

EGU22-5766

<https://doi.org/10.5194/egusphere-egu22-5766>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Sampling strategies for data-driven parameterization of gravity wave momentum transport

Lucia Yang¹ and Edwin Gerber²

¹Courant Institute of Mathematical Sciences, New York University, New York, United States of America (minah.yang@nyu.edu)

²Courant Institute of Mathematical Sciences, New York University, New York, United States of America (epg2@nyu.edu)

With the goal of developing a data-driven parameterization of unresolved gravity waves (GW) momentum transport for use in general circulation models (GCMs), we investigate neural network architectures that emulate the Alexander-Dunkerton 1999 (AD99) scheme, an existing physics-based GW parameterization. We analyze the distribution of errors as functions of shear-related metrics in an effort to diagnose the disparity between online and offline performance of the trained emulators, and develop a sampling algorithm to treat biases on the tails of the distribution without adversely impacting mean performance.

It has been shown in previous efforts [1] that stellar offline performance does not necessarily guarantee adequate online performance, or even stability. Error analysis reveals that the majority of the samples are learned quickly, while some stubborn samples remain poorly represented. We find that the more error-prone samples are those with wind profiles that have large shears– this is consistent with physical intuition as gravity waves encounter a wider range of critical levels when experiencing large shear; therefore parameterizing gravity waves for these samples is a more difficult, complex task. To remedy this, we develop a sampling strategy that performs a parameterized histogram equalization, a concept borrowed from 1D optimal transport.

The sampling algorithm uses a linear mapping from the original histogram to a more uniform histogram parameterized by $t \in [0,1]$, where $t=0$ recovers the original distribution and $t=1$ enforces a completely uniform distribution. A given value t assigns each bin a new probability which we then use to sample from each bin. If the new probability is smaller than the original, then we invoke sampling without replacement, but limited to a reduced number consistent with the new probability. If the new probability is larger than the original, then we repeat all the samples in the bin up to some predetermined maximum repeat value (a threshold to avoid extreme oversampling at the tails). We optimize this sampling algorithm with respect to t , the maximum repeat value, and the number and distribution (uniform or not) of the histogram bins. The ideal combination of those parameters yields errors that are closer to a constant function of the shear metrics while maintaining high accuracy over the whole dataset. Although we study the performance of this algorithm in the context of training a gravity wave parameterization emulator, this strategy can be used for learning datasets with long tail distributions where the rare samples

are associated with low accuracy. Instances of this type of datasets are prevalent in earth system dynamics: launching of gravity waves, and extreme events like hurricanes, heat waves are just a few examples.

[1] Espinosa, Z. I., A. Sheshadri, G. R. Cain, E. P. Gerber, and K. J. DallaSanta, 2021: A Deep Learning Parameterization of Gravity Wave Drag Coupled to an Atmospheric Global Climate Model, *Geophys. Res. Lett.*, in review. [https://edwinpgerber.github.io/files/espinosa_et al-GRL-revised.pdf]