Towards thunderstorm nowcasting by applying machine learning to a multi-sensor observation and NWP model database

Lea Beusch¹, Lorenzo Clementi¹, Loris Foresti¹, Ulrich Hamann¹, Alessandro M. Hering¹, Elena Leonarduzzi², Daniele Nerini¹, Luca Nisi¹, Marco Sassi¹, Urs Germann¹
¹Federal Office of Climatology and Meteorology MeteoSwiss, Via ai Monti 146, 6605 Locarno-Monti, Switzerland
²Institute of Environmental Engineering, ETH Zurich, Zurich, Switzerland

Motivation

Severe convective systems can lead to hail, heavy rain, gale-force wind gusts, lightning, and flash floods. Hence, forecasting them accurately is crucial. But numerical weather prediction models face difficulties in predicting the exact position and strength of these systems. As a consequence, thunderstorm nowcasting in the first few hours is mostly based on current observations. Today’s nowcasting systems primarily rely on a series of threshold tests. We plan to additionally exploit the potential of machine learning techniques to automatically extract information on the typical development of thunderstorms from a multi-sensor database and thus further improve the very short-term forecasts.

Satellite-based rainfall retrieval

Methodology

- Data: June - August 2017: Training (800 time slots), validation (400), and test set (400)
- Input Features:
  - MSG SEVIRI IR channels + differences
  - NWCSAF products
  - Local solar time, coordinates, topography, land-sea mask, satellite viewing geometry
- Model:
  - Precip detection: logistic regression (logreg), SVM
  - Rain rate: linear regression (+ probability matching)

Outlook

- Test potential of Artificial Neural Networks (+ possibly other machine learning algorithms) for rainfall retrieval
- Create thunderstorm data set in Lagrangian coordinates with variables used so far + include radar, NWP, and lightning information
- Employ machine learning methods to nowcast thunderstorm evolution
- Develop real-time, multi-sensor, seamless, quantitative, end-to-end, localized, robust, and customer-oriented products

1. Towards thunderstorm nowcasting by applying machine learning to a multi-sensor observation and NWP model database

Motivation

Severe convective systems can lead to hail, heavy rain, gale-force wind gusts, lightning, and flash floods. Hence, forecasting them accurately is crucial. But numerical weather prediction models face difficulties in predicting the exact position and strength of these systems. As a consequence, thunderstorm nowcasting in the first few hours is mostly based on current observations. Today’s nowcasting systems primarily rely on a series of threshold tests. We plan to additionally exploit the potential of machine learning techniques to automatically extract information on the typical development of thunderstorms from a multi-sensor database and thus further improve the very short-term forecasts.

Satellite-based rainfall retrieval

Methodology

- Data: June - August 2017: Training (800 time slots), validation (400), and test set (400)
- Input Features:
  - MSG SEVIRI IR channels + differences
  - NWCSAF products
  - Local solar time, coordinates, topography, land-sea mask, satellite viewing geometry
- Model:
  - Precip detection: logistic regression (logreg), SVM
  - Rain rate: linear regression (+ probability matching)

Case study: 12 UTC 09 July 2017

- Precip detection: logreg: hardly any misses, many false alarms
  - HSADF: more misses, fewer false alarms
- Rain rate: Linreg: mostly predict low values close to mean
  - HSADF: overestimation over large areas
  - Both: fine-scale structure of radar is not captured

Verification

- Scores for the test set (ts) and the case study above (ca): Left: precip detection, circles = 0.8 mio random cloudy pixels (balanced) scores; Right: rain rate retrieval, circles = 0.4 mio random rainy pixels scores; Both: crosses = case study scores, stars = optimal scores, black horizontal line = 0 skill line
- High case-to-case variability in performance of the algorithm
- Many scores strongly dependent on test set distribution -> which one to optimize?

Conclusions

- Precip detection:
  - Very satisfactory results with logistic regression
  - SVM slightly inferior performance + much larger computational time (not shown)
- Rain rate:
  - Difficult task
  - Ground truth from different instrument
- Instantaneous rain rates highly skewed
- Predictions close to mean favored: prob. matching?

Thunderstorm nowcasting

- Test potential of Artificial Neural Networks (+ possibly other machine learning algorithms) for rainfall retrieval

Outlook

- Create thunderstorm data set in Lagrangian coordinates with variables used so far + include radar, NWP, and lightning information
- Employ machine learning methods to nowcast thunderstorm evolution
- Develop real-time, multi-sensor, seamless, quantitative, end-to-end, localized, robust, and customer-oriented products