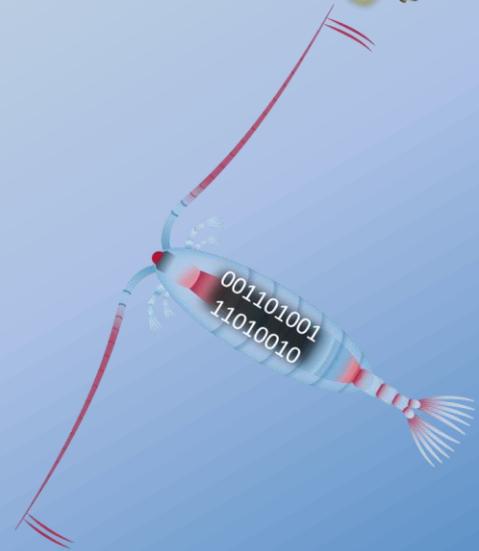


# Simulating Plankton - getting it right in the era of Digital Twins of The Ocean

## *Simplicity vs complexity*



**Flynn, Sherr, Atkinson, Lønborg,  
Mulholland, Rokitta, Rost, Schmidt,  
Stoecker, Thiele & Thingstad**



# Simulating Plankton - getting it right in the era of Digital Twins of The Ocean

## *Simplicity vs complexity*

Flynn KJ<sup>1\*</sup>, Sherr E<sup>2</sup>, Atkinson A<sup>1</sup>, Lønborg C<sup>3</sup>, Mulholland M<sup>4</sup>, Rokitta S<sup>5</sup>, Rost B<sup>5</sup>, Schmidt K<sup>6</sup>, Stoecker DK<sup>7</sup>, Thiele S<sup>8</sup>, Thingstad TF<sup>9</sup>

- 1) Plymouth Marine Laboratory, Prospect Place, Plymouth PL1 3DH, UK
- 2) Eugene, Oregon, USA
- 3) Department of Ecoscience, Aarhus University, Roskilde, Denmark
- 4) Old Dominion University, Norfolk 23529 USA
- 5) Alfred-Wegener-Institute - Helmholtz Centre for Polar and Marine Research, 27570 Bremerhaven, Germany
- 6) Earth & Environmental Sciences, University of Plymouth, PL4 8AA, UK
- 7) Horn Point Laboratory, University of Maryland Center for Environmental Science, Cambridge, Maryland, USA
- 8) Microbial Ecology, University of Vienna, Djerassiplatz 1, A-1030 Vienna, Austria
- 9) Biological Sciences, University of Bergen, Bergen 5020, Norway

\* [KJF@PML.ac.uk](mailto:KJF@PML.ac.uk)

**Citing this work**

Flynn, K. J., Sherr, E., Atkinson, A., Lønborg, C., Mulholland, M., Rokitta, S., Rost, B., Schmidt, K., Stoecker, D. K., Thiele, S. & Thingstad, T. F. (2024) *Simulating Plankton - getting it right in the era of Digital Twins of The Ocean; simplicity vs complexity*.

Zenodo <https://doi.org/10.5281/zenodo.10953197>

Please note that this work may be subject to changes during its year of publication. The doi noted in the citation details will always link to the most up-to-date version. Please check for updates.

**Acknowledgment to Funder**

This work was funded by UKRI Natural Environment Research Council (NE/X010783/1; *Simulating Plankton - getting it right in the era of Digital Twins of The Ocean*), under the direction of the lead author of this document.

**Author Contributions**

Flynn designed the overall project and this specific task reported here. Others, notably Sherr, helped to improve the task description and the template. All, except Flynn, undertook the task described in Methods, working individually except where indicated, but typically with some level of support from Flynn. Flynn undertook the initial analysis of the Results and drafted the Discussion. All contributed to finalising the report.

## Executive Summary

This work describes the outcomes from a project funded by the NERC (UK) during 2023, with the overarching aim of facilitating the construction of the next generation of plankton simulation models by engaging with experts in real plankton physiology and ecology. Over 30 experts, covering plankton from viruses to krill, contributed to various facets of the project. They were selected specifically for their empirical interests; modellers per se were not included.

Ten experts in empirical plankton science contributed to configuring aspirational plankton digital twins (PDT) of different levels of complexity, with emphasis on the required number of state variables. Those state variables could be allocated to any combination of abiotic and biotic components (invariably with units of mass  $m^{-3}$ ). The contributions were divided between those centred on the microbial loop, phototrophy, or on consumers. Most included multiple plankton functional groups, though two of the nine projects were essentially autecological models (single bacteria type using a complex description of DOM, or centred on phytoplankton phototrophy).

Key conclusions from this work were:

- i) In general, complexity was expected of the models, with simple models (typifying traditional plankton models) not considered to reflect reality sufficiently to be of use.
- ii) As a trade-off for the allocation of a given number of state variables, a preference was expressed in various models for a greater importance on describing plankton diversity rather than physiological diversity as reflected by variable stoichiometry for each of those plankton types included. While stoichiometric ecology is undoubtedly a powerful driver of trophic dynamics, that is only so if the organisms are there to be described in the first instance. To do full justice, of course, both diversity and variable stoichiometry are required.
- iii) Pragmatically, one could argue that there is no logic in having a complex model if it is not possible to parameterise it; simple models have much to commend them. However, both here and elsewhere in the project, a counter view is expressed that one can learn much from the ultimate failure of a well-founded complex model; complexity in models of plankton should not be avoided simply for want of empirical numeric data.
- iv) As much interest was expressed in the non-plankton components of the system as in the organism descriptions themselves.
- v) From (iii), and from other contributions to this project, early iterations of aspiring digital twin models of plankton would likely be useful containing rather few plankton types (ca. <10) operating in a well-defined multi-stressor scenario.

The exercise was useful not only in revealing different interests but also for the communication experience between the contributors and the task lead. The latter indicates that only through actually attempting to build and deploy plankton digital twins can progress be best made. This is because of the intense modeller-user interaction that is required to produce a useful product.

## Contents

Executive Summary .....	3
1. Introduction .....	5
1.1 Simplicity versus plausibility .....	5
1.2 Aim .....	7
2. Method .....	7
3. Results .....	8
3.1 Microbial loop target .....	8
3.2 Phototroph target .....	15
3.3 Consumer target .....	22
4. Discussion .....	29
4.1 Trade-offs in simplicity vs complexity .....	29
4.2 Pragmatism .....	29
4.3 Conclusions for building plankton digital twins .....	30
References .....	31
Appendix 1 .....	34
A1 Introducing the task .....	34
A1.1 Allocating state variables .....	34
A1.2 Constant stoichiometry? .....	35
A1.3 Allometry and functional grouping .....	35
A1.4 Ontogeny and life cycle development .....	36
A1.5 DOM .....	37
A2 Task instructions .....	37
A2.1 Working alone or in a small group .....	38
A2.2 Select the physical scenario .....	38
A2.3 Build the starting configuration .....	38
A2.4 Sequentially add 3 state variables .....	38
A2.5 And repeat! .....	39
A3 Acceptability .....	39
A4 When you are done .....	41
‘Frequently Asked Questions’ (FAQs) .....	41

## 1. Introduction

The more processing a computer has to undertake in order to achieve a given task, the greater is the computational load. Computational loads provide an important constraint on simulations; if an increased load is not to extend the period of time taken to complete a given simulation task, then more processing power is required. While simple models may complete their task within seconds with the output data being easily held in a small database, models of large complex 3D ecosystems may take many days to execute and require very large data storage facilities. Depending on the application, a digital twin is likely to be required to give a near-instantaneous response; certainly model outputs would be expected within minutes.

Complexity in a plankton model comprises combinations of detail at the organismal level, the number of organism types, and of the associated chemico-physical environment. The time taken to undertake a simulation of what could be considered as an ideal level of complexity by the user may be excessive. On the other hand, there is no logic in running a model that is so simple that it provides a dubious, or even flawed, description. A compromise needs to be struck.

Detail is held primarily within the state variables that describe abundance and concentrations of resources and plankton. The more state variables, the more detail, but also the higher the computational costs; in crude terms, a doubling in the state variable count will double the computational load. Contributors to this task were asked to consider, for a plankton scenario of their choice, the types of detail that would satisfy their own expectations and also of others who may use that model. Such a process starts with considering which organisms are to be included, and then for each organism which structural traits. Traits include resource needs and biotic factors affecting growth, features of the plankton description (such as stoichiometry, ontogenic age-at-stage, etc.), and the release of products that form the resources for other organisms. These needs must then be ranked (prioritised) to provide the most compact selection of state variables that would provide a description likely to be considered as acceptable by other plankton researchers.

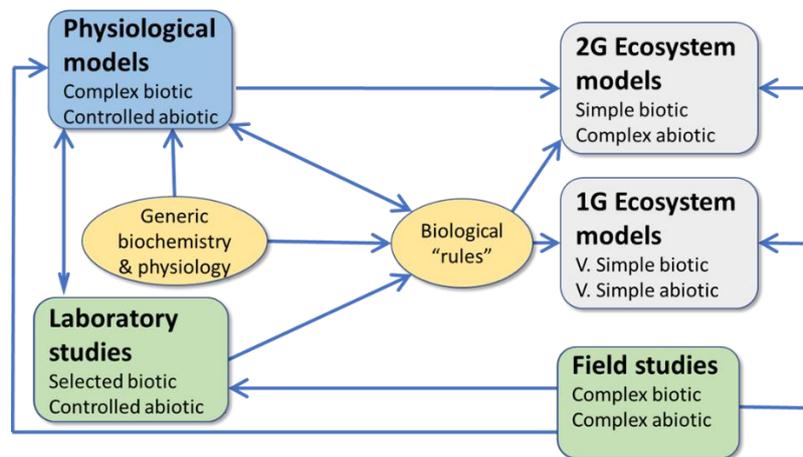
### 1.1 Simplicity versus plausibility

Simulation models are built around state variables. **A state variable is a parameter with history, such as concentration or abundance.** The numeric value of each state variable usually changes over time as the simulation progresses and, in consequence, the computational costs for operating the model are primarily associated with the state variable count. Of course, it is not just the state variable count that defines complexity; those state variables have to interact with each other in the right way. However, if you do not have the state variable, then you cannot have the interactions.

Traditionally, plankton simulation models have been very simple, not only with respect to the description of individual organisms, but also with respect to the trophic interactions. The classic nutrient-phytoplankton-zooplankton NPZ model described the N-source as either combined ammonium+nitrate, or as separate entities, and then all phototrophs were within one 'box', and all zooplankton in another 'box'. The organisms were based upon N as the currency, and the whole model contained just 3 or 4 state variables. The NPZB version added another two state variables, again in terms of N, as bacteria and DON (Fasham et al. 1990).

No one will argue that such simple NPZ style models could provide a digital twin experience. That is not to say these simple models are not of value; cutting out extraneous information is an important scientific tool. However, the usefulness of such models also requires that the reader fully appreciates the caveats, and indeed believes that the 'extraneous' information that has been omitted is indeed extraneous. As a research tool, simple models are also built to deliver to a specific task such as testing a stated hypothesis. A digital twin is a very different type of simulation model. In a digital twin, the end user may consider situations, or hypotheses, beyond those envisaged by those making the model. Building a digital twin thus requires a more complete consideration of reality than does the selective reality of the research tool.

The first generation plankton models (**Fig. 1**) made use of a few biological rules and concepts. Most obviously they exploited resource-rate interactions (such as nutrient uptake and predation) as simple saturating events, typically akin to Michaelis-Menten or Monod kinetics. Current, second generation, plankton models contain greatly enhanced descriptions of the non-biological aspects of the system, while also raising the complexity of the plankton descriptions. However, most of that increase in biological complexity has been achieved by just increasing the number of functional types that are represented; the underlying description of each organism remains for the most part extremely simple. Indeed, it took a long time before variable stoichiometry became established as a ‘must-have’. Before that, it was argued that models could just assume fixed Redfield stoichiometry (Flynn 2010), despite clear evidence to the contrary from empirical studies (now firmly established through *ecological stoichiometry* – Sterner & Elser 2002).



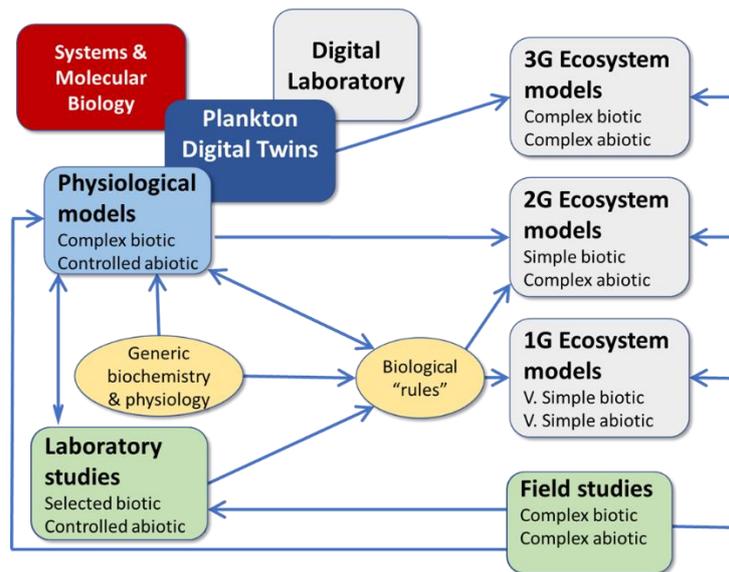
**Fig. 1** The early development of plankton models took place primarily over the 1960's-90's.

Extant plankton models, as deployed in ocean basin through to global scenarios, remain extremely simple. The arguments for this situation are usually couched in terms of the computational cost of adding complexity, and that the addition of more plankton types provides little or no advantage, and indeed often causes problems due to instabilities. Invariably, those ‘different’ plankton are described by the same core biological concept, just with different parameter values (e.g., maximum growth rates, half saturation constants etc.). From a biological point of view, such descriptions are clearly too simplistic to be of value, and there are none of the features of reality that would stabilise or otherwise affect trophic interactions.

For 20-30 years plankton modelling has remained essentially unchanged, and still exhibits fundamental problems. That is especially so with respect to the descriptions of consumers (Mitra et al. 2014), which includes not only protist and metazoan zooplankton, but mixoplankton and (arguably, as ‘consumers’) also viruses. Bacteria are also very poorly considered (typically ignored). While the scientific challenge for which most plankton models were configured was biogeochemistry, the emphasis now is shifting towards biodiversity (noting that biogeochemistry is an emergent property of biodiversity). It is becoming apparent that climate change is having profound effects on plankton community structure (Schmidt et al. 2020; Heneghan et al. 2023). For biodiversity applications, to support fisheries or HAB management, we need appropriate biodiversity-capable models. However, we still have to minimise the computational costs else the biological description cannot be run within the complexity of the chemico-physical model within a limited computational load. The bottom line is, however, that there is no sense in having a model that is too simple to describe reality, or at least something that is plausible.

## 1.2 Aim

This task sought to locate that sweet spot that balances simplicity with a model structure that gives a plausible description of reality. How we make the jump from traditional crude models to next (third) generation plankton models is where the concept of digital twins comes in (**Fig. 2**).



**Fig. 2** Developments from systems biology, coupled with different approaches to modelling and expert witness validation, will underpin the development of new plankton models to support 3<sup>rd</sup> generation marine modelling capabilities. From Flynn et al. (2022).

## 2. Method

The task was driven by the materials presented in **Appendix 1**.

Other than some stylistic changes to improve presentation, the example outputs (presented as copies of the A4 format Powerpoint template slides) are as provided by the contributors.

### 3. Results

The activities of each contributor are shown in **Figs. 3-11**, as indicated in **Table 1**.

**Table 1** Contributors and plankton types included in the model constructs deemed to be acceptable.

Fig.#	Contributor	Setting	Duration	Bacteria	Phytoplankton	Mixoplankton	Protist Z	Metazoan Z	HTL
3	Thingstad	Mesocosm	Weeks	•	•		•	•	
4	Thiele	Field	Month(s)	•	•	•	•		
5	Lønborg	Microcosm	Weeks	•					
6	Rokitta/Rost	Field	Months		•	•			
7	Mullholland	Field	Months	•	•	•	•	•	
8	Stoecker	Field	Month	•	•	•	•	•	
9	Sherr	Field	Weeks	•	•		•	•	•
10	Schmidt	Field	Year		•		•	•	•
11	Atkinson	Field	Months					•	•

The conceptual models envisaged by contributors show considerable contrast not only in the plankton descriptions but also in the need for different levels of complexity in the abiotic components, which would be required to support the operation of the model. None of the projected models had explicit requirements for 3D environmental scenarios; 0D or 1D (i.e. depth) was deemed to be sufficient.

The models have been grouped in terms of their principal focus, though (as is clear from **Table 1**), most included components from many plankton types.

It should be noted that, unless otherwise indicated, connectivity of plankton to recreate the trophic dynamics was not explicitly developed in these proposals. The emphasis was on how complex the organism and resource descriptions needed to be in terms of the main computational cost of state variable, not on how they would be connected per se (connections do not incur a cost comparable that of moving the values of state variables around in a fluid dynamics environment – see **Introduction**).

The contributions could be divided between those targeting microbial loop processes, phototrophy, or consumer processes.

#### 3.1 Microbial loop target

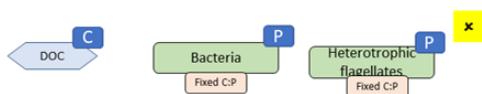
**Figs. 3-5** consider the microbial loop as their main targets. **Fig. 3** is produced by Thingstad, the only contributor with an established interest in using models. The description used is pragmatic, arguing for why a more expansive description could be counter-productive. Although a relatively simple construction of 9 state variables is considered as being quite usable, the more interesting construction is given at 16 state variables and higher, where the bacteria community is itself described as two functional groups. In this model (**Fig. 3**), copepods are noted as the only predators of large diatoms.

**Fig. 4**, by Thiele, makes an interesting contrast to **Fig. 3**. The target here is also the complexity of the prokaryote community, though now the whole description in terms of functional groups is far more expansive. The base model, deemed to be acceptable, includes 13 organism types; these range from bacteria through to metazoan zooplankton, with extensive and complex inter-linkage. This model does include heterotrophic dinoflagellates as prey for copepods and as part of microbial loop nutrient recycling, but not as major predators of phytoplankton

**Fig. 5**, by Lønborg, takes a very different stance, describing a microcosm (flask scenario) with only a single bacteria but a complex description of the substrates and environmental stressors. In contrast, **Figs. 3 & 4** describe only a single dissolved organic pool with multiple microbes. The final iterations of model development in **Fig. 5** also include pH, temperature and O<sub>2</sub>.

## Exploring microbial loop interactions

### 3 state variables

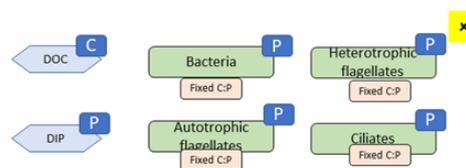


This is basically an NPZ model «in disguise». May illustrate the basic idea, but otherwise not very useful.

#### Note on fixed stoichiometry:

The «fixed C:P» for bacteria implies a fixed bacterial yield on DOC (bacterial biomass P formed per unit DOC consumed) and thus also a fixed respiration coefficient. The «fixed C:P» for heterotrophic flagellates implies a fixed yield of HF on bacteria (units of bacterial P incorporated in HF biomass)

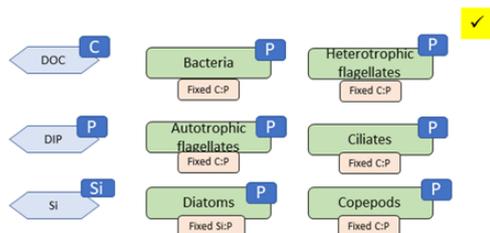
### 6 state variables



With 6 SVs you can reproduce the potential for both C-limitation and mineral nutrient limitation of bacteria and how this depends on ciliate abundance, but you cannot include the effect of nutrient competition from phytoplankton not grazed by ciliates (i.e. large diatoms)

The «fixed C:P» for AF implies a fixed (106:1) C-fixation per DIP molecule consumed.

### 9 state variables



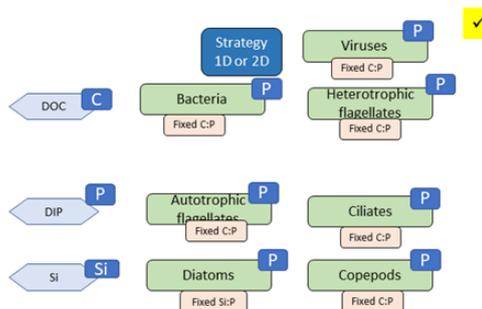
With 9 state variables one can represent the three main pathways for mineral nutrients through the microbial food web to meso-zooplankton. Our experience is that this (in many cases) can reproduce the main observed response patterns in mesocosms. In particular, one can pick up the cascading effects from variable copepod stocks and the effect of Si on glucose consumption (when large diatoms «steal» the mineral nutrients).

If we need to add special phenomena (like small diatoms eaten by ciliates) we try to add these into this structure in a manner not destroying the cascading in the minimum model.

The diatoms have a fixed C:P stoichiometry in biomass, but not in production since we couple diatom photosynthesis to diatom P-biomass, not to P-uptake. Most of the production is probably excreted as glucans so biomass C:P:Si may be relatively constant.

This is my favorite «minimum model» or «1st generation DT». Thingstad, T. F., et al. (2007). "Ability of a "minimum" microbial food web model to reproduce response patterns observed in mesocosms manipulated with N and P, glucose, and Si." *Journal of Marine Systems* 64(1-4): 15-34.

### 12 (11) state variables



With 12 SVs one can add bacterial viruses and thus the «viral loop» burning off a lot of the DOC before it reaches higher levels. This model requires a «strategy» variable defining the adaptive balance between competitive and defensive traits in the bacterial community. If both viral and predator defence is included the strategy variable becomes 2-dimensional.

Fig. 3 Contribution by Thingstad, to integrate the microbial loop in a plankton food web context. Intended to simulate a laboratory or mesocosm scenario over a few weeks.

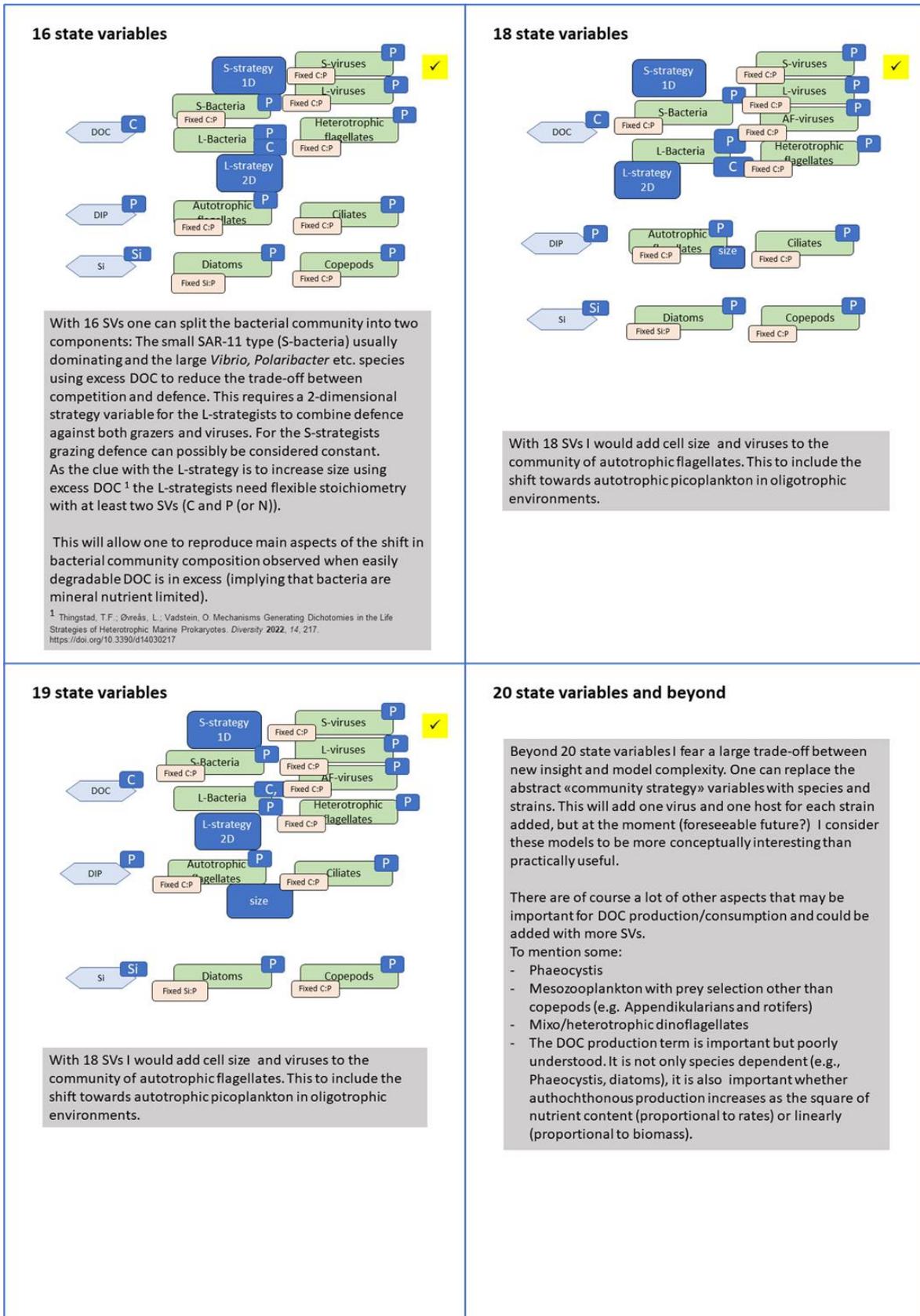


Fig. 3 continued.

## Exploring microbial-loop interactions



**Fig. 4** Contribution from Thiele describing a detailed microbial loop with protist predation on bacteria. Intended for a natural setting, operating over many weeks-months.

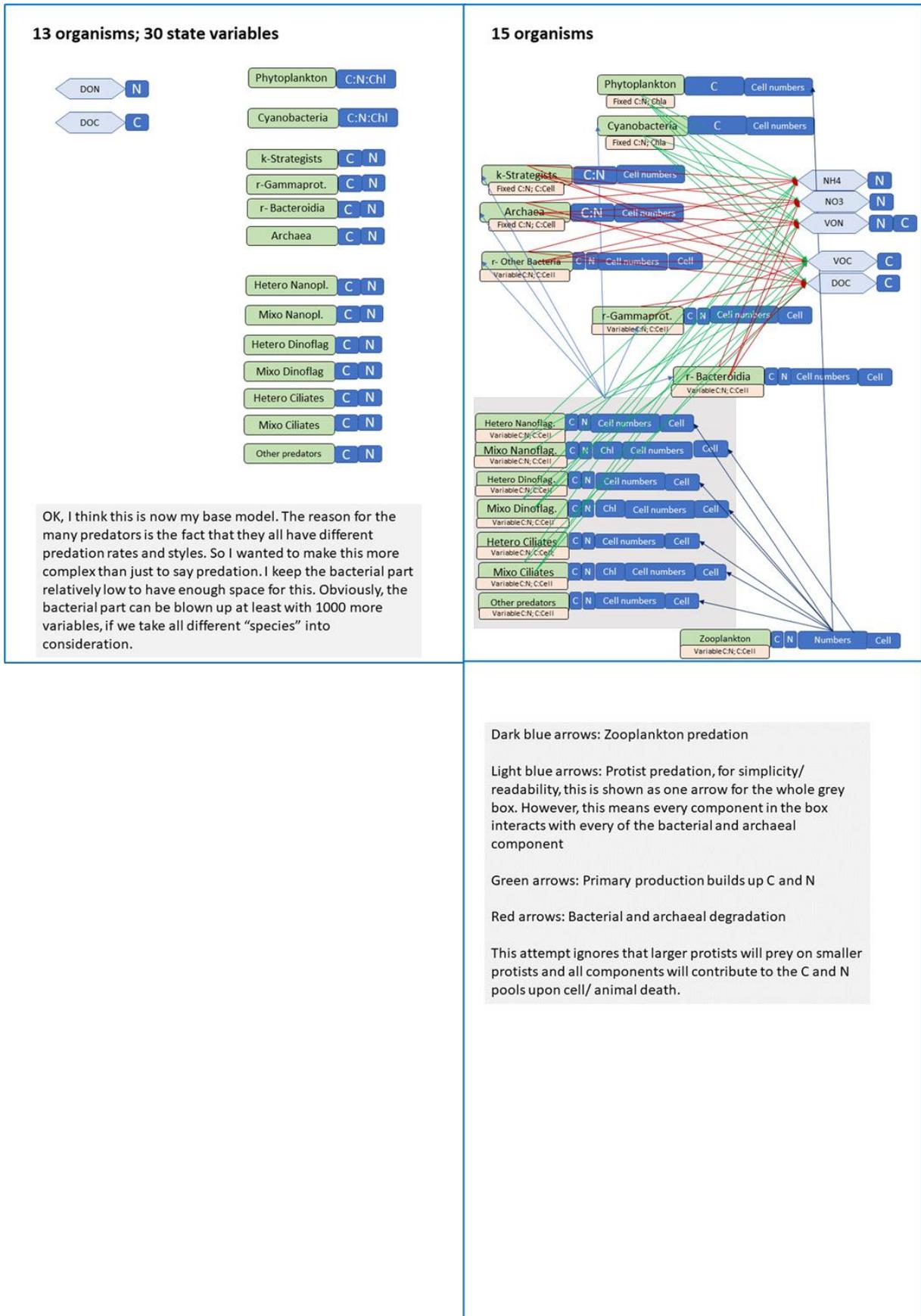
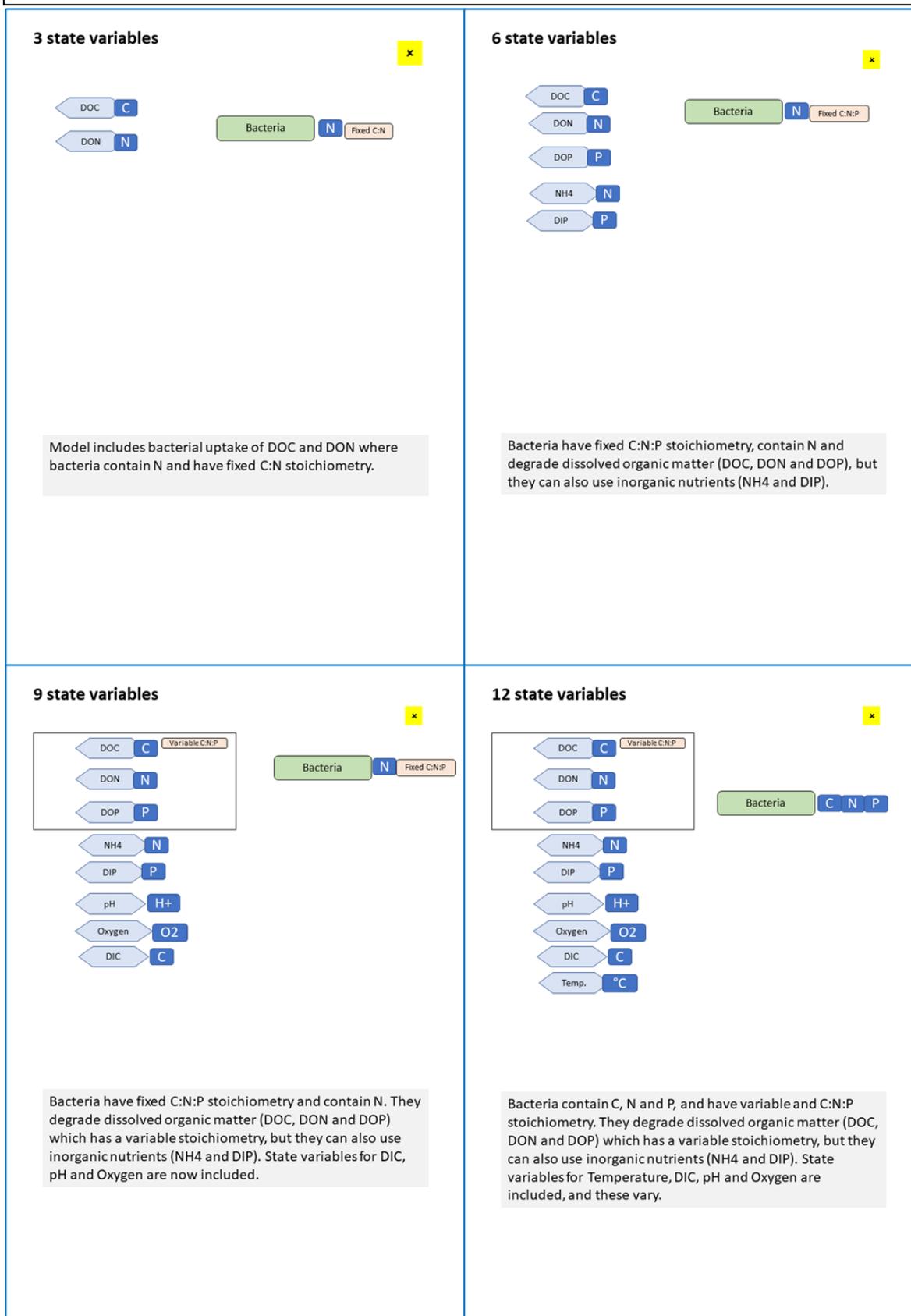


Fig. 4 continued.

## Exploring bacteria-DOM interactions



**Fig. 5** Contribution from Lønborg, describing climate-change factors (temperature, acidification,  $O_2$ , individually and combined) affecting bacterial degradation/uptake of DOM. The scenario is laboratory or microcosm with controlled populations, running over several weeks.

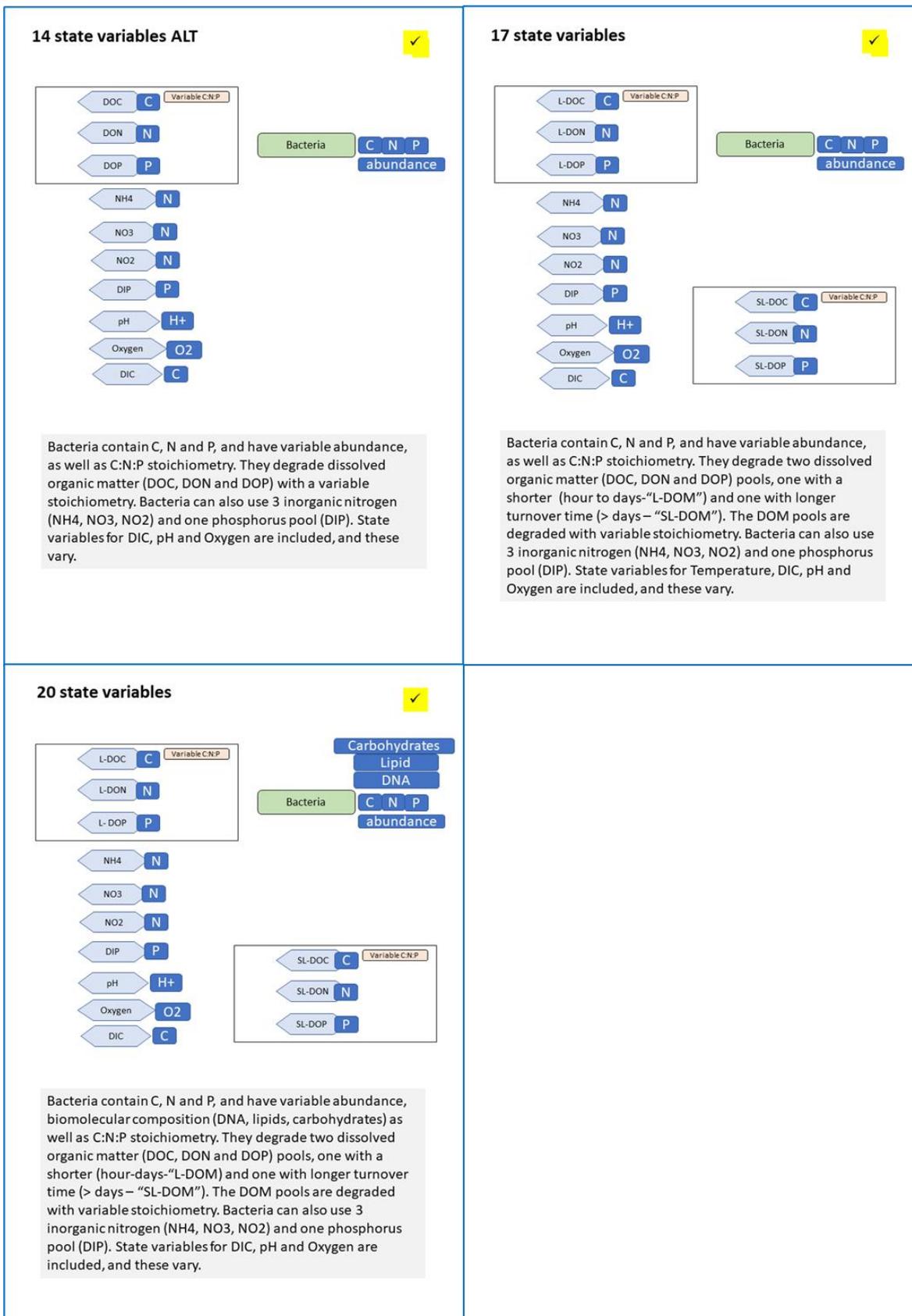


Fig. 5 continued.

### 3.2 Phototroph target

**Figs. 6-8** consider phototrophs as their main target, although for **Fig. 8**, that phototroph exploits acquired phototrophy in a mixoplankton.

**Fig. 6**, by Rokitta & Rost, considers a multi-functional type description of just the phytoplankton (accepting that some of those listed as phytoplankton are actually mixoplankton). The interest is in the detailed involvement of bottom-up factors (allometrics, multi-level biochemical stoichiometry, with nutrients, light, pH etc) affecting succession. While grazing is mentioned, the model is focused on phytoplankton diversity and nutrient sources. The application is for flasks or theoretical studies.

**Fig. 7**, by Mullholland, considers field successional interactions between diazotrophs, diatoms, non-diatoms and the harmful alga (mixoplankton), *Karenia*. Included here are roles for dissolved organics, Fe and bacteria. Like many of the models, the variable stoichiometric description of some plankton types is sacrificed to enable (for a given computational cost) a higher count of organism types.

**Fig. 8**, from Stoecker, also considers the trade-off in allocation of state variables to diversity versus complexity in variable stoichiometry. The model considers a field application of multiple plankton interactions; here that interaction is especially important as the core target of the model is generalist non-constitutive mixoplanktonic ciliates which acquire phototrophy from their prey. The contributor expressed a preference for having an enhanced biodiversity at the cost of variable stoichiometry. The argument is simple but profound; the nutritional value of a prey item is irrelevant if it is not present. Mixotrophy is obligate in most mixotrophic marine oligotrichs, (Stoecker et al. 1988; Stoecker et al 1988/1989; Maselli et al. 2020), although at least one species can grow slowly in the dark (McManus et al. 2012). Not all types of phytoplankton/mixoplankton are suitable sources of plastids and support growth. Haptophytes, chlorophytes, and cryptophytes support growth, but diatoms and dinoflagellates are generally poor foods for mixotrophic oligotrichs in culture (Stoecker et al. 1988; Stoecker et al 1988/1989; McManus et al. 2012; Maselli et al. 2020). Microscopic and molecular studies of field populations of mixotrophic oligotrichs have shown that haptophyte, stramenopile, chlorophyte, and cryptophyte plastids are retained, and that some oligotrich species appear to be generalists and others specialists in regard to plastid retention (Laval-Peuto & Febvre 1986; McManus et al. 2012; Johnson and Beaudoin 2019). These data indicate that the biodiversity of phytoplankton/mixoplankton available as prey has a major impact on the feeding and growth of mixotrophic oligotrich ciliates in the sea.

## Exploring spring-bloom interactions between phytoplankton

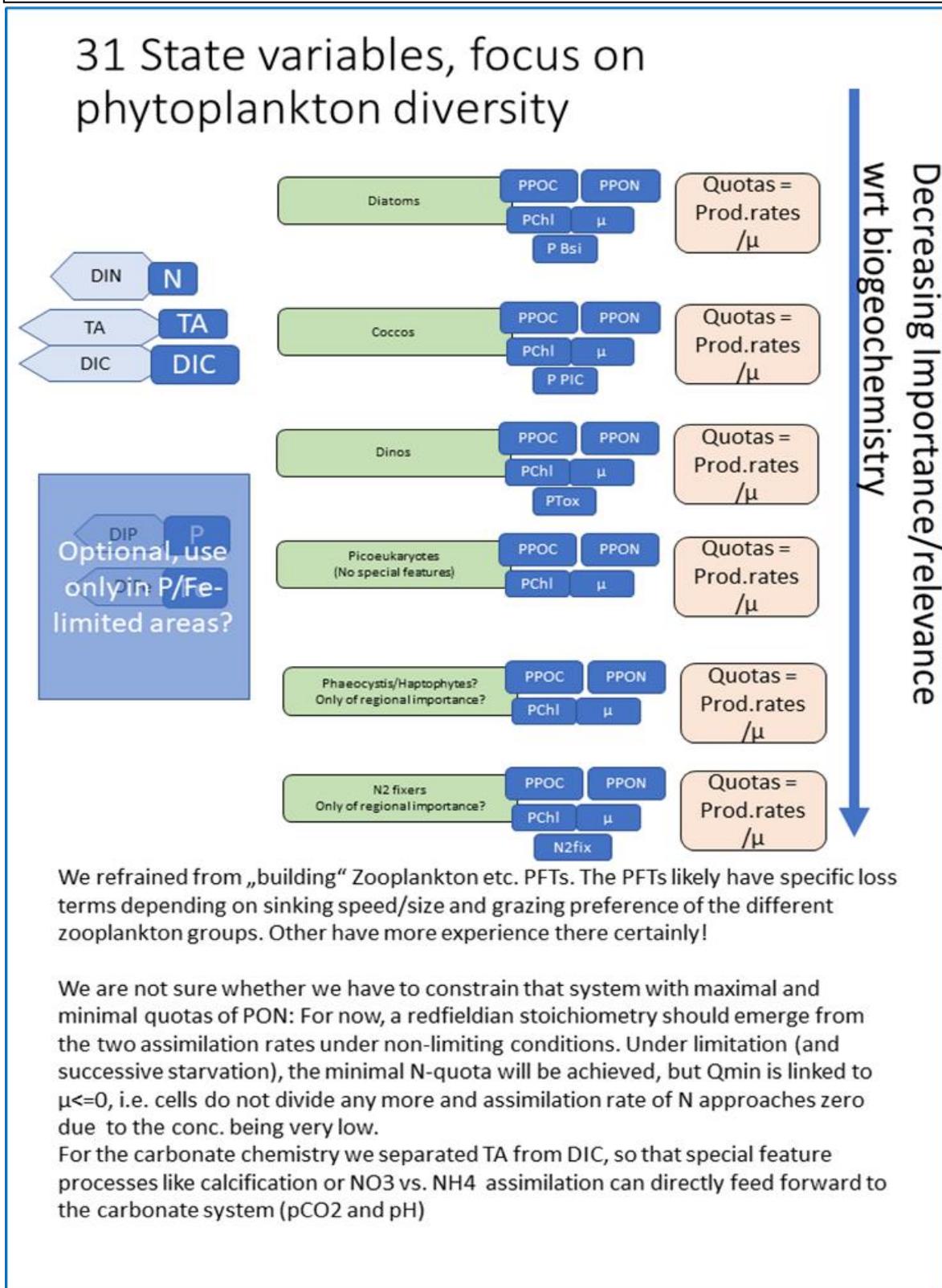
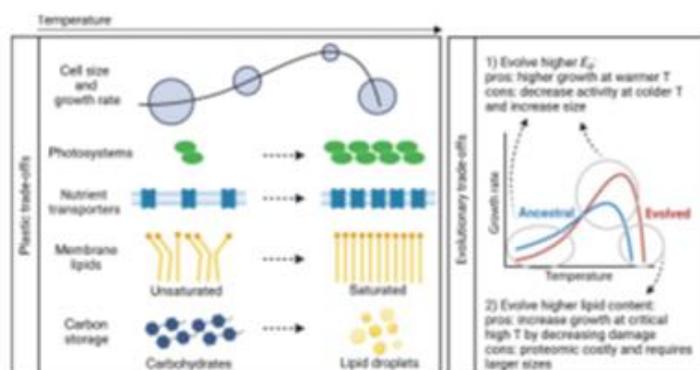


Fig. 6 Contribution by Rokitta & Rost describing interactions occurring during a natural spring bloom-like event, over several months.

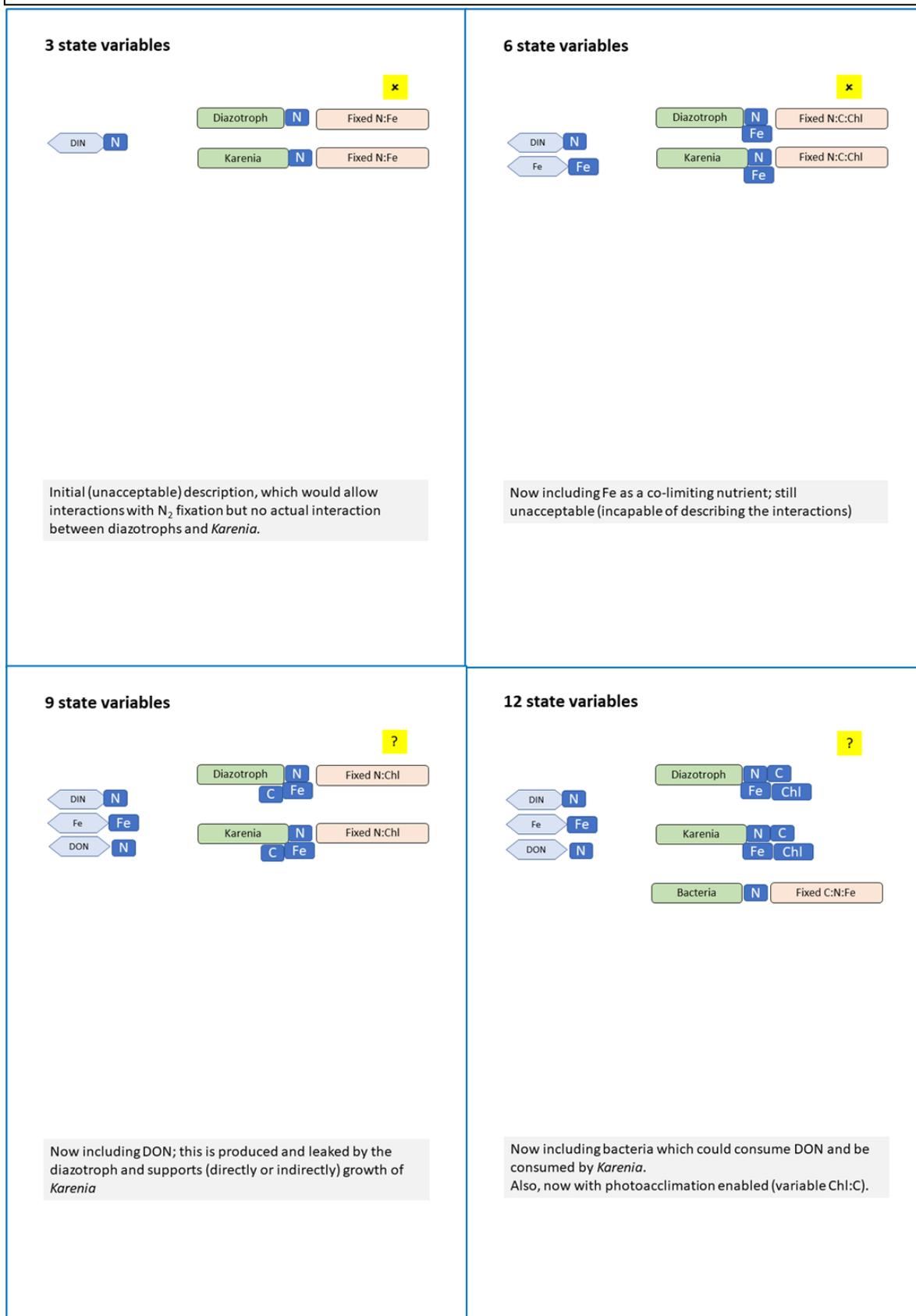
- One idea to simplify these models, is to neglect cell division as a whole and describe the PFTs as “boxes that get bigger”, This would cause a model to depict the growth of a certain biomass.
- POC<sub>Diatom</sub> should develop based on a POC production rate. POND<sub>Diatom</sub> will have its own production rate, etc.
- As a result, there should arise a POC:PON ratio of diatoms as a state variable. No cell division is accounted for, so that we do not need to bother about POC cell<sup>-1</sup> or Chl cell<sup>-1</sup> but rather stick to the core SVs and determine Chl POC<sup>-1</sup>.
- We’d love to see several PFTs in a diversified plankton model: Diatoms, Coccus (=Ehux?), Phaeocystis/Flagellates, Picoeukaryotes
- All should have POC, PON, Chl as core state variables, with addition of PIC (Coccus), biogenic Silica (BSi for Diatoms). Making these production rates independent warrants flexible stoichiometry.
- Allometry and size of cells:
  - We suggest to include the depiction of the “size of a cell” (measured as POC cell<sup>-1</sup> for example) by combining two processes: One is the (net) production of biomass modelled above, with a rate of  $\mu \times \text{pg POC} (\text{pg POC})^{-1} \times \text{d}^{-1}$ . Alternatively one could do it with N as a currency, correspondingly yielding  $\text{pg N} (\text{pg PON})^{-1} \times \text{d}^{-1}$ . The second, somewhat independent process is the cell division. It requires N and P to make nucleotides and synthesize the genome and some energy to separate daughter cells, but as long as N and P are present in sufficiently concentrations, i.e. over vast areas of the driver space, the cell division is to a large extent independent from POC production and both processes have their “own” temperature performance curves” in dependence of environmental drivers.
  - E.g., depicting the  $\mu$  over [Nutrient] likely yields a saturating function. In contrast,  $\mu$  over temperature gives a (often skewed) bell shaped function, with a  $T_{\min}$ ,  $T_{\max}$ ,  $T_{\text{opt}}$ ,  $\mu_{\max}$  and so on..
  - From our experiments we know that also net POC production (PPOC) follows a bell shape curve over a range of temperature correspondingly yielding PPOC<sub>min</sub>, PPOC<sub>max</sub>,  $T_{\text{opt}}$  and PPOC<sub>max</sub>.
  - We found that (according to Morel, J Phycol 1987) since  $\text{PPOC} = \text{POC} \times \mu$ , it follows that  $\text{PPOC}/\mu = \text{POC}$  quota.
  - The interesting thing for our project here is that by modulating these bell-shaped curves of  $\mu$  and PPOC, we obtain responses on POC quota that are emergent properties, explaining this graph I found recently in a paper by Leles & Levine (2023 <https://www.science.org/doi/pdf/10.1126/sciadv.adh8043>).



- The uppermost graph on cell size and growth rate shows the emergence of POC quotas, since towards the ends of the T-spectrum, cell division ceases, and it ceases faster than POC production. So while division ceases, there’s still a gain in POC, causing large cells towards the end.
- It could also be understood differently: If the POC production ceases more rapidly towards the ends of the T spectrum than cell division, then cells will keep dividing despite no or little gain in POC, causing smaller cells towards the ends of the T-spectrum.
- If the  $T_{\text{opt}}$  of the two processes is different, then the emerging pattern in POC quotas will approach “linear” upwards or downwards slopes.
- We are currently writing up a paper on this finding with our PhD student Linda Rehder, who has shown the above mentioned modes of this phenomenon in several Arctic species.

Fig. 6 continued.

## Exploring interactions promoting *Karenia* HAB formation



**Fig. 7** Contribution from Mulholland, describing interactions between diazotrophic cyanobacteria, diatoms and allied plankton interactions promoting *Karenia* bloom formation. The setting would be oceanographic, over months.

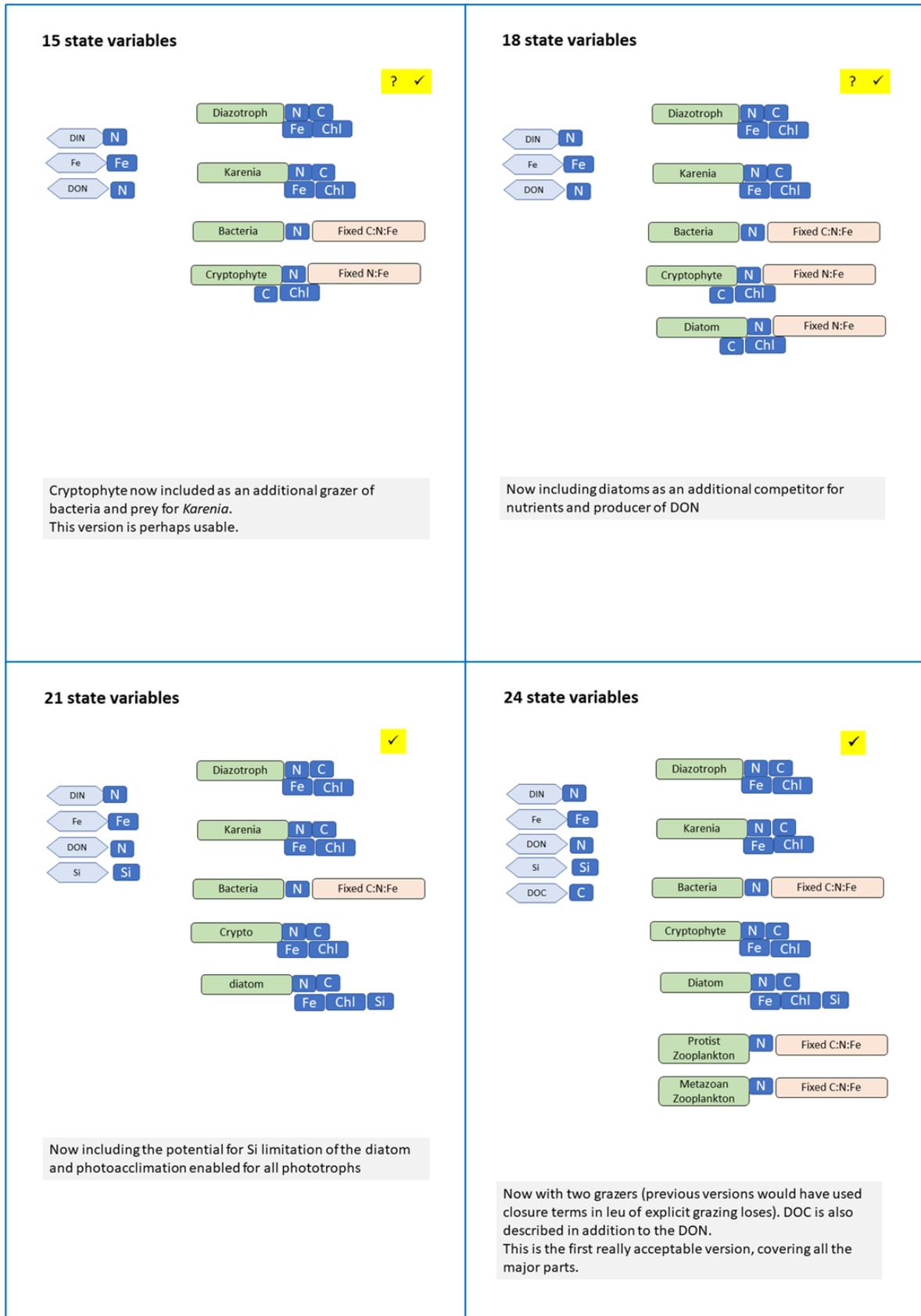


Fig. 7 continued.

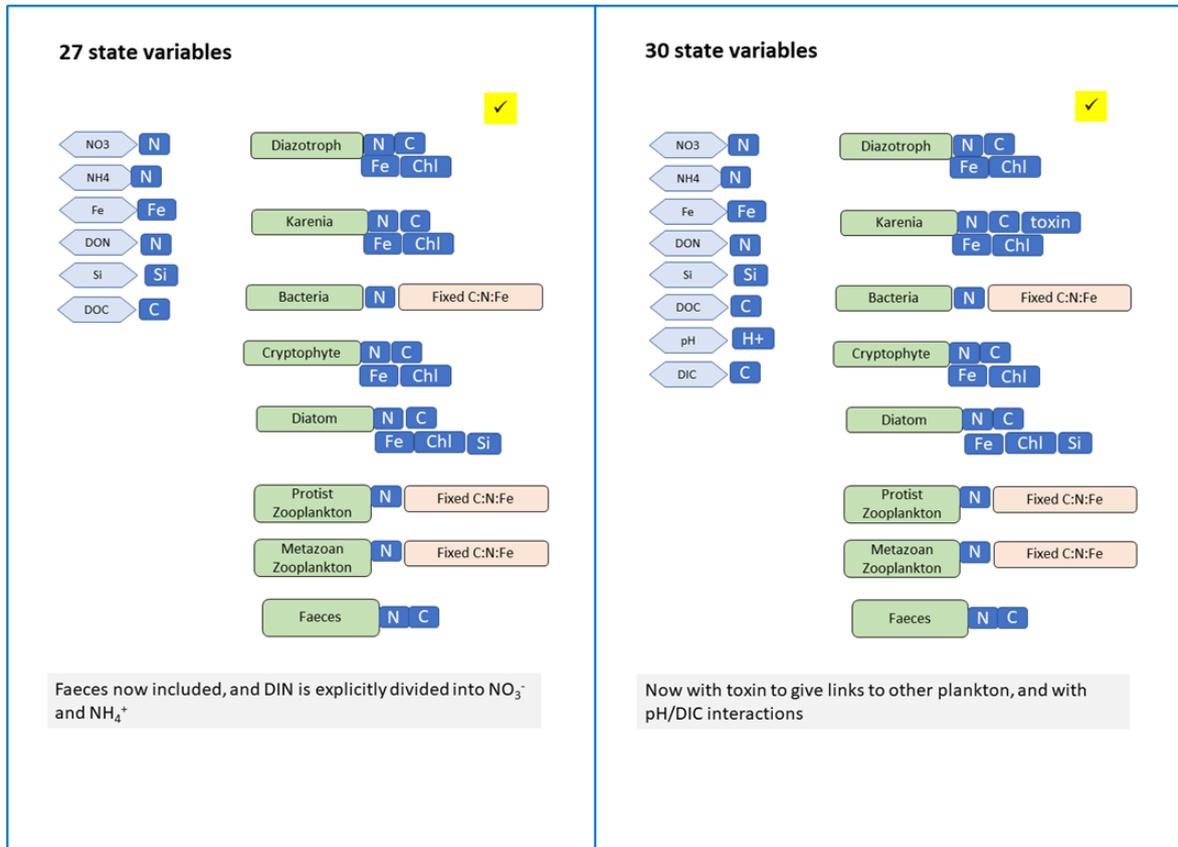
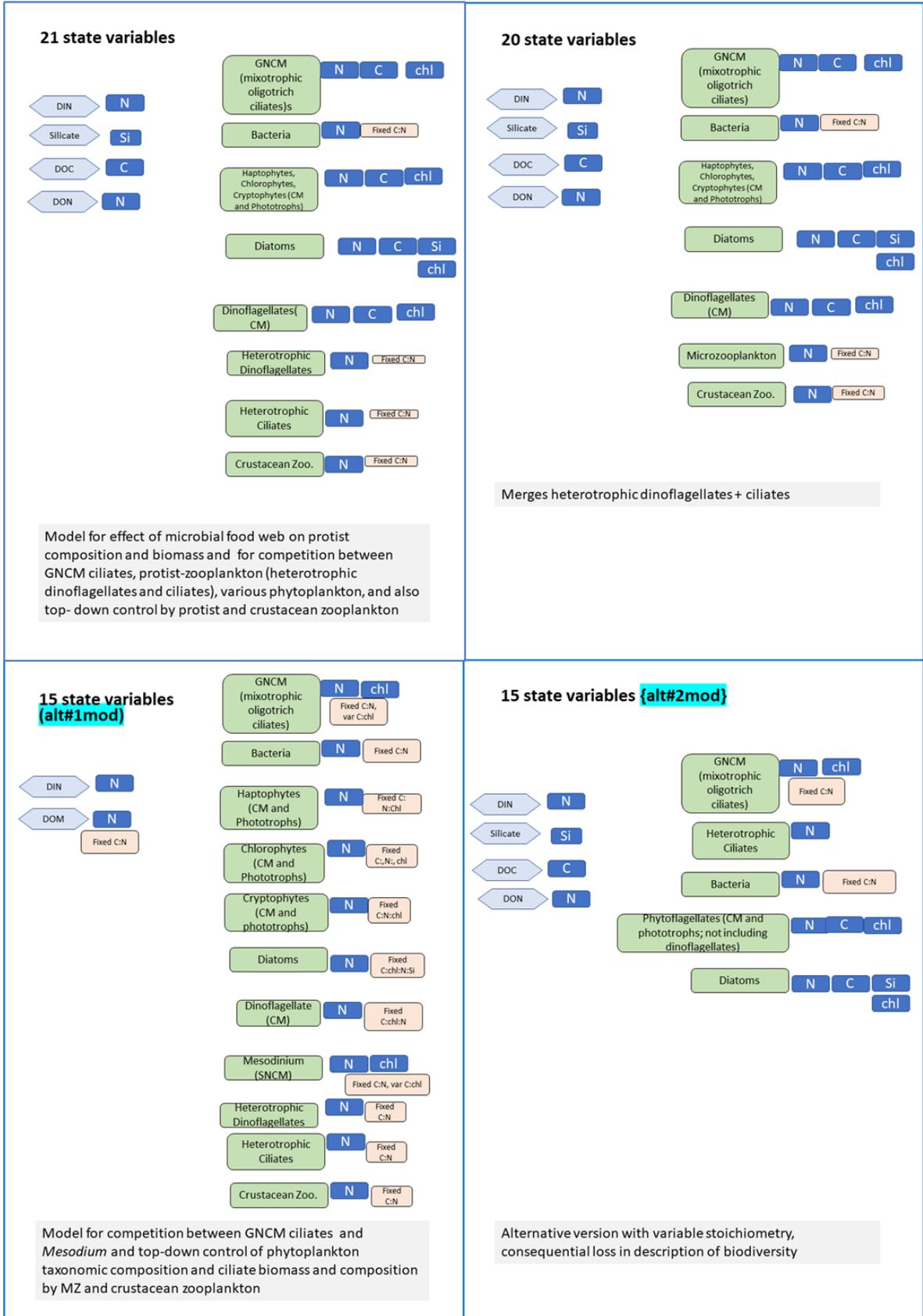


Fig. 7 continued.

## Exploring food web interactions containing GNCM-ciliate mixoplankton



**Fig. 8** Contribution from Stoecker, to explore under what conditions are GNCM (mixotrophic oligotrich ciliates) important components of the plankton in the N. Atlantic spring bloom mixed layer. The duration is of one month.

### 3.3 Consumer target

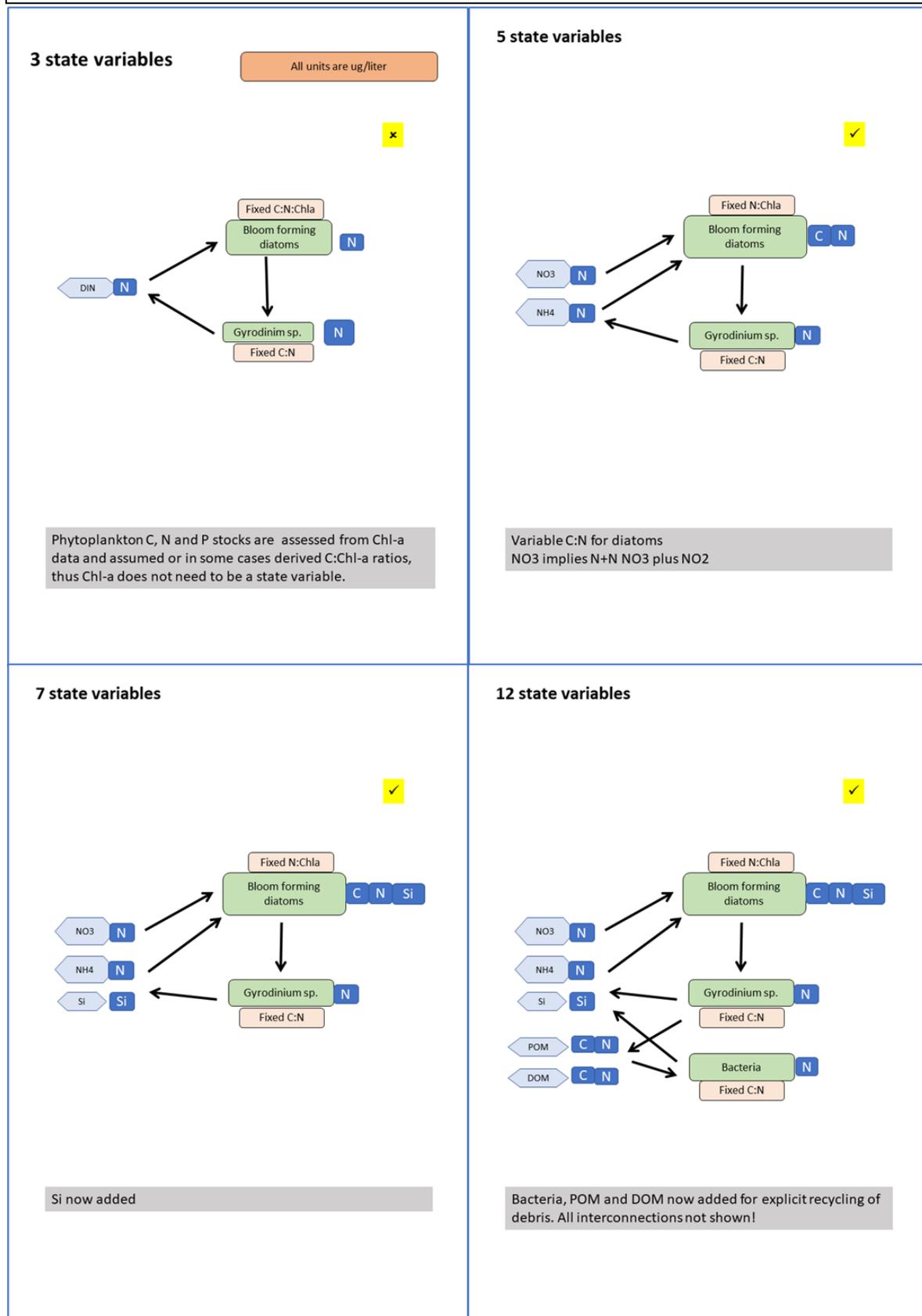
**Figs. 9-11** consider consumers as the focal point for their construction.

**Fig. 9**, by Sherr, considers the role of heterotrophic dinoflagellates, *Gyrodinium* spp, as consumers of diatoms in an upwelling system. Heterotrophic dinoflagellates, including *Gyrodinium* and *Protoberidium* species, can be as significant as mesozooplankton as consumers of production of bloom-forming diatoms (Saito et al. 2006; Sherr & Sherr 2007). This model considers is a field deployment of several weeks. Variable stoichiometry of N and P is driven by input of new nutrients via upwelling and by nutrient recycling via a bacterial-nanoflagellate microbial loop that degrades organic particles (POM) and high molecular weight dissolved organic particles (DOM) produced by blooming diatoms and by grazers. Abiotic features include: temperature, nitrate + nitrite, ammonium, phosphate, silicate, POM, DOM. Biotic features include: chlorophyll, bacterial abundance, components have been quantified in the Oregon upwelling system during diatom blooms and are available as data for model testing (Wetz et al. 2005, 2006). Data sets and papers can be downloaded at <https://ir.library.oregonstate.edu>.

**Fig. 10**, by Schmidt, is focussed on high-latitude copepods, specifically on factors affecting their seasonal dormancy and vertical migration to great depth. It is increasingly noticed that under certain internal or external setting, this migration behaviour and the associated lipid pump does not take place, e.g. seen in *Calanus finmarchicus*, *C. glacialis* and *C. hyperboreus*. Discussed factors are the lack of lipid stores, the lack of predators, attraction to food in surface waters and high buoyancy (Yayanos et al. 1978, Kvile et al. 2019, Hobbs et al. 2020, Espinasse et al. 2022). This activity is associated with variable stoichiometry, both of the prey and the copepods, and also of the sinking debris. Fixed stoichiometry is still used for higher predators.

**Fig. 11** by Atkinson, describes an overwintering scenario in the Southern Ocean, including the role of lipid in supporting that process, and the implications for Fe cycling. The lack of phototrophs, that would perhaps otherwise introduce a need for variable stoichiometry (cf. **Fig. 10**) does not negate the need for variable stoichiometry in the animals.

## Exploring *Gyrodinium* grazing on diatom blooms



**Fig. 9** Submission by Sherr, describing plankton interaction dominated by *Gyrodinium* grazing on diatoms in the Oregon (USA) upwelling systems over the first few weeks of bloom formation and decline.

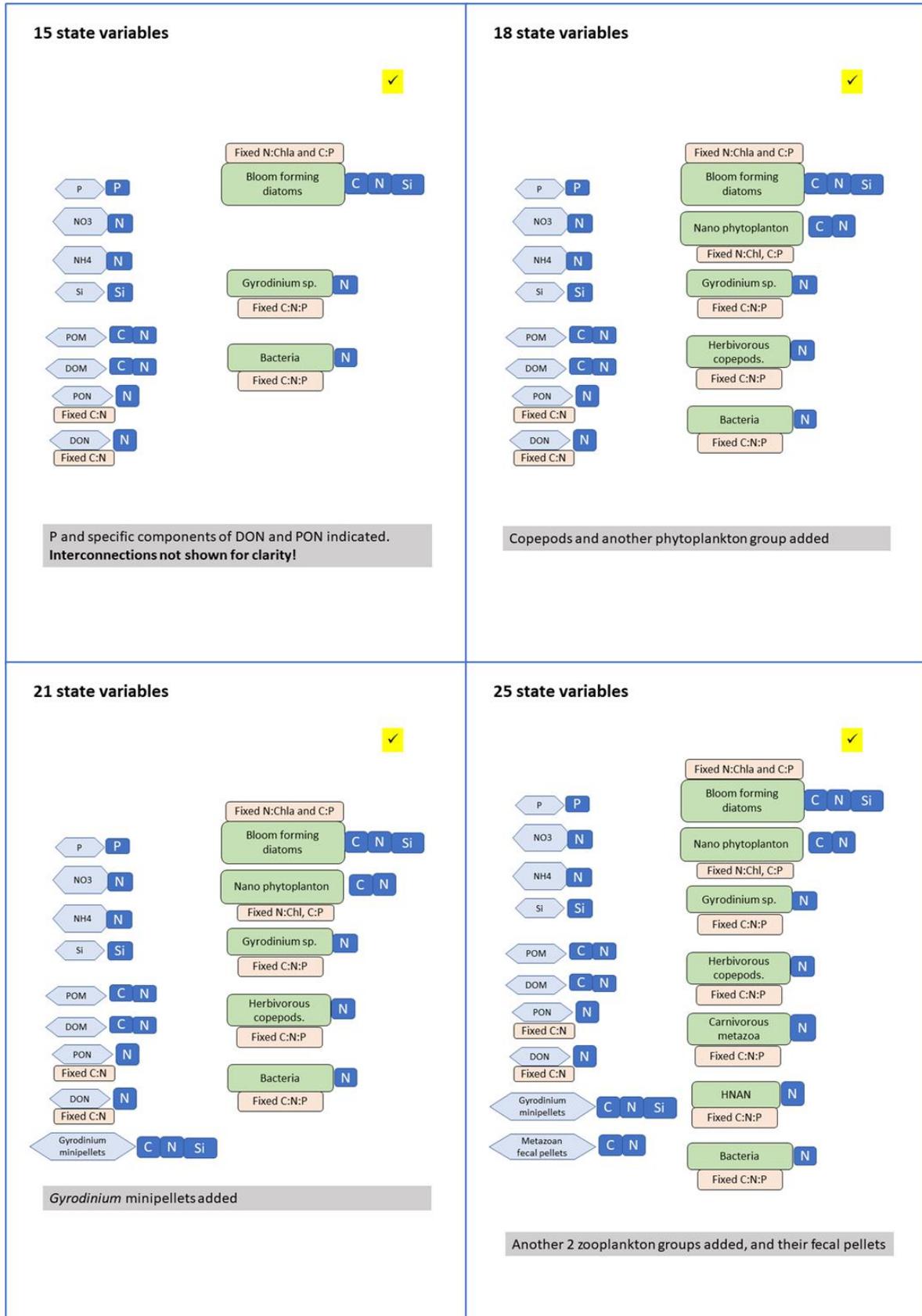
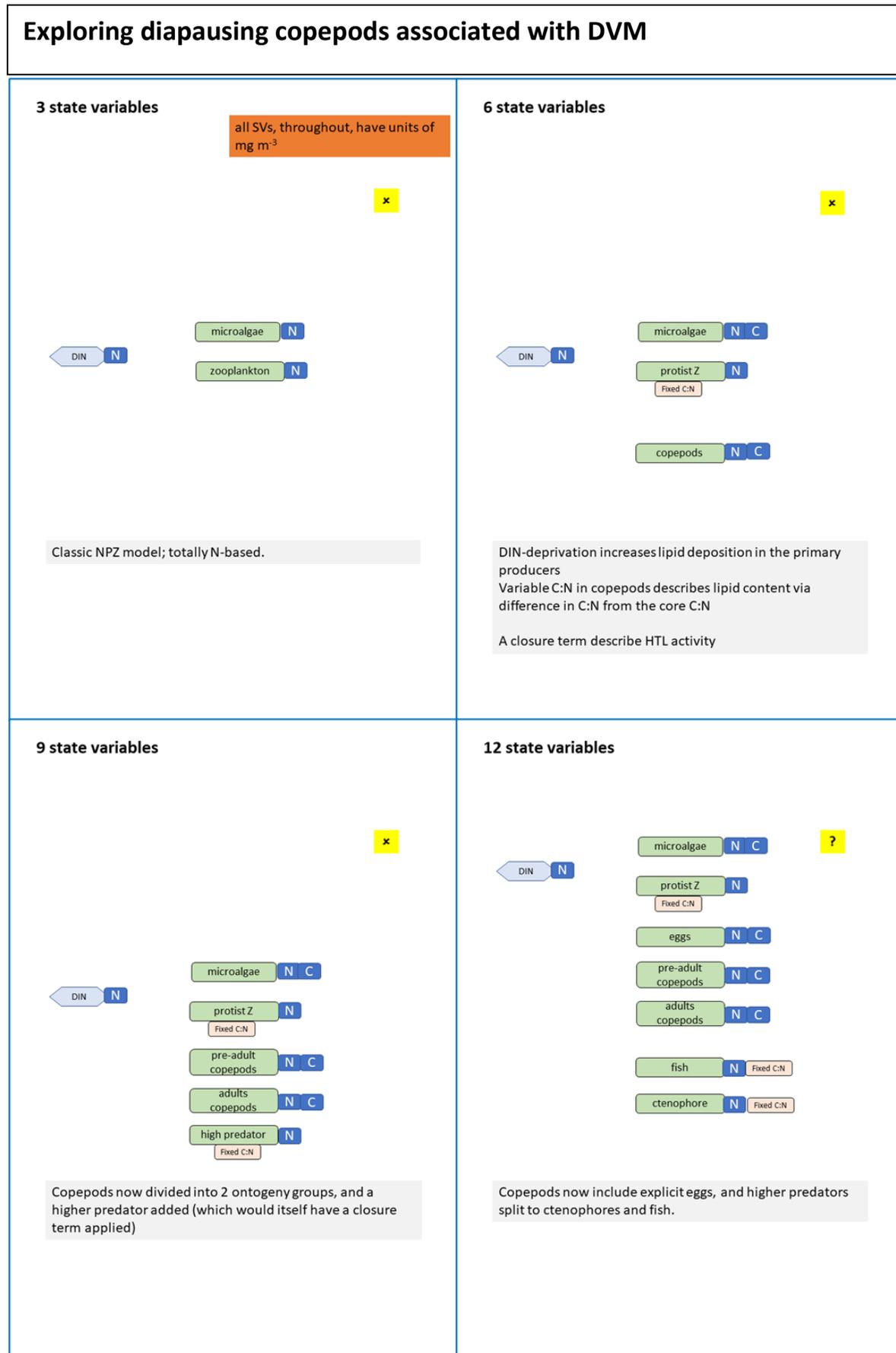


Fig. 9 continued.



**Fig. 10** Contribution of Schmidt, exploring dormant copepods (*Calanus* spp) undergoing seasonal vertical migration in the Atlantic/Arctic, linked to age and storage-buoyancy lipid preconditioning. The duration would be over a year.

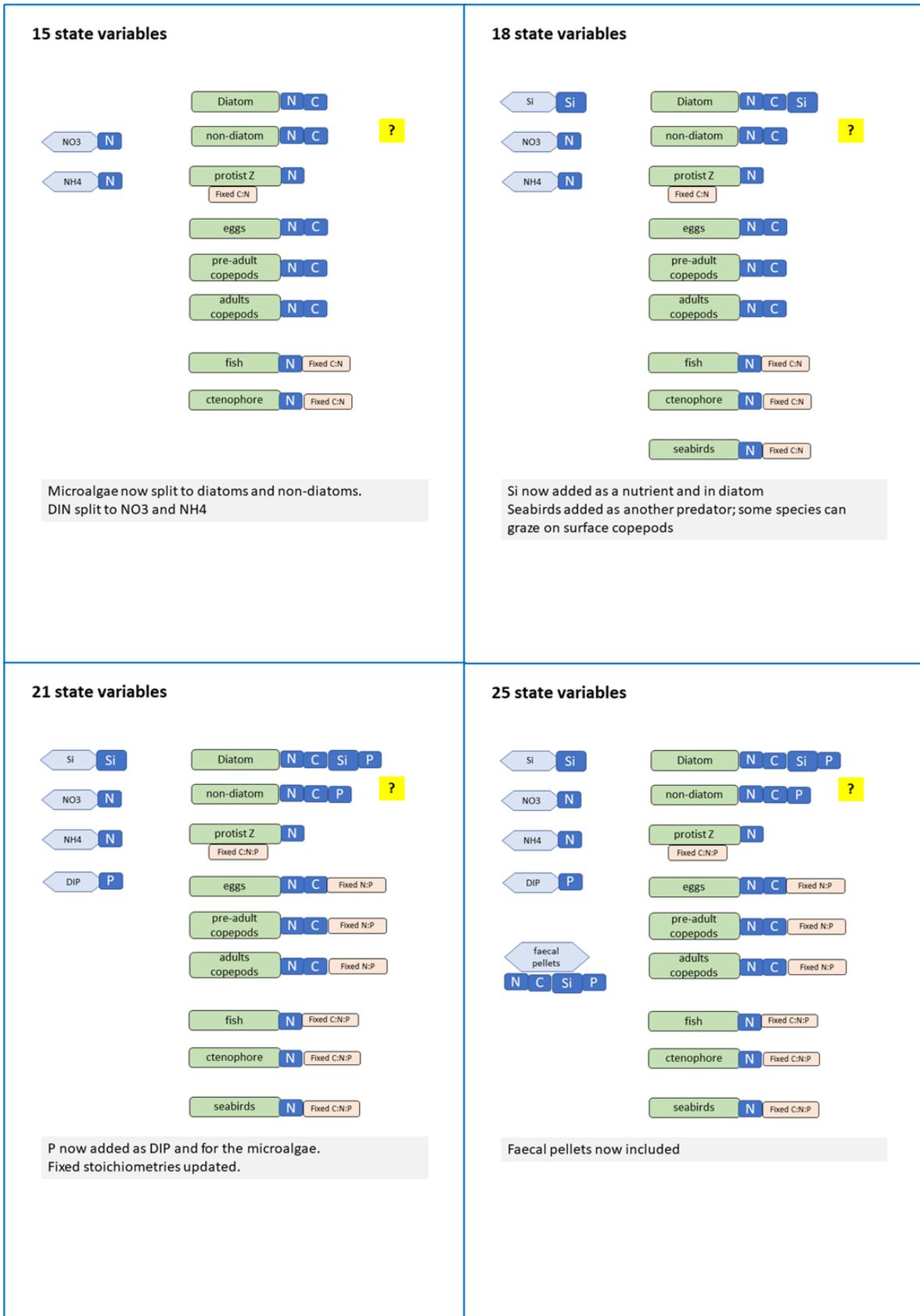
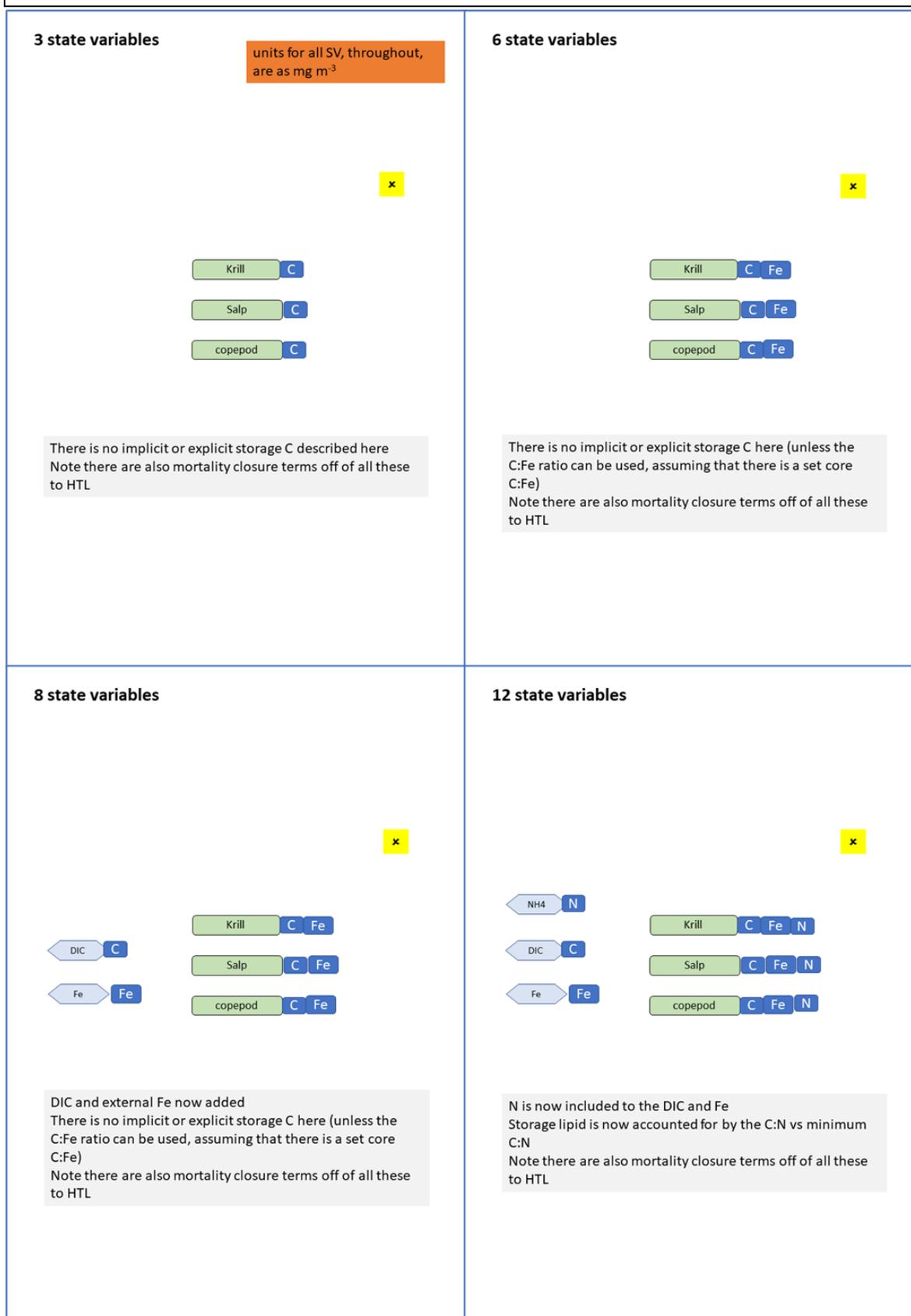


Fig. 10 continued.

## Exploring zooplankton interactions in the Southern Ocean



**Fig. 11** Submission from Atkinson describing interactions over a depth profile in the Southern Ocean with krill/copepod/salp including the lipid pump and C/Fe sequestration. The period is during non-feeding months, during which losses are mainly with respiration and death during migrations.

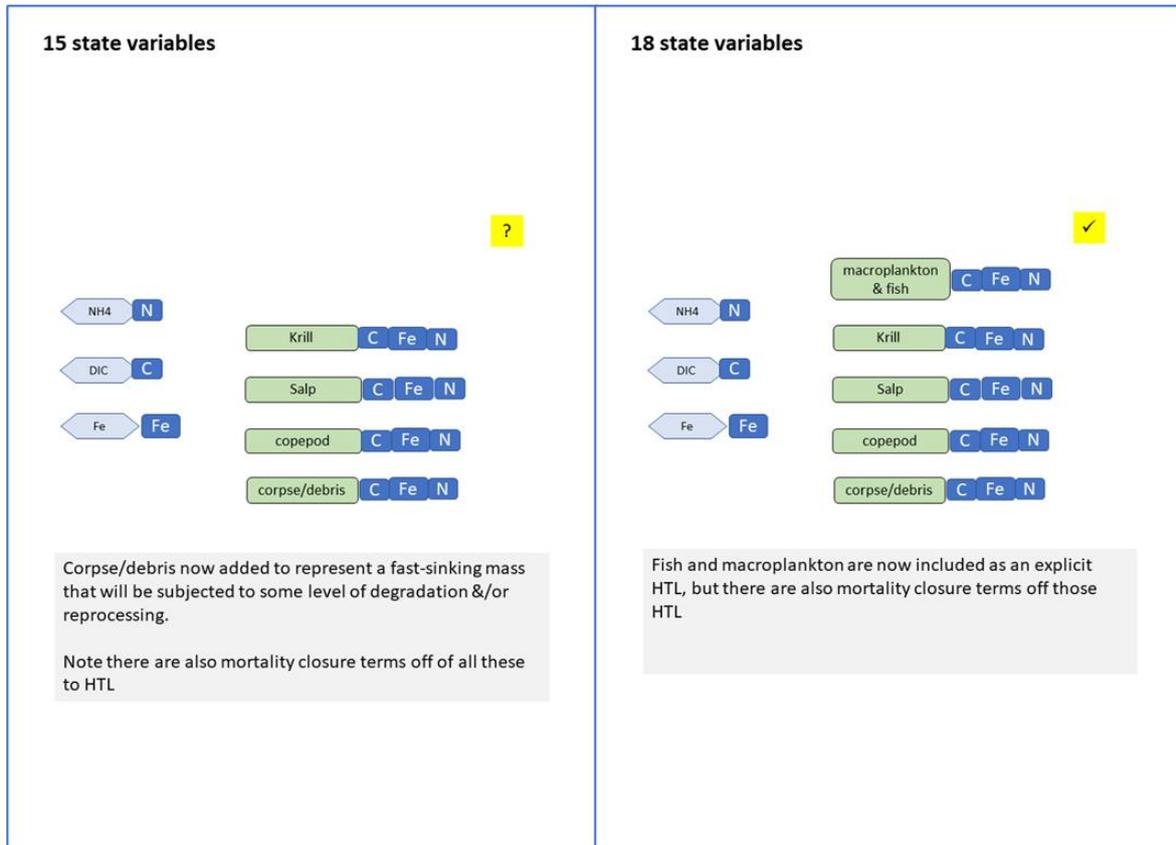


Fig. 11 continued.

## 4. Discussion

The aims of this work were to explore how experts in empirical plankton science might approach the construction of a plankton digital twin, challenged with minimising the computational cost. Elsewhere in this project (Flynn et al., 2024a, 2024b, 2024c), we have seen that the contributors to the whole project (total of 32, with interests from virus to metazoan zooplankton) view a need for complexity in reality to be reflected by complexity (within reason) in models that aspire to provide a digital twin experience.

### 4.1 Trade-offs in simplicity vs complexity

The descriptions provided here (Figs. 3-11) span single-organism autecology (Fig. 4) to highly complex trophic webs with components ranging from bacteria to metazoan.

An interesting trade-off, voiced explicitly by Stoecker (Fig. 8) but evidenced elsewhere as well, is in describing biodiversity vs variable stoichiometry. Given the availability of 3 extra state variables, do you allocate those to a C,N,P description of one organism type, or to 3 organisms assuming fixed stoichiometry? While undoubtedly it is true that food quality is an irrelevance if the item is not even present (or in this instance, not described explicitly), there is an irony and a set of circular arguments that greatly complicate the discussion.

Because of the functioning of stoichiometric ecology (Sterner & Elser 2002), with the relative retention of the components present at the lowest abundance and recycling of excesses, inclusion of variable stoichiometry in models has as a profound effect on trophic dynamics as does the presence or absence of prey items. Furthermore, these effects ripple up and down through the food web, especially as the effects of poor quality food (poor stoichiometric content) can often be non-linear (Mitra & Flynn 2005, 2006). There are also links between stoichiometry and behaviour (resource selectivity, motility, migration) and organism size and ontogeny. In short, the irony is that unless the chemical stoichiometry is indeed tightly constrained, then both biodiversity and variable stoichiometry need careful consideration for inclusion in a plankton model. To complicate matters further, biodiversity emerges in consequence of differential abilities of organisms to modulate their physiology (which is related to their chemical stoichiometry) and whether the needed resources are present (which depends on stoichiometric-ecology-linked recycling and the biodiversity).

This trade-off is also related to the functional type (FT) descriptions. For plankton, FT has often been inferred as relating to biogeochemical functioning (e.g., diazotrophy, calcifiers etc. - Le Quéré et al. 2005). The proper (original) usage of FT refers to ecological functionality (Weithoff 2003), such that organisms within that FT are de facto substitutable with respect to both bottom-up and top-down linkages. There is clear potential for such FT constructions to become increasingly challenged as complexity in resource and predator descriptions will beget complexity in FT descriptions. In a simple NPZ model, the 'P' (phytoplankton, or more accurately phototroph) box can only consume N and can only be consumed by Z.

### 4.2 Pragmatism

Of the contributors, only Thingstad has extensive experience of modelling; the contrast between the approach taken by Thingstad (Fig. 3) and the others (Figs. 4-11) is evidence of pragmatism born of experience in not only building and deploying models but also in trying to locate data to support their testing (see Flynn et al. 2024b).

That said, it is worth recalling that in other components of this project (Flynn et al. 2024a, 2024c) contributors indicated that a driving factor interesting them in plankton digital twin developments is how complexity in planktonic systems plays out. By analogy with engineering usage of models in stress analysis, a role of models for plankton science is for sensitivity testing. How important is inclusion of a given FT, or of a particular facet of behaviour, or of the value of a particular half saturation constant? Testing such matters requires that the most complex model structure used as a starting point is itself fit for purpose such that the stress test may

be considered a true reflection of reality. This is where the argument for deploying plankton digital twins perhaps becomes most powerful, as a means to establish how we may most efficiently and effectively simplify simulation models for placement in regional or global scale simulators.

It is interesting also to note at which point the contributors considered that their concepts became worthy of exploiting. The simplest levels were never considered useful. Most attained wished to continue developing their concept significantly beyond the initial '✓' (acceptable) stage, while some were never totally comfortable with applying a '✓', at best applying '?' (perhaps acceptable).

### 4.3 Conclusions for building plankton digital twins

Although the template instructions (**Appendix 1**) prompted contributors to build models perhaps as far as 48 state variables, none actually progressed much beyond half that value. Accepting that in some instances the contributors misunderstood allocations of state variables, it appears that relatively small plankton digital twin structures would be of utility, perhaps describing 10 or so organisms, or FTs.

Mirroring the outputs elsewhere in the project (Flynn et al. 2024a), contributors expect plankton models to be complex and they are not afraid of such models 'failing'; on the contrary, they expect to learn much from such failings as these indicate priority areas for further research. Such a desire, however, carries within the tacit expectation that the model fails for good reasons and not because what is there is (unknowing to them) dysfunctional (Flynn 2010).

Two other points are noteworthy from the above:

- The model structure needs to be transparent to the non-expert in computer coding, otherwise the user cannot appreciate the subtleties of the format of equations. From responses in Flynn et al. (2024a), that would best be achieved through extensive use of infographics in describing model functionality, rather than using mathematical and computer code as the sole descriptors.
- As much consideration needs to be applied in being able to define the multi-stressor abiotic components as defining the plankton themselves.

Given the contrasting levels of interest in different facets of planktonic systems, a PDT platform that allows flexibility for the end user to design scenarios to their own requirements and operate to provide outputs consistent with expectations, would likely be most attractive. Having some level of plug-and-play capability in such a platform would also be useful; this would enable the user to introduce pre-configured plankton components (as 'players') into the required physico-chemico scenario and then start the simulation.

## References

- Cunningham, B. R. & John, S. G. (2017). The effect of iron limitation on cyanobacteria major nutrient and trace element stoichiometry. *Limnology & Oceanography* 62; 846-858.
- Espinasse, B., Daase, M., Halvorsen, E., Reigstad, M., Berge, J., & Basedow, S. L. (2022). Surface aggregations of *Calanus finmarchicus* during the polar night. *ICES Journal of Marine Science*, 79(3), 803-814.
- Fasham, M. J. R., Ducklow, H. W. & McKelvie, S. M. (1990). A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Research* 48; 591–639.
- Flynn, K. J. (2010). Ecological modelling in a sea of variable stoichiometry: dysfunctionality and the legacy of Redfield and Monod. *Progress in Oceanography*, 84(1-2), 52-65.
- Flynn, K. J. (2024) *Simulating Plankton - getting it right in the era of Digital Twins of The Ocean; projection introduction and executive discussion*. Zenodo <https://doi.org/10.5281/zenodo.10953377>
- Flynn, K. J., Torres, R., Irigoien, X., & Blackford, J. C. (2022). Plankton digital twins—a new research tool. *Journal of Plankton Research*, 44(6), 805-805. <https://doi.org/10.1093/plankt/fbac042>
- Flynn, K. J., Sherr, E., Stoecker, D. K., Thingstad, T. F., Atkinson, A., Beardall, J., Boersma, M., Bowler, C., Calbet, A., Dam, H., Glibert, P. M., Hansen, P. J., Jin, P., Lomas, M. W., Lønborg, C., Menden-Deuer, S., Mock, T., Needham, D., Robinson, C., Rokitta, S., Rost, B., Saiz, E., Schmidt, K., Thiele, S. & Våge, S. (2024a). *Simulating Plankton – getting it right in the era of Digital Twins of The Ocean; building and challenging perceptions*. Zenodo <https://doi.org/10.5281/zenodo.10952555>
- Flynn, K. J., Sherr, E., Stoecker, D. K., Atkinson, A., Boersma, M., Bowler, C., Brunet, C., Lomas, M. W., Lønborg, C., Millette, N. C., Mock, T., Needham, D., Robinson, C. & Rokitta, S. (2024b) *Simulating Plankton – getting it right in the era of Digital Twins of The Ocean; data to support plankton model construction*. Zenodo <https://doi.org/10.5281/zenodo.10953276>
- Flynn, K. J., Atkinson, A., Beardall, J., Berges, J. A., Boersma, M., Brunet, C., Calbet, A., Caron, D., Dam, H., Glibert, P. M., Hansen, P. J., Jin, P., Lomas, M. W., Lønborg, C., Mayor, D., Meyer, B., Millette, N., Mock, T., Mulholland, M., Poulton, A., Robinson, C., Rokitta, S., Rost, B., Saiz, E., Scanlan, D., Schmidt, K., Sherr, E., Stoecker, D. K., Svensen, C., Thiele, S., Thingstad, T. F., Unrein, F. & Våge, S. (2024c). *Simulating Plankton - getting it right in the era of Digital Twins of The Ocean; core features of plankton digital twins*. Zenodo <https://doi.org/10.5281/zenodo.10952868>
- Geider, R. J. & LaRoche, J. (2002). Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. *European Journal of Phycology* 37, 1-17.
- Heneghan, R. F., Everett, J. D., Blanchard, J. L., Sykes, P. & Richardson, A. J. (2023). Climate-driven zooplankton shifts cause large-scale declines in food quality for fish. *Nature Climate Change*, <https://doi.org/10.1038/s41558-023-01630-7>
- Hobbs, L., Banas, N. S., Cottier, F. R., Berge, J., & Daase, M. (2020). Eat or sleep: availability of winter prey explains mid-winter and spring activity in an Arctic *Calanus* population. *Frontiers in Marine Science*, 7, 541564.
- John, E. H. & Flynn, K. J. (2002) Modelling changes in paralytic shellfish toxin content of dinoflagellates in response to nitrogen and phosphorus supply. *Marine Ecology Progress Series* 225; 147-160.
- Johnson, M. D., & Beaudoin, D. J. (2019). The genetic diversity of plastids associated with mixotrophic oligotrich ciliates. *Limnology and Oceanography*, 64(5), 2187-2201.

- Jones, R. H. & Flynn, K. J. (2005) Nutritional status and diet composition affect the value of diatoms as copepod prey. *Science* 307; 1457-1459.
- Kvile, K. Ø., Ashjian, C., & Ji, R. (2019). Pan-Arctic depth distribution of diapausing *Calanus* copepods. *The Biological Bulletin*, 237(2), 76-89.
- Laval-Peuto, M., & Febvre, M. (1986). On plastid symbiosis in *Tontonia appendiculariformis* (Ciliophora, Oligotrichina). *Biosystems*, 19(2), 137-158.
- Le Quéré, C. L., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., ... & Wolf-Gladrow, D. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, 11(11), 2016-2040.
- Makino, W., Cotner, J. B., Sterner, R. W. & Elser, J. J. (2003). Are bacteria more like plants or animals? Growth rate and resource dependence of bacterial C: N: P stoichiometry. *Functional Ecology* 17; 121-130.
- Maselli, M., Altenburger, A., Stoecker, D. K., & Hansen, P. J. (2020). Ecophysiological traits of mixotrophic *Strombidium* spp. *Journal of Plankton Research*, 42(5), 485-496.
- McManus, G. B., Schoener, D. M., & Haberlandt, K. (2012). Chloroplast symbiosis in a marine ciliate: ecophysiology and the risks and rewards of hosting foreign organelles. *Frontiers in microbiology*, 3, 321.
- Mitra, A. (2009). Are closure terms appropriate or necessary descriptors of zooplankton loss in nutrient-phytoplankton-zooplankton type models? *Ecological Modelling*, 220; 611-620.
- Mitra, A. & Flynn, K. J. (2005) Predator-prey interactions: is “ecological stoichiometry” sufficient when good food goes bad? *Journal of Plankton Research* 27; 393-399.
- Mitra, A., & Flynn, K. J. (2006). Promotion of harmful algal blooms by zooplankton predatory activity. *Biology letters*, 2(2), 194-197.
- Mitra, A., Castellani, C., Gentleman, W. C., Jónasdóttir, S. H., Flynn, K. J., Bode, A., Halsband, C., Kuhn, P., Licandro, P., Agersted, M. D., Calbet, A., Lindeque, P. K., Koppelman, R., Møller, E. F., Gislason, A., Nielsen, T. G. & St. John, M. (2014) Bridging the gap between marine biogeochemical and fisheries sciences; configuring the zooplankton link. *Progress in Oceanography* 129, 176-199 DOI: 10.1016/j.pocean.2014.04.025
- Saito, H., Ota, T., Suzuki, K., Nishioka, J., & Tsuda, A. (2006). Role of heterotrophic dinoflagellate *Gyrodinium* sp. in the fate of an iron induced diatom bloom' *Geophysical Research Letters*, 33, L09602, doi:10.1029/2005GL025366.
- Schmidt, K., Birchill, A. J., Atkinson, A., Brewin, R. J., Clark, J. R., Hickman, A. E., ... & Ussher, S. J. (2020). Increasing picocyanobacteria success in shelf waters contributes to long-term food web degradation. *Global Change Biology*, 26(10), 5574-5587.
- Sherr, E. B. & Sherr, B. F. (2007). Heterotrophic dinoflagellates: A significant component of microzooplankton biomass and major grazers of diatoms in the sea. *Marine Ecology-progress Series*, 352, 187-197, DOI:10.3354/meps07161
- Sterner, R. W. and Elser, J. J. (2002) *Ecological Stoichiometry: the Biology of Elements from Molecules to the Biosphere*. Princeton University Press, Princeton, NJ
- Stoecker, D. K., Silver, M. W., Michaels, A. E., & Davis, L. H. (1988). Obligate mixotrophy in *Laboea strobila*, a ciliate which retains chloroplasts. *Marine Biology*, 99, 415-423.

- Stoecker, D. K., Silver, M. W., Michaels, A. E. & Davis, L. H. (1988/89) Enslavement of algal chloroplasts by four *Strombidium* spp. (Ciliophora, Oligotrichida) *Mar. Microbial Food Webs* 3: 79-100
- Weithoff, G. (2003) The concepts of 'plant functional types' and 'functional diversity' in lake phytoplankton—a new understanding of phytoplankton ecology? *Freshwater Biology* 48; 1669– 1675
- Wetz, J. J., Corwith, H., & Wheeler, P. A. (2005). *Nutrient and extracted chlorophyll data from the GLOBEC Long-term observation program 1997-2004*. COAS, Oregon State University, Data Report 193, Reference 2004-1, 2004, Revised 2005.
- Wetz, J. J., Arrington, J., & Wheeler, P. A. (2006). *Particulate and dissolved organic carbon and nitrogen data from the GLOBEC Long-term Observation Program, 1997-2004*. College of Oceanic and Atmospheric Sciences, Oregon State University, Data Report 204, Reference 2006-1.
- Yayanos, A. A., Benson, A. A., & Nevenzel, J. C. (1978). The pressure-volume-temperature (PVT) properties of a lipid mixture from a marine copepod, *Calanus plumchrus*: implications for buoyancy and sound scattering. *Deep Sea Research*, 25(3), 257-268.

## Appendix 1

The following, with renumbering of subtitles and Figures, are the instructions as provided to the contributors. In addition, a Powerpoint template (in A4 format) was provided, that gave example material to copy-paste (see **Fig.A1**, below, for examples).

### A1 Introducing the task

This task requires a consideration of the minimum level of complexity likely required to achieve a plausible structure for a plankton digital twin. That plausibility will depend on the operational scenario of the DT; this will dictate the required complexity of the biological components.

In Task 1, contributors were asked to consider the construction of an autecology description of a planktonic organism of interest to them. In many instances (involving consumers or viruses) this would have also needed a consideration of interactions with other organisms. For this task there will invariably be a need to consider interactions between plankton types or species, and those interactions will bring in additional layers of complexity. While one may follow a predator-prey interaction simply by changes in numeric abundance, the actual interactions involve amongst other things the nutritional quality and physiological status of the prey (or host) and of the consumer; that may require knowledge of some components of the C:N:P:Fe:Si:lipid stoichiometry, changes in organism size and motility, etc. etc. For bacteria and other microbial plankton, there is also the question of which nutrients need to be included; for components of DOM that also requires consideration of how you wish to describe those organic fractions.

The simplest usable model of plankton is typically viewed as the 3-state-variable NPZ model mentioned in Section 1. You may, or (more likely) may not, think that such a structure provides a digital twin experience, but how would you improve it? What is the minimum level of additional complexity that you would require for a particular task? More importantly for a digital twin description, what would your fellow plankton experts expect to be present in the digital water column? Remember that a digital twin is not only for your use; it is for use by anyone who wishes to explore the scenario enabled by the digital twin.

#### A1.1 Allocating state variables

The task requires that you consider how you would allocate state variables in a model structure, starting from the NPZ level of an initial 3 state variables and then incrementing complexity by sequential additional of blocks of 3 state variables.

Why increment in 3's? It is useful to at least consider additional of blocks rather than just one state variable at a time, and with 3 you could add a {C,N,P} description of a bacteria, a {C,N,Chl} description of a phytoplankton, or a {body-C, gut-C, lipid-C} description of a copepod. Too-large an increment prevents a searching consideration of how to allocate extra state variables. Too-small an increment and the task will become perhaps painfully slow.

Remember, a state variable (SV) is something with a history, something tangible. Examples of state variables include:

- organism numeric abundance (1 SV is needed for each organism type)
- nutrient (1 SV is needed for each of nitrate, ammonium, phosphate, glucose, amino acid, O<sub>2</sub>, etc.)
- chemical content of the organism (1 SV is needed for each organism, for each of C, N, P, Fe, Si, Chl, phycocyanin, Ca, lipid, toxin, acquired phototrophic NA, etc.)
- stoichiometry of DOM and POM, for example released as waste products and used as resources by other organisms (1 SV is needed for each component of this material, for each of C, N, P, lipid, etc.)

To make things more complicated, if your prey item describes (for example) C,N,Fe, then your consumer should also describe those chemicals as well. That is not always so (you would perhaps not want to waste a state variable for Chl in a copepod, unless you were interested in pigments in the gut), but you need to think about it. And of course, you could assume a fixed C:N:Fe, or a fixed C:N, in the consumer (see below).

### A1.2 Constant stoichiometry?

Stoichiometry refers to chemical composition. In plankton models, this typically refers to C:N, C:P, though more appropriately and comprehensively it would refer to C:N:P:Fe:Si:Ca etc. Stoichiometry may also refer to compounds, such as C:Chl, C:lipid, etc. It could also refer to a constant organism (typically cell) size, as  $C \text{ cell}^{-1}$ , for example. The ratios may be described in which way suits convention (e.g., N:C is used in quota models, but often C:N is used in reports of chemical analyses). The ratios may also be described as mass or mole ratios; it does not matter for the model (and you do not need to concern yourself with this here) as long as the units are given and transformed if and when appropriate.

If you have components in the model of fixed chemical stoichiometry, or that you can reliably assume to have a tightly constrained stoichiometry, you can save on the allocation of state variables. At the extreme, this was what was done with the original NPZ models, which contained just 3 state variables for the plankton (phytoplankton, zooplankton, bacteria), all of which assumed Redfield-like fixed stoichiometry (Fasham et al. 1990). Transforms from Chl to phytoplankton biomass in such models also assume a fixed N:Chl. Extant models used for biogeochemical descriptions often assume fixed stoichiometry for certain components.

For some plankton, the assumption of fixed stoichiometries is both logical and pragmatic. Viruses and protist zooplankton, for example, may have fixed or essentially invariant C:N:P. Bacteria may be assumed to have a fixed C:N:P:Fe, though evidence suggest that actually these ratios can be highly variant (Makino et al. 2003; Cunningham & John 2017), as they are for phototrophs (Geider & LaRoche 2002).

For metazoan zooplankton, it may be appropriate to assume a fixed chemical stoichiometry for certain ontogenic stages, but not for others. It may, for example, be important to explicitly describe changes in lipid content in over-wintering stages, and in eggs.

Voided particulate organic material (VOM; i.e., faecal matter) will inevitably contain a variable chemical stoichiometry if the consumers are feeding on material of different stoichiometries, but unless you are size-fractionating this material, the wastes from all plankton can go to one set of VOM\_C, VOM\_N, VOM\_P state variables, for VOM C, N and P, respectively.

### A1.3 Allometry and functional grouping

Perhaps the fundamental difference between modelling (including for digital twins) and empirical plankton science is that the former by necessity has to group organisms according to commonalities, while the latter most often targets differences. In this task, that difference is brought into sharp focus.

There are many ways in which plankton species may be segregated or conversely grouped. Grouping according to allometry is very common, although the basis for those groups may be opaque at best. For example, in many models, all diatoms are placed together in one group; given that diatom size span ca. 5 – 500  $\mu\text{m}$  that practice is perhaps at the very least questionable. On the other hand, phytoplankton or mixoplankton of very different physiologies may be placed in size bins, such as ‘nanoflagellates’ or ‘microflagellates’.

What is clear is that you are limited in what you do with a model, and to make sense ecologically anything you do at one trophic level requires a plausible selection of options at the next, and perhaps the preceding, trophic level. There is arguably little reality achievable in describing the ‘phytoplankton’ as 50 or so separate

groups, and having them eaten by only 5 or so 'zooplankton' groups assuming a fixed 1:10 prey:predator size ratio. How will you decide the basis of your groupings?

So-called biogeochemical functional groups have provided the base for most plankton models, notably between Si-containing diatoms, Ca-containing coccolithophorids, N<sub>2</sub>-fixing diazotrophs etc. For plankton digital twins it is likely that you will wish to consider functional groups in their original meaning (Weithoff 2003), as ecological or trophic homologs. Clearly all organisms do not have strict homologs, else we would not see the biodiversity that we do, so our placement of organisms within groups will depend on how 'zoomed in' we require the digital twin. Your choice will reflect the scenario, the temporal setting and the main target of the digital twin. Your groupings will also likely change as you progress through the task, adding more state variables.

#### A1.4 Ontogeny and life cycle development

For descriptions of microbes, you may be able to justify a fixed cell size. This saves a state variable (you can transform C-biomass to numeric abundance) but also prevents you from considering changes in size not only during the diel cycle but also in consequence of nutrient stress (e.g. John & Flynn 2002); such changes can be important for trophic interactions. You may wish to include resting stages for your microbes; if you can assume a fixed chemical stoichiometry for those stages, that saves on state variables.

The real challenges are within the zooplankton descriptions. Most plankton models use an extremely simplistic description of 'zooplankton'. Often any and all zooplankton, and their life stages, are grouped into one 'box'. This zooplankton group then feeds on perhaps a few, or in some models very many, 'phytoplankton' with little or no allometric or other specialisation. Reality is very different, but reality is also very complex, and very expensive to describe in a model. You need to consider which parts of reality have most leverage on the behaviour of the model. Ironically, the best way to determine that is to conduct a sensitivity analysis on each part, which requires that you include everything, and then strip each away to test the consequences.

With respect to metazoan zooplankton ontogeny, and taking a copepod as an exemplar, one option (in line with many experimental studies) is to just consider 'female egg-laying adults'. At the other extreme, one could consider a full age-stage description where individuals (or more likely, cohorts) age each day, and if appropriate they also move into the next ontogenic stage. State variables may be saved by at least merging some of the stages together, such as {eggs + non-feeding nauplii}, {feeding nauplii + early copepodites}, {late copepodites + adults} (e.g., Flynn & Irigoien 2009). State variables may also be saved by increasing the aging bands, perhaps from 1 day to 1 week, or using other approaches to 'age' the stages.

Often these stage transitions also mark important trophic interactions, perhaps affecting prey selections, assimilation efficiencies, mortality rates, etc. The stage transitions may also necessitate additional state variables; if storage lipid is described, then that may only be needed for adults and overwintering copepodites, and perhaps for eggs.

There is also the issue of merging animals into few functional groups; what is the basis for that grouping - do they share the same feed and the same, or similar, life histories? Depending on the scenario being considered, and especially the temporal setting, more or fewer age/stage descriptions may be appropriate. As you work through the task, you will need to consider how best to handle the challenge in a way that can still be acceptable, or that at least minimises conflicts with reality.

You may think, if your primary interests are in microbial loop communities, that these are not matters that affect you. But they do. If the operation of the microbial loop depends on the activity of bacterivorous flagellates, then the success of copepods and other metazoans that consume those flagellates will impact the loop. Conversely, the success of different microbes for the acquisition of nutrients affects the prey field

required to support the growth of those metazoans. It is because of such matters that you should ideally work on this task with colleagues with interests that are some distance from your own.

### A1.5 DOM

DOM present a particular challenge that requires careful consideration for descriptions of microbial loop and microbial carbon pump activities.

Traditionally, DOM has been either ignored or described as a single component (e.g., 'DON') or very few components (e.g., 'labile DOC' vs 'semi-labile DOC'). In reality the material is poorly characterised chemically, and often (in consequence of the way that samples are prepared for analysis) includes material that is not technically dissolved (e.g., including viruses, cell structure debris, etc.).

Some parts of DOM may be considered as having a fixed stoichiometry in terms of C:N:P, while in other parts and certainly collectively, the stoichiometry is variable. There is thus a choice that you will have to make, and that choice will have to be applied both to the producers and the consumers of the DOM fractions.

1. Assume a particular component has a fixed stoichiometry – this costs 1 SV for each component (typically either as  $\text{mgC m}^{-3}$  or  $\text{mgN m}^{-3}$ )
2. Assume a variable stoichiometry – this costs 1 SV for each element being considered in the component (e.g., C, N, P)

You can combine approaches; amino acids, nucleic acids and proteins could perhaps be assumed to have the same, essentially invariant C:N content. Descriptions of semi-labile through to extreme-recalcitrant forms of DOM may perhaps be considered as a single form of DOM (distinguished by requiring extracellular digestion) with a variable C:N:P, with the rDOM being very high C:N and C:P. Cellulose-like material, which is relatively recalcitrant but contains no N or P, would be another component.

By the time you have considered labile DOC (sugars, lipid), DFAA and NA, proteins, TEP, rDOM, etc. etc., each with either fixed or variable stoichiometry, you can end up defining a great many state variables just to supply substrates for bacterial growth. You will need to decide where you are going to start and stop.

### A2 Task instructions

To help you with this task, a template is available as a PowerPoint file. As you work through the task, each successive slide should describe the same conceptual model but of increasing complexity through addition of 3 more state variables, and also (as you require) by reallocation of the previously described state variables.

Feel free to modify and develop the template if you need to, though it will greatly facilitate subsequent merging of ideas if you can largely retain the template structure. The slide format is A4; depending on how your model develops you may wish to increase the font size, or alter between landscape and portrait.

Each model-development slide will contain state variables divided between abiotic and biotic components. The first slide will most likely contain 1 abiotic nutrient, and 2 interacting biotic components.

If you have questions, please first check the 'Frequently Asked Questions' (FAQ) section of this document, or email [KJF@PML.ac.uk](mailto:KJF@PML.ac.uk)

### A2.1 Working alone or in a small group

It would be instructive to work in pairs, or as a group of 3, for this task. You should ideally work with someone with interests that are removed from your own interests by at least 1 trophic level. For example, if your interests are in viruses, with a bacteria host, work with someone with interests in protist grazers, or higher trophic levels. The reason for this is because digital twin models are, unlike traditional models, deliberately designed to be used by anyone, and that 'anyone' may likely come into a topic with a different perspective to your own.

Information of the interests of colleagues who have, at the time of writing, submitted their Task 1 information are provided in {see **Appendix 1** in Flynn 2024}; this may help you locate a suitable person to work with. There is nothing preventing you from contributing to several pairs, or of repeating the same biological construct but operating in a different scenario and/or temporal setting (see below).

If you wish to work alone, try and be especially self-critical of the structure as you develop it.

**You can involve a person from outside of the project as well – please see FAQ.**

### A2.2 Select the physical scenario

Select the scenario you wish to consider. It could, for example, be 'temperate spring water column', 'tropical oligotrophic gyre', 'summer Antarctic'. Most importantly, you need to decide whether this is an artificial setting where you control which plankton are present, or is it a natural setting where you have no control over the plankton that are present (other than perhaps by filter screening by organism size). This decision affects how you build the model. For artificial scenarios, please work through the task to involve at least 3 organisms (species, function types); for field scenarios you will have the challenge of how to group organisms.

You then need to consider the temporal setting of the scenario as a short, medium or long-term study; that period affects the types of detail or focus that you would require of your digital twin. For example, if the study period is very short, then perhaps you do not need to worry about life cycle developments in metazoan zooplankton.

### A2.3 Build the starting configuration

From your core interest, you need to decide how you are going to allocate the first 3 state variables, to what organism(s), or functional type(s), to what nutrients, and with what units (e.g. C, N, abundance). If your interests do not permit any useful form of model with just 3 state variables, you will have to compromise and close the gaps as you add more state variables (see on). Recall that a base plankton model, akin to those used for IPCC work, may contain only 1 or 2 'phytoplankton' and 1 'zooplankton'! If you assume saturating nutrients, then your 3-SV model may contain only biological components.

Provide a brief description of the structure.

### A2.4 Sequentially add 3 state variables

At each iteration, for each new slide in the template, you have to add another 3 state variables. You could enlarge or rebuild the descriptions of the components that you have already, or start to add new details. See **Fig.A1**.

NOTE: it may well be that just adding 3 state variables does not provide a useful development, and that you will have to iterate this addition several times. If that happens, please add a comment noting that that particular model state is 'incomplete', and proceed with the next iteration. However, please try to avoid this

option. If you do not need to use all 3 additions, just leave a blank state variable on the slide for use in the next iteration.

Update your description of the structure.

### A2.5 And repeat!

Repeat the process in A.4, until you reach a total of ca. 30-48 state variables, spread across the abiotic and biotic components. Depending on how you have configured your model, that may be sufficient for 10 organism groups. However, if you describe stoichiometry with greater detail (e.g., C:N:P:Si:Fe vs C:N:P), or you increase the detail for ontogeny (e.g., eggs, nauplii stages, copepodite stages, adults vs eggs, nauplii, copepodite+adult) you will use up your state variables very quickly.

**You are encouraged to request feedback on your developing project, please send the file at stage '21 state variables' to [KJF@PML.ac.uk](mailto:KJF@PML.ac.uk)**

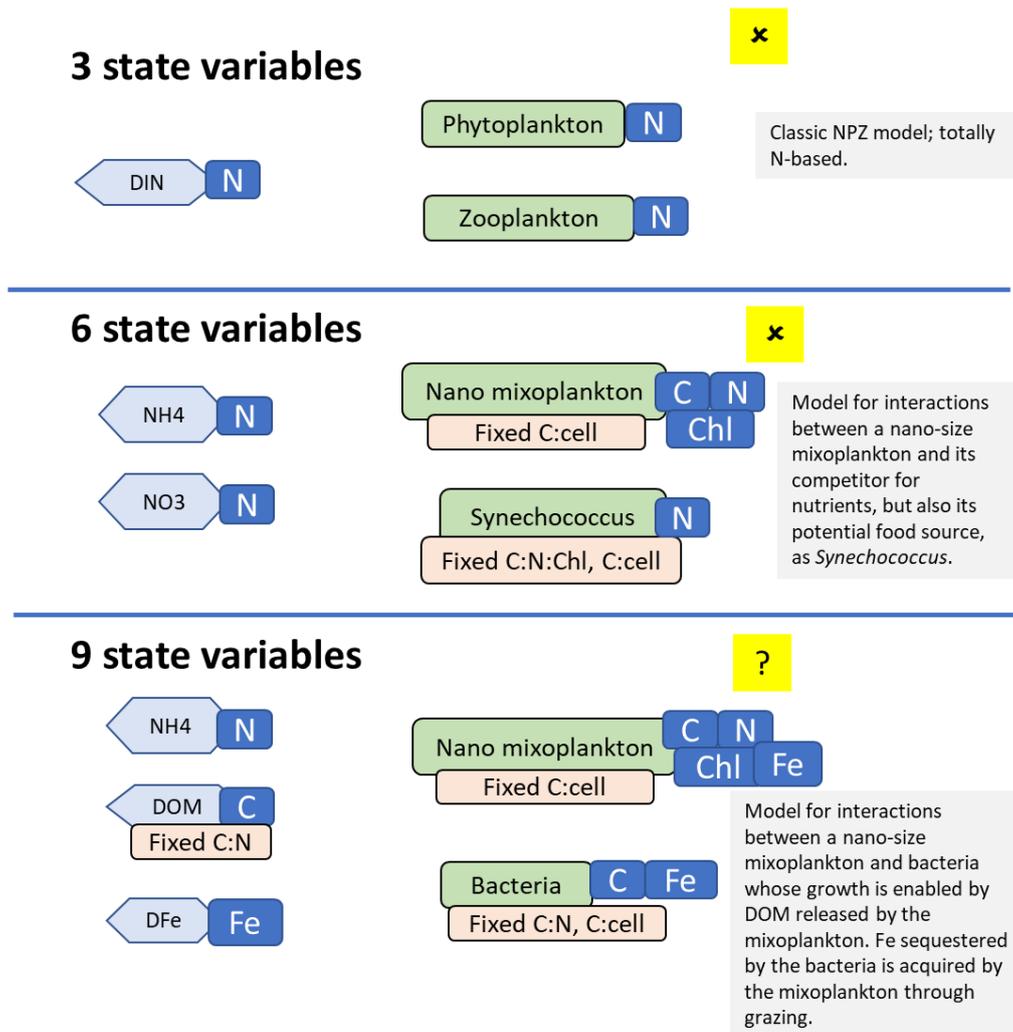
### A3 Acceptability

Once you have finished iterating A2.4 – A2.5, go back through your slides, starting from your first 3-SV model, and ask yourself whether you could ever consider that description as providing the core of a digital twin for your chosen planktonic scenario.

For each slide, in the top right-hand corner, please add the acceptability symbol for whether you find that configuration as 'Not acceptable' (✖), 'Marginally acceptable' (?), or 'Acceptable' (✓). See **Fig.A1**.

You need to be self-critical; do you think your colleagues would agree with your grading of each slide?

If you never get as far as 'Acceptable', you may wish to repeat the addition of further block of 3 state variables until you attain your goal.



**Fig. A1** Examples of model complexity, as configured using the Task Template. Note that these are not intended as developments of the same concept! The '**3 state variables**' model is a representation of a classic NPZ model, and is totally N-based. The description is not considered as acceptable as the core for a digital twin, hence the '**✘**' symbol in yellow. The '**6 state variables**' model shows a model for the interactions between a nano-size mixoplankton and its competitor for nutrients, but also its potential food source, as *Synechococcus*. Note that the *Synechococcus* is described here as having a fixed C:N:Chl stoichiometry, but is described itself in N units, consistent with the inorganic nutrients, which are also N. Organism abundance, and predator-prey encounter rates, assume fixed cell size as C cell<sup>-1</sup>. The description is not considered as acceptable as the core for a digital twin, because the *Synechococcus* is not described as having variable stoichiometry (which affects its nutritional value as prey for the mixoplankton), and it is also incapable of describing photoacclimation (for which it needs variable C:Chl), hence the '**✘**' symbol in yellow. The '**9 state variables**' model is also of a nano-size mixoplankton, but this time it provides for an interaction with bacteria. The bacteria growth is enabled by DOM released by the mixoplankton, while the Fe sequestered by the bacteria is then acquired by the mixoplankton through grazing. Here, the bacteria has variable C:Fe stoichiometry, but fixed C:N. The DOM is assumed as having fixed C:N. Organism abundance, and predator-prey encounter rates, assume fixed cell size as C cell<sup>-1</sup>. The description may be considered as potentially acceptable as the core for a digital twin, hence the '**?**' symbol in yellow.

## A4 When you are done

Please be sure that the following aspects are made clear on slide 1:

- contributors are acknowledged
- the focus of the work
- the physical setting
- the temporal setting

Please submit your final PowerPoint file (as pptx, please) to [KJF@PML.ac.uk](mailto:KJF@PML.ac.uk)

THANK YOU

## 'Frequently Asked Questions' (FAQs)

### **Can I have additional guidance and help?**

**A:** Yes, of course. It is also recommended that you send an intermediate version of your project (at around the '21 state variable' stage) to [KJF@PML.ac.uk](mailto:KJF@PML.ac.uk) for feedback.

### **Can I work with more than one other person / Can I work on my own?**

**A:** If you wish, you could work as a team of 3, but if you do, please ensure that each team member is well separated in their interests from the others. You could work alone, but if you do so, please try and ensure that you take the perspective of plankton research interests that are well distanced from your main interests.

### **I cannot find someone to work with; can I involve someone who is not listed as a member of the project?**

**A:** Yes, by all means. BUT, please only involve experts in the ecophysiology of real plankton (i.e., not pure modellers). If you wish to involve another person, please email [KJF@PML.ac.uk](mailto:KJF@PML.ac.uk) so that they can be introduced to the project.

### **Can I link up the components to make the model look like a food web?**

**A:** By all means do so - you can use the object-link tools in Powerpoint, or just use arrows, or draw lines as you wish. However, you may decide not to make links back to voided material and nutrients, as there will be a great many arrows to and from these components.

### **How can I best economise on state variables?**

**A:** This requires a combination of grouping organisms with common features both for what they need and (equally important) for who needs them as prey, and also considering how you describe forms of DOM. If you can accept an assumed fixed stoichiometry, then that will save many state variables.

**What is wrong with assuming fixed stoichiometry? Values in nature do not vary much compared to values in the extreme conditions of laboratory cultures.**

**A:** Nothing is wrong with making such an assumption provided that you have good cause to believe that it gives an adequate description and does not affect the dynamics of the interactions. Problems with assuming fixed stoichiometry stem from the non-linearity of the consequences (Mitra & Flynn 2005). For example, small differences in C:N:P can mask important changes in biochemical composition (such as production of toxins and allelopaths) that can radically change palatability of prey and/or the efficiency of conversion of prey biomass into consumer biomass (e.g., Jones & Flynn 2005). Differences in the C:N of phytoplankton growing in N-replete conditions on nitrate are often higher than those growing on ammonium; the stoichiometric consequences of that are important with respect to the lipid content of those organisms, and for the subsequent regeneration of N on consumption of those organism.

**If we use (assume) a fixed stoichiometry, can we use different values for different components? Do we have to assume Redfield?**

**A:** You should use whatever ratio is appropriate. The problems with using Redfield ratios stem from the fact that this ratio is an average and thus disguises the variability in stoichiometry between components of the system which can create serious discrepancies in modelled dynamics (Flynn 2010).

**What about non-elemental stoichiometry?**

**A:** You can include whatever chemical components that you wish. The most common non-element component described in models is Chl, but toxins and allelopaths are other obvious candidates, as is the nucleic acid package transferred for acquired phototrophy in some plastidic specialist non-constitutive mixoplankton, such as *Dinophysis*. Typically such components can be assumed to have fixed stoichiometry; for example paralytic shell-fish toxins can be assumed to have a fixed C:N and thus can be referenced as either PST-C or PST-N. An observation from Task 1 is the importance placed on lipids; a case can clearly be made for some organisms for including a state variable for 'lipid'. In some instances the importance of lipid may be handled implicitly, by including organisms that contain those lipids as essential prey for consumers, thus negating the need for a state variable for these chemicals. In other instances, you may need an explicit inclusion of 'lipid' as a SV. Perhaps the most common non-elemental stoichiometry is for organism size; if you assume constant size then you do not need a separate SV for numeric abundance.

**Do we have to describe the food web from virus to krill if we are really only interested on one or other end of the web?**

**A:** You have to start and stop somewhere. Traditionally plankton models stopped at 'zooplankton', usually emphasising 'phytoplankton'; fisheries models started at 'zooplankton', invariably assuming a grossly simplified description of 'phytoplankton'. Zooplankton have been poorly described by both groups of modellers (Mitra et al. 2014). The challenge with digital twins is that the models are deliberately designed for others (not the creators of the model) to test 'what-if?' scenarios. Likely that requires a 'health-warning' alerting the user to caveats; there is nothing wrong with that at all. You need to make a simulation configuration that is fit for the task at hand, so it is up to you to justify in your own mind, and importantly also to others, where to start and stop. And of course, where to start and stop will depend on how constrained you are by the state variable count.

**How do I describe the activity of higher trophic levels, above plankton?**

**A:** The activity of higher trophic levels are described in models using a 'closure function'. These are abundance-related equations that de facto prevent the biomass of the upper-level organisms from exceeding a certain value. As an extreme example, you could have a phytoplankton model, with no explicit zooplankton description, by just applying a closure term on the phytoplankton (Mitra 2009). Fisheries models typically work the other way around, with assumptions made on the availability of suitable phytoplankton etc. paying scant attention to whether that feed would be of suitable size and chemical composition. Closure terms do not require state variables; they just remove a fraction of the organisms to waste (i.e., with flows to state variables for 'POM' and 'DOM', with their associated chemical stoichiometries).

**How do I handle ontogeny?**

**A:** With reference to the time frame of your simulation (can you get away with assuming no ontogeny in a stimulation period of a few days, for example?), decide which stages can be plausibly merged, considering whether those can be described assuming fixed stoichiometry, and the duration of any developmental stages. These matters, together with commonality in feeding, may likely also affect which organisms you can group together.

**How do I handle DOM and regeneration?**

**A:** Decide how far down the lability scale you need to go. The more recalcitrant forms of DOM are likely only of importance in long-term simulations (ca. >6mo). More typically, you only need to consider labile (e.g. sugars, amino acids) and semi-labile (which will include cellular components requiring extracellular digestion, such as proteins and lipids). Regeneration (release of inorganics) is affected by both the activity of bacteria and also of their protistan predators (protistan zooplankton and mixoplankton), notably being promoted by differences in the stoichiometry between those organisms.