

14 November 2025

Dr. Sophie Valcke, Editor
Geoscientific Model Development

We enclose the revisions for our manuscript “Automatic reduction of ocean biogeochemical models: a case study with BFM (v5.3)” and address the reviewers’ comments in the following remarks. We have indicated our changes in the manuscript with the following color code: **Reviewer 1** and **Reviewer 2**. Deletions are shown with a strike out and the respective color, e.g., a deletion for Reviewer 2 would look like this: ~~some-deleted-text~~. In this document, we consider each reviewer’s comment and provide our response with “**Our Response:**”.

We sincerely thank the reviewers for their suggestions, and for the helpful feedback

Reviewer #1

This manuscript addresses the important and computationally challenging problem of reducing the complexity of ocean biogeochemical models. The authors adapt a model reduction technique from combustion science (DRGEP) and apply it to the 50-state-variable Biogeochemical Flux Model (BFM), generating a suite of smaller models. The topic is timely and of significant interest to the ocean modeling community.

However, while the goal is laudable, the manuscript in its current form suffers from several fundamental scientific flaws in its methodology, interpretation of results, and the substantiation of its core claims. The proposed “modified DRGEP” method is presented without sufficient theoretical justification and appears ad hoc. Its application requires manual interventions that undermine its claim of being an “automatic” process. Furthermore, the study’s conclusions are drawn from a very narrow set of environmental conditions, leading to overstated claims of generalizability. The results include spurious findings, such as a single-variable oxygen model, which stem from a misapplication of the reduction technique. Finally, the central motivation of the paper—improving computational efficiency—is never quantitatively demonstrated.

Due to these significant issues, the manuscript is not suitable for publication in its current state. A major revision is required to address the methodological weaknesses and to provide a more rigorous validation and analysis of the results.

Our Response: Thank you for the feedback. The edits made in response to your comments have strengthened the manuscript. We hope that our additions and revisions have satisfactorily addressed these concerns, and clarified the goals of the manuscript.

The original DRGEP method is rooted in the principles of chemical kinetics. This manuscript replaces the mechanistically-derived interaction coefficient with an empirical “error matrix”. This approach is problematic because: (i) The method for calculating the modified direct interaction coefficient—by removing state variables

one by one and normalizing by the maximum row-wise error—is not theoretically justified. Why is this specific normalization chosen? It could introduce biases, and its relationship to the propagation of error through a coupled, non-linear system is unclear. (ii) The authors should provide a much more robust defense of this methodological choice, perhaps by testing it on a simpler, known system to demonstrate its validity before applying it to a complex model like the BFM.

Our Response: Thank you for raising this question. This approach is intended to pragmatically calculate what the direct interaction coefficients were designed to quantify in the DRG and DRGEP methods: impact on overall production rate of one component induced by removal of another component [1, 2]. We chose to normalize by the maximum row-wise error to scale the contribution of removing state variable B to the rate of change in state variable A between $[0,1]$ to apply a single cutoff threshold. Without this normalization we cannot ensure consistency in threshold values between all state variables.

To further justify our introduction of the modified DRGEP method, we have added a short verification study (new **Section 3**, on **pages 8–9**) using GRI-Mech 3.0, a gas-phase combustion model consisting of 53 chemical species and 325 elementary reactions. The results indicate that the modified DRGEP method performs similarly to the DRGEP method in terms of size and accuracy of reduced models produced.

The authors report that they were required to manually retain ammonium, all phytoplankton chlorophyll constituents, and in one case nitrate, to prevent the model from failing. The justification is that these variables play roles “not completely captured using the interaction coefficients”. This is a critical admission that the proposed method is flawed and fails on its own to identify indisputably essential components of the ecosystem. An objective and truly automatic method should not require such a priori expert knowledge to ensure a biologically sensible outcome.

Our Response: It is true that the modified DRGEP—like the original DRG and DRGEP methods applied to chemical-kinetics models [1, 3, 4]—does not completely capture the complex roles of certain state variables with the interaction coefficients. This comes from the component–component nature of the interaction coefficients, which do not map perfectly to the global quantities of interest. The need to manually specify or “protect” components from removal, particularly inert species, is consistent with all prior applications of the DRG and DRGEP methods used in the combustion literature. Ultimately, the goal of these reduction methods is to pragmatically automate the reduction procedure, but human inputs are still required; the method is not intended to provide a theoretical quantity of component importance.

The entire analysis is based on a single set of climatological forcing data from the BATS site. The conclusion that 14 state variables (including all mesozooplankton, silicate, and semi-refractory DOM) can be removed is therefore only valid for this specific location and condition, yet it is presented as a major finding of the paper. The removal of entire functional groups like diatoms and mesozooplankton makes the reduced models fundamentally unsuitable for application in many other oceanic

regimes (e.g., upwelling systems, polar oceans) or for studying key ecosystem processes like trophic transfer and carbon export.

Our Response: The reviewer is absolutely right: the reduced models produced here may be unsuitable for other applications. The reduced models shown here are specific to the BATS conditions used, the target components specified, and the allowable error. The example reduced models produced here are for demonstration purposes only and not intended for broad use; we have added a clarification statement addressing this point to the Conclusions on **page 22**. The major contribution here is the introduction of the reduction method itself for biogeochemical models, and the major result is the success of the method.

The reduction strategy would need to be reapplied using environmental conditions from other ocean regimes to accurately represent the new system. This work is not intended to provide reduced-complexity ocean biogeochemical models for further study. Instead, the usable output is the method itself. We have added a statement emphasizing this point to the Conclusions on **page 23**.

The Oxygen3 scenario produced a model with only a single state variable, oxygen, leading the authors to conclude that peak oxygen concentration is driven more by air-sea flux than biology. This conclusion is scientifically unsound. The BFM's own equations show that oxygen is directly consumed and produced by biological processes. The reduction algorithm has simply identified the largest term in the budget at the surface and erroneously concluded that all other coupled terms are unimportant. This result should be presented as a failure of the method under these conditions, not as a valid scientific insight.

Our Response: The purpose of the reduction method is to produce a more-compact model by eliminating unimportant components, depending on the desired application. Error is unavoidable with respect to the initial, full-detail model—instead, an allowable level of error is specified, with the reduction in model size as the tradeoff.

While the reviewer is correct that biological processes directly produce and consume oxygen, if the oxygen profile is the only quantity of interest, using a model that solely represents oxygen allows capturing trends in oxygen profile throughout the water column with a 2.2% normalized root mean square error in oxygen concentration. However, this may not extend to all ecosystems or conditions, and should only be used for the conditions considered in the reduction. As mentioned above, we have added a clarification statement addressing this point to the Conclusions on **page 22**.

Furthermore, the modified DRGEP method is applied at multiple depths in the water column to consider surface, subsurface, and bottom effects while performing the model reduction analysis. The direct interaction coefficient for all state variables was chosen as the maximum across all depths to prevent bias from surface or bottom fluxes.

The manuscript is motivated entirely by the need to reduce the high computational cost of biogeochemical models. However, the paper presents zero data to support the claim that the reduced models are more efficient. The authors must provide a

quantitative analysis of the computational speed-up for each reduced model versus the full BFM50. Without this, the primary justification for the work is missing.

Our Response: Thank you for this comment. We agree that we should do a better job of demonstrating the benefits of the reduced models. To address this point, we generated reduced models and added a new subsection (5.2, on page 21) and figure (Figure 9) to the manuscript:

Subsection 5.2:

To demonstrate the potential improvement in performance by reducing biogeochemical models, we manually constructed versions of each reduced model obtained from the reduction strategies used in this study. We used Oregon State University’s high performance computing cluster to run five-year simulations of the fully-coupled biophysical models, using the same conditions as described previously. Figure 9 shows the average computational time from five simulations run for each model. Reducing the number of components by a factor of 1.4–2.2× reduces the wall-clock time by 2.6–3.6×, or between a linear and quadratic reduction in time. Larger, more-complex initial models would see further factors of reduction in computational cost. In this demonstration, transport operations contribute substantially to the overall cost, as seen in the sub-linear time reduction when reducing to a single model component.

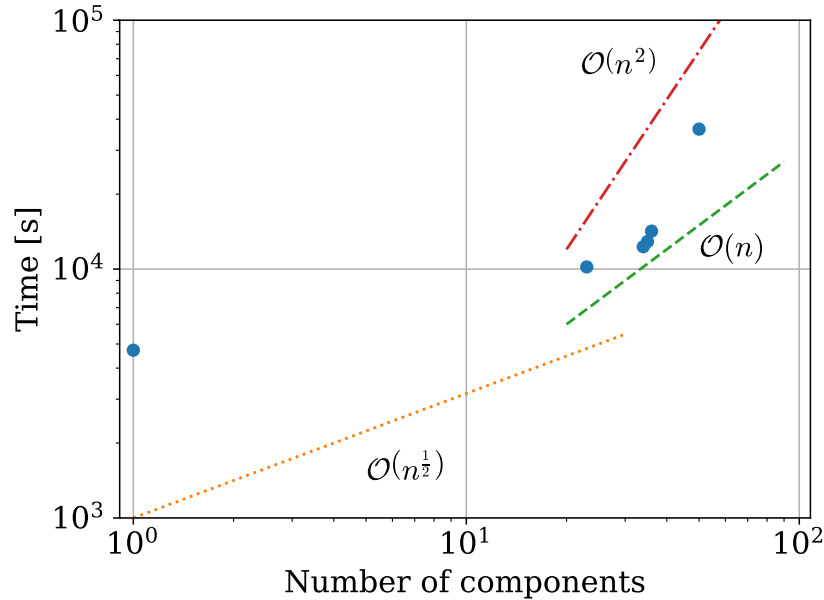


Figure 9: Wall-clock times for five-year simulations of reduced BFM models and full 50-component BFM coupled to pyPOM1D, shown as a function of component number. Trend lines indicate square-root, linear, and quadratic scaling of time with number of components.

The current BFM implementation does not allow for the full removal of state variables from the model and generation of a true reduced model, because the model structure is hardcoded. Instead, to evaluate the performance of reduced models, we use flags to zero out unimportant state variables and their associated rates to disregard their contribution to the system. This does not reduce the

size of the system of ordinary differential equations or the number of computations performed during the simulation, and so does not result in a significant change in computation time.

We also clarify this point in new subsection 5.2:

The current hardcoded structure of the BFM implementation does not allow generation of truly reduced models, with identified unimportant components and interactions fully removed. Rather, state variables identified as unnecessary for a given reduction strategy, along with all associated rate of change components, are set to zero in the simulation to eliminate their contribution to the model output. While this enables our examination of the reduction approach here, it does not reduce the size of the system of ordinary differential equations or the number of calculations performed throughout the simulation. Time complexity of iterative solvers used in implicit time integration algorithms scales between $O(N^2)$ and $O(N^3)$ in the worst case, where N represents the size of the system [5, 6]. These costs come from numerical evaluation and factorization of the Jacobian matrix.

The reduction algorithm is guided by errors in average concentration, peak concentration, and time of peak. These metrics can easily mask significant errors in the model's dynamic behavior. They do not constrain process rates (e.g., NPP, grazing) or the duration of events (e.g., a bloom). A model could match a peak value but have completely unrealistic fluxes, making it unsuitable for most scientific applications. The authors should evaluate the reduced models against a wider set of process-based metrics.

Our Response: Thank you for pointing this out. We agree that certain dynamics may be masked using the metrics we chose. However, we are concerned with the overall contribution of a state variable to the system rather than the dynamics of individual biological processes. We added clarification for this point in Section 4 on **page 10**:

The target state variables and metrics for the reduction strategies are chosen for the purpose of evaluating the BFM. Individual biological processes are not evaluated in the reduction strategies. Rather, state variables are retained or removed based on their overall contribution to the dynamics of the system.

The study's entire methodology is built upon the "modified direct interaction coefficient," which is derived from a "percent error" in the instantaneous rate of change, dCA/dt (Eq. 6). This metric is fundamentally unstable for stiff, non-linear systems like biogeochemical models. For state variables that are near quasi-steady state, the denominator, $|dCA/dt|_{\text{original}}$, will approach zero. This will cause the ratio to become numerically explosive, massively amplifying any minor numerical noise or transient fluctuation. The authors do not describe any form of regularization, time-averaging, or thresholding to handle this critical instability, which means the foundation of their graph reduction is unreliable.

After calculating the error matrix, the authors normalize each row by its maximum value to create coefficients bounded between $[0,1]$ (Eq. 7). This step erases the absolute magnitude of the sensitivities. Consequently, a state variable that is highly

sensitive to multiple perturbations becomes indistinguishable from a variable that is only weakly sensitive to all perturbations. Applying a single, global cutoff threshold, ϵ , to these normalized—and no longer comparable—values has no consistent physical or biogeochemical meaning, likely leading to the erroneous pruning or retention of state variables.

Our Response: Indeed, this is the major improvement of the DRGEP (and modified DRGEP) method over the original DRG method [2].

Normalizing the error matrix by the maximum row-wise error ensures a bounded direct interaction coefficient between $[0,1]$ for all state variables regardless of differences in magnitude.

The purpose of the direct interaction coefficient is to quantify the importance of one component to another in a normalized way, by estimating the direct impact of removal [1, 2, 7]. In this way, only direct interaction coefficients r_{AB} , r_{AC} , r_{AD} , etc., are directly comparable, because they rank the importance of different components to component A only. This is value of the path-dependent nature of the DRGEP method and the overall interaction components, based on particular overall targets of interest. Individual direct interaction components are not compared and the threshold is not applied to them; instead, the threshold is applied to the overall interaction coefficients, found as the product along pathways of the direct coefficients.

In response to this comment, we now include a new section (3) and Figure (3) that provide a verification of the modified DRGEP method:

To verify that the modified DRGEP method performs similarly as the original DRGEP method, we applied both to reduce GRI-Mech 3.0 [8], a 53-species gas-phase combustion model consisting of 325 chemical reactions. We implemented the DRGEP method using pyMARS [9], which uses the Cantera library [10] for handling chemical kinetics. To sample for the reduction process and evaluate error of reduced models, we use autoignition simulations that solve time-dependent governing equations for conservation of energy and conversation mass in a homogeneous reactor (i.e., zero-dimensional, with no spatial transport). We specified an error limit of 5% in ignition delay time, the time required for temperature to increase by 400 K. All combinations of the following conditions were used for the autoignition simulations: four initial temperatures (1000, 1200, 1400, and 1600 K); three initial pressures (1, 11, and 21 atm); and three equivalence ratios (0.5, 1.0, 1.5).

Figure 3 shows number of species in reduced models versus threshold (ϵ), along with error in ignition delay for a single ignition condition (1000 K, 1 atm, and equivalence ratio of 1.0). The results indicate that the modified DRGEP method performs qualitatively similar to the DRGEP method in terms of size and accuracy for reduced models. For a given threshold, the modified DRGEP method retains additional species, though the errors in the reduced models produced by the methods remain nearly identical until $\epsilon \approx 0.21$. These trends—including the flat error with reducing species count until a sharp spike—match those seen previously for DRGEP, and the difference between the methods is similar to differences between DRG and DRGEP implementations in past comparisons [2, 3, 11]. While DRGEP removes more species for a given threshold value than modified DRGEP, ultimately the methods produce a similarly sized final reduced model for the given upper error limit of 5%. In fact, modified DRGEP method reduces

the model to 33 species and 187 reactions with 1.07% error in ignition delay time. The DRGEP method reduces the model to 35 species and 205 reactions with 1.08% error in ignition delay time.

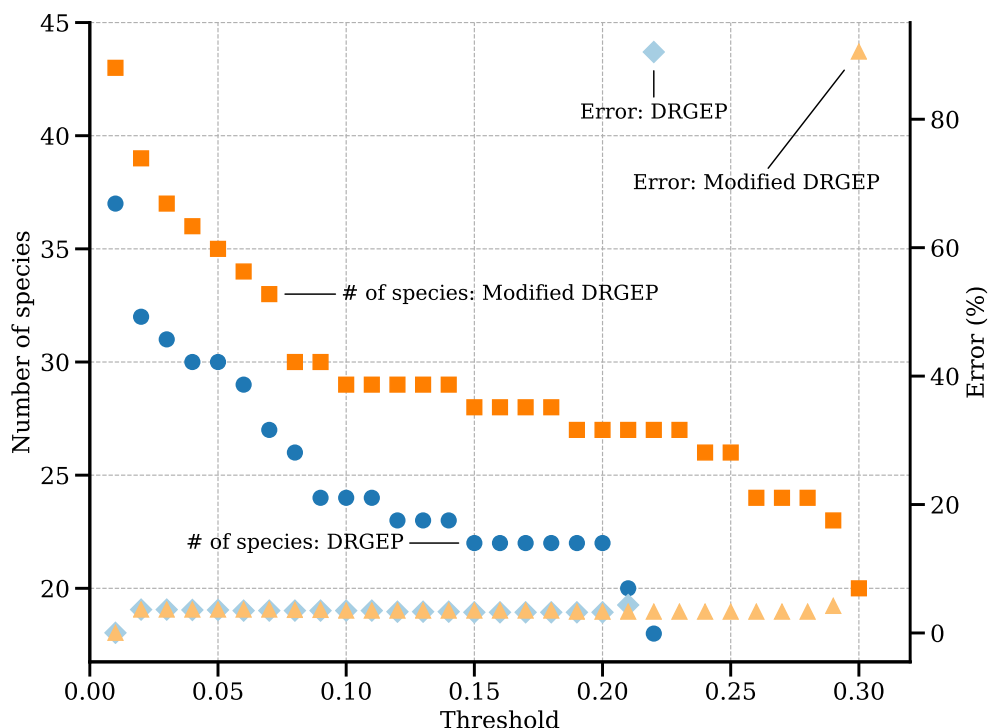


Figure 3: Number of species and percent error in ignition delay versus threshold for DRGEP and modified DRGEP for the GRI Mech 3.0 model [8]; the error in ignition delay comes from initial conditions of 1000 K, 1 atm, and an equivalence ratio of 1.0.

The tables summarizing the reduction scenarios (e.g., Tables 1-5) contain a column labeled “% Error”. However, the values reported in this column are not percentages but are given in absolute physical units (e.g., mg Chl-a m-3, mmol N m-3). This is a significant inconsistency, not a minor typographical error. It creates a fundamental ambiguity in interpreting the results and undermines the credibility of the automated reduction algorithm, which is described as terminating when a “percent error” exceeds a user-specified limit.

Our Response: Thank you for pointing out this confusing inclusion in the original manuscript; the physical units we included represented the state variables rather than the error itself. The error is given as a percentage and does not have units. We have removed these units to avoid the confusion.

Reviewer #2

The manuscript titled “Automatic reduction of ocean biogeochemical models: a case study with BFM (v5.3)” provides a comprehensive examination of simplifying the complexity of ocean biogeochemical models using the DRGEP method. The authors demonstrate how they reduced the 50-state-variable BFM model to a smaller set of variables while preserving essential system behavior. By testing various scenarios involving living organic matter, nutrients, oxygen, and carbon, the study successfully achieves model reduction without significant loss of information. The authors further suggest that this reduction approach could be applied to other biogeochemical models, potentially leading to substantial decreases in computational costs for highly complex systems.

While the manuscript effectively describes the methodology and results of the reduction process, several key issues should be addressed before publication. In particular, the study should more clearly articulate its primary research objective and emphasize the novel contributions of this work to strengthen its scientific impact.

Our Response: Thank you for your positive feedback, and helping us improve our communication around the objectives and contributions.

1. The authors have effectively conducted a reduction of the ocean biogeochemical model, particularly for the Princeton Ocean Model coupled with the BFM biogeochemical module. In the abstract, the authors mention that large-eddy simulations (LES) require high computational resources. As I understand the manuscript, the reduction of variables aims to enable the implementation of a simplified biogeochemical model within LES that explicitly resolves turbulence. However, the manuscript lacks a clear statement of the main objective behind reducing the ocean biogeochemical variables. Is the goal to develop a 1D turbulence-resolving LES or to reduce model complexity in climate-scale Earth System Models? It would be helpful for the authors to clarify in the abstract what type of model development—LES or large-scale climate modeling—the variable reduction and computational efficiency are intended to support.

Our Response: Thank you for pointing out the lack of a clear objective statement. The reviewer is correct that our main objective is producing reduced models for use in large-scale multidimensional LES that resolve turbulence and the evolution of reactive biogeochemical tracers. In this manuscript, our focus is in demonstrating a method for automatically producing such reduced models; the actual models presented here are not intended for reuse beyond the specific conditions used for their generation.

Our abstract does state

Large-scale biophysical simulations, such as high-resolution large-eddy simulations (LES) of the upper ocean, require significant computing resources to capture small-scale turbulent processes while also resolving the evolution of reactive biogeochemical tracers.

and we have now added an additional statement clarifying our primary goal on **page 3**:

Our primary goal in reducing such models is to produce more-compact versions to be used in large-scale biophysical simulations, such as high-resolution LES of the upper ocean that resolve both small-scale turbulent processes and the evolution of reactive biogeochemical tracers.

2. The authors have conducted tests using ocean biogeochemical datasets at the BATS site and compared the full version of the 1D BFM with its reduced versions across various biogeochemical variables. However, first and foremost, the full-version BFM simulation should be directly compared with the BATS observations. Even though the BFM development team may have demonstrated this in previous studies, a fundamental step in model validation is to evaluate the baseline biases between the observations and the full-variable model and understand the reduced-variable results based on these biases.

In addition, it is unclear what specific changes have been made between the full-variable configuration (BFM50) and the reduced-variable simulations (e.g., BFM23, BFM36). Therefore, I strongly recommend that the authors clearly present the differences between the full and reduced versions in Figures 5, 6, 7, and related supplementary figures. For example, in Section 4.1.2, the authors note oversaturation of nitrate and phosphate and underestimation of PON and DIC; however, these differences warrant closer examination and clearer comparison.

Our Response: The reviewer is correct that a validation of the full BFM simulation is an important step. The purpose of this work is not to provide reduced versions of the BFM to be used for further research. Recent articles from collaborators have studied the performance of BFM with respect to BATS and HOTS observational data in more detail [12]. We have added a statement addressing this point to the Conclusions on **page 22**.

The objective of this work is to demonstrate the ability of the modified DRGEP method to reduce high-complexity ocean biogeochemical models under a given set of environmental conditions. Here, we verify the method by comparing the reduced models to the full 50-state-variable BFM rather than observational data.

The state variables removed in each reduction scenario are actually presented throughout Section 4 already (**pages 11, 13, 14, 15**). The differences in the fields between the full and reduced models are attributed to the removal of living organisms and their biological rates—nutrient uptake, respiration, excretion, lysis, etc.

3. Although the reduction of BGC variables is well demonstrated using the BATS dataset, I am concerned that this reduction may lead to overfitting specific to the BATS conditions. This raises an important question regarding the sensitivity of the reduction method to physical environmental processes, including vertical diapycnal mixing, seasonal entrainment, and isopycnal mixing. While there are limited sites with long-term biogeochemical time-series, I recommend performing a similar analysis using the HOTS dataset. Alternatively, applying the reduced-variable model to other locations for validation would strengthen the robustness of the reduced BATS-based model and demonstrate its potential applicability for future model de-

velopments.

Our Response: The reviewer is absolutely correct that the reduced models presented here may be overfit/specifically apply to the BATS conditions. In general the reduced models presented in this manuscript are specific to the environmental conditions used and would need to be reapplied if such conditions were to change. We have added a clarification statement addressing this point to the Conclusions on **page 22**.

20: Recommend to add the most recent ocean biogeochemical models such as MARBL, PISCES.

Our Response: We have added mention to these additional recent ocean biogeochemical models on **page 2**.

References

- [1] T Lu and CK Law, “A directed relation graph method for mechanism reduction,” *Proceedings of the Combustion Institute*, vol. 30, no. 1, pp. 1333–1341, 2005. DOI: [10.1016/j.proci.2004.08.145](https://doi.org/10.1016/j.proci.2004.08.145).
- [2] P Pepiot-Desjardins and H Pitsch, “An efficient error-propagation-based reduction method for large chemical kinetic mechanisms,” *Combustion and Flame*, vol. 154, no. 1–2, pp. 67–81, 2008. DOI: [10.1016/j.combustflame.2007.10.020](https://doi.org/10.1016/j.combustflame.2007.10.020).
- [3] KE Niemeyer, CJ Sung, and MP Raju, “Skeletal mechanism generation for surrogate fuels using directed relation graph with error propagation and sensitivity analysis,” *Combustion and Flame*, vol. 157, no. 9, pp. 1760–1770, Sep. 2010. DOI: [10.1016/j.combustflame.2009.12.022](https://doi.org/10.1016/j.combustflame.2009.12.022).
- [4] KE Niemeyer and CJ Sung, “Reduced Chemistry for a Gasoline Surrogate Valid at Engine-Relevant Conditions,” *Energy & Fuels*, vol. 29, no. 2, pp. 1172–1185, 2015. DOI: [10.1021/ef5022126](https://doi.org/10.1021/ef5022126).
- [5] T Lu and CK Law, “Toward accommodating realistic fuel chemistry in large-scale computations,” *Progress in Energy and Combustion Science*, vol. 35, no. 2, pp. 192–215, 2009. DOI: [10.1016/j.pecs.2008.10.002](https://doi.org/10.1016/j.pecs.2008.10.002).
- [6] Y Bin, XI Yang, SJ Grauer, and RF Kunz, “Data-enabled reduction of the time complexity of iterative solvers,” *Journal of Computational Physics*, vol. 529, 2025. DOI: [10.1016/j.jcp.2025.113859](https://doi.org/10.1016/j.jcp.2025.113859).
- [7] KE Niemeyer and CJ Sung, “On the importance of graph search algorithms for DRGEP-based mechanism reduction methods,” *Combustion and Flame*, vol. 158, no. 8, pp. 1439–1443, 2011. DOI: [10.1016/j.combustflame.2010.12.010](https://doi.org/10.1016/j.combustflame.2010.12.010).
- [8] GP Smith, DM Golden, M Frenklach, NW Moriarty, B Eiteneer, M Goldenberg, CT Bowman, RK Hanson, S Song, J William C. Gardiner, VV Lissianski, and Z Qin, *GRI-Mech 3.0*, http://www.me.berkeley.edu/gri_mech/, Jul. 1999.
- [9] PO Mestas, P Clayton, and KE Niemeyer, “Pymars: Automatically reducing chemical kinetic models in python,” *Journal of Open Source Software*, vol. 4, no. 41, p. 1543, 2019. DOI: [10.21105/joss.01543](https://doi.org/10.21105/joss.01543).

- [10] DG Goodwin, RL Speth, HK Moffat, and BW Weber, *Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes*, <https://www.cantera.org>, Version 2.4.0, 2018. DOI: [10.5281/zenodo.1174508](https://doi.org/10.5281/zenodo.1174508).
- [11] L Tosatto, BAV Bennett, and MD Smooke, “Comparison of different DRG-based methods for the skeletal reduction of JP-8 surrogate mechanisms,” *Combustion and Flame*, vol. 160, no. 9, pp. 1572–1582, Sep. 2013, ISSN: 0010-2180. DOI: [10.1016/j.combustflame.2013.03.024](https://doi.org/10.1016/j.combustflame.2013.03.024).
- [12] S Kern, ME McGuinn, KM Smith, N Pinardi, KE Niemeyer, NS Lovenduski, and PE Hamlington, “Computationally efficient parameter estimation for high-dimensional ocean biogeochemical models,” *Geoscientific Model Development Discussions*, vol. 2023, pp. 1–34, 2023. DOI: [10.5194/gmd-2023-107](https://doi.org/10.5194/gmd-2023-107).