
Extreme Precipitation Seasonal Forecast Using a Transformer Neural Network

Daniel Salles Civitarese ^{*1} Daniela Szwarcman ^{*1} Bianca Zadrozny ¹ Campbell Watson ¹

Abstract

An impact of climate change is the increase in frequency and intensity of extreme precipitation events. However, confidently predicting the likelihood of extreme precipitation at seasonal scales remains an outstanding challenge. Here, we present an approach to forecasting the quantiles of the maximum daily precipitation in each week up to six months ahead using the temporal fusion transformer (TFT) model. Through experiments in two regions, we compare TFT predictions with those of two baselines: climatology and a calibrated ECMWF SEAS5 ensemble forecast (S5). Our results show that, in terms of quantile risk at six month lead time, the TFT predictions significantly outperform those from S5 and show an overall small improvement compared to climatology. The TFT also responds positively to departures from normal that climatology cannot.

1. Introduction

Extreme weather and climate events can severely impact communities, infrastructure, agriculture and other natural and human systems. The observed increase in frequency and intensity of extreme events has been directly linked to anthropogenic climate change, and the trend is expected to continue in the coming years (Herring et al., 2021). Improving our understanding of the impacts of climate change on weather and climate extremes and developing early warning systems for better preparedness is of paramount importance.

This short paper is concerned with extreme precipitation prediction, which can cause flooding, crop damage, and widespread disruption to ecosystems. The ability to confidently predict the likelihood of extreme precipitation at sub-seasonal to seasonal scales remains a significant challenge (King et al., 2020). Recent works have shown that statistical machine learning (ML) models can outperform state-of-the-art dynamical models when predicting extreme

events several months ahead (e.g., Chantry et al. (2021), Cohen et al. (2019)). These ML models tend to rely on slowly-changing variables, such as soil moisture and El Niño-Southern Oscillation (ENSO) indices (Fernando et al., 2019). Although most of these variables are publicly available, the degree of influence of each one regarding precipitation prediction varies in space and time (e.g., Strazzo et al. (2019)). Properly accounting for these effects in a ML model is a complex task.

In this work, we present an ML approach to forecast the *maximum daily precipitation in each week* up to six months ahead. Our approach uses the temporal fusion transformer (TFT) model, which combines multi-horizon forecasting with specialized components to select relevant inputs and suppress unnecessary features (Lim et al., 2019). Of particular interest to both researchers and end-users, the TFT produces multiple quantiles. These quantiles offer a global view of the interrelations between the input and output (Davino et al., 2013) and can help risk management by indicating the likely best and worst-case values of the target variable (Lim et al., 2019). To the Authors' knowledge, this is the first time the TFT has been used to forecast extreme weather.

2. Methodology

We define our task as a quantile forecast for precipitation in multiple locations. The structure is similar to the one presented in (Lim et al., 2019), where the authors consider three kinds of input: (a) static $s \in \mathbb{R}^{m_s}$, (b) known future $x \in \mathbb{R}^{m_x}$, and (c) historical $z \in \mathbb{R}^{m_z}$ (Figure 1 – yellow, pink, and blue, respectively). Static information does not change through time, e.g., location or altitude. Examples of known future information are the month-of-the-year and external predictions for the target and other variables such as precipitation and temperature. Although predictions are not “known information” and may have errors, their use follows the same procedure as future inputs. Finally, historical inputs are the actual past values for the target series or any other exogenous one, e.g., temperature or soil moisture.

We produce weekly forecasts for each location in the form $\hat{y}_{lat,lon}(q, t, \tau)$, where each time-step $t \in [0, T]$ refers to a specific week in the time series, and $\tau \in [1, 26]$ is the lead time in weeks. The index $(lat, lon) = i$ refers to the location in the globe, and it is associated with a set of static

^{*}Equal contribution ¹IBM Research, Rio de Janeiro, Brazil. Correspondence to: Daniel Salles Civitarese <sallesd@br.ibm.com>.

inputs s_i , as well as the time-dependent inputs $z_{i,t}$ and $x_{i,t}$.

We predict the quantiles 0.1, 0.5, and 0.9 of the maximum daily rainfall for each week up to six months ahead in our setup. Each quantile is represented by $\hat{y}_i(q, t, \tau) = f_q(\tau, y_{i,t-k:t}, z_{i,t-k:t}, x_{i,t-k:t+\tau}, s_i)$. Although all quantiles refer to the maximum daily rainfall in a week, we pay closer attention to quantile 0.9 to focus on extreme events.

2.1. TFT model

The primary motivation of using the TFT model for our task is its ability to handle multiple types of variables, i.e., static, historical, and future. The model also provides other features, such as simultaneous outputs for all τ time steps and a modified multi-head attention mechanism that facilitates the interpretation of the outputs. We intend to use the latter in future work.

Figure 1 shows a summarized version of the TFT structure. Level L1 creates input embeddings from both categorical and continuous values. In Section 3 we describe all the input variables used in our experiments. Level L2 processes each input type individually and selects relevant input variables at each step. Static encoders (yellow) integrate time-invariant features into the network to condition temporal dynamics. The sequence-to-sequence LSTM module learns short-term temporal relationships, whereas the multi-head attention block on level L4 captures long-term dependencies.

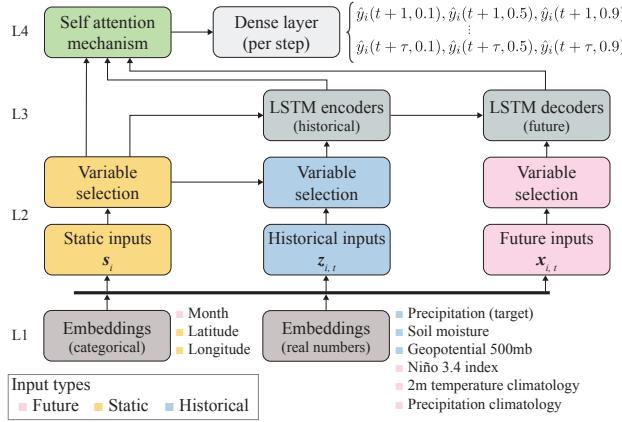


Figure 1. Overview of the TFT architecture. It shows the types of inputs: static (yellow), historical (blue), and future (pink). Main components are: variable selection, LSTM encoders and decoders, and the self-attention mechanism.

As in (Lim et al., 2019), we train the TFT model by jointly minimizing the quantile loss, summed across all quantile outputs:

$$\mathcal{L}(\Omega, \mathbf{W}) = \sum_{y_t \in \Omega} \sum_{q \in \mathcal{Q}} \sum_{\tau=1}^{\tau_{max}} \frac{QL(y_t, \hat{y}(q, t - \tau, \tau), q)}{M\tau_{max}} \quad (1)$$

$$QL(y, \hat{y}, q) = q(y - \hat{y})_+ + (1 - q)(\hat{y} - y)_+ \quad (2)$$

where Ω is the domain of the training data, which contains M samples, \mathbf{W} refers to the TFT weights, \mathcal{Q} is the set of output quantiles, and $(.)_+ = \max(0, .)$.

The normalized quantile loss (q-risk) is used for validation and testing:

$$\text{q-risk} = \frac{2 \sum_{y_t \in \tilde{\Omega}} \sum_{\tau=1}^{\tau_{max}} QL(y_t, \hat{y}(q, t - \tau, \tau), q)}{\sum_{y_t \in \tilde{\Omega}} \sum_{\tau=1}^{\tau_{max}} |t_t|} \quad (3)$$

where $\tilde{\Omega}$ is the domain of the validation or test samples.

2.2. Baseline models

Climatology. Climatology is a simple model that is commonly used as a baseline for seasonal forecast skill. Here, it is computed from historical values (1981-2010) of the variable of interest and depends on the selected temporal aggregation, e.g., monthly mean. We follow the steps in Algorithm 1 to compute the climatology quantiles for precipitation ($agg = max$, $op = quantiles$), creating predictions for each week of the year (1 to 53) for any given year.

Algorithm 1 Climatology - week

```

Input: data (1981-2010)  $x$ , aggreg.  $agg$ , operation  $op$ 
 $agg \in [max, mean, min, \dots]$ 
 $op \in [mean, quantiles, max, \dots]$ 
 $week\_x \leftarrow \text{Aggregate } x \text{ weekly, using the } agg \text{ function}$ 
Initialize array  $climo$ , size = 53
for  $i = 1$  to 53 do
     $week\_values \leftarrow \text{values for all weeks } i \text{ in } week\_x$ 
     $climo[i] \leftarrow op(week\_values)$ 
end for

```

Calibrated ECMWF SEAS5. We also compare our predictions to those from a calibrated ECMWF SEAS5 (hereafter S5) seasonal forecast 50-member ensemble (Anon, 2019). To compute the forecasted quantiles from the S5 ensemble, we first aggregate the data to obtain the weekly maximum precipitation and then calculate the quantiles across the 50 members for each week of interest.

3. Experiments

3.1. Data

We selected two regions to investigate: Rio de Janeiro (Brazil) and Florida (USA). Besides precipitation, we consider as inputs to the network: 2 m temperature, soil moisture¹, and geopotential height at 500 mbar (all from the

¹volumetric soil water from 0 to 7 cm of depth

location of the forecasts). This choice was based on past work that used these variables in the context of rainfall prediction (Fernando et al., 2019; Xu et al., 2020).

The literature also indicates that precipitation is considerably affected by ENSO in Rio and Florida (Arguez et al., 2019; Grimm & Tedeschi, 2009). To account for ENSO, we consider the Niño 3.4 index prediction as a future input. Currently, we are using Niño 3.4 index observations² as if they were predictions to train the networks, and we plan to replace these with predictions from the S5 ensemble.

The data selected for observed precipitation was CHIRPSv2, (Funk et al., 2015), a daily gridded dataset with a spatial resolution of 0.05° . For the other climate variables, we used ERA5 reanalysis data for surface (Hersbach et al., 2018b) and pressure (Hersbach et al., 2018a) variables. ERA5 is also a gridded dataset but with resolution of 0.25° . We performed a spatial max-pooling operation to convert the CHIRPSv2 data to the resolution of ERA5 (5×5 window), which guarantees that the highest precipitation values for each 5×5 window will not be smoothed out.

Precipitation, geopotential, and soil moisture are provided as historical inputs to the network: 26 weeks of past values. We consider the weekly mean for soil moisture and geopotential. For temperature, we compute the weekly mean climatology and provide it as a prediction. The precipitation climatology ($agg = max, op = mean$) is also given as a future input.

Finally, we added some categorical variables: month as a future input, and latitude and longitude as static inputs. The idea behind these static inputs is to provide some spatial information to the model. Figure 1 lists all the inputs.

3.2. Training

As the regions of interest contain numerous points (170 in Rio and 248 in Florida), we divide each region into smaller sub-regions and train a different TFT model for each. The sub-regions have around 20-45 grid points each, a decision based on preliminary results. Figure 2 shows the selected grid points in Rio ($3^\circ \times 4.5^\circ$) and Florida ($6^\circ \times 5^\circ$).

We split the data into three non-overlapping ranges to define our training (1981-2010), validation (2011-2014), and test (2015-2019) sets. We perform hyperparameter optimization with random search following the same scheme presented by the TFT authors. The optimization is carried out for 60 iterations, and the list of hyperparameters includes *dropout rate*, *hidden layer size*, *minibatch size*, *learning rate* (Adam optimizer), *maximum gradient norm*, and *number of heads*. Each TFT is trained with one Nvidia V100 GPU for 100 epochs (early-stopping patience of 5 epochs). The optimization takes 3 to 4 hours to run with this configuration.

²<http://www.jamstec.go.jp>

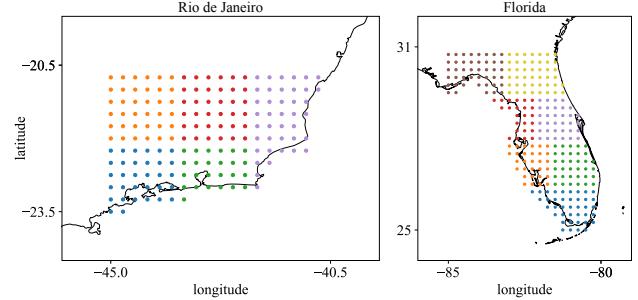


Figure 2. Selected regions and grid points in Rio de Janeiro (left) and Florida (right). The grid points are divided into smaller sub-regions, shown in different colors.

3.3. Evaluation

The TFT makes predictions for weeks 1 to 26, and here we present results only for $\tau = 26$, which means that we remove the summations over τ in Eq. 3 and replace τ by τ_{26} . as it is both challenging and directly relevant to a range of stakeholder applications. We compare the TFT predictions with both climatology and S5 predictions. Since the S5 spatial resolution is 0.4° , we use nearest neighbors interpolation to re-grid the data to the resolution of our predictions (0.25°).

4. Results and Discussion

Figure 3 shows the q-risk difference between the baseline models and TFT for quantile 0.9: climatology - TFT (top) and S5 - TFT (bottom). Positive values indicate the TFT shows superior results (lower values of q-risk are better). Compared to climatology, TFT has lower q-risks for most locations. Compared to S5, TFT exhibits considerably lower q-risk for all points – a substantial improvement.

In Table 1, we present the average q-risk difference between the reference models and TFT. Regarding climatology, TFT shows better results for all quantiles in Rio and quantile 0.9 in Florida. The average q-risk difference between S5 and TFT is significant for all quantiles in both locations, especially for quantile 0.9, which is our main focus.

Although the average q-risks for quantile 0.9 seems to show a small improvement compared to climatology, we highlight that, for seasonal time scale, climatology is exceptionally hard to improve upon (Vitart & Robertson, 2018). Furthermore, since climatology generates fixed predictions for all years, it cannot capture sudden changes or departures from normal.

Figure 4 illustrates the quantile predictions and target values of the test set in a single location in Rio de Janeiro for all models. With a relatively small q-risk improvement, one

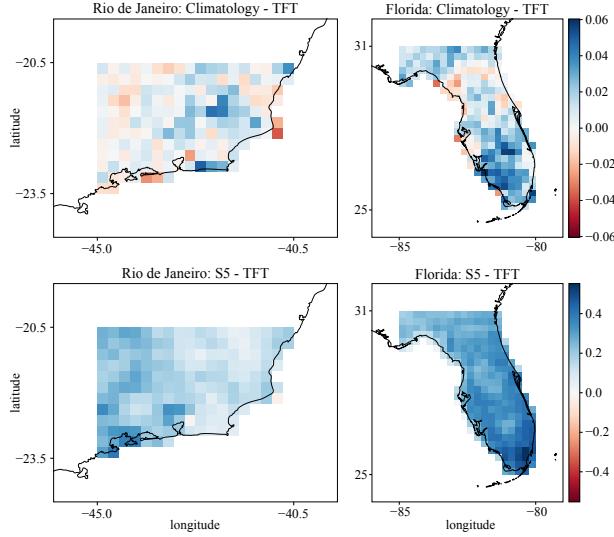


Figure 3. Q-risk comparison in Rio and Florida for quantile 0.9. Top: q-risk(Climo) - q-risk(TFT). Bottom: q-risk(S5) - q-risk(TFT).

Table 1. Average q-risk difference for each quantile.

Comparison	Region	0.1	0.5	0.9
climo - TFT	Rio	2.45%	0.90%	1.08%
	Florida	-2.16%	-0.41%	3.70%
S5 - TFT	Rio	3.71%	11.18%	29.54%
	Florida	5.70%	16.15%	41.87%

could argue that TFT is mimicking climatology. However, the TFT predictions change considerably throughout the years, contrasting with the fixed climatology behavior. We draw attention to the TFT predictions for quantile 0.9 inside the dashed box: TFT raised the quantile level during weeks of very heavy rainfall while also responding to the periods of decreased precipitation.

5. Conclusions

In this work, we used the TFT network to predict the maximum daily precipitation quantiles in a week out to a lead time of 6 months. The TFT generated significantly improved q-risks compared to the S5 model and marginal improvements compared to climatology. Comparing the 0.9 quantile prediction in one location in Rio, we have shown that TFT can accurately raise the quantile level and also respond to changes that climatology cannot. These results indicate that TFT has an interesting potential for extreme rainfall forecasting on the sub-seasonal to seasonal scale.

In the following steps, we want to modify the model's input to support 2D spatial information. That would allow us to en-

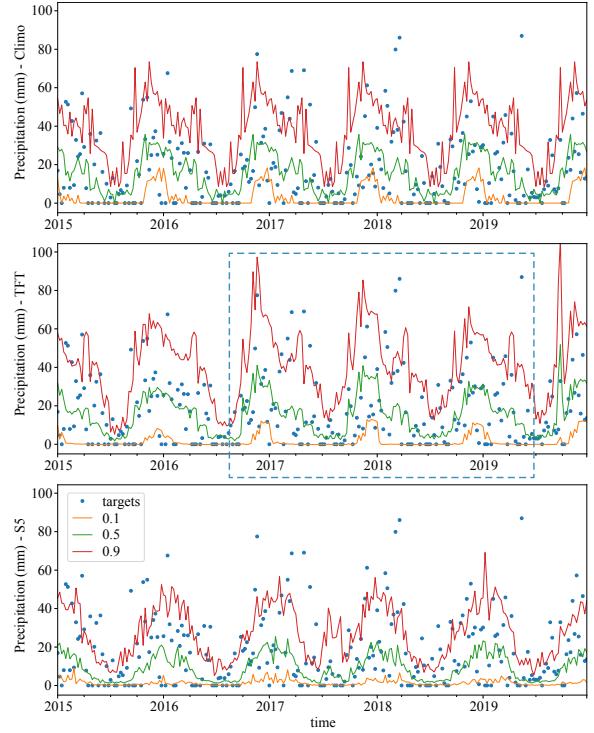


Figure 4. Predictions and targets for the test set in the location (lat -21.5, lon -41.75) in Rio. Top: Climatology. Middle: TFT. Bottom: S5. In this location, the q-risk difference for quantile 0.9 between climo and TFT was 1.9%, and S5 and TFT was 15.8%.

code neighborhood knowledge into the process. Additional pre-processing, such as POD or graph networks, could also capture teleconnections and include sparse datasets, such as sea surface temperature that is only available over the ocean. We also want to use the interpretable multi-head attention block to identify connections between the input variables and extreme rainfall.

The TFT has the ability to handle both historical data and predicted future data, providing a ready-to-use architecture for performing hybrid statistical-dynamical predictions. We intend to incorporate other input variables, such as dynamical model predictions, to investigate if they can improve TFT results.

References

- Anon. Anon. In *Science and Technology Infusion Climate Bulletin*, 2019.
- Arguez, A. et al. ENSO Normals: A New U.S. climate normals product conditioned by ENSO phase and intensity and accounting for secular trends. *Journal of Applied Meteorology and Climatology*, 58(6):1381 – 1397, 2019.
- Chantry, M., Christensen, H., Dueben, P., and Palmer, T.

Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft ai. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379 (2194):20200083, 2021.

Cohen, J. et al. S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. *WIREs Climate Change*, 10(2):e00567, 2019.

Davino, C. et al. *Quantile regression: theory and applications*, volume 988. John Wiley & Sons, 2013.

Fernando, N. et al. A process-based statistical seasonal prediction of May–July rainfall anomalies over Texas and the Southern Great Plains of the United States. *Climate Services*, 16:100133, 2019.

Funk, C. et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2(1):150066, 2015. ISSN 2052-4463.

Grimm, A. M. and Tedeschi, R. G. Enso and extreme rainfall events in south america. *Journal of Climate*, 22(7):1589 – 1609, 2009.

Herring, S. C. et al. Explaining extreme events of 2019 from a climate perspective. *Bulletin of the American Meteorological Society*, 102(1):S1 – S116, 2021.

Hersbach, H. et al. ERA5 hourly data on pressure levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2018a. Accessed on March, 2021.

Hersbach, H. et al. ERA5 hourly data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2018b. Accessed on March, 2021.

King, A. D. et al. Sub-seasonal to seasonal prediction of rainfall extremes in australia. *Quarterly Journal of the Royal Meteorological Society*, 146(730):2228–2249, 2020.

Lim, B. et al. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *arXiv preprint arXiv:1912.09363*, 2019.

Strazzo, S. et al. Application of a hybrid statistical–dynamical system to seasonal prediction of north american temperature and precipitation. *Monthly Weather Review*, 147 (2):607–625, 2019.

Vitart, F. and Robertson, A. The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Climate and Atmospheric Science*, 1:1–7, 2018.

Xu, L. et al. A data-driven multi-model ensemble for deterministic and probabilistic precipitation forecasting at seasonal scale. *Climate Dynamics*, 54(7):3355–3374, 2020.