Mesogeos: A multi-purpose dataset for data-driven wildfire modeling in the Mediterranean

Spyros Kondylatos[1,2], Ioannis Prapas[1,2], Gustau Camps-Valls[2], Ioannis Papoutsis[1]

[1] Orion Lab, Institute for Astronomy, Astrophysics, Space Applications, and Remote Sensing, National Observatory of Athens
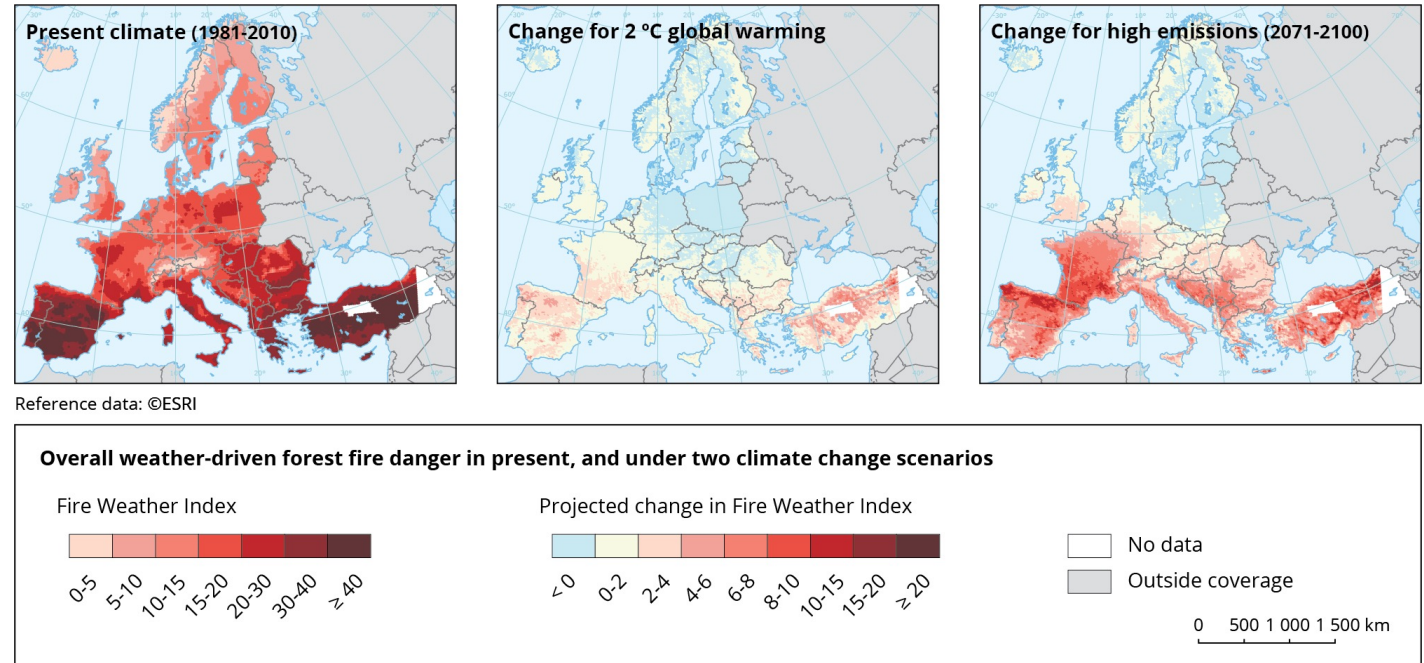
[2] Image Processing Laboratory (IPL), Universitat de València

Motivation

- **Climate change** is expected to **aggravate fire danger** in Europe

- **Mediterranean is a hotspot for climate change** [Lionello & Scarascia]

- Essential to find **new solutions** for mitigating wildfire risk

de Rigo, D., Libertà, G., Houston Durrant, T., Artés Vivancos, T., San-Miguel-Ayanz, J., 2017. Forest fire danger extremes in Europe under climate change: variability and uncertainty.

NEURAL INFORMATION PROCESSING SYSTEMS

Challenges

- Fires result from **complex climate, vegetation and humans' interactions**

- **Machine Learning** can model these dynamics

- **Lack of large-scale, public, well-curated datasets**
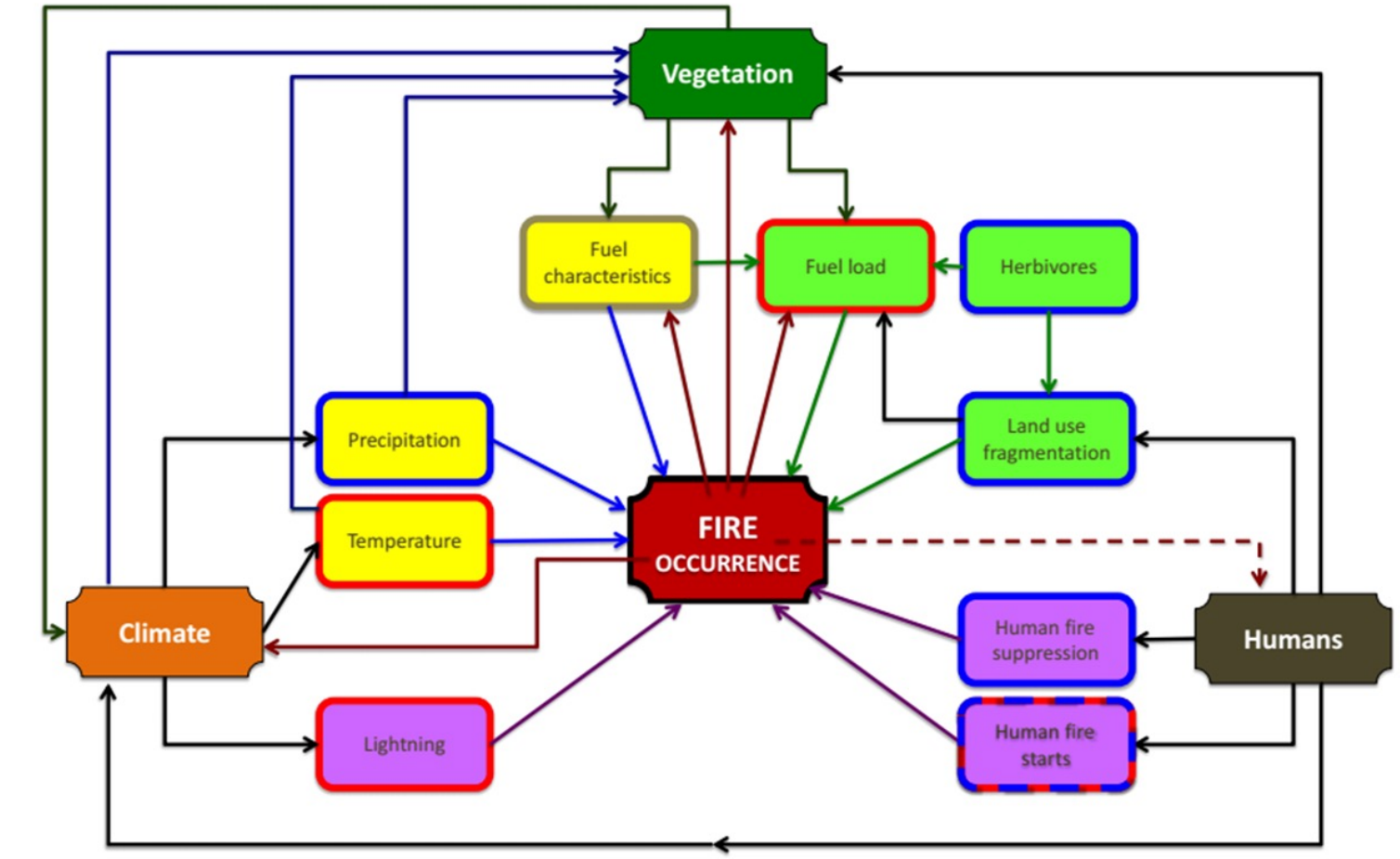
Fire Drivers. *Source: Hantson et al. "The status and challenge of global fire modelling" (2016)*

NEURAL INFORMATION
PROCESSING SYSTEMS

Related Work

Satellite fire products

- MODIS:
  - active fires [Giglio et al.]
  - Burned areas [Giglio et al., Giglio et al.]
- VIIRS [Schroeder et al.]
- FRY [Laurent et al.]
- Fire Atlas [Andela et al.]
- GlobFire [Artés et al.]
- EFFIS [San-Miguel-Ayanz et al.]

Public ML Datasets

- Tailored to single ML objective:
  - Fire Cube [Kondylatos et al.]
  - Next Day Wildfire Spread [Huot et al.]
  - WildfireDB [Singla et al.]
- Smaller scale:
  - Fires in Canada [Sayad et al.]
- Coarse resolution:
  - SeasFire Cube [Prapas et al.]

NEURAL INFORMATION
PROCESSING SYSTEMS

Introducing Mesogeos

Mesogeos: a **spatio-temporal dataset** covering the **Mediterranean**. This dataset consists of **various analysis ready variables** that influence fire ignition and spread alongside **historical burned areas**. Mesogeos serves as a **resource for extracting ML-ready datasets** tailored to various wildfire applications

NEURAL INFORMATION
PROCESSING SYSTEMS

Mesogeos

- ## Mesogeos characteristics:
  - **Large-scale:** covers the Mediterranean (4714 km x 1753 km)
  - **Versatile:** easy access and manipulation
  - **Multi-purpose:** cater to a multitude of wildfire-related tasks

- ## Properties:
  - Datacube structure: **Spatio-temporal grid**
  - Resolution: **1km x 1km x 1-day**
  - Years: **2006-2022**
  - Size**: 3.2 TB**
  - **27** variables from different source: **meteorology** (Copernicus ERA5-Land), **satellite** data from MODIS (vegetation, land surface temperature), **topography**, **anthropogenic factors**,
  - 25722 fire events: **burned areas** & **sizes**, **ignition points** from EFFIS
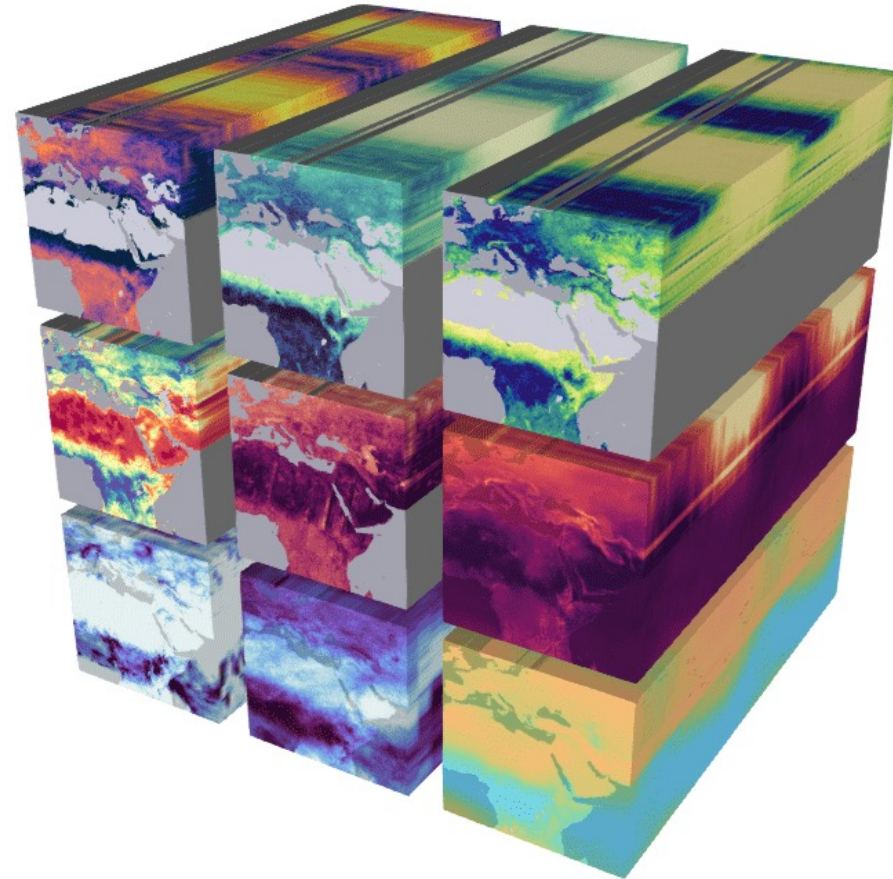
NEURAL INFORMATION
PROCESSING SYSTEMS
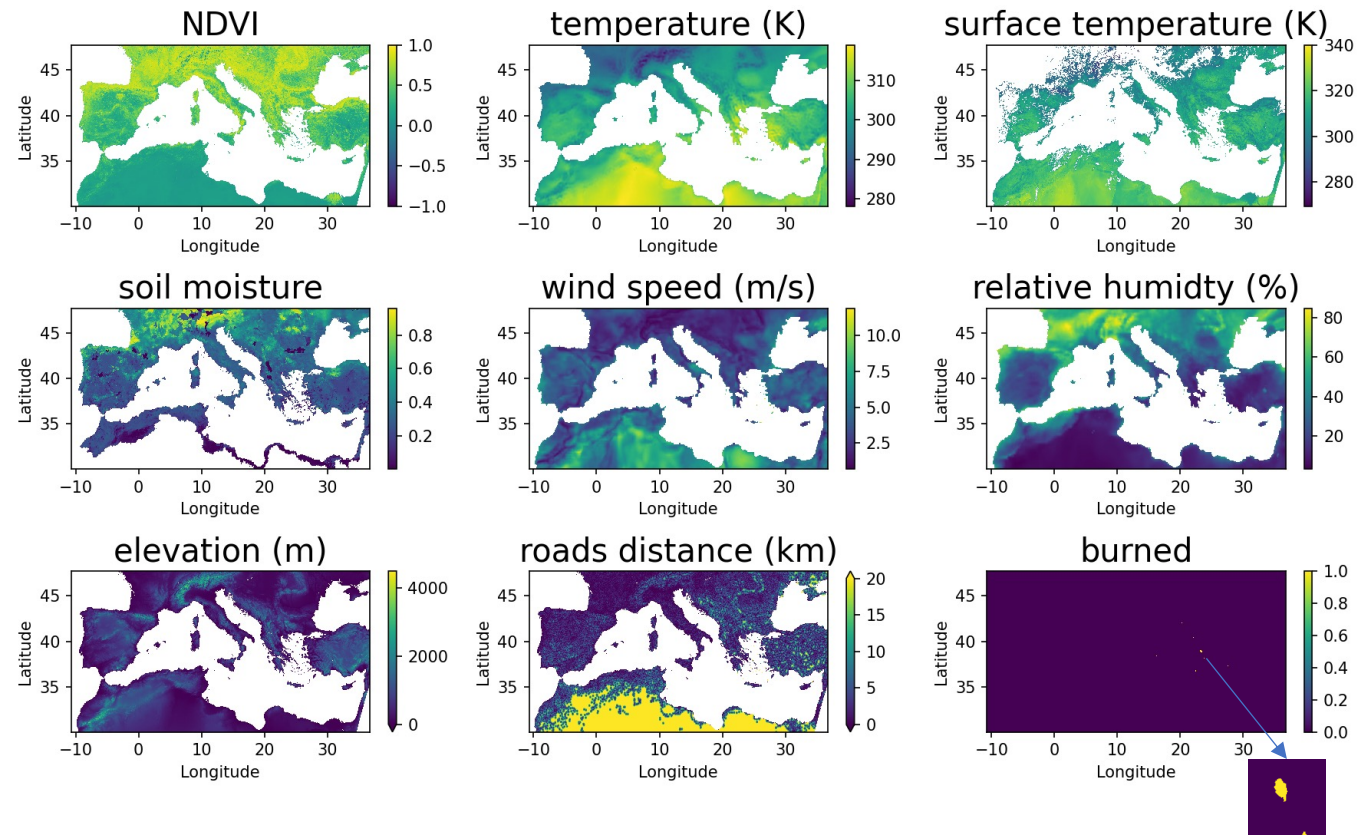
Pipeline for the creation of the datacube
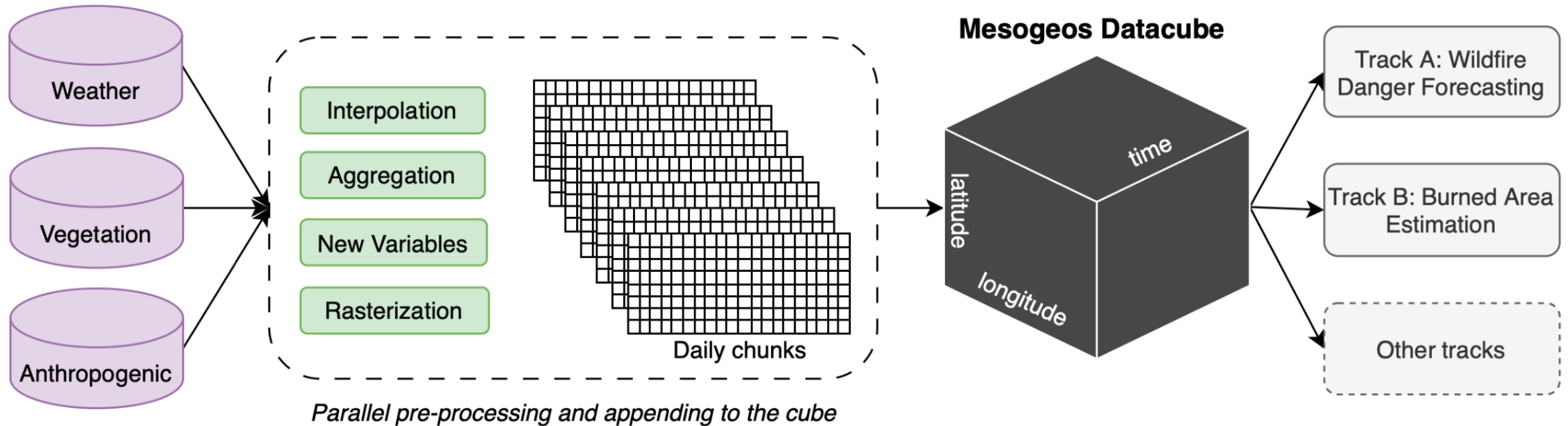
Datacube structure advantages

- Cloud-optimized
- Offers **spatio-temporal metadata**
- This structure facilitates:
  - **Data retrieval & calculation**
  - **Extraction** of ML-ready datasets
  - **Expansion** of the datacube with more variables

Mahecha, Miguel et al. "Earth System Data Cubes Unravel Global Multivariate Dynamics."

NEURAL INFORMATION PROCESSING SYSTEMS

Track A: Wildfire danger forecasting

- Task: **Predict** next day's **fire danger**
- Setup: **time-series** classification task
- ML models provide **good performance** & high-resolution maps

Predicted fire danger for 01-06-2022
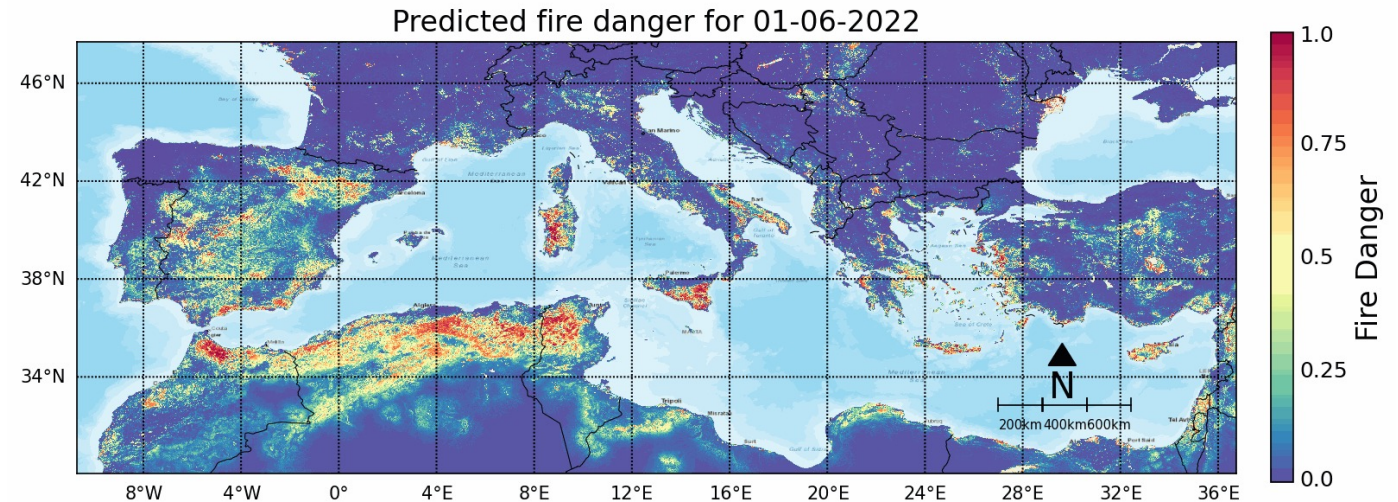
Model	Precision	Recall	F1-Score	AUPRC
LSTM	0,763	**0,812**	**0,786**	0,853
Transformer	**0,802**	0,759	0,780	0,856
Gated Transformer	0,781	0,790	**0,786**	**0,858**

NEURAL INFORMATION PROCESSING SYSTEMS

Track B: Final Burned Area Prediction

- **Predict the likely extent of a burned area**
- Setup: binary **segmentation**
- **Hard task**: Including fire drivers slightly increases performance
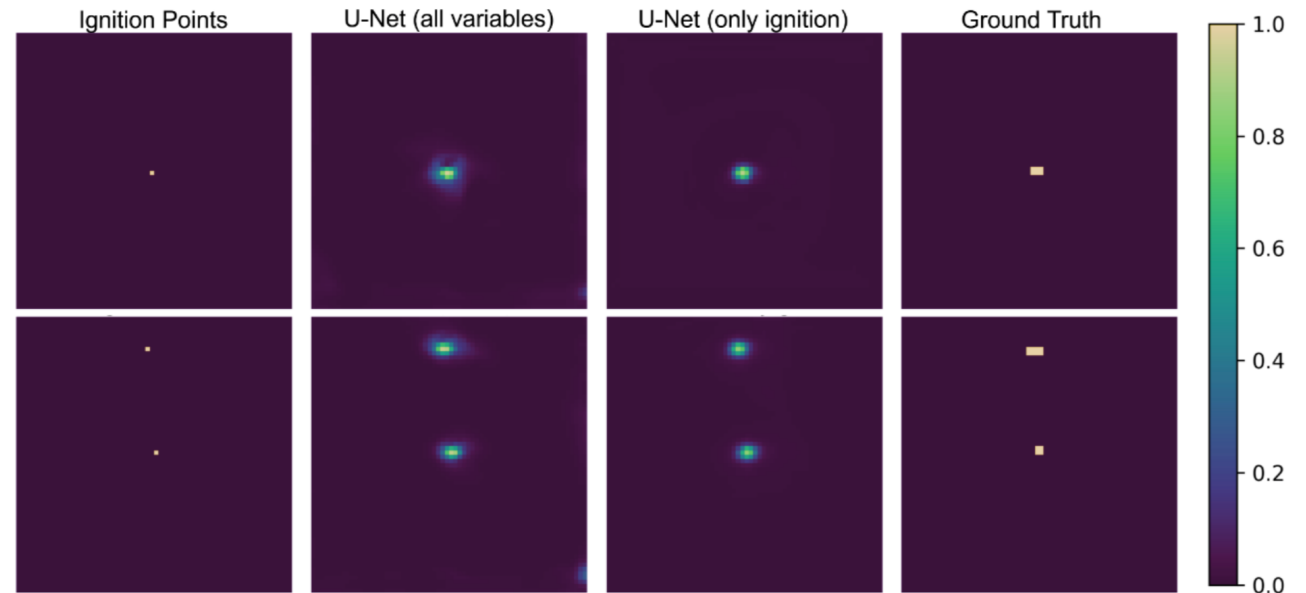
Model	CE Loss	AUPRC
U-Net (only ignitions)	0,0177	0,394
U-Net (all variables)	**0,0166**	**0,418**

Other tasks to explore

1. Fire size prediction

2. Extreme events forecasting

3. Wildfire susceptibility mapping

4. Self-supervised learning

5. Modeling at different spatial and temporal scales

6. Beyond traditional ML (e.g. causality)

NEURAL INFORMATION PROCESSING SYSTEMS

Summary

- Machine Learning holds promise for modeling wildfires in a data-driven way

- Mesogeos: the largest, harmonized dataset for data-driven wildfire modeling

- Mesogeos facilitates the development of diverse applications in wildfire research

NEURAL INFORMATION
PROCESSING SYSTEMS

Contribute with new ML-ready datasets & tracks

- Mesogeos datacube, derived datasets, models, and code are **openly accessible**
- **GitHub** link for reference:
  - https://github.com/Orion-AI-Lab/mesogeos
- **Contribute** with models and tracks and take a place in the leaderboard:
  - https://orion-ai-lab.github.io/mesogeos/

Poster Session: 17:00 – 19:00
Poster Number: 104