GUARDINGER ADVANCED TECHNOLOGIES

AI-based Enhanced Weather Forecasting using Weather Radar and Open-source Data

Model Evaluation Strategy

GUARDINGER <> ASTRA

What is happening right now:

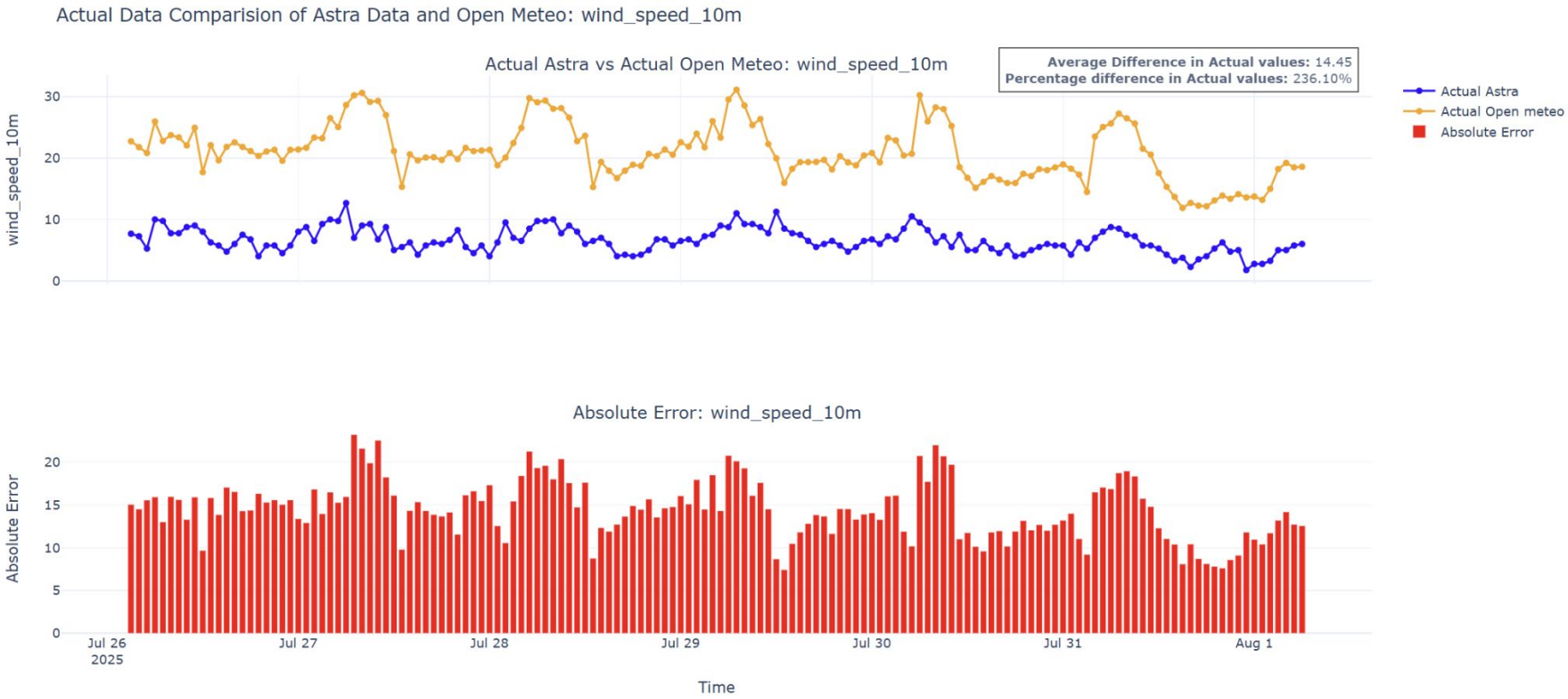
We're analyzing the data we received on August 1st from Astra, which covers the period from July 26th to August 1st. After carefully analyzing the data we've identified a major discrepancy in the data: the model's training and validation datasets come from different sources, leading to inaccurate validation and unreliable results

Specifically, we've found three key issues:

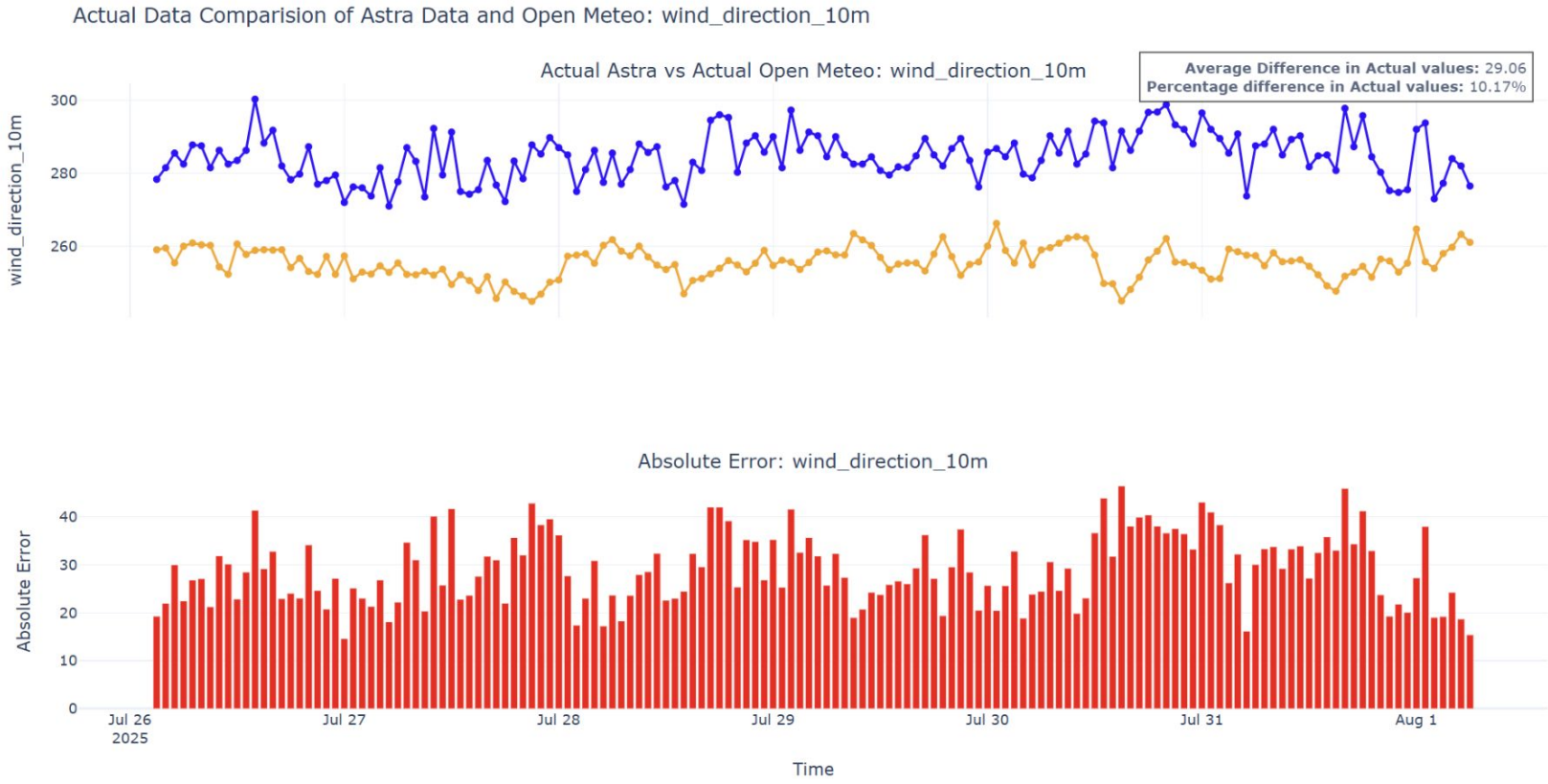
- The training data (from Open Meteo) and the validation data (from IMD, Astra) are from **different sources**
- The geographical **coordinates for the sites are not identical** across both datasets, which introduces further inaccuracies
- Both datasets were **sampled at different time intervals**, creating a mismatch that prevents a fair comparison

To ensure meaningful validation, the model must be trained and tested on consistent, source-aligned data.

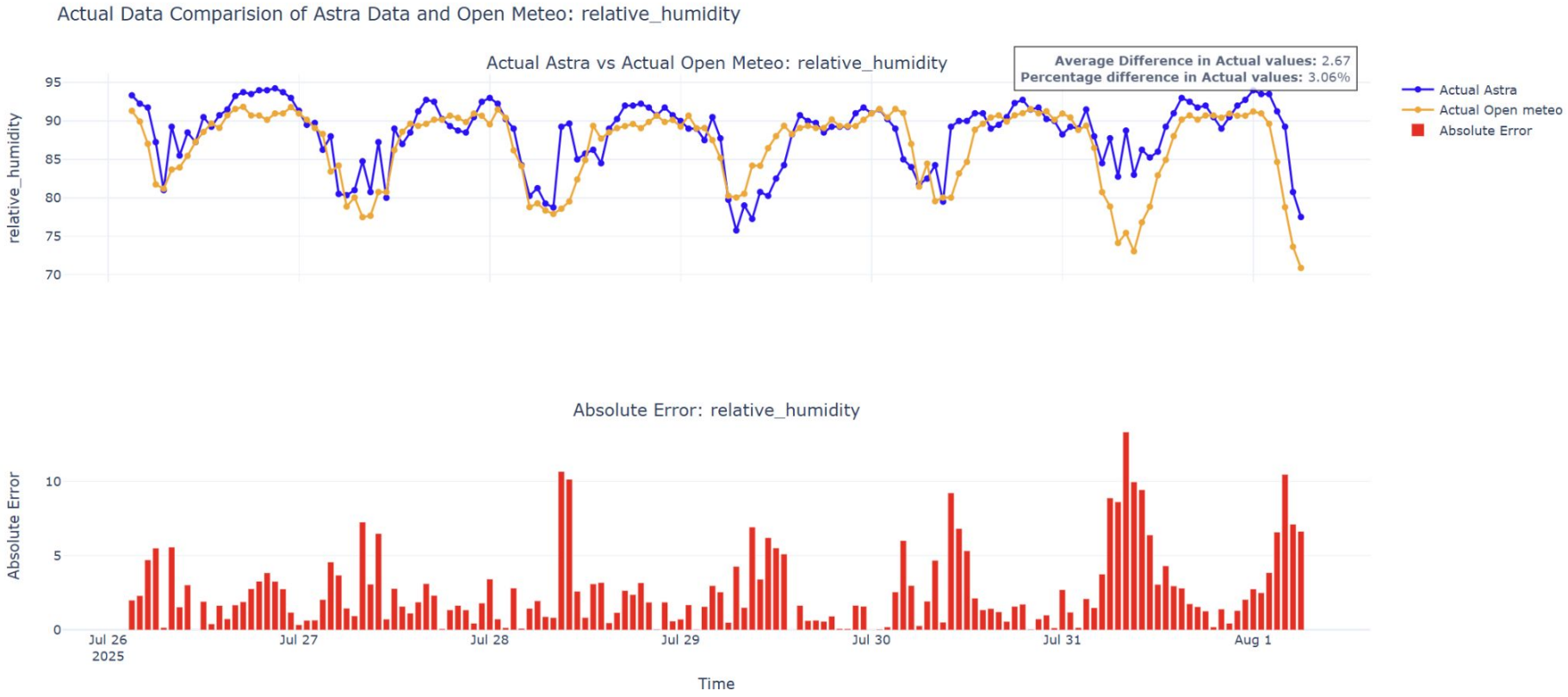
In next slides you can clearly see the differences between actual values for the same data parameters and for the same location just because the data is from the different sources



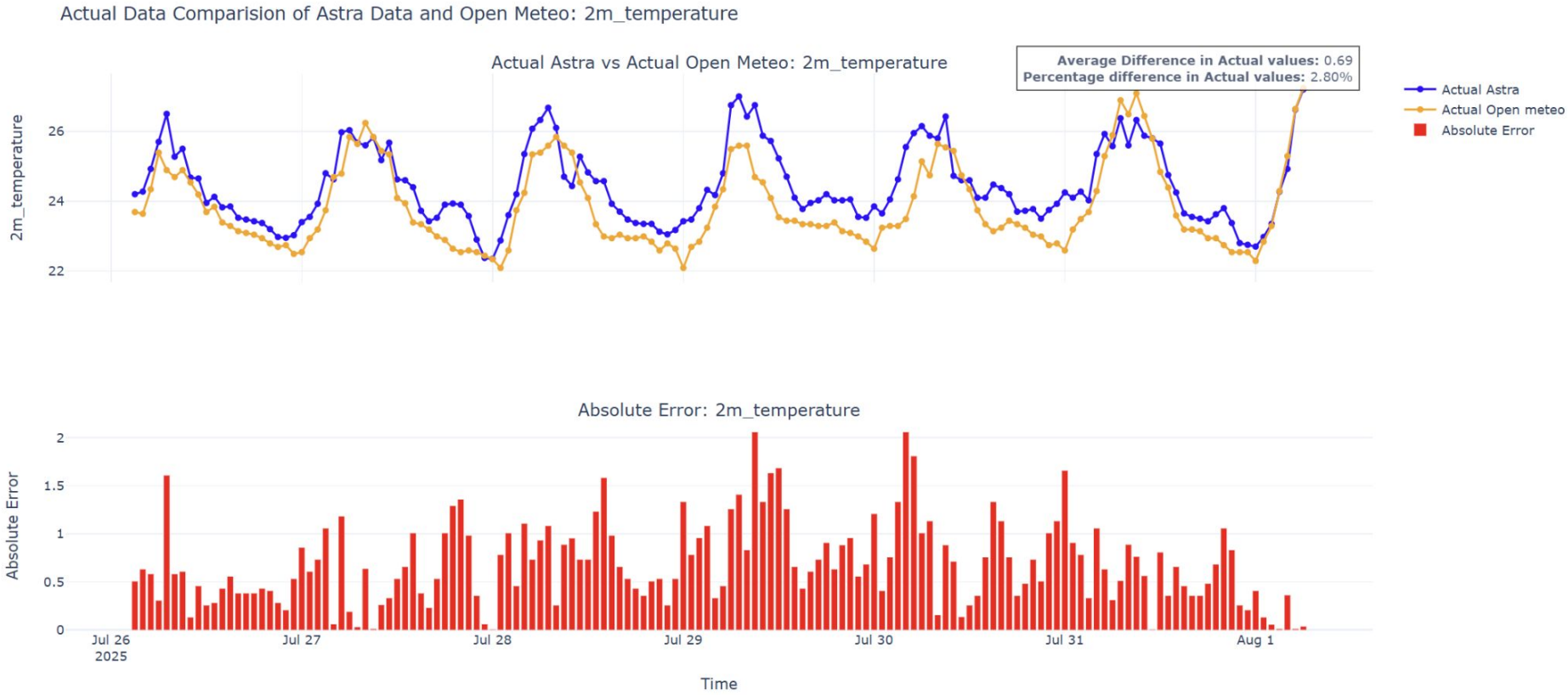
The above plots shows the difference between actual data from IMD shared by Astra and Open-Meteo, On top right corner we can see the average difference between these 2 data sources and percentage difference in the actual values of wind speed



The above plots shows the difference between actual data from IMD shared by Astra and Open-Meteo, On top right corner we can see the average difference between these 2 data sources and percentage difference in the actual values of wind direction



The above plots shows the difference between actual data from IMD shared by Astra and Open-Meteo, On top right corner we can see the average difference between these 2 data sources and percentage difference in the actual values relative humidity



The above plots shows the difference between actual data from IMD shared by Astra and Open-Meteo, On top right corner we can see the average difference between these 2 data sources and percentage difference in the actual values 2m_temperature

Correct Model Evaluation Strategy:

As a solution for this situation,

- we must ensure that the training and validation datasets are sourced from the same place (that means both must be from either open meteo or IMD, Astra) then only it can be comparable, use identical coordinates, and are sampled at the same time intervals to achieve accurate and reliable model validation.
- To implement this, we require data spanning for at least 13 days. This will allow us to train the model on the first 10 days of data and use the remaining 3 days for accurate forecasting and validation.

Thank You