Ecological Modelling 498 (2024) 110870

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

# When to add a new process to a model – and when not: A marine biogeochemical perspective

Adrian P. Martin [a,*], Angela Bahamondes Dominguez [b,c], Chelsey A. Baker [a], Chloé M.J. Baumas [d], Kelsey M. Bisson [e], Emma Cavan [f], Mara Freilich [g], Eric Galbraith [h], Martí Galí [i], Stephanie Henson [a], Karin F. Kvale [j], Carsten Lemmen [k], Jessica Y. Luo [l], Helena McMonagle [m], Francisco de Melo Viríssimo [a,n], Klas Ove Möller [o], Camille Richon [p], Iyyappan Suresh [q,r], Jamie D. Wilson [s], Matthew S. Woodstock [t], Andrew Yool [a]

[a] National Oceanography Centre, European Way, Southampton, SO14 3ZH, UK
[b] NIWA, 301 Evans Bay Parade, Hataitai, Wellington 6021, New Zealand
[c] Victoria University of Wellington, Kelburn, Wellington 6012, New Zealand
[d] Earth System Science, Stanford University, USA
[e] Ocean Biology and Biogeochemistry Program, NASA Headquarters, Washington D.C, USA
[f] Silwood Park Campus, Department of Life Sciences, Imperial College London, SL5 7PY, UK
[g] Brown University, 69 Brown St, Providence, RI 02912, USA
[h] Earth and Planetary Science, McGill University, Montreal H3A 0E8, Canada
[i] Institut de Ciències del Mar (ICM-CSIC), Barcelona, Catalonia, Spain
[j] GNS Science 1 Fairway Dr. Lower Hutt 5010, New Zealand
[k] Institute of Coastal Systems - Analysis and Modeling, Helmholtz-Zentrum Hereon, 21502 Geesthacht, Germany
[l] NOAA Geophysical Fluid Dynamics Laboratory, Princeton NJ 08540, USA
[m] School of Aquatic and Fishery Sciences, University of Washington, Seattle 98105, USA
[n] Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, London, WC2A 2AE, UK
[o] Institute of Carbon Cycles, Helmholtz-Zentrum Hereon, 21502 Geesthacht, Germany
[p] Sorbonne Université, CNRS, IRD, MNHN, Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques, Institut Pierre Simon Laplace (LOCEAN-IPSL), 75005 Paris, France
[q] Digital University Kerala, Trivandrum, Kerala, India
[r] CSIR National Institute of Oceanography, Goa, India
[s] Department of Earth, Ocean and Ecological Sciences, University of Liverpool. Liverpool, L69 3GP, UK
[t] Woods Hole Oceanographic Institution, 266 Woods Hole Road, Woods Hole, MA, 02543, USA

A R T I C L E I N F O

*Keywords:*
Model development
Parameterisation
Subjectivity
Decision-making
Experimental design
Ocean biogeochemistry

A B S T R A C T

Models are critical tools for environmental science. They allow us to examine the limits of what we think we know and to project that knowledge into situations for which we have little or no data. They are by definition simplifications of reality. There are therefore inevitably times when it is necessary to consider adding a new process to a model that was previously omitted. Doing so may have consequences. It can increase model complexity, affect the time a model takes to run, impact the match between the model output and observations, and complicate comparison to previous studies using the model. How a decision is made on whether to add a process is no more objective than how a scientist might design a laboratory experiment. To illustrate this, we report on an event where a broad and diverse group of marine biogeochemists were invited to construct flowcharts to support making the decision of when to include a new process in a model. The flowcharts are used to illustrate both the complexity of factors that modellers must consider prior to making a decision on model development and the diversity of perspectives on how that decision should be reached. The purpose of this paper is not to provide a definitive protocol for making that decision. Instead, we argue that it is important to

* Corresponding author.
*E-mail addresses:* adrian.martin@noc.ac.uk (A.P. Martin), chelsey.baker@noc.ac.uk (C.A. Baker), kelsey.bisson@nasa.gov (K.M. Bisson), mara_freilich@brown.edu (M. Freilich), eric.galbraith@mcgill.ca (E. Galbraith), mgali@icm.csic.es (M. Galí), s.henson@noc.ac.uk (S. Henson), k.kvale@gns.cri.nz (K.F. Kvale), carsten.lemmen@hereon.de (C. Lemmen), jessica.luo@noaa.gov (J.Y. Luo), f.de-melo-virissimo@lse.ac.uk (F.M. Viríssimo), klas.moeller@hereon.de (K.O. Möller), camille.richon@univ-brest.fr (C. Richon), i.suresh@duk.ac.in (I. Suresh), matthew.woodstock@whoi.edu (M.S. Woodstock).

https://doi.org/10.1016/j.ecolmodel.2024.110870
Received 21 May 2024; Received in revised form 12 August 2024; Accepted 8 September 2024
Available online 24 September 2024

acknowledge that there is no objectively “best” approach and instead we discuss the flowcharts created as a means of encouraging modellers to think through why and how they are doing something. This may also hopefully guide observational scientists to understand why it may not always be appropriate to include a process they are studying in a model.

## 1. Introduction

Computer models are routinely used to simulate aspects of the natural world, from the sub-cellular (e.g. de Jong et al., 2017) to the planetary scale (e.g. Hattab et al. 2014; Fennel et al., 2022). They span a huge range in complexity and scope but also in application; from theoretical understanding to operational forecasting. They provide a means of capturing our current understanding of an environmental system in a mathematical description that allows us to illuminate fundamental dynamics, to frame and test hypotheses, to guide data collection and to extrapolate to other times and situations (Epstein, JASSS 2008).

Like the natural world they seek to reflect, models are highly diverse. The potential breadth of this diversity is illustrated by Janssen et al. (2015) who analysed the diversity of aquatic ecosystem models, further identifying characteristics by which they can be categorised such as the modelling approach, the topic represented, and implementation details. Jakeman et al. (2024) provide a broader perspective, analysing models for a diversity of domains. Here we focus on just two fundamental characteristics that can be used to group models. The first is scale. Each model is designed to represent a limited range of space and time scales that reflect their purpose; from a test tube to the global ocean, from cell division times to glacial cycles. The second is complexity, the level of detail at which a model represents nature, reflected by the number of simulated processes. Although both characteristics require a decision by the modeller that introduces subjectivity, for the second this decision is, perhaps unsurprisingly, itself more complex. Whether the intention is to model the uptake of nutrients by individual microscopic cells (e.g. Bonachela et al., 2011, Orth et al. 2010) or material flow through the full Earth system with interacting oceans, atmosphere, land and ice (e.g. Sarmiento and Toggweiler, 1984; Ziehn, T., et al. 2017), any model will only reproduce a subset of the true processes involved. The complexity of a model is intrinsically linked to the three fundamental controls on model performance: the external data which are required (e.g. for initial conditions or forcing); the structure of the model (i.e. which processes are represented and how, including spatial resolution); and the extent to which parameters for the model can be constrained (Claussen et al., 2002).

Ocean biogeochemistry is one example of a field where many modelling approaches have been used (Janssen et al. 2015; Fennel et al., 2022), from simple single-use models to ones that are continuously developed over long periods. The Earth system models that contribute climate projections to Assessment Reports (ARs) of the Intergovernmental Panel on Climate Change (IPCC 2023) are a good example of the latter. These require significant person-years of time, funding and computing resources to build, maintain, and run. While they are known to omit processes of interest (Bopp et al., 2022; Rohr et al., 2023) and to have biases in their performance (Tagliabue et al., 2021), such models represent an established base upon which further - and traceable - development can occur, and which would be very costly and time-consuming to construct from scratch.

Many models evolve with time, generally becoming more complex. First, they may be modified to re-assess the question they were originally built for by adding a new process (e.g. enhancing an existing fisheries stock model, Punt et al. 2020). For the purpose of this article, we regard adding a new variable to a model as equivalent to adding a new process since it will be necessary to add functions linking it to the existing model. Second, there may be situations where the model provides a useful starting point to address questions related to but differing from the original one. For example, an ocean biogeochemical model may have originally been designed to investigate global carbon cycling (Schmittner et al., 2008) but subsequently modified to allow study of the dynamics of low oxygen zones in the ocean (Niemeyer et al., 2017).

There are therefore three key development stages at which decisions are made affecting the complexity of a model: (1) when the model is originally designed and built; (2) when revising a model that is still being used for the original question; (3) when the model is being adapted to address a new question. In all cases, such a decision often concerns whether to include a specific process. For the purposes of this manuscript, “include” also refers to situations where an existing representation of a process is changed. It is often the case that we have a limited understanding of many biogeochemical processes, which means that we rely on ad hoc parameterisations until further data or theoretical insight allows them to be improved.

There are several potential drivers for the inclusion of a new process. For example, more information may have become available. This could be new data that provide greater spatial or seasonal coverage or offer the opportunity to refine the representation of some processes by providing a more stringent test of how the model reproduces reality. More fundamentally, it might be that our theoretical understanding of key processes at the heart of a model has changed and aspects of the original model are now known to be inaccurate or insufficient. Increases in computing power or advancement in software development may also enable inclusion of a process that was previously impractical. A good example is the gradual increase in computing power allowing an increase in spatial resolution of ocean physics (Chassignet and Xu, 2021) or the number and complexity of processes represented in biogeochemical models (Wu et al. 2022).

These drivers toward greater complexity need to be balanced against a variety of potential constraints. First, mechanistic and observational uncertainty regarding the environmental system (arising from initial conditions, forcing, parameter values, and the inevitably simplified representation of the natural system) often inhibits our ability to create robust model estimates if poorly known processes are incorporated. Put another way, would we require significant new data and understanding to support the inclusion of a new process? Second, model development can be expensive both with respect to labour and computational resources. Third, in the case of state-of-the-art Earth system models, the decision for model improvements will also need to be consistent with the longer-term, broader strategy of the team developing the model and international modelling frameworks (Seferian et al., 2020).

In practice, making this decision is not straightforward, or objective. There have been attempts to codify an optimal approach to how this decision should be made (e.g. Kyker-Snowman et al., 2021; Moore, 2022) but all such approaches are inevitably subjective to some degree. This subjectivity is not just the consequence of the experience and biases of the modeller. For example, the decision is influenced by the broader philosophical approach of the particular model user/developer. This is most clearly illustrated by the debate around parsimony. How can one compare the relative benefits of a simple model, where cause and effect can be clearly inferred, with a more complex one which represents more of the structure and diversity of the natural system but within which causality can be hard to understand? A good example of contrasting approaches on this topic is provided by a previous discussion on the level of detail with which phytoplankton communities should be represented in ocean biogeochemical models (Anderson, 2005; Le Quéré, 2006; Flynn, 2006; Friedrichs et al., 2007). There is also subjectivity in assessing the value of including a new process. Although one may try to quantify objectively whether or not adding a process improves model performance against a selection of observationally-based metrics, there is often a subjective choice in those metrics and different performance metrics may give a different answer (Kriest et al., 2010, 2020). From a related context of population modelling, a similar discussion can be found in Levins (1966) – who emphasised the tradeoffs between model generality, realism and accuracy –and subsequent papers (e.g. Orzack and Sober, 1993).

The focus of this manuscript is on encouraging researchers to think more deeply about how they are individually deciding whether to include a new process. It does not set out to tell people how to do this. We report on an event (see Method) in which researchers were invited to build a flowchart to answer the question: "How to decide when to include a new process in a model?". Input was sought from a range of researchers; not just model developers and users, but also people who analyse and use model output. They were not required to have a particular level of modelling experience, or indeed any. They were asked to carry out this task of building a flowchart jointly while discussing and defending their additions with other researchers. The intent was both to capture varying perspectives and to encourage discussion of priorities. In short, the event was aimed at fostering debate, self-analysis and reflection on this important topic.

It should be noted that, although this paper focuses on the situation where an existing model is modified, the same factors and trade-offs are involved when building a model anew. While a distinction with the latter is that the same considerations apply simultaneously to all processes that may or may not form part of the model, this multiplies the task but doesn't change the nature of the underlying decisions. Hence what follows can equally be applied to the task of building a new model too. It should not be thought that the issues discussed here only apply to large complex models. Historically even some of our simplest models have attracted debate (Hall 1988a, 1988b).

Although we do provide a single flowchart which has been synthesised from the input to the event, it is not presented as an objective and authoritative tool. The use of this flowchart by two different researchers may still result in two different decisions due to the different perspectives they will have; for example on how much new data is "enough". It is instead offered as an illustration and starting point for a modeller to use their own judgement on how to tackle this issue.

Even if the sole consequence of reading this manuscript is that a reader subsequently thinks more deeply about their approach to making this decision and why it is the appropriate one, then this paper has achieved what it set out to. Although this paper represents the thoughts of researchers involved in modelling ocean biogeochemistry, the issues it addresses are fundamental to all models.

## 2. Methods

"How to decide when to include a new process in a model" was an event run by the UN Ocean Decade Programme JETZON (Joint Exploration of the Twilight Zone Ocean Network – https://jetzon.org). This took place from 16 to 17 September 2021, as a virtual "satellite event" contribution to A Predicted Ocean Laboratory which was run by the Federal Ministry of Education and Research, Germany as a contribution to the UN Ocean Decade for Sustainable Development.

A Google document was set up to allow contributors to draft a flowchart reflecting their perspective on how the decision of whether to include a new process in a model should be reached. This document was originally blank to avoid influencing the outcome and a new blank sheet was available for every 6 h period across the 24 h for which the event ran, to make sure no contributor felt constrained in what they could add. Contributors were free to add additional sheets too – which they did. No contributor provided input for more than one sheet. Contributors could also discuss aspects with each other either within the Google document or in a Wonder.me room set up for the purpose. A spreadsheet was also set up to allow contributors to list key questions in the decision process and to vote on whether they thought those listed were important. In practice both the Wonder.me room and spreadsheet remained unused. A free text document was found to be more successful in gathering opinions.

This event was advertised widely through the UN Ocean Decade, the Predicted Ocean Laboratory team and the network of informal contacts of researchers across a variety of social media platforms. It was entirely open, allowing input from anyone globally with an interest and access to these platforms. To achieve this, contributions were received via Google documents which had global write access for the duration of the event. This allowed contributions to be anonymous if preferred to remove any bias that might arise from perceptions of how a comment may be received. The fully open approach additionally meant that time-zone was no barrier to contributing, as the event ran, and documents were open, for 24 h. Additionally, 7 people with considerable modelling experience were online at regular intervals of 2–6 h during the day, to stimulate the discussion if necessary. They were also free to contribute themselves.

Over the course of 24 h, multiple people contributed to the activity. It is not possible to say how many because input was allowed to be anonymous if required. Those who identified themselves ranged from PhD students to senior scientists, with experience in modelling ranging from four years to 24 years and from at least 8 different countries including Australia, France, Germany, India, New Zealand, Spain, UK and USA. The contributors also represented a wide span in the types of models used from individual cell models to global coupled climate models. At least one person had no modelling experience. At least 11 had experience of observational approaches to science. The following section summarises the results. The full original responses can be found in the Appendix.

## 3. Results

Five flowcharts were produced during the event. They are reproduced as originally created during the event in the Appendix. Across the 5 flowcharts, there was considerable variability in responses and the contributions during the event captured a wide range of issues influencing decisions and choreographed them in different ways, with some issues present in some flowcharts but not others and varying in their position in the decision sequence within the flowcharts they do appear in.

Specific questions highlighted by contributors as part of their flowcharts are summarised in Table 1. Note that they are simply reported as contributed but wording may have changed slightly from the original if the same question was raised multiple times and a description capturing all instances was required. In total, 23 unique questions were posed in creating the 5 flowcharts. Table 1 also allocates the questions to one of four categories. Although it could be argued that several questions may be relevant to multiple categories this was done as a crude examination of whether a question was related to: (i) Understanding - the need for theoretical knowledge of the process; (ii) Data - the need for observational, including laboratory-sourced, information on the process or its impact; (iii) Implementation - the technical challenge of incorporating a representation of the process in a model; (iv) Performance - the key test of how the model performs with the process included. The distribution across these four categories is also represented as the relative frequency of occurrence in Fig. 1. The Understanding and Implementation categories were equally most frequent, corresponding to 35 % of questions each. Performance was represented a little less frequently at 22 % of issues and only 8 % of issues related to Data. It should be noted that the number of issues in each category does not in any way correlate to the relative importance of the categories when adding a process to a model.

The structure of the five flowcharts (see Appendix) varies both in the number of questions they include as part of a decision but also in their degree of connectivity and in particular how they lead the user of the flowchart to a decision. Table 2 shows summary information that captures some of this variability. First, it shows the number of questions

**Table 1**
Questions affecting a decision to add a process to a model contributed during the event.

| Category | Question |
|---|---|
| Understanding | Is there evidence that there is a flux or behaviour missing from the model that is essential to address your question? |
| Understanding | Is there evidence that the new process is quantitatively significant compared to current model processes that are known to be important? |
| Understanding | Are functional relationships with existing model constituents known / hypothesised? |
| Understanding | Is there an analogue or proxy from a different environment? |
| Understanding | Is it possible to conduct an idealised sensitivity study first, and see how big the effect is? |
| Understanding | Is there reason to think the process has a non-negligible effect on the system? |
| Understanding | Is there reason to think the process is relevant to your research question? |
| Understanding | Is there a strong theoretical understanding of the new process? |
| Data | Are data sufficiently (in space and time depending on the model resolutions) available to develop a parameterisation of the new process ? |
| Data | Do you have or can you get data to validate the process once you put it in? |
| Implementation | Is it relatively easy to add from a coding perspective? |
| Implementation | Are you prepared and able to re-tune the model after adding in the new process? |
| Implementation | Do other compensating or related processes need to be added/ adjusted? |
| Implementation | Can you reduce the scope / domain of your study? |
| Implementation | Does it simplify the representation of other processes (e.g. fewer and better constrained parameters)? |
| Implementation | Can you develop the parameterisation of the new process in a reasonable amount of time? |
| Implementation | Could the intended impact of the new process be captured equally well by existing model equations, within their parameter uncertainty? |
| Implementation | Is your model complexity level suitable for addition of the new process? (e.g. is it too simple, track the wrong element?) |
| Performance | Does the new process make a significant change in model behaviour? |
| Performance | Does the new process improve or degrade model skill /behaviour? |
| Performance | Is the current model right for the wrong reasons? |
| Performance | Does the new process substantially change model results for potential future applications? |
| Performance | Does the new process increase computing resources needed (time or processors)? |

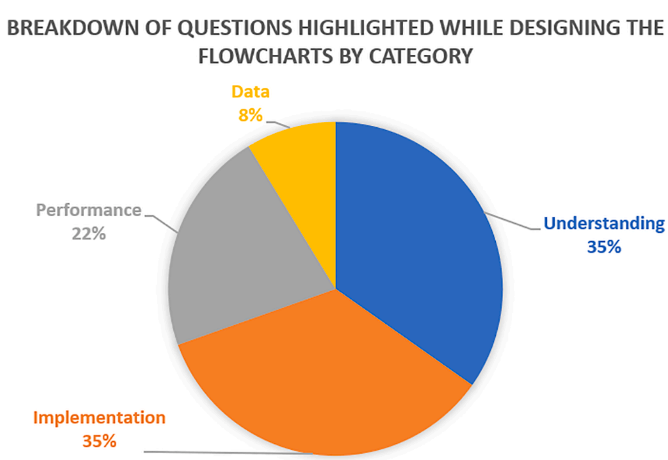


**Fig. 1.** breakdown of questions identified during the event into four categories.

included in each flowchart. Although it is a little difficult to see at first glance (as some flowcharts have comments and advice as well as questions) the number of questions included varies between 4 and 16, but the flowchart with 16 is the outlier with the others having 4 (in three cases)

**Table 2**
flowchart structure and connectivity.

| Flowchart Time period | Number of questions | Number of routes | Number of ways to 'yes' | Fraction leading to 'yes' |
|---|---|---|---|---|
| 0001–0600 CEST | 4 | 10 | 1 | 0.10 |
| 0601–0900 CEST | 6 | 15 | 6 | 0.40 |
| 0901–1200 CEST | 4 | 5 | 4 | 0.80 |
| 1201–1800 CEST | 16 | 58 | 20 | 0.34 |
| 1801–2400 CEST | 4 | 7 | 2 | 0.29 |
| Merged ( Fig. 2) | 12 | 13 | 2 | 0.15 |

and 6. The number of possible routes through the flowchart and the number of those that lead to a 'yes, include the process' is also shown, together with the fraction that lead to 'yes'. (The numbers of ways and fraction leading to 'no' can simply be calculated by subtracting the 'yes' value from total number of routes or 1, respectively.) There is considerable variability. Notably, the flowchart with 16 questions has a greater fraction of routes leading to 'yes' than two of the three flowcharts with just 4 questions. It is also the case that taking just one flowchart and restructuring it to put the same questions in a different order could lead to a different number of ways to 'yes'. It should not therefore be thought that increasing the number of questions automatically decreases the likelihood of getting a 'yes'. This is important to be aware of as it may subconsciously influence the construction of a flowchart if the reverse was true.

Finally, Fig. 2 shows an attempt at combining the 5 flowcharts (and the thinking behind each flowchart) in the same format to give a clearer overview of the issues that were raised and how they interlink. This flowchart was created with the assumption that someone is trying to make a decision at a specific point in time. Hence, although it indicates some of the approaches that could be taken if a 'no' is reached, it does not attempt to represent the typical iterative nature of model development. This is deliberate as the latter then raises the additional question of how long one should keep testing/trying - which probably warrants a paper in itself. This editorial approach is why the merged flowchart has fewer questions and routes (Table 2) than some of the contributed flowcharts. It is also assumed that no one is embarking on modifying their model until they have funding/time to allow them to do so.

## 4. Discussion

Although the flowchart-generation event involved only people from a marine biogeochemistry background, the discussion and issues should be familiar to any modeller. The intrinsic tightly-connected complexity of the Earth system and its components means that there will rarely be a situation where the decision on whether to include a model is independent of other factors.

The focus here has been on when to include a new process in an existing model. A 'process' in this context could take many forms: it could be modifying an existing functional form already present in a model; it could be adding a new compartment, variable or sub-model; it could be adding a new link between existing model variables. However, a similar approach could be more widely applicable to other situations including: what to include when building a new model, how to merge two models (Nisbet et al., 2012) and what to remove if it is necessary to simplify an existing one (Ward et al., 2013; Galbraith et al., 2015).

The combined flowchart presented in Fig. 2 is intended to help modellers engage in structured reflection, rather than providing a definitive guide to the decision process. There is no perfect protocol for objectively deciding whether to include a process in a model. We would

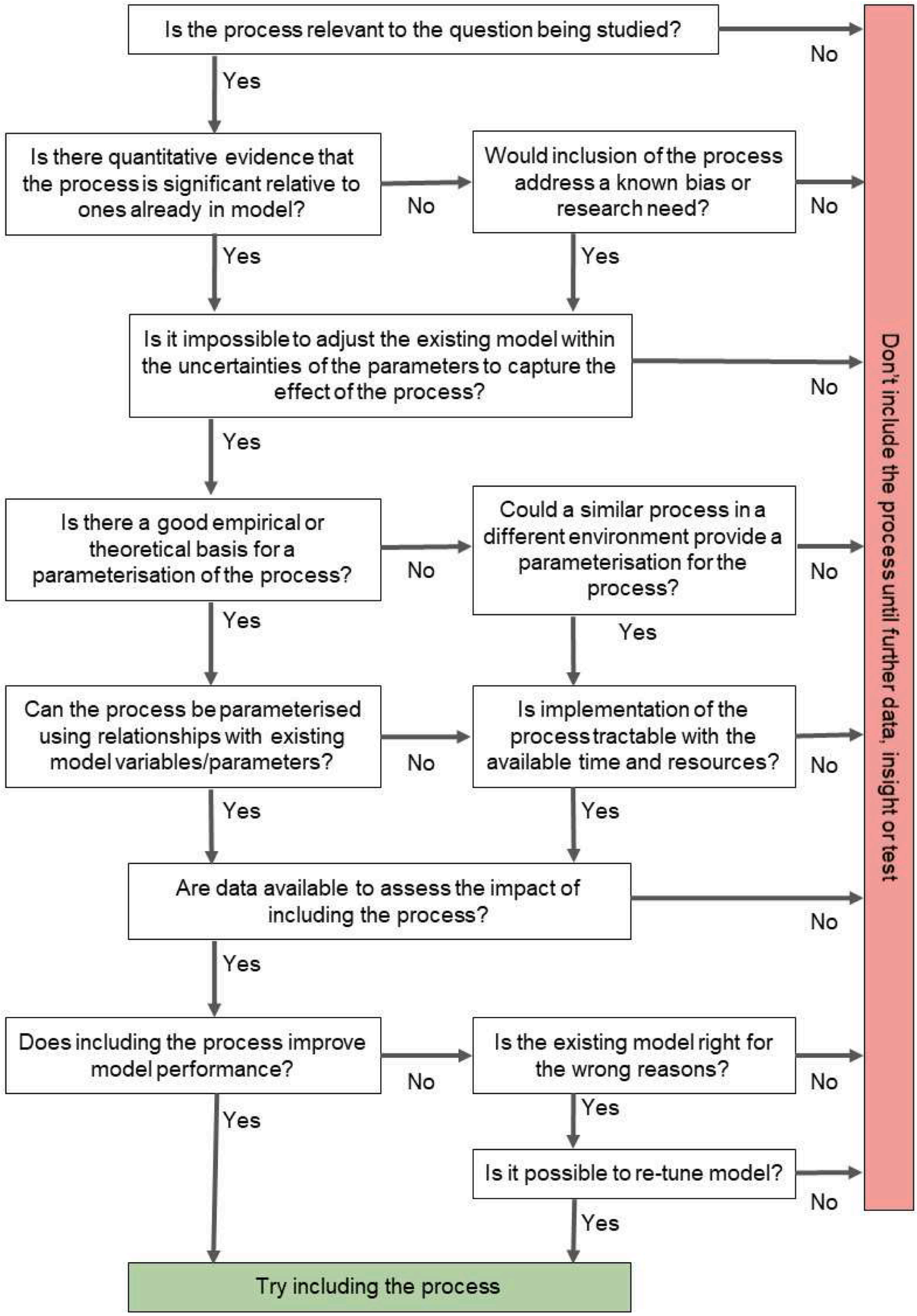


**Fig. 2.** Illustrative flowchart based on a merger of the 5 flowcharts created during the event.

suggest that it may be better to create a flowchart oneself, rather than rely on a generic one, as the process of analysis and reflection is valuable in making the ultimate decision and any flowchart is bound to be missing elements. For example, only one of the contributed flowcharts explicitly asked whether the new process was central to the science question(s) the modeller sought to answer. It is unlikely that the omission was because it was thought to be unimportant; rather that it was implicitly assumed. To identify and address any potential biases we might have we need to be explicit about everything we are assuming. Regardless of one's best intentions and adherence to the scientific method, every scientist will bring their own perspectives and biases to any such decision, whether consciously or subconsciously. Even if we did, unwisely, tout our merged flowchart as the definitive one to use, different users would interpret it in different ways, making different judgements on important aspects (e.g. what is the metric and related threshold for "improved model performance"? how is an empirical or a theoretical basis assessed as "good"? How many datapoints are enough?) and reaching different decisions for the same process and model. A different choice of base model may also lead to different outcomes.

Although it did not appear in any of the flowcharts, an additional important question is whether the model to which you wish to add a new process is best suited to the task. There may be a tendency to address new science questions by adapting existing models, but in some cases it may be better to start afresh with a new model.

A modeller needs to be clear on what type of model is suited to their task: ecological or biogeochemical approach? relatively coarse spatial resolution global model or regional high-resolution model? whether the model should be spatial at all? Not all models will be well-suited to incorporating a particular process, even if that new process meets all the other criteria. The programming language and the modelling framework in which a model is written may be additional factors in this as they can have a practical bearing on the ease with which a new process can be incorporated, balancing factors such as level of detail, speed of running, and simplicity of diagnosis. Simple models will be well-suited for some situations and more complex ones for others. Also, despite positive moves in this direction, it is still true that not all models are easily accessible and usable by others (Preisig et al., 2006; Patel et al. 2020); many suffer from lack of high-performance portability or insufficient software sustainability (Crouch et al. 2013; Siewertsen et al. 2013). At the more 'heavy-duty' end of the model scale, it is not usually feasible to add significant extra complexity to a coupled climate model unless there is strong evidence that a new process is very important to climate feedback. Furthermore, with higher spatial resolution models, the strength of the argument has to increase if supercomputing time is limited. It is also possible that when a new process is included because it is central to a particular research question this might slightly degrade the model performance regarding other parameters/processes. The extent to which this influences a decision might vary according to the longer-term and/or wider use of the model. Additionally, not all aspects of model choice are science based. It might partly be a matter of strategy (e.g. if a lot of resources and time have already gone into developing a global model). It might also be affected by resources e.g. access to a supercomputer, or by the experience of the modeller. It might even be irrational, as highlighted by the "sunk cost effect" (Olivola, 2018).

It should also be noted that Fig. 2 was based on a single event with a relatively limited pool of contributors and it is possible that some questions that other people might view as essential have not been identified and included. This should not be surprising. The flowchart event was not carried out in a way calculated to be exhaustive. Even with 5 groups working on separate flowcharts there are still likely to be questions that were not raised but which might have been with different participants. In particular, in situations with more specific objectives there may be additional questions and/or constraints. One example of this is the move towards 'seamless forecasting' in operational settings (e.g. Dirmeyer and Ford, 2020) where the model under consideration for a new process may need to be combined with other ones covering different scales to provide a unified forecasting system, so broader compatability may be necessary. Another example is in a policy setting, such as Environmental Impact Assessments, where arguments have been made for standardised suites of models (e.g. Forbes, 2024). There are also subtler facets of a question that may be sufficiently important to be separate questions in some cases. Take the case of model performance. Given the ongoing debate about the relationship between stability and complexity (Denman 2003; Hannah et al., 2010; Allesina and Tang, 2012) it may be felt that a question on the stability consequences of adding a process is needed. Similarly, it may be felt that a question about whether the original model may be over-fitted, so biasing the behaviour when a new process is added, might be useful.

It is also worth clarifying that the combined flowchart in Fig. 2 is very specific in the sense that it represents a set of issues to consider in a situation where all the information is available to make a final decision. As several of the flowcharts created during the event highlight, making a decision on whether to include a process is often likely to involve iteration, with feedback, reflection and further tests or data being sought after a 'no', rather than abandoning the idea of including the process. A valuable attribute of a flowchart may be the light it sheds on the user's understanding of the process and the model in the course of making the decision. Several flowcharts took a broader view of how a decision should be made, incorporating feedback loops from the observational community as central to the model development process. These potentially lead to the modeller revisiting their thought process and making adjustments, which is crucial to the scientific process but also time-consuming. It is also far from trivial. Modellers often struggle to gather data they feel is necessary for building, testing and running their models. This is for numerous reasons. First, observationalists may be less experienced in identifying which data are most lacking and/or valuable for models. Second, while they may be interested in the perspective offered by models, observationalists may have different priorities for the type of data they collect. Third, and more fundamentally, there may be significant issues in reconciling what can be observed and what is required for models (Strzepek et al. 2022). More informed discussions between both communities to co-design dedicated experiments/fieldwork and models would be a very valuable outcome of using a flowchart, particularly in research areas with large discrepancies between models and observations (Rohr et al., 2023).

This iterative approach is itself susceptible to external influences which will vary with the situation. In particular, the time needed to include a new process in a model is not all one needs to consider; even once coded, verified and run, there is a good chance that the new model version will need to be tuned again to account for the impact of the change on the existing model components. Constraints on project funding and/or time may also mean that seeking additional data is impossible. In that case, including processes that are highly uncertain could still be worthwhile if one of the goals of the model is to identify parameters that the model output is most sensitive to. This could be useful for prioritising future observational work and building off the model in the future.

Every one of the flowcharts created had 'no' as a possible outcome. Allocating effort wisely sometimes requires choosing not to do things, even if they are interesting and important in some contexts. The fraction of situations that would lead to such an outcome varied between the flowcharts (Table 2) but there are always situations where the user can be led to not including a process even though the process is known to occur. This does not mean that observationalists should be encouraged to stop studying the process or collecting related measurements. Processes (and observations of them) will have more than one context so even if a process might not significantly impact one aspect of the environment it might be important for another. For example, spatial variability in organic carbon fluxes may not have a particularly large impact on carbon storage, but this variability should still be measured because it is important for global patterns of productivity, e.g. as a food source for mesopelagic (Shea et al., 2023) and benthic organisms (Yool et al., 2017) and for dissolved tracers like oxygen (Davila et al., 2023).

Even in situations where a flowchart leads to a conclusion of 'no', it does not necessarily mean that nothing can be gained by incorporating the new process anyway. In some cases, adding a new process/feature to a model might be worthwhile in order to quantitatively demonstrate that it has no impact. In these cases, the model provides the mechanistic explanation for why (usually in a more global context) there is no impact, which is still a valid scientific finding (e.g. Gurgacz et al., 2023). Obviously it is not grounds for retaining the process in the model though.

The flowchart event successfully crowd-sourced opinions from a wide range of professional experience, research disciplines, and geographic locations. Although there was no prerequisite for participation, all groups were able to converge on situations where a modeller should and should not add a new process to a model. Additionally, the groups had a fair amount of convergence in ideas, suggesting there is some consensus among the population of opinions provided. Gathering ideas from a diverse collection of people allows for the inclusion of different backgrounds and experiences in the overall study results, which reduces survey bias. The most open approach is 'citizen science' where anyone can contribute (Kosmala et al., 2016). Using an open call for input from a wide range of researchers has successfully been used elsewhere to provide expert assessment (e.g. Henson et al., 2024). With the appropriate amount of organisation from project leaders, this survey structure could be used by societies and consortia to quickly (in ~24 h) obtain opinions from a global audience. During the flowchart event, documents were open for 24 h (so as not to be biased towards specific timezones) and allowed anonymous input. Nevertheless it is inevitable that biases remain. The people who took part are self-selected and may already be those who think about these issues. Restrictions on the use of the Google platform in some countries will also have had an effect. Even though a fresh document was available every few hours the presence of other contributions may influence what people suggested or did not suggest. For the reasons given above though this does not significantly influence the message of this manuscript. The approach was successful in gathering different perspectives and highlighting that there is no single way of tackling the question of when to include a new process in a biogeochemical model.

## 5. Conclusion

There is no objective way to decide when to include a new process in a model. However, by identifying the issues that need to be addressed to do so, the underlying assumptions and limitations become more apparent and the criteria for a decision are exposed to greater scrutiny. Rather than present a definitive approach, we instead highlight different perspectives and examine the influences that may give rise to them. We provide a combined flowchart for helping to guide a decision, though we do not recommend its uncritical use. Instead we encourage modellers to develop their own flowcharts - borrowing from the examples shown here if useful - as a means of gaining a deeper understanding of what they seek to achieve by adding a process, and to encourage any such decision to be iterative and draw in a broad range of expertise. In this way, we hope that modelling efforts can be more productively allocated to increase complexity where it is most useful, while avoiding the accumulation of model complexity that is not aligned with scientific questions, theoretical understanding, observational constraints, or computational capacity.

## Funding

This was a contribution in-kind to the UN Ocean Decade on behalf of the Joint Exploration of the Twilight Zone Ocean Network (JETZON) programme. It formed a satellite event for the Predicted Ocean Laboratory event run on 16 September 2021 as part of the United Nations Decade for Sustainable Development of the Ocean. APM acknowledges support from CLASS (NE/R015953/1) and ATLANTIS (NE/Y005589/1) NC projects, NERC CUSTARD (NE/P021247/2) and EU SUMMER (817806) grants. KFK acknowledges support from the New Zealand Ministry of Business, Innovation and Employment (MBIE) within the Antarctic Science Platform, grant ANTA1801. KFK also acknowledges support from the New Zealand Ministry of Business, Innovation and Employment through the Global Change through Time programme (Strategic Science Investment Fund, contract C05X1702). SAH was supported by a European Research Council Consolidator grant (GOCART, agreement number 724416), co-funding from the European Union, Horizon Europe Funding programme and UK Research and Innovation (OceanICU, agreement number 101083922). MG acknowledges support from the Spanish government through the OPERA project (PID2019–107952GA-I00) and a "Severo Ochoa" Centre of Excellence grant to ICM-CSIC (CEX2019–000928-S). FdMV acknowledges support from an European Research Council Consolidator grant (GOCART, agreement number 724416), a Large Grant from the UK Natural Environment Research Council (COMICS, agreement numbers NE/M020835/1 and NE/M020835/2), a UK Natural Environment Research Council RoSES programme grant (CUSTARD, agreement number NE/P021247/2), and a UK Natural Environment Research Council grant (ODESSS, agreement number NE/V011790/1). C.R. acknowledges support from the Institut des Sciences du Calcul et des Données (ISCD) of Sorbonne University (IDEX SUPER 11-IDEX-0004) through the support of the sponsored project-team FORMAL (From ObseRving to Modelling oceAn Life). H.M. was supported by the National Science Foundation's Graduate Research Fellowship Program. C.A.B. was supported by the Natural Environment Research Council (NERC) funding for the CLASS project through grant NE/R015953/1. MSW is supported by a Woods Hole Oceanographic Institution Postdoctoral Scholarship. K.O.M and C. L. were supported by the Changing Coasts programme of the Helmholtz Association of German Research Centers; C.L. received support through the MuSSeL grant 03F0862A by the German Ministry of Education and Research (BMBF). JYL acknowledges funding from the NOAA Marine Ecosystems Tipping Points initiative, and we thank Xiao Liu for her internal review of this manuscript.

## Competing interests statement

None of the authors have any competing interests.

## CRediT authorship contribution statement

**Adrian P. Martin:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Angela Bahamondes Dominguez:** Writing – review & editing, Writing – original draft, Formal analysis. **Chelsey A. Baker:** Writing – review & editing, Writing – original draft, Formal analysis. **Chloé M.J. Baumas:** Writing – review & editing, Writing – original draft, Formal analysis. **Kelsey M. Bisson:** Writing – review & editing, Writing – original draft, Formal analysis. **Emma Cavan:** Writing – review & editing, Writing – original draft, Formal analysis. **Mara Freilich:** Writing – review & editing, Writing – original draft, Formal analysis. **Eric Galbraith:** Writing – review & editing, Writing – original draft, Formal analysis. **Martí Galí:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Stephanie Henson:** Writing – review & editing, Writing – original draft, Formal analysis. **Karin F. Kvale:** Writing – review & editing, Writing – original draft, Formal analysis. **Carsten Lemmen:** Writing – review & editing, Writing – original draft, Formal analysis. **Jessica Y. Luo:** Writing – review & editing, Writing – original draft, Formal analysis. **Helena McMonagle:** Writing – review & editing, Writing – original draft, Formal analysis. **Francisco de Melo Viríssimo:** Writing – review & editing, Writing – original draft, Formal analysis. **Klas Ove Möller:** Writing – review & editing, Writing – original draft, Formal analysis. **Camille Richon:** Writing – review & editing, Writing – original draft, Formal analysis. **Iyyappan Suresh:** Writing – review & editing, Writing – original draft, Formal analysis. **Jamie D. Wilson:** Writing – review & editing, Writing – original draft, Formal analysis. **Matthew S. Woodstock:** Writing – review & editing, Writing – original draft, Formal analysis. **Andrew Yool:** Writing – review & editing, Writing – original draft, Formal analysis.

## 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.

## Data availability

All data discussed in the manuscript is presented in the manuscript

## Appendix. : Flowcharts produced by each workgroup

See Figs. A1-A5.

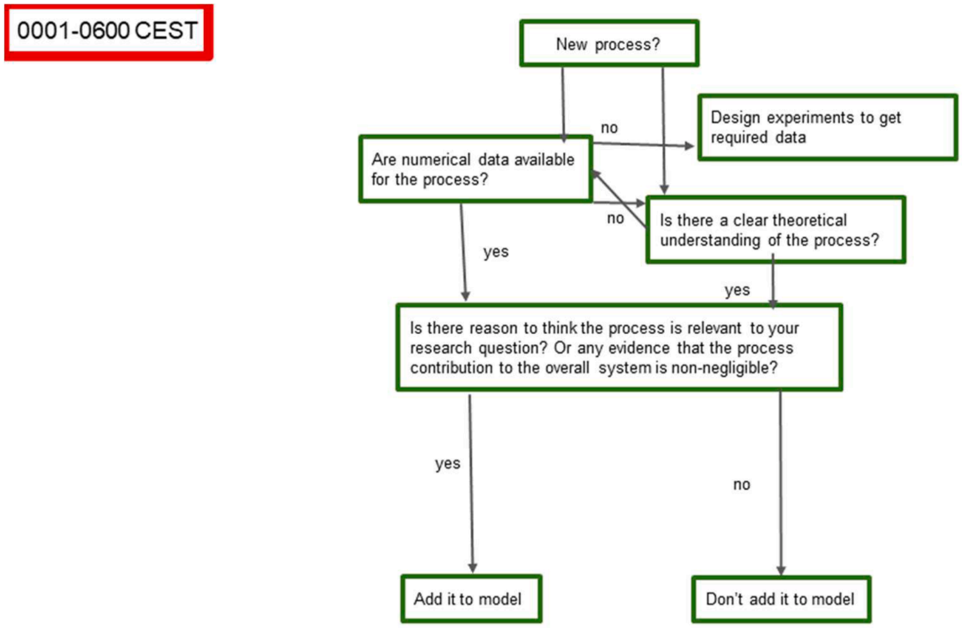


**Fig. A1.** Flowchart summarising the discussions during 00:01–06:00 CEST workgroup.

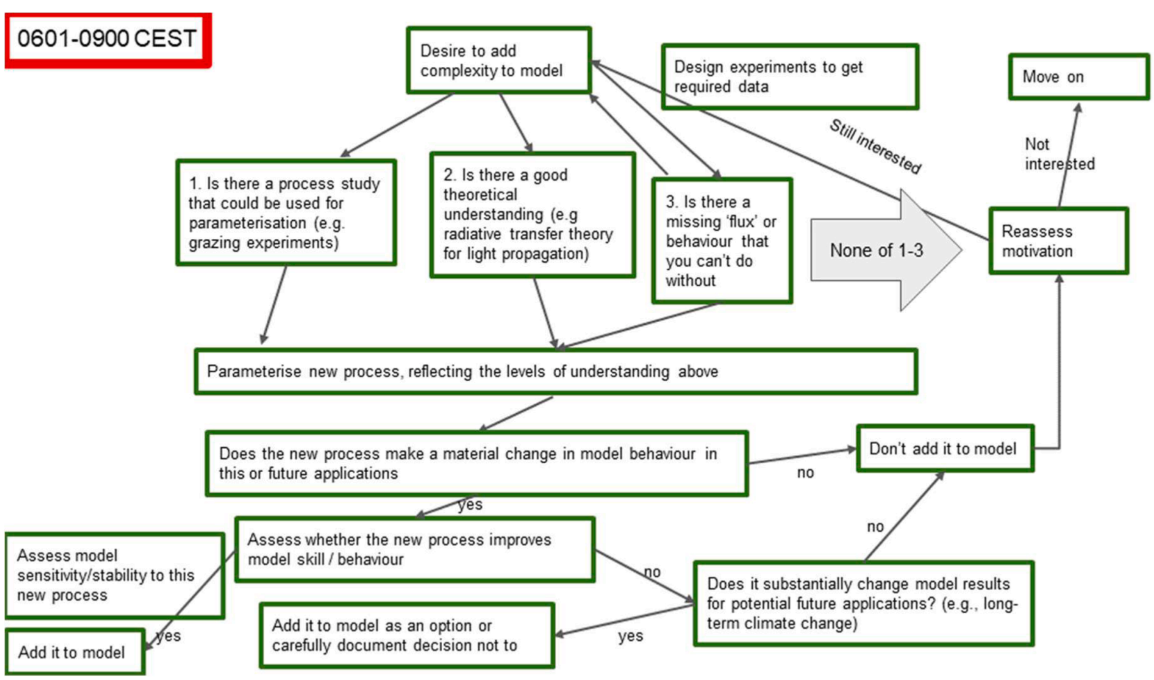


**Fig. A2.** As Fig. A1 but for 06:01–09:00 CEST workgroup.

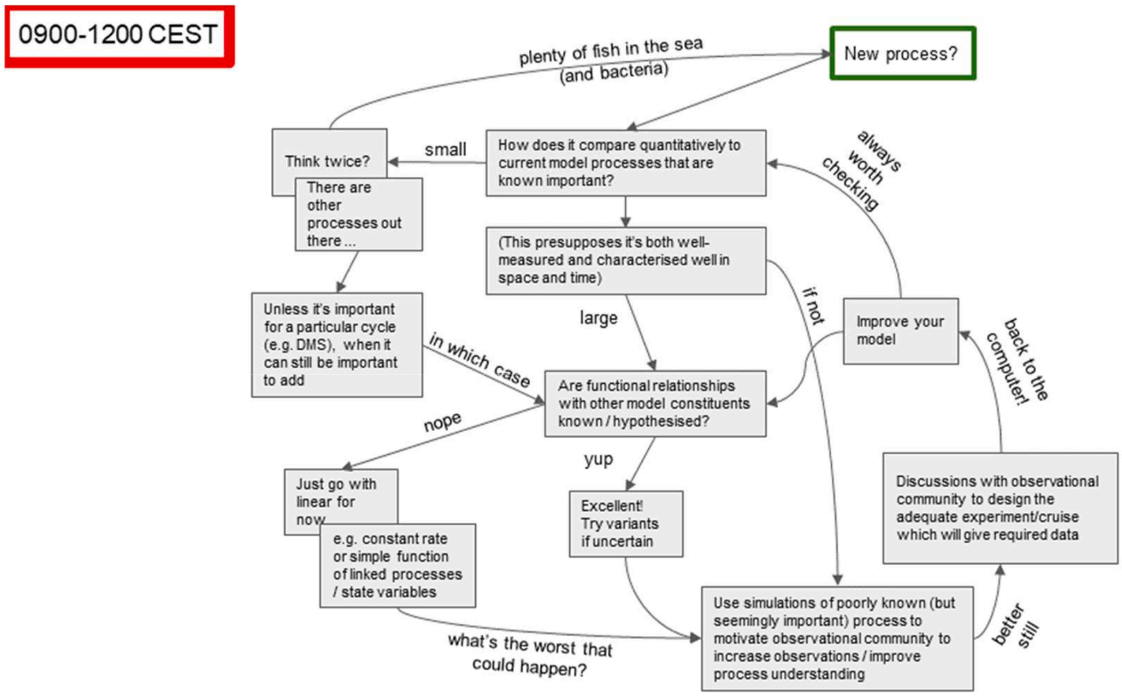


**Fig. A3.** As Fig. A1 but for 09:01–12:00 CEST workgroup.

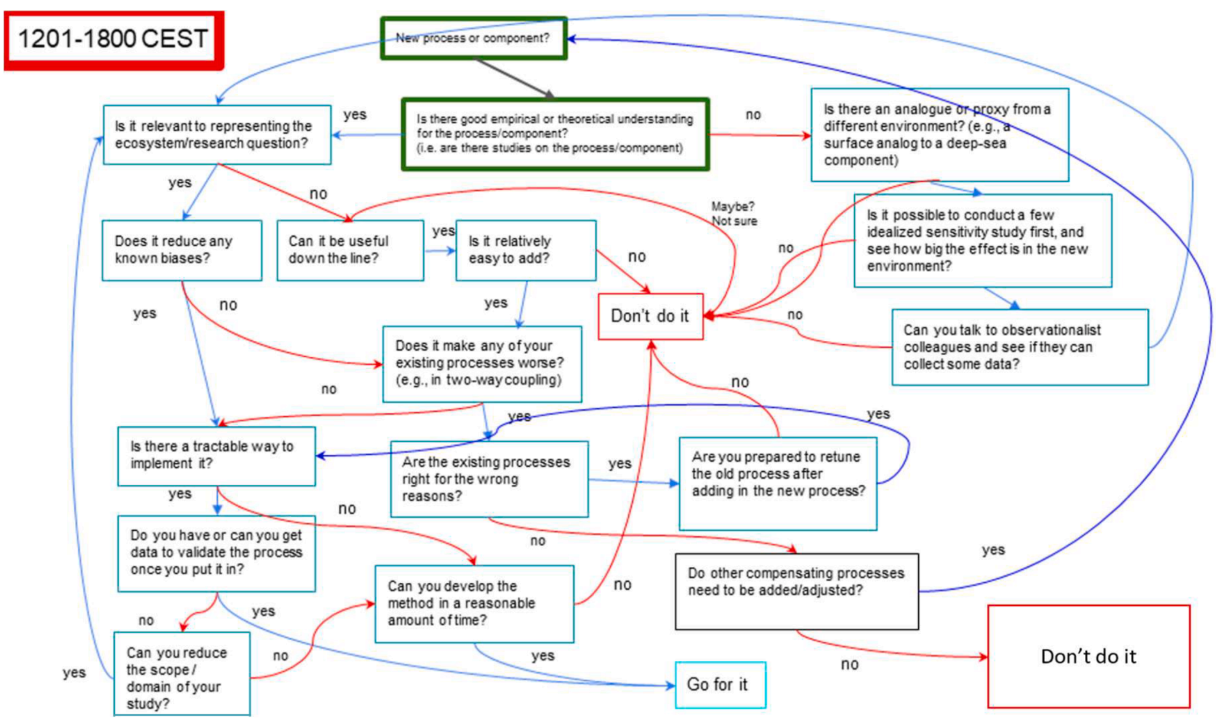


**Fig. A4.** As Fig. A1 but for 12:01–18:00 CEST workgroup.

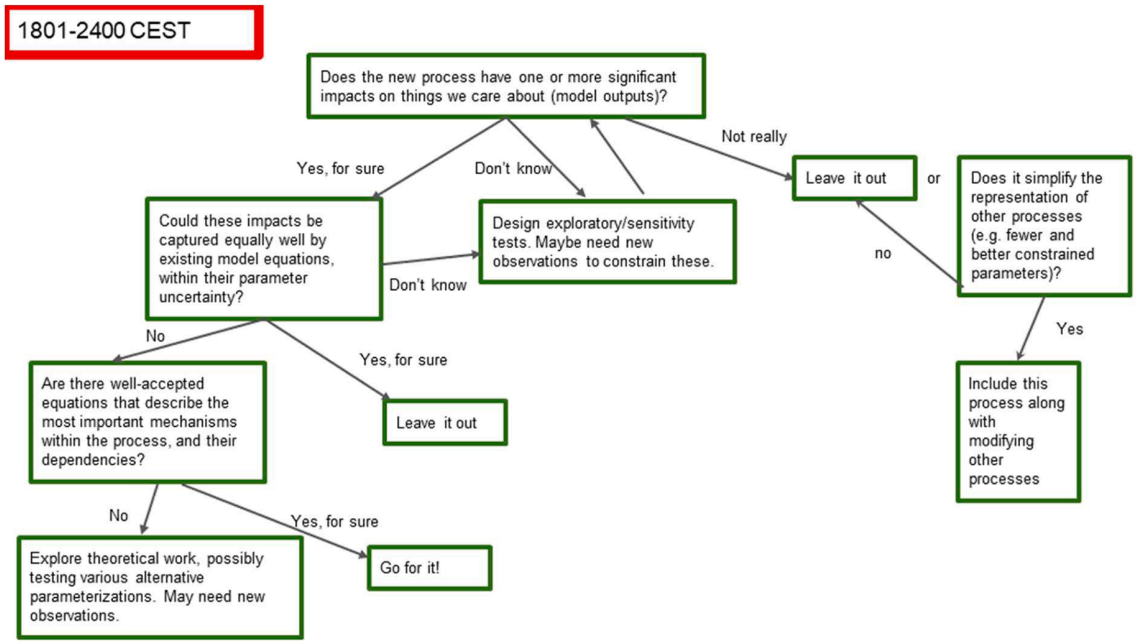


**Fig. A5.** As Fig. A1 but for 18:01–24:00 CEST workgroup.

## References


Allesina, S., Tang, S., 2012. Stability criteria for complex ecosystems. Nature 483 (2012), 205–208. https://doi.org/10.1038/nature10832.

Anderson, T.R., 2005. Plankton functional type modelling: running before we can walk? J. Plankton Res. 27 (11), 1073–1081. https://doi.org/10.1093/plankt/fbi076.

Bonachela, J.A., Raghib, M., Levin, S.A., 2011. Dynamic model of flexible phytoplankton nutrient uptake. In: Proceedings of the National Academy of Sciences, 108, pp. 20633–20638. https://doi.org/10.1073/pnas.1118012108.

Bopp, L., et al., 2022. Diazotrophy as a key driver of the response of marine net primary productivity to climate change. Biogeosciences 19 (17), 4267–4285. https://doi.org/10.5194/bg-19-4267-2022.

Chassignet, E.P., Xu, X., 2021. On the importance of high-resolution in large-scale ocean models. Adv. Atmos. Sci. 38 (10), 1621–1634.

Claussen, M., et al., 2002. Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models. Clim. Dyn. 18, 579–586. https://doi.org/10.1007/s00382-001-0200-1.

Crouch, S., Hong, N.C., Hettrick, S., Jackson, M., Pawlik, A., Sufi, S., Carr, L., De Roure, D., Goble, C., Parsons, M., 2013. The software sustainability institute: changing research software attitudes and practices. Comput. Sci. Eng., 15 (6), 74–80. https://doi.org/10.1109/MCSE.2013.133.

de Jong, H., et al., 2017. Mathematical modelling of microbes: metabolism, gene expression and growth. J. R. Soc., Interface 14, 20170502. https://doi.org/10.1098/rsif.2017.0502.

Davila, X., et al., 2023. On the origins of open ocean oxygen minimum zones. J. Geophys. Res.: Oceans 128, e2023JC019677. https://doi.org/10.1029/2023JC019677.

Denman, K.L., 2003. Modelling planktonic ecosystems: parameterizing complexity. Prog. Oceanogr. 57 (3–4), 429–452. https://doi.org/10.1016/S0079-6611(03)00109-5.

Dirmeyer, P.A., Ford, T.W., 2020. A technique for seamless forecast construction and validation from weather to monthly time scales. Mon. Wea. Rev. 148, 3589–3603. https://doi.org/10.1175/MWR-D-19-0076.1.

Epstein, J.M., 2008. Why model? J. Artificial Societ. Social Simulation 11 (4), 12. www.jasss.org/11/4/12.html.

Fennel, K., Mattern, J.P., Doney, S.C., et al., 2022. Ocean biogeochemical modelling. Nat. Rev. Methods Primers 2, 76. https://doi.org/10.1038/s43586-022-00154-2.

Flynn, K., 2006. Reply to horizons article 'Plankton functional type modelling: running before we can walk' Anderson (2005): II. Putting trophic functionality into plankton functional types. J. Plankton Res. 28 (9), 873–875. https://doi.org/10.1093/plankt/fbl015.

Forbes, V.E., 2024. The need for standardization in ecological modeling for decision support: lessons from ecological risk assessment. Ecol. Modell. 492, 110736. https://doi.org/10.1016/j.ecolmodel.2024.110736.

Friedrichs, M.A.M., et al., 2007. Assessment of skill and portability in regional marine biogeochemical models: role of multiple planktonic groups. J. Geophys. Res.: Oceans 112 (C8). https://doi.org/10.1029/2006JC003852.

Galbraith, E.D., et al., 2015. Complex functionality with minimal computation: promise and pitfalls of reduced-tracer ocean biogeochemistry models. J. Adv. Model. Earth Syst. 7, 2012–2028. https://doi.org/10.1002/2015MS000463.

Gurgacz, N.S., et al., 2023. Impact of plastic pollution on atmospheric carbon dioxide. FACETS 8, 1–7. https://doi.org/10.1139/facets-2023-0061.

Hall, C.A., 1988a. An assessment of several of the historically most influential theoretical models used in ecology and of the data provided in their support. Ecol. Modell. 43 (1–2), 5–31.

Hall, C.A., 1988b. What constitutes a good model and by whose criteria? Ecol. Modell. 43 (1–2), 125–127.

Hannah, C., Vezina, A., John, M.S., 2010. The case for marine ecosystem models of intermediate complexity. Prog. Oceanogr. 84 (1–2), 121–128. https://doi.org/10.1016/j.pocean.2009.09.015.

Hattab, T., et al., 2014. Towards a better understanding of potential impacts of climate change on marine species distribution: a multiscale modelling approach. Global Ecol. Biogeogr. 23 (12), 1417–1429. https://doi.org/10.1111/geb.12217.

Henson et al., (preprint). Knowledge gaps in quantifying the climate change response of biological storage of carbon in the ocean. https://essopenarchive.org/doi/full/10.22541/essoar.170473205.55245231/v1.

IPCC, 2023. In: Lee, H., Romero, J. (Eds.), Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel On Climate Change. IPCC, Geneva, Switzerland, p. 184. https://doi.org/10.59327/IPCC/AR6-9789291691647 [Core Writing Team].

Jakeman,A.J. et al. (In Press). Towards normalizing good practice across the whole modeling cycle: its instrumentation and future research topics. Socio-Environ. Syst. Modell.

Janssen, A.B.G., et al., 2015. Exploring, exploiting and evolving diversity of aquatic ecosystem models: a community perspective. Aquatic Ecol. 49, 513–548. https://doi.org/10.1007/s10452-015-9544-1.

Kosmala, M., et al., 2016. Assessing data quality in citizen science. Front. Ecol. Environ 14, 551–560. https://doi.org/10.1002/fee.1436.

Kriest, I., Khatiwala, S., Oschlies, A., 2010. Towards an assessment of simple global marine biogeochemical models of different complexity. Prog. Oceanogr. 86, 337–360. https://doi.org/10.1016/j.pocean.2010.05.002.

Kriest, I., et al., 2020. One size fits all? Calibrating an ocean biogeochemistry model for different circulations. Biogeosciences 17, 3057–3082. https://doi.org/10.5194/bg-2020-9.

Kyker-Snowman, E., et al., 2022. Increasing the spatial and temporal impact of ecological research: a roadmap for integrating a novel terrestrial process into an Earth system model. Glob. Chang. Biol. 28 (2), 665–684. https://doi.org/10.1111/gcb.15894.

Le Quéré, C, 2006. Reply to Horizons Article 'Plankton functional type modelling: running before we can walk' Anderson (2005): I. Abrupt changes in marine ecosystems? J. Plankton Res. 28 (9), 871–872. https://doi.org/10.1093/plankt/fbl014.

Levins, R., 1966. The strategy of model building in population biology. Am. Sci. 54 (4), 421–431.

Moore, D.J.P., 2022. A framework for incorporating ecology into earth system models is urgently needed. Glob. Chang. Biol. 28 (2), 343–345. https://doi.org/10.1111/gcb.15915.

Niemeyer, D., et al., 2017. A model study of warming-induced phosphorus–oxygen feedbacks in open-ocean oxygen minimum zones on millennial timescales. Earth Syst. Dynam. 8 (2), 357–367. https://doi.org/10.5194/esd-8-357-2017.

Nisbet, R.M., 2012. Integrating dynamic energy budget (DEB) theory with traditional bioenergetic models. J. Exp. Biol. 215, 892–902. https://doi.org/10.1242/jeb.059675.
Olivola, C.Y., 2018. The Interpersonal Sunk-Cost Effect. Psychol. Sci. 29 (7), 1072–1083. https://doi.org/10.1177/0956797617752641.
Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28 (3), 245–248. https://doi.org/10.1038/nbt.1614.
Orzack, S.H., Sober, E., 1993. A critical assessment of Levins's the strategy of model building in population biology (1966). Q. Rev. Biol. 68 (4), 533–546.
Patel, R, 2020. Why software is not accessible: technology professionals' perspectives and challenges. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–9. https://doi.org/10.1145/3334480.3383103.
Preisig, H.A. et al. (2006). On model portability. In W. Marquardt & C. Pantelides (Eds.), Computer Aided Chemical Engineering. 21, https://doi.org/10.1016/S1570-7946(06)80092-1.
Punt, A.E., et al., 2020. Essential features of the next-generation integrated fisheries stock assessment package: a perspective. Fish. Res. 229, 105617. https://doi.org/10.1016/j.fishres.2020.105617.
Rohr, T., et al., 2023. Zooplankton grazing is the largest source of uncertainty for marine carbon cycling in CMIP6 models. Nature Communicat. Earth Environ. 4 (1), 212. https://doi.org/10.1038/s43247-023-00871-w.
Sarmiento, J.L., Toggweiler, J.R., 1984. A new model for the role of the oceans in determining atmospheric P CO2. Nature 308 (5960), 621–624. https://doi.org/10.1038/308621a0.
Séférian, R., et al., 2020. Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6. Curr. Clim. Change Rep. 6, 95–119. https://doi.org/10.1007/s40641-020-00160-0.
Schmittner, A., et al., 2008. Future changes in climate, ocean circulation, ecosystems, and biogeochemical cycling simulated for a business-as-usual CO2 emission scenario until year 4000 AD. Global Biogeochem. Cycles 22 (1). https://doi.org/10.1029/2007GB002953.
Shea, C.H., et al., 2023. Small particles and heterotrophic protists support the mesopelagic zooplankton food web in the subarctic northeast Pacific Ocean. Limnol. Oceanogr. 68, 1949–1963. https://doi.org/10.1002/lno.12397.
Siewertsen, E., Piwonski, J., Slawig, T., 2013. Porting marine ecosystem model spin-up using transport matrices to GPUs. Geosci. Model. Dev. 6 (1), 17–28. https://doi.org/10.5194/gmd-6-17-2013.
Strzepek, R.F., 2022. The ongoing need for rates: can physiology and omics come together to co-design the measurements needed to understand complex ocean biogeochemistry? J. Plankton Res. 44, 485–495. https://doi.org/10.1093/plankt/fbac026.
Tagliabue, A., et al., 2021. Persistent uncertainties in ocean net primary production climate change projections at regional scales raise challenges for assessing impacts on ecosystem services. Frontiers in Climate 3. https://doi.org/10.3389/fclim.2021.738224.
Ward, B.A., et al., 2013. When is a biogeochemical model too complex? Objective model reduction and selection for North Atlantic time-series sites. Prog. Oceanogr. 116, 49–65. https://doi.org/10.1016/j.pocean.2013.06.002.
Wu, Z., et al., 2022. Single-cell measurements and modelling reveal substantial organic carbon acquisition by Prochlorococcus. Nat. Microbiol. 7 (12), 2068–2077. https://doi.org/10.1038/s41564-022-01250-5.
Yool, A., et al., 2017. Big in the benthos: future change of seafloor community biomass in a global, body size-resolved model. Glob Chang Biol 23, 3554–3566. https://doi.org/10.1111/gcb.13680.
Ziehn, T., et al., 2017. The carbon cycle in the australian community climate and earth system simulator (ACCESS-ESM1) – Part 2: historical simulations. Geosci. Model. Dev. 10 (7), 2591–2614. https://doi.org/10.5194/gmd-10-2591-2017.