GVA PROMETEO AI4CS
AI for Complex Systems:
Earth, Brain and Social Systems
Workshop on AI for Complex Systems

Gustau Camps | Emiliano Diaz | Vassilis Sitokonstatinou
Universitat de València
@isp_uv_es · http://isp.uv.es
https://isp.uv.es/projects/prometeo/

WIFI: ESPACIOS / Sirvete01=!

VNIVERSITAT DE VALÈNCIA

GENERALITAT VALENCIANA
Conselleria d'Educació, Cultura i Esport

Welcome!

Gustau

Welcome!

Did you know...?

Lluis Vives (1493–1540) was a Valencian scholar and Renaissance humanist who promoted his insight into early medical practice and natural phenomena and pioneered relevant perspectives on emotions, memory and learning.

He is considered the father of modern psychology and believed that the best and distinct part of the human wits is the ability to understand, remember, reason, evaluate and judge the facts and the world from our limited senses.

The project

Gustau

Why

What

Who

How

When

The funding program

- The PROMETEO program of Generalitat Valenciana funds excellence science groups
- Projects (or rather awards) should aim to consolidate ongoing lines of research, improving collaboration at the national and international level and the dissemination of knowledge
- In practice:
    - Bunch of friends do science together
    - 4 years, 600k€
    - Activities under a common umbrella (in this case “AI for Complex Systems”)

Prometheus carrying the fire (light) to the Humanity

AI4CS: AI for complex systems - Brain, Earth, Climate, Society

Gustau Camps-Valls, María Piles (PIs)

Image Processing Laboratory (IPL)

Universitat de València

4-year project (Jan 2021- Dec 2025)

Research teams







G Camps[1] (PI) (75, 4) M Piles[1] (PI), (25, 2) J Malo[1], (24, 5) L Gómez[1], (40, 4) J Muñoz[1], (33, 3) J Amorós[1], (17, 2)

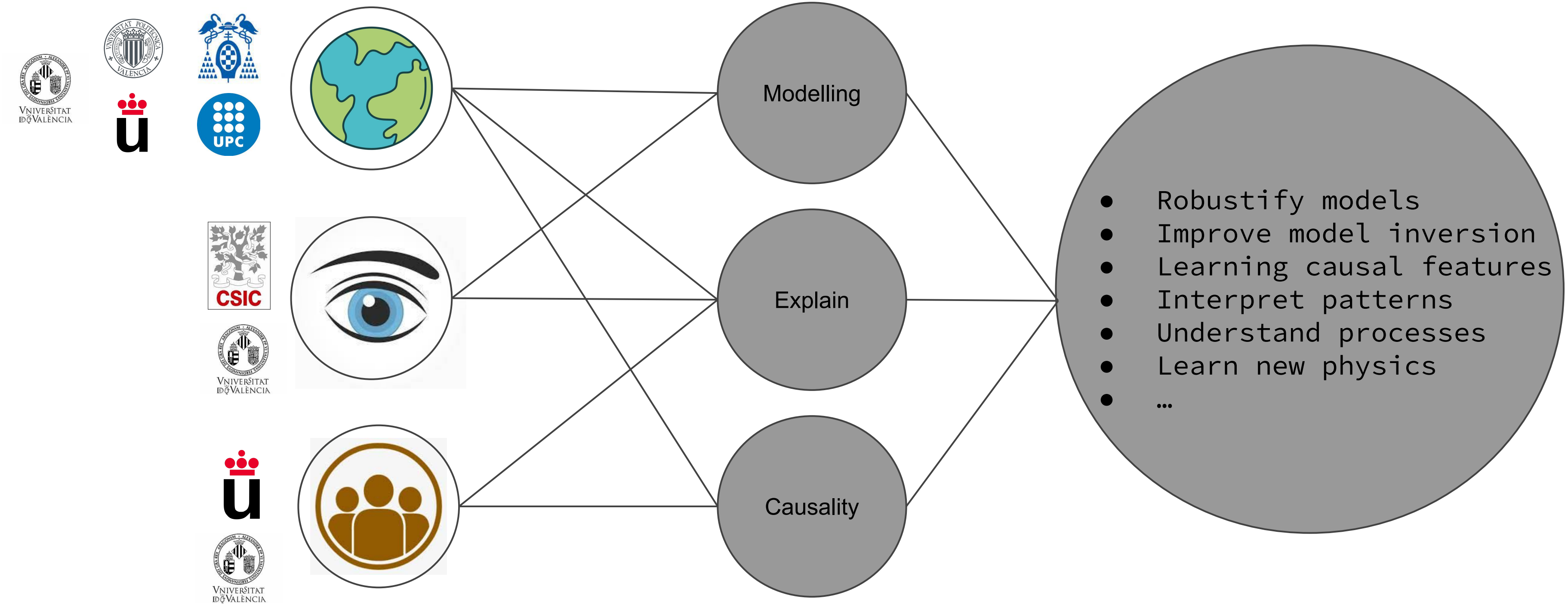






M Bertalmio[2], (35, 3) A Ruescas[1] (12, 1) J García[1] (33, 5) J Esteve[1] (16, 2) S Salcedo[3] (45, 4) A Camps[4] (54, 4) L Guanter[5] (59, 3)

Interactions



Vision

AI4CS aims to develop novel artificial intelligence methods to model and understand complex systems, and more specifically the visual brain, Earth and climate systems and the biosphere-anthroposphere interactions

What?

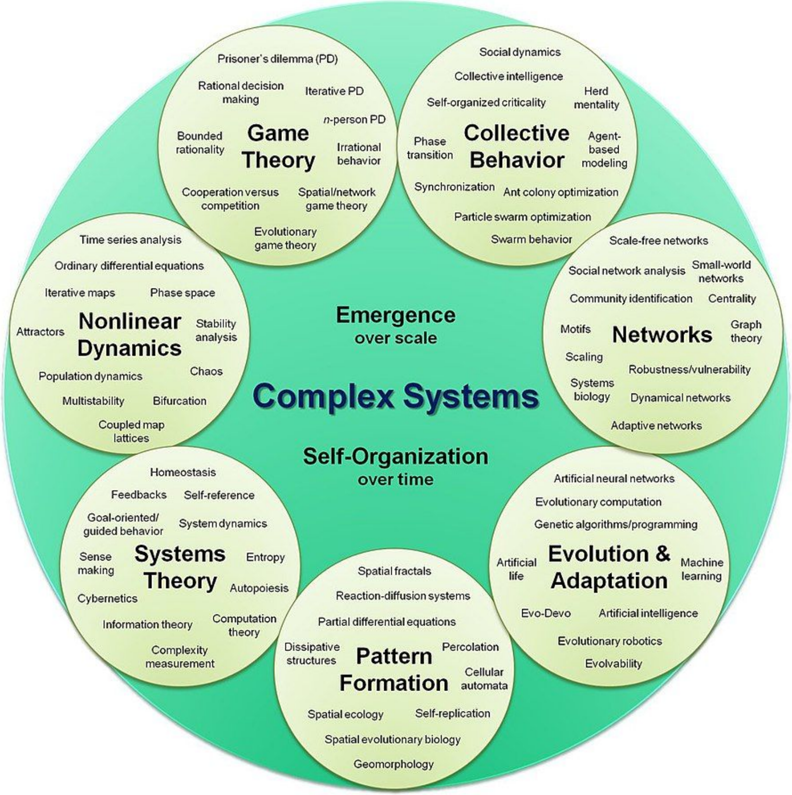
- Modeling
- Understanding
- Intervening

How?

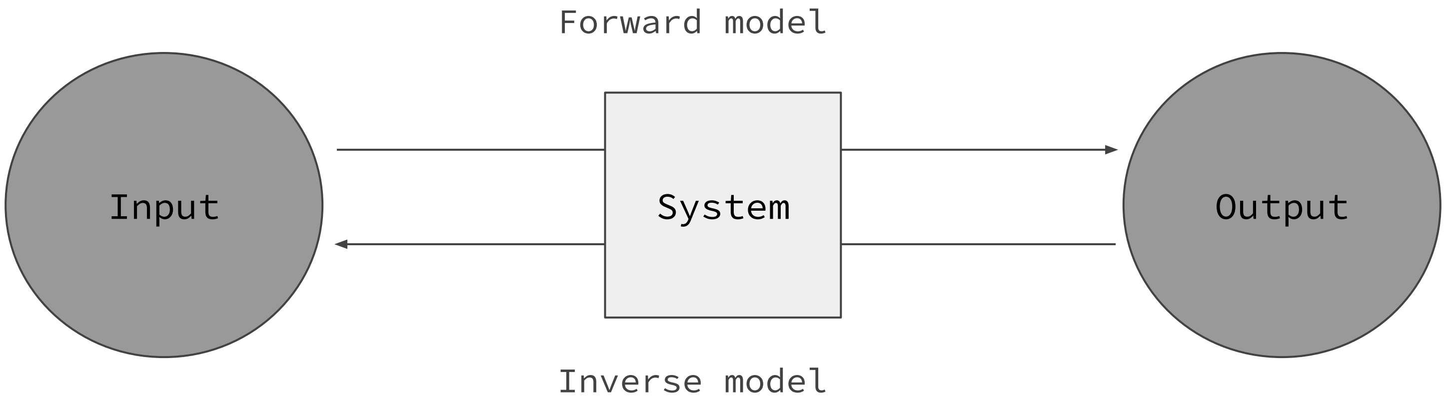
- Hybrid ML modeling
- Causal discovery and XAI
- Causal inference

Why the 'complex systems' framework?

- **Def:** *A complex system is a system composed of many components which may interact with each other*
- **Ex:** climate, organisms, brain, power grid, comm. systems, software & electronic systems, social and economic orgs., ecosystem, a living cell, and … the entire universe
- **RQ1:** How AI help deal with modeling, understanding and interactions?
- **RQ2:** Commonalities in methods and challenges in brain, climate and social systems?

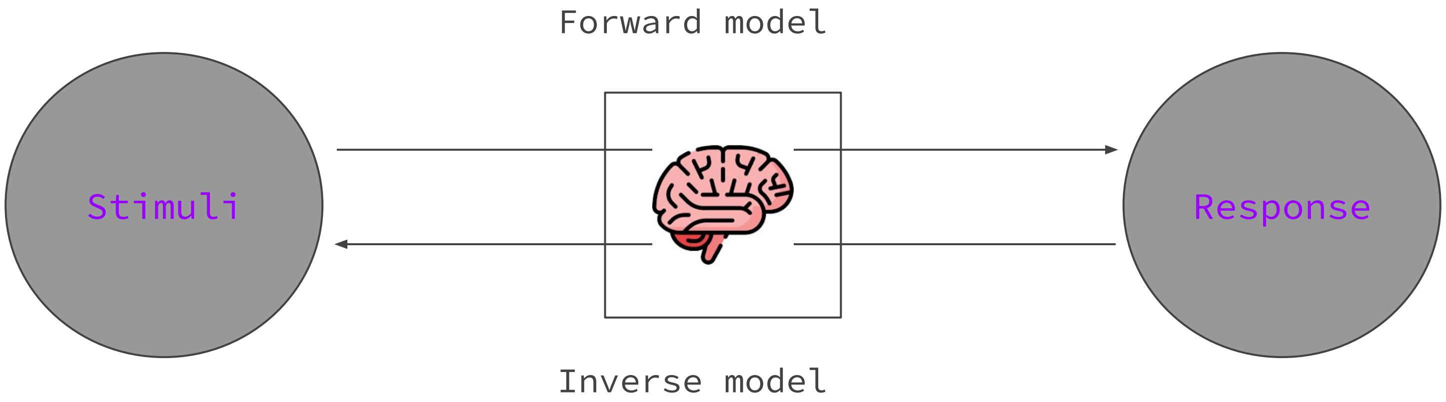


A general forward-inverse system's modeling approach

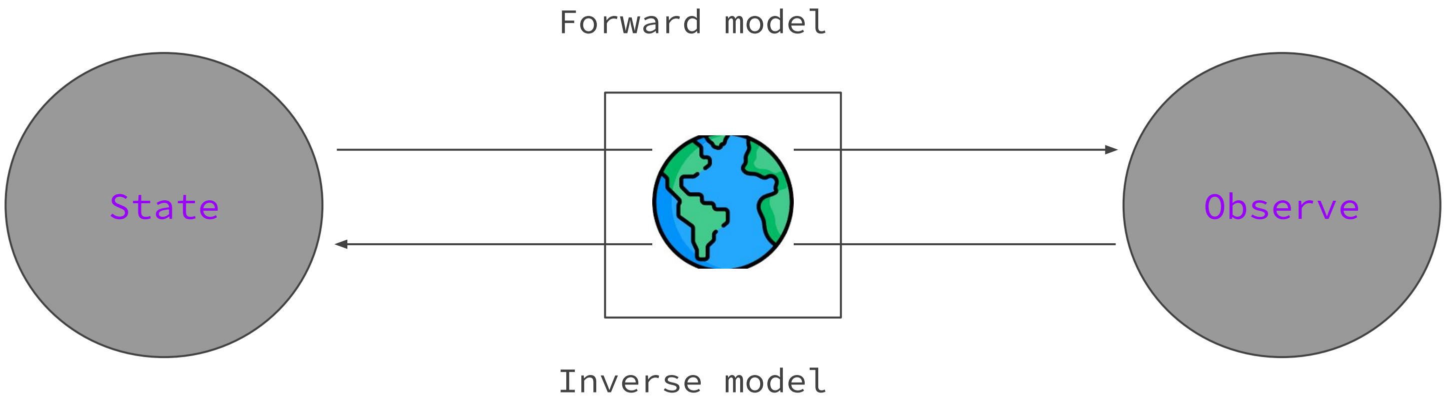


Goals: learn to model, forecast, invert, retrieve features, understand and explain relations, learn to intervene, and assess effects from actions or interventions

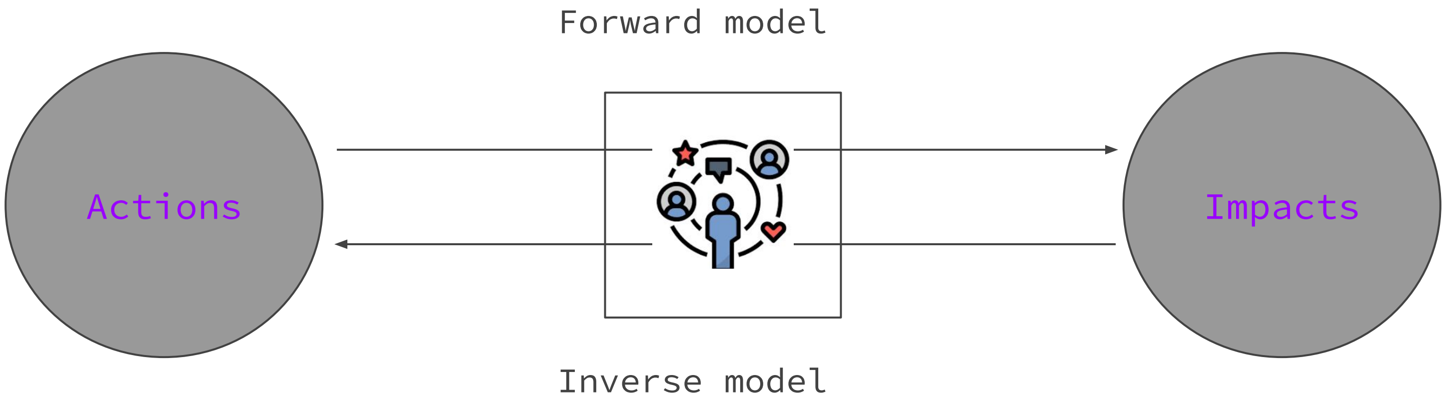
A general forward-inverse system's modeling approach



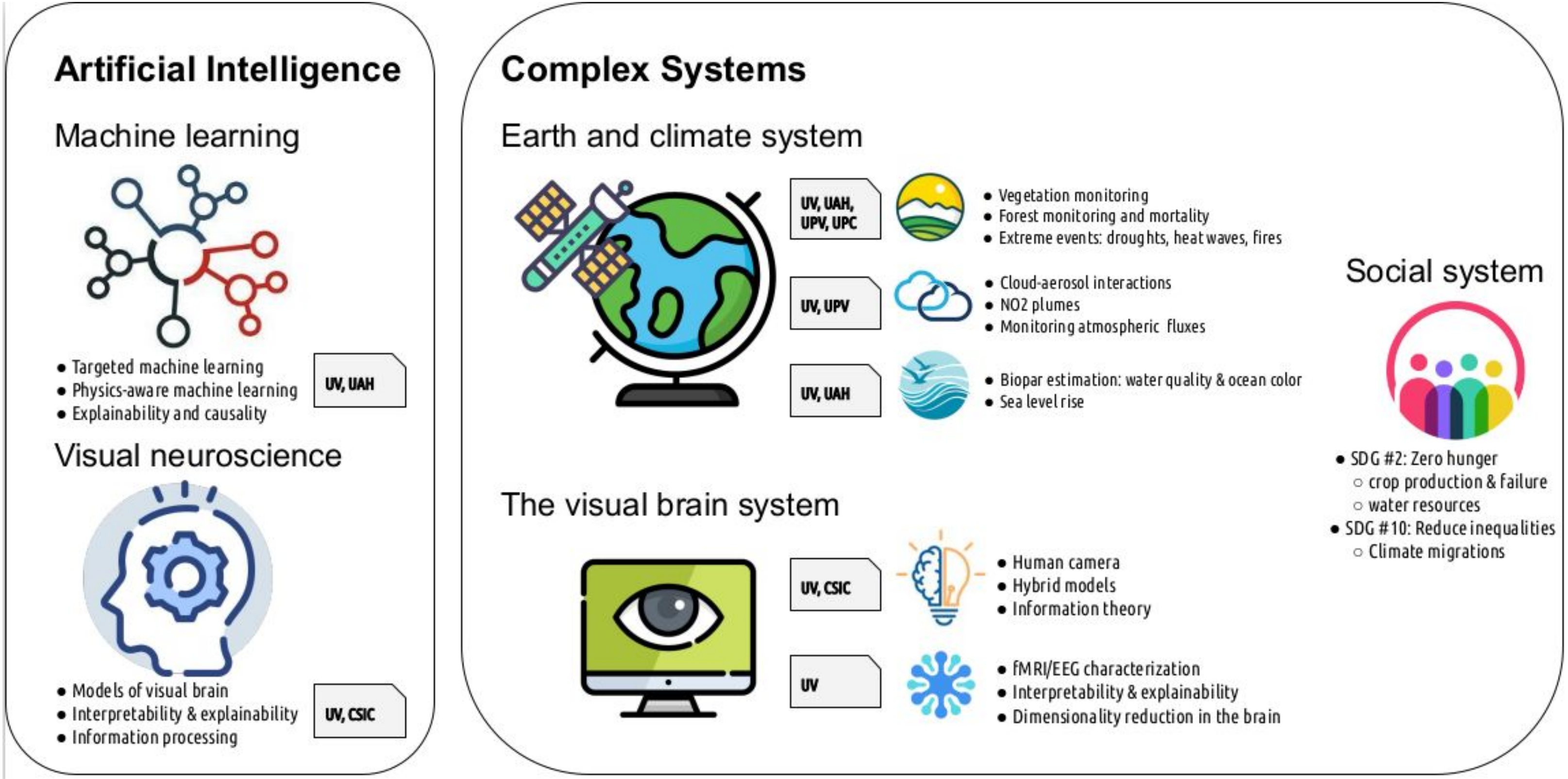
A general forward-inverse system's modeling approach



A general forward-inverse system's modeling approach



The overarching structure



Housekeeping

Emiliano & Vassilis

Organization

Logistics

Housekeeping

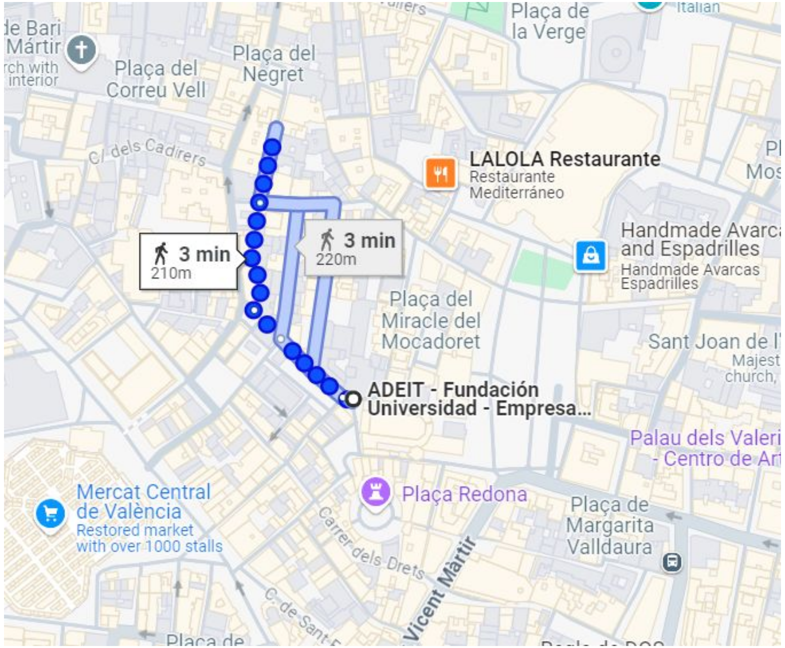
Navigating ADEIT ...

- Rooms - Aulas 1.1 + 1.2
- Coffee and lunch breaks
- Breakouts & walk & talk sessions
- Want advice?
- Need help?

Walk & Talks

Dinner

14-01 - Time: 21.00



Coffee and Lunch

- Thermos Coffee, 1 Milk, and 1 Water with Infusions
- Coffee break every two hours
- 14.01 - 1:30 PM – 3:30 PM: Group Lunch & group photo

Workshop overview

- 2 day workshop on *AI for Complex Systems*
- 6 sessions, 16 talks

  Thematic areas:
    - Earth systems: Land, Water, Atmosphere
    - Socio-environmental systems
    - Vision
    - Machine Learning
- Talks: 20 minute presentation + 10 minute questions
- Breakout sessions
  - 1 h in-depth discussions on selected topics

Agenda

Day 1 – Monday, 13th January

- **2:00 PM – 2:30 PM: Introduction to AI4CS and workshop goals**
  - Overview of AI4CS, objectives, and key challenges
- **2:30 PM – 4:00 PM: Session 1 (S1) – Earth-Land Systems**
  - Presentations: (20+10 min).
- **4:00 PM – 4:30 PM: Coffee Break**
- **4:30 PM – 5:30 PM: Session 2 (S2) – Earth-Water Systems**
  - Presentations: (20+10 min each).
- **5:30 PM – 6:00 PM: Session 3 (S3) – Earth-Atmosphere Systems**
  - Presentations: (20+10 min).

Agenda

Day 2 – Tuesday, 14th January

- **9:00 AM – 11:00 AM: Session 3 (S3) – Earth-Atmosphere Systems**
  - Presentations: (20+10 min each).
  - Breakout Discussions (1 hour): Topics entered and voted using Slido.
- **11:00 AM – 11:30 AM: Coffee Break**
- **11:30 AM – 1:00 PM: Session 4 (S4) – Socio-Environmental Systems**
  - Presentations: (20+10 min each).
- **1:00 PM – 1:30 PM: Session 5 (S5)– Vision**
  - Presentations: (20+10 min each).
- **1:30 PM – 3:30 PM: Group Lunch & group photo**
- **3:30 PM – 4:30 PM: Session 5 (S5)– Vision**
  - Presentations: (20+10 min each)
- **4:30 PM – 5:00 PM: Coffee Break**
- **5:00 PM – 6:00 PM: Breakout Discussions**

Agenda

Day 3 – Wednesday, 15th January

- **9:00 AM – 11:00 AM: Session 6 (S6) –** Machine learning
  - Presentations: (20+10 min each).
  - Breakout Discussions: Topics entered and voted using Slido.
- **11:00 AM – 11:30 AM: Coffee Break**
- **11:30 AM – 1:00 PM: Session 7 (S7) – Conclusions and Wrap-up**

Breakout sessions

- Put your suggested topic on slido
- Vote your favorite topic
- Top voted topics are selected breakout sessions
  - Person who suggested the topic is the moderator of the discussion
  - One volunteer for notes

slido

Join at

slido.com

#6104 203

Q&A

Polls

Type your question

Popular

Recent

1 question

V

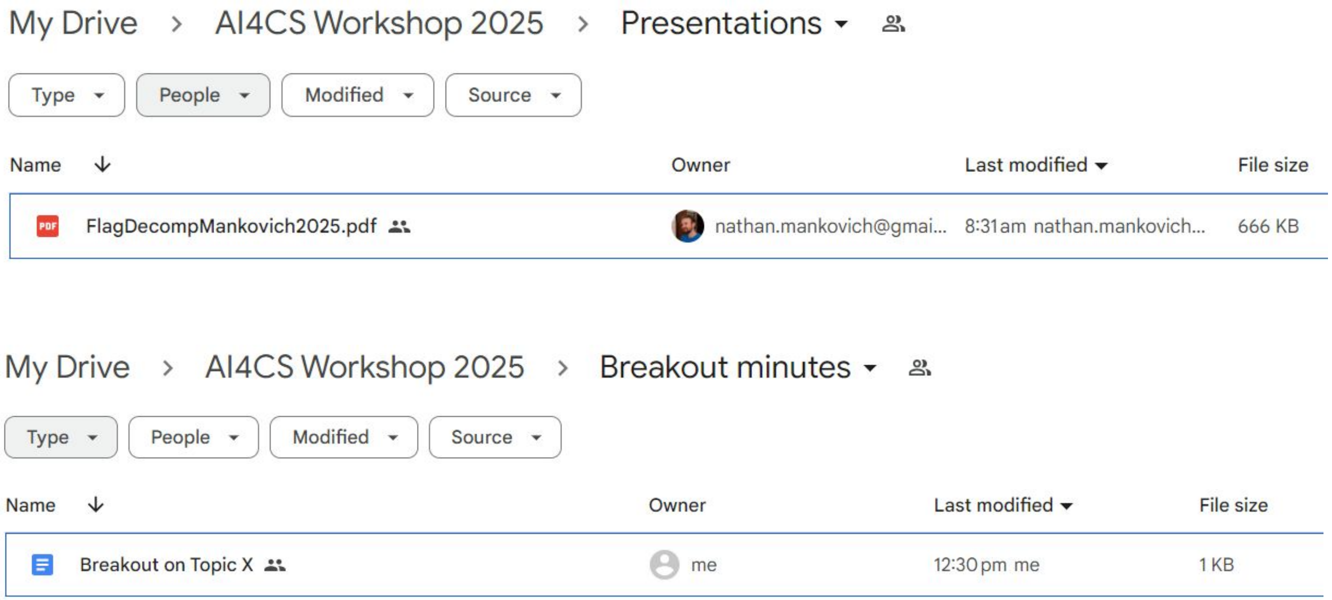
Vassilis

19 minutes ago

1

What are the commonalities in models used for the Earth and the Brain?

Google drive



- Upload slide on *complexity*
- Copy and paste example docs on breakout minutes

Final remarks

Gustau

What for

Legacy

Thanks

Overall goal

- Aim for a serendipitous, open-minded & fun workshop
- Learn from methods and challenges in different domains
- Ask the unquestionable, answer the bold remark
- **Legacy:** Write something out of this? Start collabs? Reframing problems?

Serendipity in Earth Sciences

Gustau Camps-Valls
Image Processing Laboratory (IPL), Universitat de València
gustau.camps@uv.es · https://isp.uv.es

January 11, 2025

Abstract

Throughout history, the most profound advances in science have often been the product of serendipity rather than meticulous planning or directed investment [1, 2]. This phenomenon is particularly evident in Earth sciences. From the origins of geoscience to the latest advancements in environmental monitoring and climate understanding, serendipity has played a decisive role in steering the course of scientific progress [3, 4]. However, the new industry players, available tools, and pace at which science has been done in the last decades challenge it [5, 6]. We discuss the new place and structure of serendipity in the Earth sciences and anticipate new ways to catalyze future breakthroughs.

Table 1: Categories of Serendipity in Science

Name	Definition	How to?
1. Discovery through anomalies	Unexpected findings from errors	Investigate anomalies
2. Repurposing technology	Adapting existing tools for new fields	Apply existing tools creatively
3. Connections across fields	Insights from linking unrelated disciplines	Collaborate across disciplines
4. Fortunate timing/context	Discoveries from chance observations or encounters	Be open to random opportunities
5. Persistence in exploration	Breakthroughs through persistence with ignored phenomena	Continue investigation despite challenges
6. Experimental design	Discoveries from experiments with a different hypothesis	Explore beyond initial hypothesis
7. Environmental serendipity	Discoveries from uncontrolled events or natural settings	Observe natural or field events
8. Insight from unrelated work	Inspiration from past or unrelated research	Reflect on past work or unrelated activities
9. Collaboration serendipity	Discoveries from casual discussions or meetings	Engage in informal scientific dialogue
10. Just luck	Unexpected discoveries without a clear cause	Be ready for random, fortunate occurrences

Thanks!

- Gustau Camps
- Emiliano Diaz
- Vasileios Sitokonstantinou
- Iván Sánchez

name.surname@uv.es

Presentation highlights

- Day 1

Session1: Earth-Land Systems

- Variables such as VOD include components at different temporal scales (phenology, biomass, water-content).
- Challenging to disaggregate temporal scales
- Challenging to aggregate species in a way that properly represents the behavior of a community of individuals of mixed species within a system
- The forward and inverse problem: upscaling result of forward modeling vs. downscaling remote sensing data.
- High complementarity between process based and AI approach (cf. LFMC modeling)
- Fusion of remote sensing, in-situ data and phylogenetic sources of information

Session2: Earth-Water Systems

- Modeling of water systems complicated by the heterogeneity of data sources available, both in terms of types and quality.
- There exists a resistance to AI approaches in the field of remote sensing of water systems.
- Using only the mean to summarize the traits of communities of trees is insufficient. Using additional statistics has proven to be beneficial
- Crowd-sourced data successful in improving representativity of vegetation trait maps

Presentation highlights

- Day 1 & 2

Session3: Earth-Atmospheric Systems

- Clustering + Genetic algorithm approach leads to an interpretable model for extreme forecasting (heatwaves): easily identify relevant spatial clusters, important lags for different drivers
- Methane Plume detection from multi-spectral data difficult because of noise. Possible after denoising channels using multi-temporal, multi-spectral indices.
- Methane application underscores the importance of quality label data in Earth Observation.
- Analogue method in latent space (for heatwave modeling) may be a way to (partially) overcome issues of non-stationarity
- Simulation data from physical models greatly improves the analogue approach by enriching the “library” of similar “historical” points available

Session4: Socio-environmental systems

- In the agriculture domain DL techniques are not state of the art for now. This is due to the size of the datasets available.
- Process guided approach that separate genetic, environmental and management components have shown promising potential.
- Non-stationarity also an important challenge in this domain
- Quality and heterogeneity of datasets an important limiting factor in modeling food insecurity
-

Presentation highlights

- Day 2 & 3

Session5: Vision

- For modeling the visual system in the brain parametric models have proven successful in obtaining easier to train and more interpretable models. Perhaps this extends beyond study of the human visual system to EO?
- Document understanding may provide some clues for EO regarding how to deal with multimodality
- Flag manifolds may be useful when there are hierarchies among sets of variables. Can this be used for VOD modeling that includes modes of variability at 3 temporal scales

Session6: Machine Learning

- Clouds are a source of great uncertainty in Climate Change projections
- Huge data imbalance for the study of clouds using vertical profiles due to the long revisit time of profiling instruments. Self-supervision paradigm useful in this context
- ML can help in multimodality approaches and instrument to instrument translation
- Chaos and Stochasticity are two sources of complexity for Earth System sciences
- Ocean, atomoshphere ocean have both sources, stochastic element less well accounted for
- First order model (Koopman operators) + Diffusion for resolving small scale physics. Successful for kolmogorov flow and shallow water equation

Presentation highlights

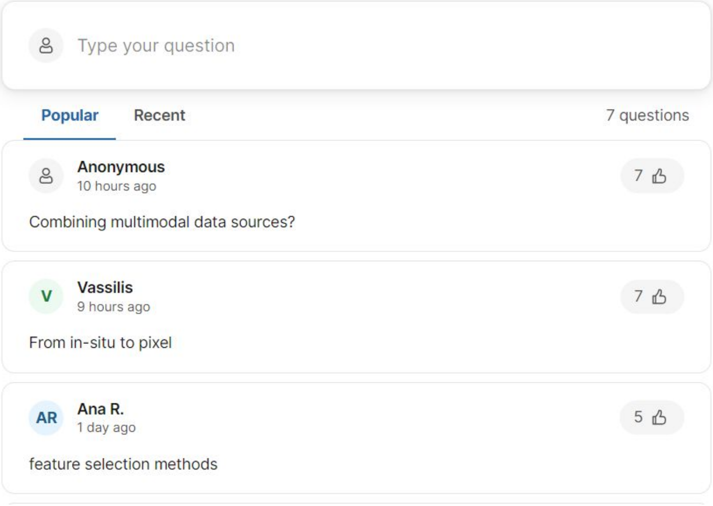
- Day 3

Session6: Machine Learning

- Latent models help represent manifold of ecological parameters
- Use of ML tools within a physical modeling approach that assimilates data useful for inversion and upscaling
- Complex data cycle/AI environments imply different risks
- Explainability vs Interpretability spectrum describes complexity of AI models and increasing ethical issues regarding their deployment

Breakout sessions

- Day 2: 3 topics
  - Combining multimodal data
  - From in-situ to pixel
  - Feature selection/extraction



Breakout sessions

- Day 2: 3 topics
  - Combining multimodal data
    -

Breakout sessions

- Day 2: 3 topics
  - Combining multimodal data

Uncertainty across data modalities:

- Different data modalities exhibit different types of uncertainty, which deep learning (DL) often fails to address adequately.
- The community appears to pay limited attention to this issue, despite its significance.

Data assimilation:

- Data assimilation techniques excel at managing and incorporating uncertainty, a challenge that has been addressed in this field for a long time.
- Representing uncertainty in raw data itself remains a fundamental challenge.

Causal meaning:

- DL is highly effective in summarizing large, heterogeneous datasets but is less suited for uncovering causal relationships and providing causal understanding.

Breakout sessions

- Day 2: 3 topics
  - Combining multimodal data

Foundation models

- **Self-supervised learning** is seen as a promising direction, particularly when building foundation models.
- Combining different modalities often assumes correlations between them. Problems arise when this assumption breaks down, highlighting a need for better handling of uncorrelated modalities.
- There is a growing debate on evaluation criteria for foundation models. Should evaluations prioritize "fit-for-purpose" assessments rather than general benchmarks?

Complexity and Memorization in Models:

- As models grow in complexity, their function shifts from mere memorization to finding the most efficient ways to encode and recall information.

Breakout sessions

- Day 2: 3 topics
  - Feature selection/extraction

- **Feature Selection and Interpretability:**
  - Feature selection is closely linked to the interpretability of complex models, particularly in understanding their behavior.
  - Pre-processing datasets, such as removing temporal or seasonal trends (e.g., temporal indices for anomaly detection), can significantly enhance model accuracy and interpretability.
- **Common Feature Selection Methods:**
  - Widely used methods include:
    - Principal Component Analysis (PCA) for dimensionality reduction.
    - Feature importance scores and relevance measures.
    - SHAP (SHapley Additive exPlanations) for model-agnostic interpretability, though it struggles with high-dimensional data.
    - UMAP (Uniform Manifold Approximation and Projection), which extends PCA for low- to high-dimensionality transitions.

Breakout sessions

- Day 2: 3 topics
  - Feature selection/extraction

- **Challenges in Interpretability:**
  - Interpretability varies significantly across models:
    - Logistic regression is inherently interpretable.
    - Random forests (RF) require additional methods to improve explainability.
    - Deep learning models are the least explainable and often need external interpretability tools.
  - Explainability methods like SHAP sometimes yield less accurate or less consistent results, especially in classification tasks.
- **Role of Explainable AI (XAI):**
  - Explainable AI (XAI) overlaps with interpretable AI, emphasizing both understanding the "how" (interpretability) and "why" (explainability) behind predictions.
  - Model-agnostic interpretation methods like SHAP offer flexibility, as they can be applied across models but may face challenges with certain data types or tasks.

Breakout sessions

- Day 2: 3 topics
  - Feature selection/extraction

- **Specific Use Cases:**
  - **High-dimensional data challenges:** SHAP is less effective for autonomous driving tasks or high-dimensional datasets.
  - **Extreme climatological events:** Interpretability and attribution are critical for understanding the causes and impacts.
  - SHAP can be used to explain prediction models but is less useful for classification tasks unless results are aggregated by class.
- **Pipeline Design for Interpretability:**
  - Designing processing pipelines with interpretability in mind ensures clearer outputs and more accurate feature selection.
  - Training models on outputs derived from SHAP values is a potential avenue for improving model explainability and performance.

Breakout sessions

- Day 2: 3 topics
  - Feature selection/extraction

Conclusions:

1. **Explainable AI (XAI):** Explains why a model made a prediction, while interpretability focuses on how it made the prediction (e.g., logistic regression is inherently interpretable; RF and deep learning need explainability tools).
2. **Feature Selection (FS):** Critical for enhancing interpretability in complex models and is a precursor to accurate ML pipelines.
3. **Traditional Methods:** PCA, feature importance, and relevance measures remain staples for feature selection.
4. **Emerging Techniques:** Increasing use of UMAP and SHAP for dimensionality reduction and prediction problems with high spatial or temporal dimensions.
5. **Explainability in Feature Selection Models:** The best feature selection method is often the one that is interpretable and explainable.
6. **Model-Agnostic Methods:** These provide flexibility, allowing interpretation across a wide range of models, though they may have specific limitations depending on the application.

Breakout sessions

- Day 2: 3 topics
  - In-situ to pixel

1. **Spatializing Point Data:**
   a. Transitioning from point-level (in situ) data to pixel-level data involves significant challenges in scaling and representativeness.
   b. Key question: Can processes learned at the field scale remain valid when applied at the pixel scale?
2. **Key Questions and Challenges:**
   a. **Averaging Challenges:**
      i. Is weighted averaging appropriate for upscaling point data? Many natural processes have non-linear properties, making simple averaging insufficient.
   b. **Causality Across Scales:**
      i. While causal relationships often hold across scales, identifying which processes dominate at different scales remains an open challenge.
   c. **Representativeness of Measurements:**
      i. There are inherent limits to how well point data can represent pixel-level variability. Can we identify these limits and work within them effectively?

Breakout sessions

- Day 2: 3 topics
  - In-situ to pixel

1. **Sources of Variability:**
   a. Variability manifests at different spatial and temporal scales and may not be fully captured by limited observational data.
   b. Examples highlight extreme variability at very fine scales:
      i. **Cameroon:** Tree phenology variability observed even at the branch level.
      ii. **Croplands:** Regional-level phenology studies using anonymized positional data.
2. **Ideas and Proposed Approaches:**
   a. **Modeling Across Scales:**
      i. Develop models conditioned on spatial scale by representing information from both point and pixel levels in a shared latent space. Flags or indicators could potentially help manage scale-specific variability.
   b. **Analogy with Image Compression/Decompression:**
      i. Leverage redundancy and perceptual properties to model relationships between point and pixel scales.
   c. **Weak Learning for Pixel-to-Point Disaggregation:**
      i. Explore methods for learning processes from pixel data and disaggregating them back to point-level information.
3. **Broader Considerations:**
   a. Bridging point-to-pixel information requires understanding the limits of representativeness, variability, and the transferability of processes across scales.

Breakout sessions

- Day 2: 3 topics
  - In-situ to pixel

Key Questions for Future Exploration:

1. How can we design models that effectively bridge the point-to-pixel gap while preserving the underlying variability and causality of processes?
2. What are the limits of weighted averaging, and what alternative methods can better capture non-linear natural processes?
3. Can tools like latent space representations and image compression-inspired methods provide robust frameworks for scaling data?
4. How do we identify and quantify the limits of representativeness in point-to-pixel transformations?
5. Validation of cloud profiles using in situ measurements: how to aggregate across time to deal with non representativity of pixel by pointwise observations
6. Can we learn aggregation functions/statistics for vegetation traits representing communities of trees. A first approach upscaling to plot scale would be appropriate.