Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-319-RC1, 2020

Geoscientific Model Development Discussions

Interactive comment on "Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz96 case study (v1.0)" *by* Stephan Rasp

Anonymous Referee #1

Received and published: 20 January 2020

This manuscript describes a method which hybridises machine learning with traditional numerical methods for simulating the Earth System that could avoid issues of numerical stability that impacted earlier attempts. The idea is to run a high-resolution model in parallel with the low-resolution, machine learning-hybridised model and to repeatedly retrain the machine learning algorithm based on how the high-resolution model evolves. It is essential to keep the high-resolution model synchronised to the low-resolution model through nudging, so that both are simultaneously situated in equivalent regions of their respective attractors. The idea is demonstrated in a toy model context, with linear regression and a neural net, and is also discussed for a real Earth System model. In the latter case, the author suggests how this technique could be applied when the

"high-resolution" model is literally a high-resolution configuration of the low-resolution model, but they also discuss how existing superparametrized models could be taken advantage of.

Although the technique is shown to work well for a toy model, it is unknown whether it can also work in a realistic setup. On this I have no strong intuition either way, and I can't think of any obvious reasons why it wouldn't work. I therefore enthusiastically recommend the manuscript for publication, subject to the corrections below which are mostly editorial.

1 Minor comments

- 32: "So far, three studies..." Chevallier at ECMWF conducted a number of studies on radiation parametrization 20 years ago which as far as I'm aware also implemented all three steps. See "Use of a neural-network-based long-wave radiative-transfer scheme in the ECMWF atmospheric model", Chevallier et al. , QJRMS (2000)
- 240 (whole paragraph): This doesn't necessarily warrant a change to the manuscript, but reading this paragraph made me think of the incremental 4D-Var algorithm used for data assimilation at various NWP centres, including ECMWF. There too, one needs to frequently interpolate between high- and low-resolution grids. This is for "online" use, not simply postprocessing. The innovations (observation - background) are computed using a high-resolution nonlinear model. These are then interpolated to a low-resolution grid so that the cost function gradient can be computed with low-resolution tangent-linear and adjoint models during minimisation. The resulting analysis increment is then upscaled for application to the high-resolution model, and so on. The fact that they successfully use high-/low-resolution models with a difference of $\sim$5 in grid-spacing makes me not so

concerned about the issues the author discusses. Though admittedly, I have no idea how they actually do the interpolation.

2 Technical corrections

- 3: "a machine learning algorithm"
- 4: "the trained algorithm was"
- 31: "and is coupled to the"
- 64: "second fix involves a multi-time-step"
- 77: "since they do not"
- Figure 2: $\Delta\phi_{\text{nudging}} = \Delta t_{\text{LR}} \Delta\phi / \tau_{\text{nudging}}$
- Equation 2: define overbar (especially since it is redefined in Algorithm 2)
- Equation 3: $(Y_{j+2,k} - Y_{j-1,k})$
- Algorithm 1: "Store $\Delta\mathbf{X}_{\text{HR-internal}}\cdots$" — I think these should be nonbold with a k subscript
- 150: "Δt_{HR} and Δt_{LR}" — I guess "Δt_{ML}" was a typo
- 157: "because the batch, which has size m", no?
- Figure 4: To make the reading more smooth, please refer to this figure in the text, for example on line 152
- 168: It took me a while to understand this sentence, until I replaced "learn" with "train". Then it made perfect sense. I recommend you make this change

- 168: "a much more complicated"
- 169: a clarification — "gradually approaches the one learned offline *using the training data generated with the correct parameters* F, h, c *and* b." I hope I've understood this correctly...
- 174: "issues that require a coupled"
- 181: this is, as far as I can see, the first mention of Algorithm 2. Please add a brief description, on line 178, for example
- Figure 5: Again, I couldn't find a reference to this in the text. Please add one
- 200: I had to look up the meaning of "SP-CAM". Please define the acronym here
- 217: "physical constraints"
- 225: I don't understand "offline performance". Do you mean the computational performance is slower because you need to add extra operations to ensure conservation? If so why does this affect offline but not online performance? Or do you mean the technique was only tested offline and was found to be less accurate (performance = accuracy then) than with no conservation checking?
- 293: I don't think the HR and LR acronyms are necessary here
- 297: "best guess of the truth would be a reanalysis"

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-319, 2020.