Deep-Learning Augmented Data Assimilation: Reconstructing Missing Information With Convolutional Autoencoders

Yueya Wang*[1], Xiaoming Shi†[1,2], and Jimmy Fung[1,3]

[1]Division of Environment and Sustainability [HKUST] – Division of Environment and Sustainability, The Hong Kong University of Science and Technology Clear Water Bay, Kowloon, Hong Kong, Hong Kong SAR China

[2]Department of Civil and Environmental Engineering [HKUST] – Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology Clear Water Bay, Kowloon, Hong Kong, Hong Kong SAR China

[3]Department of Mathematics [HKUST] – Department of Mathematics, The Hong Kong University of Science and Technology Clear Water Bay, Kowloon, Hong Kong, Hong Kong SAR China

Abstract

Remote sensing data play a critical role in improving numerical weather prediction (NWP), especially for regions with no *in situ* observations available. However, Data voids often exist in physical space (e.g., no-precipitation region for radar reflectivity) which undermines the accuracy of initial conditions from data assimilation, negatively impacting NWP. Here, we use the barotropic vorticity equation (BVE) to demonstrate the great potential of deep learning (DL) in augmenting data assimilation, by reconstructing spatially complete observation fields from incomplete observations. The training data set for deep learning is a long-term BVE simulation on a coarse resolution (T63). By training a convolutional autoencoder (CAE), we obtained a deep-learning approximation of a 'reconstruction operator,' which maps spatially incomplete observation to reconstruction with full spatial coverage. The reconstruction operator was tested with the streamfunction generated from a high-resolution (T85) simulation and exhibited satisfactory performance. We further evaluated the impact of DL reconstruction on assimilation and forecast with four groups of observations from the T85 benchmark simulation, including ones with complete coverage (FullObs), incomplete observation (PartObs), DL reconstruction (RecObs), and a combination of PartObs and RecObs (MixObs). The data assimilation using RecObs and MixObs generates significantly more accurate initial conditions than those using PartObs, close to the accuracy when using FullObs. Predictions based on the initial conditions from the DL-augmented data assimilation are also considerably better than the predictions relying on digesting spatially incomplete observations only.

Keywords: Assimilation, EnSRF, Deep learning, Convolutional Autoencoders

*Speaker

†Corresponding author: shixm@ust.hk