



Using machine learning to estimate Inertia-Gravity Waves properties from a coarse-grained description of the flow

Mozhgan Amirajadi (1,2), Riwal Plougonven (2), Albert Hertzog (2), Ali R. Mohebalhojeh (1), and Mohammad Mirzaei (1)

(1) University of Tehran, Institute of Geophysics, Space Physics, Tehran, Iran, Islamic Republic Of (m.amjadi@ut.ac.ir), (2) Laboratoire de Météorologie Dynamique du CNRS, École Polytechnique, Palaiseau, France (riwal.plougonven@lmd.polytechnique.fr)

This survey explores the use of machine learning to estimate a subgrid process from knowledge of a coarse-grained description of the flow in the troposphere. The process studied is atmospheric gravity waves which need to be parameterized in climate models. In these parameterizations, an essential yet poorly constrained component is the launch spectrum of non-orographic waves in the troposphere or lowermost stratosphere. We make use of the fact that current Numerical Weather Prediction models, with a spatial resolution of only a few tens of kilometers, describe well the variations of the gravity wave field, although it remains only partly unresolved (Jewtoukoff et al 2015). Hence the ERA5 reanalyses at full resolution provide us knowledge of the subgrid-scale process to estimate (the target): gravity waves at 100 hPa, quantified by the variance of vertical velocity or the associated momentum fluxes. We employ a 5-month data-set (May to September 2018), for a mid-latitude region of the Indian Ocean far from land masses and orographic obstacles. Explanatory variables are Low-resolution ERA5 data ($2.5^\circ \times 2.5^\circ$) through the troposphere (from 1000 to 100 hPa) including wind speed, relative vorticity, temperature gradients, and the Fraction of Cloud cover.

As a first exploration of the capacity to reconstruct information on gravity waves from a low-resolution description of the tropospheric flow, we have used Random Forests as a non-parametric statistical model which readily provides information on the relative importance of the different explanatory variables. The statistical model is trained and tested using 10-fold cross-validation. The gravity wave diagnostics are reconstructed from knowledge of the tropospheric flow with promising agreement: correlation coefficients between the target and the reconstructed time-series are 0.7 or better. We also performed another examination to train the model over the Indian Ocean and test it on another area over Indian and Southern Pacific Oceans. The statistical evaluation shows reasonable results again. This exploration provides insights for the development of gravity wave parameterizations: first, it indicates a lower bound on how much of the gravity wave signal can be reconstructed, given the low-resolution information on the tropospheric flow. Second, it ranks the tropospheric variables by order of relevance, revealing notably the importance of upper-level winds.