Fast Machine Learning for Science
Imperial College London
Real-time and accelerated ML for fundamental sciences
25-28 September 2023

Contribution ID: **6** Type: **Standard Talk**

A hybrid data-driven and data assimilation operational model for long term spatiotemporal forecasting: Global and regional PM2.5 forecasting

Tuesday 26 September 2023 16:45 (15 minutes)

Accurate and reliable long-term operational forecasting is of paramount importance in numerous domains, including weather prediction, environmental monitoring, early warning of hazards, and decision-making processes. Spatiotemporal forecasting involves generating temporal forecasts for system state variables across spatial regions. Data-driven methods such as Convolutional Long Short-Term Memory (ConvLSTM) are effective in capturing both spatial and temporal correlations, but they suffer from error accumulation and accuracy loss in long-term forecasting due to the nonlinearity and uncertainty in physical processes. To address this issue, we propose to combine data-driven and data assimilation methods for explainable long-term operational forecasting. Data assimilation is used for updating the predictive results by merging measurements into operational modelling. However, it is computationally impossible to conduct online data assimilation and provide real-time reanalysis data for large-scale problems. Recent advancements in neural network techniques have offered new opportunities for improving data assimilation and operational forecasting. Neural networks excel at capturing nonlinear relationships and are computationally efficient at leveraging large-scale datasets, making them particularly well-suited for handling the intricate dynamics of operational forecasting systems. Here we propose a hybrid-ConvLSTM and DA model for accurate and efficient long term forecasting. This proposed hybrid ConvLSTM-DA method is demonstrated through hourly/daily PM2.5 forecasting globally and regionally (in China), which is a challenging task due to the complexity of geological and meteorological conditions in the region, the need for high-resolution forecasting over a large study area, and the scarcity of observations. The results show that the ConvLSTM-DA method outperforms conventional methods and can provide satisfactory hourly PM2.5 forecasting in the following 27 days with spatially averaged RMSE below 40 ug/m¬3 and correlation coefficient (R) above 0.7. In addition, the ConvLSTM-DA method shows a substantial reduction in CPU time when compared to the commonly used NAQPMS model, up to three orders of magnitude. Overall, the use of data driven modelling provides efficient prediction and also speeds up data assimilation. This hybrid ConvLSTM-DA is a novel operational forecasting technique for spatiotemporal forecasting and used in real spatiotemporal forecasting for the first time.

Primary authors: Ms CAI, Shengjuan (Imperial College London); Dr FANG, Fangxin (Imperial College London); Dr ZHU, Jiang (Institute of Atmospheric and Physics); Dr WANG, Yanghua (Imperial College London)

Presenter: Dr FANG, Fangxin (Imperial College London)

Session Classification: Contributed Talks

Track Classification: Contributed Talks