Advancing extreme rainfall warnings in Belgium through radar QPE-driven pySTEPS-BE ensemble nowcasting

F. Erdmann[1], L. De Cruz[1,2], M. Reyniers[1], R. Reinoso-Rondinel[3,1], D. R. Poelman[1], M. Van Ginderachter[1], L. Bonnefoy[1]
[1] Royal Meteorological Institute of Belgium (RMI), Brussels, Belgium
[2] Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel, Brussels, Belgium
[3] Civil Engineering, Hydraulics & Geotechnics, KU Leuven, Leuven, Belgium

Objective: To validate an operational real-time municipal-scale alert system for extreme rainfall in Belgium based on ensemble nowcasts and radar QPE products

1. PySTEPS-BE ensemble nowcast

Based on STEPS [1, 2], built in the open-source framework pySTEPS [3]

Input:
- Observations: radar QPE fields, 1km resolution, 5min frequency
- NWP: ALARO+AROME at 1.3km, 5min accumulations

Output:
- **24-member** ensemble, every 15min (dev. 5min)
- Forecast time step of 5min for up to **+6 hours lead time**

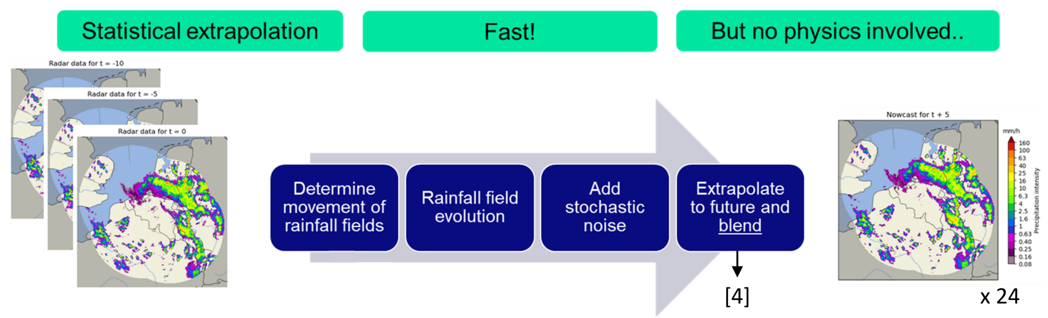

Fig. 1: PySTEPS methodology

2. Computation of warning thresholds

PyRainWarn = **py**STEPS-based **rain**fall **warn**ings

Warning thresholds [Fig. 2] derived from rainfall **return periods** using spatial Generalized Extreme Value (GEV) models based on long-term time series of rain gauge data [5] and from exceedance probabilities with

$$\text{Exceedance probability} = \frac{\text{n(members > threshold anywhere in the municipality)}}{\text{ensemble size}}$$

- Computed for accumulation times of 10min, 30min, 1h, and 2h
- Aggregation to municipal scale: highest level reached within borders
- Where needed, pyRainWarn **combines observed and predicted rainfall** to compute the accumulations.

Exceed. probability	5 years	20 years	50 years	100 years
100%	Moderate	Moderate	High	High
75%	Moderate	Medium	High	High
50%	Low	Moderate	Moderate	High
25%	Low	Low	Moderate	Moderate

No risk — Return period

Fig. 2: Warning threshold definitions

3. Validation

A: Generation of a validation dataset on municipal scale
- **Radar QPE** observations are compared to **rainfall thresholds** to create a base map of possible cases (**evidence level 1**), likely cases (**evidence level 2**), and confirmed cases (**evidence level 3**)
- Highest pixel level defines the municipality evidence level
- Gathering of **complementary data of confirmed impacts** to enhance the radar-based cases and **increase evidence levels**
  - Smartphone app + automated weather station reports of heavy rainfall (low impact) and flash floods (high impact)
  - Emergency intervention statistics (currently Brussels Region only)
  - Media reports of flash flood events in Belgium
  - Drone footage and other flood-indicating imagery (not yet)

B: Municipal scale validation
- Evaluation per observed evidence level (1, 2, 3) and per pyRainWarn forecast threshold (1: yellow, 2: orange, 3: red)
- Hits, misses, false alarms to define skill scores

C: Event-based validation
- **Event** = a period of consecutive time steps with exceeded threshold within one given municipality
- Pre-warn time statistics and event duration-based analysis

4. PyRainWarn output

- **Interactive map** [Fig. 3] to monitor the alerts
- Details as a popup window for each municipality
- **JSON** file list the municipality alerts to enable reuse in other applications.
- **Scan me!**

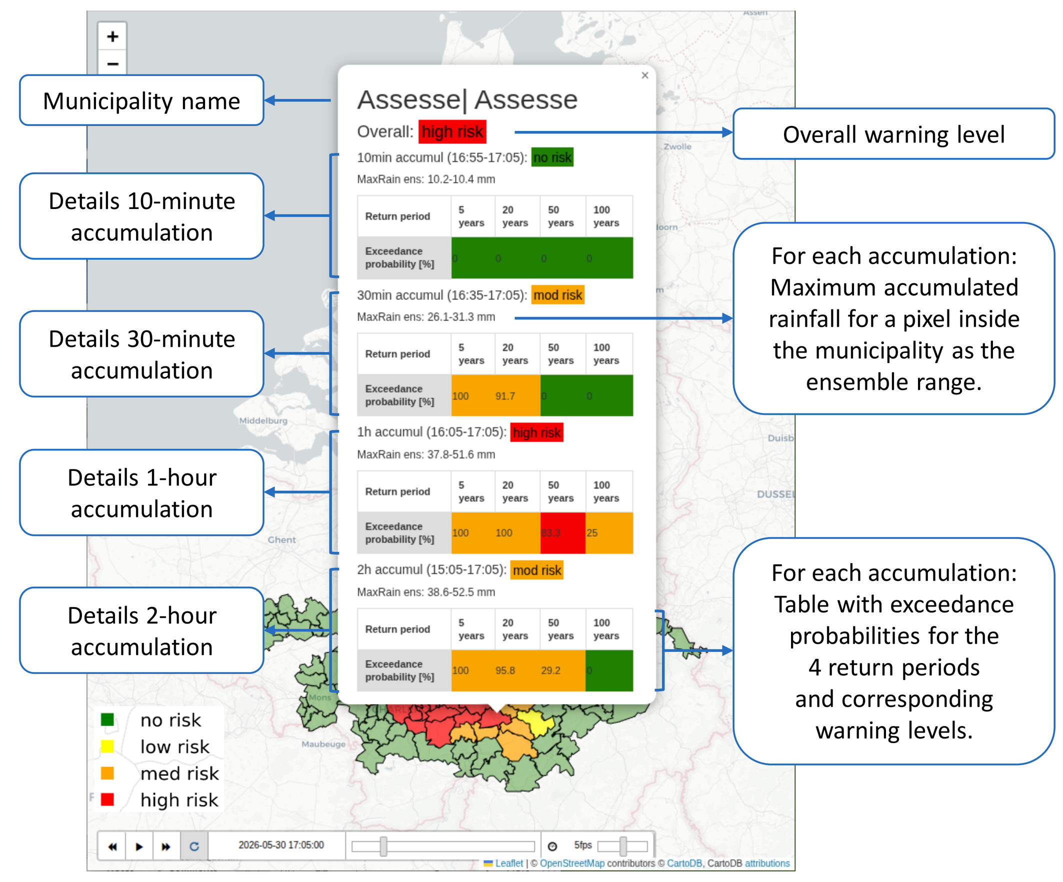

Fig. 3: PyRainWarn map with explanations

5. Preliminary results

Municipal scale: Scores (all cases, 43 runs)
- Probability of Detection (**POD**), False Alarm Ratio (**FAR**), Critical Success Index (**CSI**), and Frequency Bias Index (**FBI**) [Fig. 4, 5]
- Higher pyRainWarn warning thresholds have better FAR but worse POD
- High POD for municipalities with evidence level of 3 [Fig. 5]

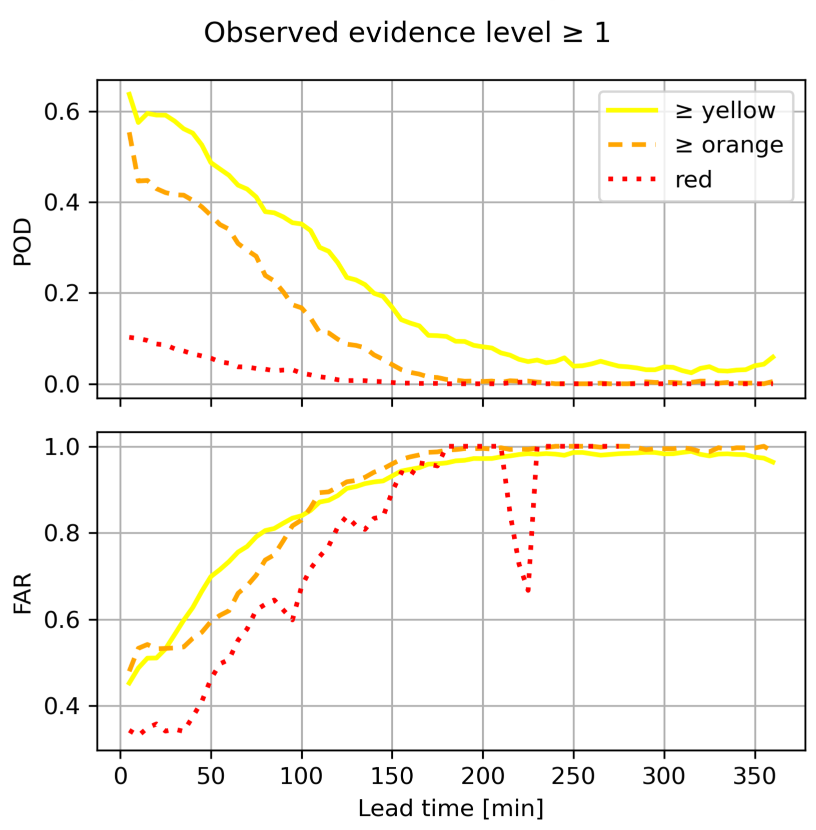

Fig. 4: POD and FAR as functions of lead time for observed evidence level ≥ 1 and the 3 forecast thresholds (colors)

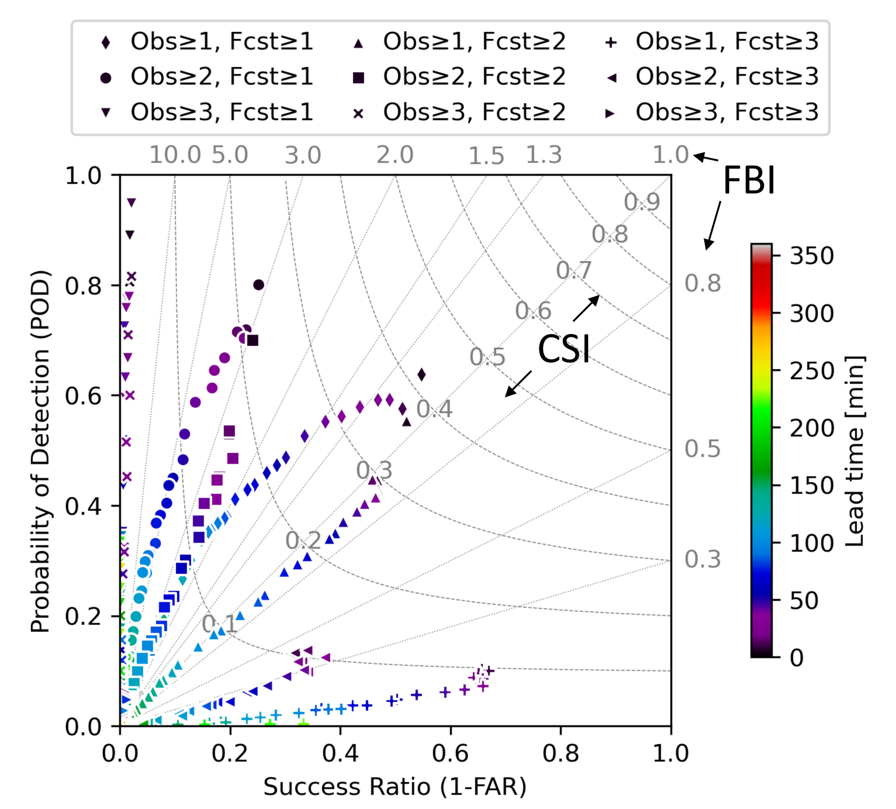

Fig. 5: Roebber performance diagram for pyRainWarn lead times (colors) per observed evidence level (Obs) and forecast threshold (Fcst)

Event-based: Pre-warn times (unique events, 43 runs)
- Events have a start and end time, and a duration (in the municipality).
- **Positive timing offset**: The event start time was correctly predicted before it started (in advance).
- **Negative timing offset**: The event was detected late, but before it ended (still a positive lead time but pyRainWarn missed the start).
- **Missed events**: Non of the event times received an active forecast
- Mean **pre-warn times of about 30 min**, even higher for events with evidence level 3 [Fig. 6c] that often had confirmed impact.

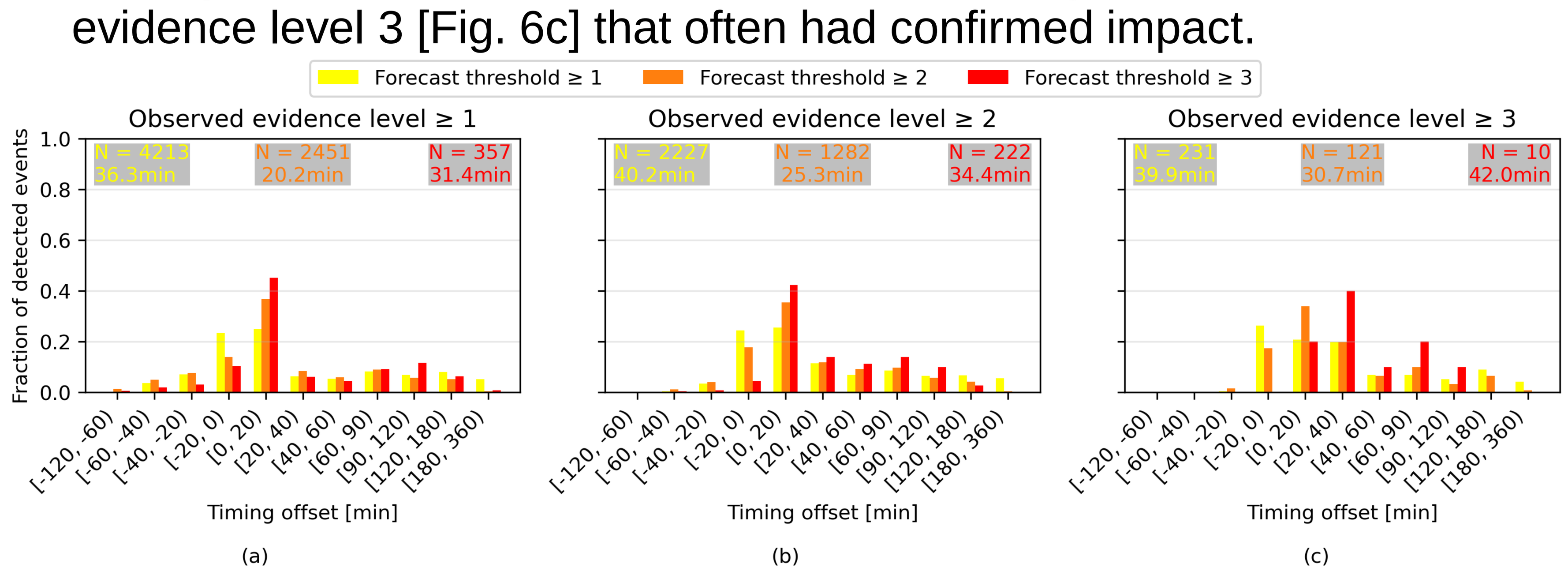

Fig. 6: Timing offset of detected events for observed evidence levels (a) ≥ 1, (b) ≥ 2, (c) ≥ 3 and the 3 pyRainWarn forecast thresholds (color). Number of events (N) and mean timing offset as colored numbers on top in each diagram.

Conclusion: First promising results that encourage to refine the validation approach, grow the validation dataset, and fine tune the pyRainWarn warning thresholds for operational use of the alerts.

References
[1] Bowler, N.E., Pierce, C.E. and Seed, A.W. (2006): https://doi.org/10.1256/qj.04.100
[2] Seed, A.W., Pierce, C.E., and Norman, K. (2013): https://doi.org/10.1002/wrcr.20536
[3] Pulkkinen, S. et al. (2019): https://doi.org/10.5194/gmd-12-4185-2019
[4] Imhoff, R.O. et al. (2023): https://doi.org/10.1002/qj.4461
[5] Van de Vyver, H. (2012): https://doi.org/10.1029/2011wr011707

Acknowledgements
The project is funded by the Service public de Wallonie (SPW) and also supported by BELSPO and Vlaams Brabant.
We thank all our colleagues who contributed to establishing pySTEPS-BE as the operational ensemble nowcasting system at the RMI and to run pyRainWarn as an operational product in real time. Thanks to Ruben Imhoff who designed Fig. 1.