Identifying Spatio-Temporal Drivers of Extreme Events

Mohamad Hakam Shams Eddin and Jürgen Gall

[University of Bonn] [Lamarr Institute]

QR code

Contact: shams@iai.uni-bonn.de
gall@iai.uni-bonn.de

NEURAL INFORMATION PROCESSING SYSTEMS

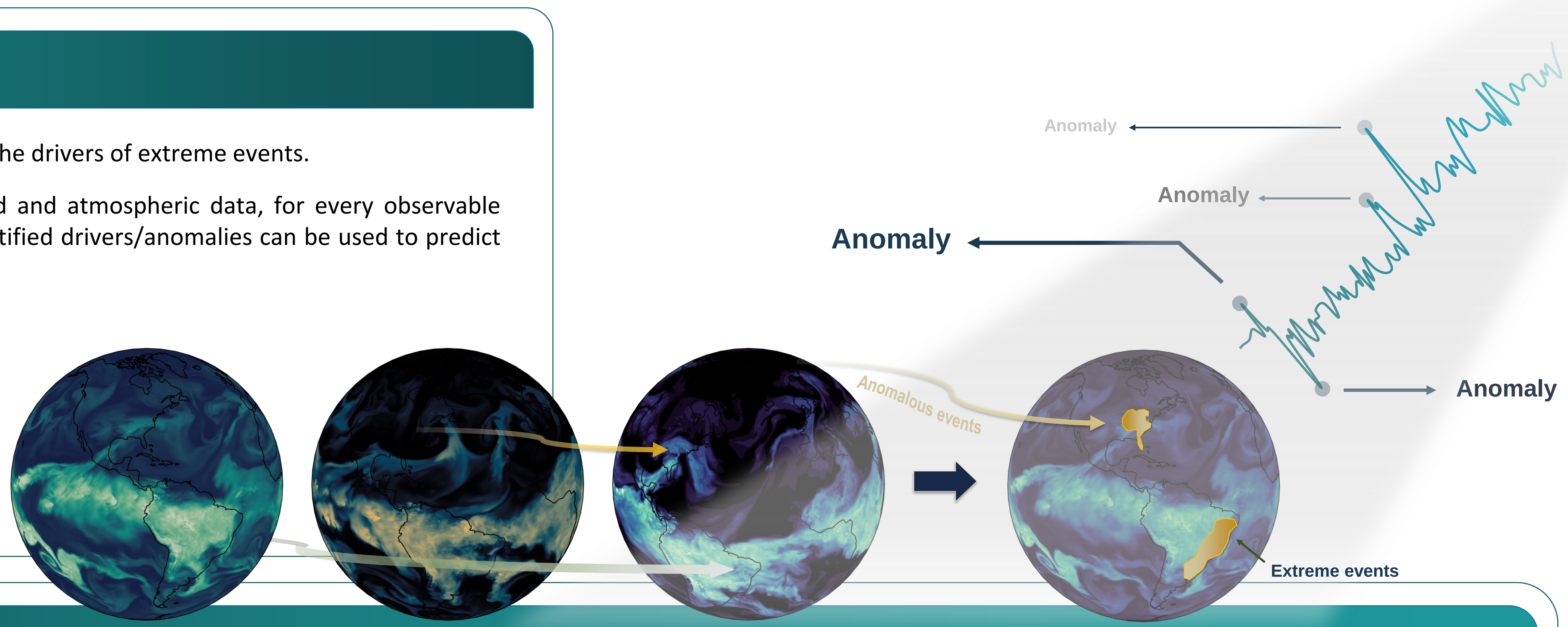
CRC 1502 DETECT

LAMARR INSTITUTE FOR MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

UNIVERSITÄT BONN

Introduction

- We introduce a deep learning model, designed to leverage climate data to identify the drivers of extreme events.
- We assume that there exist precursor drivers, primarily as anomalies in the land and atmospheric data, for every observable impact of extremes. We verify our method by measuring to which degree the identified drivers/anomalies can be used to predict extreme agricultural droughts.
- **Key hypothesis** is that the model will be able to:
  - I. identify spatio-temporally drivers and anomalies in climate data
  - II. quantify the impact of anthropogenic drivers on anomalous events
  - III. identify sets of variables that are currently not investigated for predicting extremes using statistical methods.



Method

- **Input:** Reanalysis or Earth observation data $\mathbf{X} \in \mathbb{R}^{V\times C\times T\times Lat\times Lon}$ (input variables; Input channels)
  + Valid regions $\mathbf{S} \in \mathbb{Z}_2^{T\times Lat\times Lon}$
- **Output:** Identified drivers and anomalous events $\mathbf{Q} \in \mathbb{Z}_2^{V\times T\times Lat\times Lon}$ at timesteps $-\Delta t_7 \rightarrow \Delta t_0$.
  + Predicted extreme agricultural drought events $\mathbf{E} \in \mathbb{Z}_2^{Lat\times Lon}$ at the timestep Δt_0.
- **Model components:**

  feature extractor $f_\theta: \mathbf{X} \rightarrow \mathbf{Z}$, vector quantizer $q_\phi: z \subset \mathbf{Z} \rightarrow z_q$, classifier $g_\psi: \mathbf{Z}_q \rightarrow \mathbf{E}$

  where $z_q = \mathrm{Linear}\big(\mathrm{sign}(z_l)\big) = \mathrm{Linear}\big(-\mathbb{1}_{\{z_l\leq 0\}} + \mathbb{1}_{\{z_l>0\}}\big)$,

  $q = \mathbb{1}_{\{z_l>0\}}, \quad z_l \subset \mathbf{Z}_l = \mathrm{Linear}(\mathbf{Z}) \in \mathbb{R}^{V\times 1\times T\times Lat\times Lon}$.
- **Objective function:**

$$\min_{\theta,\phi,\psi} \underbrace{\mathcal{L}_{(extreme)}(\mathbf{E},\hat{\mathbf{E}},\mathbf{S})}_{\text{predicts extremes from drivers}} + \underbrace{\mathcal{L}_{(quantize)}(\mathbf{Z}_l)}_{\text{encourages confident quantization}} + \underbrace{\mathcal{L}_{(driver)}(,\mathbf{Z}_q\, \hat{\mathbf{E}}_t,\mathbf{S},\mathbf{Z}_{q=0})}_{\text{assigns drivers to the same code in the codebook}},$$

$$\mathcal{L}_{(extreme)} = -\sum_v^{V+1} \big(\hat{\mathbf{E}} \log(\mathbf{E}_v) + (1-\hat{\mathbf{E}}) \log(1-\mathbf{E}_v)\big)\, \mathbf{S}$$

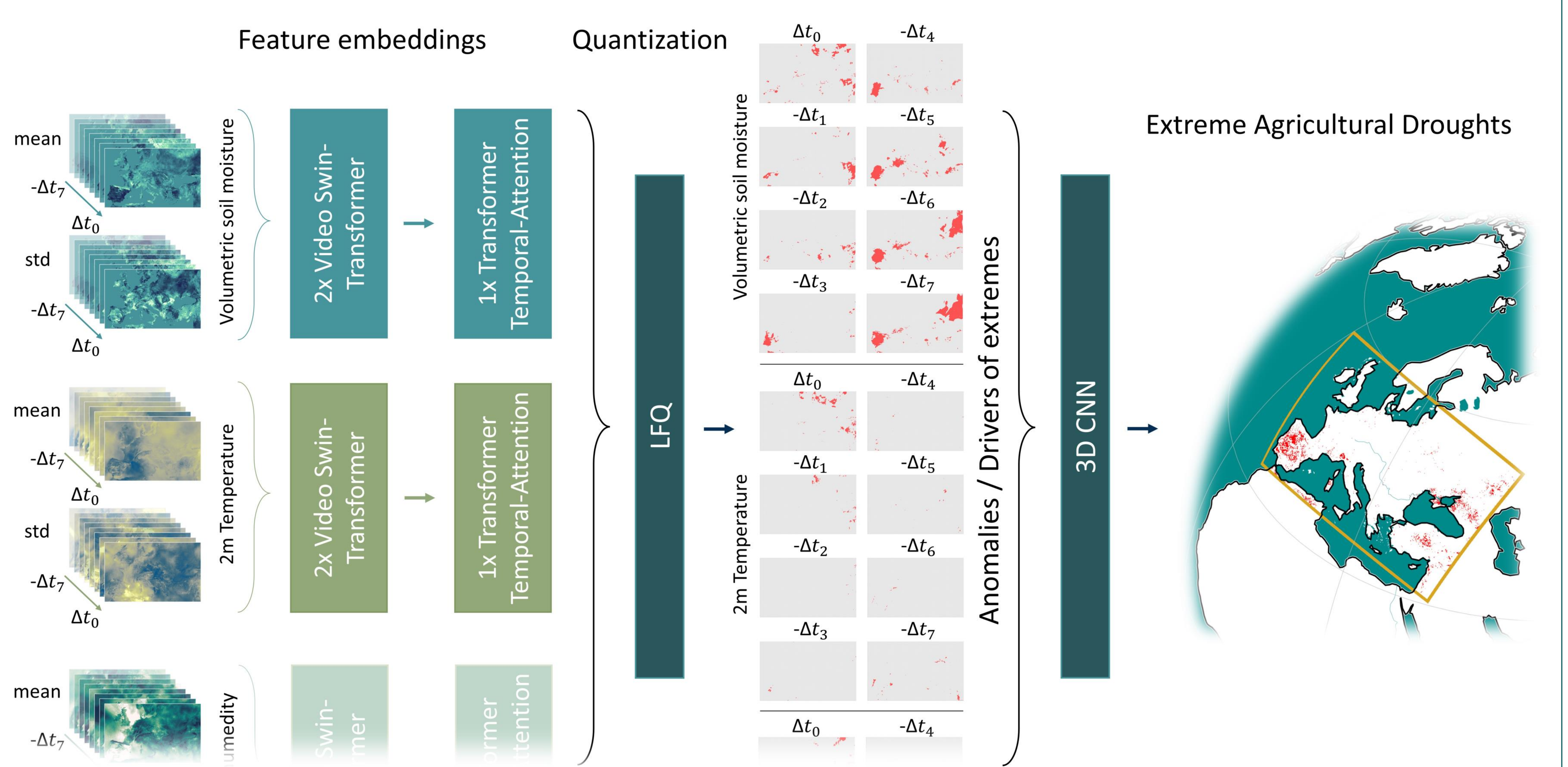
ground truth; predicted extremes from variable v. $\mathbf{E}_{v=0}$ is the multivariate prediction; mask of valid pixels

$$\mathcal{L}_{(quantize)} = \lambda_c \,\|\mathbf{Z}_l - \mathrm{sg}(\mathbf{Z}_l)\|_2^2 + \lambda_e\, \mathbb{E}\big[H\big(\mathrm{sign}(\mathbf{Z}_l)\big)\big] - \lambda_d\, H[\mathbb{E}(\mathrm{sign}(\mathbf{Z}_l))]$$

weight; stop gradient; entropy

$$\mathcal{L}_{(driver)} = \lambda_a \left|\mathbf{Z}_q - \mathrm{sg}(\mathbf{Z}_{q=0})\right| \big(1-\hat{\mathbf{E}}_t\big)\, \mathbf{S}$$

quantization code of the normal data; union of extremes at all time steps



Synthetic data

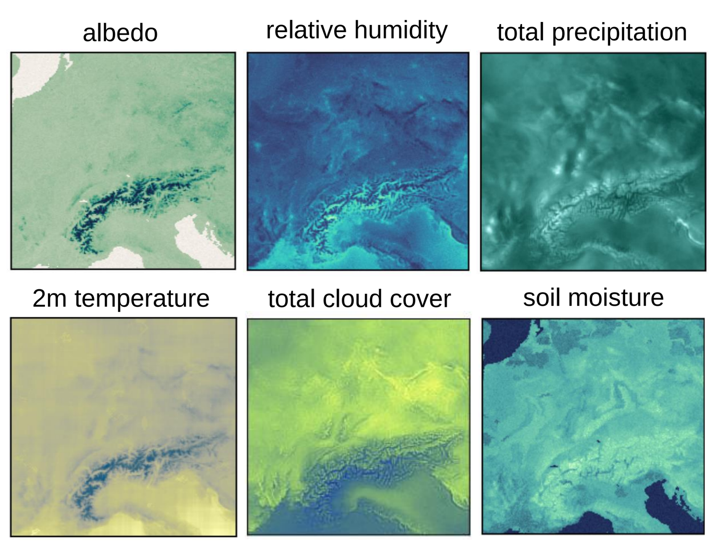
I. First, we generate the normal base signal $\mathbf{B} \in \mathbb{R}^{V\times T\times Lat\times Lon}$ from the climatology.

II. Next, we induce binary extreme events $\mathbf{E}^{ex} \in \mathbb{Z}_2^{T\times Lat\times Lon}$ and track their exact locations.

III. Based on $\mathbf{E}^{ex}$ and a randomly predefined coupling matrix $\mathbf{M}$, we generate the binary anomalous events $\mathbf{E}^a$.

IV. We generate random anomalous events $\mathbf{E}^r \in \mathbb{Z}_2^{V\times T\times Lat\times Lon}$. Unlike $\mathbf{E}^a$, $\mathbf{E}^r$ is uncorrelated with the extreme.

V. Finally, we sample noise signals $\mathbf{N} \in \mathbb{R}^{V\times T\times Lat\times Lon}$ and the artificial signals $\mathbf{\Phi}_{(v,t,lat,lon)}$ are generated as:



Generated artificial dataset

$$\mathbf{\Phi}_{(v,t,lat,lon)} = \mathbf{B}_{(v,t,lat,lon)} + \mathbf{\Lambda}_{(v,t,lat,lon)} \cdot \mathbf{\Theta}_{(v,t,lat,lon)},$$

induced anomalies; controls the sign

$$\mathbf{\Theta}_{(v,t,lat,lon)} = \mathbf{B}_{(v,t,lat,lon)} \cdot \Big(2^{(kb.(\mathbf{E}^r_{(v,t,lat,lon)} \vee \mathbf{E}^a_{(v,t,lat,lon)}))} - 1\Big) + \mathbf{N}_{(v,t,lat,lon)} \cdot 2^{(kn.(\mathbf{E}^r_{(v,t,lat,lon)} \vee \mathbf{E}^a_{(v,t,lat,lon)}))} + ks.\,(\mathbf{E}^r_{(v,t,lat,lon)} \vee \mathbf{E}^a_{(v,t,lat,lon)})).\,\sigma_N,$$

control parameter; standard deviation of the noise

$$\mathbf{\Lambda}_{(v,t,lat,lon)} = \begin{cases} \dfrac{\mathbf{\Delta}_{(v,t,lat,lon)}}{\delta}, & \text{if } \mathbf{E}^a_{(v,t,lat,lon)} = 1, \\ +1, & \text{otherwise,} \end{cases}$$

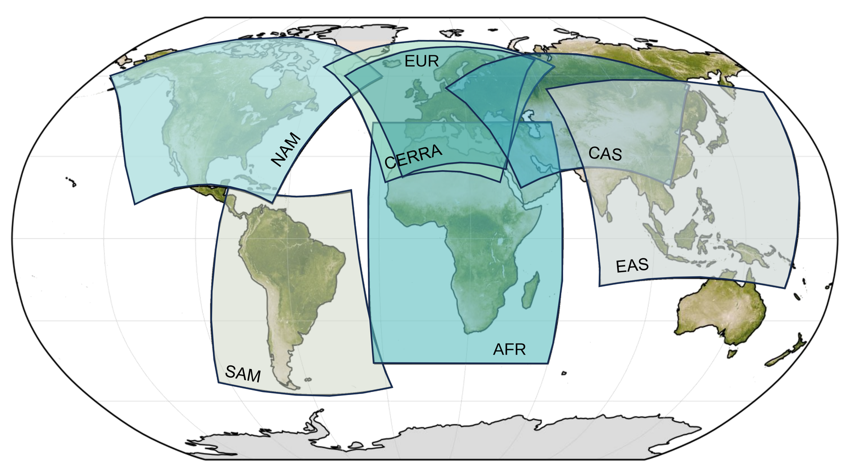
predefined coupling sign from $\mathbf{M}$

$$\mathbf{\Delta}_{(v,t,lat,lon)} = -\mathbb{1}\{\mathbf{\Theta}_{(v,t,lat,lon)} \leq 0\} + \mathbb{1}\{\mathbf{\Theta}_{(v,t,lat,lon)} > 0\}.$$

Reanalysis data

- Reanalysis provides a reconstruction of the historical Earth system state as close to reality as possible.
- We conducted the experiments on two real-world reanalysis (ERA5-Land[1] and CERRA[2]) including data from five continents.

- **ERA5-Land Reanalysis:**
  - 1981 – 2024
  - ~11 km hourly
  - Global coverage



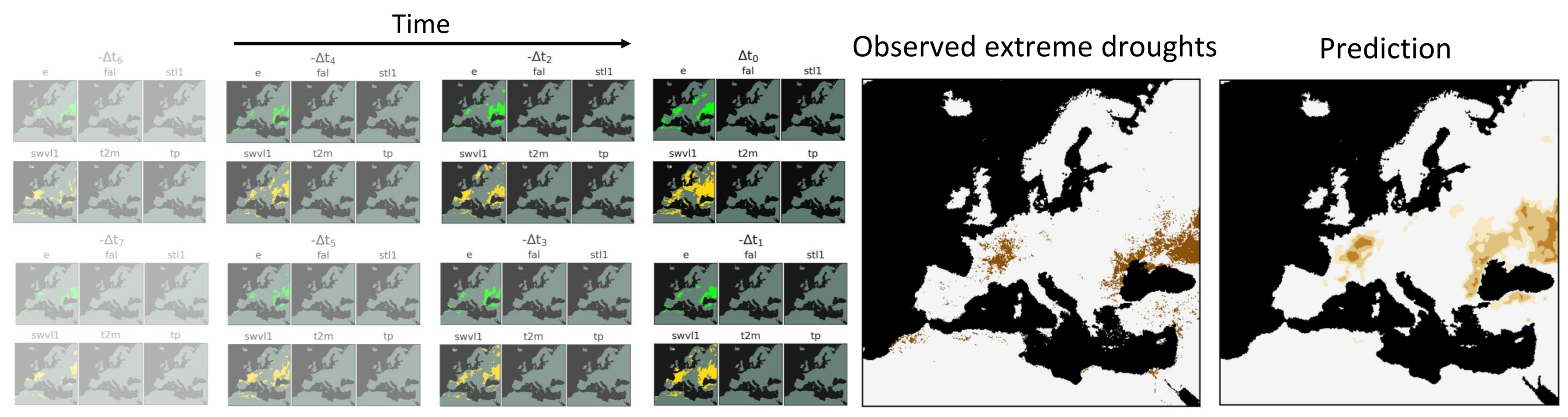
- **CERRA Reanalysis:**
  - 1984 – 2021
  - ~5.5 km 3-hourly
  - Europe

- Reanalysis data include variables such as:

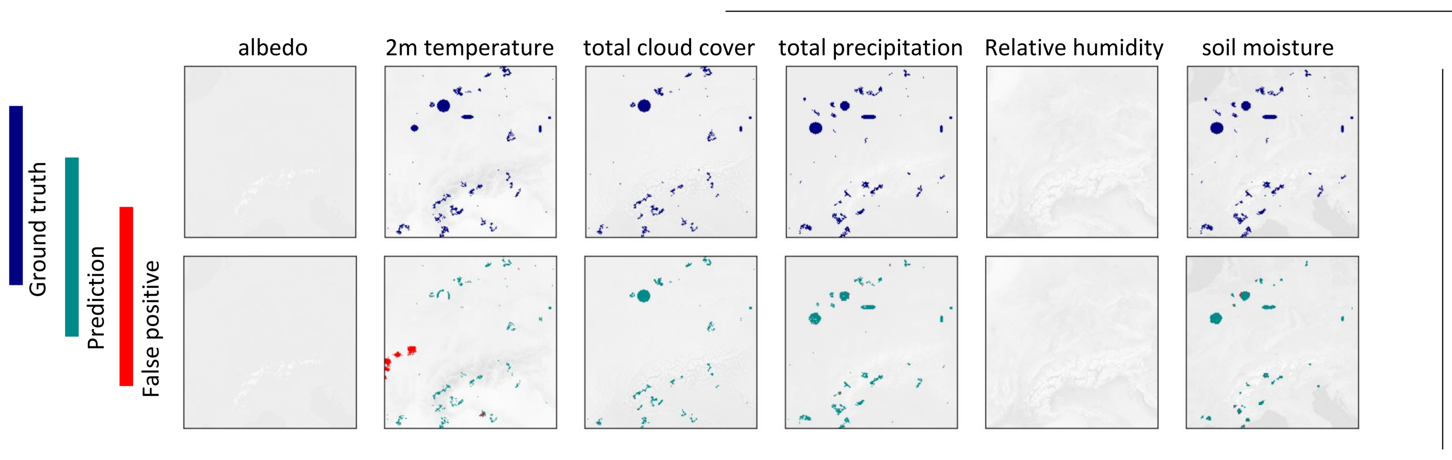
2-meter temperature (t2m)	total precipitation (tp)	albedo (al)
2-meter relative humidity (r2)	skin temperature (skt)	soil temperature (stl)
volumetric soil moisture (swv)	total evaporation (e)	surface pressure (sp)
total cloud cover (tcc)	2-meter dewpoint temperature (d2m)	

Results

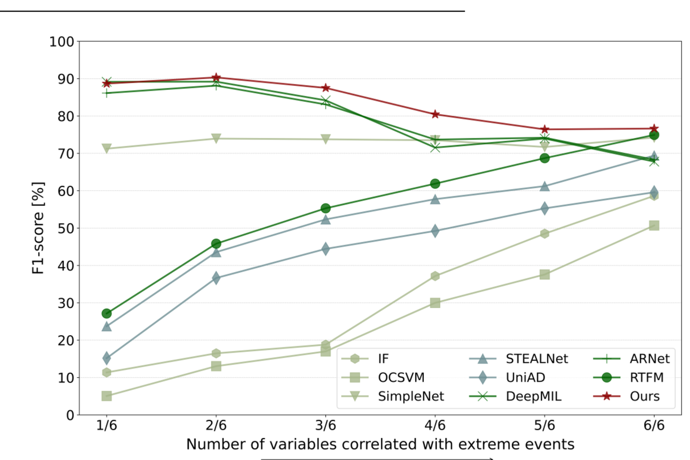
- Qualitative results on ERA5-Land reanalysis data.
- Shown is the prediction for the week 35 and year 2020 over Europe.



Identified drivers and anomalous events localized spatio-temporally 7 weeks before the extreme agricultural drought events.



Qualitative results for anomaly detection on the synthetic CERRA Reanalysis.



Comparison with anomaly detection baselines on the synthetic CERRA Reanalysis.

Conclusions

- We present a first approach and benchmarks to identify the spatio-temporal relations between extreme climate events and their drivers.
- Total evaporation (e) and soil moisture (swv) are the most relevant variables to detect drivers of extreme agricultural droughts.
- An unsolved research problem is identifying the cause of an anomaly. With this respect, we aim to use the model to analyze which anthropogenic drivers can cause a statistically significant increase of anomalous events.

Acknowledgment

The framework to generate synthetic data is inspired by Flach et. al (Multivariate anomaly detection for Earth observations: a comparison of algorithms and feature extraction techniques, Earth Syst. Dynam., 8, 677–696, 2017).

[1] Muñoz-Sabater et .al (ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, 2021).

[2] CERRA sub-daily regional reanalysis data for Europe on single levels from 1984 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS).

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1502/1–2022 – project no. 450058266 within the Collaborative Research Center (CRC) for the project Regional Climate Change: Disentangling the Role of Land Use and Water Management (DETECT) and by the Federal Ministry of Education and Research (BMBF) under grant no. 01IS24075C RAINA.