EGUsphere, referee comment RC2
https://doi.org/10.5194/egusphere-2022-797-RC2, 2022

Comment on egusphere-2022-797

Anonymous Referee #2

Referee comment on "Multifidelity Monte Carlo estimation for efficient uncertainty quantification in climate-related modeling" by Anthony Gruber et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-797-RC2, 2022

Review for Multifidelity Monte Carlo Estimation for Efficient Uncertainty Quantification in Climate-Related Modeling

The authors describe a "multi-fidelity" Monte Carlo method for expected value estimation that is shown to improve convergence/efficiency compared to traditional Monte Carlo methods. This multi-fidelity approach is different to existing techniques in that expected values are computed using simulations run at differing resolutions and/or incorporating various levels of dynamical approximation. The test cases presented here employ hierarchies of progressively coarsened meshes to provide low-cost estimators, as well as an interpolation-based approach. The authors present results for two shallow-water benchmarks and one ice-sheet test-case, and demonstrate significant speed-ups compared to the traditional MC scheme for various maximum-value estimation problems. The speed-ups demonstrated for the MFMC method are very attractive, especially in the context of climate-model estimators which are often extremely expensive to obtain. There are a number of questions I have relating MFMC performance to problem nonlinearity though, in addition to various minor corrections.

Comments:

- The three test-cases analysed appear to involve estimators evaluated after relatively short integration periods, and, rather than fully nonlinear flows/dynamics, the response of the system may be relatively linear over these short horizons as a result --- is this expected to influence the effectiveness of the MFMC results presented? For example, TC5 from Williamson does not become strongly nonlinear until approx. 20-days, with 50-days being the typical analysis window at which turbulence is fully developed. The SOMA case described in Wolfram (2015) is typically spun-up over several years, and then analysed over 30-day windows. In this work, it appears the TC5 case is analysed after 10 days, and the barotropic-gyre-version-of-SOMA after 3-days (restarted from a 15-day spin-up). Are the MFMC results robust when the duration/nonlinearity of the test-cases is increased?

- Is it possible to estimate the relative "multi-fidelity" contributions to the accuracy of the MFMC estimator? For example, is the overall accuracy governed more by the small number of high resolution runs, the large number of low resolution runs, or something in-between? Considering the more linear (or at least non-turbulent) nature of solutions studied, how would a conventional MC estimator compare if run only using lower-resolution simulations? In other words, is the good performance of the MFMC method due to the solution being well-resolved even on the coarser meshes, or is the multi-resolution hierarchy effective in estimating behaviour resolvable only at high-resolution? If it is the former, I wonder whether the problems studied are sufficiently nonlinear at the grid-scale. If it is the latter, this may be a nice result to highlight further.

- I believe the gradient terms in the shallow-water system (12) should be grad(1/2*|u|^2) + g*grad(h + h_b) rather than the grad(rho) included currently. Here p = rho_0*g*h is used to simplify the linear 1/rho_0 * grad(p) shallow-water pressure gradient, consistent with e.g. Ringler et al (2010).

- The SWE runtimes noted in 3.1.1 and 3.2.1 appear to be quite slow --- requiring 100's of seconds to advance a single time-step using relatively small O(<= 100,000) cell meshes? Are these runtimes for the full multi-day simulations instead, or for all ensemble members perhaps?

- While the MFMC methods presented here are clearly different in that they leverage varying resolution simulations, is it fair to compare against only the "historical" MC method, which is known to be uncompetitive in terms of efficiency? Significant work on alternative MC methods has been conducted by various authors in which a variety of accelerated techniques have been proposed. Are the large gains reported for MFMC expected to be replicated compared to e.g. MCMC (Markov Chain Monte Carlo) approaches more frequently used in climate model estimation?

Minor comments:

- The SOMA test case (Simulating Ocean Mesoscale Activity) typically refers to simulations using the multi-layer primitive equations, in which mesoscale eddies form due to 3d interactions between the momentum, density and forcing tendencies. In this shallow-water configuration with rho = const., it appears to be a wind-driven barotropic gyre that's studied instead, which is typically less turbulent, as per the smooth flow features in fig. 2. If so, it's suggested to label this test case as a wind-driven gyre.

- Wallis (2012) reference appears to be missing.

- ln 76: Is saying "no guesswork involved" too strong a statement? The systematic nature of the MFMC approach is attractive, but is it *the* provably optimal sampling strategy, or more of an effective heuristic?

- ln 72: ...also uses cheaper to obtain...

- fig. 3 labelling: left-right vs top-bottom.

- ln 308: Is this an expression for the free surface height or the layer thickness --- h appears to be thickness in the shallow-water system (12).

- The Gruber (2022) paper referenced here appears to be an arXiv preprint, that in-turn references this GMD submission??