Can we trust the weather forecast

Marc Baumholz
Matrikelnummer 6606376
marc.baumholz@student.uni-tuebingen.de

Lennard Berger
Matrikelnummer 6668183
lennard.berger@student.uni-tuebingen.de

David Schlaak
Matrikelnummer 6631578
david.schlaak@student.uni-tuebingen.de

Abstract

Weather forecasts are heavily beeing relied upon and an important factors for life decision. In this paper, we analyse the trustworthiness of the weather forceast based upon forecasts by the Deutsche Wetterdienst. We analysed two highly relevant parameters: temperature and precipitation. To evaluate the trustworthiness we computed a regression model between ground truth data of the past six months and hourly collected forecasts of the next 240 hours (10 day forecast). 10 fold cross-validation was used to provide a rigorous estimate of the predictive accuracy of weather forecasts. Overall, the results indicated a higher trustworthiness for temperature compared to precipitation.

1 Introduction

Have you ever left your umbrella at home because the weather app said it would not rain, only to end up drenched? Weather forecasts significantly influence our daily decisions, from clothing choices to event planning. Despite advanced technology, accurately predicting weather remains challenging due to the atmosphere's complexity and chaotic nature. Our study investigates the reliability of temperature and rainfall forecasts in weather apps, focusing on the question: "How much can we trust the temperature and rainfall forecasts our weather apps give us for the next few days?"

We utilize data from the MOSMIX system, run by the German Weather Service (DWD) [1], which is based on the ICON (ICOsahedral Non-hydrostatic) modeling framework [2]. ICON creates a global grid of 2,949,120 triangles, each covering approximately 173 square kilometers. The weather forecast produced by MOSMIX tunes the ICON output based on weather input data from stations located inside the area covered by each triangle and supplemented by globally available weather data (atmospheric pressure, solar intensity, sea temperature, etc.). This geometry offers statistical advantages but has limitations:

> All scalar prognostic model variables (e.g. temperature, density, moisture quantities) are located in the circumcenter of the triangles, whereas the edge-normal wind components are located in the edge midpoints.

Weather stations, chosen strategically, overlap with these triangles. However, stations near triangle edges may suffer in forecast quality.

To investigate the real-world implications of this, we are looking at MOSMIX forecast data hour by hour for the next 10 days (240 hours) from different 10 weather stations.

Our goal is to see if we can spot any patterns or trends in how accurate these forecasts are, getting a better idea of how much one can rely on their weather apps.

2 Methods

2.1 Dataset

We conducted a research study using data from the Deutscher Wetterdienst (DWD), specifically the MOSMIX weather forecast accessed through their official API. Our study focused on two key parameters: temperature (Kelvin, 2m above surface) and precipitation (millimeters, total for the last hour), as seen in Figure 1. Data was collected hourly over six months, with forecasts extending 240 hours in advance, from 10 selected weather stations. This approach was chosen to validate our hypothesis on weather forecast reliability, with particular relevance to the official iPhone weather app, which relies on the DWD forecasts.

The Data format was the following:

- Station ID: Unique identifier for the weather station.
- Timestamp: Time of the forecast issuance.
- Forecast Time: The time the forecast predicts weather conditions.
- Parameter: Meteorological variable (e.g., temperature, precipitation).
- Value: Forecasted value for the corresponding parameter.

The raw data extraction involved developing a specialized parsing script to extract relevant parameters, in this case temperature and precipitation measurements.

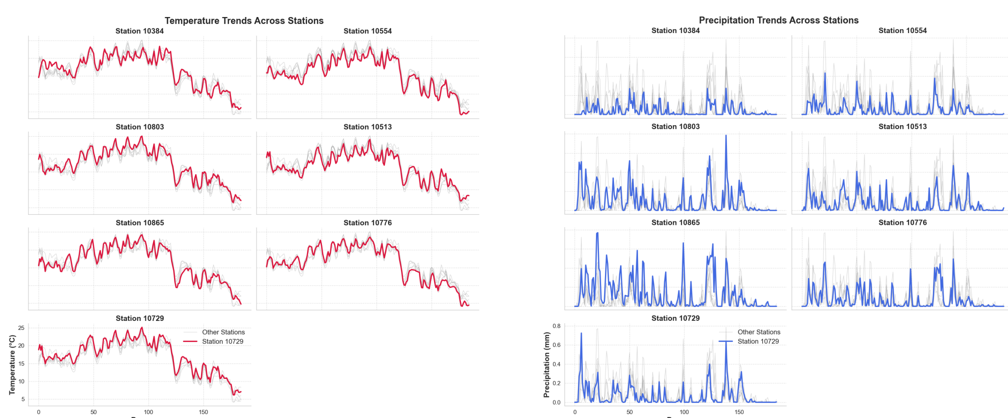


Figure 1: Temperature and Precipitation Trends over all datapoints

2.2 Data Preparation

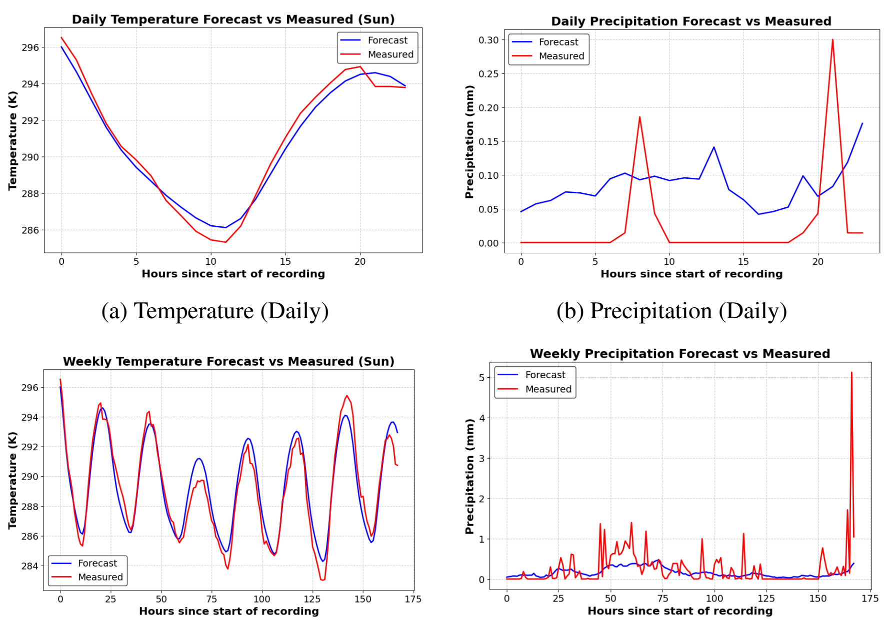
Our analysis commenced with a comprehensive examination of the dataset's various parameters, encompassing tasks such as identifying missing data, performing data cleaning operations, detecting outliers, converting temperature scales, and conducting correlation analyses. Initially, we pruned the dataset by removing hours that lacked the full complement of 240 datapoints, which affected the initial measurement period, constituting 135 out of 4417 hours measured $\approx 3\%$. Additionally, we filtered out weather stations exhibiting incomplete data, as well as those with erroneous readings, ultimately resulting in a refined dataset comprising 7 stations. These 7 stations were mapped in correspondence between real weather stations, supplying ground truth data, and forecasts stations (the DWD provides some overlap between those two, but not all of them). This meticulous data preparation process yielded a dataset structure featuring 240 datapoints per hour across 7 stations, amounting to a total of 1680 measurements for each hourly interval and parameter.

2.3 Regression

Given the nature of our data involving continuous variables like temperature, we selected a regression model, which is particularly well-suited for predicting trends and relationships between variables over time. Our regression model mapped the 1680 datapoints collected hourly to a single ground truth value. To ensure model robustness, we employed a standard data splitting technique with 10-fold cross-validation, allocating 90 percent of data to the training set and 10 percent for testing. The dependent variable was the average hourly value computed across different weather stations, which helped smooth out local variations and provide a more representative overall forecast.

3 Results

Our analysis of the weather forecast data revealed distinct patterns in the reliability of temperature and precipitation predictions. The temperature forecasts demonstrated high accuracy, with predicted values closely aligning with actual measurements across various timeframes. This reliability was visually confirmed through time series daily and weekly plots (Figure 2), which showed minimal deviation between forecasted and observed temperature values. In contrast, the precipitation forecasts exhibited lower accuracy, with significant discrepancies between predicted and actual rainfall measurements. These results indicate that while temperature forecasts are highly dependable, precipitation forecasts are less reliable and require further refinement.



(a) Temperature (Daily)

(b) Precipitation (Daily)

(c) Temperature (Weekly)

(d) Precipitation (Weekly)

Figure 2: Daily and Weekly Temperature and Precipitation Comparisons

This results also shown up in our numerical analysis. The table below shows all relevant results of the regression analysis and further 10-fold-cross-validation.

The regression analysis for temperature forecasts demonstrates a strong alignment between forecasted and observed values, with an r-value of 0.9746, indicating near-perfect correlation. The slope of 1.0328 suggests that the forecasts closely track the observed temperature trends with only a slight deviation. Moreover, the 10-fold cross-validation results show a mean accuracy of 0.8908, reflecting consistently reliable predictive performance across the dataset. This consistency is further

Metric	Temperature	Precipitation
Slope (± Std Error)	1.0328 ± 0.0036	1.3512 ± 0.0423
R-value	0.9746	0.4387
R-squared	0.9498	0.1925
Accuracy (10-fold, ± Std Error)	0.8908 ± 0.0336	0.1555 ± 0.1347

Table 1: Regression results for temperature and precipitation

supported by the relatively narrow range of cross-validation scores (± Std Error 0.0336), indicating low variability in predictive performance for temperature forecasts. In contrast, rainfall forecasts exhibit a weaker relationship with observations, as evidenced by an r-value of 0.4387. While statistically significant, this moderate correlation highlights the challenges in accurately modeling precipitation, a parameter known for its high variability. When validating the accuracy via 10-fold cross-validation we obtained a mean accuracy of 0.1555 (± Std Error 0.1347). We repeated this analysis with ridge regularization to obtain near identical results.

4 Discussion

The findings of this study highlight the varying degrees of reliability in weather forecasting, particularly when comparing temperature and precipitation predictions. Temperature forecasts demonstrated high accuracy, as reflected by the strong correlation and consistent predictive performance across all validation metrics. This reliability can be attributed to the relatively straightforward physical processes that govern temperature changes, which are well-captured by forecasting models. In contrast, precipitation forecasts showed significantly lower accuracy and greater variability. This disparity underscores the inherent challenges in modeling precipitation, a phenomenon influenced by complex, localized, and often stochastic meteorological processes. Thus the hypothesis that weather forecasts for temperature and rainfall are reliable is only partially supported. While temperature forecasts proved trustworthy, precipitation forecasts fell short of similar standards. Several limitations of this study must be acknowledged. First, the temporal scope was limited to six months of data, potentially excluding seasonal and long-term variations in weather patterns. Second, the dataset was confined to 10 weather stations, which may limit the generalizability of the findings to other regions or weather conditions. Third, the analysis focused solely on temperature and precipitation, excluding other meteorological parameters that could provide additional context or insights. Finally, the use of simple statistical approaches, such as linear regression, while effective for capturing general trends, may not fully account for the complexities inherent in meteorological phenomena. To address these limitations and improve the reliability of weather forecasting, several future steps are recommended. Extending the analysis over a longer time frame, such as multiple years, would capture seasonal and inter-annual variability. Incorporating data from a larger and more diverse set of weather stations would enhance the regional and global applicability of the findings. Employing advanced analytical methodologies, such as machine learning models (e.g., random forests, neural networks) and time series analysis, could better capture non-linear and temporal dependencies in meteorological data. Expanding the range of parameters to include wind speed, humidity, and atmospheric pressure would provide a more comprehensive understanding of weather dynamics. Additionally, collaborating with meteorological agencies to refine forecasting models and validate results on a larger scale could significantly advance the field. The results of this study demonstrate that temperature forecasts are a reliable tool for decision-making, reflecting the robustness of current models in capturing temperature trends. However, the less reliable precipitation forecasts highlight the need for methodological advancements. By addressing these challenges through the outlined future directions, weather forecasting systems can be made more accurate and reliable, ultimately enhancing public trust in their predictions.

5 Statement of Contributions

From the initial brainstorming phase, each team member contributed unique perspectives that shaped the project's conceptual framework and guided the statistical analyses. During execution, Lennard Berger played a key role in implementing and optimizing the regression model, while all members collaborated on drafting, refining methodologies, and synthesizing results for the final report.

References

[1] Deutscher Wetterdienst. *Verfahrensdokumentation für das MOSMIX-Vorhersagesystem*. Deutscher Wetterdienst, Offenbach am Main, 2024. Stand: 10.06.2024.

[2] Günther Zängl, Daniel Reinert, Pilar Rípodas, and Michael Baldauf. The icon (icosahedral non-hydrostatic) modelling framework of dwd and mpi-m: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, 141(687):563–579, 2015.