12th European Conference on Severe Storms, Utrecht, The Netherlands, 17-21. November 2025

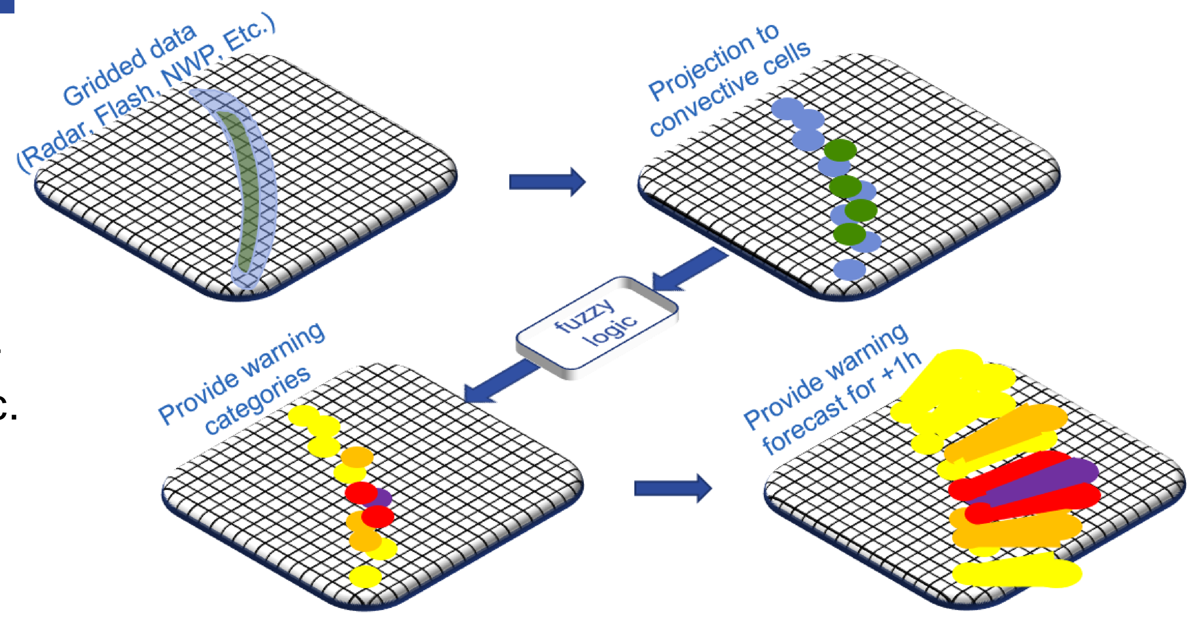
Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Towards a Machine-Learning Enhanced Nowcasting Tool for Storm Severity Analysis and Prediction

Gergely Bölöni, Paul James Michael Debertshäuser and Susanne Theis

NowCastMIX in a nutshell

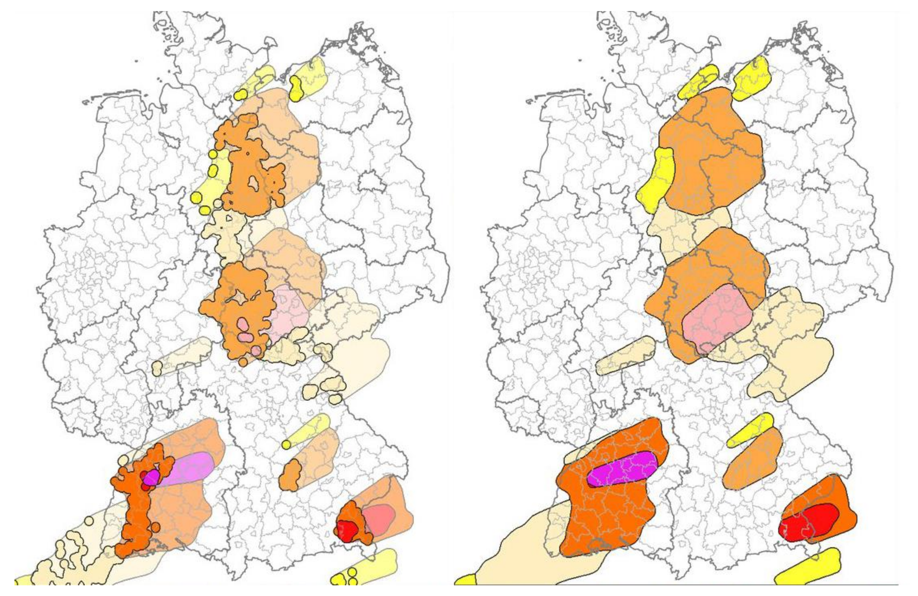
- Nowcasting system at the German Weather Service DWD
- Basis for severe weather warnings
- Input: remote sensing and in-situ measurements, NWP, etc.
- Output: warning polygons for the next hour
- Warning attribution based on fuzzy logic
- Full description: James et al., 2018, Wea. Forecasting



Schematic description of NowCastMIX: remote sensing data are used to detect and forecast convective cells including severe weather warning attributes using fuzzy logic

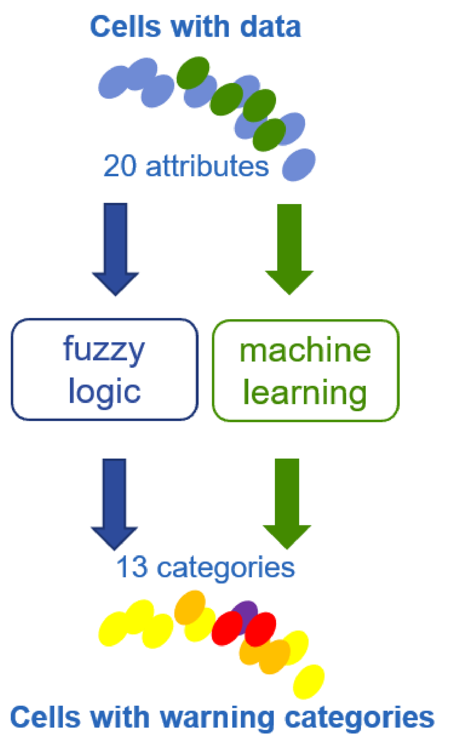
ii	Gusts Bft.	Rain mm/h	Hail
31	≤ 7	<15	
33	≤ 10	<15	
34	≤ 7	15-25	
36	≤ 10	15-25	
38	≤ 10	15-25	✓
40	≤ 12+	<15	
42	≤ 9	25-40	
46	≤ 10	25-40	✓
95	≤ 9	>40	✓
48	≤ 12+	25-40	✓
61	-	15-25	-
62	-	25-40	-
66	-	>40	-

Table of current warning categories issued by NowCastMIX



NowCastMIX in practice: forecasted warning polygons are smoothed in space and time. Warning polygons are directly editable by duty forecasters and can be projected to administrative borders.

Analysis step: fuzzy logic vs. machine learning



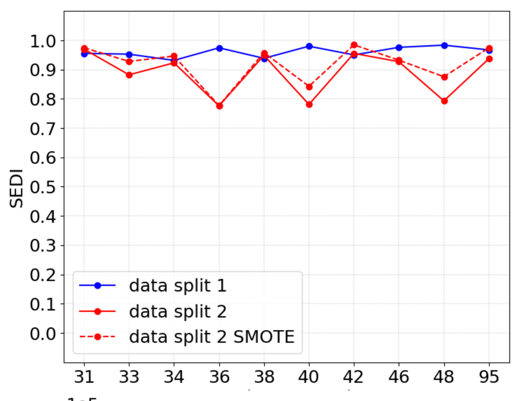
Fuzzy logic → (flexibility) → Machine Learning experiments

Fuzzy logic

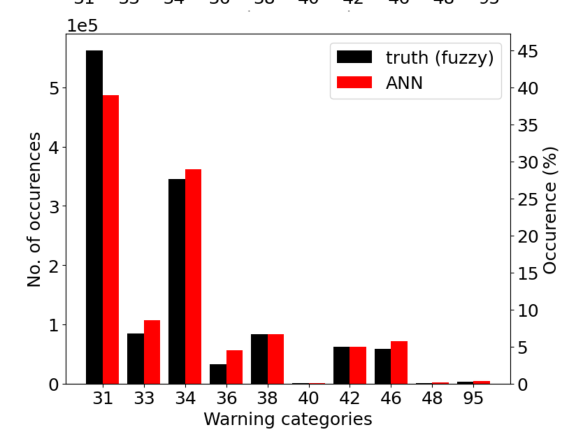
Works well but elaborate and costly tuning is needed if input data systematically change: radar & NWP upgrades, new obs., e.g. satellites, or new products.

Machine Learning experiments

- Data: $2*10^6$ instances of training set (seasons 2019, 2020), 10^6 instances of validation set (season 2021)
- Tool: scikit-learn (CPU only)
- Multi-class classification with ANN (MLP), decision trees, random forest, support vector machine, Gaussian process
- Verification: POD, FAR, SEDI, distribution of warning categories



Sensitivity of the warning category estimate to data splitting to training and validation set



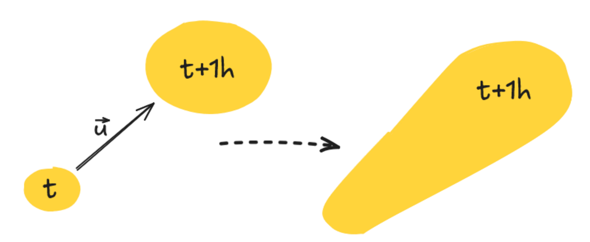
Histogram showing the distribution of warning categories with fuzzy logic and ANN

Conclusions

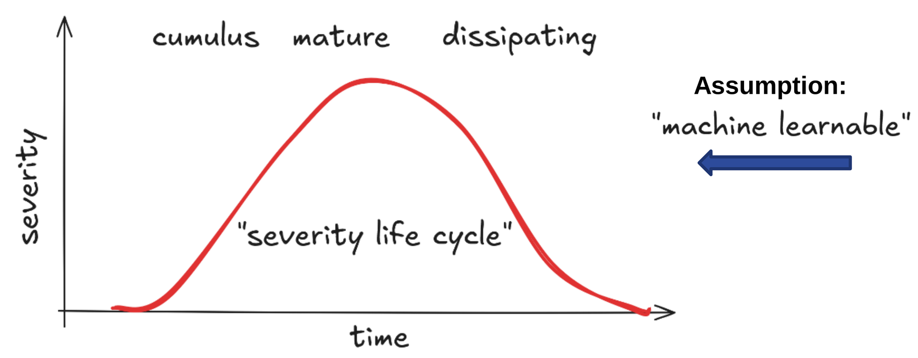
- ANN (MLP) works better than any other classification methods in scikit-learn
- Low sensitivity to ANN architecture and setup
- Results not robust ←→ huge data imbalance, i.e. $f_{cat40} = 10^5 * f_{cat31}$
- Down-sampling or up-sampling (SMOTE) do not help
- More data needed

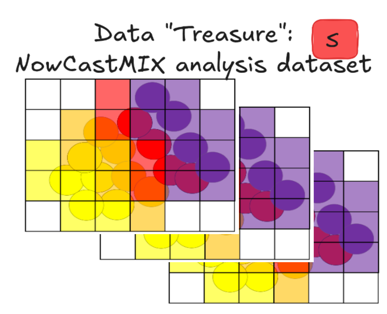
Prediction step: linear advection vs. machine learning

Prediction step as it is now →

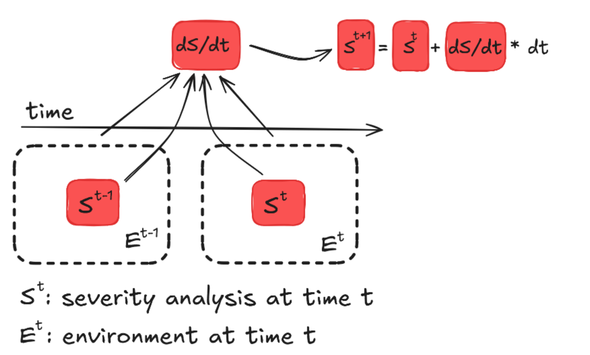


Prediction step with machine learning





Schematic of the ML-model



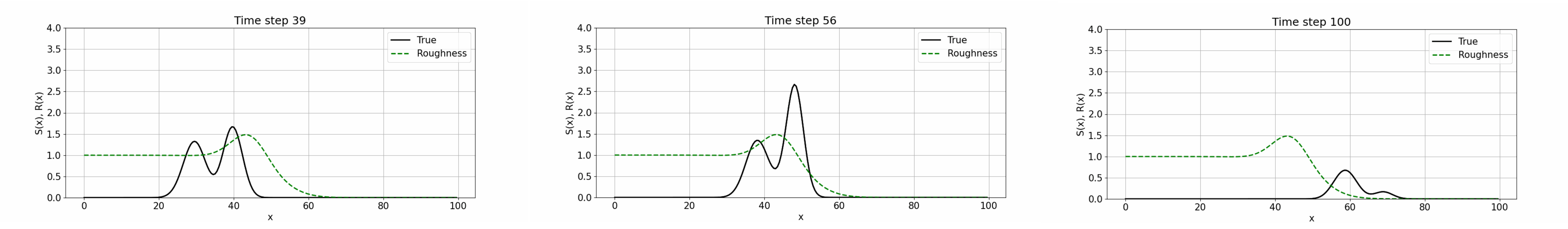
Synthetic idealized testbed

$$\frac{dS}{dt} = Q$$

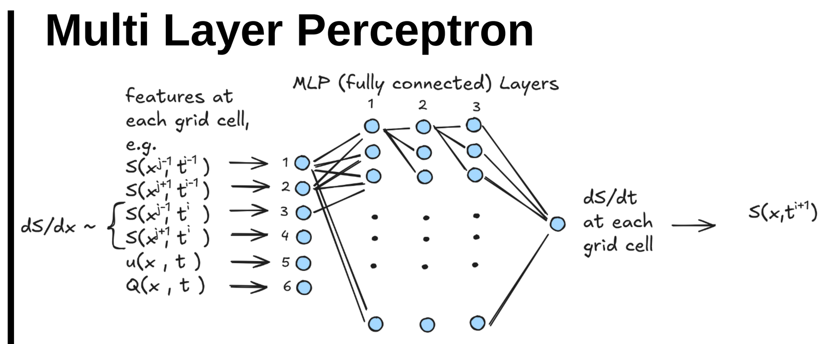
$$\frac{\partial S}{\partial t} = -\underbrace{u\frac{\partial S}{\partial x}}_{\text{numerical}} + \underbrace{Q}_{\text{ML}}$$

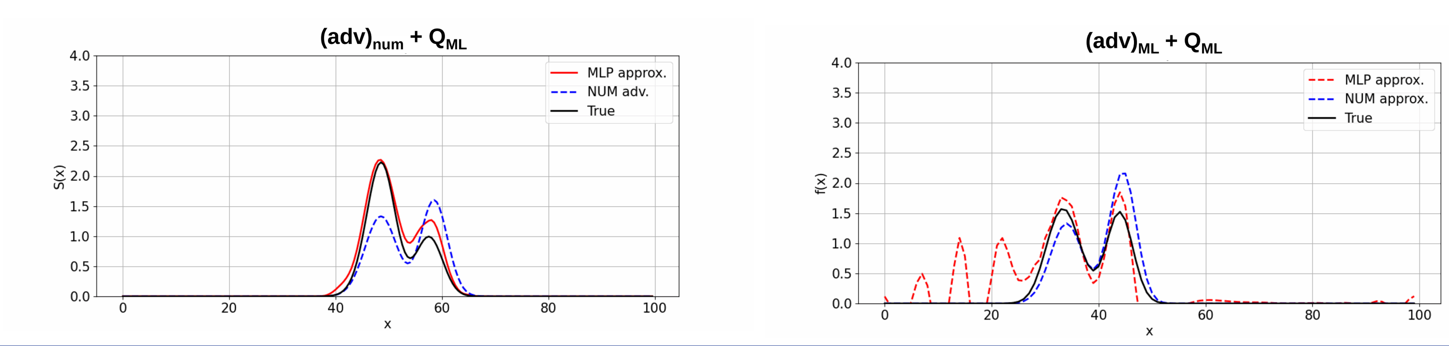
with forcing:

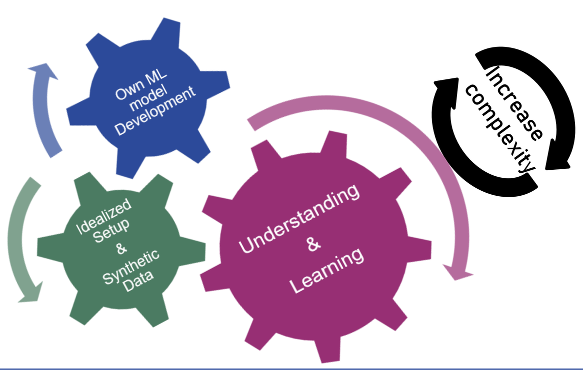
$$Q \rightarrow S(t) = S(t)_{adv} \cdot R$$



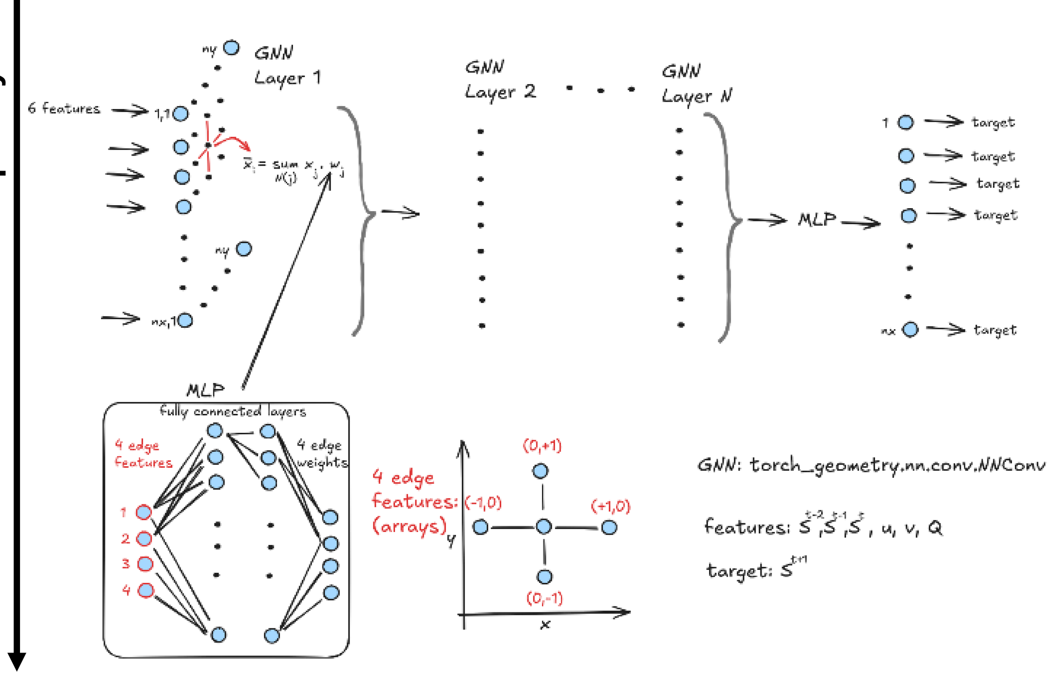
Multi Layer Perceptron

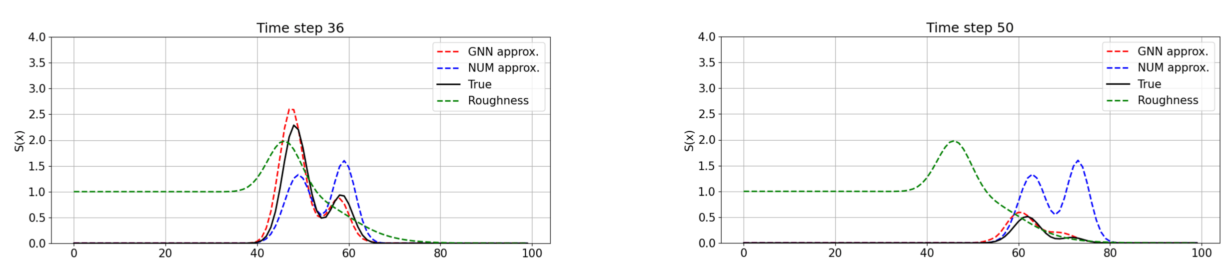


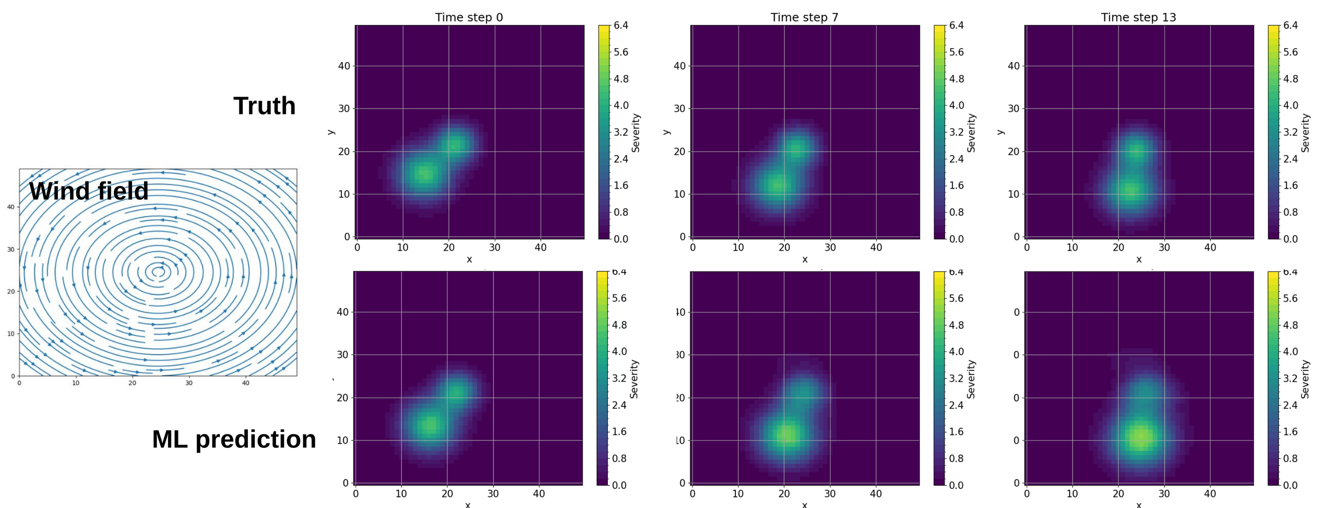




Graph Neural Network







Complexity

Measures to improve GNN for 2D

- Residual treatment: $x^{l+1} = x^l + GNN(x^l)$ with x^l being features at layer l
- Loss function: $loss = loss_{mse} + \alpha\, loss_{corr} + \beta\, loss_{grad}$ with the 2 last terms penalizing correlation- and gradient-differences wrt. truth
- Gating: $x^{l+1} = x^l + gate(x^l)\, GNN(x^l)$ with $gate(x)$ being a (MLP) learned function to weight updates by the GNN

DWD, FE22, Gergely Bölöni, Gergely.Boeloeni@dwd.de