Dear reviewers,

Thank you for your interest in our paper and your feedback. In the following, we will directly respond to your comments.

Best regards,
Jan Gärtner

Addressing Comments of Referee #1

Reviewer Comment
The manuscript compares a demonstration case by Mehlmann et al (2021) with a stationary wind field and states that narrow features are observed. Are these right or wrong?

Author Response
As described in section 2.2 ("code verification"), we conducted a comparative analysis between Veris and the MITgcm under a variety of initial conditions and forcing fields, including those used to produce the sea ice field shown in Fig. 1. By comparing the differences between Veris and the MITgcm with the differences between the MITgcm and a MITgcm-rerun with perturbed initial and forcing conditions, we demonstrate the accuracy of Veris' dynamic component. We have added the following difference plot to the manuscript, showing that the differences between Veris and the MITgcm are within acceptable limits:

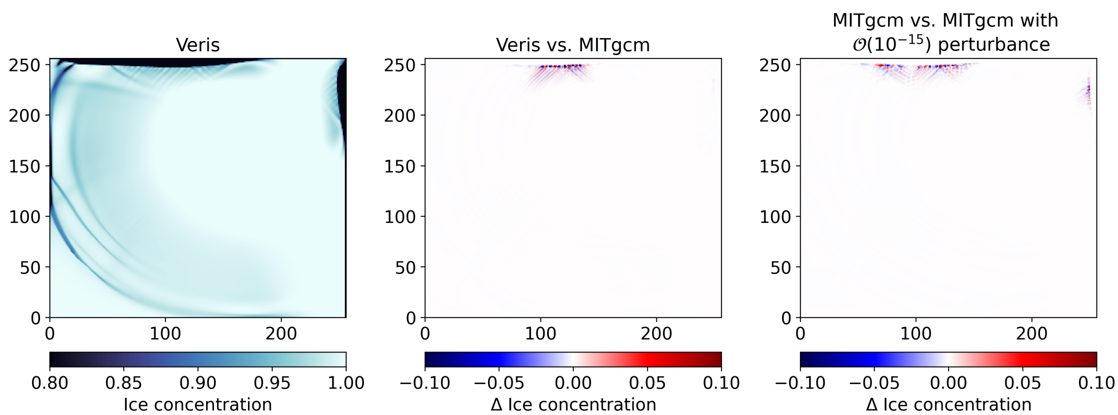

Figure A1: Example benchmark simulation used for verifying the dynamic component of Veris. Left panel: Sea ice concentration simulated by Veris after 1000 10-minute time steps using forcing fields designed to induce ice fracturing. Middle panel: Comparison of the simulated sea ice between Veris and the MITgcm employing the same initial and forcing conditions. Right panel: As a reference for acceptable divergence, the difference between the MITgcm and a rerun of the MITgcm with perturbed initial and forcing conditions is shown. This figure is representative of the results of different dynamic benchmark comparisons using different grid sizes and initial and forcing conditions.

Additionally, we have added this note in the caption of Fig. 1:

A reference simulation with the MITgcm, employing the same forcing, produced the same features (Fig. A1).

Reviewer Comment
Then it simulates Antarctic in order to show that the coupled system works. But again no real validation or discussion.

Author Response
Fig. 3 serves primarily as a proof of concept, showing that Veris can be integrated with Veros to form a fully Python-based coupled sea ice-ocean model. The accuracy of Veris itself was established through the direct comparison with the MITgcm presented in section 2.2. The simulation result of the coupled Veris-Veros model is mostly determined by the ocean model Veros. Consequently, a thorough analysis of the Antarctic sea ice would primarily evaluate Veros rather than Veris.

Reviewer Comment
The largest focus is on optimization and they show scaling based on increased domain size. This Scaling is normally shown

with increasing number of processors (including the timing of running on 1 processor). For climate models the limitation should be usage of the full bandwidth, which is not necessarily reached.

Author Response
In the first version of the submitted manuscript, we presented only the scaling of model performance with domain size, not with process count. This was because the initial version of the model could not run in parallel using the JAX backend. This limitation has now been resolved. In the updated version of Veris, the model infrastructure has been redesigned around JAX's sharded arrays, enabling full parallelization on both CPUs and GPUs. Accordingly, the revised manuscript now includes an analysis of model performance scaling with process count.

Reviewer Comment
Line 15: This is a subjective thing. I don't think that Python is easier to read or maintain. I think that this depends on the programmer. I agree that Python is easier accessible as it does not require a compiler and that more people have used this.

Author Response
In the introduction, we highlight that the advantages of Python over Fortran, particularly in terms of ease of use and maintainability, are especially beneficial for new researchers or those with limited experience in low-level programming.

Reviewer Comment
Line 77: I assume this is theoretical speaking and if all resources are used.

Author Response
We have revised the manuscript as follows, with the changes marked in italics:

Due to their higher core counts, GPUs are much better suited for parallel computing than CPUs*, provided that the available computational and memory resources are fully utilized.* As a result, they are more energy-efficient for parallel tasks such as climate simulations, as they complete these simulations faster than CPUs.

Reviewer Comment
Section 2.2 Validation: What is the setup for the dynamic test? 1000 iterations is a few days Is this long enough?

Author Response
We have revised the manuscript as follows:

For the dynamic component, a two-dimensional benchmark following Mehlmann et al. (2021) was conducted on rectangular grids with resolutions of 128×128*,* 256×256*,* 512×512*, and* 1024×1024 *grid cells. As in the thermodynamic tests, varying initial conditions for ice thickness, snow thickness, and ice concentration were employed, along with variations in atmospheric and oceanic forcing.*

In addition to the 1000-time step benchmarks, a longer integration of 3 months (12960 time steps) was performed to assess Veris's long-term behavior (Rosinski and Williamson, 1997), which remained consistent with the MITgcm.

Reviewer Comment
2.3 Parallelization. EVP cannot use JAX. How does the model speed up the dynamics in the JAX parallelization cases in the figure if it only do dynamics? This is normally the most expensive part.

Author Response
Veris and its EVP solver can use JAX. Furthermore, in the updated version of the model presented in the revised manuscript, Veris is also able to run in parallel on both CPU and GPU based architectures using JAX's sharded arrays.

Reviewer Comment
Section 3.1. Is this the same test as the dynamics? Described in section 2.2?

Author Response
In section 2.2, several tests were conducted using varying initial and forcing conditions, including, but not limited to, the simulation presented in section 3.1, Fig. 1.

Addressing Comments of Referee #2

Thank you for your interest in our paper and for your valuable feedback!

Over the past months, we have worked on parallelizing Veris by redesigning its model infrastructure around JAX's sharded arrays. In this updated version, Veris is able to run in parallel on both CPU- and GPU-based architectures. As a result, we have substantially revised the manuscript, which now includes an analysis of model performance scaling with process count, as well as an expanded discussion of the limitations and the potential of JAX for climate modeling.

Reviewer Comment
Sec 2.2: if I understood right, the section describes regression testing using MITgcm as the reference. This is code verification not validation. I am sorry for being strict here, but for a community that is trying to modernize itself, it is timely to use appropriate terminology.

Author Response
We have changed the section title to "Code Verification".

Reviewer Comment
sec 2.2: authors indicate relative differences of 10^{-7} and 10^{-3} for the quantities under testing and conveniently mention that these differences are acceptable without any justification. Typically, one perturbs the initial condition to generate ensembles to get an estimate of expected errors. The observed error is then subjected to these expected errors . I suggest authors to do a literature review on the subject of testing. One good example from the community is "THE ACCUMULATION OF ROUNDING ERRORS AND PORT VALIDATION FOR GLOBAL ATMOSPHERIC MODELS, JAMES M. ROSINSKI AND DAVID L. WILLIAMSON" and the from outside the community: "Verification of Codes and Calculations Patrick J. Roache*"

Author Response
The code verification section was rewritten and now includes a comparison between Veris and the MITgcm based on the suggested study and using a perturbed reference simulation.

Reviewer Comment
Sec 3.1: The results section is does not have any reference. For a non-expert in sea-ice modelling, it is difficult to evaluate correctness of the results.

Author Response
In the revised manuscript, we have added the following note in the figure caption:

A reference simulation with the MITgcm, employing the same forcing, produced the same features (Fig. A1).

Furthermore, we have added a plot comparing the simulated sea ice field between Veris and the MITgcm (Fig. A1), showing that the MITgcm simulates the same features. We refer to Mehlmann et al. (2021) for a comprehensive discussion of the simulated linear kinematic features:

These linear kinematic features are discussed in Mehlmann et al. (2021) and show that Veris's dynamics are on par with state-of-the-art sea-ice dynamics codes.