# SALAMA 1D: Identification of thunderstorm occurrence from convection-permitting ensemble forecasts using deep learning

Kianusch Vahid Yousefnia Christoph Metzl Tobias Bölle

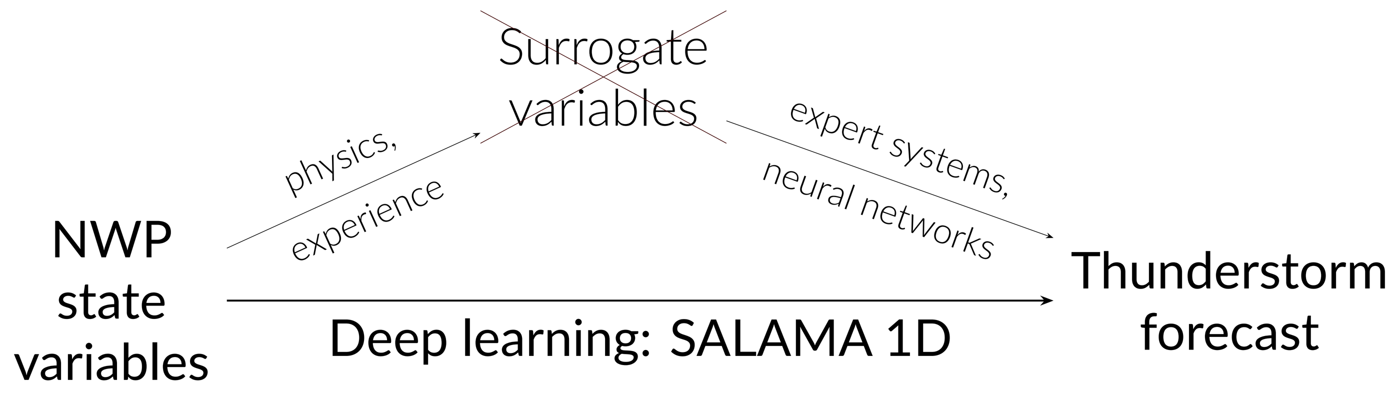
Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

## Motivation

Thunderstorm forecasts based on numerical weather prediction (NWP) often rely on single-level surrogate predictors (e.g. CAPE), derived from vertical profiles of more fundamental variables.

In this study, we show that our deep learning model SALAMA 1D, which infers thunderstorm occurrence from vertical NWP profiles,

- outperforms thunderstorm inference based on derived single-level features,
- allows for interpretability.




## Dataset compilation

We train our model on two summers (2021, 2022) of operational forecasts of ICON-D2-EPS, a convection-permitting NWP ensemble model for Central Europe with a horizontal resolution of $\sim 2$ km and 65 vertical levels. Ground-based lightning observations by the LINET network serve as the ground truth.

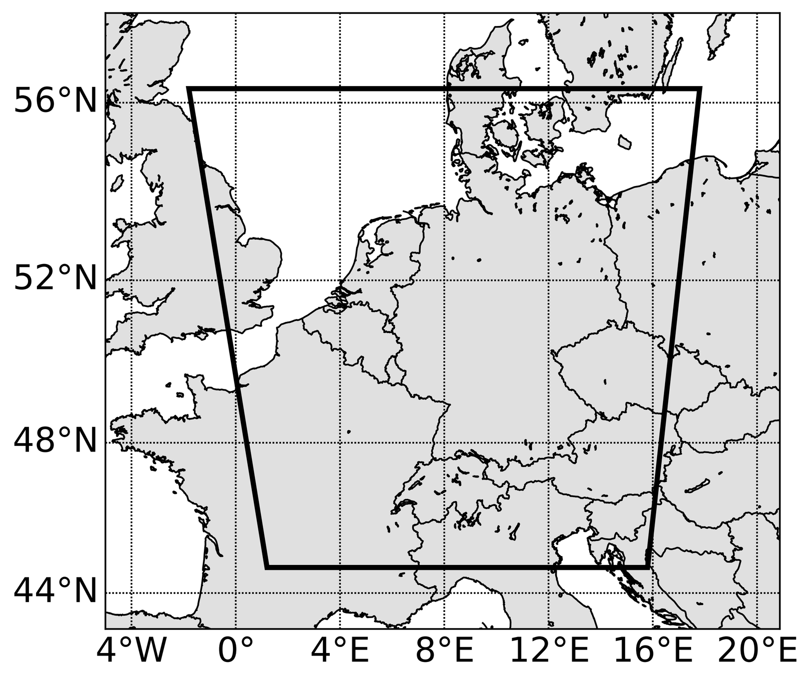



Figure 1. Study region.

| |
|---|
| Zonal wind speed (U) |
| Meridional wind speed (V) |
| Temperature (T) |
| Pressure (P) |
| Specific humidity (QV) |
| Cloud water mixing ratio (QC) |
| Cloud ice mixing ratio (QI) |
| Graupel mixing ratio (QG) |
| Cloud cover (CLC) |
| Vertical wind speed (W) |

Table 1. NWP variables.

## Physically-motivated architecture

A sparse layer reduces parameter size by encouraging interactions at similar height levels, while a shuffling layer prevents the model from learning patterns tied to the vertical NWP grid structure.

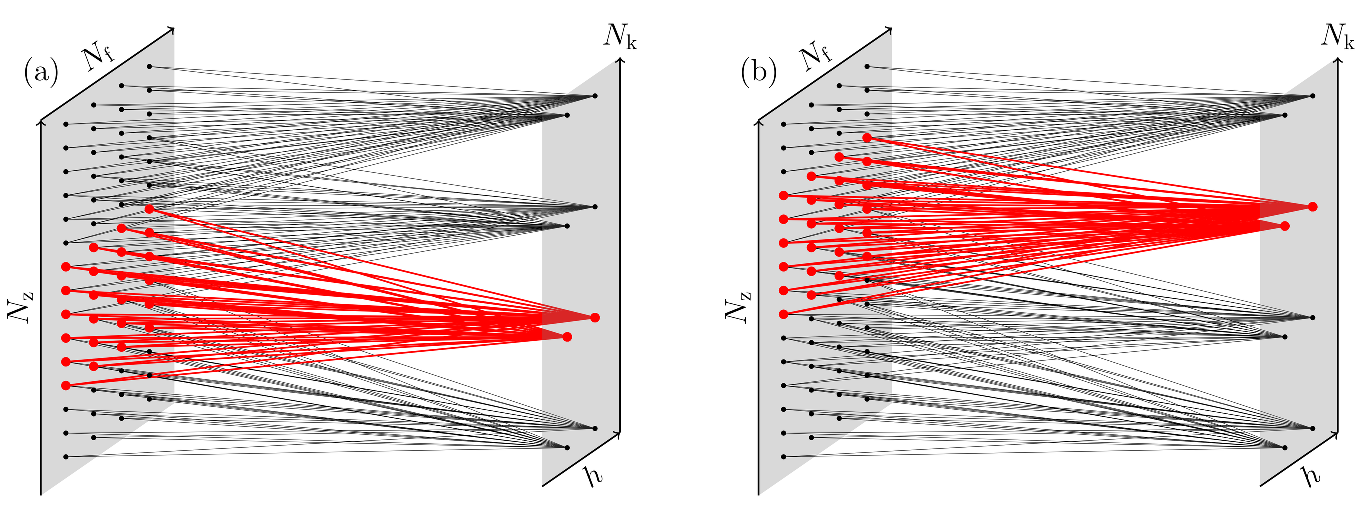
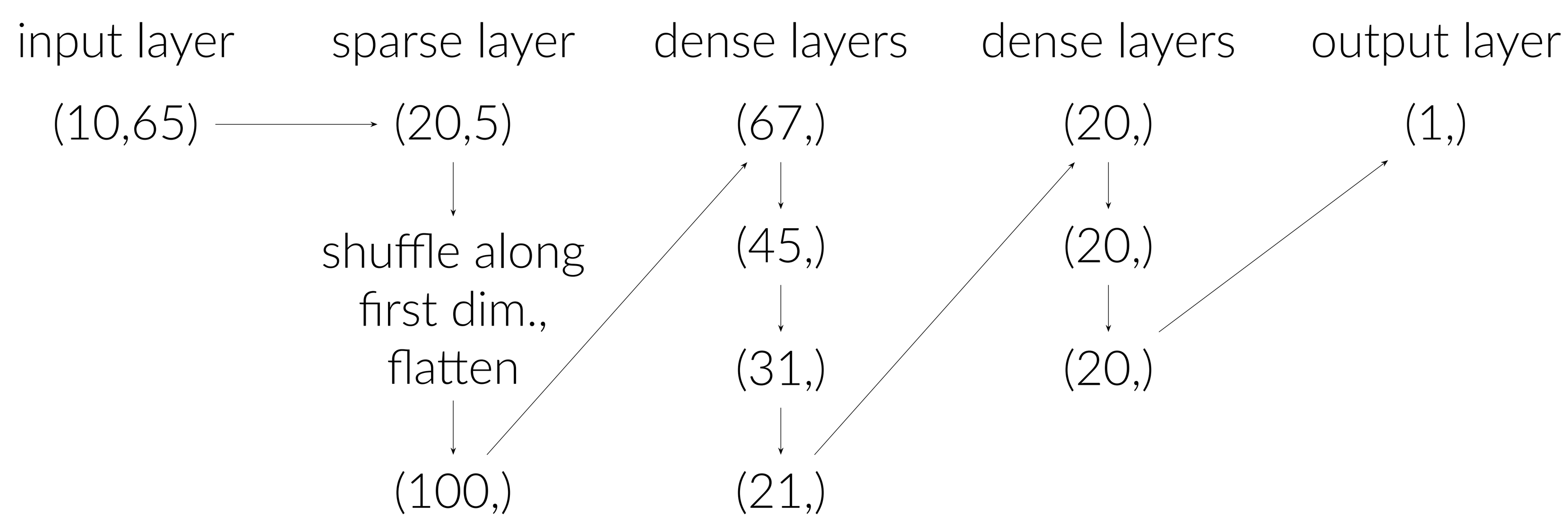



Figure 2. Illustration of the sparse connections between the input layer and the following layer.




## Higher skill when considering vertical profiles

We compare SALAMA 1D to SALAMA 0D, which is an ML model identifying thunderstorm occurrence from 21 traditional derived features, such as CAPE, model reflectivity, and relative humidity at 700 hPa.

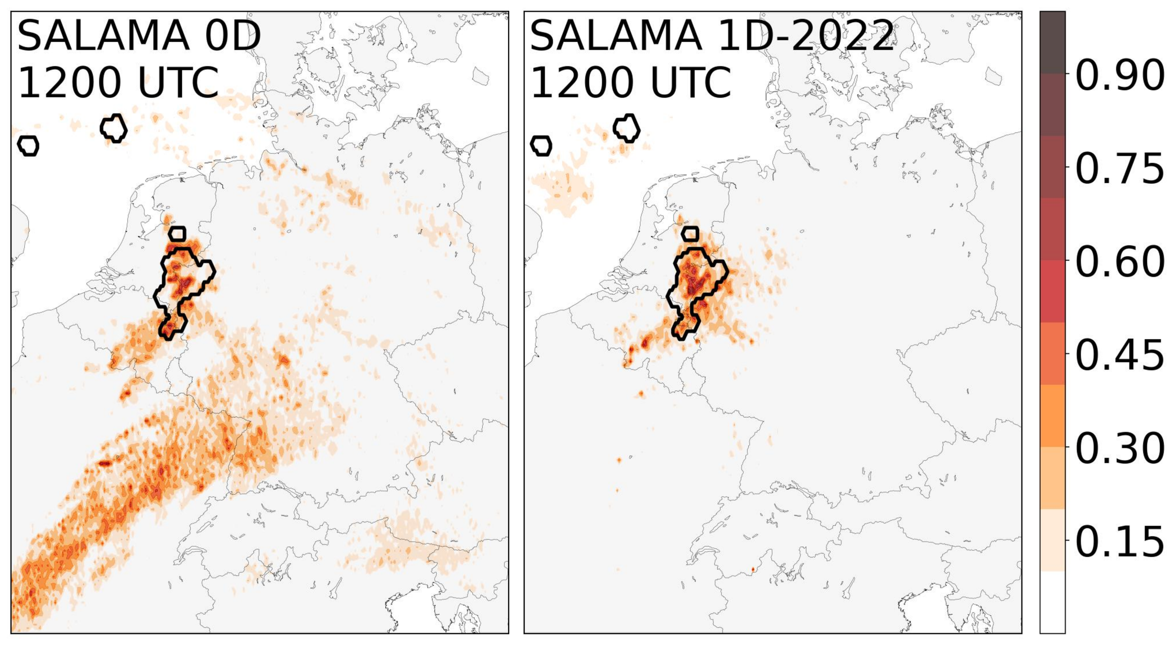



Figure 3. Model probability of thunderstorm occurrence for August 2, 2023, 1200 UTC. Lightning observations are shown as black contours.

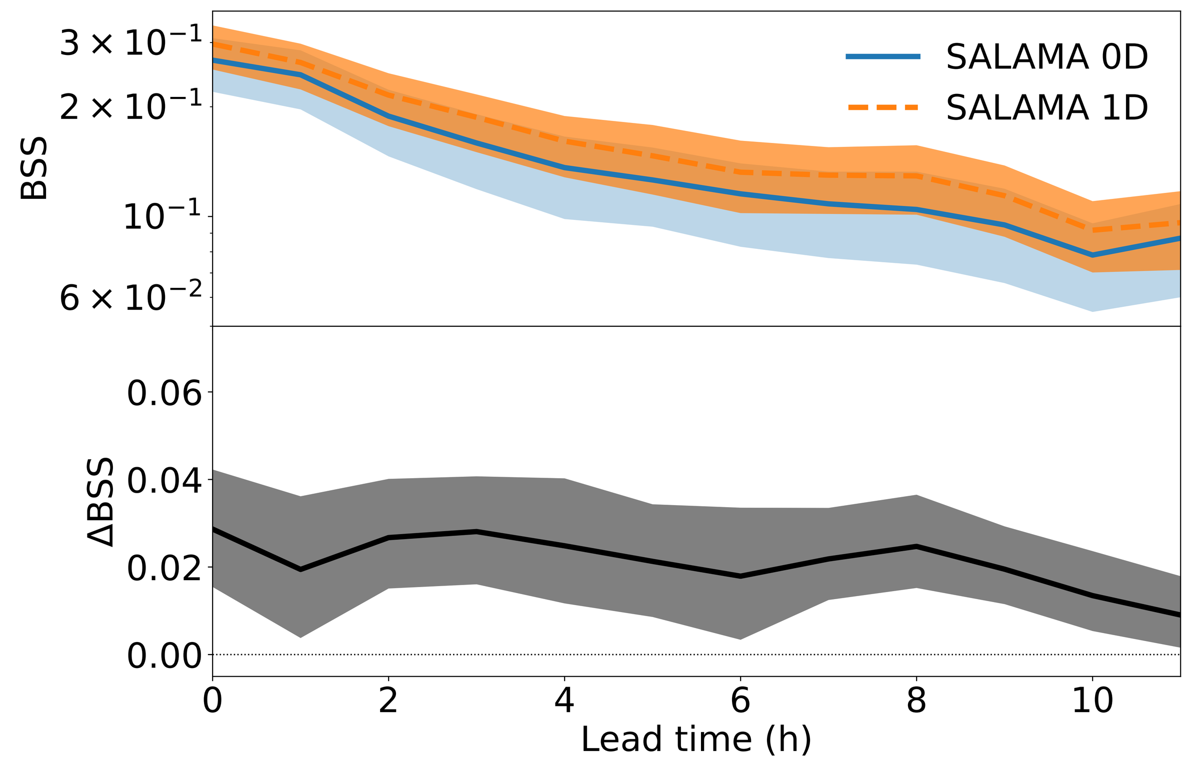



Figure 4. Lead time dependence of skill in terms of the Brier Skill Score (BSS) for the two ML models.

## Interpretable and consistent with physical understandings

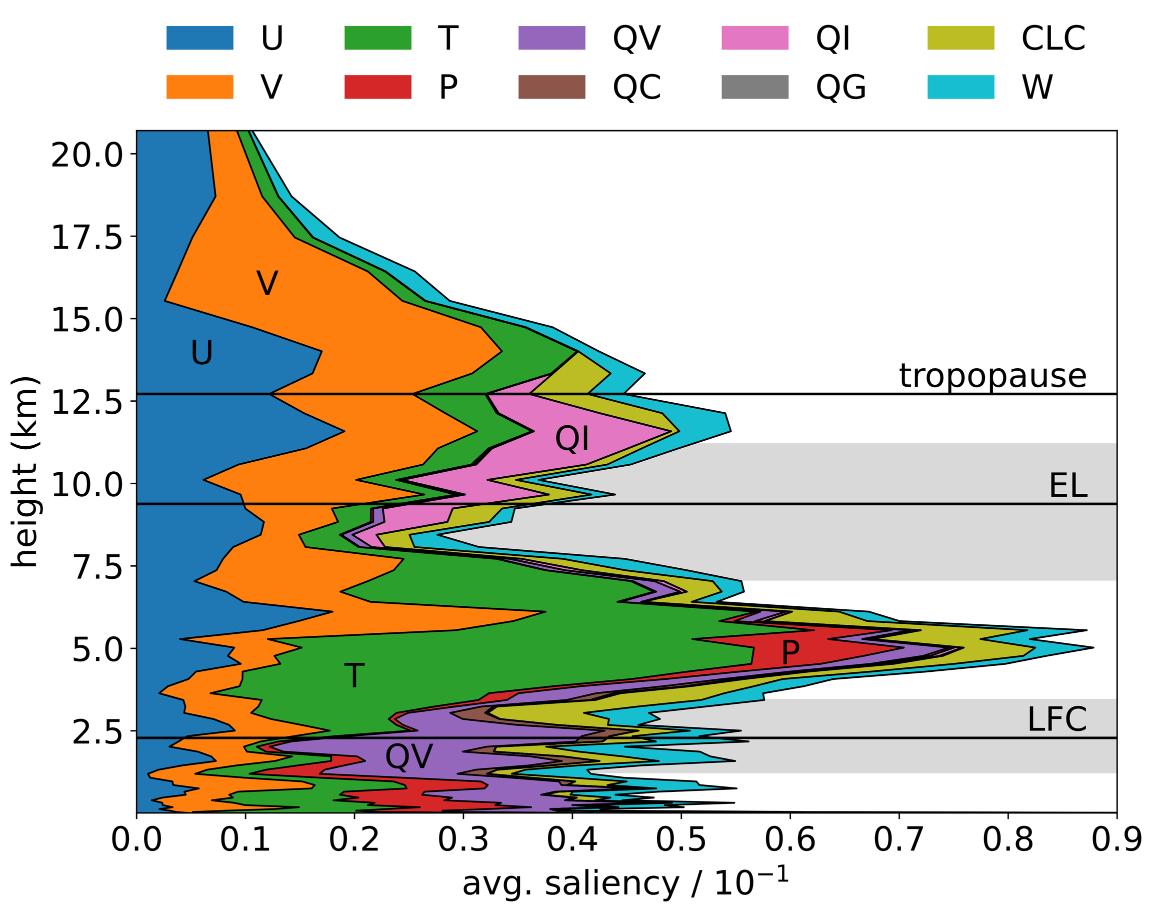



Figure 5. Average feature importance (saliency) for the top probability percentile of test set examples *(left)*, and corresponding average vertical profiles *(remaining panels)*. Saliencies for the different fields are stacked on top of each other in the order given in Table 1.

## More information

K. Vahid Yousefnia et al (2025): Inferring Thunderstorm Occurrence from Vertical Profiles of Convection-Permitting Simulations: Physical Insights from a Physical Deep Learning Model. *Artif. Intell. Earth Syst.*, 4, 240096, doi: 10.1175/AIES-D-24-0096.1

The code and data for the SALAMA models is available from https://github.com/kvahidyou/SALAMA

Deutsches Zentrum für Luft- und Raumfahrt
German Aerospace Center

*github.com/kvahidyou/SALAMA*
*kianusch.vahidyousefnia@dlr.de*