Metabolic constraints on the body size scaling of extreme population densities

Angel Segura[1] and Gonzalo Perera[2]

[1]Facultad de Ciencias
[2]Universidad de la República, Centro Universitario Regional Este, Modelización y Análisis de Recursos Naturales, Uruguay

November 29, 2022

Abstract

Pest outbreaks, harmful algal blooms, and population collapses are extreme events with critical consequences for ecosystems, highlighting the importance of deciphering the driving ecological mechanisms underlying extreme events. By combining the generalized extreme value (GEV) theory from statistics and the hypothesis of a resource-limited metabolic restriction to population abundance, we evaluated theoretical predictions on the size-scaling and variance of extreme population abundance. Phytoplankton data from the L4 station in the English Channel showed a negative size scaling of the expected value of maxima, whose confidence interval included the predicted metabolic scaling (a = –1). We showed a humped pattern in variance with maxima at intermediate sizes. These results are consistent with the bounded abundance of small-sized populations that are subjected to strong grazing and with the expected decrease in variance towards large sizes. This approach provides unbiased return times, thereby improving the prediction accuracy of the timing of bloom formation, and describes a coherent framework in which to explore extreme population densities in natural communities.

Metabolic constraints on the body size scaling of extreme population densities

Segura A.M.[1*] & Perera, G.[1,2]

[1] Modelización Estadística de Datos e Inteligencia Artificial (MEDIA), Centro Universitario Regional del Este, Universidad de la República, Rocha, Uruguay.

[2] Instituto de Matemática y Estadística Rafael Laguarda (IMERL), Facultad de Ingenieria, Universidad de la República

To be submitted to: Ecology letters

*Corresponding author

Angel M. Segura

e-mail: asegura@cure.edu.uy

CURE-Rocha, Ruta 9 km 210. PC 27000

Type= Letter

# of words: 3684 in abstract: 147

 # Figures: 4

 # Citations: 44

Author`s contribution: AMS designed research, generate R codes and results and write a first draft. GP provided analytical tools and contributed to the writing of the final draft.

Data accessibility statement:

Should the manuscript be accepted, the data supporting the results will be archived in Zenodo and the code for downloading and analizing data will be deposited in the same repository. It will be available i) the raw data files that were analyzed to produce the statistics and figures reported in the paper, ii) a metadata file explaining what the data are in each column and what rows represent. The units of measurement and a verbal description in English of the data included in each column, iii) the computer code used to produce the figures in R. Station L4 data are archived at the British Oceanographic Data Centre ***www.bodc.ac.uk*** under DOI:10.5285/c9386b5c-b459-782f-e053-1

6c86abc0d129and are freely available at https://www.bodc.ac.uk/data/published_data_library/catalogue/10.5285/c9386b5c-b459-782f-e053-6c86abc0d129/. Western Channel Observatory (L4) nutrient concentration profiles (2000 – 2020) doi:10.5285/bc3cf5ce-d18c-1f42-e053-6c86abc02e29

Abstract

Pest outbreaks, harmful algal blooms, and population collapses are extreme events with critical consequences for ecosystems, highlighting the importance of deciphering the driving ecological mechanisms underlying extreme events. By combining the generalized extreme value (GEV) theory from statistics and the hypothesis of a resource-limited metabolic restriction to population abundance, we evaluated theoretical predictions on the size-scaling and variance of extreme population abundance. Phytoplankton data from the L4 station in the English Channel showed a negative size scaling of the expected value of maxima, whose confidence interval included the predicted metabolic scaling ($\alpha=-1$) and a humped pattern in variance. These results are consistent with the bounded abundance of small-sized populations that are subjected to strong grazing and with the expected decrease in variance towards large sizes. This approach provides unbiased return times, thereby improving the prediction accuracy of the timing of bloom formation under a coherent framework.

Keywords: Extreme events, generalized extreme value distribution, metabolic limit, scaling

INTRODUCTION

Extreme events are relevant for ecosystem functioning in terms of both the impact of abiotic extremes in the abundance of natural populations (*e.g.*, hurricanes, tsunamis) and in the extremes caused by intrinsic population dynamics (May 1973; Bjornstad 2001). Species collapse, pest outbreaks, and harmful algal blooms are pertinent examples of events with important consequences for the environment, human health, and conservation. Descriptions of the dynamics of extreme events in hydrology date back to the early eighteenth century; today, a statistical theory of extreme values is well developed (e.g. Embrechts *et al.* 2008). For decades, extreme environmental effects have been of interest to both population and community ecology (Gaines & Denny 1993; Katz *et al.* 2005), and more recent studies have advanced our understanding of the distribution of extreme population dynamics (Segura *et al.* 2013; Batt *et al.* 2017). However, there continues to lack a formal test of the mechanisms driving the limits of extreme population abundances. Two aspects deserve special attention; first, GEV distribution provides an unbiased estimation of return period of extreme events. This is critical to predict the timing of population outbursts, because the use classical distributions (e.g., Gaussian) could lead to a large underestimation of the average time between extreme population's abundance (*i.e.* return times, Fig. 1). This bias could lead to wrong management actions, or the sense of unpredictability of the extreme events. Second, the analysis of extremes is conceptually rooted on defining an upper limit to abundance in populations or communities, but the theoretical basis is not well developed (Lawton 1989; Barneche *et al.* 2016).

The limits to maximum population abundance are set by the ability of a population to transform resources into offspring and minimize losses, while embedded in a complex network of interactions (Lawton 1989; Bjornstad 2001). Macroscopic approaches are useful to understand the structure and dynamics of ecosystems because they simplify the complexity of natural systems by describing a set of general patterns using observable variables (*e.g.* body volume) linked by power scaling laws (Brown *et al.* 2004; Marquet 2005). Body size is related to metabolism, growth,

abundance, and predator defence, among other processes and it is considered a master trait in ecology (Brown *et al.* 2004; Zaoli *et al.* 2017; Hatton *et al.* 2019). The metabolic limit to maximum population abundances dependens on body size, and this relationship has been suggested as a macroscopic approach to this problem (Agustí *et al.* 1987; Belgrano *et al.* 2002; Deng *et al.* 2012).

The formal link between the scaling of population abundances with body size and extreme value theory provides a fruitful avenue to explore metabolic limits of extreme population abundances in natural communities. In the present paper, under the hypothesis of a metabolic restriction to maximum populations abundance, we developed a theoretical model and tested specific formal predictions on the distribution of extreme population abundances using one of the most comprehensive time series of abundance in a natural community.

Model formulation

Metabolic limits to maximum population abundance

Without loss of generality, we developed a theoretical model for aquatic primary producers; however, the principles and formulations are generalizable to other ecosystems and taxonomic groups. The size density relationship (SDR) presents a power law relationship between body-size (M) and maximum population density *max*(N) (Belgrano *et al.* 2002; Brown *et al.* 2004; Barneche *et al.* 2016; Segura & Perera 2019):

$$max(N) = a\,M^{-\alpha} \qquad (1)$$

where the power α is the scaling of the metabolic rate and body size and a is a temperature and resource normalization constant. This relationship is based on the assumption that total resource use (R_{tot}) in a local ecosystem (e.g., light or nutrients in the case of autotrophs) is equal to the sum of the population-level rates of resource use per unit area or volume, R_i, across S cohabiting species.

R_i, in turn, is proportional to the product of the metabolic rate (B_i) and the population density per unit area or volume (N_i):

$$R_i = N_i B_i \quad (2)$$

Our interest is in understanding the upper limit to population density, *max*(N_i). Assuming a dominant species can use no more than γ of total resources (R_{tot}) and that metabolic rate is a function of body size and temperature (Brown *et al.* 2004)

$$B_i = b_0 e^{\frac{-E}{kT}} M_i^{\alpha} \quad (3)$$

Where b_0 is a taxon-dependent normalization constant, E is the activation energy (eV) and *k* is the Boltzmann constant (1.6×10^{-5} eVK^{-1}). Replacing eq. 3 in eq. 2, and then substituting γR_{tot} for R_i, it is possible to explore the relationship for maximum population density (Belgrano *et al.* 2002; Deng *et al.* 2012):

$$max(N_i) = \gamma b_0^{-1} R_{tot} e^{\frac{E}{kT}} M_i^{-\alpha} \quad (4)$$

Taking the logarithm in both sides, following Segura and Perera (2019), we can define:

$$n_{max} = \log(max(N_i)) = [\log(\gamma) + \log(b_0^{-1}) + \log(R_{tot}) + \frac{E}{kT}] - \alpha \ \log(M_i) \quad (5)$$

Here, we explore the role of nutrients in the fraction of resources used by dominant species by assuming that γ is resource- and size-dependent $\gamma = f(R_{tot}, M)$ and bounded between 0 and 1. The size-dependent fraction of total resources captured by a given dominant species can be modelled as a Monod-like function (Huisman *et al.* 1999):

$$\gamma \propto \frac{R_{tot}}{R_{tot} + k_s} \quad (6)$$

where k_s is the half saturation constant.

Two resources are fundamental for primary producers, nutrients, and light. The concentration of dissolved inorganic nitrogen (DIN) forms, which is the sum of nitrate (NO_3), nitrite (NO_2), and ammonia (NH_4), are limiting in marine ecosystems and theoretical biophysical arguments on the dynamics of exchange of resources across the cell membrane suggest a specific size-scaling of k_s ($k_s \sim M^{0.33}$) (Aksnes & Egge 1991). In the case of phytoplankton, an empirical power law scaling with body size supporting this biophysical theoretical model has been found (k_s= 0.14 $M^{0.33}$; Edwards *et al.* 2012). Light limitation is a complex process but exhibits a saturating curve with respect to light, and it has been modelled using a Monod formulation with an invariant half saturation constant (k_{PAR} = 1.51 mol photons m^{-2} day^{-1}) in phytoplankton (Huisman *et al.* 1999; López-Urrutia *et al.* 2006). Under steady state, metabolism is limited by the most limiting rate (dictated by nutrients or light); therefore, we define a general k_R that represents either nutrient or light limitation. Including the theoretical formulation for the fraction of resources used by dominant species (eq. 6) into eq. 5:

$$n_{max} = log(max(N_i)) = [log(\frac{R_{tot}}{R_{tot}+k_R}) + log(b_0^{-1}) + log(R_{tot}) + \frac{E}{kT}] - \alpha \ \ log(M_i) \qquad (7)$$

Equation 7 allows the generation of testable predictions about the relationship between size, resources, temperature, and extreme population abundance.

The generalized extreme value theory

Taking the maximum of a finite sequence of independent and identically distributed (*iid*) random variables leads to an asymptotic distribution satisfying the “max-stability property” (Katz *et al.* 2005; Embrechts *et al.* 2008). There are three nominal distributions of extreme values, namely, the Weibull, Gumbel, and Fréchet distributions. They result from taking maxima from *iid* bounded (*e.g.*, Uniform), unbounded with light tailed (*e.g.*, Gaussian), and unbounded with fat-tailed (*e.g.*, Cauchy) distributions. The general extreme value (GEV) distribution is a family of continuous probability distributions, which includes the three mentioned canonical extremal distributions. Its cumulative function (F) is described as:

$$F(x,\mu,\sigma,\xi)=\begin{cases}\exp\{-[1+\xi(x-\mu)/\sigma]^{-1/\xi}\}, & 1+\xi(x-\mu)/\sigma>0, \quad \xi\neq 0\\ \exp\{-\exp(-(x-\mu)/\sigma)\}, & \xi=0\\ 0 \quad , & \xi(x-\mu)/\sigma\leq -1 \quad \xi>0\\ 1 \quad , & \xi(x-\mu)/\sigma\leq -1 \quad \xi<0\end{cases} \tag{8}$$

Where μ is the location, σ > 0 is the scale, and ξ is a shape parameter. The location specifies the centre of the distribution, the scale is the spread, and the shape is a relevant parameter that describes the behaviour of the tails. The shape parameter (ξ) is critical to understand ecological dynamics, as it characterizes whether the distribution of maxima is bounded (*i.e.* ξ < 0), has a light tail (*i.e.* ξ = 0), or has a heavy tail (*i.e.*, ξ > 0). There are many ecological mechanisms that can drive the outburst of species abundance and, thus heavy tailed extremal distributions (ξ > 0), as suggested by previous analysis (Keitt & Stanley 1998; Segura *et al.* 2013; Batt *et al.* 2017), but other ecological mechanisms may keep populations constrained within a bounded range of abundances (Bjornstad 2001; Segura *et al.* 2013).

From the metabolic dependence of maximum population density (eqs. 4 and 7) we derive a set of specific predictions amenable to testing under the GEV theory. The size-scaling of the expected valued of the GEV will match the scaling of metabolism (α), which in marine phytoplankton was found empirically to be one (α=1; López-Urrutia *et al.* (2006)). The variance of population fluctuations will decrease with body size (Segura & Perera 2019). However, a strong match of populations abundance driven by predation (as it occurs in the lower size fractions in phytoplankton) will exhibit deviations from this pattern. Using the SDR framework and the GEV theory, we test explicit predictions using one of the longest time series of abundance in a species-rich community of plankton in the English Channel L4 station.

MATERIALS AND METHODS

Time series description

Phytoplankton abundance and individual volume information were obtained from data collected in

the Western English Channel, 10 nautical miles off Plymouth (50°15' N, 4°13' W), at the L4 station and provided by the Western Channel Observatory (http://www.westernchannelobservatory.org.uk/l4/Phytoplankton) (Widdicombe & Harbour 2021). This is an extensive data set of more than 200 species and consisted of 28 years' worth (October 1992 to June 2020) of weekly sampled phytoplankton abundance (N weeks = 1183). Some weeks were not sampled due to weather issues. Detailed information on sampling design, taxonomic identification and counting methodologies can be found in Widdicombe *et al.* (2010) and Widdicombe and Harbour (2021). Average individual volume (μm^3/org) was used as a proxy for body size.

Temperature was measured *in situ* using CTD. Water samples were kept refrigerated until nutrient concentration (μM) was analysed using a 5-channel Bran and Luebbe segmented flow colorimetric autoanalyser (Norderstedt, Germany) (Woodward & Harris 2021). Dissolved inorganic nitrogen (DIN) is defined as the sum of the nitrate, nitrite, and ammonia concentrations (NO_3, NO_2, and NH_4) and provide an estimate of the generally limiting resource in the L4 station (Smyth *et al.* 2010). Dissolved inorganic nitrogen concentrations below the half saturation constant (k_s) were defined as limiting. As the saturation constant was defined to scale allometrically, each phytoplankton size class exhibited a different limitation threshold. The fraction of weeks where DIN was below k_s was calculated for each phytoplankton size class.

Parameter estimation

Phytoplankton populations were combined in size bins (in $\log_2$) classes from 5 to 19 every two units ($\log_2(\mu m^3\ org^{-1})$) (Segura *et al.* 2011), and for each bin we estimated maximum population abundance every week. Maximum population abundance per week was estimated for an average of 27 species per size-bin.

We employed two complementary approaches to evaluate theoretical predictions on the size scaling of extremes: First, for each size class, a GEV model was fitted GEV using the block maxima

approach (Embrechts *et al.* 2008) by maximum likelihood as implemented in the functions *gev.fit* from the {ismev} package (Port & Stephenson 2018) The relationship of the expected value (E[x]) and variance (Var(x)) with body size was evaluated *a posteriori* by means of linear regression. E[x] and Var(x) were estimated from the location (μ), scale (σ) and shape (ξ) parameters according to the following equations:

$$
\begin{aligned}
&E[x]=\mu+\sigma\,\frac{(g_1-1)}{\xi} && if\quad \xi\neq 0 \quad \xi<1\\
&E[x]=\mu+\sigma\gamma && if\quad \xi=0\\
&Var(x)=\sigma^2\,\frac{(g_2-g_1^2)}{\xi^2} && if\quad \xi\neq 0 \quad \xi<1/2\\
&Var(x)=\sigma^2\pi/6 && if\quad \xi=0
\end{aligned}
\tag{9}
$$

where γ is Euler's constant and $g_k = \Gamma(1- k\xi)$ where Γ is the gamma function.

The second approach to fit GEV models was to unify in a single data frame weekly maximum abundance per body size class, body size, average nutrients, average incident light, and average water temperature as covariates and fit models by maximum likelihood using the function *fevd* from the {extReme} package (Gilleland & Katz 2016). We did not include covariates with significant correlations in the same model in order not to inflate variance (see table 1 for model definition). We fitted a set of nested models: a null model in which there was no dependence of GEV parameters with any covariate (*e.g.*, size, DIN, etc.) and models following eq. 7 including size, nutrient or light and the inverse of temperature effects in the GEV parameters. The inverse of temperature (in Kelvin) was divided by the Boltzmann constant and so the estimate represent the activation energy (E). Size class was used as a categorical covariate when associated to GEV shape parameter to avoid convergence problems. Models that did not converged were not considered.

The use of the two statistical approaches ensured the use of all information, as resource concentration (DIN) and incident light was not available for the whole time series. The data set with resources and temperature were not available for the whole time series, it began in 2000 and ended in 2011. Models were ranked using Akaike information criteria (AIC) and the smalles AIC choosen.

The selected model was inspected for residuals and goodness of fit using standard metrics (quantile-quantile plot and the distribution of residuals).

RESULTS

The total abundance of phytoplankton fluctuated along the year with maxima in summer and a tendency to decrease in winter, while temperature and DIN forms showed marked seasonal cycles (Fig. 2). A significant negative correlation between weekly average incident light and *in situ* measured DIN was found (Pearson's r = -0.75, df = 437, p = $2e^{-16}$). DIN concentration was limiting for all size classes, but the proportion of cases DIN was below k_s increased from the smallest size class to the largest size class (0.37, 0.42, 0.47, 0.52, 0.58, 0.67 and 0.82 from classes 5, 7, 9, 11, 13, 15, and 17 in $\log_2(\mu m^3\ org^{-1})$, respectively). Daily averaged incident light was always higher than the half-saturation constant for light.

GEV models were fitted for each size class. The expected value (E[x]= log(Abundance)) showed a significant negative scaling with the logarithm of size class ($\log_2(M)$). The linear model showed negative slope and a high coefficient of determination (Fig. 3A; $E(x) = 9.68 - 0.82 \log_2(M)$; $R^2 = 0.74$; $p < 0.01$; N = 6)). The 95% confidence interval of the slope was -1.38 to -0.26. The variance showed a humped pattern, increasing from the smallest size classes to the middle sizes classes, and then decreasing (Fig. 3B). The shape parameter (mean[standard error]) was negative in the smallest size classes (ξ = -0.25 [0.01]) and closer to zero in larger size classes, indicating a tendency from Weibull to Gumbell distributions. Only size class 15 showed a positive shape parameter (ξ = 0.066[0.02]) indicative of Frechét distribution.

Under the second statistical approach, the model including the logarithm of size, nutrient concentration, the interaction between both and the inverse of temperature affecting location, the logarithm of size on scale and size and nutrients on shape parameters presented lower AIC with respect to the remaining models (model #0 in Table 1). The selected model showed that the fitted

coefficient representing the scaling of body size was negative (μ_1 = -0.75) and the logrithm of DIN presented a negative effect on location (μ_2 = -0.52) while the interaction between both was positive (μ_3 = 0.017). The inverse of temperature showed a negative scaling (μ_4 = -0.58) close to activation energy. Scale parameter showed a negative dependence with size (σ_1= -0.265). The shape was negative for all (ξ_i < 0) except the larger size class (ξ_i > 0). Residuals from the selected model showed a good fit to the data (Fig. 4). The results were congruent between the two statistical approaches employed.

DISCUSSION

We demonstrated the use of GEV theory to test an explicit hypothesis about the metabolic limits of maximum population density in phytoplankton. Using an extraordinary time series, we explored how metabolic limits to population abundance were shaped by body size and resources. The rigorous test of the metabolic limit to extreme population density using GEV theory opens a field for the analysis of extreme events in natural species-rich populations.

The use of the GEV approach is well developed in many fields, especially in hydrology and finance (Embrechts *et al.* 2008), but it is less explored in ecology (Katz *et al.* 2005; Batt *et al.* 2017). Assuming a Gaussian distribution of errors associated to the maximum abundance of species has been done (Belgrano *et al.* 2002; Cermeño *et al.* 2006), but it can lead to serious bias in the interpretation of results. The Gaussian distribution would lead to systematic bias in the estimation of the return periods, which is the average time between extreme events, an important metric for understanding the probability of occurrence of extreme events. For example, under a Gaussian model for the residuals, the return time of medium-sized phytoplankton organisms (size class 13 to 15 ~ 8000 to 32,700 μm^3 org^{-1}) would yield 6 years to reach an extreme abundance of ~64 cells mL^{-1}. In contrast, the estimated return time using the GEV approximation is 2 years. We remark that this

cell concentration represents a large amount of biomass of primary producers ($5e^5$ to $2e^6$ μm^3 mL^{-1}). Blooms of diatoms of intermediate size fuels aquatic food webs in temperate regions, channelling energy towards larger crustacean predators and ultimately fish, favouring the export of organic carbon (Widdicombe *et al.* 2010). The difference in the estimation of the return period between Gaussian and GEV approaches can have profound consequences for the understanding and managing of aquatic ecosystems, such as the estimation of harmful algal blooms. Appropriate characterization of the distribution of extreme values is therefore critical to forecast bloom formation and to track changes in the patterns induced by global change.

The limits imposed by metabolic restrictions to maximum population density are expected to operate in the dominant species (Lawton 1989). The use of the GEV theory to characterize this upper limit showed excellent performance. Under this framework, we were able to confirm the expected negative scaling of body size with extreme population abundance, supporting the hypothesis of metabolic restrictions, as found previously (Deng *et al.* 2012; Barneche *et al.* 2016). Moreover, the explicit inclusion of resources was instrumental to capture deviations from the expected size-abundance scaling value when resource are limiting (Enquist *et al.* 2007). Resource limitation could help to explain the otherwise contradictory evidence on size-density scaling presented in previous studies (e.g. Isaac *et al.* 2011). Deviations from the linear fit (Fig. 3) showed peaks coincident with the position of clumps detected for the planktonic community in this ecosystem (Segura *et al.* 2011), which suggests that other ecological forces are able to mould the pattern that arises from metabolic restrictions. Including the effects of ecological mechanisms and the role of resource co-limitation on the theoretical limits to population abundance is an interesting avenue to explore. Similarly, traits other than size (e.g. Silica wall in diatoms) can help to explain differences from the expected value caused by the limitation caused by other resources.

The GEV theory can provide further insight into the dynamics of natural populations. The Weibull distribution is the result of taking the maxima of a random variate that follows a bounded

distribution. The negative shape parameter representative of a Weibull distribution in small-sized populations is indicative of a strong population control of small sized organisms subjected to strong predator-prey interactions within the microbial food web (Calbet & Landry 2004) and consistent with a previous analysis of population fluctuations (Segura *et al.* 2013). However, medium-sized organisms showed higher variance (dictated by larger value of the scale parameter) and a shape parameter value close to zero, indicative of a Gumbel distribution. A Gumbel distribution in the logarithm of abundance ($log(N)$) implies that population abundance (N) follows a *Fréchet* distribution, which is a fat tail distribution consistent to previous observations (Allen *et al.* 2001; Segura *et al.* 2013) and recent theoretical arguments (Segura & Perera 2019). This change in variance is mostly explained by the effect of the allometric scaling of the half-saturation constant, which exerts a stronger resource limitation to large-sized organisms when resources are limiting, but allows for a large increase in the upper limit when resources are not limiting. The scope for variability registered in medium- and large-sized organisms allows for outburst of abundance, thereby allowing escape from predatory control (Irigoien *et al.* 2005). The increase of abundance in medium-sized species fuels the herbivorous food web, where primary producers are fed upon by metazoan plankton consumed by large invertebrate or vertebrate predators (*e.g.*, fish larvae). This phenomenon has been described in temperate latitudes (Kiørboe 1993; Mann & Lazier 2006) and the English Channel (Widdicombe *et al.* 2010). Exploring these patterns beyond aquatic ecosystems could help to shed light on general principles regarding the size-structure and dynamics of natural populations.

Present results suggest nutrients and body size are two fundamental variables defining structure and variability of extreme abundance in the phytoplankton populations, and the most parsimonious model included size and nutrients but did not include temperature nor incident light (PAR; Table 1). However, the effect of light and temperature could not be discarded completely. First, there are significant correlations between the studied environmental variables (DIN, PAR, Temperature) that

render it difficult to tease out specific effects under this correlative approach. Experimental designs (*in vitro* or model-based) to isolate the effect of each variable could help delineate specific responses. The use of an invariant half-saturation constant for light dependence (López-Urrutia *et al.* 2006) omit possible size dependences of light acquisition processes (Finkel *et al.* 2004; Mei *et al.* 2009; Key *et al.* 2010). Theoretical size dependence of light acquisition (e.g. Mei *et al.* 2009) could be explicitly included in future model formulations to derive and test explicit predictions. Incident light reflects the amount of radiation reaching the water surface, while an underwater light environment is dependent on turbidity and depth of the thermocline which were not considered here (Huisman *et al.* 1999). Temperature is a key driver of metabolic processes (Brown *et al.* 2004), but it was not included in the selected model. Activation energy in primary producers is generally lower than in heterotrophic organisms (López-Urrutia *et al.* 2006; Segura *et al.* 2018), which implies a reduced impact on metabolic rates, together with the relatively cold and stable water temperature registered in the English Channel L4 station (T = 9–19 °C). The indirect effects of temperature may be more pronounced than its direct effect on metabolism. The changes in resource supply to the euphotic layer mediated by strong stratification or changes in the light landscape by deepening of the thermocline must be characterized. We showed that resource limitation is common for phytoplankton in the English Channel a pattern which could be greatly altered by human behaviour. Changes in thermal stratification and resource inflow to coastal zones caused by extreme weather effects (Smyth *et al.* 2010) are expected to increase in magnitude and frequency (Stockwell *et al.* 2020), with marked effects in extreme population abundances.

In summary we tested explicit predictions on the scaling of the distribution of extreme population abundances of phytoplankton linking a well developed general extreme value theory with a metabolic resource limited model. We showed a marked effect of body size and resources in extreme population abundance and introduce a robust approach to explore community dynamics in natural species-rich ecosystems.

Acknowledgements

We thank Plymouth Marine Laboratory (PML) for generating and kindly providing data. We would like to thank Eric Gilleland for advise in fitting extreme models and suggestions in a previous version of the manuscript. We would like FVF for a grant to AMS and ANII for a grant to AMS (code FCE_3_2020_1_162710). SNI and PEDECIBA is also acknowledged for partial financial support. We also specially thank two reviewers for their constructive suggestions which greatly improved the manuscript.

REFERENCES

Agustí, S., Duarte, C. & Kalff, J. (1987). Algal cell size and the maximum density and biomass of phytoplankton. *Limnol. Oceanogr.*, 32, 983–986.

Aksnes, D.L. & Egge, J.K. (1991). A theoretical model for nutrient uptake in phytoplankton. *Marine Ecology Progress Series*, 70, 65–72.

Allen, A.P., Li, B.-L. & Charnov, E.L. (2001). Population fluctuations, power laws and mixtures of lognormal distributions. *Ecology Letters*, 4, 1–3.

Barneche, D.R., Kulbicki, M., Floeter, S.R., Friedlander, A.M. & Allen, A.P. (2016). Energetic and ecological constraints on population density of reef fishes. *Proceedings of the Royal Society B: Biological Sciences*, 283, 20152186.

Batt, R.D., Carpenter, S.R. & Ives, A.R. (2017). Extreme events in lake ecosystem time series: Extreme events in lake ecosystem time series. *Limnol. Oceanogr.*, 2, 63–69.

Belgrano, A., Allen, A.P., Enquist, B.J. & Gillooly, J.F. (2002). Allometric scaling of maximum population density: a common rule for marine phytoplankton and terrestrial plants. *Ecology letters*, 5, 611:613.

Bjornstad, O.N. (2001). Noisy Clockwork: Time Series Analysis of Population Fluctuations in Animals. *Science*, 293, 638–643.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M. & West, G.B. (2004). Toward a metabolic theory of ecology. *Ecology*, 85, 1771–1789.

Calbet, A. & Landry, M.R. (2004). Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. *Limnol. Oceanogr.*, 49, 51–57.

Cermeño, P., Marañón, E., Harbour, D. & Harris, R.P. (2006). Invariant scaling of phytoplankton abundance and cell size in contrasting marine environments. *Ecology Letters*.

Deng, J., Zuo, W., Wang, Z., Fan, Z., Ji, M., Wang, G., *et al.* (2012). Insights into plant size-density relationships from models and agricultural crops. *Proceedings of the National Academy of Sciences*, 109, 8600–8605.

Edwards, K.F., Thomas, M.K., Klausmeier, C.A. & Litchman, E. (2012). Allometric scaling and taxonomic variation in nutrient utilization traits and maximum growth rate of phytoplankton. *Limnology and Oceanography*, 57, 554–566.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (2008). *Modelling extremal events: for insurance and finance*. Stochastic modelling and probability. 4. corr. printing and 8. printing. Berlin.

Enquist, B.J., Kerkhkoff, A.J., Stark, S.C., Swenson, N.G., McCarthy, M.C. & Price, C.A. (2007). A general integrative model for scaling plant growth, carbon flux, and functional trait spectra. *Ecology Letters*, 449.

Finkel, Z.V., Irwin, A.J. & Schofield, O.M.E. (2004). Resource limitation alters the 3⁄4 size scaling of metabolic rates in phytoplankton. *Marine ecology progress series*, 273, 269–279.

Gaines, S.D. & Denny, M.W. (1993). The Largest, Smallest, Highest, Lowest, Longest, and Shortest: Extremes in Ecology. *Ecology*, 74, 1677–1692.

Gilleland, E. & Katz, R.W. (2016). extRemes 2.0: An Extreme Value Analysis Package in R. *Journal of Statistical Software*, 72, 1–39.

Hatton, I.A., Dobson, A.P., Storch, D., Galbraith, E.D. & Loreau, M. (2019). Linking scaling laws across eukaryotes. *Proc Natl Acad Sci USA*, 116, 21616–21622.

Huisman, J., Jonker, R.R., Zonnevel, C. & Weissing, J. (1999). Competition for light between phytoplankton species: experimental test of mechanistic theory. *Ecology*, 80, 211–222.

Irigoien, X., Flynn, K. & Harris, R. (2005). Phytoplankton blooms: a 'loophole' in microzooplankton grazing impact? *Journal of Plankton Research*, 27, 313–321.

Isaac, N.J.B., Storch, D. & Carbone, C. (2011). Taxonomic variation in size-density relationships challenges the notion of energy equivalence. *Biology Letters*, 7, 615–618.

Katz, R.W., Brush, G.S. & Parlange, M.B. (2005). Statistics of extremes: modeling ecological disturbances. *Ecology*, 86, 1124–1134.

Keitt, T.H. & Stanley, H.E. (1998). Dynamics of North American breeding bird populations. *Nature*, 393, 257–260.

Key, T., McCarthy, A., Campbell, D.A., Six, C., Roy, S. & Finkel, Z.V. (2010). Cell size trade-offs govern light exploitation strategies in marine phytoplankton. *Environmental Microbiology*, 12, 95–104.

Kiørboe, T. (1993). Turbulence, Phytoplankton Cell Size, and the Structure of Pelagic Food Webs. In: *Advances in Marine Biology*. pp. 1–72.

Lawton, J.H. (1989). What Is the Relationship between Population Density and Body Size in Animals? *Oikos*, 55, 429.

López-Urrutia, A., San Martin, E., Harris, R.P. & Irigoien, X. (2006). Scaling the metabolic balance of the oceans. *Proccedings of the National Academy of Sciences of the United States of America*, 103, 8739–8744.

Mann, K.H. & Lazier, J.R.N. (2006). *Dynamics Of Marine Ecosystems*.

Marquet, P.A. (2005). Scaling and power-laws in ecological systems. *Journal of Experimental Biology*, 208, 1749–1769.

May, R.M. (1973). *Stability and complexity in model ecosystems*. Princeton Landmarks in Biology. 2nd edn. London.

Mei, Z.-P., Finkel, Z.V. & Irwin, A.J. (2009). Light and nutrient availability affect the size-scaling of growth in phytoplankton. *Journal of Theoretical Biology*, 259, 582–588.

Port, O.S. functions written by J.E.H. with R. & Stephenson, R. documentation provided by A.G. (2018). *ismev: An Introduction to Statistical Modeling of Extreme Values*.

Segura, A.M., Calliari, D., Kruk, C., Conde, D., Bonilla, S. & Fort, H. (2011). Emergent neutrality drives phytoplankton species coexistence. *Proccedings of the Royal Society B*, 278, 2355–2361.

Segura, A.M. & Perera, G. (2019). The Metabolic Basis of Fat Tail Distributions in Populations and Community Fluctuations. *Frontiers in Ecology and Evolution*.

Segura, A.M., Sarthou, F. & Kruk, C. (2018). Morphology-based differences in the thermal response of freshwater phytoplankton. *Biology Letters*.

Segura, Calliari, D., Fort, H. & Lan, B.L. (2013). Fat tails in marine microbial population fluctuations. *Oikos*, 122, 1739–1745.

Smyth, T.J., Fishwick, J.R., AL-Moosawi, L., Cummings, D.G., Harris, C., Kitidis, V., *et al.* (2010). A broad spatio-temporal view of the Western English Channel observatory. *Journal of Plankton Research*, 32, 585–601.

Stockwell, J.D., Doubek, J.P., Adrian, R., Anneville, O., Carey, C.C., Carvalho, L., *et al.* (2020). Storm impacts on phytoplankton community dynamics in lakes. *Glob Change Biol*, 26, 2756–2784.

Widdicombe, C.E., Eloire, D., Harbour, D., Harris, R.P. & Somerfield, P.J. (2010). Long-term phytoplankton community dynamics in the Western English Channel. *Journal of plankton research*, 32, 643–655.

Widdicombe, C.E. & Harbour, D. (2021). Phytoplankton taxonomic abundance and biomass time-series at Plymouth Station L4 in the Western English Channel, 1992-2020.

Woodward, E.M.S. & Harris, C. (2021). Micromolar Nutrient concentration profiles from the long term time series at Station L4 in the Western English Channel from 2000 to 2020.

Zaoli, S., Giometto, A., Maritan, A. & Rinaldo, A. (2017). Covariations in ecological scaling laws fostered by community dynamics. *Proceedings of the National Academy of Sciences*, 114, 10672–10677.

Table 1.- Generalized Extreme Value (GEV) distribution models fitted to phytoplankton maximum population density in the English channel L4 station. Body size (logM) was fitted as a continuous covariate and also as a categorical covariate (classM) where subscript *i* denotes a specific parameter for each size class. The model following eq. 7 including the logarithm of Disolved Nitrogen Concentration (logDIN) and Incident light (PAR) improved model fitting as evidenced by the lowest AIC (ΔAIC) and did not showed convergence problems. The "interaction" between the two additive variables in the location parameter is showed (interaction).

#	Location (μ)	Scale (σ)	Shape (ξ)	ΔAIC
0	$\mu+\mu_1$logM$+\mu_2$logDIN$+\mu_3$interaction$+\mu_4$invTemp	$\sigma + \sigma_1$ logM	$\xi+\xi_2$ logDIN$+ \xi_i$ classM	0
1	μ	σ	ξ	2021
2	$\mu+\mu_1$logM	$\sigma + \sigma_1$ logM	ξ	619
3	$\mu+\mu_1$logM	$\sigma + \sigma_1$ logM	ξ	376
4	$\mu+\mu_1$logM	$\sigma + \sigma_1$ logM	$\xi + \xi_i$ classM	168
5	$\mu+\mu_1$logM$+\mu_2$logDIN$+\mu_3$interaction	$\sigma + \sigma_1$ logM	$\xi+\xi_2$ logDIN$+ \xi_i$ classM	13
6	$\mu+\mu_1$logM$+\mu_2$invTemp	$\sigma + \sigma_1$ logM	$\xi+ \xi_i$ classM	98
7	$\mu+\mu_1$logM$+\mu_2$logDIN$+\mu_3$interaction	$\sigma + \sigma_1$ invTemp	$\xi+ \xi_i$ classM	431

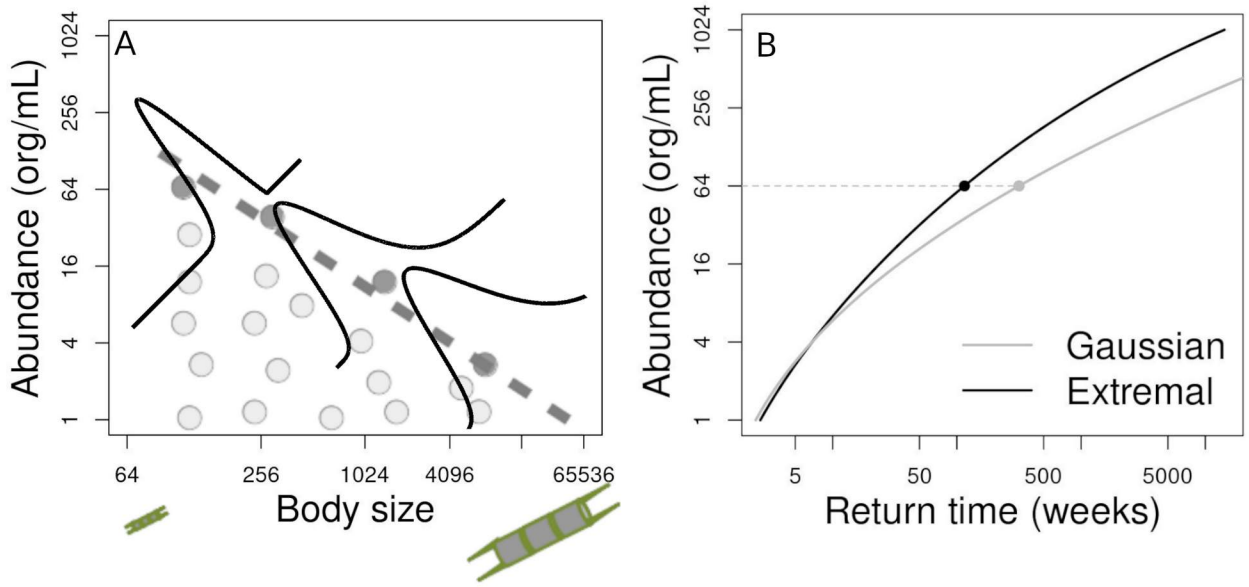

Figure 1.- Schematic representation of the A) Size abundance distribution showing two patterns: i) the expected negative relationship (grey dashed line with negative slope) and ii) the generalized extremal distribution (GEV) expected in each size class for the residuals of taking the weekly maxima (darkgrey dots) of species abundance (grey dots). B) the average time to observe (*i.e.* return time) an extreme event differ importantly between using the appropriate GEV distribution and assuming a Gaussian approximation. Note axis are in logarithmic scale.

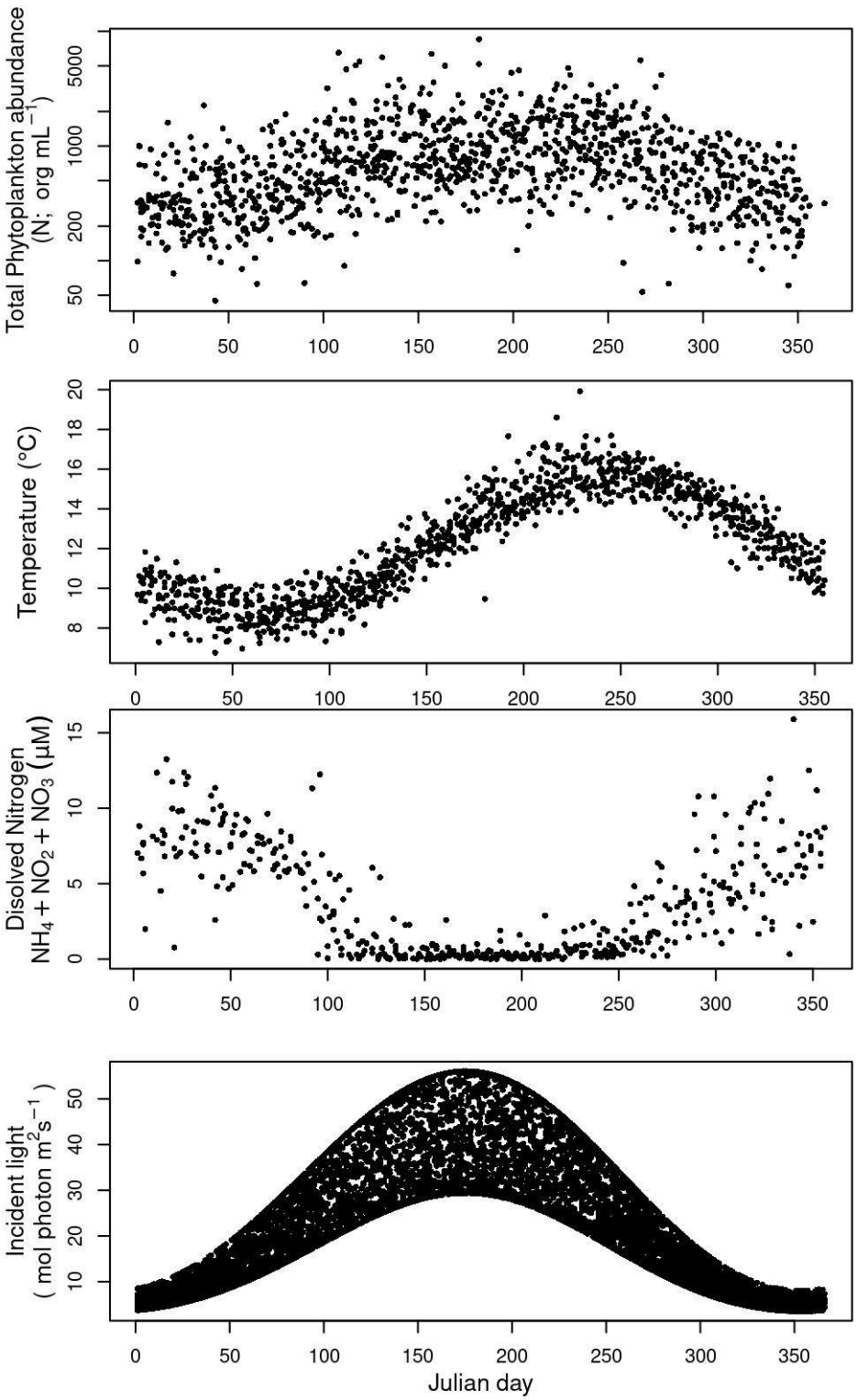

Figure 2.- Temporal dynamics of phytoplankton abundance, temperature, nutrients and incident light measured in the English channel L4 station from 1992 to 2009.

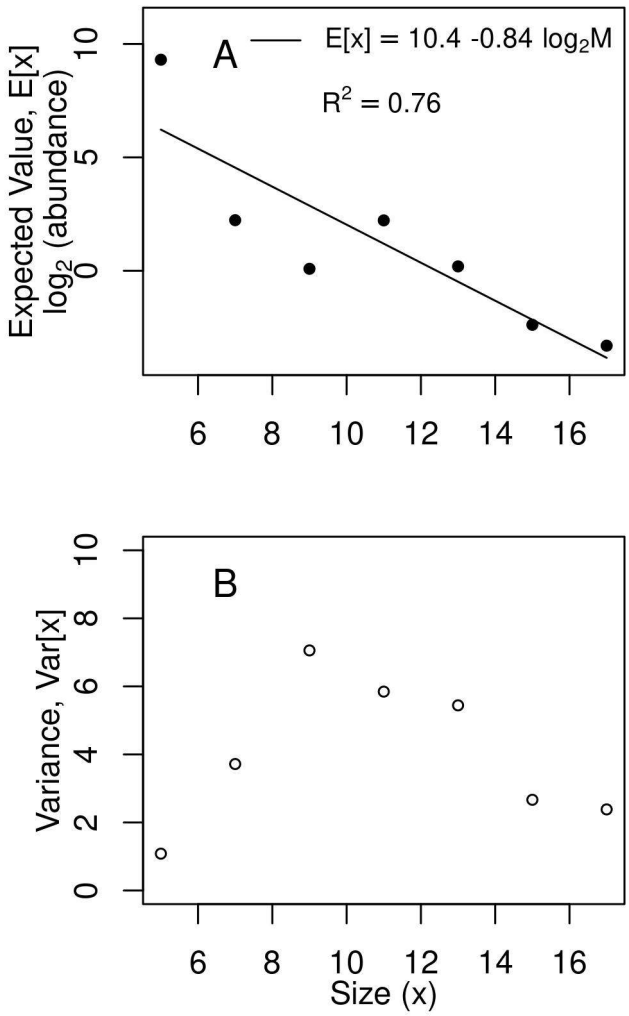

Figure 3.- Size scaling of the A) expected value (E[*x*]) and B) the variance (Var[*x*]) of the fitted GEVs distributions. The value of the negative relationship (Average [95%CI]= -0.84 [-1.38–0.30]) between size and the expected value supports the theoretical predictions of a negative scaling related to metabolism. Phytoplankton size classes are in logarithmic classes (in $\log_2$).

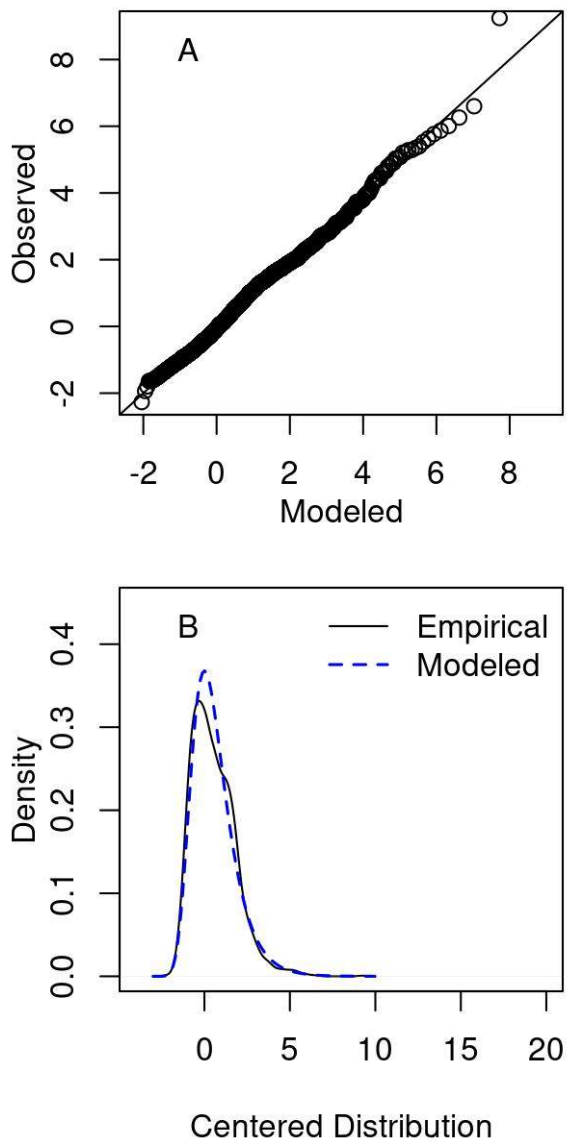

Figure 4.- Summary plots for the best fitted GEV model (#0 in Table 1 in agreement with theoretical eq. 7) with size, resources and the inverse of temperature as covariates to explain maximum population abundance data in the L4 station in the English channel. A) quantile-quantile plot and B) residual density in the observed and modelled data.