IMPERIAL

Spatiotemporal Forecasting via Machine Learning and Data Assimilation: Applications to Pollution and Flood Risk

S. Cai, M. Cheng,, J. Zhou, L. Li, J. Zheng, C. C. Pain, Y. Wang, **F. Fang*** (Imperial College London),
J. Zhu. X. Tan (IAP, China),
I.M. Navon (FSU,USA),
V. Henri Peuch, M. Alexe (ECMWF)
* f.fang@imperial.ac.uk

Case Study 1: Flooding Forecasting

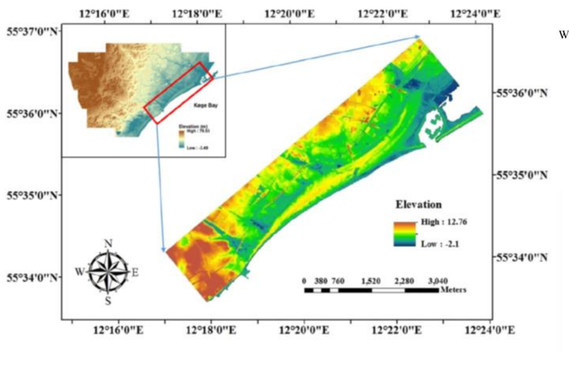

Fig. 1. Study area in Greve, Denmark, affected by flood disasters induced by extreme sea-level events along its coast.

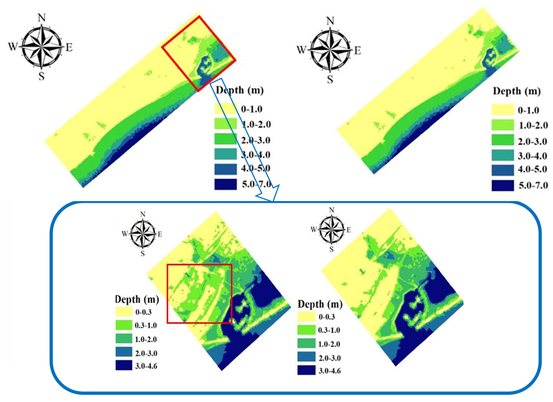

Fig. 2. Comparison of the spatial distribution of predictive water depth obtained from the DCGAN (left) and MIKE 3 FM (right).

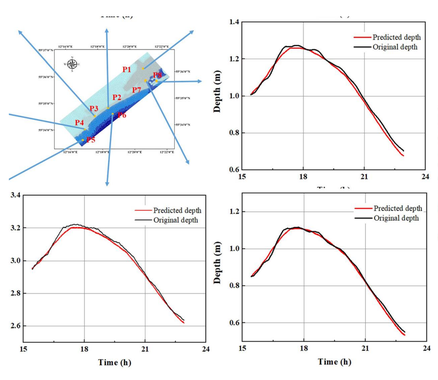

Fig. 3. Comparison of the predicted water depth from the DCGAN and observations at monitors

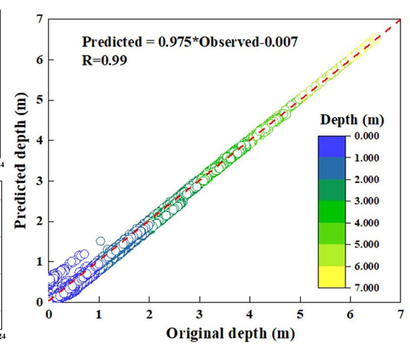

Fig. 4. The correlation coefficients of water depth solutions between the DCGAN and MIKE 3 FM.

Case description: Flooding on 13 Oct. 1760, The area is 2.3 x 7.5 km. A 100-year return period extreme event is considered. The training period [1, 15] h and the predictive period is [15, 24] h.

Datasets: Simulation results by running the high fidelity model (MIKE 3 FM).

ML model: DCGAN-Deep convolutional generative adversarial network.

Case study 2: PM2.5 Regional Forecasting in China

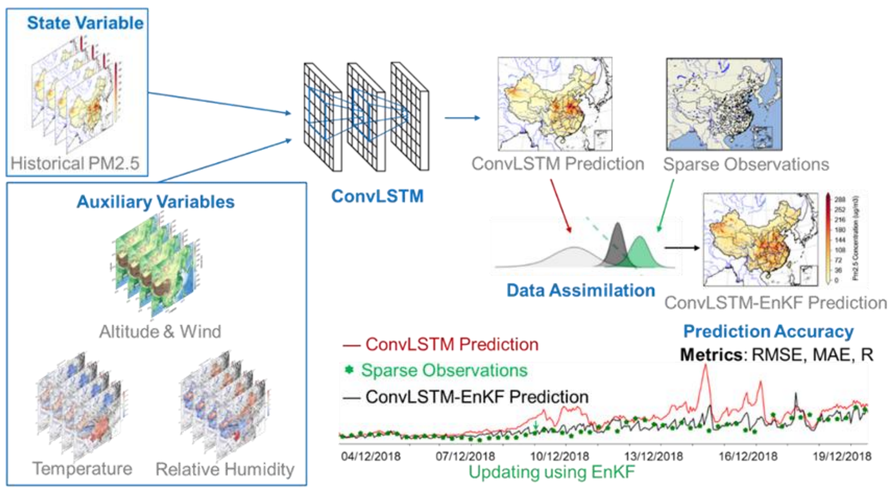

Fig. 5. Illustration of the proposed hybrid ConvLSTM-EnKF method for operational PM2.5 concentration forecasting in China. In this case study, PM2.5 concentrations are used as the state variable, and topographical and meteorological fields, including U- and V-components of the wind field, temperature, relative humidity, and altitude, are used as auxiliary variables. Forecasting is conducted using a trained three-layer ConvLSTM model while EnKF is used to improve the forecast accuracy by assimilating the monitoring data into the trained ConvLSTM model.

Datasets:

- **Reanalysis data** (a high spatial resolution of15km x 15km, 339x430 nodes, a temporal resolution of1h) - the Nested Air Quality Prediction Modeling Systems (NAQPMS) combined with EnKF. Meteorological conditions come from WRF
- **Observations:**1436 air quality monitoring stations in China.

Training (90%) + validation (10%) : 2013-2017, Predicting: 2018; Data assimilation frequency: 6h.

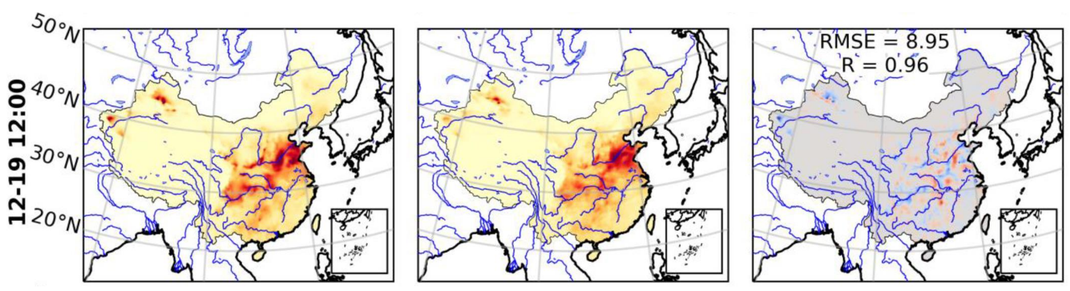

Fig. 6. Comparison of the PM2.5 concentrations obtained from the reanalysis dataset (left), hourly ConvLSTM-EnKF forecasts (middle), and differences between them (right) at12:00 UTC+08:00 on Dec.19, 2018.

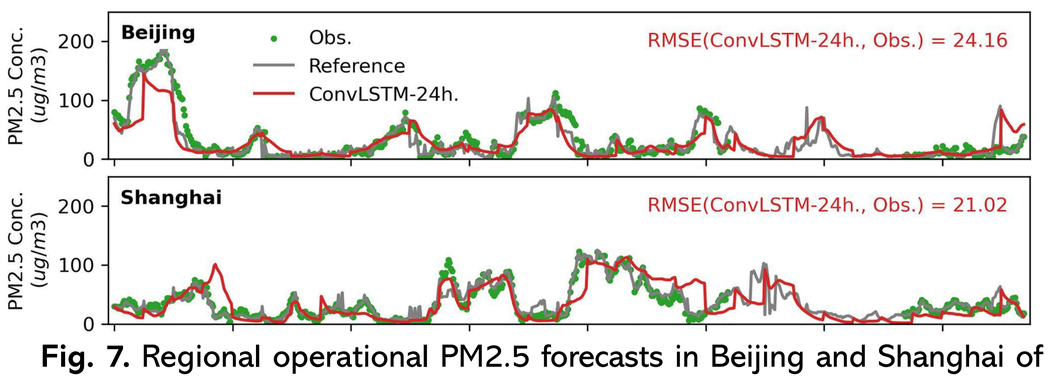

Fig. 7. Regional operational PM2.5 forecasts in Beijing and Shanghai of China during December 1-21, 2018. Data assimilation (DA) is conducted with 100 ensemble members every 24 hours. RMSE (ConvLSTM-EnKF., Obs.) denotes the temporally averaged RMSE between the ConvLSTM-EnKF forecasts and observations for each city.

Case study 3: PM2.5 Global Forecasting

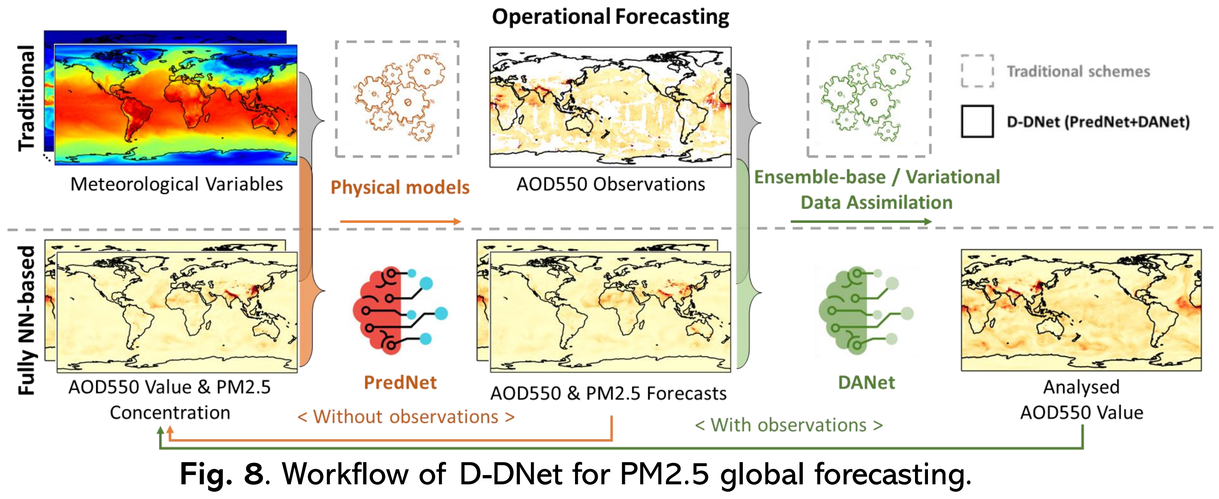

Fig. 8. Workflow of D-DNet for PM2.5 global forecasting.

Datasets:

- **EAC4 reanalysis data** (80km x 80km, 60000 nodes, a temporal resolution of 3 h): Integrating satellite data and physical simulation (0.4° x 0.4° and a temporal resolution of 1-hour)- advanced Copernicus Atmosphere Monitoring Service (CAMS), operated by the European Centre for Medium-Range Weather Forecasts (ECMWF).
- **Satellite observations:** MOD08 and MYD08, both part of NASA's MODIS (Moderate Resolution Imaging Spectroradiometer) data collection , Sparse data: 3258.
- **CAMS-GLOB-ANT emission dataset (**2000-2023 period at a spatial resolution of 0.1° x 0.1°): monthly global emissions data for 36 compounds. The compounds include key air pollutants and greenhouse gases across 17 sectors.

Training (90%) + validation (10%) : 2013-2018, Predicting: 2019; DA frequency: 12h.

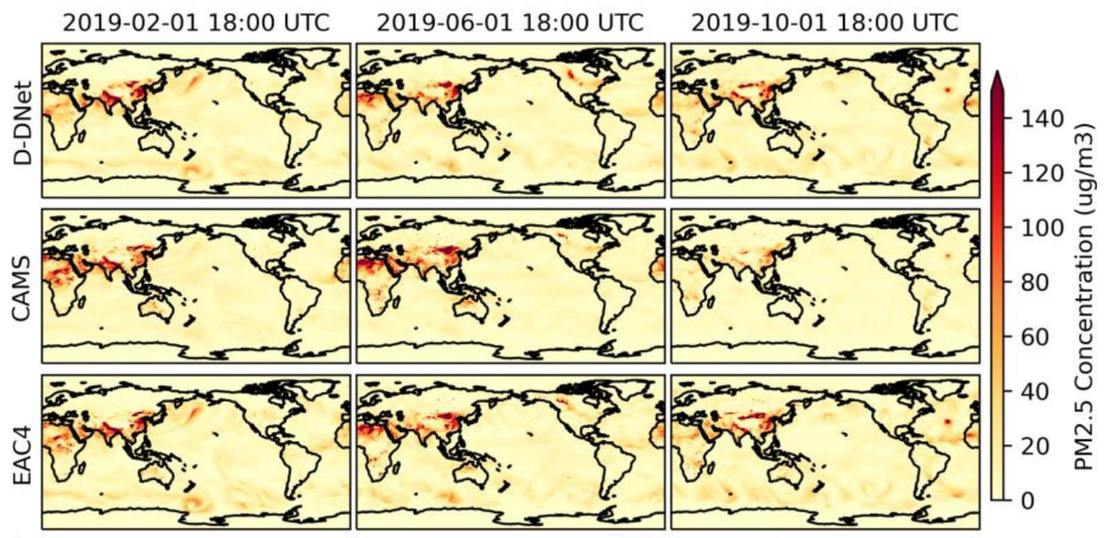

Fig. 9. Spatial distribution of PM2.5 concentrations from the D-DNet forecasts, CAMS 4D-Var forecasts, and EAC4 reanalysis ("ground truth") on 1 Feb. 1 June, and 1 Oct. 2019.

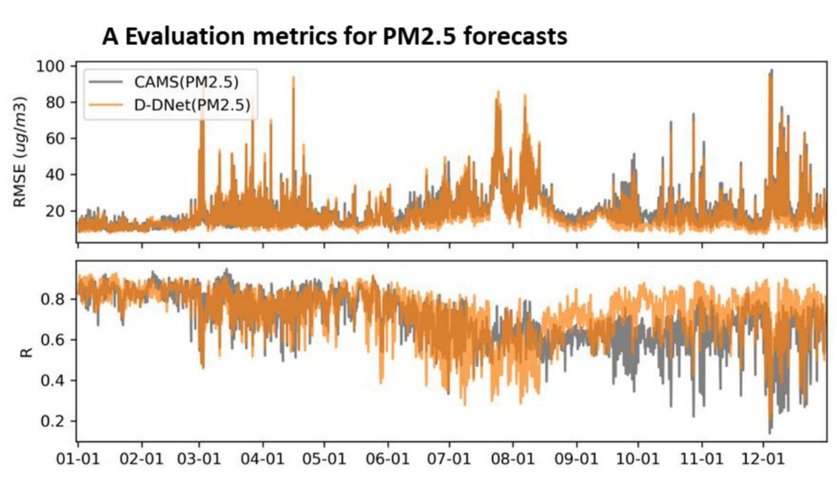

Fig. 10. Global operational forecast evaluation metrics (RMSE and R) compared to the "ground truth" in 2019. Operational forecasts from D-DNet are compared with those from the CAMS 4D-Var model, a renowned system for global operational atmospheric composition forecasting. Both D-DNet and the CAMS 4D-Var model deliver operational forecasts with a DA frequency of 12 hours.

References: 1. Cheng, M., Fang, F., Navon, I.M. and Pain, C.C., 2021. A real-time flow forecasting with deep convolutional generative adversarial network: Application to flooding event in Denmark. *Physics of Fluids*, *33*(5).
1. Cai, S., Fang, F., Tang, X., Zhu, J. and Wang, Y., 2024. A hybrid data-driven and data assimilation method for spatiotemporal forecasting: PM2. 5 forecasting in China. *Journal of Advances in Modeling Earth Systems*, *16*(2).
2. Cai, S., Fang, F., Peuch, V.H., Alexe, M., Navon, I.M. and Wang, Y., 2024. Advancing operational PM2. 5 forecasting with dual deep neural networks (D-DNet). *arXiv preprint arXiv:2406.19154*.

Imperial College London ECMWF EPSRC Engineering and Physical Sciences Research Council