



Data Assimilation

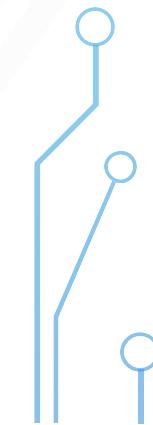
1 March 2023

Dr Rossella Arcucci

Department of Earth Science & Engineering,
Data Learning working group,
Data Science Institute,
AI network speaker at ICL (~260 academics),
World Meteorological Organization wg-member,

r.arcucci@imperial.ac.uk

<https://www.imperial.ac.uk/people/r.arcucci>



CONTENT

- Uncertainty quantification, backward error analysis and propagation of errors in data-driven models
- Data Assimilation for uncertainty minimization and data learning for effectively merging heterogeneous data



What is AI?

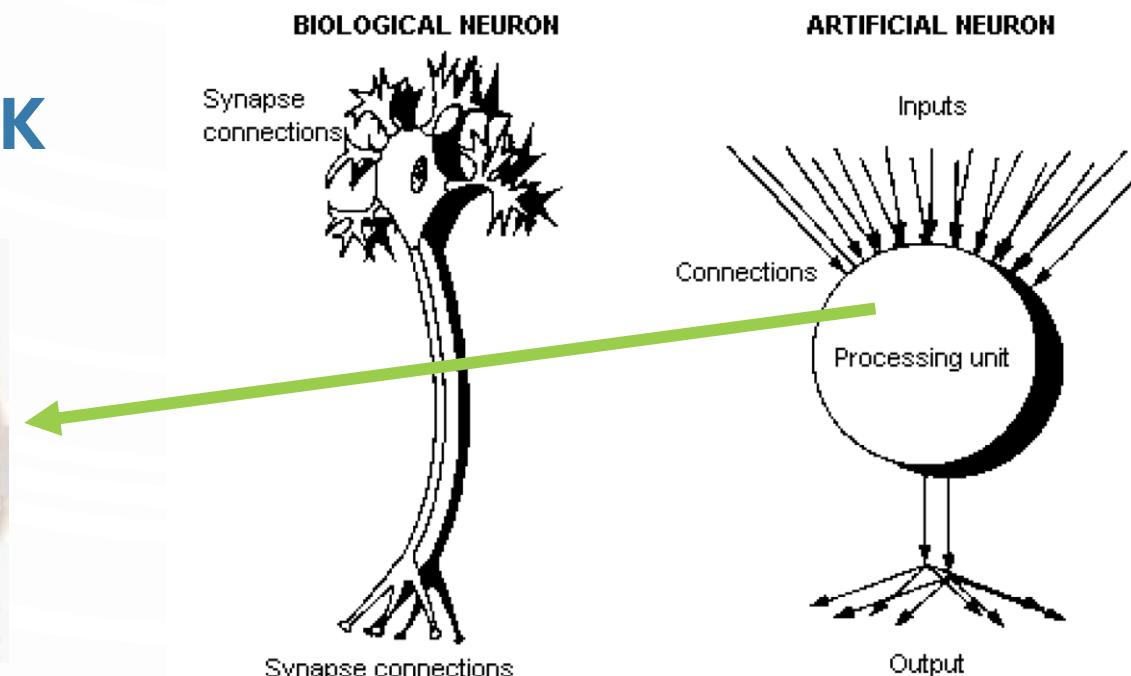
Artificial Intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.



What is AI?

Artificial Intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.

ARTIFICIAL NEURAL NETWORK



What is AI?

Artificial Intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.

DATA

SYNONYMS

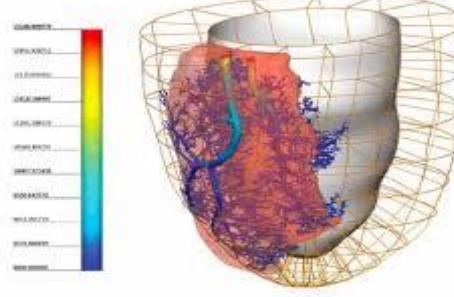
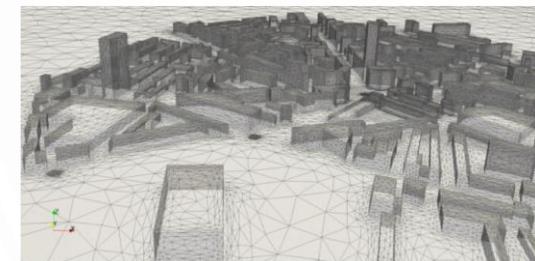
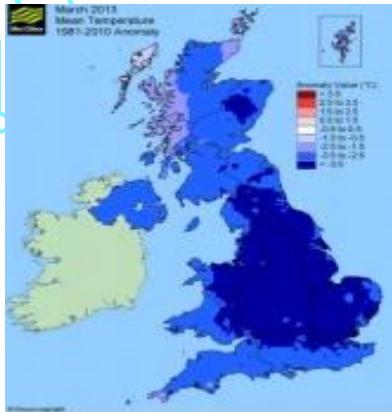
facts, figures, statistics, details, particulars, specifics, features
information, evidence, intelligence, material, background, input
proof, fuel, ammunition
statement, report, return, dossier, file, documentation, archive, archives





... the era of the data!

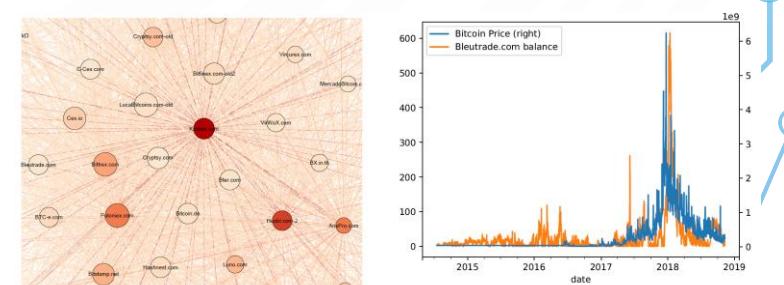
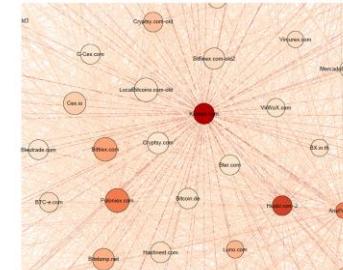
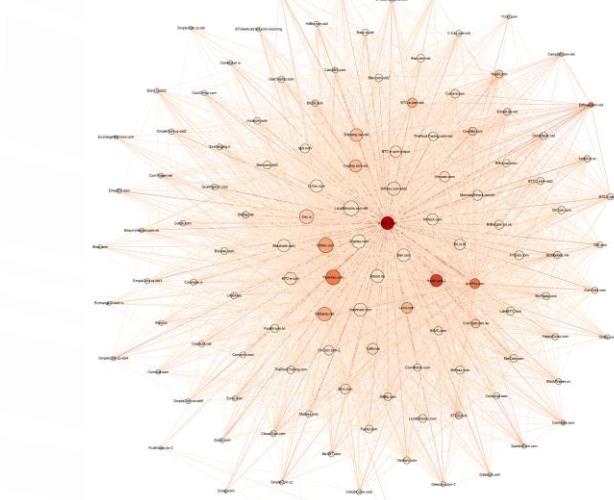
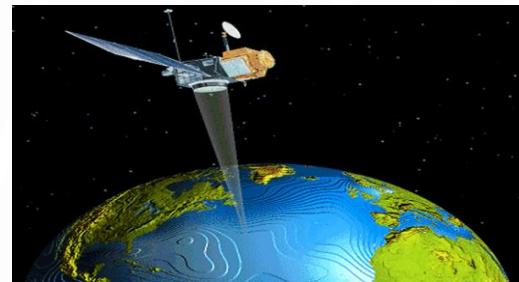
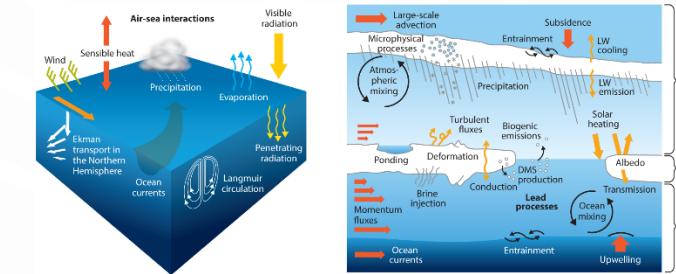
High resolution
Models...



facebook



Real observations...



TYPE OF DATA ANALYSIS

- Descriptive analytics** : What happened?
- Diagnostic analytics** : Why did it happen?
- Predictive analytics** : What is likely to happen in the future?
- Prescriptive analytics** : What is the best course of action to take?

SOME EFFECTS OF THE ERRORS



Input data
(initial conditions)



$$\boxed{M(P): \begin{cases} L(u(t))=f \\ u(t_0)=u_0 \end{cases}}$$

Solution
(forecast)



The numerical solution (the prediction of the system) is affected by errors:

- Simplifications of the model
- Discretization of the model
- Implementation by finite precision computations
- Model's approximation errors
- Discretization errors
- Round-off errors



errors on data

SOME EFFECTS OF THE ERRORS



Input data
(initial conditions)



M(P):

$$\begin{cases} L(u(t))=f \\ u(t_0)=u_0 \end{cases}$$

Solution
(forecast)



The numerical solution (the prediction of the system) is affected by errors:

(absolute error) = (approximate value) – (true value)

(relative error) = $\frac{\text{(absolute error)}}{\text{(true value)}} = \frac{\text{(approximate value)}}{\text{(true value)}} - 1$

SOME EFFECTS OF THE ERRORS



Input data
(initial conditions)



$$\boxed{M(P): \begin{cases} L(u(t))=f \\ u(t_0)=u_0 \end{cases}}$$

Solution
(forecast)



The numerical solution (the prediction of the system) is affected by errors:

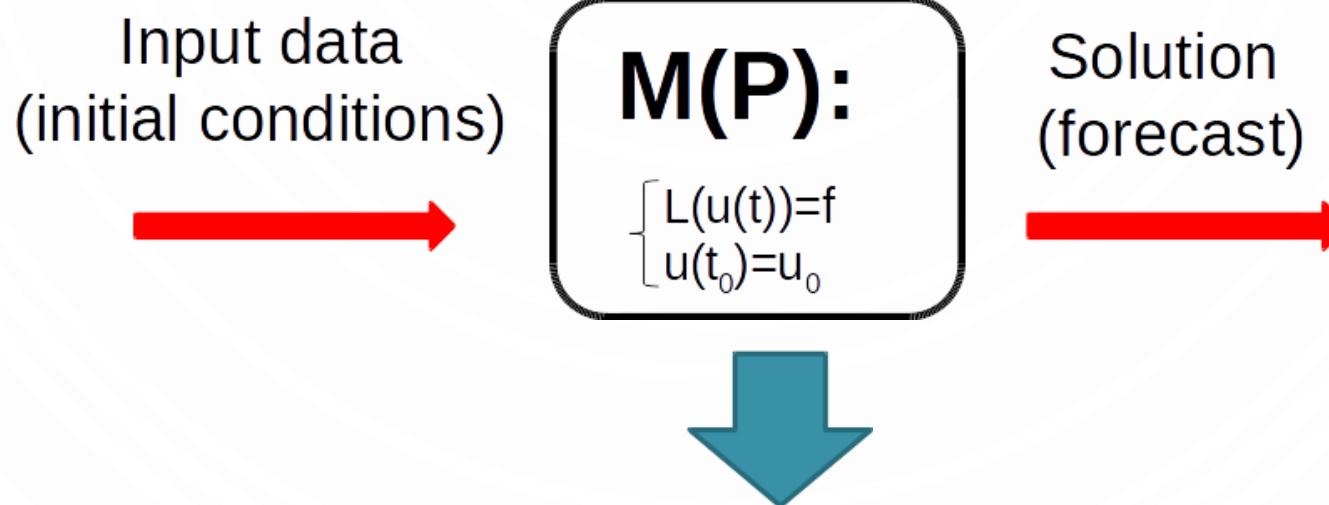
Fundamental difficulty with measuring error: for many problems we cannot compute the exact answer, we can only approximate it!

Often, the best we can do is *estimate* the error!

SOME EFFECTS OF THE ERRORS



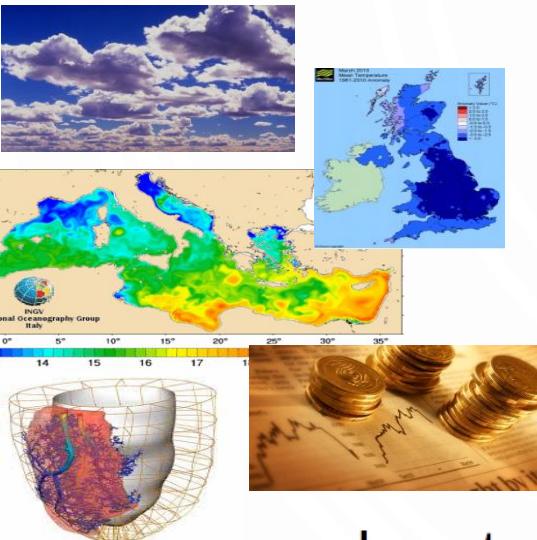
backward error analysis:



any perturbation on the data (δ) propagates on the solution (σ) by an amplification factor given by c = a parameter that depends on the algorithm and μ = the Condition number of the problem

Error on the solution: $\sigma = \mu c \delta$

data at $t=0$ (initial condition)



Input data
(initial conditions)



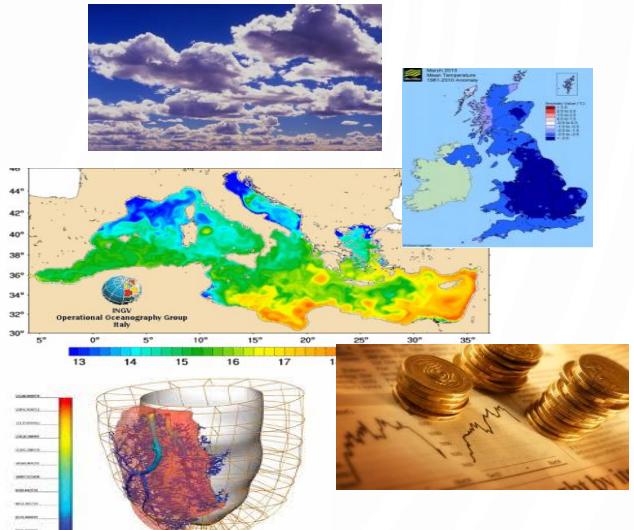
M(P):

$$\begin{cases} L(u(t)) = f \\ u(t_0) = u_0 \end{cases}$$

Solution
(forecast)

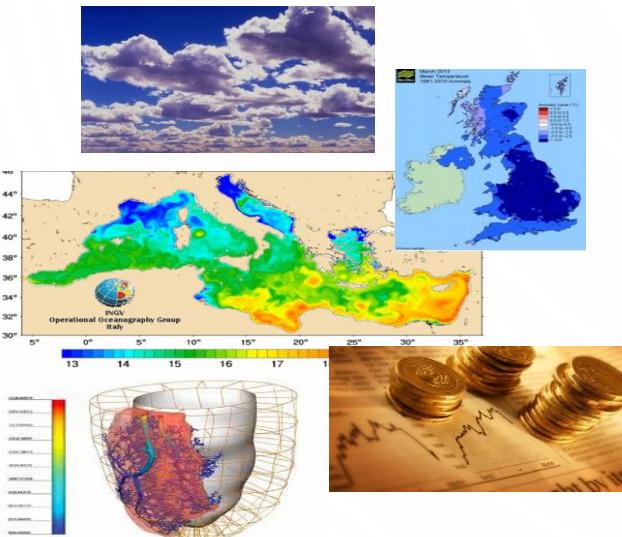


data at $t=t+k$

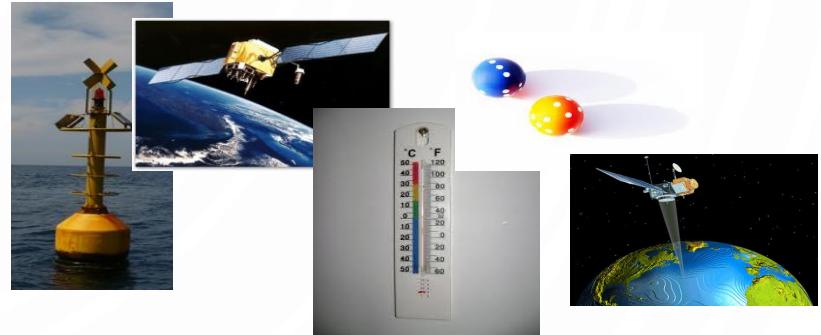


EXTRA DATA

data (initial condition)



observed data



INVERSE ILL POSED
PROBLEM

A BIG DATA PROBLEM

A NON LINEAR
PROBLEM

A DATA FILTERING
PROBLEM

Data Assimilation: try to google it!

AN UNCERTAINTY
QUANTIFICATION
PROBLEM

A LARGE SIZE
COMPUTATIONAL
PROBLEM

A COMPUTATIONAL INTENSIVE
PROBLEM

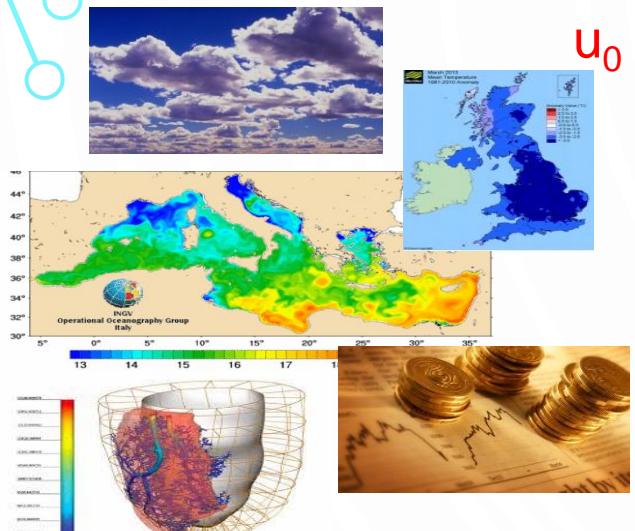
everyone faces it from a point
of view...
what is ours?



https://www.youtube.com/watch?v=JUB6_x6-d4

MERGE THE DATA

data (initial condition)



observed data



y_0

u_0

D.A.

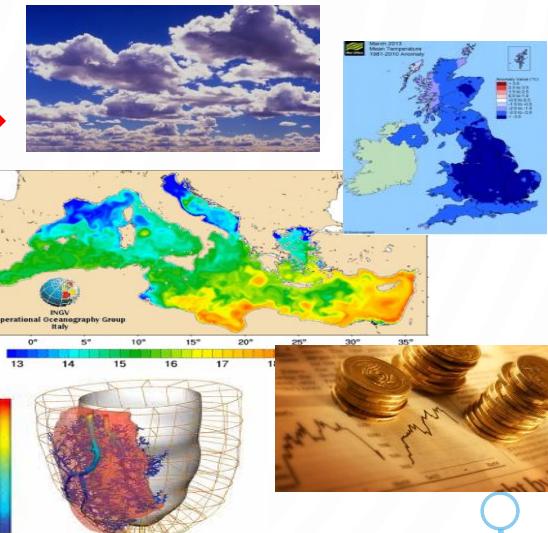
u_0

M(P):

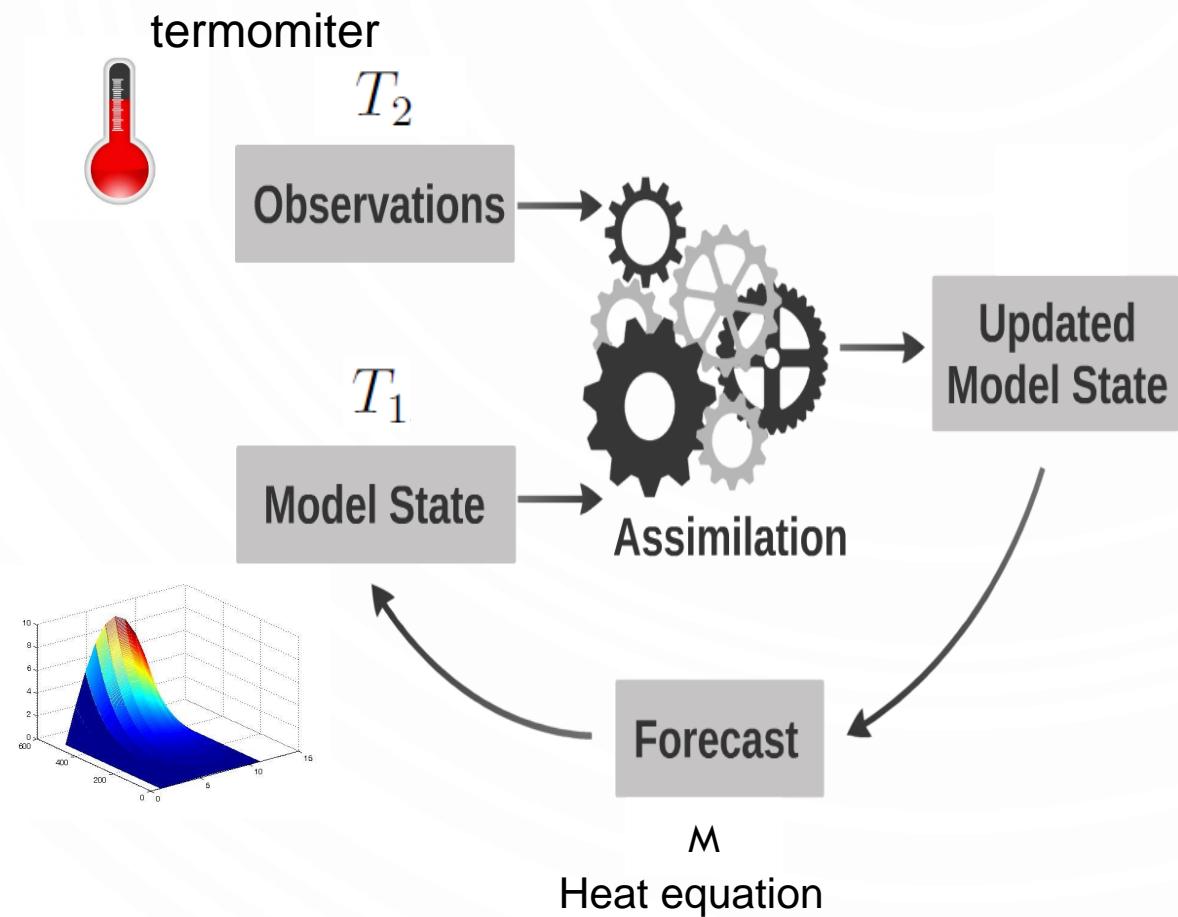
$$\begin{cases} L(u(t))=f \\ u(t_0)=u_0 \end{cases}$$



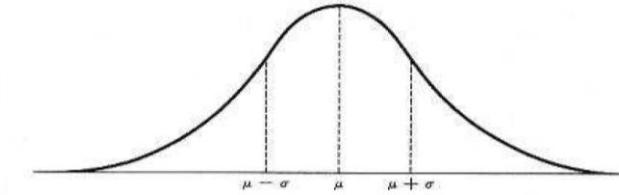
$u(t), t > t_0$
forecast



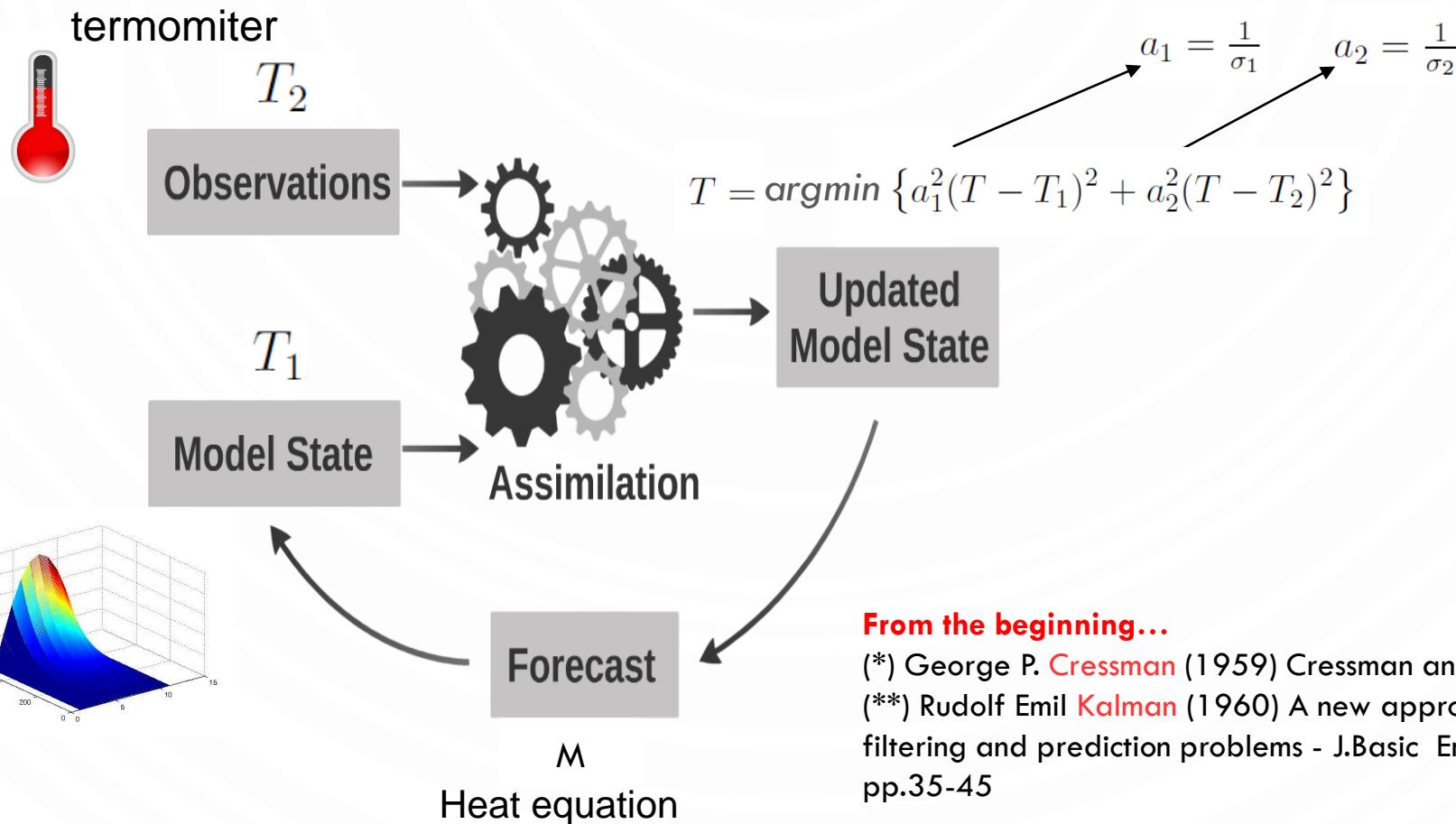
Data Assimilation: example...



Data Assimilation: example...



Background and observation
error covariance assumed to be Gaussian (*, **)



From the beginning...

(*) George P. Cressman (1959) Cressman analysis

(**) Rudolf Emil Kalman (1960) A new approach to linear filtering and prediction problems - J.Basic Emg. vol.82D, pp.35-45

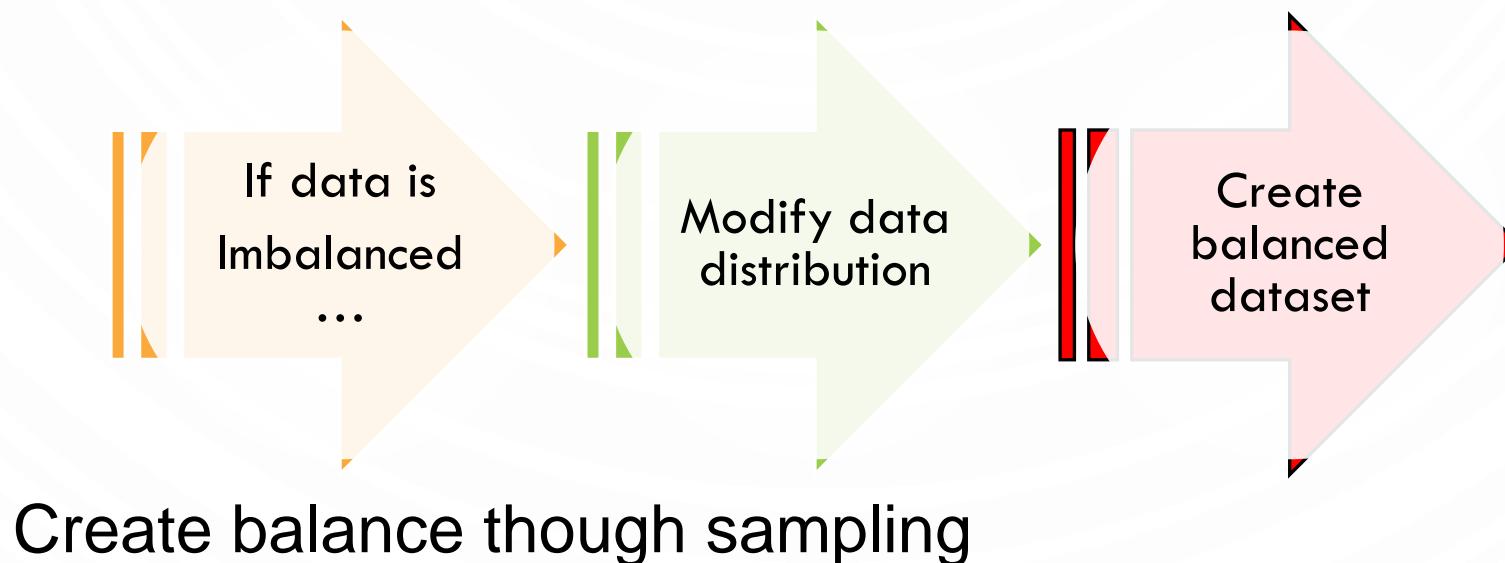
Intrinsic and extrinsic imbalance

Intrinsic:

- Imbalance due to the nature of the dataspace

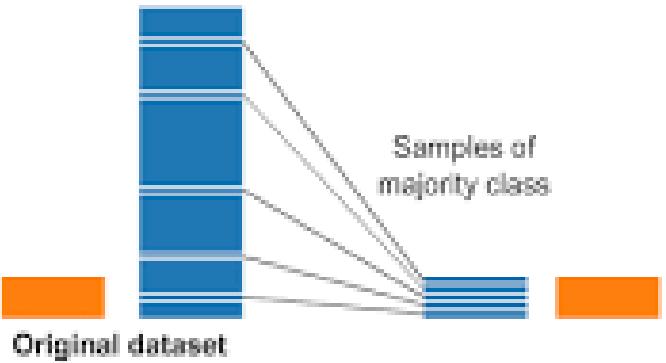
Extrinsic:

- Imbalance due to time, storage, and other factors

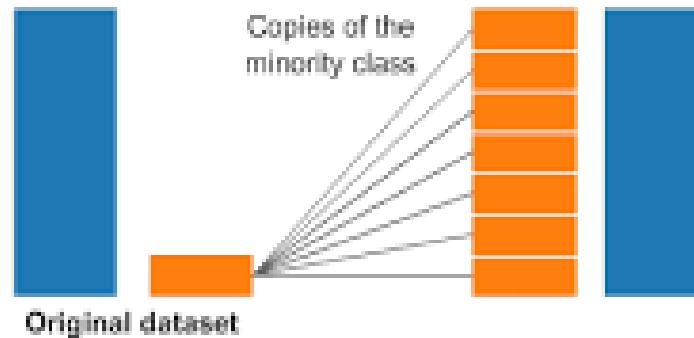


Undersampling and Oversampling

Undersampling



Oversampling



Loss of important information



Possible overfitting

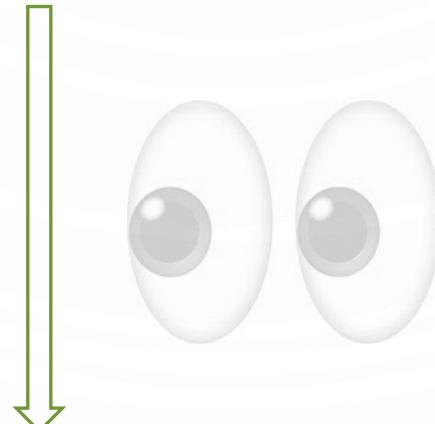
Undersampling and Oversampling

Unsupervised (undersampling):

use random subsets of the majority class to create balance and form multiple classifiers

Supervised (undersampling or oversampling):

iteratively (or not) create balance and pull out (or in) redundant (novel) samples in majority (minority) class to form a final classifier



Trying to reduce the error!!!

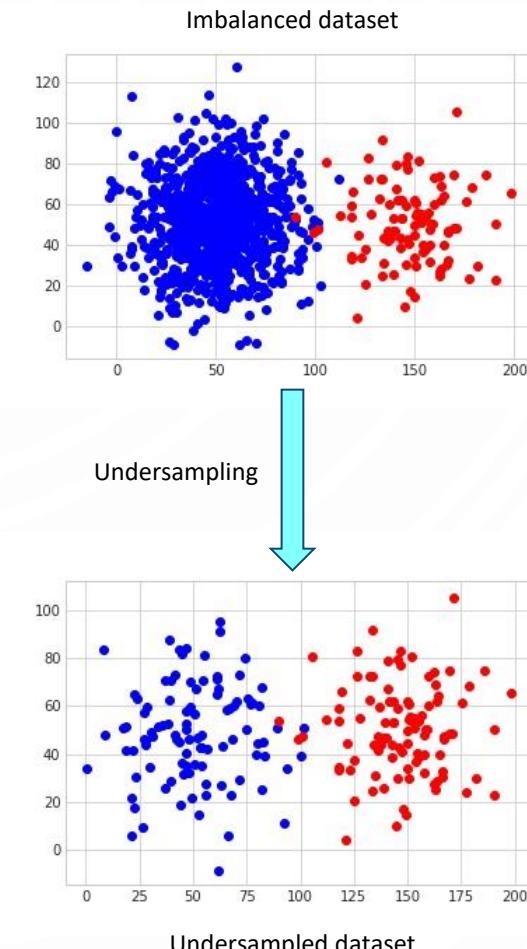
Undersampling the majority class

CORRECTING FOR IMBALANCED DATASETS

Undersampling the majority class

The idea here is to create a new dataset to train our algorithm that is not imbalanced. This is typically achieved by creating a new dataset where we undersample (pick with a lower probability) data points that belong to the majority class.

1. We keep all samples from the minority class. Let N be their number.
2. We draw ("sample") uniformly at random N data points from the majority class.
3. We use the new dataset, containing $2N$ data points (N for each class), to fit our model.



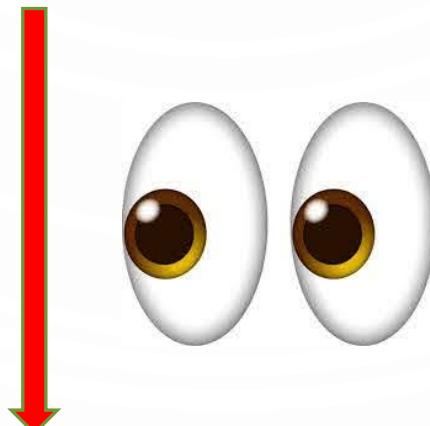
Undersampling and Oversampling

Unsupervised (undersampling):

use random subsets of the majority class to create balance and form multiple classifiers

Supervised (undersampling or oversampling):

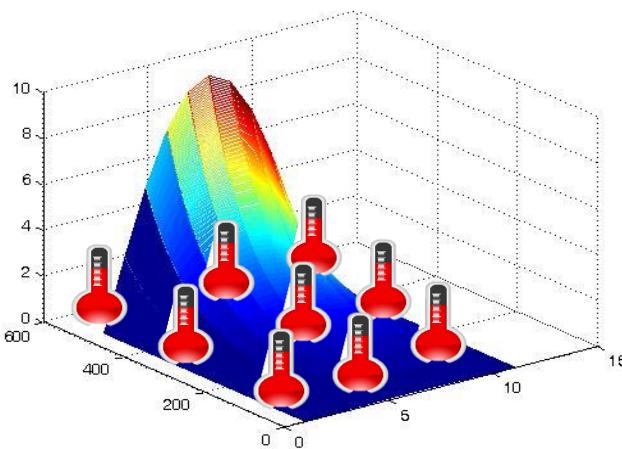
iteratively (or not) create balance and pull out (or in) redundant (novel) samples in majority (minority) class to form a final classifier



Trying to reduce the error!!!

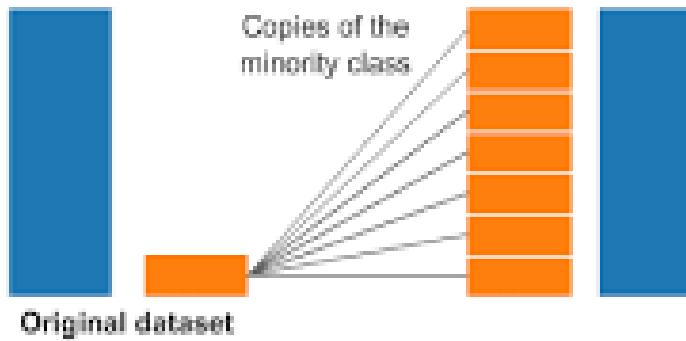
Oversampling

thermometer



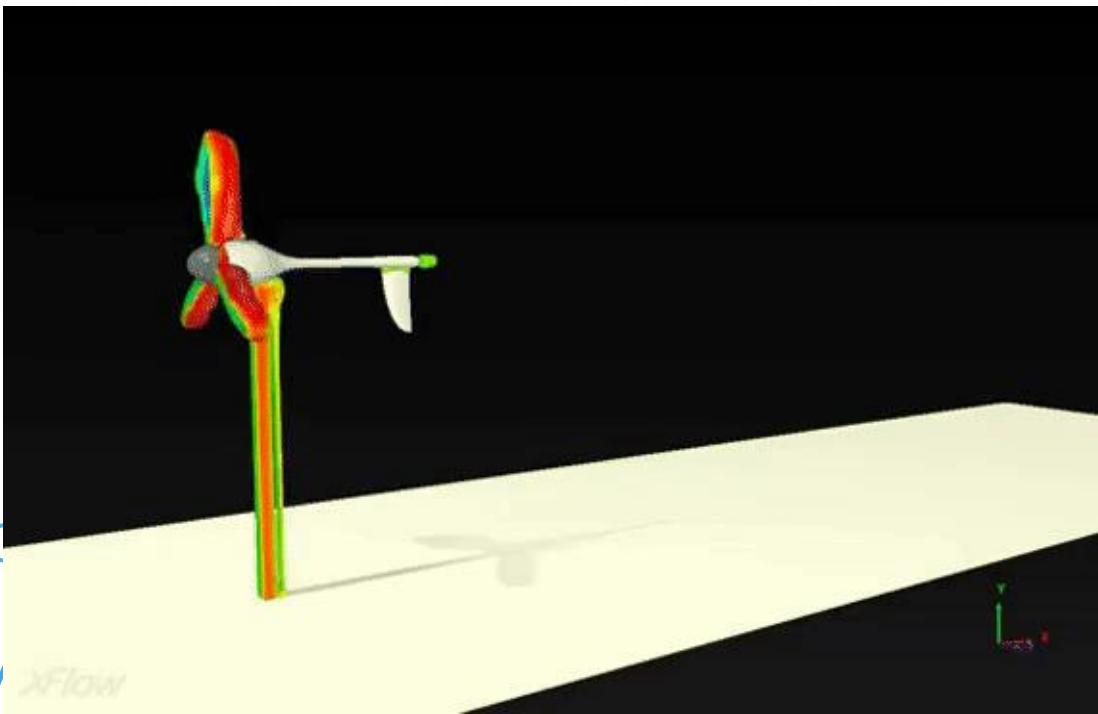
Computational model

Oversampling



observations

When working with Numerical Models...



- Finite precision introduces error
- Does not capture randomness of real situations
- Very sensitive to initial conditions

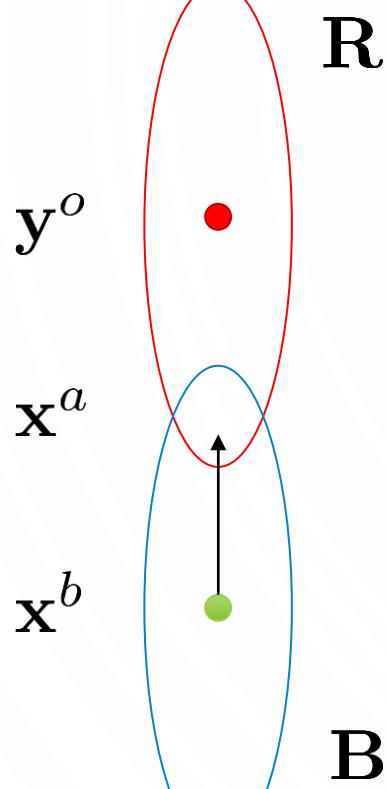


What is Data Assimilation

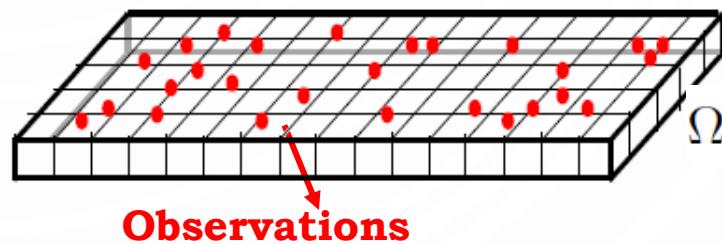
Data assimilation is a technique where observational data is incorporated

with the output of a numerical model or simulation to produce an optimal

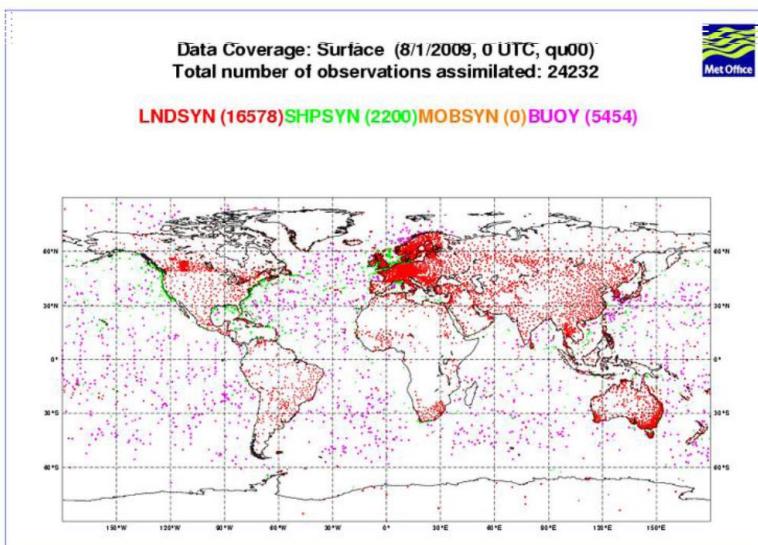
state of an evolving dynamic system



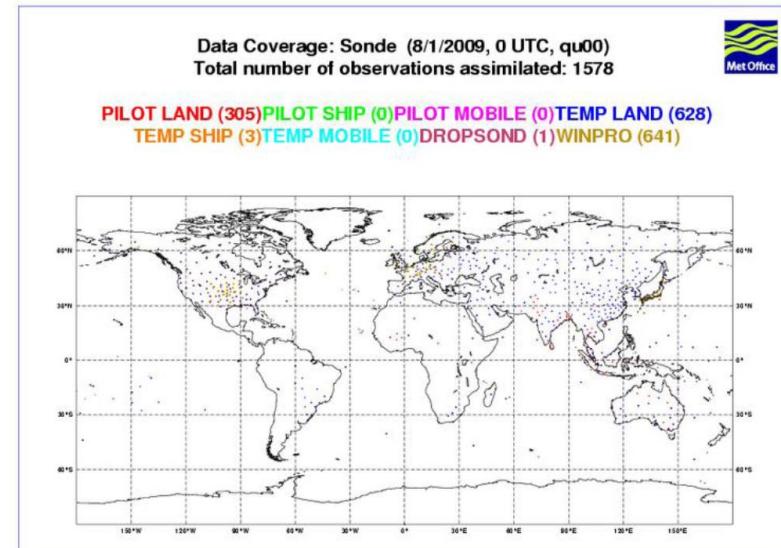
Given a set of observations, y^o with known error R , correct the background state or model output x^b with known error B , to output an analysis, x^a



Observations: why not just use these data?



Surface



Radiosonde

The data are not "dense" enough and the math problem to compute the solution "everywhere" is an ill posed inverse problem



The Tikhonov-regularized formulation



MODEL EQUATIONS

Full Model

$$\mathbf{x}_{k+1} = \mathcal{M}_{k+1} \mathbf{x}_k$$

\mathcal{M}_k = Model Operator at time k

$$\mathbf{y}_k^o = \mathcal{H}_k \mathbf{x}_k + \epsilon_k^o$$

\mathcal{H}_k = Observation Operator at time k

Tangent Linear Model

$$\mathbf{x}_{k+1} = \mathbf{M}_{k+1} \mathbf{x}_k$$

$$\mathbf{M}_{k+1} = \text{Linearized } \frac{\delta \mathbf{x}_{k+1}}{\delta \mathbf{x}_k}$$

$$\mathbf{y}_k^o = \mathbf{H}_k \mathbf{x}_k + \epsilon_k^o$$

$$\mathbf{H}_{k+1} = \text{Linearized } \frac{\delta \mathbf{y}_k^o}{\delta \mathbf{x}_k}$$

The Tikhonov-regularized formulation

FULL FIELD COST FUNCTION (TIKHONOV REGULARISATION)

$$J(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + \frac{1}{2} (\mathcal{H}\mathbf{x} - \mathbf{y}^o)^T \mathbf{R}^{-1} (\mathcal{H}\mathbf{x} - \mathbf{y}^o)$$

$$\nabla J(\mathbf{x}) = \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) - \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y}^o - \mathbf{H}\mathbf{x}) = \cancel{\mathbf{0}}$$

\mathbf{B} = Background Error covariance matrix

Use L-BFGS Method to
Iteratively Solve

\mathbf{R} = Observation Error covariance matrix

- Difficult to minimise directly
- Multiple local minima

The Tikhonov-regularized formulation – 4D

FULL FIELD COST FUNCTION (TIKHONOV REGULARISATION)

$$J(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + \frac{1}{2} (\mathbf{G}\mathbf{x} - \mathbf{y}^o)^T \mathbf{R}^{-1} (\mathbf{G}\mathbf{x} - \mathbf{y}^o)$$

diag [$\mathbf{H}_0, \mathbf{H}_1\mathbf{M}_{0,1}, \dots, \mathbf{H}_{N-1}\mathbf{M}_{N-2,N-1}$]

From the Tikhonov-regularized formulation



NOT considering the temporal variable **considering the temporal variable**

$$\min \{ \|Hu - v\|_R^2 + \|u - u_0\|_B^2 \}$$

$$\min \{ \|Gu - v\|_R^2 + \|u - u_0\|_B^2 \}$$

Variational approach

$$J(u) = \|Hu - v\|_R^2 + \|u - u_0\|_B^2$$

$$u^{DA} = \min_u \{ J(u) \}$$

3DVariational

$$J(u) = (Hu - v)^T R^{-1} (Hu - v) + (u - u_0)^T B^{-1} (u - u_0)$$

Normal equations

$$S = \|Hu - v\|_R^2 + \|u - u_0\|_B^2$$

$$\frac{\partial S}{\partial u} = 0$$

**Kalman Filter,
Optimal Interpolation**

$$u^{DA} = u_0 + K(Hu - v)$$

$$K = RH^T(H^T R H + B)^{-1}$$

Variational approach

$$J(u) = \|Gu - v\|_R^2 + \|u - u_0\|_B^2$$

$$u^{DA} = \min_u \{ J(u) \}$$

4DVariational

$$J(u) = (Gu - v)^T R^{-1} (Gu - v) + (u - u_0)^T B^{-1} (u - u_0)$$

Normal equations

$$S = \|Gu - v\|_R^2 + \|u - u_0\|_B^2$$

$$\frac{\partial S}{\partial u} = 0$$

**Ensemble/Extended
Kalman Filter**

$$u^{DA} = u_0 + K(Gu - v)$$

$$K = RG^T(G^T R G + B)^{-1}$$

Data Assimilation: a Big Data problem

1. **High resolution** models
2. The domain of validity of **the linear hypothesis**
shrink with increasing resolution and integration length

So **high resolution linearized models**
require time-consuming
computations.

Data Assimilation: a Big Data problem

DA is a large size computational problem

that should be solved in near **real-time**.

1. **High resolution** models
2. The domain of validity of **the linear hypothesis**
shrink with increasing resolution and integration length

So **high resolution linearized models**
require time-consuming
computations.

Data Assimilation: a Big Data problem

DA is a large size computational problem

that should be solved in near real-time.

1. **High resolution** models
2. The domain of validity of **the linear hypothesis**
shrink with increasing resolution and integration length

So **high resolution linearized models**
require time-consuming
computations.



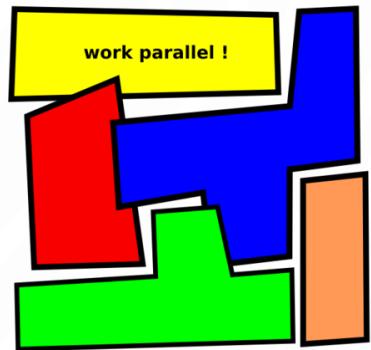
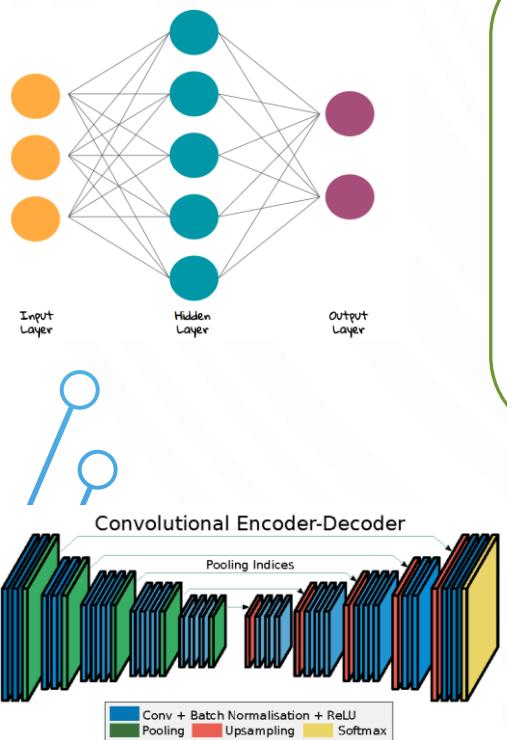
due to the computation complexity of the DA model,
it is impossible to solve a DA problem
by using high resolutions for the model.

Two main “possibilities”

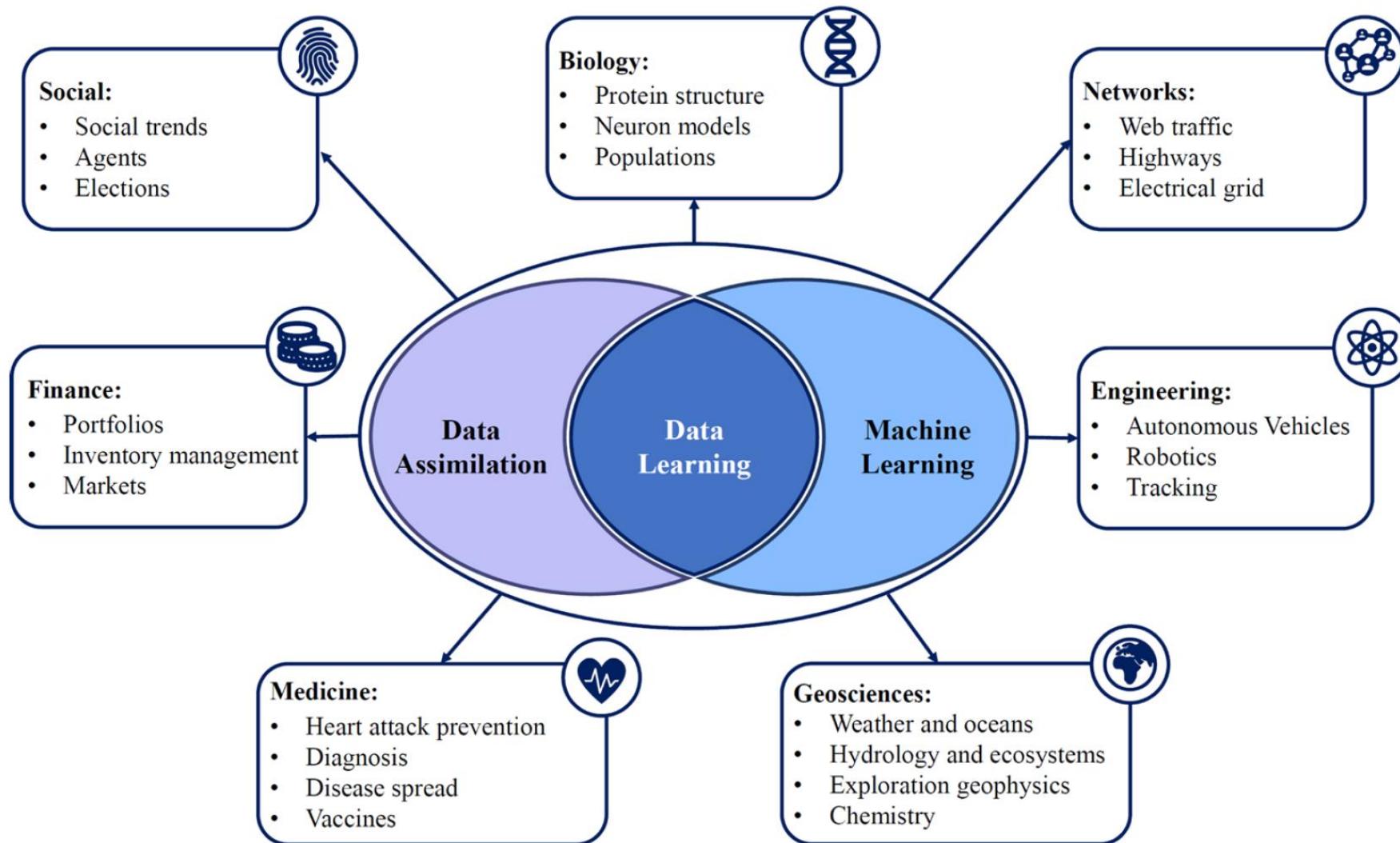
To use some simplifications,
(e.g. reduced models,
nested models,
etc..)

To use some decompositions,
to exploit the high performance of
emerging computing architectures

to develop scalable algorithms



Data Assimilation + Machine Learning = Data Learning



Data Learning: a modular approach

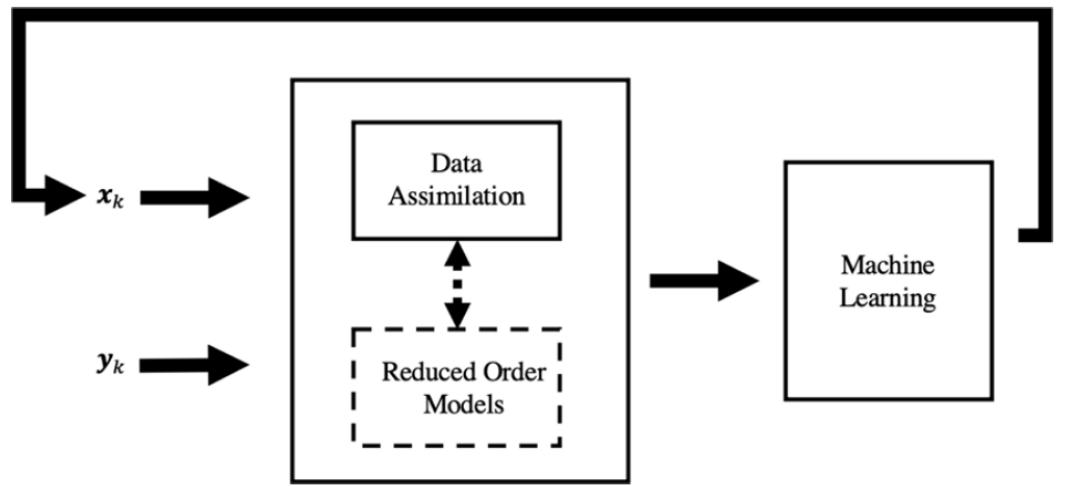
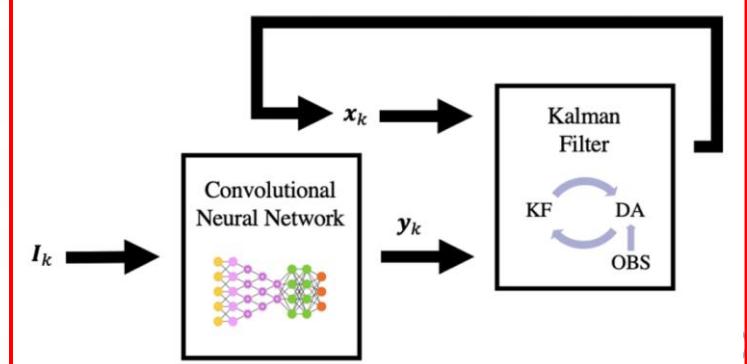
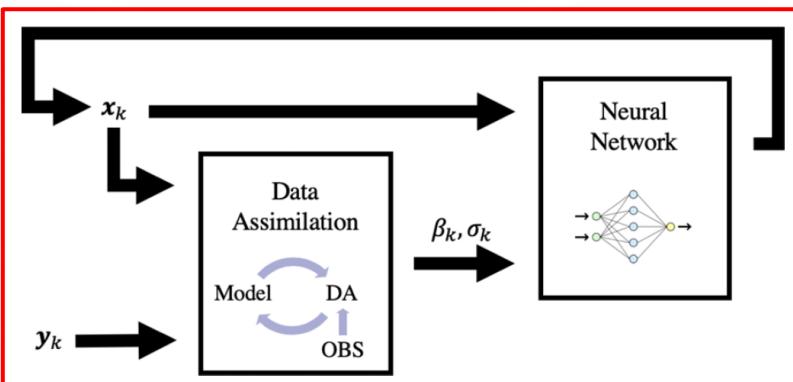
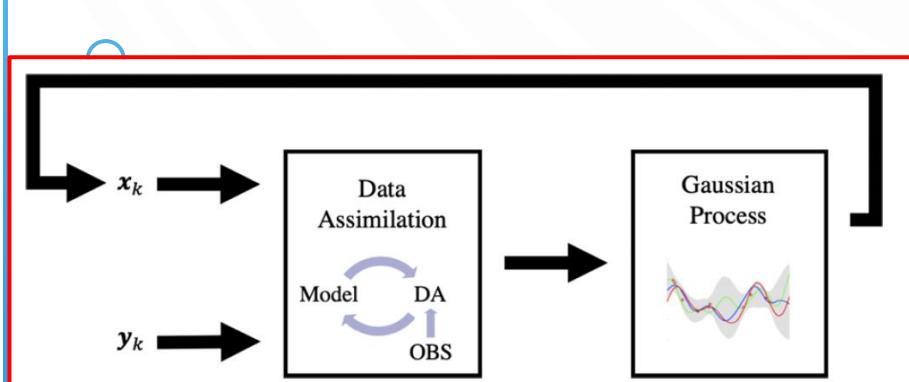
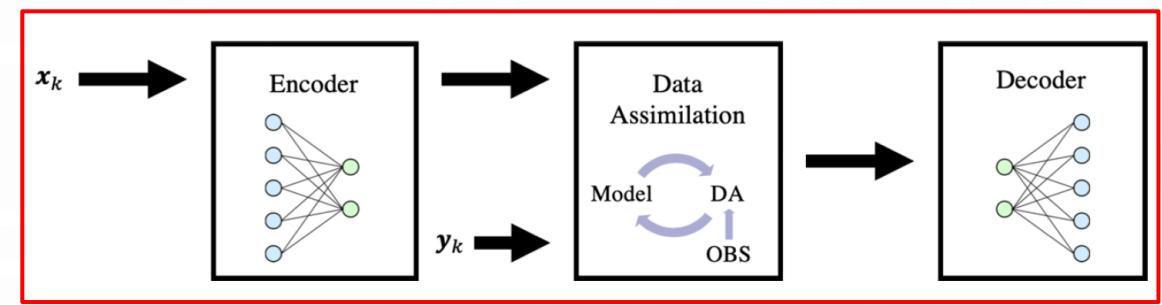
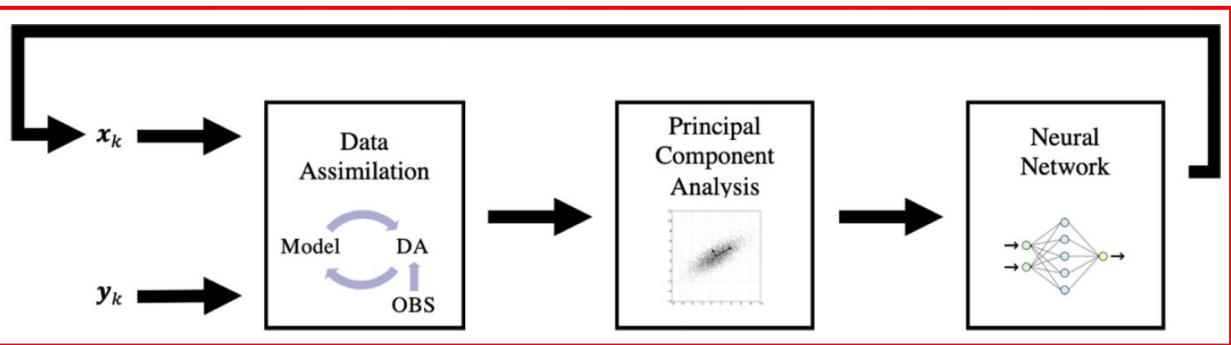
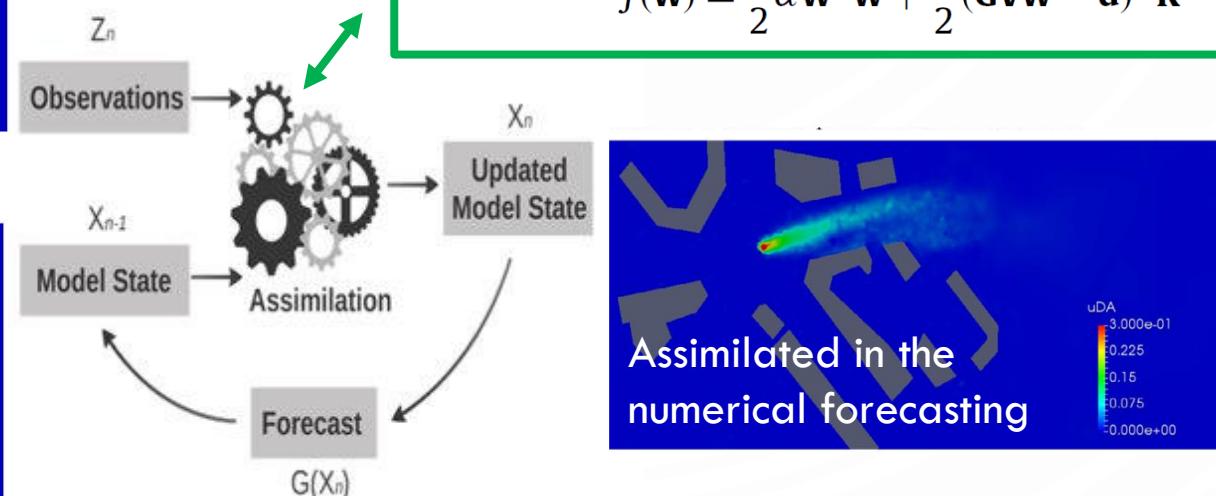
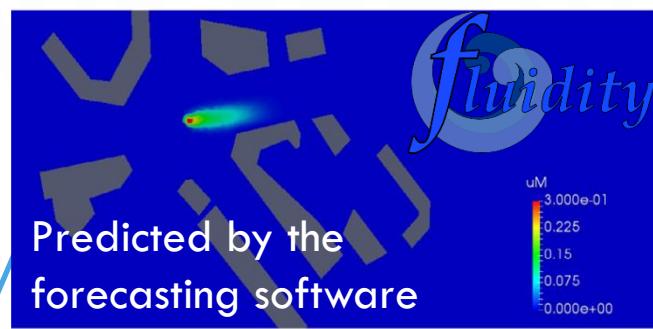
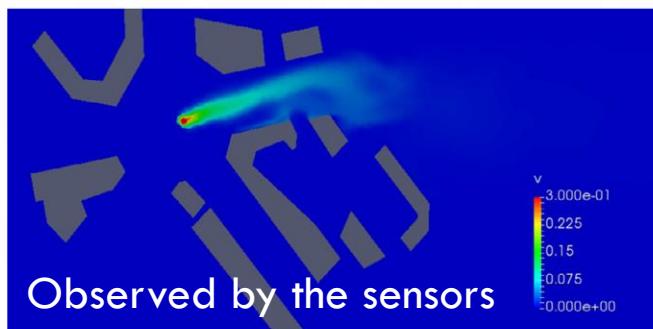


Fig. 1. General Data Learning framework.





Data Assimilation or Latent Assimilation?



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

3DVar in the control space

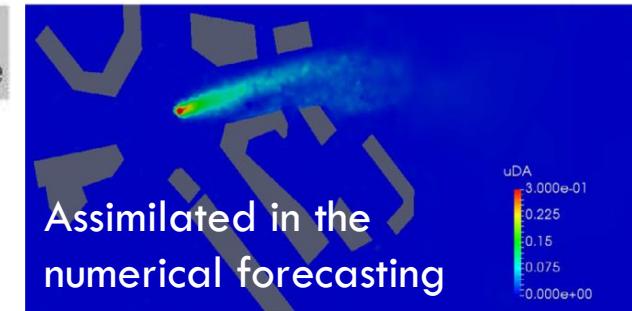
$$\mathbf{w} = \mathbf{V}^+ \delta \mathbf{u} \quad \mathbf{B} = \mathbf{V} \mathbf{V}^T \quad \text{Reduced space, TSVD}$$

$$\mathbf{w}^{DA} = \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^{NP \times N}} J(\mathbf{w})$$

with

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

$$\sigma = \mu c \delta$$



EPSRC



MAGIC

Envisaging a world with greener cities

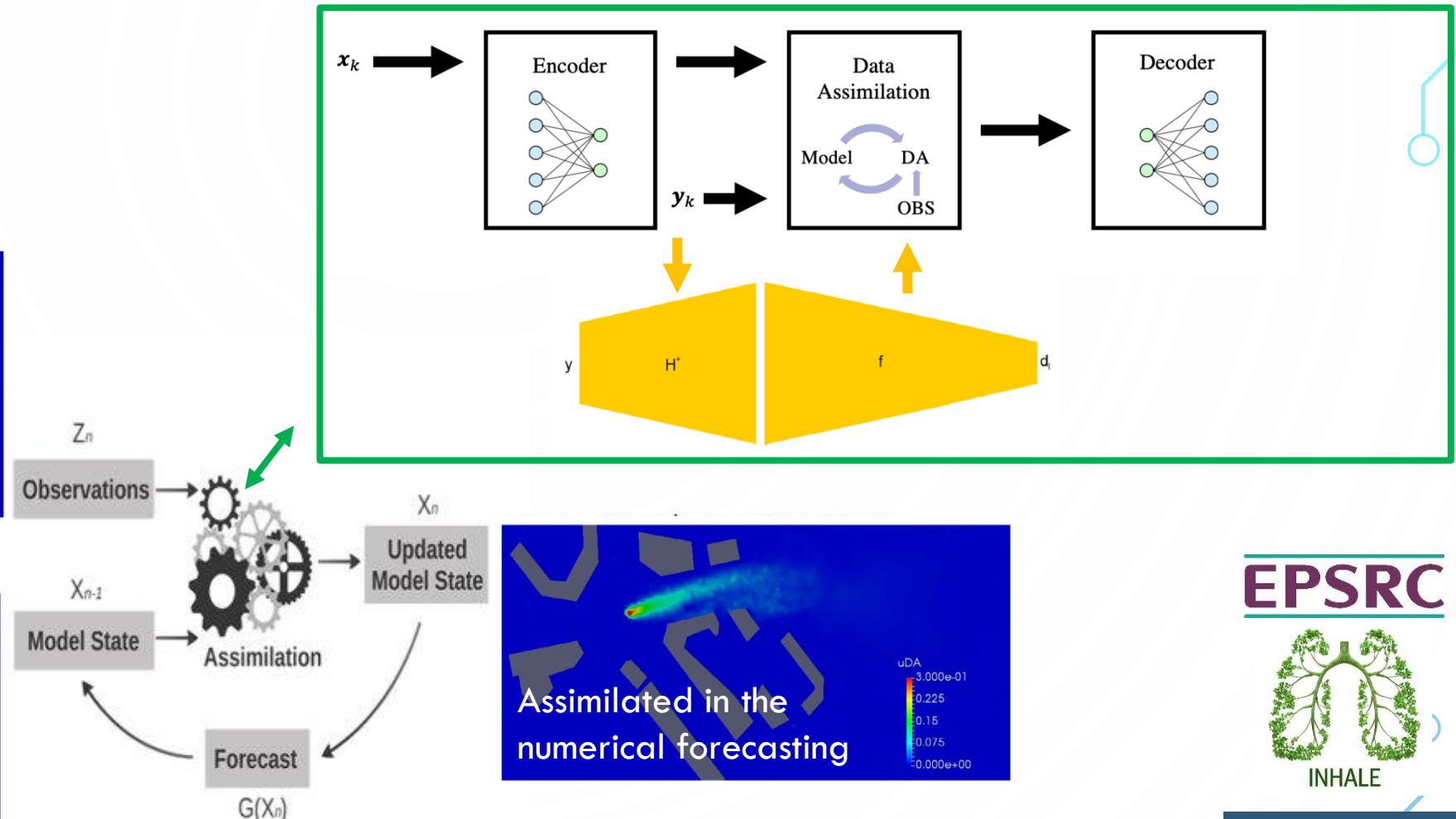
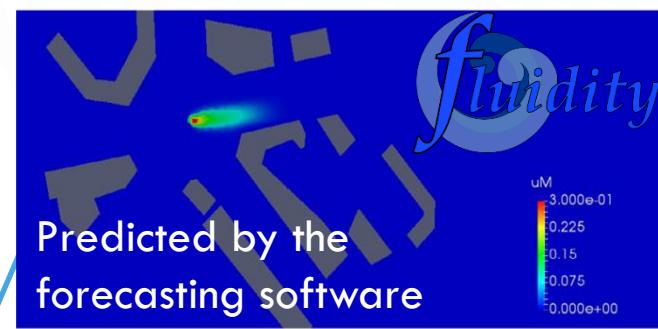
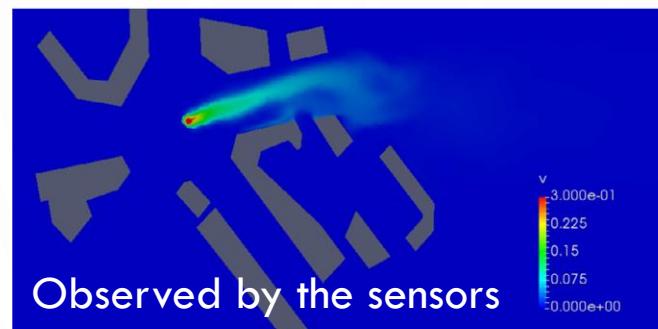
[*] R. Arcucci, L. Mottet, C. Pain and Y. Guo - **Optimal reduced space for Variational Data Assimilation** -Journal of Computational Physics

[**] 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 or Latent Assimilation?



EPSRC



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

Assimilating Images

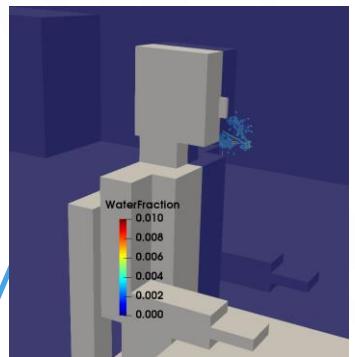
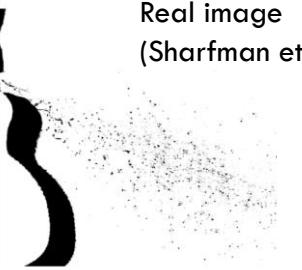


Assimilating Images

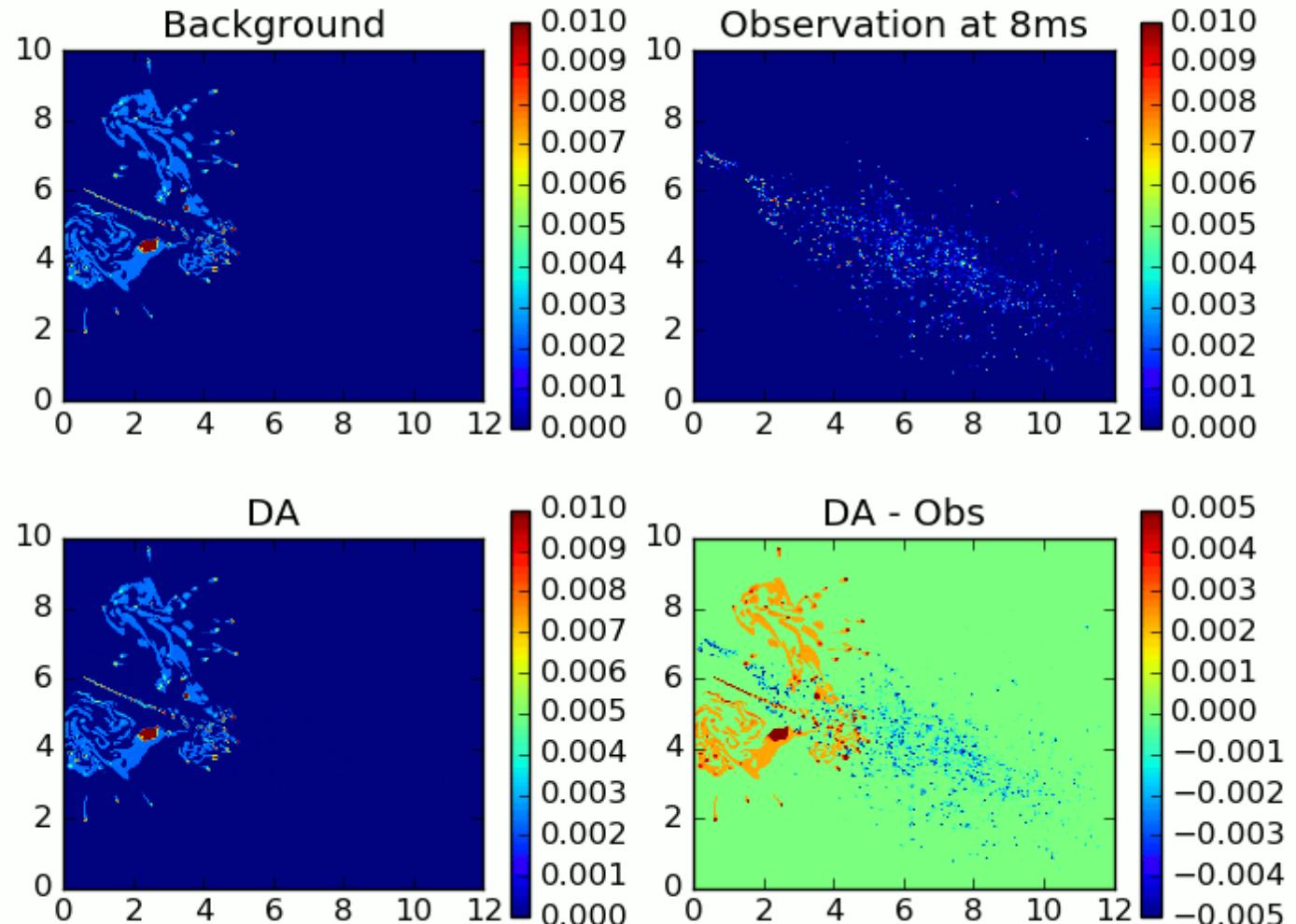
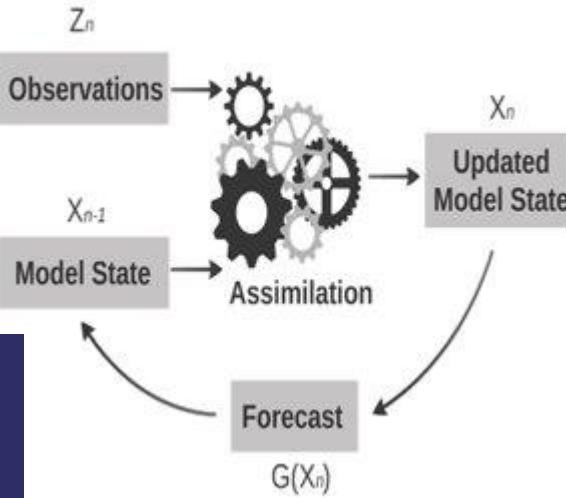
Observed data: experiments

Real image
(Sharfman et al, 2016)

time=8ms



CFD simulation or Surrogate model



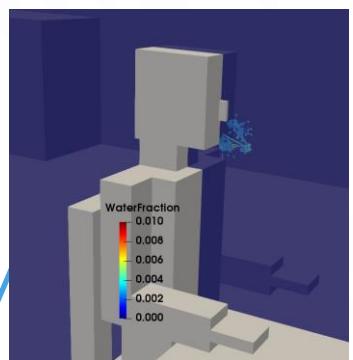
Assimilating Images

COVID-19: GENERIC CASE - SIMULATION OF DISPERSION IN A CLASS

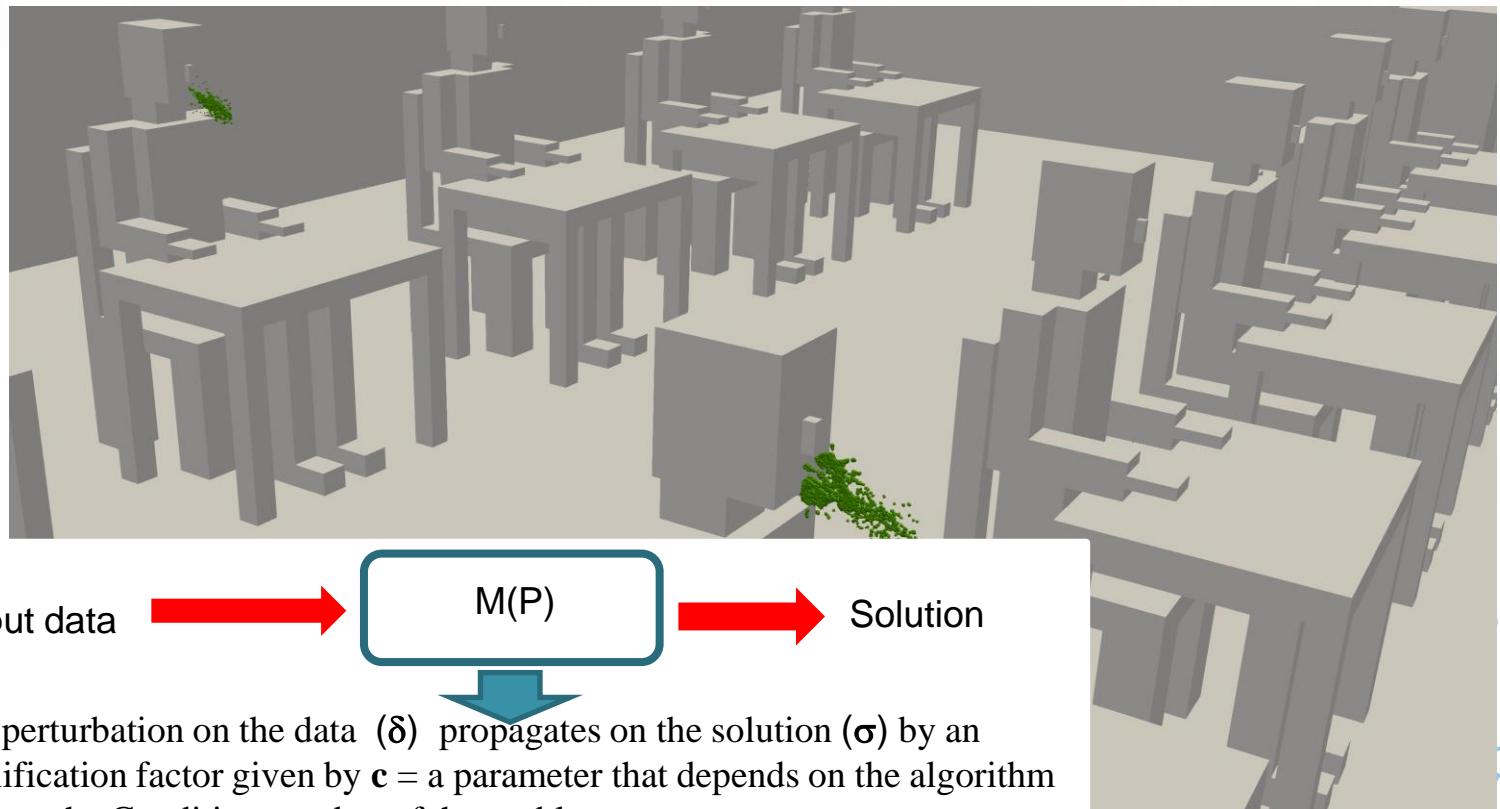
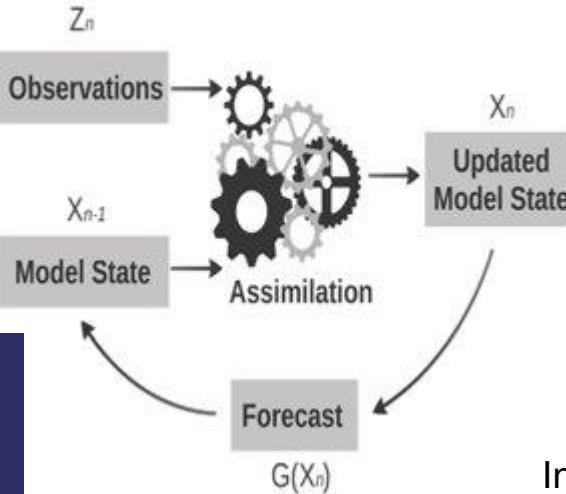
Green iso-surface shows dispersion of passive tracer.

Observed data: experiments

Real image
(Sharfman et al, 2016)



CFD simulation or Surrogate model



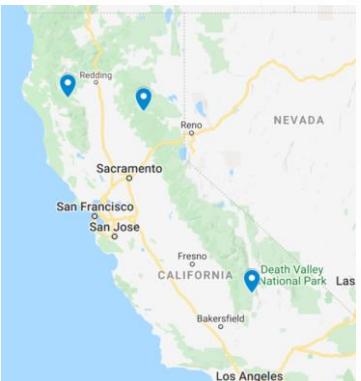
any perturbation on the data (δ) propagates on the solution (σ) by an amplification factor given by $c = \mu$ a parameter that depends on the algorithm and $\mu = \text{the Condition number of the problem}$

$$\text{Error on the solution: } \sigma = \mu c \delta$$

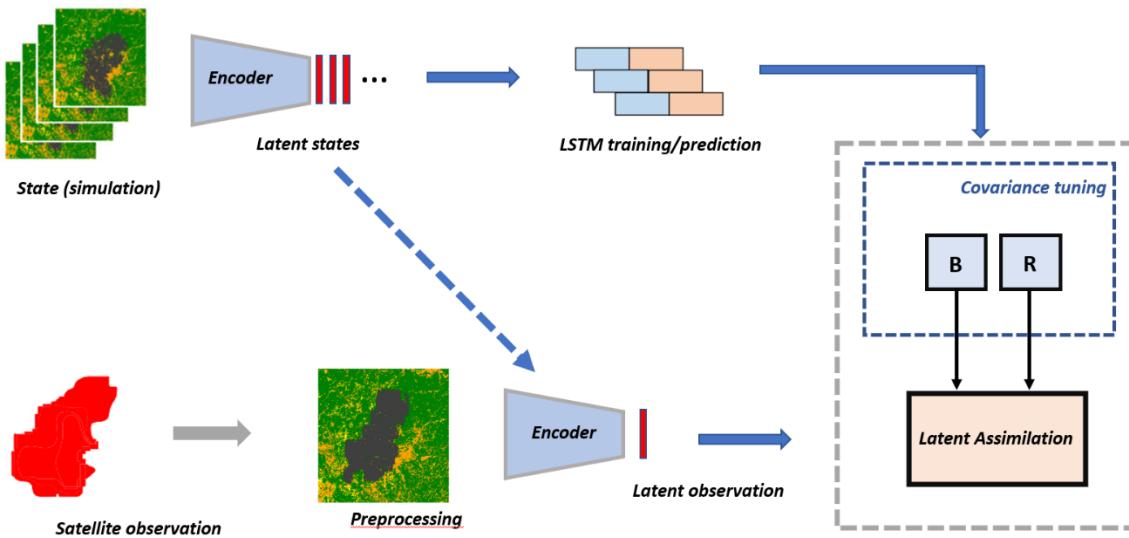
Wildfire forecasting



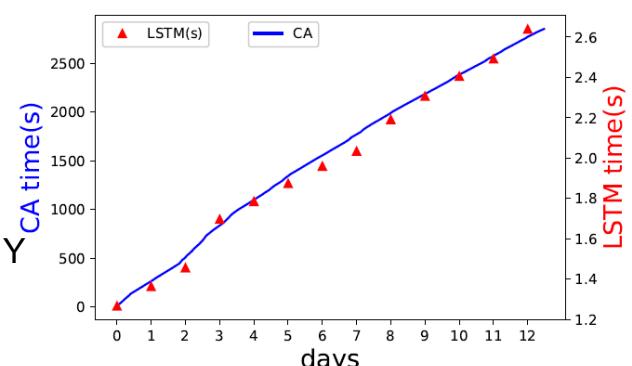
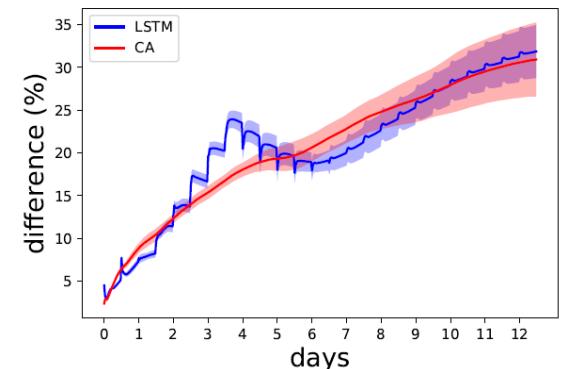
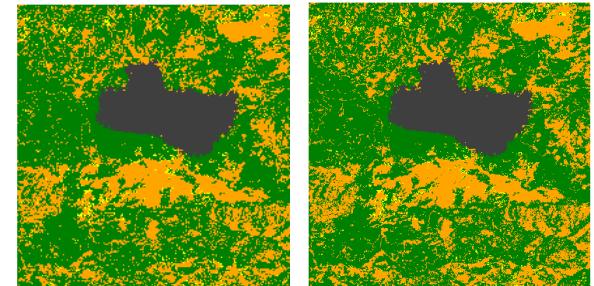
MODIS: every 1-2 days at 1km resolution



- Learning from simulation data
- Using satellite observations to validate/assimilate



observation prediction



[*] Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting S Cheng, IC Prentice, Y Huang, Y Jin, YK Guo, R Arcucci - Journal of Computational Physics, 111302

[**] Parameter Flexible Wildfire Prediction Using Machine Learning Techniques: Forward and Inverse Modelling S Cheng, Y Jin, SP Harrison, C Quilodrán-Casas, IC Prentice, YK Guo, ..., R. Arcucci - Remote Sensing 14 (13), 3228

Human Sensors!



In 2021, the number of mobile devices operating worldwide stood at almost **15 billion**



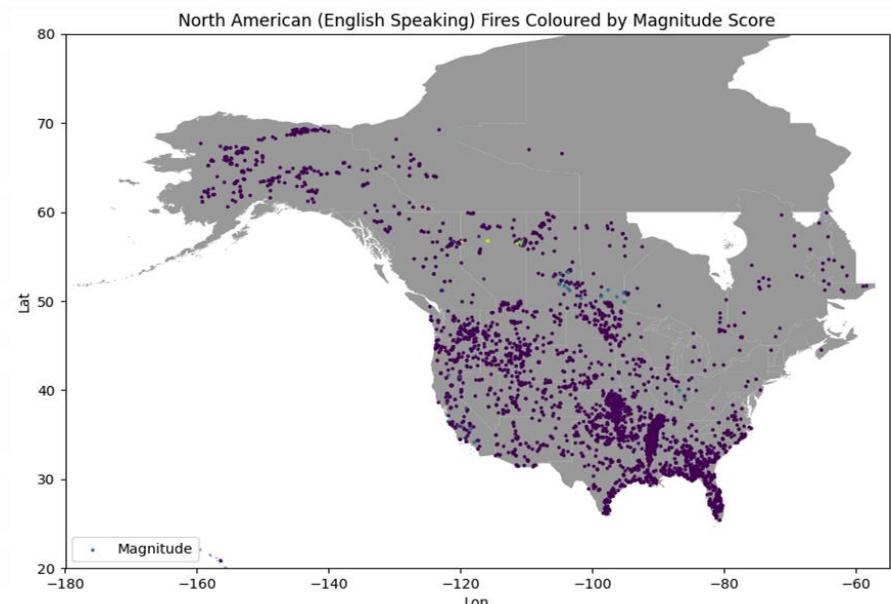
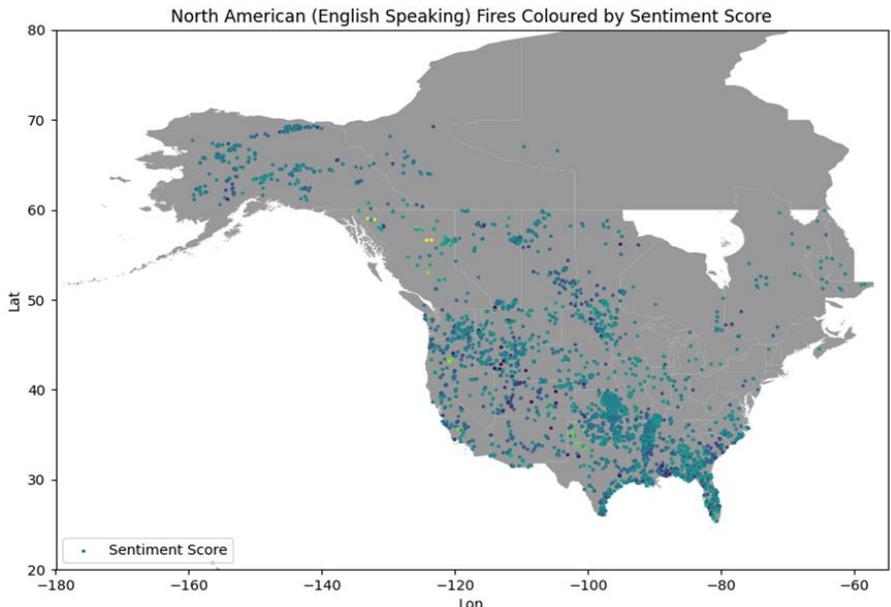
In 2020, **over 3.6 billion people** were using social media worldwide, a number projected to increase to almost 4.41 billion in 2025.



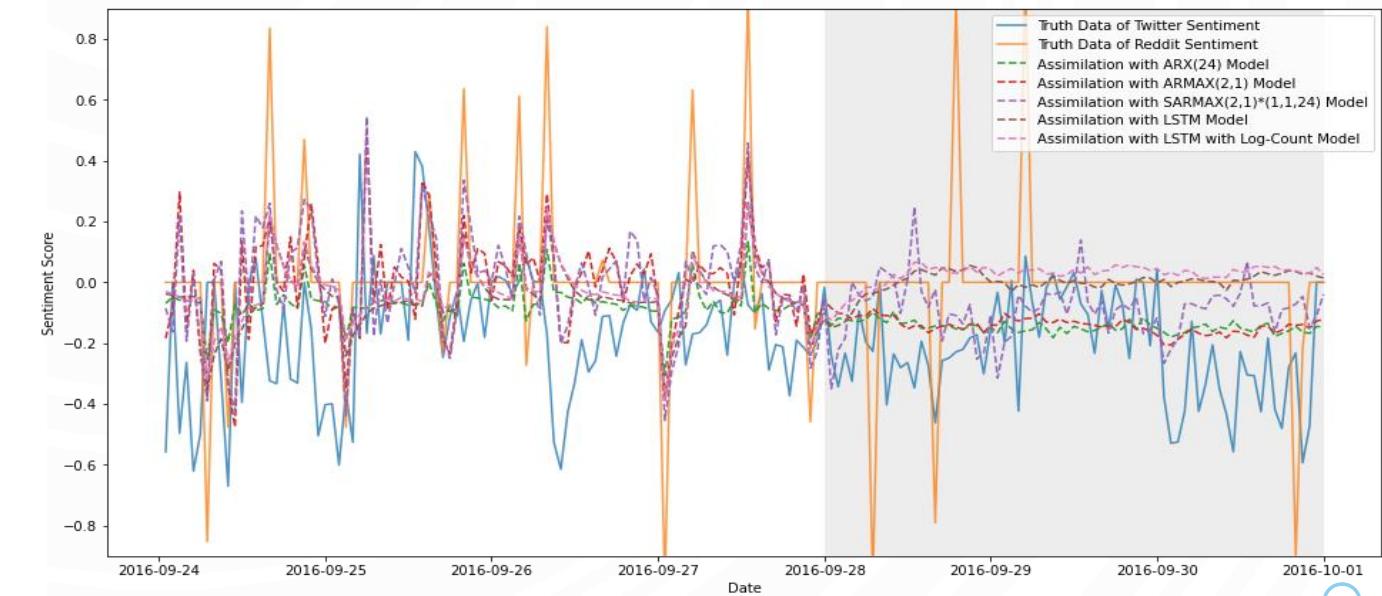
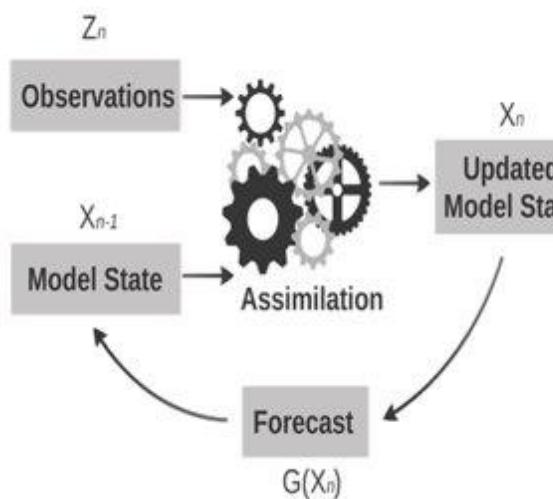
Social data and satellites: Wildfires System



- How do people perceive wildfires? Can this be measured or modelled?
- Idea: can we collect many subjective opinions on certain natural events, and are these opinions reflective of the size and severity of the event?
- Social media and Twitter - human sensors; Sentiment analysis - Converting emotional leaning in a passage of text into a numerical value, evaluating the positivity (or negativity) of the emotions expressed.



Social data and satellites: Wildfires System



[*] Social Data Assimilation of Human Sensor Networks for Wildfires

J Lever, R Arcucci, J Cai - Proceedings of the 15th International Conference PETRA

[**] Sentimental wildfire: a social-physics machine learning model for wildfire nowcasting

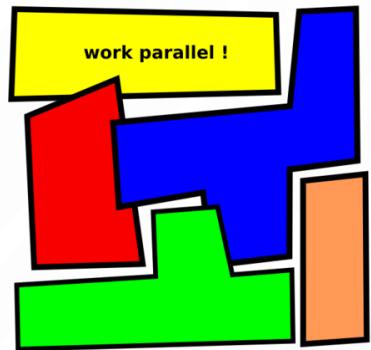
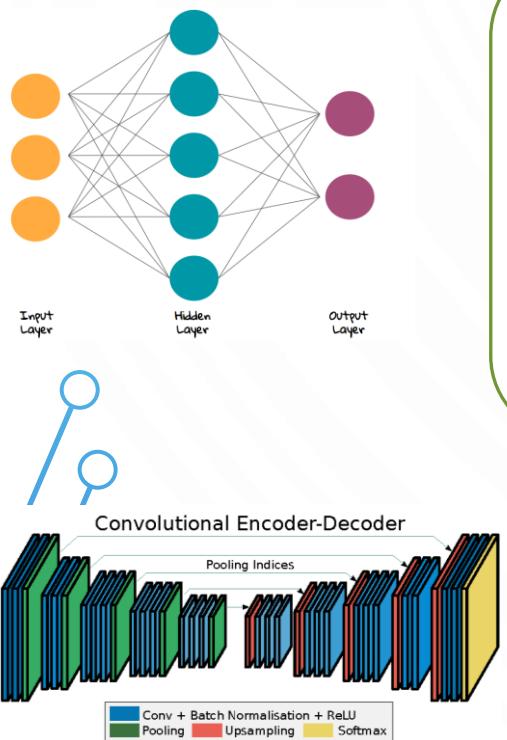
J Lever, R Arcucci - Journal of Computational Social Science, 1-39

Two main “possibilities”

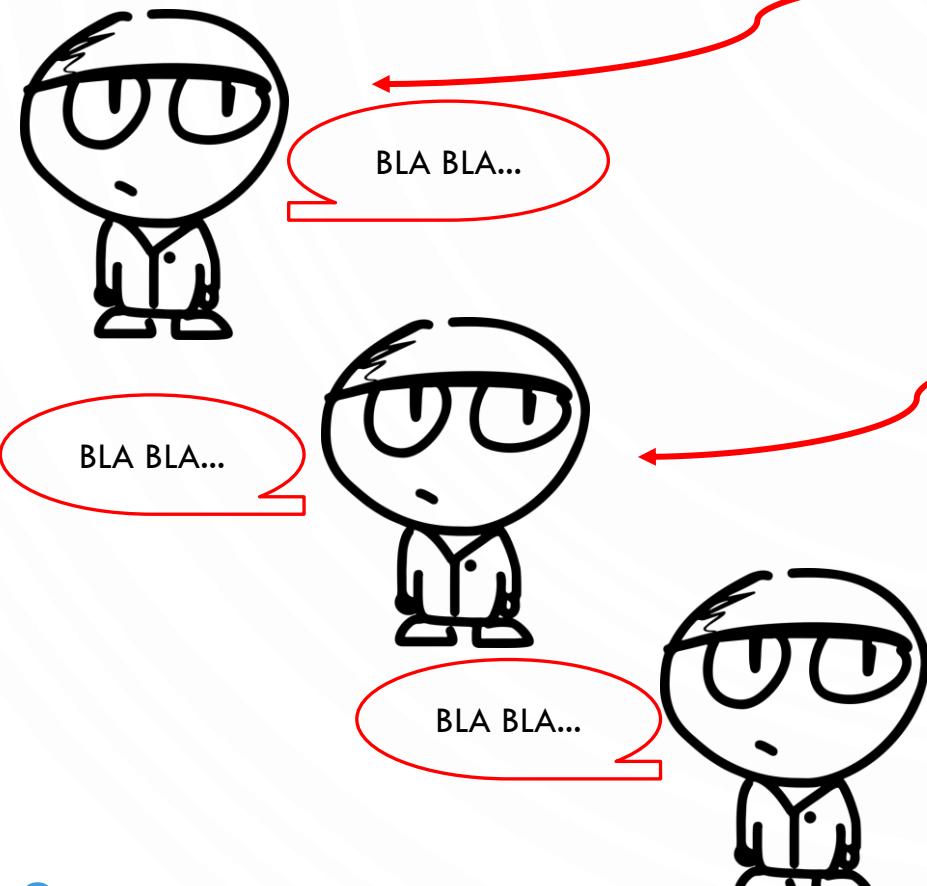
To use some simplifications,
(e.g. reduced models,
nested models,
etc..)

To use some decompositions,
to exploit the high performance of
emerging computing architectures

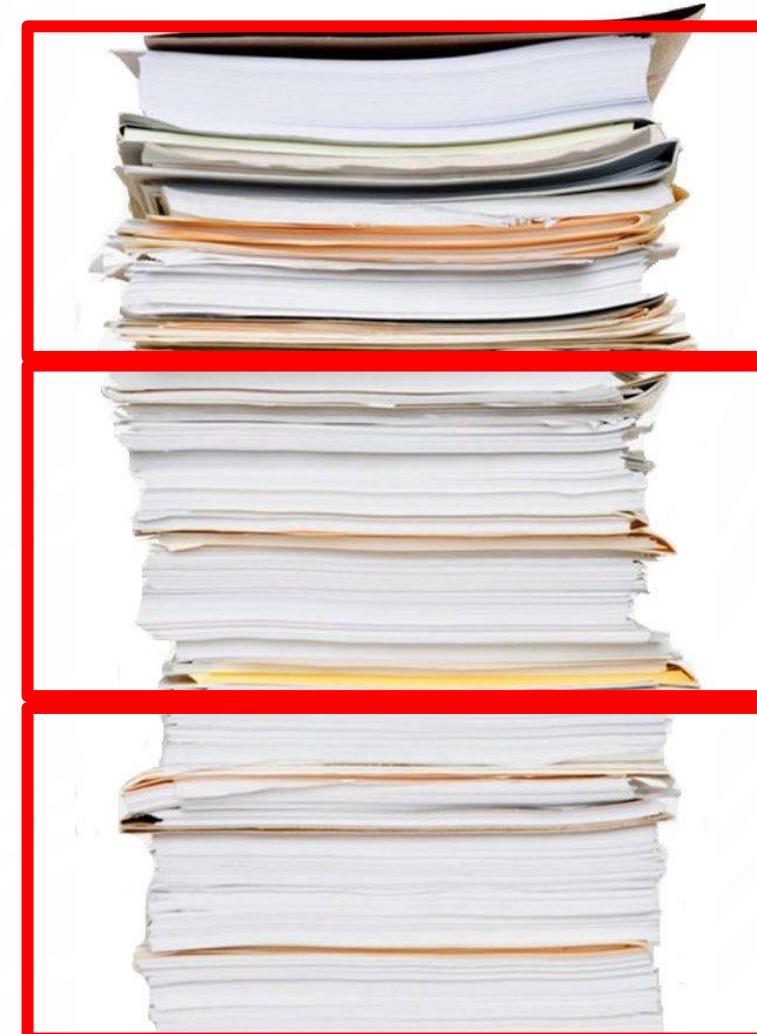
to develop scalable algorithms



... to develop scalable algorithms



Parallel computing

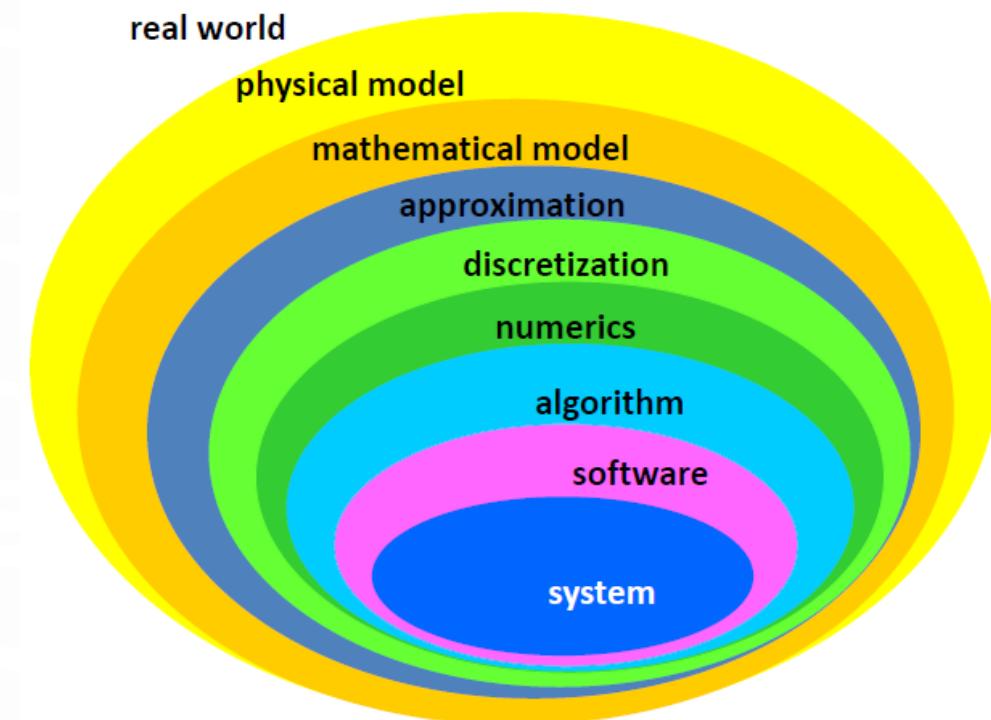


Domain Decomposition

Domain Decomposition for Data Assimilation



Mathematical Stack

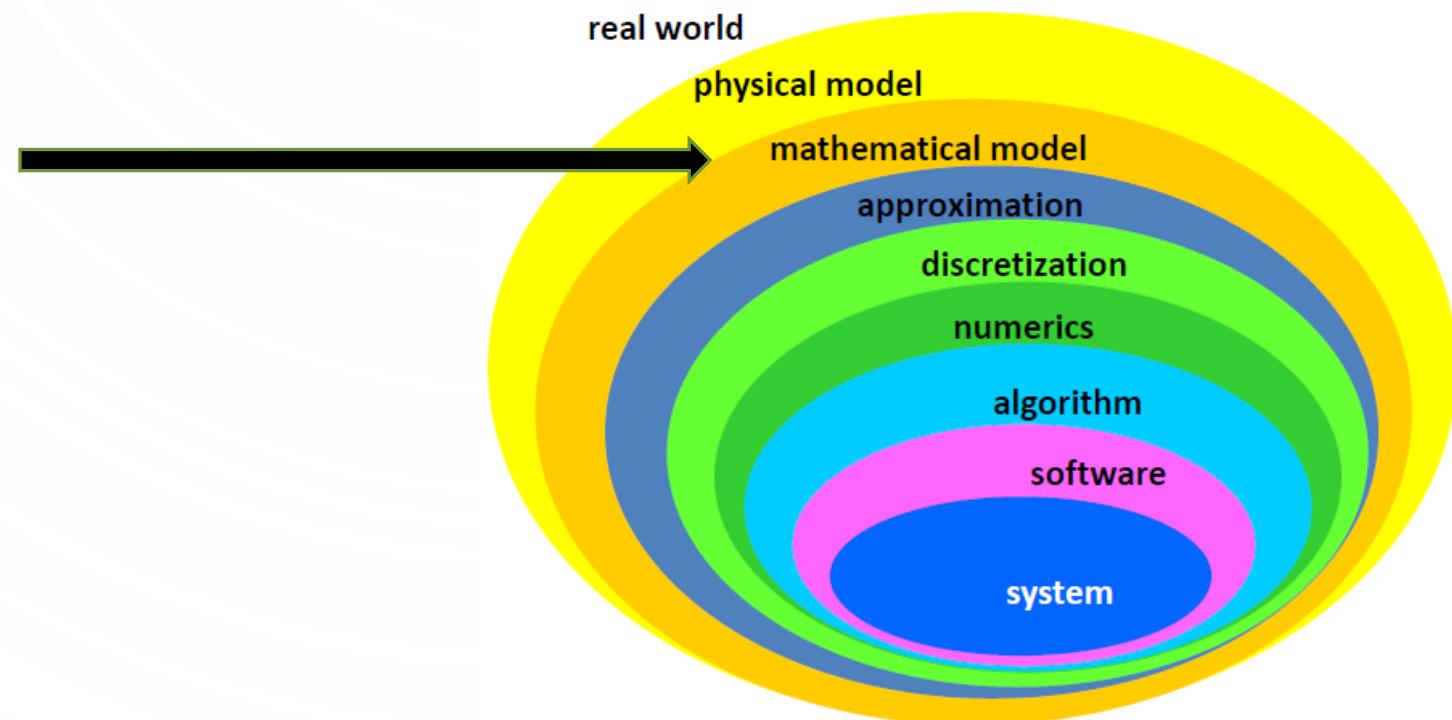


Domain Decomposition for Data Assimilation

«Adapting old programs to fit new machines usually means adapting new machines to behave like old ones.»

Alan Perlis

Parallelization should be introduced at the beginning...



Domain Decomposition for Data Assimilation

Let $\Omega \subset \mathbb{R}^S$ be decomposed into a sequence of p overlapping sub-domains $\Omega_i \subset \mathbb{R}^{r_i}$, $r_i \leq N$, $i = 1, \dots, p$ such that

$$\Omega = \bigcup_{i=1}^p \Omega_i$$

where

$$\Omega_i \cap \Omega_j = \Omega_{ij} \neq \emptyset$$

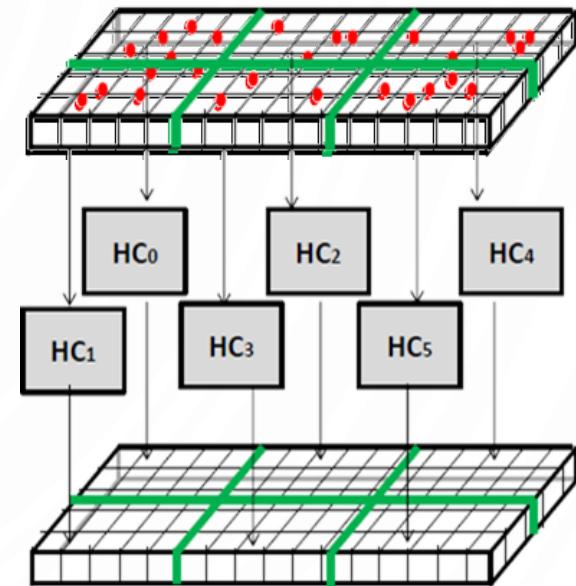
if the sub-domains are adjacent.

$$\mathbf{u}_i^{DA} = \operatorname{argmin}_{\mathbf{u}_i} \left\{ J/\Omega_i + O/\Omega_{ij} \right\}$$

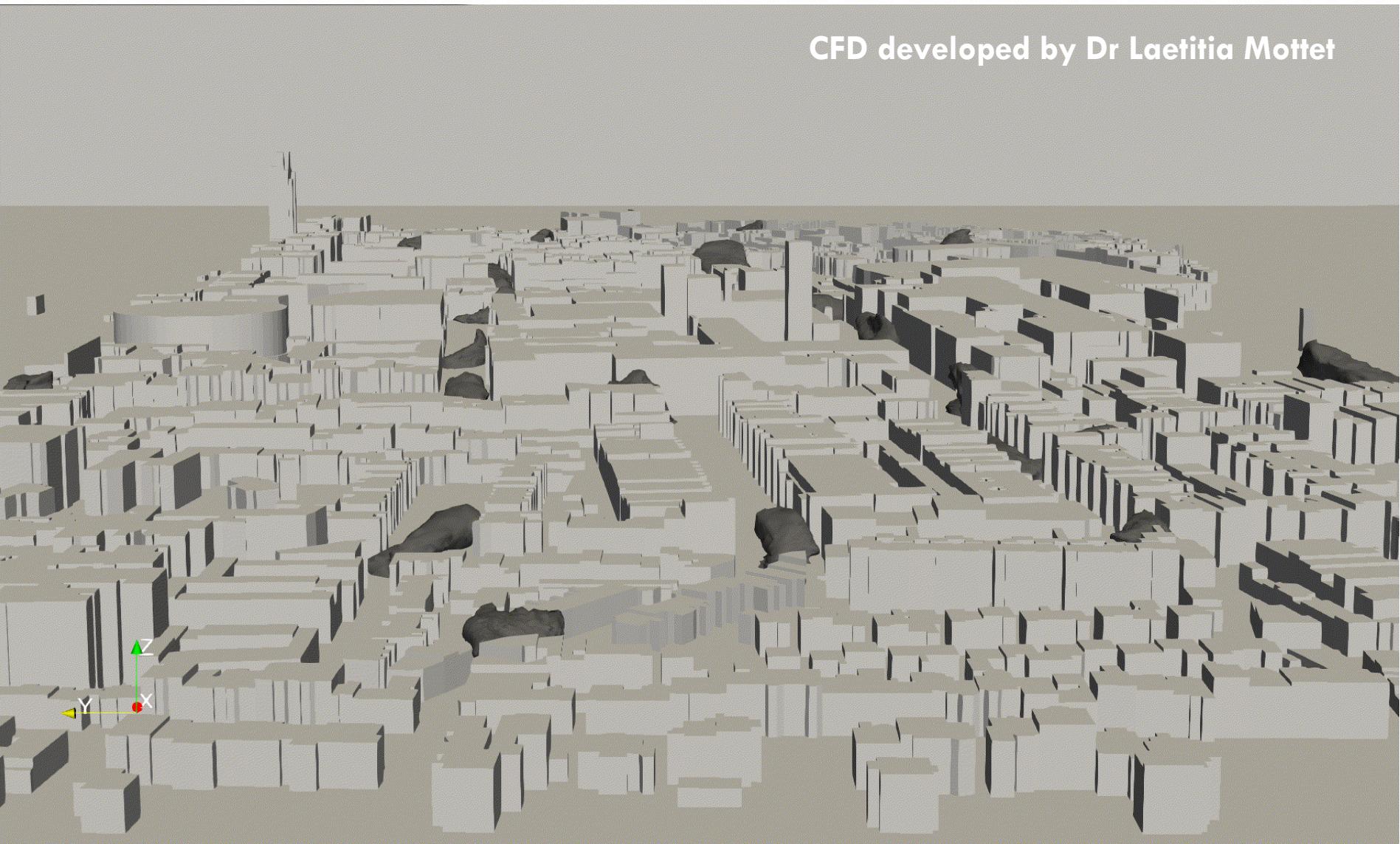
where O/Ω_{ij} is called overlapping operator.



$$J(\mathbf{u}) = \left\{ \|\mathbf{G}\mathbf{u} - \mathbf{v}\|_R^2 + \|\mathbf{u} - \mathbf{u}_0\|_B^2 \right\}$$

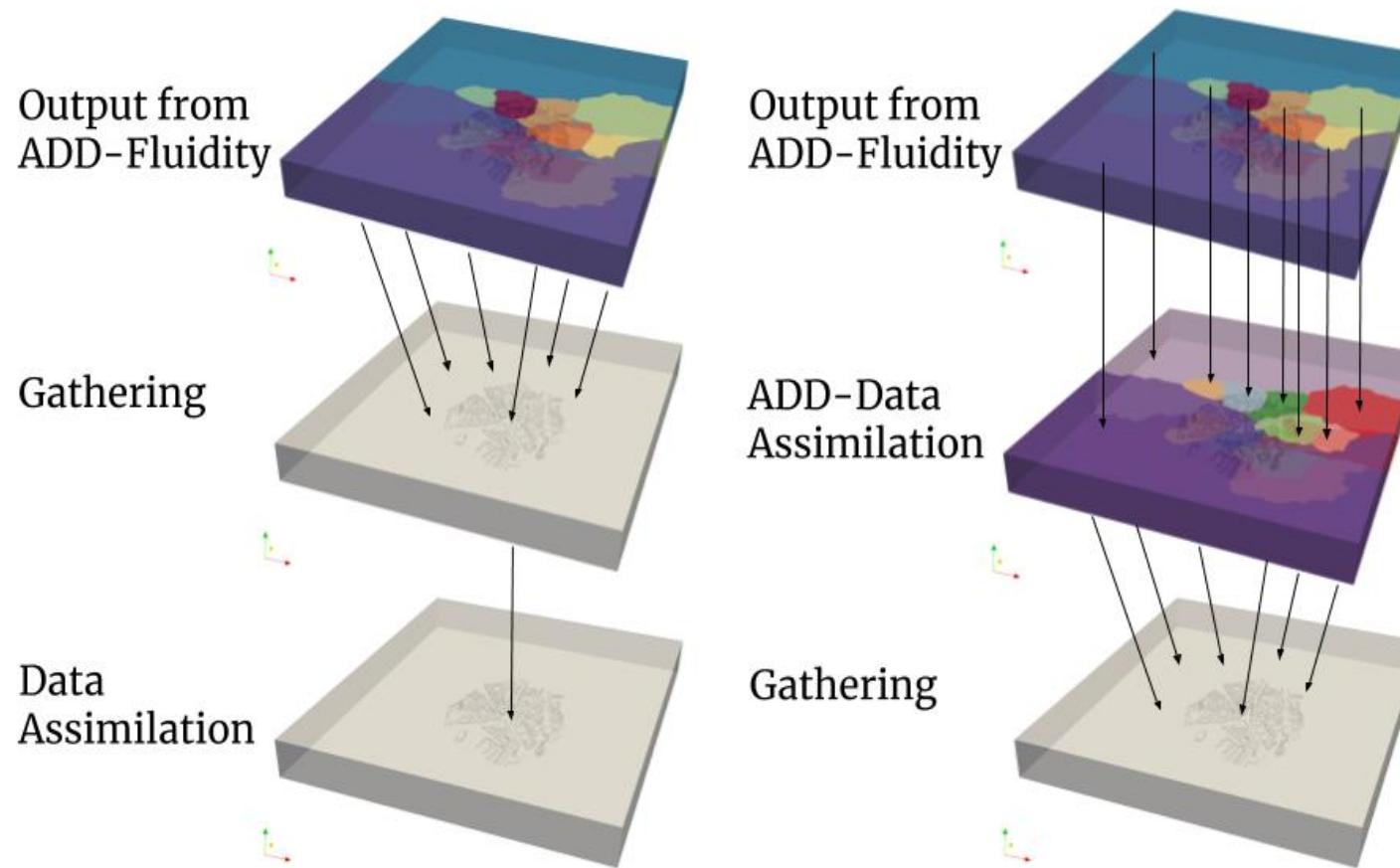


CFD developed by Dr Laetitia Mottet



Domain Decomposition for Data Assimilation

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



Domain Decomposition for Data Assimilation

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



Predicted by the forecasting software



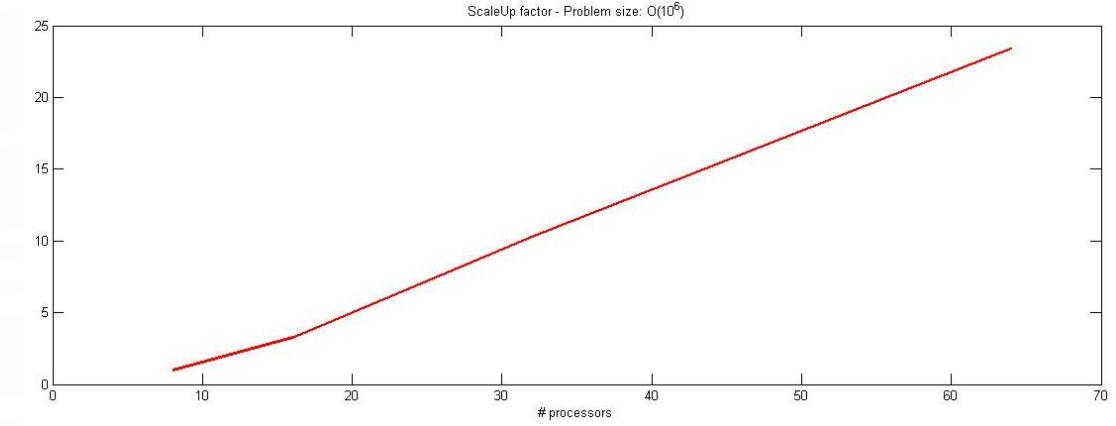
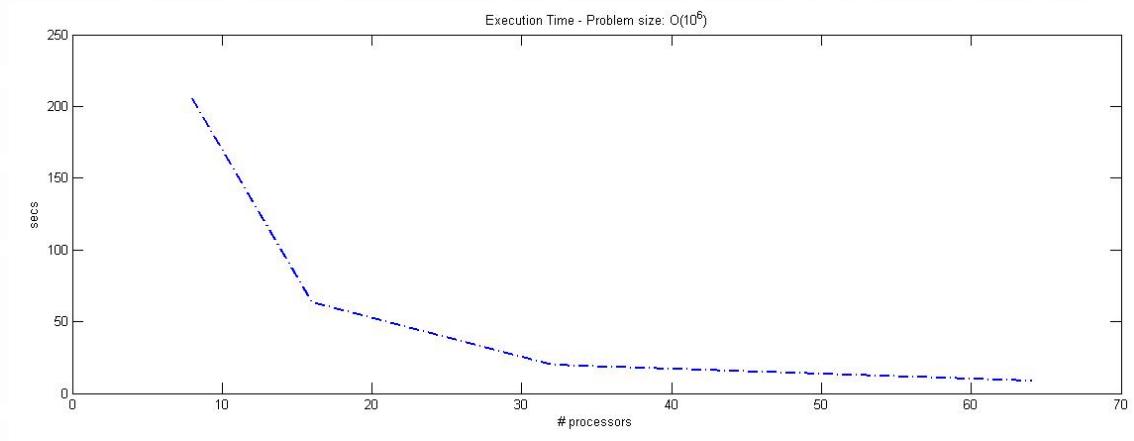
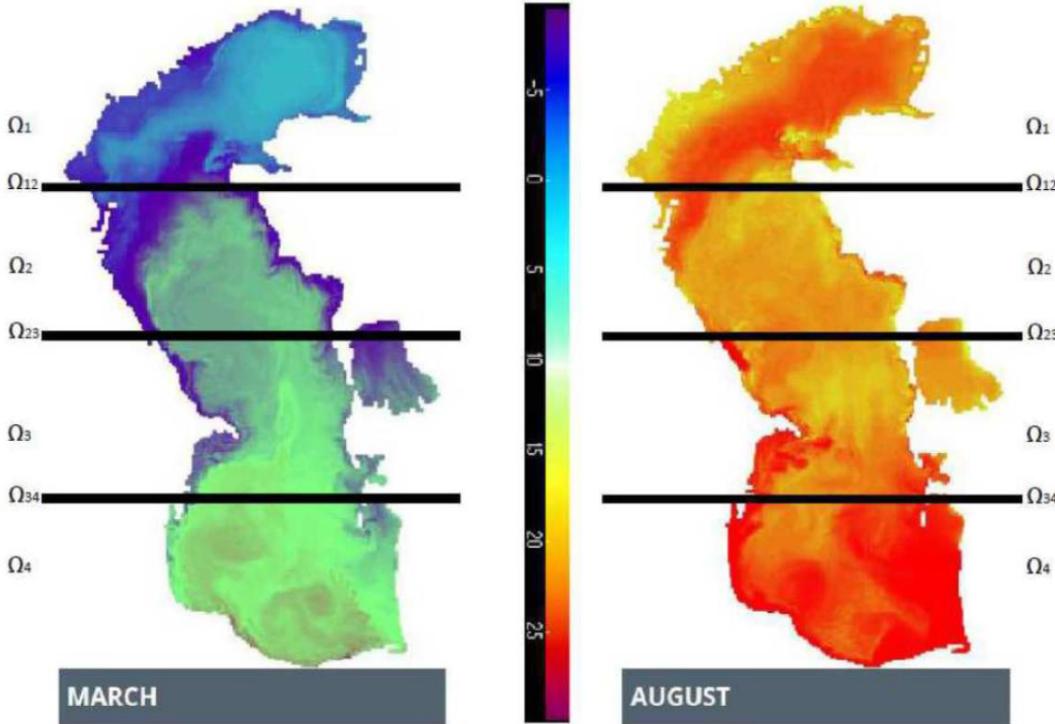
Observed by the sensor



Assimilated in the numerical forecasting

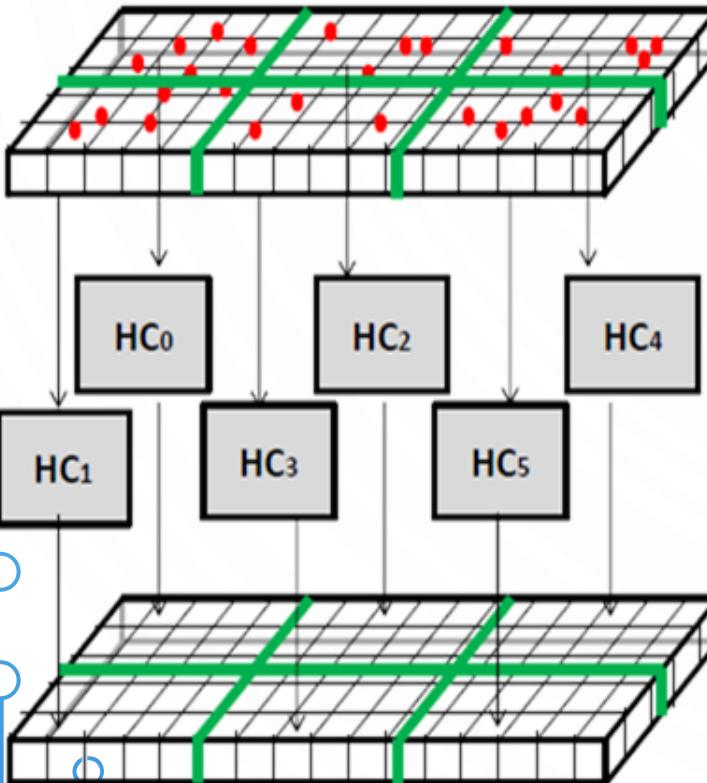
Iso-surface, in white, of the pollutant concentration computed in parallel with 10 processors

Domain Decomposition for Data Assimilation



Domain Decomposition for Data Assimilation

... big domains



total of 10,649,600 CPU cores across the entire system





Data Assimilation:

Examples and code (<https://github.com/DL-WG/DA-tutorial-/>)

... we are happy to share

Weekly meetings with invited speakers from other universities or companies:

We meet every Tuesday at 4pm (UK time) on Zoom

Our mailing list:

<https://mailman.ic.ac.uk/mailman/listinfo/datalearning>



All the talks are recorded and uploaded
on our YouTube Channel – Data Learning



International Conference:

Every year, the DataLearning group organises a workshop on **Machine Learning and Data Assimilation for Dynamical Systems (MLDADS)**, as part of the International Conference on Computational Science (ICCS).



Prague – ICCS 2023 London - ICCS 2022 Poland - ICCS 2021 Amsterdam - ICCS 2020 Faro, Portugal - ICCS 2019

Sharing contents with our community worldwide:

To get access to our **codes**: Our GitHub <https://github.com/DL-WG>



There is nothing measured that doesn't exist.

Thank you!

Some other papers and applications:

<https://sites.google.com/view/rossella-arcucci/datalearning>