
453 bathymetric depth. Panel **h** zooms in on the distribution tail in **g** (note horizontal axis limits).
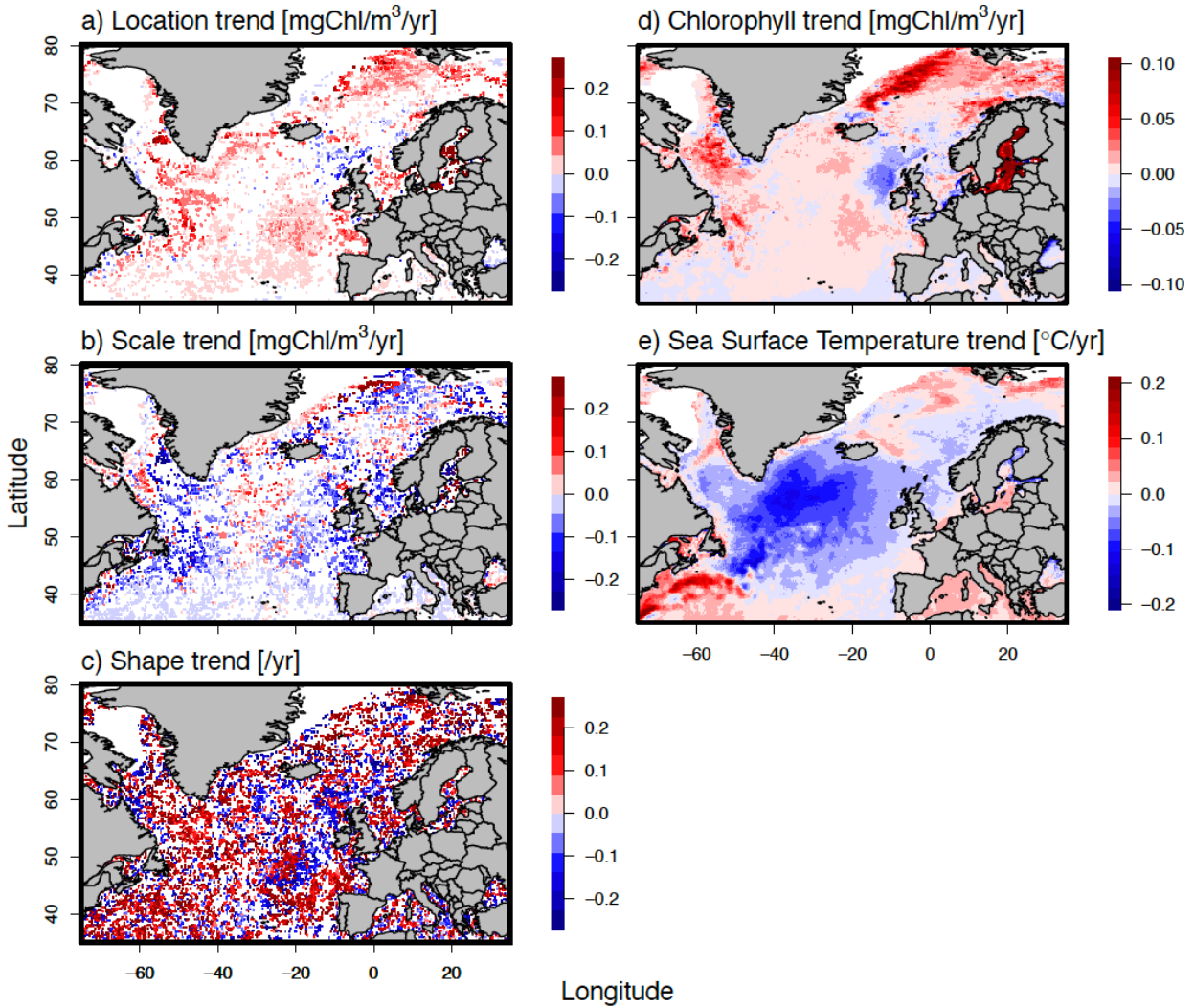
454
455 Figure 3. Trends in GEVD parameters in the context of basin-scale trends in chlorophyll and sea
456 surface temperature. Panels **a**, **b**, and **c** give the estimated parameter rate of change for location,
457 scale, and shape parameters, respectively. Rates of change are only shown where model selection
458 favored the nonstationary distribution (see text for details). Panels **d** and **e** give the linear trend in
459 surface chlorophyll and sea surface temperature, respectively, estimated from MODIS
460 observations since 2002.

461 **Appendix A**
462
463 The expected value of the GEVD is given by
464

$$\mathbf{E}[x] = \begin{cases} \mu + (g(1) - 1)\dfrac{\sigma}{\xi} & if\ \xi \neq 0, \xi < 1 \\ \mu + \sigma\gamma & if\ \xi = 0 \\ \infty & if\ \xi \geq 1 \end{cases}$$

465

466
467 where $g(k) = \Gamma(1 - k\xi)$, $\Gamma(r)$ is the gamma function, and γ is Euler's constant ($\gamma \cong 0.5772$).
468 The variance of the GEVD is
469

$$\mathbf{Var}[x] = \begin{cases} (g(2) - g(1)^2)\dfrac{\sigma^2}{\xi^2} & if\ \xi \neq 0, \xi < \dfrac{1}{2} \\ \dfrac{\sigma^2\pi^2}{6} & if\ \xi = 0 \\ \infty & if\ \xi > \dfrac{1}{2} \end{cases}$$

470

471
472 The mode of the GEVD is
473

$$\mathbf{Mode}[x] = \begin{cases} \mu + \dfrac{\sigma}{\xi}((1 + \xi)^{-\xi} - 1) & if\ \xi \neq 0 \\ \mu & if\ \xi = 0 \end{cases}$$

474

475

Supplemental Information:

Extreme value distributions describe interannual variability in the seasonal North Atlantic phytoplankton bloom

Gregory L. Britten
Program in Atmospheres, Oceans, and Climate
Massachusetts Institute of Technology, Cambridge, MA, USA
gregleebritten@gmail.com
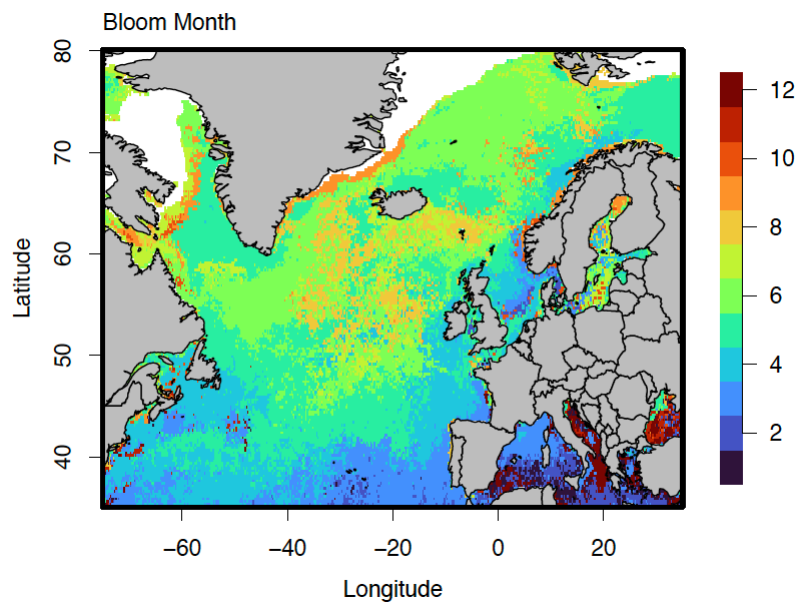
Supplementary Figure 1. The most frequent month when the annual maxima occurred calculated using MODIS observations. January is month #1.
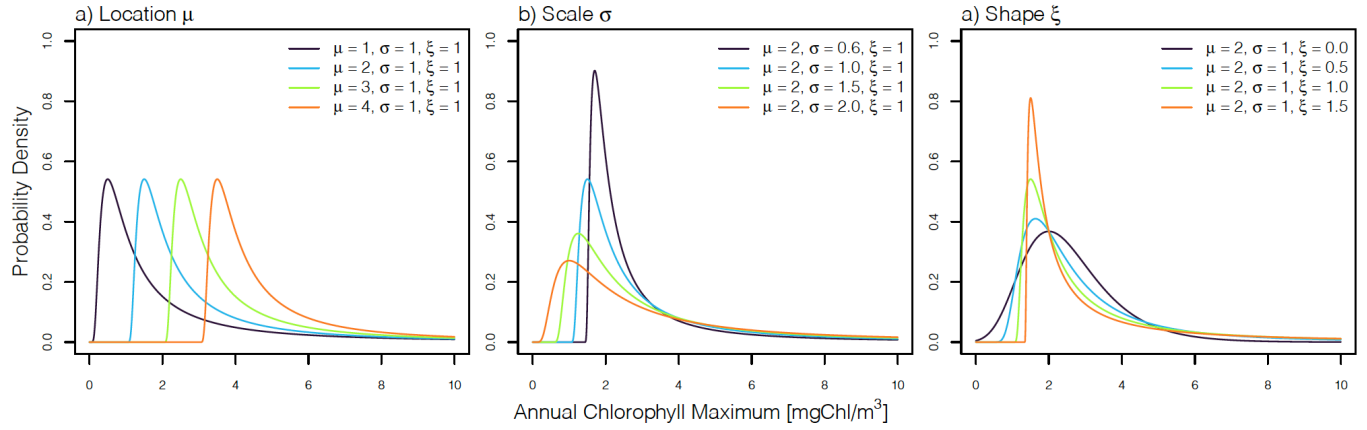
Supplementary Figure 2. Examples of the GEVD distribution for different values of the location (**a**), scale (**b**), and shape (**c**) parameters.
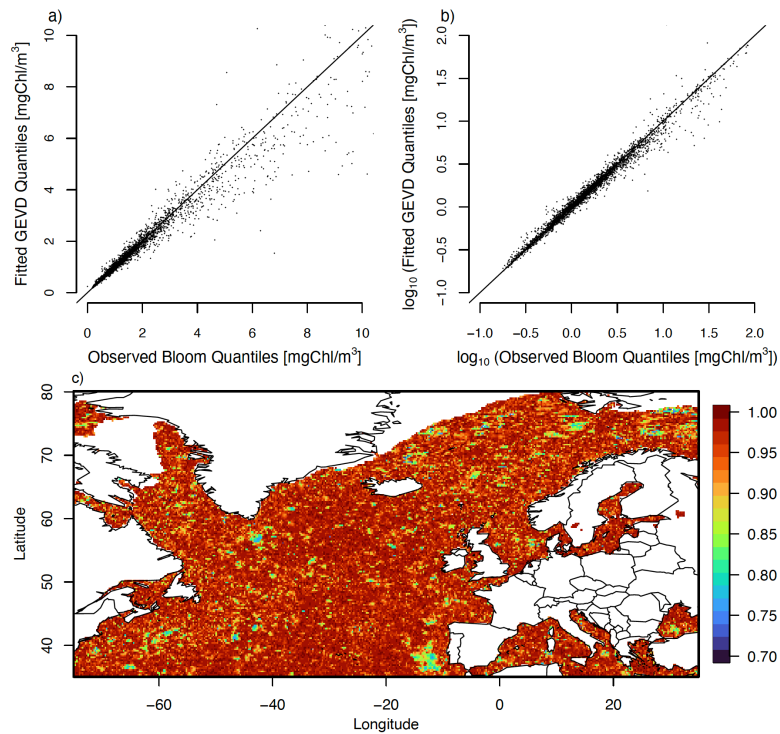
Supplementary Figure 3. Observed quantiles of North Atlantic annual bloom extremes vs. those predicted from the fitted GEVDs. Panel **a** gives the relationship in arithmetic space; panel **b** gives the relationship in log (base ten) space. A random sample of 5000 quantiles were drawn in order to visualize the relationship in **a** and **b**. The correlations in **a** and **b** are 0.97 and 0.98, respectively. Panel **c** gives the spatial distribution of quantile-quantile correlations. Note the correlation color bar extends from 0.7 to 1.0.
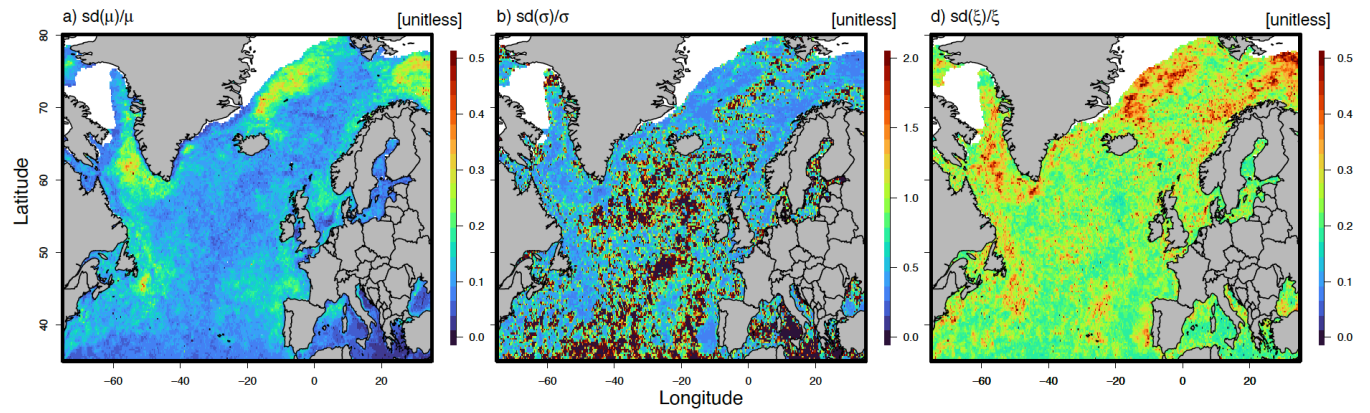
Supplementary Figure 4. Uncertainty in fitted GEVD parameters expressed as the coefficient of variation (CV). The CV is calculated as the ratio of the uncertainty standard deviation divided by the mean of the fitted parameter.
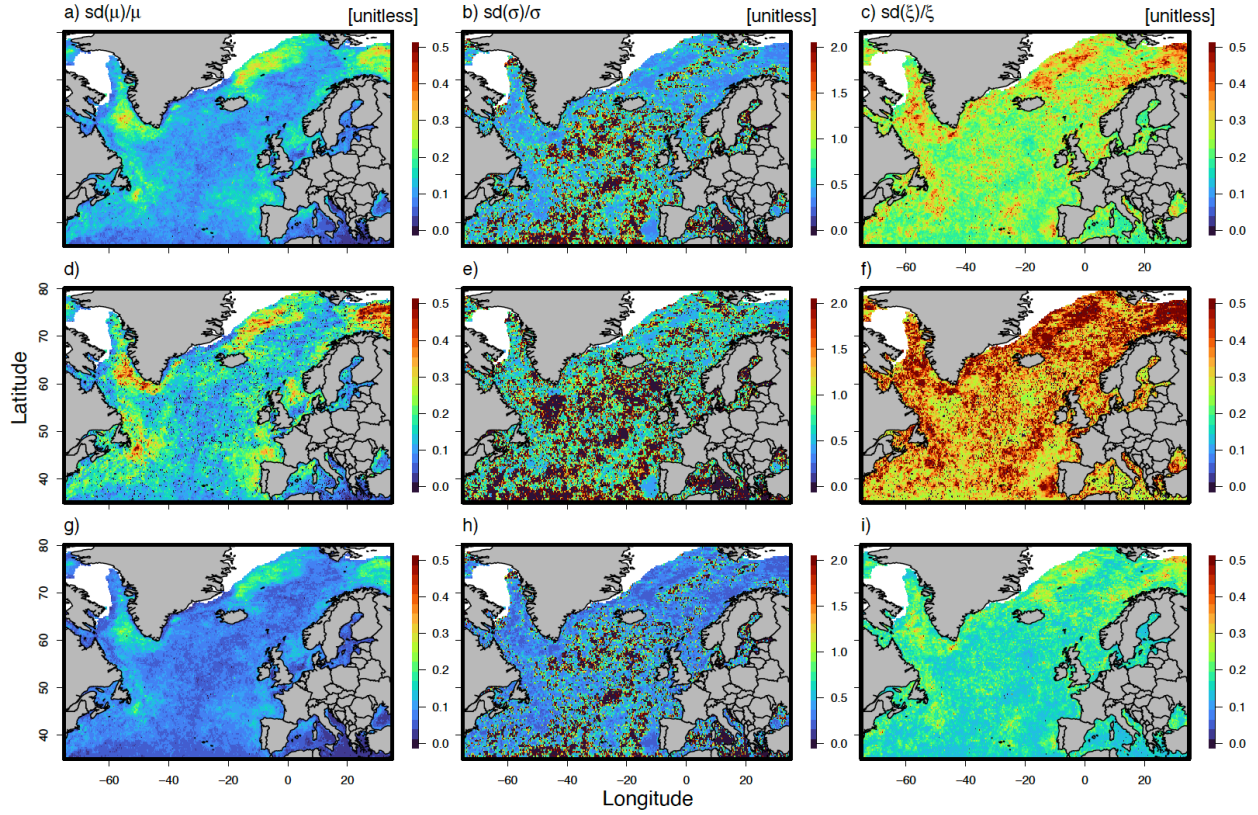
Supplementary Figure 5. Effect of sample size on parameter uncertainty. Top row (**a-c**) gives the coefficients of variation as in Supplementary Figure 4. Second (**d-f**) and third (**g-i**) rows give the coefficients of variation when the number of observations was halved and doubled, respectively. The first nine years were used when halving the observations. The time series was doubled by repeating the time series a second time after the last observed year.
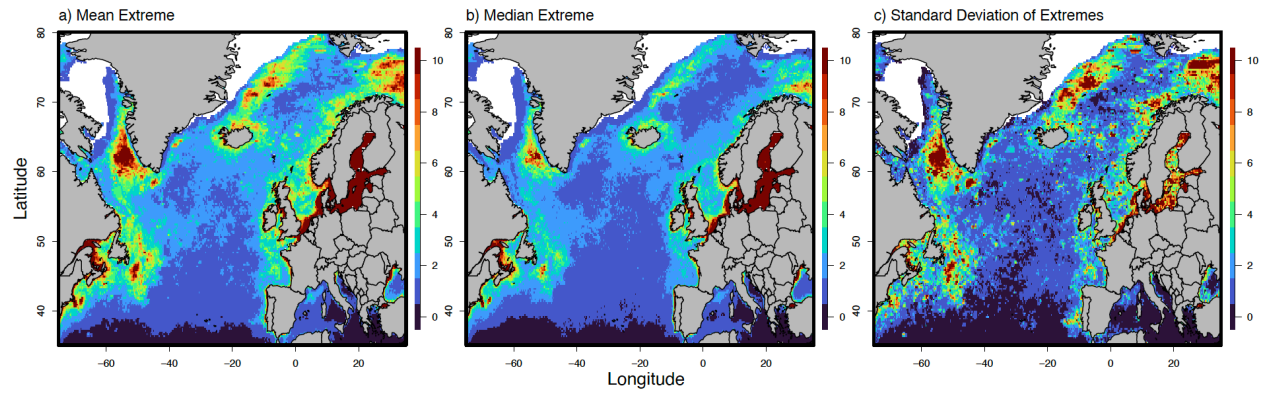
Supplementary Figure 6. Maps giving the means (**a**), medians (**b**), and standard deviation (**c**) of the observed extremes.
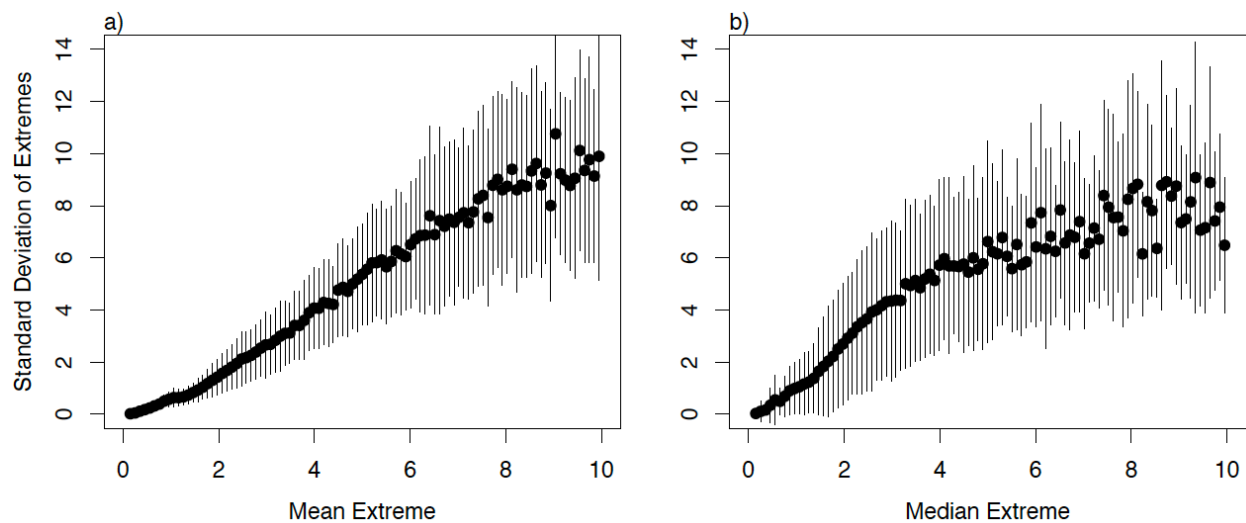
Supplementary Figure 7. Scaling relationships between the mean extreme, median extreme, and standard deviation of the extremes across the North Atlantic basin.
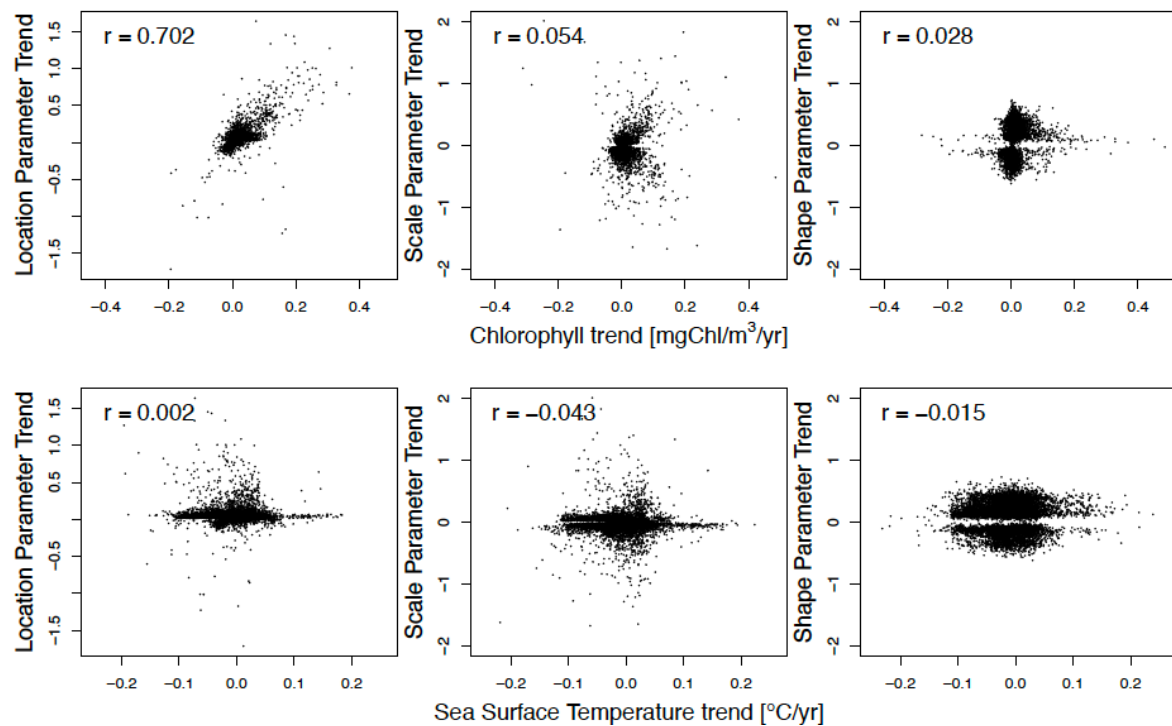
Supplementary Figure 8. Scatterplot relationships derived from results of Figure 3 in the main text. Top row gives the estimated trend in location, scale, and shape parameters vs. the trend in background chlorophyll concentrations. Bottom row gives the estimated trend in location, scale, and shape parameters vs. the trend in SST. Correlation coefficient for each relationship is given in the top left of each panel. Trends are only shown were ΔBIC > 2 for the nonstationary relationships.
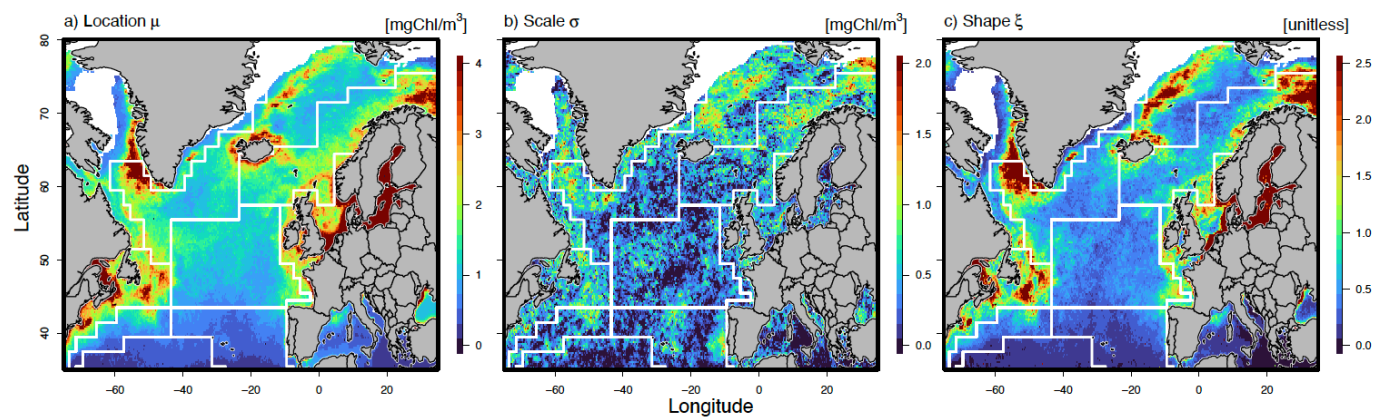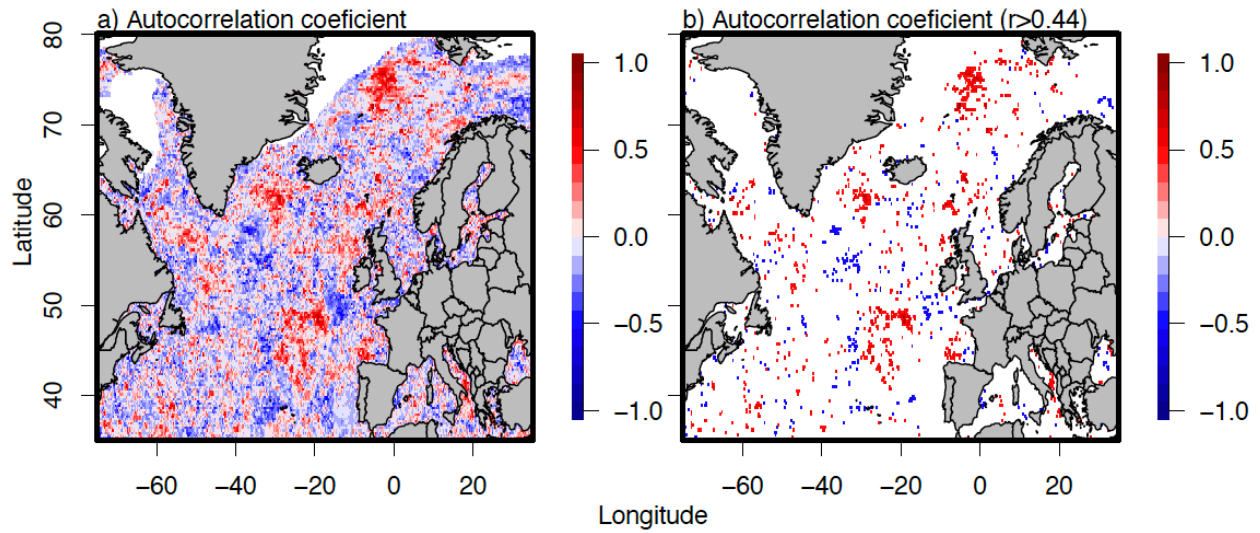
Supplementary Figure 9. Figure 3 of the main text with Longhurst biogeochemical provinces overlaid.
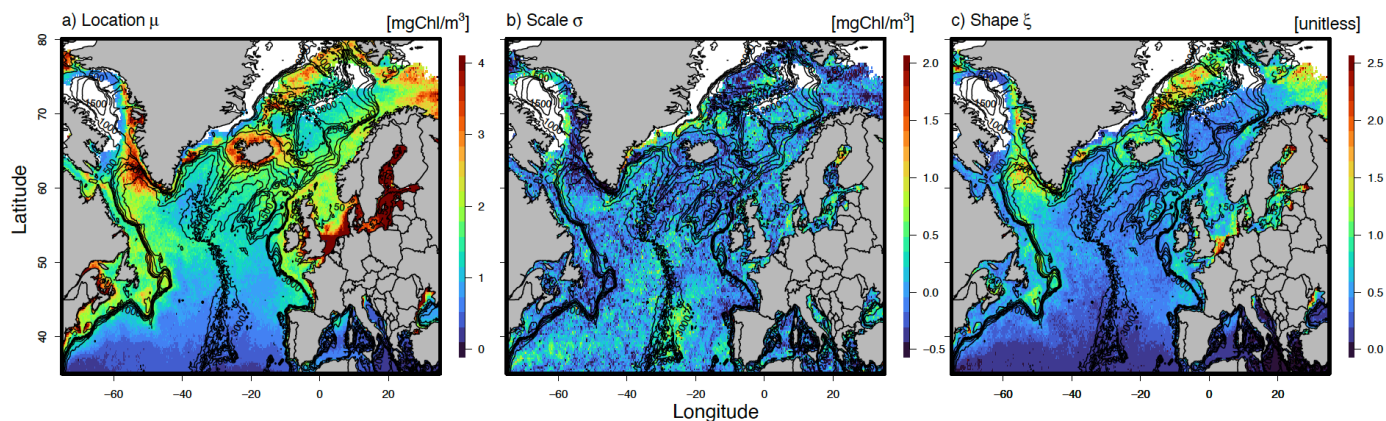
Supplementary Figure 10. As in Figure 1 of the main text but using OC-CCI chlorophyll estimates.
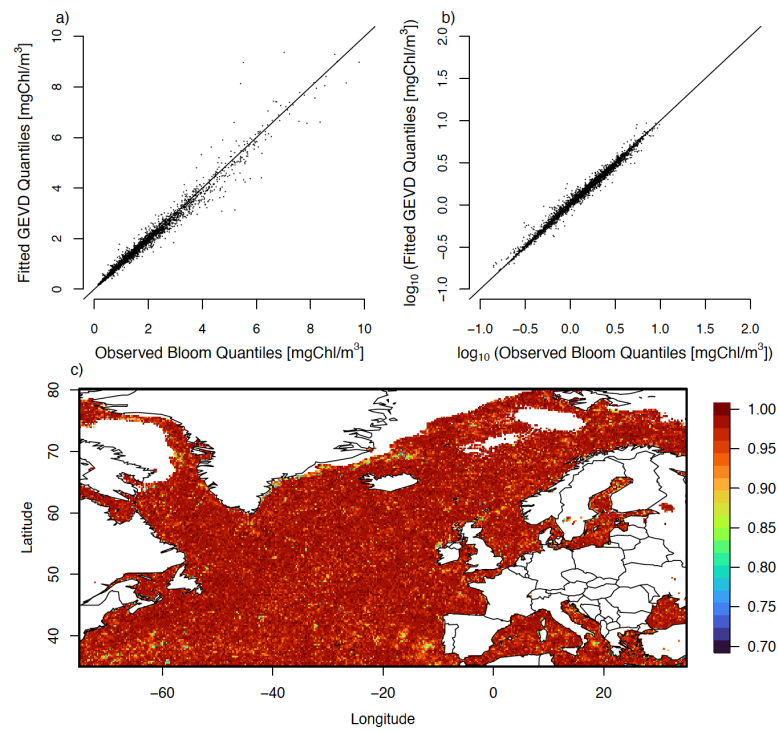
Supplementary Figure 11. As in Supplementary Figure 3 but using OC-CCI chlorophyll estimates.

Supplementary Figure 12. Lag one autocorrelation coefficient for the extreme time series. Panel (a) gives all estimated coefficients. Panel (b) gives those that are considered statistically significant at the 95% confidence level, found by solving the t statistic equation $1.96 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$.