Fine-scale modelling of the dynamics of phytoplankton communities

Coralie Picoche

université de BORDEAUX

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR

DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET D'INFORMATIQUE

par **Coralie PICOCHE**

SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES ET CALCUL SCIENTIFIQUE

Modélisation à fine échelle de la dynamique des communautés phytoplanctoniques

---

Date de soutenance : 14 septembre 2022

Devant la commission d'examen composée de :

Bedr'Eddine AINSEBA	Professeur, Université de Bordeaux	Président
Frédéric BARRAQUAND	Chargé de Recherche, CNRS	Directeur
Frederik DE LAENDER	Professeur, Université de Namur	Rapporteur
Christèle ETCHEGARAY	Chargée de Recherche, INRIA	Examinatrice
Bart HAEGEMAN	Chargé de Recherche, CNRS	Examinateur
Jon PITCHFORD	Professeur, Université de York	Rapporteur

– 2022 –

Titre Modélisation à fine échelle de la dynamique des communautés phytoplanctoniques

Résumé Les communautés de phytoplancton, composées d'algues photosynthétiques, peuvent comprendre plusieurs centaines d'espèces nécessitant des ressources similaires, alors que les modèles théoriques prédisent le plus souvent que le nombre d'espèces coexistant ne peut être beaucoup plus grand que le nombre de ressources. De nombreuses explications à ce « paradoxe du plancton » ont été proposées, souvent basées sur les mêmes hypothèses : les interactions sont uniquement compétitives, la démographie n'intègre qu'un seul stade de vie, celui de l'individu flottant dans l'eau (le stade pélagique), et les individus sont distribués de façon homogène dans l'espace, avec des espèces mélangées. Dans cette thèse, deux modèles indépendants ont été construits afin de modifier ces hypothèses.

Dans un premier temps, un modèle de dynamiques de communautés avec deux stades de vie a permis de prendre en compte le stade dormant (la 'graine') dont la probabilité de survie est plus importante que celle du stade pélagique, particulièrement dans des conditions environnementales défavorables. Dans ce modèle, le stade pélagique permet de circuler entre l'océan et la côte, tandis que les individus en dormance s'accumulent dans une banque de graines côtière. Le réseau d'interactions est inspiré par l'analyse de séries temporelles réelles et comprend de la facilitation apparente en plus de la compétition. Nous montrons que la présence d'une banque de graines permet le maintien des espèces spécialistes, voire évite l'extinction totale de la communauté dans des conditions environnementales difficiles. La facilitation ne promeut par ailleurs pas la coexistence.

Dans le volet spatial de la thèse est développé un modèle individu-basé avec des processus démographiques et hydrodynamiques à micro-échelle. La réplication d'un modèle existant en deux dimensions a permis d'expliciter son analyse numérique et mathématique, donnant les fondations pour un modèle en trois dimensions. Dans ce modèle, les évènements de naissance et de mort sont représentés par un processus de branchement, chaque individu est transporté par une marche aléatoire correspondant à la diffusion, et soumis à une advection issue d'un modèle simplifié de la turbulence. Le modèle est paramétré à partir des caractéristiques observées du phytoplancton. Nous montrons qu'à des distances inter-individuelles pour lesquelles les interactions sont possibles, les petits organismes (nanophytoplancton) sont principalement entourés par des individus de la même espèce, ce qui favorise leur coexistence, tandis que les espèces plus grandes (microphytoplancton) sont plus mélangées, ce qui favorise la compétition inter-espèces. Nous discutons les mécanismes supplémentaires pouvant expliquer la maintenance de la diversité du microphytoplancton.

Mots-clés coexistence, compétition, dormance, équations de moments spatiaux, facilitation, modèle individu-basé, phytoplancton, processus de points

Title Fine-scale modelling of the dynamics of phytoplankton communities

Abstract Phytoplankton communities, made of photosynthetic algae, can include up to hundreds of species requiring similar resources. Classical population dynamics models, however, often predict that the number of coexisting species cannot be much larger than the number of resources. Numerous explanations to this "paradox of the plankton" have been proposed, often based on the same hypotheses: interactions are competitive, population dynamics are based on a single life stage, corresponding to the organism floating in the water column (the pelagic stage), and organisms are distributed homogeneously in space, all species being perfectly mixed in the environment. In this thesis, we build two independent models which enable us to relax these hypotheses.

Firstly, we establish a community dynamics model with two life stages, involving a dormant one (a 'seed'), which has a higher survival probability than the pelagic stage, especially in adverse environmental conditions. In this model, pelagic organisms can move between the ocean and the coast while dormant individuals remain in a coastal seed bank. The structure of interactions is inspired by field data, and comprises facilitation in addition to competition. The presence of a seek bank allows specialist species to survive in the community, and prevents the extinction of all species in harsh environmental conditions. Facilitation does not seem to promote coexistence.

In the spatial section of the thesis, we present an individual-based model including hydrodynamic and demographic processes at the microscale. The replication of an existing single-species model in two dimensions allowed us to develop the numerical and analytical methods which serve as a foundation for a three-dimensional, multispecies model. In this model, birth and death events are modeled by a branching process, and organisms are displaced by a random walk representing diffusion, and by a simplified model of turbulence. Parameter values are based on phytoplankton characteristics. We show that, for distances between individuals allowing interactions to happen, small organisms (nanophytoplankton) are mostly surrounded by individuals of the same species, which can favour coexistence, while larger species (microphytoplankton) are more mixed, which favours interspecific competition. We then discuss other potential mechanisms that could explain microphytoplankton diversity maintenance.

Keywords coexistence, competition, dormancy, facilitation, individual-based model, phytoplankton, spatial moment equations, spatial point processes

Laboratoire d'accueil Institut de Mathématiques de Bordeaux UMR 5251, Université de Bordeaux 351, cours de la Libération - F 33 405 TALENCE

Résumé détaillé en français

Contexte

Le phytoplancton est constitué d'algues photosynthétiques qui forment la base de la chaîne alimentaire dans les milieux aquatiques et représentent environ 50% de la production primaire mondiale [121]. Ces organismes sont caractérisés par une grande diversité, qui couvre plusieurs embranchements comprenant des centaines voire des milliers d'espèces. Cette thèse est principalement inspirée par les Bacillariophyceae (diatomées) et les Dinophyceae (dinoflagellés). Les observations des milieux aquatiques, que ce soit en eau douce ou salée, révèlent qu'en un seul lieu, plus d'une centaine d'espèces appartenant à ces classes peuvent cohabiter. En un seul prélèvement de quelques dizaines de mL, jusqu'à 70 espèces peuvent être identifiées [331].

Ces observations semblent contredire les modèles classiques de dynamique des communautés, qui indiquent que le nombre d'espèces coexistant ne peut être beaucoup plus grand que le nombre de ressources pour lesquelles les organismes sont en compétition. Cette opposition est souvent nommée le « paradoxe du plancton » [168]. De nombreuses propositions ont été formulées au cours des années pour expliquer ce phénomène. Les fluctuations environnementales font partie des hypothèses les plus précoces [168], mais il a été plus tard montré que des mécanismes supplémentaires étaient nécessaires pour qu'elles assurent effectivement le maintien de la biodiversité [125, 211] : le taux de croissance moyen des espèces doit être une fonction non-linéaire et/ou non-additive des variables environnementales. Une pierre angulaire de la théorie moderne de la coexistence de Chesson [74, 76], l'effet de stockage (*storage effect*), couple les fluctuations environnementales à d'autres mécanismes pour expliquer le maintien de la biodiversité. Il nécessite en effet une réponse différente du taux de croissance de chaque espèce aux variations environnementales, la covariation des effets de la compétition et de l'environnement, ainsi que leur sous-additivité qui amortit les fluctuations du taux de croissance. Cependant, d'autres modèles montrent que la diversité peut se maintenir même dans un environnement constant ; les fluctuations endogènes, parfois chaotiques, émergent alors des interactions entre organismes, que ce soit dans des modèles comprenant uniquement du phytoplancton et quelques ressources [165, 166], ou intégrant aussi des prédateurs [91]. La prédation peut aussi contrôler la dynamique du phytoplancton via un processus de 'suppression du gagnant' (*Kill The Winner*) : le prédateur transfère sa consommation principale vers l'espèce la plus abondante [326]. Cela évite que la force de compétition de l'espèce dominante pousse toutes les autres espèces à l'extinction.

Une partie des mécanismes décrits précédemment est associée, de façon plus ou moins explicite, à la théorie de la niche [206]. Si différents mécanismes se combinent pour que la régulation soit plus importante au sein d'une même espèce qu'entre les espèces, celles-ci peuvent coexister. Cela se traduit par des interactions intraspécifiques plus fortes que les interactions

interspécifiques [29], bien que des relations plus complexes puissent apparaître entre coefficients d'interactions pour des communautés diverses [28]. Or, les modèles étudiant des communautés de taille réaliste (au moins une dizaine d'espèces) sont moins fréquents que ceux focalisant sur deux ou trois espèces ou groupes fonctionnels. Bien que ces derniers étudient souvent plus finement les mécanismes qu'ils présentent, ils risquent d'ignorer les propriétés émergeant avec la multiplicité des espèces.

En plus de la dimensionalité, certaines hypothèses sur lesquelles s'appuient régulièrement les modèles peuvent être discutées à partir des observations de terrain. Ainsi, un seul stade de vie du phytoplancton est utilisé, le stade pélagique, représentant l'organisme en suspension dans l'eau. On sait néanmoins depuis longtemps que le phytoplancton a aussi un stade dormant (l'équivalent d'une 'graine') qui peut lui permettre de survivre plus longtemps et dans des conditions environnementales plus difficiles que le stade pélagique [111]. La prise en compte de plusieurs stades de vie a montré plusieurs fois son intérêt pour la coexistence dans d'autres écosystèmes [131, 239] ; son application dans des communautés phytoplanctoniques diverses reste à faire. Une seconde hypothèse prévaut dans la majorité des modèles : les organismes phytoplanctoniques sont censés n'interagir que via la compétition pour les ressources. Nos études de séries temporelles d'abondance de phytoplancton en milieu côtier ont néanmoins permis d'inférer des interactions facilitatrices entre les espèces, formant parfois la majorité des interactions apparentes ([32] et [271], Appendice A). La facilitation reste un phénomène relativement peu étudié dans les modèles de coexistence au sein d'un même niveau trophique. Cela peut s'expliquer par la complexité de ses effets : elle permet aux espèces de se maintenir ensemble, mais peut aussi déstabiliser la communauté via la rétroaction positive des espèces les unes sur les autres [228]. Il est donc important d'évaluer son impact sur un modèle de dynamique de communautés. Enfin, les variables d'état les plus souvent utilisées sont les concentrations, biomasses ou abondances des populations, qui supposent que les individus sont omniprésents et bien mélangés. Cependant, de nombreuses observations montrent que la distribution spatiale des individus est au contraire fragmentée : le phytoplancton forme des agrégats (ou clusters) qui se retrouvent à toutes les échelles spatiales. Cette hétérogénéité spatiale du phytoplancton peut s'expliquer par le type d'environnement dans lequel il évolue : mesurant quelques dizaines de µm, les organismes perçoivent un environnement principalement visqueux dans lequel les mouvements sont limités alors que la reproduction locale participe à la formation de clusters. L'agrégation intraspécifique favorise alors la coexistence en encourageant la compétition intraspécifique.

Nous avons donc défini trois hypothèses dont le relâchement pourrait affecter le maintien de la diversité du phytoplancton. Nous étudions l'effet de ces modifications dans deux modèles séparés. Un premier volet de la thèse couvre un modèle de dynamique des communautés avec une dizaine d'espèces, des interactions positives et une banque de graines. Dans un deuxième volet, nous développons un modèle individu-basé en deux puis trois dimensions à micro-échelle, et étudions numériquement et analytiquement l'agrégation intraspécifique en résultant.

Modèle incluant une banque de graines

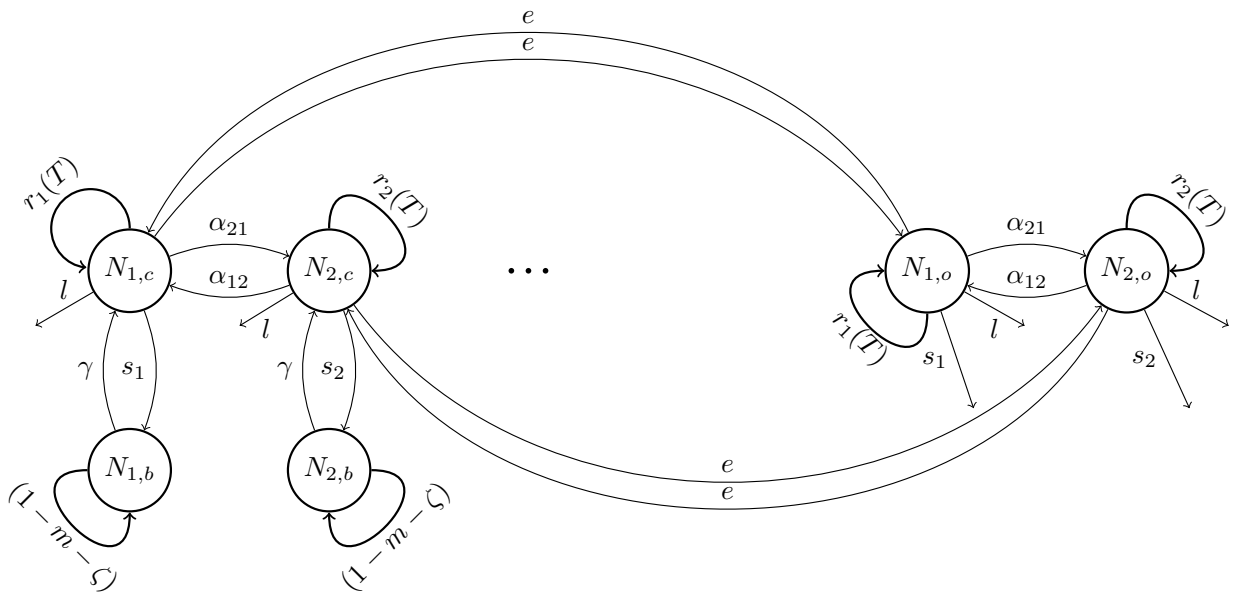

Figure 1: Structure du modèle incluant une banque de graines : les espèces phytoplanctoniques (dont les abondances sont indiquées par des cercles) sont présentes dans leur forme pélagique le long de la côte (indice c) et dans l'océan (o), et dans leur forme dormante dans la banque de graines (b). Les paramètres contrôlant les processus principaux sont indiqués par leurs symboles dans le schéma et définis dans le texte ci-dessous. Deux espèces sont montrées dans ce schéma, mais onze espèces sont présentes dans le modèle.

Modèle Nous étudions dans le Chapitre 2 une communauté phytoplanctonique avec un modèle en temps discret représentant des échanges entre la côte, l'océan, et la banque de graines (Fig. 1). Pendant un pas de temps d'un jour, les abondances N évoluent selon un modèle de croissance Beverton-Holt multispécifique (Eq. 1, 2) puis les organismes sont déplacés entre les différents compartiments (Eq. 3). Deux modèles ont été retenus pour représenter les interactions : le premier (modèle I) est basé sur la loi d'action de masse et s'écrit

$$\begin{cases} N_{t',i,c} &= \frac{\exp(r_i(T))N_{t,i,c}}{1+\sum_j \alpha_{ij}N_{t,j,c}} - lN_{t,i,c} \\ N_{t',i,o} &= \frac{\exp(r_i(T))N_{t,i,o}}{1+\sum_j \alpha_{ij}N_{t,j,o}} - lN_{t,i,o} \\ N_{t',i,b} &= N_{t,i,b}(1-m-\zeta) \end{cases} . \tag{1}$$

L'augmentation de l'abondance $N_{t',i,c/o}$ de l'espèce i le long de la côte ou dans l'océan est déterminée par son taux de croissance maximal $\exp(r_i(T))$, qui dépend de la température et des préférences thermiques de l'espèce, réduit par l'effet des interactions, où α_{ij} représente la force per capita de l'interaction de l'espèce j sur l'espèce i. Le taux de perte par mortalité naturelle, parasitisme ou prédation est représenté par le paramètre l.

Le deuxième modèle (modèle II) prend en compte un effet de saturation dans les interactions

inspiré par [279] :

$$N_{t',i,c/o} = \frac{\exp(r_i(T))N_{t,i,c/o}}{1 + \sum_{j\in\mathbb{C}} \frac{a_C N_{t,j,c/o}}{H_{ij}+N_{t,j,c/o}} + \sum_{j\in\mathbb{F}} \frac{a_F N_{t,j,c/o}}{H_{ij}+N_{t,j,c/o}}} - lN_{t,i,c/o} \tag{2}$$

où a_C et a_F sont les effets maximaux des interactions compétitives $\mathbb{C}$ et facilitatrices $\mathbb{F}$, respectivement, et H_{ij} est la densité de demi-saturation de l'espèce j affectant l'espèce i.

La banque de graines est uniquement soumise à la mortalité des graines m et à leur enfouissement ζ.

Après la phase de croissance est modélisée la phase d'échanges, durant laquelle

$$\begin{cases} N_{t+1,i,c} &= (1 - s_i - e)N_{t',i,c} + \gamma N_{t',i,b} + eN_{t',i,o} \\ N_{t+1,i,o} &= (1 - s_i - e)N_{t',i,o} + eN_{t',i,c} \\ N_{t+1,i,b} &= (1 - \gamma)N_{t',i,b} + s_i N_{t',i,c} \end{cases} . \tag{3}$$

Le passage des individus de la banque de graines côtière à la surface est lié à la germination et la remise en suspension γ tandis que le flux inverse est paramétré par s_i, qui représente à la fois l'encystement et le taux de plongée spécifique à l'espèce i. Enfin, la côte et l'océan échangent via les marées avec un taux e.

Paramètres Les valeurs des paramètres sont basées sur la littérature de terrain, principalement en milieu côtier. Ces valeurs étant fortement variables, une analyse de sensibilité a été effectuée en prenant en compte plusieurs valeurs observées, et a confirmé que les résultats ne changent pas significativement tant que l'on reste dans une gamme réaliste.

La communauté modélisée est inspirée par l'une des communautés analysées dans Picoche & Barraquand, 2020 [271] (voir Appendice A). Les onze espèces les plus fréquentes ont servi de base pour les onze morphotypes modélisés, incluant leurs préférences thermiques et leurs statuts de généralistes ou de spécialistes. Les interactions calculées dans [271], caractérisées par une forte régulation intraspécifique comparée aux interactions interspécifiques, et 70% de facilitation apparente, ont été adaptées à une formulation de type Beverton-Holt (avec ou sans saturation), puis re-calibrées pour atténuer les différences entre structures de modèles.

Scénarios et résultats principaux Après avoir confirmé la capacité du modèle à produire des dynamiques de population relativement proches des observations (survie de toutes les espèces et évolution cyclique avec des pics d'abondance saisonniers), deux scénarios ont été testés avec et sans banque de graines pour en évaluer l'importance :

1. Les forces des interactions compétitives et facilitatrices ont été séparément diminuées puis augmentées d'un ordre de magnitude pour représenter l'impact de variations biotiques, ou de modifications des ressources dans l'environnement.

2. La moyenne et la variance de la température ont été modifiées pour approximer des scénarios extrêmes de réchauffement climatique.

Les deux modèles d'interactions ont des dynamiques similaires. En l'absence d'une banque de graines, même sans modification des paramètres, la diversité chute de onze à quatre espèces. Les espèces qui survivent sont les généralistes dont les abondances ont l'amplitude de variation la plus faible, et la valeur minimale la plus élevée. L'effet de la banque de graines n'est pas notable quand on modifie les forces d'interactions, tandis que, dans des conditions environnementales difficiles (température élevée), la formation de graines permet le maintien d'au moins 9 espèces, même quand l'absence de banque de graines mènerait à l'extinction totale de toutes les espèces.

La variation des forces d'interactions permet de révéler l'effet déstabilisant de la facilitation. Quand celle-ci est augmentée, ou quand la compétition est diminuée, la communauté est déstabilisée et peut perdre des espèces, plus particulièrement dans le modèle II. L'équitabilité de la distribution des espèces, mesurée par l'inverse de l'indice de Simpson, diminue aussi, jusqu'à être divisée par trois, quand le poids de la facilitation s'accentue.

Ces résultats confirment l'importance de la banque de graines pour le maintien des espèces qui seraient plus susceptibles de disparaître dans des conditions environnementales inadaptées à leur niche thermique. Ils montrent aussi la nécessité de mieux comprendre l'effet de la facilitation sur la dynamique des communautés.

Perspectives Le modèle que nous présentons dans le Chapitre 2 permet d'évaluer l'importance de la banque de graines quand on prend en compte un seul niveau trophique. Les interactions entre organismes correspondent alors à un ensemble de phénomènes implicites : compétition pour les nutriments, mais aussi régulation par les ennemis naturels, qui peut faire émerger une facilitation apparente [1]. Il peut donc être intéressant d'ajouter un niveau trophique supplémentaire, représent l'exploitation par la prédation et / ou les parasites afin d'étudier leurs effets sur la coexistence. La modification éventuelle des interactions modélisées pourrait alors éviter la forte déstabilisation du système observée quand on augmente la force de la facilitation dans la communauté. Le rajout de ce niveau trophique peut se faire de façon implicite en ajoutant des délais dans la formulation de la dynamique des populations, qui indiquerait le décalage entre la croissance du phytoplancton et celle des prédateurs ou des parasites en réponse à une augmentation d'abondance du phytoplancton.

L'article correspondant a été publié sous la référence suivante : C. Picoche & F. Barraquand (2022) Seed banks can help to maintain the diversity of interacting phytoplankton species. *Journal of Theoretical Biology*, 538, 111020.

Modèle spatialisé à micro-échelle

Modèle : simulations numériques et analyse mathématique Nous avons développé un modèle spatial à partir du Brownian Bug Model (BBM, qui pourrait être traduit peu élégamment par 'Modèle de Microbe Brownien') initialement présenté par Young *et al.*, 2001 [345]. Ce modèle individu-basé en temps et espace continus permet de suivre le devenir d'organismes de concentration C qui se reproduisent (avec un taux de naissance λ) et meurent (avec un taux de mortalité μ) dans un environnement visqueux et turbulent à l'échelle du plus petit tourbillon possible dans un milieu océanique. Les organismes sont déplacés par la diffusion (de diffusivité D) et la turbulence qui sépare deux organismes selon une loi exponentielle de paramètre $d\gamma$ où d est le nombre de dimensions de l'espace. Pour les simulations numériques, le modèle est discrétisé dans le temps : les évènements de naissance et de mort sont représentés par un processus de branchement, chaque individu est transporté par une marche aléatoire représentant la diffusion, et soumis à une advection issue d'un modèle simplifié de la turbulence [274].

La publication [345] présente les résultats du modèle en deux dimensions et pour une seule espèce. De plus, l'obtention des formules analytiques n'est pas explicitée. Dans un article de réplication (Chapitre 3), nous avons retrouvé les résultats originaux, mais aussi re-démontré les équations permettant de caractériser la distribution spatiale des organismes. Ainsi, l'agrégation des organismes est quantifiée par la fonction de corrélation de paires g. Celle-ci peut être définie en considérant $C^2 g(r,t) d\mathbf{x}_A d\mathbf{x}_B$ la probabilité de trouver un individu dans la sphère $d\mathbf{x}_A$ et un individu dans la sphère $d\mathbf{x}_B$, la distance entre les centres de $d\mathbf{x}_A$ et $d\mathbf{x}_B$ étant égale à r ([169], page 219). Quand les individus sont agrégés, $g > 1$. Nous montrons que la variation dans le temps de g en d dimensions suit

$$\frac{\partial g}{\partial t} = 2Dr^{1-d}\frac{\partial}{\partial r}\left(r^{d-1}\frac{\partial g}{\partial r}\right) + 2(\lambda-\mu)g + \gamma r^{1-d}\frac{\partial}{\partial r}\left(r^{d+1}\frac{\partial g}{\partial r}\right) + \frac{2\lambda}{C}\delta(\boldsymbol{\xi}). \tag{4}$$

À partir de la formule précédente, et en supposant que la population est stable ($\lambda = \mu$), nous avons calculé les formules analytiques de g avec et sans turbulence, et confirmé l'agrégation des individus due à la reproduction locale et la limitation des mouvements dans l'environnement.

Dans un deuxième temps, afin de pouvoir étudier la coexistence du phytoplancton avec des paramètres plus réalistes, nous avons adapté le BBM en trois dimensions et pour plusieurs espèces (Chapitre 4). Nous avons aussi mis en place un autre indice dérivable à partir de g mais plus facile à interpréter écologiquement : l'indice de dominance locale $\mathcal{D}_i$ caractérise la composition du voisinage d'un individu de l'espèce i dans la sphère de rayon r l'entourant à partir du ratio entre le nombre moyen d'individus de la même espèce (conspécifiques) et le nombre total d'individus. Quand les individus sont principalement entourés de leurs conspécifiques, $\mathcal{D}_i \to 1$ tandis qu'il tend vers la fraction relative de l'abondance de l'espèce i dans l'environnement $\frac{C_i}{C_{\text{tot}}}$ quand les individus sont mélangés sans distinction d'espèce. En l'absence d'advection,

$$\mathcal{D}_i(r,t) = \frac{C_i\left[\frac{4}{3}\pi r^3 + \frac{\lambda}{C_i D}\left(\frac{r^2}{2} - \frac{1}{2}\operatorname{erf}(\frac{r}{\sqrt{8Dt}})(r^2-4Dt) - \frac{\sqrt{2Dt}r}{\sqrt{\pi}}e^{-r^2/8Dt}\right)\right]}{\sum_{j=1}^{S} C_j\frac{4}{3}\pi r^3 + \frac{\lambda}{D}\left(\frac{r^2}{2} - \frac{1}{2}\operatorname{erf}(\frac{r}{\sqrt{8Dt}})(r^2-4Dt) - \frac{\sqrt{2Dt}r}{\sqrt{\pi}}e^{-r^2/8Dt}\right)} \tag{5}$$

tandis qu'en présence d'advection,

$$\mathcal{D}_i(r) = \frac{C_i\left[\frac{4}{3}\pi r^3 + \frac{\lambda}{3C_i D}\left(r^2 + \frac{\sqrt{\gamma}r^3\arctan(\sqrt{\frac{\gamma}{2D}}r)}{\sqrt{2D}} + \frac{D\log\left(\gamma\frac{r^2}{2D}+1\right)}{\gamma} - \frac{\sqrt{\gamma}\pi r^3}{2\sqrt{2D}}\right)\right]}{\sum_{j=1}^{S} C_j\frac{4}{3}\pi r^3 + \frac{\lambda}{3D}\left(r^2 + \frac{\sqrt{\gamma}r^3\arctan(\sqrt{\frac{\gamma}{2D}}r)}{\sqrt{2D}} + \frac{D\log\left(\gamma\frac{r^2}{2D}+1\right)}{\gamma} - \frac{\sqrt{\gamma}\pi r^3}{2\sqrt{2D}}\right)}. \tag{6}$$

La solution avec advection correspond au système à l'état stable $\left(\frac{\partial g}{\partial t} = 0\right)$. S'il est possible de calculer une formule équivalente en l'absence d'advection, la convergence vers l'état stable se fait dans un temps trop long pour que l'hypothèse de stabilité reste fondée. C'est pourquoi nous présentons dans l'équation 5 une formule dépendant du temps.

Afin de faire le lien entre l'indice de dominance locale et des processus biologiques, nous considérons que les interactions entre individus sont uniquement dues à la compétition, et nous définissons une sphère d'influence autour de l'individu dans laquelle celui-ci peut diminuer la concentration de ressources par absorption. Le chevauchement des sphères d'influence de deux organismes permet de supposer qu'une interaction est possible. Le rayon d'interactions potentielles correspond à dix fois le diamètre d'un individu [187]. On s'intéresse alors à la valeur de l'indice de dominance à ce seuil afin de voir si un individu est plus susceptible d'entrer en compétition avec des conspécifiques ou des hétérospécifiques (individus d'une espèce différente).

Le volume de l'espace dans lequel les organismes évoluent correspond approximativement à la taille du plus petit tourbillon (échelle de Kolmogorov) dans un milieu océanique, c'est-à-dire quelques cm. L'effet de l'environnement est défini par la présence, ou non, du mixage par la turbulence dont l'intensité est calculée à partir du nombre de Reynolds, égal à 1 à l'échelle de Kolmogorov.

Nous représentons séparément deux types de communautés phytoplanctoniques : le microphytoplancton, ayant un diamètre de 50 µm, un taux de croissance de 1 jour^{-1} et une concentration de 10^4 cellules/L en moyenne, et le nanophytoplancton, ayant un diamètre de 3 µm, un taux de croissance de 2.5 jour^{-1} et une concentration de 10^6 cellules/L en moyenne. Nous calculons la diffusivité des organismes grâce à la formule de Stokes-Einstein ([110], cité par Dusenbery [104]). Au sein d'une même communauté, toutes les espèces ont la même taille et le même taux de croissance. Nous évaluons d'abord la distribution spatiale dans une communauté simplifiée avec trois espèces et la même abondance moyenne, puis nous nous intéressons à une communauté de 10 espèces avec une répartition d'abondances asymétrique, plus proche des observations. Les indices décrivant la distribution spatiale des individus calculés à partir des simulations sont très proches des valeurs prédites par l'analyse mathématique.

Le mélange des espèces par la turbulence est confirmé : en l'absence d'advection, les espèces

sont beaucoup plus agrégées entre elles qu'en sa présence. Néanmoins, il existe toujours un rayon (parfois très petit) pour lequel l'individu n'est entouré que de conspécifiques, c'est-à-dire qu'il y a une agrégation intraspécifique. Les résultats sont plus nuancés à plus grande échelle. Les espèces de nanophytoplancton tendent à rester agrégées au sein du volume d'interactions potentielles ($\mathcal{D}_i \geq 0.8$ dans une communauté simplifiée), tandis que la composition du voisinage du microphytoplancton est plus mixte ($\mathcal{D}_i \geq 0.37$ dans une communauté simplifiée). Ce résultat est confirmé dans une communauté de 10 espèces, dans laquelle l'indice de dominance du microphytoplancton pour une distance égale au rayon maximal d'interactions potentielles varie entre 0.34 et 0.033 tandis qu'il est compris entre 0.2 et 0.54 pour le nanophytoplancton.

On peut supposer à partir de ces constats que la coexistence du nanophytoplancton peut s'expliquer par l'agrégation intraspécifique qui permet à l'espèce de s'auto-réguler via la compétition plus qu'elle ne régule les autres espèces. En revanche, pour le microphytoplancton, la moindre dominance observée semble indiquer que le maintien de la diversité ne peut s'expliquer seulement par la compétition liée à la distribution spatiale des individus.

Perspectives Le modèle que nous présentons dans les Chapitres 3 et 4 ne s'intéresse qu'à des individus flottant dans la colonne d'eau sans capacités de mouvement volontaire, et sans interactions. Il peut être utilisé comme un modèle nul auquel comparer les modèles caractérisant le phytoplancton de manière plus réaliste. Parmi les traits pouvant agir sur l'agrégation et la force des interactions intra- versus interspécifiques, la colonialité d'un grand nombre d'espèce encourage fortement l'agrégation intraspécifique. La formation de colonies est renforcée par la capacité du phytoplancton à se mouvoir de façon volontaire dans son environnement, potentiellement vers ses conspécifiques, que ce soit dans les trois dimensions ou uniquement à la verticale grâce au contrôle de sa flottabilité.

Les interactions que nous avons envisagées au niveau individuel sont uniquement dues à la compétition pour les nutriments. Cependant, il existe au moins un autre mécanisme de partage de ressources qui pourrait modifier l'interdépendance entre organismes : les individus sont entourés d'une phycosphère, une zone dans laquelle des échanges ont lieu entre phytoplancton et bactéries, souvent pour des éléments favorables à leurs croissances respectives (*cross-feeding*). Ainsi, une balance entre compétition et mutualisation de ressources pourrait émerger du rapprochement des individus.

L'ensemble des mécanismes présentés peut bénéficier à la coexistence des espèces, particulièrement dans le cas du microphytoplancton, ce qui permettrait d'atténuer les différences observées entre nano- et microphytoplancton.

L'article de réplication a été publié sous la référence suivante : C. Picoche, W.R. Young & F. Barraquand (2022) [Re] Reproductive pair correlations and the clustering of organisms. *ReScience C*, 8.

Le manuscrit correspondant au Chapitre 4 est en cours de rédaction, et vise *Journal of Mathematical Biology.*

Remerciements

Mes premiers remerciements ne peuvent aller que vers mon directeur de thèse, Frédéric Barraquand, dont le niveau de connaissances, la disponibilité et la pédagogie m'ont permis de comprendre et apprendre bien plus que je ne l'aurais espéré. Fred, je souhaite à tous tes futurs étudiants et post-doctorants que ta capacité à partager ton savoir, ton exigence et ton honnêteté intellectuelles leur apportent autant qu'à moi ; la recherche ne peut sortir que grandie de ton mentorat. Je te remercie sincèrement de ton aide, ton soutien et ton indulgence, même dans des conditions parfois compliquées. Je n'aurais pu faire cette thèse sous une meilleure direction.

Je remercie aussi les membres de mon comité de thèse, Bedr'Eddine Ainseba, Soizic Morin et Marc Sourisseau qui, pendant cette thèse 'un peu' confinée, ont apporté un point de vue extérieur et toujours positif sur mon travail. Ce retour a toujours su relancer mes foulées dans le marathon qu'est la thèse. Tout particulièrement, je remercie Bedr'Eddine d'avoir répondu présent en dernière minute pour faire partie de mon jury de thèse. Merci pour ta disponibilité et ta compréhension.

I would also like to thank all the members of my thesis defense committee for accepting to spend a part of their summer with this manuscript. I am especially grateful for the sensible comments of my referees, and the judicious questions of everyone during my defense, which helped me deepen my own reflection about my work.

A special word needs to be said about Dr. William R. Young, who, despite his very busy schedule and my *very* naive approach to hydrodynamics and physics in general, found the time and patience to explain (and sometimes, re-explain), verify and correct my work.

Parce que le travail de thèse est nourri par plus que le travail, je ne peux ensuite qu'étendre ma gratitude envers tous ceux qui ont fait de mon temps libre des escapades dans d'autres univers. Sans ordre particulier, Steph, Vhan, Hélène, Anthony, Raphie, et surtout mon Quentin... qui m'ont sorti au moins un temps de ces années folles.

Marion, merci d'être encore là après toutes ces années, malgré la vie qui s'immisce.

Alix, Bordeaux a un peu perdu de sa couleur à ton départ, mais je te remercie pour toutes nos conversations qui continuent à m'inspirer.

Il reste enfin les piliers de ma vie. Cette page n'est pas l'endroit pour vous dire tout ce que vous m'avez apporté. Je tiens néanmoins à vous citer.

Mon grand-père, Élie Picoche, que j'espère rendre fier à travers ces quelques pages. Tu ne pourras sans doute pas entièrement les comprendre, mais je te les dois quand même en partie.

Ma mère, Véronique Picoche, dont le soutien sans faille a maintenu cette thèse. Je te dois les qualités dont je peux être fière, et l'exemple pour corriger mes défauts.

Quentin, ma lumière et ma chance, j'espère que tu sais déjà toute la force que tu m'as donnée pendant les trois années passées... et vivement les années futures !

Contents

Organisation

List of papers

This thesis manuscript is constituted of three main articles written during my doctoral work.

- C. Picoche & F. Barraquand (2022) Seed banks can help to maintain the diversity of interacting phytoplankton species. *Journal of Theoretical Biology*, 538. (Chapter 2)
- C. Picoche, W.R. Young & F. Barraquand (2022) [Re] Reproductive pair correlations and the clustering of organisms. *ReScience C*, 8. (Chapter 3)
- C. Picoche, W.R. Young & F. Barraquand. Local intraspecific aggregation in phytoplankton model communities: spatial scales of occurrence and implications for coexistence. (draft, Chapter 4)

A paper, first submitted before, and considerably improved during, my doctoral work, until its publication in April 2020, is also provided ([271], Appendix A) to help the reader understand some of the statements and assumptions we make in the rest of the manuscript.

- C. Picoche & F. Barraquand (2020). Strong self-regulation and widespread facilitative interactions in phytoplankton communities. *Journal of Ecology*, 108.

Each paper is written with ecologists in mind as the first readership. Supplementary Information often provide further analyses useful for other, more mathematically-oriented, readers. We therefore include the Supplementary Information immediately after each article. This is especially relevant for Chapter 3, where one of the main additions to the replication presented is detailed in the Appendices.

Contributions

Chapters	1	2	3	4	5	A
Project design	–	CP-FB	CP-FB	CP-FB	–	FB-CP
Computer code	–	CP	CP	CP	–	CP
Mathematical analyses	–	–	FB-CP-WY	CP-FB-WY	–	FB-CP
Manuscript	CP	CP-FB	CP-FB-WY	CP-FB-WY	CP	FB-CP

FB: F. Barraquand
CP: C. Picoche
WY: W.R. Young

Chapter 1

Introduction

1.1 Coexistence in phytoplankton communities

1.1.1 Phytoplankton ecology

Phytoplankton organisms are autotrophic entities which spend most of their life in quasi-suspension in aquatic environments [285]. In the ocean, they sustain most of their ecosystem food web and are responsible for approximately half the primary production on Earth [121]. The broad definition of these primary producers covers both prokaryote and eukaryote domains, the latter being separated between 12 phyla, as well as several orders of magnitude in dimension, from 0.1 to 10^3 µm. Among eukaryotes, diatoms (Bacillariophyta) and dinoflagellates (Dinophyta) are present in both freshwater and marine systems and sometimes amount to 84 % of total diversity observed in samples [155]. We focus here on the microphytoplankton size fraction (i.e., the portion of organisms between 20 and 200 µm), with a digression towards nanophytoplankton (2-20 µm) at the end of this thesis. Microphytoplankton is of specific interest, as it can be observed, and more importantly, identified by simple optical microscopy, which enables us to draw from decades of observations and long-term monitoring when modeling phytoplankton communities.

Phytoplankton diversity is surprisingly high. Based on the number of species already described at the global scale, diatom diversity is estimated between 30,000 and 100,000 species [221], and the number of known dinoflagellate species reaches 2,377 [135]. The recent TARA Ocean survey, a 4-year, ocean-scale sampling which coupled metabarcoding and morphology-based identification, revealed between 3,875 and 41,381 different diatom operational taxonomic units (OTUs, a proxy for species in molecular surveys) in the microphytoplankton size range [220, 66], compared to 256 morphology-based identifications performed on the same samples [66]. With no size filtering, dinoflagellate's OTU number reached around 18,000, and diatom's, around 15,000, with morphology-based values about 10 times lower [273].

This global diversity is also reflected at the local scale. In France, the National Phytoplankton and Phycotoxin Monitoring Network [284], also called REPHY monitoring, has been following the abundance and diversity of microphytoplankton on a biweekly basis along the continental French coast for 20 to 30 years. Within most sites, 100 different phytoplankton genera, or groups of genera, have been identified locally, among which at least 10 groups of genera have remained present, and abundant, consistently through the years [271]. In a single 10 mL-sample, up to 60 taxa (from species to groups of genera) could be identified, with an average of

22 [284]. Further north, between the English Channel and the North Sea, 47 different genera of diatoms, and 30 genera of dinoflagellates were found in 19 years, with 8 taxa dominating the community [155]. In England, a 28 year-long time series of phytoplankton abundances at Plymouth Station L4 has enabled the identification of 111 diatom species (including 31 species for a single genus, *Chaetoceros*) and 61 dinoflagellate species, ignoring organisms that could not be identified below the genus level - which means that the total number of different species and genera could be even higher [331]. In a single 200 mL sample, up 70 species of diatoms and dinoflagellates could be identified in this monitoring, with an average of 35. At the global scale, a compilation of 16 long-term environmental monitoring in 9 different countries showed that a single sample could contain between 15 and 39 microphytoplankton species on average [73].

Phytoplankton abundance time series also reveal specific patterns of phytoplankton dynamics: population abundances have a cyclic behaviour, correlated with seasonal variation. They reach their minimum during winter, to the point that certain species seem to disappear, either due to a real absence from the water column, or to sampling effects. When spring comes, abundances increase up to potentially very high numbers. The peaks in abundances several orders of magnitude above average values are called blooms; they can happen as early as spring (or even late winter), last a few days, and are characteristic of phytoplankton communities in lakes, polar and subpolar oceans, and coastal environments [38]. While all species do not bloom, communities in these places always comprise a few species that do, and that can constitute up to 90% of the total phytoplankton abundance during a bloom [155].

Such dynamics depend on both abiotic and biotic resources, whose fluctuations are correlated with the variations in phytoplankton abundances. Being autotrophic organisms, the first resource requirement of diatoms is light, which can also be used as an energy source by about 50% of dinoflagellates, more precisely the ones that contain plastids [135]. Among the many elements required to maintain phytoplankters' growth, nitrogen (N), phosphorus (P) and iron (Fe), as well as silicon (S) for diatoms, whose cell walls (frustules) are made of silicic acid, are the most well-known limiting factors [285]. These resources' availability tends to be cyclic due to light duration and nutrient stock dynamics, especially in coastal environments where nutrient inputs vary with time. The hydrodynamic environment, which also depends on seasonal phenomena, adds to variations in resource availability: water column mixing can help phytoplankters remain in an illuminated zone and mix nutrients in the surface layer, but the following stratification period can lead to nutrient depletion within the same layer, even though sunlight maintains the potential for high phytoplankters' growth rates [38, 39]. In addition to these shared abiotic factors, phytoplankton populations are regulated by similar enemies, either predators (mostly zooplankters, which can consume up to three-quarters of phytoplankton production [69]), or parasites such as viruses [63] or fungi [129]. While bloom formation may be partly initiated by light and nutrient meeting phytoplankton requirements, the increase in abundance can trigger a delayed increase in predators' and parasites' abundances, which can in turn reduce phytoplankton populations. The balance between bottom-up (nutrient and light

availability) and top-down (predation) control is debated, as different studies lead to different conclusions regarding the most significant level in the trophic network [38]. Obviously, both processes do matter in regulating phytoplankton dynamics.

This highly schematic description of phytoplankton population dynamics highlights one of the major wonders of phytoplankton and coexistence studies, deemed the paradox of the plankton: *"how it is possible for a number of species to coexist in a relatively isotropic or unstructured environment all competing for the same sorts of materials"* [168]. Indeed, early population dynamics models show that S species cannot coexist on much less than S resources, niches or limiting factors, depending on the focus of the model (among the earliest works, see [330], [219]), a finding which has later been called the 'competitive exclusion principle', or Gause's principle [145]. Even though the competitive exclusion principle has been debated ever since its formulation [145, 16], the coexistence of such a high number of species still remains surprising for ecologists. Several hypotheses have therefore been formulated to explain the maintenance of the observed phytoplankton diversity.

1.1.2 Coexistence hypotheses

Hutchinson himself, in his seminal paper on the paradox of the plankton [168], proposed temporal fluctuations as the main mechanism behind diversity maintenance. More precisely, he suggested that it is an absence of equilibrium that allows all species to survive, which happens when the time to extinction (i.e., the time necessary for competitive replacement in a constant environment), remains comparable to the time for significant seasonal changes. Since then, Fox (2013) [125] has recapitulated the arguments of many studies showing that temporal fluctuations on their own are not sufficient to maintain diversity, but that species' average per-capita growth rates need to be nonlinear and/or nonadditive functions of fluctuating variables. Such relationships have been found in experiments showing complex interactions between environmental effects and growth rates. For instance, phytoplankton growth is affected by interactions between nutrient availability and the presence of temperature fluctuations [134], and there is a saturating relationship between growth and irradiance, which also varies with temperature [108].

The 'modern coexistence theory', developed by P. Chesson and colleagues, has built on the role of environmental fluctuations, highlighting the interactions between environment and competition [74, 76]. In this framework, the survival of a species in a community is characterized by its ability to grow from low densities (invade) in the presence of the rest of the community. Responding to Hutchinson's theory, Li & Chesson (2016) [211] have shown that the focus on time to extinction should be displaced to time for resource depletion (i.e., to resource dynamics, see box 1.1). In this context, fast resource depletion leads to a stronger interaction between environmental fluctuations and competition. Such mechanism is linked with the storage effect, a cornerstone of the modern coexistence theory, which requires species-specific responses to the environment, covariance between environmental and competitive effects, and buffered

population growth; all conditions which have been found in experiments on a phytoplankton community [96]. Interestingly enough, the storage effect does not necessarily rely on temporal fluctuations: spatial variability can also create a spatial storage effect when the effect of competition is greater for a species in a more favourable environment (there is a positive covariance between environment quality and strength of competition [146]). Contrary to Hutchinson's initial statement, phytoplankton densities are indeed not homogeneous: organisms tend to form aggregates [24], potentially encouraging a spatial storage effect.

Box 1.1: Environmental fluctuation-based model

The response of Li & Chesson to Hutchinson's solution to the paradox of the plankton [211] is described in eq. 1.1 (using the primary-producer version shown in the Appendix of [211]).

$$\begin{aligned} \frac{dP_j}{dt} &= S_P P_j b_j \left(a_j(t)R - m_j\right) \\ \frac{dR}{dt} &= S_R \left(I - R\left[D + \Sigma_j a_j(t) P_j\right]\right) \end{aligned} \tag{1.1}$$

where P_j is the biomass of phytoplankton species j, R is the resource, b_j is the conversion of resource uptake to production of new consumers of species j, $a_j(t)$ is the species' resource uptake rate, m_j, its maintenance rate, I is the resource supply rate, and D is its decay rate. Note that environmental changes are represented in the time-dependence of $a_j(t)$, not directly in R parameters. S_P and S_R are speed modifiers that can increase population and resource turnover, respectively.

Ignoring speed modifiers for now, the environmental response $E(t)$ is computed as the ratio between the resource uptake of species j and its average over time $(a_j(t)/\bar{a}_j)$ and the magnitude of competition $C(t)$ is a measure of resource shortage $(I - RD)$. The growth rate $r_j = \frac{1}{b_j \bar{a}_j P_j}\frac{dP_j}{dt}$ can therefore be written as

$$r_j = E_j\left(\frac{I - C}{D}\right) - \mu_j \tag{1.2}$$

where $\mu_j = m_j/\bar{a}_j$.

We note that this model is a good example of the storage effect, as it shows differential responses to the environment (through values of $a_j(t)$ in $E(t)$), covariance between the environment and competition, and a buffered growth rate $\frac{\partial^2 r_j}{\partial E \partial C} < 0$.

The invasibility criterion, which defines the ability of a species to survive in a community, is defined as $\bar{r}_i$, the average growth rate of a species invading the community. The growth rate $\bar{r}_i$ has to be positive for a species to maintain. It is defined by

$$\bar{r}_i = (\mu_s - \mu_i) + (\mathrm{Cov}(E_s, C) - \mathrm{Cov}(E_i, C)) \tag{1.3}$$

where $(\mu_s - \mu_i)$ measures the difference in average fitness between a long-term resident (subscript s) and an invader (subscript i), predicting the winner at equilibrium. The difference $\Delta\text{Cov} = \text{Cov}(E_s, C) - \text{Cov}(E_i, C)$ is the difference in coupling between competition and species-specific environmental responses. Covariance between competition and environment is therefore a key parameter to ensure the coexistence of a species, even when this species should disappear in stable conditions (and in this case, $\Delta\text{Cov} > \mu_i - \mu_s$ to counterbalance the negative average fitness difference).
Li & Chesson prove that the time to exclusion as defined by Hutchinson is proportional to the inverse of the phytoplankton population dynamics speed S_P, and does not depend on resource dynamics speed S_R. They also show that ΔCov is most affected by resource speed, as it directly impacts resource increase or decrease (and therefore, availability), while phytoplankton speed first impacts population growth, and, in a second phase, resource availability. According to them, coexistence is therefore mostly impacted by resource speed, not by the time to exclusion linked to phytoplankton speed. These results are robust to different types of fluctuations, either a white noise or a sinusoidal ('seasonal') variation.

However, observations of a natural plankton community showed that, even in a constant abiotic environment (e.g., constant input of nutrients), most of the species, and their cyclic dynamics, could maintain [42]. Interactions such as competition or predation within a community can indeed create cylic or chaotic dynamics [232], especially, in a competitive system, when intermediate resource requirements are coupled with high consumption rates ([165], see box 1.2). Chaos also emerged in a planktonic predator-prey model [91], where the variability in species composition, and overall maintenance of species over the years could be seen as a seasonally entrained chaos corresponding to observed planktonic dynamics. Endogenous and exogenous fluctuations are therefore non exclusive, but seem to both participate in coexistence.

Box 1.2: Endogenous fluctuation-based model

Huisman & Weissing (1999) [165] have designed a model in which fluctuation-based coexistence mechanisms emerge from a constant environment. Assuming a community of S phytoplankton species and k resources, they write

$$\begin{aligned} \frac{dP_i}{dt} &= P_i\left(\min\left[p_{1i}(R_1), \ldots, p_{ki}(R_k)\right] - m_i\right) \\ \frac{dR_j}{dt} &= D(S_j - R_j) - \Sigma_i c_{ji} P_i\left(\min\left[p_{1i}(R_1), \ldots, p_{ki}(R_k)\right]\right) \end{aligned} \quad \text{with } p_{ji} = \frac{r_i R_j}{K_{ji} + R_j} \tag{1.4}$$

where P_i is the abundance of phytoplankton species i, R_j is the availability of resource j, r_i is the maximum specific growth rate of species i, K_{ji} is the half-saturation constant of species i for resource j, m_i is the specific loss rate of species i, D is the ressource turnover rate, S_j is the supply of resource and c_{ji} is the content of resource j in species i (or efficiency of consumption). $K_{j1} < K_{j2}$ means that species 1 is a better competitor than species 2 for resource j, and $c_{j1} > c_{j2}$ means that species 1 consumes more of R_j than species 2.

The competitive exclusion principle states that S species can coexist only if $S \leq k$. With $S = k = 3$, Huisman & Weissing (2001) [166] show that, if $c_{.i}$ is maximal for intermediate value of $K_{.i}$), the system can either exhibit limit cycles (fixed-frequency, 'permanent coexistence') or heteroclinic cycles ('impermanent coexistence'). In the latter case, the community cycles through a succession of quasi monocultures (only one abundant species), until species that are too close to 0 finally disappear due to demographic stochasticity - the identity of the survivor is thus difficult to predict. With $S = k = 5$, a similar situation leads to competitive chaos, and all species can coexist.

Huisman & Weissing [165] show that $S > k$ species can maintain ('supersaturated coexistence') thanks to competitive chaos, as long as there are specific relationships between parameters, summarized as "each species is an intermediate competitor for the resources that most limit its growth rate".

When parameters are random, supersaturation rarely happens, and only few species can maintain [299]. However, new relationships were introduced in [164], including a trade-off between competitive abilities ($0 < K < 1$ and mean(K)=0.5), and a cyclic relationship between competitive abilities and resource contents (high $c_{i.}$ leads to low $K_{(i+1).}$). In this case, up to 115 species could survive on 3 resources with an average of 7 species. Trade-offs between consumption and competitive abilities are likely [214], which argues in favour of at least *some* relationships between parameters.

While these models remain mostly theoretical (parameters are not based on real data [299], and some assumptions might seem a bit artificial), they still show that environmental fluctuations are not necessary to maintain coexistence, which can also be due to endogenous fluctuations and chaotic competition.

Environmental effects on population dynamics can be broadly classified into two types: bottom-up effects, due to resource dynamics, which has supported the majority of the hypotheses presented above, and top-down effects, mostly due to regulation by natural enemies including grazing by zooplankton. Three types of zooplankton responses to multiple resources have been distinguished [133]: (a) absence of switching, i.e., preferential feeding on a resource is density-independent, (b) passive switching, i.e., selection is density-dependent and associated with the response to single resources due to, for example, resource vulnerability, handling time,

nutritional content, and (c) active switching, i.e., active selection as behaviours change with relative abundances of resources, such as rejection of the least abundant prey, or search activity focused on high-density patches. Diet switching, i.e., the fact that the proportion of the phytoplankton species in a zooplankter's diet shifts from less than expected to more than expected when its proportion in the environment increases, can enforce coexistence by providing a refuge for rarer species, but it also controls the most abundant species, therefore keeping it from exerting too strong a control on other phytoplankton species through competition. Preference, active switching and maximal ingestion (strong top-down control) can be combined in a so-called Kill-The-Winner response (example in box 1.3), which has been shown to sustain the highest number of phytoplankton species compared to other zooplankton responses in a world-scale model [326], switching being the prevailing process for diversity [14].

Box 1.3: Top-down control: Kill-The-Winner hypothesis

Active prey-switching and maximal feeding are not always well-represented by classical multi-species functional responses. Vallina et al. [326] have designed a functional response including both mechanisms. Defining G_j as the ingestion rate upon species j, and G as the total ingestion rate upon all prey items,

$$\begin{aligned} G_j &= V_{\max} \frac{\rho_j p_j^2}{\Sigma_i \rho_i p_i^2} \frac{(\Sigma_i \rho_i p_i)^\beta}{k_{\text{sat}}^\beta + (\Sigma_i \rho_i p_i)^\beta} \\ G &= \Sigma_j G_j = V_{\max} Q \end{aligned} \tag{1.5}$$

where $V_{\max}$ is the maximum ingestion rate, ρ_j is the constant preference towards species j, p_j is the abundance of species j, β is the Hill coefficient which determines the shape of the feeding probability (Type II or Type III response; in [326], $\beta = 2$) and k_{sat} is the half-saturation constant.
This model is able to maintain much more phytoplankton species than other functional responses (up to 48 species out of 64).

Looking at the bigger picture, two main frameworks of coexistence should be considered. Firstly, in the neutral theory of biodiversity [162, 288], communities are considered as ensembles of identical individuals which differ only by their species labels. More precisely, equivalence in fitness emerges from trade-offs in life-history traits; competition is therefore the same within and between species, and coexistence is maintained by immigration, speciation and ecological drift. While attempts have been made to find neutrality within diatom and dinoflagellate communities based on data from the English station L4 [248], we showed the contrary in several French stations in [32] and [271] (provided in Appendix A). Secondly, the opposed, more traditional approach, is based on the definition of species-specific niches, i.e., the characteristic set of optimal environmental conditions which maximize the growth rate of a particular species. For example, phytoplankton species either have slightly different responses to light

wavelengths [296, 10], differ in their macronutrient metabolism pathways [6], and/or zooplankters have different preference profiles for them [251]. Different responses to the environment constitute one of the conditions for coexistence in [74, 76] mentioned above. Niche differentiation also leads to an increase in intraspecific interaction strengths relative to interspecific interaction strengths, as organisms of the same species (conspecifics) have more similar regulating factors than organisms of different species (heterospecifics): a stronger self-regulation ensures that species control their own population more than they control, or outcompete, the others. A high intra-to-interspecific interaction strength ratio is therefore a classical requirement in coexistence models [28, 29], and is a usual find in phytoplankton systems: in Picoche & Barraquand, 2020 [271] (provided in Appendix A), we estimated interaction strengths based on phytoplankton abundance time series, and found intraspecific interaction strengths an order of magnitude higher than interspecific ones, as well as intra-to-inter ratios between 5 and 50 in other phytoplankton studies. These estimates, however, only reflect apparent interaction strengths, without knowing which mechanisms and dimensions of the niche (e.g., which specific requirements/regulating factors) are at work.

While neutral and niche models seem to be complete opposite, a middle ground has appeared, coined emergent neutrality [160, 298]. In this framework, species do not evolve strictly separate niches, but organize as several clumps along a niche axis, with niche differentiation being the main mechanism maintaining different clumps, and neutrality governing dynamics within a clump. This organization emerges as a very long transient in models [298, 301], but also appears in observed phytoplankton communities [302]. Using parameters based on phytoplankton communities, we showed in a previous work that clumps along a given niche axis (temperature), where quasi-neutrality is at play, could be maintained especially in the presence of strong niche differentiation over other, non explicitly modeled, traits, when environmental fluctuations were seasonal [270]. Whether the emergent neutrality hypothesis stands on its own, or is a product of hidden niches, is still in debate [27], which calls for additional exploration of coexistence mechanisms.

Some models incorporate most of the hypotheses presented above to describe the whole dynamics of ecosystems, by taking into account nutrients, phytoplankton, zooplankton, and sometimes even recycling through a detritus compartment. Such models are called NPZD models (see box 1.4) and can either be local (no spatialisation) or in one to three dimensions. Recent models distinguish diatoms, dinoflagellates and nanoflagellates [235], or separate different niches related to, for example, light use [10], or resource consumption (uptake and saturation) and edibility/nutritional value for zooplankton in addition to light [312]. In the latter model, Spaak *et al.* showed that top-down control has a stronger effect than resource competition and phytoplankton niche differences on community diversity, but final richness was only between 1 and 5 species, even though simulations started with 20 phytoplankton species. However, this model lacked temporal fluctuations. A world-scale model including a Kill-The-Winner response and seasonal fluctuations of nutrients was able to maintain 48 species

of phytoplankton on average, out of the intial 64 [326].

The role of hydrodynamic variations was shown in [209, 210], where a regional-scale model was used to describe eddies and fronts at the mesoscale ($\approx$ 100 km) in addition to nutrient, light and zooplankton niches. Mesoscale eddies acted as shelters for less competitive species, increasing the diversity in the regions they passed through, but eddy cores were less diverse, as keeping species in a smaller space led to more competitive exclusion. Meanwhile, fronts were diversity hot-spot. Overall, while local diversity increased when taking into account eddy advection, regional diversity remained the same, and turbulence-induced variability could only create different diversity patterns in space. The model was not able to maintain all of its species: on average, only 20 out of 100 phytoplankton species remained at the end of the simulations. In a model taking into account the same types of niches, the tidal cycle was introduced to induce high-frequency temporal variability, leading to the survival of 45 species out 120 [68]. As the timescales associated with the tidal cycle and phytoplankton growth are comparable, this tends to support the hypothesis of diversity being strongly linked with non-equilibrium conditions.

Box 1.4: Nutrient-Phytoplankton-Zooplankton-Detritus model

The basis for many phytoplankton models is the Nutrient-Phytoplankton-Zooplankton-Detritus (NPZD) model [126, 282, 171], which schematically writes as eq. 1.6. It can either be used in isolation, for theoretical purposes, or constitute the biogeochemical / ecological part of a coupled physical-biological model of the ocean at global or regional scales.

$$\begin{aligned} \frac{dN}{dt} &= -f(I)g(N)P + k(D)D \\ \frac{dP}{dt} &= f(I)g(N)P - h(P)Z - i(P)P \\ \frac{dZ}{dt} &= \gamma Z h(P) - j(Z)Z \\ \frac{dD}{dt} &= i(P)P + j(Z)Z + (1-\gamma)h(P)Z - k(D)D \end{aligned} \tag{1.6}$$

where N, P, Z, D represent nutrient, phytoplankton, zooplankton and detritus concentrations respectively, $f(I)$ is the phytoplankton response to irradiance (irradiance can be replaced by another environmental parameter characterising the environment), $g(N)$ is the uptake of nutrients by phytoplankton, $k(D)$ is the degradation of detritus, $h(P)$ is the grazing of phytoplankton by zooplankton, $i(P)$ and $j(Z)$ are loss terms due to excretion, death and predation by other organisms for phytoplankton and zooplankton respectively, and γ is the growth rate of zooplankton. It should be noted that the NPZ formulation (without detritus) tends to be more used than the NPZD one, which we mention here for the sake of completeness.

While NPZ(D) models can sustain a reasonably high number of species, this diversity is still much lower than the observed one, by at least one order of magnitude. This might be due to difficulties in parameterisation, but also to unknown phenomena or debatable assumptions that could be relaxed in order to better understand phytoplankton dynamics.

1.1.3 Discussing common assumptions

All models need to rely on a series of simplifying assumptions, some of them being more discutable than others in the model framework. In this thesis, we do not intend to question the results of models presented in the literature up to now, but to build upon them and relax some of these assumptions based on ecological considerations.

Population dynamics are usually considered with mean-field models: individuals are supposed to be everywhere and well-mixed, so that simulations can be based on average concentrations (or biomasses, abundances, etc.). These concentrations can vary in space when models are spatialised, but remain homogeneous at the model scale. However, the heterogeneous nature of phytoplankton spatial distributions has been shown for centuries. One of the first observations of 'streaks' of a different colour in the ocean dates back to 1773, by Captain James Cook (as reported by Bainbridge, 1957 [24]); they could have been the result of the accumulation of either diatoms or flagellates. The formation of 'bands, streams, lanes' is a recurrent observation in sailing accounts, including by Darwin (1839). These elongated coloured shapes would measure from 10 to 10^3 mi^2, with streaks of only a few ft embedded within them [24]. New techniques confirmed these first observations. Satellite data can now identify streaks of less than 1 km in their longest dimension [252]. These patches can maintain for a while, and follow the currents: satellites have been able to follow the displacement of a large phytoplankton patch (approximately 50 km wide) which maintained for several months along an $\approx$ 1,500 km track in the ocean [204], and another, smaller one, which lasted for several days and moved along a few-km coast [100]. At smaller scales, sampling has allowed the identification of circular patches measuring approximately 2 km in diameter [124]. Patches also appear along the vertical axis (the water column). Precise *in situ* fluorescence revealed 10-cm wide patches of higher abundances separated by tens of cm along the first m of the water column in the ocean [101, 237]. Going even further down the length scale, samples of bacterioplankton reveal not only microscale abundance patchiness, but also a patchy distribution of species: patches can be highly dissimilar in their species composition, even along a few mm [217]. All these observations, with smaller and smaller patches embedded in larger patches, show that phytoplankton organisms seem to have a somehow fractal distribution: the variance of concentrations follow a power-law of the observation scale [95] (see Section 1.3.1). As spatial models usually cover large distances, especially NPZD models which tend to focus on global or regional scales covering up to thousands of km, these models cannot reach the cm-scale discreteness of phytoplankton, at which competition for resource patches, or differential distributions of species leading to differ-

ential interactions, can occur. Non-spatial models, or spatial models at too large a scale, which rely on mean-field dynamics, can have very different results from spatial models focusing on the individual scales and potential patchiness due to limited dispersal: even considering a single species, population abundance can either increase or decrease drastically when comparing a non-spatial and a small-scale spatial models [201]. This motivates us to model multispecies communities at the microscale.

A second common assumption in phytoplankton literature is that all species compete together for different resources. Many models only consider competition for nutrients, which might be too restrictive for different reasons: (a) nutrients may not always be limiting, (b) other processes can modify observed interactions and (c) interactions are not necessarily competitive. Firstly, the assumption that nutrients are limiting may depend on the type of environment considered. Although there is some evidence for iron deficiency in the ocean [58], and for -rare- phosphorus and nitrogen limitations in few oligotrophic lakes (Soto et al. 1994, Diaz and Pedrozo 1996, cited by Reynolds, 2006 [285], p. 162), we have shown for a coastal environment, benefiting from terrestrial inputs such as artificial agricultural inputs, that nutrients do not limit phytoplankton growth [32]. Secondly, most of the models presented above focused on competition. They either considered interactions through nutrient or light depletion, when they modeled an explicit compartment for the limiting resource, or through phenomenological interaction parameters reducing a species' growth rate in response to other species' abundance or biomass increase, in absence of a resource compartment. This latter way of modeling relies on so-called 'apparent interactions', which means that modeled interactions are an aggregate of all processes taking place that could lead to a species affecting another species' growth rate. Among these processes, apparent interactions between phytoplankton species may result as much from predation or parasitism as they do from competition. We mentioned above models including predators which ended up with more diverse communities when taking into account predator switching diet. Indeed, a predator switching its consumption towards the most abundant species alleviates the pressure on less abundant species: what we finally observe is that an increase in the abundance of species 1, which leads to a predator consuming more of species 1, lessens predation on a less abundant species 2, meaning that species 1 has a positive effect on species 2 [1]. This leads to the third point: apparent interactions can be facilitative. This goes against the usual understanding of coexistence within a single trophic level: while facilitative and mutualistic interactions are known to exist, they are often excluded from coexistence theory, or only swiftly touched upon, so much so that many models do not consider positive interactions between species to be possible. For instance, multivariate autoregressive models, a common type of statistical models used to infer interaction strengths from long-term time series of phytoplankton abundances, usually force interactions to be negative by setting the positive estimates to 0 (among others, [175, 142]). When removing this assumption, we found in previous works that there were positive apparent interactions between species in all phytoplankton communities, and that these facilitative interactions could reach 70% of all in-

teractions [32, 271] (the latter is provided in Appendix A). Coexistence might therefore be dependent upon facilitation and mutualism, in addition to competition.

The third common assumption that we want to challenge here is based on demography. Models in which a species' population is defined through a single state variable (the biomass, concentration or abundance of the species) presuppose that organisms are always in the same physiological state, and have the same behaviour. This is, in fact, not true: the life cycle of a phytoplankton organism actually comprises several steps. Recent reviews have shown that a dormant stage, defined as a stage during which cells minimize their metabolic activity, often to resist harsh environmental conditions, is widespread in phytoplankton life histories [21, 111]. The existence of this stage is not a new discovery: dormant cells of most common species have been found for many years [234, 323] but their prevalence might have been overlooked, and their longevity was most likely underestimated, as recent studies show that thousand year-old cells can still be revived [294]. The accumulation of dormant cells at the bottom of the water column creates a so-called 'seed bank', an analogy with terrestrial plants which form stocks of seeds acting as a buffer through time. Phytoplankton dormant stages can be produced either by sexual reproduction or asexual processes [111]. Sexual reproduction might be surprising for organisms which spend most of their life reproducing by fission, but is actually common. In addition to contributing to evolution through genetic recombination and mutation, sexual reproduction is crucial for diatoms, whose silicon-based structure shrinks with mitotic divisions, until they need to reproduce with another diatom to form a spore benefiting from the material of both parents [105]. Sexual reproduction also happens as part of the classic life cycle of dinoflagellates, creating cysts [136]. Asexual cyst and spore formation can happen in response to environmental conditions (change in nutrients, irradiance, or temperature [136]), but also without known triggers [111]. All types of dormant stages can have long-term effects on the dynamics and survival of the populations: diverting part of the growing population to the seed bank reduces short-term fitness, but ensures that part of the dormant cells will be able to develop later on. This supports the maintenance of genetic material and diversity over long periods of time, as well as ecosystem resilience. In a system where a species whose pelagic cells have undergone competitive exclusion can reoccur in the environment after germination of its spores or cysts, diversity is more likely to maintain than in a system where a species is extinct when its pelagic cells have disappeared.

In this thesis, I aim to alleviate the aforementioned assumptions. Since they are of distinct nature, the following chapters are not directly related to each other. In Chapter 2, I question the effect of the seed bank on diversity maintenance, in a community where interaction strengths are inspired by our previous statistical analyses of time-series of observed abundances ([271], in Appendix A), and include facilitation. This model assumes mean-field dynamics and consider population abundances rather than individuals. In Chapters 3 and 4, I study only the fraction of organisms in suspension in the water column, and build a three-dimension, individual-based

model of organisms being displaced by oceanic currents, in order to reproduce their patchiness at the microscale. I then examine the potential effects of the spatial distribution on interactions between individuals, and possibly on coexistence of different species.

In the following, I provide an overview of what is known about seed bank models, and spatial, individual-based models, in theoretical ecology in general, and in phytoplankton communities in particular. The aim is to familiarize the readers with concepts that are used, but less detailed, in the following chapters.

1.2 Seed bank effect on coexistence

1.2.1 Diversity in terrestrial plant studies including a seed bank

Coexistence studies have shown multiple times, in many systems, the beneficial effect of modeling different stages of the life cycle on diversity (e.g., [239, 131]). Different terrestrial plant models have been designed to study the dormant life stage specifically; some of their main results are described here. The seed stage is of specific interest for terrestrial plants because seeds' ecophysiology enables plants to maintain through time and disperse to invade new spaces.

Dormancy duration is key for many species. As defined in [5], we consider here that a seed bank is formed when seeds can germinate several growing seasons after their formation, which means that they outlast the generation time of the aboveground vegetation. In this case, aboveground populations are the result of seed production through several previous seasons. This introduces lags in the population dynamics, which, for a species taken in isolation, can in turn stabilize the population by smoothing out the impact of any given year [261]. Plants that divert part of their seed production to several growing seasons reduce their short-term fitness and risk seed decay over time, but also increase their resistance to bad environmental conditions during which the survival of the aboveground vegetation is compromised. This process is called bet-hedging [81, 64]. Additional processes related to dormancy emerge in multispecies communities, as diversity is not the simple result of isolated species maintenance. Seed bank diversity arises from variations of regional aboveground diversity during potentially long periods of time, which means that the species composition of a seed bank does not necessarily synchronize with the composition of the aboveground vegetation [139, 94]. The seed bank is also less sensitive to long-term changes in community composition than its aboveground counterpart [94]. This allows the survival of species in their seed form, even though they might be filtered out at emergence when environmental conditions do not match their niches, or competition from already installed species is too strong. Models show that a species with a seed bank competing with a species without a seed bank has a higher relative fitness, and is therefore able to survive despite other less competitive traits [5, 207]. In communities where all species form a seed bank, results are more ambiguous: while Pacala [261] found that dormancy does not

increase the invasibility potential of a species in a community when compared to a case without dormancy, Ellner showed that different investements in seed production and survivorship may maintain coexistence [112], which is another way life history strategies encourage diversity. Furthermore, dormancy is a major player in the presence of environmental fluctuations, which foster the storage effect. Field experiments show that the buffering effect of the seed bank, higher germination in years of reproductive success (i.e., covariance between the effects of competition and environment) and differential species responses to the environment are all present for both annual and perennial plants [264, 117, 13], meeting the conditions for the storage effect to maintain diversity through time. The spatial dimension of coexistence, however, should not be forgotten.

For sessile organisms such as plants, the ability to colonize new environments via seed dispersal (before they are stored in the soil) is crucial to survive. Most populations are organized in metacommunities (a set of communities which are linked together by movement, with different, potentially interacting, species) in which competition and dispersal are local, i.e., have limited ranges of action. The competition-colonization hypothesis posits that coexistence can emerge when the strongest competitor has a low colonizing ability [50]. The dominant species, however, can have more reserves and take over a specific space once established. This is supported by the observed trade-off between seed size and seed production: a smaller seed disperses more but has less reserves, which makes it less costly to produce in great quantity, but also a lesser competitor [283]. In many plant communities, the 'patch dynamics' is able to explain the community stability at the regional scale as an aggregate of non-equilibria at the patch scale. A species which has been moved in an empty space, for example, would persist in a refuge without suffering competitive effects from heterospecifics, even though, at the regional scale, it is considered to cohabit with them [306, 11].

Dormancy can be seen as a temporal dispersal mechanism; the interaction of dormancy and dispersal is therefore expected to be complex. Wisnoski and colleagues [336, 337] built a metacommunity model which uncovered interactive, non-linear effects of temporal and spatial dispersal on metacommunity diversity. For instance, under equal competition, local (alpha-) diversity was a dome-shaped function of dispersal when seed survival was high, but remained low with low survival. With niche differences, local diversity remained a saturating function of dispersal, regardless of dormancy. In both competition regime, regional (gamma-) diversity was a saturating decreasing function of dispersal but increased with seed survival [337]. Empirical data show that dispersers that can enter dormancy are present in more sites than dispersers lacking this ability [336], but also that different types of organisms have specific competition structure, dispersal and dormancy abilities, which may lead to different diversity patterns [337].

Taking into account dormancy in terrestrial plant communities allows us to consider bet-hedging strategies, buffering of environmental conditions, potential fostering of the storage effect, and interaction with dispersal, all phenomena which can explain diversity at local or regional scales. Similar mechanisms can be examined for aquatic plants such as phytoplankton.

1.2.2 A short review of phytoplankton models with a seed bank

A few phytoplankton models including a seed bank exist, which are summarized here[1]. Most of these models, however, tend to focus on a single species, sometimes on low-diversity communities. To my knowledge, they also tend to be less focused on theoretical mechanisms, and to provide instead more applied studies which aim to capture the spatial and temporal dynamics of toxic species forming harmful blooms.

Some single-species models are used to track the displacements of the pelagic cells after seed germination, and highlight preferential zones for seed accumulation [229, 12, 340]. A bloom may be able to emerge even when only a seed bank is present at the beginning of a simulation (no pelagic cell in the water column) [340]. The removal of the seed bank, however, may lead to the disparition of a species, confirming its role as an 'insurance' against environmental conditions which can drive the pelagic population to extinction [152, 116]. A model of cyanobacteria was also used to show a covariation between environmental conditions and germination: the participation of newly germinated cells from the seed bank increased when environmental quality decreased and there were less pelagic cells in the water column [153].

These results, however, do not give a lot of information on the way diversity may be maintained (or not) when many species compete. The multispecific phytoplankton models with a seed bank describe only a few large functional groups, far from the diversity observed in field data. In line with terrestrial plant studies, some of these models consider only one species with a seed bank, and the others without, and they either have very different temperature, light and nutrient niches [341] or plugged-in stabilizing interactions [116]. The contribution of the seed bank to diversity is unclear because these models would likely have allowed for coexistence even in the absence of any seed bank. In the NPZ+Virus model of [122], there are three different groups of phytoplankton: two of them, including the only one that can form cysts, can be exploited by a virus whereas the third one cannot. This model is used to explain bloom dynamics; specifically, the seed bank balances the effect of viral infection. Phytoplankton functional types are nevertheless too different from one another to conclude on coexistence mechanisms when all species have a seed bank. In [203], all three functional groups (diatoms, dinoflagellates and cyanobacteria) have resting stages and compete for the same pool of nutrient. Diatoms are characterized by a higher growth rate and lower germination rate than dinoflagellates. As temperature increases, the time between germination and growth of dinoflagellates decreases, allowing them to colonize the environment more efficiently than the diatoms. This model therefore makes use of the seed bank existence and germination over time as an explanation for the species composition change with global warming. However, these functional groups remain too large to study coexistence of more similar species.

There are common results in terrestrial and aquatic plant studies: the buffering effect of

[1] I use the word seed to qualify cyst (for dinoflagellates) and spore (for diatoms) in the following paragraphs, in order to keep in mind the parallel with other terrestrial plants.

the seed bank, and the potential coexistence of species having different investments in their seed bank. Meanwhile, they lack a focus on coexistence in diverse communities: phytoplankton diversity can reach tens of species in any given sample, which is never reflected in models including a seed bank. This is the goal of our first study (see Chapter 2).

1.3 Models of phytoplankton spatial distribution

This section is dedicated to phytoplankton spatial modeling: an overview of large-scale approaches and their limitations is used to support the potential of spatial individual-based models operating at much smaller scales. Spatial statistics tools used to analyse them are introduced. Existing spatial individual-based models designed for phytoplankters are then summarized.

1.3.1 Macro- and meso-scale approaches

Spatial phytoplankton models have described phytoplankton patches for more than half a century at macro- (hundreds to thousands of km), meso- (tens of km) and submeso- (under a km) scales. In any case, they take into account the balance between biological processes (e.g., growth) and hydrodynamics (turbulence observed in aquatic environment, see box 1.5) to represent the distribution of the phytoplankton biomass.

There are two main ways of describing fluid dynamics, here used for phytoplankton concentrations. The first approach is the Eulerian description, where concentrations are seen as fields which vary in space and time. The other technique is the Lagrangian method which focuses on the behaviour of a given phytoplankton parcel (an organism, or a collection of them which share the same characteristics - in this case, we describe more specifically a Lagrangian-Ensemble model) moving around. Macro- and mesoscale models are, most of the time, Eulerian descriptions where phytoplankton concentrations are continuous functions of space and time (but see [338, 265] for examples of Lagrangian-Ensemble models at the (sub)mesoscale). Among the first attempts to model phytoplankton patchiness, the KISS model(s) (named after its authors, Kierstead, Slobodkin and Skellam [308, 190]) has been used to define minimal sizes for patches to maintain (see box 1.6), which vary with the processes taken into account (diffusion-only, advection and diffusion, predation, etc.). The classical KISS model characterizes the size of a single patch in the horizontal dimension only. The vertical structure, however, adds yet another dimension to the complexity of phytoplankton distribution patterns [103]. Phytoplankton patchiness along the vertical dimension is well-known: the so-called 'thin layers' are horizontal strata of higher phytoplankton concentrations, whose thickness varies between a few cm and a few m. Hydrodynamic processes, including vertical velocities, stratification and shears, are among the mechanisms used in models to explain phytoplankton growth and concentration increase at certain depths [103].

Generally speaking, the distribution of phytoplankton concentrations is well described in

macro- and mesoscale models coupling physical and biological (e.g., temperature-dependent growth rate, competition) processes. In these cases, concentrations most of the time follow the abiotic patterns in the environment (see for instance [124, 230]; in [229], a decomposition of the growth rate dependence on abiotic factors also shows this pattern), which is confirmed by observations at the mesoscale [224]. Still, the dynamics of these models are not necessarily adapted to much finer scale variability. While observations [276] have shown that the spectrum of variance of chlorophyll (as a proxy for the variance of phytoplankton biomass) is a -5/3 power law of the wavenumber at the submesoscale, similar to the power law of turbulence, theoretical considerations, supported by other data, show otherwise [95]. Denman & Platt [95] have indeed uncovered a discontinuity in the variance spectrum, where the power coefficient should fall from -1 to values between -2 and -3 for larger wavenumbers. The break happens at a critical wavelength estimated between 0.2 and 20 km in the upper ocean, with an average of 1 km. The authors have suggested that the switch between turbulence and demography control may happen at different critical wavelengths for species with different growth rates, which would create spatial niches in favour of coexistence. They also consider that phytoplankton growth rates have more control on spatial variability at larger scales. However, reviewing a series of observations in the ocean, Daly and Smith [92] concluded that "biological processes may be more important at smaller scales where behavior such as vertical migration and predation may control the location and production of plankton". When modeling at the microscale (i.e., at the scale of the individual), encounter rates with predators vary with the relative speed of the organisms and the intensity of turbulence [289, 192], reproduction balances out separation by turbulence [345], phytoplankton voluntary movement may exceed passive advection or interact with turbulence to create patches [102, 59]. Moreover, chlorophyll or biomass measures cannot account for species diversity, nor can their variability distinguish intra- from interspecific aggregation. We therefore need to model individuals, not only average concentrations or biomasses as a function of space. This is the goal of individual-based models.

Box 1.5: Turbulence and viscosity

Big whirls have little whirls that feed on their velocity, and little whirls have lesser whirls and so on to viscosity.

Lewis Fry Richardson [286]

While this is not the place for a comprehensive review of fluid hydrodynamics (the 1600 pages of [240] should deter from such attempt), we need to define, in very broad terms, the most common concepts and numbers used to characterize the environment in which phytoplankton dwells.

Turbulent and laminar regimes can be seen as the two ends of the fluid dynamics spectrum; they are not opposed, but depend on the scale of reference used to study a fluid. Ocean currents, such as the Gulf Stream, observed at large scales, are turbulent: they are made of eddies formed under the sheer stress of mechanic energy applied at the air-water interface by the wind. Turbulence is characterized by its randomness, diffusivity (transport of mass, heat, or momentum due to movement at scales several orders of magnitude above molecular diffusion), and the presence of three-dimensional vorticity fluctuations [318]. At the hydrodynamic level, its length and time scales of variation are much larger than any molecular scale: turbulence is considered as a continuum, allowing continuous equations of fluid mechanics to hold. The viscosity of the environment opposes the movement imposed by turbulence, and mechanical energy eventually dissipates into internal energy of the fluid (heat). The dissipative nature of turbulence is another of its main characteristics, studied through the Kolmogorov spectrum. The kinetic energy generating turbulence decreases with the scale at which the fluid is considered, following a -5/3 power law of the wavenumber, as larger eddies cascade into smaller and smaller eddies until they are overwhelmed by viscosity. When viscosity takes over, the fluid is mostly laminar, which can be understood as parallel layers of fluid moving in the same direction without mixing.

The ratio between turbulent and viscous forces is defined by the Reynolds number

$$\mathrm{Re} = \frac{UL}{\nu}, \tag{1.7}$$

where U is the flow speed (m s^{-1}), L is the characteristic dimension (m) and ν is the kinematic viscosity of the fluid (m^2 s^{-1}). At 20°C, sea water viscosity is around 10^{-6} m^2 s^{-1}. Viscosity takes over for a Reynolds number equal or below 1, for dimensions equal or below the smallest eddy size. The corresponding length scale is called the Kolmogorov scale

$$\eta = \left(\frac{\nu^3}{\epsilon}\right)^{1/4}, \tag{1.8}$$

where ϵ is the dissipation rate of turbulent kinetic energy ($m^2\ s^{-3}$). In the open ocean, ϵ varies between 10^{-10} and 10^{-6} $m^2\ s^{-3}$ [33] (with extremes between 10^{-11} and 10^{-5} $m^2\ s^{-3}$ [130]), leading to Kolmogorov scales between 1 mm and 1 cm (Fig. 1.1).
The size of microphytoplankton, between 20 and 200 µm, is much smaller than the smallest possible eddy in their environment, and their Reynolds number varies between 10^{-2} and 10^{-3} [187]. They therefore perceive the world as mostly viscous, while being displaced by advective currents at a much larger scale. The movements and spatial distributions of individuals are therefore determined within the referential of the smallest eddy, keeping in mind that it is then moved around within larger eddies. Phytoplankton organisms are still affected by turbulent conditions, whether it be through its effect on predation, competition, or nutrient uptake [156, 269].
Concentration variations of nutrient, or any solute, happen at much lower scales than η. The smallest length of a nutrient patch, before diffusion, is the Batchelor scale

$$\eta_b = \left(\frac{\nu D_m^2}{\epsilon}\right)^{1/4}, \tag{1.9}$$

where D_m is the molecular diffusivity of the solute ($m^2\ s^{-1}$), which is around 10^{-9} $m^2\ s^{-1}$, depending on the molecule considered. Batchelor scales in the ocean vary between 20 and 300 µm.
As mentioned before, turbulence can increase the transport of a solute; the Péclet number Pe is used to quantify the effectiveness of advective transport compared to diffusive transport for a given length scale r [187].

$$Pe = \frac{Ur}{D_m} \tag{1.10}$$

The Sherwood number Sh is then defined as the ratio between convective transfer and diffusive transfer at a boundary surface and can be computed from the Péclet number [187].

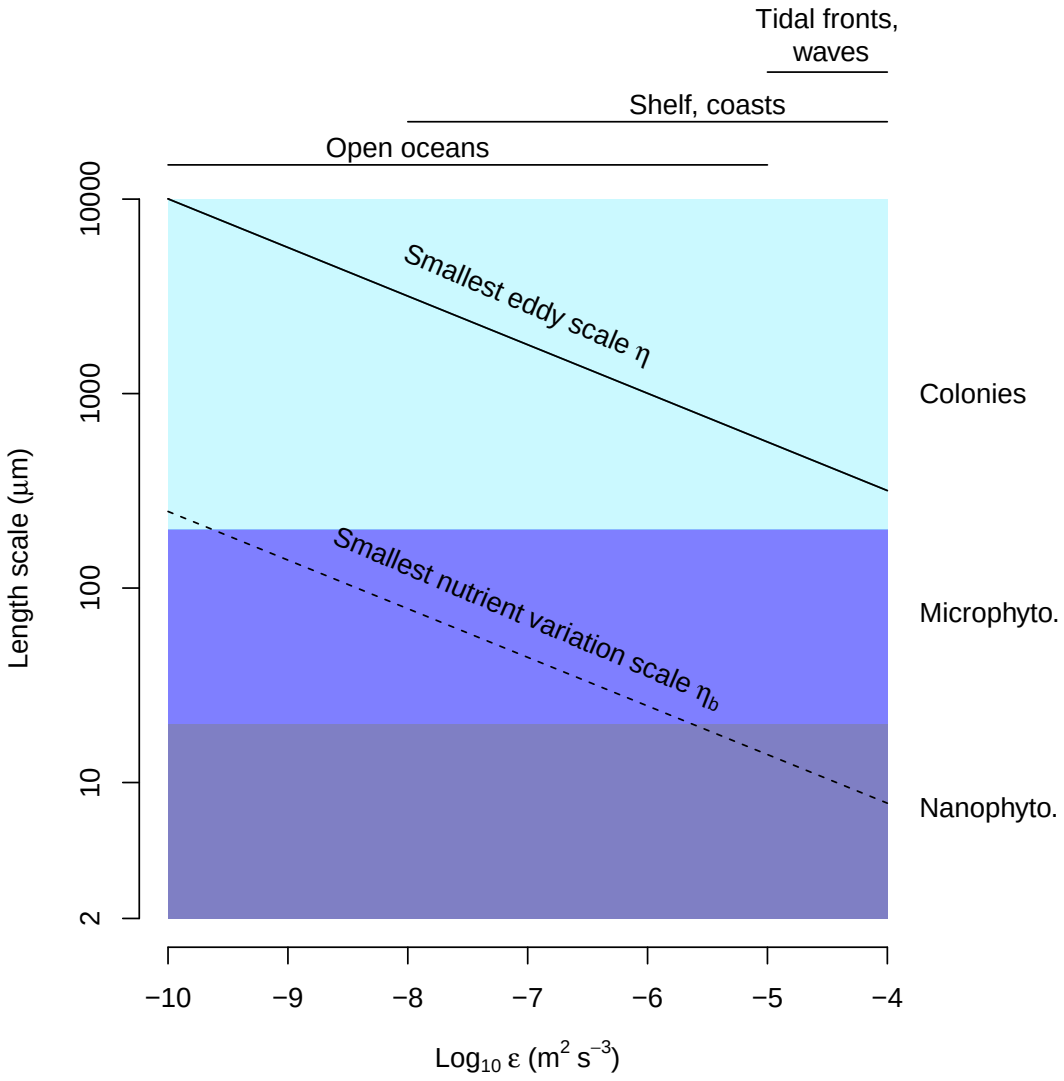

Figure 1.1: Length scales associated with the Kolmogorov (η) and Batchelor (η_b) scales as a function of turbulent kinetic energy dissipation rates ϵ, in different marine environments, compared to phytoplankton size classes. (Reproduction of Fig. 1 in [33], with characteristic values of turbulence in different environments from [130])

Box 1.6: KISS model(s)

One of the first approaches of phytoplankton patchiness is the KISS model (named after its authors, H. Kierstead & L.B. Slobodkin [190] and J.G. Skellam [308]), which focuses on the minimal size of a viable patch. The framework is simple. In a one-dimensional model, assume a patch of size L where there is a concentration P of phytoplankton and where diffusion and growth take place while advection is ignored. The patch is assumed to be surrounded by unsuitable water ($P(0) = P(L) = 0$); this could be due to resource limitation or advection of phytoplankton outside of the patch. Starting at $t = 0$ with $P(x) = f(x)$ ($f(x)$ non specified), the spatial dynamics in one dimension follow

$$\frac{\partial P}{\partial t} = D\frac{\partial^2 P}{\partial x^2} + \alpha P \tag{1.11}$$

where D is the diffusivity of the environment and α is the phytoplankton growth rate. With this model, a phytoplankton patch can only maintain despite diffusion if $L \geq \pi\sqrt{\frac{D}{\alpha}}$. In two dimensions, for a circular patch of radius R, $R \geq 2.4048\sqrt{\frac{D}{\alpha}}$. Scale-dependent diffusivity can be added to the model, which means that the intensity of the diffusivity parameter depends on the scale of the observation. This leads to minimum patch dimensions around 2 km for $\alpha = 1$ division per day (Steele 1976, cited in [257]), which correspond to observations at large scales (see Section 1.1.3).

This model is obviously highly simplified. If advection is added, using a fluid velocity U, eq. 1.11 can be rewritten as

$$\frac{\partial P}{\partial t} = D\frac{\partial^2 P}{\partial x^2} + U\frac{\partial P}{\partial x} + \alpha P. \tag{1.12}$$

If the flow converges to the center of the patch, this can greatly reduce the minimal size of the patch, but if the flow diverges, even small velocity values (a few cm per second) could prevent the patch maintenance [257].

Predatory control on the population can also be introduced to this model [258], as

$$\begin{array}{rcl} \frac{\partial P}{\partial t} & = & D_P\frac{\partial^2 P}{\partial x^2} + \alpha_P P - \beta_P PZ \\ \frac{\partial Z}{\partial t} & = & D_Z\frac{\partial^2 Z}{\partial x^2} - \alpha_Z Z + \beta_Z PZ \end{array} \tag{1.13}$$

with α_p the phytoplankton growth rate, β_P its loss to predation by zooplankton, α_Z the loss rate of zooplankton and β_Z its growth rate due to phytoplankton consumption.

In this case, the system can either converge to the equilibrium densities they would have had without diffusion (α_Z/β_Z and α_P/β_P for P and Z respectively), or to spatially uniform temporal oscillations (no patch). This remains true in the presence of advection, nonlinear interactions, and different initial conditions or boundary conditions (absorbing or reflecting).

Finally, in the case of a community with S species, eq. 1.11 is modified to eq. 1.14 [258].

$$\frac{\partial P_i}{\partial t} = D_i \frac{\partial^2 P_i}{\partial x^2} + G_i(P_j)P_i \quad i,j = 1,2,...,S \tag{1.14}$$

where D_i is the species-specific diffusivity and $G_i(P_j)$ is the growth rate of species i in presence of species j with concentration P_j.
This leads to a minimal size $L = \pi\sqrt{\frac{D_{\min}}{G_{\max}(0)}}$ with $G_{\max}(0)$ the maximal growth rate among the community in the absence of competitors.
Note that defining minimal patch sizes does not give information on the distribution of organisms within a patch.

1.3.2 (Spatial) individual-based models: definition and analysis

Definition Before focusing on individual-based models (IBMs) for phytoplankton communities, and the additional information they may provide, I need to define what IBMs are and how their results compare with those of mean-field models in other classical systems.

An individual-based model, also known as an agent-based model, is a model which focuses on the behaviour and/or dynamics of each organism forming part of a community, instead of studying the dynamics of an ensemble of supposedly identical individuals. In an IBM, each individual has its own properties, which usually include size, nutrient content, reproductive characteristics – possibilities are endless, depending on the main interest of the study. Most of the time, the goal remains to eventually study the system emerging at the population level (abundance, biomass or density) from individuals' stochastic functioning [150].

While the IBM approach has been debated for its potential lack of generality in ecological theory, many studies have proposed situations justifying the rise of this relatively new approach [167, 185, 216, 150, 151]. In the specific case of microbes, including phytoplankton, a review of 46 articles has shown that the main reasons for choosing individual-based modeling were population heterogeneity (intra-population variability of characteristics such as individual growth rate or nutrient content), and emergence of new properties in the population from individual behaviour (e.g., intracellular mechanics explaining variability and resource requirements at the population level) [150]. IBMs are also supposed to better handle the chaotic behaviours of ecological systems and the lack of population mixing leading to different interactions between individuals (assuming interactions are local) [185]. A spatial dimension (from one to three) is often introduced in IBMs, to consider distances between individuals and limited movements, and therefore the effect of spatial heterogeneity.

Comparisons of spatial IBMs and non-spatial models have shown that spatial heterogeneity could sometimes lead to very different, sometimes opposite, population dynamics. In Law *et al.*, 2003 [201], a spatial IBM where individuals could only disperse over small distances and

compete with organisms close to them (i.e., small dispersion and competition ranges) predicted the extinction of a population, which was not expected to happen in a non-spatial setting. On the contrary, increasing the dispersal range and maintaining a small competition range led to a higher equilibrium abundance than in a non-spatial model. In Birch & Young, 2006 [47], the strength of dispersal has also emerged as a key factor; in their case, weak dispersion was associated with higher abundances than in an equivalent mean-field model. Spatial heterogeneity can therefore partly determine the fate of a population. This heterogeneity can be quantified with spatial statistics belonging to the field of point process theory.

Spatial statistics applied to IBMs I present here the basic elements of spatial statistics applied to spatial IBMs. This section introduces tools that are reminded in the Methods in Chapter 4, with additional details regarding the computation of some of these statistics (see box 1.7).

The field of spatial statistics is well adapted to analyse spatial stochastic IBMs. The state of a system generated by a spatial stochastic IBM can be described as a collection of S populations, where the population of species i is made of k_i individuals randomly distributed in space, with positions $\boldsymbol{X}_i = [\boldsymbol{x}_{1,i}, \boldsymbol{x}_{2,i}, ...\boldsymbol{x}_{k_i,i}]$. This corresponds to the realization of a spatial point process, defined by Illian *et al.*, 2008 [169] as a "mathematical model that describes the arrangement of objects that are irregularly or randomly distributed in the plane or in space" (random distributions can also be a function of time). There are many metrics developed to analyse a spatial point process. One of the most common methods uses the moment measures (hereafter, moments) of the process. These moments are the spatial statistic equivalents of the moments of random variables: the first moment corresponds to the average number of points in space, the second moment, to the variance of the positions in space, and so on. Evaluating moments of a spatial point process can give a complete view of the expected behaviour of the model.

The moments of a spatial point process can often be derived analytically. The specific case of an individual-based model including stochastic birth, death and movements satisfying a Markovian property (i.e., rates of change at any given time depend only on the current state, not on past states), constitutes a classical model in ecology, and the focus of Chapters 3 and 4. The dynamics of the patterns of the corresponding model can be characterized by a master equation, the rate of change of the probability density for a given distribution [99]. Moments can then be derived from this equation. Other examples of moment derivations applied to IBMs with birth-death-movements are numerous in the ecological literature (e.g., [52, 47, 254, 31, 275]), and help to summarize the emerging patterns in spatial distributions [51].

In many models, the strength of competition between individuals (more precisely, the impact on mortality, fecundity or establishment due to local densities) is a function of the distance between points, often called the kernel of interactions [52, 201]. This is why metrics close to spatial moments, which characterize the relative positions of particles, can be useful to understand population dynamics. Ripley's K-function $K(r)$ (hereafter, K-function, or K) and

the pair correlation function (hereafter, pcf, or g) are among the most well-known (for an example of derivation, see box 1.7). Firstly, the K-function measures the expected number of points that are located within a distance r of a given particle, i.e.,

$$\forall r \geq 0,\ K(r) = \frac{1}{C}\mathbb{E}\left(N\left(b(o,r)\backslash\{o\}\right)\right) \tag{1.15}$$

where C is the expected point concentration, and $N\left(b(o,r)\backslash\{o\}\right)$ is the number of points of the process N in the (hyper)sphere of radius r centered on o, not counting o itself.

The K-function, being an integral of all points within a volume, can be sensitive to specific regions of high or low density, which can eclipse smaller variations at specific scales. This is why the pcf $g(r)$, is recommended [169, 188], since it focuses on the pairs of points that are at a distance exactly r from each other. It is defined as

$$\forall r \geq 0, g(r) = \frac{K'(r)}{db_d r^{d-1}}) \tag{1.16}$$

where b_d is the volume of the unit ball in $\mathbb{R}^d$ [169]. Organisms are said to be aggregated (respectively, segregated) if the pcf of their distribution is above (respectively, below) 1, the expected value under the uniform distribution assumption.

Box 1.7: Derivation of K and g for a classical point process

We use here the derivation of the formula for Ripley's K-function and the pair correlation function in three dimensions, for the Thomas point process, as a didactic example to familiarize the reader with these metrics. A Thomas distribution is generated in two steps: (a) 'parents' are distributed uniformly in space, with intensity λ_p, then (b) 'daughter' points are distributed around each parent point, their locations following a Gaussian distribution centered on the parent with standard deviation σ. The number of daughter points per parent follows a Poisson distribution with intensity $\bar{c}$.

Ripley's K-function In [88], the K-function for a Neyman-Scott point process such as the Thomas one is defined as

$$K(r) = \frac{\pi^{d/2} r^d}{\Gamma(1+\frac{d}{2})} + \frac{E(N(N-1))F(r)}{\lambda_p \bar{c}^2} \tag{1.17}$$

where $E(N) \equiv \bar{c}$, $F(r)$ is the cumulative distribution function (cdf) of the distance between two events in the same cluster, and $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$.
First, we have $E(N(N-1)) = E(N^2) - E(N) = Var(N) + (E(N))^2 - E(N)$. With a Poisson distribution, $Var(N) = \bar{c}$, so $E(N(N-1)) = \bar{c}^2$, which leads to

$$K(r) = \frac{\pi^{d/2} r^d}{\Gamma(1 + \frac{d}{2})} + \frac{F(r)}{\lambda_p}. \tag{1.18}$$

Thus, in three dimensions,

$$K(r) = \frac{4}{3}\pi r^3 + \frac{F(r)}{\lambda_p}. \tag{1.19}$$

Let us compute the mean squared distance between two points belonging to the same cluster p_1 and p_2 at $(x_1, y_1, z_1)'$ and $(x_2, y_2, z_2)'$. Following the definition of the Thomas point process, $x_\cdot \sim \mathcal{N}(0, \sigma^2)$, $y_\cdot \sim \mathcal{N}(0, \sigma^2)$ and $z_\cdot \sim \mathcal{N}(0, \sigma^2)$. As variances are additive, $(x_1 - x_2)^2 \sim \mathcal{N}(0, \sigma^2 + \sigma^2)$, i.e., $\frac{(x_1 - x_2)^2}{2\sigma^2} \sim \mathcal{N}(0, 1)$. The reasoning is the same for y and z. We can define $w = \frac{1}{\sqrt{2}\sigma}\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$ with $W \sim \chi(3)$ the corresponding variable; the distance between p_1 and p_2 is $r = \sqrt{2}\sigma w$. The cdf for W, $F_3(w)$, is defined as

$$F_3(w) = \frac{1}{\Gamma(3/2)}\gamma(3/2, w^2/2) \tag{1.20}$$

where $\gamma(a, x) = \int_0^x t^{a-1}e^{-t}dt$ is the incomplete gamma function. Thus,

$$F_3(w) = \frac{2}{\sqrt{\pi}} \int_0^{w^2/2} \sqrt{t} e^{-t} dt. \tag{1.21}$$

Let us define $z^2 = t \Rightarrow dt = 2zdz$, then integrate eq. 1.21.

$$\begin{aligned}
F_3(w) &= \tfrac{4}{\sqrt{\pi}} \textstyle\int_0^{w/\sqrt{2}} \sqrt{z^2} e^{-z^2} z dz \\
&= \tfrac{4}{\sqrt{\pi}} \textstyle\int_0^{w/\sqrt{2}} z \left(z e^{-z^2}\right) dz \\
&= \tfrac{4}{\sqrt{\pi}} \left(\left[-\tfrac{1}{2} z e^{-z^2}\right]_0^{w/\sqrt{2}} + \tfrac{1}{2} \textstyle\int_0^{w/\sqrt{2}} e^{-z^2} dz\right) \\
&= \tfrac{4}{\sqrt{\pi}} \left(-\tfrac{1}{2}\tfrac{w}{\sqrt{2}} e^{-w^2/2} + \tfrac{1}{2}\tfrac{\sqrt{\pi}}{2} \operatorname{erf}\left(\tfrac{w}{\sqrt{2}}\right)\right) \\
&= \operatorname{erf}\left(\tfrac{w}{\sqrt{2}}\right) - \tfrac{\sqrt{2}}{\sqrt{\pi}} w e^{-w^2/2}
\end{aligned} \tag{1.22}$$

where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz$ is the error function.
The cdf of the distance r can then be defined as

$$F(r) = \operatorname{erf}\left(\frac{r}{2\sigma}\right) - \frac{1}{\sigma\sqrt{\pi}} r e^{-(r/2\sigma)^2}. \tag{1.23}$$

Combining all results,

$$K(r) = \frac{4}{3}\pi r^3 + \frac{1}{\lambda_p \sigma \sqrt{\pi}} \left(\sigma\sqrt{\pi} \operatorname{erf}\left(\frac{r}{2\sigma}\right) - r e^{-\left(\frac{r}{2\sigma}\right)^2}\right). \tag{1.24}$$

Pair correlation function By definition, $g(r) = \frac{K'(r)}{4\pi r^2}$. Derivation is simple, and we obtain

$$g(r) = 1 + \frac{1}{\lambda_p} \frac{1}{(4\pi\sigma^2)^{3/2}} e^{-\left(\frac{r^2}{4\sigma^2}\right)}. \tag{1.25}$$

While eqs. 1.15 and 1.16 are written for a single population (distances are computed between all pairs of individuals regardless of their species), the metrics can also be written as bivariate functions, which then characterize the distribution of individuals of a given species around an individual of another species. This is particularly interesting if we are to compare intraspecific and interspecific spatial patterns.

Keeping the hypothesis of distance-dependent interaction strengths, computation of expected spatial associations can in turn be related to classical coexistence theory, especially if distance distributions between conspecifics and heterospecifics differ. When computing K or g leads to predictions of aggregation of conspecifics, but not of heterospecifics, we might infer stronger intraspecific interactions than interspecific interactions, corresponding to a spatial niche (however, see [246, 245] for a more nuanced view, if interaction ranges differ between heterospecifics and conspecifics).

I have described here a class of mathematical models that are able to represent individuals distributed in space, going through stochastic processes, and the tools to analyse such models. I now need to determine how these models have been applied in phytoplankton communities to potentially pinpoint missing information regarding diversity maintenance.

1.3.3 Application of spatial IBMs to phytoplankton modelling

Some spatial individual-based models have been developed to study specifically phytoplankton dynamics and distributions in their environment at the microscale (corresponding to the size of microphytoplankton) in two or three dimensions. They can be seen as Lagrangian approaches of phytoplankton distributions, as opposed to the Eulerian description often used in larger-scale models. A particularity of IBMs for phytoplankton is that they must describe their ever-changing hydrodynamic environment: turbulence (advection and viscosity) always affects the particles' movements. The main differences between spatial phytoplankton IBMs then emerge from the way they model hydrodynamics and their focus on different ecological processes, which we summarize here. While different species are sometimes modeled, monospecific models are the most common. All models show at least some kind of particle clustering.

Hydrodynamics Hydrodynamic processes include both fluid velocity (advection) and viscosity of the environment, and can determine the way particles are displaced. There are two main ways to describe the velocity field, either based on the exact Navier-Stokes equations,

or on a simpler, less computationally costly, approximation of a turbulent flow (synthetic turbulence, or kinematic simulation, first proposed by Kraichnan [199]). In the first case, direct numerical simulations can describe the dynamics of the flow [59, 53]. In the second case, the turbulent velocity field is represented as a truncated Fourier series of wave vectors with amplitudes chosen to follow the turbulent energy spectrum, a method which can produce similar results to the Navier-Stokes equation [59]. At the lowest end of the turbulent energy spectrum, the Kolmogorov length scale, where turbulence collapses into viscosity, advection by turbulence can be approximated by a single mode, corresponding to the smallest wavelength [274, 345].

Individuals can also be displaced by diffusion, following a Brownian motion [345]; this is actually the only hydrodynamic process modeled in [47, 55]. Advection by synthetic turbulence and species-specific diffusion are modeled in Chapters 3 and 4, based on Young *et al.*, 2001 [345].

Organisms are displaced passively by the mechanisms described above. However, the organism's voluntary movements and/or inertia can interact with the fluid velocity; in this case, the Maxey-Riley equation is used to compute its acceleration [41, 53, 19]. Most of the models which consider phytoplankton motility do not take into account ecological processes: organisms do not reproduce nor die. In this case, aggregation emerges either from common responses to the environment (e.g., buoyancy increases or decreases when turbulence increases in [53, 19]), or from active movements. Organisms can indeed modify their orientation in response to other particles' presence in their detection range, and aggregation increases when the strength of interaction increases [59]. Phytoplankton motility itself can have an 'unmixing effect' (increasing patchiness) [102]. Phytoplankton patchy distribution can therefore be at least partially attributed to movement regulation only.

Ecological processes Some models take into account the demography of phytoplankton with different degrees of precision, including reproduction and death only [345], competition modeled as the increase in death rate with local density [47, 55], and potential cooperation between organisms [56]. Most of the time, local reproduction, competition and dispersal lead to aggregation.

In the set of single-species models proposed by Bouderbala and colleagues, where both interactions and motility are taken into account, the balance between chemo-attraction and competition leads to different patterns in organism aggregation. The intensity of competition itself is an important factor: in [55], weak-to-medium competition entices clustering while high competition prevents aggregates from maintaining. Implicit top-down effects are assumed in [56], where local densities reduce division due to competition, but decrease mortality due to predation (the assumption being that predators can more easily ingest small cells rather than large aggregates). Chemo-attraction then encourages cluster formation while reduction in mortality maintains the cluster. Results are similar when adding self-shading and diel vertical migration [57].

In a multi-species model coupling inertial forces in a chaotic flow to birth and death, Benczik

and colleagues [41] show that up to five species can coexist using only one resource. Species with lower fitnesses (modeled by higher death rates) still maintain thanks to chaotic flows displacing them differently: organisms of different species (characterized by their inertia) therefore occupy specific regions of the environment. This study is the only example of a community of more than two coexisting groups including hydrodynamics and ecological processes.

What can we learn on coexistence mechanisms from these models? Most models focus on the distribution of a single population, and therefore give no information on the way different aggregation or segregation patterns between species could affect a community. Models which consider different functional groups, and find intragroup aggregation, often describe completely opposite hydrodynamic traits: increased vs decreased buoyancy with turbulence [53, 19] or motile vs non-motile organisms [102] (see [41] for a slightly more refined description of inertia-related traits). However, field observations of rich communities show that traits such as size, related to hydrodynamics properties, are similar between species [302]; and coexistence of similar species is still not studied.

Overall, results are numerous for single-species patchiness, but an analytically-supported study of aggregation patterns in a rich community at the microscale, taking into account both hydrodynamics and demographics, is still missing. This is the endeavour of Chapters 3 and 4.

Chapter 2

Seed banks can help to maintain the diversity of interacting phytoplankton species

Coralie Picoche[1,*] & Frédéric Barraquand[1]
[1]Institute of Mathematics of Bordeaux, University of Bordeaux & CNRS
[*]corresponding author: coralie.picoche@u-bordeaux.fr

Abstract

Seed formation is part of the reproductive cycle, leading to the accumulation of resistance stages that can withstand harsh environmental conditions for long periods of time. At the community level, multiple species with such long-lasting life stages can be more likely to coexist. While the implications of this process for biodiversity have been studied in terrestrial plants, seed banks are usually neglected in phytoplankton multispecies dynamics models, in spite of widespread empirical evidence for such seed banks. In this study, we build a metacommunity model of interacting phytoplankton species, including a resting stage supplying the seed bank. The model is parameterized with empirically-driven growth rate functions and field-based interaction estimates, which include both facilitative and competitive interactions. Exchanges between compartments (coastal pelagic cells, coastal resting cells on the seabed, and open ocean pelagic cells) are controlled by hydrodynamical parameters to which the sensitivity of the model is assessed. We consider two models, i.e., with and without a saturating effect of the interactions on the growth rates. Our results are consistent between models, and show that a seed bank allows to maintain all species in the community over 30 years. Indeed, a fraction of the species are vulnerable to extinction at specific times within the year, but this process is buffered by their survival in their resting stage. We thus highlight the potential role of the seed bank in the recurrent re-invasion of the coastal community, and of coastal environments in re-seeding oceanic regions. Moreover, the seed bank enables populations to tolerate stronger interactions within the community as well as more severe changes to the environment, such as those predicted in a climate change context. Our study therefore shows how resting stages may help phytoplanktonic diversity maintenance.

Keywords: dormancy; phytoplankton; coexistence; competition; facilitation
Published in Journal of Theoretical Biology (2022) doi:10.1016/j.jtbi.2022.111020

2.1 Introduction

How the high biodiversity of primary producers maintains is still an unresolved question for both experimental and theoretical ecology. Terrestrial plants and phytoplanktonic communities can present hundreds of species relying on similar resources, a situation where Gause's principle implies that a handful of species should outcompete the others. Some degree of niche differentiation, perhaps hidden to the human observer, is generally expected for coexistence to maintain [74]. However, a complex life-history structure can further increase the likelihood of coexistence (e.g., [239, 131]), and so does the response of life history traits to variation in the environment [77, 161].

Analyses of coexistence in terrestrial plant communities sometimes take into account several life stages (e.g., [5, 85, 79]), though many models consider only a single life-stage (see, among others, [112, 207, 226]). When considering at least two stages, seeds/seedlings and adults, several mechanisms that can contribute to long-term coexistence in spatially and/or temporally fluctuating environment have been uncovered [306, 79].

The storage effect, a major paradigm in modern coexistence theory [74, 76], which involves a positive covariance between the strength of competition and favourable environmental variation as well as buffered population growth, is one of them. It has been shown to arise from a combination of interspecific competition and a long-lived dormant stage, together with a temporally variable recruitment rate [67], which often arises when recruitment is a direct function of environmental conditions [75]. The presence of a seed bank may therefore lead to a storage effect [13], though not systematically [5]. However, although the storage effect usually captures theoreticians' attention, the contribution of seeds to coexistence may be much larger than their potential contribution to the storage effect. A long-lived seed bank can help coexistence by other, simpler means. For instance, in the meta-community model of Wisnoski *et al.* [336], when dormancy and dispersal are present (without seed dispersal), local diversity increases in temporally fluctuating environments. In their model, adding a dormant stage could increase species diversity both at the local and regional scales. These results suggest that considering a seed stage in dynamical models can profoundly alter our understanding of community persistence (see also [176]).

Although there is some awareness of the role of cryptic life stages in shaping terrestrial plant coexistence, the effect of such dormant life stages on aquatic plant communities, and more specifically that of phytoplanktonic algae, is often ignored. Such a gap is even more surprising that phytoplankton organisms constitute one of the most important photosynthetic groups on Earth, being responsible for half the global primary production [121], and are the very basis of marine food webs. The classical view behind phytoplankton dynamics is that their blooms (peaks in abundances several orders of magnitude above their baseline level) are due to seasonal variation in light, temperature and nutrients, as well as hydrodynamic processes [285]. In this mindset, differential responses to environmental signals ensure the coexistence

of multiple species [223, 309], while always assuming that vegetative cells are already present in the environment, or immigrating from a nearby water mass. Momentary disappearances of a species are viewed as sampling issues at low density. However, a complementary hypothesis suggests that resuspension and germination of phytoplanktonic resting cells is another major player allowing re-invasion from very low or locally zero population densities [268, 222]. This long-standing hypothesis is supported by recent reviews [21, 111] which confirm that life history strategies including dormant individuals are widespread in phytoplankton (see [234, 323] for extensive lists of species with known resting stages). Formation of a resting cell can either be part of the life cycle of phytoplankton species and result from sexual reproduction or result from asexual processes triggered by specific environmental conditions [111]. Hereafter, when using the term 'seed banks', we refer to the accumulation of all types of resting stages in the seabed, either sexual (dinoflagellate cysts) or asexual (diatom spores or resting cells). A variety of models have endeavoured to explain and predict amplitude, timing and/or spatial distribution of blooms by explicitly modeling multiple stages in the life cycle of a particular species, but without interactions with other organisms (see for example [229, 153, 152, 340]). Two-to-four species [341, 116, 203] models also exist, but include a single species or compartement having a resting stage. This state of affairs means that we currently have no clear understanding of how the resting stage might help maintaining biodiversity in species-rich communities. In the present paper, we demonstrate the potential role of seed banks using a phytoplankton community dynamics model.

Phytoplankton communities in coastal environments may benefit from seed banks even more than the oceanic communities (see, for example, [229]), as the distance to the sea bottom is smaller, which favours recolonization from the sea bottom, something that is impossible in the deep ocean. Moreover, 'horizontal' exchanges between oceanic and coastal pelagic phytoplanktonic communities are usually observed. A flow from the ocean to coastal communities has been noticed for dinoflagellates especially [319, 34]. Conversely, in many other bloom-forming species, the shallower coastal areas might function as a reservoir for biodiversity in the ocean. Indeed, resting cells are able to germinate again after dozens of years [233, 111] or even thousands of years [294] of dormancy. Therefore, we consider in this study three interlinked compartements: the coastal pelagic environment, the seed bank, and the pelagic open ocean. The coastal pelagic environment acts as a bridge between the seed bank and the open ocean.

Our model is parameterized from field data (growth and interaction rates within the phytoplankton community), and includes biotic and abiotic constraints (e.g., particle sinking). In our analyses, we examine how seed banks may influence the maintenance of biodiversity, including under changing biotic interactions or changing environmental conditions. We either add or remove the dormant compartment, which allows to pinpoint its contribution to coexistence. We find that the presence of resting stages prevents the extinction of several species. Seed banks also allow a community to maintain its richness even with strong disturbances of its interaction network, unless facilitative interactions completely eclipse competitive interactions.

Changes in the environment, here represented by an increase in the mean temperature, can also be buffered by seed banks. Finally, we discuss what information would be required to further more accurate modeling of resting stages in phytoplankton community dynamics.

2.2 Methods

Models

Our models (Fig. 2.1) build atop recent models developed by Shoemaker & Melbourne [307] and Wisnoski *et al.* [336], although they diverge in several aspects developed below (e.g., possibility for facilitative interactions). These discrete-time models are designed for metacommunities with multiple interacting populations. Any discrete-time model requires an ordering of events; in our models, these unfold as follows: first, populations grow or decline according to a Beverton-Holt (BH) multispecies density-dependence (eqs. 2.1 and 2.3), and then, in a second step, exchanges occur between the different compartments or patches constituting the metacommunity (eq. 2.4).

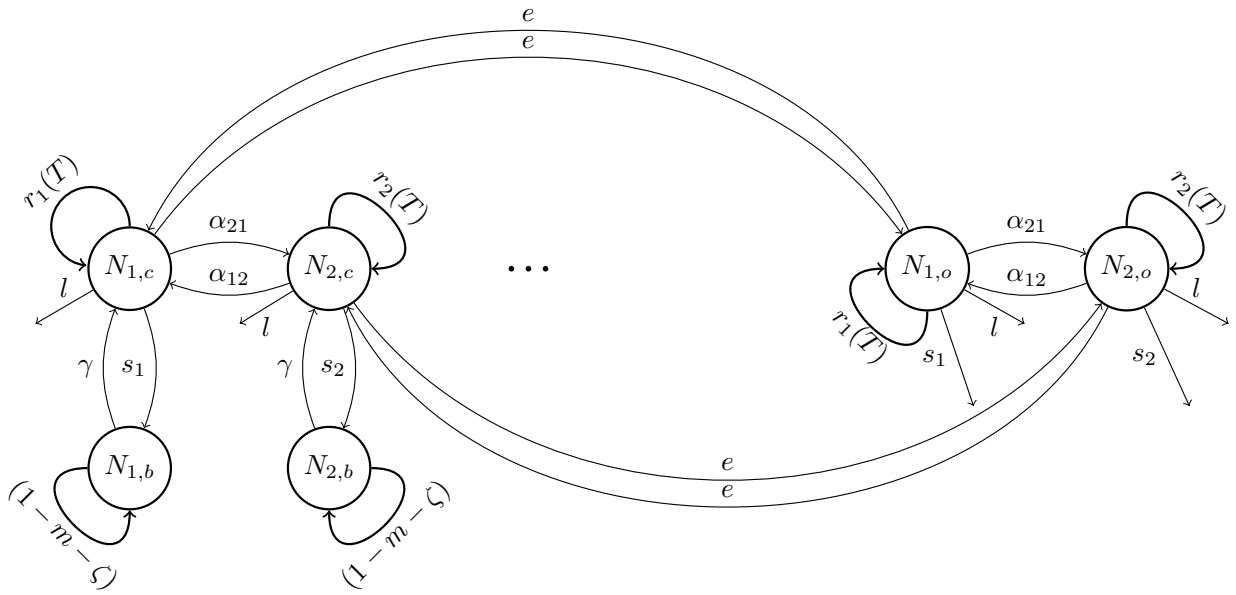

Figure 2.1: Structure of the model. Phytoplanktonic species (represented by circles) are present in the coast (subscript c), the ocean (o) and the seed bank (b). Parameters governing demography, interactions between organisms and exchanges between compartments are defined in Table 2.1. Only two species are shown here for the sake of simplicity but 11 species are present in the model.

In this paper, individuals are phytoplanktonic cells that move between the upper layer of coastal water, its bottom layer where a seed bank accumulates in the sediment, and the oceanic zone surrounding the coastal water masses (hereafter "the ocean"). Only oceanic and coastal pelagic cells are subject to BH-density dependence. Resting cells in the seed bank are only

affected by mortality m and burial due to sedimentation ζ. Parameters and state variables are defined in Table 2.1.

Parameter	Name	Value (unit)	Status
$N_{t,i,c/o/b}$	Abundance of species i at time t in the coast (c), ocean (o), or coastal bank (b)	NA (Number of cells)	Dynamic
T	Temperature	NA (K)	Dynamic
$r_i(T)$	Intrinsic growth rate of species i	NA (day^{-1})	Dynamic
b_i	Thermal decay of species i	Field-based (K^3)	Calibrated
T_i^{opt}	Optimal temperature for species i	Field-based (K)	Calibrated
d	Daylength	50 (%)	Fixed
α_{ij}	Interaction strength of species j on i in model I	Field-based (Cells^{-1})	Calibrated
a_C/a_F	Maximum competitive/facilitative interaction strength in model II	Field-based (NA)	Calibrated
H_{ij}	Half-saturation for the interaction strength of species j on i in model II	Field-based (Cells)	Calibrated
s_i	Sinking rate of species i's resting cell	$\{0.1; \mathbf{0.3}; 0.5\} \times$ Beta$(0.55, 1.25)$ (day^{-1})	Fixed
e	Exchange rate between ocean and coast	$\{0; \mathbf{0.4}; 0.9\}$ (day^{-1})	Fixed
l	Loss rate of pelagic phytoplanktonic cells	$\{0.04; 0.1; \mathbf{0.2}\}$ (day^{-1})	Fixed
m	Resting cell mortality rate	$\mathbf{10^{-5}}$ (day^{-1})	Fixed
ζ	Resting cell burial rate	$\{10^{-3}, \mathbf{10^{-2}}, 10^{-1}\}$ (day^{-1})	Fixed
γ	Germination (γ_1) $\times$ Resuspension (γ_2) rate	$\{10^{-3}, \mathbf{10^{-2}}, 10^{-1}\} \times \{10^{-5}, 10^{-3}, \mathbf{10^{-1}}\}$ (day^{-1})	Fixed

Table 2.1: Definition of main state variables and model parameters. State variables and fluctuating parameters are indicated in the last column as "Dynamic". Parameters that are constant through time are either "Fixed" (directly obtained from literature) or "Calibrated" (obtained through model fitting, with initial values arising from previous studies at the study site). When a range of values is given, the bold numbers indicate the reference values while the others are used for sensitivity analysis. Beta(0.55,1.25) is the Beta distribution with parameters 0.55 and 1.25. For γ, germination values for sensitivity analysis were multiplied by the reference value for resuspension, and conversely.

The Beverton-Holt (BH) formulation of multispecies population dynamics, sometimes called Leslie-Gower [90], is a Lotka-Volterra competition equivalent for discrete-time models, and is often used to represent terrestrial plant community dynamics. In our implementation of the model, the population growth rate is modified by both competitive and facilitative interactions, which translates into positive and negative α_{ij} coefficients, respectively. We present two different interaction models. We first use the classical multispecies BH model (model I, eq. 2.1), also present in the original models of Shoemaker & Melbourne [307] and Wisnoski *et al.* [336]. However, the high number of facilitative interactions characterizing the modeled phytoplankton community [271] combined to the mass-action assumption could have very irrealistic destabilizing consequences, which have been likened to an "orgy of mutual benefaction" [228]: populations grow to infinity because there is no saturation of beneficial effects when density increases. In model I, we therefore forbid the realized growth rate to go above the intrinsic growth rate (its theoretical limit), by setting a minimum value of 1 to the denominator of the BH formulation. We subsequently define saturating interactions, inspired by Qian & Akçay [279], in our model II (eq. 2.3). Model II provides a more process-orientated solution to the issue of excessive mutual benefaction (but at the cost of added parameters). Setting a miminum value of 1 to the denominator is still required for large increases or decreases in interaction strengths.

In our framework, the first step of model I can be written as

$$\begin{cases} N_{t',i,c} &= \frac{\exp(r_i(T))N_{t,i,c}}{1+\sum_j \alpha_{ij}N_{t,j,c}} - lN_{t,i,c} \\ N_{t',i,o} &= \frac{\exp(r_i(T))N_{t,i,o}}{1+\sum_j \alpha_{ij}N_{t,j,o}} - lN_{t,i,o} \\ N_{t',i,b} &= N_{t,i,b}(1-m-\zeta) \end{cases} \tag{2.1}$$

where the intrinsic growth rate $r_i(T)$ is a species-specific function of the temperature (see eq. 2.2), the interaction coefficients α_{ij} are per capita effects of species j on species i, and the loss term l accounts for lethal processes such as natural mortality, predation or parasitism. First estimates of interaction coefficients are inferred from a previous study of coastal community dynamics with Multivariate AutoRegressive (MAR) models [271]. We later calibrate these coefficients for model I, since MAR models were applied at a different timescale.

The intrinsic growth rate $r_i(T)$ is defined through a modified version of the formula used by Scranton & Vasseur [301] (eq. 2.2), which classically decomposes $r_i(T)$ in two parts: the species-independent metabolism part $E(T)$ and the species-specific niche part $f_i(T)$:

$$\begin{aligned} r_i(T) &= E(T)f_i(T) \\ \text{where } E(T) &= d \times 0.81e^{0.0631(T-273.15)} \\ \text{and } f_i(T) &= \begin{cases} \exp(-|T-T_i^{opt}|^3/b_i), & T \leq T_i^{opt} \\ \exp(-5|T-T_i^{opt}|^3/b_i), & T > T_i^{opt}. \end{cases} \end{aligned} \tag{2.2}$$

The metabolism part describes the maximum achievable intrinsic growth rate based on Bissinger *et al.* [48]. This maximum daily intrinsic growth rate is weighted by the daylength d as no growth occurs at night. The realized niche part $f_i(T)$ describes the decrease in growth rate due to the difference between the temperature in the environment and the species-specific thermal optimum T_i^{opt}, and is controlled by the species-specific thermal decay b_i, which depends on the niche width. It is important to note that unlike $E(T)$, which models direct effects of temperature on metabolism, $f_i(T)$ is a phenomenological construct including all indirect effects of temperature, mediated by environmental variables correlated to temperature (light, nutrient inputs, predation, among other environmental conditions). In other words, $f_i(T)$ corresponds to a realized niche, pertaining to a given environment, which can be much more narrow than the fundamental thermal niche. Parameterisation is further detailed in Section 2.S1 of the SI.

In model II, oceanic and coastal dynamics are governed by eq. 2.3:

$$N_{t',i,c/o} = \frac{\exp(r_i(T))N_{t,i,c/o}}{1+\sum_{j\in\mathbb{C}} \frac{a_C N_{t,j,c/o}}{H_{ij}+N_{t,j,c/o}} + \sum_{j\in\mathbb{F}} \frac{a_F N_{t,j,c/o}}{H_{ij}+N_{t,j,c/o}}} - lN_{t,i,c/o} \tag{2.3}$$

where a_C and a_F are the maximum competition and facilitation strengths, respectively, with $\mathbb{C}$ and $\mathbb{F}$ the sets of competitors and facilitators of species i. We use here similar notations to Qian & Akçay [279], but use different parameters that vary between species. Indeed, the half-

saturation coefficients H_{ij} vary between species, as opposed to the maximum rates in Qian & Akçay [279]. It did not make sense biologically for H_{ij} to be fixed (e.g., in a resource competition context, different species are expected to feel resource limitations at different concentrations of nutrients and at different numbers of competitors). How to shift from MAR- to BH-interaction matrices in model I, and to use the parameter estimates of model I to specify parameters in model II is described in Section 2.S2 of the SI.

After growth and mortality processes occur, exchanges take place between the three compartments, which constitutes the second step of the model (eq. 2.4):

$$\begin{cases} N_{t+1,i,c} &= (1-s_i-e)N_{t',i,c} + \gamma N_{t',i,b} + eN_{t',i,o} \\ N_{t+1,i,o} &= (1-s_i-e)N_{t',i,o} + eN_{t',i,c} \\ N_{t+1,i,b} &= (1-\gamma)N_{t',i,b} + s_i N_{t',i,c}. \end{cases} \tag{2.4}$$

Each compartment (ocean, coast, coastal seed bank) contains 10^3 cells at the beginning of the simulation, and the dynamics are run for 30 years with a daily time step. We model the temperature input as a noisy sinusoidal signal with the same mean and variance as the empirical data set described below: the amplitude of the sinusoid is 12.4°C and the standard deviation of the noise is 0.25°C.

Parameterization of the models

Literature-derived parameter values

Loss rate The loss rate l of vegetative cells can be attributed to natural mortality, predation or parasitism. This rate is quite variable in the literature: the model of Scranton & Vasseur [301] considered a rate around 0.04 day^{-1} while a review by Sarthou *et al.* [295] indicates a grazing rate of the standing stock between 0.2 and 1.8 day^{-1} and an autolysis rate between 0.005 and 0.24 day^{-1} (in the absence of nutrients, or because of viral charge). A maximum value of 0.2 is fixed for the model, as a balance between using a high loss rate (probable because of predation) and keeping all species in the community in the reference model (see Section 2.S3 of the SI for more details).

Sinking rate Phytoplanktonic particles have a higher density than water and cannot swim to prevent sinking (although they are able to regulate their buoyancy [285]). Sinking is mostly affected by hydrodynamics, but at the species-level, size, shape, density-regulation and colony-formation capabilities are key determinants of the particle floatation. In this model, the sinking rate of each species resting cells is drawn from a Beta distribution with a mean value of 9%, and a maximum ($S_{\max}$) around 30%, that is $s \backsim 0.3\beta(0.55, 1.25)$ (see Fig. 2.S4), adapted from observations by Passow [267] and Wiedmann *et al.* [332].

Exchange rate The exchange rate e between the ocean and the coast depends on the shape and location of the coast (estuary, cape, ...). Our calibration site is located at an inlet. The flow at the inlet leads to a renewal time of the coastal area water evaluated between 1 and 2.5 days [20], which corresponds to a daily exchange rate between 40 and 100 %.

Mortality and burial in the seed bank Loss from the seed bank is the result of cells' mortality m and their burial by sedimentation ζ. Mortality values range between 10^{-5} and 10^{-4} per day (more details on the approximation of mortality rates from [233] are given in Section 2.S3 of the SI). However, burial by sedimentation is the prevailing phenomenon. Indeed, once resting cells have been buried, they are not accessible for resuspension even if they could still germinate. Burial depends on the hydrodynamics of the site, but also on biotic processes (i.e., bioturbation) and anthropogenic disturbances such as fishing or leisure activities (e.g., jet skiing). This parameter is thus heavily dependent on the environmental context and varies here between 0.001 and 0.1 per day.

Germination/resuspension Both germination (γ_1) and resuspension (γ_2) are needed for resting cells to contribute to the vegetative pool in the water column (γ = resuspension×germination). As actual rates of germination are not easily deduced from the literature, a set of credible values has been tested (1%, 0.1%, 0.01%). Similarly, resuspension values are seldom computed for phytoplanktonic cells, but models for inorganic particles can be used (see Section 2.S3 of the SI for literature and details). In this paper, we explore values between 10^{-5} (stratified water column) to 0.1 (highly mixed environment).

Initial interaction matrix

Initial values of interaction strengths between species are based on Multivariate AutoRegressive (MAR) model estimates [271]. MAR(1) models relate the log-abundance of each of the S phytoplankton species at time $t+1$ to log-abundances of all species at time t, through an interaction matrix, and effects of abiotic variables at time $t+1$ (see Section 2.S2 of the SI). Interactions are estimated only between species within the same trophic level, and are independent from the environmental variables that were included in the MAR model estimates as covariates, such as temperature in our case. This allows to remove at least some of the confounding factors, such as seasonality. A phylogeny-based interaction matrix resulted in a better fit to model the community dynamics, i.e., pennate/centric diatoms only interact with other pennate/centric diatoms, respectively, and dinoflagellates only interact with dinoflagellates.

The MAR model can only estimate apparent interaction strengths: complex processes can be at work behind values of competition and facilitation, either from abiotic (e.g., hydrodynamics) or biotic (e.g., consumption by predators or parasites) variables. For more consideration on apparent interactions detected in phytoplankton communities with the MAR model, we refer the reader to Barraquand *et al.* [32]. In the present model, while we tune interaction strength

values (see below), the type of interaction (competition, facilitation or absence of interaction) remains the same as computed in Picoche & Barraquand [271].

Parameter calibration

As explained above, we use initial interaction estimates from our previous time series modelling ([271], see Section 2.S3 of the SI for the equations), which are then calibrated to the time series (thus possibly re-estimated, see below), to take into account the differences in model structure and timescale between this study and Picoche & Barraquand [271]. In Section 2.S2, we present the formulas relating the MAR interaction coefficients to the Jacobian matrices of the Beverton-Holt multispecies models, as these formulas allow to obtain proper α_{ij} coefficient values.

The calibration procedure consists in launching 1000 simulations, each characterized by a specific set of interaction coefficients. More precisely, for each simulation, an interaction coefficient (α_{ij} in model I, H_{ij} in model II) has probability 1/5 to keep its present value, probability 1/5 to increase by 10% , 1/5 to decrease by 10%, 1/5 to be halved and 1/5 to be doubled. The numbers of coastal pelagic cells (which are the ones measured empirically) are then extracted over the last 2 years of the simulation, and compared to observations using the following summary statistics:

- average abundance $f_1 = \sqrt{\frac{1}{S}\sum_i^S \left(\bar{n}_{i,obs} - \bar{n}_{i,sim}\right)^2}$ where S is the number of taxa and $\bar{n_i}$ is the logarithm of the mean abundance of taxon i.
- amplitude of the cycles $f_2 = \sqrt{\frac{1}{S}\sum_i^S \left[(\max(n_{i,obs}) - \min(n_{i,obs})) - (\max(n_{i,sim}) - \min(n_{i,sim}))\right]^2}$ where n_i is the logarithm of the abundance of taxon i.
- period of the bloom. The year is divided in 3 periods, i.e., summer, winter and the spring/autumn group (as taxa blooming in these periods can appear in either or both seasons). We give a score of 0 if the simulated taxon blooms in the same period as its observed counterpart and 1 otherwise.

Simulations with taxon extinction (i.e., the taxon is absent for more than 6 months in a compartment) are discarded, as extinctions are not observed in the field data. Parameter sets are then ranked according to their performance for each summary statistic, and we select the set of interactions minimizing the sum of the ranks.

Sensitivity analysis

Parameters taken from the literature may be site- or model- specific, or vary over several orders of magnitude in the literature, e.g., rates of sinking s, resuspension/germination γ, seed burial ζ, and loss of pelagic cells l. We therefore performed a sensitivity analysis to these highly uncertain parameters. The set of tested values for each parameter is given in Table 2.1. We used average abundances and amplitudes at the community and taxon levels for the last 2 years of simulations as the major model diagnostics.

Empirical dataset used for calibration

The models are calibrated using time series of phytoplanktonic abundances that have been monitored biweekly for 21 years in the Marennes-Oléron Bay, on the French Atlantic Coast (the Auger site analysed in [271]). We stress that we are not trying to model precisely this particular community, but rather to constrain our models with an empirically-derived interaction network and species-specific thermal niches, which helps to produce lifelike patterns of phytoplankton community dynamics resembling observable data (seasonal dynamics, high-amplitude blooms, differences in average abundances matching data).

Scenarii

The effect of the seed bank on biodiversity and community dynamics can be evaluated through the response to disturbance with and without the resting-stage compartment. Mortality in the seed bank is set to 100% to effectively remove the compartment. We evaluate two main disturbances:

1. increase or decrease in interaction strength
2. temperature change, either in mean value or variability

In the first scenario, interaction strengths are multiplied or divided by a factor ranging between 1 and 10. In order to differentiate the effects of facilitative and competitive interactions on coexistence, we vary only one type of interactions at a time. Here, both intra and interspecies interactions are modified; we present in Section 2.S6 of the SI additional simulations with a change in interspecies interactions only.

In the second scenario, five different climate change trajectories are assessed. In the first three, the average temperature is increased by 2, 5, or 7°C [54]. In the next two, keeping the reference average temperature, the total variance of the temperature, including seasonality and noise, is either decreased or increased by 25%. Each climate change trajectory is run 5 times to account for the intrinsic stochasticity of the temperature signal.

In both scenarii, simulations are run for 30 years for both population growth models, with and without a seed compartment, and only the last 2 years are considered to evaluate effects of changes in parameters and in temperature. The code for all simulations is to be found at `https://github.com/CoraliePicoche/SeedBank`.

2.3 Results

Phytoplankton dynamics

The classical mass-action (model I) and saturating interaction (model II) formulations of multi-species dynamics both reproduced the main characteristics of observed phytoplankton dynam-

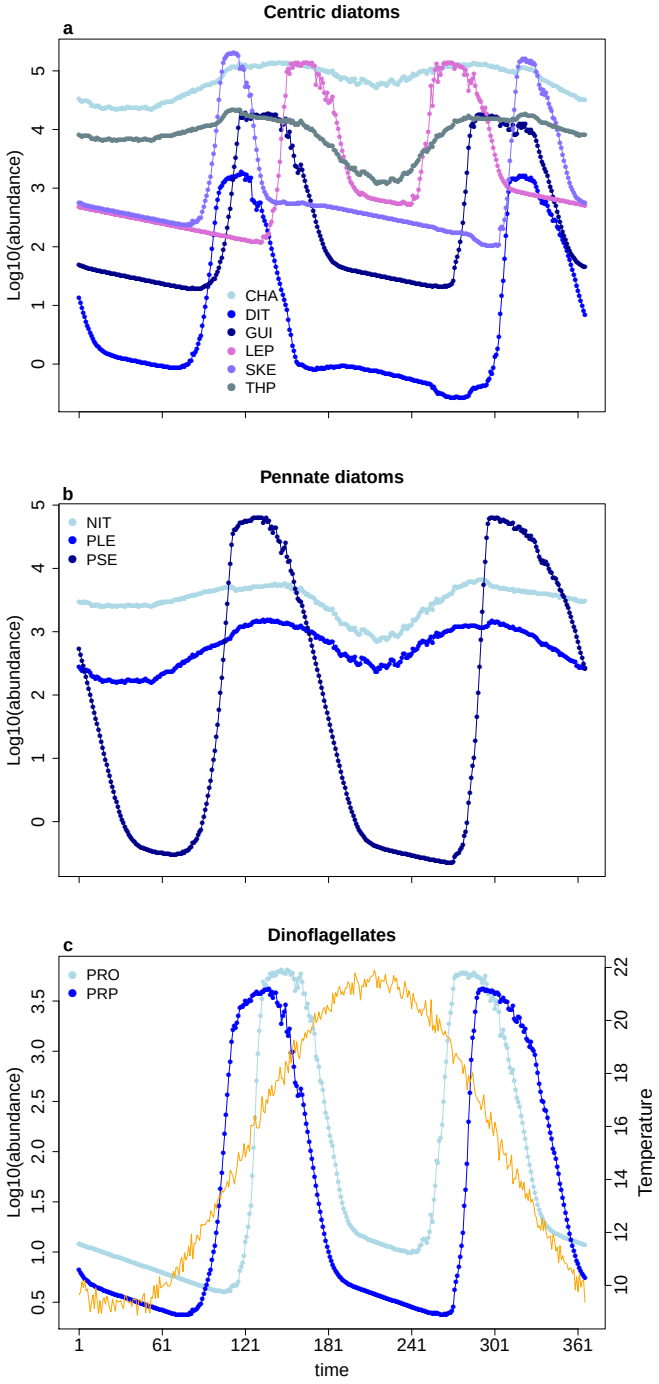

Figure 2.2: Simulated phytoplankton dynamics for a year in model I (time in days). Each panel corresponds to a cluster of interacting taxa: centric diatoms (a), pennate diatoms (b) and dinoflagellates (c). Taxa only interact within their cluster of related taxa (see Methods). The orange line in the third panel indicates the temperature.

ics. They produced one or two blooms during the year and a range of abundances covering several orders of magnitude, with the right timing of the blooms. At the Auger site that was used for calibration, abundances increase in spring and can last over part of summer, or start a new bloom in autumn, which is what we observed as well in the models. Annual mean abundance of the various species was also well reproduced. That said, in some cases, abun-

dances could be lower than expected and the variation in abundances due to seasonality was underestimated (Fig. 2.2). In all cases, saturating interactions led to higher abundances than mass-action interactions throughout the year (Fig. 2.S5).

Sensitivity to uncalibrated parameters

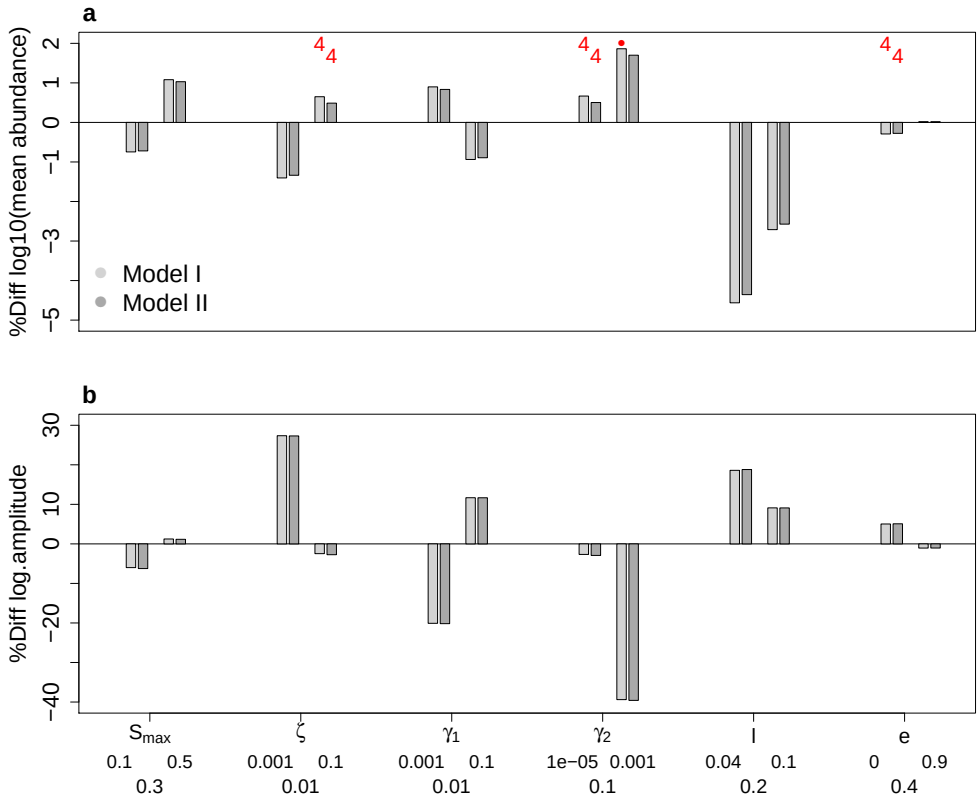

Figure 2.3: Sensitivity of the model to variation in parameters, measured as the difference between the reference simulation metric and the metric for the simulation including a change in parameter. The two metrics used were the decimal log average abundance (a) and the decimal logarithm of the ratio between maximum and minimum abundance (b) of the coastal phytoplanktonic community. The x-axis is divided in three rows: the symbol of the parameter defined in Table 1 (first row), values used in the sensitivity analysis (second row) and values used in the reference simulation (third row). Numbers in red in the top panel are the final number of species in the ocean and dots correspond to simulations in which at least one species abundance reached 0 at one point but the species did not disappear.

Phytoplankton abundances were not strongly affected by changes in the parameter values (Fig. 2.3). As parameters were varied in their plausible range, the average change in mean abundance on the coast between the reference simulation and the sensitivity simulations varied between -4.6 and 1.9% for model I and between -4.4 and 1.7% for model II, with similar deviations (same sign and magnitude) in the two models.

In the two models, the decrease in mortality rate of vegetative cells l had the highest impact on the final average abundance, leading to an increase in abundances. The exchange

rate between the ocean and the coast had a much lower effect on the coastal average abundance.

On the other hand, the amplitude (i.e., decimal logarithm of the maximum to minimum ratio of abundance) was more affected by changes in parameters and could vary by -39.6 to 18.8% in model I, and between -39.6% and 18.8% in model II. Results were qualitatively the same in the two models, with a decrease in resting-stage burial being the main driver of the decrease in amplitude, and a decrease in resuspension leading to an increase in amplitude.

In three cases (burial rate set to 0.1, resuspension set to 10^{-5} or exchange rate set to 0), the final richness of the oceanic community decreased from 11 to 4. Extant species were the same in all simulations (CHA, THP, NIT, PLE, i.e., temperature generalist species; the correspondence between codes and groups of species is given in Section 2.S4 of the SI). When resuspension was set to 0.001, a species periodically disappeared from the ocean, to be subsequently re-seeded by the coastal population.

For all parameters, except the sinking rate, an increase in mean abundance was associated to a decrease in amplitude.

Scenarii of environmental change

Two scenarii were designed to test the buffering effect of the seed bank against disruption. In both cases, it consisted in removing the seed bank by setting resting-stage mortality to 100% per day. Without any other disturbance to the system, this led to a decrease in richness from 11 to 4 species at the end of the simulation (Fig. 2.4) while the total abundance of phytoplankton was not strongly affected (around 10^5 in all cases). The inverse of the Simpson index (the second Hill number) decreased from approximately 3 to 1, showing that the disappearance of the seed bank did not affect only the rarest species.

Biotic effects

Our first hypothesis was that the absence of the seed bank would cause the community to be more affected by a higher competition strength. Counter-intuitively, our results (Fig. 2.4) showed that an increase in competition strength only had negative effects with model I, and for high competition values (6 times the reference ones at least), shifting from 4 to 3 species in the oceanic compartment of a community without a seed bank. By contrast, an increase in competition strength did not affect the richness of a community with a seed bank. On the contrary, a decrease in competition (from a factor 0.5 and lower) or an increase in facilitation (starting from a factor 2 and higher) led to much smaller communities in model II in the presence of a seed bank.

The inverse of the Simpson index was also affected by the changes in interaction strengths, with similar patterns to richness, as it was lowest for high facilitation or low competition. Some species reached very high growth rates in these scenarios, which then fed back onto community dynamics, generating lower diversity in the end.

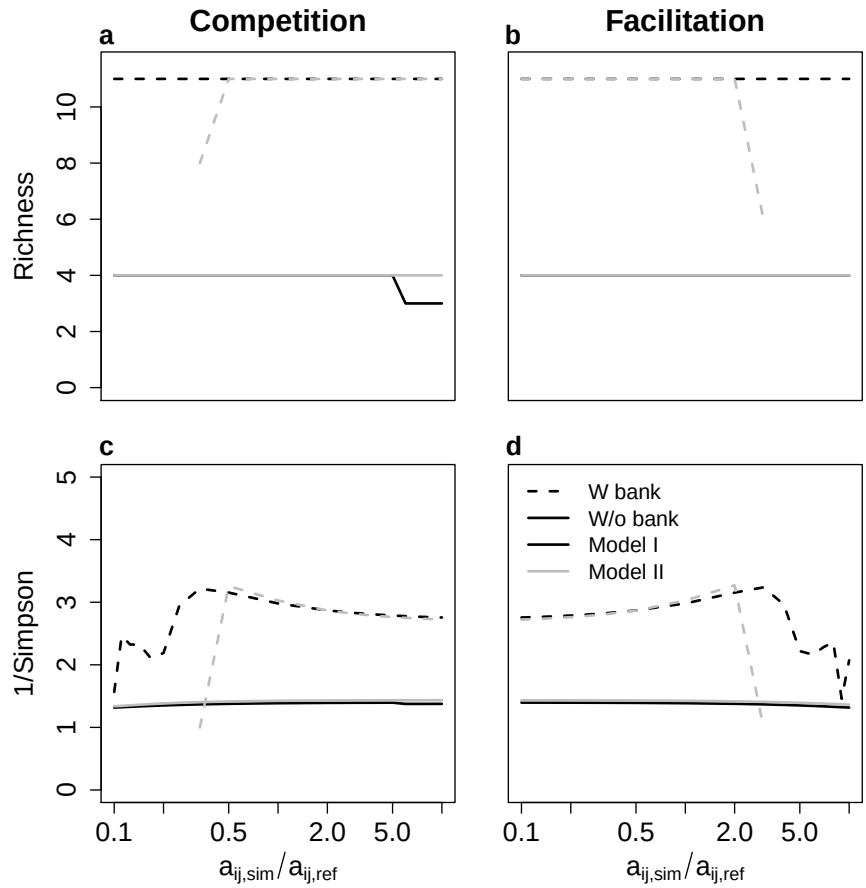

Figure 2.4: Measures of biodiversity in the ocean at the end of the simulation: a-b) richness and c-d) inverse of the Simpson index, with (dashed line) and without (solid line) a seed bank, as a function of the strength of competition and facilitation with a classical Beverton-Holt (black lines) or a saturating interaction (grey lines) formulation. The x-axis shows the factor by which each interaction coefficient was multiplied, e.g., the value 0.1 indicates that the interaction strengths in the simulation are 10 times lower than the interaction strengths in the reference simulation. Note the logarithmic scale.

Species which disappeared were characterized by a lower minimum abundance, a higher amplitude of fluctuations and a small niche (Fig. 2.5). However, their interactions were not qualitatively different from the other species.

Abiotic effects

Our second hypothesis was that the absence of a seed bank would reduce the ability of a community to withstand changes in its abiotic environment, here represented by variation in the temperature. This was true for both models (Fig. 2.6), as the communities without a seed bank could not maintain their richness with an increase in temperature above 2°C, as opposed to communities with a seed bank, which could only be affected by a 7°C increase (scenario SSP5 8.5 in [54]). In all cases however, the total abundances were not strongly affected. Indeed, the total abundance of a community is driven by a small number of numerically dominant species,

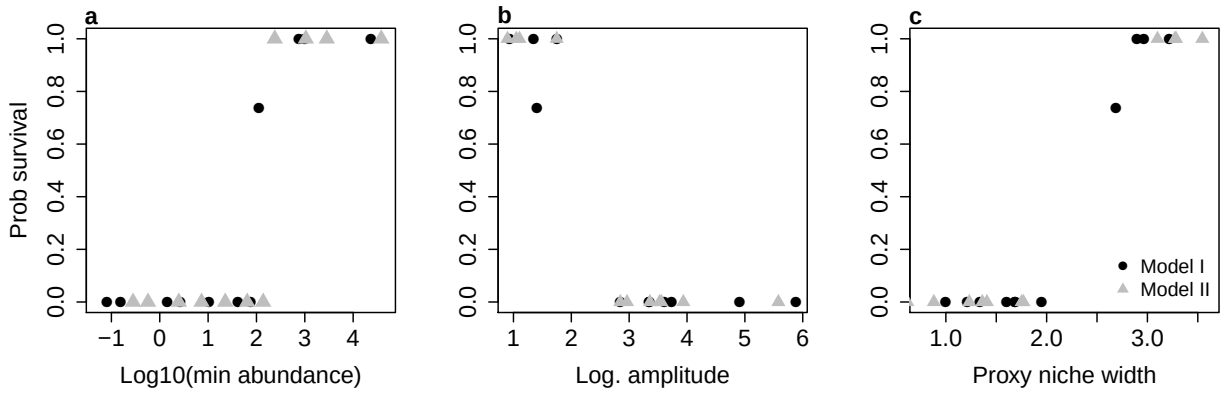

Figure 2.5: Probability of survival of species along a gradient of competition strength and in the absence of a seed bank, as a function of their population dynamics characteristics (minimum abundance, logarithm of amplitude and niche width) in the reference parameter set. Each dot corresponds to a species, and its survival using model I (black) or model II (grey).

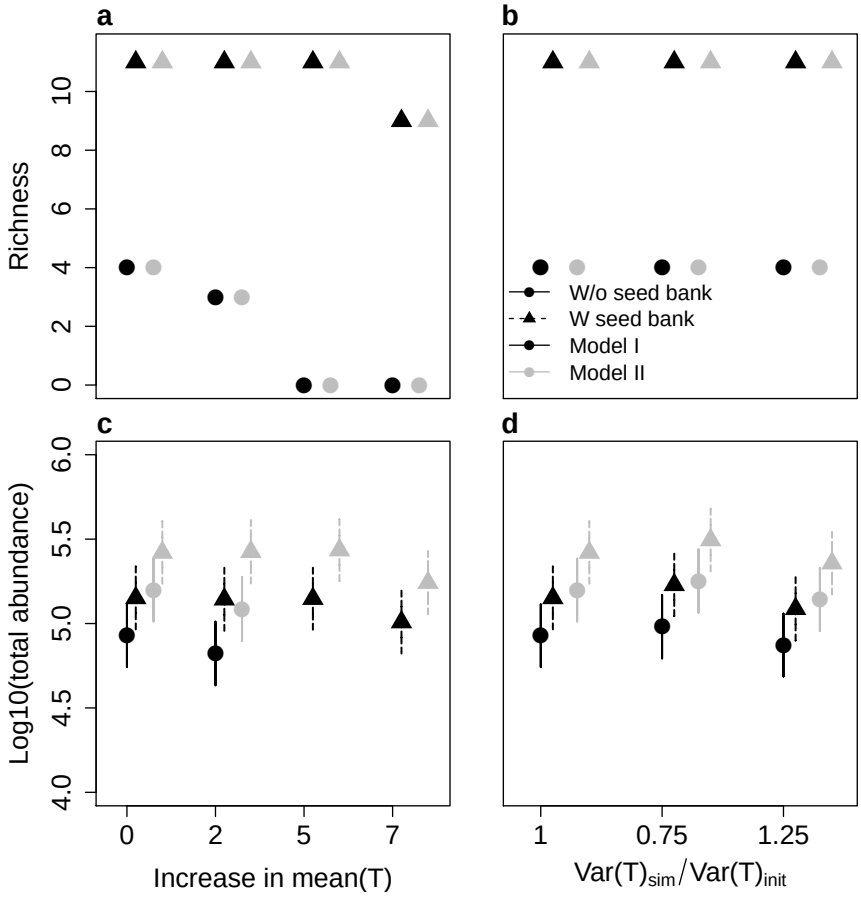

Figure 2.6: Variation in richness and total abundance with and without a seed bank as a function of the mean (left) and variance (right) of the temperature with a mass-action (black line, model I) or a saturating interaction (grey line, model II) formulation.

which did not disappear. High total abundances tended to correspond to the abundance of only one or two species. Model II consistenly led to higher abundances, as was already the case

in the reference simulations.

The variance of the temperature did not affect richness nor total abundance of communities with a seed bank. This was also true without a seed bank. The presence of resting stages did increase total abundance though.

Some additional simulations were made to better understand the functioning of the model, during which the temperature was kept constant (equal to the average temperature of the fluctuating environment). For a constant and average temperature, the presence of a seed bank did not change the final richness of the communities. There were 9 extant species at the end of the constant-environment simulations, whether there was a seed bank or not. The two specialist species whose thermal optima were farthest from the constant temperature went extinct even when the seed bank was present, since the environmental fluctuations allowing them to benefit from cold or warm temperature were removed. The other species benefited from the constant environment, as 9 species instead of 4 persisted in the case where the seed bank was removed (see below for a discussion). The high persistence in a constant environment (9 out of 11) translates the stabilizing, data-driven interaction structure of our model. For both models I and II, the dynamical system reached a fixed point equilibrium in absence of the fluctuations induced by a variable temperature.

2.4 Discussion

Using a meta-community model which accounts for exchanges between the ocean and the coast, as well as movements between the top and the bottom of the coastal water column, we were able to show that a resting stage, leading to the formation of a seed bank, can help maintain biodiversity. Our phytoplanktonic community dynamics model was parameterized based on literature, field-based phenology, and interaction strength estimates. It was then calibrated on phytoplankton community time series. Our model was able to simulate realistic community dynamics (both mean abundances and temporal patterns), while including the effects of both positive and negative interactions on community dynamics. When removing the seed bank, biodiversity decreased drastically. The total abundance of the community decreased as well. This was true for the reference parameter values, as well as when species interaction strengths and environmental fluctuation levels were altered, in which cases the seed bank's buffering influence disappeared. Moreover, when faced with a biotic or abiotic perturbation, communities where species could divert part of their population to a dormant stage were less prone to species loss and could maintain their biomass through the years. These results were consistent for the two interaction models that we considered, with and without progressive saturation in interaction strengths. Our results therefore demonstrate the major potential role of phytoplanktonic resting stages in maintaining biodiversity. These results align with the findings of previous theoretical studies, that have put forward similar effects of dormant stages in other taxa, such as plants [207, 176], invertebrates [336] or (smaller) microbes [184].

The positive effect of a seed bank on species diversity in our model is contingent upon environmental fluctuations. Indeed, in a constant environment, the absence of a seed bank did not alter persistence, and final richness was higher than in a fluctuating environment (9 species in a constant environment vs. 4 species in a variable environment, both without a seed bank). The absence of abiotic perturbations of the intrinsic growth rates enabled species which were not too far from their thermal optima to maintain. This might be surprising for readers acquainted with Hutchinson's nonequilibrium theory [168]. However, the theoretical literature on persistence in variable environments has shown that additional mechanisms (reviewed in [125]) are needed to maintain diversity in the long run, such as relative nonlinearities or the storage effect, which is at least absent from the pelagic part of our model due to the absence of buffered growth (see Section 2.S8 in SI). There is a clear destabilizing effect of environmental variation in our model without the seed bank, which is probably heightened by the fact that environmental variation is positively autocorrelated through seasonality [270, 300]. Although the storage effect could manifest itself when combining pelagic and seed bank parts of our model, it is very unlikely to be at play here as environmental conditions do not impact recruitment rates in our model, while this is usually a requirement for the storage effect to manifest in such models [75].

Even in the absence of a storage effect, however, the high longevity of the resting stage itself can explain the effect of the seed bank in a fluctuating environment. Such longevity is due to dormancy, which has long been observed in field and experimental data, including for phytoplanktonic organisms [109], and has been theorized to be an important and neglected process in the wider microbiology literature [215, 184]. It allows recolonization of a community where counts of pelagic cells alone would suggest that some species have gone extinct. This colonization-in-time may of course combine with present recolonization from other spatial areas, as is known for plants [306]. In our case, our focus on phytoplankton led us to assume that organisms moved between the coast and the ocean, which were largely synchronous environments. Spatial recolonization was therefore less important than temporal recolonization; the relative importance of the two processes may vary depending on the organisms and the degree of spatial synchrony of their environment (in plankton, see [242, 12]).

The specificities of phytoplankton resting stages, that usually fall to the ocean bottom in coastal areas, led us to assume that only the "vegetative" stage (here, the classic pelagic form of planktonic cells) disperse. In some other metacommunity models with dormant seed banks (e.g., [336]), the dormant stage can disperse as well. This would be true for most plants too (and perhaps some phytoplankters in situations where they are transported by animals). However, the restriction about which stage can move did not change the general conclusion already stated by Wisnoski *et al.* [336]: the combination of spatial dispersal and dormancy through seed banks greatly helps biodiversity maintenance. In our study, this main result was also robust to changes in exchange parameters and mean interaction values in the community.

Species persistence varied between the coastal and oceanic compartment. Temporary species

disappearance could happen periodically in the ocean but species were able to reinvade from the coast, thanks to the connection of the coast to the seed bank. Local extinction in the absence of a seed bank confirms conclusions from Hellweger *et al.* [152] on a single species. This suggests that some species may be locally transient: they are filtered out from certain patches, but can reinvade more or less periodically the environment (as found by [139], for grasslands).

Survival probabilities of species within the community in the absence of a seed bank also depended on their characteristics – and therefore species rescued by the seed bank also have specific characteristics. While extinct species growth rates tended to be less affected by interactions for average densities (Fig. 2.S7 in SI), such species had higher amplitudes of population variation as well as a smaller niche width. Specialist species, growing at specific temperatures, therefore typically benefit the most from a seed bank.

Despite the evidence for seed bank effects that we and others uncovered, phytoplanktonic community models designed to explain biodiversity usually avoid modelling seed banks. In our view, this may decrease the possibility of spontaneous re-colonization at the coast (at very low densities initially), which can then spill to the open ocean by progressive dispersal by the currents. Ignoring the seed bank may also decrease overall persistence, as fitting models I and II without a seed bank does not allow all species to persist here. If the goal of a community-level model is very short-term prediction (days, weeks), this recolonization can probably be neglected, and there are no risks of extinction over these short time frames. However, over multiple years, ignoring cryptic life stages allowing recolonization could strongly bias downwards our view of long-term coexistence. Long-term phytoplankton coexistence modelling (over multiple decades or more) likely requires that we take into account resting stages, whose influence may become only more important as the timescale increases, due to the very long possible dormancies that have been evidenced [111, 294]. When modelling different stages of the life cycle in a detailed manner – as done here – is impractical, the recolonization could perhaps be simplified to a stochastic immigration term (as done in [313] in a single-species context). This technical suggestion certainly extends to models of (terrestrial) plant community dynamics.

More research on dormant stages may be needed to parameterize truly predictive mechanistic phytoplankton models with multiple life stages, in particular to inform parameters such as the sinking rate of resting cells, as well as burial and resuspension parameters. These parameters are all linked to hydrodynamics [341, 340] and may locally vary. Sinking rates show an interesting conflict between short- and long-term survival: in coastal areas, a fraction of sinking cells contribute to the seed bank, increasing the odds of species long-term survival at the cost of short-term individual cell survival. But high sinking rates are essentially "wasted" in the open ocean – whether different sinking rates can be selected, to some degree, by such different environments could be quite revealing. How cells get up rather than down in the water column might be as interesting but more difficult to study. The likely idiosyncratic nature of recolonization by resting cells – due to the contingency on local hydrodynamics – means that experimentation might be the only manner in which the frequency of reinvasion can be assessed.

Currently, one of the only recolonization-related parameters measured in the field is the rate of survival of the cells found in the sediment [241, 310]. While very important, this parameter is a necessary not sufficient condition for reinvasion of the population at future times. We need more information about the abilities of resting cells buried in the sediment to come up to the pelagic zone, which is required for recolonization to actually occur. Many factors may contribute: bottom currents, benthic animals,... We therefore encourage both experiments and field observations to follow actual seed trajectories, in order to help us understand this cryptic part of the diversity maintenance process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contribution statement

CP and FB designed the project and the models. CP wrote the computer code and produced the figures. CP and FB interpreted the results and wrote the manuscript.

Acknowledgements

We thank Nathan Wisnoski and Frank Jabot for useful feedback. FB and CP were supported by grants ANR-10-LABX-45 and ANR-20-CE45-0004. CP was supported by a PhD grant from the French Ministry of Research.

2.S Supplementary Information

2.S1 Intrinsic growth rate modeling

The intrinsic growth rate can be decomposed in 2 elements:

$$r_i(T) = E(T) f_i(T) \tag{2.S1}$$

with $E(T)$ the response to temperature common to all species and $f_i(T)$ the species-specific response. This section provides detailed information regarding $E(T)$ and $f_i(T)$ estimates.

Common response to temperature

Phytoplanktonic growth rates cover a broad range of values: between 0.2 and 1.78 day^{-1} for diatoms in Reynolds [285], even reaching 3 day^{-1} in the meta-analysis of 308 experiments by Edwards *et al.* [107]. These values are often computed from measurements on isolated species or on small communities in laboratory conditions, in a constant environment. A broader perspective is therefore necessary to understand general responses to changes in the environment [48, 108], especially temperature.

We first used the equation by Scranton & Vasseur [301] as a starting point, but it was not able to reproduce the observed values found in Edwards *et al.* [107] (see low values of the growth rate in Fig. 2.S1a, between 0.05 and 0.8, in place of values between 0.2 and 3 day^{-1}). In this context, we decided to use the formula by Bissinger *et al.* [48] to compute the maximum growth rate response to the temperature. There are two reasons for this choice. First, their model is a general function that can be applied to all species. Second, Bissinger *et al.* [48] is an update of the seminal work of Eppley [114] (which was used in [301], but might be outdated).

The relationship between temperature and growth rate is then $E(T) = 0.81 \exp^{0.0631(T-273.15)}$, with T in Kelvin degrees. 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 daily growth rates need to be proportional to the daylength as no growth occurs at night: we therefore divide $E(T)$ by two in our models.

Species-specific response to temperature

The niche part of the growth rate $f_i(T)$ is mainly defined by two parameters which drive the phenology of a species: the thermal optimum T_i^{opt} and a proxy of the niche width b_i (eq. 2.S2).

$$f_i(T) = \begin{cases} \exp(-|T - T_i^{opt}|^3 / b_i), & T \leq T_i^{opt} \\ \exp(-5|T - T_i^{opt}|^3 / b_i), & T > T_i^{opt} \end{cases} \tag{2.S2}$$

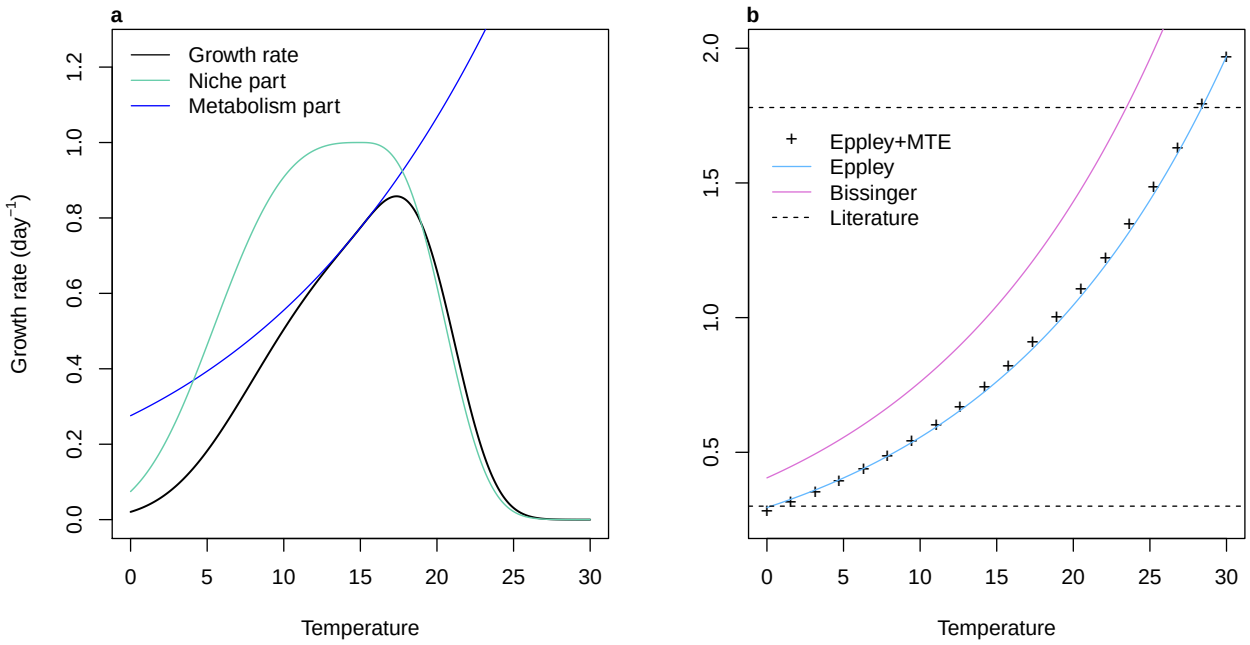

Figure 2.S1: Decomposition of the Scranton & Vasseur [301] growth rate formula (a). The black line indicates the final growth rate with their model. The blue line corresponds to the species-specific response to temperature for a thermal optimum of 15°C and the green line is the maximum achievable growth rate, a composite of the metabolic theory of ecology (MTE) and the formula by Eppley [114]. This formula is shown by black crosses in (b) and compared to the Eppley [114] curve in blue and Bissinger *et al.* [48] formula in purple. Horizontal lines show limits found in the literature [285].

The annual dynamics of phytoplanktonic organisms is usually 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 these variables being more or less correlated to the seasonal rythm, it is reasonable to restrain this study to the effect of one variable, temperature. Thus, the niche computed here is not the fundamental thermal niche, but a composite of all environmental conditions covarying with the temperature and promoting population growth. Such environmental conditions of course include solar irradiance which is strongly correlated to temperature in the field, but could also account for other factors, such as predation, that have a seasonal rythm which is partly captured by temperature. The niche part of the growth rate $f_i(T)$ therefore describes the *realized* niche of species i, not its fundamental niche.

We base our estimates of T_i^{opt} and b_i on field observations. For each taxon and each year, we define the beginning of the bloom as the date when the taxon 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 during which 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 the duration of its cumulated bloom days over a year is above the average duration of all blooms (137 days) for at least 15 years over the 20 years of the time series, and as a specialist if it falls below this threshold.

In the models, we assume that generalists have a niche width between 15 and 30° and specialists, between 5 and 10°. In order to compute b_i, we make the additional assumption that temperatures outside of this range lead to a growth rate at least 10 times inferior to the growth rate obtained at their thermal optimum ($\exp(-|7.5|^3/b_i = 0.1$ for a niche width of 15°). 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 within these boundaries. Meanwhile, taxa are ordered as a function of the mean cumulated bloom length and larger niche values are attributed to longer mean bloom length, i.e., $\sum \overline{L}_i > \sum \overline{L}_j \Rightarrow b_i > b_j$ where $\overline{L}$ is the mean over 20 years of the annual cumulated bloom lengths.

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

It should be noted that a variation in niche width also affects the final shape of thermal preferences. Indeed, when b_i increases, the niche term f_i has smaller variation in values around the thermal optimum. In this case, the final value of the growth rate is driven by the metabolism part of the equation (Fig. 2.S2).

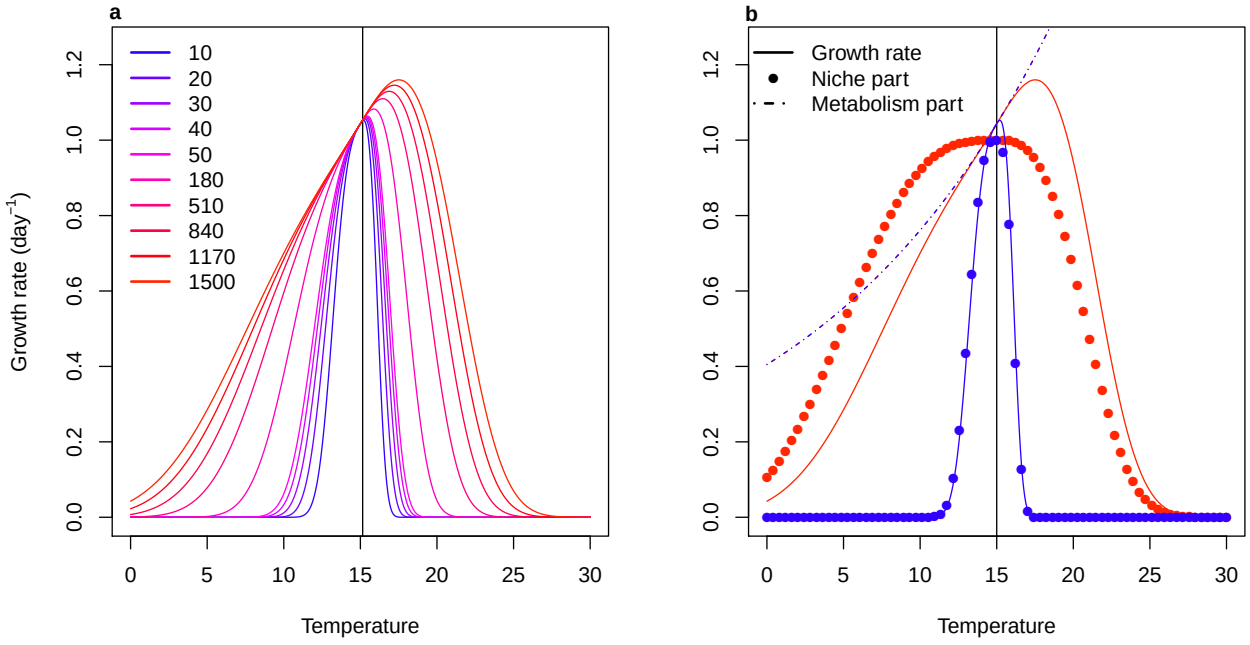

Figure 2.S2: Relationship between daily growth rates and temperature with different values of niche width b (whose values are indicated in the legend, corresponding to specialist and generalist species) and the same thermal optimum, 15°C, indicated by the solid black line (a). On the right panel, only the two extreme values of b (10 and 1500) are shown in blue and red respectively. Solid lines then correspond to the final growth rate, points correspond to $f_i(T)$ values (see eq. 2.S1) and the dotted line corresponds to $E(T)$ values.

We show the growth rates $r_i(T)$ as a function of temperature for each modeled taxon below (Fig 2.S3).

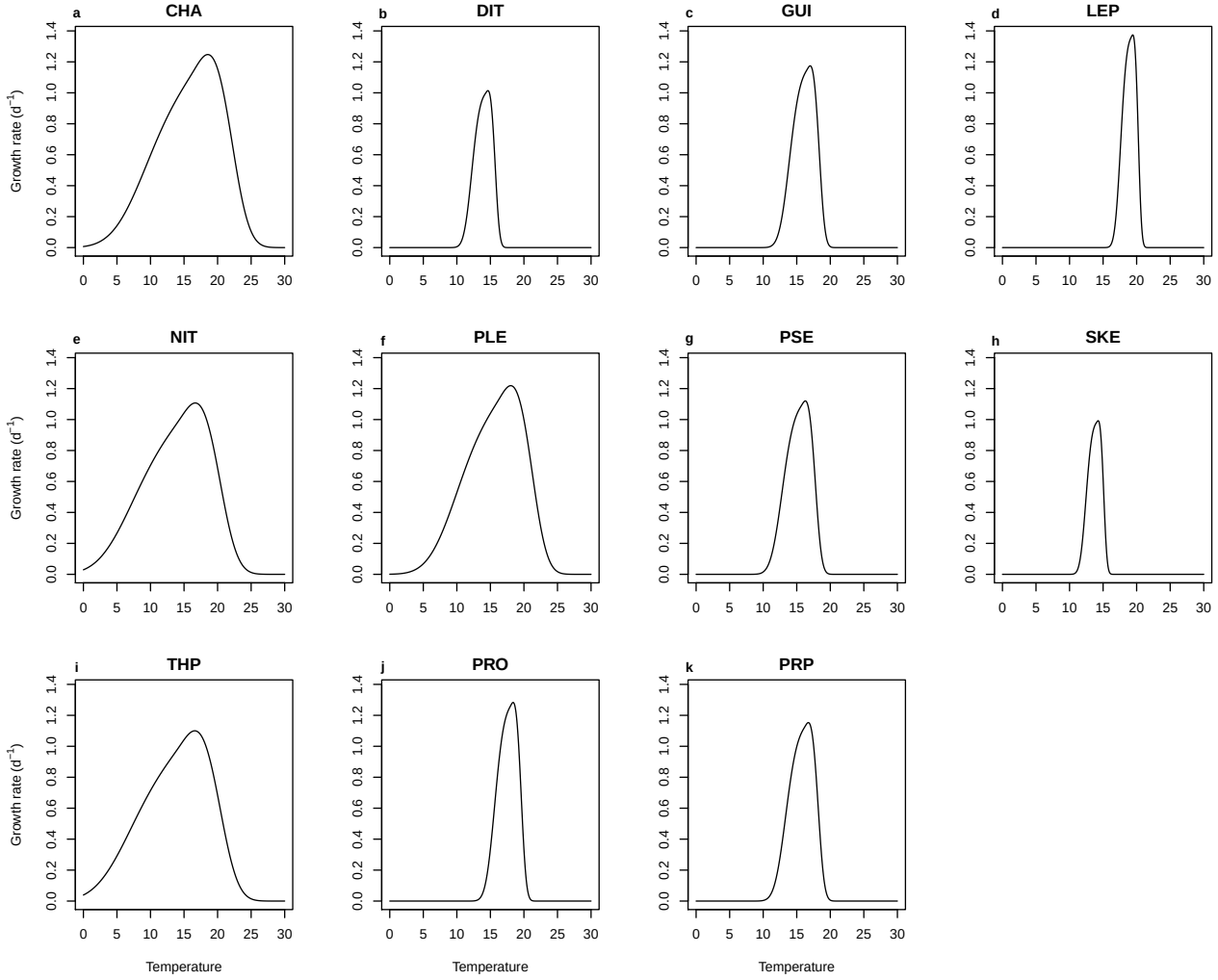

Figure 2.S3: Growth rate as a function of temperature for the taxa used in the model (names are defined in Table 2.S1).

2.S2 Initial estimates of interaction values

Model I: Lotka-Volterra interactions

Interactions between taxa have previously been computed with a Multivariate AutoRegressive model (eq. 2.S3, [271]).

$$\mathbf{n}_{t+1} = \mathbf{B}\mathbf{n}_t + \mathbf{C}\mathbf{u}_{t+1} + \mathbf{e}_t, \mathbf{e}_t \sim \mathcal{N}_{\mathcal{S}}(0, \mathbf{Q}) \tag{2.S3}$$

where S is the number of taxa, $\mathbf{n}_t$ is the $S \times 1$ vector of log-abundance of phytoplankton taxa, $\mathbf{B}$ is the $S \times S$ interaction matrix with elements b_{ij} (the effect of taxon j on taxon i), $\mathbf{C}$ is the $S \times V$ environment matrix describing the effects of variables $\mathbf{u}_{t+1}$ on growth rates and $\mathbf{e}_t$ is a $S \times 1$ noise vector following a multivariate normal distribution with a variance-covariance matrix $\mathbf{Q}$. The interaction model we use in the present paper is a Beverton-Holt multispecies model (eq. 2.1 in main text), also called at times Leslie-Gower. In the Supporting Information of [271], we showed that MAR and BH interaction coefficients, respectively b_{ij} and α_{ij}, could map once abundances at equilibrium N_i^* are defined.

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

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

$$\tilde{b}_{ij}(1 + f_A(i)) = -\alpha_{ij} N_j^*.$$

We then sum on columns (on j):

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

$$\Leftrightarrow -f_A(i)(1 + \sum_j \tilde{b}_{ij}) = \sum_j \tilde{b}_{ij}$$

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

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

$$\Leftrightarrow \alpha_{ij} = -\frac{1}{N_j^*}\frac{\tilde{b}_{ij}}{1 + \sum_j \tilde{b}_{ij}}.$$

This gives an exact correspondence between α_{ij} and b_{ij}. In the multispecies BH model, the presence of mutualistic interactions can lead to an orgy of mutual benefaction [228]. We impose

a minimum value of 1 to the denominator of the BH formulation, meaning that the growth rate cannot be higher than the maximum growth rate calculated, $r_i(T)$.

Model II: saturating interactions

We now move to a model with saturating interactions between taxa:

$$N_{t+1,i} = \frac{e^{r_i(T)} N_{t,i}}{1 + \sum_{j/a \in \mathbb{C}} \frac{a_C N_{t,j}}{H_{ij} + N_{t,j}} + \sum_{j/a \in \mathbb{F}} \frac{a_F N_{t,j}}{H_{ij} + N_{t,j}}} \tag{2.S4}$$

where coefficients a_C and a_F are the maximum interaction strengths for competition and facilitation, respectively, H_{ij} is the abundance of taxon j to reach half of the maximum effect of taxon j on taxon i, and $\mathbb{C}$ and $\mathbb{F}$ are the sets of competitive and facilitative interactions. This formula can be linked to the Unique Interaction Model by Qian & Akçay [279], i.e., each taxon provides a unique type of benefit or disadvantage to the focus taxon.

There is no single solution for matching the **B** matrix of the MAR model to model II including H_{ij}, a_C and a_F. We approximate the maximum interaction strength a_C as the average sum of all taxon effects $\alpha_{ij} N_j$ exerted on a given taxon if all interactions were competitive (eq. 2.S5). To compute a_F, we make two assumptions: on average, a) there is 70% facilitation in our dataset and b) the realized growth rate on a log-scale should not exceed $r_i(T)$, as in model I. We consider that the relationships that apply to individual interactions α_{ij} should also apply to the saturation point (eq. 2.S6), so that:

$$a_C = \frac{1}{S} \sum_i \left(\sum_j |\alpha_{ij}| N_{j,\max} \right) \tag{2.S5}$$

$$(1 - 0.7) a_C + 0.7 a_F = 0 \tag{2.S6}$$

where $N_{j,\max}$ is the maximum observed abundance of species j. We use $N_{j,\max}$ and the absolute value of interactions $|\alpha_{ij}|$ (i.e., all interactions are considered competitive in this case) to make sure that we maximise a_C.

At low abundances, we can consider that interactions are far from saturation. Taking the tangent of the function at this point, H_{ij} can be approximated by $f \frac{a_{C/F}}{\alpha_{ij}}$, where $f = 2$ is a correction factor that takes into account the fact that the slope at origin for the type II response is likely higher than the slope for a linear effect of density.

2.S3 Choice of parameters derived from literature

This section contains additional information on "fixed" parameter definitions (i.e., parameters not estimated from field data) and their chosen values. Note that these values are then subjected to sensitivity analysis, where by definition they are modified.

Loss rate The loss rate corresponds to multiple mortality processes. The Lotka-Volterra model of Scranton & Vasseur [301] considered a rate around 0.04 day^{-1}. In Jewson *et al.* [182], washout (0.5%), parasitism (4% of cells are infested and die) and grazing still remained low (about 0.05%) when compared to growth rates. Li *et al.* [212] found values between 0.02 and 0.1 day^{-1} for natural mortality only, while a review by Sarthou *et al.* [295] indicated a loss of daily primary productivity around 45% due to grazing only, while cell autolysis only can lead to a loss rate between 0.005 and 0.24 day^{-1} (in the absence of nutrients, or because of viral charge). Cumulating both natural mortality (cell autolysis) and grazing, we know that maximum rates should be above 0.24 day^{-1}. Trying to make a compromise emerge from the literature above, given that we are not always at the maximal loss rates, we set the reference value to 0.2 day^{-1}. This strikes a balance between acknowledging all sources of mortality and the need to keep all species in the reference model.

Sinking rate Among the hydrodynamics processes that drive the sinking rate, turbulence and eddies – themselves driven by tidal currents, the shape of the coast or wind conditions – are influential in keeping the cells at the top of the water column. For that reason, laboratory experiments on sinking rates are not sufficient to calibrate a field-based model. We therefore chose sinking rate values from field studies. In the Gotland Basin (central Baltic Sea), Passow [267] 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 highlighted, with a small number of genera that sank more than the rest of the community. The mean sinking 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 rates in Kowe *et al.* [197] in a river and Wiedmann *et al.* [332] in an estuary (mouth of Adventfjorden). When estimating changes of the sinking rate over time, values between 4 and 50% were obtained [182]. We therefore chose to represent the sinking rates with a Beta-distribution (Fig. 2.S4) which accounts for observed maximum and mean values, while still allowing a highly skewed distribution of sinking rates between species. High sinking rates are attributed to the morphotypes corresponding to *Chaetoceros* (CHA) and *Thalassiosira* (THP).

Mortality and burial in the seed bank McQuoid *et al.* [233] present maximum and mean depth of sediment 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 resting cells before death. Assuming m is the probability of mortality and survival follows a geometric law, the life expectancy of a resting cell is $\frac{1}{m} \Leftrightarrow m = \frac{1}{L_{\text{mean}}}$ where L_{mean} is the average duration of the dormant cell viability. Another way to look at the process is that life expectancy L follows the distribution $p(L > l) = e^{-ml}$. We arbitrarily chose that for the oldest dormant cells (i.e., the ones buried the deepest), $p(L > l_{\text{max}}) = 0.05$. Following this, $m = -\frac{\ln(0.05)}{l_{\text{max}}}$ where l_{max} derives from maximum

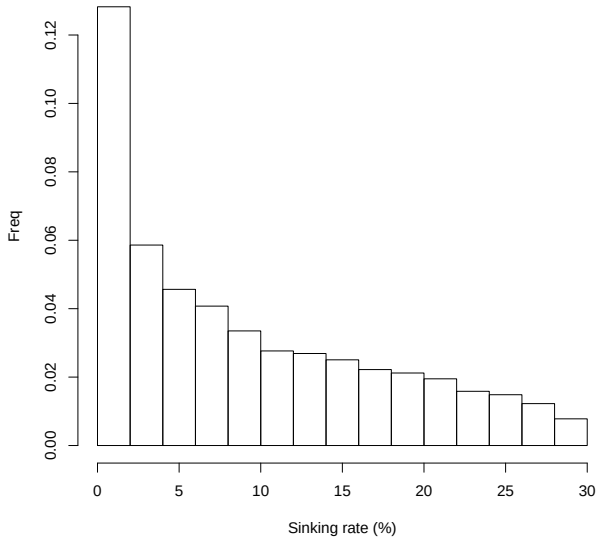

Figure 2.S4: Assumed Beta-distribution of sinking rates

depth values (for each species) in McQuoid *et al.* [233]. For both methods, m varies between 10^{-5} day^{-1} and 10^{-4} day^{-1} for all species considered.

As we highlight in the main text, burial is a very important process that controls the availability of resting cells, conditional to their survival in the sediment. However, burial rate is almost entirely dependent on the local sedimentation and no generally applicable literature could be found. We varied the burial rate ζ between 0.001 and 0.1 per day.

In scenarios where we remove the seed bank, we set $m + \zeta$ to 100% (for simplicity, we do not eliminate resting stage formation, only resting stage survival).

Resuspension As mentioned in the main text, resuspension values are mostly taken from models or data for inorganic particles. Rates vary greatly from one publication to another: in Fransz & Verhagen [127], 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 relationship between resuspension and the light extinction coefficient. In Kowe *et al.* [197], the resuspension rate of diatoms is evaluated around 1.9×10^{-5} day^{-1}. In Le Pape *et al.* [202], 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 burial, sinking rate and resuspension are all highly contingent upon the local hydrodynamics and therefore are intermingled processes.

2.S4 Phytoplankton taxa at the calibration site

Code	Taxa
CHA	*Chaetoceros*
DIT	*Ditylum*
GUI	*Guinardia*
LEP	*Leptocylindrus*
NIT	*Nitzschia+Hantzschia*
PLE	*Pleurosigma+Gyrosigma*
PRO	*Prorocentrum*
PRP	*Protoperidinium+Archaeperidinium+Peridinium*
PSE	*Pseudo-nitzschia*
SKE	*Skeletonema*
THP	*Thalassiosira+Porosira*

Table 2.S1: Name and composition of the phytoplanktonic groups used in main text, see Picoche & Barraquand [271] for more information on these taxa.

2.S5 Phytoplankton time series with model II

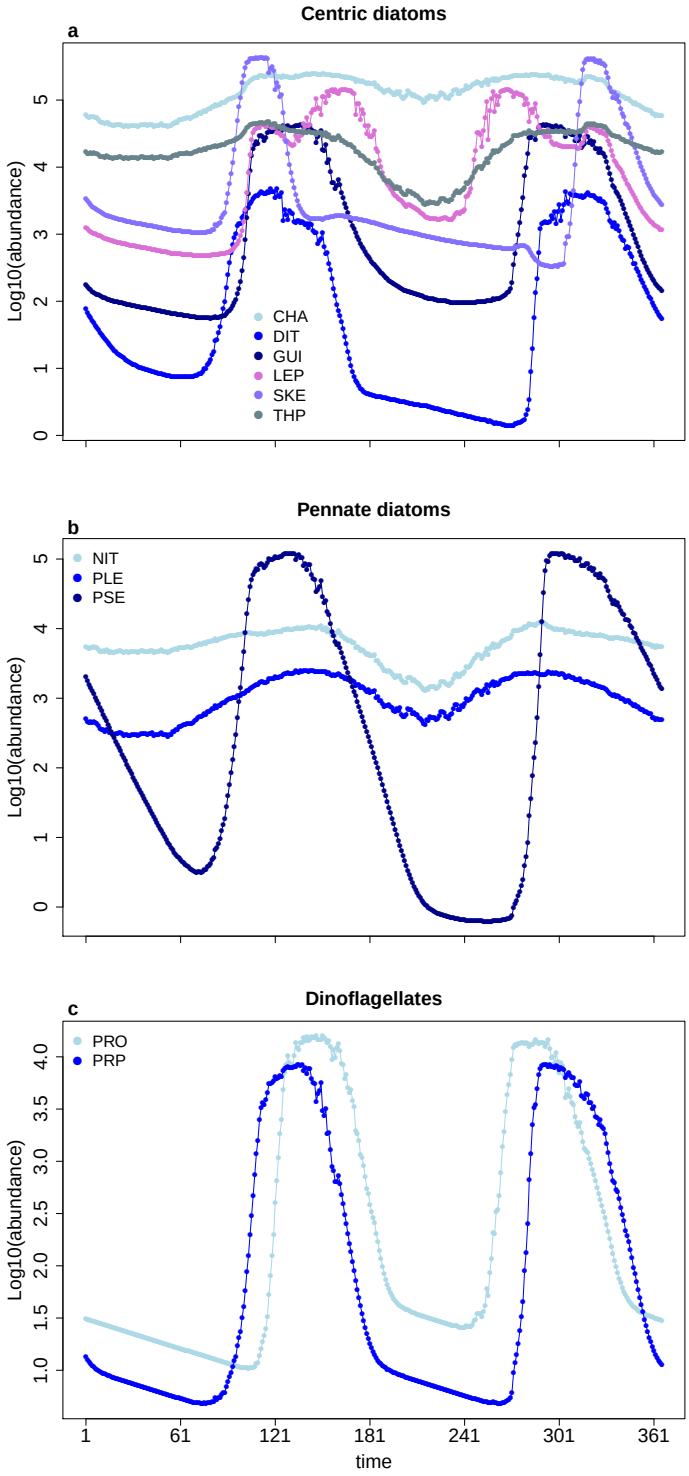

Figure 2.S5: Simulated phytoplankton time series for a year in model II (with saturating interactions). Each panel corresponds to a cluster of interacting taxa: centric diatoms (a), pennate diatoms (b) and dinoflagellates (c). Taxa only interact within their cluster (see Methods in main text).

2.S6 Scenario: changing interspecific interaction strengths but not intraspecific interaction strengths

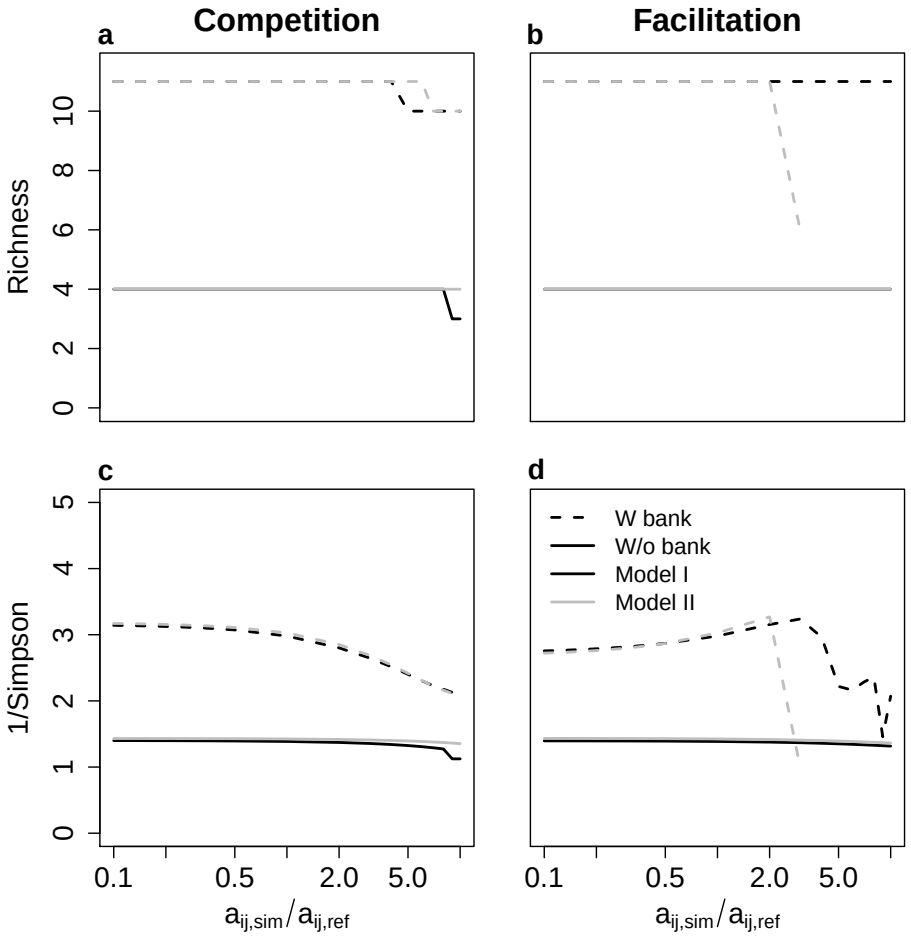

Figure 2.S6: Measures of biodiversity in the ocean at the end of the simulation: a-b) richness and c-d) inverse of the Simpson index, with (dashed line) and without (solid line) a seed bank, as a function of the strength of competition and facilitation with a classical Beverton-Holt (black lines) or a saturating interaction (grey lines) formulation, keeping the same values for the intraspecific interaction strengths. The x-axis shows the factor by which each interaction coefficient was multiplied, e.g., the value 0.1 indicates that the interaction strengths in the simulation are 10 times lower than the interactions strengths in the reference simulation. Note the logarithmic scale.

When intraspecific interaction strengths do not change, reducing the values of interspecific competition has very little effect on both richness and the inverse of the Simpson index. Increasing facilitation finally destabilizes the community and leads to the observed diversity decrease in the absence of a seed bank. We do stress, however, that this scenario is more of a thought experiment changing niche differentiation rather than anything mimicking an environmental perturbation, where the factors that change the degree of competition between species will likely change competition within species too.

2.S7 Growth rates and survival without a seed bank

In order to investigate the relationships between population dynamics and survival probabilities, we computed the realized per capita growth rates ($\text{PCGR}_i = \exp(r_i(T))/C_i$, where the competition C_i is defined in Section 2.S8) of each species in simplified conditions (Fig. 2.S7):

- when all species abundances are set to 1, so that there is nearly no competition
- when all species abundances are set to their average abundance in the environment, as a proxy of competition endured by each species throughout the year
- when temperature is set to the average temperature of the environment
- when temperature is either optimal for each species, or too high for all species

The population grows when $\text{PCGR}_i > 1 + l$, with l the loss rate.

It should be noted that the realized growth rate is not computed using long-term simulation, but for a set of fixed environmental and competition values, in the abovementioned conditions. Growth rates are then related to the probability of survival in the ocean without a seed bank when varying interaction strengths (scenario 1 in the main text). Probability of survival is itself computed over different values of interaction strengths.

Extinction in the absence of the seed bank is mostly due to a narrow niche: extinct species are always specialists (Fig. 2.5), even though these species should be able to invade a typical environment (Fig. 2.S7 a) and tend to be less regulated than surviving species (Fig. 2.S7 c, d).

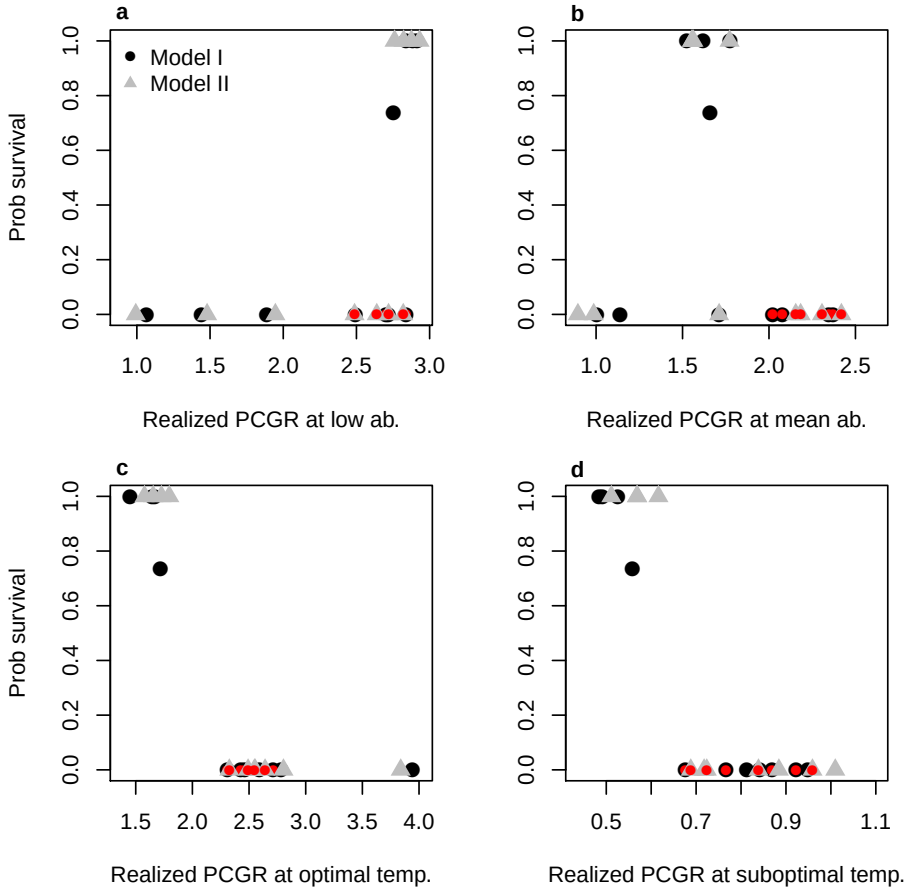

Figure 2.S7: Probability of survival in the absence of a seed bank, as a function of the realized per capita growth rates (PCGR) of each species in different conditions. We consider (a) mean temperature and low densities, i.e., temperature is set to 15°C and all species densities are 1 cell/L, or (b) mean temperature (15°C) and average densities, i.e., all species are at their simulated average density values; or (c) optimal temperature for each species and average densities, or (d) suboptimal temperatures (30°C) and average densities. Red points correspond to species which have a high growth rate at low density but still go extinct.

2.S8 Absence of a storage effect in the model

The first step of our model (eqs. 2.1 and 2.3 in main text) describes the increase in abundance of coastal and oceanic populations due to both environmental fluctuations and interactions with other organisms. The formula may be interpreted by certain readers as already including a storage effect, as shown by Miller & Klausmeier [236] for a similar continuous-time model, or due to its apparent similarity to the model of Chesson [75] in discrete time. We show below why the results of the cited analyses do not apply here and that the intra-compartment growth in discrete-time model does not lead to a storage effect.

The storage effect requires three elements: (a) a positive covariance between good environmental conditions and competition, (b) species-specific environmental responses and (c)

subadditivity of environmental and competitive effects on the growth rate [75]. Condition (a) is met as covariance may be created by temporally correlated fluctuations in the environmental signal [300]. Condition (b) is also met, as shown in Figure 2.S3. Condition (c), however, is not met. Subadditivity of environmental and competitive effects can be mathematically expressed as $\frac{\partial}{\partial E}\left(\frac{\partial g}{\partial C}\right) < 0$ where g is the growth rate, and E and C are the effects of the environment and competition on the growth rate respectively.

In model I, the growth rate of a population is defined by the following equation:

$$N_{t+1,i} = \frac{\exp(r_i(T))N_{t,i}}{1+\sum_j \alpha_{ij}N_{t,j}} - lN_{t,i}.$$

Here $E_i(t) = \exp(r_i(T))$ and $C_i(t) = 1 + \sum_j \alpha_{ij}N_{t,j}$. Therefore,

$$\begin{aligned} g_i &= \log\left(\frac{N_{t+1,i}}{N_i}\right) \\ &= \log\left(\frac{\exp(r_i(T))}{1+\sum_j \alpha_{ij}N_{t,j}} - l\right) \\ &= \log\left(\frac{E}{C} - l\right). \end{aligned}$$

Miller & Klausmeier [236] consider instead a per capita growth rate (fitness) in continuous time $\frac{1}{N_i}\frac{dN_i}{dt}$.

While the discrete-time formula is at first sight very similar to the one in Chesson [75], we wish to draw the reader's attention to the fact that, contrary to Chesson [75], interaction strengths do not depend on environmental effects, nor are they linearly correlated.

If we take derivatives with regards to competition effect C:

$$\begin{aligned} \frac{\partial g}{\partial C} &= -\frac{1}{C^2}\frac{E}{\frac{E}{C}-l} \\ &= -\frac{E}{CE-lC}. \end{aligned}$$

Finally, we take a derivative with respect to E:

$$\begin{aligned} \frac{\partial}{\partial E}\left(\frac{\partial g}{\partial C}\right) &= -\frac{CE-lC-EC}{(CE-lC)^2} \\ &= \frac{lC}{(CE-lC)^2}. \end{aligned}$$

As $C = 1 + \sum_j \alpha_{ij}N_{t,j}$ in the first model, or $C = 1 + \sum_{j\in\mathbb{C}} \frac{a_C N_{t,j,c/o}}{H_{ij}+N_{t,j,c/o}} + \sum_{j\in\mathbb{F}} \frac{a_F N_{t,j,c/o}}{H_{ij}+N_{t,j,c/o}}$ in the second model, which is never negative and always above 1, we always have $\frac{\partial}{\partial E}\left(\frac{\partial g}{\partial C}\right) > 0$. There is no buffered growth in the effective growth rate itself, and therefore no storage effect in the coastal or oceanic compartment dynamics.

Now we highlight a more delicate point, in models that include the seed bank compartment. The storage effect could also be due to the presence of the seed bank, whose exchanges with other compartments are described in the second step of the model (eq. 2.4 in main text). The presence of a seed bank is indeed often associated with a storage effect maintaining coexistence [13], even though models show that this association is not systematic [5]. However, for the

storage effect to happen due to the seed bank, theory currently highlights that environmental variations have to affect the recruitment rate, i.e., in our case, the germination and resuspension rates [75]. In our model, neither environmental nor competition effects affect the exchanges with the seed bank: the sinking and germination/resuspension rates are fixed and independent from both past and current environmental conditions. The storage effect is therefore unlikely to happen in our model. A thorough analysis of the storage effect based on simulations, as described by Ellner *et al.* [113], would however be needed to conclude with complete certainty, as the presence of three distinct compartments may create non-obvious non-additivities or non-linearities.

Chapter 3

Replication: Reproductive pair correlations and the clustering of organisms

3.1 Foreword: the reproducibility crisis

We present in this chapter the replication of Young *et al.*, 2001 [345]. Such type of studies is becoming more and more important as questions regarding reproducibility intensify. I recapitulate here some of the arguments that have been made in favor of reproduction in science.

The 'reproducibility crisis' qualifies the increasing wariness stemming from the inability to repeat the findings of published research (note that the mere existence, or, at least magnitude, of this crisis remains disputed [179, 119]). Reproducibility is a pillar of science: observations and experiments should be "inter-subjectively testable" [277], in order to avoid as much as possible personal bias, potential mistakes, or outright voluntary deceit. Two types of reproductions can be distinguished[1]: 'from-scratch' *replications* which require the production of new data and computer program (if necessary) following the methods described in the original paper, or *reproductions* which use data and software provided by the original authors to ensure final results are the same [290]. The reader should however be aware that these definitions of replication and reproduction remain disputed between fields.

The failure to reproduce or replicate a published study is a common experience: large-scale surveys show that approximately 70% of scientists are confronted at least once to this difficulty [46, 26], with more than 50% of scientists not being able to replicate their own experiments [26]. When formal replication projects are developed, results are striking: for instance, only 11% of oncology [37] and 40% of psychology studies [25] could be successfully replicated. The implications of such results are grim, as the potential distrust towards non-reproduced articles finally extends to the rest of scientific research (the provocative title 'Why Most Published Research Findings Are False' [170] is an extreme example – the content of the article itself being more constructive).

There can be many reasons behind reproduction or replication failure[2]. Debatable experimental design or analysis by the original laboratory, either due to lack of means, experimental or methodological formation, and/or questionable research practices [128], constitute the core

[1] An adjacent type of reproduction puts the focus on the conclusions of a study, which can be either confirmed or falsified by a different experiment. It therefore does not require to repeat the focus study.

[2] While 'pressure to publish' is often cited as a factor contributing independently to poor reproducibility [46, 26], it could be argued that it constitutes an underlying phenomenon for most of the other factors presented here.

of the issues according to the majority of scientists [26]. In many scientific fields, the focus has been put on null hypothesis significance testing and the (over-)use of p-values to establish findings, from its methodological limitations and potentially misinformed uses [255, 147], to the emergence of practices such as 'p-hacking' [149, 266]. This has become a subject in and of itself which we need to be mindful of, but which exceeds, by far, the scope of this section. Missing information from published studies (encompassing unavailable raw data, insufficient method description – from missing parameter values to assumptions or details deemed futile or obvious at the time of writing –, or absence of code when applicable) is another important obstacle. While it seems easier to prevent, it can still constitute up to 45% of the replication difficulties [46]. In ecology for instance, data and code are not always available: in a survey of 174 articles in 2018-2019, if 79% of the articles provided their data, only 30% of them gave access to the corresponding code despite code-sharing policies of a rising number of journals [89]. Even when codes are available, computational reproducibility is hindered by the use of proprietary software and the lack of maintenance and user support [157]. Finally, 'honest mistakes', cases where the code used in the analysis does not correspond to the written (and intended) methods, are probably another common source of replication failure.

In light of this assessment, more work should be put into replicating and reproducing published results. This is however not encouraged by the current state of science. As written in [259], "innovative findings produce rewards of publication, employment, and tenure; replicated findings produce a shrug". Even when accepting to spend time on a less 'rewarding' replication, researchers may be faced with many barriers when contacting journals to address errors [9]. And when falsification is proven, this does not mean that the original research and its conclusions will not remain a classical reference as a so-called 'zombie paper', retracted but still cited as proof [43]; replicable publications also tend to be cited less than nonreplicable ones [37, 303].

In recent years, tools have been developed to overcome these issues. Among other checklists, the 'Tools for Transparency in Ecology and Evolution' offer guidelines for data, methods and material archiving, as well as reporting of methods, pre-registration of studies and analyses, and replication [266]. Reproducibility initiatives emerge, such as the collaboration between different scientific journal and tools[3], or joint challenges such as the 'Ten Years Reproducibility Challenge'[4], during which scientists were encouraged to try and re-run their own code 10 years after publication. In addition to 'journals of negative results', meant to encourage publications that would not nurture the bias towards positive results [118], journals dedicated to replication now exist: ReScience C and ReScience X are both academic journals which only publish replications and reproductions for computational and experimental research respectively [290]. Having a place to publish partly defuses the feeling of an 'unrewarding' work, as there is a publication to highlight the work done. Finally, the 'Retraction Watch' organization[5] has started a database

[3] http://validation.scienceexchange.com/reproducibility-initiative, last accessed June, 17th 2022
[4] https://rescience.github.io/ten-years/, last accessed June, 20th 2022
[5] https://retractionwatch.com, last accessed June, 20th 2022

of retracted works that can be used as a qualitative survey of current issues.

In the replication we present, we not only confirm the results of Young *et al.*, 2001 [345], but also make the corresponding program available for future users, and provide further information on mathematical analyses. These mathematical proofs were indeed absent from the original paper and had to be performed again.

[Re]: Reproductive pair correlations and the clustering of organisms

Coralie Picoche[1,*], William R. Young[2] & Frédéric Barraquand[1]
[1]Institute of Mathematics of Bordeaux, University of Bordeaux & CNRS
[2]Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California, USA
[*]corresponding author: coralie.picoche@u-bordeaux.fr

Published in ReScience C (2022) doi:10.5281/zenodo.6546488

3.2 Introduction

In the present work, we replicate the results of Young et al. 2001 "Reproductive pair correlations and the clustering of organisms" [345], an analysis of the formation of aggregates in an otherwise homogeneous environment mimicking marine small-scale hydrodynamics. Using an individual-based model of independent, random-walking particles (also called "Brownian bugs"), they show that reproduction by fission in a turbulent [318] and viscous flow leads to the formation of elongated clusters. Spatial patterns therefore depart from the usual, homogeneous solution of the advection-diffusion-reaction equation for a large population.

Due to their size, phytoplankton organisms experience a mostly viscous environment in a laminar shear field, with random but homogeneous changes in directions due to turbulence [104, 269]. Reproduction and limited movement of daughter cells, which occur at the phytoplankton scale, interact with these hydrodynamic processes and can lead to aggregates. In this context, a better understanding of the interactions between demography and small-scale hydrodynamics could provide further explanation for observed spatial distribution of phytoplankton species, and perhaps even their coexistence. This motivated us to revisit Young et al. 2001 [345].

In addition to replicating the numerical and mathematical results of Young et al. 2001, we also wished to present the mathematical derivations that were missing from the original paper, which should make this replication article more accessible to most readers, especially those without a fluid mechanics background.

3.3 Brownian bug model

The Brownian bug model is defined as an individual-based model in continuous space and time, here presented in its 2D formulation. For efficient computer simulation, it is implemented in discrete time [345]. Each particle is characterized by the vector of its Cartesian coordinates

$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and its original position on the y-axis at $t = 0$ (a child particle inherits this attribute), this last characteristic being used only for representation purposes. Space is a $L \times L$ square with periodic boundary conditions. Each timestep, of duration τ, is divided into three substeps: (1) demographic processes, (2) diffusion, and (3) advection.

Demographic processes take place during the first substep (1). Each organism has a fixed probability (p) of reproducing, dying (q), or remaining unchanged ($1-p-q$). When an individual reproduces, a new organism appears on top of the parent. In the following, $p = q = 0.5$. Diffusion is then modeled as a Brownian motion (2), i.e., $\boldsymbol{x'}(t) = \boldsymbol{x}(t) + \delta\boldsymbol{x}(t)$ where each component of $\delta\boldsymbol{x}(t)$ follows a Gaussian distribution $\mathcal{N}(0, \Delta^2)$ where $D = \frac{\Delta^2}{2\tau}$ is the diffusivity. The discrete-time Markov chain presented here approximates the continuous-time Brownian bug model, which can be thought of as a spatial birth-death or branching process (described in Section 3.S1 in the Supplementary Information), step (1) being referred to in Young et al. [345] as a Galton-Watson process. Finally, (3) the turbulent flow governing advective stirring follows the Pierrehumbert random map [274].

$$x_1(t+\tau) = x_1'(t) + (U\tau/2)\cos[kx_2'(t) + \phi(t)] \tag{3.1}$$

$$x_2(t+\tau) = x_2'(t) + (U\tau/2)\cos[kx_1(t+\tau) + \theta(t)] \tag{3.2}$$

where $\phi(t)$ and $\theta(t)$ are random phases uniformly distributed between 0 and 2π, $k = 2\pi/L$ and U is the stretching parameter.

Unless otherwise specified, each simulation is initialized with $N_0 = 20,000$ particles uniformly distributed in a 1×1 square and run for 1000 timesteps.

Pair density function $G(r,t)$

The pair density function $G(\boldsymbol{x}_i, \boldsymbol{x}_j, t)$ is defined so that $G(\boldsymbol{x}_i, \boldsymbol{x}_j, t)dA_1dA_2$ is the probability of finding a pair of Brownian bugs with one member in the area dA_1 around $\boldsymbol{x}_i$ and the other in the area dA_2 around $\boldsymbol{x}_j$. Defining $\boldsymbol{\xi} = \boldsymbol{x}_i - \boldsymbol{x}_j$, $G(\boldsymbol{\xi}, t)$ is actually called the pair correlation function in [345]. The radial density function $g(r,t)$ is defined as $G(\boldsymbol{\xi}, t) = C^2 g(r,t)$ where C is the concentration of bugs and $r = |\boldsymbol{\xi}|$. As the pair correlation disappears when $r \to \infty$, $g \to 1$.

Derivation of $G(r,t)$

All details of the derivation of $G(r,t)$ are to be found in Section 3.S1 in the Supplementary Information. We finally obtain:

$$\frac{\partial G}{\partial t} = 2Dr^{1-d}\frac{\partial}{\partial r}\left(r^{d-1}\frac{\partial G}{\partial r}\right) + 2(\lambda-\mu)G + \gamma r^{1-d}\frac{\partial}{\partial r}\left(r^{d+1}\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{\xi}) \tag{3.3}$$

where d is the number of dimensions, λ is the birth rate ($p = \lambda\tau$) and μ is the death rate ($q = \mu\tau$).

We focus on the case $d = 2$ and $\lambda = \mu$, which means Eq. 3.3 can be reduced to

$$\frac{\partial G}{\partial t} = \frac{2D}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{\gamma}{r}\frac{\partial}{\partial r}\left(r^3\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{\xi}). \tag{3.4}$$

The value of γ is computed from simulations (see Section 3.S2 in the Supplementary Information).

Analytical solution with advection

CP and FB could only find the analytical solutions of $G(r,t)$ with and without advection with the indications of WY. In the presence of advection ($\gamma \neq 0$), a steady-state solution can be found; without advection, there is no steady-state and the solution changes through time. Let us first examine the steady-state solution, given by

$$\begin{aligned}
&\frac{2D}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{\gamma}{r}\frac{\partial}{\partial r}\left(r^3\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{\xi}) = 0 \\
\Leftrightarrow\ & 2\pi r\left(\frac{2D}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{\gamma}{r}\frac{\partial}{\partial r}\left(r^3\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{\xi})\right) = 0 \\
\Leftrightarrow\ & 2\pi\left(2D\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \gamma\frac{\partial}{\partial r}\left(r^3\frac{\partial G}{\partial r}\right)\right) + 2\pi r 2\lambda C\delta(\boldsymbol{\xi}) = 0.
\end{aligned} \tag{3.5}$$

We can then integrate Eq. 3.5 over a small area centered on a particle, with radius ρ. Let us first note that

$$\begin{aligned}
&\int_{\mathbb{R}^2}\delta(\boldsymbol{x})d\boldsymbol{x} = 1 \\
\Leftrightarrow\ &\int_0^{2\pi}\int_0^{\rho}\delta(r)\delta(\theta)r\,dr\,d\theta = 1 \\
\Leftrightarrow\ &2\pi\int_0^{\rho}\delta(r)r\,dr = 1.
\end{aligned} \tag{3.6}$$

Using Eq. 3.5 and 3.6, we can integrate between 0 and ρ,

$$\begin{aligned}
0 =\ & 2\pi\left(2D\rho\frac{\partial G}{\partial \rho} + \gamma\rho^3\frac{\partial G}{\partial \rho}\right) + 2\lambda C \\
\Leftrightarrow\ & \frac{\partial G}{\partial \rho} = -\frac{1}{2\pi}\frac{2\lambda C}{2D\rho + \gamma\rho^3}.
\end{aligned} \tag{3.7}$$

Eq. 3.7 can now be integrated between ρ and ∞, knowing that $G(\infty) = C^2$:

$$C^2 - G(\rho) = -\frac{1}{2\pi}\int_\rho^\infty \frac{2\lambda C}{2Dr + \gamma r^3} dr. \tag{3.8}$$

Using the variable change $u = \frac{2D}{r^2} + \gamma$, with $du = \frac{-4D}{r^3}dr$, the integral writes

$$\begin{aligned} G(\rho) &= C^2 + \tfrac{\lambda C}{\pi}\int_{\frac{2D}{\rho^2}+\gamma}^{\gamma} \tfrac{1}{r^3 u}\tfrac{r^3}{-4D} du \\ &= C^2 - \tfrac{\lambda C}{4\pi D}\left[\ln u\right]_{\frac{2D}{\rho^2}+\gamma}^{\gamma} \\ &= C^2 - \tfrac{\lambda C}{4\pi D}\left(\ln(\gamma) - \ln\left(\tfrac{2D}{\rho^2} + \gamma\right)\right) \\ &= C^2 + \tfrac{\lambda C}{4\pi D}\ln\left(\tfrac{2D+\gamma\rho^2}{\gamma\rho^2}\right). \end{aligned} \tag{3.9}$$

Finally, the pair correlation function $g = G/C^2$ is defined as

$$g(r) = \frac{\lambda}{4\pi DC}\ln\left(\frac{2D + \gamma r^2}{\gamma r^2}\right) + 1. \tag{3.10}$$

Analytical solution without advection

When $U = 0$, $\gamma = 0$ and there is no steady solution in 2D. We can get back to Eq. 3.4:

$$\frac{\partial G}{\partial t} = \frac{2D}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{\xi}). \tag{3.11}$$

Assuming an isotropic environment (and switching to the Cartesian coordinate system), this means

$$\frac{\partial G}{\partial t} - 2D\Delta G = 2\lambda C\delta(\boldsymbol{\xi}) \tag{3.12}$$

where $\Delta = \nabla^2$ is the Laplacian operator.

We therefore have

$$\mathcal{L}G(\boldsymbol{\xi}, t) = 2\lambda C\delta(\boldsymbol{\xi}) \tag{3.13}$$

where $\mathcal{L}$ is the linear differential operator $\partial_t - 2D\Delta$.

We can use a Green's function H, defined with $\mathcal{L}H = \delta(\boldsymbol{\xi}, t) = \delta(\boldsymbol{\xi})\delta(t)$.

By definition, we know that $G(y) = \int H(y, s)2\lambda C\delta(s)ds$ (where $y = (\boldsymbol{\xi}, t)$) is a solution to Eq. 3.13.

$$\begin{aligned} G(\boldsymbol{\xi}, t) =& 2\lambda C\int_{\mathbb{R}^2}\int_0^t H(\boldsymbol{\xi} - \boldsymbol{\xi}', t')\delta(\boldsymbol{\xi}')d\boldsymbol{\xi}'dt' \\ =& 2\lambda C\int_0^t H(\boldsymbol{\xi}, t')dt'. \end{aligned} \tag{3.14}$$

Eq. 3.14 can be used in Eq. 3.11:

$$\frac{\partial}{\partial t}\left(2\lambda C\int_0^t H(\boldsymbol{\xi},t')dt'\right) = 2D2\lambda C\Delta\int_0^t H(\boldsymbol{\xi},t')dt' + 2\lambda C\delta(\boldsymbol{\xi}) \quad (3.15)$$

$$\Leftrightarrow \int_0^t \left(\frac{\partial H(\boldsymbol{\xi},t')}{\partial t'} - 2D\Delta H(\boldsymbol{\xi},t')\right)dt' = \delta(\boldsymbol{\xi}) \quad (3.16)$$

$$\Leftrightarrow \int_0^t \delta(\boldsymbol{\xi})\delta(t')dt' = \delta(\boldsymbol{\xi}) \quad (3.17)$$

which is true.

A solution for the Green's function using $\mathcal{L} = \partial_t - 2D\Delta$ in 2 dimensions is

$$H(r,t) = \frac{1}{4\pi 2Dt}\exp\left(\frac{-r^2}{4\times 2Dt}\right). \quad (3.18)$$

$G(r,t)$ can then be computed:

$$G(r,t) = 2\lambda C\left[\frac{E_1\left(\frac{r^2}{8Dt'}\right)}{8D\pi}\right]_0^t \quad (3.19)$$

where $E_1(x) = \int_x^\infty \frac{e^{-t}}{t}dt$ is the exponential integral. Using $G(r,0) = C^2$ and $\lim_{x\to+\infty} E_1 = 0$ in Eq. 3.19, we finally obtain

$$G(r,t) = \lambda C\frac{E_1\left(\frac{r^2}{8Dt}\right)}{4D\pi} + C^2 \quad (3.20)$$

$$\Leftrightarrow g(r,t) = \frac{\lambda}{C}\frac{E_1\left(\frac{r^2}{8Dt}\right)}{4D\pi} + 1. \quad (3.21)$$

3.4 Results

We were able to reproduce the three figures of Young et al. [345] highlighting the spatial distributions of Brownian bugs.

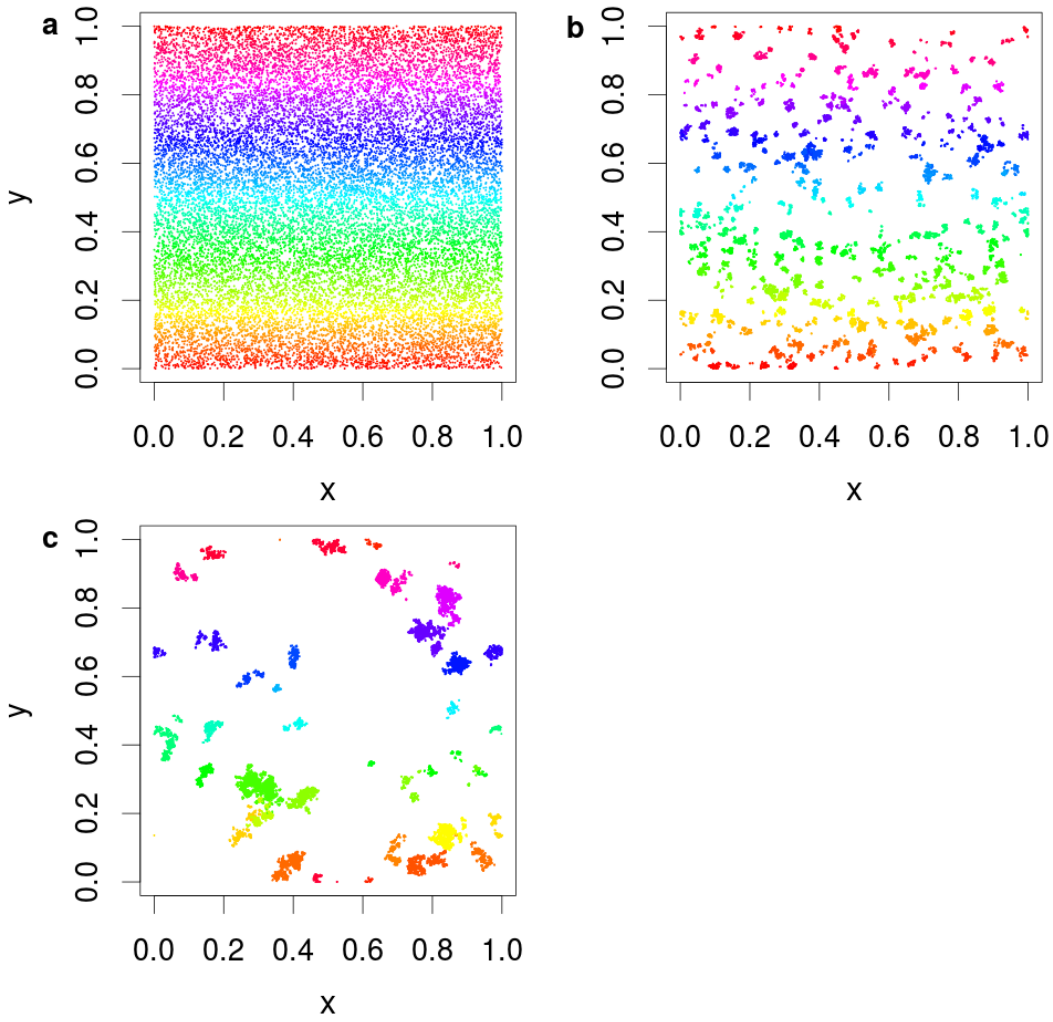

Figure 3.1: Distribution of Brownian bugs at different times in a simulation with $\Delta = 10^{-3}$ and $U = 0$: initial conditions with a Poisson spatial distribution (a), $t = 100\tau$ (b) and $t = 1000\tau$ (c). Each particle is identified by a color which corresponds to the initial position on the y-axis of its ancestor at $t = 0$.

In Fig. 3.1, the model has been run without the advection component and we can see the clumping of organisms due to reproduction. In Fig. 3.2 a), the model has been run without its demographic component, but with advection and diffusion, confirming that hydrodynamics alone cannot ensure cluster formation, while in Fig. 3.2 b), advection, diffusion and demography are present, in which case organisms form elongated aggregates.

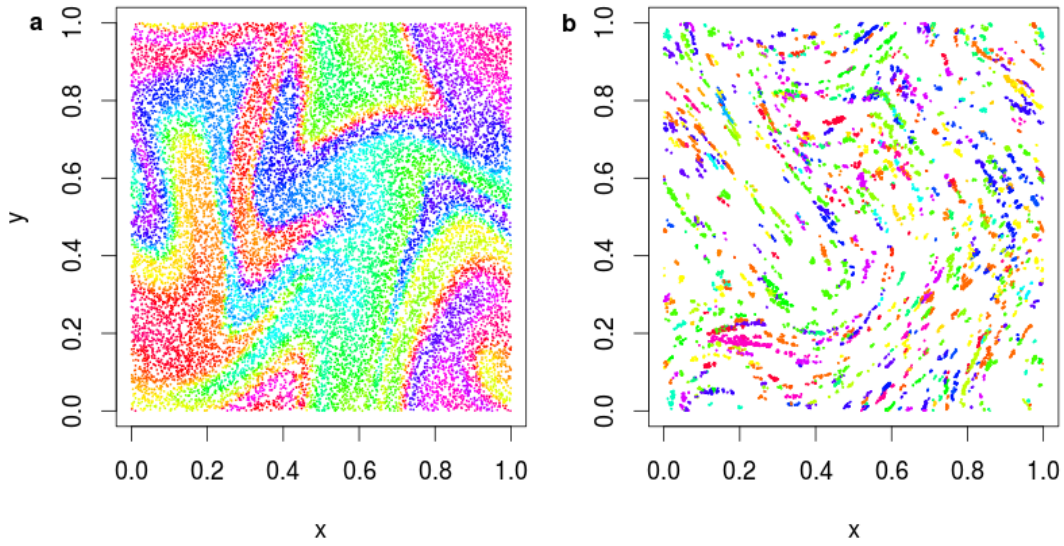

Figure 3.2: Distribution of Brownian bugs in a simulation with advection and $\Delta = 10^{-3}$, $U\tau/2 = 0.1$: without demographic processes at $t = 30\tau$ (a), and with demographic processes at $t = 1000\tau$ (b). Each particle is identified by a color which corresponds to the initial position on the y-axis of its ancestor at $t = 0$.

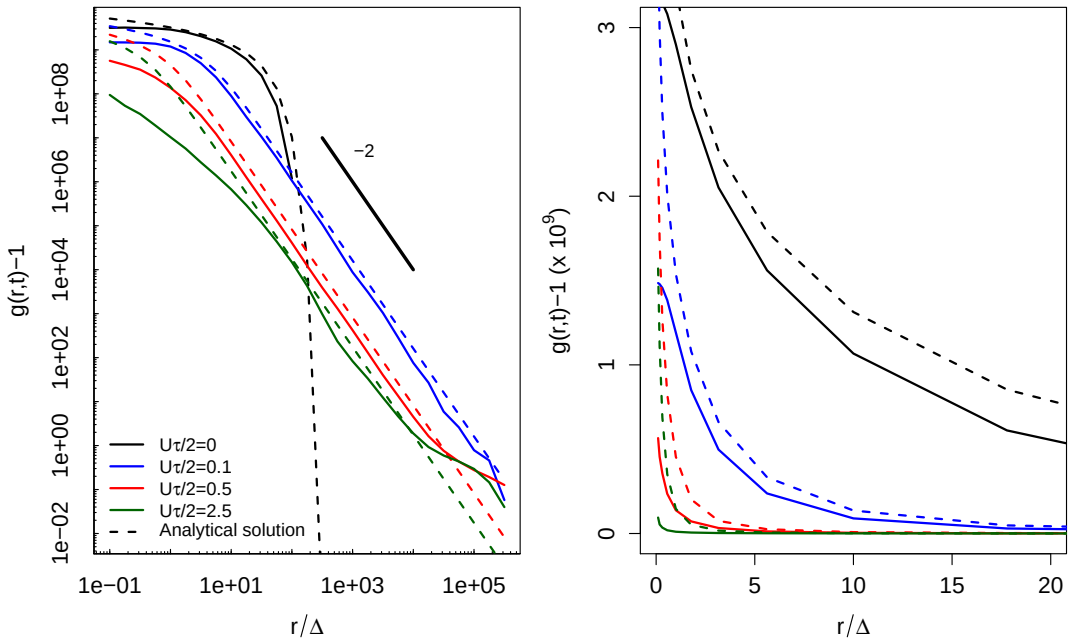

Figure 3.3: Logarithmic (a) and linear (b) plots of $g(r,t)$ versus r/Δ, with $\Delta = 10^{-7}$ and $U\tau/2 = 0, 0.1, 0.5, 2.5$ at $t = 1000\tau$. To compute simulation-based values of $g(r)$ with a large number of points ($N_0 = 200,000$) and avoid sampling issues, we replicated the 1×1 square 10 times, so that its length is $\sqrt{10}$ while keeping $L = 1$ and $k = 2\pi/L$ in eq. 3.1 and 3.2. Solid lines result from simulations, dotted lines correspond to analytical solutions and the solid grey line indicates the r^{-2} scaling predicted by Eq. 3.3.

Fig. 3.3 proved much more challenging. Retrieving the analytical solutions of Eq. 3.4 was difficult as there was no other equation than Eq. 3.3 in the original paper. We also encountered issues when computing the pair correlation functions on simulations: for large values of r/Δ, we observed zero values (absent pairs) of the pcf when $U = 0$. This is a sampling effect as pcf values get very low for large distances without advection, even though we multiplied the study area by 10 to produce Fig. 3.3 and counter such effects. Despite the missing values, we can confirm that simulated and analytical pcf match (Fig. 3.3), with a slight underestimation by the simulations. The numerical pcf also got closer to 1 than the analytical predictions for large values of r.

3.5 Discussion

We successfully replicated both the numerical results and analytical solutions of Young et al. [345]. Even though stochasticity prevents us from replicating exactly the same spatial point patterns as those seen in the original Fig. 3.1 and Fig. 3.2, we considered the patterns to be close enough to validate the replication. Fig. 3.3 was also very close to the one shown in the original article, despite a slight underestimation of the pcf in simulated data.

The most challenging part of the replication was actually not to replicate the numerical results, but to find back the analytical expression of the pair density function $G(r, t)$ dynamics from first principles. How to derive such dynamics was indeed briefly explained in words in the original article, but the many intermediate steps involved (see Supplementary Material) make the additional mathematical derivations presented here worthwhile in our opinion. In addition to providing critical information to CP and FB regarding how to obtain the pair correlation dynamics, WY also communicated the required mathematical steps to find back the analytical solutions for $G(r, t)$ plotted in Fig. 3.3. We hope that the additional material on the derivation of $G(r, t)$ dynamics (as Supplementary Material) as well as the provided analytical solutions of such dynamics (now presented in the main text) will help readers through both the original and replication articles.

The original article did not provide quantitative values exactly matching marine microbes ecology; we thus wondered about the time and spatial scales that could be used for a realistic phytoplankton model. The length of the square side, L, is defined roughly as the Kolmogorov scale [318], the scale at which viscosity starts dominating turbulence (we use $L = 1$ cm as an upper bound). Here, we consider k the smallest wavenumber corresponding to the largest length scale L, i.e., $k = 2\pi/L$. The chosen length scale defines the Reynolds number which then allows to obtain U, the velocity difference between two points separated by a distance L.

$$\begin{aligned} \mathrm{Re} &= \frac{U}{k\nu} && (3.22)\\ \Rightarrow 1 &= \frac{UL}{2\pi\nu} && (3.23)\\ \Leftrightarrow U &= \frac{2\pi\nu}{L} && (3.24) \end{aligned}$$

where $\nu = 10^{-6}$ m^2 s^{-1} is the kinematic viscosity for water. These numerical values lead to $U = 6.3 \times 10^{-4}$ m s^{-1}. Note that U is the speed in the frame of reference of the small square area considered here, which might itself be embedded within larger spatial structures (e.g., large eddies) moving at higher speeds in the ocean or any large waterbody.

To determine the diffusivity of small organisms, we use the Stokes-Einstein equation [110]:

$$D = \frac{RT}{N_A}\frac{1}{6\pi\eta a} \quad (3.25)$$

where $R = 8.314$ J K^{-1} mol^{-1} is the molar gas constant, $T = 293$ K is the temperature of the environment, $N_A = 6.0225 \times 10^{23}$ is Avogadro's number, $\eta = 10^{-3}$ m^{-1} kg s^{-1} is the viscosity of water and a is the radius of the organism considered. We apply this formula to microphytoplankton organisms of diameter 50 µm, keeping τ outside of the equation for now:

$$\begin{aligned} \Delta &= \sqrt{2D\tau} && (3.26)\\ &= \sqrt{\frac{RT}{N_A}\frac{\tau}{3\pi\eta a}} && (3.27)\\ &= \sqrt{\frac{8.314 \times 293}{6.0225 \times 10^{23}}\frac{1}{3\pi \times 10^{-3} \times 25 \times 10^{-6}}}\sqrt{\tau} && (3.28)\\ &= 1.3 \times 10^{-7}\sqrt{\tau}\ \mathrm{m}. && (3.29) \end{aligned}$$

To compute τ, we can consider a phytoplankton doubling rate of 1 d^{-1} [48], which means, with $p = 0.5$, that $\tau = 0.5$ d.

This leads to $U\tau/2 \approx 5.4 \times 10^3$ cm d^{-1} and $\Delta \approx 5 \times 10^{-5}$ cm. These two values are much higher than those used in Fig. 3.3 ($0.1 < U\tau/2 < 2.5$ and $\Delta = 10^{-7}$). A thorough discussion of the parameters is therefore necessary before extrapolating these results to real phytoplanktonic systems.

As the Brownian bug model is currently fairly theoretical in its 2D formulation, a logical next step would be to consider similar dynamics in a 3D model, which would render the comparison to real data easier. Using actual concentrations of phytoplanktonic organisms (e.g., diatoms),

between 10^3 and 10^6 cells L^{-1}, this would lead to 1 to 10^3 organisms if we kept $L = 1$ cm. We might therefore need to increase the size of the considered square, or apply the model to small bacteria only. With a closer match between field and simulated concentrations, the model could provide us with a better picture of the likely fine-scale spatial structure of phytoplanktonic populations.

Acknowledgements

We are grateful to Francesco Turci for comments and to Rajesh Singh for detailed feedback and code suggestions.

3.S Supplementary Information

3.S1 Derivation of $G(r,t)$

Diffusion and birth/death processes

In this section, we aim to find back from first principles Eq. 2 in Young et al. [345], i.e., Eq. 3.3 in our manuscript. We will first focus on the diffusion and birth/death processes, corresponding to the evolution equation for the pair density:

$$\frac{\partial G}{\partial t} = 2Dr^{1-d}\frac{\partial}{\partial r}\left(r^{d-1}\frac{\partial G}{\partial r}\right) + 2(\lambda-\mu)G + 2\lambda C\delta(\boldsymbol{\xi}). \tag{3.S1}$$

We first define an ensemble of k identical Brownian bugs in a d-dimensional space. The bug number p is located at $\boldsymbol{x}_p = [x_1, x_2, ...x_d]$. At time t, the space is defined by (a) the number of Brownian bugs k and (b) the vector of their locations $\boldsymbol{X}_k = [\boldsymbol{x}_1, \boldsymbol{x}_2, ...\boldsymbol{x}_k]$. This is also called the Fock space [47].

The probability distribution over the state space is given by the functions $\mathcal{P}_k(\boldsymbol{X}_k, t)$ such that

$$\mathcal{P}_k(\boldsymbol{X}_k, t)d\boldsymbol{X}_k = \Pr\{k \text{ bugs, with a bug in } d\boldsymbol{x}_1, \text{a bug in } d\boldsymbol{x}_2, \text{ etc.}\}. \tag{3.S2}$$

As bugs are indistinguishable, we can exchange $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$ (permutation symmetry):

$$\mathcal{P}_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p, \ldots, \boldsymbol{x}_q, \ldots, \boldsymbol{x}_k, t) = \mathcal{P}_k(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q, \ldots, \boldsymbol{x}_p, \ldots, \boldsymbol{x}_k, t). \tag{3.S3}$$

The normalization is

$$\mathcal{P}_0(t) + \int \mathcal{P}_1(\boldsymbol{X}_1, t)d\boldsymbol{x}_1 + \int\int \mathcal{P}_2(\boldsymbol{X}_2, t)d\boldsymbol{x}_1 d\boldsymbol{x}_2 + \ldots + \int_{\mathbb{R}^{2k}} \mathcal{P}_k(\boldsymbol{X}_k, t)d\boldsymbol{X}_k + \ldots = 1, \tag{3.S4}$$

because having k individuals at time t defines a partition of the sample space for $k = 0, 1, 2, ...$

We define $b_k(\boldsymbol{x}, t) = \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{X}_{k-1}, t)d\boldsymbol{X}_{k-1}$, i.e., $b_k(\boldsymbol{x}, t)d\boldsymbol{x}$ is the probability that there are k bugs and bug number 1 is in $d\boldsymbol{x}$.

The density of points is defined as

$$\rho(\boldsymbol{x}, t) = \sum_{k=1}^{\infty} k b_k(\boldsymbol{x}, t) = \sum_{k=1}^{\infty} k \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{X}_{k-1}, t)d\boldsymbol{X}_{k-1}. \tag{3.S5}$$

The pair correlation function is then

$$G(\boldsymbol{x}, \boldsymbol{y}, t) = \sum_{k=2}^{\infty} k(k-1) \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t)d\boldsymbol{X}_{k-2}. \tag{3.S6}$$

We define the two particle distribution functions $c_k(\boldsymbol{x}, \boldsymbol{y}, t) = \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t) d\boldsymbol{X}_{k-2}$. Note that $G(\boldsymbol{x}, \boldsymbol{y}, t) = \sum_{k=2}^{\infty} k(k-1) c_k(\boldsymbol{x}, \boldsymbol{y}, t)$.

Proposition

The time derivative of c_k is given by

$$\frac{\partial c_k}{\partial t}(\boldsymbol{x}, \boldsymbol{y}, t) = D\nabla_2^2 c_k \tag{3.S7a}$$
$$- k(\lambda + \mu) c_k \tag{3.S7b}$$
$$+ (k+1)\mu c_{k+1} \tag{3.S7c}$$
$$+ 2\frac{\lambda}{k}\delta(\boldsymbol{x} - \boldsymbol{y}) b_{k-1}(\boldsymbol{x}) + \lambda \frac{(k-2)(k+1)}{k} c_{k-1}. \tag{3.S7d}$$

Proof

We write the evolution of $\mathcal{P}_k(\boldsymbol{X}_k, t)$:

$$\frac{\partial \mathcal{P}_k}{\partial t} = D\nabla_k^2 \mathcal{P}_k \tag{3.S8a}$$
$$- k(\lambda + \mu)\mathcal{P}_k \tag{3.S8b}$$
$$+ (k+1)\mu \int \mathcal{P}_{k+1}(\boldsymbol{X}_k, \boldsymbol{y}, t) d\boldsymbol{y} \tag{3.S8c}$$
$$+ \frac{\lambda}{k} \sum_{p=1}^{k} \sum_{q=1, q\neq p}^{k} \delta(\boldsymbol{x}_p - \boldsymbol{x}_q) \mathcal{P}_{k-1}(\boldsymbol{X}_{k|p,t}) \tag{3.S8d}$$

where $\nabla_k^2 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \ldots + \frac{\partial^2}{\partial x_k^2} + \frac{\partial^2}{\partial y_k^2}$ in two dimensions and $\boldsymbol{X}_{k|p} = \boldsymbol{X}_k$ without $\boldsymbol{x}_p$.

Here, Eq. 3.S8a is the diffusion part of the process, Eq. 3.S8b corresponds to the rate at which realizations with k bugs lose a bug by mortality or gain a bug through birth, Eq. 3.S8c corresponds to the rate at which realizations with $k+1$ bugs lose a bug. Finally, a realization with k bugs can also be produced by a birth in a realization with $k-1$ bugs.

Combining the time derivative of $\mathcal{P}_k(\boldsymbol{X}_k, t)$ with the definition of c_k, we obtain

$$\frac{\partial c_k}{\partial t}(\boldsymbol{x}, \boldsymbol{y}, t) = D \int \nabla_k^2 \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t) d\boldsymbol{X}_{k-2} \tag{3.S9a}$$
$$- k(\lambda + \mu) \int \mathcal{P}_k(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, t) d\boldsymbol{X}_{k-2} \tag{3.S9b}$$
$$+ (k+1)\mu \int \mathcal{P}_{k+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{X}_{k-2}, \boldsymbol{z}, t) d\boldsymbol{X}_{k-2} d\boldsymbol{z} \tag{3.S9c}$$
$$+ \frac{\lambda}{k} \int \sum_{p=1}^{k} \sum_{q=1, q\neq p}^{k} \delta(\boldsymbol{x}_p - \boldsymbol{x}_q) \mathcal{P}_{k-1}(\boldsymbol{X}_{k|p,t}) d\boldsymbol{X}_{k-2}. \tag{3.S9d}$$

We will treat one term after the other.

Diffusion term (3.S9a)

$$D\int \nabla_k^2 \mathcal{P}_k(\boldsymbol{x},\boldsymbol{y},\boldsymbol{X}_{k-2},t)d\boldsymbol{X}_{k-2} = D\nabla_k^2 \int \mathcal{P}_k(\boldsymbol{x},\boldsymbol{y},\boldsymbol{X}_{k-2},t)d\boldsymbol{X}_{k-2} \tag{3.S10a}$$

$$= D\nabla_2^2 \int \mathcal{P}_k(\boldsymbol{x},\boldsymbol{y},\boldsymbol{X}_{k-2},t)d\boldsymbol{X}_{k-2} \tag{3.S10b}$$

$$= D\nabla_2^2 c_k \tag{3.S10c}$$

because we already integrate over $k-2$ coordinates (and thus Laplacians for these coordinates are zero).

Second term (3.S9b)

$$k(\lambda+\mu)\int \mathcal{P}_k(\boldsymbol{x},\boldsymbol{y},\boldsymbol{X}_{k-2},t)d\boldsymbol{X}_{k-2} = k(\lambda+\mu)c_k. \tag{3.S11}$$

Death term (3.S9c)

$$(k+1)\mu\int \mathcal{P}_{k+1}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{X}_{k-2},\boldsymbol{z},t)d\boldsymbol{X}_{k-2}d\boldsymbol{z} = (k+1)\mu\int \mathcal{P}_{k+1}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{X}_{k-1},t)d\boldsymbol{X}_{k-1} \tag{3.S12a}$$

$$= (k+1)\mu c_{k+1}. \tag{3.S12b}$$

Birth term (3.S9d)

In this section, we assume $\boldsymbol{x} = \boldsymbol{x}_1$ and $\boldsymbol{y} = \boldsymbol{x}_2$ (the reasoning is the same for different positions of $\boldsymbol{x}$ and $\boldsymbol{y}$ due to permutation symmetry).

We can decompose the double sum, starting with $p = 1$.

$$\sum_{q=2}^{k} \delta(\boldsymbol{x}_1-\boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_2,\dots,\boldsymbol{x}_k,t) = \delta(\boldsymbol{x}_1-\boldsymbol{x}_2)\mathcal{P}_{k-1}(\boldsymbol{x}_2,\dots,\boldsymbol{x}_k,t) \tag{3.S13a}$$

$$+ \delta(\boldsymbol{x}_1-\boldsymbol{x}_3)\mathcal{P}_{k-1}(\boldsymbol{x}_2,\dots,\boldsymbol{x}_k,t) \tag{3.S13b}$$

$$+ \dots \tag{3.S13c}$$

$$+ \delta(\boldsymbol{x}_1-\boldsymbol{x}_k)\mathcal{P}_{k-1}(\boldsymbol{x}_2,\dots,\boldsymbol{x}_k,t). \tag{3.S13d}$$

We integrate over the last $k-2$ coordinates.

$$\int \sum_{q=2}^{k} \delta(\boldsymbol{x}_1-\boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_2,\dots,\boldsymbol{x}_k,t)d\boldsymbol{x}_3\dots d\boldsymbol{x}_k \tag{3.S14a}$$

$$= \int \delta(\boldsymbol{x}_1-\boldsymbol{x}_2)\mathcal{P}_{k-1}(\boldsymbol{x}_2,\dots,\boldsymbol{x}_k,t)d\boldsymbol{x}_3\dots d\boldsymbol{x}_k \tag{3.S14b}$$

$$+ \int \delta(\boldsymbol{x}_1-\boldsymbol{x}_3)\mathcal{P}_{k-1}(\boldsymbol{x}_2,\dots,\boldsymbol{x}_k,t)d\boldsymbol{x}_3\dots d\boldsymbol{x}_k + \dots \tag{3.S14c}$$

$$+ \int \delta(\boldsymbol{x}_1-\boldsymbol{x}_k)\mathcal{P}_{k-1}(\boldsymbol{x}_2,\dots,\boldsymbol{x}_k,t)d\boldsymbol{x}_3\dots d\boldsymbol{x}_k \tag{3.S14d}$$

which leads to

$$\int \sum_{q=2}^{k} \delta(\boldsymbol{x}_1 - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{3.S15a}$$

$$= \delta(\boldsymbol{x}_1 - \boldsymbol{x}_2) \int \mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{3.S15b}$$

$$+ \int \mathcal{P}_{k-1}(\boldsymbol{x}_2, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_4 \ldots d\boldsymbol{x}_k + \ldots \tag{3.S15c}$$

$$+ \int \mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_1, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_1 \tag{3.S15d}$$

$$=\delta(\boldsymbol{x}_1 - \boldsymbol{x}_2) \int \mathcal{P}_{k-1}(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{3.S15e}$$

$$+ \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{X}_{k-3}, t)d\boldsymbol{X}_{k-3} + \ldots \tag{3.S15f}$$

$$+ \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{X}_{k-3}, t)d\boldsymbol{X}_{k-3} \tag{3.S15g}$$

$$=\delta(\boldsymbol{x}_1 - \boldsymbol{x}_2)b_{k-1}(\boldsymbol{x}_2) + (k-2)c_{k-1} \tag{3.S15h}$$

By symmetry, if $p = 2$, we obtain $\delta(\boldsymbol{x}_1 - \boldsymbol{x}_2)b_{k-1}(\boldsymbol{x}_1) + (k-2)c_{k-1}$.

Now, we need to use $p \geq 3$.

$$\int \sum_{q=1, q\neq p}^{k} \delta(\boldsymbol{x}_p - \boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{3.S16a}$$

$$= \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_1)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{3.S16b}$$

$$+ \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_2)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k + \tag{3.S16c}$$

$$+ \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_3)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k + \ldots \tag{3.S16d}$$

$$+ \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_k)\mathcal{P}_{k-1}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_k \tag{3.S16e}$$

$$= \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_1)d\boldsymbol{x}_p \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_{p-1}d\boldsymbol{x}_{p+1} \ldots d\boldsymbol{x}_k \tag{3.S16f}$$

$$+ \int \delta(\boldsymbol{x}_p - \boldsymbol{x}_2)d\boldsymbol{x}_p \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_{p-1}d\boldsymbol{x}_{p+1} \ldots d\boldsymbol{x}_k \tag{3.S16g}$$

$$+ \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_p, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_k, t)d\boldsymbol{x}_4 \ldots d\boldsymbol{x}_k + \ldots \tag{3.S16h}$$

$$+ \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p-1}, \boldsymbol{x}_{p+1}, \ldots, \boldsymbol{x}_p, t)d\boldsymbol{x}_3 \ldots d\boldsymbol{x}_{k-1} \tag{3.S16i}$$

$$=2 \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{X}_{k-3}, t)d\boldsymbol{X}_{k-3} \tag{3.S16j}$$

$$+ (k-3) \int \mathcal{P}_{k-1}(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{X}_{k-3}, t)d\boldsymbol{X}_{k-3} \tag{3.S16k}$$

$$=2c_{k-1} + (k-3)c_{k-1} \tag{3.S16l}$$

$$=(k-1)c_{k-1}. \tag{3.S16m}$$

Thus, $\sum_{p=3}^{k}\sum_{q\neq p}\int \delta(\boldsymbol{x}_p-\boldsymbol{x}_q)\mathcal{P}_{k-1}(\boldsymbol{X}_{k|p})d\boldsymbol{X}_{k-2} = (k-2)(k-1)c_{k-1}$.

Finally, the birth term is

$$\frac{\lambda}{k}\Big(2\delta(\boldsymbol{x}-\boldsymbol{y})b_{k-1}(\boldsymbol{x})+2(k-2)c_{k-1}+(k-2)(k-1)c_{k-1}\Big) = 2\frac{\lambda}{k}\delta(\boldsymbol{x}-\boldsymbol{y})b_{k-1}(\boldsymbol{x})+\lambda\frac{(k-2)(k+1)}{k}c_{k-1}. \tag{3.S17}$$

Combining all terms, we obtain the expected result:

$$\begin{aligned}
\frac{\partial c_k}{\partial t}(\boldsymbol{x},\boldsymbol{y},t) &= D\nabla_2^2 c_k && \text{(3.S18a)}\\
&- k(\lambda+\mu)c_k && \text{(3.S18b)}\\
&+ (k+1)\mu c_{k+1} && \text{(3.S18c)}\\
&+ 2\frac{\lambda}{k}\delta(\boldsymbol{x}-\boldsymbol{y})b_{k-1}(\boldsymbol{x}) + \lambda\frac{(k-2)(k+1)}{k}c_{k-1}. && \text{(3.S18d)}
\end{aligned}$$

Proposition

In Cartesian coordinates, the pair density admits the following evolution equation:

$$\frac{\partial G}{\partial t}(\boldsymbol{x},\boldsymbol{y},t) = D\nabla_2^2 G + 2(\lambda-\mu)G + 2\lambda\delta(\boldsymbol{x}-\boldsymbol{y})\rho(\boldsymbol{x}). \tag{3.S19}$$

Proof

Using the definition of $G(\boldsymbol{x},\boldsymbol{y},t)$ and Eq. 3.S7a-3.S7d:

$$\begin{aligned}
\frac{\partial G}{\partial t}(\boldsymbol{x},\boldsymbol{y},t) &= \sum_{k=2}^{\infty} k(k-1)D\nabla_2^2 c_k && \equiv T1 && \text{(3.S20a)}\\
&- \sum_{k=2}^{\infty} k(k-1)k(\lambda+\mu)c_k && \equiv T2 && \text{(3.S20b)}\\
&+ \sum_{k=2}^{\infty} k(k-1)(k+1)\mu c_{k+1} && \equiv T3 && \text{(3.S20c)}\\
&+ \sum_{k=2}^{\infty} k(k-1)2\frac{\lambda}{k}\delta(\boldsymbol{x}-\boldsymbol{y})b_{k-1}(\boldsymbol{x}) && \equiv T4 && \text{(3.S20d)}\\
&+ \sum_{k=2}^{\infty} \lambda k(k-1)\frac{(k-2)(k+1)}{k}c_{k-1} && \equiv T5 && \text{(3.S20e)}
\end{aligned}$$

T1 $T1 = D\nabla_2^2\sum_{k=2}^{\infty}k(k-1)c_k = D\nabla_2^2 G_k$.

T3

$$T3 = \sum_{k=2}^{\infty} k(k-1)(k+1)\mu c_{k+1} \tag{3.S21a}$$

$$= \sum_{k'=3}^{\infty} (k'-1)(k'-2)k'\mu c_{k'} \qquad \text{with } k' = k+1 \tag{3.S21b}$$

$$= \sum_{k'=2}^{\infty} (k'-1)(k'-2)k'\mu c_{k'} \qquad \text{if } k' = 2, k'-2 = 0 \tag{3.S21c}$$

T4

$$T4 = 2\lambda\delta(\boldsymbol{x}-\boldsymbol{y}) \sum_{k=2}^{\infty} (k-1) b_{k-1}(\boldsymbol{x}) \tag{3.S22a}$$

$$= 2\lambda\delta(\boldsymbol{x}-\boldsymbol{y}) \sum_{k''=1}^{\infty} k'' b_{k''}(\boldsymbol{x}) \qquad \text{with } k'' = k-1 \tag{3.S22b}$$

$$= 2\lambda\delta(\boldsymbol{x}-\boldsymbol{y})\rho(\boldsymbol{x}) \tag{3.S22c}$$

T5

$$T5 = \lambda \sum_{k=2}^{\infty} k(k-1)\frac{(k-2)(k+1)}{k} c_{k-1} \tag{3.S23a}$$

$$= \lambda \sum_{k''=1}^{\infty} k''(k''-1)(k''+2)c_{k''} \qquad \text{with } k'' = k-1 \tag{3.S23b}$$

$$= \lambda \sum_{k''=2}^{\infty} k''(k''-1)(k''+2)c_{k''} \qquad \text{if } k'' = 1, k''-1 = 0 \tag{3.S23c}$$

T2+T3+T5

$$T2 + T3 + T5 = \sum_{k=2}^{\infty} -k(k-1)k(\lambda+\mu)c_k + (k-1)(k-2)k\mu c_k + k(k-1)(k+2)\lambda c_k \tag{3.S24a}$$

$$= \sum_{k=2}^{\infty} k(k-1)c_k(-k(\lambda+\mu) + (k-2)\mu + (k+2)\lambda) \tag{3.S24b}$$

$$= 2(\lambda-\mu) \sum_{k=2}^{\infty} k(k-1)c_k \tag{3.S24c}$$

$$= 2(\lambda-\mu)G_k \tag{3.S24d}$$

T1+T2+T3+T4+T5 Combining all terms, we have

$$\frac{\partial G}{\partial t}(\boldsymbol{x},\boldsymbol{y},t) = D\nabla_2^2 G \tag{3.S25a}$$

$$+ 2(\lambda-\mu)G_k \tag{3.S25b}$$

$$+ 2\lambda\delta(\boldsymbol{x}-\boldsymbol{y})\rho(\boldsymbol{x}). \tag{3.S25c}$$

We can now write the diffusion term in Eq. 3.S25a, $D\nabla_2^2 G$ where $\nabla_2^2 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial y_2^2}$, with a more classical Laplacian operator. Hereafter, we use the notations x and y for coordinates, as opposed to positions $\boldsymbol{x}$ and $\boldsymbol{y}$. We define two points at positions $\mathbf{p}_1 = (x_1, y_1)^T$ and $\mathbf{p}_2 = (x_2, y_2)^T$, and the vectors $\boldsymbol{\xi} = \mathbf{p}_1 - \mathbf{p}_2$ and $\mathbf{X} = (\mathbf{p}_1 + \mathbf{p}_2)/2$. Along x, the correspondence between coordinates is as follows:

$$\begin{cases} \xi_x &= x_1 - x_2 \\ X_x &= \frac{x_1+x_2}{2} \end{cases} \Leftrightarrow \begin{cases} x_1 &= X_x + \frac{\xi_x}{2} \\ x_2 &= X_x - \frac{\xi_x}{2} \end{cases}, \tag{3.S26}$$

which can be derived as

$$\frac{\partial}{\partial x_1} = \frac{\partial}{\partial \xi_x} + \frac{1}{2}\frac{\partial}{\partial X_x} \Rightarrow \frac{\partial^2}{\partial x_1^2} = \frac{\partial^2}{\partial \xi_x^2} + \frac{1}{4}\frac{\partial^2}{\partial X_x^2} + \frac{\partial^2}{\partial \xi_x \partial X_x} \tag{3.S27}$$

and

$$\frac{\partial}{\partial x_2} = -\frac{\partial}{\partial \xi_x} + \frac{1}{2}\frac{\partial}{\partial X_x} \Rightarrow \frac{\partial^2}{\partial x_2^2} = \frac{\partial^2}{\partial \xi_x^2} + \frac{1}{4}\frac{\partial^2}{\partial X_x^2} - \frac{\partial^2}{\partial \xi_x \partial X_x}. \tag{3.S28}$$

Therefore

$$\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} = 2\frac{\partial^2}{\partial \xi_x^2} + \frac{1}{2}\frac{\partial^2}{\partial X_x^2}. \tag{3.S29}$$

With the same arguments,

$$\frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial y_2^2} = 2\frac{\partial^2}{\partial \xi_y^2} + \frac{1}{2}\frac{\partial^2}{\partial X_y^2} \tag{3.S30}$$

and, finally,

$$\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial y_2^2} = 2\left(\frac{\partial^2}{\partial \xi_x^2} + \frac{\partial^2}{\partial \xi_y^2}\right) + \frac{1}{2}\left(\frac{\partial^2}{\partial X_x^2} + \frac{\partial^2}{\partial X_y^2}\right). \tag{3.S31}$$

As the environment is homogeneous, $\frac{\partial G}{\partial X_x} = \frac{\partial^2 G}{\partial X_x^2} = \frac{\partial G}{\partial X_y} = \frac{\partial^2 G}{\partial X_y^2} = 0$.

We can thus write $D\nabla_2^2 G(\mathbf{p}_1, \mathbf{p}_2, t) = 2D\nabla^2 G(\boldsymbol{\xi}, t)$. The Laplacian formulation we have used above corresponds to Cartesian coordinates. Its equivalent in polar coordinates is $\nabla^2 G = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 G}{\partial \theta^2}$. As the process is isotropic, $\frac{\partial^2 G}{\partial \theta^2} = 0$. Finally, we can write

$$2D\nabla^2 G(\boldsymbol{\xi}, t) = \frac{2D}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right) \tag{3.S32}$$

with $r = |\boldsymbol{\xi}|$.

Advection process

The stretching of line elements considered by Kraichnan [200] leads to the following term for the advection process:

$$\frac{\partial G}{\partial t} = \gamma r^{1-d}\frac{\partial}{\partial r}\left(r^{d+1}\frac{\partial G}{\partial r}\right). \tag{3.S33}$$

We show below that this equation corresponds indeed to the convection of passive scalars in the Batchelor regime envisioned by Kraichnan [200]. Let r(t) be the distance between two points as a function of time, which follows a geometric Brownian motion as we are in the Batchelor regime, and $q(t) = \log(r(t)/r(0))$. Kraichnan defines $Q(q)$ as the probability distribution of q. We have $Q = r^d G$, leading to

$$\frac{\partial G}{\partial r} = \frac{\partial}{\partial r}\left(Q/r^d\right) = \frac{1}{r^d}\frac{\partial Q}{\partial r} + Q(-d)r^{-d-1}. \tag{3.S34}$$

Now

$$r^d\frac{\partial G}{\partial t} = \frac{\partial Q}{\partial t} = \gamma r\frac{\partial}{\partial r}\left(r\frac{\partial Q}{\partial r} - dQ\right). \tag{3.S35}$$

We have $\frac{dq}{dr} = \frac{1}{r}$ so that using $\frac{\partial Q}{\partial r} = \frac{\partial Q}{\partial q}\frac{dq}{dr}$ we obtain

$$\frac{\partial Q}{\partial t} = \gamma\frac{\partial}{\partial q}\left(\frac{\partial Q}{\partial q} - dQ\right) = \gamma\frac{\partial^2 Q}{\partial q^2} - \gamma d\frac{\partial Q}{\partial q} \tag{3.S36}$$

which is the Fokker-Planck equation with diffusion coefficient γ (noted c in [200], and also called the stretching parameter) and drift γd (noted $< a >$ in [200]). Eq. 3.S36 leads to Kraichnan's solution for pair separation (Eq. 2.31 in [200]), with drift and diffusion coefficients linked by Eq. 2.27 and 2.33 in [200].

Combining Eq. 3.S33 with Eq. 3.S1, we obtain Eq. 3.3.

3.S2 Stretching parameter γ

γ is computed with simulations, using the formula $r(t) \propto \exp(\gamma dt) \rightarrow \frac{1}{2}\ln(r(t)) = \gamma t$ if $d = 2$, with r the separation between pairs of particles. γ is estimated as the slope of

$$\frac{1}{2}\langle\ln(r(t))\rangle = f(t)$$

with $\langle \ln(r(t)) \rangle$ the average obtained from 800 pairs of particles.

$$\forall t, \langle \ln(r(t)) \rangle = \frac{1}{800} \sum_{p=1}^{800} \ln(r(\boldsymbol{x}_{1p}(t) - \boldsymbol{x}_{2p}(t)))$$

where $r(\boldsymbol{x}_{1p}(t) - \boldsymbol{x}_{2p}(t))$ is the distance between a particle $1p$ at position $\boldsymbol{x}_{1p}$ and its counterpart $2p$, initialized with $r(0) = 10^{-7}$ $\forall p$ (see Fig. 3.S1 for γ estimates).

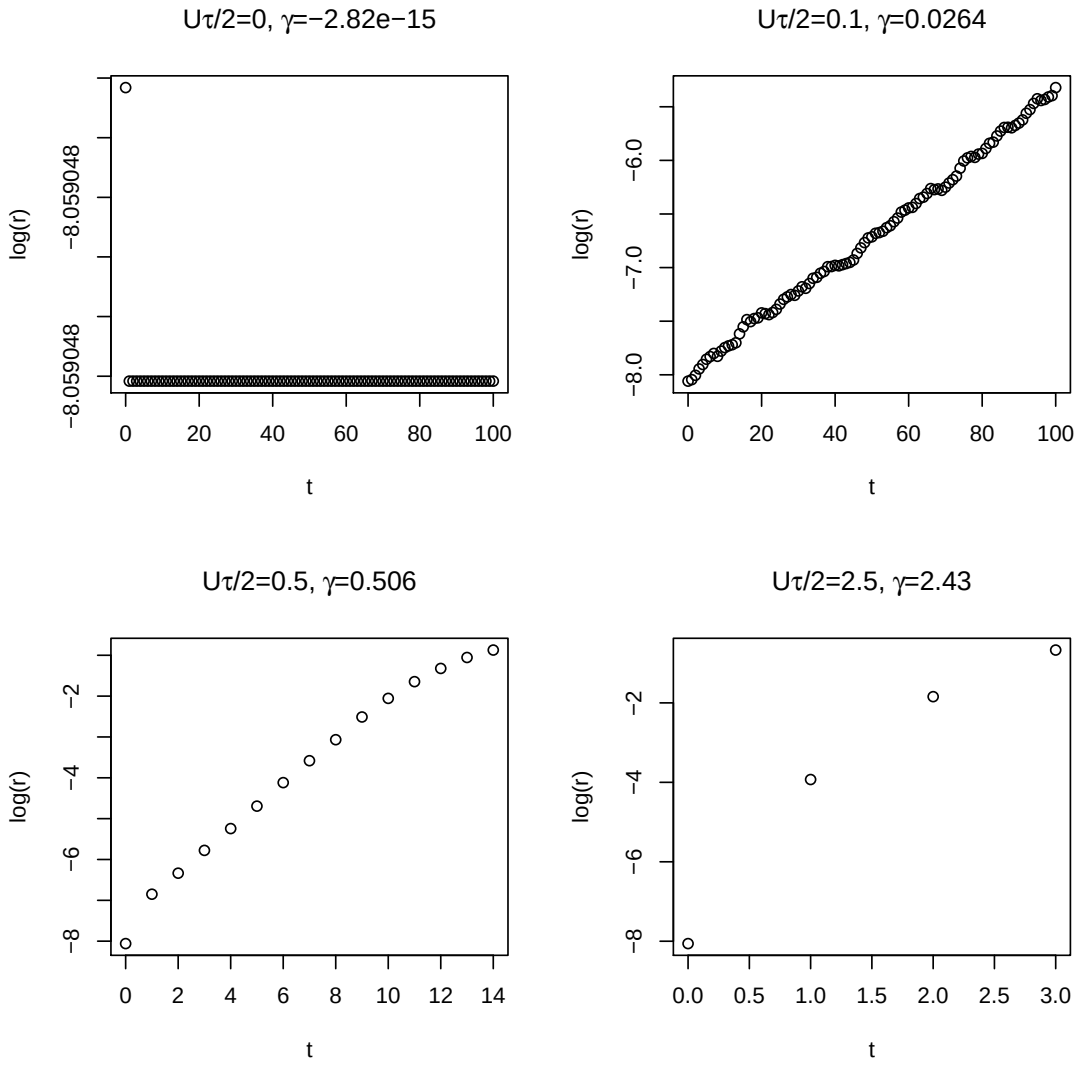

Figure 3.S1: Estimates of γ for different $U\tau/2$.

Chapter 4

Local intraspecific aggregation in phytoplankton model communities: spatial scales of occurrence and implications for coexistence

Coralie Picoche[1,*], William R. Young[2] & Frédéric Barraquand[1]
[1]Institute of Mathematics of Bordeaux, University of Bordeaux & CNRS, Talence, France
[2]Scripps Institution of Oceanography, La Jolla, California, USA
*corresponding author: coralie.picoche@u-bordeaux.fr

Abstract

The coexistence of multiple phytoplankton species despite their reliance on similar resources is often explained with mean-field models assuming mixed populations. In reality, observations of phytoplankton indicate spatial aggregation at all scales, including at the scale of a few individuals. Local spatial aggregation can hinder competitive exclusion since individuals then interact mostly with other individuals of their own species, rather than competitors from different species. To evaluate how microscale spatial aggregation might explain phytoplankton diversity maintenance, an individual-based, multispecies representation of cells in a hydrodynamic environment is required. We formulate a three-dimensional and multispecies individual-based model of phytoplankton population dynamics at the Kolmogorov scale. The model is studied through both simulations and the derivation of spatial moment equations, in connection with point process theory. The spatial moment equations show a good match between theory and simulations. We parameterized the model based on phytoplankters' ecological and physical characteristics, for both large and small phytoplankton. Defining a zone of potential interactions as the overlap between nutrient depletion volumes, we show that local species composition—within the range of possible interactions—depends on the size class of phytoplankton. In large phytoplankton, individuals are surrounded by cells from other species, while in small phytoplankton, individuals remain in mostly monospecific clusters. Spatial structure therefore favours intra- over inter-specific interactions for small phytoplankton, which likely contributes to coexistence mechanisms. Other factors behind diversity maintenance must be examined for large phytoplankton.

Keywords: aggregation; coexistence; individual-based model; phytoplankton; spatial moment equations; spatial point process

4.1 Introduction

Phytoplankton communities are among the most important photosynthetic groups on Earth, being at the bottom of the marine food chain, and responsible for approximately half the global primary production [121]. Their contribution to ecosystem functions is only matched by their contribution to biodiversity. Indeed, phytoplankton communities are characterized by a surprisingly high number of species. For example, a single sample as small as a few mL can contain up to seventy species [284, 331]. This observation is usually called the "paradox of the plankton" [168], which refers to the conflict between the observed diversity of species competing for similar resources in a seemingly homogeneous environment, and models predicting that only a few species will persist by outcompeting the others [219, 165, 299]. Phytoplankton models for coexistence are now almost as diverse as their model organisms [282], but they often describe only a handful of species, which does not correspond to the diversity observed in the field. When modeling rich communities ($>$ 10 species), classical answers to the plankton paradox involving temporal fluctuations (e.g., [211, 76]) are not sufficient to maintain a realistic diversity. For instance, we found that a phytoplankton community dynamics model with environmental fluctuations and storage effect still requires extra niche differentiation for coexistence, which manifests in stronger intraspecific than interspecific interactions [270]. However, it is not clear that we should resort to hidden niches to explain phytoplankton coexistence, as most models also make hidden simplifying assumptions that could be relaxed. One that we relax here is mean-field dynamics at the microscale. Indeed, field observations have revealed phytoplankton patchiness for more than a century [24, 314], from the macro- to the micro-scale [205, 101, 124].

Phytoplankton patchiness can at least be partly explained by the hydrodynamics of their environment: the size of these organisms is mostly below the size of the smallest eddy (i.e., the Kolmogorov scale). In a typical aquatic environment such as the ocean, phytoplankton individuals are embedded in viscous micro-structures [269] while phytoplankton populations are displaced by a turbulent flow at slighly larger scales [225, 278]. Phytoplankton organisms therefore live in an environment where fluid viscosity dominates at the scale of an individual but turbulent dispersion dominates on length scales characteristic of a small population of those individuals [115, 278].

This leads us to consider demography in the context of this environmental variation created by hydrodynamic processes. Individual-based models provide a convenient depiction of population dynamics and movement at the microscale [150]. In this framework, population growth is a result of individual births and deaths. Aggregation of individuals can emerge from local reproduction coupled with limited dispersal, which can happen in a fluid where turbulence and diffusion are not strong enough to disperse kin aggregates [345]. The resulting local aggregation can then affect the community dynamics at larger spatial scales, even when all competitors are equivalent (i.e., with equal interaction strengths irrespective of species identity). Indeed, the combination of local dispersal after reproduction and local interactions leads to stronger

intraspecific interactions than interspecific interactions at the population level [97]. This mechanism stabilizes the community as a high intra-to-interspecific interaction strength ratio makes a species control its abundance more than it controls the abundance of other species, which is associated with coexistence in theoretical models [206, 29] and often observed in the field at the population level [4, 271]. Therefore, the microscale spatial distribution of individuals likely affects the interaction structure within a community [140], and may sustain diversity.

Existing models of phytoplankton populations near the Kolmogorov scale — between 1 mm and 1 cm in an oceanic environment [33] — focus on a single species and the clustering of its individuals [345, 47, 55, 59]. They share similarities to dynamic point process models [201, 52, 275] developed initially with larger organisms in mind. When phytoplankton individual-based models consider multiple types of organisms, they focus for now on how organisms with opposite characteristics (e.g., increase versus decrease in density with turbulence in [53, 19]) segregate spatially, or on coexistence for species that have contrasting trait values (e.g., size in [41]). This is useful as an explanation of how species with marked differences might coexist. The difficulty of the coexistence problem, however, is that we have to explain how closely related species or genera (e.g., within diatoms), many of whom have similar size, buoyancy, chemical composition, etc., manage to coexist within a single trophic level. This requires modelling *similar* species in a spatially realistic environment and objectively quantifying whether they aggregate or segregate in space.

To do so, we build a multispecies version of the Brownian Bug Model (BBM) of Young *et al.* [345], an individual-based model which includes an advection process mimicking a turbulent fluid flow, passive diffusion of organisms, as well as stochastic birth and death processes. The initial version of this model [345] coupled limited dispersal and local reproduction with ocean-like microscale hydrodynamics, and showed spatial clusters of individuals of the same species. The original BBM was limited to a single species and was illustrated with two-dimensional simulations. The model was not strongly quantitative [272] in the sense that parameters were not informed by current knowledge on phytoplankton biology (numbers of cells per liter, diffusion characteristics, etc.). As phytoplankton organisms live in a three-dimensional environment, informing the model with more realistic parameters requires us to shift to three dimensions. We also extend the model to multiple species, and consider two size classes for our phytoplankton communities, which are either made of nanophytoplankton (3 µm diameter, $\approx 10^6$ cells L^{-1}) or microphytoplankton (50 µm, $\approx 10^4$ cells L^{-1}). We populate each community with 3 to 10 different species.

The Brownian Bug model (in its original single-species form as in the multispecies version considered here) is related to spatial branching processes. Without advection, it combines a continuous-time, discrete-state model for population growth and a continuous-time, continuous-space Brownian motion for particle diffusion [47]. It is further complexified by a turbulent flow in Young *et al.* [345], Picoche *et al.* [272] as well as here. In spite of this complexity, it remains possible to derive the dynamics of pair density functions, which quantify the degree of intra- and

interspecific clustering of organisms, via correlations between positions of organisms (see next section). Thus we can understand emergent spatial structure in analytic detail and compare these predictions to the results from three-dimensional simulations. Furthermore, because we do not consider direct interactions between organisms, the multispecies spatial point process that represents the stable state of the BBM is a random superposition of spatial point processes for each species [169]. This enables us to derive, in addition to pair correlation functions, analytical formulas for the species composition in the neighbourhood of an individual, which are more readily ecologically interpreted than pair density or correlation functions.

4.2 Model and spatial statistics

Brownian Bug Model

The Brownian Bug Model (BBM) describes the dynamics of individuals in a turbulent and viscous environment, including demographic processes. The model is continuous in space and time. Here we extend the mostly two-dimensional, monospecific version in Young *et al.* [345], to three dimensions and S species.

Each individual is characterized by its species identity i and its position $\mathbf{x}^T = (x, y, z)$. The population dynamics are modeled by a linear birth-death process with birth rate λ_i and death rate μ_i. Each individual independently follows a Brownian motion with diffusivity D_i, and is advected by a common stochastic and chaotic flow modelling turbulence. The model applies in the Batchelor regime, which means that the separation $s(t)$ between two individuals k and l grows exponentially with time with stretching parameter γ, i.e. $s(t) = \ln(|\boldsymbol{x}_k - \boldsymbol{x}_l|(t)) \propto 3\gamma t$ [200, 345].

Within a given community (the set of all individuals of the S species), all species share the same parameters: λ_i, μ_i and D_i values can change between communities, as we later consider small and large phytoplankton, but are set to common values within a community. On the contrary, γ describes the environment and is not community-specific, i.e., all individuals are displaced by the same turbulent stirring. For numerical simulations, time needs to be discretized (this is required for diffusion and advection modelling). The approximated model advances through time in small steps of duration of τ. During each interval, events unroll as follows:

1. Demography: each individual can either reproduce with probability $p_i = \lambda_i\tau$ (forming a new individual of the same species i at the same position $\mathbf{x}$ as the parent), die with probability $q_i = \mu_i\tau$, or remain unchanged with probability $1 - p_i - q_i$.

2. Diffusion: each individual moves to a new position $\mathbf{x}(t') = \mathbf{x}(t) + \delta\mathbf{x}(t)$, with $t < t' < t+\tau$. The random displacement $\delta\mathbf{x}(t)$ is drawn from a Gaussian distribution $\mathcal{N}(0, \Delta_i^2)$ with $D_i = \Delta_i^2/2\tau$ the diffusivity. This diffusive step separates the initially coincident pairs produced by reproduction in step 1 above.

3. Turbulence: each individual is displaced by a turbulent flow, modeled with the Pierrehumbert map [274], adapted to three dimensions following Ngan & Vanneste [253]. Thus given the position at time t' the updated position at time $t+\tau$ is

$$\begin{aligned} x(t+\tau) &= x(t') + \tfrac{U\tau}{3}\cos\left(ky(t') + \phi(t)\right) \\ y(t+\tau) &= y(t') + \tfrac{U\tau}{3}\cos\left(kz(t') + \theta(t)\right) \\ z(t+\tau) &= z(t') + \tfrac{U\tau}{3}\cos\left(kx(t+\tau) + \psi(t)\right). \end{aligned} \tag{4.1}$$

Above, U is the velocity of the flow, $k = 2\pi/L_s$ is the wavenumber for the flow at the length scale L_s (see below) and $\phi(t)$, $\theta(t)$, $\psi(t)$ are random phases drawn from a uniform distribution between 0 and 2π; these phases remain constant during the interval between t and $t+\tau$. The shift from continuous to discrete-time turbulence modelling is described in Section 4.S1 in the Supplementary Information. The velocity U is related to γ. As the separation between two points grows exponentially with parameter 3γ due to turbulence, the exponent γ can be estimated as the slope of $1/3\,\langle\ln(s(t))\rangle = f(t)$ in the absence of diffusion and demography [345, 272].

Individuals are distributed in a cube of side length L, with periodic boundary conditions. The cube dimensions are determined to balance computing costs and realistic concentrations of individuals; they represent the accumulation of a few volumes of scale L_s.

Characterization of the spatial distribution

Let W be the observation window (in our case, the whole cube, which we never subsample hereafter). The state of the system at time t can be described as a collection of S populations, where the population of species i is made of n_i individuals randomly distributed in W, with positions $\boldsymbol{X}_i(t) = [\boldsymbol{x}_{1,i}(t), \boldsymbol{x}_{2,i}(t), ...\boldsymbol{x}_{n_i,i}(t)]$. $\boldsymbol{X}(t) = [\boldsymbol{X}_1(t), \ldots, \boldsymbol{X}_S(t)]$ arises from a stochastic and spatial individual-based model changing through time, but can also be analyzed as a spatial point process at time t. We note that the point distributions remain the same for all spatial translations $\boldsymbol{\xi}$ (i.e., the point process described by the set $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ...\boldsymbol{x}_k]$ is the same as $\boldsymbol{X}_{\boldsymbol{\xi}} = [\boldsymbol{x}_1 + \boldsymbol{\xi}, \boldsymbol{x}_2 + \boldsymbol{\xi}, ...\boldsymbol{x}_k + \boldsymbol{\xi}]$): the process is stationary.

A useful method to characterize a spatial point process is the use of spatial moments (illustrated in Section 4.S2 of the SI for simple spatial point processes). These can be theoretically derived and used to check simulations. The spatial moments of a process are, however, merely statistical indicators which then need to be related to more easily ecologically interpretable quantities. This is the role of the dominance index, which we present below.

Spatial moments

The first-order moment is the intensity of the process, or mean concentration of individuals, whose empirical estimate is $C_i = \frac{\widehat{N_i(W)}}{V(W)}$, where $\widehat{N_i(W)}$ is the empirical number of individuals of species i in the cube W and $V(W) = L^3$ is the volume of the cube; it does not give any information regarding the spatial distribution of individuals, and their spatial correlations.

The second-order product density, or pair density $G(r,t)$, is the expected density of pairs of points separated by a distance r [201]. A similar characteristic can be used for marked spatial point process. In our case, the marks are the species' identities, and we can define $G_{ij}(r,t)$, so that $G_{ij}(r,t)d\mathbf{x}_A d\mathbf{x}_B$ is the probability of finding an individual of species i in volume $d\mathbf{x}_A$ and an individual of species j in volume $d\mathbf{x}_B$, with the distance between the centers of $d\mathbf{x}_A$ and $d\mathbf{x}_B$ equal to r (pages 219 and 325 in [169]). We define $\boldsymbol{\xi}$ as the vector connecting the center of $d\mathbf{x}_A$ to the center of $d\mathbf{x}_B$, while $r = |\boldsymbol{\xi}|$ is the radial distance. We show in Picoche *et al.* [272] that the intraspecific pair density $G_{ii}(r,t)$, in three dimensions, is a solution of

$$\frac{\partial G_{ii}}{\partial t}(r,t) = \frac{2D_i}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial G_{ii}}{\partial r}\right) + \frac{\gamma}{r^2}\frac{\partial}{\partial r}\left(r^4\frac{\partial G_{ii}}{\partial r}\right) + 2(\lambda_i - \mu_i)G_{ii} + 2\lambda_i C_i \delta(\boldsymbol{\xi}). \tag{4.2}$$

The pair correlation function $g_{ij}(r,t)$, or pcf, can be derived from the pair density and is defined as

$$g_{ij}(r,t) = \frac{G_{ij}(r,t)}{C_i C_j}. \tag{4.3}$$

The pcf is equal to one when the spatial distribution of species i individuals is random relative to species j individuals. To compute the intraspecific pcf $g_{ii}(r,t)$ at steady state, considering a population at equilibrium, we integrate Eq. 4.2 (see Appendices, Eqs. 4.19-4.30) with $\lambda_i = \mu_i$ and obtain

$$g_{ii}(r) = 1 + \frac{\lambda_i}{4\pi D_i C_i \ell_{\mathrm{B},i}}\left(\frac{\ell_{\mathrm{B},i}}{r} + \arctan\left(\frac{r}{\ell_{\mathrm{B},i}}\right) - \frac{\pi}{2}\right), \tag{4.4}$$

where $\ell_{\mathrm{B},i} = \sqrt{2D_i/\gamma}$ approximates the Batchelor scale for species i.

The system converges rapidly to the solution in Eq. 4.4 in the presence of advection. However, when there is no turbulent advection, convergence is much slower, to the point that an equilibrium assumption requires unrealistically long timeframes (see Section 4.S3 in the SI). We therefore need a time-dependent formula for the pcf in the absence of advection, which can be obtained in the case where $\gamma = 0$ using a Green's function (see derivation in the Appendices, Eqs. 4.31-4.37),

$$g_{ii}(r,t) = 1 + \frac{\lambda_i}{4\pi r D_i C_i}\left\{1 - \mathrm{erf}\left(\frac{r}{\sqrt{8D_i t}}\right)\right\}. \tag{4.5}$$

The above equations match when $\gamma \to 0$ and $t \to +\infty$.

As populations of different species do not directly interact, each population is an independent realization of a point process, which means that the distribution of all individuals within the community at time t is a random superposition of stationary point processes and thus

$g_{ij}(r,t) = 1$ if $i \neq j$ ([169], p. 326, eq. 5.3.13).

Related to the pair correlation function is Ripley's K-function $K(r)$. Using its marked version, $C_j K_{ij}(r)$ is the average number of points of species j surrounding an individual of species i within a sphere of radius r [169], i.e.,

$$\forall r \geq 0,\ K_{ij}(r) = \frac{1}{C_j}\mathbb{E}_i\left(N_j\left(b(o,r)\backslash\{o\}\right)\right), \tag{4.6}$$

where $\mathbb{E}_i$ is the expectation with respect to individuals of species i and $N_j\left(b(o,r)\backslash\{o\}\right)$ is the number of individuals of species j in the sphere of radius r centered on o, not counting o itself. $K_{ij}(r)$ is related to $g_{ij}(r)$ as

$$g_{ij}(r) = \frac{K'_{ij}(r)}{4\pi r^2}. \tag{4.7}$$

Combining Eq. 4.7 and, when $U > 0$, Eq. 4.4, we can show that (see Appendices, Eqs. 4.38-4.44)

$$K_{ii}(r) = \frac{4}{3}\pi r^3 + \frac{\lambda_i r^3}{3D_i C_i \ell_{\mathrm{B},i}}\left(\frac{\ell_{\mathrm{B},i}}{r} + \frac{\ell_{\mathrm{B},i}^3 \log\left(\frac{r^2}{\ell_{\mathrm{B},i}^2}+1\right)}{2r^3} + \arctan\left(\frac{r}{\ell_{\mathrm{B},i}}\right) - \frac{\pi}{2}\right). \tag{4.8}$$

When $U = 0$, we need a time-dependent solution corresponding to our simulation duration, i.e. (see Appendices, Eq. 4.46-4.51)

$$K_{ii}(r,t) = \frac{4}{3}\pi r^3 + \frac{\lambda_i r^2}{C_i D_i}\left(\frac{1}{2} - \frac{1}{2}\operatorname{erf}\left(\frac{r}{\sqrt{8D_i t}}\right)\left(1 - \frac{4D_i}{r^2}t\right) - \frac{\sqrt{2D_i t}}{\sqrt{\pi}r}e^{-\frac{r^2}{8D_i t}}\right). \tag{4.9}$$

For random superposition of stationary point processes, $K_{ij}(r,t) = \frac{4}{3}\pi r^3$ if $i \neq j$ ([169], p. 324, eq. 5.3.5).

Dominance index

The dominance index (defined in Table S1 in the Supporting Information of [333]) is the ratio between the number of conspecifics and the number of individuals of all species surrounding a given individual.

Let $M_{ij}(r)$ be the average number of individuals of species j within a circle of radius r around an individual of species i, which can also be written with Ripley's K-function as $M_{ij}(r) = C_j K_{ij}(r)$. $M_{ii}(r)$ corresponds to the conspecific neighbourhood and $M_{io}(r) = \sum_{j=1, j\neq i}^{S} M_{ij}(r)$ corresponds to individuals of all other species. We can then define $\mathcal{D}_i$ as

$$\begin{aligned}\mathcal{D}_i(r) &= \frac{M_{ii}(r)}{M_{ii}(r)+M_{io}(r)} \\ &= \frac{C_i K_{ii}(r)}{\sum_{j=1}^{S} C_j K_{ij}(r)}.\end{aligned} \tag{4.10}$$

When individuals of the same species i tend to cluster, $\mathcal{D}_i(r)$ tends to 1 while it tends to the proportion of individuals of species i in the whole community when the distribution is uniform (Section 4.S2 of the SI).

Using Eq. 4.9 and 4.10, we obtain the formula for the dominance index in the presence of advection as

$$\mathcal{D}_i(r) = \frac{\frac{\lambda_i}{3D_i\ell_{\mathrm{B},i}}\left(\frac{\ell_{\mathrm{B},i}}{r} + \frac{\ell_{\mathrm{B},i}^3 \log\left(\frac{r^2}{\ell_{\mathrm{B},i}^2}+1\right)}{2r^3} + \arctan\left(\frac{r}{\ell_{\mathrm{B},i}}\right) - \frac{\pi}{2}\right) + \frac{4}{3}\pi C_i}{\frac{\lambda_i}{3D_i\ell_{\mathrm{B},i}}\left(\frac{\ell_{\mathrm{B},i}}{r} + \frac{\ell_{\mathrm{B},i}^3 \log\left(\frac{r^2}{\ell_{\mathrm{B},i}^2}+1\right)}{2r^3} + \arctan\left(\frac{r}{\ell_{\mathrm{B},i}}\right) - \frac{\pi}{2}\right) + \sum_{j=1}^{S}\frac{4}{3}\pi C_j}. \tag{4.11}$$

In the absence of advection ($U = 0, \gamma = 0$), we use the time-dependent dominance index, computed similarly:

$$\mathcal{D}_i(r,t) = \frac{\frac{\lambda_i}{D_i r}\left(\frac{1}{2} - \frac{1}{2}\operatorname{erf}\left(\frac{r}{\sqrt{8D_i t}}\right)\left(1 - \frac{4D_i}{r^2}t\right) - \frac{\sqrt{2D_i t}}{\sqrt{\pi}r}e^{-\frac{r^2}{8D_i t}}\right) + \frac{4}{3}\pi C_i}{\frac{\lambda_i}{D_i r}\left(\frac{1}{2} - \frac{1}{2}\operatorname{erf}\left(\frac{r}{\sqrt{8D_i t}}\right)\left(1 - \frac{4D_i}{r^2}t\right) - \frac{\sqrt{2D_i t}}{\sqrt{\pi}r}e^{-\frac{r^2}{8D_i t}}\right) + \sum_{j=1}^{S}\frac{4}{3}\pi C_j}. \tag{4.12}$$

Parameters

We model two types of organisms: microphytoplankton (defined by a diameter between 20 and 200 µm, here 50 µm) and nanophytoplankton (defined by a diameter between 2 and 20 µm, here 3 µm). These two groups are characterized respectively by a low diffusivity, slow growth and lower concentration vs. high diffusivity, fast growth and higher concentration. Organisms are displaced by a turbulent fluid whose velocity defines the time scale of the discretized model: we give here the reasoning behind parameter values, keeping in mind that our model parameters are only approximate. Main parameter definitions and values are given in Table 4.1.

Advection

We first consider the advection process, due to the turbulence of the environment. We only consider the Batchelor-Kolmogorov regime, i.e., the size of the volume W is below the size of the smallest eddy, but above the smallest length scale of fluctuations in nutrient concentrations. The defining scale of the environment therefore corresponds to a Reynolds number

$$\mathrm{Re} = \frac{U}{k\nu} \approx 1 \tag{4.13}$$

where $\nu = 10^{-6}$ m^2 s^{-1} is the kinematic viscosity for water. The smallest wavenumber k corresponds to the largest length scale L_s (Kolmogorov scale), i.e., $k = 2\pi/L_s$, with $L_s \approx 1$ cm

in the ocean [33]. The definition of the Reynolds number leads to

$$\begin{array}{rcl} 1 & \approx & \frac{UL_s}{2\pi\nu} \\ \Leftrightarrow U & \approx & \frac{2\pi\nu}{L_s}. \end{array} \tag{4.14}$$

This means that $U = 6.3 \times 10^{-4}$ m s^{-1}=5.4 $\times 10^3$ cm d^{-1}. Using $U\tau/3 = 0.5$ cm as in Young *et al.* [345], we have $\tau = 2.8 \times 10^{-4}$ d = 24 s. When $U\tau/3 = 0$, the environment is only diffusive, we keep the same value for τ. For $U\tau/3 = 0.5$ cm, $\gamma = 1231$ d^{-1}.

Diffusion

If we use the Stokes-Einstein equations ([110], cited from [104]), diffusivity can be computed with

$$D_i = \frac{RT}{N_A} \frac{1}{6\pi\eta a_i} \tag{4.15}$$

where $R = 8.314$ J K^{-1} mol^{-1} is the molar gas constant, $T = 293$ K is the temperature, $N_A = 6.0225 \times 10^{23}$ is Avogadro's number, $\eta = 10^{-3}$ m^{-1} kg s^{-1} is the dynamic viscosity of water and a_i is the radius of the organism considered.

Using $D_i = \frac{\Delta_i^2}{2\tau}$, we find that

$$\begin{array}{rcl} \Delta_i & = & \sqrt{2\tau D_i} \\ \Leftrightarrow \Delta_i & = & \sqrt{\frac{RT}{N_A} \frac{\tau}{3\pi\eta a_i}}. \end{array} \tag{4.16}$$

We consider $a_n = 1.5$ µm for nanophytoplankton individuals and $a_m = 25$ µm for microphytoplankton individuals, which allows us to compute Δ_n and Δ_m (see Table 4.1).

Ecological processes

We study the community at equilibrium, with the birth rate equal to the death rate, i.e., $p_i = q_i \, \forall i$. We use a microphytoplankton doubling rate of 1 d^{-1} [48] and consider the fastest-growing nanophytoplankton species, corresponding to a diameter of 3 µm [36], for which the doubling rate is between 2 and 3 d^{-1} (set to 2.5 d^{-1} here).

Parameter	Definition	Value
p_m, q_m	Probability of reproducing/dying for microphytoplankton individuals	2.8×10^{-4}
p_n, q_n	Probability of reproducing/dying for nanophytoplankton individuals	6.9×10^{-4}
U	Turbulent advection speed	{0, 0.06} cm.s^{-1}
Δ_m	Diffusion parameter for microphytoplankton individuals	6.4×10^{-5} cm
Δ_n	Diffusion parameter for nanophytoplankton individuals	2.6×10^{-4} cm

Table 4.1: Definitions and values of the main parameters used in the three-dimensional BBM, assuming the duration of a time step τ is 24 seconds.

Range of interaction

As we examine individual aggregation and its potential effects on interactions between species, we have to ascertain the volume in which an individual can be affected by the presence of other individuals, or affect other individuals. We only consider here interactions due to competition for nutrients, and therefore need to define a nutrient depletion volume. We approximate this volume as the sphere of radius r where $C(r) \leq 90\%C_\infty$ with C_∞ the background concentration of the nutrient. The radius of this nutrient depletion volume is maximized when the individual is in stagnant water so that diffusion is the only hydrodynamic process. In this case, the depletion radius corresponds to 10 times the radius of the individual [186, 187]. We define the maximum distance which allows for potential interactions (due to competition for resources) between two individuals of radius a_i and a_j as $d_{\text{threshold}}$, and the corresponding volume of potential interactions around an organism as $V_{\text{int}} = 4/3\pi d^3_{\text{threshold}}$ with

$$d_{\text{threshold}} = 10a_i + 10a_j. \tag{4.17}$$

We consider this maximum value as our baseline, keeping in mind that turbulence reduces the size of the nutrient depletion volume and increases the nutrient flux to the cell [18]. We caution that determination of the shape of the nutrient depletion volume in the presence of turbulence is too complex to be addressed here [187].

We consider a total volume of 1000 cm^3 for microphytoplankton and 10 cm^3 for nanophytoplankton (volumes are adapted to balance realistic concentrations and computation time) with periodic boundary conditions. Individuals are uniformly distributed in the cube at the beginning of the simulation. We run an idealized simulation with 3 species with an even abundance distribution of about 10^4 cells L^{-1} for microphytoplankton [271] and 10^6 cells L^{-1} for nanophytoplankton individuals [106]. We then model a more realistic community with 10 species having a skewed abundance distribution (between 55,000 and 400 cells L^{-1} for microphytoplankton, according to observations of field abundance distributions in [271], and multiplied by 10^2 for nanophytoplankton). All simulations are run for 1000 time steps of duration τ (corresponding to approximately 6h40). The computation of g and K of simulated distributions is explained in Section 4.S4 of the SI. The code for all simulations and analyses can be found at https://github.com/CoraliePicoche/brownian_bug_3D.

4.3 Results

We show an example of nanophytoplankton spatial distributions with and without advection at the end of a simulation in Fig. 4.1: clustering is not visible to the naked eye, even when zooming in on the observation volume, in the presence of advection, but removing turbulence helps visualising small aggregates of conspecifics. Microphytoplankton distributions are not so easy to analyse as no clusters can be detected from visual observations (although it may actually be present), whether advection is included or not (Section 4.S5 of the SI). Statistics are therefore needed to go further in detecting patterns of aggregation.

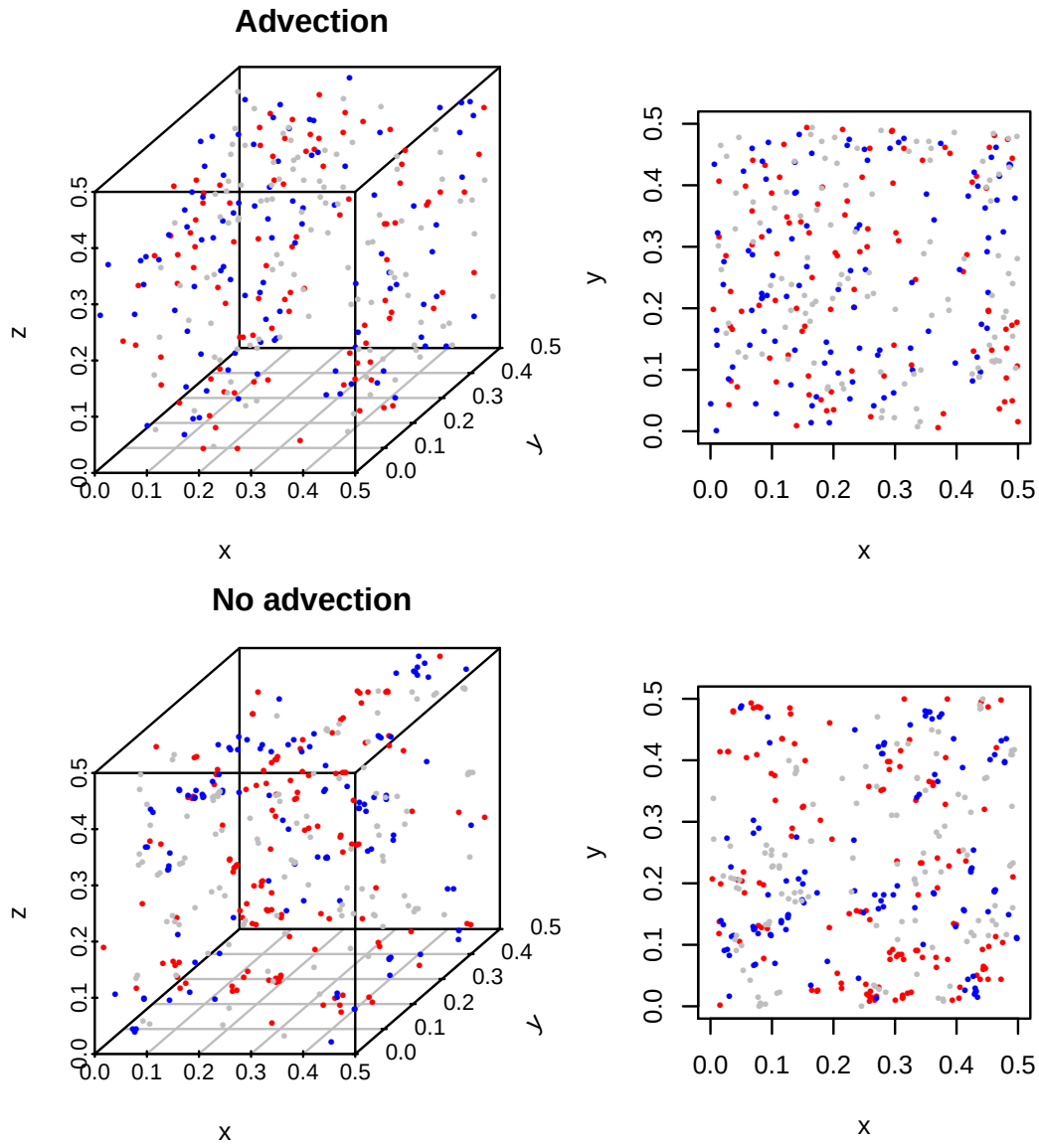

Figure 4.1: Spatial distributions of a 3-species community of nanophytoplankton with and without advection with density $C = 10^3$ cells cm^{-3} after 1000 time steps. Each color corresponds to a different species. On the left-hand side, only a zoom on a $0.5 \times 0.5 \times 0.5$ cm^3 cube is shown, and its projection on the x-y plane is shown on the right-hand side.

Ripley's K-functions extracted from numerical simulations match theoretical formula (Fig. 4.2) for both types of organisms, which also indicates that dominance indices extracted from the simulations match theoretical expectations.

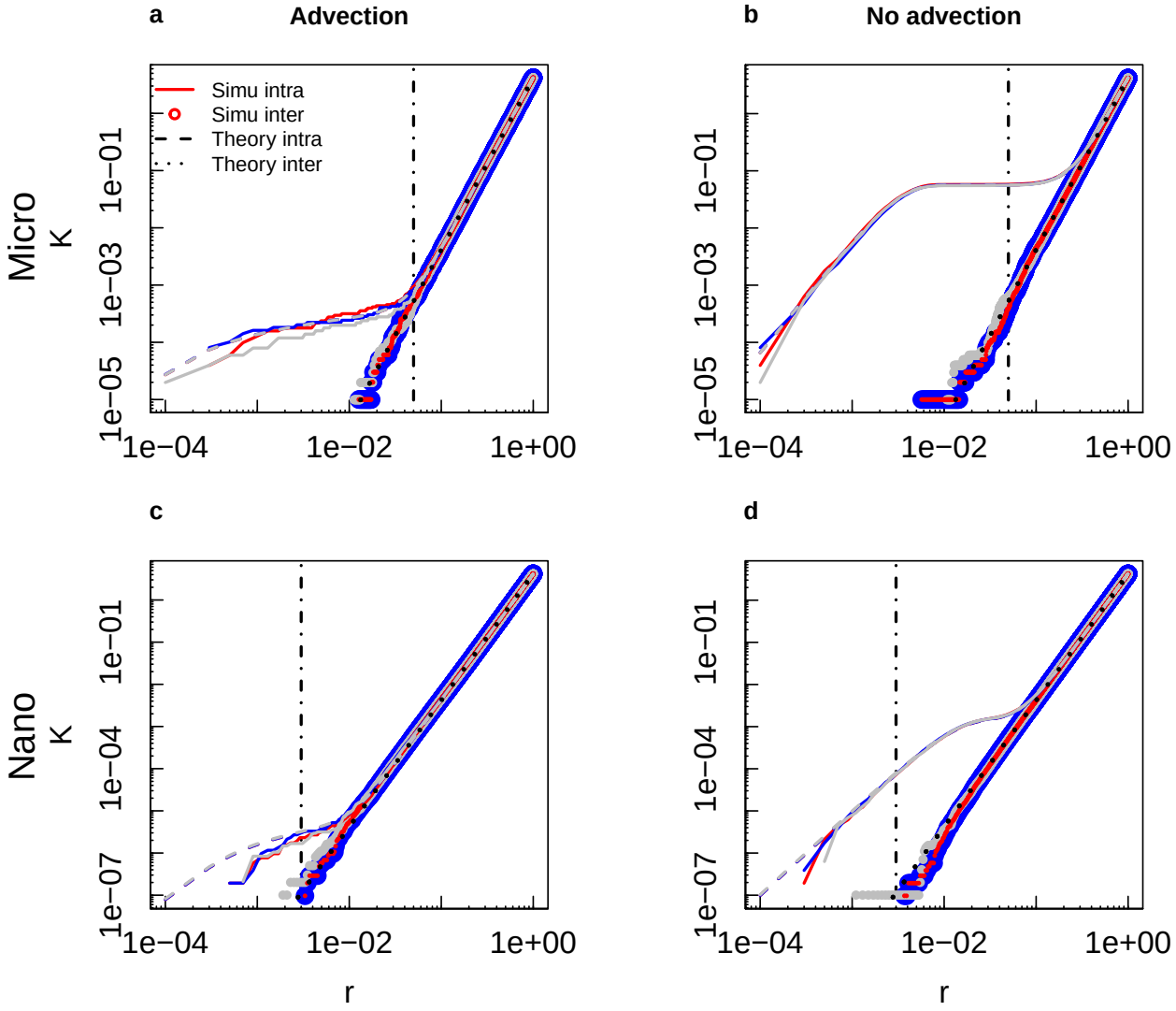

Figure 4.2: Comparison of theoretical and simulated Ripley's K-functions as a function of distance (in cm) for microphytoplankton (a-b) and nanophytoplankton (c-d) in a 3-species community with even abundance distributions after 1000 timesteps, with (a, c) and without (b, d) advection. Each color represents a different species. Intraspecific K-functions are shown with dashed (theoretical values) and solid (simulated values) lines. Interspecific K-functions are shown with dotted lines (theoretical values) and circles (simulated values). The black dash-dotted line corresponds to the threshold considered as the maximum distance for nutrient-based competition.

Dominance indices all follow a similar pattern (Fig. 4.3 and 4.4). The dominance index is close to 1 for small distances: there is always a scale at which an organism is surrounded almost only by conspecifics. The index then decreases sharply to converge at large distances (close to 1 cm) to the proportion of the focus species in the whole community, as it would for a uniform

spatial distribution. Patterns differ at intermediate ranges of distances between organisms.

In the presence of advection, the dominance index starts decreasing for a distance between 5 and 10 times smaller than when advection is absent, which indicates that organisms are closer to heterospecifics when their environment is turbulent. A quasi-uniform distribution is also reached for smaller distances with advection than without. Microphytoplankton species start mixing for distances larger than for nanophytoplankton species irrespective of the hydrodynamic regime surrounding them.

In a 3-species community with the same initial abundances, in the presence of advection, microphytoplankton dominance indices are between 0.37 and 0.47 at the distance threshold for potential interactions, while they are between 0.80 and 0.94 for nanophytoplankton species. In the absence of turbulence, dominance indices are all above 0.98 when the distance threshold is reached (Fig. 4.3). Microphytoplankton organisms are therefore as likely to share their depletion volume with conspecifics as they are with heterospecifics, but only when turbulent advection is accounted for, whereas nanophytoplankton organisms always have almost only conspecifics around them.

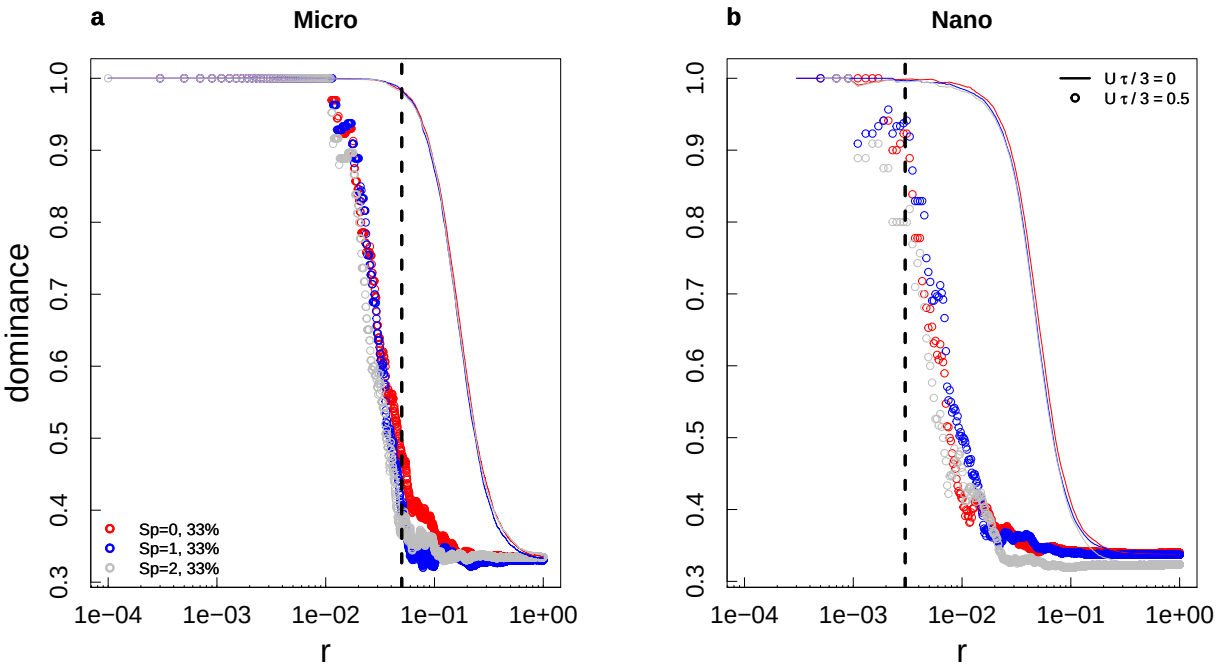

Figure 4.3: Dominance indices as a function of distance (in cm) for microphytoplankton (a) and nanophytoplankton (b) in a 3-species community with even abundance distributions (final proportions in the community are indicated in the figure) after 1000 timesteps, with (circles) and without (lines) advection. Each color represents a different species. The black dashed line corresponds to the threshold considered as the maximum distance for nutrient-based competition.

More mixing in microphytoplankton than nanophytoplankton, and more mixing with advection, also holds when considering a 10 species-community with a skewed abundance distribution (Fig. 4.4), but dominance indices are overall lower in communities with more species and with

less even abundances. In the presence of advection, microphytoplankton dominance indices at the distance threshold are between 0.34 (for the most abundant species) and 0.033 (for one of the least abundant species), while they are between 0.90 and 0.85 when advection is not taken into account. Nanophytoplankton species, too, are more mixed than in the 3 species-community: dominance indices vary between 0.54 and 0.2 when the depletion threshold is reached (with an exception of 0 for one particular species which had no conspecific for distances below 10^{-2} cm) when organisms are displaced by turbulence, while the same quantity is between 1 and 0.97 when they are only subject to diffusion.

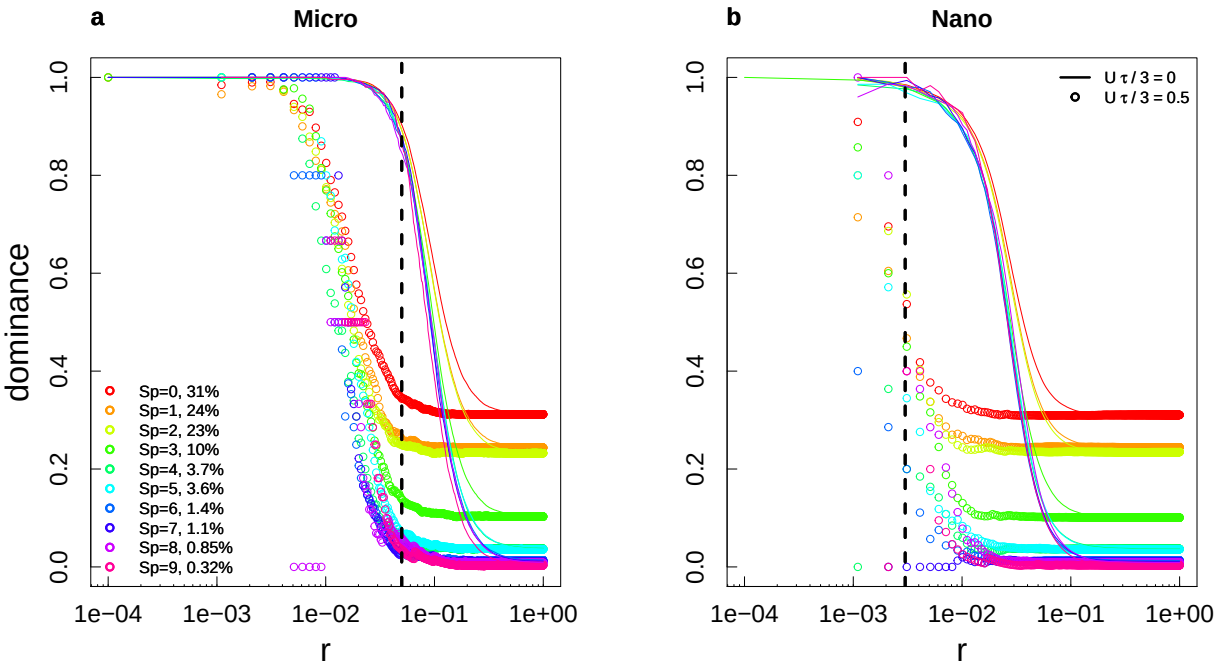

Figure 4.4: Dominance indices as a function of distance (in cm) for microphytoplankton (a) and nanophytoplankton (b) in a 10-species community with a skewed abundance distribution (final proportions in the community are indicated in the figure) after 1000 timesteps, with (circles) and without (lines) advection. Each color represents a different species. The black dashed line corresponds to the threshold considered as the maximum distance for nutrient-based competition.

Differences in spatial distributions are not only due to organism sizes, which determine their demographic and hydrodynamic properties, but also to their abundances (here set through initial values). In the presence of turbulence, the threshold distance at which dominance falls below 95% is smaller for more abundant species (Fig. 4.5 a-b). Abundant species tend to be present nearly everywhere when they are mixed in the environment. Therefore, they are also more likely to be close to a heterospecific, but still have more conspecifics close to them than the less abundant species ($\mathcal{D}\left(d_{\text{threshold}}\right)$ increases with abundance, Fig. 4.5 c-d). However, this increase is less marked for nanophytoplankton than for microphytoplankton (Fig. 4.5 c-d). When turbulence is absent, the relationships with abundance are unclear, possibly affected by sampling effects, and we refrain from interpreting them.

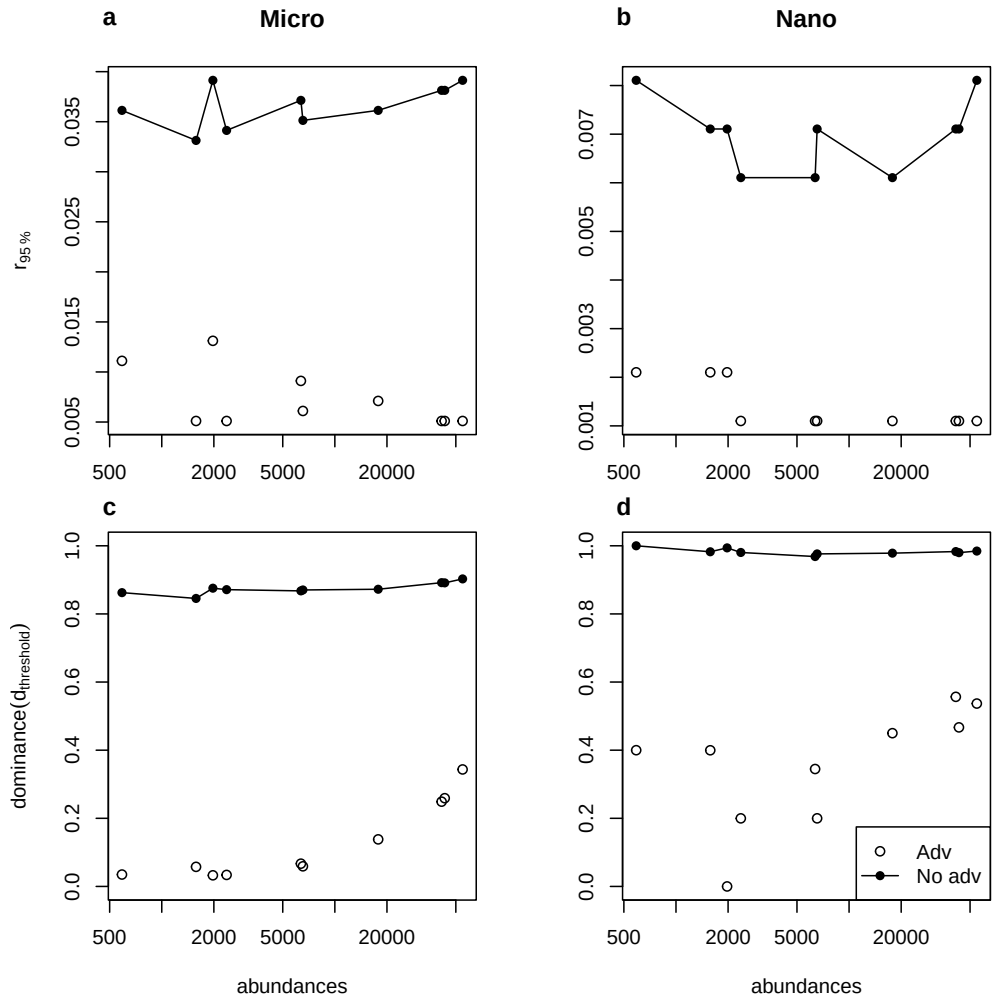

Figure 4.5: Minimum distances (in cm) between points for dominance to drop below 95% (a and b) and dominance at a distance corresponding to the threshold for competition (c and d) as a function of abundances (note the logarithmic scale on the x-axis) for microphytoplankton and nanophytoplankton. We consider cases with and without advection in a 10-species community with a skewed abundance distribution.

4.4 Discussion

We designed a stochastic, three-dimensional, individual-based model of the spatial distribution of multiple species in a viscous and turbulent flow. We conducted both mathematical analyses and numerical simulations to quantify spatial correlations in the distribution of organisms. We focused on the pair correlation function and Ripley's K-function, for which numerical and theoretical analyses showed a good agreement, and extracted a more ecologically-oriented metric from them, i.e., the dominance index. This statistic is the *local* average ratio of conspecifics, i.e., the number of organisms of the focal species in the neighbourhood of an individual of the same species, divided by the total number of organisms in that neighbourhood. Intraspecific clustering corresponds to a dominance index close to 1, which decreases when interspecific

mixing increases. The choice of this index was motivated by two reasons: (1) it is at its core a proportion of a focus species in a certain volume, i.e. a scale-dependent, localized metric bounded between 0 and 1 as opposed to other statistics whose values are less directly interpreted, and (2) it is easy to relate to coexistence theory as it describes the environment of an organism in terms of heterospecifics and conspecifics, which can, under certain hypotheses that we discuss below, be related to interspecific and intraspecific interactions. Comparing the distributions of organisms of different sizes, we showed that the presence of turbulence always increased mixing (results are robust to slight modifications in the computation of advection velocity U, shown in Section 4.S7 of the SI). The species composition around an organism depended on its size, which mechanically determines its hydrodynamic properties (diffusivity), and is linked with its ecological characteristics (growth rate and density). Microphytoplankters (20 to 200 µm), larger cells with lower diffusivity, growth rate and abundance, were on average further away from other cells, due to their lower concentrations (Figure 4.S10 of the SI), than nanophytoplankters (2 to 20 µm). However, they were surrounded by more heterospecifics than conspecifics within a volume of potential interactions, whose radius is defined as the maximum distance for which nutrient depletion volumes of two different individuals may overlap. If we consider that interactions between species (not modelled directly here because of timescale issues, see below) could occur with equal probability at all distances within the volume of potential interactions, we would conclude that microphytoplankters are more likely to interact with individuals from other species than with individuals of their own species. This affirmation is, however, conditional upon interactions at 10 cell diameters from an individual being equally likely than at 1 diameter from an individual. If we keep in mind that interactions are more likely or stronger at very short distances, microphytoplankters may still experience more frequent effects of conspecifics than heterospecifics.

To see this, let us first focus on the smallest distances between organisms. The nearest neighbour of an organism was always an organism of the same species, and the minimum distance between conspecifics was always lower than expected for a uniform distribution (Section 4.S6 of the SI). The dominance index remained close to 1 for distances below 10^{-2} cm or 10^{-3} cm for microphytoplankton and nanophytoplankton respectively. There was therefore always *some* intraspecific aggregation, i.e. conspecifics were always closer than heterospecifics at the smallest distances. This is due to the prevalence of demographic processes at individual scales, because an individual acts as a source point for other organisms of the same species, and hydrodynamic processes do not separate conspecifics fast enough to prevent aggregation. If we consider that interaction strengths are a smoothly decaying function of distance, a common assumption in spatial coexistence models (e.g., [52, 201]), this implies that population-level intraspecific interactions could be stronger than interspecific interactions due to intraspecific micro-scale aggregation. However, the mechanisms of competition at this scale are poorly known, likely relying on multiple types of resources with different distributions in the environment, effects on the cell, uptakes, etc. Rather than weighting much more heavily the potential interactions

with the closest neighbour(s) through an interaction kernel, we therefore chose conservatively to define a maximum distance for two organisms to possibly affect the concentrations of elements in the environment of each other. We consider that, at all distances below this threshold, interactions could happen between organisms. We continue the discussion with that simplification in mind, and explicitly mention when it is relaxed.

Dominance indices began to decrease at distances above 10^{-3} cm, still below the maximum distance for interactions. At this distance and above, the balance between heterospecifics and conspecifics was much more sensitive to different phytoplankters' demographic and hydrodynamic traits. The species composition of an organism's neighbourhood depended on its size: nanophytoplankton organisms mainly shared their volume of potential interactions with conspecifics (the dominance index remained close to 1, even near the distance threshold, i.e. the maximum distance for the overlap of nutrient depletion volumes) while microphytoplankton organisms could affect both conspecifics and heterospecifics (the dominance index was often below 0.5 at the distance threshold, i.e. an individual's depletion zone probably overlapped with more heterospecifics' than conspecifics'). Microphytoplankters were therefore more likely to share their depletion volume with heterospecifics than nanophytoplankters. The rate of production of new microphytoplankton conspecifics was not sufficient to compensate for the mixing induced by turbulence and diffusivity, even though the diffusivity of microphytoplankters was smaller than that of nanophytoplankters. There may therefore be different mechanisms at play at the community level for microphytoplankton and nanophytoplankton to maintain coexistence. For nanophytoplankton, the spatial structure likely leads to more interactions between conspecifics than between heterospecifics. The spatial distribution of microphytoplankton species, on the contrary, encourages more interactions between heterospecifics. If we consider that local interaction strengths are equal within the volume of potential interactions, scaling to the population level, we would likely observe stronger intra- over interspecific interactions for nanophytoplankton (a key factor in coexistence theory, [29]) but not necessarily so for microphytoplankton. Using a timescale separation argument, we show in Section 4.S8 in the SI how stronger interactions at population level than individual level may arise in a Lotka-Volterra model whose spatial structure is summed up by the dominance indices evidenced here. Stronger intra- than interspecific competition may arise at population level even when assuming that all local interaction strengths between individuals are equal, regardless of the identity of competitors.

All of the above discussion is based on a microphytoplankter's neighbourhood in its nutrient depletion volume. To simplify the computation, we used maximum volumes of potential interactions, corresponding to a diffusive-only flow of nutrient particles. But when fluid turbulence increases, nutrient uptake increases, and the size of the depletion zone decreases [187]. The proportion of change in the depletion volume increases with the size of organisms: a 10 µm-diameter organism might not experience any change, while the uptake of a 100 µm-diameter organism would increase by at least 50% [187]. Therefore the volume of potential interactions shrinks

in the presence of turbulence for microphytoplankton, but not necessarily for nanophytoplankton. This could be one additional reason why microphytoplankers might still be surrounded by conspecifics at ecologically meaningful distances and interacting more frequently with them.

Up to now, we have only focused on the dominance index, a localized proportion of conspecifics. However, interactions also depend on the absolute densities of individuals. Mechanically, when density decreases, the distances between neighbours increase, which explains that the distances between the low-abundance microphytoplankters tended to be greater than distances between the more abundant nanophytoplankters (Section 4.S6 of the SI). Explicit mathematical models using pair densities to express interaction rates (e.g., [201, 275]) may be able to incorporate those effects; however, as we highlight below, the timescales and spatial correlations that are seen in such models may not necessarily represent faithfully phytoplankton community dynamics.

Contrary to other similar models (e.g., [47, 55]), we did not consider explicit effects of local density on survival and fertility rates. Outside of simply maintaining analytical tractability, we had another, more biological reason to do so: we cannot be sure that these local density-dependencies make sense in our phytoplankton context. To understand why, consider that even if a species abundance is locally tripled, competition might not directly ensue at the time scales covered by our model ($\approx$ 7h), if nutrient depletion has not had time to set in yet. Even if we considered longer time frames, we would need lagged local density-dependencies, which are to our knowledge not leading to tractable spatial branching or dynamic point processes. We could, of course, directly model nutrients, perhaps as resource "points" with a dynamics of their own [244, 254], which in turn change the reproduction or death rate of individuals. If the resource points risk being depleted, this entails a negative spatial correlation between organisms and their resources [244, 30]. And that is where such models might be inadequate. The phycosphere, a micro-environment at the periphery of a phytoplankton organism where communities of bacteria interact [304], can also impact phytoplankton fitness, both positively (cross-feeding) and negatively (algicidal activities of bacteria). This can sometimes lead to an accumulation of key resources close to the phytoplankter. This will lead to positive spatial correlations between consumers and their resources, and we currently do not have theoretical models to represent this process (short of modelling precisely the spatial distribution of these bacteria).

Our model should be viewed as a first model of spatial distributions of multiple phytoplankton species in a realistic, three-dimensional environment at the microscale, describing only basic hydrodynamic and demographic processes. Using this model, we were able to predict whether phytoplankters could be in contact with individuals of their own or other species, and emit reasonable conjectures regarding potential intra vs interspecific interactions between species, emerging at the population level through spatial distributions [97]. It is worthwhile to keep in mind that there are many remaining features of phytoplankton physiology and life histories which we do not address here, but which may affect spatial distributions. Many phytoplankters

are able to move actively in three dimensions, which can favour cluster formation [59]. Even those who are believed to move passively actually often move along the vertical dimension by regulating their buoyancy [285], and can at times aggregate to form pairs [123]. Finally, a part of spatial structure is explained by the partially colonial nature of microphytoplankton [193]. This clearly calls for viewing our model as a null model to which more complex mechanistic models and their spatial outputs can be compared.

Acknowledgments

FB and CP were supported by the grant ANR-20-CE45-0004. CP was supported by a PhD grant from the French Ministry of Research.

4.5 Appendices

Derivation of the spatial characteristics of the Brownian Bug Model

We show here how to compute the monospecific pair correlation function and Ripley's K-function of the Brownian Bug Model (see [345] and [272] for a detailed derivation of the master equation). As these formula only apply to intraspecies pairs, we ignore species' index in the following for the sake of clarity. Similar formula for well-known spatial point processes are given in the Supplementary Information, for readers who want to understand better the properties of these spatial statistics.

Proof of Eq. 4.4 and Eq. 4.5

In three dimensions, when the birth rate λ is the same as the mortality rate μ, the pair density $G(r)$ is a solution of

$$\frac{\partial G}{\partial t} = \frac{2D}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial G}{\partial r}\right) + \frac{\gamma}{r^2}\frac{\partial}{\partial r}\left(r^4\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{\xi}). \tag{4.18}$$

Steady-state solution We first compute the steady-state solution, *i.e.*

$$\begin{aligned}
0 &= \frac{2D}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial G}{\partial r}\right) + \frac{\gamma}{r^2}\frac{\partial}{\partial r}\left(r^4\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{\xi}) \\
0 &= 4\pi r^2\left(\frac{2D}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial G}{\partial r}\right) + \frac{\gamma}{r^2}\frac{\partial}{\partial r}\left(r^4\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{\xi})\right) \\
0 &= 4\pi\left(2D\frac{\partial}{\partial r}\left(r^2\frac{\partial G}{\partial r}\right) + \gamma\frac{\partial}{\partial r}\left(r^4\frac{\partial G}{\partial r}\right)\right) + 4\pi r^2 2\lambda C\delta(\boldsymbol{\xi}).
\end{aligned} \tag{4.19}$$

We can then integrate Eq. 4.18 over a small sphere centered on an individual, with radius ρ. Let us first note that

$$\begin{aligned} \int_{\mathbb{R}^3} \delta(\boldsymbol{\xi}) d\boldsymbol{\xi} &= 1 \\ \Leftrightarrow \int_0^{2\pi} \int_0^{\pi} \int_0^{\rho} \delta(r)\delta(\phi)\delta(\theta) r^2 \sin(\phi) dr d\phi d\theta &= 1 \\ \Leftrightarrow 4\pi \int_0^{\rho} \delta(r) r^2 dr &= 1. \end{aligned} \tag{4.20}$$

Using Eq. 4.19 and 4.20,

$$\begin{aligned} 0 &= 4\pi \left(2Dr^2 \frac{\partial G}{\partial r} + \gamma r^4 \frac{\partial G}{\partial r} \right) + 2\lambda C \\ \Leftrightarrow \frac{\partial G}{\partial r} &= -\frac{1}{4\pi} \frac{2\lambda C}{2Dr^2 + \gamma r^4}. \end{aligned} \tag{4.21}$$

We can integrate Eq. 4.21 between ρ and ∞. As $G(\infty) = C^2$,

$$C^2 - G(\rho) = -\frac{\lambda C}{2\pi} \int_{\rho}^{\infty} \frac{1}{2Dr^2 + \gamma r^4} dr. \tag{4.22}$$

We first compute the primitive $A = \int \frac{1}{2Dr^2 + \gamma r^4} dr$.

$$A = \int \frac{1}{r^2 (2D + \gamma r^2)} dr \tag{4.23}$$

$$= \int \frac{1}{2Dr^2} - \frac{\gamma}{2D (2D + \gamma r^2)} dr \tag{4.24}$$

$$= -\frac{1}{2Dr} - \frac{\gamma}{2D} \int \frac{1}{2D \left(1 + \left(\sqrt{\frac{\gamma}{2D}} r \right)^2 \right)} dr. \tag{4.25}$$

With a change of variable $u = \sqrt{\frac{\gamma}{2D}} r$, using $\int \frac{1}{1+u^2} = \arctan(u)$, we have

$$A = -\frac{1}{2Dr} - \frac{\sqrt{\gamma} \arctan\left(\frac{\sqrt{\gamma} r}{\sqrt{2D}} \right)}{2\sqrt{2} D \sqrt{D}} + K \tag{4.26}$$

where K is a constant. We can know compute $B = [A]_{\rho}^{\infty}$.

$$B = -\frac{\sqrt{\gamma}\pi}{4\sqrt{2}D\sqrt{D}} + \frac{1}{2D\rho} + \frac{\sqrt{\gamma} \arctan\left(\frac{\sqrt{\gamma}\rho}{\sqrt{2D}} \right)}{2\sqrt{2}D\sqrt{D}}. \tag{4.27}$$

This leads to

$$G(\rho) = C^2 + \frac{\lambda C}{2\pi} B \tag{4.28}$$

$$= C^2 + \frac{\lambda C}{2\pi}\left[\frac{1}{2D\rho} + \frac{\sqrt{\gamma}\arctan\left(\frac{\sqrt{\gamma}\rho}{\sqrt{2D}}\right)}{2\sqrt{2}D\sqrt{D}} - \frac{\sqrt{\gamma}\pi}{4\sqrt{2}D\sqrt{D}}\right]. \tag{4.29}$$

Finally, the pair correlation function $g = G/C^2$ is defined as

$$g(\rho) = \frac{\lambda}{4\pi CD}\left(\frac{\sqrt{\gamma}\arctan\left(\frac{\sqrt{\gamma}\rho}{\sqrt{2D}}\right)}{\sqrt{2D}} + \frac{1}{\rho} - \frac{\pi\sqrt{\gamma}}{2\sqrt{2D}}\right) + 1. \tag{4.30}$$

Time-dependent solution In the absence of advection by turbulent diffusion ($U = 0$, $\gamma = 0$), convergence to the steady-state solution can be very slow (more than a week, see Section 4.S3 in the SI). In order to keep a realistic timeframe, we need to compute a time-dependent solution. We can get back to Eq. 4.18 with $\gamma = 0$, which yields

$$\frac{\partial G}{\partial t} = \frac{2D}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial G}{\partial r}\right) + 2\lambda C\delta(\boldsymbol{\xi}). \tag{4.31}$$

Assuming an isotropic environment, this means

$$\frac{\partial G}{\partial t} - 2D\Delta G = 2\lambda C\delta(\boldsymbol{\xi}) \tag{4.32}$$

where $\Delta = \nabla^2$ is the Laplacian operator. We therefore have

$$\mathcal{L}G(\boldsymbol{\xi}, t) = 2\lambda C\delta(\boldsymbol{\xi}) \tag{4.33}$$

where $\mathcal{L}$ is the linear differential operator $\partial_t - 2D\Delta$.

Using the Green's function theory, we know that $G(y) = \int H(y,s)2\lambda C\delta(s)ds$ where $H(y,s) = H(y-s)$ is the Green kernel (heat kernel). We can therefore write

$$\begin{array}{rcl} G(\boldsymbol{\xi}, t) & = & 2\lambda C \int_{\mathbb{R}^3}\int_0^t H(\boldsymbol{\xi} - \boldsymbol{\xi'}, t')\delta(\boldsymbol{\xi'})d\boldsymbol{\xi'}dt' \\ \Leftrightarrow G(\boldsymbol{\xi}, t) & = & 2\lambda C \int_0^t H(\boldsymbol{\xi}, t')dt'. \end{array} \tag{4.34}$$

A solution for the Green's function using $\mathcal{L} = \partial_t - 2D\Delta$ in 3 dimensions is $H(r,t) = \left(\frac{1}{8\pi Dt}\right)^{3/2}\exp(\frac{-r^2}{8Dt})$. $G(r,t)$ can then be computed as

$$G(r,t) = 2\lambda C\left(\frac{-\operatorname{erf}\left(\frac{r}{\sqrt{8tD}}\right)}{8\pi Dr} + K\right) \tag{4.35}$$

where erf is the error function. Using $G(r,0) = C^2$ and $\lim_{x\to+\infty} \mathrm{erf}(x) = 1$ in Eq. 4.35,

$$\begin{array}{rcl} C^2 & = & 2\lambda C\left(\frac{1}{8\pi D r} + K\right) \\ \Leftrightarrow \frac{C}{2\lambda} - \frac{1}{8\pi D r} & = & K. \end{array} \tag{4.36}$$

We can finally compute $G(r,t)$:

$$\begin{array}{rcl} G(r,t) & = & 2\lambda C\left(-\frac{\mathrm{erf}\left(\frac{r}{\sqrt{8tD}}\right)}{8\pi D r} + \frac{C}{2\lambda} + \frac{1}{8D\pi r}\right) \\ & = & \frac{\lambda C}{4\pi D r}\left\{1 - \mathrm{erf}\left(\frac{r}{\sqrt{8Dt}}\right)\right\} + C^2 \\ \Leftrightarrow g(r,t) & = & \frac{\lambda}{4D\pi r C}\left\{1 - \mathrm{erf}\left(\frac{r}{\sqrt{8Dt}}\right)\right\} + 1. \end{array} \tag{4.37}$$

Proof of Eq. 4.8 and Eq. 4.9

We can integrate the pcf formula to compute Ripley's K-function, as $g(r) = \frac{K'(r)}{4\pi r^2}$.

Steady-state solution From Eq. 4.30,

$$K(\rho) = 4\pi \int_0^\rho r^2 + \frac{\lambda}{2\pi C}\left[\frac{r}{2D} + \frac{\sqrt{\gamma} r^2 \arctan\left(\frac{\sqrt{\gamma} r}{\sqrt{2D}}\right)}{2\sqrt{2}D\sqrt{D}} - \frac{\sqrt{\gamma}\pi r^2}{4\sqrt{2}D\sqrt{D}}\right] dr. \tag{4.38}$$

We define $A = \int_0^\rho r^2 dr$, $B = \int_0^\rho \frac{r}{2D} dr$, $C = \int_0^\rho r^2 \arctan\left(\frac{\sqrt{\gamma} r}{\sqrt{2D}}\right) dr$ and $E = \int_0^\rho \frac{\sqrt{\gamma}\pi r^2}{4\sqrt{2}D\sqrt{D}} dr$.

$$\begin{array}{rcl} A = & \frac{1}{3}\rho^3. \\ B = & \frac{\rho^2}{4D}. \\ E = & \frac{\sqrt{\gamma}\pi\rho^3}{12\sqrt{2}D\sqrt{D}}. \end{array} \tag{4.39}$$

We can also compute $C = \int_0^\rho r^2 \arctan\left(\frac{\sqrt{\gamma} r}{\sqrt{2D}}\right) dr$. We first change variable, with $u = \frac{r}{\sqrt{2D}}$, $dr = \sqrt{2D} du$, and obtain

$$C = (2D)^{3/2} \int_0^{\rho/\sqrt{2D}} u^2 \arctan(\sqrt{\gamma} u) du. \tag{4.40}$$

We can integrate by parts, with $f = \arctan(\sqrt{\gamma} u)$ and $g' = u^2$, which leads to

$$C = (2D)^{3/2}\left(\frac{\rho^3}{3(2D)^{3/2}} \arctan(\sqrt{\frac{\gamma}{2D}}\rho) - \frac{\sqrt{\gamma}}{3} \int_0^{\rho/\sqrt{2D}} \frac{u^3}{(\gamma u^2+1)} du\right). \tag{4.41}$$

We then substitute $v = \gamma u^2 + 1$, $du = \frac{1}{2\gamma u} dv$, and have

$$\begin{array}{rcl} \int_0^{\rho/\sqrt{2D}} \frac{u^3}{(\gamma u^2+1)} du & = & \frac{1}{2\gamma^2} \int_1^{\gamma\rho^2/2D+1} \frac{v-1}{v} dv \\ & = & \frac{1}{2\gamma^2} \int_1^{\gamma\rho^2/2D+1} 1 - \frac{1}{v} dv \\ & = & \frac{1}{2\gamma^2}(\gamma\frac{\rho^2}{2D} - \log(\gamma\frac{\rho^2}{2D} + 1)). \end{array} \tag{4.42}$$

Going back to C, we obtain

$$\begin{aligned} C &= \frac{\rho^3 \arctan(\sqrt{\frac{\gamma}{2D}}\rho)}{3} - (2D)^{3/2}\frac{\sqrt{\gamma}}{3}\frac{1}{2\gamma^2}\left(\frac{\gamma}{2D}\rho^2 - \log(\gamma\frac{\rho^2}{2D}+1)\right) \\ &= \frac{\rho^3 \arctan(\sqrt{\frac{\gamma}{2D}}\rho)}{3} - \frac{\sqrt{2D}}{6\sqrt{\gamma}}\rho^2 + \frac{\sqrt{2}D^{3/2}}{3\gamma^{3/2}}\log(\gamma\frac{\rho^2}{2D}+1). \end{aligned} \tag{4.43}$$

Combining all equations,

$$\begin{aligned} K(\rho) &= \frac{4}{3}\pi\rho^3 + \frac{2\lambda}{C}\left(\frac{\rho^2}{4D} + \frac{\sqrt{\gamma}\rho^3\arctan(\sqrt{\frac{\gamma}{2D}}\rho)}{6\sqrt{2}D^{3/2}} - \frac{\rho^2}{12D} + \frac{\log\left(\gamma\frac{\rho^2}{2D}+1\right)}{6\gamma} - \frac{\sqrt{\gamma}\pi\rho^3}{12\sqrt{2}D\sqrt{D}}\right) \\ &= \frac{4}{3}\pi\rho^3 + \frac{\lambda}{3C}\left(\frac{\rho^2}{D} + \frac{\sqrt{\gamma}\rho^3\arctan(\sqrt{\frac{\gamma}{2D}}\rho)}{\sqrt{2}D^{3/2}} + \frac{\log\left(\gamma\frac{\rho^2}{2D}+1\right)}{\gamma} - \frac{\sqrt{\gamma}\pi\rho^3}{2\sqrt{2}D\sqrt{D}}\right). \end{aligned} \tag{4.44}$$

Note that in the absence of advection,

$$\begin{aligned} g(r) &= \frac{\lambda}{4\pi CDr} + 1 \\ \Rightarrow K'(r) &= \frac{\lambda r}{CD} + 4\pi r^2 \\ \Leftrightarrow K(r) &= \frac{\lambda r^2}{2CD} + \frac{4}{3}\pi r^3. \end{aligned} \tag{4.45}$$

Time-dependent solution In the absence of advection ($U = 0, \gamma = 0$), we need to compute a time-dependent solution. From eq. 4.37,

$$\begin{aligned} K(\rho) &= \frac{\lambda}{DC}\int_0^\rho r\left\{1 - \operatorname{erf}\left(\frac{r}{\sqrt{8Dt}}\right)\right\} + 4\pi r^2 dr \\ &= \frac{\lambda}{CD}\left(\frac{\rho^2}{2} - \int_0^\rho r \times \operatorname{erf}\left(\frac{r}{\sqrt{8Dt}}\right)dr\right) + \frac{4}{3}\pi\rho^3. \end{aligned} \tag{4.46}$$

We first compute the primitive for $\int_0^\rho r \times \operatorname{erf}\left(\frac{r}{\sqrt{8Dt}}\right)dr$. We define $u = \frac{r}{\sqrt{8Dt}}$, $dr = \sqrt{8Dt}du$, then

$$\int_0^\rho r \times \operatorname{erf}\left(\frac{r}{\sqrt{8Dt}}\right)dr = 8Dt\int_0^{\rho/\sqrt{8Dt}} u \times \operatorname{erf}(u)\,du. \tag{4.47}$$

We can integrate by parts, with $f = \operatorname{erf}(u)$ and $g' = u$, and obtain

$$8Dt\int_0^{\rho/\sqrt{8Dt}} u \times \operatorname{erf}(u)\,du = 8Dt\left(\frac{\rho^2}{2}\frac{1}{8Dt}\operatorname{erf}(\frac{\rho}{\sqrt{8Dt}}) - \frac{1}{\sqrt{\pi}}\int_0^{\rho/\sqrt{8Dt}} u^2 e^{-u^2}du\right). \tag{4.48}$$

We integrate by parts again, this time with $f = u$ and $g' = ue^{-u^2}$, which leads to

$$\int u^2 e^{-u^2}du = -\frac{ue^{-u^2}}{2} + \frac{1}{2}\int e^{-u^2}du = -\frac{ue^{-u^2}}{2} + \frac{\sqrt{\pi}\operatorname{erf}(u)}{4}. \tag{4.49}$$

If we use Eq. 4.49 in Eq. 4.48,

$$\begin{aligned} 8Dt\int_0^{\rho/\sqrt{8Dt}} u \times \operatorname{erf}(u)\,du &= 8Dt\left(\frac{\rho^2}{2}\frac{1}{8Dt}\operatorname{erf}(\frac{\rho}{\sqrt{8Dt}}) - \frac{\operatorname{erf}(\frac{\rho}{\sqrt{8Dt}})}{4} + \frac{1}{2\sqrt{\pi}}\frac{\rho}{\sqrt{8Dt}}e^{-\frac{\rho^2}{8Dt}}\right) \\ \Leftrightarrow \int_0^\rho r \times \operatorname{erf}\left(\frac{r}{\sqrt{8Dt}}\right)dr &= \frac{1}{2}\operatorname{erf}(\frac{\rho}{\sqrt{8Dt}})(\rho^2 - 4Dt) + \frac{\sqrt{2Dt}}{\sqrt{\pi}}\rho e^{-\frac{\rho^2}{8Dt}}. \end{aligned} \tag{4.50}$$

We can now compute $K(\rho)$:

$$K(\rho) = \frac{\lambda}{CD}\left(\frac{\rho^2}{2} - \frac{1}{2}\,\mathrm{erf}(\frac{\rho}{\sqrt{8Dt}})(\rho^2 - 4Dt) - \frac{\sqrt{2Dt}\rho}{\sqrt{\pi}}e^{-\frac{\rho^2}{8Dt}}\right) + \frac{4}{3}\pi\rho^3. \quad (4.51)$$

4.S Supplementary Information

4.S1 Derivation of the turbulent map

We show here how to derive a discrete-time map for turbulence from the continuous-time formula. We consider that the velocity field $\mathbf{u}^T = (u_x, u_y, u_z)$ at position $\boldsymbol{x}^T = (x, y, z)$ alternates between the three dimensions during a period τ, so that

$$\mathbf{u}^T(\boldsymbol{x}, t) = \begin{cases} (U\cos(ky+\phi), 0, 0) & \text{for } n\tau \le t < (n+\frac{1}{3})\tau \\ (0, U\cos(kz+\theta), 0) & \text{for } (n+\frac{1}{3})\tau \le t < (n+\frac{2}{3})\tau \\ (0, 0, U\cos(kx+\psi)) & \text{for } (n+\frac{2}{3})\tau \le t < (n+1)\tau. \end{cases} \tag{4.S1}$$

The discrete-time map can be obtained by computing the displacement over a period, between $t = n\tau$ and $t+1 = (n+1)\tau$, with $\boldsymbol{x}(t+1) = \boldsymbol{x}(t) + \int_{n\tau}^{(n+1)\tau} \mathbf{u}(\boldsymbol{x}, t)dt$, and knowing the initial position $\mathbf{x}(t)$. This can be solved in three steps (eqs. 4.S2, 4.S3 and 4.S4). We start with

$$\begin{aligned} x(t+\tau/3) &= x(t) + \tfrac{U\tau}{3}\cos(ky(t)+\phi) \\ y(t+\tau/3) &= y(t) \\ z(t+\tau/3) &= z(t). \end{aligned} \tag{4.S2}$$

Then,

$$\begin{aligned} x(t+2\tau/3) &= x(t+\tau/3) \\ y(t+2\tau/3) &= y(t) + \tfrac{U\tau}{3}\cos(kz(t)+\theta) \\ z(t+2\tau/3) &= z(t). \end{aligned} \tag{4.S3}$$

And finally,

$$\begin{aligned} x(t+\tau) &= x(t+\tau/3) \\ y(t+\tau) &= y(t+2\tau/3) \\ z(t+\tau) &= z(t) + \tfrac{U\tau}{3}\cos(kx(t+\tau)+\psi). \end{aligned} \tag{4.S4}$$

In the third step, we need z to be a function of $x(t+\tau)$, not $x(t)$, so that the volume is conserved (the determinant of the Jacobian matrix is equal to 1).

4.S2 Characteristics of standard spatial point processes

In order to get the reader acquainted with the spatial point process metrics that we use in the main text, we present here the analytical formulas and corresponding figures (Fig. 4.S1 and 4.S2) for the pair correlation function, Ripley's K-function and dominance index for well-known point processes. We focus on the uniform distribution, i.e. the Poisson point process, and a clustered distribution, the Thomas point process. The Thomas point process is the result of a two-stage mechanism: a Poisson point process generates "parent points" around which "daughter points" are scattered, their locations following a Gaussian distribution centered on

the parent location, with standard deviation σ. The numbers of parents and daughters per parent follow two Poisson distributions with mean N_p and N_d respectively. All solutions are given for three-dimensional spatial distributions.

Pair correlation function

In the case of a Poisson point process,

$$\forall r \geq 0,\ g_{ii}(r) = 1. \tag{4.S5}$$

For a Thomas point process, the expected value of the pcf is

$$g_{ii}(r) = 1 + \frac{1}{C_p} \frac{1}{(4\pi\sigma^2)^{3/2}} e^{-\left(\frac{r^2}{4\sigma^2}\right)} \tag{4.S6}$$

where $C_p = N_p/V$ is the concentration/intensity of the parent process in the volume V.

For a random superposition of stationary point processes with marks (species) i and j, $\forall i \neq j, \forall r \geq 0,\ g_{ij}(r) = 1$ [169, p. 326, eq. 5.3.13].

Ripley's K-function

In the case of a Poisson point process,

$$\forall r \geq 0,\ K_{ii}(r) = \frac{4}{3}\pi r^3. \tag{4.S7}$$

For a Thomas point process,

$$K_{ii}(r) = \frac{4}{3}\pi r^3 + \frac{1}{C_p \sigma\sqrt{\pi}} \left(\sigma\sqrt{\pi}\,\mathrm{erf}\left(\frac{r}{2\sigma}\right) - r e^{-\left(\frac{r}{2\sigma}\right)^2} \right). \tag{4.S8}$$

For a random superposition of stationary point processes, $K_{ij}(r) = \frac{4}{3}\pi r^3$ [169, p. 324, eq. 5.3.5].

Dominance index

In the Poisson point process, $K_{ii}(r) = K_{ij}(r)$, which means that the dominance index can be reduced to ratios of concentrations:

$$\mathcal{D}_i(r) = \frac{C_i}{\sum_{j=1}^{S} C_j}. \tag{4.S9}$$

In the Thomas process, using eq. 4.S8,

$$\mathcal{D}_i(r) = \frac{C_i\left(\frac{4}{3}\pi r^3 + \frac{F(r)}{C_{p,i}}\right)}{C_i\frac{F(r)}{C_{p,i}} + \sum_j C_j \frac{4}{3}\pi r^3} \tag{4.S10}$$

with $F(r) = \frac{1}{\sigma\sqrt{\pi}}\left(\sigma\sqrt{\pi}\,\mathrm{erf}\left(\frac{r}{2\sigma}\right) - re^{-\left(\frac{r}{2\sigma}\right)^2}\right)$.

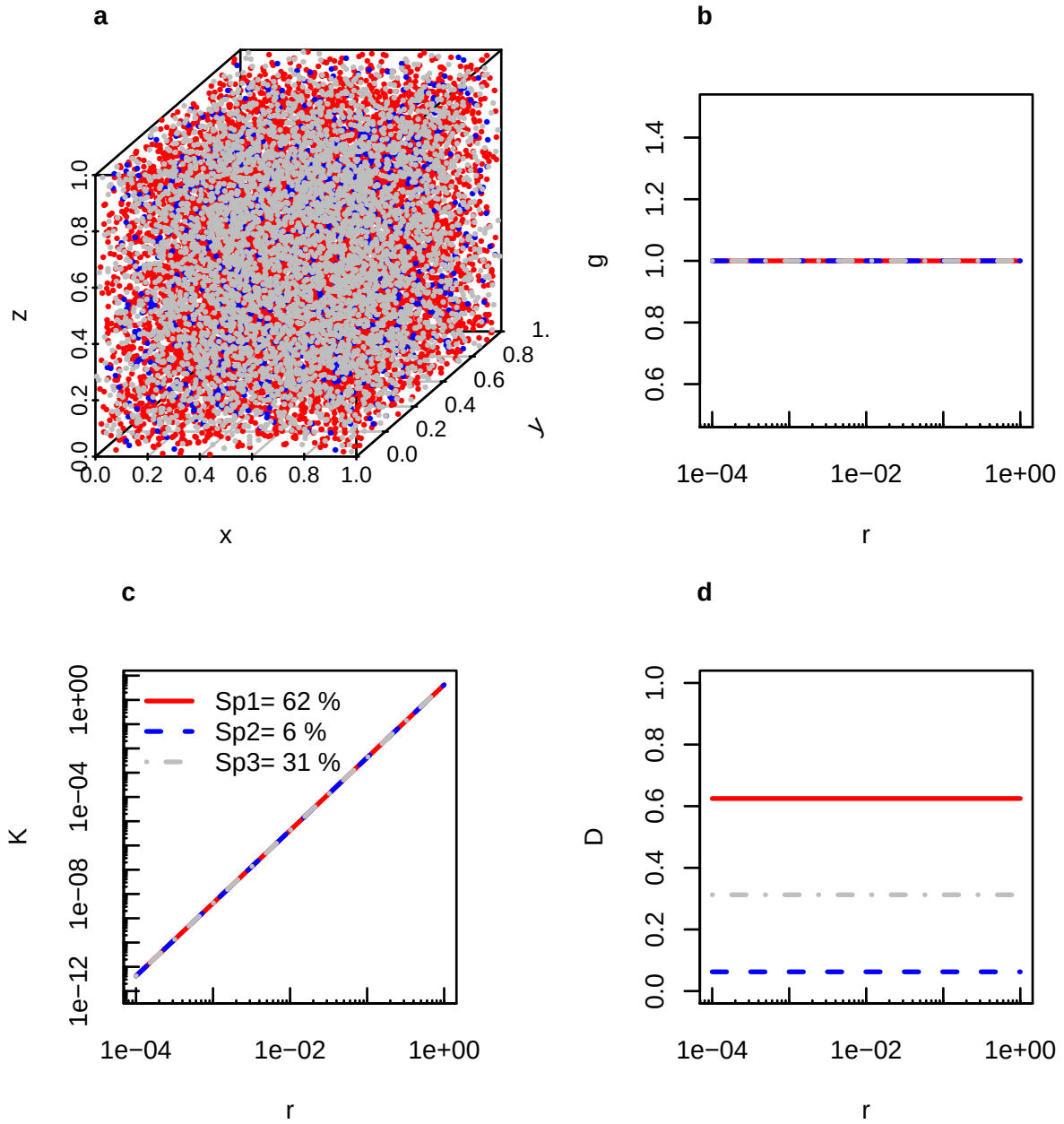

Figure 4.S1: Example of spatial distribution (a) and theoretical pair correlation function (b), Ripley's K-function (c) and dominance index (d) for a Poisson point process in 3-species communities with different intensities (10000 cm^{-3}, 1000 cm^{-3}, 5000 cm^{-3}; proportions in the community are given in the figure).

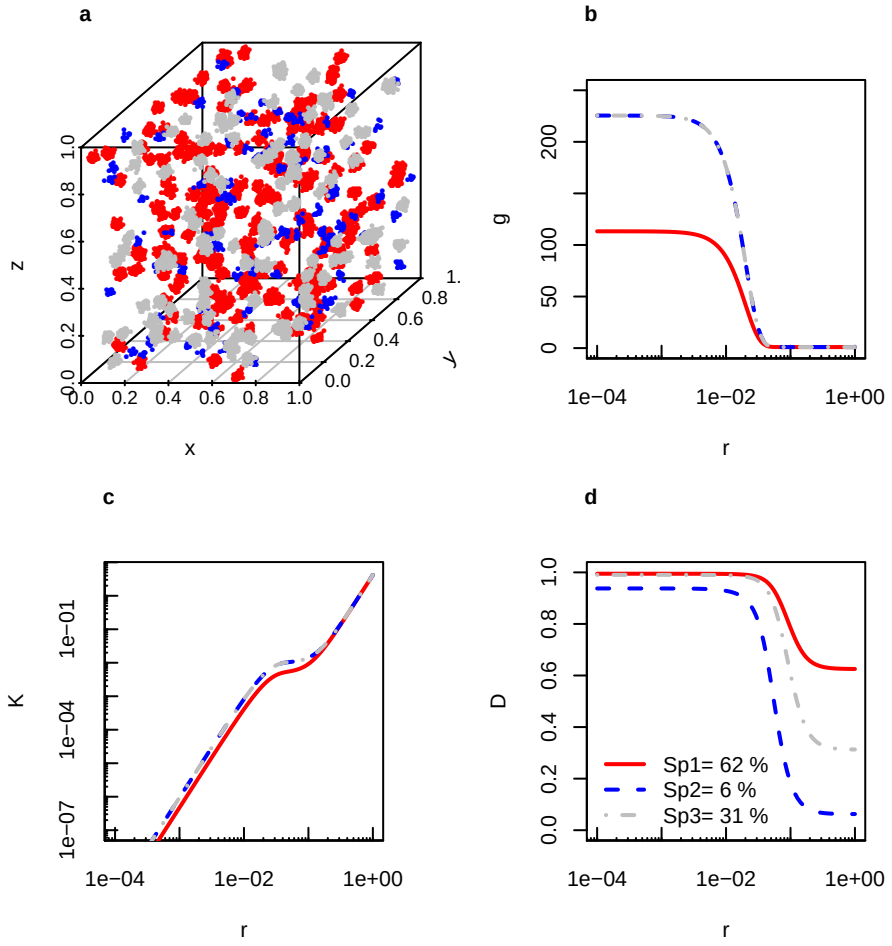

Figure 4.S2: Example of spatial distribution (a) and theoretical pair correlation function (b), Ripley's K-function (c) and dominance index (d) for a Thomas point process in 3-species communities with different parent intensities (200 cm^{-3}, 100 cm^{-3}, 100 cm^{-3}), and different children per parent intensities (50, 10, 50; final proportions in the communities are given in the figure), with $\sigma = 0.01$.

4.S3 Convergence in time of the spatial characteristics of the BBM

The theoretical formulas of g, K and $\mathcal{D}$ can be used to study the behaviour of the BBM. In the absence of advection, convergence cannot be reached in a reasonable timeframe: even a week is not long enough for the steady-state solution to be reached (see blue line in Fig. 4.S3). However, the population-at-equilibrium hypothesis that we use cannot hold for such a long amount of time, which led us to use the time-dependent formulas shown in Eqs. 4.5, 4.9 and 4.12 in the main text.

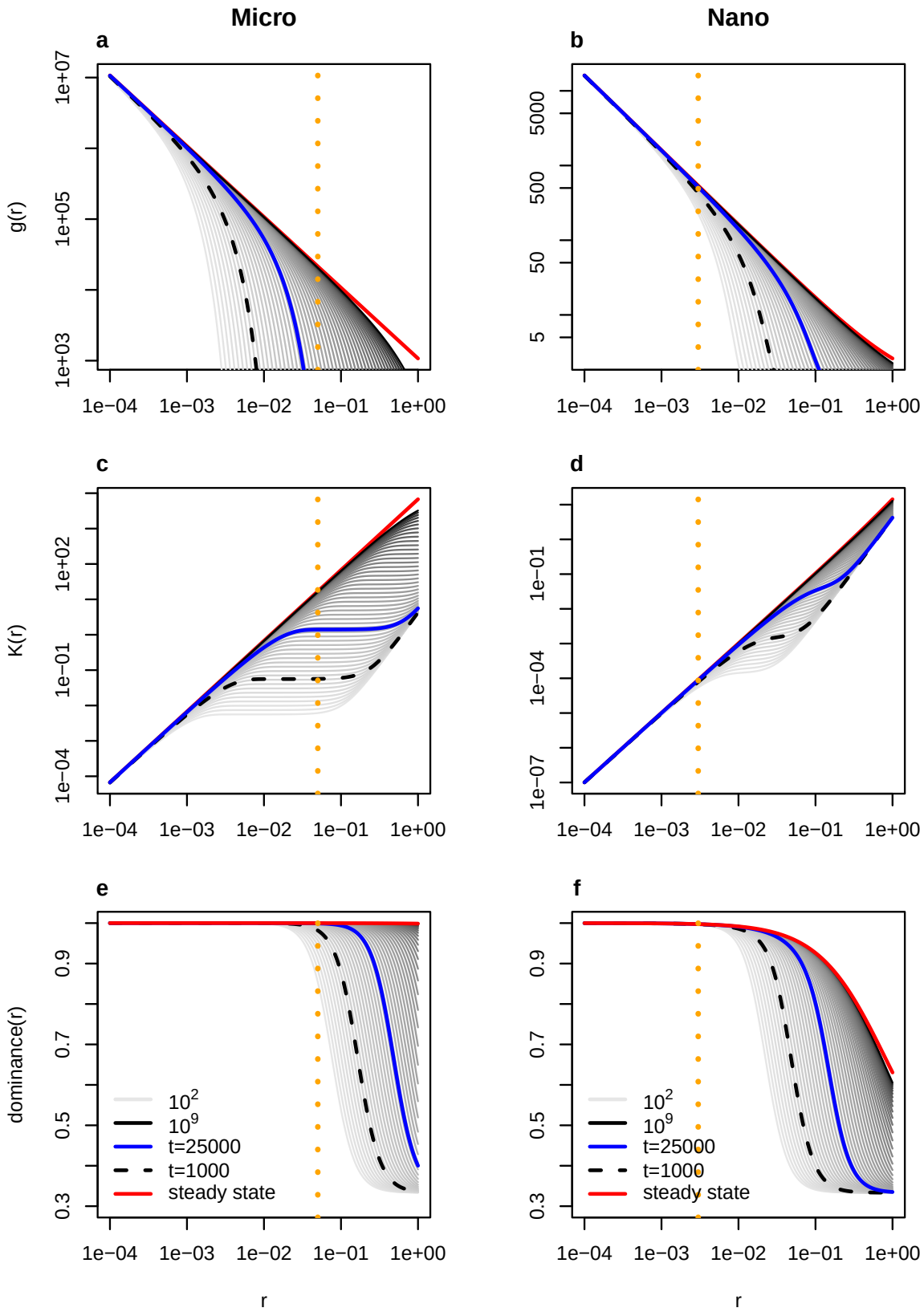

Figure 4.S3: (Caption next page.)

Figure 4.S3: Intraspecific pair correlation function (a, b), Ripley's K-function (c,d) and dominance index (e,f) as a function of distance (in cm) for microphytoplankton and nanophytoplankton in the absence of advection, for a single species in a 3-species community with an even abundance distribution. Shorter timeframes are shown with light grey lines while longer ones are shown with darker shades. The theoretical value at steady state is shown in red. The duration currently used in the simulations ($t = 1000\tau$) is shown with dashed black lines. A duration corresponding to a week is shown with solid blue lines. Dotted orange lines correspond to the distance threshold for interaction.

In a similar fashion, we can show with the dominance index (Fig. 4.S4) the progressive clustering of individuals with time when advection is absent, and compare it to the steady state this time *with* advection. We see that even after a short period of time ($t = 100\tau$), the dominance index without advection is larger than with advection, and this spatial aggregation only grows with time in absence of turbulent advection.

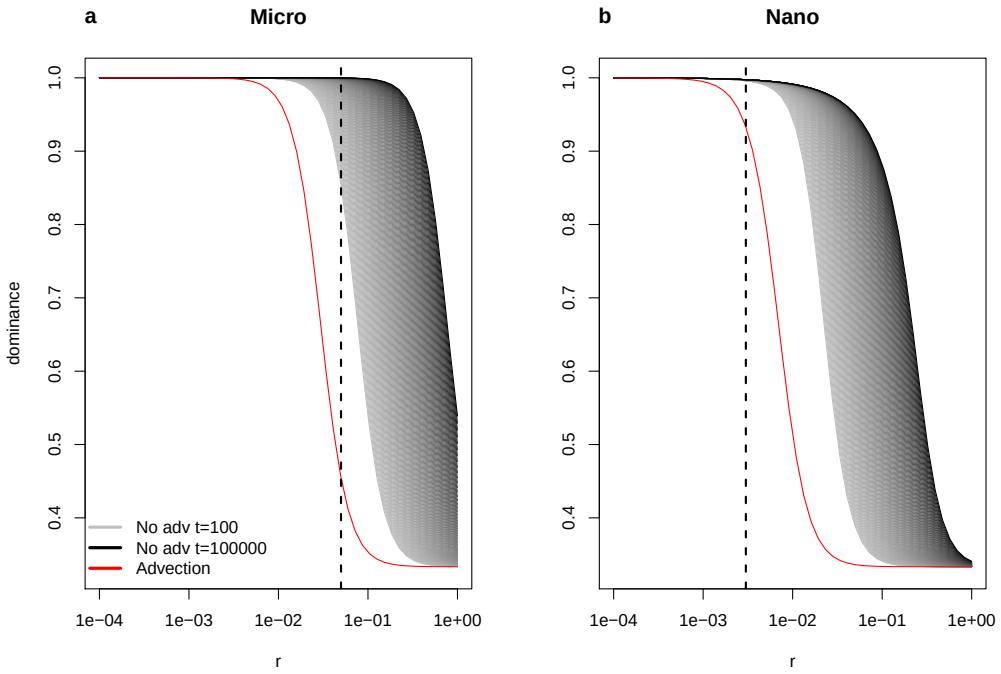

Figure 4.S4: Theoretical dominance indices as a function of the distance (in cm) from a particle of a given species, for a microphytoplankton (a) and nanophytoplankton (b) 3-species community with an even abundance distribution, with (red line) and without (grey to black lines, with darker lines for longer simulations) advection. The vertical dashed line corresponds to the distance threshold for interaction.

4.S4 Computation of the pair correlation function and Ripley's K-function

The algorithm for the pcf computation was adapted from the function `pcf3est` in spatstat 2.2-0 [22] and slightly modified to compute the interspecific pcf (i.e., the pcf for marked point processes).

The pcf estimate $\hat{g}_{ij}(r)$ is computed via the use of the Epanechnikov kernel κ_E with bandwidth δ, i.e.

$$\hat{g}_{ij}(r) = \frac{1}{\hat{C}_i}\frac{1}{\hat{C}_j}\frac{1}{4\pi r^2}\sum_{k\in i}\sum_{l\in j}\kappa_E(r-||\boldsymbol{x}_k-\boldsymbol{x}_l||)w(\boldsymbol{x}_k,\boldsymbol{x}_l) \tag{4.S11}$$

where $w(\boldsymbol{x}_k,\boldsymbol{x}_l)$ is the Ohser translation correction estimator [256] and the kernel is defined as follow.

$$\kappa_E(x) = \begin{cases} \frac{3}{4\delta}\left(1-\frac{x^2}{\delta^2}\right) & \text{for } -\delta \leq x \leq \delta \\ 0 & \text{otherwise.} \end{cases} \tag{4.S12}$$

The estimate $\hat{g}_{ij}(r)$ is therefore very sensitive to the bandwidth: if it is too small, the estimate is noisy and may even be missing several pairs of points; if it is too large, the smoothing might be so important that values are strongly underestimated. In spatstat 2.2-0 [22], the bandwidth default value is $\delta = 0.26C^{-1/3}$. The pcf computation function was first tested on standard distributions (with the default bandwidth), then on the Brownian Bug Model (with different bandwidths, see Fig. 4.S7).

Estimates of the Ripley's K-function were also computed with the Ohser translation correction estimator but did not require any kernel smoothing. The same computation could be done using wrapped-around boundary conditions (for pcf estimation; for simulation we always consider periodic boundary conditions).

Standard point processes

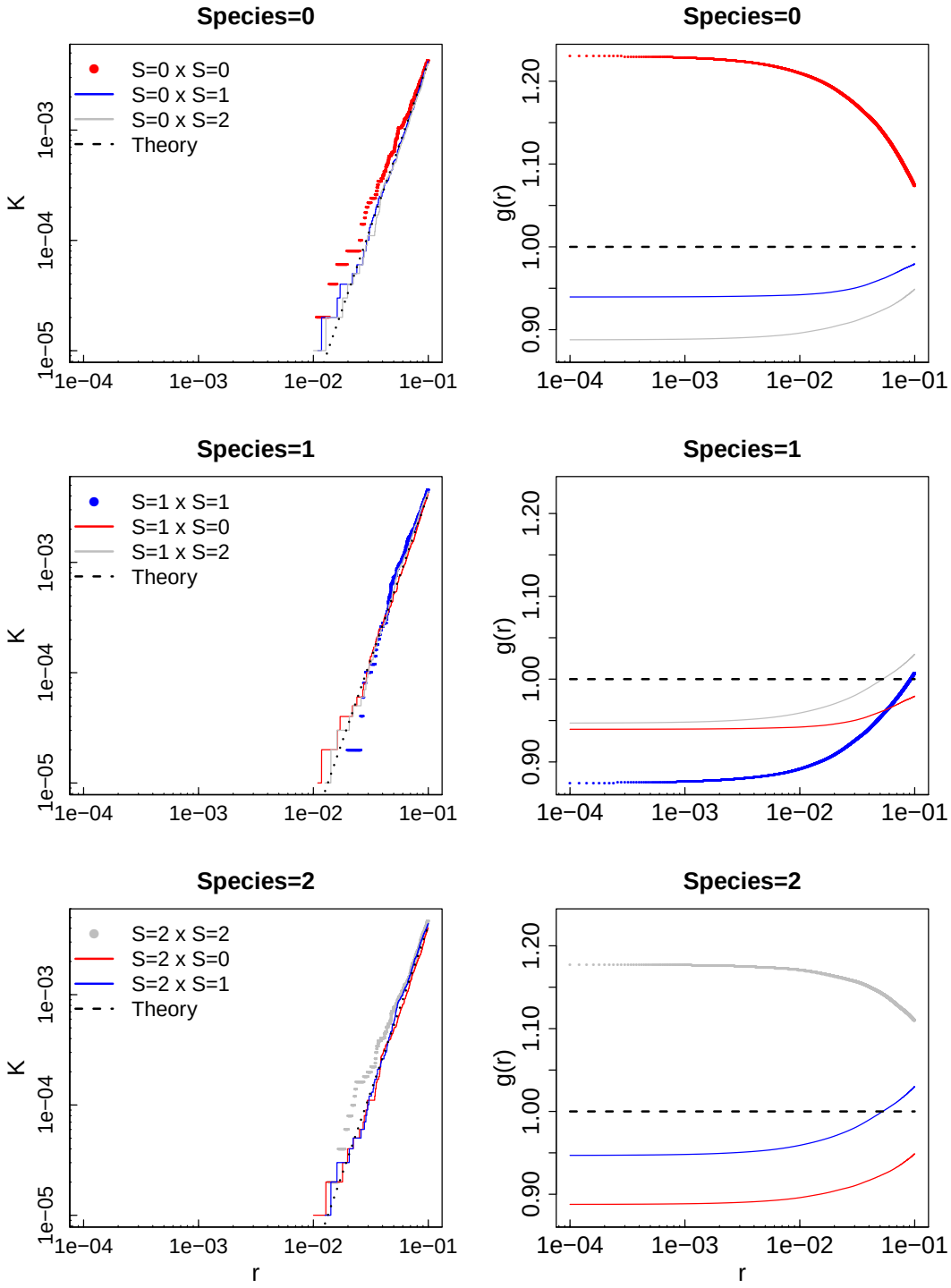

Figure 4.S5: Intra- and inter-specific Ripley's K-function and pair correlation function values as a function of distance (in cm) for 3 species following a Poisson process with intensity 10 cm^{-3}, in a volume of 1000 cm^3. Values computed from our simulations (circles and solid lines for intra- and interspecific values, respectively) are compared with theoretical formulas (dotted lines). Note that theoretical values are the same for intra and interspecific indices for the Poisson distribution. Colors correspond to the different species (red for species 0, blue for species 1, grey for species 2).

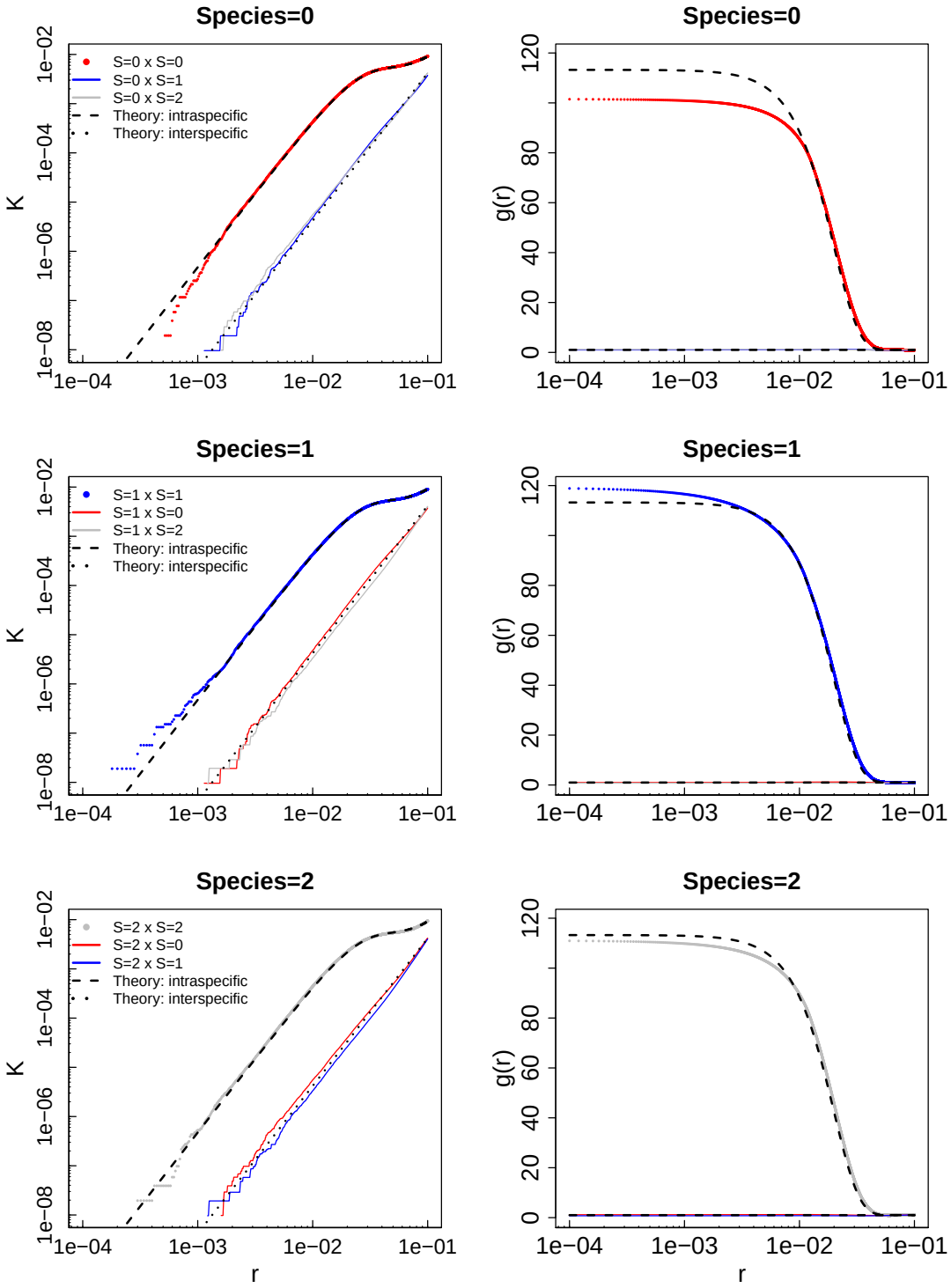

Figure 4.S6: Intra- and inter-specific Ripley's K-function and pair correlation function values as a function of distance (in cm) for 3 species following a Thomas process with parent intensity $C_p = 200$ cm^{-3}, number of children per parent $N_c = 50$, in a volume of 1 cm^3, $\sigma = 0.01$ and $\delta \approx 0.012$. Values computed from our simulations (circles and solid lines for intra- and interspecific values, respectively) are compared with theoretical formulas (dashed and dotted lines for intra- and interspecific values, respectively). Colors correspond to the different species (red for species 0, blue for species 1, grey for species 2).

Brownian Bug Model

While the pcf was one of the first indices that we intended to use, we quickly realized that the combination of the large range of distances we wanted to explore (from 10^{-4} to 1 cm) and the low density of individuals, at least for microphytoplankton, made the estimation difficult as the choice of the bandwidth was critical. We give an example of the sensitivity of the pcf computation to the bandwidth below (Fig. 4.S7).

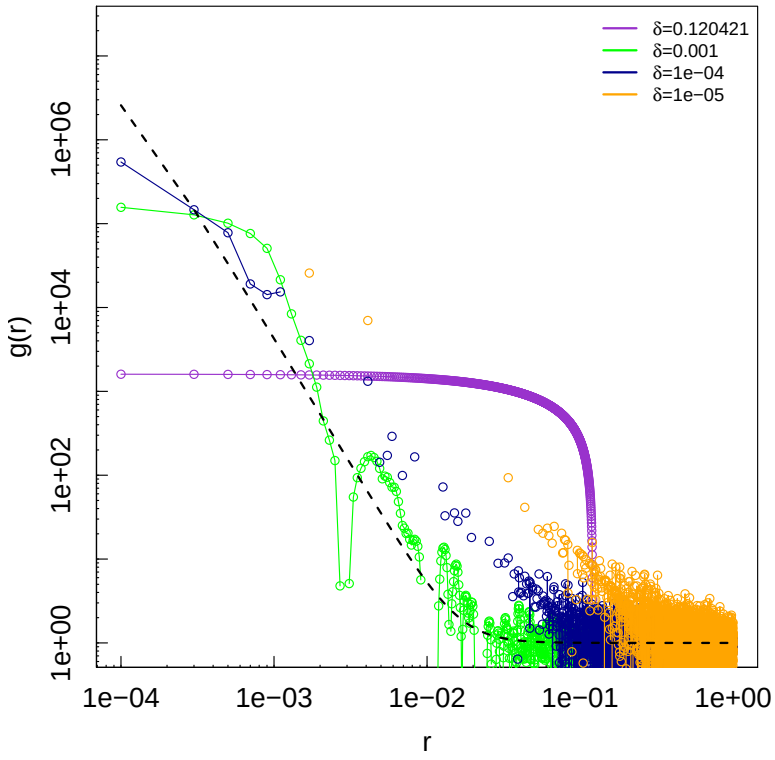

Figure 4.S7: Intraspecific pair correlation function as a function of distance (in cm) computed for the Brownian bug model with microphytoplankton individuals, after 1000 time steps, with different values of the bandwidth δ. The dashed line indicates the theoretical pcf.

We decided, realizing that it would be very challenging to obtain a non-noisy pcf curve matching the theoretical expectation, to focus on Ripley's K-function whose cumulative nature helps the estimation process, which enabled us to compute the dominance index without having to calibrate a bandwidth beforehand.

4.S5 Spatial distributions

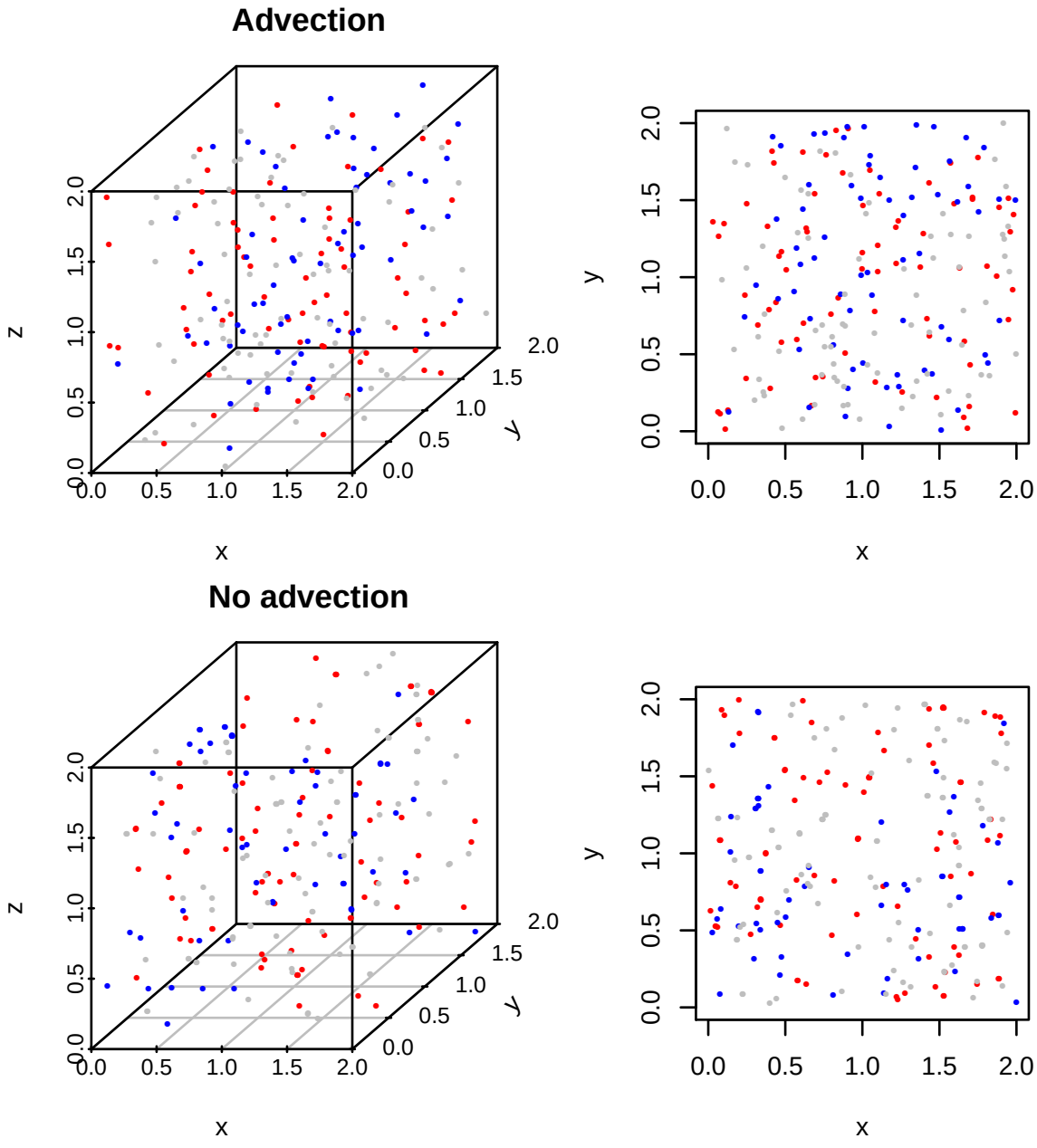

Figure 4.S8: Spatial distributions of a 3-species community of microphytoplankton with and without advection with density $C = 10$ cells cm^{-3} after 1000 time steps. Each color corresponds to a different species. On the left-hand side, only a zoom on a $2 \times 2 \times 2$ cm^3 cube is shown, and its projection on the x-y plane is shown on the right-hand side.

4.S6 Minimum distances between individuals

Theory

One of the reasons why estimating $K(r)$, and even more so $g(r)$, is difficult is that for small distances (below 10^{-2}), we can find very few observations of pairs of points. In order to better understand at which distance ranges we should expect some estimation difficulties, we wanted to compute the minimum expected distance between points (distance to the nearest neighbour, DNN) when they are uniformly distributed.

In d dimensions, the probability distribution of the distance r to the nearest-neighbour follows $f(r) = db_d C r^{d-1} \exp(-b_d r^d C))$ where C is the intensity of the process. If we want to find the distribution of the minimum DNN between n realized points of a Poisson process with intensity C, we can write

$$\begin{aligned} \mathbb{P}(\min(R_1, ..., R_n) > r) &= \mathbb{P}(R_1 > r, ..., R_n > r) \\ &= \Pi_i^n \mathbb{P}(R_i > r) \\ &= \Pi_i^n \exp(-b_d r^d C) \\ &= \exp(-b_d r^d \Sigma_i^n C). \end{aligned} \tag{4.S13}$$

We can then conclude that the distribution of the minimum distance follows the same distribution as the DNN, but with intensity nC.

Clark & Evans [80] show that a variable with probability distribution (with notations changed to fit our own) $f(r) = \frac{dC\pi^{d/2}r^{d-1}}{\Gamma(\frac{d}{2}+1)} \exp(-\frac{C\pi^{d/2}r^d}{\Gamma(\frac{d}{2}+1)}) = dCb_d r^{d-1} \exp(-Cb_d r^d)$ has an expected value of $\mu_d = \frac{\left(\Gamma(\frac{d}{2}+1)\right)^{1/d}\Gamma(\frac{1}{d}+1)}{C^{1/d}\pi^{1/2}}$.

With intensity nC, we can write $\frac{\left(\Gamma(\frac{d}{2}+1)\right)^{1/d}\Gamma(\frac{1}{d}+1)}{(nC)^{1/d}\pi^{1/2}}$.

In three dimensions,

$$\begin{aligned} \mu_d &= (nC)^{-1/3} \frac{\left(\Gamma(\frac{3}{2}+1)\right)^{1/3}\Gamma(\frac{1}{3}+1)}{\pi^{1/2}} \\ &= (nC)^{-1/3} \left(\tfrac{3}{2}\Gamma(3/2)\right)^{1/3} \tfrac{1}{3}\Gamma(1/3)\tfrac{1}{\pi^{1/2}} \\ &\approx 0.554 \tfrac{1}{(nC)^{1/3}}. \end{aligned} \tag{4.S14}$$

This needs to be taken into account when defining C. For microphytoplankton, using $C = 10$ cells cm^{-3} and $n \approx 10^4$, the expected smallest NN distance for a uniform distribution is 1.2×10^{-2} cm. For nanophytoplankton, using $C = 10^3$ cells cm^{-3} and $n \approx 10^4$, it is reduced to 2.6×10^{-3} cm.

Simulations

We can compute the simulated distance to the nearest neighbour in the BBM and compare it to what we should obtain with a uniform distribution: the simulated mean distance to the

nearest organism, regardless of its species, is close to the expected value for a uniform spatial distribution, but the minimum distance to a conspecific is much lower than expected (Fig. 4.S9 for microphytoplankton, results are similar for nanophytoplankton).

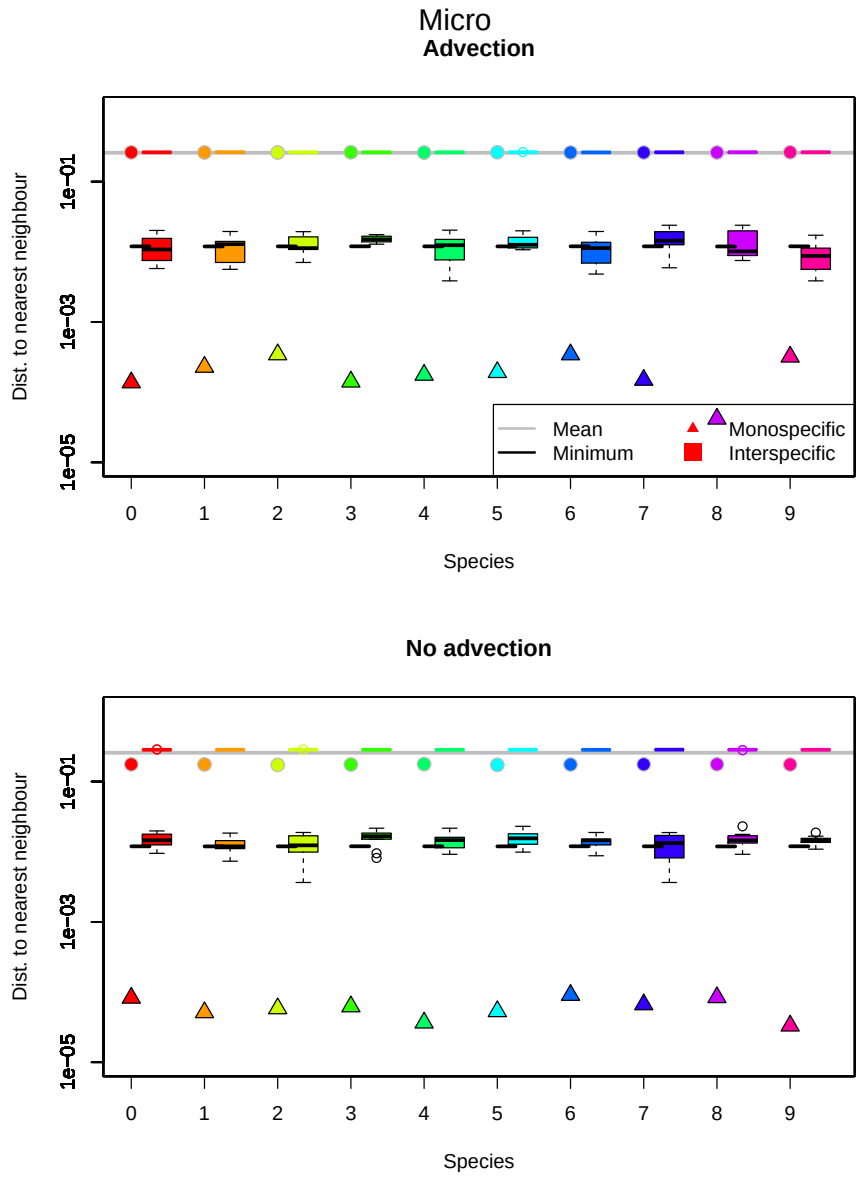

Figure 4.S9: Mean and minimum distance (in cm) to the nearest neighbour for 10 microphytoplankton species with density $C = 10$ cells cm^{-3}, with and without advection, after 1000 time steps, compared to predictions for a uniform distribution. Horizontal lines show the average distance to the nearest neighbour (grey line) and the expected minimum distance to the nearest neighbour with the actual number of realizations (black line). Circles and triangles represent mean and minimum distance to a conspecific, respectively. Boxplot corresponds to the distribution of mean (grey outlines) and minimum (black outlines) distances to a heterospecific. Colors correspond to different species.

Relationship with densities

In the case of a uniform distribution, an increase in density leads to a decrease in distance to the nearest neighbour (eq. 4.S14). Mechanically, we can indeed expect that if the number of particles increases within the same volume, they likely get closer to each other. We confirm that this is also the case in the Brownian Bug Model.

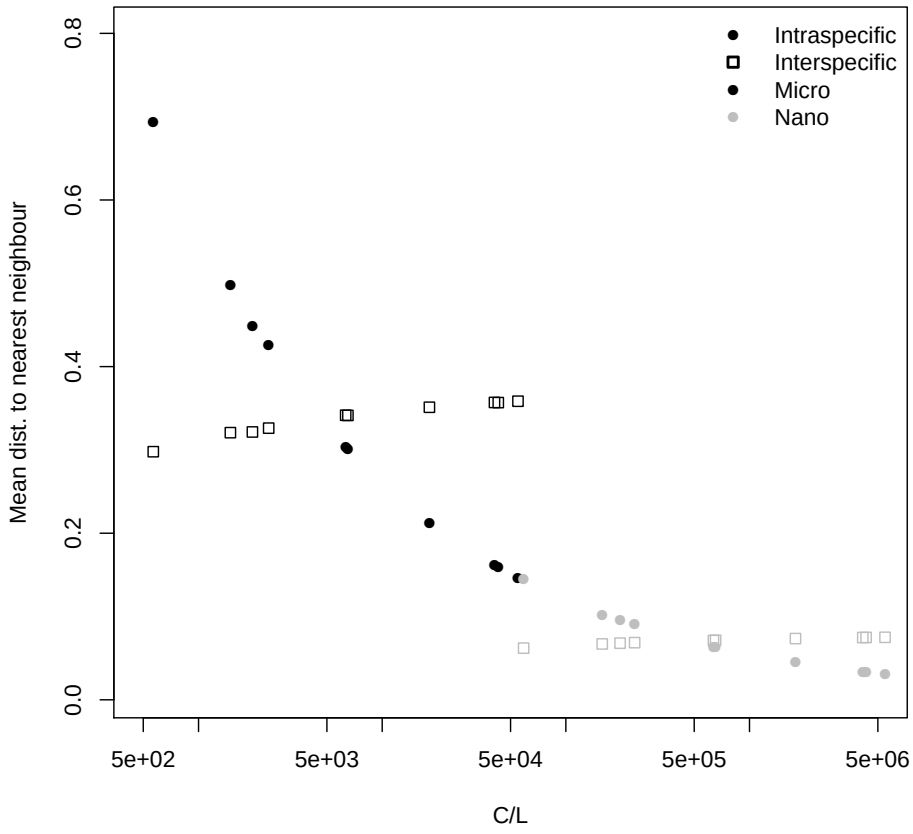

Figure 4.S10: Mean distance (in cm) to the nearest conspecific (filled circle) or heterospecific (empty square) as a function of density in the environment for both microphytoplankton (black) and nanophytoplankton (grey) communities with a skewed abundance distribution, in the presence of advection.

4.S7 Sensitivity to the computation of the advection parameter

To compute the value of the maximum velocity of an organism in our model at the Kolmogorov scale, we used the formula $\mathrm{Re} = U/k\nu \approx 1$ where k is the smallest wavenumber associated with turbulence. However, we could compute the Reynolds number with another, slightly different formula, using the equivalent sphere diameter (L_v) of our system (eq. 4.S15). In this case,

$\mathrm{Re} = UL_v/\nu$ and

$$\begin{array}{rcrcl} & \frac{4}{3}\pi\left(\frac{L_v}{2}\right)^3 & = & L_c^3 & \\ \Leftrightarrow & L_v & = & 2L_c\left(\frac{3}{4\pi}\right)^{1/3} & \\ \Leftrightarrow & L_v & = & 1.24 \text{ cm}. & \end{array} \tag{4.S15}$$

If we use $U \approx \nu/L_v$, $U \approx 8.1\times10^{-5}$ m s^{-1}. Using $U\tau/3 = 0.5$ cm, we have $\tau = 185$ s $= 2.1\times10^{-3}$ d. This means that $\gamma = 164$ d^{-1}. As could be expected, when the flow velocity decreases, mixing decreases (Fig. 4.S11).

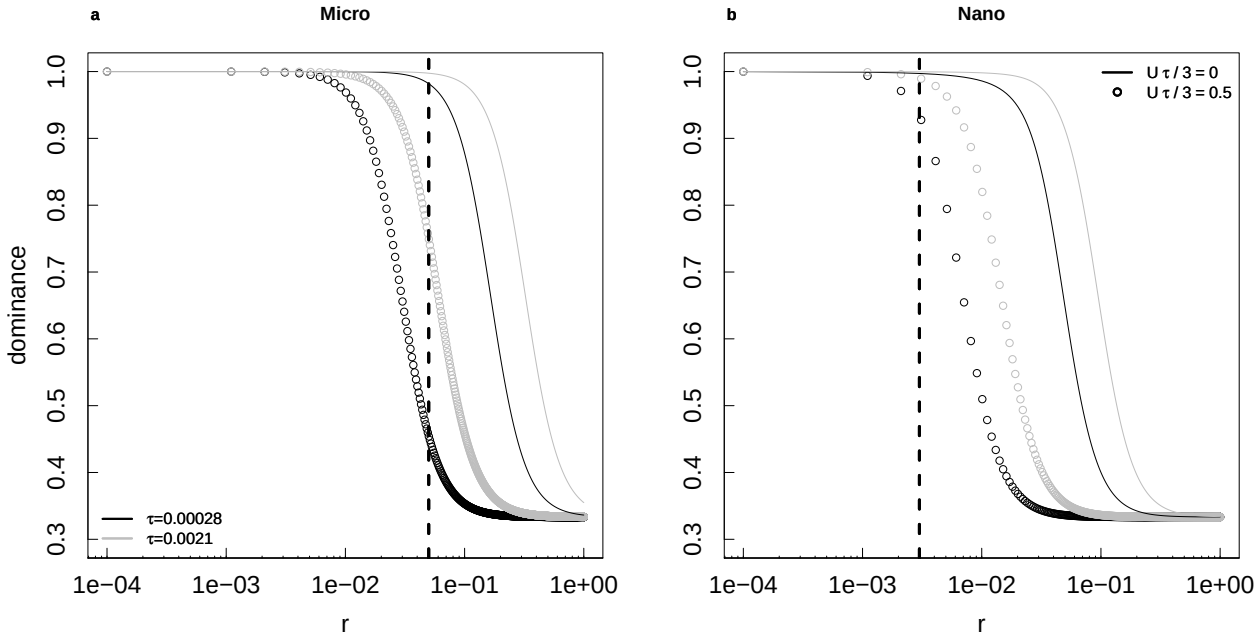

Figure 4.S11: Dominance indices as a function of distance (in cm) for one species in a microphytoplankton (a) and nanophytoplankton (b) 3-species community with even distributions after 1000 timesteps with (circles) and without (lines) advection for different durations of the timesteps, with reference parameters (black) and lower flow velocity (grey).

4.S8 Relationship between the dominance index, relative strengths of interactions and coexistence in Lotka-Volterra models

In this section, we evaluate the potential relationship between local dominance, ratios of intra-to-interspecific interaction strengths observed at the population level, and their consequences in a spatial, dynamic point process Lotka-Volterra framework. Let us define μ_i the average growth rate of a typical individual of species i. Assuming it is linearly dependent on the abundances of the individual's conspecifics and heterospecifics within a neighbourhood of radius r,

$$\begin{array}{rcl} \mu_i(r,t) & = & b_i + \beta_{ii}K_{ii}(r,t)C_i(t) + \beta_{io}\Sigma_{j\neq i}K_{ij}(r,t)C_j(t) \\ & = & b_i + \beta_{ii}M_{ii}(r,t) + \beta_{io}M_{io}(r,t) \end{array} \tag{4.S16}$$

where b_i is the intrinsic individual growth rate, and β_{ii}/β_{io} are individual-level interaction coefficients with a conspecific / heterospecific, respectively. $C_j(t)K_{ij}(r,t)$ is the expected number of individuals of species j around a typical individual of species i within a sphere of radius r centered on the focal individual at time t. $M_{ii}(r,t) = K_{ii}(r,t)C_i(t)$ and $M_{io}(r,t) = \Sigma_{j\neq i}K_{ij}(r,t)C_j(t)$.

If we are close to an equilibrium at the local scale, and intra- and interspecific interaction strengths are equal *at the individual level* ($\beta_{ii} = \beta_{io} = \beta$), on average,

$$b_i + \beta M_{ii}(r,t) + \beta M_{io}(r,t) \approx 0. \tag{4.S17}$$

We can now focus on the dynamics at the community level. We denote α_{ij} the interactions at population level (by contrast to β_{ij} at individual level, as in 334). Assuming that all *interspecific* population-level interactions are similar to one another so that $\alpha_{ij} = \alpha_{io}$ if $j \neq i$, the per capita growth rate

$$\frac{1}{C_i}\frac{dC_i(t)}{dt} = b_i + \alpha_{ii}C_i(t) + \alpha_{io}C_o(t) \approx 0. \tag{4.S18}$$

We can then write the approximate equalities $\alpha_{ii} \approx \beta\frac{M_{ii}(r,t)}{C_i(t)}$ and $\alpha_{io} \approx \beta\frac{M_{io}(r,t)}{C_o(t)}$ by matching Eqs. 4.S16 and 4.S18, and obtain the population-level interaction strength ratio

$$\frac{\alpha_{io}}{\alpha_{ii}} \approx \frac{M_{io}(r,t)}{M_{ii}(r,t)}\frac{C_i(t)}{C_o(t)}. \tag{4.S19}$$

Using the formulation for the dominance index

$$\begin{aligned} \mathcal{D}_i(r,t) &= \frac{M_{ii}(r,t)}{M_{ii}(r,t)+M_{io}(r,t)} \\ \Leftrightarrow \frac{M_{io}(r,t)}{M_{ii}(r,t)} &= \frac{(1-\mathcal{D}_i(r,t))}{\mathcal{D}_i(r,t)}. \end{aligned} \tag{4.S20}$$

Thus the population-level interaction strength ratio can be written out as a function of the dominance index and of the ratio of conspecific to heterospecific density:

$$\frac{\alpha_{io}}{\alpha_{ii}} \approx \frac{(1-\mathcal{D}_i(r,t))}{\mathcal{D}_i(r,t)}\frac{C_i(t)}{C_o(t)}. \tag{4.S21}$$

Let us first focus on microphytoplankton in a 3-species community with an even distribution of abundances. We know that $\mathcal{D}(d_{\text{threshold}}) \approx 0.4$ at equilibrium. In this case, $\frac{(1-\mathcal{D}_i(r,t))}{\mathcal{D}_i(r,t)}\frac{C_i(t)}{C_o(t)} = 0.75$. For nanophytoplankton, $\mathcal{D}(d_{\text{threshold}}) \approx 0.9$, thus $\frac{\alpha_{io}}{\alpha_{ii}} \approx 0.06$. Both ratios of population-level interaction strength are below 1, in spite of the $\beta_{ii} = \beta_{io} = \beta$ assumption, and therefore meet a necessary condition for diversity maintenance in a Lotka-Volterra model. Similar calculations for the 10-species communities, combining small dominance indices to low average concentrations, lead to $\frac{\alpha_{io}}{\alpha_{ii}} \ll 1$, compatible with coexistence [28].

Chapter 5

Conclusion and perspectives

The models built in this thesis were intended to alleviate some of the common hypotheses used in classical phytoplankton community dynamics models, i.e., competitive-only interactions, single life-stage demography and homogeneous spatial distribution. These assumptions were not challenged all at once, leading to two separate models with different levels of precision in the life cycle and state variable (population or individual), as well as different time and spatial scales. In both cases, we endeavoured to model a semi-realistic biodiversity by describing more than two or three species, as is often done in other coexistence models. While these smaller models are helpful to decompose diversity maintenance mechanisms, they may nevertheless miss phenomena emerging from the increased dimensionality we described. In Chapter 2, we used a classical, mean-field, community dynamics model but added a dormant stage to the phytoplankton life cycle, as well as facilitative interactions. We showed that the presence of a dormant stage enabled specialists to maintain and allowed more species to coexist in adverse environmental conditions. Facilitation, on the contrary, tended to destabilize communities. We then replicated the results of the Brownian Bug Model, an individual-based model of phytoplankton in a turbulent and viscous environment at the microscale, for a single species and in two dimensions in Chapter 3, to later develop a three-dimension, multispecies version in Chapter 4. This model focused on relative positions of organisms in the smallest possible eddy in a typical oceanic environment, but did not explicitly represent interactions or life stages. Spatial heterogeneity emerging from local reproduction and limited-range movement gave us some insight into the way phytoplankton distribution could partially explain coexistence. As is always the case, these models are far from definitive depictions of reality. In this section, we discuss some of the other, 'real-life' mechanisms that are absent from this thesis but could have a significant effect on model behaviours and our understanding of coexistence in phytoplankton communities.

Natural enemies Phytoplankton populations exist within a complex food web, which is often taken as a major driver in NPZ models. The community dynamics model developed with a dormant stage in Chapter 2 only models phytoplankton, i.e., presents only one trophic level. Intra and interspecific regulation by competition for nutrients and by natural enemies are implicitly included in 'apparent' interaction strengths. However, a slightly more mechanistic model could be considered, as apparent interactions are the result of different processes, including consumption / exploitation by natural enemies. Predation, for instance, can regulate most abundant species via the 'kill-the-winner' mechanism [14]. In addition to the exploiters

usually modeled (mainly zooplankton, sometimes larger filter-feeders), viruses are also part of phytoplankton natural enemies, are numerous in marine environments [60] and can affect population dynamics through host lysis. They interact with other trophic levels by diverting part of the phytoplankton biomass from upper trophic levels (e.g., zooplankton) to dissolved organic matter (a mechanism called 'viral shunt' [335]). The contrasting functioning of viruses and zooplankton may lead to different effects on the observed co-variation of phytoplankton abundances. Viruses have a faster growth rate than zooplankton ($\approx$ days vs weeks), and therefore a shorter delay in their response to phytoplankton abundance increase. They also tend to be specific to certain hosts [63] and may therefore have a higher effect on apparent intraspecific regulation while zooplankton, with their generalist profile, may have a similar impact on all phytoplankton species. A more mechanistic model could allow us to explore the effects of these differences on phytoplankton.

Exploiters can also be implicitly modeled by introducing delays in phytoplankton dynamics in the seed bank model. Considering this modification raises two main questions: (1) how impactful is exploitation on apparent interactions between organisms? and (2) what is the effect of the exploiter's specificity? The two ends of the exploitation spectrum (though this is a somewhat exaggerated presentation), virus infection and zooplankton predation, offer a good framework to study these questions. The answer to the first question, the effect of exploitation on observed interactions, determines the way parameters can be modified in the model. The effect of natural enemies can indeed be described in at least two ways: (1a) they can directly define interactions between organisms, i.e., interactions are considered to emerge from exploitation only, which could be observed, for example, in a nutrient-rich environment, or (1b) they can alter loss rate, i.e., interactions are due to other processes, such as competition or auxotrophy, and phytoplankton biomass is lost to exploitation in addition to natural mortality. In both cases, either strength of interactions or loss rate would be modified by the value of phytoplankton in the past, that is the value which has actually impacted zooplankton or virus abundances. Thus, N_{t+1} depends both on N_t (the current phytoplankton population) and $N_{t-\tau}$ (which determines the kill rate of the exploiter), with τ depending on the speed of the dynamics of the exploiter (τ being smaller for viruses than for zooplankton). Natural enemies can also be (2a) specialists or (2b) generalists. In the second case, an exploiter can consume all species in the community and is therefore sensitive to the total abundance of the community $N_{t,T} = \Sigma_j N_{t,j}$. We can therefore propose approximate formulas for four different models accounting for exploitation: interaction strength or loss rate of the focus species i depends either on $N_{t-\tau,i}$ (specialist exploiter) or $N_{t-\tau,T}$ (generalist exploiter).

Remember that, in model II of the seed bank model (Eq. 2.3),

$$N_{t+h,i} = \frac{\exp(r_i(T))N_{t,i}}{1 + \sum_{j \in \mathbb{C},\mathbb{F}} \frac{a.N_{t,j}}{H_{ij}+N_{t,j}}} - lN_{t,i} \tag{5.1}$$

where $\sum_{j\in\mathbb{C},\mathbb{F}} \frac{a_.}{H_{ij}+N_{t,j}} N_{t,j}$ quantifies the saturating effect of competitive $\mathbb{C}$ and facilitative $\mathbb{F}$ interactions on the growth rate of species i, and l is the total loss-rate.

If interactions are only related to species-specific parasitism,

$$N_{t+h,i} = \frac{\exp(r_i(T))N_{t,i}}{1 + \sum_{j\in\mathbb{C},\mathbb{F},j\neq i} \frac{a_. N_{t,j}}{H_{ij}+N_{t,j}} + \frac{a_C}{H_{ii}+N_{t,i}} N_{t-\tau,i}} - lN_{t,i}. \tag{5.2}$$

If interactions are only related to generalist predation,

$$N_{t+h,i} = \frac{\exp(r_i(T))N_{t,i}}{1 + \frac{a_C}{H+\sum_j N_{t,j}} N_T(t-\tau)} - lN_{t,i}. \tag{5.3}$$

If exploitation impacts the loss rate, a new formulation can be considered, i.e.,

$$l(t, t-\tau) = \begin{cases} \frac{l_{\max} N_{t-\tau,i}}{H_l+N_{t-\tau,i}} & \text{for species } i\text{-specialists} \\ \frac{l_{\max} N_{t-\tau,T}}{H_l+N_{t-\tau,T}} & \text{for generalists} \end{cases}. \tag{5.4}$$

The delayed models are meant to express that a more exploitation-driven view of interactions may have led to different representations than those of Chapter 2. Such models would require to search for new values of interaction coefficients. As predation can lead to apparent facilitation [1], this might be of interest in the community we modeled based on observations, where up to 70% of apparent interactions were facilitative ([271], provided in Appendix A). Without actually running simulations, however, it is difficult to predict if the populations would be more resistant to facilitation increase than in the case we presented in Chapter 2. Dependence on delayed abundance could also accentuate the phytoplankton cycles that are currently observed. While we do think dynamics would be modified and give us information on the different effects of exploiters' dynamics, we do not expect an increase or decrease in community resistance to harsh environmental conditions.

There already exists a variation of a NPZ-model which takes into account nutrients, natural enemies (viruses and zooplankton) and cysts [122]. This model confirms that interactions between (grazing and lysis of phytoplankton) and within (competition for nutrients) trophic levels regulate the blooming events (duration and intensity) which characterize phytoplankton. It also indicates that viruses control phytoplankton succession and that cysts provide a refuge for phytoplankton types which would otherwise go extinct due to parasitism. Oscillations and chaotic behaviours emerge in a constant environment as endogenous fluctuations take over. However, as is usually the case in this type of model, the phytoplankton community only comprises three functional types (with only one type being able to form cysts), therefore does not give much information on the way closely related species can survive together in a diverse community. It may be useful in the future to compare the results of these two models, as we use 11 species, all with a seed bank but with slightly different thermal niches. We could thus study different resistances to environmental conditions in addition to endogenous (exploitation-

related) fluctuations.

The spatial component of this thesis gives a mostly physics-based depiction of organisms in an aquatic environment. Spatial distributions are the product of passive displacements of individuals. While this offers a first approach of phytoplankton distributions at the scale of a few individuals, this lacks many biological traits of phytoplankton, some of which could have a great impact on spatial distribution. We propose that the version of the Brownian Bug Model presented here could be seen as a point of comparison (a 'null' model) to test the effect of other mechanisms.

Coloniality The first trait that should be taken into account when modeling phytoplankton aggregation is their ability to combine in colonies, defined in [136] as protist bodies "consisting of a group of cells held together by mucilage or cell wall material". Colonies often take the shape of a filament, or 'chain'. Approximately 25-30% of microphytoplankton (diatoms and dinoflagellates) and nanophytoplankton taxa are colonial ([315, 281], but this number reached 70% in experimental cultivations of 63 species of coastal diatoms [61]). Chains increase the apparent size of phytoplankters, sometimes by several orders of magnitude (an extreme –and rare– example is the 50 cm-long tube formed by $\approx$ 30 µm-long *Navicula thallodes* [196]). Colonies seem to be most abundant during blooms [189], but studies of environmental conditions encouraging chain formation do not always concur. Predation and nutrient availability are among the most common factors mentioned in relation with coloniality. The size increase resulting from aggregation in colonies is often thought to reduce consumption by copepods but it also increases the predation risk as detection is facilitated [35, 293]. Some species may therefore decrease their chain length in response to grazer cues [44, 49], which partially contradicts the "defence against predation" hypothesis (but see [218], which summarizes some observations of grazer-induced colony formation). Nutrient availability may also be a driver of chain formation: diffusive nutrient transport is reduced around a chain when compared to a solitary cell [263, 35], which makes long chain formation more likely in nutrient replete conditions. This is true in experiments but not necessarily in the field [316], and some observations either show no relationship between the number of chains and nutrient concentrations [189], or longer chains in depleted nutrient conditions in the field [311]. In addition to diffusive transport, turbulence around a chain can also affect advective transport of nutrients. Strong turbulence may balance the limitation in diffusive transport and enable chain maintenance even with low nutrient concentrations [93]. Due to their size, colonies occupy a slightly different hydrodynamics regime than single cells (Fig 1.1), which may encourage coexistence of single life and colonial forms [223]. Finally, chain formation has an effect on the group buoyancy, even if it is not clearly resolved: chains may sink faster due to their size, but their sinking speed is slower than expected for single cells of the same volume [285].

Colony formation is therefore a key process in phytoplankton ecology. By encouraging

aggregation, it likely increases intraspecific competition for nutrients, except when turbulence is strong enough to balance out diffusion limitation. However, aggregation partially protects the individuals from other loss mechanisms, due to predation (note that it might not be the case for viruses, which likely benefit from clustered hosts). In our model, adding a mechanism of chain formation would likely increase intraspecific aggregation while not necessarily affecting interspecific distances. It would thus augment the dominance of a colonial species at scales corresponding to the expected chain length. We expect difficulties in modeling this process, as chains only extend in one dimension. Heterospecifics can easily get close to the cells along the other two dimensions, and partly balance intraspecific aggregation along the chain. Turbulence would have to be modified around these chains, and the diffusivity of individuals in a colony should be reduced, at least along the chain, due to their stickiness (chain flexibility should actually interact with turbulence, but this would require much more changes in the model, maybe switching to a different representation of the diatom, from points to solids, e.g., [247, 344]).

Motility Motility, sometimes modeled in other IBMs, is another important feature of phytoplankton. In our model, we followed the simplest, classical idea that microphytoplankton organisms are passively transported in their environment. However, many observations show that this is not the case, as species have evolved numerous ways to regulate their movements (e.g., gliding, swimming, regulation of their buoyancy). Motile diatoms and dinoflagellates can represent more than half of the taxa observed in the field [315, 281]. Movements can be motivated by photosensitivity [83], or the presence of toxic elements [84]. It can also be coupled to coloniality, enabling organisms to come together. Phytoplankters can indeed sense and affect each other's movement. As mentioned in Section 1.3.3, there are some models of organisms interacting in this way, eventually leading to patchy distributions [59, 56]. Buoyancy modification in response to turbulence, which can be considered as a movement regulation along the vertical axis, can also increase clustering of organisms [53, 19]. Motility characteristics were not shown as continuous variables in these models (e.g., no displacement speed variation, but rather 'motile or not'), and therefore the 'communities' considered did not exceed two species. Thus, we have no information on the intensity of intraspecific clustering compared to interspecific distributions in a rich community.

Adding motility to our model would be of interest if taking into account incentives: it can either be by adding resource points, or, implicitly, by biasing the random walk of individuals towards their conspecifics (in the case of colonial organisms). This would likely require computing the local intensity of attraction exerted on every individual by surrounding neighbours, as a function of their distance. Technically, this means that kernels of attraction should be computed at every time step, a highly computation intensive process (see paragraph 'Computational optimization' below). Buoyancy regulation poses another issue by differentiating the vertical dimension from the horizontal dimensions: up to now, dimensions in our model

are commutable without affecting the properties of the system. With these changes, we expect intraspecific aggregation to become stronger, as results from other models always show aggregation due to motility (e.g., [102, 59, 56]).

Phycosphere When trying to infer potential interactions between organisms based on their distance from each other in Chapter 4, we only took into account competition for nutrients. The separation between individuals was compared to the maximum size of the concentration boundary layer in a diffusive environment, and we assumed that competition might happen if distances were lower than this threshold. Phytoplankton organisms, however, are able to modify their direct environment in more than one way; the formation of a phycosphere is one of these fundamental interfaces [304]. The phycosphere is the volume surrounding an individual where exchanges of many chemicals take place between the phytoplankton cell and bacteria. To summarize multifarious interactions comprising commensalism, mutualism, competition, antagonism and parasitism, phytoplankton leak, but also actively exude, organic compounds used by bacteria, which in turn release both detrimental (e.g., algicidal substances) and beneficial compounds (e.g., vitamins). Competition for nutrients does happen, but bacteria can also increase the affinity for limiting nutrients (e.g., iron). Overall, cross-feeding seems to dominate: association with bacteria mostly contributes to phytoplankton growth [304, 183]; conversely, phytoplankton dynamics partly control bacteria community composition in their surroundings [305]. The 'marketplace' [304] that represents the phycosphere is therefore another volume of resources around a phytoplankter that we can take into consideration. The major change from the nutrient depletion zone is that the gradient of resources is reversed (Fig. 5.1): whereas nutrients become more depleted when getting closer to the surface of the cell, the phycosphere resource concentrations likely increase as phytoplankton exude elements, and bacteria are often fixed on the cells. In the case of nutrient depletion, nutrient and phytoplankton concentrations are negatively correlated, but in the phycosphere case, a positive correlation between phytoplankton and resources is likely. A higher density of phytoplankton increases the concentrations of substrates used by bacteria, potentially creating a positive feedback loop. Clusters, either made of close singular cells or colonies, may thus promote growth instead of increasing density-dependent competition. As bacteria community composition seems to partially depend on their host's species, aggregation may be more important for conspecifics than for heterospecifics [40]. The second difference between the nutrient boundary layer and the phycosphere lies in the effect of turbulence. Although strong turbulence increases nutrient uptake [187], it can thin the phycosphere layer to the point of mixing its compounds in the environment [304].

The effect of the phycosphere, though, varies with the size of the organism. The phycosphere volume is partly determined by the volume of the focal individual [177]: larger cells leak more organic matter, which increases the distance of detection by a bacteria, and therefore their encounter rate with phytoplankton. Nanophytoplankton, however, have a nearly negligible phycosphere which remains undetectable by bacteria [304]. This is not to say that

nanophytoplankton phycospheres do not exist, but they might be less dense, and less impactful on their dynamics. The phycosphere is therefore an additional mechanism that differentiates microphytoplankton and nanophytoplankton, and that may explain the divergence between size classes observed in potential interaction distributions.

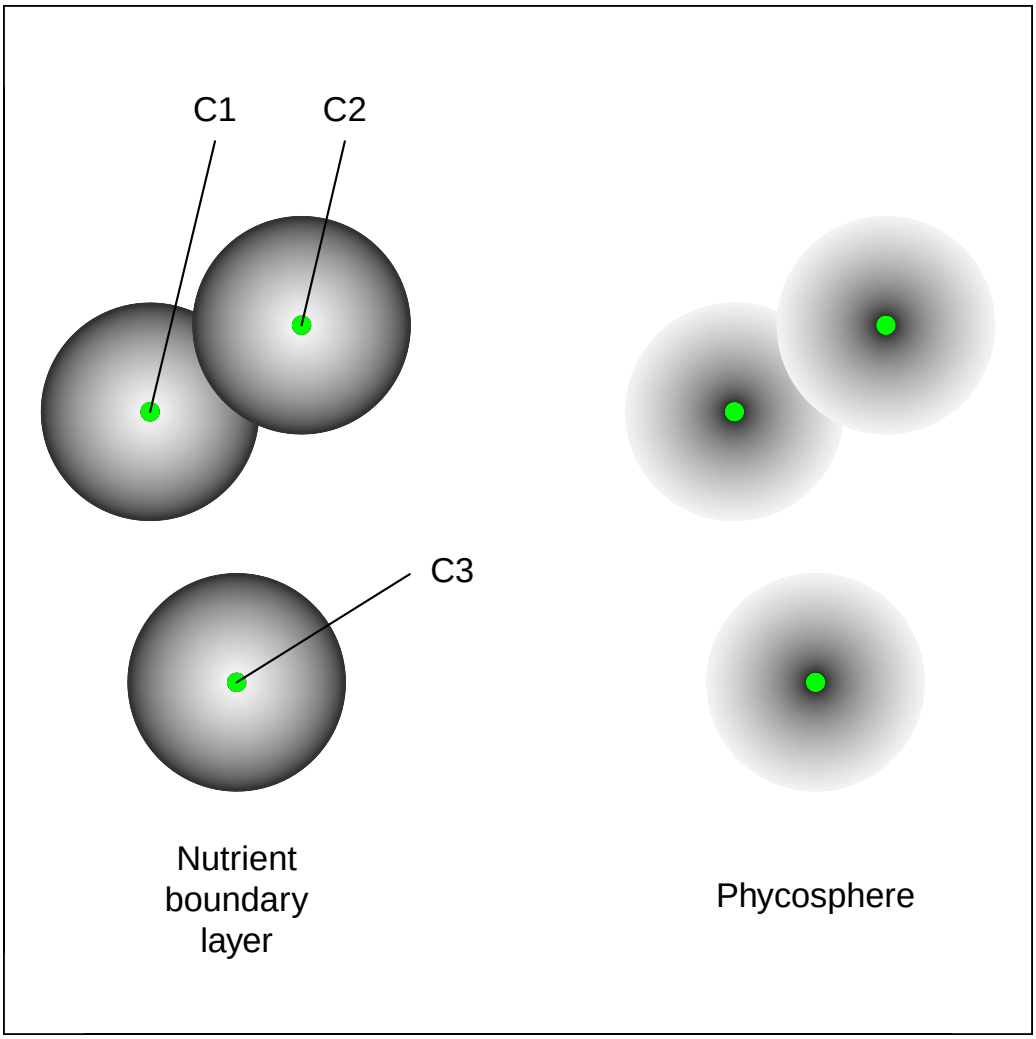

Figure 5.1: Schematic view of phytoplankton cells (green circles) with their spheres of influence (nutrient boundary layer on the left-hand side and phycosphere on the right-hand side). Lighter and darker shades of grey correspond to lower and higher concentrations of elements, respectively. Cells C1 and C2 share their spheres of influence and may be able to affect each other: an interaction is possible. Cell C3 is too far from the other cells to be affected by them.

I have reviewed until now mechanisms that encourage intraspecific aggregation and likely modify intraspecific interactions. Indeed, when motile organisms sense conspecifics nearby, they can be motivated to get closer and form colonies. Increased interactions may ensue, especially as organisms both compete for nutrients and potentially share elements from their phycospheres, which result in both positive and negative effects on organism growth. I now turn to more technical questions: how to define a more complex turbulence structure, and how to overcome potential computational limitations that could hinder model modifications.

Larger scales The BBM was first built with the Kolmogorov scale in mind, and approximates the distribution of individuals within the smallest possible eddy in the ocean. This allowed us to use a single wavenumber model of the flow. However, we could consider bigger volumes and slightly higher Reynolds numbers (as done in [102, 59], with $20 \leq Re \leq 101$). To do so, we could complexify the Pierrehumbert map with N wavenumbers, leading to

$$\begin{array}{ll} x(t+\tau) & = x(t') + \frac{U\tau}{3}\Sigma_{n=1}^{N} a_n \cos\left(k_n y(t') + \phi_n(t)\right) \\ y(t+\tau) & = y(t') + \frac{U\tau}{3}\Sigma_{n=1}^{N} a_n \cos\left(k_n z(t') + \theta_n(t)\right) \\ z(t+\tau) & = z(t') + \frac{U\tau}{3}\Sigma_{n=1}^{N} a_n \cos\left(k_n x(t+\tau) + \psi_n(t)\right) \end{array} . \tag{5.5}$$

In this case, $k_n = n\frac{2\pi}{L_S}$ with L_S corresponding to the smallest wavelength: this means that all k_n would be multiples of the wavenumber we used in Chapters 3 and 4. Amplitudes a_n can be specified as $a_n = k_n^p$, where p is selected to model any spectral slope we want to represent (usually in oceanography, a -5/3 power law). This model remains extremely simplistic, but would at least turn the observation window from a single eddy embedded in the ocean to a collection of eddies of different sizes. Both observations and numerical simulations at the mesoscale ($\approx$ 1-10 km) have shown that eddies could generate a structure in organism sizes: larger organisms tend to be kept outside of the circulation gyre while smaller organisms are concentrated in its center [124]. At the submesoscale ($\approx$ 100 m), differences in inertia and buoyancy allow phytoplankton to remain within an eddy, while denser zooplankton are drawn out of it [339]. Thin layers are also often observed at smaller scales (few cm) along the vertical axis [103]. For such structure in size to be observed, we would however need to add inertial effects to our model, and thus strongly modify its formulation. Above a few cm, the environment is not isotropic anymore: light, a crucial element for photosynthetic organisms, strongly decreases with depth. Attenuation of light by water and self-shading can affect the growth ability of the organisms located at the bottom of the water column. Motility, in this case, is even more important as individuals can regulate their floatability in response to light cues [83]. The model might therefore show a patchy distribution preferentially along the vertical axis; the ability to swim would certainly be a major trait explaining differential distributions and potentially aggregation. Similarly to motility, this would require to differentiate the vertical dimension with the horizontal dimensions.

Computational optimization Increasing the size of the space woud lead to an increase in the number of individuals considered, as would modeling blooming conditions, even in a small environment. In any case, this would drastically increase the computational cost of numerical simulations. Furthermore, when discussing the effect of motility, we mentioned that movements could be determined by computing every distances between pairs of individuals at every time step. A naive algorithm would run in polynomial time – $O(n^2)$ with n the number of individuals –, which can become very time consuming if taking into account 10^6 cells L^{-1} for a microphytoplankton bloom, maybe more for nanophytoplankton. The mechanisms we could

model are therefore also a function of the available computing power and time. There exists ways to overcome this issue, however. One of them would be to try and compute the master equation corresponding to the new model, and work mostly with theoretical formulas. This might prove difficult, as introducing density-dependences involve the introduction of the third moment in the second moment equation, the fourth moment in the third moment equation, and so on; developing moment closures are needed to approximate the final formula [201, 31]. If using numerical simulations, we would advocate for slight changes in the model formulation instead of trying to optimise the computation of the distance matrix at every time step, at the cost of potential gross approximations. Finally, a Lagrangian-Ensemble description could also be considered. In this type of model, the focus 'individual' (hereafter called agent to avoid any confusion with an actual individual corresponding to a single organism) is made of a collection of organisms. The agent is therefore characterized, among other properties, by its size/number of organisms which change over time [338, 265]. In [291], agents are aggregates, which can grow, divide, joint, and move: this allows to compute an aggregate size-spectrum, and a patchy spatial distribution. If modeling a bigger volume, we might want to use this method to see how a community (= an agent) in a given eddy may be moved around by larger eddies, and analyse the corresponding spatial distributions. This would obviously only offer a much coarser model than the one we were aiming for in this thesis.

Links between models The community dynamics model for mean concentrations of interacting species with a daily time step presented in Chapter 2 is extremely different from the individual-based model at the scale of a few seconds in the smallest possible eddy for independent particles described in Chapters 3 and 4. One might then wonder if there is any possible link that can be made between the two models. One lesson that can be drawn from the seed bank model is the fact that populations are partly maintained by immigration from the seed bank. This can help justify the 'population at equilibrium' hypothesis used in Chapter 3 and 4. Conversely, the heterogeneity observed in spatial distributions at the microscale can be used to explain why intraspecific interactions can be much stronger than interspecific interactions, which is what we parameterized in the seed bank model. If we were to consider an explicit coupling or, at least, some interactions between models, we would need to work on the spatio-temporal differences between models. Currently, many processes forbid direct coupling. For instance, the seed bank turn-over is much longer than the time frame we study in the spatial model: it would not make sense to introduce explicit spatial movements through seed formation and immigration in the IBM. Also, the tidal exchanges in the seed bank model (and the corresponding changes in turbulence) modify the equilibrium hypothesis of the spatial model. For now, we can only inform one model with the other, but they cannot interact directly.

We hope nonetheless that these models will encourage others to challenge some of the hypotheses of classical phytoplankton models, as we did, and to build upon our results to analyse new coexistence mechanisms, at their specific spatial and timescales.

Appendix A

Strong self-regulation and widespread facilitative interactions in phytoplankton communities

Coralie Picoche[1] & Frédéric Barraquand[1,*]
[1]Institute of Mathematics of Bordeaux, University of Bordeaux & CNRS
[*]corresponding author: frederic.barraquand@u-bordeaux.fr

Abstract

1. The persistence of phytoplanktonic diversity in spite of competition for basic resources has long been a source of wonder and inspiration to ecologists. To sort out, among the many coexistence mechanisms suggested by theory and experiments, which ones actually maintain diversity in natural ecosystems, long-term field studies are paramount.
2. We analysed a large dataset of phytoplankton abundance time series using dynamic, multivariate autoregressive models. Phytoplankton was counted and identified down to the genus level, every two weeks over twenty years, at ten sites along the French coastline. Multivariate autoregressive models allowed to estimate biotic interaction networks, while also accounting for abiotic variables that may drive part of the phytoplankton fluctuations. We then analysed the ratio of intra- to inter-taxa interactions (a measure of niche differentiation), the frequency of negative vs positive interactions, and how stability metrics (both at the network and genus level) relate to network complexity and genus self-regulation or abundance.
3. We showed that a strong self-regulation, with competition strength within a taxon (genus) an order of magnitude higher than between taxa, was present in all phytoplanktonic interaction networks. This much stronger intragenus competition suggests that niche differentiation - rather than neutrality - is commonplace in phytoplankton. Furthermore, interaction networks were dominated by positive net effects between phytoplanktonic taxa (on average, more than 50% of interactions were positive). While network stability (*sensu* resilience) was unrelated to complexity measures, we unveiled links between self-regulation, intergenera interaction strengths and abundance. The less common taxa tend to be more strongly self-regulated and can therefore maintain in spite of competition with more abundant ones.
4. We demonstrate that strong niche differentiation, widespread facilitation between phytoplanktonic taxa and stabilizing covariances between interaction strengths should be common features of coexisting phytoplankton communities in the field. These are structural

properties that we can expect to emerge from plausible mechanistic models of phytoplankton communities. We discuss mechanisms, such as predation or restricted microscale movement, that are consistent with these findings, which paves the way for further research.

Keywords: phytoplankton; coexistence; facilitation; mutualism; niche theory; time series; networks

Published in Journal of Ecology (2020) doi:10.1111/1365-2745.13410

A.1 Introduction

How species or close genera can coexist together in spite of competition is one of the main puzzles of community ecology, especially for primary producers that seemingly share the same basic resources [168]. Many theoretical studies of competition models have shown that competitive exclusion is likely in those circumstances, unless mechanisms involving spatial or temporal variation are at play [15, 16, 78, 166, 211, 76]. Neutral theory models, assuming that all individuals have equal birth and death rates and exert equal competitive pressure on conspecifics and heterospecifics alike, produce instead a non-equilibrium coexistence maintained by dispersal from a regional pool. They have been proposed as a solution to the puzzle presented by highly diverse communities [162, 288].

However, the evidence gathered from terrestrial plant communities starts to suggest that, in fact, niche rather than neutral processes may be paramount to explain coexistence, with intraspecific competition dwarfing interspecific competition in most cases [2, 4]; see also [329]. Whether these conclusions drawn mostly from studies of terrestrial plants apply to other ecosystems and taxa is currently little known (but see [249]).

Moreover, competition may not be the rule: the meta-analysis by Adler *et al.* [4] reported a large number of facilitative interactions (30%) and several reviews [62, 231, 191] have highlighted that facilitation may be much more widespread than ecologists usually tend to think. Although some theoretical studies suggest that facilitative interactions can be destabilizing (*sensu* resilience) and therefore undermine coexistence in Lotka-Volterra models [87], multiple other modelling [138, 279] and empirical [62, 70] studies have suggested that facilitative interactions can to a large degree benefit coexistence, especially when multiple interaction types are considered simultaneously [243, 132, 279].

Here, we analyse a spatially replicated, long-term community-level dataset, consisting of ten multivariate time series of phytoplankton abundance along the French coastline. We do so using multivariate autoregressive (MAR) models, that allow to estimate interactions between genera. Although many ecological studies focus on interactions between species, competition has been shown experimentally to occur between different genera of phytoplankton [321, 96]. The genus level is also a rather fine taxonomic scale for phytoplankton interaction studies, as most studies are restricted to interactions between different classes or even phyla [175, 142, 137]. Studying

interactions between different genera of phytoplankton therefore both makes empirical sense in light of competition experiments and allows to estimate better-resolved networks. We focus here on genera that belong mostly to diatoms and dinoflagellates. To put our results into a more general context, we then compare our interaction strength estimates to previously published interaction networks produced under the same statistical framework, both in plankton and other empirical systems.

A.2 Material and methods

Sampling methods

All phytoplankton samples were collected by Ifremer coastal laboratories as part of the National Phytoplankton and Phycotoxin Monitoring Network [284]. Since 1987, this monitoring program has required the sampling of 26 sites along the French coastline every 2 weeks within 2 hours of high tide to document both biotic (phytoplankton counts) and abiotic (water temperature, salinity) variables. We focused on sites which had the longest time series. We also excluded time series which had missing data for over 6 months or an average delay between sampling dates above 20 days. This reduced the number of study sites to 10 sites nested within 4 regions (Brittany, Oléron, Arcachon and the Mediterranean Sea; Fig. A.S1 and Table A.S1 in the Supporting Information).

Abiotic variables (temperature, salinity) were measured directly from the boat during the sampling process while water samples for biotic analyses were fixed with a Lugol's solution and examined later. Phytoplankton cells above 20 µm were identified at the lowest possible taxonomic level and counted with the Utermöhl method using an optical microscope [325]. Throughout the years and sites, more than 600 taxa were identified at different taxonomic levels. We aggregated them at the genus (or group of genera when not possible) level based on previous work (Table A.S2; [154, 32]), except for cryptophytes and euglenophytes in Arcachon, which could not be identified below the family level. Although the taxonomic resolution used here may seem coarse in comparison to land plants, it is in fact more refined than 86% of the MAR(1) studies of phytoplankton listed in Table A.S4.

For each region, the MAR(1) analysis focused on the most abundant and most frequently observed genera to avoid most of the gaps in the time series. When gaps did not exceed a month, missing values were linearly interpolated; remaining missing values were replaced by a random number between 0 and half of the lowest observed abundance [143]. Time series are plotted in Fig. A.S2. We tested extensively this and other methods to deal with missing data in a previous publication on a subset of this dataset [32]. All time series were scaled and centered before MAR analyses.

MAR(1) models

Multivariate autoregressive (MAR) models are used to determine the interspecific interactions and abiotic effects shaping a community's dynamics [175]. MAR(1) models are based on a stochastic, discrete-time Gompertz equation which relates the log-abundance of each of the S taxa at time $t+1$ to log-abundances of the whole community at time t, with possible interactions between taxa, and effects of V abiotic variables at time $t+1$. These assumptions are encapsulated in eq. A.1:

$$\mathbf{n}_{t+1} = \mathbf{B}\mathbf{n}_t + \mathbf{C}\mathbf{u}_{t+1} + \mathbf{e}_t, \mathbf{e}_t \sim \mathcal{N}_S(0, \mathbf{Q}) \tag{A.1}$$

where $\mathbf{n}_t$ is the $S \times 1$ vector of log-abundance of phytoplankton taxa, $\mathbf{B}$ is the $S \times S$ community (interaction) matrix, $\mathbf{C}$ is the $S \times V$ environment matrix describing the effects of V variables (stacked in vector $\mathbf{u}_{t+1}$) on growth rates, with $V = 2$ in our case (temperature and salinity). The noise $\mathbf{e}_t$ is a $S \times 1$ vector, following a multivariate normal distribution with a variance-covariance matrix $\mathbf{Q}$. $\mathbf{Q}$ is diagonal and we have previously showed that this parsimonious choice did not affect qualitatively the results [32]. We used the MARSS package v3.9 [159, 158], in R v3.3.2 [280], to estimate parameters with a maximum likelihood procedure.

Our previous analysis of the Arcachon region, for which more covariables were available [32], revealed that hydrodynamics and hydrology had more influence on phytoplankton dynamics than nutrients on the two-week timescale. Because temperature and salinity, in addition to their direct effects, sum up seasonal changes in light and hydrology (salinity is inversely related to freshwater inflow), they represent the two key drivers needed to account for abiotic influences [297]. They are therefore used to summarize the abiotic environment in the remainder of the article.

The analysis of real data in Barraquand *et al.* [32] was complemented by that of simulated data mimicking the study design, which confirmed the ability of MAR(1) models to infer biotic interactions and abiotic forcings. Fitting a more sophisticated model (threshold autoregressive model) did not reveal extra non-linearities or a storage effect in the Arcachon subset of the data [32]. Other aspects of the MAR(1) modelling are likewise quite robust: using two abiotic variables (temperature and salinity) in this study rather than the full set used in Barraquand *et al.* [32] led to almost identical covariate effects and interaction estimates for the Arcachon study sites. Even if some departures from the true data-generating model may not always be detectable through MAR(1) diagnostics (e.g., residuals), the analysis of nonlinear simulations has showed that MAR(1) models are in general robust to nonlinearities if the inference focuses on interaction sign and order of magnitude of model coefficients [71], which is how these models are used here. For ease of interpretation of MAR(1) interaction coefficients, we also highlight how intra- and inter-taxa interaction strengths in a MAR(1) model map to their counterparts in a multispecies Beverton-Holt model, i.e., a discrete-time Lotka-Volterra model [90], in the Supporting Information.

In this study, the number of phytoplankton taxa (S) and the community composition vary slightly between regions but sites share on average 67% of their taxa. In order to have comparable models across sites, we keep the same 2 covariates, water temperature and salinity, that were measured at all study sites. Therefore, the dimension of the dynamical system depends on the (square of the) number of phytoplankton taxa we study, which ranges between 7 (Mediterranean Sea) and 14 (Brittany). The smallest system still requires 63 parameters to be estimated (49 for the 7×7 interaction matrices and 14 for the 7×2 environment matrices) if we consider all possible interactions between taxa. To reduce this dimensionality and remove unnecessary parameters, we built different 'interaction scenarios' based on known phylogenetic information (as suggested in [328, 250]). The null interaction scenario assumed no interaction between genera (diagonal interaction matrix) and was compared to four other interaction scenarios. The first interaction scenario assumed that interactions could only occur between phylogenetically close organisms, i.e., within a class (groups were then diatoms, dinoflagellates, and other phytoplanktonic organisms) while the second interaction scenario further differentiated pennate and centric diatoms. The third interaction scenario considered the reverse hypothesis, that only unrelated organisms could interact (i.e., a diatom could only interact with a dinoflagellate or a cryptophyte, but not with another diatom), and the last interaction scenario did not constrain the interactions at all (full interaction matrix). We selected the best scenario by comparing BIC (Fig. A.S3), which proved to be satisfactory in our previous analyses of both real data and similar simulated datasets ([32], Appendix 2). The second interaction scenario, hereafter called the pennate-centric scenario, had the lowest BIC for all sites (Fig. A.S3). This parsimonious scenario was therefore chosen as the basis for further investigations of network structure.

Analysis of interaction strengths

The interaction matrix obtained from MAR(1) analyses can be used to determine the stability of a discrete-time dynamical system [172, 175]. To investigate stability-complexity relationships, we compared the maximum modulus of the eigenvalues of the pennate/centric matrices for each site to network descriptors. The maximum modulus is analogous to the real part of the leading eigenvalue for continuous time models, and measures resilience while still accounting for some variability properties [172]. However, because most theory on stability-complexity has been developed in continuous time (e.g., [8]), we numerically checked that the maximum modulus of the eigenvalues in a discrete-time interaction matrix and its continuous-time model counterpart yield similar information in the Supporting Information. We then compared this resilience measure to complexity metrics, such as the interaction strength distribution (sign, mean and variance) and weighted connectance [45]. Weighted connectance is a measure of the proportion of realized links compared to all possible links, taking into account the shape of the flux distribution. This metric is adapted to weighted interaction matrices but cannot accommodate for both positive and negative coefficients: we therefore chose to focus on interaction

strength only (absolute values of the coefficients), irrespective of interaction sign. In contrast, mean and variance of the off-diagonal coefficients, which can affect the stability of a community [8], are computed on raw values of the coefficients. Interaction coefficient variance is multiplied by the number of taxa, according to theory [8].

In addition to these network-level metrics, we also computed the average vulnerability (average effect of other taxa on a focal taxon, eq. A.S5) and the average impact (average effect of a focal taxon on other taxa, eq. A.S6), on both raw and absolute values of the interaction coefficients. Such vulnerability and impact scores can be related to in-strength and out-strength in the meta-analysis of Kinlock [191]. We then compared these to the regulation a focal species exerted on itself. Vulnerability computed on raw coefficient values indicates the average effect that can be expected on the growth rate of a taxon from the rest of the community (i.e., is the effect of others mostly positive or negative?), while vulnerability computed on absolute coefficient values characterises the strength of all types of interactions on a taxon (i.e., is a taxon strongly affected by the others?). A similar reasoning applies to the impact score.

Finally, we compared the observed ratio between mean self-regulation (intrataxon interaction strength) and mean intertaxa interaction strength to other published studies based on a MAR(1) model. A list of references is given in Table A.S4. Authors usually reported only coefficients that were significant with a 5% significance level, thus ignoring potentially many weak effects, which we had to set to 0. There are therefore two ways of computing the mean intertaxa interactions, i.e., taking the mean value of all coefficients outside of the matrix diagonal, including zeroes (which decreases the estimated mean intertaxa interaction strength, Fig. A.4), or taking the mean value of statistically significant intertaxa coefficients only (which increases the estimated mean intertaxa interaction strength, Fig. A.S9). We considered both; a detailed description of these different ways to compare intra- and inter-taxa interactions can be found in the Supporting Information.

A.3 Results

Interaction estimates

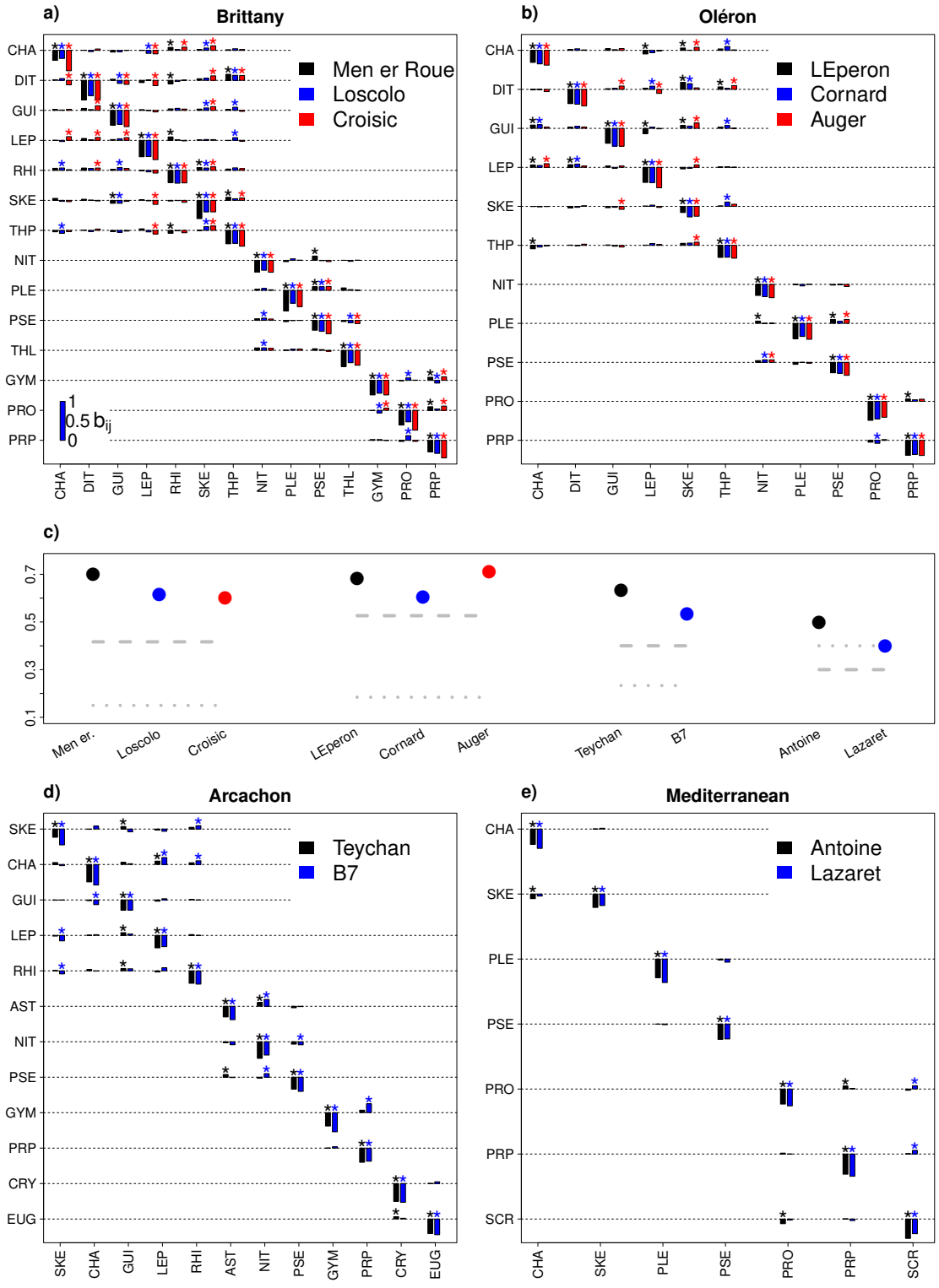

Figure A.1: (Caption next page.)

Figure A.1: **Interaction matrices estimated at 10 sites along the French coastline**. The sites are distributed in 4 regions: (a) Brittany, (b) Oléron, (d) Arcachon, from north to south on the Atlantic coast and (e) Mediterranean (see Supporting Information for a map). Taxon j (in columns) has an effect on taxon i's growth rate (in rows) proportional to the bar height, which corresponds to the $\mathbf{B} - \mathbf{I}$ matrix (community composition in Table A.S2, most parsimonious interaction scenario presented). The scale for the coefficient values is given at the bottom left of panel (a). Coefficients significantly different from 0 ($\alpha = 5\%$) are marked by asterisks (*). The fraction of positive interactions in each matrix is given by points in (c) while the dashed (resp., dotted) line represents the ratio of interactions remaining positive (resp., negative) for all sites of a given region.

Using MAR(1) autoregressive models, we produced interaction matrices [175, 141] – i.e., Jacobian community matrices on the logarithmic abundance scale [175]. Best-fitting models corresponded to a phylogenetically-structured interaction scenario, where interactions only occurred betwen closely related genera (Fig. A.S3). This led to sparse, modular matrices that have two main features. First, we observed a strong self-regulation for all sites (Fig. A.1, diagonal elements of all matrices), a feature that we had previously highlighted in a more detailed analysis on one of the considered study regions [32]. The ratio of mean intragenus to intergenera interaction coefficients varied between 6 and 10, not counting coefficients set to 0 before the estimation process. When we included the zeroes in the interaction matrix in the computation of the intra/inter mean interaction strength (see the Supporting Information for details of that computation), the ratio rose to 21-43. Therefore, intragenus interactions were on average one order of magnitude stronger than intergenera interactions.

Second, although the percentage of facilitative interactions varied among sites (between 40% and 71% of interactions in the selected models), facilitation remained predominant in 9 sites out of 10 (only Lazaret, in the Mediterranean Sea, has 60% negative interactions). Our observational setup being nested, with sites within regions, we could examine whether locally positive interactions remain positive in a regional context: the percentage of consistently positive interactions at the regional level varied between 30% and 53%, higher than the percentage of similarly defined negative interactions (between 15% and 40%), except for sites in the Mediterranean Sea.

We found that the percentage of true mutualism (+/+) was substantial: averaged over all sites, 32% of all interactions were (+/+) while only 12% of them were (-/-), see also Fig. A.S5. The sign correspondence was not always maintained between regions: the only interaction that was non-zero in the 10 sites (CHA/SKE) was mutualistic in Men er Roue only (Brittany) and mixed (+/-) in all other sites. Within the same region, however, interactions measured in different sites tended to keep the same sign. In the 3 sites of Oléron, for instance, there were 4 interactions which remained positive for both taxa involved (CHA/GUI, DIT/GUI, LEP/THP, SKE/THP), 3 of them being also mutualistic in some of the Brittany sites. This contradicts previous observations that mutualistic interactions tend to be more context-dependent than

competitive interactions [72].

Interaction network analysis

The stability (*sensu* resilience, Ives & Carpenter 173) of all interaction matrices was not strongly affected by the percentage of positive interactions or the mean and variance of the interactions between taxa (Fig. A.2). There was a slight increase in stability with weighted connectance, with a drop in eigenvalue modulus for weighted connectances between 0.09 and 0.1. The maximum modulus of the interaction matrix eigenvalues remained between 0.65 and 0.80.

Given that a direct complexity-stability (*sensu* resilience) link was not obvious, we investigated whether the matrix coefficients had some particular structure that could help theoretical ecology to make better null models of joint community dynamics and interactions [180]. Relations between intra- and inter-taxa interactions emerged (Fig. A.3): genera that were more self-regulating also had a higher average vulnerability score. Those two influences are likely to trade-off: a high degree of self-regulation somehow buffers the effect of outside influences on population dynamics. Taxa that were less self-regulating were also more likely to have a stronger effect onto other taxa. As these genera tended to be more abundant (Fig. A.S7), this could be mediated by the average density of a genus. It is important to note, however, that these trends are weak and there is therefore a considerable amount of randomness dominating the interaction matrix: many scenarios of self-regulation vs limitation by others are therefore possible.

Aside from the trade-offs of Fig. A.3, we found no remarkable patterns of covariation between matrix elements (Fig. A.S5) other than a mean-variance scaling of interaction coefficients (Fig. A.S6).

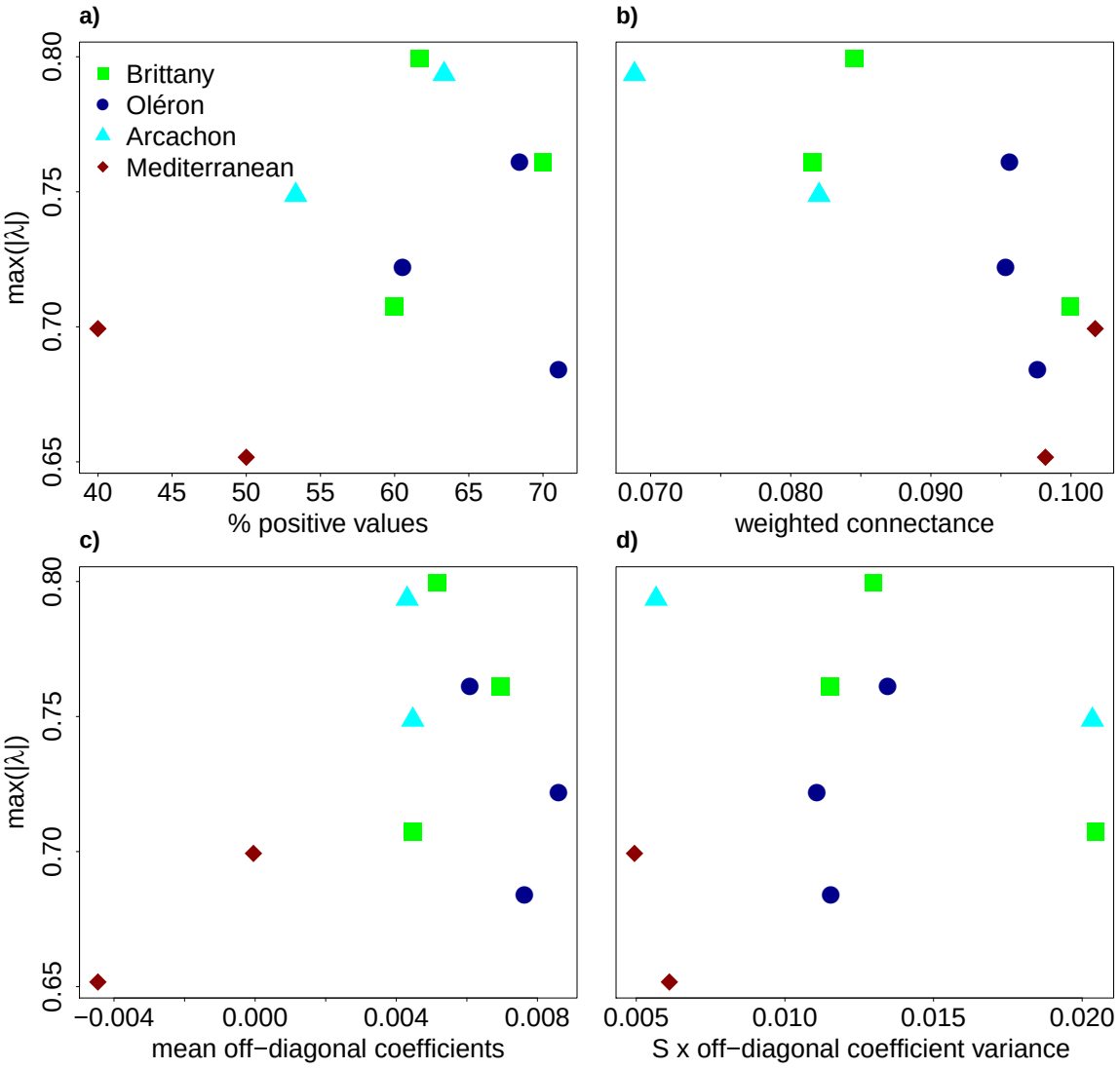

Figure A.2: **Relation between stability and complexity of the interaction networks.** The maximum modulus of the eigenvalues of the interaction matrix **B**, plotted on the y-axis, indicates stability *sensu* resilience. Complexity metrics are (a) the fraction of positive coefficients, (b) weighted connectance, (c) the mean intertaxa coefficient and finally (d) the variance of intertaxa coefficients, scaled by the number of taxa. Off-diagonal coefficient variance is multiplied by the dimension of the network, that is the number of species in the region. Each color or shape corresponds to a given region. The formula for weighted connectance is given in the Supporting Information.

Literature comparison

Finally, we sought to put these results in a broader context by compiling the mean intrataxa vs mean intertaxa interaction estimates of previous MAR(1) studies of long-term observational count data (listed in Table A.S4). We found that the order of magnitude of intra/inter interaction strengths considered here is not particularly above those found for most planktonic systems to which MAR(1) models have been fitted, considering that our systems are relatively

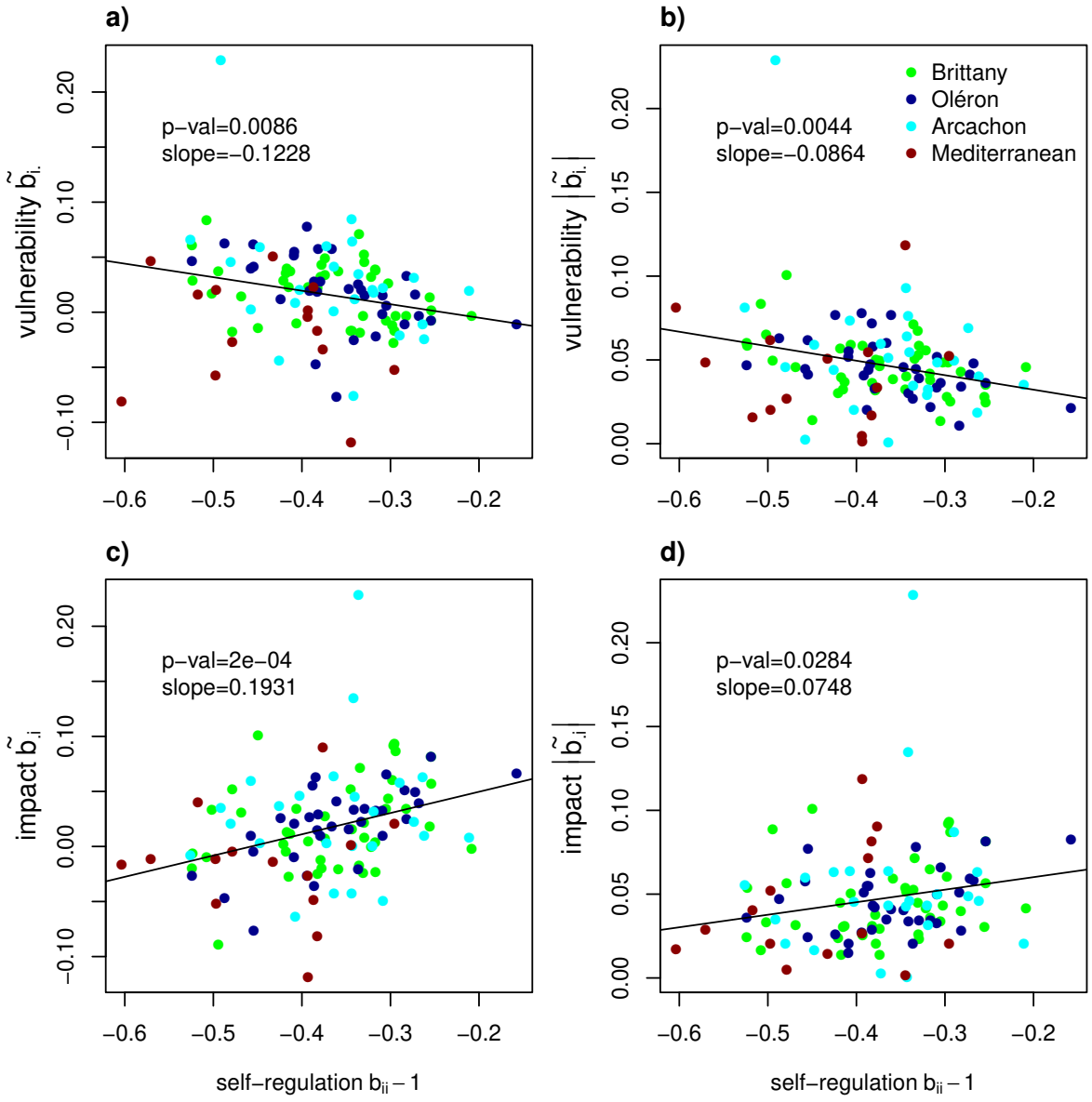

Figure A.3: **Relation between vulnerability/impact and self-regulation.** Average vulnerability (effects of others on the focal taxon growth rate, a and b) and impact (effects of the focal taxon on others' growth rates, c and d), as well as self-regulation, are computed for untransformed (a-c) or absolute (b-d) values of the coefficients of the interaction matrix $(\mathbf{B}-\mathbf{I})$ for the 10 study sites. Each color corresponds to a given region (Fig A.S1). Linear regressions are shown as black lines.

high-dimensional and that the higher the number of taxa, the larger the intraspecific regulation [29]. We included in Fig. A.4 not only plankton studies but also a couple of vertebrate or insect studies on less diverse communities, where interactions are stronger, in order to provide lower bounds for the intra/inter ratio. The conclusion from this comparison seems to be that, unlike small communities that can be tight-knit, any diverse field system of competitors and facilitators has evolved large niche differences making on average intrataxon competition much larger in magnitude than intertaxa interactions.

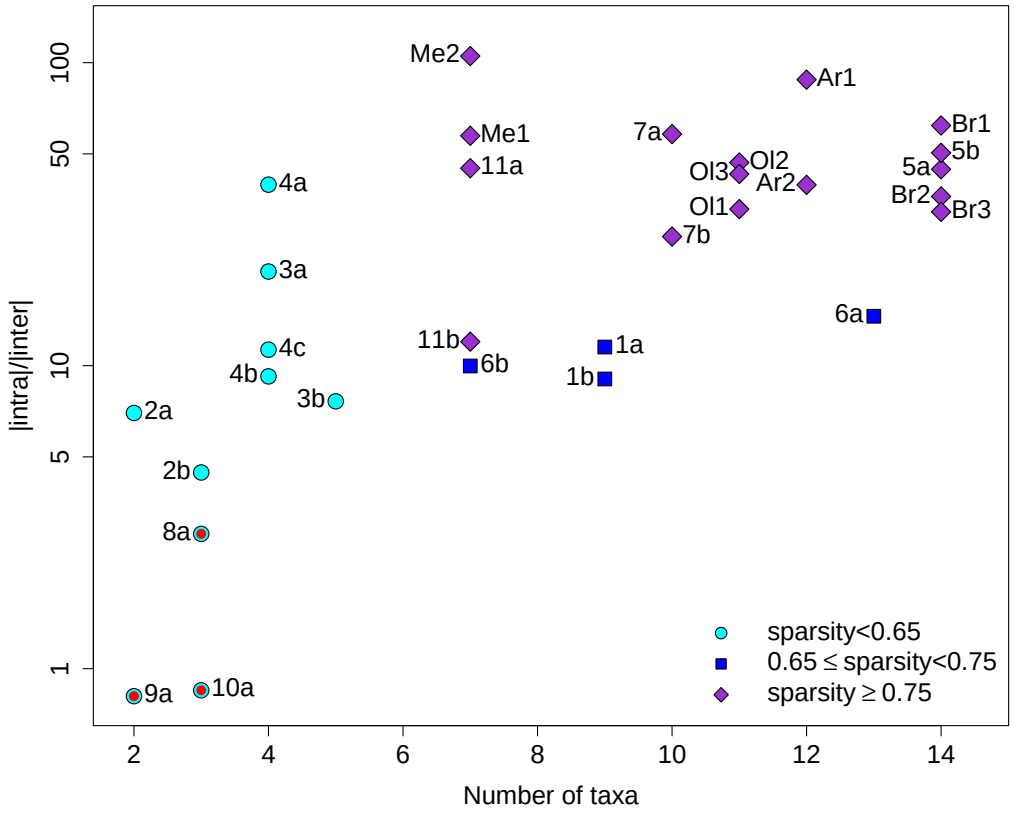

Figure A.4: **Ratios of mean intrataxa to mean intertaxa interaction strength in Multivariate AutoRegressive (MAR) models.** The reference for each study is given in Table A.S4. Codes beginning with letters correspond to the present study (Ar: Arcachon; Ol: Oléron; Br: Brittany; Me: Mediterranean Sea). The symbol color and shape correspond to the sparsity of the interaction matrix (e.g., the proportion of null interactions in the matrix). Red dots correspond to terrestrial and/or low dimension predator-prey systems, giving a lower bound for the intra/inter ratio. Intertaxa interactions were set to 0 when they were not specified in the articles (in most cases, authors removed non-significant interactions at the 5% level; Fig. A.S9 is the same figure taking into account only significant interactions).

A.4 Discussion

Strong self-regulation and facilitation

We found very large niche differences between genera, translating into much higher intragenus than intergenera effects on growth rates, together with a high degree of facilitative net interactions.

The intra/intertaxa interaction strength ratio [206] that we found, from 6–10 to above 20, depending on whether one includes interactions set to zero before the estimation process, could appear very high in light of previous intra/interspecific competition strength estimates of 4 to 5 by Adler *et al.* [4]. Additional estimates using the unconstrained interaction matrix yielded ratios between 8 and 11 depending on the site (Table A.S3 and Fig. A.S8 in the Supporting Information), but weak intertaxa effects are likely to be inflated in the full model. Therefore,

a intra/inter ratio of 10 seems like a conservative estimate. It is twice that of Adler *et al.* [4] who use a different model, i.e., a Lotka-Volterra competition model. We outline how to relate a MAR(1) model to a discrete-time Lotka-Volterra equivalent in the Supporting Information; even though there is a relationship between intra/inter ratios in both models, the relationship is not trivial when abundances vary greatly between species. Hence, to some degree, intra/inter ratios can differ between model frameworks or ways of measuring density-dependencies (e.g., a high measurement error due to using proxies of densities for plants can result in bias in interaction coefficient estimates [98]). However, a ratio intra/inter at least twice larger than the ones previously found may call for other explanations. One could also argue that our high intra/inter ratio arises because we consider the genus as our baseline taxonomic unit, rather than the species. It is logical that niche differentiation increases as one gets up the phylogenetic tree, and that getting down to the species level could slightly decrease that ratio (but see [250], in which phylogenetic closeness decreases competition strength). However, taxonomic resolution is unlikely to be the sole explanation for the high intra/inter ratio of interaction strength found here, for two reasons. First, phytoplankton species belonging to different genera are often found to compete in experiments [321, 320, 96]. In the field-based dataset studied here, the same genera that are considered in experiments are found not to compete (or only weakly), hence there must be some niche differentiation occurring in the field but not in the lab. Second, the only other study that managed to provide MAR(1) estimates down to the species level for phytoplankton, that of Huber & Gaedke [163], provides an intra/interspecific strength ratio similar to ours (point 7a in Fig. A.4). Strong self-regulation seems therefore a genuine feature of field phytoplanktonic communities. We discuss below possible mechanistic interpretations.

Another main finding of our study is the large frequency of positive interactions, with 30% truly mutualistic (+/+) interactions and between 40 and 70% facilitative effects. Although a seasonal environment can generate some positive covariation between taxa, those effects have already been filtered out by the inclusion of our 2 abiotic covariates (Fig. A.S4). The facilitative effects shown here are therefore residual effects, once abiotic trends are accounted for. Between 40 and 70% of facilitation can be compared to the meta-analysis by Adler *et al.* [4] who also found facilitative interactions, but less than here (≈30%). However, Adler *et al.* [4]'s review contains many experiments while the plant literature is replete with field examples of facilitation [62, 231], so that plant facilitation could be more prevalent in the field. At the moment, it is therefore unknown how the predominance of facilitative interactions that we found in phytoplankton compares to facilitation in terrestrial plants. We note that several authors using MAR(1) models previously forbade positive interactions within the same trophic level, so that the fraction of facilitative interactions in plankton cannot be computed from literature-derived MAR(1) estimates.

The large niche differences and facilitative interactions that arise when considering a single trophic level are an emergent property, resulting from hidden effects of resource or predator partitioning/sharing [76]. In our previous publication investigating in detail the Arcachon study

sites [32], we have argued that for phytoplankton, the strong intrataxon density-dependence could arise from effects of natural enemies [148]. Natural enemies could also very well create apparent mutualism between prey species [1, 292]. We believe this to be likely for the present study, given that the study regions (Arcachon, Oléron, Brittany, Mediterranean) have similar predators (zooplankton, e.g., [181, 238, 322]) and parasites (viruses, e.g., [260]; fungi). Though natural enemies are good candidates to explain the observed niche differences and emerging facilitation, one must bear in mind that other known drivers of phytoplankton dynamics such as allelopathy [120], auxotrophy [317] or hydrodynamics [208] can all, in theory, help create different niches and an emerging facilitation (see last subsection of the Discussion). Finally, resources that are usually considered limiting for all species might in fact not always be: Burson *et al.* [65] show that phytoplanktonic taxa specialize on different components of the light spectrum. This constitutes an example of fine-scale resource partitioning of one resource, light, that all species and genera are usually thought to compete for.

No complexity-stability relationship but connections between self-regulation and interactions between taxa

There was no relation between the complexity of the communities (measured as either the weighted connectance or the interaction coefficient variance) and their stability (measured by the largest modulus of the eigenvalues, which quantifies the return time to a point equilibrium, i.e., resilience). This result is conditional upon our model being a good approximate description of the system (i.e., no multiyear limit cycles or chaotic attractors as the mapping between eigenvalues and actual stability is distorted in that case [71]). However, we already showed on a subset of this data that a fixed point in a MAR(1) model, perturbed by seasonality and abiotic variables, is an accurate description of the system [32]. Therefore, we are confident that the absence of complexity-resilience relationship found here is not a mere artefact of an inadequate model. This absence of direct link between complexity and stability could be an actual feature of empirical systems, as shown previously by Jacquet *et al.* [178] using a different technique. This result seems to contradict theory based on random matrices, especially for competitive and/or mutualistic networks [7]. However, one must bear in mind that such result could also be generated by the limited size of our networks, as random matrix theory relies on asymptotics [8]. We should also mention that our interaction matrices (based on a discrete-time model) are not strictly analogous to the ones used most frequently in theoretical ecology (continuous-time model), though the spectral radius (largest modulus) is here tightly related to the real part of the lead eigenvalue in equivalent continuous-time models (see Supporting Information). Thus while the jury is still out regarding the absence of complexity-resilience relation found here, it may well be a genuine absence. We also found that the percentage of mutualistic interactions, that is thought to affect the stability of a network, either positively or negatively [243, 87, 132], does not in fact have a major impact on our networks' resilience.

In addition to weighted connectance and interaction variance, indices at the genus level (vulnerability and impact) approximate the average effects exerted and sustained by any given taxa in the different study sites. While network structure did not affect resilience, a relation emerged between self-regulation, necessary for coexistence, and genus-level indices. We found that the more a genus is self-regulated, the more it tends to be vulnerable to other genera's impacts and the less it impacts other genera. We examined whether vulnerability and impact could be affected by phylogenetic correlations; they were not, as on Fig. A.3, points were not clustered according to genus, family or phylum. High self-regulation usually indicates large niche differences with the rest of the community, and it makes therefore sense that a species/genus whose needs strongly differ from the others only marginally impacts the resources of the other coexisting species. This is what we expect under strong niche partitioning. A low self-regulation was also correlated with high average abundance, which echoes findings by Yenni *et al.* [343] who demonstrated that rare species usually show stronger self-regulation. This correlation between relative rarity and self-regulation could explain the lesser impact of highly self-regulated species/genus: a taxon which dominates the community composition can have a major effect on the others, especially as they usually cover more space, while it is harder for the less common taxa to have large impacts. In contrast, it was more difficult to explain the relationship between self-regulation and vulnerability: a genus that is more self-regulated and less common was found here to be on average more vulnerable to other genera's increases in densities. Such relation implies greater stability (*sensu* resilience [175], and also invariability [17]) for the network as a whole, because the taxa that are the more vulnerable to other taxa's impacts are also those whose dynamics are intrinsically more buffered. By which mechanisms this could happen is so far unclear. As a final note on relationships between interaction matrix coefficients, we caution that the trends evidenced are all relatively weak: considerable stochasticity still dominates the distribution of interaction matrix coefficients.

Ghosts of competition past and present

Overall, the dominance of niche differentiation in observational plankton studies – based on our analysis of the REPHY dataset and re-analysis of the MAR(1) literature – is similar to what has been recently found in plant community studies [4] or empirically parameterized food webs including horizontal diversity [29]. Large niche differences might be due to the ghost of competition past, i.e., competition has occurred in the past, leading to strong selection and subsequent evolution, and then to progressive niche separation. In this scenario, species have evolved niches that allow them not to compete or to interact only weakly. The likely predator effects that we highlighted above could be comprised within such niche differentiation *sensu largo*: specialized predators can make strong conspecific density-dependence emerge [23, 86], while switching generalists can also promote diversity [326]. Both predators and resources have often symmetrical effects and can therefore contribute almost equally to such past niche

differentiation [76].

An intriguing new possibility, dubbed the "ghost of competition present" [324], suggests by contrast that spatial distributions in relation to abiotic factors might have a large impact on the interaction strengths inferred from temporal interaction models such as ours. Recent combinations of model fitting and removal experiments have shown that model fitting usually underestimates the effect of competitors that are uncovered by removal experiments [324, 3]. This could occur for instance if species are spatially segregated (at a small scale) because each species only exists within a domain where it is relatively competitive (Pacala's spatial segregation hypothesis [262]), while a focal species could spread out if competitors were removed. This means that a species can be limited by competitors, but act so as to minimize competition (a little like avoidance behaviour in animals) and maximize opportunities for positive interactions, which implies that competition is in effect hard to detect when all species are present. This mechanism would require some spatial segregation between phytoplankton species at the scale of interactions, i.e., at the microscale. At the moment, while it is known that fine-scale hydrodynamics generate inhomogeneities at the microscale [33, 59] it is yet quite unclear how they might affect multivariate spatial patterns of species distributions (*sensu* Bolker & Pacala [52] or Murrell & Law [246]). Moreover, even with some microscale spatial segregation between species, a "ghost of competition present" mechanism might not work in phytoplankton as in terrestrial plants, because the turbulent, ever-changing aquatic environment imposes additional constraints on the spatial distribution of organisms.

Acknowledgements

This study was only made possible by the dedication of the members of the REPHY program by Ifremer [284], providing invaluable data through years of fieldwork. We are grateful to David Murrell for his careful reading and suggestions, and to Peter Adler for helpful exchanges. We also want to thank Bérangère Péquin for constructive comments, as well as Gyuri Barabás whose thoughtful suggestions improved the connection made to continuous-time stability theory. This study was supported by the French ANR through LabEx COTE (ANR-10-LABX-45).

Authors' contributions

CP and FB contributed equally to the project design and the methodology. The computer code was written by CP. The authors jointly interpreted the results and co-wrote the manuscript after an early draft by FB.

Data accessibility

The REPHY dataset has already been published [284]. All computer codes and datasets needed for analyses are available online in a GitHub repository (https://github.com/CoraliePicoche/REPHY-littoral) and archived at Zenodo (http://doi.org/10.5281/zenodo.3743249).

A.S Supplementary Information

A.S1 Sampling

Study sites are shown on Fig. A.S1 and the characteristics of each site are given in Table A.S1; they are part of the REPHY monitoring network by Ifremer coastal laboratories [284]. The mean temperature at each site mostly depends on its latitude, reflecting the climate of the region, while salinity is also influenced by more local factors such as evaporation and terrestrial inputs from rivers.

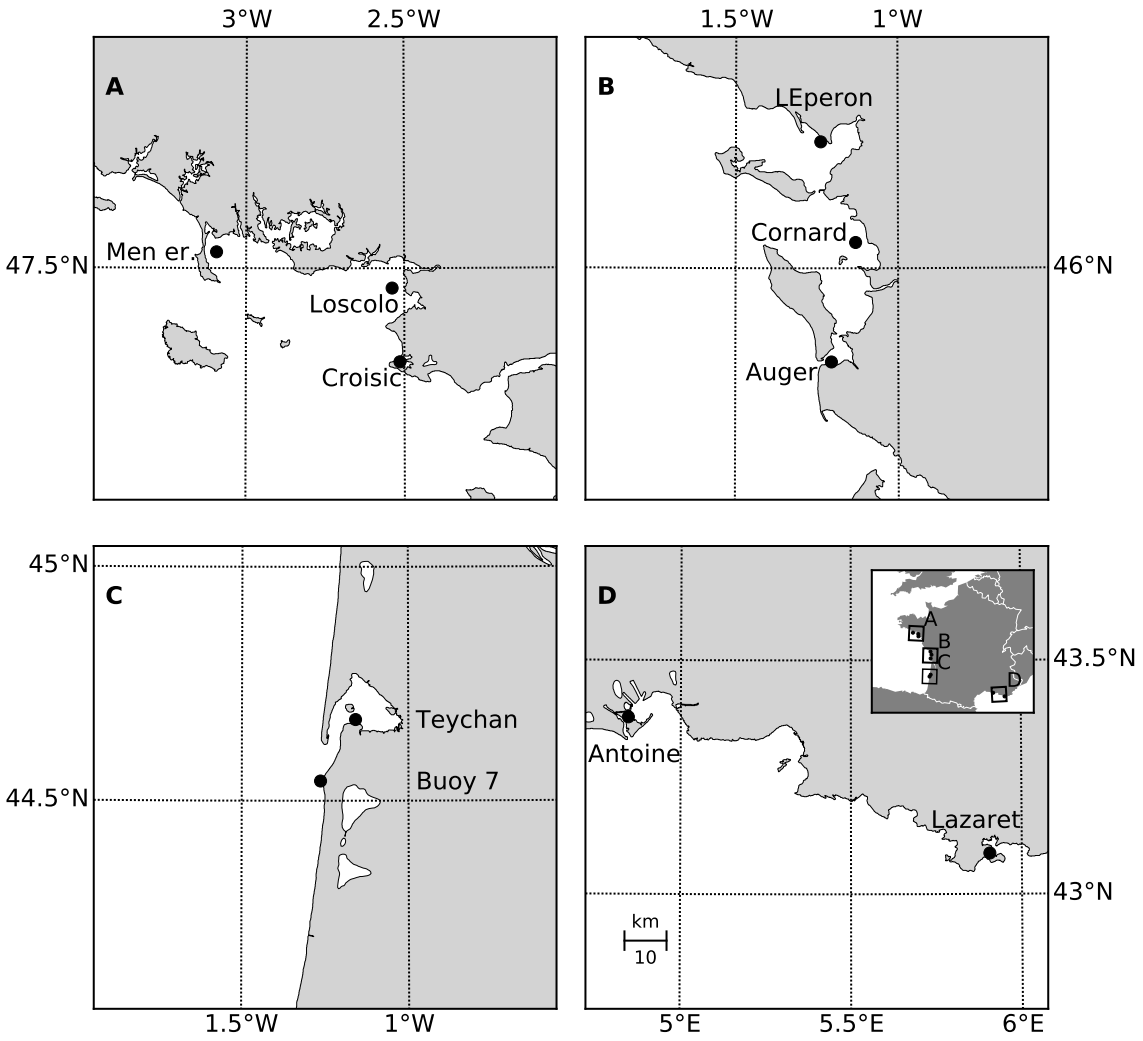

Figure A.S1: **Location of each study site in their region**: Brittany (A), Oléron (B), Arcachon (C) and the Mediterranean Sea (D). The common scale of the panels is given in the left corner of D.

At each site, water samples were collected below surface (between 0 and 1m depth) in a Niskin bottle, from which 10mL were taken for phytoplanktonic counts. We focus on phyto-

planktonic taxa between 20 and 200 µm, the so-called microphytoplankton fraction [285]. The scaling between the volume sampled and the size of individual organisms means that approximately up to 1000 cells could be lined up (in a thought experiment) along one single dimension of an equivalent cubic sample. In other words, the volume sampled is approximately 1000 body sizes to the power three. We therefore consider samples to be representative of local community competition, if not of the spatial structure of the community.

Name of site	**Location**	**Region**	**N. samples**	**Temperature (° C)**	**Salinity (g/L)**
Men Er Roue	47°32' N / 3°5' W	Brittany	503	14.4 +/- 3.7	33.5 +/- 1.9
Loscolo	47°27' N / 2°32' W	Brittany	463	14.9 +/- 4.0	32.0 +/- 3.0
Croisic	47°18' N / 2°30' W	Brittany	500	14.7 +/- 3.9	31.8 +/- 3.1
L'Eperon	46°16' N / 1°14' W	Oléron	460	15.3 +/- 4.8	32.1 +/- 3.2
Cornard	46°3' N / 1°7' W	Oléron	491	15.6 +/- 4.8	32.7 +/- 2.4
Auger	45°47' N / 1°12' W	Oléron	524	15.4 +/- 4.4	32.7 +/- 1.8
Buoy 7	44°32' N / 1°15' W	Arcachon	311	15.2 +/- 3.8	34.7 +/- 0.7
Teychan	44°40' N / 1°9' W	Arcachon	494	15.5 +/- 4.6	32.5 +/- 1.9
Antoine	43°22' N / 4°50' E	Mediterranean Sea	539	16.8 +/- 5.1	32.3 +/- 3.9
Lazaret	43°5' N / 5°54' E	Mediterranean Sea	512	17.4 +/- 4.2	35.9 +/- 2.4

Table A.S1: **Summary of study sites characteristics**, including the mean and standard deviation of the two main environmental parameters (temperature and salinity).

A.S2 Phytoplankton dynamics

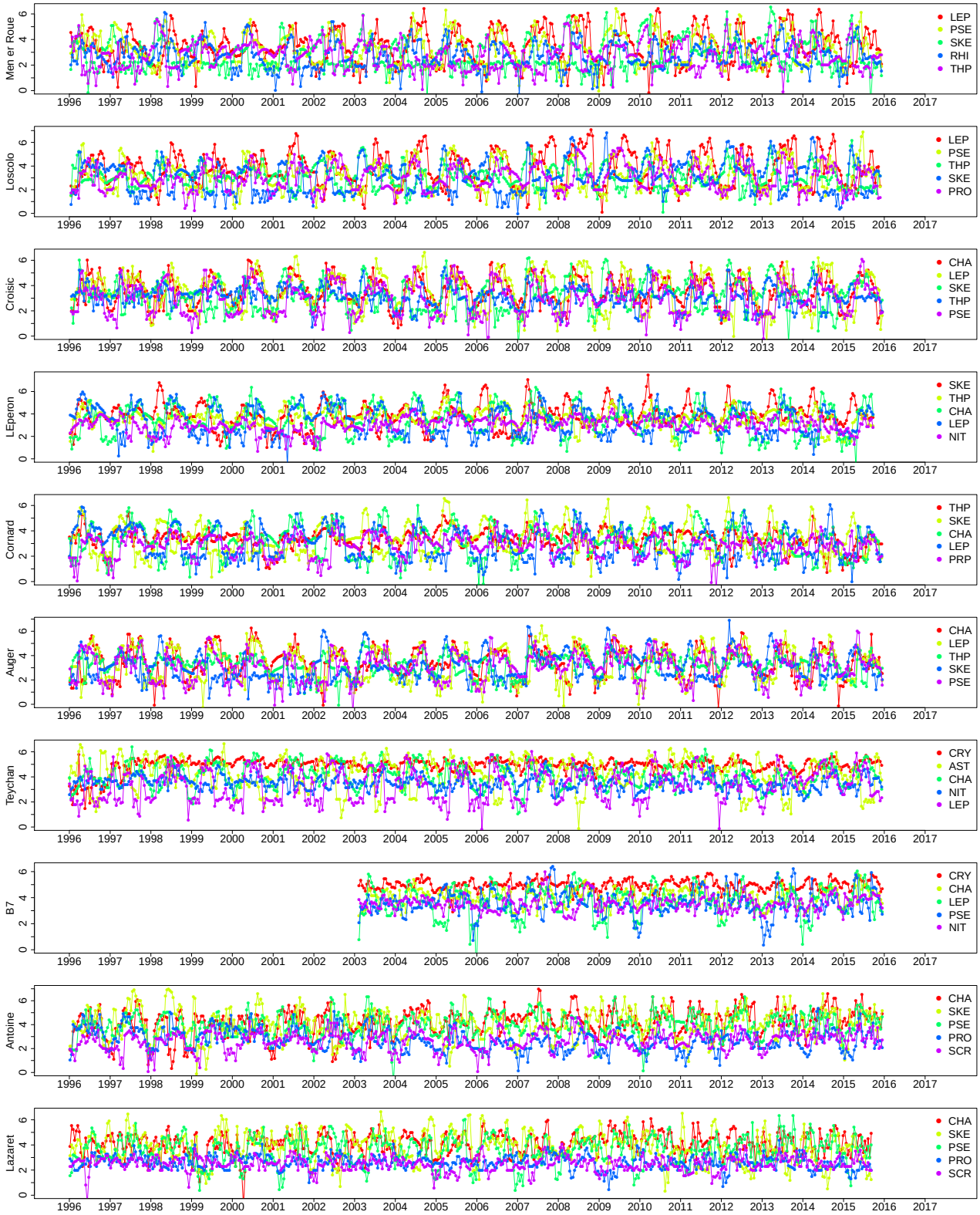

Figure A.S2: **Time series of the 5 most abundant phytoplanktonic genera at each site.**

Code	Taxa
AST	*Asterionella+Asterionellopsis+Asteroplanus*
CHA	*Chaetoceros*
CRY	Cryptophytes
DIT	*Ditylum*
EUG	Euglenophytes
GUI	*Guinardia*
GYM	*Gymnodinium+Gyrodinium*
LEP	*Leptocylindrus*
NIT	*Nitzschia+Hantzschia*
PLE	*Pleurosigma+Gyrosigma*
PRO	*Prorocentrum*
PRP	*Protoperidinium+Archaeperidinium+Peridinium*
PSE	*Pseudo-nitzschia*
RHI	*Rhizosolenia+Neocalyptrella*
SCR	*Scrippsiella+Ensiculifera+Pentapharsodinium+Bysmatrum*
SKE	*Skeletonema*
THL	*Thalassionema+Lioloma*
THP	*Thalassiosira+Porosira*

Table A.S2: **Name and composition of each phytoplanktonic group.** We generally refer to those groups as *taxa*, though we sometimes use *genera* to emphasize that they are above the species level (even though some are stricto sensu groups of genera). This taxonomic grouping is based on Hernández Fariñas *et al.* [154].

A.S3 MAR(1) models

We selected the most parsimonious interaction matrices using BIC. Not only the best-ranking model, but also the overall ranking of interaction scenarios were similar for most sites (Fig. A.S3). Based on these results, we considered the model selection to be quite robust, and focused on the pennate-centric scenario to analyze interaction matrices, as this scenario corresponded to the best fitted models that still allowed interactions between taxa. However, in order to be as comprehensive as possible, we also present results for the full/unconstrained matrix in a section below ("Comparison with a full interaction matrix").

We inspected the Hessian matrices (i.e., Hessians of the negative log-likelihood which is the observed Fisher Information Matrix) for interaction parameters in order to check that results presented in Fig. A.3 of the main text could not be due to statistical constraints. Correlation matrices (derived from the Hessian matrices) showed that the mean absolute value of the correlation was 0.02 (hence on average there were no correlations), few of those were noticeable (75% were below 0.1 in only 1 site, even lower in the 9 other sites) and even the very few large ones were never above 0.5 in absolute value.

For all coefficients, significance was estimated by parametric bootstrapping with 1000 samples.

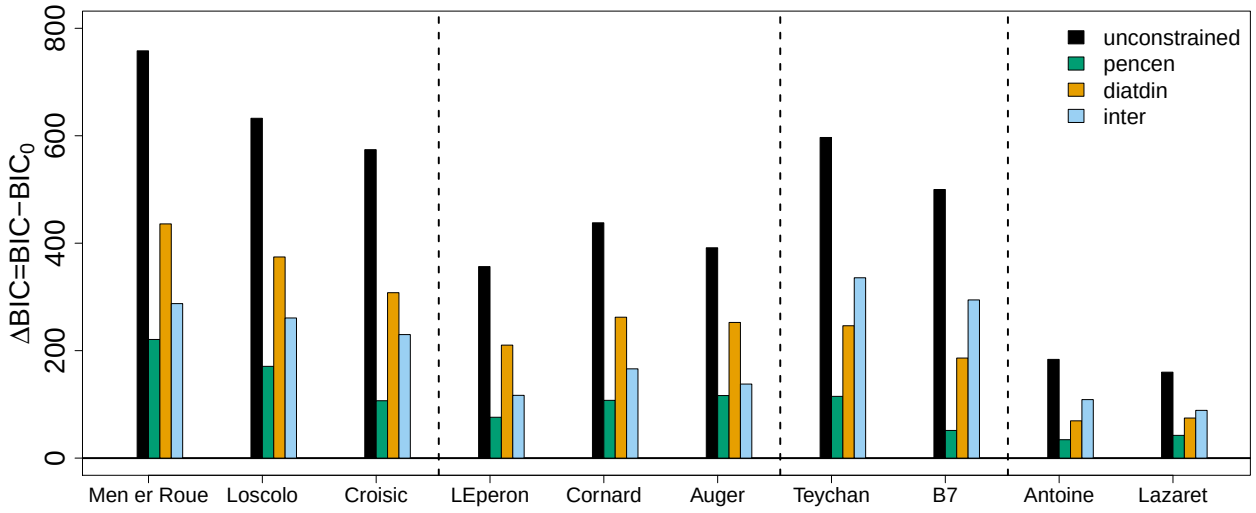

Figure A.S3: **Comparison of the BIC of different interaction scenarios**, compared to the null scenario (diagonal interaction matrix, allowing only intrataxon (intragenus) interactions), at 10 sites in 4 different regions, separated by dashed lines (Brittany, Oléron, Arcachon and the Mediterranean Sea). Different interaction matrices may allow interactions between all taxa (unconstrained), only interactions within pennate diatoms, centric diatoms, dinoflagellates, or other phytoplanktonic taxa (pencen), only interactions within diatoms, dinoflagellates or other taxa (diatdin), or only interactions between taxa belonging to these different groups (inter). As time series lengths differ between sites/regions, bars of the same color but from different sites should not be compared.

In addition to the coefficients of the interaction matrix, MAR(1) models allowed us to estimate the effect of environmental variables. The regression coefficients reveal abiotic effects such as phenology (temperature, related to insolation) or responses to hydrological changes such as salinity variation (Fig. A.S4). Overall, temperature tended to have more effect on phytoplankton dynamics than salinity, which is logical since it integrates seasonal variation in received solar energy. The absolute effect of temperature was on average 3.5 times higher than salinity effects and temperature coefficients were significant at the 5% level for 68% of all estimates, as opposed to 16% for salinity effects. Temperature had a positive effect in 80% of all estimates while salinity had a negative effect for 66% of them. The sign of significant temperature effects on a given species remained the same between regions, except for SKE, which was negatively affected by temperature in Brittany and Oléron but positively affected by temperature in the Mediterranean Sea.

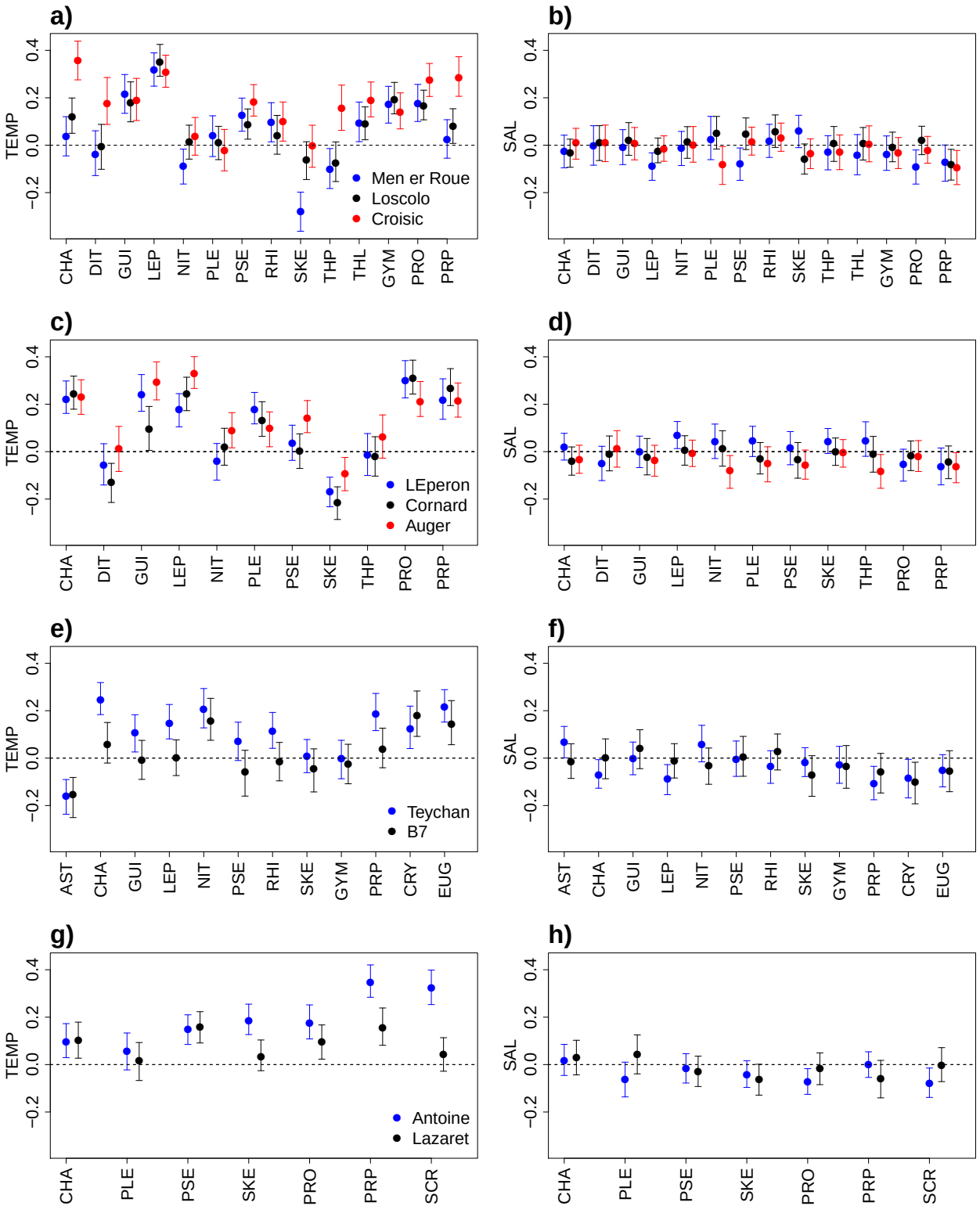

Figure A.S4: **Effects of environmental variables (temperature, TEMP or salinity, SAL) on phytoplankton genera** in Brittany (a, b), Oléron (c, d), Arcachon (e, f) and in the Mediterranean Sea (g, h). Each color corresponds to a different site. Error bars correspond to the 95% confidence interval around the estimated coefficient. All variables were normalized before estimation.

A.S4 Network analysis

Computation of intra/inter-taxa interaction strength ratios from interaction coefficients

We present in the main text ratios between intra and inter-taxa mean interaction strengths, $\kappa = \frac{\text{mean}|b_{ii}|}{\text{mean}|b_{ij,j\neq i}|}$, in several places. These are computed from three sets of intertaxa interaction coefficients:

- b^*_{ij}, coefficients significantly different from 0 at the 5% level
- $b^{\bar{*}}_{ij}$, coefficients not significantly different from 0
- b^0_{ij}, coefficients set to 0 before the estimation process.

Some coefficients are indeed set to 0 before the estimation process, in order to reduce the dimensionality of the system (see section above). There are therefore different estimates of κ, depending on the set of b_{ij} chosen for the denominator.

- $\{b_{ij}\} = \{b^*_{ij}\} \cup \{b^{\bar{*}}_{ij}\}$, leading to an observed ratio κ between 6 and 10.
- $\{b_{ij}\} = \{b^*_{ij}\} \cup \{b^{\bar{*}}_{ij}\} \cup \{b^0_{ij}\}$, leading to a ratio κ between 21 and 43.
- $\{b_{ij}\} = \{b^*_{ij}\} \cup \{b^{\bar{*}}_{ij}\} \cup \{b^0_{ij}\}$ where $b^{\bar{*}}_{ij} := 0$ for comparisons with literature-based MAR(1) coefficients, as results in the literature rarely differentiate between non-significant coefficients and coefficients that were set to 0 beforehand. This corresponds to Fig. A.4 in the main text.
- $\{b_{ij}\} = \{b^*_{ij}\}$, illustrated in Fig. A.S9.

Unless otherwise stated, all analyses presented in the main text and below are based on $\{b_{ij}\} = \{b^*_{ij}\} \cup \{b^{\bar{*}}_{ij}\}$.

Interaction types

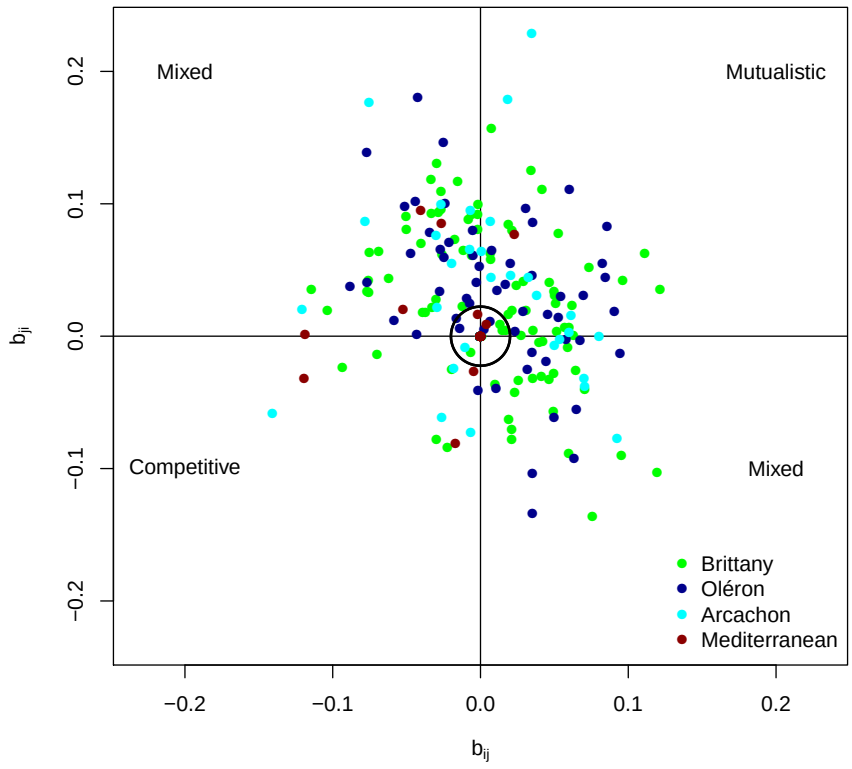

Figure A.S5: **Pairs of coefficients at each study site**. The effect of species i on j is given as a function of the effect of species j on species i. The black circle indicates the first quartile of the interaction values, under which we can assume that the effect is weak or null. Beyond the circle, they are competitive, mutualistic or mixed (+/-) links. Close to the vertical and horizontal lines, one of the coefficient might be essentially zero.

Metrics

We characterised each interaction network with 3 quantitative descriptors: the mean and variance of the intra- and inter-taxa coefficients (i.e., on and off the matrix diagonal) and the weighted connectance of $\mathbf{B} - \mathbf{I}$. The mean of absolute values of intrataxon coefficients was approximately 8 times higher than the mean of absolute values of intertaxa interactions. The intrataxon interactions' variance was about 4 times higher than the variance of intertaxa interactions (Fig. A.S6).

The intrataxon interaction strength could be related to the mean abundance of each taxon (i.e., each genus or groups of genera) as the most self-regulated genera were also the least abundant (Fig. A.S7).

Weighted connectance is described in Bersier *et al.* [45]. It is based on the average of vulnerability and generality in the network. More precisely, diversity measures of the interactions

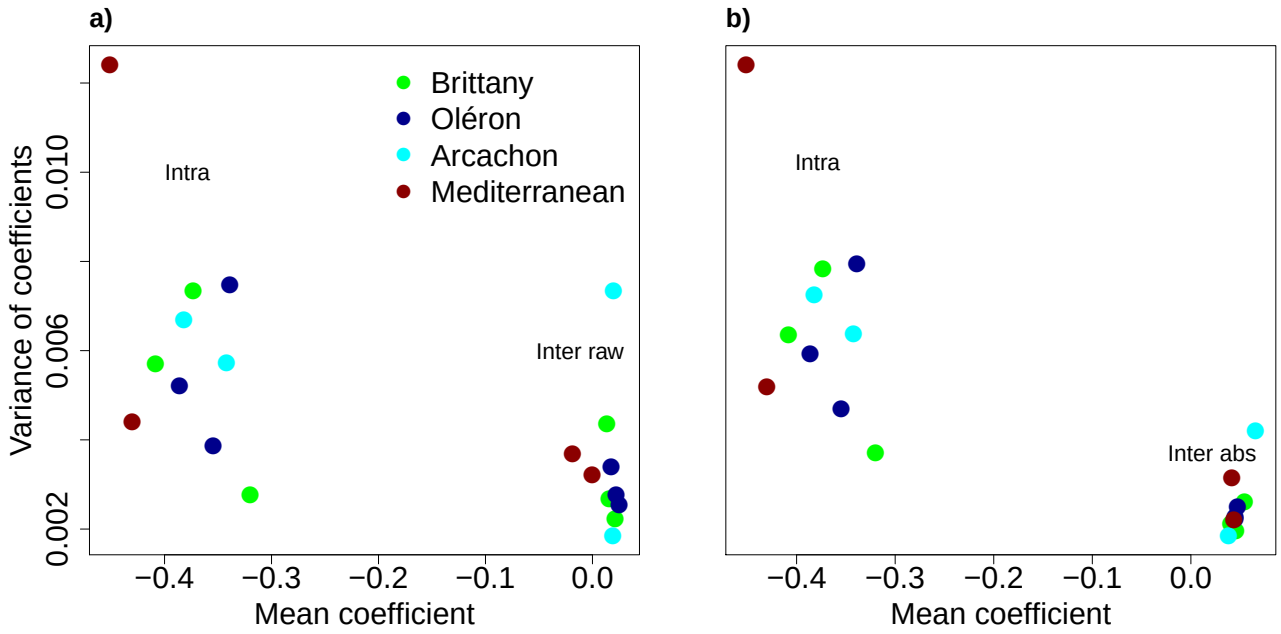

Figure A.S6: **Relationship between mean and variance of the intra- and inter-taxa interaction coefficients**. The variance of the coefficients in the interaction matrix (**B**–**I**) increases with the mean, at 10 sites in 4 regions, with a model allowing interactions only within clads (within centric or pennate diatoms, dinoflagellates, or other taxa). The mean-variance relation was either computed with raw values of intertaxa interactions (a) or absolute values of the intertaxa coefficients (b). We did not take the absolute value of intrataxon coefficients since they were all negative.

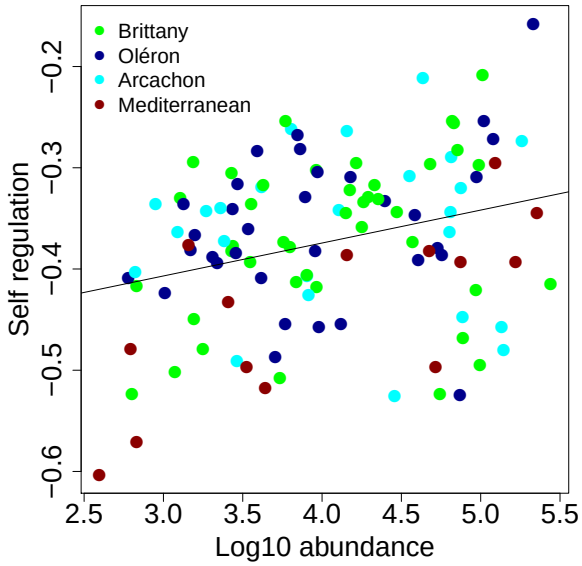

Figure A.S7: **Relationship between abundance and self-regulation** (intragenus interaction coefficients). Mean abundance is computed for each taxon/genus at 10 sites in 4 regions and intragenus interaction strengths are the diagonal coefficients of the interaction matrix (**B**–**I**).

from ($H_{P,k}$) and to ($H_{N,k}$) the phytoplanktonic taxon k can be computed as:

$$H_{N,k} = -\sum_{i=1}^{S} \frac{b_{ik}}{b_{\bullet k}} \log_2 \left(\frac{b_{ik}}{b_{\bullet k}} \right) \tag{A.S1}$$

$$H_{P,k} = -\sum_{i=1}^{S} \frac{b_{ki}}{b_{k\bullet}} \log_2 \left(\frac{b_{ki}}{b_{k\bullet}} \right) \tag{A.S2}$$

where b_{ik} is a coefficient of the interaction matrix (**B**–**I**), $b_{k\bullet} = \sum_{i=1}^{S} b_{ki}$ is the sum of all coefficients over row k and S is the number of taxa in the network. These indices are then averaged for the whole network as the linkage density LD (eq. A.S3).

$$LD = \frac{1}{2} \left(\sum_{k=1}^{S} \frac{b_{\bullet k}}{b_{\bullet\bullet}} 2^{H_{N,k}} + \sum_{k=1}^{S} \frac{b_{k\bullet}}{b_{\bullet\bullet}} 2^{H_{P,k}} \right) \tag{A.S3}$$

where $b_{\bullet\bullet} = \sum_{j=1}^{S} \sum_{i=1}^{S} b_{ji}$ is the sum of all coefficients of the interaction matrix (**B**–**I**).

Weighted connectance C is then defined as:

$$C = \frac{LD}{S} \tag{A.S4}$$

Contrary to linkage density, weighted connectance accounts for the dimension of the interaction matrix and can be used to compare network in different regions, with different dimensions.

In addition to this network-level metric, we also considered metrics for each phytoplanktonic taxon. We measured both its average vulnerability score (mean strength of the interactions that are applied to a taxon, eq. A.S5) and its average impact score (mean strength of the interactions the taxon applies to other taxa, eq. A.S6) in each network.

$$v_k = \frac{1}{\sum_{i=1}^{S} \mathbf{1}_{b_{ki} \neq 0}} \sum_{i=1}^{S} b_{ki} \tag{A.S5}$$

$$e_k = \frac{1}{\sum_{i=1}^{S} \mathbf{1}_{b_{ik} \neq 0}} \sum_{i=1}^{S} b_{ik} \tag{A.S6}$$

where $\sum_{i=1}^{S} \mathbf{1}_{b_{ki} \neq 0}$ is the number of interactions which are different from 0 in row k.

A.S5 Comparison with a full interaction matrix

We checked that, by choosing the model with the lowest BIC, we did not miss interactions which would have changed our conclusions. To do so, we examined the full (unconstrained) model results for all study sites. We present those results below (Fig. A.S8 and Table A.S3).

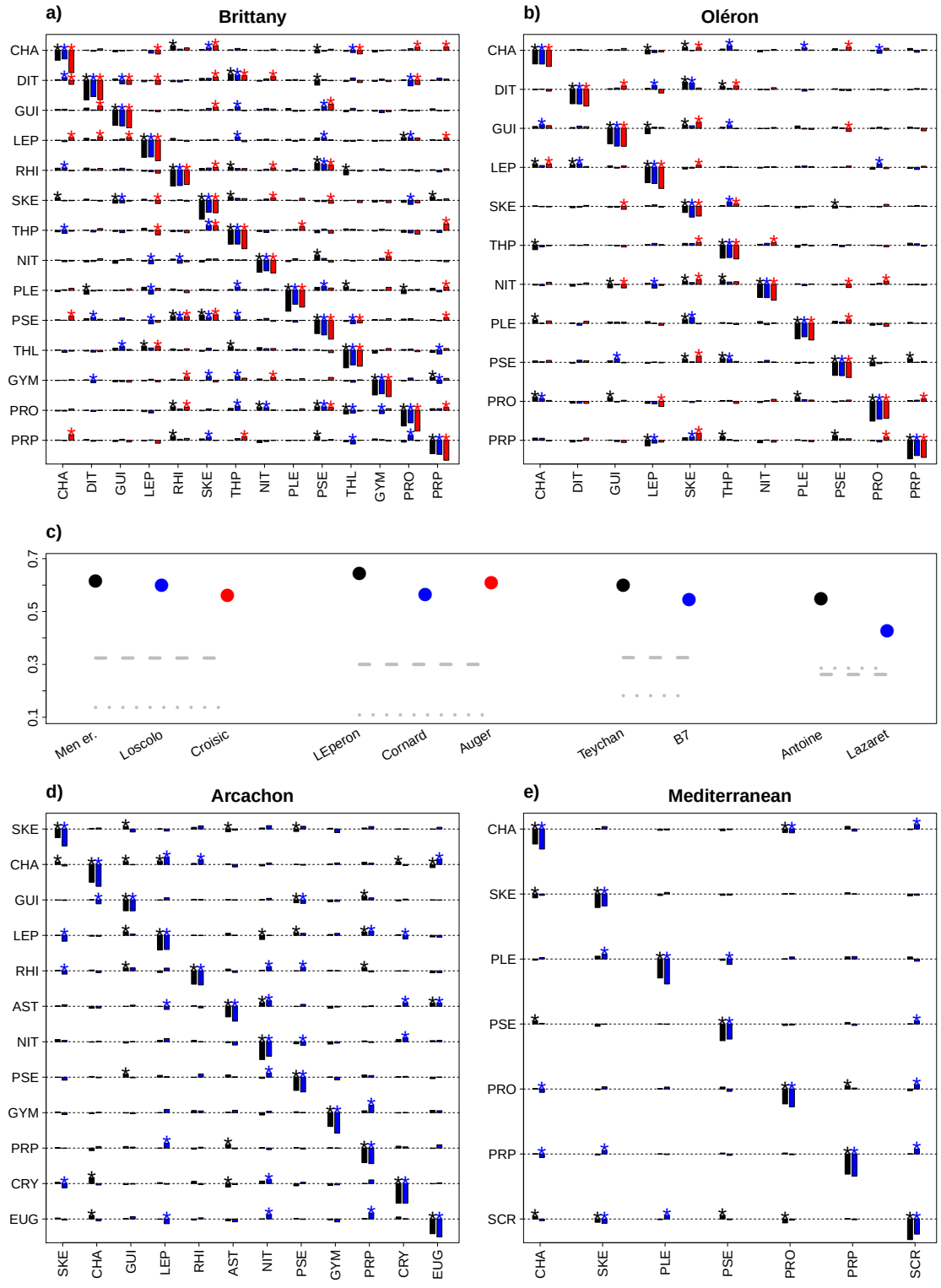

Figure A.S8: **Interaction matrices estimated at 10 sites along the French coastline**. Four regions are distinguished: Brittany (a), Oléron (b), Arcachon (d) and the Mediterranean Sea (e). There is no constraint on the structure (modularity) of the interaction matrices. The fraction of positive interactions in each matrix is given by points in c) while the dashed (resp., dotted) line represents the ratio of interactions remaining positive (resp., negative) for all sites of a given region.

	signif outside	positive	ratio intra/inter	ratio in/out block	transfo sign
Men er Roue	0.09	0.57	11.06	3.03	0.04
Loscolo	0.14	0.56	8.42	2.65	0.07
Croisic	0.13	0.52	10.15	2.89	0.09
LEperon	0.13	0.59	8.78	2.75	0.04
Cornard	0.08	0.51	10.32	3.64	0.06
Auger	0.07	0.55	9.66	3.10	0.06
Antoine	0.10	0.47	11.18	5.21	0.00
Lazaret	0.18	0.37	8.67	4.30	0.00
Teychan	0.11	0.55	10.46	3.68	0.14
B7	0.11	0.50	8.29	3.59	0.14

Table A.S3: Descriptors of coefficients in unconstrained interaction matrices and comparison to best-fitting pennate-centric structures: ratio of coefficients significantly different from 0 outside of the pennate-centric blocks vs total number of coefficients in the unconstrained matrix, proportion of positive interactions in the unconstrained matrix, ratio of mean intrataxon interaction strength and mean intertaxa interaction strength in the unconstrained matrix, ratio of mean interaction strengths inside the pennate-centric modules vs outside the pennate-centric modules in the unconstrained matrix, and proportion of interactions changing sign between the two matrix structures.

Thus, even if we chose to select the full interaction model, there would be no difference in our main conclusions: intrataxon interactions are much stronger than intertaxa interactions and positive interactions are still the rule. There is at most 18% of interactions significantly different from zero outside of the pennate and centric blocks and those interactions are on average 3.5 times lower than the interactions inside the pennate and centric blocks (Table A.S3).

A.S6 MAR references and analysis

We present here the MAR references we used to compare the effects of intra- and intertaxa interactions (Table A.S4, Fig. A.S9). We add information on the biome and taxa used in the study as they tend to be linked to the estimated parameters (Fig. A.S10). We should mention two potential biases associated with this comparison across the published literature. First, low-dimensional matrices tended to be more complete (less sparse) than high-dimensional matrices, as these small interaction matrices were used to study known interactions (observed predation between organisms, for instance). In fact, we add a handful of predator-prey systems (in red) mainly to give a scale to the plot. Conversely, the number of parameters to estimate increases as the square of the number of interacting taxa, leading most authors to reduce this set before the estimation process for large interaction matrices. There is therefore a positive correlation between sparsity and dimensionality (Fig. A.S10). A second caveat is that, while we informed our model selection by phylogeny, several authors have instead reduced the number of estimated parameters by an automated procedure, usually based on the comparison of hundreds of randomly chosen interaction matrices by AIC [174]. The latter choice is likely to bias high

non-zero interactions in the literature. This is why we decided to present in the main text intra/inter ratios including intertaxa coefficients set to zero (see Fig. A.4 in the main text), which should be less sensitive to the model selection method and therefore make comparisons across studies possible. In Fig. A.S9, mean interaction strengths were computed as the mean absolute value of the set of coefficients which were deemed significant at the 5% level in the **B**–**I** matrix (see the "Computation of intra/inter-taxa ratios" section above for details).

Code	Ref	Dimension	Type of organisms	Taxonomic level	System	T
1a	Ives *et al.* [174], CLS	9	Zooplankton	Species and functional groups	Lake	100
1b	Ives *et al.* [174], TLS	9	Zooplankton	Species and functional groups	Lake	100
2a	Klug *et al.* [195]	2	Phytoplankton	Phylum	Lake	100
2b	Klug *et al.* [195]	3	Zooplankton	Species	Lake	50
3a	Klug & Cottingham [194]	4	Functional groups of plankton	NA	Lake	300
3b	Klug & Cottingham [194]	5	Taxonomic groups of plankton	Phylum/division	Lake	300
4a	Ives *et al.* [175]	4	Plankton	Zooplankton v. phytoplankton, size classes	Lake	100
4b	Ives *et al.* [175]	4	Plankton	Zooplankton v. phytoplankton, size classes	Lake with high planktivory	100
4c	Ives *et al.* [175]	4	Plankton	Zooplankton v. phytoplankton, size classes	Lake with low planktivory	100
5a	Hampton & Schindler [144]	14	Plankton	Phylum (phytoplankton), genus (zooplankton)	Lake	300
5b	Hampton & Schindler [144]	14	Plankton, growing season	Phylum (phytoplankton), genus (zooplankton)	Lake	200
6a	Hampton *et al.* [143]	13	Plankton	Phyum (phytoplankton), genus (zooplankton)	Lake	400
6b	Hampton *et al.* [143]	7	Simpler web, plankton	Phylum (phytoplankton), genus (zooplankton)	Lake	400
7a	Huber & Gaedke [163]	10	Ciliates	Genus and species	Lake	300
7b	Huber & Gaedke [163]	10	Phytoplankton	Genus and species	Lake	300
8a	Yamamura *et al.* [342]	3	Insects	Species	Terrestrial	50
9a	Vik *et al.* [327]	2	Lynx/Hare	Species	Terrestrial	100
10a	Lindegren *et al.* [213]	3	Fish	Species	Baltic Sea	30
11a	Griffiths *et al.* [137]	7	Phytoplankton	Phylum	Coastal site	1000
11b	Griffiths *et al.* [137]	7	Phytoplankton	Phylum	Offshore site	700

Table A.S4: Studies used when comparing |intra|/|inter| ratios in Fig. A.4 in main text. T is the approximate number of sampling dates in each time series.

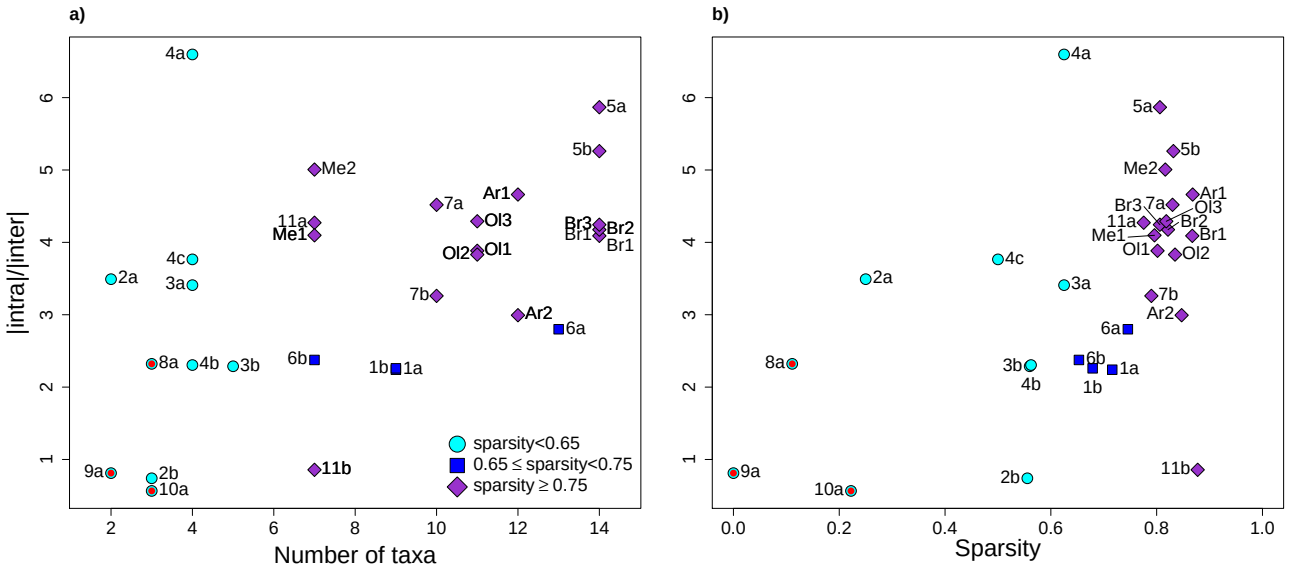

Figure A.S9: **Ratios of intra- to inter-taxa mean interaction strength in MAR(1) studies.** Only significant values are taken into account and missing values in the matrix are not considered (e.g., not replaced by 0 as they are in the main text). The color and shape of each point are a function of the sparsity of the interaction matrix $\mathbf{B}-\mathbf{I}$ and the relation between the ratio and the sparsity of the matrix is given in the right panel. Red dots correspond to terrestrial and/or low dimension, predator-prey systems, giving a lower bound for the intra/inter ratio. Corresponding studies are described in Table A.S4. Annotations beginning with letters correspond to the present study.

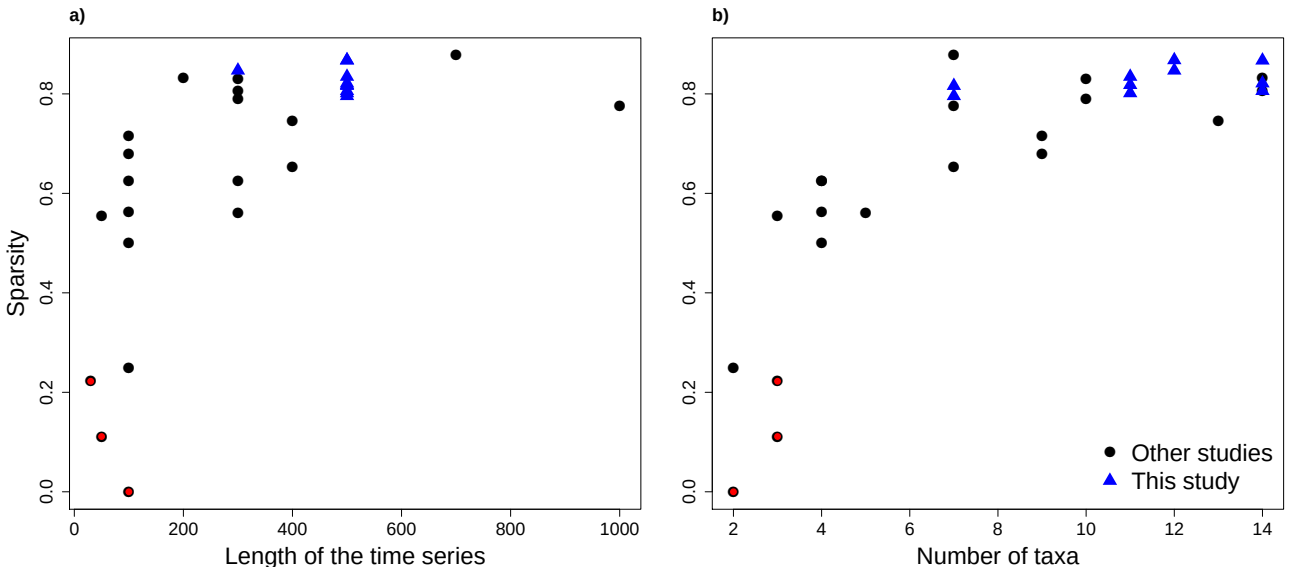

Figure A.S10: **Relationship between interaction sparsity and study design** in studies described in Table A.S4. Red dots correspond to terrestrial and/or low dimension, predator-prey systems, giving a lower bound for the intra/inter ratio. Blue triangles correspond to the present study.

A.S7 Connection to continuous-time models

The relation between complexity and stability in community models has been debated for decades in theoretical circles [227, 8]. In theoretical ecology, random matrix theory has been mostly applied to continuous-time interaction models (Allesina & Tang [8], but see Cohen & Newman [82]). Here, we intend to connect our statistical discrete-time models to the continuous-time models of stability theory. The discrete-time log-linear model writes $\mathbf{x}_{t+1} = \mathbf{B}\mathbf{x}_t$ in the main text. This model can only approximate continuous-time, possibly non-linear dynamics. There are at least two ways to relate discrete-time models to continuous-time dynamics.

The first approach is to linearize continuous-time dynamics ($\mathrm{d}\mathbf{x} = \mathbf{A}\mathbf{x}\mathrm{d}t$ where $\mathbf{A}$ is the continuous-time community matrix) and to integrate the system over time. In this case, the map from one time-step to the next can be written $\mathbf{x}_{t+1} = \mathrm{e}^{\mathbf{A}}\mathbf{x}_t$. The discrete-time equivalent of the community matrix $\mathbf{A}$ is then $\log(\mathbf{B})$ where $\mathrm{e}^{\mathbf{A}}$ is a matrix exponential and $\log(\mathbf{B})$ the reciprocal of $\mathrm{e}^{\mathbf{A}}$ with $\mathbf{B}$ defined as above.

The second approach is to first integrate a continuous-time model over a time-step and then linearize the system. In this case, the equivalent matrix $\mathbf{A} = \mathbf{B} - \mathbf{I}$ because it describes the effects of densities on population growth rates (by contrast $\mathbf{B}$ describes effects of log-densities at time t on log-densities at time $t + 1$). The second approach is illustrated in more detail in the next section of the Supporting Information.

Moreover, the measure of resilience differs in discrete- and continuous-time models. In discrete-time models, and therefore in this study, resilience is measured as the maximum modulus of the eigenvalues of the community matrix ($\max(|\lambda_B|)$), through the dominant eigenvalue of $\mathbf{B}$. In continuous-time models, resilience is linked to the maximum real part of the eigenvalues ($\max(\mathrm{Re}(\lambda_A))$), i.e., the real part of the leading eigenvalue of $\mathbf{A}$. There is therefore a link to be made between these metrics. We present in Fig. A.S11 the relationship between resilience metrics $\max(|\lambda_B|)$ and $\max(\mathrm{Re}(\lambda_A))$ to make sure that our results in discrete-time are consistent with continuous-time theory.

We see in Fig. A.S11 that:

- leading eigenvalues of $\mathbf{A}$ are similar for $\mathbf{A} = \mathbf{B} - \mathbf{I}$ and $\mathbf{A} = \log(\mathbf{B})$ (the difference in real parts is around 0.04 for values between -0.45 and -0.2). Hence, $\mathbf{B} - \mathbf{I}$ is a simpler approximation of $\mathbf{A}$ (Fig A.S11 a),
- the modulus of the dominant eigenvalue of $\mathbf{B}$ is strongly correlated to the real part of the leading eigenvalue of $\mathbf{A} = \mathbf{B} - \mathbf{I}$ and $\mathbf{A} = \log(\mathbf{B})$ (> 0.99 in both cases), which means that our results are compatible with continuous-time theory (Fig A.S11 a)
- there is no apparent relationship between the continuous-time equivalent metric for stability/resilience and complexity (the latter being measured as the number of taxa times the variance of the intertaxa interaction coefficients, Fig A.S11 b).

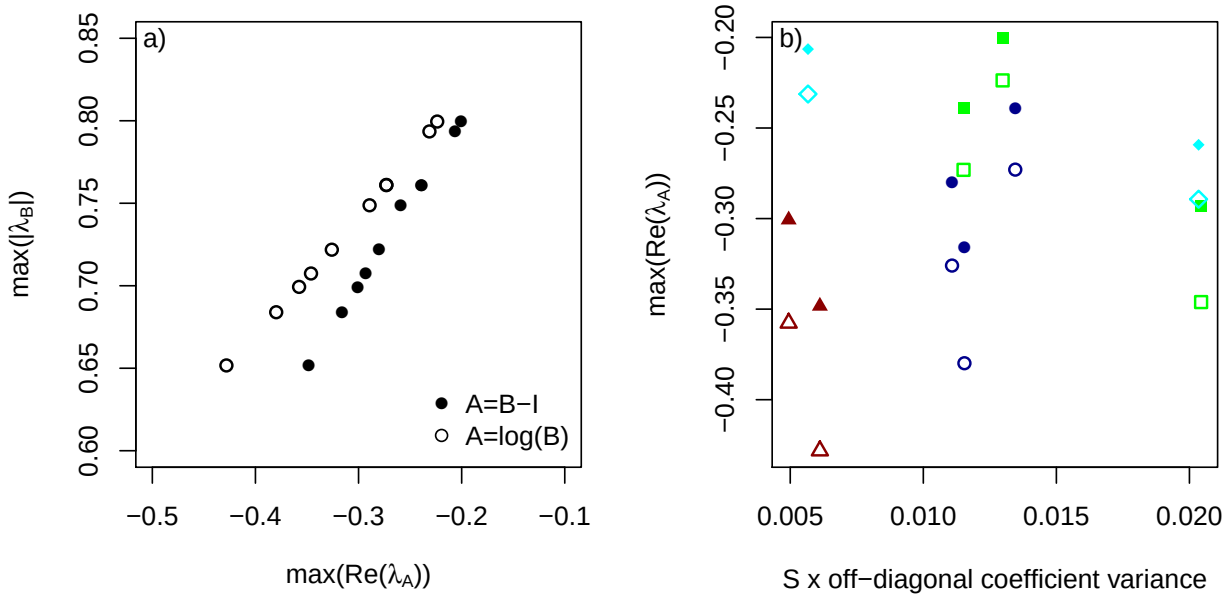

Figure A.S11: **Stability metrics in discrete vs continuous time.** Relationship between discrete-time and continuous-time stability metrics in (a); relationship between $\max(\mathrm{Re}(\lambda_A))$ and an index of complexity in (b), that is, the variance of the off-diagonal coefficients weighted by the number of taxa in each community.

The link between stability and complexity that we study here is therefore of a similar nature to that studied in continuous-time stability theory, and the relation between stability and complexity was found to be non-existent using both discrete-time metrics and equivalent metrics mimicking continuous-time. The absence of relationship between complexity and stability *sensu* resilience therefore appears to be genuine.

A.S8 Connection to Lotka-Volterra competition dynamics

The Beverton-Holt multispecies competition model, whose variants are widely used for modelling plant community dynamics [206, 198], is the closest discrete time equivalent to the continuous-time Lotka-Volterra model (see [90]; although the mapping is not perfect for $S \geq 3$ [287]). The Beverton-Holt multispecies competition model writes

$$N_{i,t+1} = \frac{e^{r_i} N_{i,t}}{1 + \sum_j \alpha_{ij} N_{j,t}} \tag{A.S7}$$

where $N_{i,t}$ is the abundance of species i at time t, r_i is its growth rate and α_{ij} is the effect of species j on species i. Here, we show how the interaction strengths α_{ij} map to those of the MAR(1) models used in the main text.

We start with a 2 species model for simplicity, and we note the equilibrium values of species 1 and 2 as N_1 and N_2 (without time subscript). We re-write the model at equilibrium.

$$\begin{cases} R_1 = & \alpha_{11}N_1 + \alpha_{12}N_2 \\ R_2 = & \alpha_{21}N_1 + \alpha_{22}N_2 \end{cases}, \text{with} \begin{cases} R_1 = & e^{r_1} - 1 \\ R_2 = & e^{r_2} - 1 \end{cases} \tag{A.S8}$$

$$\Leftrightarrow \begin{cases} \alpha_{21}R_1 = & \alpha_{21}\alpha_{11}N_1 + \alpha_{21}\alpha_{12}N_2 \\ \alpha_{11}R_2 = & \alpha_{11}\alpha_{21}N_1 + \alpha_{11}\alpha_{22}N_2 \end{cases} \tag{A.S9}$$

$$\Leftrightarrow \begin{cases} N_1 = & \frac{\alpha_{12}R_2 - \alpha_{22}R_1}{\alpha_{12}\alpha_{21} - \alpha_{22}\alpha_{11}} \\ N_2 = & \frac{\alpha_{21}R_1 - \alpha_{11}R_2}{\alpha_{12}\alpha_{21} - \alpha_{22}\alpha_{11}}. \end{cases} \tag{A.S10}$$

Setting $n = \log(N)$, eq. A.S7 is equivalent to

$$\begin{cases} n_{1,t+1} = & r_1 + n_{1,t} - \ln(1 + \alpha_{11}N_{1,t} + \alpha_{12}N_{2,t}) \\ n_{2,t+1} = & r_2 + n_{2,t} - \ln(1 + \alpha_{21}N_{1,t} + \alpha_{22}N_{2,t}). \end{cases} \tag{A.S11}$$

We want to compute J, the log-scale Jacobian matrix of the model.
Let us note $X = \ln(1 + \alpha_{11}N_1 + \alpha_{12}N_2)$ and $Y = \ln(1 + \alpha_{21}N_1 + \alpha_{22}N_2)$.

$$J = \begin{pmatrix} 1 - \frac{\partial X}{\partial n_1} & -\frac{\partial X}{\partial n_2} \\ -\frac{\partial Y}{\partial n_1} & 1 - \frac{\partial Y}{\partial n_2} \end{pmatrix}. \tag{A.S12}$$

We have $\frac{\partial X}{\partial n_1} = \frac{\partial X}{\partial N_1}\frac{\partial N_1}{\partial n_1} = \frac{\alpha_{11}}{1+\alpha_{11}N_1+\alpha_{12}N_2}e^{n_1} = \frac{\alpha_{11}}{1+\alpha_{11}N_1+\alpha_{12}N_2}N_1$, which leads to

$$J - I = \begin{pmatrix} -\frac{\alpha_{11}N_1}{1+\alpha_{11}N_1+\alpha_{12}N_2} & -\frac{\alpha_{12}N_2}{1+\alpha_{11}N_1+\alpha_{12}N_2} \\ -\frac{\alpha_{21}N_1}{1+\alpha_{21}N_1+\alpha_{22}N_2} & -\frac{\alpha_{22}N_2}{1+\alpha_{21}N_1+\alpha_{22}N_2} \end{pmatrix}. \tag{A.S13}$$

For this demonstration, we consider diffuse competition, that is

$$\begin{cases} \alpha_{ii} = & k\alpha \\ \alpha_{ij} = & \alpha, \forall i \neq j \end{cases}. \tag{A.S14}$$

If we combine eq. A.S10 and eq. A.S14, we have

$$\begin{cases} N_1 = & \frac{\alpha R_2 - k\alpha R_1}{\alpha^2 - k^2\alpha^2} \\ N_2 = & \frac{\alpha R_1 - k\alpha R_2}{\alpha^2 - k^2\alpha^2} \end{cases} \tag{A.S15}$$

and

$$J - I = \begin{pmatrix} -\frac{k\alpha N_1}{1+k\alpha N_1+\alpha N_2} & -\frac{\alpha N_2}{1+k\alpha N_1+\alpha N_2} \\ -\frac{\alpha N_1}{1+\alpha N_1+k\alpha N_2} & -\frac{k\alpha N_2}{1+\alpha N_1+k\alpha N_2} \end{pmatrix} \tag{A.S16}$$

$$\Rightarrow \begin{cases} (J-I)_{11} = & \frac{-\frac{k}{1-k^2}(R_2 - kR_1)}{1+\frac{k}{1-k^2}(R_2-kR_1)+\frac{1}{1-k^2}(R_1-kR_2)} \\ (J-I)_{12} = & \frac{-\frac{1}{1-k^2}(R_1 - kR_2)}{1+\frac{k}{1-k^2}(R_2-kR_1)+\frac{1}{1-k^2}(R_1-kR_2)} \end{cases} \Rightarrow (J-I)_{11} = k\frac{R_2 - kR_1}{R_1 - kR_2}(J-I)_{12}. \tag{A.S17}$$

By symmetry, we can also write

$$(J-I)_{22} = k\frac{R_1 - kR_2}{R_2 - kR_1}(J-I)_{21}. \tag{A.S18}$$

The same reasoning can actually be applied with S species as the Jacobian has a similar form:

$$\begin{aligned}(J-I)_{ii} = & \frac{-\alpha_{ii}N_i}{1+\sum_j \alpha_{ij}N_j} \\ = & \frac{-k\alpha N_i}{1+k\alpha N_i + \sum_{j,j\neq i}\alpha N_j}\end{aligned} \tag{A.S19}$$

and

$$\begin{aligned}(J-I)_{ij} = & \frac{-\alpha_{ij}N_j}{1+\sum_l \alpha_{il}N_l} \\ = & \frac{-\alpha N_j}{1+k\alpha N_i + \sum_{l,l\neq i}\alpha N_l} \\ = & (J-I)_{ii}\frac{\alpha N_j}{k\alpha N_i} \\ = & \frac{1}{k}(J-I)_{ii}\frac{N_j}{N_i}.\end{aligned} \tag{A.S20}$$

However, in the main text, we use the ratio $\kappa = \frac{\overline{|(J-I)_{ii}|}}{\overline{|(J-I)_{ij}|}}$, with average quantities in the numerator and the denominator. To build a relationship between k and κ, we denote $b_{ij} = (J-I)_{ij}$, $N_T = \sum_{i=1}^{S} N_i$. In this simple diffuse competition case, all b_{ij} have the same negative sign. Therefore,

$$\begin{aligned}\kappa = & \frac{\overline{b_{ii}}}{\overline{b_{ij}}} \\ \text{where } \overline{b_{ii}} = & \frac{1}{n}\sum_{i=1}^{S} b_{ii} \\ \text{and } \overline{b_{ij}} = & \frac{1}{n(n-1)}\sum_{i=1}^{S}\sum_{j=1,j\neq i}^{S} b_{ij}.\end{aligned} \tag{A.S21}$$

Using eq. A.S20, we obtain

$$\begin{aligned}
\overline{b_{ij}} = & \frac{1}{kS(S-1)} \sum_{i=1}^{S} \sum_{j=1, j\neq i}^{S} b_{ii} \frac{N_j}{N_i} \\
= & \frac{1}{kS(S-1)} \sum_{i=1}^{S} \frac{b_{ii}}{N_i} \left(\left(\sum_{j=1}^{S} N_j \right) - N_i \right) \\
= & \frac{1}{kS(S-1)} \sum_{i=1}^{S} \frac{b_{ii}}{N_i} (N_T - N_i) \\
= & \frac{1}{k(S-1)} \left(\frac{1}{S} \sum_{i=1}^{S} b_{ii} \frac{N_T}{N_i} - \frac{1}{S} \sum_{i=1}^{S} b_{ii} \right) \\
= & \frac{1}{k(S-1)} \left(\frac{1}{S} \sum_{i=1}^{S} b_{ii} \frac{N_T}{N_i} - \overline{b_{ii}} \right).
\end{aligned} \tag{A.S22}$$

Coupling eq. A.S21 and A.S22, this leads to

$$\frac{\overline{b_{ij}}}{\overline{b_{ii}}} = \frac{1}{\kappa} = \frac{1}{k(S-1)} \left(\frac{1}{S} \sum_{i=1}^{S} \frac{b_{ii}}{\overline{b_{ii}}} \frac{N_T}{N_i} - 1 \right). \tag{A.S23}$$

Thus, even in the simple case of diffuse competition, the ratio intra/inter might change to some degree between MAR(1) and BH competition, as a function of relative abundances. For even communities, the mapping between Lotka-Volterra and MAR(1) interaction strength ratios is good; the combined effect of variance in self-regulation strength and realistic levels of community uneveness may change this. Also, diffuse competition is a simplification: mapping such interaction strength ratios in the general many-species case is a non-trivial endeavour, and further deviations between the two frameworks could be expected (see [71], for the two-species case).

Bibliography

[We are] dwarfs perched on the shoulders of giants [...]
We see more and farther than our predecessors, not because we have keener vision or greater height, but because we are lifted up and borne aloft on their gigantic stature.
(B. de Chartres, cited by J. of Salisbury, 1159)

[1] Abrams, P.A., Holt, R.D. & Roth, J.D. (1998). Apparent competition or apparent mutualism? Shared predation when populations cycle. *Ecology*, 79, 201–212.

[2] Adler, P.B., Ellner, S.P. & Levine, J.M. (2010). Coexistence of perennial plants: an embarrassment of niches. *Ecology letters*, 13, 1019–1029.

[3] Adler, P.B., Kleinhesselink, A., Hooker, G., Taylor, J.B., Teller, B. & Ellner, S.P. (2018). Weak interspecific interactions in a sagebrush steppe? Conflicting evidence from observations and experiments. *Ecology*, 99, 1621–1632.

[4] Adler, P.B., Smull, D., Beard, K.H., Choi, R.T., Furniss, T., Kulmatiski, A., Meiners, J.M., Tredennick, A.T. & Veblen, K.E. (2018). Competition and coexistence in plant communities: intraspecific competition is stronger than interspecific competition. *Ecology Letters*, 21, 1319–1329.

[5] Aikio, S., Ranta, E., Kaitala, V. & Lundberg, P. (2002). Seed bank in annuals: competition between banker and non-banker morphs. *Journal of Theoretical Biology*, 217, 341–349.

[6] Alexander, H., Jenkins, B.D., Rynearson, T.A. & Dyhrman, S.T. (2015). Metatranscriptome analyses indicate resource partitioning between diatoms in the field. *Proceedings of the National Academy of Sciences*, 112, E2182–E2190.

[7] Allesina, S. & Tang, S. (2012). Stability criteria for complex ecosystems. *Nature*, 483, 205–208.

[8] Allesina, S. & Tang, S. (2015). The stability-complexity relationship at age 40: a random matrix perspective. *Population Ecology*, 57, 63–75.

[9] Allison, D.B., Brown, A.W., George, B.J. & Kaiser, K.A. (2016). Reproducibility: a tragedy of errors. *Nature*, 530, 27–29.

[10] Álvarez, E., Lazzari, P. & Cossarini, G. (2022). Phytoplankton diversity emerging from chromatic adaptation and competition for light. *Progress in Oceanography*, 204, 102789.

[11] Amarasekare, P. (2003). Competitive coexistence in spatially structured environments: a synthesis. *Ecology Letters*, 6, 1109–1122.

[12] Anderson, D.M., Stock, C.A., Keafer, B.A., Bronzino Nelson, A., Thompson, B., McGillicuddy, D.J., Keller, M., Matrai, P.A. & Martin, J. (2005). *Alexandrium fundyense* cyst dynamics in the Gulf of Maine. *Deep Sea Research Part II: Topical Studies in Oceanography*, 52, 2522–2542.

[13] Angert, A.L., Huxman, T.E., Chesson, P. & Venable, D.L. (2009). Functional tradeoffs determine species coexistence via the storage effect. *Proceedings of the National Academy of Sciences*, 106, 11641–11645.

[14] Archibald, K.M., Sosik, H.M. & Neubert, M.G. (2019). Preference and switching in the kill-the-winner functional response: diversity, size structure, and synergistic grazing in plankton models. *bioRxiv*. doi:10.1101/848564.

[15] Armstrong, R.A. & McGehee, R. (1976). Coexistence of species competing for shared resources. *Theoretical Population Biology*, 9, 317–328.

[16] Armstrong, R.A. & McGehee, R. (1980). Competitive exclusion. *The American Naturalist*, 115, 151–170.

[17] Arnoldi, J.F., Loreau, M. & Haegeman, B. (2019). The inherent multidimensionality of temporal variability: how common and rare species shape stability patterns. *Ecology Letters*, 22, 1557–1567.

[18] Arnott, R.N., Cherif, M., Bryant, L.D. & Wain, D.J. (2021). Artificially generated turbulence: a review of phycological nanocosm, microcosm, and mesocosm experiments. *Hydrobiologia*, 848, 961–991.

[19] Arrieta, J., Jeanneret, R., Roig, P. & Tuval, I. (2020). On the fate of sinking diatoms: the transport of active buoyancy-regulating cells in the ocean. *Philosophical Transactions of the Royal Society A*, 378, 20190529.

[20] Ascione Kenov, I., Muttin, F., Campbell, R., Fernandes, R., Campuzano, F., Machado, F., Franz, G. & Neves, R. (2015). Water fluxes and renewal rates at Pertuis d'Antioche/Marennes-Oléron Bay, France. *Estuarine, Coastal and Shelf Science*, 167, 32–44.

[21] Azanza, R.V., Brosnahan, M.L., Anderson, D.M., Hense, I. & Montresor, M. (2018). The role of life cycle characteristics in harmful algal bloom dynamics. In: *Global Ecology and Oceanography of Harmful Algal Blooms* (eds. Glibert, P.M., Berdalet, E., Burford, M.A., Pitcher, G.C. & Zhou, M.). Springer International Publishing, Cham, vol. 232, pp. 133–161.

[22] Baddeley, A., Rubak, E. & Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R.* Chapman and Hall/CRC Press, London.

[23] Bagchi, R., Gallery, R.E., Gripenberg, S., Gurr, S.J., Narayan, L., Addis, C.E., Freckleton, R.P. & Lewis, O. (2014). Pathogens and insect herbivores drive rainforest plant diversity and composition. *Nature*, 506, 85–88.

[24] Bainbridge, R. (1957). The size, shape and density of marine phytoplankton concentrations. *Biological Reviews*, 32, 91–115.

[25] Baker, M. (2015). Over half of psychology studies fail reproducibility test. *Nature News*, 27.

[26] Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533, 452–454.

[27] Barabás, G., D'Andrea, R., Rael, R., Meszéna, G. & Ostling, A. (2013). Emergent neutrality or hidden niches? *Oikos*, 122, 1565–1572.

[28] Barabás, G., Michalska-Smith, M.J. & Allesina, S. (2016). The effect of intra- and interspecific competition on coexistence in multispecies communities. *The American Naturalist*, 188, E1–E12.

[29] Barabás, G., Michalska-Smith, M.J. & Allesina, S. (2017). Self-regulation and the stability of large ecological networks. *Nature Ecology & Evolution*, 1, 1870–1875.

[30] Barraquand, F. & Murrell, D.J. (2012). Evolutionarily stable consumer home range size in relation to resource demography and consumer spatial organization. *Theoretical Ecology*, 5, 567–589.

[31] Barraquand, F. & Murrell, D.J. (2012). Intense or spatially heterogeneous predation can select against prey dispersal. *PLOS ONE*, 7, e28924.

[32] Barraquand, F., Picoche, C., Maurer, D., Carassou, L. & Auby, I. (2018). Coastal phytoplankton community dynamics and coexistence driven by intragroup density-dependence, light and hydrodynamics. *Oikos*, 127, 1834–1852.

[33] Barton, A.D., Ward, B.A., Williams, R.G. & Follows, M.J. (2014). The impact of fine-scale turbulence on phytoplankton community structure. *Limnology and Oceanography: Fluids and Environments*, 4, 34–49.

[34] Batifoulier, F., Lazure, P., Velo-Suarez, L., Maurer, D., Bonneton, P., Charria, G., Dupuy, C. & Gentien, P. (2013). Distribution of *Dinophysis* species in the Bay of Biscay and possible transport pathways to Arcachon Bay. *Journal of Marine Systems*, 109-110, S273–S283.

[35] Beardall, J., Allen, D., Bragg, J., Finkel, Z.V., Flynn, K.J., Quigg, A., Rees, T.A.V., Richardson, A. & Raven, J.A. (2009). Allometry and stoichiometry of unicellular, colonial and multicellular phytoplankton. *New Phytologist*, 181, 295–309.

[36] Bec, B., Collos, Y., Vaquer, A., Mouillot, D. & Souchu, P. (2008). Growth rate peaks at intermediate cell size in marine photosynthetic picoeukaryotes. *Limnology and Oceanography*, 53, 863–867.

[37] Begley, C.G. & Ellis, L.M. (2012). Raise standards for preclinical cancer research. *Nature*, 483, 531–533.

[38] Behrenfeld, M.J. & Boss, E.S. (2014). Resurrecting the ecological underpinnings of ocean plankton blooms. *Annual Review of Marine Science*, 6, 167–194.

[39] Behrenfeld, M.J. & Boss, E.S. (2017). Student's tutorial on bloom hypotheses in the context of phytoplankton annual cycles. *Global Change Biology*, 24, 55–77.

[40] Behringer, G., Ochsenkühn, M.A., Fei, C., Fanning, J., Koester, J.A. & Amin, S.A. (2018). Bacterial communities of diatoms display strong conservation across strains and time. *Frontiers in Microbiology*, 9.

[41] Benczik, I.J., Károlyi, G., Scheuring, I. & Tél, T. (2006). Coexistence of inertial competitors in chaotic flows. *Chaos*, 16, 043110.

[42] Benincà, E., Huisman, J., Heerkloss, R., Jöhnk, K.D., Branco, P., Van Nes, E.H., Scheffer, M. & Ellner, S.P. (2008). Chaos in a long-term experiment with a plankton community. *Nature*, 451, 822–825.

[43] Berenbaum, M.R. (2021). On zombies, struldbrugs, and other horrors of the scientific literature. *Proceedings of the National Academy of Sciences*, 118, e2111924118.

[44] Bergkvist, J., Thor, P., Jakobsen, H.H., Wängberg, S.A. & Selander, E. (2012). Grazer-induced chain length plasticity reduces grazing risk in a marine diatom. *Limnology and Oceanography*, 57, 318–324.

[45] Bersier, L.F., Banasek-Richter, C. & Cattin, M.F. (2002). Quantitative descriptors of food-web matrices. *Ecology*, 83, 2394.

[46] American Society for Cell Biology (2015). ASCB member survey on reproducibility. https://www.ascb.org/wp-content/uploads/2015/11/final-survey-results-without-Q11.pdf [last accessed: June, 24th 2022].

[47] Birch, D.A. & Young, W.R. (2006). A master equation for a spatial population model with pair interactions. *Theoretical Population Biology*, 70, 26–42.

[48] Bissinger, J.E., Montagnes, D.J.S., 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.

[49] Bjærke, O., Jonsson, P.R., Alam, A. & Selander, E. (2015). Is chain length in phytoplankton regulated to evade predation? *Journal of Plankton Research*, 37, 1110–1119.

[50] Bolker, B., Pacala, S. & Neuhauser, C. (2003). Spatial dynamics in model plant communities: What do we really know? *The American Naturalist*, 162, 135–148.

[51] Bolker, B.M. & Pacala, S.W. (1997). Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical Population Biology*, 52, 179–197.

[52] Bolker, B.M. & Pacala, S.W. (1999). Spatial moment equations for plant competition: understanding spatial strategies and the advantages of short dispersal. *The American Naturalist*, 153, 575–602.

[53] Borgnino, M., Arrieta, J., Boffetta, G., De Lillo, F. & Tuval, I. (2019). Turbulence induces clustering and segregation of non-motile, buoyancy-regulating phytoplankton. *Journal of the Royal Society Interface*, 16, 20190324.

[54] Boucher, O., Servonnat, J., Albright, A.L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.Y., Guenet, B., Guez Lionel, E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A.K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N. & Vuichard, N. (2020). Presentation and evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS002010.

[55] Bouderbala, I., El Saadi, N., Bah, A. & Auger, P. (2018). A 3D individual-based model to study effects of chemotaxis, competition and diffusion on the motile-phytoplankton aggregation. *Acta Biotheoretica*, 66, 257–278.

[56] Bouderbala, I., El Saadi, N., Bah, A. & Auger, P. (2019). A simulation study on how the resource competition and anti-predator cooperation impact the motile-phytoplankton groups' formation under predation stress. *Ecological Modelling*, 391, 16–28.

[57] Bouderbala, I., El Saadi, N., Bah, A. & Auger, P. (2020). Understanding how the collective behaviour of phytoflagellates is affected by light attenuation and diel vertical migration using individual-based modelling. *Journal of Theoretical Biology*, 494, 110241.

[58] Bowie, A.R., Maldonado, M.T., Frew, R.D., Croot, P.L., Achterberg, E.P., Mantoura, R.C., Worsfold, P.J., Law, C.S. & Boyd, P.W. (2001). The fate of added iron during a mesoscale fertilisation experiment in the Southern Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 48, 2703–2743.

[59] Breier, R.E., Lalescu, C.C., Waas, D., Wilczek, M. & Mazza, M.G. (2018). Emergence of phytoplankton patchiness at small scales in mild turbulence. *Proceedings of the National Academy of Sciences*, 115, 12112–12117.

[60] Breitbart, M. (2012). Marine viruses: truth or dare. *Annual Review of Marine Science*, 4, 425–448.

[61] Breton, E., Christaki, U., Sautour, B., Demonio, O., Skouroliakou, D.I., Beaugrand, G., Seuront, L., Kléparski, L., Poquet, A., Nowaczyk, A., Crouvoisier, M., Ferreira, S., Pecqueur, D., Salmeron, C., Brylinski, J.M., Lheureux, A. & Goberville, E. (2021). Seasonal variations in the biodiversity, ecological strategy, and specialization of diatoms and copepods in a coastal system with *Phaeocystis* blooms: the key role of trait trade-offs. *Frontiers in Marine Science*, 8, 1178.

[62] Brooker, R.W., Maestre, F.T., Callaway, R.M., Lortie, C.L., Cavieres, L.A., Kunstler, G., Liancourt, P., Tielbörger, K., Travis, J.M.J., Anthelme, F., Armas, C., Coll, L., Corcket, E., Delzon, S., Forey, E., Kikvidze, Z., Olofsson, J., Pugnaire, F., Quiroz, C.L., Saccone, P., Schiffers, K., Seifan, M., Touzard, B. & Michalet, R. (2008). Facilitation in plant communities: the past, the present, and the future. *Journal of Ecology*, 96, 18–34.

[63] Brussaard, C.P.D. (2004). Phytoplankton cell lysis and its ecological implications. *Journal of Eukaryotic Microbiology*, 51, 125–138.

[64] Bulmer, M.G. (1984). Delayed germination of seeds: Cohen's model revisited. *Theoretical Population Biology*, 26, 367–377.

[65] Burson, A., Stomp, M., Greenwell, E., Grosse, J. & Huisman, J. (2018). Competition for nutrients and light: testing advances in resource competition with a natural phytoplankton community. *Ecology*, 99, 1108–1118.

[66] Busseni, G., Caputi, L., Piredda, R., Fremont, P., Hay Mele, B., Campese, L., Scalco, E., de Vargas, C., Bowler, C., d'Ovidio, F., Zingone, A., Ribera d'Alcalà, M. & Iudicone, D. (2020). Large scale patterns of marine diatom richness: drivers and trends in a changing ocean. *Global Ecology and Biogeography*, 29, 1915–1928.

[67] Cáceres, C.E. (1997). Temporal variation, dormancy, and coexistence: a field test of the storage effect. *Proceedings of the National Academy of Sciences*, 94, 9171–9175.

[68] Cadier, M., Gorgues, T., L'Helguen, S., Sourisseau, M. & Memery, L. (2017). Tidal cycle control of biogeochemical and ecological properties of a macrotidal ecosystem. *Geophysical Research Letters*, 44, 8453–8462.

[69] Calbet, A. & Landry, M.R. (2004). Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. *Limnology and Oceanography*, 49, 51–57.

[70] Cavieres, L.A. & Badano, E.I. (2009). Do facilitative interactions increase species richness at the entire community level? *Journal of Ecology*, 97, 1181–1191.

[71] 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.

[72] Chamberlain, S.A., Bronstein, J.L. & Rudgers, J.A. (2014). How context dependent are species interactions? *Ecology Letters*, 17, 881–890.

[73] Chang, C.W., Miki, T., Ye, H., Souissi, S., Adrian, R., Anneville, O., Agasild, H., Ban, S., Be'eri-Shlevin, Y., Chiang, Y.R., Feuchtmayr, H., Gal, G., Ichise, S., Kagami, M., Kumagai, M., Liu, X., Matsuzaki, S.I.S., Manca, M.M., Nōges, P., Piscia, R., Rogora, M., Shiah, F.K., Thackeray, S.J., Widdicombe, C.E., Wu, J.T., Zohary, T. & Hsieh, C.h. (2022). Causal networks of phytoplankton diversity and biomass are modulated by environmental context. *Nature Communications*, 13, 1140.

[74] Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual review of Ecology and Systematics*, 31, 343–366.

[75] Chesson, P. (2003). Quantifying and testing coexistence mechanisms arising from recruitment fluctuations. *Theoretical population biology*, 64, 345–357.

[76] Chesson, P. (2018). Updates on mechanisms of maintenance of species diversity. *Journal of Ecology*, 106, 1773–1794.

[77] Chesson, P. & Huntly, N. (1988). Community consequences of life-history traits in a variable environment. *Annales Zoologici Fennici*, 25, 5–16.

[78] Chesson, P. & Huntly, N. (1997). The roles of harsh and fluctuating conditions in the dynamics of ecological communities. *The American Naturalist*, 150, 519–553.

[79] Chu, C. & Adler, P.B. (2015). Large niche differences emerge at the recruitment stage to stabilize grassland coexistence. *Ecological Monographs*, 85, 373–392.

[80] Clark, P.J. & Evans, F.C. (1979). Generalization of a nearest neighbor measure of dispersion for use in k dimensions. *Ecology*, 60, 316–317.

[81] Cohen, D. (1966). Optimizing reproduction in a randomly varying environment. *Journal of Theoretical Biology*, 12, 119–129.

[82] Cohen, J. & Newman, C. (1984). The stability of large random matrices and their products. *The Annals of Probability*, 12, 283–310.

[83] Cohn, S.A., Halpin, D., Hawley, N., Ismail, A., Kaplan, Z., Kordes, T., Kuhn, J., Macke, W., Marhaver, K., Ness, B., Olszewski, S., Pike, A., Rice, E., Sbarboro, J., Wolske, A. & Zapata, Y. (2015). Comparative analysis of light-stimulated motility responses in three diatom species. *Diatom Research*, 30, 213–225.

[84] Cohn, S.A. & McGuire, J.R. (2000). Using diatom motility as an indicator of environmental stress: effects of toxic sediment elutriates. *Diatom Research*, 15, 19–29.

[85] Comita, L.S., Muller-Landau, H.C., Aguilar, S. & Hubbell, S.P. (2010). Asymmetric density dependence shapes species abundances in a tropical tree community. *Science*, 329, 330–332.

[86] Comita, L.S., Queenborough, S.A., Murphy, S.J., Eck, J.L., Xu, K., Krishnadas, M., Beckman, N. & Zhu, Y. (2014). Testing predictions of the Janzen-Connell hypothesis: a meta-analysis of experimental evidence for distance- and density-dependent seed and seedling survival. *Journal of Ecology*, 102, 845–856.

[87] Coyte, K.Z., Schluter, J. & Foster, K.R. (2015). The ecology of the microbiome: networks, competition, and stability. *Science*, 350, 663–666.

[88] Cressie, N. (1993). *Statistics for spatial data.* John Wiley & Sons.

[89] Culina, A., van den Berg, I., Evans, S. & Sánchez-Tójar, A. (2020). Low availability of code in ecology: a call for urgent action. *PLOS Biology*, 18, e3000763.

[90] Cushing, J.M., Levarge, S., Chitnis, N. & Henson, S.M. (2004). Some discrete competition models and the competitive exclusion principle. *Journal of Difference Equations and Applications*, 10, 1139–1151.

[91] Dakos, V., Benincà, E., van Nes, E.H., Philippart, C.J.M., Scheffer, M. & Huisman, J. (2009). Interannual variability in species composition explained as seasonally entrained chaos. *Proceedings of the Royal Society B: Biological Sciences*, 276, 2871–2880.

[92] Daly, K.L. & Smith, W. (1993). Physical-biological interactions influencing marine plankton production. *Annual Review of Ecology and Systematics*, 24, 555–585.

[93] Dell'Aquila, G., Ferrante, M.I., Gherardi, M., Lagomarsino, M.C., d'Alcalà, M.R., Iudicone, D. & Amato, A. (2017). Nutrient consumption and chain tuning in diatoms exposed to storm-like turbulence. *Scientific Reports*, 7, 1828.

[94] DeMalach, N., Kigel, J. & Sternberg, M. (2021). The soil seed bank can buffer long-term compositional changes in annual plant communities. *Journal of Ecology*, 109, 1275–1283.

[95] Denman, K.L. & Platt, T. (1976). The variance spectrum of phytoplankton in a turbulent ocean. *Journal of Marine Research*, 34, 593–601.

[96] Descamps-Julien, B. & Gonzalez, A. (2005). Stable coexistence in a fluctuating environment: an experimental demonstration. *Ecology*, 86, 2815–2824.

[97] Detto, M. & Muller-Landau, H.C. (2016). Stabilization of species coexistence in spatial models through the aggregation-segregation effect generated by local dispersal and nonspecific local interactions. *Theoretical Population Biology*, 112, 97–108.

[98] Detto, M., Visser, M.D., Wright, S.J. & Pacala, S.W. (2019). Bias in the detection of negative density dependence in plant communities. *Ecology Letters*, 22, 1923–1939.

[99] Dieckmann, U. & Law, R. (2000). Relaxation projections and the method of moments. In: *The Geometry of Ecological Interactions* (eds. Dieckmann, U., Law, R. & Metz, J.A.J.). Cambridge University Press, pp. 412–455. 1st edn.

[100] Doglioli, A.M., Grégori, G., Marrec, P., Thyssen, M., Wagener, T., Rougier, G., Bhairy, N., Andre, J.M., Berline, L., Cyr, F., Deverneil, A., Duforet-Gaurier, L., Duhaut, T., Estournel, C., Goutx, M., Marsaleix, P., Meriaux, X., Nencioli, F., Petrenko, A., Pinazo, C., Ross, O.N.N., Rousselet, L., Yohia, C. & Zakardjian, B. (2016). OSCAHR (Observing Submesoscale Coupling At High Resolution) cruise report. Research report, MIO-Mediterranean Institute of Oceanography, Aix-Marseille Université.

[101] Doubell, M.J., Seuront, L., Seymour, J.R., Patten, N.L. & Mitchell, J.G. (2006). High-resolution fluorometer for mapping microscale phytoplankton distributions. *Applied and Environmental Microbiology*, 72, 4475–4478.

[102] Durham, W.M., Climent, E., Barry, M., De Lillo, F., Boffetta, G., Cencini, M. & Stocker, R. (2013). Turbulence drives microscale patches of motile phytoplankton. *Nature Communications*, 4, 2148.

[103] Durham, W.M. & Stocker, R. (2012). Thin phytoplankton layers: characteristics, mechanisms, and consequences. *Annual Review of Marine Science*, 4, 177–207.

[104] Dusenbery, D.B. (2009). *Living at micro scale: the unexpected physics of being small*. Harvard University Press.

[105] Edlund, M.B. & Stoermer, E.F. (1997). Ecological, evolutionary and systematic significance of diatom life histories. *Journal of Phycology*, 33, 897–918.

[106] Edwards, K.F. (2019). Mixotrophy in nanoflagellates across environmental gradients in the ocean. *Proceedings of the National Academy of Sciences*, p. 201814860.

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

[108] Edwards, K.F., Thomas, M.K., Klausmeier, C.A. & 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.

[109] Eilertsen, H.C. & Wyatt, T. (2000). Phytoplankton models and life history strategies. *South African Journal of Marine Science*, 22, 323–337.

[110] Einstein, A. (1905). Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der physik*, 4.

[111] Ellegaard, M. & Ribeiro, S. (2018). The long-term persistence of phytoplankton resting stages in aquatic 'seed banks'. *Biological Reviews*, 93, 166–183.

[112] Ellner, S. (1987). Alternate plant life history strategies and coexistence in randomly varying environments. *Vegetatio*, 69, 199–208.

[113] Ellner, S.P., Snyder, R.E. & Adler, P.B. (2016). How to quantify the temporal storage effect using simulations instead of math. *Ecology Letters*, 19, 1333–1342.

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

[115] Estrada, M., Alcaraz, M. & Marrasé, C. (1987). Effects of turbulence on the composition of phytoplankton assemblages in marine microcosms. *Marine Ecology Progress Series*, 38, 267–281.

[116] Estrada, M., Solé, J., Anglès, S. & Garcés, E. (2010). The role of resting cysts in *Alexandrium minutum* population dynamics. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57, 308–321.

[117] Facelli, J.M., Chesson, P. & Barnes, N. (2005). Differences in seed biology of annual plants in arid lands: a key ingredient of the storage effect. *Ecology*, 86, 2998–3006.

[118] Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLOS ONE*, 5, e10068.

[119] Fanelli, D. (2018). Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*, 115, 2628–2631.

[120] Felpeto, A.B., Roy, S. & Vasconcelos, V.M. (2018). Allelopathy prevents competitive exclusion and promotes phytoplankton biodiversity. *Oikos*, 127, 85–98.

[121] Field, C.B., Behrenfeld, M.J., Randerson, J.T. & Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, 281, 237–240.

[122] Flynn, K.J., Mitra, A., Wilson, W.H., Kimmance, S.A., Clark, D.R., Pelusi, A. & Polimene, L. (2022). 'Boom-and-busted' dynamics of phytoplankton-virus interactions explain the paradox of the plankton. *New Phytologist*, 234, 990–1002.

[123] Font-Muñoz, J.S., Jeanneret, R., Arrieta, J., Anglès, S., Jordi, A., Tuval, I. & Basterretxea, G. (2019). Collective sinking promotes selective cell pairing in planktonic pennate diatoms. *Proceedings of the National Academy of Sciences*, 116, 15997–16002.

[124] Font-Muñoz, J.S., Jordi, A., Tuval, I., Arrieta, J., Anglès, S. & Basterretxea, G. (2017). Advection by ocean currents modifies phytoplankton size structure. *Journal of the Royal Society Interface*, 14, 20170046.

[125] Fox, J.W. (2013). The intermediate disturbance hypothesis should be abandoned. *Trends in Ecology & Evolution*, 28, 86–92.

[126] Franks, P.J.S. (2002). NPZ models of plankton dynamics: their construction, coupling to physics, and application. *Journal of Oceanography*, 58, 379–387.

[127] Fransz, H.G. & Verhagen, J.H.G. (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.

[128] Fraser, H., Parker, T., Nakagawa, S., Barnett, A. & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLOS ONE*, 13, e0200303.

[129] Frenken, T., Alacid, E., Berger, S.A., Bourne, E.C., Gerphagnon, M., Grossart, H.P., Gsell, A.S., Ibelings, B.W., Kagami, M., Küpper, F.C., Letcher, P.M., Loyau, A., Miki, T., Nejstgaard, J.C., Rasconi, S., Reñé, A., Rohrlack, T., Rojas-Jimenez, K., Schmeller, D.S., Scholz, B., Seto, K., Sime-Ngando, T., Sukenik, A., Van de Waal, D.B., Van den Wyngaert, S., Van Donk, E., Wolinska, J., Wurzbacher, C. & Agha, R. (2017). Integrating chytrid fungal parasites into plankton ecology. research gaps and needs. *Environmental Microbiology*, 19.

[130] Fuchs, H.L. & Gerbi, G.P. (2016). Seascape-level variation in turbulence- and wave-generated hydrodynamic signals experienced by plankton. *Progress in Oceanography*, 141, 109–129.

[131] Fujiwara, M., Pfeiffer, G., Boggess, M., Day, S. & Walton, J. (2011). Coexistence of competing stage-structured populations. *Scientific Reports*, 1.

[132] García-Callejas, D., Molowny-Horas, R. & Araújo, M. (2018). The effect of multiple biotic interaction types on species persistence. *Ecology*, 99, 2327–2337.

[133] Gentleman, W., Leising, A., Frost, B., Strom, S. & Murray, J. (2003). Functional responses for zooplankton feeding on multiple resources: a review of assumptions and biological dynamics. *Deep Sea Research Part II: Topical Studies in Oceanography*, 50, 2847–2875.

[134] Gerhard, M., Koussoroplis, A.M., Hillebrand, H. & Striebel, M. (2019). Phytoplankton community responses to temperature fluctuations under different nutrient concentrations and stoichiometry. *Ecology*, 100, e02834.

[135] Gómez, F. (2012). A quantitative review of the lifestyle, habitat and trophic diversity of dinoflagellates (Dinoflagellata, Alveolata). *Systematics and Biodiversity*, 10, 267–275.

[136] Graham, L.E., Graham, J.M., Wilcox, L.W. & Cook, M.E. (2016). *Algae*. LJLM press.

[137] Griffiths, J.R., Hajdu, S., Downing, A.S., Hjerne, O., Larsson, U. & Winder, M. (2015). Phytoplankton community interactions and environmental sensitivity in coastal and offshore habitats. *Oikos*, 125, 1134–1143.

[138] Gross, K. (2008). Positive interactions among competitors can produce species-rich communities. *Ecology Letters*, 11, 929–936.

[139] Guittar, J., Goldberg, D., Klanderud, K., Berge, A., Boixaderas, M.R., Meineri, E., Töpper, J. & Vandvik, V. (2020). Quantifying the roles of seed dispersal, filtering, and climate on regional patterns of grassland biodiversity. *Ecology*, 101, e03061.

[140] Haegeman, B. & Rapaport, A. (2008). How flocculation can explain coexistence in the chemostat. *Journal of Biological Dynamics*, 2, 1–13.

[141] Hampton, S.E., Holmes, E.E., Scheef, L.P., Scheuerell, M.D., Katz, S.L., Pendleton, D.E. & Ward, E.J. (2013). Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. *Ecology*, 94, 2663–2669.

[142] Hampton, S.E., Izmest'Eva, L.R., Moore, M.V., Katz, S.L., Dennis, B. & Silow, E.A. (2008). Sixty years of environmental change in the world's largest freshwater lake - Lake Baikal, Siberia. *Global Change Biology*, 14, 1947–1958.

[143] Hampton, S.E., Scheuerell, M.D. & Schindler, D.E. (2006). Coalescence in the Lake Washington story: interaction strengths in a planktonic food web. *Limnology and Oceanography*, 51, 2042–2051.

[144] Hampton, S.E. & Schindler, D.E. (2006). Empirical evaluation of observation scale effects in community time series. *Oikos*, 113, 424–439.

[145] Hardin, G. (1960). The competitive exclusion principle. *Science*, 131, 1292–1297.

[146] Hart, S.P., Usinowicz, J. & Levine, J.M. (2017). The spatial scales of species coexistence. *Nature Ecology & Evolution*, 1, 1066–1073.

[147] Hartig, F. & Barraquand, F. (2022). The evidence contained in the P-value is context dependent. *Trends in Ecology & Evolution*, 37, 569–570.

[148] Haydon, D. (1994). Pivotal assumptions determining the relationship between stability and complexity: an analytical synthesis of the stability-complexity debate. *The American Naturalist*, 144, 14–29.

[149] Head, M.L., Holman, L., Lanfear, R., Kahn, A.T. & Jennions, M.D. (2015). The extent and consequences of P-hacking in science. *PLOS Biology*, 13, e1002106.

[150] Hellweger, F.L. & Bucci, V. (2009). A bunch of tiny individuals–individual-based modeling for microbes. *Ecological Modelling*, 220, 8–22.

[151] Hellweger, F.L., Clegg, R.J., Clark, J.R., Plugge, C.M. & Kreft, J.U. (2016). Advancing microbial sciences by individual-based modelling. *Nature Reviews Microbiology*, 14, 461–471.

[152] Hellweger, F.L., Kravchuk, E.S., Novotny, V. & Gladyshev, M.I. (2008). Agent-based modeling of the complex life cycle of a cyanobacterium (*Anabaena*) in a shallow reservoir. *Limnology and Oceanography*, 53, 1227–1241.

[153] Hense, I. & Beckmann, A. (2006). Towards a model of cyanobacteria life cycle-effects of growing and resting stages on bloom formation of N_2-fixing species. *Ecological Modelling*, 195, 205–218.

[154] Hernández Fariñas, T., Bacher, C., Soudant, D., Belin, C. & Barillé, L. (2015). Assessing phytoplankton realized niches using a French national phytoplankton monitoring network. *Estuarine, Coastal and Shelf Science*, 159, 15–27.

[155] Hernández-Fariñas, T., Soudant, D., Barillé, L., Belin, C., Lefebvre, A. & Bacher, C. (2014). Temporal changes in the phytoplankton community along the French coast of the eastern English Channel and the southern Bight of the North Sea. *ICES Journal of Marine Science*, 71, 821–833.

[156] Hill, P.S., Nowell, A.R.M. & Jumars, P.A. (1992). Encounter rate by turbulent shear of particles similar in diameter to the Kolmogorov scale. *Journal of Marine Research*, 50, 643–668.

[157] Hocquet, A. & Wieber, F. (2021). Epistemic issues in computational reproducibility: software as the elephant in the room. *European Journal for Philosophy of Science*, 11, 38.

[158] Holmes, E., Ward, E. & Scheuerell, M. (2014). *MARSS: Multivariate Autoregressive State-Space Modeling*. R package version 3.9 https://CRAN.R-project.org/package=MARSS.

[159] Holmes, E.E., Ward, E.J. & Wills, K. (2012). MARSS: Multivariate autoregressive state-space models for analyzing time-series data. *The R Journal*, 4, 30.

[160] Holt, R. (2006). Emergent neutrality. *Trends in Ecology & Evolution*, 21, 531–533.

[161] Huang, Z., Liu, S., Bradford, K.J., Huxman, T.E. & Venable, D.L. (2016). The contribution of germination functional traits to population dynamics of a desert plant community. *Ecology*, 97, 250–261.

[162] Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press.

[163] Huber, V. & Gaedke, U. (2006). The role of predation for seasonal variability patterns among phytoplankton and ciliates. *Oikos*, 114, 265–276.

[164] Huisman, J., Johansson, A.M., Folmer, E.O. & Weissing, F.J. (2001). Towards a solution of the plankton paradox: the importance of physiology and life history. *Ecology Letters*, 4, 408–411.

[165] Huisman, J. & Weissing, F.J. (1999). Biodiversity of plankton by species oscillations and chaos. *Nature*, 402, 407–410.

[166] Huisman, J. & Weissing, F.J. (2001). Biological conditions for oscillations and chaos generated by multispecies competition. *Ecology*, 82, 2682–2695.

[167] Huston, M., DeAngelis, D. & Post, W. (1988). New computer models unify ecological theory. *BioScience*, 38, 682–691.

[168] Hutchinson, G.E. (1961). The paradox of the plankton. *The American Naturalist*, 95, 137–145.

[169] Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. John Wiley & Sons.

[170] Ioannidis, J.P.A. (2005). Why most published research findings are false. *PLOS Medicine*, 2, e124.

[171] Ismail, K. & Al-Shehhi, M.R. (2022). Reviews and syntheses: assessment of biogeochemical models in the marine environment. *Biogeosciences Discussions*.

[172] Ives, A., Gross, K. & Klug, J. (1999). Stability and variability in competitive communities. *Science*, 286, 542–544.

[173] Ives, A.R. & Carpenter, S.R. (2007). Stability and diversity of ecosystems. *Science*, 317, 58–62.

[174] Ives, A.R., Carpenter, S.R. & Dennis, B. (1999). Community interaction webs and zooplankton responses to planktivory manipulations. *Ecology*, 80, 1405–1421.

[175] Ives, A.R., Dennis, B., Cottingham, K.L. & Carpenter, S.R. (2003). Estimating community stability and ecological interactions from time-series data. *Ecological monographs*, 73, 301–330.

[176] Jabot, F. & Pottier, J. (2017). Macroecology of seed banks: the role of biogeography, environmental stochasticity and sampling. *Global Ecology and Biogeography*, 26, 1247–1257.

[177] Jackson, G.A. (1987). Simulating chemosensory responses of marine microorganisms. *Limnology and Oceanography*, 32, 1253–1266.

[178] Jacquet, C., Moritz, C., Morissette, L., Legagneux, P., Massol, F., Archambault, P. & Gravel, D. (2016). No complexity-stability relationship in empirical ecosystems. *Nature Communications*, 7.

[179] Jager, L.R. & Leek, J.T. (2014). An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics*, 15, 1–12.

[180] James, A., Plank, M.J., Rossberg, A.G., Beecham, J., Emmerson, M. & Pitchford, J.W. (2015). Constructing random matrices to represent real ecosystems. *The American Naturalist*, 185, 680–692.

[181] Jamet, J.L., Boge, G., Richard, S., Geneys, C. & Jamet, D. (2001). The zooplankton community in bays of Toulon area (northwest Mediterranean Sea, France). *Hydrobiologia*, 557, 155–165.

[182] 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.

[183] Johansson, O.N., Pinder, M.I.M., Ohlsson, F., Egardt, J., Töpel, M. & Clarke, A.K. (2019). Friends with benefits: exploring the phycosphere of the marine diatom *Skeletonema marinoi*. *Frontiers in Microbiology*, 10.

[184] Jones, S.E. & Lennon, J.T. (2010). Dormancy contributes to the maintenance of microbial diversity. *Proceedings of the National Academy of Sciences*, 107, 5881–5886.

[185] Judson, O.P. (1994). The rise of the individual-based model in ecology. *Trends in Ecology & Evolution*, 9, 9–14.

[186] Jumars, P.A., Deming, J., Hill, P., Karp-Boss, L., Yager, P. & Dade, W. (1993). Physical constraints on marine osmotrophy in an optimal foraging context. *Marine Microbial Food Webs*, 7, 121–159.

[187] Karp-Boss, L., Boss, E. & Jumars, P.A. (1996). Nutrient fluxes to planktonic osmotrophs in the presence of fluid motion. *Oceanography and Marine Biology: An Annual Review*, 34, 71–107.

[188] Keil, P., Wiegand, T., Tóth, A.B., McGlinn, D.J. & Chase, J.M. (2021). Measurement and analysis of interspecific spatial associations as a facet of biodiversity. *Ecological Monographs*, 91, e01452.

[189] Kenitz, K.M., Orenstein, E.C., Roberts, P.L.D., Franks, P.J.S., Jaffe, J.S., Carter, M.L. & Barton, A.D. (2020). Environmental drivers of population variability in colony-forming marine diatoms. *Limnology and Oceanography*, 65, 2515–2528.

[190] Kierstead, H. & Slobodkin, L.B. (1953). The size of water masses containing plankton blooms. *Journal of Marine Research*, 12, 141–147.

[191] Kinlock, N.L. (2019). A meta-analysis of plant interaction networks reveals competitive hierarchies as well as facilitation and intransitivity. *The American Naturalist*, 194, 640–653.

[192] Kiørboe, T. (2018). *A mechanistic approach to plankton ecology*. Princeton University Press.

[193] Kiørboe, T., Andersen, K.P. & Dam, H.G. (1990). Coagulation efficiency and aggregate formation in marine phytoplankton. *Marine Biology*, 107, 235–245.

[194] Klug, J.L. & Cottingham, K.L. (2001). Interactions among environmental drivers: community responses to changing nutrients and dissolved organic carbon. *Ecology*, 82, 3390–3403.

[195] Klug, J.L., Fischer, J.M., Ives, A.R. & Dennis, B. (2000). Compensatory dynamics in planktonic community responses to pH perturbations. *Ecology*, 81, 387–398.

[196] Kociolek, J.P. & Wynne, M.J. (1988). Observations on *Navicula thallodes* (Bacillariophyceae), a blade-forming diatom from the Bering Sea. *Journal of Phycology*, 24, 439–441.

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

[198] Kraft, N.J.B., Godoy, O. & Levine, J.M. (2015). Plant functional traits and the multidimensional nature of species coexistence. *Proceedings of the National Academy of Sciences*, 112, 797–802.

[199] Kraichnan, R.H. (1970). Convergents to turbulence functions. *Journal of Fluid Mechanics*, 41, 189–217.

[200] Kraichnan, R.H. (1974). Convection of a passive scalar by a quasi-uniform random straining field. *Journal of Fluid Mechanics*, 64, 737–762.

[201] Law, R., Murrell, D.J. & Dieckmann, U. (2003). Population growth in space and time: spatial logistic equations. *Ecology*, 84, 252–262.

[202] 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.

[203] Lee, S., Hofmeister, R. & Hense, I. (2018). The role of life cycle processes on phytoplankton spring bloom composition: a modelling study applied to the Gulf of Finland. *Journal of Marine Systems*, 178, 75–85.

[204] Lehahn, Y., d'Ovidio, F., Lévy, M., Amitai, Y. & Heifetz, E. (2011). Long range transport of a quasi isolated chlorophyll patch by an Agulhas ring. *Geophysical Research Letters*, 38.

[205] Leonard, C.L., Bidigare, R.R., Seki, M.P. & Polovina, J.J. (2001). Interannual mesoscale physical and biological variability in the North Pacific Central Gyre. *Progress in Oceanography*, 49, 227–244.

[206] Levine, J.M. & HilleRisLambers, J. (2009). The importance of niches for the maintenance of species diversity. *Nature*, 461, 254–257.

[207] Levine, J.M. & Rees, M. (2004). Effects of temporal variability on rare plant persistence in annual systems. *The American Naturalist*, 164, 350–363.

[208] Lévy, M., Franks, P.J.S. & Smith, K.S. (2018). The role of submesoscale currents in structuring marine ecosystems. *Nature Communications*, 9, 4758.

[209] Lévy, M., Jahn, O., Dutkiewicz, S. & Follows, M.J. (2014). Phytoplankton diversity and community structure affected by oceanic dispersal and mesoscale turbulence: dispersal impact on plankton diversity. *Limnology and Oceanography*, 4, 67–84.

[210] Lévy, M., Jahn, O., Dutkiewicz, S., Follows, M.J. & d'Ovidio, F. (2015). The dynamical landscape of marine phytoplankton diversity. *Journal of The Royal Society Interface*, 12, 20150481.

[211] Li, L. & Chesson, P. (2016). The effects of dynamical rates on species coexistence in a variable environment: the paradox of the plankton revisited. *The American Naturalist*, 188, E46–E58.

[212] 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.

[213] Lindegren, M., Möllmann, C., Nielsen, A. & Stenseth, N.C. (2009). Preventing the collapse of the Baltic cod stock through an ecosystem-based management approach. *Proceedings of the National Academy of Sciences*, 106, 14722–14727.

[214] Litchman, E., Klausmeier, C.A., Schofield, O.M. & Falkowski, P.G. (2007). The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level. *Ecology Letters*, 10, 1170–1181.

[215] Locey, K.J. (2010). Synthesizing traditional biogeography with microbial ecology: the importance of dormancy. *Journal of Biogeography*, 37, 1835–1841.

[216] Łomnicki, A. (1999). Individual-based models and the individual-based approach to population ecology. *Ecological Modelling*, 115, 191–198.

[217] Long, R.A. & Azam, F. (2001). Microscale patchiness of bacterioplankton assemblage richness in seawater. *Aquatic Microbial Ecology*, 26, 103–113.

[218] Lürling, M. (2021). Grazing resistance in phytoplankton. *Hydrobiologia*, 848, 237–249.

[219] MacArthur, R. & Levins, R. (1964). Competition, habitat selection and character displacement in a patchy environment. *Proceedings of the National Academy of Sciences*, 51, 1207–1210.

[220] Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., de Vargas, C., Bittner, L., Zingone, A. & Bowler, C. (2016). Insights into global diatom distribution and diversity in the world's ocean. *Proceedings of the National Academy of Sciences*, 113, E1516–E1525.

[221] Mann, D.G. & Vanormelingen, P. (2013). An inordinate fondness? The number, distributions, and origins of diatom species. *Journal of Eukaryotic Microbiology*, 60, 414–420.

[222] Marcus, N.H. & Boero, F. (1998). Minireview: The importance of benthic-pelagic coupling and the forgotten role of life cycles in coastal aquatic systems. *Limnology and Oceanography*, 43, 763–768.

[223] Margalef, R. (1978). Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanologica acta*, 1, 493–509.

[224] Marrec, P., Grégori, G., Doglioli, A.M., Dugenne, M., Della Penna, A., Bhairy, N., Cariou, T., Hélias Nunige, S., Lahbib, S., Rougier, G., Wagener, T. & Thyssen, M. (2018). Coupling physics and biogeochemistry thanks to high-resolution observations of the phytoplankton community structure in the northwestern Mediterranean Sea. *Biogeosciences*, 15, 1579–1606.

[225] Martin, A.P. (2003). Phytoplankton patchiness: the role of lateral stirring and mixing. *Progress in Oceanography*, 57, 125–174.

[226] Martorell, C. & Freckleton, R.P. (2014). Testing the roles of competition, facilitation and stochasticity on community structure in a species-rich assemblage. *Journal of Ecology*, 102, 74–85.

[227] May, R.M. (1972). Will a large complex system be stable? *Nature*, 238, 413–414.

[228] May, R.M. (1981). *Theoretical Ecology: Principles and Applications*. Oxford University Press UK.

[229] McGillicuddy, D.J., Anderson, D.M., Lynch, D.R. & Townsend, D.W. (2005). Mechanisms regulating large-scale seasonal fluctuations in *Alexandrium fundyense* populations in the Gulf of Maine: results from a physical-biological model. *Deep Sea Research Part II: Topical Studies in Oceanography*, 52, 2698–2714.

[230] McGillicuddy, D.J. & Franks, P.J.S. (2019). Models of plankton patchiness. In: *Encyclopedia of Ocean Sciences*. Elsevier, pp. 536–546.

[231] McIntire, E.J.B. & Fajardo, A. (2014). Facilitation as a ubiquitous driver of biodiversity. *New Phytologist*, 201, 403–416.

[232] McPeek, M.A. (2022). *Coexistence in Ecology: A Mechanistic Perspective*. Princeton University Press.

[233] 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.

[234] McQuoid, M.R. & Hobson, L.A. (1996). Diatom Resting Stages. *Journal of Phycology*, 32, 889–902.

[235] Ménesguen, A., Desmit, X., Dulière, V., Lacroix, G., Thouvenin, B., Thieu, V. & Dussauze, M. (2018). How to avoid eutrophication in coastal seas? A new approach to derive river-specific combined nitrate and phosphate maximum concentrations. *Science of The Total Environment*, 628-629, 400–414.

[236] Miller, E.T. & Klausmeier, C.A. (2017). Evolutionary stability of coexistence due to the storage effect in a two-season model. *Theoretical Ecology*, 10, 91–103.

[237] Mitchell, J.G., Yamazaki, H., Seuront, L., Wolk, F. & Li, H. (2008). Phytoplankton patch patterns: seascape anatomy in a turbulent ocean. *Journal of Marine Systems*, 69, 247–253.

[238] Modéran, J., Bouvais, P., David, V., Le Noc, S., Simon-Bouhet, B., Niquil, N., Miramand, P. & Fichet, D. (2010). Zooplankton community structure in a highly turbid environment (Charente estuary, France): Spatio-temporal patterns and environmental control. *Estuarine, Coastal and Shelf Science*, 88, 219–232.

[239] Moll, J. & Brown, J. (2008). Competition and coexistence with multiple life-history stages. *The American Naturalist*, 171, 839–843.

[240] Monin, A.S. & Yaglom, A. (2007). *Statistical fluid mechanics, Volume 1 & 2*. MIT Press.

[241] Montresor, M., Di Prisco, C., Sarno, D., Margiotta, F. & Zingone, A. (2013). Diversity and germination patterns of diatom resting stages at a coastal Mediterranean site. *Marine Ecology Progress Series*, 484, 79–95.

[242] Montresor, M., Zingone, A. & Sarno, D. (1998). Dinoflagellate cyst production at a coastal Mediterranean site. *Journal of Plankton Research*, 20, 2291–2312.

[243] Mougi, A. & Kondoh, M. (2012). Diversity of interaction types and ecological community stability. *Science*, 337, 349–351.

[244] Murrell, D.J. (2005). Local spatial structure and predator-prey dynamics: counterintuitive effects of prey enrichment. *The American Naturalist*, 166, 354–367.

[245] Murrell, D.J. (2010). When does local spatial structure hinder competitive coexistence and reverse competitive hierarchies? *Ecology*, 91, 1605–1616.

[246] Murrell, D.J. & Law, R. (2003). Heteromyopia and the spatial coexistence of similar competitors. *Ecology Letters*, 6, 48–59.

[247] Musielak, M.M., Karp-Boss, L., Jumars, P.A. & Fauci, L.J. (2009). Nutrient transport and acquisition by diatom chains in a moving fluid. *Journal of Fluid Mechanics*, 638, 401–421.

[248] Mutshinda, C.M., Finkel, Z.V., Widdicombe, C.E. & Irwin, A.J. (2016). Ecological equivalence of species within phytoplankton functional groups. *Functional Ecology*, 30, 1714–1722.

[249] Mutshinda, C.M., O'Hara, R.B. & Woiwod, I.P. (2009). What drives community dynamics? *Proceedings of the Royal Society B: Biological Sciences*, 276, 2923–2929.

[250] Narwani, A., Bentlage, B., Alexandrou, M.A., Fritschie, K.J., Delwiche, C., Oakley, T.H. & Cardinale, B.J. (2017). Ecological interactions and coexistence are predicted by gene expression similarity in freshwater green algae. *Journal of Ecology*, 105, 580–591.

[251] Naustvoll, L.J. (2000). Prey size spectra and food preferences in thecate heterotrophic dinoflagellates. *Phycologia*, 39, 187–198.

[252] Nazeer, M., Wong, M.S. & Nichol, J.E. (2017). A new approach for the estimation of phytoplankton cell counts associated with algal blooms. *Science of The Total Environment*, 590-591, 125–138.

[253] Ngan, K. & Vanneste, J. (2011). Scalar decay in a three-dimensional chaotic flow. *Physical Review E*, 83, 056306.

[254] North, A. & Ovaskainen, O. (2007). Interactions between dispersal, competition, and landscape heterogeneity. *Oikos*, 116, 1106–1119.

[255] Nuijten, M.B., Hartgerink, C.H.J., van Assen, M.A.L.M., Epskamp, S. & Wicherts, J.M. (2016). The prevalence of statistical reporting errors in psychology (1985-2013). *Behavior Research Methods*, 48, 1205–1226.

[256] Ohser, J. (1983). On estimators for the reduced second moment measure of point processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 14, 63–71.

[257] Okubo, A. (1978). Horizontal dispersion and critical scales for phytoplankton patches. In: *Spatial pattern in plankton communities* (ed. Steele, J.H.). Springer Science & Business Media.

[258] Okubo, A., Hastings, A. & Powell, T. (2001). Population dynamics in temporal and spatial domains. In: *Diffusion and Ecological Problems: Modern Perspectives* (eds. Okubo, A. & Levin, S.A.). Springer.

[259] Open Science Collaboration (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7, 657–660.

[260] Ory, P., Hartmann, H.J., Jude, F., Dupuy, C., Del Amo, Y., Catala, P., Mornet, F., Huet, V., Jan, B., Vincent, D., Sautour, B. & Montanié, H. (2010). Pelagic food web patterns: do they modulate virus and nanoflagellate effects on picoplankton during the phytoplankton spring bloom? *Environmental Microbiology*, 12, 2755–2772.

[261] Pacala, S.W. (1986). Neighborhood models of plant population dynamics. 4. Single-species and multispecies models of annuals with dormant seeds. *The American Naturalist*, 128, 859–878.

[262] Pacala, S.W. (1996). Dynamics of plant communities. In: *Plant Ecology* (ed. Crawley, M.). John Wiley & Sons, Ltd, chap. 15, pp. 532–555.

[263] Pahlow, M., Riebesell, U. & Wolf-Gladrow, D.A. (1997). Impact of cell shape and chain formation on nutrient acquisition by marine diatoms. *Limnology and Oceanography*, 42, 1660–1672.

[264] Pake, C.E. & Venable, D.L. (1996). Seed banks in desert annuals: implications for persistence and coexistence in variable environments. *Ecology*, 77, 1427–1435.

[265] Paparella, F. & Vichi, M. (2020). Stirring, mixing, growing: microscale processes change larger scale phytoplankton dynamics. *Frontiers in Marine Science*, 7, 654.

[266] Parker, T.H., Forstmeier, W., Koricheva, J., Fidler, F., Hadfield, J.D., Chee, Y.E., Kelly, C.D., Gurevitch, J. & Nakagawa, S. (2016). Transparency in ecology and evolution: real problems, real solutions. *Trends in Ecology & Evolution*, 31, 711–719.

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

[268] Patrick, R. (1948). Factors effecting the distribution of diatoms. *Botanical Review*, 14, 473–524.

[269] Peters, F. & Marrasé, C. (2000). Effects of turbulence on plankton: an overview of experimental evidence and some theoretical considerations. *Marine Ecology Progress Series*, 205, 291–306.

[270] Picoche, C. & Barraquand, F. (2019). How self-regulation, the storage effect, and their interaction contribute to coexistence in stochastic and seasonal environments. *Theoretical Ecology*, 12, 489–500.

[271] Picoche, C. & Barraquand, F. (2020). Strong self-regulation and widespread facilitative interactions between genera of phytoplankton. *Journal of Ecology*, 108, 2232–2242.

[272] Picoche, C., Young, W.R. & Barraquand, F. (2022). [Re] Reproductive pair correlations and the clustering of organisms. *ReScience C*, 8.

[273] Pierella Karlusich, J.J., Ibarbalz, F.M. & Bowler, C. (2020). Phytoplankton in the *Tara* Ocean. *Annual Review of Marine Science*, 12, 233–265.

[274] Pierrehumbert, R.T. (1994). Tracer microstructure in the large-eddy dominated regime. *Chaos, Solitons & Fractals*, 4, 1091–1110.

[275] Plank, M.J. & Law, R. (2015). Spatial point processes and moment dynamics in the life sciences: a parsimonious derivation and some extensions. *Bulletin of Mathematical Biology*, 77, 586–613.

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

[277] Popper, K. (2002 - Original work published in 1934). *The logic of scientific discovery*. Routledge.

[278] Prairie, J.C., Sutherland, K.R., Nickols, K.J. & Kaltenberg, A.M. (2012). Biophysical interactions in the plankton: a cross-scale review. *Limnology and Oceanography: Fluids and Environments*, 2, 121–145.

[279] Qian, J.J. & Akçay, E. (2020). The balance of interaction types determines the assembly and stability of ecological communities. *Nature Ecology & Evolution*, 4, 356–365.

[280] R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[281] Ramond, P., Siano, R. & Sourisseau, M. (2018). Functional traits of marine protists. *SEANOE*.

[282] Record, N.R., Pershing, A.J. & Maps, F. (2014). The paradox of the "paradox of the plankton". *ICES Journal of Marine Science*, 71, 236–240.

[283] Rees, M., Condit, R., Crawley, M., Pacala, S. & Tilman, D. (2001). Long-term studies of vegetation dynamics. *Science*, 293, 650–655.

[284] REPHY (2017). *REPHY dataset - French Observation and Monitoring program for Phytoplankton and Hydrology in coastal waters. 1987-2016 Metropolitan data.* SEANOE. https://www.seanoe.org/data/00361/47248/, doi:10.17882/47248 (2017 version).

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

[286] Richardson, L.F. (2007). *Weather prediction by numerical process.* Cambridge university press.

[287] Roeger, L.I.W. & Allen, L.J.S. (2004). Discrete May-Leonard competition models I. *Journal of Difference Equations and Applications*, 10, 77–98.

[288] Rosindell, J., Hubbell, S.P. & Etienne, R.S. (2011). The unified neutral theory of biodiversity and biogeography at age ten. *Trends in Ecology & Evolution*, 26, 340–348.

[289] Rothschild, B. & Osborn, T. (1988). Small-scale turbulence and plankton contact rates. *Journal of Plankton Research*, 10, 465–474.

[290] Rougier, N.P., Hinsen, K., Alexandre, F., Arildsen, T., Barba, L.A., Benureau, F.C.Y., Brown, C.T., Buyl, P.d., Caglayan, O., Davison, A.P., Delsuc, M.A., Detorakis, G., Diem, A.K., Drix, D., Enel, P., Girard, B., Guest, O., Hall, M.G., Henriques, R.N., Hinaut, X., Jaron, K.S., Khamassi, M., Klein, A., Manninen, T., Marchesi, P., McGlinn, D., Metzner, C., Petchey, O., Plesser, H.E., Poisot, T., Ram, K., Ram, Y., Roesch, E., Rossant, C., Rostami, V., Shifman, A., Stachelek, J., Stimberg, M., Stollmeier, F., Vaggi, F., Viejo, G., Vitay, J., Vostinar, A.E., Yurchak, R. & Zito, T. (2017). Sustainable computational science: the ReScience initiative. *PeerJ Computer Science*, 3, e142.

[291] Rudnicki, R. & Wieczorek, R. (2006). Phytoplankton dynamics: from the behavior of cells to a transport equation. *Mathematical Modelling of Natural Phenomena*, 1, 81–97.

[292] de Ruiter, P.C. & Gaedke, U. (2017). Emergent facilitation promotes biological diversity in pelagic food webs. *Food Webs*, 10, 15–21.

[293] Ryderheim, F., Hansen, P.J. & Kiørboe, T. (2022). Predator field and colony morphology determine the defensive benefit of colony formation in marine phytoplankton. *Frontiers in Marine Science*, 9, 829419.

[294] Sanyal, A., Larsson, J., van Wirdum, F., Andrén, T., Moros, M., Lönn, M. & Andrén, E. (2022). Not dead yet: diatom resting spores can survive in nature for several millennia. *American Journal of Botany*, 109, 1–16.

[295] 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.

[296] Sathyendranath, S., Lazzara, L. & Prieur, L. (1987). Variations in the spectral values of specific absorption of phytoplankton. *Limnology and Oceanography*, 32, 403–415.

[297] Scheef, L.P., Hampton, S.E. & Izmest'Eva, L.R. (2013). Inferring plankton community structure from marine and freshwater long-term data using multivariate autoregressive models. *Limnology and Oceanography: Methods*, 11, 475–484.

[298] Scheffer, M. & van Nes, E.H. (2006). Self-organized similarity, the evolutionary emergence of groups of similar species. *Proceedings of the National Academy of Sciences*, 103, 6230–6235.

[299] Schippers, P., Verschoor, A.M., Vos, M. & Mooij, W.M. (2001). Does "supersaturated coexistence" resolve the "paradox of the plankton"? *Ecology Letters*, 4, 404–407.

[300] Schreiber, S.J. (2021). Positively and negatively autocorrelated environmental fluctuations have opposing effects on species coexistence. *The American Naturalist*, 197, 405–414.

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

[302] Segura, A.M., Kruk, C., Calliari, D., García-Rodriguez, F., Conde, D., Widdicombe, C.E. & Fort, H. (2013). Competition drives clumpy species coexistence in estuarine phytoplankton. *Scientific Reports*, 3, 1037.

[303] Serra-Garcia, M. & Gneezy, U. (2021). Nonreplicable publications are cited more than replicable ones. *Science Advances*, 7, eabd1705.

[304] Seymour, J.R., Amin, S.A., Raina, J.B. & Stocker, R. (2017). Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nature Microbiology*, 2, 17065.

[305] Shao, Q., Lin, Z., Zhou, C. & Yan, X. (2020). Bacterioplankton assembly and interspecies interactions follow trajectories of *Gymnodinium* - diatom bloom. *Marine Environmental Research*, 160, 105010.

[306] Shmida, A. & Ellner, S. (1984). Coexistence of plant species with similar niches. *Vegetatio*, 58, 29–55.

[307] Shoemaker, L.G. & Melbourne, B.A. (2016). Linking metacommunity paradigms to spatial coexistence mechanisms. *Ecology*, 97, 2436–2446.

[308] Skellam, J.G. (1951). Random dispersal in theoretical populations. *Biometrika*, 38, 196–218.

[309] Smayda, T.J. & Reynolds, C.S. (2001). Community assembly in marine phytoplankton: application of recent models to harmful dinoflagellate blooms. *Journal of Plankton Research*, 23, 447–461.

[310] Solow, A.R., Beet, A.R., Keafer, B.A. & Anderson, D.M. (2014). Testing for simple structure in a spatial time series with an application to the distribution of *Alexandrium* resting cysts in the Gulf of Maine. *Marine Ecology Progress Series*, 501, 291–296.

[311] Sonnet, V., Guidi, L., Mouw, C.B., Puggioni, G. & Ayata, S.D. (2022). Length, width, shape regularity, and chain structure: time series analysis of phytoplankton morphology from imagery. *Limnology and Oceanography*.

[312] Spaak, J.W., Adler, P. & Ellner, S. (2022). Modeling phytoplankton-zooplankton interactions: opportunities for species richness and challenges for modern coexistence theory. *bioRxiv*. doi:10.1101/2022.03.24.485680.

[313] Stock, C.A., McGillicuddy, D.J., Solow, A.R. & Anderson, D.M. (2005). Evaluating hypotheses for the initiation and development of *Alexandrium fundyense* blooms in the western Gulf of Maine using a coupled physical-biological model. *Deep Sea Research Part II: Topical Studies in Oceanography*, 52, 2715–2744.

[314] Stocker, R. (2012). Marine microbes see a sea of gradients. *Science*, 338, 628–633.

[315] Svensson, F., Norberg, J. & Snoeijs, P. (2014). Diatom cell size, coloniality and motility: trade-offs between temperature, salinity and nutrient supply with climate change. *PLoS ONE*, 9, e109993.

[316] Takabayashi, M., Lew, K., Johnson, A., Marchi, A., Dugdale, R. & Wilkerson, F.P. (2006). The effect of nutrient availability and temperature on chain length of the diatom, *Skeletonema costatum*. *Journal of Plankton Research*, 28, 831–840.

[317] Tang, Y.Z., Koch, F. & Gobler, C.J. (2010). Most harmful algal bloom species are vitamin B_1 and B_{12} auxotrophs. *Proceedings of the National Academy of Sciences*, 107, 20756–20761.

[318] Tennekes, H. & Lumley, J.L. (1972). *A first course in turbulence*. MIT press.

[319] Tester, P.A. & Steidinger, K.A. (1997). *Gymnodinium breve* red tide blooms: initiation, transport, and consequences of surface circulation. *Limnology and Oceanography*, 42, 1039–1051.

[320] Tilman, D., Kilham, S. & Kilham, P. (1982). Phytoplankton community ecology: the role of limiting nutrients. *Annual Review of Ecology and Systematics*, 13, 349–372.

[321] Titman, D. (1976). Ecological competition between algae: experimental confirmation of resource-based competition theory. *Science*, 192, 463–465.

[322] Tortajada, S., Niquil, N., Blanchet, H., Grami, B., Montanié, H., David, V., Glé, C., Saint-Béat, B., Johnson, G.A., Marquis, E., Del Amo, Y., Dubois, S., Vincent, D., Dupuy, C., Jude, F., Hartmann, H.J. & Sautour, B. (2012). Network analysis of the planktonic food web during the spring bloom in a semi enclosed lagoon (Arcachon, SW France). *Acta Oecologica*, 40, 40–50.

[323] Tsukazaki, C., Ishii, K.I., Matsuno, K., Yamaguchi, A. & Imai, I. (2018). Distribution of viable resting stage cells of diatoms in sediments and water columns of the Chukchi Sea, Arctic Ocean. *Phycologia*, 57, 440–452.

[324] Tuck, S.L., Porter, J., Rees, M. & Turnbull, L. (2018). Strong responses from weakly interacting species. *Ecology Letters*, 21, 1845–1852.

[325] Utermöhl, H. (1958). Zur Vervollkommnung der quantitativen Phytoplankton-Methodik. *Internationale Vereinigung für theoretische und angewandte Limnologie: Mitteilungen*, 9, 1–38.

[326] Vallina, S.M., Ward, B.A., Dutkiewicz, S. & Follows, M.J. (2014). Maximal feeding with active prey-switching: a kill-the-winner functional response and its effect on global diversity and biogeography. *Progress in Oceanography*, 120, 93–109.

[327] Vik, J.O., Brinch, C.N., Boutin, S. & Stenseth, N.C. (2008). Interlinking hare and lynx dynamics using a century's worth of annual data. *Population Ecology*, 50, 267–274.

[328] Violle, C., Nemergut, D.R., Pu, Z. & Jiang, L. (2011). Phylogenetic limiting similarity and competitive exclusion. *Ecology Letters*, 14, 782–787.

[329] Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2009). Inferring species interactions in tropical forests. *Proceedings of the National Academy of Sciences*, 106, 13854–13859.

[330] Volterra, V. (1928). Variations and fluctuations of the number of individuals in animal species living together. *ICES Journal of Marine Science*, 3, 3–51.

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

[332] Wiedmann, I., Reigstad, M., Marquardt, M., Vader, A. & Gabrielsen, T.M. (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.

[333] Wiegand, T., Gunatilleke, C.V.S., Gunatilleke, I.A.U.N. & Huth, A. (2007). How individual species structure diversity in tropical forests. *Proceedings of the National Academy of Sciences*, 104, 19029–19033.

[334] Wiegand, T., Wang, X., Anderson-Teixeira, K.J., Bourg, N.A., Cao, M., Ci, X., Davies, S.J., Hao, Z., Howe, R.W., Kress, W.J., Lian, J., Li, J., Lin, L., Lin, Y., Ma, K., McShea, W., Mi, X., Su, S.H., Sun, I.F., Wolf, A., Ye, W. & Huth, A. (2021). Consequences of spatial patterns for coexistence in species-rich plant communities. *Nature Ecology & Evolution*, 5, 965–973.

[335] Wilhelm, S.W. & Suttle, C.A. (1999). Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. *BioScience*, 49, 781–788.

[336] Wisnoski, N.I., Leibold, M.A. & Lennon, J.T. (2019). Dormancy in metacommunities. *The American Naturalist*, 194, 135–151.

[337] Wisnoski, N.I. & Shoemaker, L.G. (2022). Seed banks alter metacommunity diversity: the interactive effects of competition, dispersal and dormancy. *Ecology Letters*, 25, 740–753.

[338] Woods, J.D. (2005). The Lagrangian Ensemble metamodel for simulating plankton ecosystems. *Progress in Oceanography*, 67, 84–159.

[339] Woodward, J.R., Pitchford, J.W. & Bees, M.A. (2019). Physical flow effects can dictate plankton population dynamics. *Journal of the Royal Society Interface*, 16.

[340] Yñiguez, A.T., Cayetano, A., Villanoy, C.L., Alabia, I., Fernandez, I., Palermo, J.D., Benico, G.A., Siringan, F.P. & Azanza, R.V. (2012). Investigating the roles of intrinsic and extrinsic factors in the blooms of *Pyrodinium bahamense* var. *compressum* using an individual-based model. *Procedia Environmental Sciences*, 13, 1462–1476.

[341] Yamamoto, T., Seike, T., Hashimoto, T. & Tarutani, K. (2002). Modelling the population dynamics of the toxic dinoflagellate *Alexandrium tamarense* in Hiroshima Bay, Japan. *Journal of Plankton Research*, 24, 33–47.

[342] Yamamura, K., Yokozawa, M., Nishimori, M., Ueda, Y. & Yokosuka, T. (2006). How to analyze long-term insect population dynamics under climate change: 50-year data of three insect pests in paddy fields. *Population Ecology*, 48, 31–48.

[343] Yenni, G., Adler, P.B. & Ernest, S.K.M. (2017). Do persistent rare species experience stronger negative frequency dependence than common species? Asymmetric NFD and persistent rare species. *Global Ecology and Biogeography*, 26, 513–523.

[344] Young, A.M., Karp-Boss, L., Jumars, P.A. & Landis, E.N. (2012). Quantifying diatom aspirations: mechanical properties of chain-forming species. *Limnology and Oceanography*, 57, 1789.

[345] Young, W.R., Roberts, A.J. & Stuhne, G. (2001). Reproductive pair correlations and the clustering of organisms. *Nature*, 412, 328–331.