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
… the era of the data!

High resolution Models…

Coupled models…

Real observations…

facebook
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)
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?
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.

https://www.imperial.ac.uk/data-science/research/research-themes/datalearning/
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 $S = \{u_j\}_{j=1}^n$,

1. Compute $B_0$ from $u_0$ and $S$
2. Initialize iteration $k = 1$
3. while $k < n$ do
   4. Compute $u_k = M u_{k-1}$
   5. Compute $B_k$ from $u_k$ and $S$
   6. Compute
      \[
      u^{DA}_k = \text{argmin}_u \left\{ \|u - u_k\|_{B_k^{-1}} + \|v_k - H_k u\|_{R_k^{-1}} \right\}
      \]
   7. Count up $k$ for the next iteration
4. 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 \( S = \{u_j\}_{j=1, \ldots, n} \),

1. Compute \( B_0 \) from \( u_0 \) and \( S \)
2. Initialize iteration \( k = 1 \)
3. while \( k < n \) do
   4. Compute \( u_k = M u_{k-1} \)
   5. Compute \( B_k \) from \( u_k \) and \( S \)
   6. Compute
      \[
      u_k^{DA} = \arg\min_u \left\{ \|u - u_k\|_{B_k^{-1}} + \|v_k - H_k u\|_{R_k^{-1}} \right\}
      \]
   7. Count up \( k \) for the next iteration
4. 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)

predicted by the forecasting software

assimilated in the numerical forecasting

Observed by the sensors

\[ J(u) = \alpha \|u - u_0\|_{B^{-1}}^2 + \|Gu - v\|_{R^{-1}}^2 \]

3DVar in the control space

\[ w = V^+ \delta u \]
\[ B = V V^T \]

\[ w^{DA} = \text{argmin}_{w \in \mathbb{R}^{NP \times N}} J(w) \]

\[ J(w) = \frac{1}{2} \alpha w^T w + \frac{1}{2} (GVw - d)^T R^{-1} (GVw - d) \]

DA function

Reduced space, TSVD

[*] 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)

- **Observer**: Observed by the sensors
- **Predictor**: Predicted by the forecasting software

**Data Assimilation**
- Assimilated in the numerical forecasting
- Observed by the sensors
- Predicted by the forecasting software

**CFD (Computational Fluid Dynamic)**
- Predicted by the forecasting software

**Observation Operator**
- Learning from real data to make CFD predictions more accurate and realistic

**Data Assimilation**
- Assimilated in the numerical forecasting
- Observed by the sensors
- Predicted by the forecasting software

**CFD (Computational Fluid Dynamic)**
- Predicted by the forecasting software

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

$J(w_i) = \frac{1}{2}w_i$

3DVar in the latent space

Latent Assimilation

*with Maddalena Amendola - 2020

Measurement Update ("Correct")

1. Compute the Kalman Gain
   \[ \hat{K}_{t+1} = \hat{H} \hat{H}^T (\hat{H} \hat{H}^T + \hat{R}_{t+1})^{-1} \]

2. Update estimate with measurement
   \[ \hat{h}_{t+1}^a = \hat{h}_{t+1} + \hat{K}_{t+1} (\hat{h}_{t+1} - \hat{H} \hat{h}_{t+1}) \]

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

[*] 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

Clarence Centre building

Floorplan

[*] 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

Measurements of carbon dioxide (CO2) in a room

Table 6.11: Values of MSE of x_t^c in the Physical space for different values of the observations errors covariance matrix R with the Structured dataset.

<table>
<thead>
<tr>
<th>R</th>
<th>cov-matrix from data</th>
<th>0.01 I</th>
<th>0.001 I</th>
<th>0.0001 I</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>3.356e-02</td>
<td>6.933e-04</td>
<td>1.211e-04</td>
<td>2.691e-06</td>
</tr>
<tr>
<td>Time (sec)</td>
<td>3.191e+00</td>
<td>2.899e+00</td>
<td>2.896e+00</td>
<td>2.896e+00</td>
</tr>
</tbody>
</table>

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
What if the observation are not available?

**Deep Data Assimilation (DDA)**

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.


Deep Data Assimilation (DDA)

What if the observation are not available?

DDA ... learning the Data Assimilation process

Reduced Order Deep Data Assimilation (RODDA)


Deep Data Assimilation (DDA)

What if the observation are not available?

DDA ... learning the Data Assimilation process

Reduced Order Deep Data Assimilation (RODDA)


Reduced Order Deep Data Assimilation (RODDA)

*with Cesar Quilodran Casas

Same accuracy but RODDA is 1000 times faster than DA


Shallow water models using Reduce Order Models and Machine Learning

CFD flow past a cylinder simulation using Thetis an unstructered 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
Reduction 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
How to face a computationally expensive problem (such as Data Assimilation and Machine Learning) on a domain such as a BIG city

Iso-surface of the pollutant concentration computed in parallel

Predicted by the forecasting software

Observed by the sensors

Assimilated in the numerical forecasting

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

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

[*] 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
«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.

\[
\hat{\phi}(t) = f_{\phi_o} (W_{HO_o} h(t-1))
\]

\[
h(t) = f_H (W_{HH} \bar{x}(t-1) + W_{HH} h(t-1))
\]

\[
\bar{x}(t) = f_{\phi_f} (W_{HO_f} h(t))
\]

[26x491] Neural Assimilation - Lecture Notes in Computer Science book series Volume 12142

[*] 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.

• 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)

Accurate prediction of pollutant concentration → High-fidelity simulations of systems → High-dimensional data (250×180) → Dimension reduction

Expensive monitoring experiments → Small dataset in irregular pattern (10) → Transfer learning

Neural Assimilation in latent space

The CFD software is FLUIDITY

*with Yiwen Xu - 2020
Neural Assimilation

Pollutant concentration London (UK)

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

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

Without transfer learning
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
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!

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 error by including information with no actual physical meaning from

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

Observations $y_1$ $y_\ast$ $y_c$
Gaussian field $f_1$ $f_\ast$ $f_c$
Inputs $x_1$ $x_2$ $x_\ast$ $x_c$
Assimilated Observations $o_1$ $o_2$ ...

Outdoor... South London [*]

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Real Mean</th>
<th>Estimated Mean</th>
<th>MSE($x^M$)</th>
<th>MSE($x^{DA}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Algorithm</td>
<td>2.4662e-01</td>
<td>1.9598e-01</td>
<td>2.24e-01</td>
<td>5.25e-02</td>
</tr>
<tr>
<td>Data Learning (GP+DA)</td>
<td>2.4662e-01</td>
<td>2.2771e-01</td>
<td>1.77e-01</td>
<td>3.35e-02</td>
</tr>
<tr>
<td>Random</td>
<td>2.4662e-01</td>
<td>2.3900e02</td>
<td>6.54e00</td>
<td>8.90e-01</td>
</tr>
</tbody>
</table>


Assimilating the **seven optimally positioned sensors**, the error of the predictive model, i.e. Fluidity, is reduced by up to three order of magnitude: \( \text{MSE}(\text{C}^n) = 0.17 \) and \( \text{MSE}(\text{C}^\text{DA}) = 0.0005 \).

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: \( \text{MSE}(\text{C}^\text{DA}) = 0.023 \).
