EGU21-10475, updated on 09 Aug 2022
https://doi.org/10.5194/egusphere-egu21-10475
EGU General Assembly 2021

A machine learning approach to the observation operator for satellite radiance data assimilation

Jianyu Liang, Koji Terasaki, and Takemasa Miyoshi
RIKEN Center for Computational Science, Kobe, Japan

The 'observation operator' is essential in data assimilation (DA) to derive the model equivalent of the observations from the model variables. For satellite radiance observations, it is usually based on complex radiative transfer model (RTM) with a bias correction procedure. Therefore, it usually takes time to start using new satellite data after launching the satellites. Here we take advantage of the recent fast development of machine learning (ML) which is good at finding the complex relationships within data. ML can potentially be used as the 'observation operator' to reveal the relationships between the model variables and the observations without knowing their physical relationships. In this study, we test with the numerical weather prediction system composed of the Nonhydrostatic Icosahedral Atmospheric Model (NICAM) and the Local Ensemble Transform Kalman Filter (LETKF). We focus on the satellite microwave brightness temperature (BT) from the Advanced Microwave Sounding Unit-A (AMSU-A). Conventional observations and AMSU-A data were assimilated every 6 hours. The reference DA system employed the observation operator based on the RTTOV and an online bias correction method.

We used this reference system to generate 1-month data to train the machine learning model. Since the reference system includes running a physically-based RTM, we implicitly used the information from RTM for training the ML model in this study, although in our future research we will explore methods without the use of RTM. The machine learning model is artificial neural networks with 5 fully connected layers. The input of the ML model includes the NICAM model variables and predictors for bias correction, and the output of the ML model is the corresponding satellite BT in 3 channels from 5 satellites. Next, we ran the DA cycle for the same month the following year to test the performance of the ML model. Two experiments were conducted. The control experiment (CTRL) was performed with the reference system. In the test experiment (TEST), the ML model was used as the observation operator and there is no separate bias correction procedure since the training includes biased differences between the model and observation. The results showed no significant bias of the simulated BT by the ML model. Using the ECMWF global atmospheric reanalysis (ERA-interim) as a benchmark to evaluate the analysis accuracy, the global-mean RMSE, bias, and ensemble spread for temperature in TEST are 2% higher, 4% higher, and 1% lower respectively than those in CTRL. The result is encouraging since our ML can emulate the RTM. The limitation of our study is that we rely on the physically-based RTM in the reference DA system, which is used for training the ML model. This is the first result and still preliminary. We are currently considering other methods to train the ML model without using the RTM at all.