Imperial College London

A reduced order deep data assimilation for CFD forecast

César Quilodrán Casas, Rossella Arcucci, and Yi-Ke Guo

Data Science Institute, Department of Computing, Imperial College London

The Problem

In the last few decades, computers have permitted us to develop consistently more accurate numerical models. To improve these, Data Assimilation (DA) is a technique that allows the input of real observations into these models to correct the forecast from the numerical solution. However, the high dimensionality of the models makes this a computationally expensive task.
Here, we present a Reduced Order Deep Data Assimilation (RODDA) model to tackle the computational cost of the forecast correction.

The Solution

The RODDA workflow (Fig. 1) can be divided in three parts:

Step 1: Data Assimilation
The numerical solution of a pollutant tracer from a CFD solver (Thetis and Fluidity) is assimilated with observations using 3Dvar. This is repeated sequentially. At each iteration, the background state (u^M) from the CFD and the Data Assimilated state (u^{DA}) are saved to be used in the next steps.

Step 2: Dimension Reduction
We reduce the dimensions of u^M and u^{DA} by several orders of magnitude using Principal Components Analysis (PCA). PCA is a unsupervised learning method that simplifies high-dimensional data by transforming it into fewer dimensions.

Step 3: Long Short-term Memory (LSTM) Neural Network (NN)
The reduced order versions of u^M and u^{DA} are used as input and output of the LSTM NN (Fig. 2), respectively. This means that the data assimilated state will be predicted from the background state and that the forecast is corrected without the need of performing DA. Additionally, future observations are implicit in the model. The LSTM prediction is now u^{RODDA}.

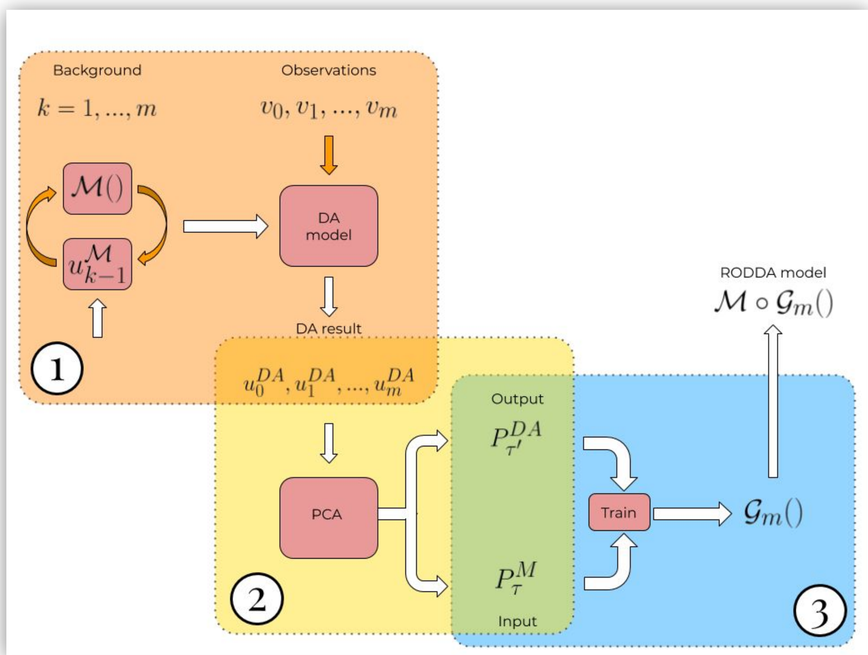

Fig. 1: Schematic of the Reduced Order Deep Data Assimilation workflow

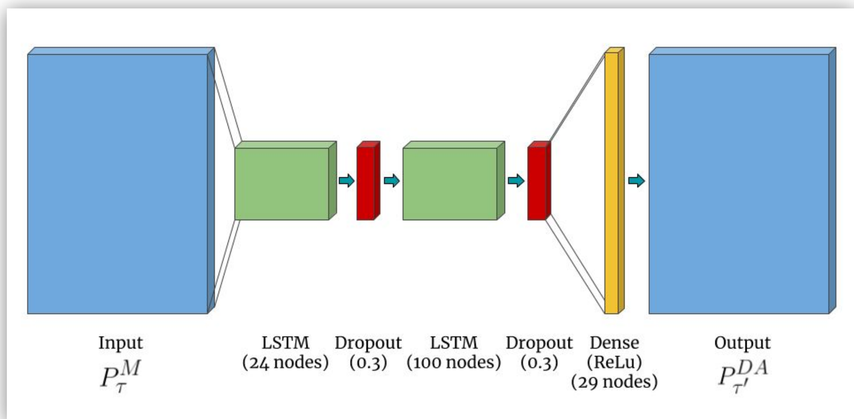

Fig. 2: Schematic of the LSTM NN

Results

RODDA was applied to two different simulations: a flow past the cylinder 2d simulation with Re = 3000 (Fig. 3), and an air pollution CFD of South London.

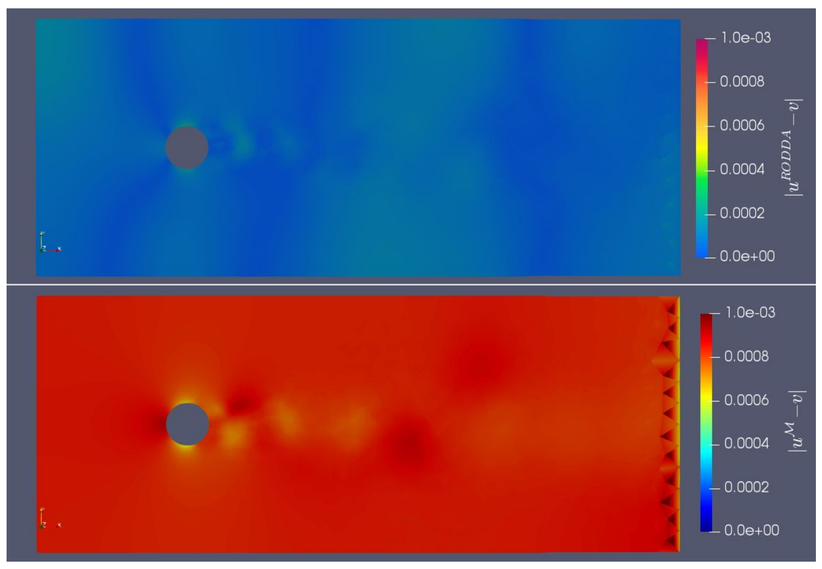

Fig. 3: Elevation absolute error between Model and RODDA solutions with respect to observations with Re = 5000.

RODDA, once trained, produces a corrected forecast 80 times faster than running a new simulation in Fluidity + DA with similar accuracy, and sometimes outperforming it, as shown in Fig. 3 and Fig.4.

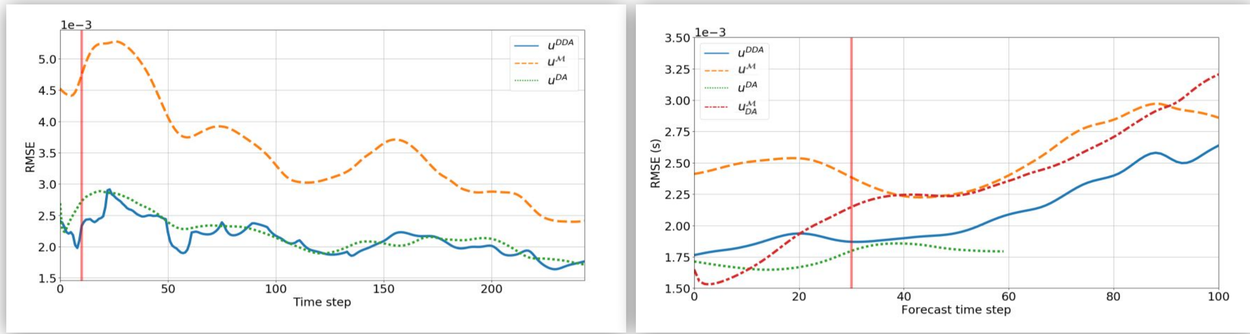

Fig. 4: Root mean squared error of uM, uDA, and uRODDA against observations.
Left: Training data, Right: Test data

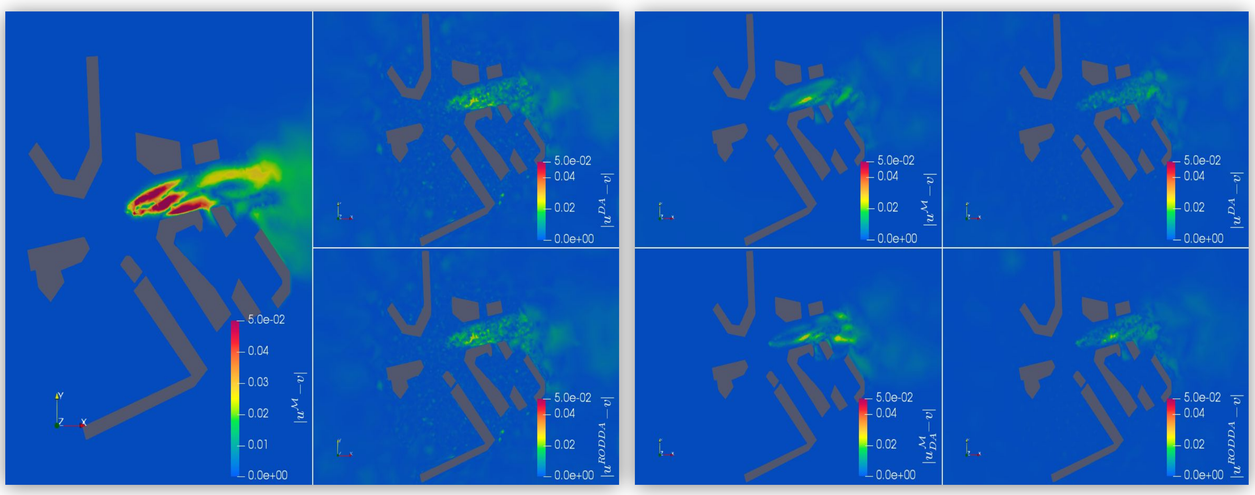

Fig. 5 : Absolute Errors at z=10m of the pollutant tracer in the training dataset.
Left: at t=10 (training), Right: at t=30 (test)

Summary and Future Work

The RODDA framework was presented showing the integration of machine learning, dimension reduction techniques and data assimilation. The framework was validated using two CFD simulations as observations. It is shown that RODDA is able to replace the traditional methods for forecasts without the assumptions and computational costs, once the LSTM network is trained. The error correction by RODDA during the forecast outperforms the forecast by the CFD simulations. Additionally, RODDA is 80 times faster than any of the other methods presented.

RODDA is not exclusive to these solutions and the workflow could be applied to different physical models where sufficient temporal data is available. Future work includes to scale RODDA into bigger domains in order to study the interaction between pollutant exposure and the effects on lungs on an individual as part of the upcoming multidisciplinary EPSRC Inhale project.

EPSRC Engineering and Physical Sciences Research Council

CMPH Centre for Mathematics of Precision Healthcare

MAGIC Envisaging a world with greener cities

INHALE

References

J. Zhu, S. Hu, R. Arcucci, C. Xu, J. Zhu, Y.-k. Guo, Model error correction in data assimilation by integrating neural networks, Big Data Mining and Analytics 2 (2019) 83–91.