How Well Are Atmospheric Rivers Captured in Medium-Range Machine Learning Weather Predictions?

This study examines how well medium-range forecasts from machine learning weather prediction (MLWP) models represent atmospheric rivers (ARs). The study is motivated by the significant impacts of ARs in terms of flooding and contributions to water resources. There is also a pressing need to better understand the strengths and weaknesses of MLWP models as more of them are developed and put into use. ARs provide a useful test case for MLWP models because they are filamentary features exhibiting large gradients in thermodynamic and kinematic fields that may not be well represented within the machine learning loss function framework. We evaluate and compare forecasts s from the two leading MLWP models, Graphcast and PanguWeather, with two operational numerical weather prediction (NWP) models, NOAA's Global Forecast System (GFS) and ECMWF's Integrated Forecast System (IFS) for wintertime ARs in the North Pacific Ocean Basin over lead-times from one to ten days, using ECMWF ERA5 reanalysis data for verification. We employ a combination of grid point-based measures of forecast performance and novel, object-based methods. The object-based attributes evaluated include the spatial extent of an AR, the orientation angle and curvature of an AR, the magnitude of the IVT within an AR, the propagation speed of an AR, as well as the presence/proximity of an extratropical cyclone center. We find that MLWP models can represent ARs well, with some caveats. Firstly, the ARs produced by MLWP models tend to be slightly too weak and too small compared to both reanalysis and NWP models. And secondly, MLWP models completely fail to capture more ARs than their NWP counterparts.