Accuracy in Short-Term Precipitation Prediction

Robin Uhrich [*1] **Lilli Diederichs** [*2] **Mathias Neitzel** [*3] **Samuel Maier** [*4]

Abstract

In this report, we discuss the performance of numerical weather prediction models for precipitation from the European Center for Medium Range Weather Forecasting (ECMWF), selected and provided by the German Weather Service (DWD). We collected 72 hourly and 56 3-hourly forecasts from 2023-12-09 to 2024-01-25. The forecasts are compared with hourly measurements from 36 rain gauges distributed over Baden-Württemberg. From the analysis, the deviation of the precipitation amount and the accuracy of the forecast are derived. We find that the 3-day forecast is higher in accuracy ¿85% and deviates less in the amount by which it predicted ill. In the 10-day forecast the global trend is more dominant. We observed higher conformity with the forecast of each station but greater errors and lower accuracy of 70% to 55%. Analysing the input features with a correlation matrix, we find that the location in our region is uncorrelated with the forecast and the deviation of the forecast. We assume that location does not play a role in the forecast in our dataset. Using recorded data from a flooded area in Germany, we were able to show that the amount of rainfall was not well predicted. Even the 3-day forecast could not give enough weight to local conditions to predict the extreme rainfall.

1. Introduction

When we look at the weather forecast, we often question its reliability. The quality of a forecast is relevant in personal life, in agriculture planning and even vital in measuring the risk of floods. Which become according to (**?**) more and more prominent as we know for example from the Ahrtal valley in 2021 (**?**) or most recent flooding event in Germany at around Christmas 2023 (**?**).

In this report we discuss the quality of precipitation forecasting in Baden-Württemberg using the German Weather Service (DWD). We look at the quality of the 3-day and 10-day forecasts from 36 stations to check for over- and under-prediction of precipitation and calculate the accuracy of the forecasts. To further understand the predictions and the limitations of our data, we look at the correlation of our input features and evaluate the dependence of a station's location on its forecast. A very local weather phenomenon combined with a local lack of drainage can lead to flooding. The short term cumulative rainfall can indicate such risks, which we look at last (**?**).

fig/fig_bawu_map.pdf

Figure 1. Map of Baden-Württemberg. Each selected station is one dot on the map. The blue dot is Tübingen (which has no data), the red dot is Dachsberg-Wolpadingen

2. Data

Our analysis is based on data provided by the DWD. We collected forecasts on a daily basis for every weather station marked in *Figure* **??** from the Open-Data dataset. The dataset is divided in two categories:

- **Reference Precipitation**:
  Selected DWD weather stations measure precipitation with a rain gauge called rain[e]H3 (**?**). The DWD provides the recent and historical measurement of precipitation with free access in the CDC dataset. We use these measurements as a reference to the forecast of each station.

- **Predictions**:
  The weather forecasts of all stations marked in *Figure*

[*]Equal contribution. University of Tübingen, Baden-Württemberg, Germany [1]matriculation number 6651884, robin.uhrich@student.uni-tuebingen.de [2]matriculation number 6638382, lilli.diederichs@gmail.com [3]matriculation number 4243096, mathias-neitzel@student.uni-tuebingen.de [4]matriculation number 4243096, sam.maier@student.uni-tuebingen.de.

The corresponding GitHub repository can be found here.

?? are collected once a day at 00:10. The total precipitation is forecast every hour for the next 3 days, called forecast 1 and every three hours for up to ten days called forecast 2.

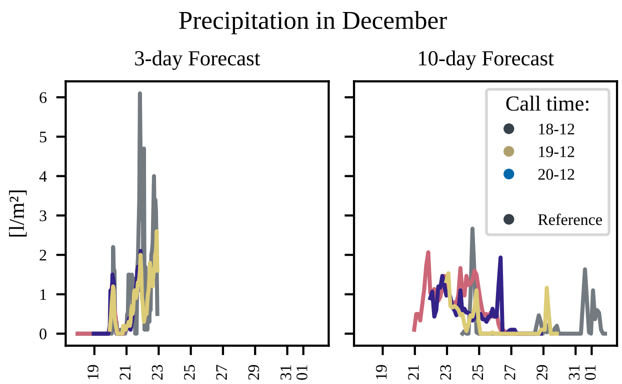

Figure 2. Raw Data for station in Dachsberg-Wolpadingen (red marked in Figure **??**. Forecasts from the 2023-12-18 and 2023-12-19 in color. Reference data in gray. The left plot is for the 3 day and the right plot is for 10 day forecast starting after day 3.

We began our collection in 2023-12-09 and ended on 2024-01-25 resulting in 225,792 forecast samples. Because our reference data is limited up to the 2024-01-25 we have in total XXXXX samples.
that time is forecast from multiple call times. A section of the available data can be seen in Figure **??**. Here we can see the measured precipitation a reference in gray and colored the different call times. The different call times start on their call day and predict for the following 3 days and thereafter 7 days. Therefore we get multiple forecasts for one time.

3. Methods

As declared before our data is split into two categories: forecasts $\hat{\mathbf{X}}$ and reference data $\mathbf{X}$. Where $\hat{\mathbf{X}}_{t,s,c,i}$ is the precipitation forecast at station s, at time t queried at time c, and $i \in \{1, 2\}$ for three day and 10 day forecast. Respectively $\mathbf{X}_{t,s}$ is the reference precipitation data at station s and at time t.
An alternative notion for forecast data is: $\hat{\boldsymbol{\mathcal{X}}}_{\Delta t,s,c,i}$ where Δt is the time delta between the timestamp of the query time and the time the forecast is for.

We denote N as the number of stations, T as the number of all possible timestamps with forecast and reference data, C is the number of queries executed and Q as the number of time steps ahead the forecast i is predicting.

3.1. Difference Measurement

To access insight into the quantity how wrong the forecast actually is we define two difference metrics. The mean error in *Equation* **??** and mean absolute error *Equation* **??**.

$$ME(\hat{\boldsymbol{\mathcal{X}}}, \mathbf{X})_{\Delta t,i} = \frac{1}{NC} \sum_{s,c} \hat{\boldsymbol{\mathcal{X}}}_{\Delta t,s,c,i} - \mathbf{X}_{c+\Delta t,s} \quad (1)$$

$$MAE(\hat{\boldsymbol{\mathcal{X}}}, \mathbf{X})_{\Delta t,i} = \frac{1}{NC} \sum_{s,c} |\hat{\boldsymbol{\mathcal{X}}}_{\Delta t,s,c,i} - \mathbf{X}_{c+\Delta t,s}| \quad (2)$$

3.2. Accuracy

3.3. Accuracy Sam

In rough approximation of what the DWD is doing to judge the quality of their forecast[1] , we calculate the accuracy $a_{\Delta t}^{\Delta T,\theta}$ of predicting that it rains for different thresholds θ and different timesteps Δt into the future, where a forecast of more precipitation than θ over ΔT is interpreted as rain. The DWD does this with $\Delta T = 12$h, $\Delta t = 1$day and $\theta = 0$. We will the show this for some values of θ and for every Δt in the 10-day-forecast, and we will mean it over all call times and all stations we observed. So we will have a value every hour for the first 3 days and afterwards every 3 hours.

We will use a slightly different notation to the previous chapter. Each forecast will describe the duration $[t, t+\Delta T)$, and will be denoted with $\hat{\mathbf{X}}_{\Delta t,s,c}^{\Delta T}$, where $\Delta t, s, c$ have the same meaning as before. Reference data will be described with $\mathbf{X}_{\Delta t,s,c}^{\Delta T}$ accordingly.

First we accumulate the reference precipitation, which was given for every $\Delta\tilde{T} = 1$h window to cover each $\Delta T = 12$h window:

$$\mathbf{X}_{\Delta t,s,c}^{\Delta T} = \sum_{[\Delta\tilde{t}, \Delta\tilde{t}+\Delta\tilde{T}) \subset [\Delta t, \Delta t+\Delta T)} \mathbf{X}_{\Delta\tilde{t},s,c}^{\Delta\tilde{T}} \quad (3)$$

Accordingly we accumulate forecast precipitation $\hat{\mathbf{X}}_{\Delta t,s,c}^{\Delta T}$.

We consider the forecast to predict rain if $\hat{\mathbf{X}}_{\Delta t,s,c}^{\Delta T} > \theta$. For the reference we always consider $\hat{\mathbf{X}}_{\Delta t,s,c}^{\Delta T} > 0.0\frac{\text{mm}}{\text{m}^2\text{h}}$ as rain.

To get the accuracy (*Trefferquote*) we count the occurrences where we the forecast and prediction according to our considerations agree for all call times c and stations s, and divide it by the number of samples $C \cdot S$. To denote the count we use the indicator function I:

$$a_{\Delta t}^{\Delta T,\theta} = \frac{1}{CS} \sum_{s,c} I\left(\left(\hat{\mathbf{X}}_{\Delta t,s,c}^{\Delta T} > \theta\right) \wedge \left(\mathbf{X}_{\Delta t,s,c}^{\Delta T} > 0\right)\right) \quad (4)$$

[1] https://www.dwd.de/DE/wetter/schon_gewusst/qualitaetvorhersage/qualitaetvorhersage_node.html#doc446682bodyText2 Abbildung 2, right

3.4. Cumulative Sum of Precipitation

To have a look into the total amount the forecast was of by the end of measurements we define the cumulative precipitation per station in *Equation* **??** and the cumulative mean forecast in *Equation* **??**.

$$\widehat{CP}(\hat{\mathbf{X}})_{s,t,i} = \sum_{\tilde{t}=t_0}^{t} \frac{1}{C_{\tilde{t}}} \sum_{c} \mathbf{X}_{\tilde{t},s,c,i} \tag{5}$$

$$CP(\mathbf{X})_{s,t} = \sum_{\tilde{t}=t_0}^{t} \mathbf{X}_{\tilde{t},s} \tag{6}$$

t_0 denotes the begin and $t_{\max}$ the end of measurements. $C_{\tilde{t}}$ is the number of forecasts available for a specific timestamp.

4. Results

So far we looked at one station and two call days, to further handle the data we need break down this tensor of stations, call day, time, and forecast type. Note that a mean over call times inherits an error by the deviation of the forecasts from the different call days. This deviation of the call days forecasting different values for the same day we examine in the Figure **??**.

According to Eq.:**??** find the evolution of the forecasts for different call times by calculating the mean over time steps into the future Δt for all days. We can see on the left plot (Figure:**??**) the mean deviation over Δt ahead for all stations and the trend of the mean of the stations for the 3-Day forecast (Figure: **??**). On the right hand side the same metric is plotted for the 10-day forecast.
Remarkably the deviations of the stations are convergent, especially for the 10-day forecast the difference of forecast to precipitation oscillates similarly for most stations. However there is one station Dachsberg-Wolpadingen (marked in red) with significant higher amplitude. This station lies in the small region of flooded areas, which we can see in the cumulative sum of precipitation later Figure:**??**.

For the 3-day forecast the trend is positive, hence predicts in sum more than what poured down. The 10-day forecast predict in average less precipitation than occurred the further the time it predicts ahead. But looking at the absolute mean error of 10-day forecast we get a positive trend that is higher than the 3-day forecast trend. So the absolute deviation of the forecast increases the further we predict ahead.

Processing the data according to **??** we can see in Figure: **??** in the background in gray the distribution of the data and in color the metrics for both the 3-day and the 10-day forecast. The convergence of our errors is due to the metric. The higher the threshold we still accept as ”no rain” the lower the true positive and false negative

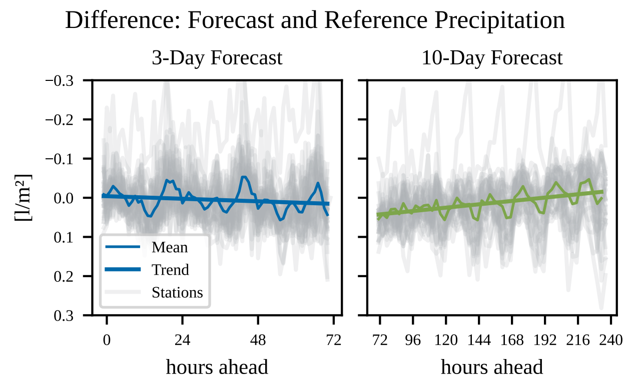

Figure 3. The difference of forecast precipitation and reference precipitation for time steps ahead for each station (gray), the mean over all stations (blue) and the trend of the mean (orange). The left plot is for forecast 1 and the right for forecast 2

rates will be. We can see the distribution of our data is accordingly. For the 3-day forecast the true negative rate saturates faster at a higher level than the 10-day forecast. The false positive rate decreases faster and the false negative declines more rapidly compared to the 10-day forecast.

The confusion statistics of metric: **??** are shown in Figure: **??** where

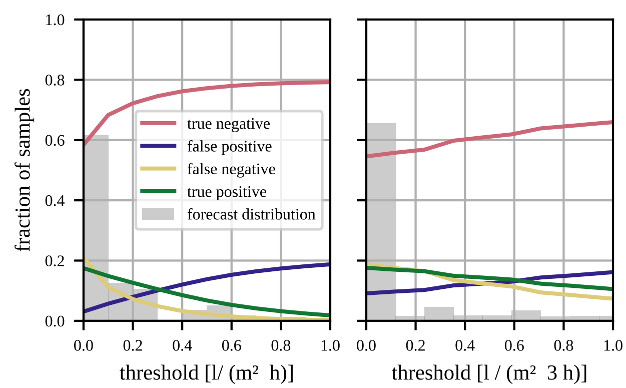

Figure 4. Error statistics

By averaging over all forecast that predict Δt into the future we assume, that the forecast per day is independently identically distributed. While this seems physically unrealistic, our forecast seems to be IID as we found the day is uncorrelated to the forecast and the difference of the forecast to the reference value. The location does not have an influence on the forecast either. The stations features only correlate with their own: Longitude, Latitude and Altitude.
The forecast seems to be mainly be correlated by the reference precipitation.

By looking at the time frame the floods occurred in south

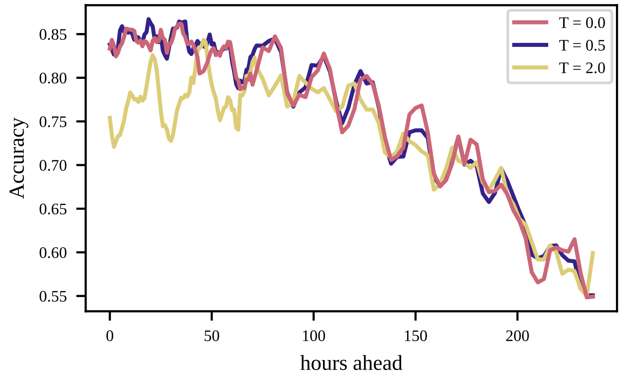

Figure 5. TODO SAM

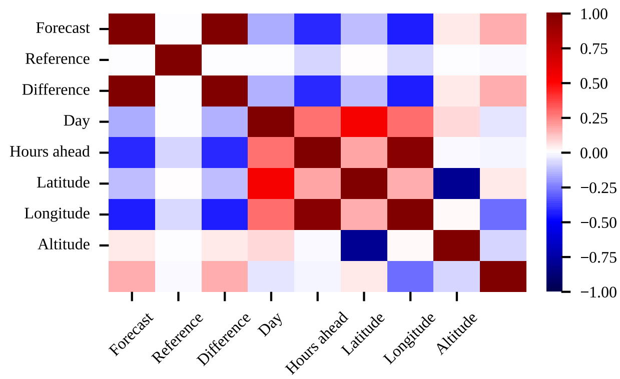

Figure 6. Correlation Matrix of input features

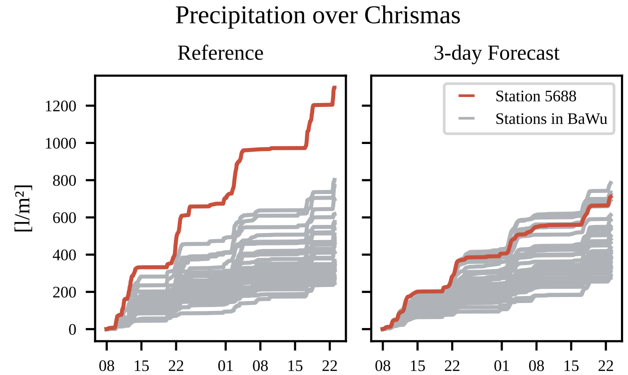

Figure 7. Unforcasted extreme precipitation in Dachsberg-Wolpadingen, marked red in figure **??**. Cumulative Sum for the sanity check.

Germany and calculating the accumulated sum, we estimate how much water came down in this time frame and how much was expected by the 3-Day forecast. In this plot we can see, for one station the amount of water surpassed $700[l/m^2]$ by the end of December when the forecast predicted around than $400[l/m^2]$. The mean precipitation for December in Baden-Württemberg is $114[l/m]$ **??** which was exceeded in only 3 days.

5. Discussion & Conclusion

Forecast predictions depend on many factors. With our data we could get an idea of the forecast of the DWD, which contributes to most forecasts in Germany. The data was collected in a rainy season and only for a Baden-Württemberg. Our results indicate only tendencies for 3-day and 10-day forecasts. For a thorough analysis we would have needed to record for at least a whole year and larger regions. Meteorological effects, large weather phenomenon or global warming, couldn't be derived from 49 days of forecast recording. Nonetheless we find our results could, still help you to read the forecast more carefully.

Our results indicate that the 3-day forecast is more accurate in terms of timing and amount of precipitation. Looking at the difference between forecast and reference precipitation plotted in *Figure:***??**, the mean trend does increase for hours ahead but is very low. Therefore in mean the amount of precipitation is slightly lower than was predicted. The trend for the absolute difference **??** is higher, but still rather close to zero. So in mean the amount of rain predicted, was similar to the reference precipitation. The 10-day forecast has a steeper mean trend in the difference of precipitation, with a negative trend indicating that the amount of rain was underestimated. The trend of the absolute difference is steeper compared to the absolute trend of the 3-day forecast, indicating that the 10-day forecast predicts less accurate in amount and timing. This could be related to the fact that after 4 to 5 days the global weather derived by the ICON 13 model ws applied, which forecasts weather phenomenon globally (**?**).

The accuracy underlines this relation. The hourly accuracy reached reliable values of $> 80\%$ for the one hour ahead which is in the well aligned with the 12h-accuracy of the DWD **??**. However the forecast's accuracy does not reflect the extent of deviation in the event of rainfall. The accuracy of the forecast is potentially misleading as it does not consider the amount of rainfall.

For example the extreme rainfall in Dachsberg-Wolpadingen was not expected. Looking at the Figure **??** we can see the accumulated amount of rain was especially high for Dachsberg-Wolpadingen. The average rainfall in December 2023 in Baden-Württemberg was 114 $[l/m]$ which Dachsberg-Wolpadingen surpassed in only 4 days.**??** The high precipitation together with the lack of drain and dry surfaces caused a flood in the region of Dachsberg-Wolpadingen..**??** This high precipitation could be due to very local meterological effect. In that case the location of the station would give an insight to the deviation of the forecast.

But looking at the correlation matrix **??** the locations of the stations in Baden-Württemberg showed almost no correlation to the forecast or its difference to the reference data. In the Figure **??** we see that there are deviations in the difference of the forecast to the reference precipitation, however the behavior of the stations were similar. Even for Dachsberg-Wolpadingen the oscillations had been around the same mean trend. We can see that the deviations were especially big for Dachsberg-Wolpadingen but as the correlation matrix indicates no correlation of the location and it's error, we suspect that the high oscillation are due to the heavy precipitation events and not generally higher for Dachsberg-Wolpadingen. For definite result we would need a collection time that is not dominated by extreme precipiatation events.

To sum up our analysis we came to the conclusion that weather is a highly complex topic and quantifying the performance of a forecast is similar difficult with many more aspects and features we haven't even done the analysis for yet. Therefore we cannot make definite claims with the amount of data we gathered until now, but we were able to show that our analysis does relate to the previous analysis published by the DWD.

We found that the forecast does worsen the further it predicts into the future. Looking 3 days ahead the forecast tends to predict a little more rain and with the highest accuracy. The 10-day forecast does underestimate the amount of rain and worsens in accuracy dropping from 80 % for 4 days ahead down to 55% for 10 days ahead. Additionally it does happen that a forecast underestimates the amount of rainfall drastically but this is usually due to additional circumstances. There is no general area in Baden-Württemberg that correlates with it's precipitation or forecasting error.

Acknowledgements

We would like to thank the DWD to provide public access to forecast and reference data. Especially we would like to thank Axel Kuschnerow for given support. Also we thank our Tutor, Emilia Magnani for help and advice at any time.

Contribution Statement

Samuel Maier and Robin Uhrich wrote the Python code base to collect and prepared the data. Together with Lilli Diederichs they performed the data analysis. In parallel Mathias Neitzel assisted by formulating the analysis into formal statements. Lilli Diederichs was responsible for the visualizations. All authors jointly wrote the text of the report.

References

Dienst, D. W. Deutschlandwetter im dezember 2023, 2023. URL https://www.dwd.de/DE/presse/pressemitteilungen/DE/2023/20231229_deutschlandwetter_dezember2023_news.html?nn=16210. Accessed: 2024-01-25.

Lambrecht. rain[e] weighing precipitation sensor. URL https://www.lambrecht.net/en/products/precipitation/weighing-precipitation-sensor-rain-e. Accessed: 2024-01-25.

Najibi, N. and Devineni, N. Recent trends in the frequency and duration of global floods. Earth System Dynamics, 9(2):757–783, 2018. doi: 10.5194/esd-9-757-2018. URL https://esd.copernicus.org/articles/9/757/2018/.

Pink S., S. J. Das Wetter ist politisch – Starkregen, Hochwasser und Flut vor der Bundestagswahl 2021. Springer, 2023.