Seed-bank contribution in maintaining phytoplanktonic biodiversity

April 27, 2020

Growth calibration

This section is dedicated to the design of the growth function used in the first step of the models, to the phenology parameters meant to characterize the seasonal dynamics of the morphotypes we aim to mimick, and to the interactions that may reduce or enhance this growth.

Growth response to the environment

(need to rephrase entirely the parts belown beginning by Growth rate: variation with temperature)

- intrinsic variability and need to take into account the effect of temperature

- scranton and vasseur proposition (decomposition in niche and metabolism, this part being separated into Eppley and the Metabolic Theory of Ecology) and its failure to represent values previously found in the literature

- introduction of the Bissinger equation: maybe put here the "issues" we have previously detected (i.e. variation in maximum growth rate due to the width of the niche)

Extraction of taxon phenology from field data

In eq. **??**, the niche part of the growth rate is mainly defined by two parameters: the thermal optimum T_i^{opt} and a proxy of its niche width b_i, which drive the phenology of the taxon. Each year, the dynamics of phytoplanktonic organisms is characterized by a blooming period and a lower concentration during the rest of the year. The bloom can be triggered by a combination of nutrient and light input, as well as a sufficient temperature. All parameters being more or less dependent on seasonality, it is reasonable to restrain this study to the effect of temperature.

We base estimates of T_i^{opt} and b_i on field observations. For each taxon and each year, the beginning of the bloom of a given taxon is defined by the date at which its abundance exceeds its median abundance over the year. The duration of the bloom is the number of days between the beginning and the date where abundance falls below the median value. Taxa are then separated into two groups. In the field, generalists are characterized by one long bloom in the year or several blooms when the abundances oscillate around their median. Specialists tend to appear only once or twice in the year for shorter amounts of time. A genus is therefore defined as a generalist if its cumulated blooms last more than the average duration of all blooms (137 days) for at least 15 years over the 20 years of the time series.

In the model, we assume that generalists have a niche width between 15 and 30°C and specialists, between 5 and 10°C. We assume that temperatures outside of this range leads to a growth rate at least 10 times inferior to the growth rate obtained at their thermal optimum. This leads to values of b_i between 180 and 1500 for generalists, and 7 and 55 for specialists. A set of b values is drawn from a uniform distribution with these boundaries. Meanwhile,

taxa are ordered as a function of the mean sum of their bloom length, i.e. $\sum \overline{L}_{i,b} > \sum L_{j,b} \Rightarrow A_i > A_j$ where $\overline{L}$ is the mean over 20 years of the annual cumulated lengths of the bloom. Larger niche values are attributed to longer mean bloom length.

The thermal optimum T_i^{opt} was first defined as the mean temperature at the beginning of the first annual bloom throughout the whole time series. However, this value led to blooms occuring mainly in the winter and needed to be increased by 5°C in order to simulate realistic phytoplankton cycles.

Interaction computation

Fixed parameters

This section contains additional information of parameter definition, estimation and value.

Loss rate The loss rate corresponds to multiple mortality processes. Scranton & Vasseur (2016) set this rate around 0.04 day^{-1}. In Jewson *et al.* (1981), washout (0.5%), parasitism (4% of infested cells) and grazing still remained low when compared to growth rates. Li *et al.* (2000) found values between 0.02 and 0.1 for natural mortality only, while a review by Sarthou *et al.* (2005) indicates a loss of daily primary productivity between 0.45 and 1.1 due to grazing only, 0.13 being potentially due to cell autolysis (in the absence of nutrients, or because of viral charge). From this overview, the reference values for the models, 0.2, is a trade-off between sometimes high level of mortality and the maintenance of the model.

Sinking rate Among the hydrodynamics processes that drive sinking rate, turbulence and eddies can retain the cells at the top of the water column, themselves driven by tidal currents, shape of the estuary or wind conditions. This is why lab experiments on sinking rates are not sufficient to calibrate a field-based model. Credible values for sinking rates are therefore taken from field studies. In the Gotland Basin (central Baltic Sea), Passow (1991) measured a large variability in sinking rates, even within the same genus (e.g., between 1 and 30% for *Chaetoceros* spp.). However, a pattern could be distinguished, with a small number of genera that sinked more than the rest of the community. The mean rate for *Chaetoceros* and *Thalassiosira* was around 10% while it was around 1% for the other species. Sinking rate values around 10% are consistent with the loss rate values in Kowe *et al.* (1998) in a river and Wiedmann *et al.* (2016) in an estuary (mouth of Adventfjorden). The chosen beta-distribution (Fig. S1) therefore accounts for both maximum and mean values, while still allowing a highly skewed distributions. High sinking rates are kept for the morphotypes corresponding to *Chaetoceros* (CHA) and *Thalassiosira* (THP). Variations of the sinking rate over time give values between 4 and 50% (Jewson *et al.*, 1981).

However, it should be noted that the sinking rate measured in these articles represent the sinking of all cells, therefore including dead cells. In our model, it should represent the exchange of live cells between the coast and the seed bank, or a loss of cells in the ocean. The maximum sinking rate used in this model might therefore need to be decreased to avoid modeling dead cells already taken into account in the loss term.

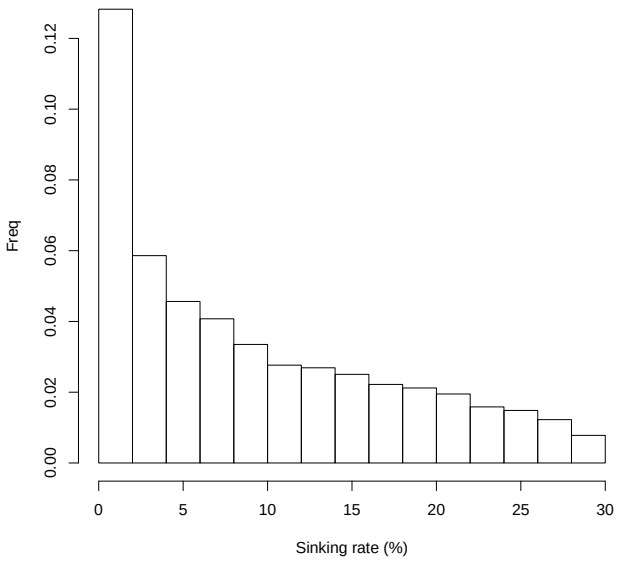

Figure 1: Possible beta-distribution of sinking rates

Cyst mortality McQuoid *et al.* (2002) present maximum and mean depth at which germination of diatoms and dinoflagellates could still occur when incubated. The authors also present sediment datation according to depth. Depth can therefore be related to maximum and mean age of phytoplankton cysts before death.

Assuming m is the probability of mortality, m follows a geometric law, i.e., m is the probability distribution of the number of days needed for a phytoplankon cyst to die. The expectancy for the life duration (the number of days without dying) is $\frac{1}{m} \Leftrightarrow m = \frac{1}{L_{mean}}$where L_{mean} is the average life duration.

Another way to look at the process is that life expectancy L follows the distribution $p(L > l) = e^{-ml}$.With maximum values, we can arbitrarily choose that for these values $p(L > l_{max}) = 0.05$. In this, $m = -\frac{ln(p(L>l_{max}))}{l_{max}}$.

In both cases, $m \propto 10^{-4}\mathrm{d}^{-1}$.

Resuspension As mentioned in the main text, resuspension values are mostly taken from model or data for inorganic particles. Rates vary from one publication to another: in Fransz & Verhagen (1985) in a coastal area, the resuspension rate of sediments is evaluated around 5.10^{-5} day^{-1} in winter and decreases in summer, with a link between resuspension and light extinction coefficient. In Kowe *et al.* (1998), the resuspension rate of diatoms is evaluated around $1.9.10^{-5}$ day^{-1}. In Le Pape *et al.* (1999), resuspension rate of sediments and dead diatoms is 0.002 day^{-1}. In this paper, we explore values between 10^{-5} (stratified water column) to 0.1 (highly mixed environment).

Finally, it should be noted that cyst burial, sinking rate and resuspension are all dependent on the hydrodynamics of the place and are therefore, at least in biological terms, not identifiable.

[[[[END OF REFORMULATION]]]]

Growth rate: variation with temperature

Phytoplanktonic growth rates are highly variable, in situ or in experimental conditions. An example of such variability appears in Balzano *et al.* (2011). For ten strains of one genus only (*Skeletonema*), and in the same experimental conditions, Balzano *et al.* (2011) have been able to detect growth rates between 0.5 and 1.25 day^{-1},

which corresponds more generally to the values found in the literature (between 0.2 and 1.78 for diatoms in Reynolds (2006), even reaching 3 in the meta-analysis of 308 experiments by Edwards *et al.* (2015); this can be much lower for dinoflagellates). These growth rates are maximum, fixed values for isolated species in laboratory conditions. Most of the time, they correspond to fixed temperature conditions, or to only a small set of values. These observations therefore cannot accomodate realistic, seasonal environment.

In this context, Bissinger *et al.* (2008) based their study on a seminal work by Eppley (1972) to compute the maximum possible growth rate depending on the temperature. The relationship between temperature and maximum growth rate, evaluated on a large database, is then $r(T) = 0.81\mathrm{e}^{0.0631T}$,with T in °C. In this case, growth rates vary between 0.81 and 3.9 day^{-1} for temperatures between 0 and 25°C, in line with previous observations. However, these values only illustrate a maximum, exponential growth which cannot be realistic for species with different niche temperatures.

The equation of Scranton & Vasseur (2016) distinguishes different niches based on optimal temperature. However, it needed to be modified to be closer to the shape in other works (Eppley, 1972; Bissinger *et al.*, 2008; Edwards *et al.*, 2016).

Figure 2: Comparison of growth rate formula. The SV solid grey line corresponds to the original Scranton and Vasseur (2016) formula. The solid black line includes both a new niche area and a gain in growth rate, and the shape of the growth rate as a function of temperature is given for a low (blue) and high (red) values of thermal optimum with dotted lines. The Bissinger and Eppley lines correspond to the formula of maximum growth rates in the studies from eponymous authors. Finally, horizontal lines show limits taken in the literature.

Growth rate: generalists vs specialists

From eq. **??**, let us note $r_i(T) = a_r(\tau_0)e^{E_r \frac{(T-\tau_0)}{kT\tau_0}} f_i(T) = E(T)f_i(T)$. Hereafter, we will call $f_i(T)$ the niche part of the equation and $E(T)$ the metabolism part. The metabolism part is driven by both metabolism theory ecology and by observations in Eppley (1972) study of maximum growth rates for phytoplankton.

The simplest way to define generalists and specialists is to remove the constraint on the fixed niche area A. In this case, the only parameter that defines the niche width is b. However, increasing b does not only lead to increases in niche width, but also to a shift towards larger values to attain the maximum growth rate. This means that for the same theoretical T_i^{opt}, the observed thermal optimum is higher.

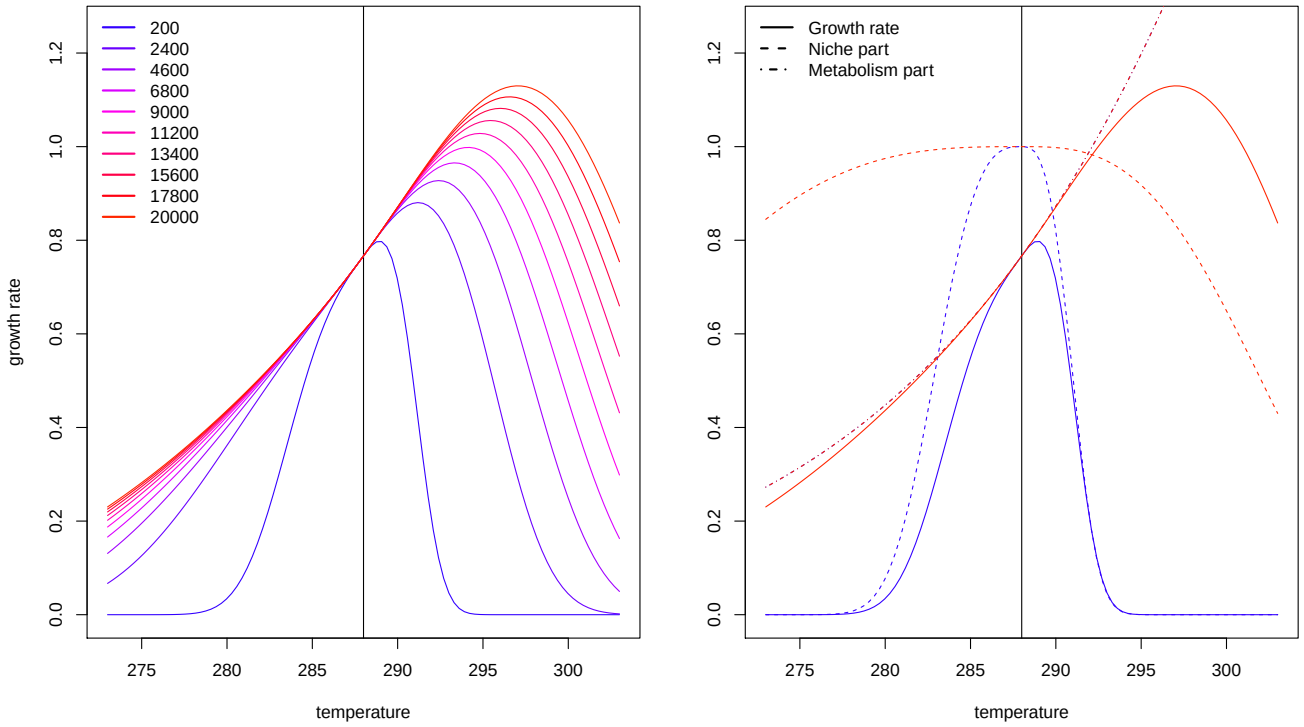

Figure 3: Relationship between daily growth rates and temperature with different values of niche width b (whose values are indicated in the legend) and the same thermal optimum, 15°C, indicated by the solid black. In the left panel, only the two extreme values of b (200 and 20 000) are shown in blue and red respectively. Coloured lines then correspond to the final growth rate, dashed lines correspond to $f_i(T)$ values (see eq. **??**) and dotted lines correspond to $E(T)$ values.

When b increases, the niche term f_i shows smaller variations in values around the thermal optimum and the final value of the growth rate is driven by the metabolism part of the equation. This can lead to confusions on the observed thermal optimum that we wish to obtain.

Even if we only keep the maximum growth rate computed in Bissinger *et al.* (2008) (as an update of Eppley's curve), we obtain slightly higher growth rate values but the same shift.

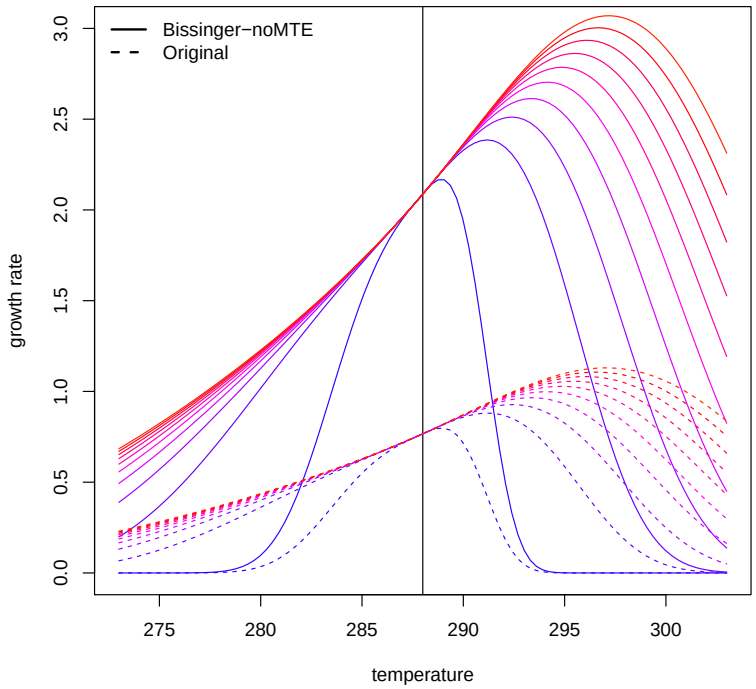

Figure 4: Relationship between daily growth rates and temperature with different values of niche width b (legend for the values of b is give in the figure above). Dashed lines correspond to the original formulation of Scranton & Vasseur (2016) while solid lines correspond to a formulation of the temperature-dependent maximum growth computed in Bissinger *et al.* (2008).

Community matrix: correspondence between Multivariate Autoregressive and Beverton-Holt models

Certain *et al.* (2018) showed that MAR and BH interaction coefficients, respectively b_{ij} and α_{ij}, could map once abundances at equilibrium N_i^* are defined.

$$\begin{cases} b_{ii} - 1 = & \frac{-\alpha_{ii} N_i^*}{1+\sum_l \alpha_{ij} N_l^*} \\ b_{ij,i\neq j} = & \frac{-\alpha_{ij} N_j^*}{1+\sum_l \alpha_{il} N_l^*} \end{cases}$$

Let's define $\tilde{b_{ij}}$ with $\tilde{b_{ii}} = b_{ii} - 1$, and $f_A(i) = \sum_l \alpha_{ij} N_l^*$.

$$b_{ij}(1 + f_A(i)) = -\alpha_{ij} N_j^*$$

We then sum on columns (on j).

$$\sum_j [b_{ij}(1 + f_A(i))] = -f_A(i)$$

$$\Leftrightarrow -f_A(i)(1 + \sum_j b_{ij}) = \sum_j b_{ij}$$

$$\Leftrightarrow f_A(i) = -\frac{\sum_j b_{ij}}{(1 + \sum_j b_{ij})}$$

$$\Leftrightarrow \alpha_{ij} = -\frac{1}{N_j^*} b_{ij} (1 - \frac{\sum_j b_{ij}}{1 + \sum_j b_{ij}})$$

$$\Leftrightarrow \alpha_{ij} = -\frac{1}{N_j^*} \frac{b_{ij}}{1 + \sum_j b_{ij}}$$

This gives an exact correspondance between α_{ij} and b_{ij}.

Quadratic programming

Maynard *et al.* (2019) recently applied quadratic programming (Bazaraa *et al.*, 2013) to ecological data, which improved the calibration of certain parameters and led to more realistic simulations. Even though we could directly use the values obtained in the literature and in previous studies (Picoche & Barraquand, 2020), the switch from a fifteen-day MAR model to a daily BH-model and the high uncertainty for all parameters due to the use of proxies, and high variability in phytoplanktonic genera on their own, are motivations to calibrate the model more precisely. Following the example of Maynard *et al.* (2019) with the package limSolve in R [ref, version], we intend[1] to apply quadratic programming to our interaction matrices. Contrary to Maynard *et al.* (2019), we do not broaden the calibration process to growth rates as ours are not fixed values in a linear model.

The quadratic programming algorithm aims at finding $\mathbf{x}$ that minimizes $||\mathbf{Cx} - \mathbf{d}||^2$ under the constraints $\mathbf{Ex} = \mathbf{f}$ and $\mathbf{Gx} >= \mathbf{h}$.

Here, $\mathbf{C} = \mathbf{I}$, $\mathbf{d} = [\text{vec}(\mathbf{A}^T)^\frown \mathbf{r}']$ where $\mathbf{A}$ is the interaction matrix, $\mathbf{r}' = -(\mathrm{e}^{\mathbf{r}} - \mathbf{1})$ is the vector of growth rates, $\mathbf{E}$ is built so that we verify the equality $\mathbf{AN}^* + \mathbf{r}' = 0$ where $\mathbf{N}^*$is the vector of abundance at equilibrium (more precisely, here, average abundance values over the whole time series), and $\mathbf{G}, \mathbf{h}$ so that $\mathbf{r} > 0$ (genera have a positive growth rate when taken in isolation) and $\forall i, a_{ii} > 0$ (negative density-dependence, individuals from the same genus compete with each other).

NOTE: This may not be useful as only the mean growth rate can be used to adjust the model (due to its dependence on the temperature), and the calibration of the interaction matrix tends to lead to a stable community, that is not able to represent the yearly cycles due to seasonality.

Calibration

In addition to quadratric programming which aims to get closer to observed average abundances, calibration is necessary to approximate other features of the system. Several indicators can be used:

- species-abundance distribution within each month/each semester/each season (see SAD in output/modelev1.1/)
- phenology can be checked with the date of the beginning of the bloom for each year (and its duration, just to check that it doesn't last the whole year)
- dynamics themselves can be used. To avoid artefacts created by average abundances over the year, an idea would be stabilize the model on synthetic temperatures, then use the actual observed temperature to observe the simulated dynamics year after year.

In all cases, a simple measure of |obs-sim| might be enough. Calibration would then be a multi-criterion process.

[1] I use this term because I am really not sure this is what we want to keep. I need to know how this method can be applied to cycles. You mean, whether it can handle something else than a fixed point equilibrium? I'm pretty sure it can.

References

Balzano, S., Sarno, D. & Kooistra, W.H.C.F. (2011). Effects of salinity on the growth rate and morphology of ten Skeletonema strains. *Journal of Plankton Research*, 33, 937–945.

Bazaraa, M.S., Sherali, H.D. & Shetty, C.M. (2013). *Nonlinear programming: theory and algorithms*. John Wiley & Sons.

Bissinger, J., Montagnes, D., Harples, J. & Atkinson, D. (2008). Predicting marine phytoplankton maximum growth rates from temperature: Improving on the Eppley curve using quantile regression. *Limnology and Oceanography*, 53, 487–493.

Certain, G., Barraquand, F. & Gåardmark, A. (2018). How do MAR(1) models cope with hidden nonlinearities in ecological dynamics? *Methods in Ecology and Evolution*, 9, 1975–1995.

Edwards, K., Thomas, M., Klausmeier, C. & Litchman, E. (2015). Light and growth in marine phytoplankton: allometric, taxonomic, and environmental variation. *Limnology and Oceanography*, 60, 540–552.

Edwards, K., Thomas, M., Klausmeier, C. & Litchman, E. (2016). Phytoplankton growth and the interaction of light and temperature: A synthesis at the species and community level. *Limnology and Oceanography*, 61, 1232–1244.

Eppley, R. (1972). Temperature and phytoplankton growth in the sea. *Fishery Bulletin*, 70, 1063–1085.

Fransz, H. & Verhagen, J. (1985). Modelling research on the production cycle of phytoplankton in the Southern Bight of the North Sea in relation to riverborne nutrient loads. *Netherlands Journal of Sea Research*, 19, 241–250.

Jewson, D.H., Rippey, B.H. & Gilmore, W.K. (1981). Loss rates from sedimentation, parasitism, and grazing during the growth, nutrient limitation, and dormancy of a diatom crop. *Limnology and Oceanography*, 26, 1045–1056.

Kowe, R., Skidmore, R., Whitton, B. & Pinder, A. (1998). Modelling phytoplankton dynamics in the River Swale, an upland river in NE England. *Science of The Total Environment*, 210, 535–546.

Le Pape, O., Jean, F. & Ménesguen, A. (1999). Pelagic and benthic trophic chain coupling in a semi-enclosed coastal system, the Bay of Brest (France): a modelling approach. *Marine Ecology Progress Series*, 189, 135–147.

Li, M., Gargett, A. & Denman, K. (2000). What determines seasonal and interannual variability of phytoplankton and zooplankton in strongly estuarine systems? *Estuarine, Coastal and Shelf Science*, 50, 467–488.

Maynard, D.S., Wootton, J.T., Serván, C.A. & Allesina, S. (2019). Reconciling empirical interactions and species coexistence. *Ecology Letters*, 22, 1028–1037.

McQuoid, M.R., Godhe, A. & Nordberg, K. (2002). Viability of phytoplankton resting stages in the sediments of a coastal Swedish fjord. *European Journal Phycology*, 37, 191–201.

Passow, U. (1991). Species-specific sedimentation and sinking velocities of diatoms. *Marine Biology*, 108, 449–455.

Picoche, C. & Barraquand, F. (2020). Strong self-regulation and widespread facilitative interactions between genera of phytoplankton. preprint, bioRxiv.

Reynolds, C.S. (2006). *The ecology of phytoplankton*. Cambridge University Press.

Sarthou, G., Timmermans, K.R., Blain, S. & Tréguer, P. (2005). Growth physiology and fate of diatoms in the ocean: a review. *Journal of Sea Research*, 53, 25–42.

Scranton, K. & Vasseur, D.A. (2016). Coexistence and emergent neutrality generate synchrony among competitors in fluctuating environments. *Theoretical Ecology*, 9, 353–363.

Wiedmann, I., Reigstad, M., Marquardt, M., Vader, A. & Gabrielsen, T. (2016). Seasonality of vertical flux and sinking particle characteristics in an ice-free high arctic fjord-Different from subarctic fjords? *Journal of Marine Systems*, 154, 192–205.