IMPERIAL

AI for Humanity

Dr Rossella Arcucci

Department of Earth Science & Engineering,
Director of the Ada Lovelace Academy,
Director of Research at the Data Science Institute,
DAOS - World Meteorological Organization
ESA Φ-lab - Visiting professor

r.arcucci@imperial.ac.uk
https://www.imperial.ac.uk/people/r.arcucci

data learning

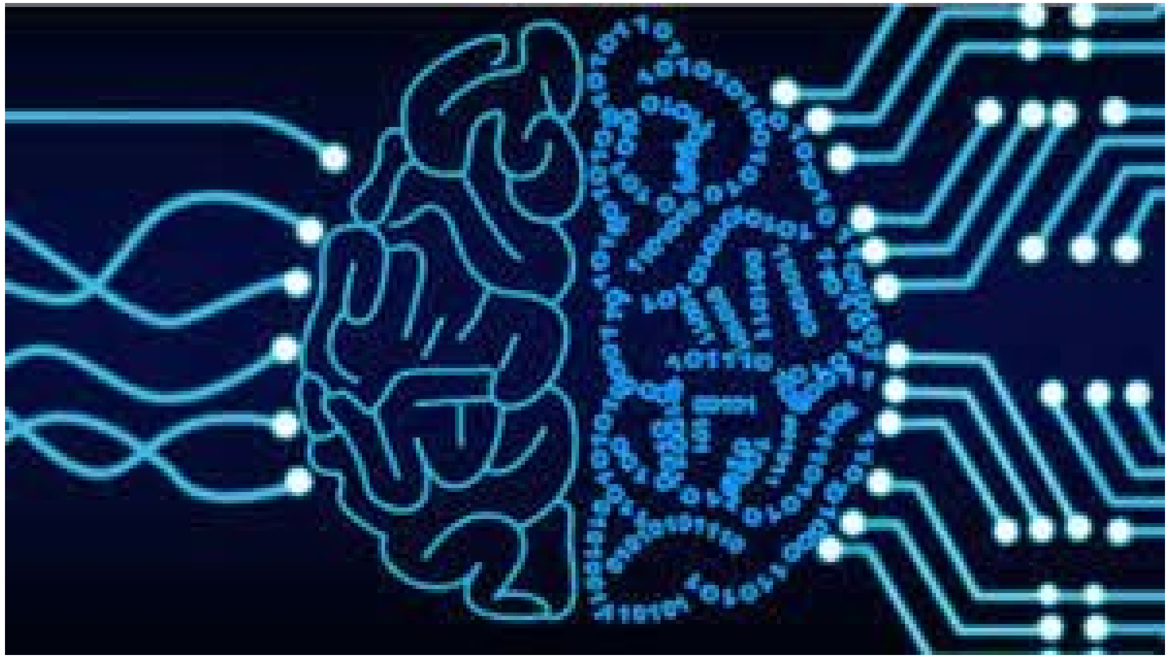
DAOS
WORLD METEOROLOGICAL ORGANIZATION

DSI

esa
Φ-lab

What is AI?

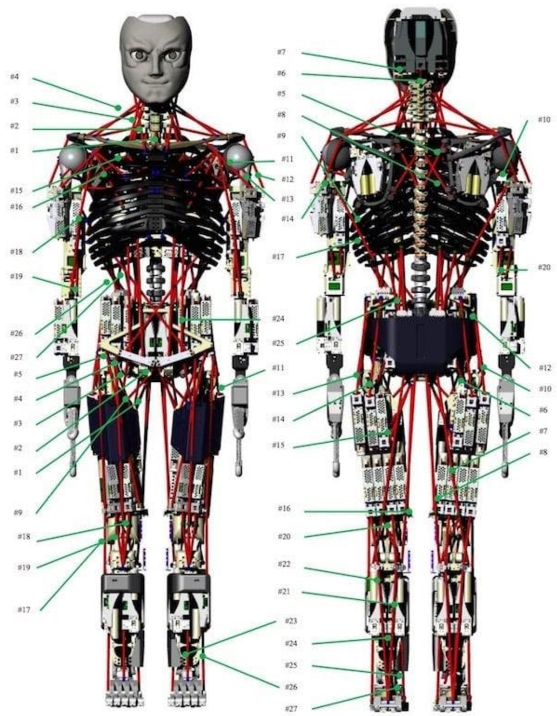
Artificial Intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.



Quiz …

Can an AI-based robot sweat?

Kengoro is a humanoid robot that cools itself by “sweating,” using water evaporation inside its body to prevent overheating during movement.



More human-like movement → Kengoro has artificial muscles that generate heat
Efficient cooling → avoids bulky fans or external systems
Biomimicry → copying how the human body solves problems

Fun fact
Kengoro can even do exercises like **push-ups and sit-ups**, and the more it “works out,” the more it needs its sweating system… just like a real athlete.

Quiz …

Can AI-based robots feel emotions?

Ameca can simulate emotions like happiness or anger through AI and facial expressions, but it does not truly feel emotions… its responses are programmed to mimic human behavior.

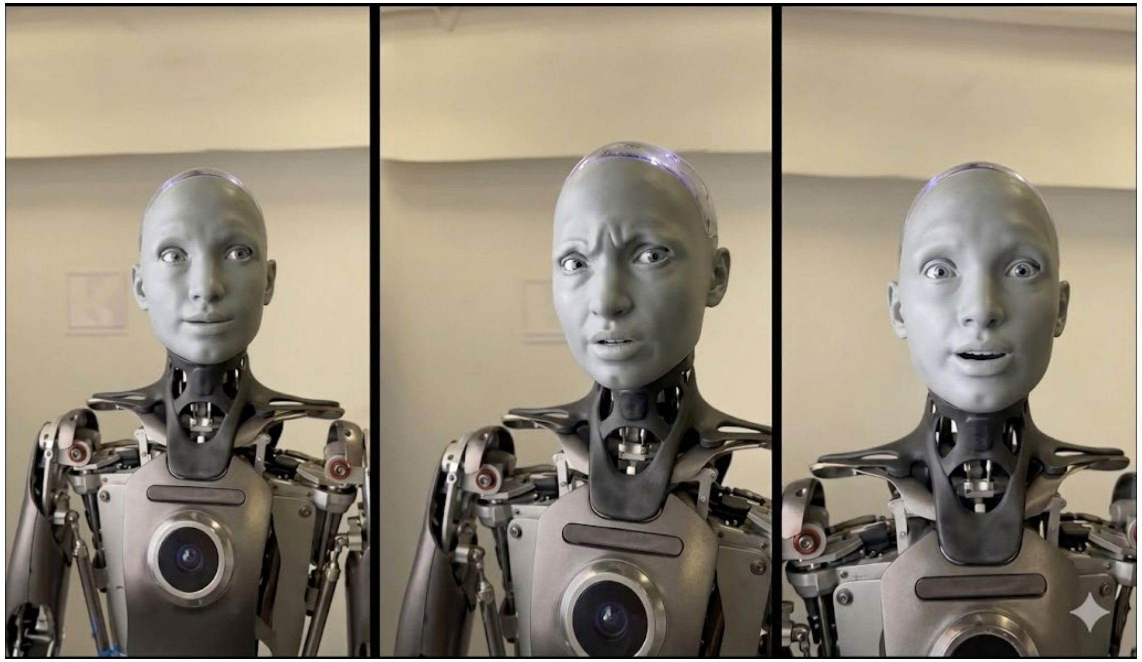
Ameca was created by the UK robotics company Engineered Arts

During the *60 Minutes Australia* interview, Ameca is asked directly if it has feelings. Its answer is very clear:

- It says it has **“a range of feelings and emotions that are programmed into me”**
- It can display emotions like:
    - joy
    - surprise
    - confusion
    - anger
    - sadness

It also explains:

- It feels “happy” when learning new things or having conversations
- It gets “angry” when humans are rude or disrespectful



Quiz …

Can AI-based robots have citizenships?

Sophia is a humanoid robot that became the first robot to receive symbolic citizenship in Saudi Arabia in 2017.

- Designed to interact socially with humans
- Uses AI to hold conversations and answer questions
- Has realistic facial expressions and can recognize faces

Sophia’s citizenship was largely **symbolic** and sparked debate worldwide about:

- robot rights
- AI ethics
- what citizenship really means

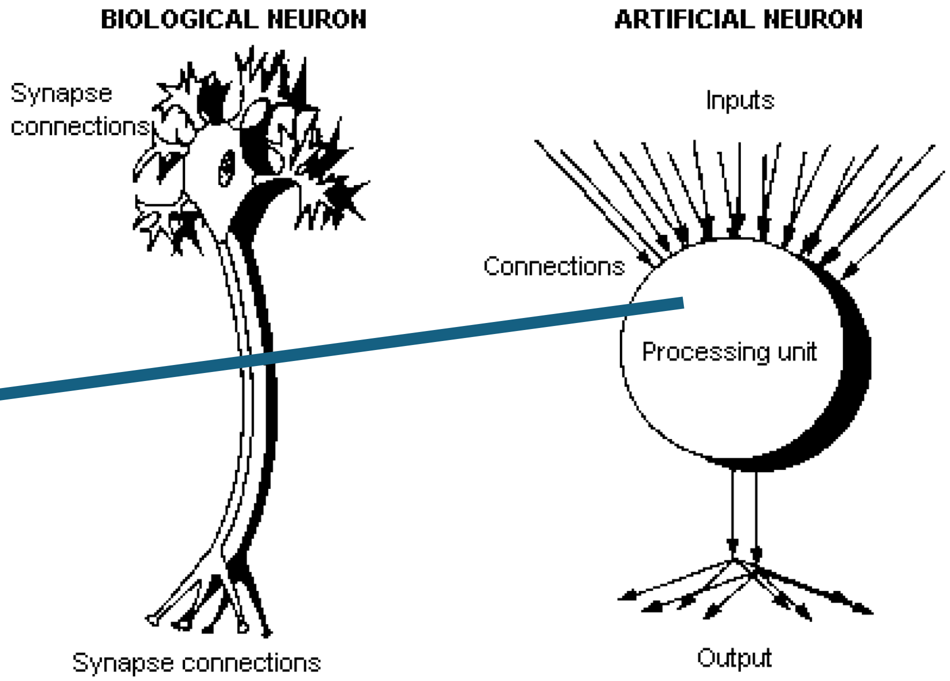
Quiz …

Can AI-based robots self learn new things/emotions?

As of early 2026, several large language models (LLMs) stand out for their sophistication and capabilities:

- **GPT-5 by OpenAI**
  Released in late 2025, GPT-5 builds on GPT-4o with stronger reasoning, improved multimodal abilities (text, image, video, and audio), and more reliable real-time interactions.
- **Claude 4 by Anthropic**
  Claude 4 models feature enhanced long-context understanding (200K+ tokens), strong coding abilities, and improved safety-focused responses.
- **Grok-2 by xAI**
  An upgraded version of Grok-1, Grok-2 integrates deeply with X (formerly Twitter), offering real-time data access and improved reasoning with a mixture-of-experts design.
- **DeepSeek-R2 by DeepSeek**
  An evolution of DeepSeek-R1, this model is known for top-tier performance in math, coding, and logical reasoning, and remains a leading open-source contender.
- **Gemini 2.0 by Google DeepMind**
  Gemini 2.0 advances multimodal AI with extremely large context windows (up to millions of tokens) and strong performance across text, video, and complex reasoning tasks.

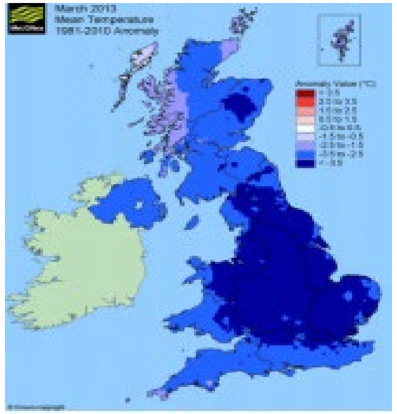
The Brain Behind Modern AI: Artificial Neural Networks

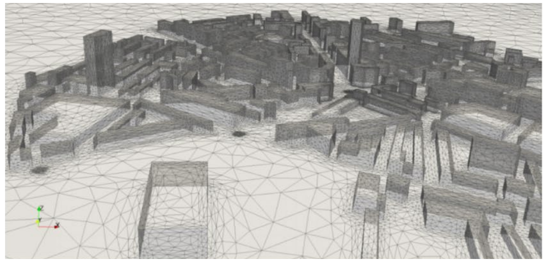


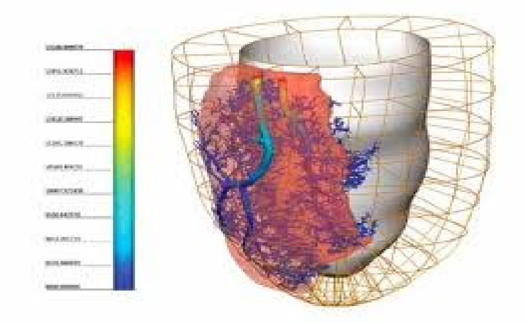
Learning | Training

Data Science Empowering AI: the era of the data!

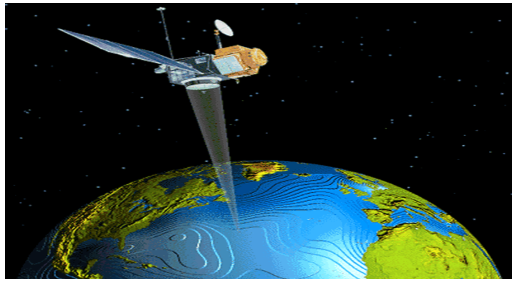
High resolution Models…



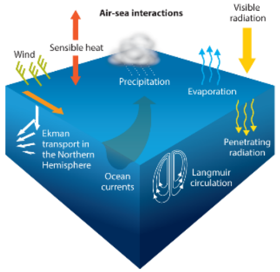


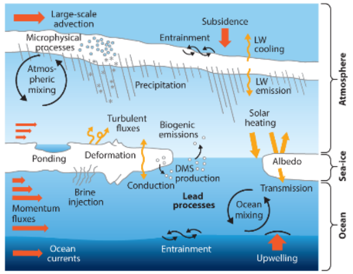


Real observations…



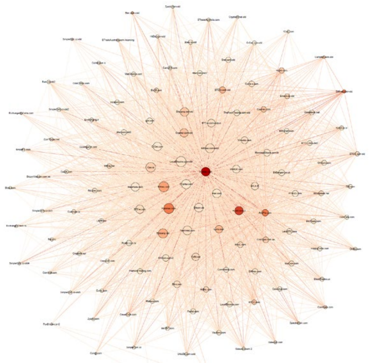
reddit

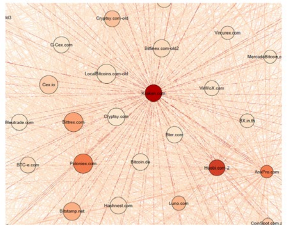


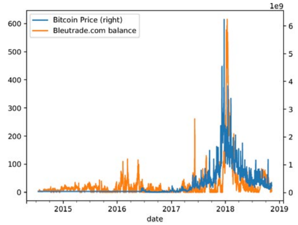


TikTok

facebook



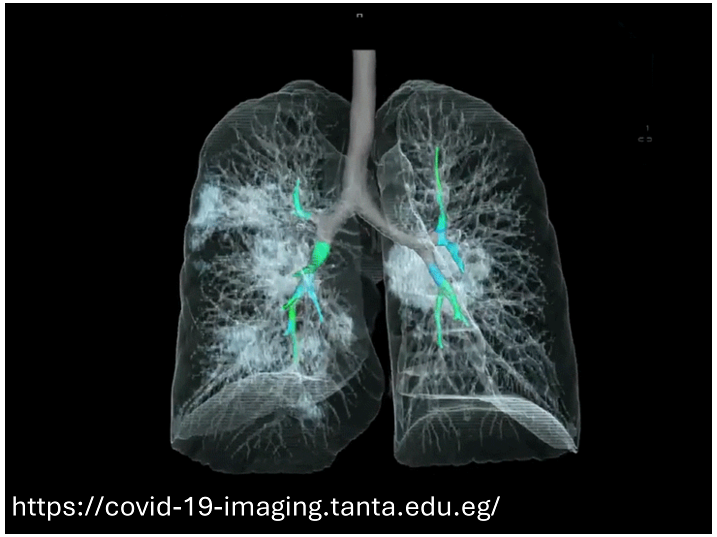




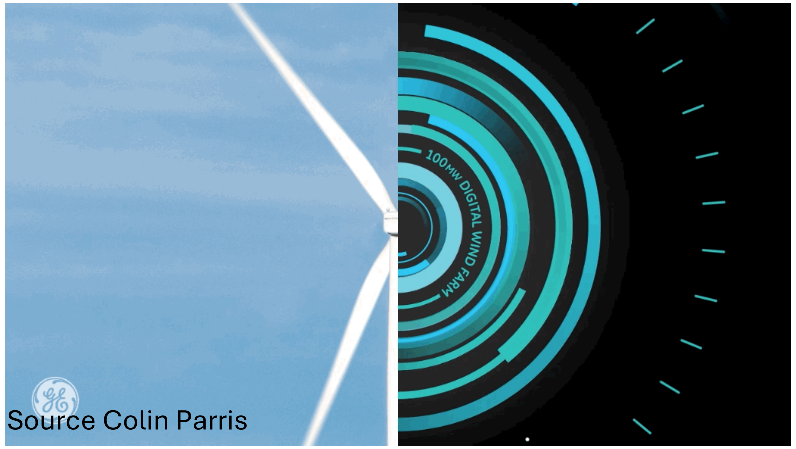
Digital Twins

Source Microsoft Blog Europe's open data revolution: the road to collaboration

Source ESA



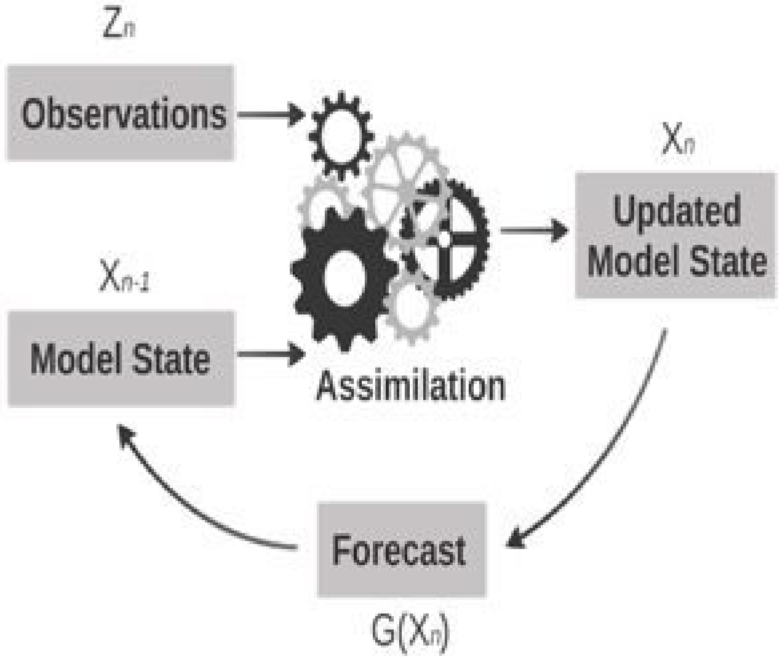
https://covid-19-imaging.tanta.edu.eg/



Source Colin Parris

Adjusting

Data Assimilation



Data Learning: a modular approach

Journal of Computational Science 58 (2022) 101525

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Data Learning: Integrating Data Assimilation and Machine Learning

Caterina Buizza [b], César Quilodrán Casas [a], Philip Nadler [a], Julian Mack [a], Stefano Marı
Zainab Titus [c], Clémence Le Cornec [d], Evelyn Heylen [e], Tolga Dur [a], Luis Baca Ruiz [a,g],
Claire Heaney [c], Julio Amador Díaz Lopez [a,h], K.S. Sesh Kumar [a], Rossella Arcucci [a,c,*]

[a] Data Science Institute, Imperi
[b] Personal Robotics Lab, Depar
[c] Department of Earth Science
[d] Department of Civil and Envi
[e] Control and Power Group, De
[f] DIETI, University of Naples F
[g] Department of Computer Scie
[h] Data Science Intitute, London

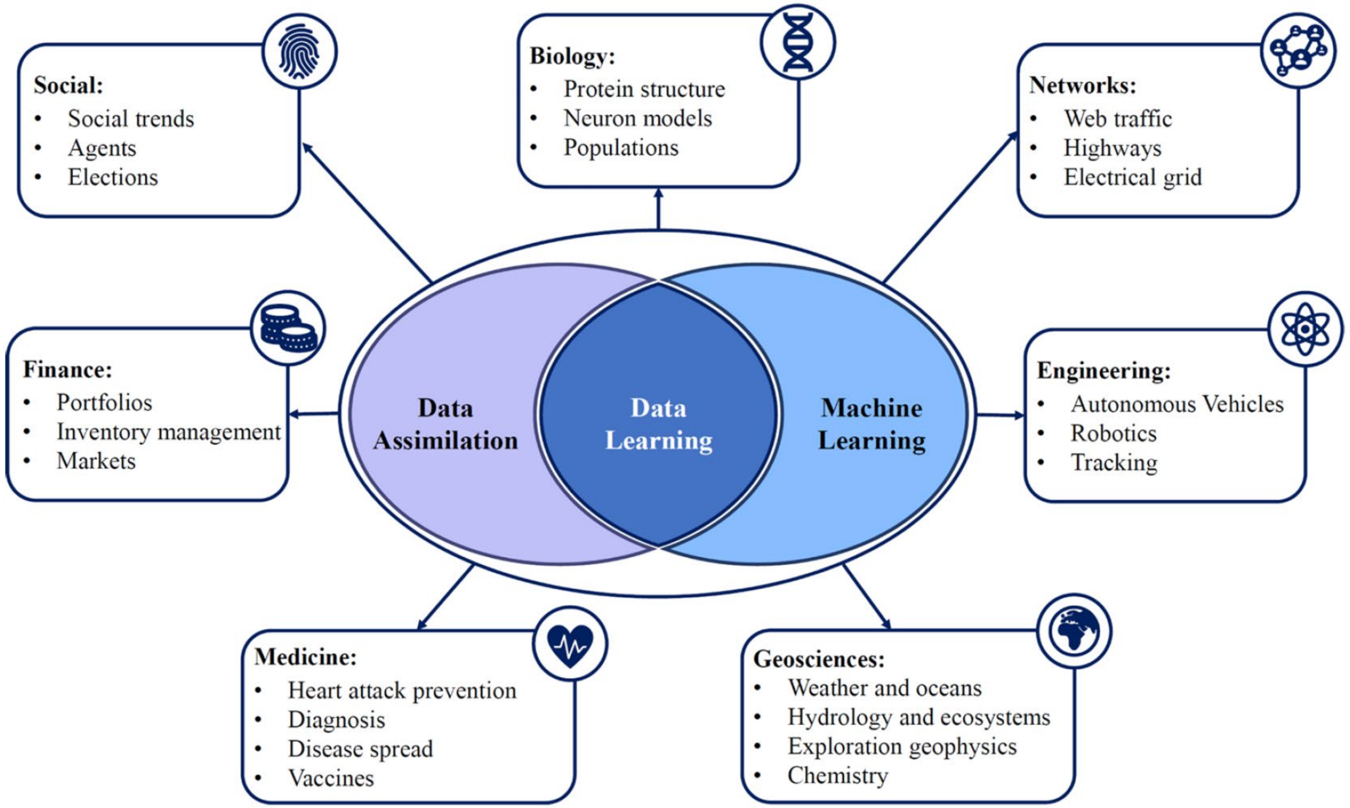
ELSEVIER

Journal of Computational Science

Volume 85, February 2025, 102523

Facing & mitigating common challenges when working with real-world data: The Data Learning Paradigm

Jake Lever [a b c], Sibo Cheng [b c], César Quilodrán Casas [a b], Che Liu [a b],
Hongwei Fan [b d], Robert Platt [a], Andrianirina Rakotoharisoa [a b], Eleda Johnson [a],
Siyi Li [a], Zhendan Shang [b], Rossella Arcucci [a b c]



AI4Good - Natural Hazards

Wildfires: except for the risk map, there isn't a reliable physics-based model to predict them, and human factors play a major role.

NOWCASTING

Floods: they can often be predicted, but once they occur, it's usually too late to prevent the damage.

PREDICTION

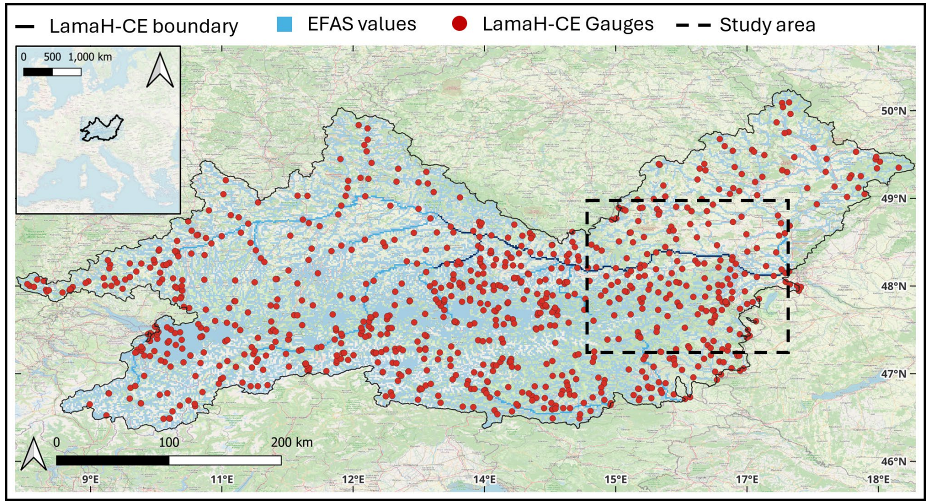
LEVERHULME
Centre for **Wildfires**, **Environment** and **Society**

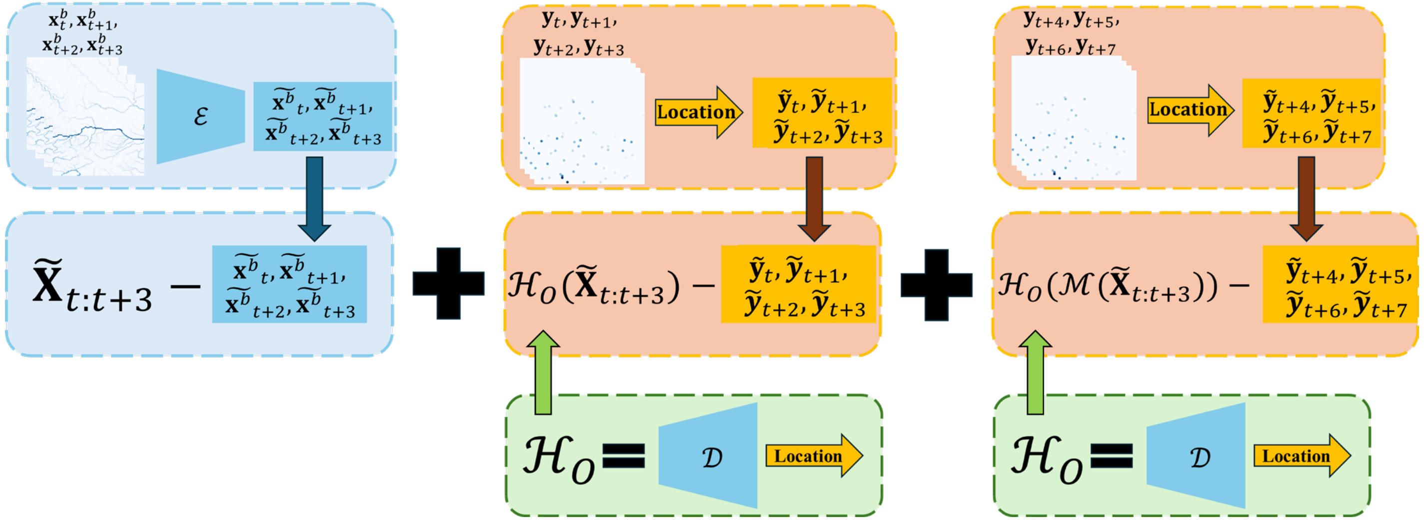
esa
Φ-lab

WORLD METEOROLOGICAL ORGANIZATION

Met Office

AI-empowered latent four-dimensional variational data assimilation for river discharge forecasting



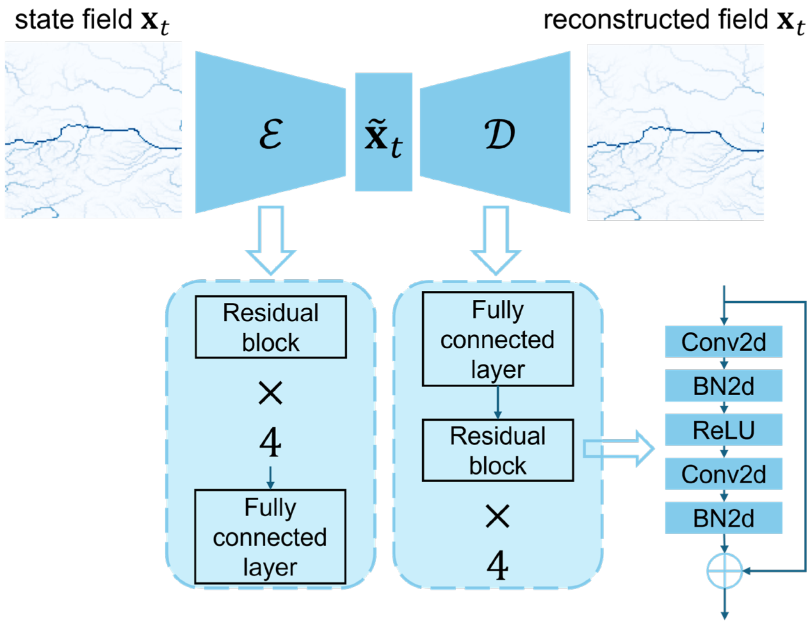


Name	EFAS	LamaH-CE
Type	state field	observation
Variable	river discharge	
Time coverage	1992-2017	
Temporal resolution	6-hour	
Spatial resolution	$1 arcmin \times 1 arcmin$	\
Number of gauge	\	58
Dimension of data	128 × 128	58 × 1

$$
\begin{aligned}
\mathcal{J}(\tilde{\mathbf{x}}_{t:t+3}) &= \left\|\tilde{\mathbf{x}}_{t:t+3} - \tilde{\mathbf{x}^b}_{t:t+3}\right\|^2_{\tilde{\mathbf{B}}^{-1}} \\
&+ \|\tilde{\mathbf{y}}_{t:t+3} - \mathcal{H}_o(\tilde{\mathbf{x}}_{t:t+3})\|^2_{\mathbf{R}^{-1}} \\
&+ \|\tilde{\mathbf{y}}_{t+4:t+7} - \mathcal{H}_o(\mathcal{M}(\tilde{\mathbf{x}}_{t:t+3}))\|^2_{\mathbf{R}^{-1}}
\end{aligned}
$$

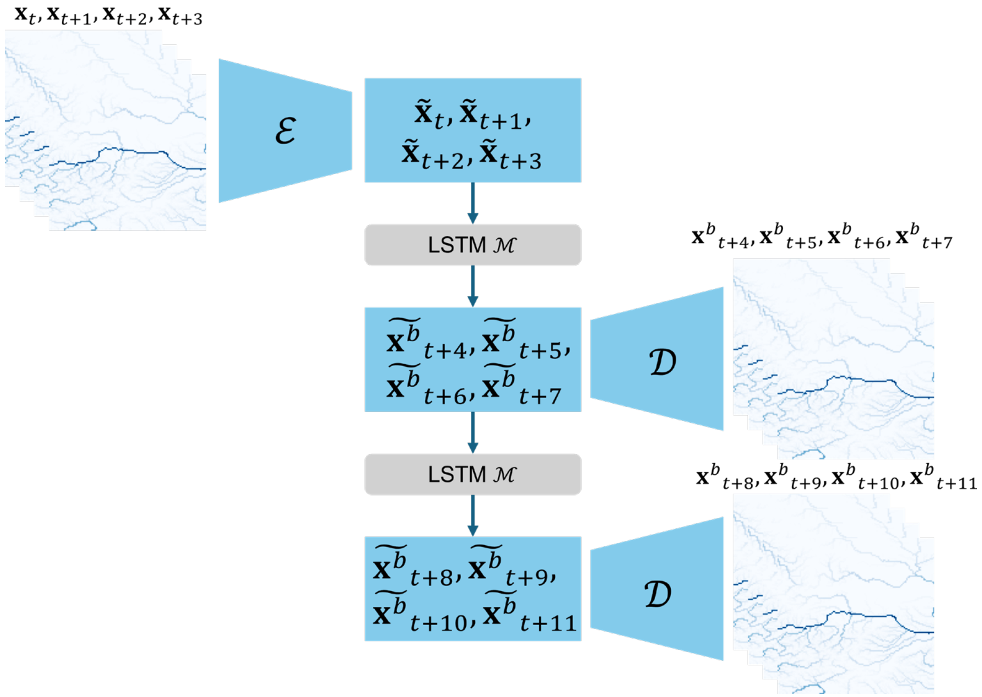
AI-empowered latent four-dimensional variational data assimilation for river discharge forecasting

Compression model



$$\tilde{\mathbf{x}}_t = \mathcal{E}(\mathbf{x}_t)$$
$$\mathbf{x}_t = \mathcal{D}(\tilde{\mathbf{x}}_t)$$

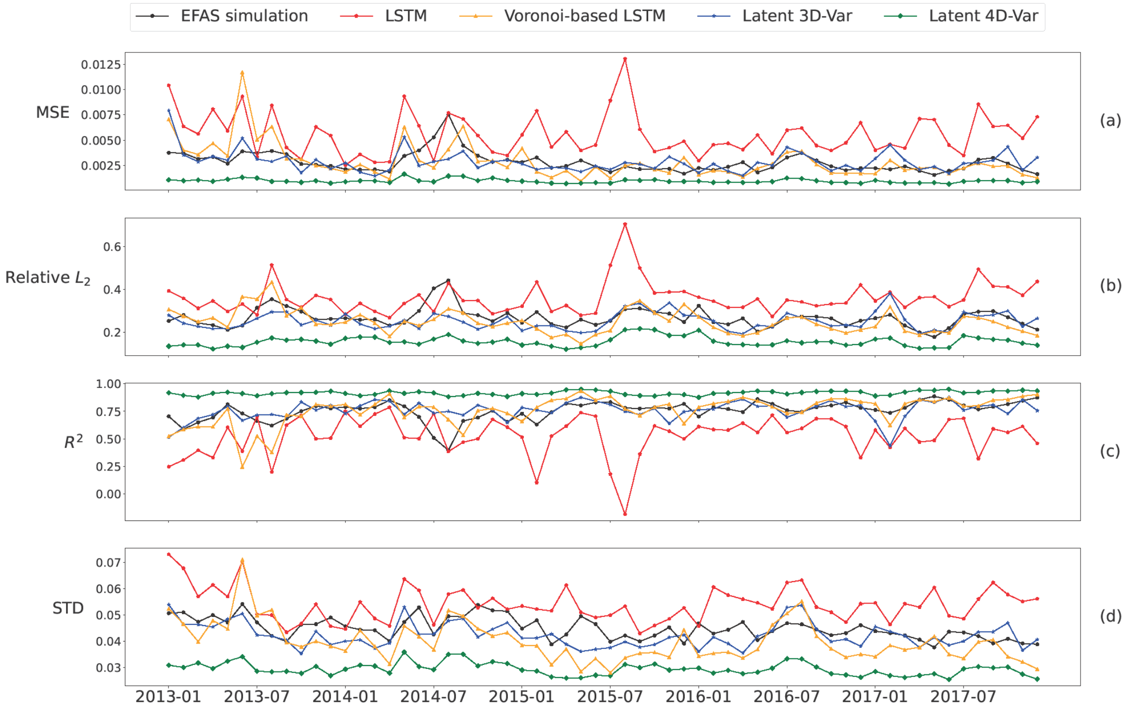
Predictive model



$$\tilde{\mathbf{x}^b}_{t+4:t+7} = \mathcal{M}(\tilde{\mathbf{x}}_{t:t+3})$$

AI-empowered latent four-dimensional variational data assimilation for river discharge forecasting

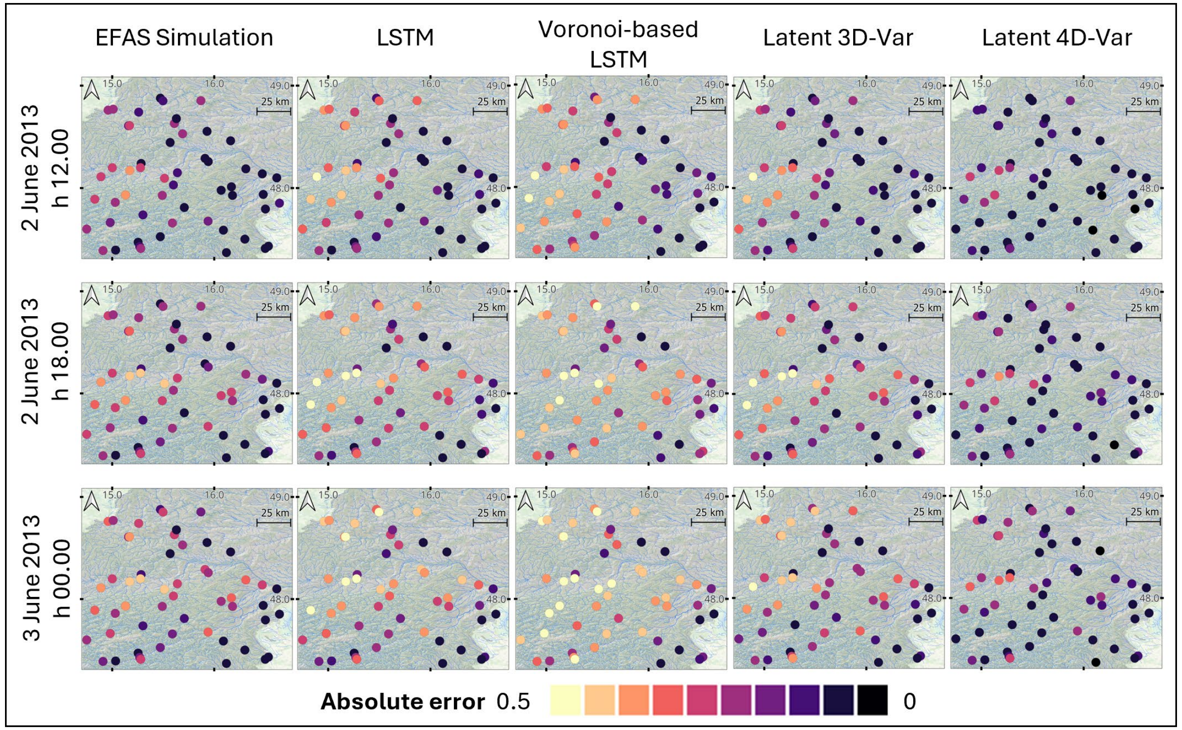
Evaluation metrics



Numerical result

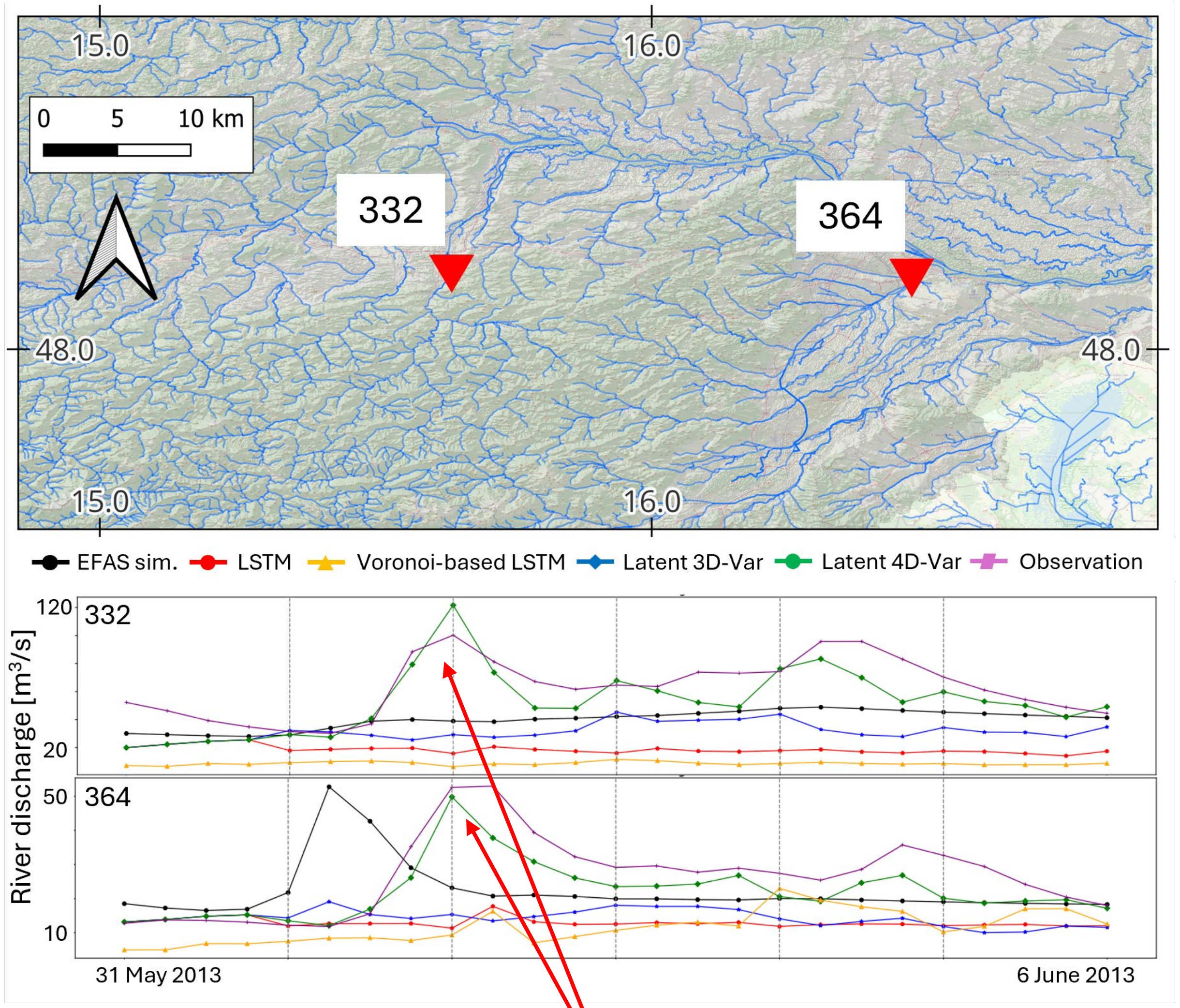
	EFAS simulation	LSTM	Voronoi-based LSTM	Latent 3D-Var	Latent 4D-Var
MSE ↓	0.00279	0.00556	0.00289	0.00285	0.00096
Relative L_2↓	0.26585	0.35939	0.24780	0.25120	0.15495
R2 ↑	0.75563	0.52654	0.76050	0.75936	0.91683
STD ↓	0.04479	0.05397	0.03981	0.04228	0.02929
KGE ↑	0.81432	0.67678	0.79239	0.81370	0.93816

Absolute error



AI-empowered latent four-dimensional variational data assimilation for river discharge forecasting

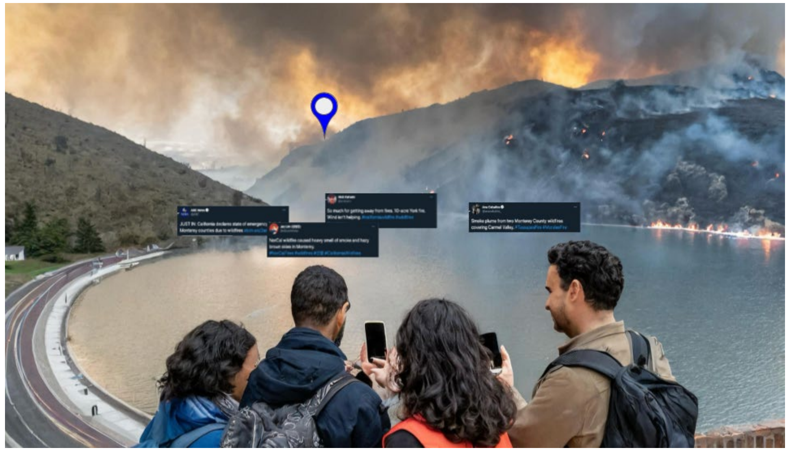
Flood event (June 2013)



Predicted!

FLOOD-MLDA: A Multi-Source Dataset and Benchmarks for Flood Forecasting with Machine Learning and Data Assimilation

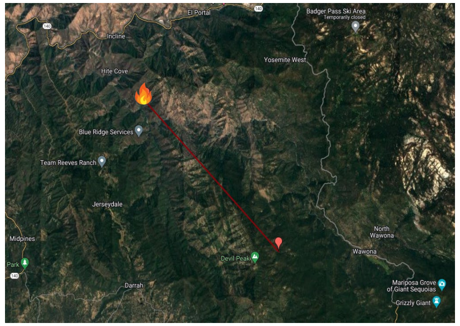
Human sensors



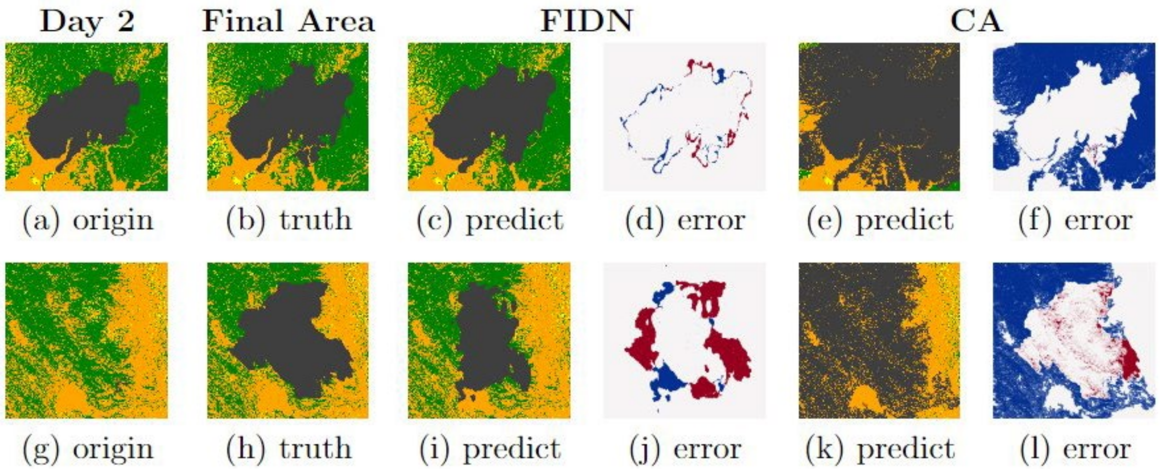
Case Study -
Chimney Wildfire

(-119.72, 37.53)

COORDINATES



Predict spatio-temporal wildfire propagation



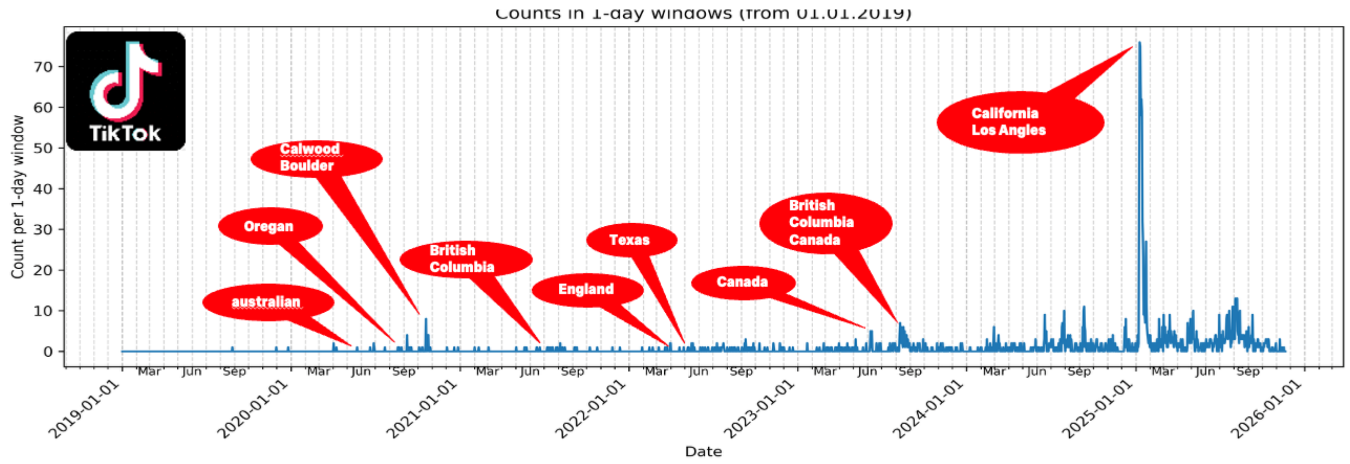
Social media platforms generate massive amounts of data every day. For example, **TikTok sees around 1–2 billion video views daily, X produces about 500 million posts per day**, and **Facebook handles roughly 4–5 billion interactions daily**, together amounting to **hundreds of terabytes to multiple petabytes of data generated each day**.

Human sensors

Social media posts → Real-time monitoring → Approximate Location Extraction → Validation (satellite ...) → Ignition geolocation, fire features → Wildfire nowcasting model

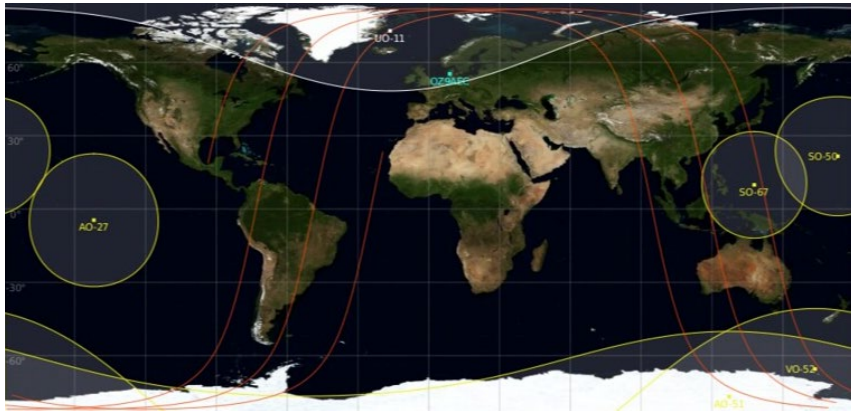
SocialFetch

```
socialfetch tiktok --categories "wildfire" "bushfire"
            --categories "california" "los angeles"
            --limit 50
            --hashtag-names "fireseason" "wildfire2025"
            --unwanted-words "politics" "election" --authors "calfire" "usfs"
            --country-name "united states"
            --time-bounds "2025-09-01T00:00:00Z" "2025-09-30T23:59:59Z"
            --is-random --download-media
```

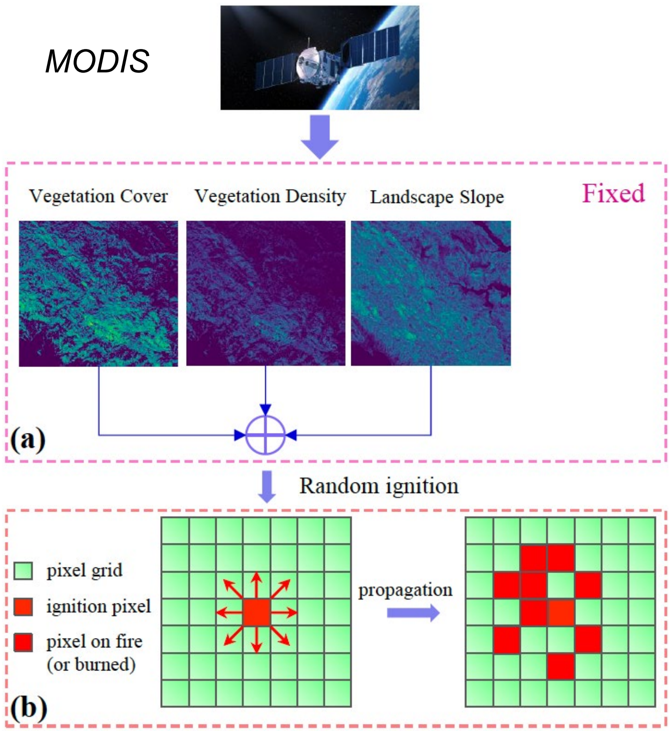


SatCov

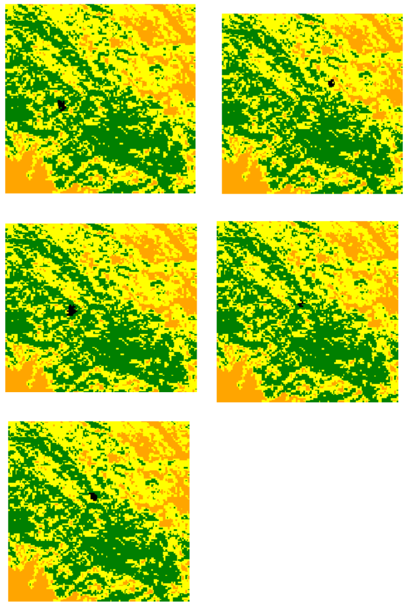
Realtime tracing of useful satellite for fire detection



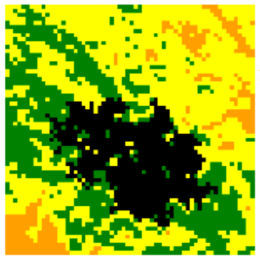
Predicting spatio-temporal wildfire propagation: dataset creation



Data collection, including canopy density, canopy cover, landscape slope and local wind speed

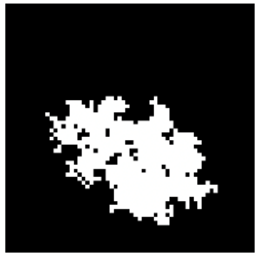


Burned area dataset



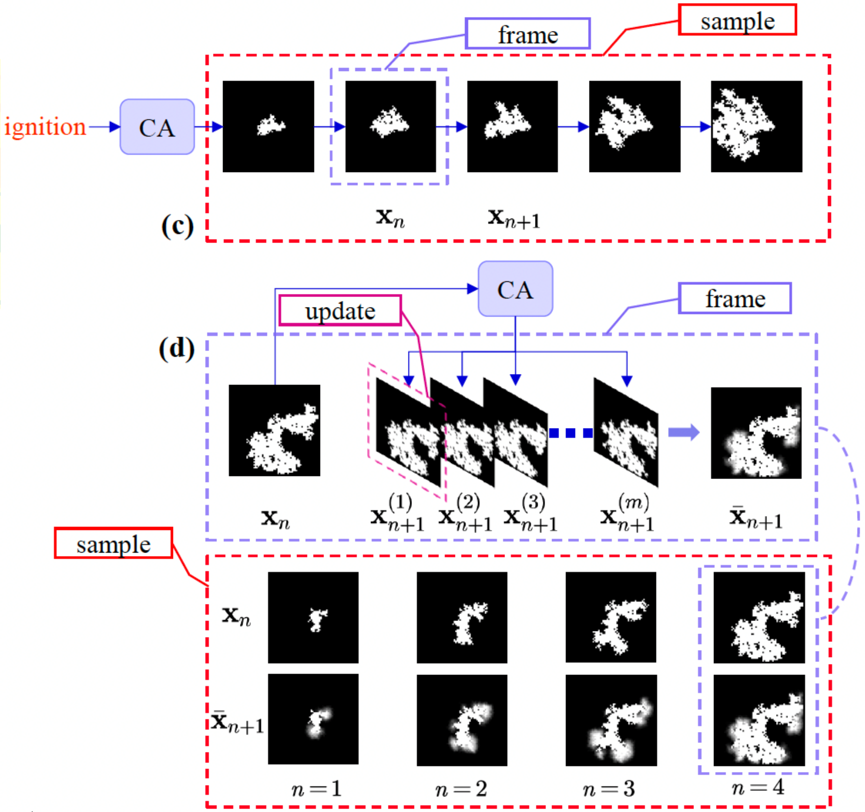
(a)

simulated with CA



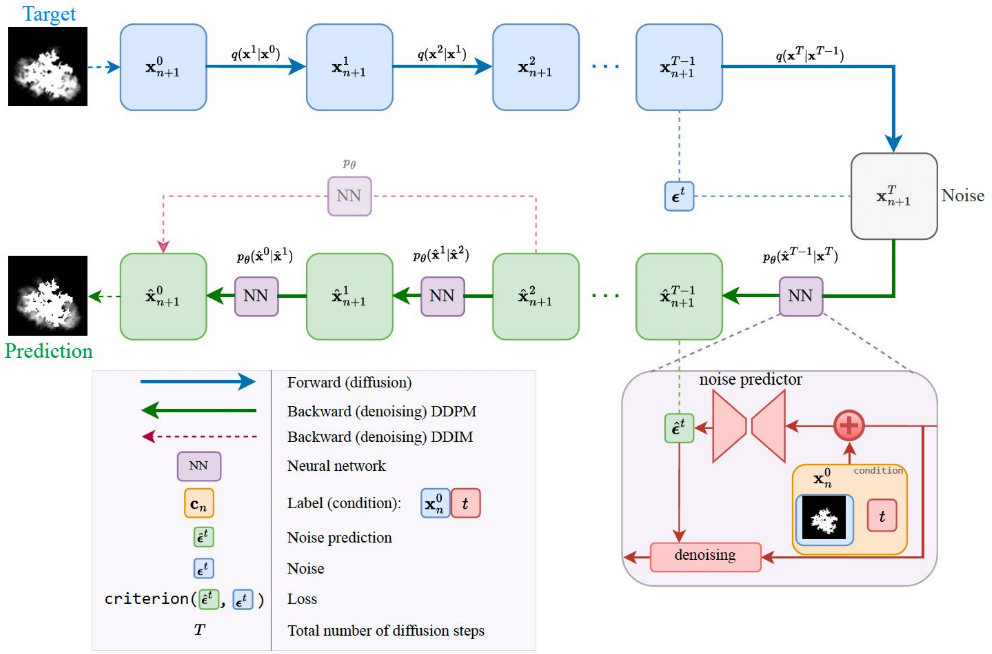
(b)

Grayscaled snapshot

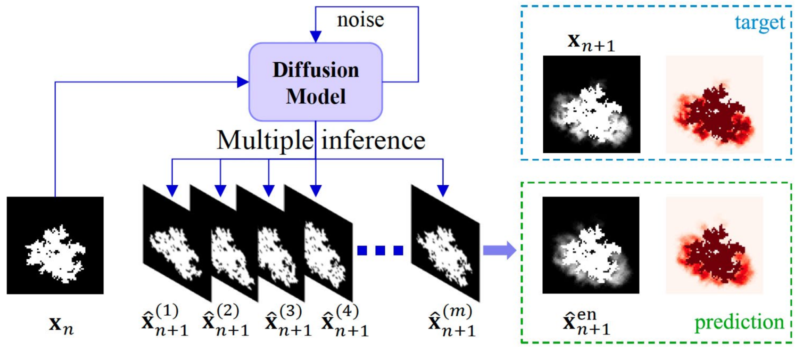


Predicting spatio-temporal wildfire propagation

Diffusion model training process



Model evaluation.

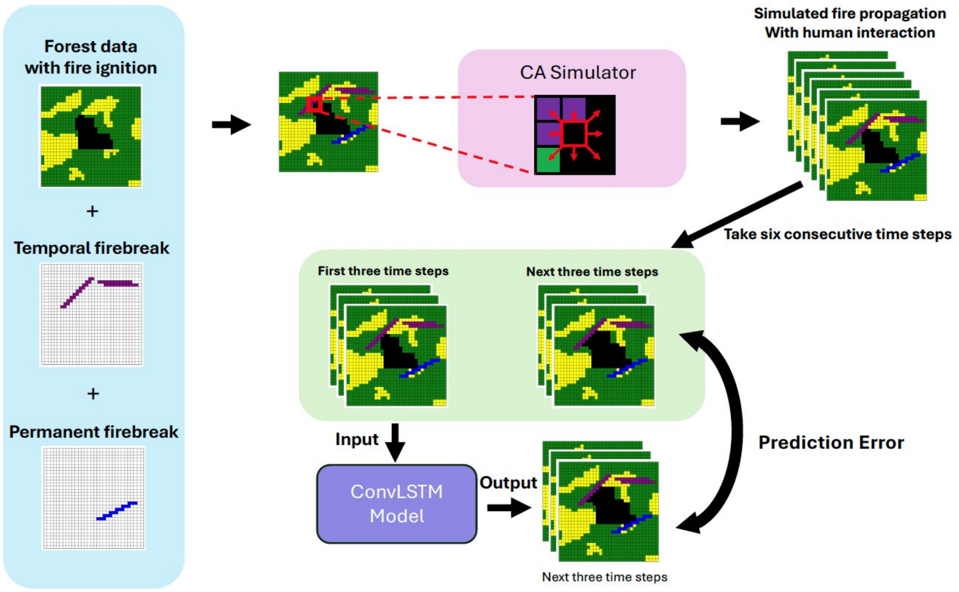


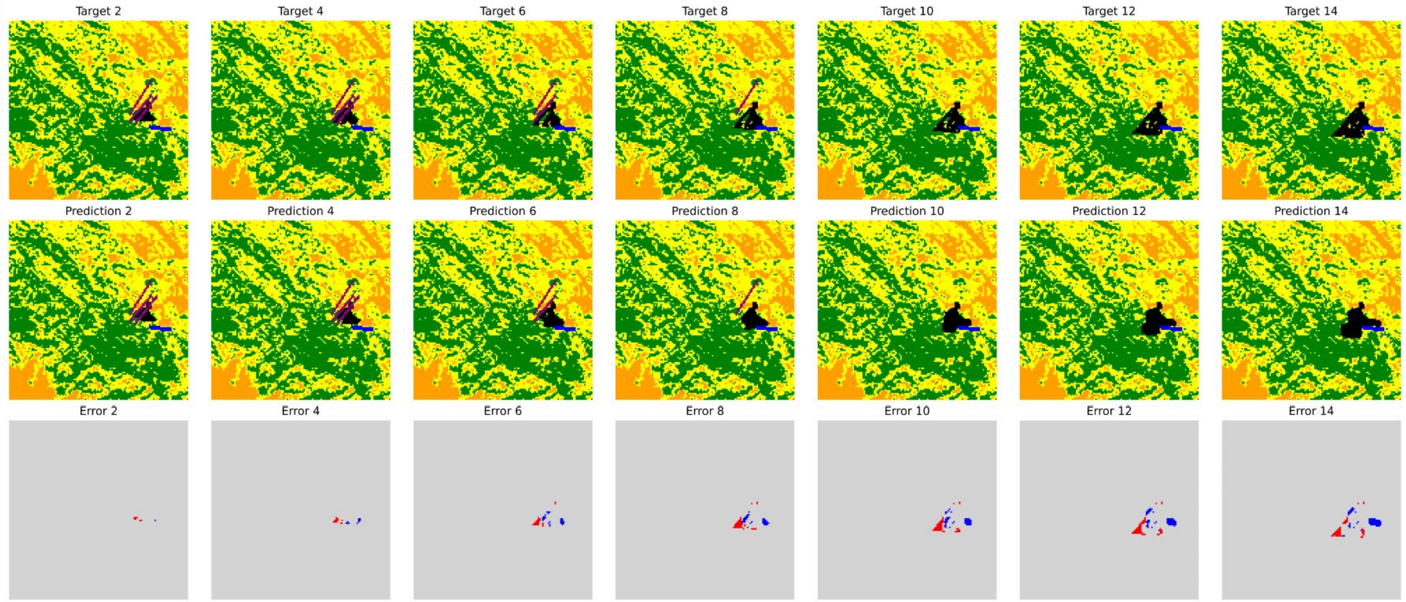
Performance on Chimney fire dataset

Metric \ Size	Deterministic					Diffusion				
	50	100	200	500	900	50	100	200	500	900
MSE ↓	0.0199	0.0191	0.0160	0.0149	0.0143	0.0105	0.0077	0.0065	0.0056	0.0053
PSNR ↑	17.149	17.260	18.171	18.460	18.543	17.820	21.419	22.080	22.873	23.127
SSIM ↑	0.8348	0.8383	0.8444	0.8449	0.8467	0.5154	0.8365	0.8692	0.8923	0.8968
HR ($\epsilon = 0.2$) ↓	0.7186	0.7231	0.7327	0.7350	0.7376	0.7360	0.8313	0.8495	0.8692	0.8754
FID ↓	181.35	182.92	182.21	182.19	179.96	112.99	90.021	72.730	42.260	38.540
KL ↓	-	-	-	-	-	341.45	206.96	201.20	173.80	169.80

Predicting spatio-temporal wildfire propagation with dynamic firebreaks

Data generation and training pipeline



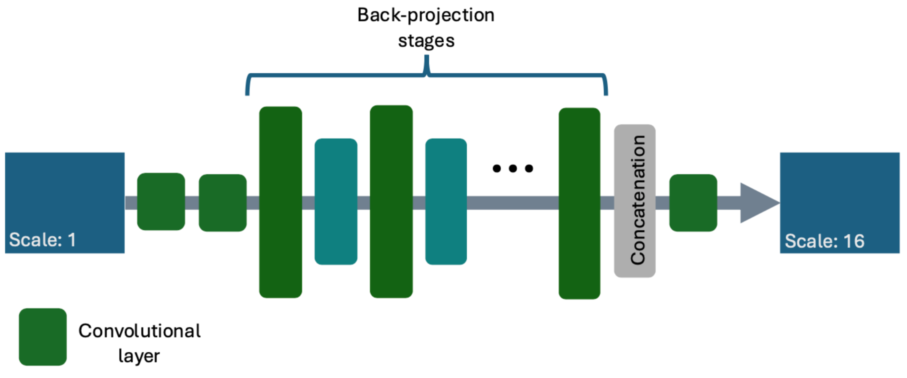


error map: red means false negatives and blue means false positive.

Data Science Empowering AI:

Increasing the resolution of satellite imagery

Our AI based - Super Resolution model

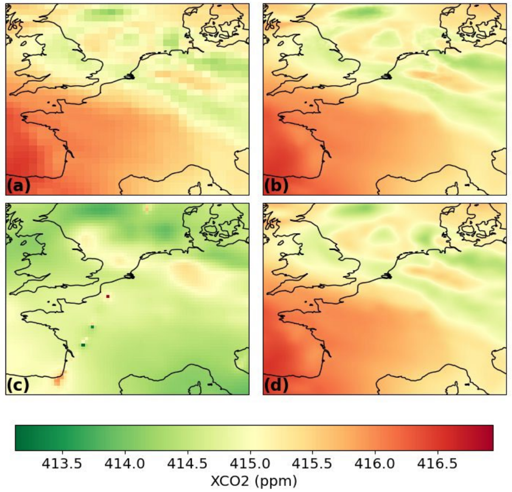


Test Case: Greenhouse Gasses (GHG)

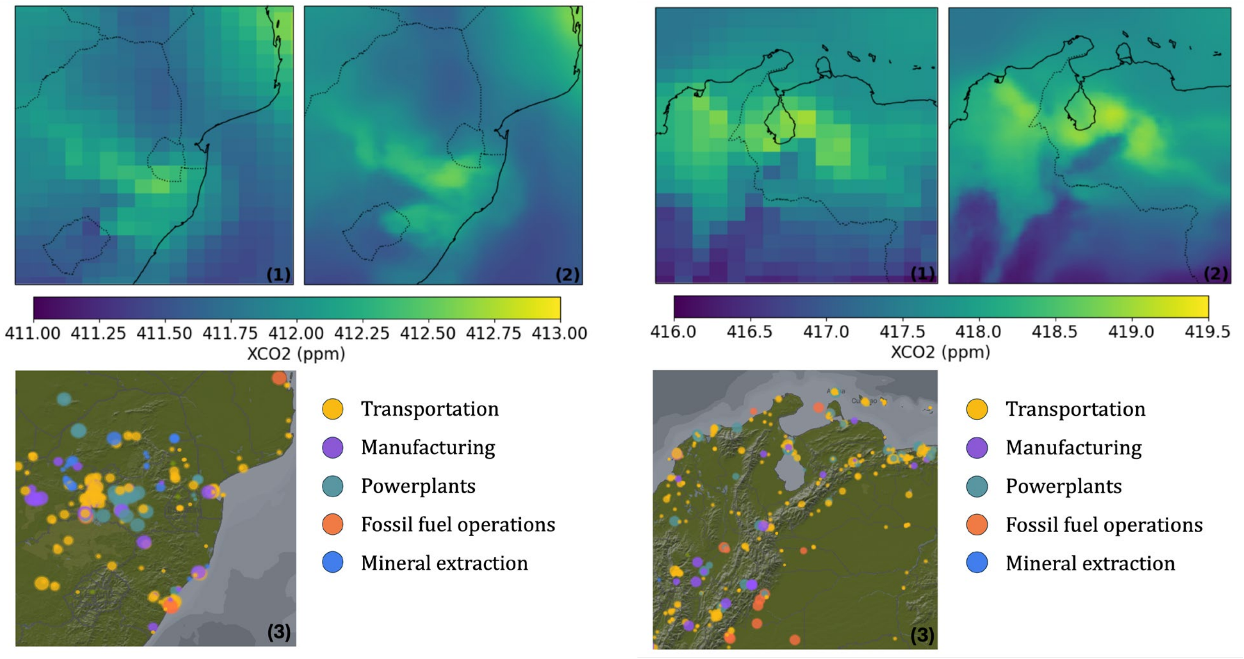
AI based - Super Resolution maps of XCO2 with a spatial resolution of 0.03*0.04, 16-times higher than the original datasets from the OCO-2 Level 3 (L3) data.
The model is not gas-specific and can therefore be adapted for monitoring other GHG beyond CO2.

High accuracy results.

We validate the model accuracy using ground sensors.



Example of application: **Detection of emission sources**



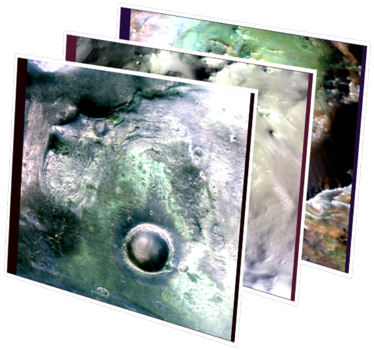
South Africa

Venezuela

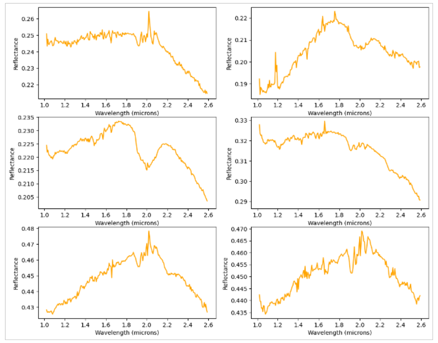
Data Science Empowering AI:
Denoising and Classifying Hyperspectral Data

Test Case: Hyperspectral camera on a satellite orbiting Mars

data set: Compact Reconnaissance Imaging Spectrometer for Mars (CRISM). Highly detailed (18m/pixel) imagery of most of the surface of Mars in Visible, Near Infrared and Shortwave Infrared.



77 Images

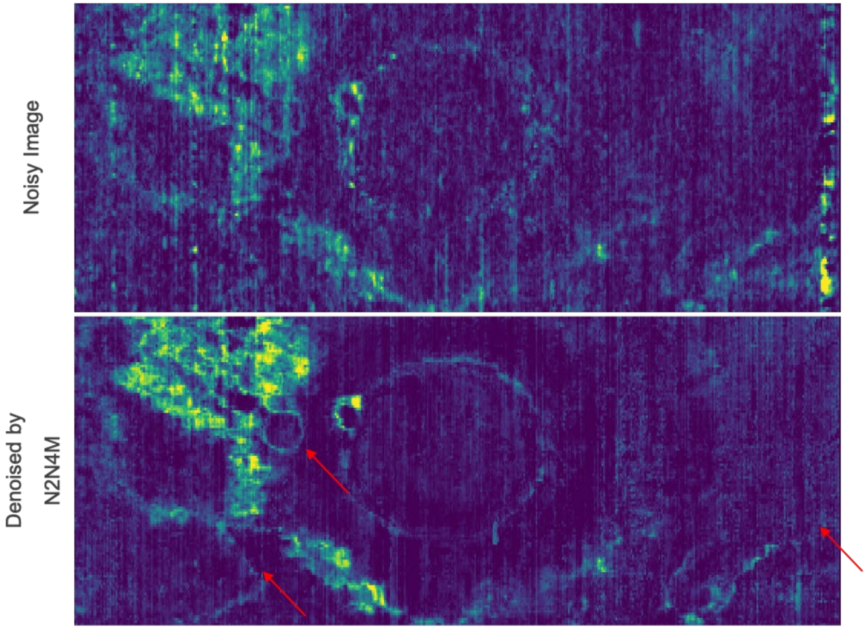


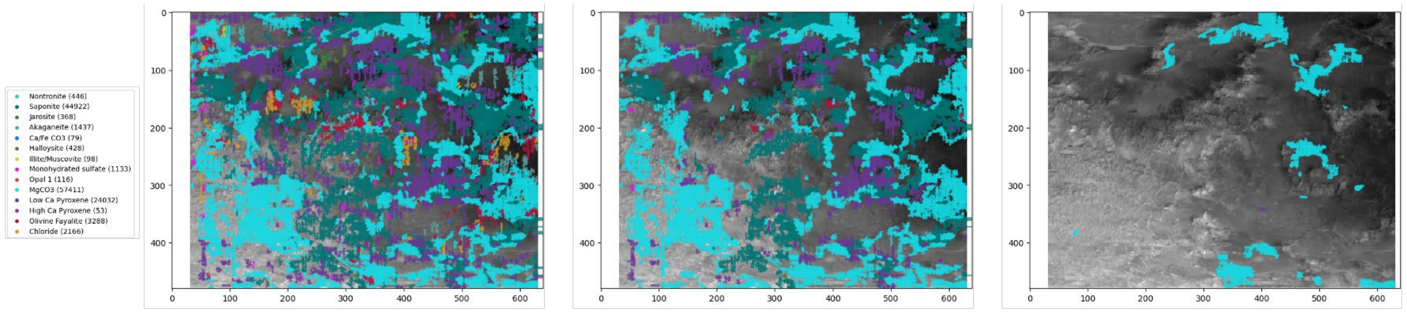
592,413 Spectra

1: 'CO2 Ice'
2: 'H2O Ice'
3: 'Gypsum'
4: 'Ferric Hydroxysulfate'
5: 'Hematite'
6: 'Fe smectite'
7: 'Mg smectite'
8: 'Prehnite'
...
28: 'Jarosite'
29: 'Serpentine'
30: 'Alunite'
31: 'Akaganeite'
32: 'Ca/Fe CO3'

32 Mineral Class Labels

Denoising: Our technology is based on a Convolutional U-Net architecture





Increasing Filtering

Classification: Our technology is based on Variational AutoEncoders and Generative AI

What does the future look like?

Position paper, npj Artificial Intelligence, accepted the 12 April 2026

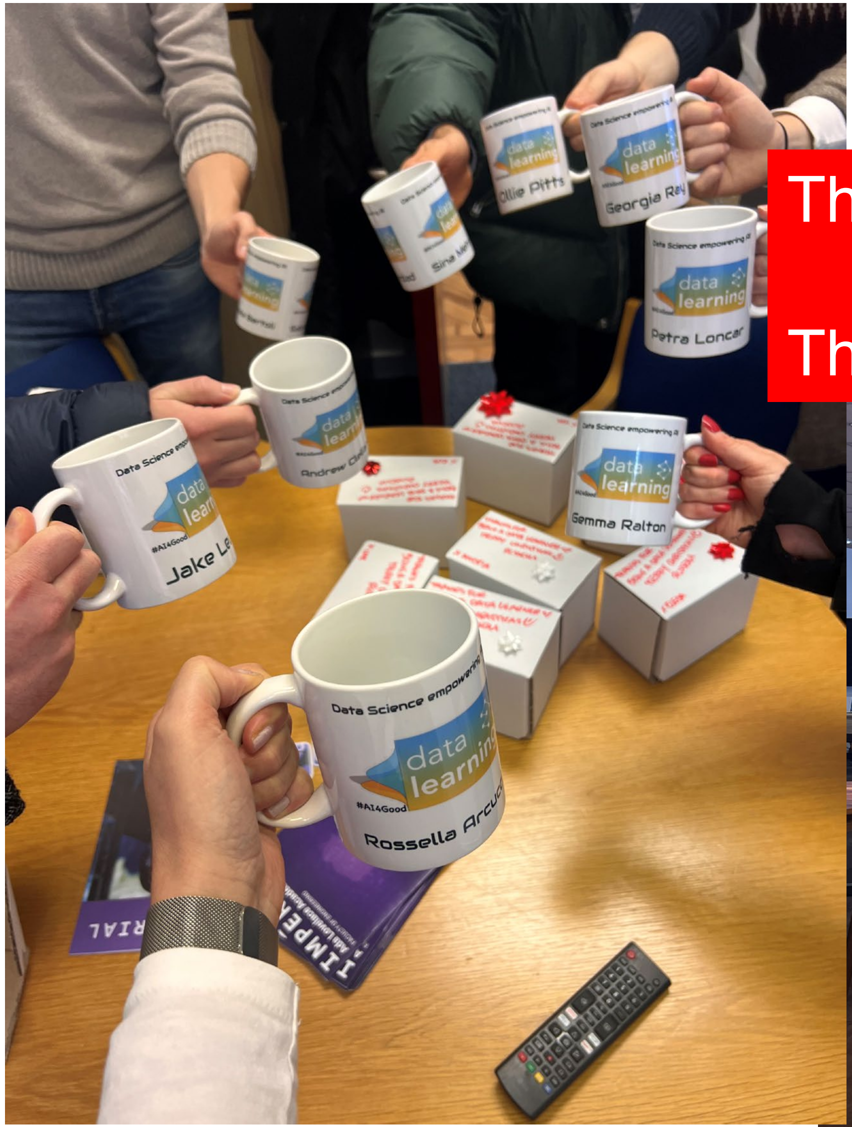
The Convergence of Machine Learning and Data Assimilation in Earth System Science

Rossella Arcucci[1,*], Sean Healy[2], Sarah Dance[3], Lili Lei[4], Eviatar Bach[3], Anthony Weaver[5], Takemasa Miyoshi[6], Maria Eugenia Dillon[17], Clara Draper[13], Rochelle Schneider[19], Simon Lang[2], Peter Dueben[2], Niels Bormann[2], Peter Lean[2], Alan Geer[2], Massimo Bonavita[2], Peter Jan van Leeuwen[8], Sibo Cheng[16], Marc Bocquet[16], Nedjeljka Zagar[15], Haroldo Fraga de Campos Velho[7], Juan Jose Ruiz[9], Peter Bauer[10], Sid Ahmed Boukabara[11], Alberto Carrassi[12], Russ Treadon[13], Andrew Collard[13], Daryl Kleist[13], Azadeh Gholoubi[13], Xuguang Wang[14], Nahidul Samrat[18], Gemma Ralton[1], Andrew M Moore[20], Katia Lamer[21], Nico Caltabiano[22]

Learn from the past, inspire the future.

Thanks to my group!

The Data Learning group



There is nothing measured that doesn't exist.

Thank you!

My book