Sub-seasonal daily precipitation forecasting based on Long Short-Term Memory (LSTM) models

Claudia Bertini[1], Gerald Corzo[1], Schalk Jan van Andel[1], Dimitri Solomatine[1]
[1] IHE Delft Institute for Water Education, Westvest 7, P.O. Box 3015, 2601 DA Delft, Netherlands

Introduction

Water managers need accurate rainfall forecasts for a wide spectrum of applications, ranging from water resources evaluation and allocation, to flood and drought predictions. In the past years, several frameworks based on Artificial Intelligence have been developed to improve the traditional Numerical Weather Prediction (NWP) forecasts, thanks to their ability of learning from past data, unravelling hidden relationships among variables and handle large amounts of inputs. Among these approaches, Long Short-Term Memory (LSTM) models emerged for their ability to predict sequence data, and have been successfully used for rainfall [1] and flow forecasting [2], mainly with short lead-times.

Problem Description & Research Objective

- Local water managers need reliable forecasts of daily precipitation to monitor precipitation deficit and forecast meteorological droughts.
- According to local end-users, currently precipitation forecasts are reliable up to a lead time of 15 days.
- Accurate daily precipitation forecasts with a lead time up to 30 days are needed to plan drought mitigation interventions.

This research aims to explore the use of different LSTM models to predict daily precipitation for the upcoming 30 days, using both local atmospheric and global climate data.

Case study

Fig. 1. : The study area of Rijnland (green) within the Netherlands

The area of Rijnland, the Netherlands, is located at the end of the Rhine delta. Recently, it has been affected by summer droughts, which have been occurring more frequently in the past years. From meteorological perspective, summer droughts are characterised by high evapotranspiration rates, that often exceed the amount of rainfall. The local water authority, i.e. Rijnland Water Board, monitors the cumulative precipitation deficit (difference between precipitation and evapotranspiration) to monitor drought evolution and plan mitigation interventions.

Methodology

Input variables selection

Input variables pre-selection: local climate, literature review [3,4]

→ Data analysis and preparation: Cross-correlation analysis, climatology, Principal Component Analysis (PCA) for global variables

Local atmospheric (ERA 5):
- Precipitation (P)
- Relative Humidity (RH)
- Total Cloud Cover (TCC)
- U and V Wind (UW, UV)
- Temperature (T)

Global climate (ERA 5):
- Mean Sea Level Pressure (MSLP)
- Geopotential Height 500 hPa (Z500)

LSTM model experimental set-up

- Vanilla: LSTM
- Stacked: $LSTM_1$, $LSTM_2$
- Bidirectional: LSTM

→ Univariate: 1 Day, 30 Days
→ Multivariate: 1 Day, 30 Days

Performance evaluation

- Mean Absolute Error (MAE) between the N observations y_i and N predictions $\hat{y}_i$ (Eq. 1)

$$MAE = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N} \quad (1)$$

- Comparison against climatology.

Results

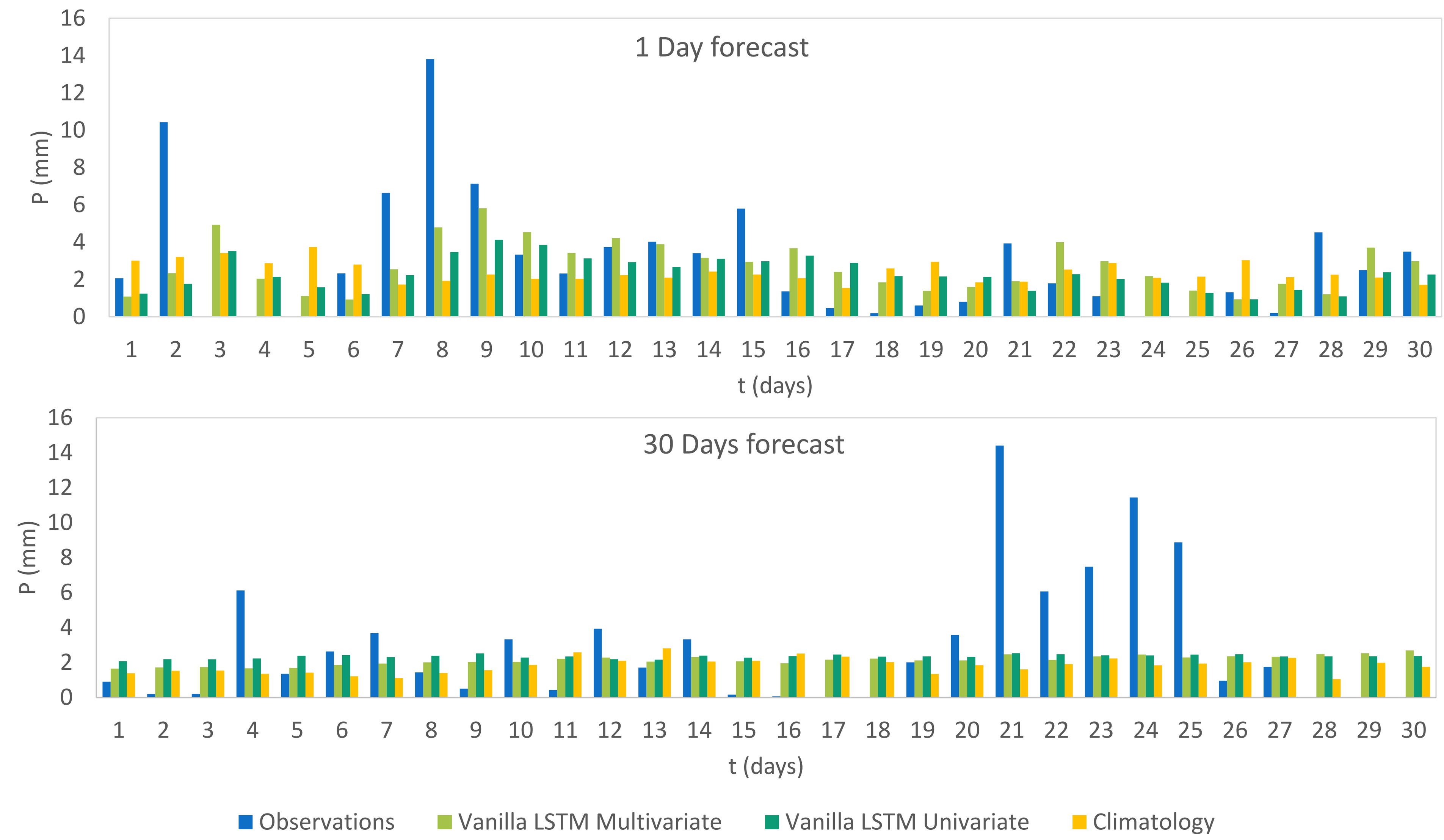

Fig. 2. : a) Results of 1 day forecasts obtained with Vanilla LSTM (univariate and multivariate), plotted against climatology and observations b)) Results of 30 days forecasts obtained with Vanilla LSTM (univariate and multivariate), plotted against climatology and observations

- For 1 day forecasting, LSTM models perform slightly better than climatology (MAE = 2.06 mm for multivariate LSTM, MAE = 2.5 mm climatology), but do not capture the extremes.
- For 30 days forecasting, all the LSTM models perform similarly to climatology. MAE is overall low (<1 mm), but LSTM models predict always values around the average daily precipitation (Fig. 2 b)
- Vanilla LSTM model performs overall better than the other LSTM models, with better results for the multivariate cases.

Conclusions

- LSTM models could not accurately predict daily precipitation, despite the lead-time and model architecture selected. Better results are shown for predicting one day rainfall, but peaks are not well captured, but still not satisfactory.
- The possibility of including existing precipitation forecasts in the input variables will be explored in the future, using LSTM to post-process existing forecasts rather than generating ne predictions.

References

[1] Zhu, K., Yang, Q., Zhang, S., Jiang, S., Wang, T., Liu, J., & Ye, Y. (2023). Long lead-time radar rainfall nowcasting method incorporating atmospheric conditions using long short-term memory networks. Frontiers in Environmental Science.

[2] Cheng, M., Fang, F., Kinouchi, T., Navon, I. M., & Pain, C. C. (2020). Long lead-time daily and monthly streamflow forecasting using machine learning methods. Journal of Hydrology, 590, 125376.

[3] Felsche, E., & Ludwig, R. (2021). Applying machine learning for drought prediction using data from a large ensemble of climate simulations. Natural Hazards and Earth System Sciences Discussions, 2021, 1-20.

[4] Bosso, F., Bertini, C., Giuliani, M., Solomatine, D., and van Andel, S. J.: Improving sub-seasonal drought forecasting via machine learning to leverage climate data at different spatial scales, EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023

Contact: c.bertini@un-ihe.org