Imperial College London

Data Learning:
Integrating Data Assimilation and Machine Learning

Dr. Rossella Arcucci

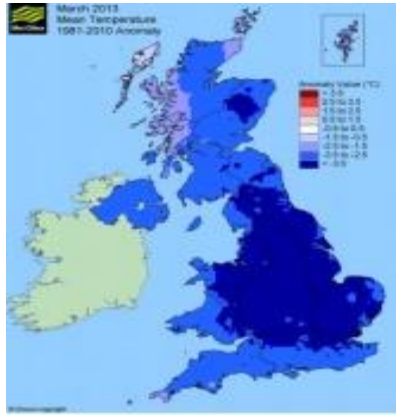
Data Learning working group,
Data Science Institute, Imperial College London
Leonardo Centre, Imperial College Business School

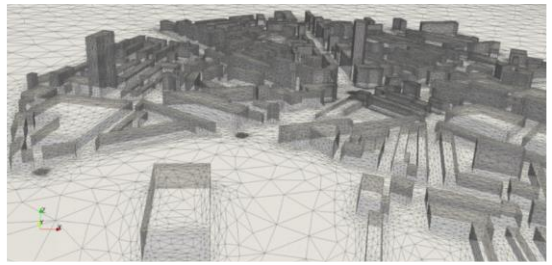
17 May 2021 - ECMWF/OceanPredict workshop on Advances in Ocean Data Assimilation

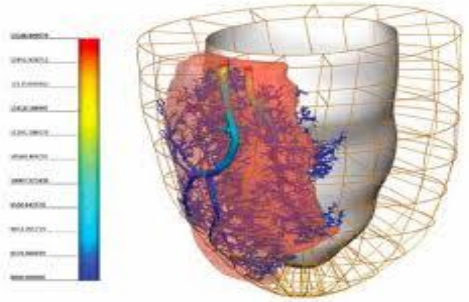
Data Science Institute

… the era of the data!

High resolution Models…

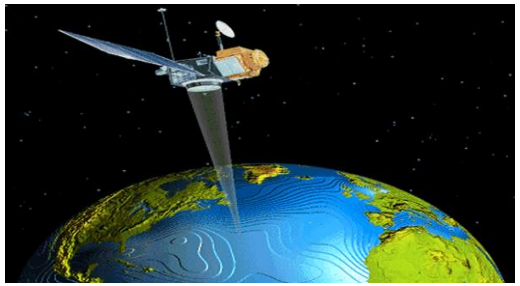


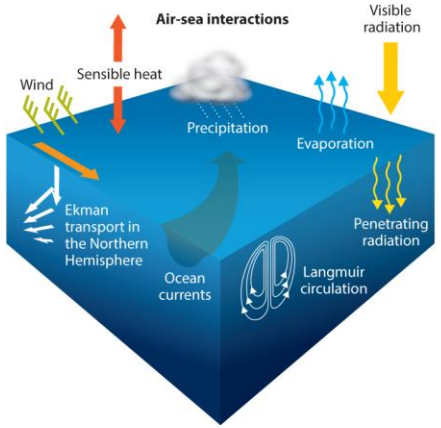


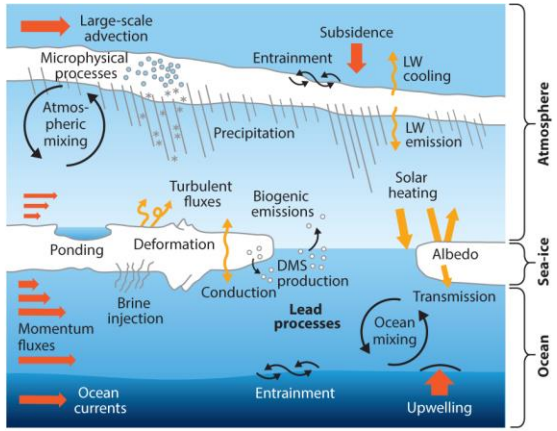


facebook

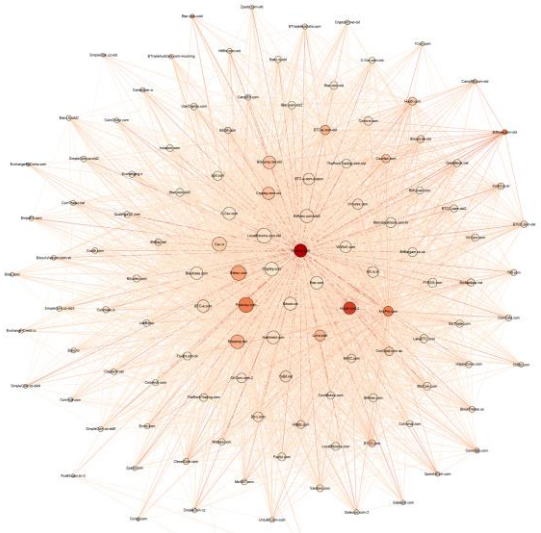
Coupled models…

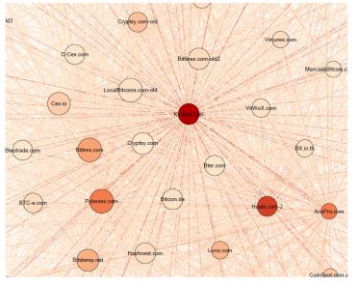


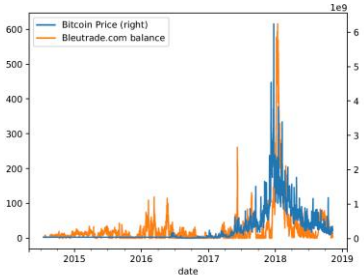




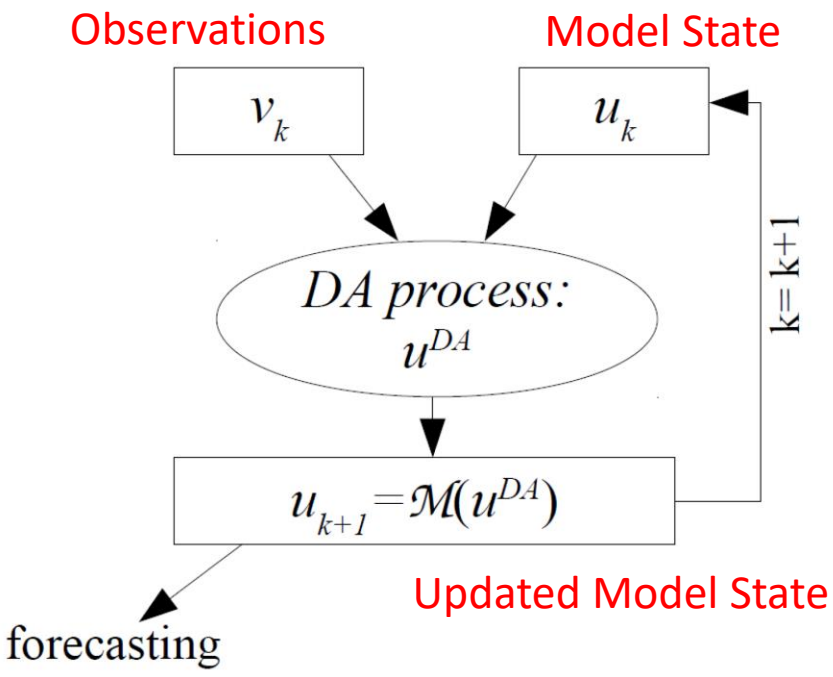
Real observations…







Content



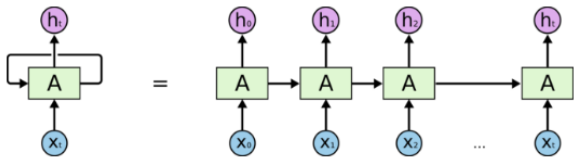
To have the expected benefit from data assimilation, we need:

- **Good efficiency** (accelerate the execution time)
- **Good accuracy** (reduce the errors propagation in the models)
- **Real world scenarios** (develop models that can be used in real world scenarios – face **Big Data** problems)
- **Good quality data** (optimal placements/locations)

Artificial Neural Networks at the service of DA

Conventional methods for DA have increased in sophistication to better fit their application requirements and circumvent their implementation issues. Nevertheless, these approaches are incapable of fully overcoming some unrealistic assumptions.

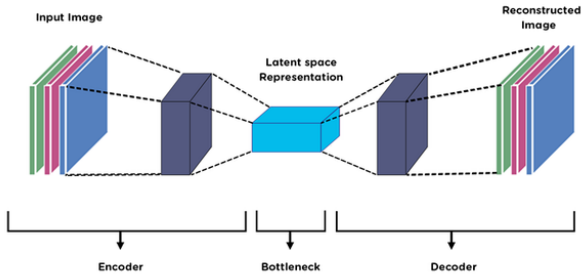
- **Nonlinearities:** Can we **replace** part of the **DA** process (or the entire DA process) **by a ML** technology?
- **Heavy background error covariance matrices**: can we compute them in the **latent space** of an Auto-Encoder?
- **Computationally expensive CFD software and difficulties to compute the adjoints of the models**: can we use **surrogate models** in the DA process?



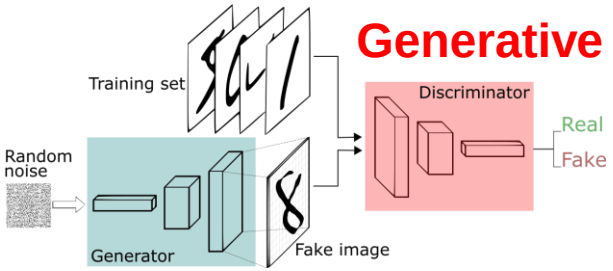
An unrolled recurrent neural network.

Recurrent NN

Encoder-Decoder



Generative Adversarial Networks



Data Learning

All the models and the technologies which have been developed at DataLearning working group are completely general and applied to a lot of different real world applications.

Grants:

EPSRC

UKRI Engineering and Physical Sciences Research Council

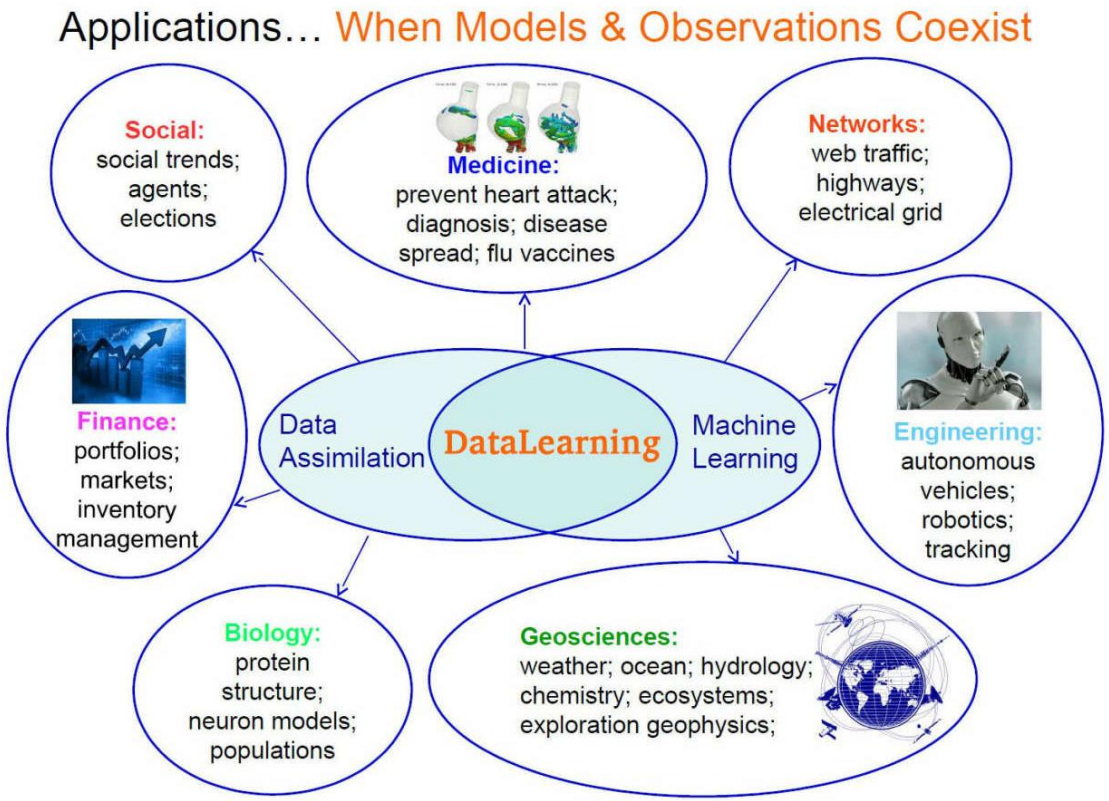
INHALE

MAGIC
Envisaging a world with greener cities

PREMIERE

RELIANT

LEVERHULME
Centre for Wildfires, Environment and Society



Our Academic Collaborations:

UNIVERSITY OF CAMBRIDGE

UNIVERSITY OF SURREY

University of Glasgow

THE UNIVERSITY OF EDINBURGH

The Alan Turing Institute

UCL

KING'S College LONDON

https://www.imperial.ac.uk/data-science/research/research-themes/datalearning/

Data Assimilation

Algorithm 1: DA

Input: for $k = 0, \ldots, m$ temporal steps: observations y_k, matrices R_k, model M, function H_k, background u_0, historical data $S = \{u_j\}_j = 1, \ldots, n$,

Compute B_0 from u_0 and S

Initialize iteration $k = 1$

while $k < n$ **do**

  Compute $u_k = M\ u_{k-1}$

  Compute B_k from u_k and S

  Compute

$$u_k^{DA} = argmin_u \left\{ \|u - u_k\|_{B_k^{-1}} + \|v_k - H_k u\|_{R_k^{-1}} \right\}$$

  Count up k for the next iteration

end

Output: u^{DA}

Data Assimilation with Machine Learning

Algorithm 1: DA

Input: for $k = 0, \ldots, m$ temporal steps: observations y_k, matrices R_k, model M, function H_k, background u_0, historical data $S = \{u_j\}_j = 1, \ldots, n$,

Compute B_0 from u_0 and S
Initialize iteration $k = 1$
while $k < n$ **do**
  Compute $u_k = \boxed{M} u_{k-1}$
  Compute $\boxed{B_k}$ from u_k and S
  Compute

$$u_k^{DA} = argmin_u \left\{ \|u - u_k\|_{B_k^{-1}} + \|v_k - H_k u\|_{R_k^{-1}} \right\}$$

  Count up k for the next iteration
end
Output: u^{DA}

efficiency

machine learning to accurately **model the dynamic systems** reducing the CPU time

accuracy

machine learning models to **reduce the errors** in the assimilated data

Learning from real data to make CFD predictions more accurate and realistic

*CFD (Computational Fluid Dynamic)

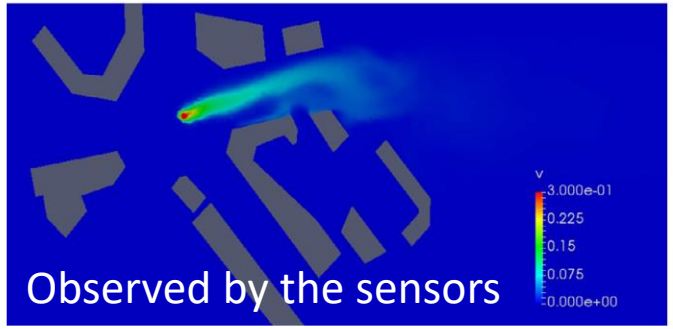
$$J(\mathbf{u}) = \alpha \|\mathbf{u} - \mathbf{u}_0\|^2_{\mathbf{B}^{-1}} + \|\mathbf{G}\mathbf{u} - \mathbf{v}\|^2_{\mathbf{R}^{-1}}$$ ← DA function

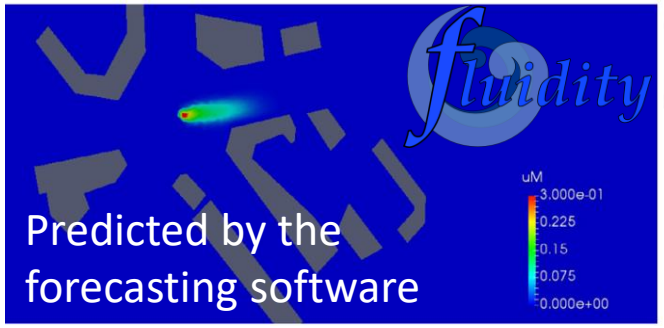
3DVar in the control space $\mathbf{w} = \mathbf{V}^{+}\delta\mathbf{u}$, $\mathbf{B} = \mathbf{V}\mathbf{V}^{T}$ ← Reduced space, TSVD

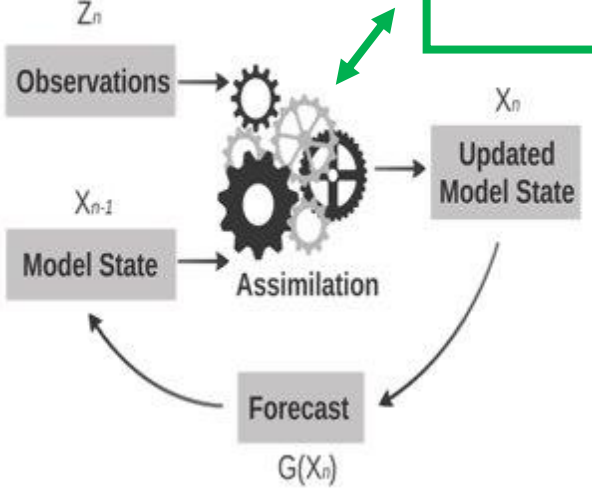
$$\mathbf{w}^{DA} = argmin_{\mathbf{w}\in\Re^{NP\times N}} J(\mathbf{w})$$

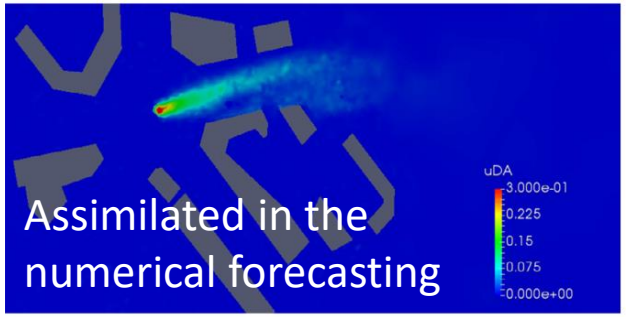
with

$$J(\mathbf{w}) = \frac{1}{2}\alpha\mathbf{w}^T\mathbf{w} + \frac{1}{2}(\mathbf{G}\mathbf{V}\mathbf{w} - \mathbf{d})^T\mathbf{R}^{-1}(\mathbf{G}\mathbf{V}\mathbf{w} - \mathbf{d})$$









EPSRC

INHALE

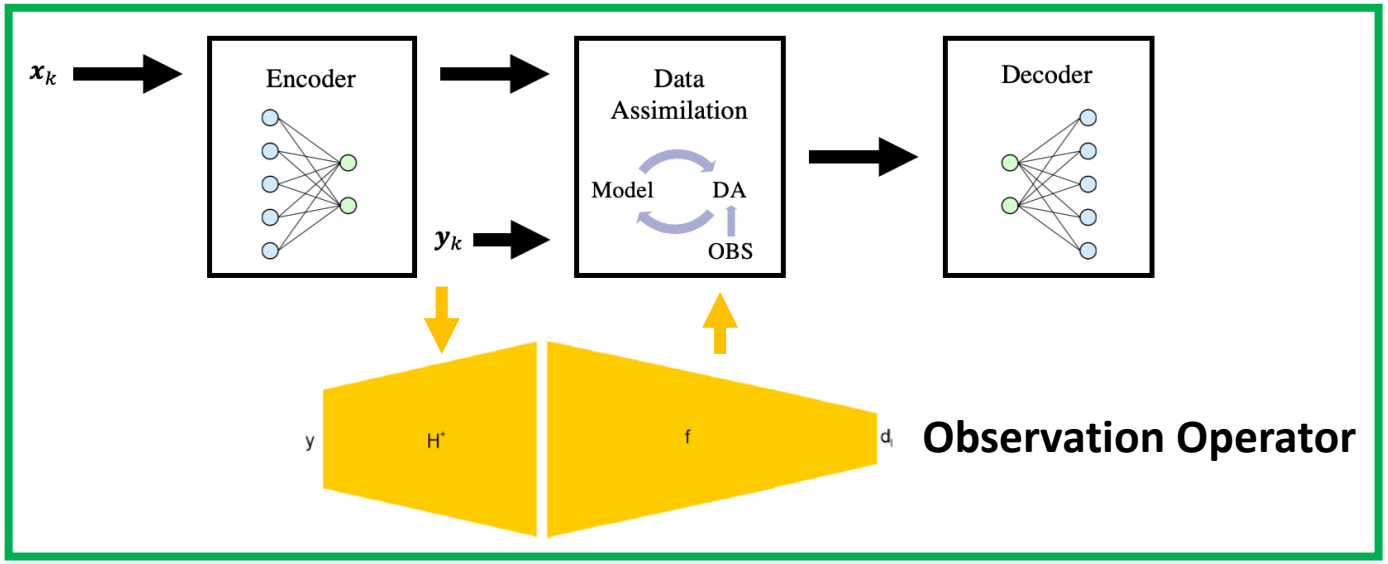
MAGIC
Envisaging a world with greener cities

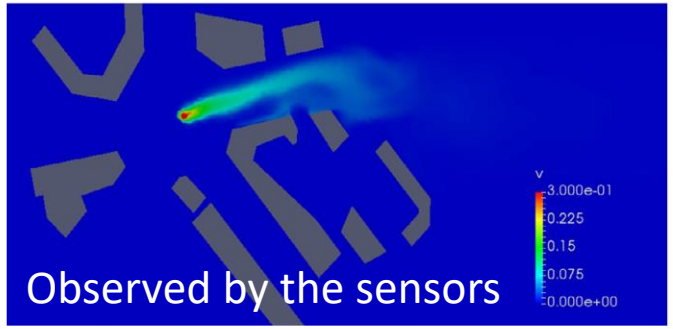
[*] R. Arcucci, L. Mottet, C. Pain and Y. Guo - **Optimal reduced space for Variational Data Assimilation** -Journal of Computational Physics

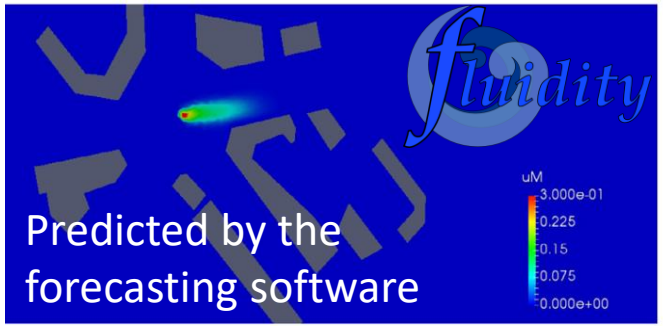
[**] R. Arcucci, C. Pain, Y. Guo, **Effective variational data assimilation in air-pollution prediction**, Big Data Mining and Analytics

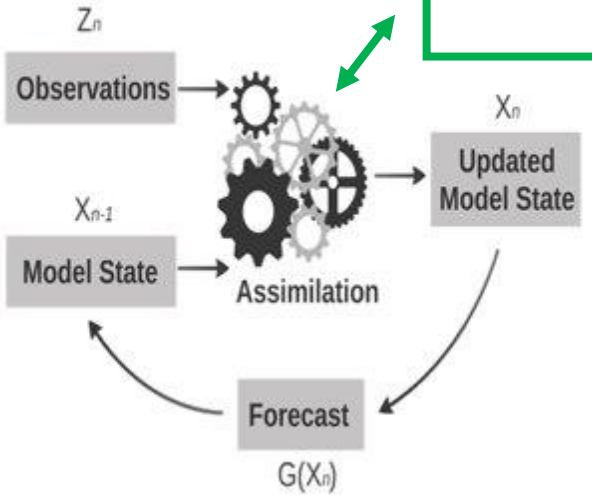
Learning from real data to make CFD predictions more accurate and realistic

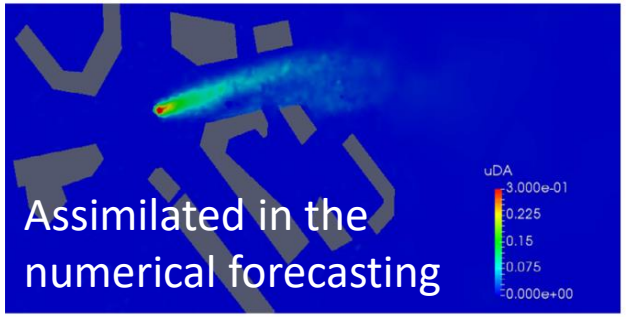
*CFD (Computational Fluid Dynamic)











EPSRC

INHALE

MAGIC
Envisaging a world with greener cities

[*] R. Arcucci, L. Mottet, C. Pain and Y. Guo - **Optimal reduced space for Variational Data Assimilation** -Journal of Computational Physics

[**] R. Arcucci, C. Pain, Y. Guo, **Effective variational data assimilation in air-pollution prediction**, Big Data Mining and Analytics

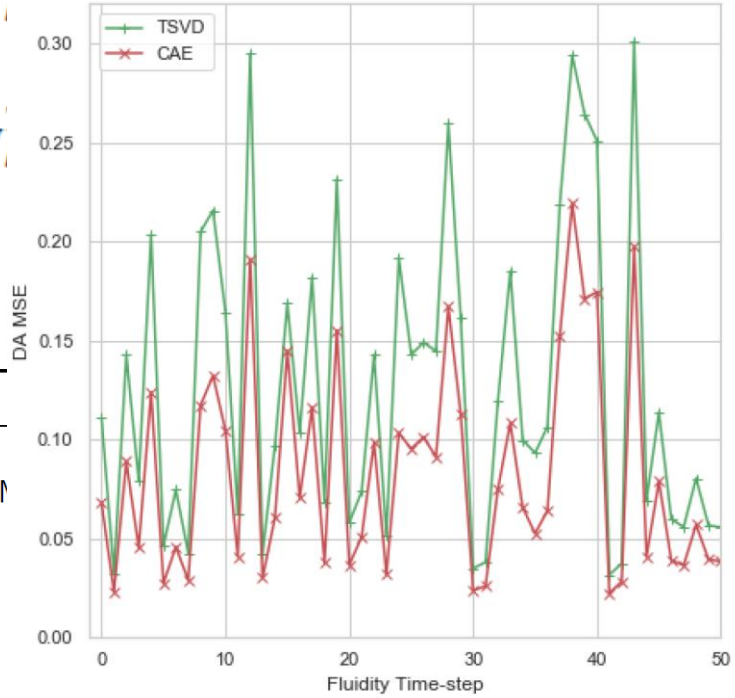
[***] Mack, J., Arcucci, R., Molina-Solana, M., & Guo, Y. K. (2020). **Attention-based Convolutional Autoencoders for 3D-Variational Data Assimilation**. *Computer Methods in Applied Mechanics and Engineering*

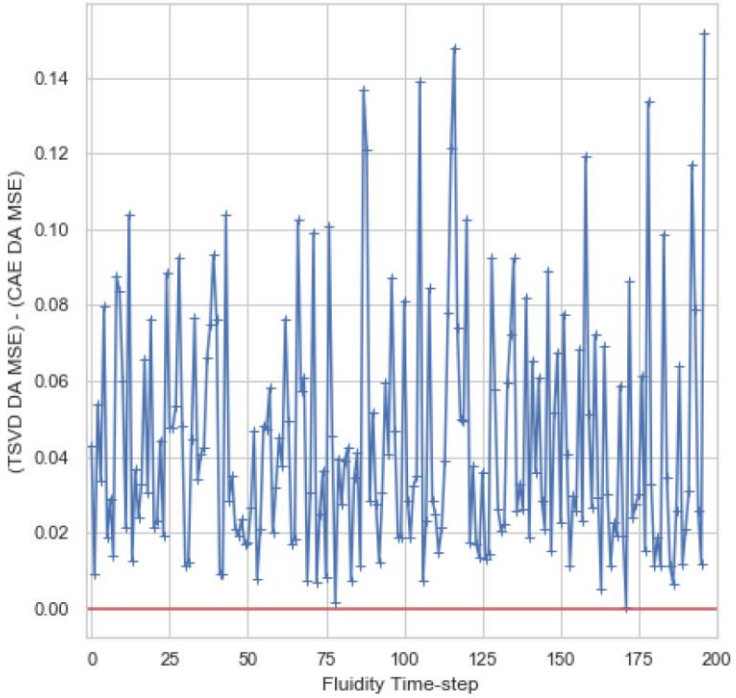
Data Assimilation in a latent space

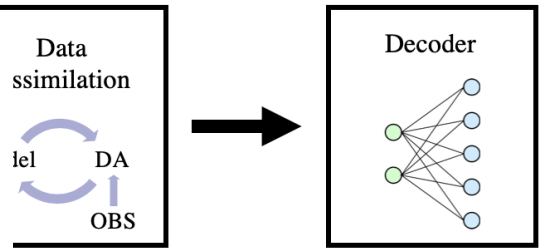
***with Julian Mack - 2019**

3DVar in the latent space

$$J(\mathbf{w}_l) = \frac{1}{2}\mathbf{w}$$







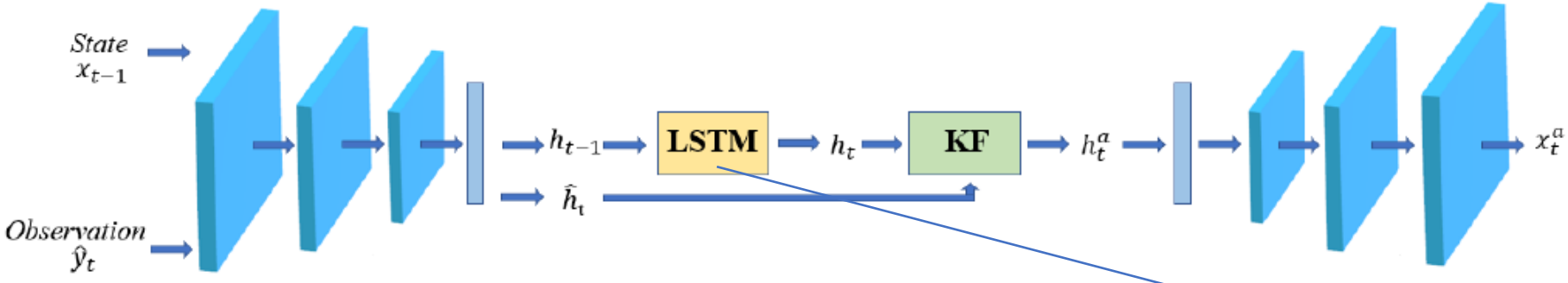
Model
Backbone
ResNeXt3-27-1-vanilla+CBAM
RDB3-27-4-vanilla+CBAM
RAB-4-NeXt
GRDN-NeXt+CBAM
Tucodec-NeXt
Tucodec-vanilla

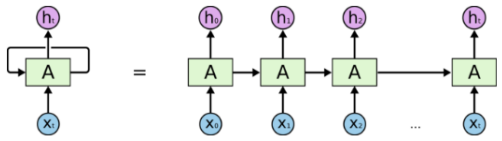
	MSE	Execution Time (s)
	1.0001	-
	0.1270	1.8597
	0.1270	0.2627
	0.1334	0.0443
	0.1680	0.0390
eXt	0.0787	0.0537

[*] Mack, J., Arcucci, R., Molina-Solana, M., & Guo, Y. K. (2020). **Attention-based Convolutional Autoencoders for 3D-Variational Data Assimilation.** *Computer Methods in Applied Mechanics and Engineering*, *372*, 113291.

Latent Assimilation

***with Maddalena Amendola - 2020**





An unrolled recurrent neural network.

Time Update ("Predict")

(1) Project the state ahead

$$h_{t+1} = l(h_{t,q})$$

Measurement Update ("Correct")

(1) Compute the Kalman Gain

$$\widehat{K}_{t+1} = \widehat{Q}\widehat{H}^T(\widehat{H}\widehat{Q}\widehat{H}^T + \widehat{R}_{t+1})^{-1}$$

(2) Update estimate with measurement

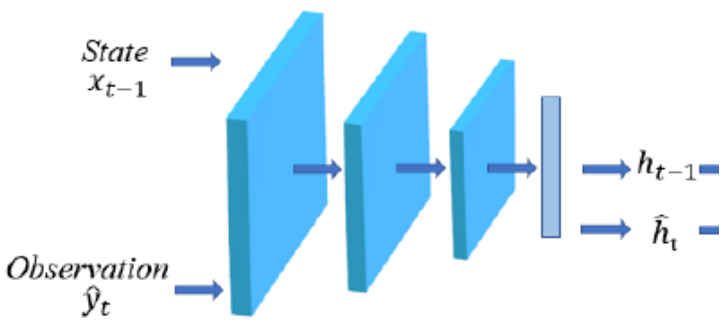
$$h_{t+1}^a = h_{t+1} + \widehat{K}_{t+1}(\widehat{h}_{t+1} - \widehat{H}\widehat{h}_{t+1})$$

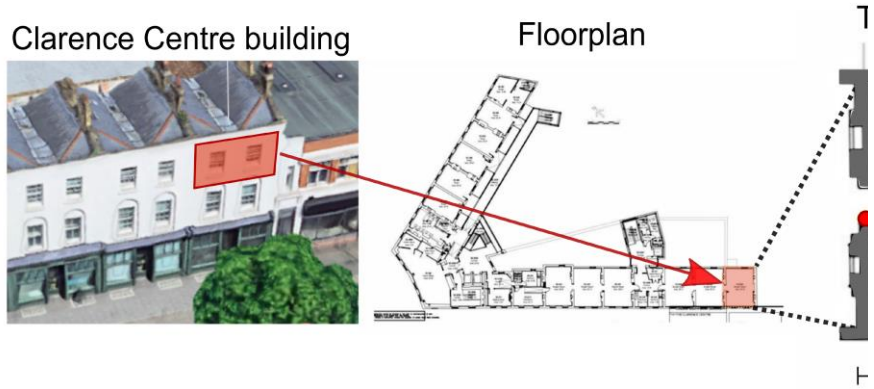
$$f_t = \sigma_g(W_f * x_t + U_f * h_{t-1} + V_f \circ c_{t-1} + b_f)$$
$$i_t = \sigma_g(W_i * x_t + U_i * h_{t-1} + V_i \circ c_{t-1} + b_i)$$
$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c * x_t + U_c * h_{t-1} + b_c)$$
$$o_t = \sigma_g(W_o * x_t + U_o * h_{t-1} + V_o \circ c_t + b_o)$$
$$h_t = o_t \circ \sigma_h(c_t)$$

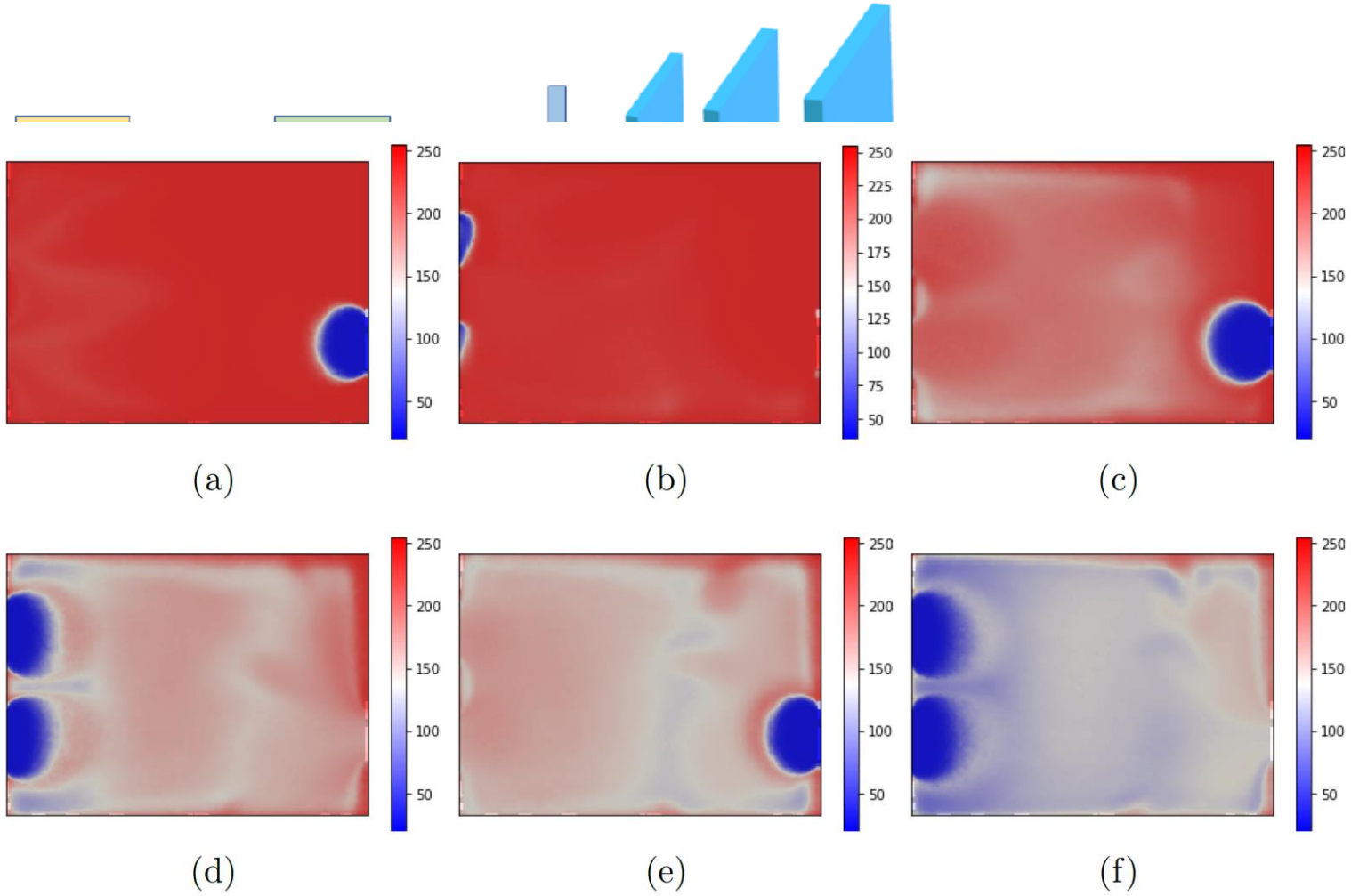
[*] M Amendola, R Arcucci, L Mottet, CQ Casas, S Fan, C Pain, P Linden, Y Guo - **Data Assimilation in the Latent Space of a Neural Network**

Latent Assimilation

***with Maddalena Amendola - 2020**



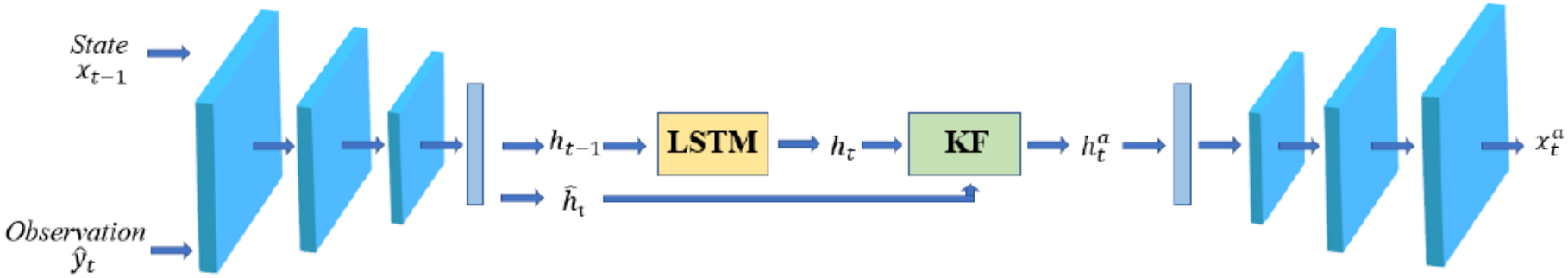




[*] M Amendola, R Arcucci, L Mottet, CQ Casas, S Fan, C Pain, P Linden, Y Guo - **Data Assimilation in the Latent Space of a Neural Network**

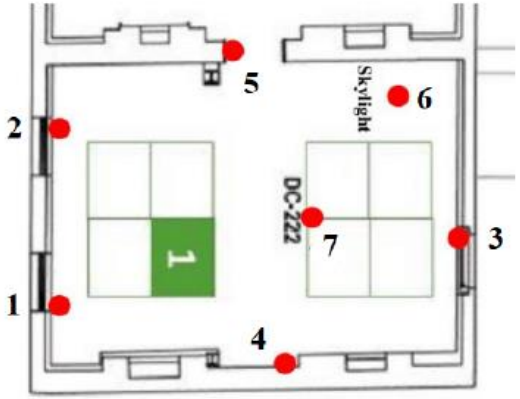
Latent Assimilation

***with Maddalena Amendola - 2020**

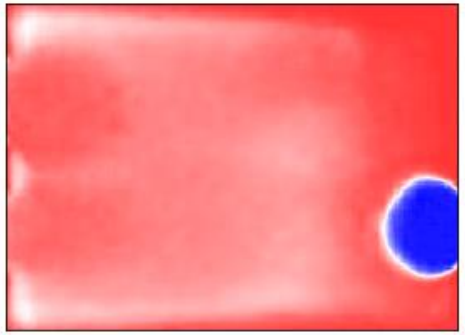


Measurements of carbon dioxide (CO2) in a room

MAGIC
Envisaging a world with greener cities



sensors



CFD

$\hat{R}$	cov-matrix from data	0.01 I	0.001 I	0.0001 I
MSE	3.356e-02	6.933e-04	1.211e-04	2.691e-06
Time (sec)	3.191e+00	2.899e+00	2.896e+00	2.896e+00

Table 6.11: Values of MSE of of x_t^a in the Physical space for different values of the observations errors covariance matrix $\hat{R}$ with the Structured dataset.

R	cov-matrix from data	0.01 I	0.001 I	0.0001 I
MSE	5.179e-02	6.928e-03	6.928e-03	6.997e-03
Time (sec)	2.231e+03	2.148e+03	2.186e+03	2.159e+03

Table 6.12: Standard Assimilation in the physical space performed by a KF

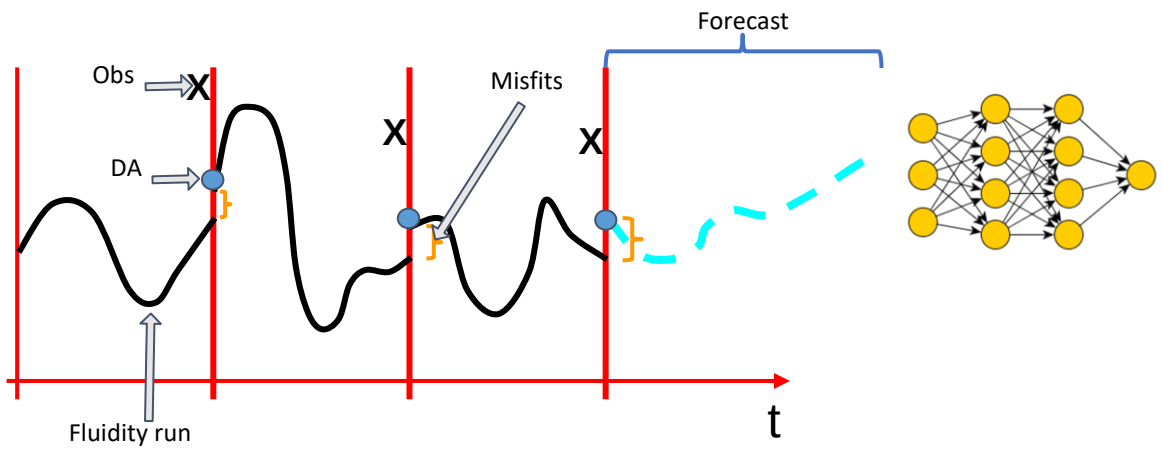
[*] M Amendola, R Arcucci, L Mottet, CQ Casas, S Fan, C Pain, P Linden, Y Guo - **Data Assimilation in the Latent Space of a Neural Network**

Deep Data Assimilation (DDA)

What if the observation are not available?

DDA … learning the Data Assimilation process

The idea:
Data Assimilation at each time step give us a misfit (DA - fluidity background), the saved misfits are trained using a Long short-term memory (LSTM) network and used for future forecasts.



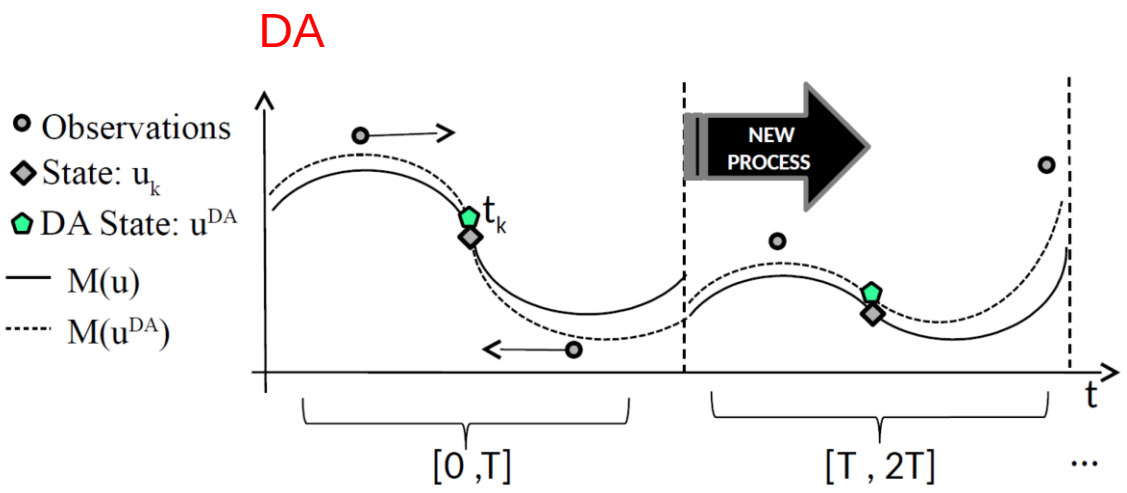
[*] R. Arcucci, J. Zhu, S. Hu, YK Guo, **Deep data assimilation: Integrating deep learning with data assimilation** - Applied Sciences
[**] C. Quilodran Casas, R. Arcucci**,** P. Wu, C. Pain, Y. Guo - **A Reduced Order Deep Data Assimilation model** – Physica D: nonlinear phenomena

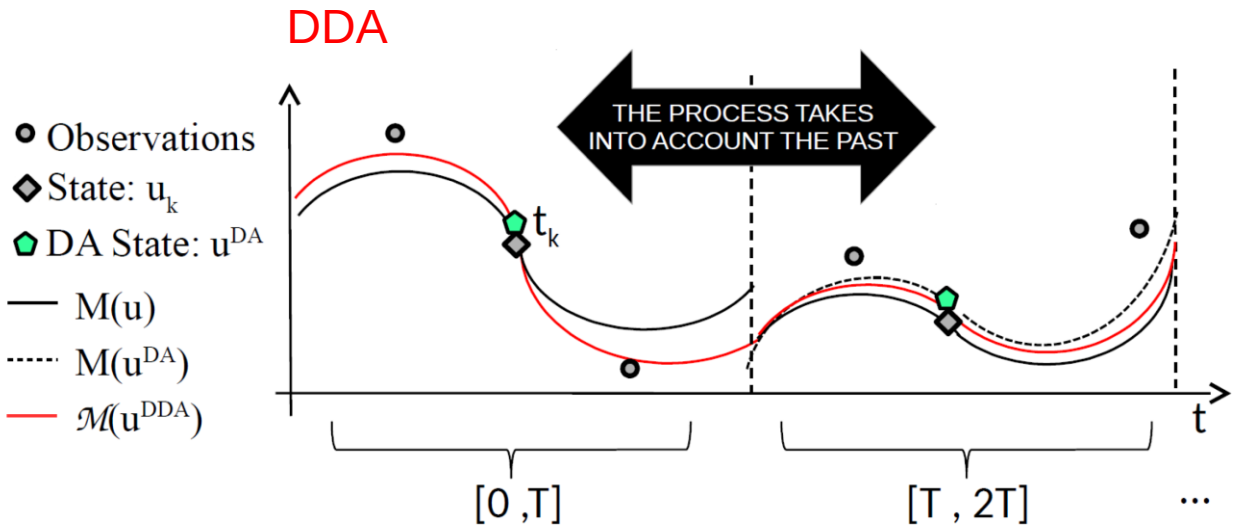
Deep Data Assimilation (DDA)

What if the observation are not available?

DDA … learning the Data Assimilation process

The idea:
Data Assimilation at each time step give us a misfit (DA - fluidity background), the saved misfits are trained using a Long short-term memory (LSTM) network and used for future forecasts.



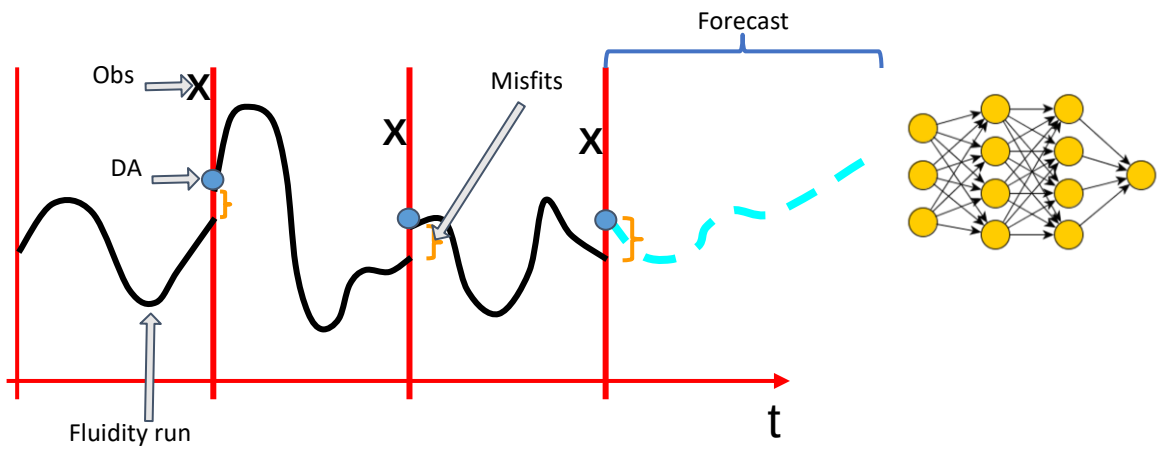


[*] R. Arcucci, J. Zhu, S. Hu, YK Guo, **Deep data assimilation: Integrating deep learning with data assimilation** - Applied Sciences
[**] C. Quilodran Casas, R. Arcucci, P. Wu, C. Pain, Y. Guo - **A Reduced Order Deep Data Assimilation model** – Physica D: nonlinear phenomena

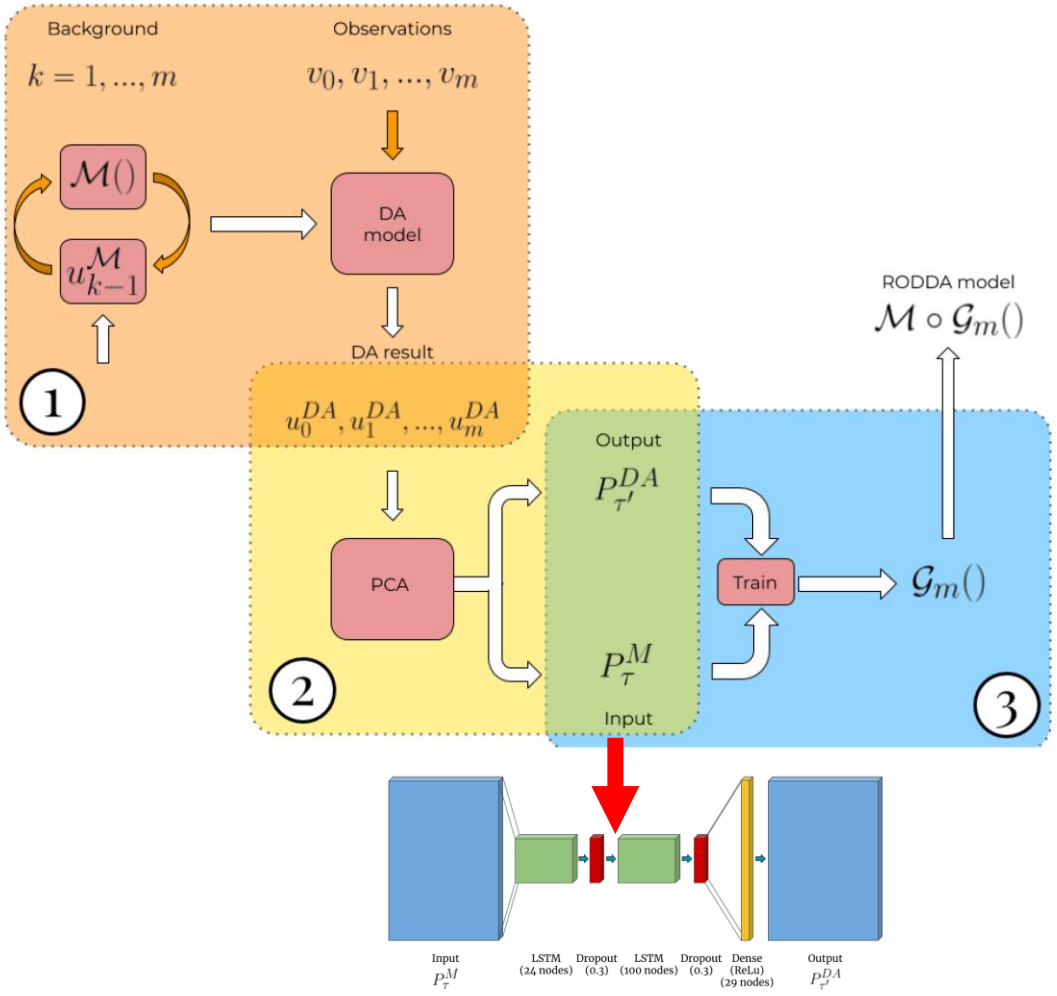
Deep Data Assimilation (DDA)

What if the observation are not available?

DDA … learning the Data Assimilation process



Reduced Order Deep Data Assimilation (RODDA)



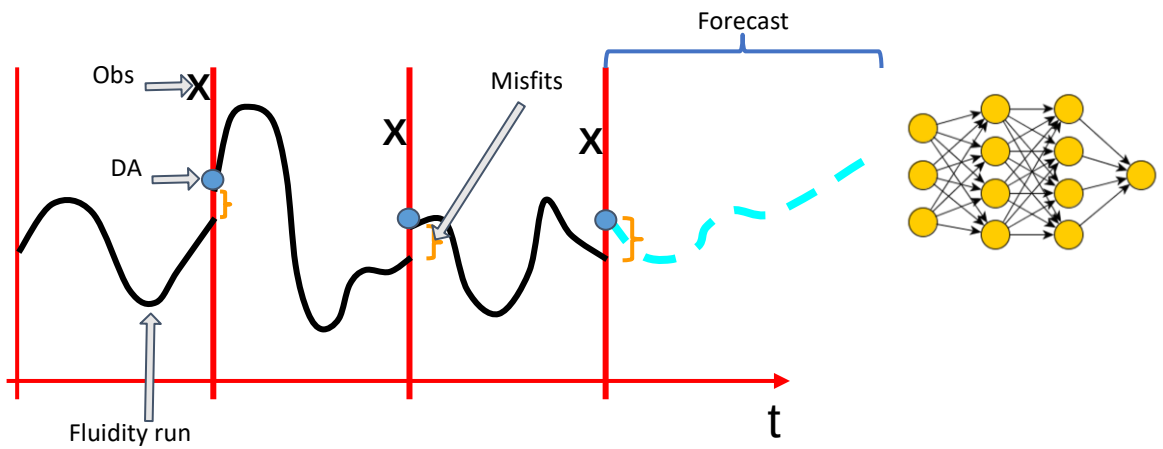
[*] R. Arcucci, J. Zhu, S. Hu, YK Guo, **Deep data assimilation: Integrating deep learning with data assimilation** - Applied Sciences

[**] C. Quilodran Casas, R. Arcucci, P. Wu, C. Pain, Y. Guo - **A Reduced Order Deep Data Assimilation model** – Physica D: nonlinear phenomena

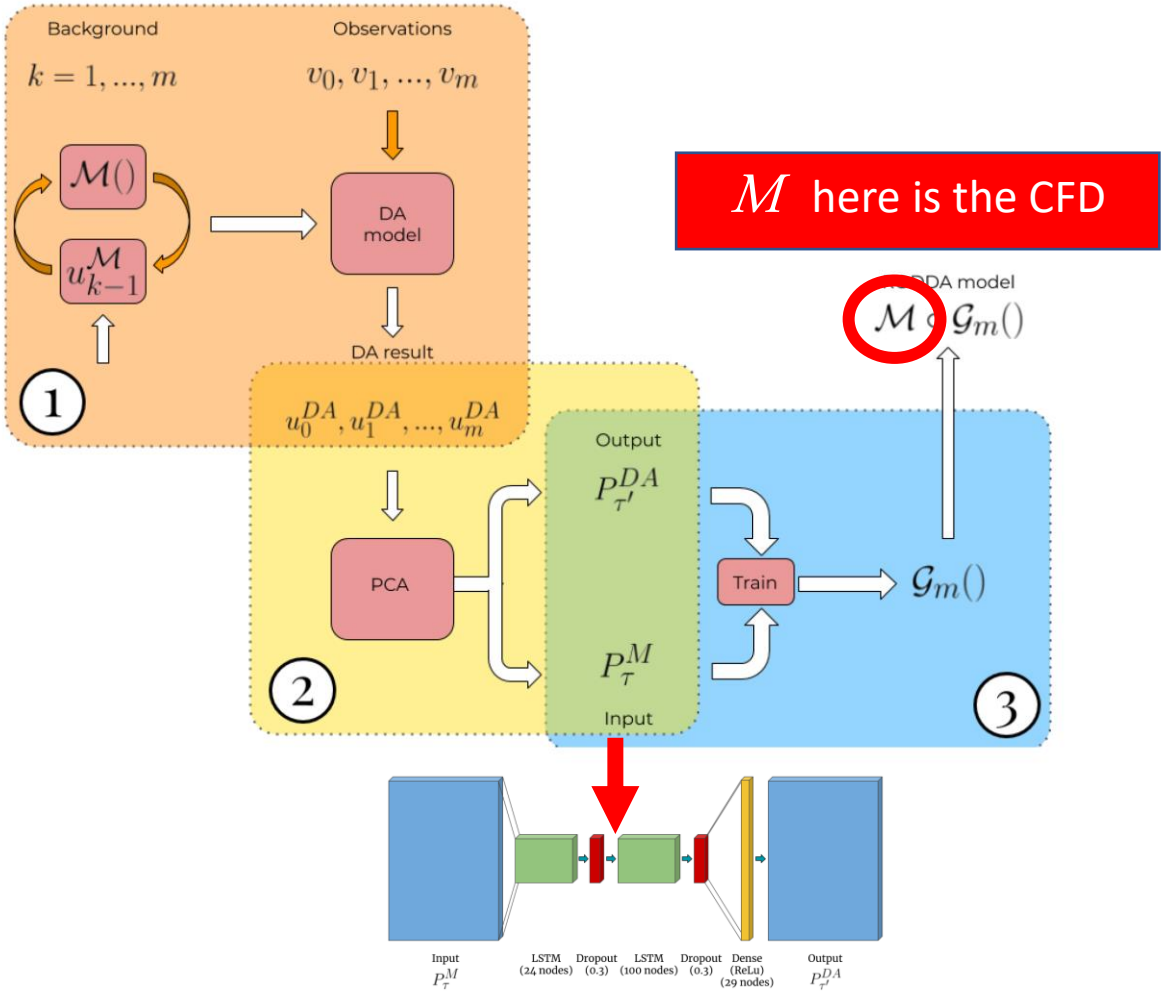
Deep Data Assimilation (DDA)

What if the observation are not available?

DDA … learning the Data Assimilation process



Reduced Order Deep Data Assimilation (RODDA)



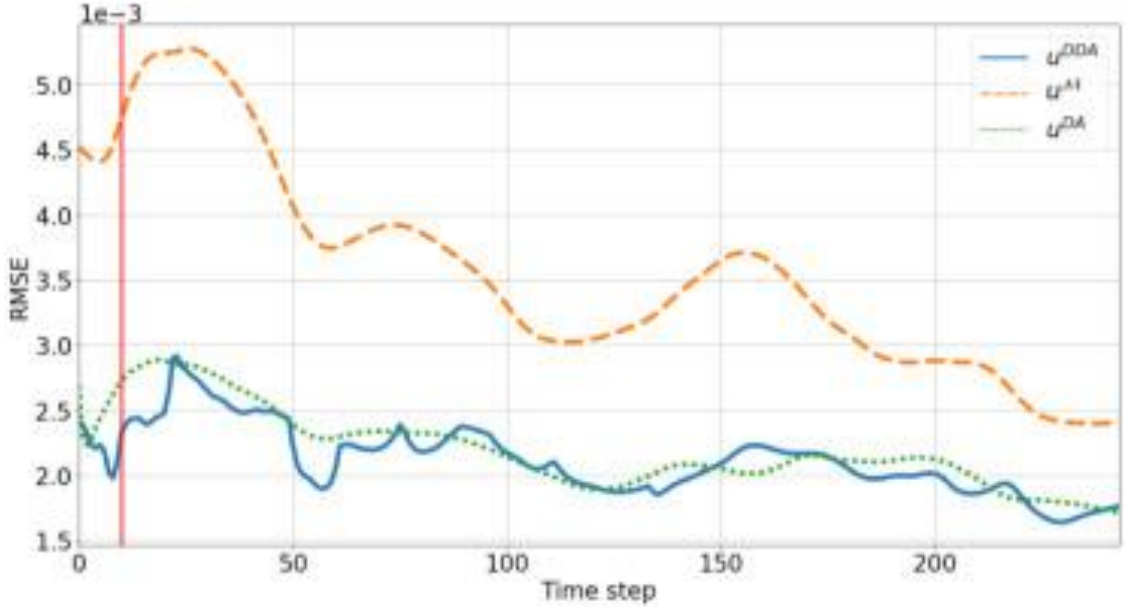
M here is the CFD

[*] R. Arcucci, J. Zhu, S. Hu, YK Guo, **Deep data assimilation: Integrating deep learning with data assimilation** - Applied Sciences

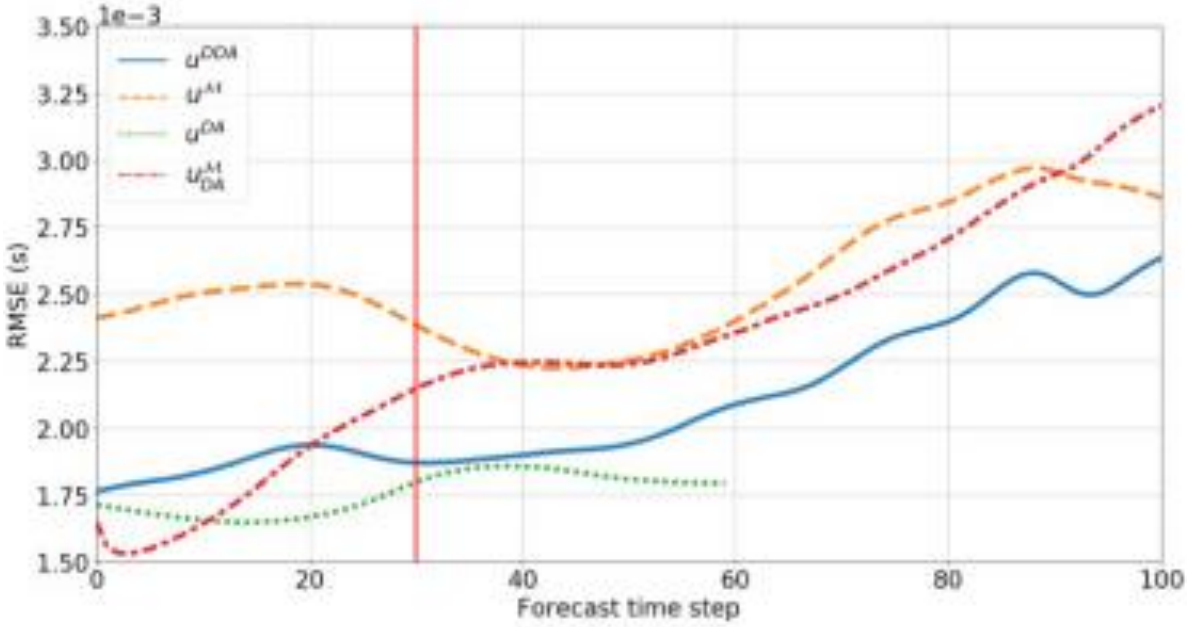
[**] C. Quilodran Casas, R. Arcucci, P. Wu, C. Pain, Y. Guo - **A Reduced Order Deep Data Assimilation model** – Physica D: nonlinear phenomena

Reduced Order Deep Data Assimilation (RODDA)

***with Cesar Quilodran Casas**



(a) RMSE time-series during the training



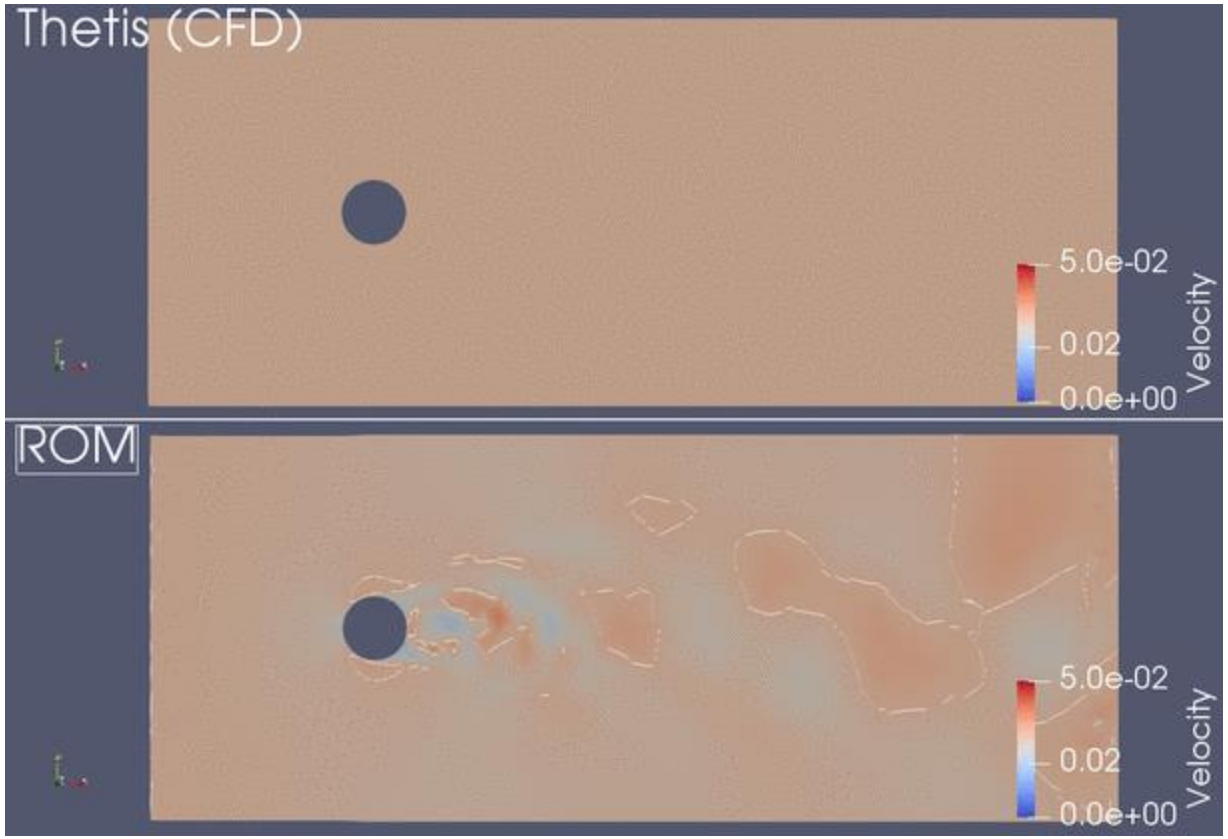
(b) Forecast RMSE time-series.

Same accuracy but RODDA is 1000 times faster than DA

[*] R. Arcucci, J. Zhu, S. Hu, YK Guo, **Deep data assimilation: Integrating deep learning with data assimilation** - Applied Sciences

[**] C. Quilodran Casas, R. Arcucci**,** P. Wu, C. Pain, Y. Guo - **A Reduced Order Deep Data Assimilation model** – Physica D: nonlinear phenomena

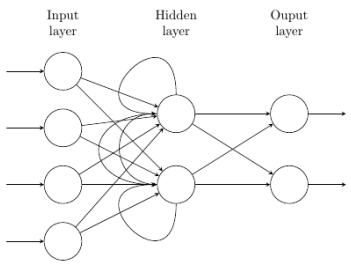
Shallow water models using Reduce Order Models and Machine Learning



CFD flow past a cylinder simulation using Thetis an unstructured grid coastal ocean model built using the Firedrake finite element framework

Time: 0.6h

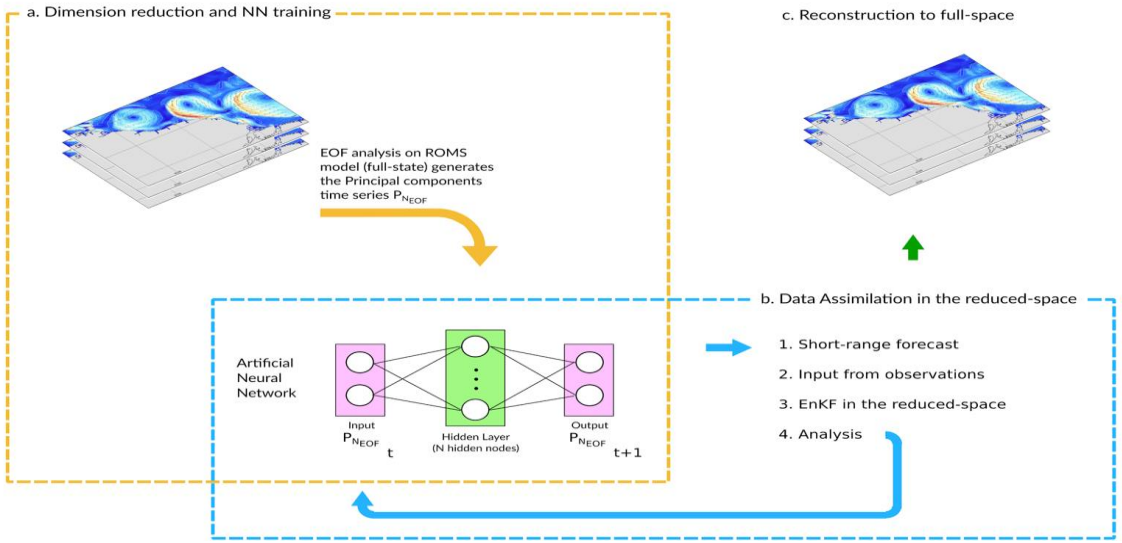
Reduced-order model (ROM) is needed to speed up the simulation

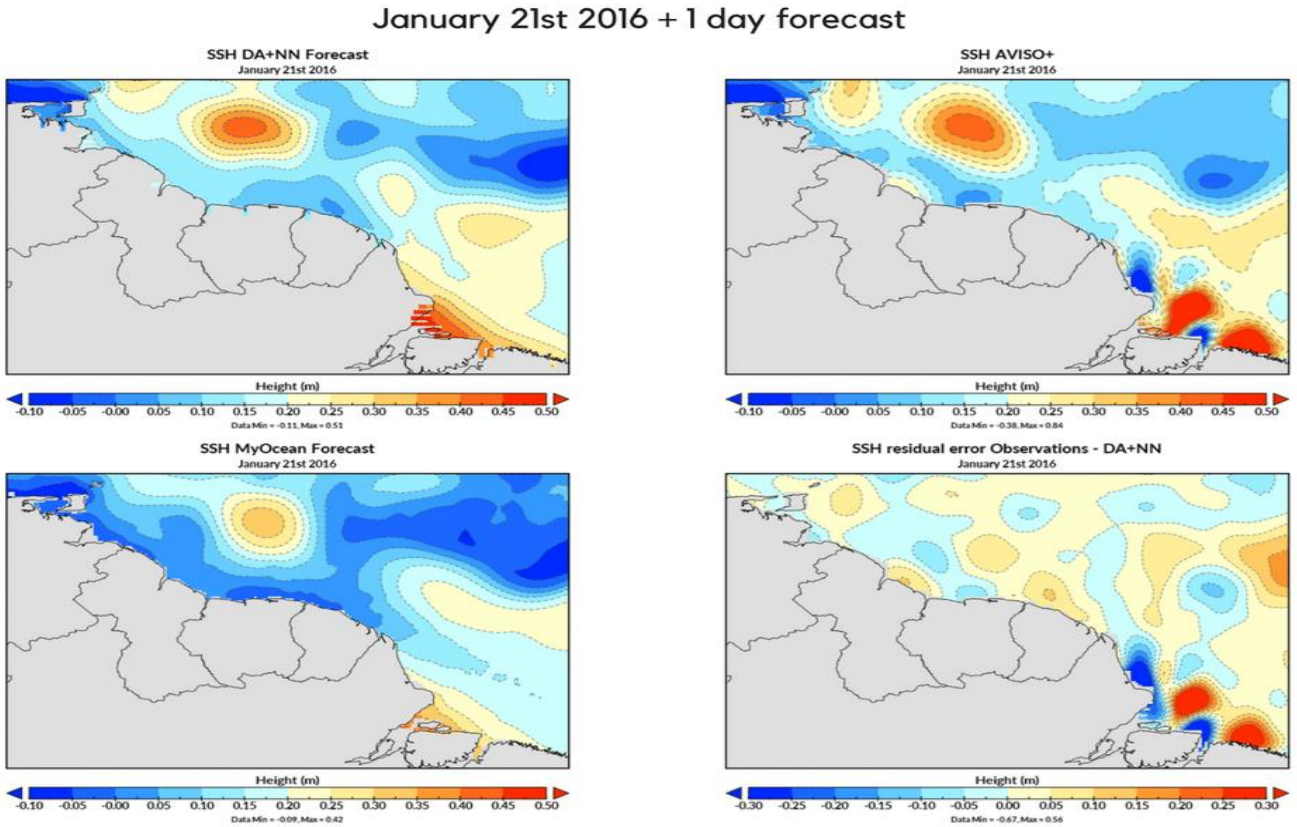


~20 PCs were used (~99% variance)

Time: 0.001h ~ 3.6secs

Reduced Order Modelling with Data Assimilation and Machine Learning





6-day forecasts of sea surface height (SSH) in the North Brazil Current: **Fast ocean data assimilation and forecasting using a neural-network reduced-space regional ocean model of the north Brazil current**, C. Quilodran Casas

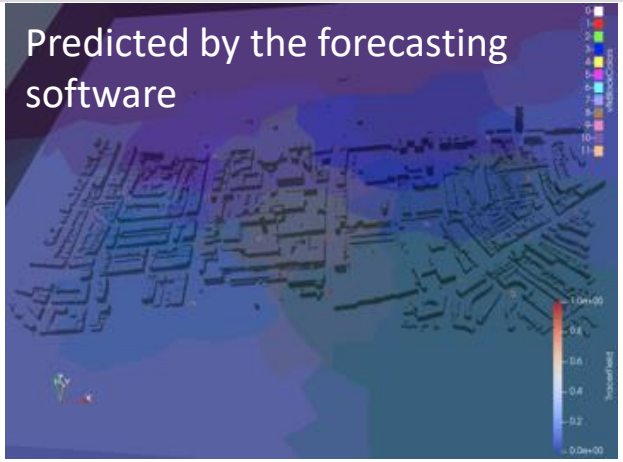
… to make all this doable in a real case scenarios

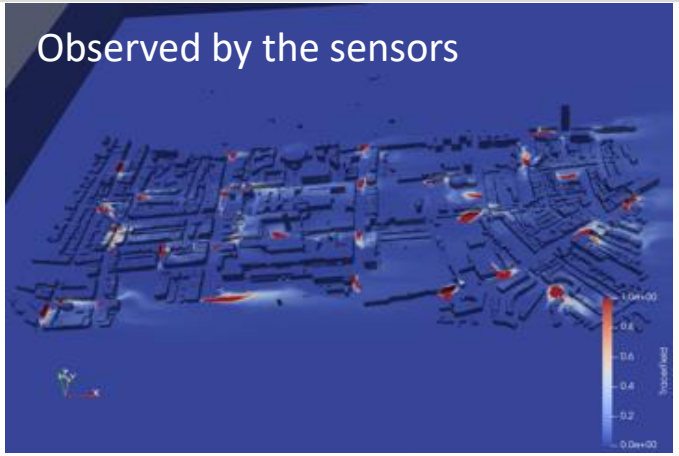
(facing Big Data problems): **Domain Decomposition**

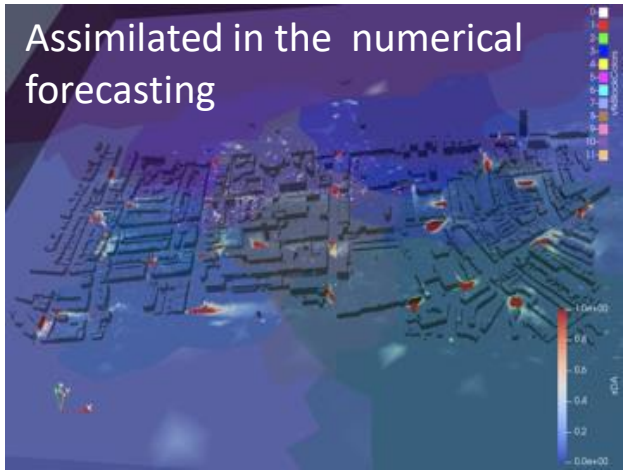
How to face a computationally expensive problem (such as Data Assimilation and Machine Learning) on a domain such as a BIG city

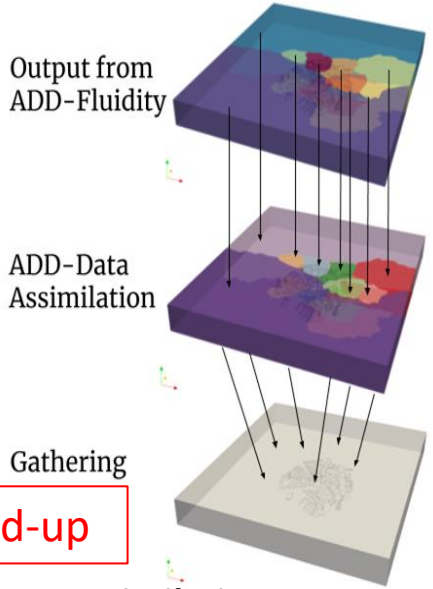
South Kensington, Imperial College Campus, CFD mesh made of 12,8 millions of nodes

Iso-surface of the pollutant concentration computed in parallel









This approach is **fully scalable**, the reduction in terms of execution time has a linear speed-up

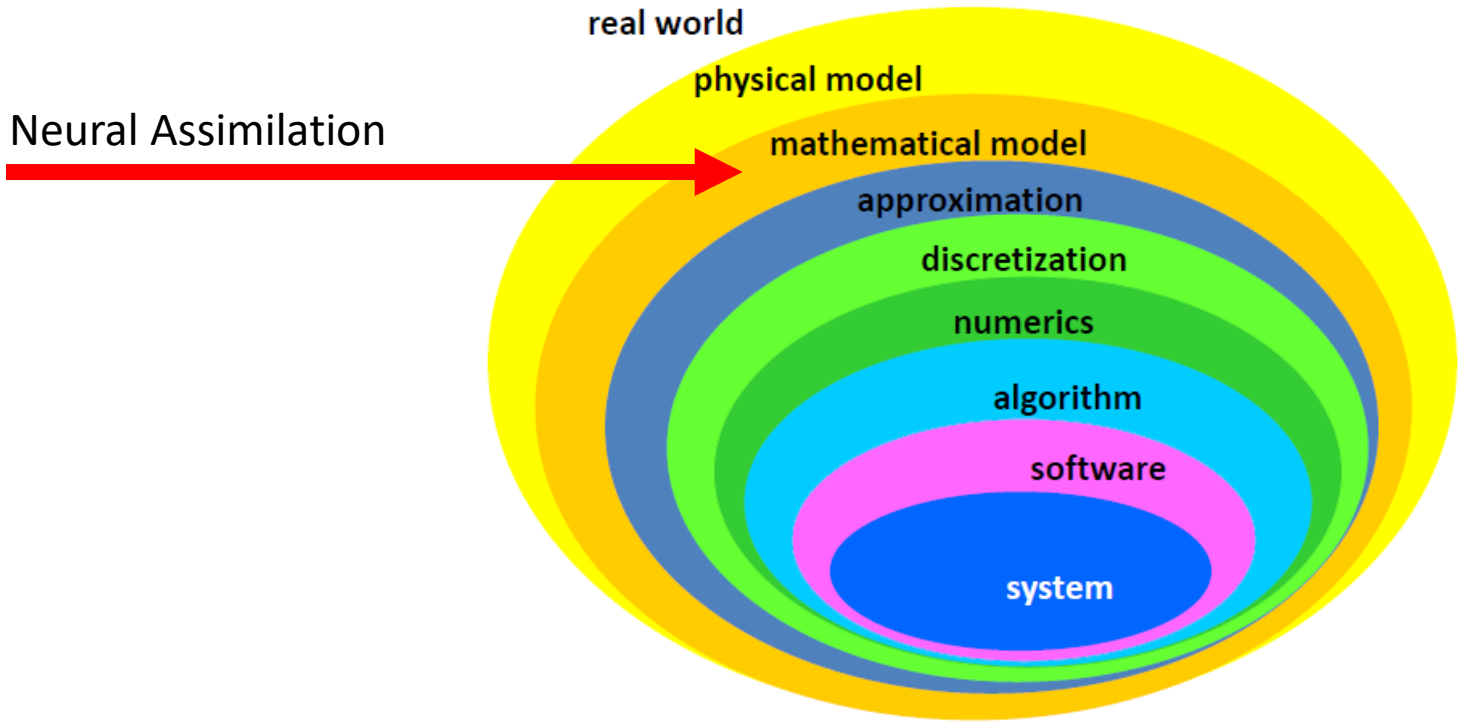
[*] Arcucci R, Mottet L, Casas CAQ, Guitton F, Pain C, Guo YK et al., 2020, **Adaptive Domain Decomposition for Effective Data Assimilation**

[**] Arcucci R, Casas CQ, Xiao D, Mottet L, Fang F, Wu P, Pain C, Guo YKet al., 2020, **A domain decomposition reduced order model with data assimilation**

Data Assimilation with Machine Learning…

«Adapting old algorithms to fit new data sets usually means adapting new data sets to behave like old ones.»
Rossella Arcucci

For data scientists



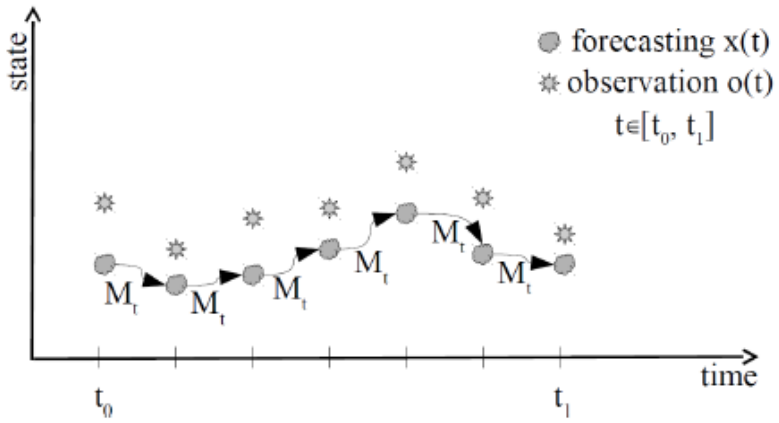
Neural Assimilation

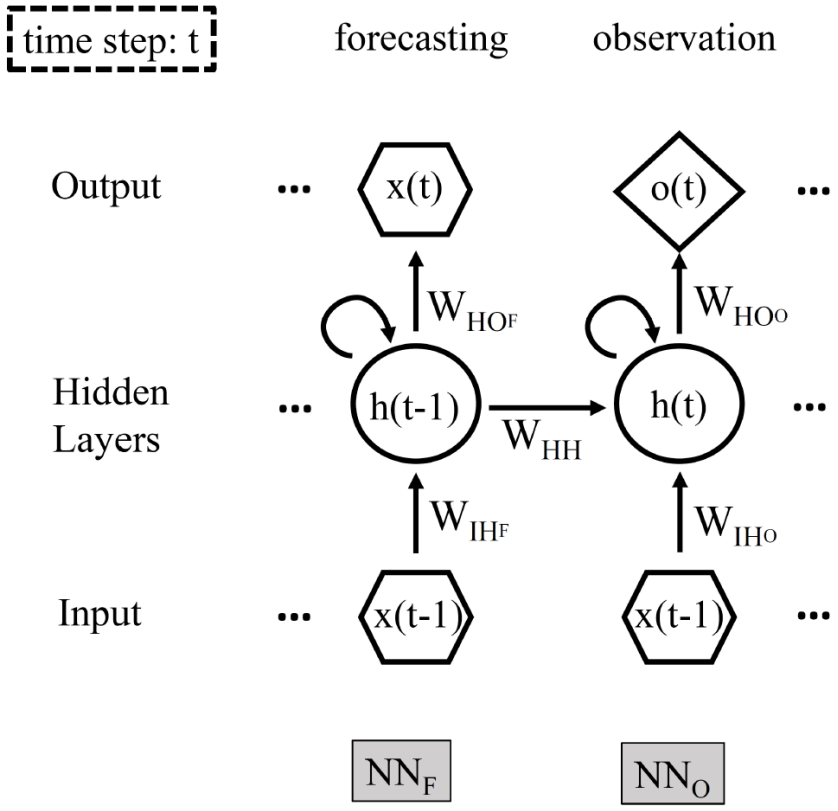
… a coupled neural network made of two Recurrent Neural Networks trained on forecasting data and observed data respectively.

$$\bar{o}(t) = f_{O_O}\left(W_{HO_O} h(t-1)\right)$$

$$h(t) = f_H\left(W_{IH}\bar{x}(t-1) + W_{HH} h(t-1)\right)$$

$$\bar{x}(t) = f_{O_F}\left(W_{HO_F} h(t)\right)$$



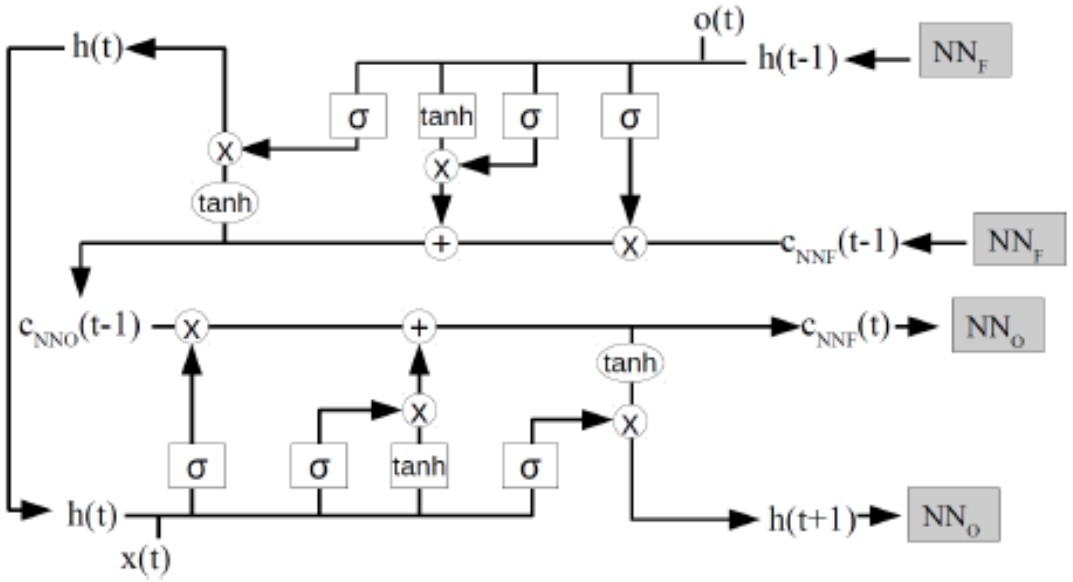


[*] R. Arcucci, L. Moutiq, Y. Guo - Neural Assimilation - Lecture Notes in Computer Science book series Volume 12142

… Neural Assimilation

… adopting Long short-term memory (LSTM) architecture for the two RNNs.

Why LSTM???

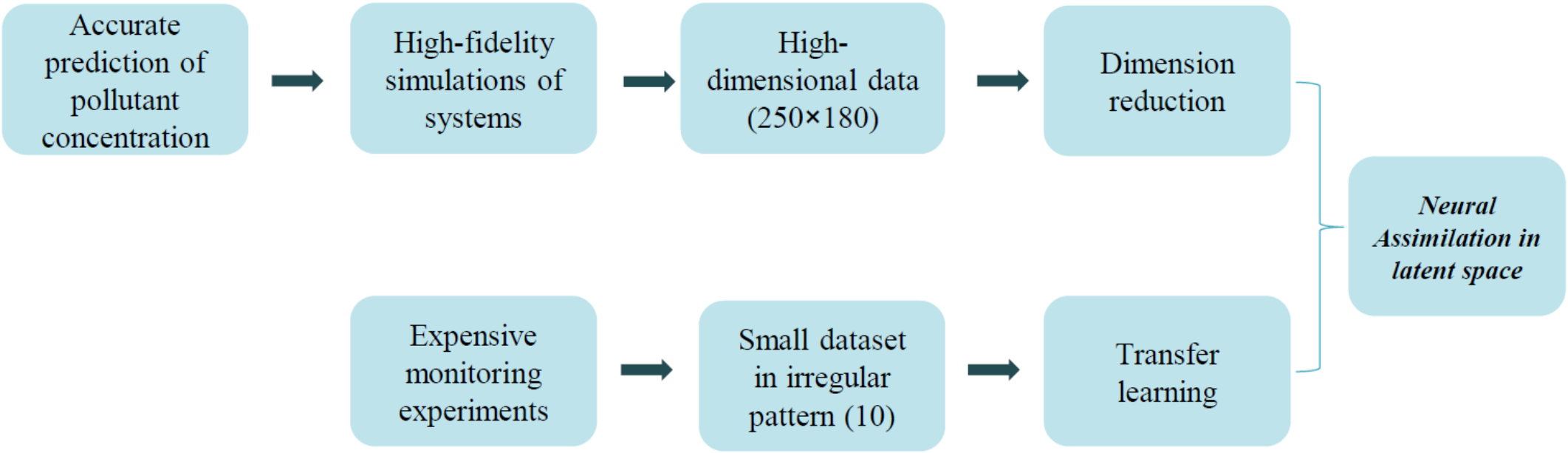


- LSTMs are suitable to contain information outside the normal flow of the recurrent network so it is easier to plug two networks together
- LSTMs allow to preserve the error that can be backpropagated through time and layers which is a very important point for discrete forecasting models.

Neural Assimilation

Pollutant concentration London (UK)

***with Yiwen Xu - 2020**



INHALE

MAGIC
Envisaging a world with greener cities

EPSRC
Engineering and Physical Sciences Research Council

The CFD software is FLUIDITY

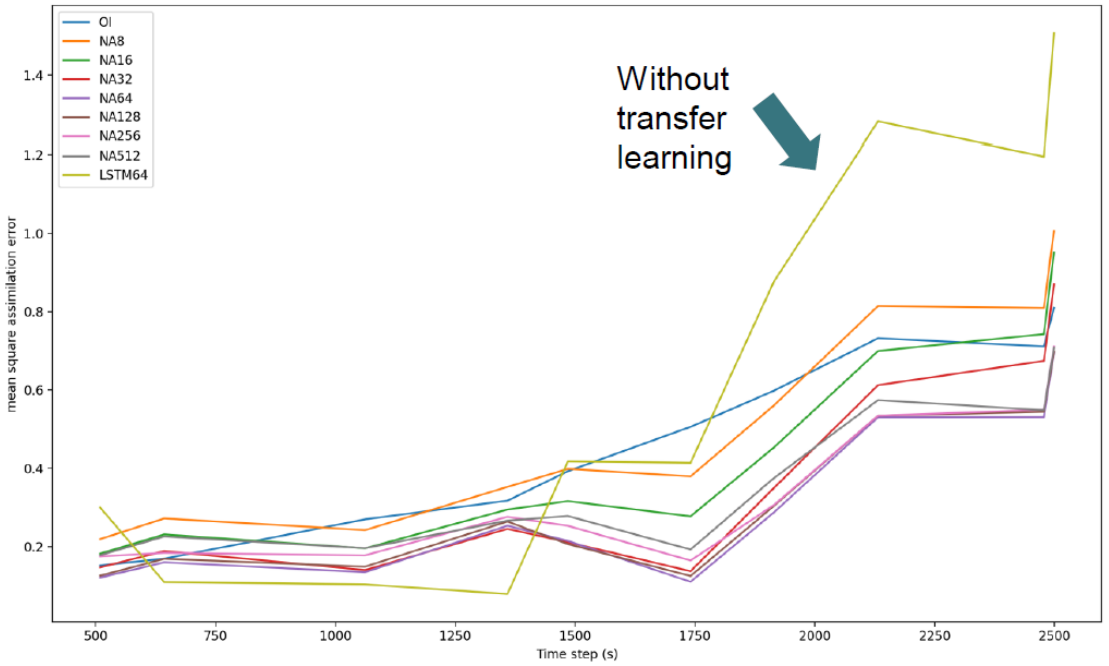
Neural Assimilation

Pollutant concentration London (UK)

***with Yiwen Xu - 2020**

reduced method	NN_F training(s)	NN_O training(s)	NA prediction(s)	error
8 PCA	73.04	10.64	0.032	0.39
16 PCA	69.87	9.53	0.032	0.33
32 PCA	69.43	9.54	0.032	0.26
64 PCA	70.44	9.51	0.033	0.24
128 PCA	71.39	9.12	0.033	0.26
256 PCA	73.46	11.68	0.033	0.27
512 PCA	79.85	16.4	0.034	0.28
8 PCAE	63.42	12.68	0.023	0.51
64 Autoencoder	112	9.09	0.014	0.15
128 Autoencoder	120	9.72	0.015	0.16
256 Autoencoder	108	7.86	0.014	0.53

Reference OI: 0.34
Reference OI: 267.43 (s) per time step



INHALE

MAGIC
Envisaging a world with greener cities

EPSRC
Engineering and Physical Sciences Research Council

… we are happy to share

To get access to our **codes**:

- Our GitHub https://github.com/DL-WG
- Send me an email r.arcucci@imperial.ac.uk
- Google “DataLearning Data Science Institute Imperial College London”… you will find our DataLearning group https://www.imperial.ac.uk/data-science/research/research-themes/datalearning/

Event: Third edition of the **Workshop on Machine Learning and Data Assimilation for Dynamical Systems**

MLDADS 2021 - ICCS 2021
Virtual event 16-18 June 2021

MLDADS

Machine Learning and Data Assimilation for Dynamical Systems

Thank You

r.arcucci@imperial.ac.uk

Imagine it, then do it!

The true sign of Intelligence is not Knowledge but Imagination. Albert Einstein

Intelligence / Artificial intelligence ? … Imagination is the key!

research.reading.ac.uk/dare/2018/10/18/machine-learning-and-data-assimilation/

University of Reading

HOME ABOUT US RESEARCH EVENTS BLOG

MACHINE LEARNING AND DATA ASSIMILATION

18th October 2018 By Jessica Gardner Comment Closed

by Rossella Arcucci

Imagine a world where it is possible to accurately predict the weather, climate, storms, tsunami and other computational intensive problems in real time from your laptop or even mobile phone – if one has access to a supercomputer then to be able to predict at unprecedented scales/detail. This is the long term aim of our work on Data Assimilation with Machine Learning at the Data Science Institute (Imperial College London, UK) and as such, we believe, it will be a key component of future Numerical Forecasting systems.

We proved that the integration of machine learning with Data assimilation can increase the reliability of prediction, reducing errors by including information with an actual physical meaning from

Search …

RECENT POSTS

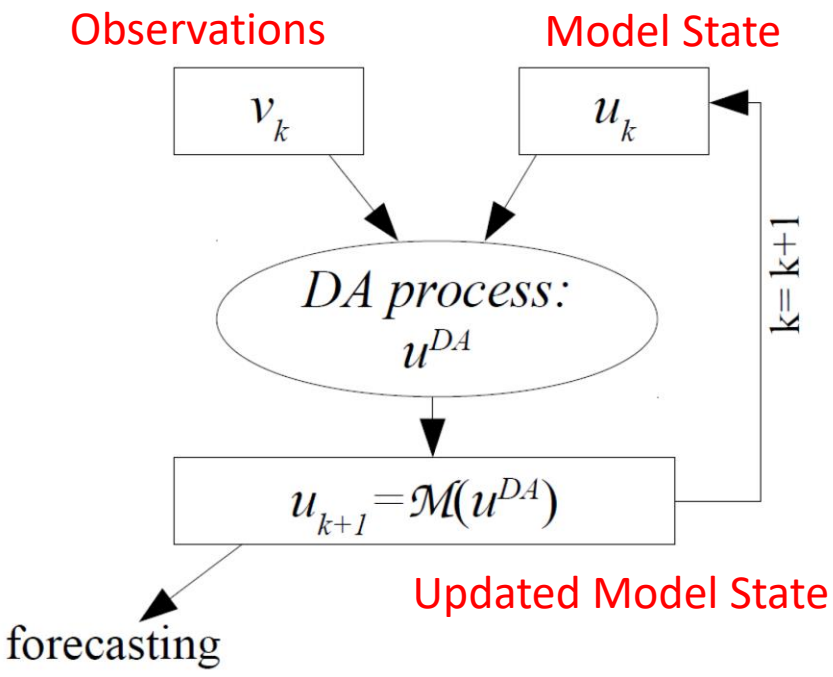
DARE contributes evidence to national flooding inquiry

Dr Joanne Waller wins Royal Meteorological Society LF Richardson Prize

The Ensemble Club

https://research.reading.ac.uk/dare/2018/10/18/machine-learning-and-data-assimilation/

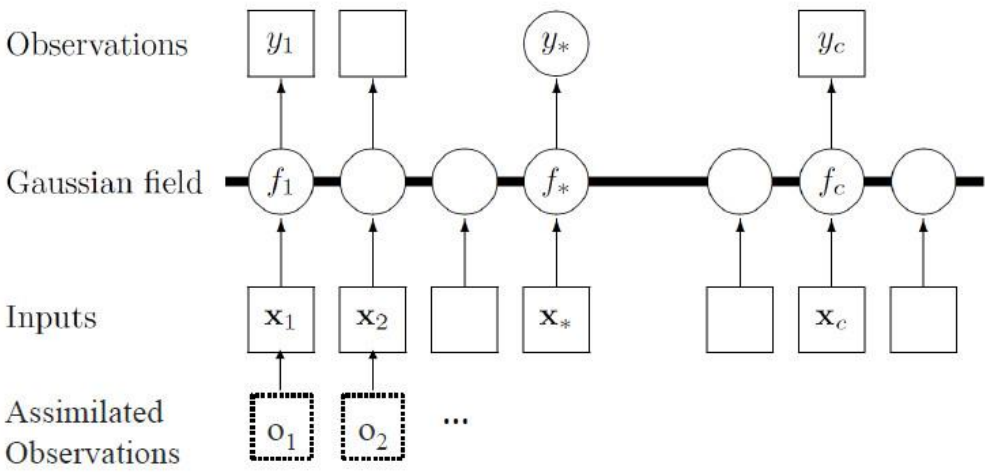
Content

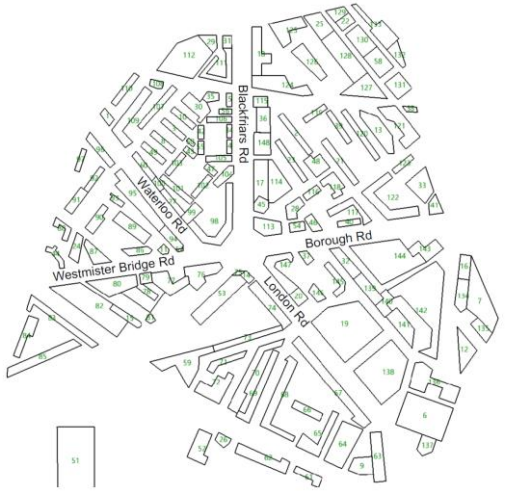


To have the expected benefit from data assimilation, we need:

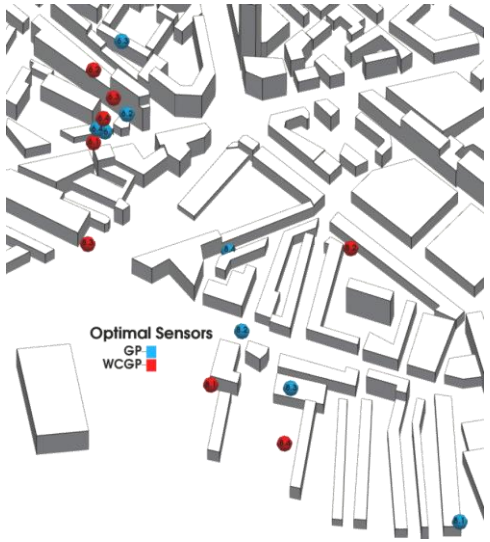
- **Good efficiency** (accelerate the execution time)
- **Good accuracy** (reduce the errors propagation in the models)
- **Real world scenarios** (develop models that can be used in real world scenarios – face **Big Data** problems)
- **Good quality data** (optimal placements/locations)

DA+ML … for optimal sensor placement or optimal data collection





Outdoor... South London [*]

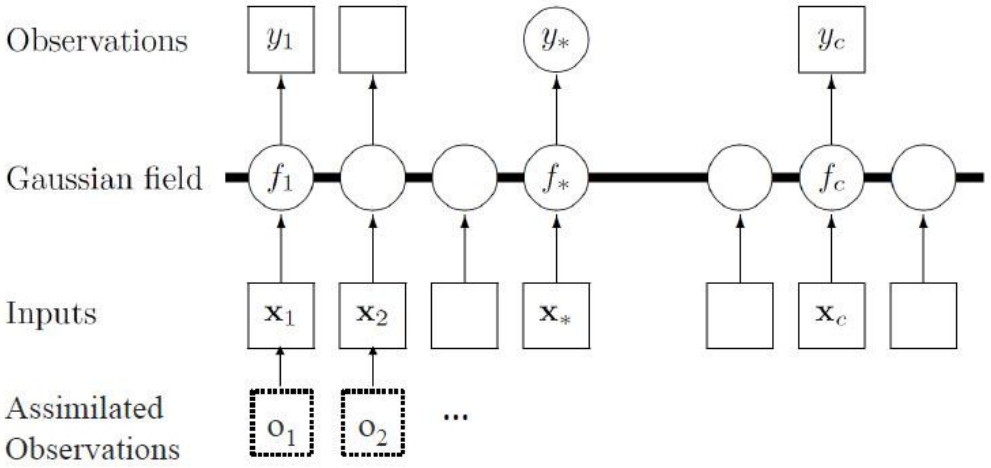


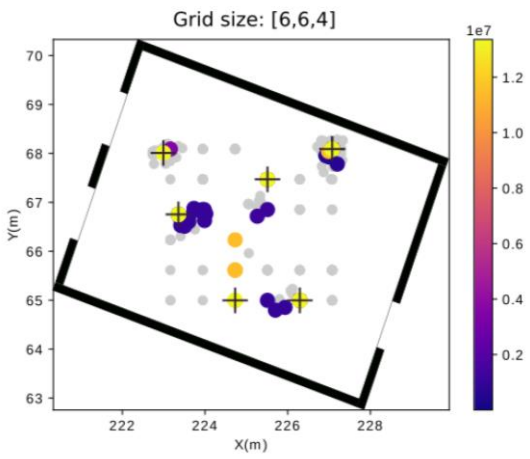
	Real Mean	Estimated Mean	$\mathrm{MSE}(\mathbf{x}^{\mathcal{M}})$	$\mathrm{MSE}(\mathbf{x}^{\mathrm{DA}})$
Original Algorithm	2.4662e-01	1.9598e-01	2.24e-01	5.25e-02
Data Learning (GP+DA)	2.4662e-01	2.2771e-01	1.77e-01	3.35e-02
Random	2.4662e-01	2.3900e02	6.54e00	8.90e-01

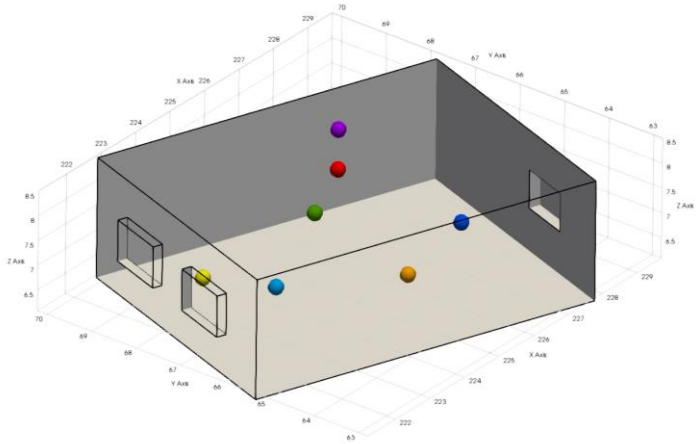
[*] T. Dur, R. Arcucci, L. Mottet, M. Molina Solana, C. Pain, Y. Guo - **Weak Constraint Gaussian Process for optimal sensor placement** - Journal of Computational Science vol 42,pag.101-110 DOI:10.1016/j.jocs.2020.101110

[**] G. Tajnafoi, R. Arcucci, L. Mottet, C. Vouriot, Molina Solana, C. Pain, Y. Guo - **Variational Gaussian Processes for optimal sensor placement** - Journal of Applied Mathematics

DA+ML … for optimal sensor placement or optimal data collection







Assimilating the **seven optimally positioned sensors**, the error of the predictive model, i.e. Fluidity, is reduced by up to three order of magnitude: MSE(C^n) = 0,17 and MSE(C^{DA}) = 0,0005.

Moreover, this error is up to two order of magnitude lower than the ones computed using random sensors placement. In one of the worst random case scenario: MSE(C^{DA}) = 0,023.

[*] T. Dur, R. Arcucci, L. Mottet, M. Molina Solana, C. Pain, Y. Guo - **Weak Constraint Gaussian Process for optimal sensor placement** - Journal of Computational Science vol 42,pag.101-110 DOI:10.1016/j.jocs.2020.101110

[**] G. Tajnafoi, R. Arcucci, L. Mottet, C. Vouriot, Molina Solana, C. Pain, Y. Guo - **Variational Gaussian Processes for optimal sensor placement** - Journal of Applied Mathematics