Fraunhofer IWES

10/10/2024

Wind Data Gap Filling – Is it worth it?

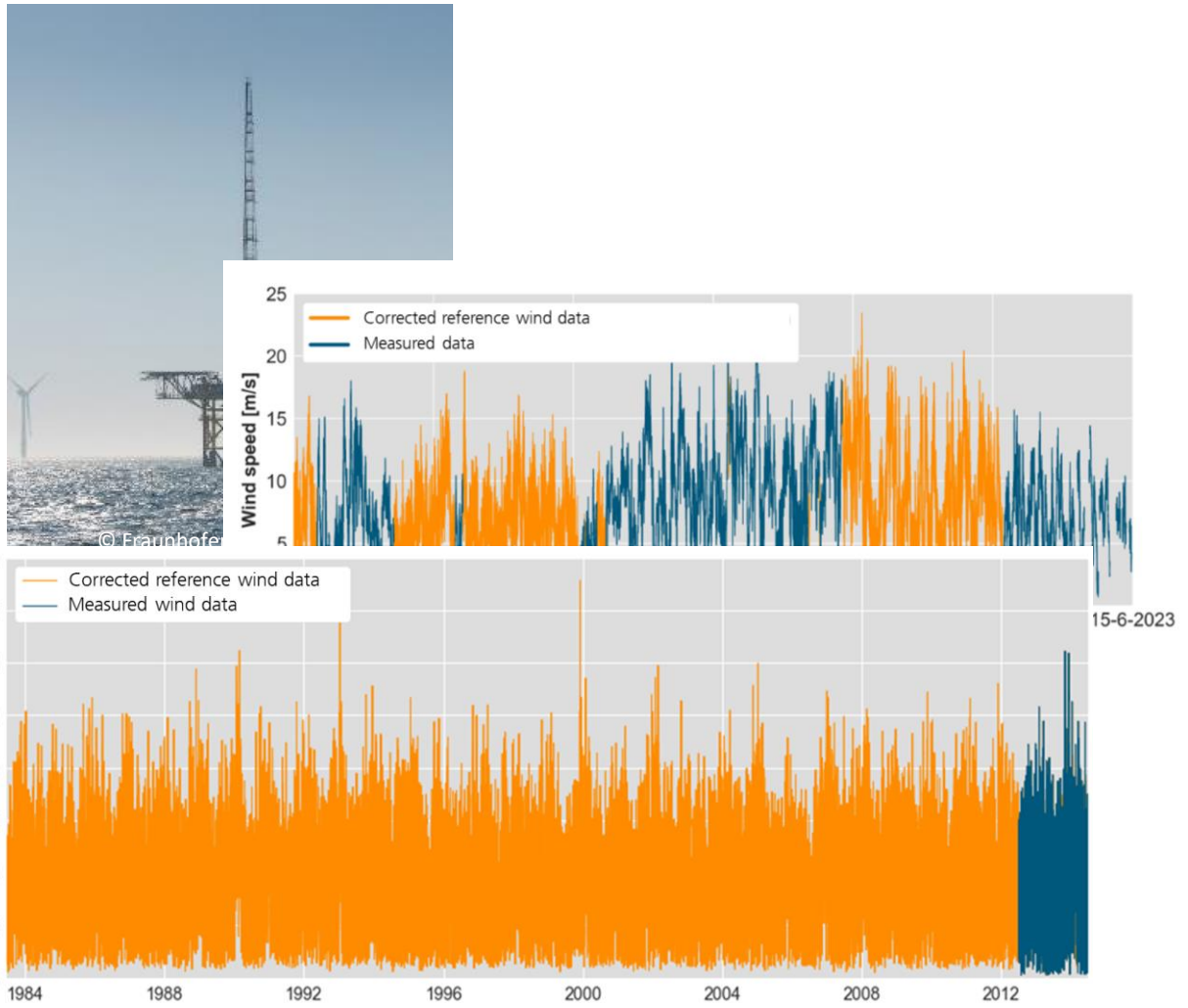
Martin Jonietz Alvarez
martin.georg.jonietz.alvarez@iwes.fraunhofer.de

From wind measurement to long-term wind resource

- **Wind measurement**
  - Data usually contains gaps
  - Short term (1-5 years)

- **Data gap filling**
  - Numerical algorithm (linear, ML…)
  - Model data input (ERA5, NEWA…)

- **Long-term extrapolation (20-30 years)**
  - Numerical algorithm (linear, ML…)
  - Model data input (ERA5, NEWA…)



Fraunhofer IWES

The research question

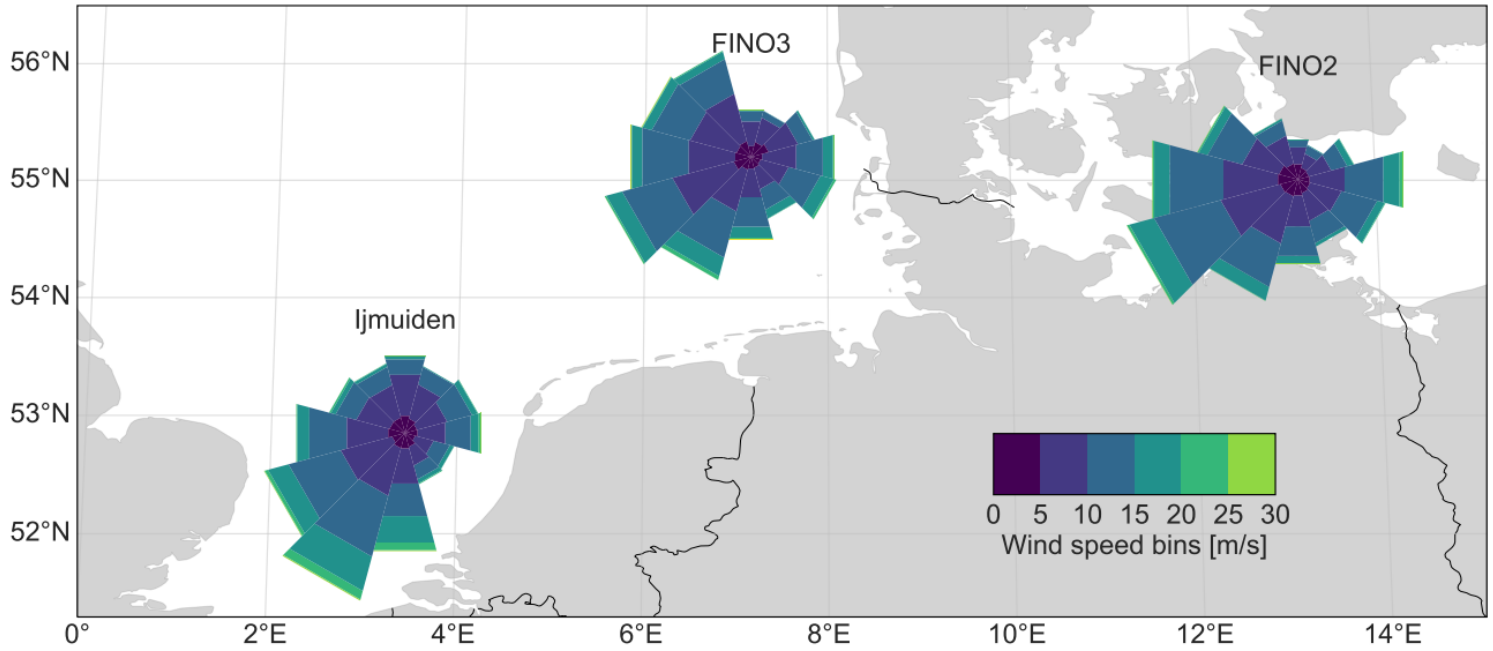
Since training data is the same…

Is gap filling giving us any additional information?

Data base

1-hour wind speed and direction data at 100m

- **„Measurement campaign“ data from FINO2, FINO3 und Ijmuiden:**
  - 2 years (2012 – 2014)
  - Without gaps
  - No nearby wind farms
- **ERA5 reference data**



M. Jonietz et al., “Understanding the impact of data gaps on long-term offshore wind resource estimates,” *Wind En. Sc*, 2023, 10.5194/wes-2023-127

Fraunhofer IWES

How to choose a numerical method

Train-test-split

- **K-fold validation**
  - **Split** measurement into **train** and **test** subsets
  - **Create algorithm** that generates data in **training** period
  - **Apply algorithm** to generate data in **test** period
  - **Compare** generated data **with actual measurement** in test subset (RMSE, R^2, Distribution…)
  - **Repeat** for other test subsets and **average** score over all test subset results

- **Train-test splitting strategies:**

Fraunhofer IWES

Do we have a winner?

Which algorithm gives the best k-fold validation results?

- **Tested methods:**
  - KNN (K=3)
  - KNN (K=300)
  - linear interpolation
- **Tested statistics:**
  - Mean wind speed
  - Mean wind direction
  - Wind speed distributions
- **<u>Optimal algorithm depends on test subset distribution</u>**

Random split

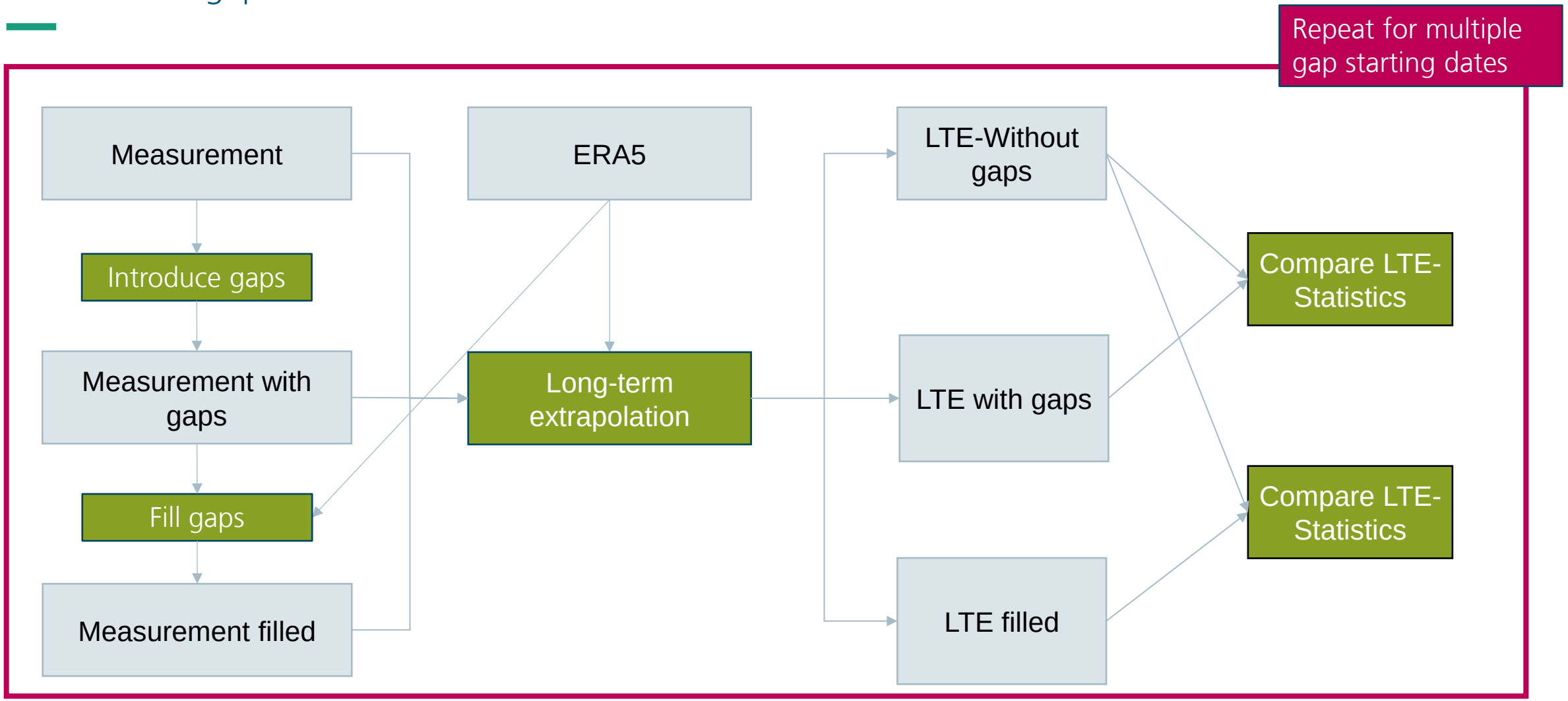
	$\overline{WS}$	$\overline{Dir}$	WS $Distribution$
FINO3	KNN (K = 3)	KNN (K = 3)	KNN (K = 3)
FINO2	KNN (K = 3)	KNN (K = 3)	KNN (K = 3)
Ijmuiden	KNN (K = 3)	KNN (K = 3)	KNN (K = 3)

Coherent split

	$\overline{WS}$	$\overline{Dir}$	WS $Distribution$
FINO3	KNN (K = 300)	KNN (K = 300)	KNN (K = 300)
FINO2	KNN (K = 300)	linear	linear
Ijmuiden	linear	KNN (K = 300)	linear

Evaluating gap filling effect on the long-term extrapolation

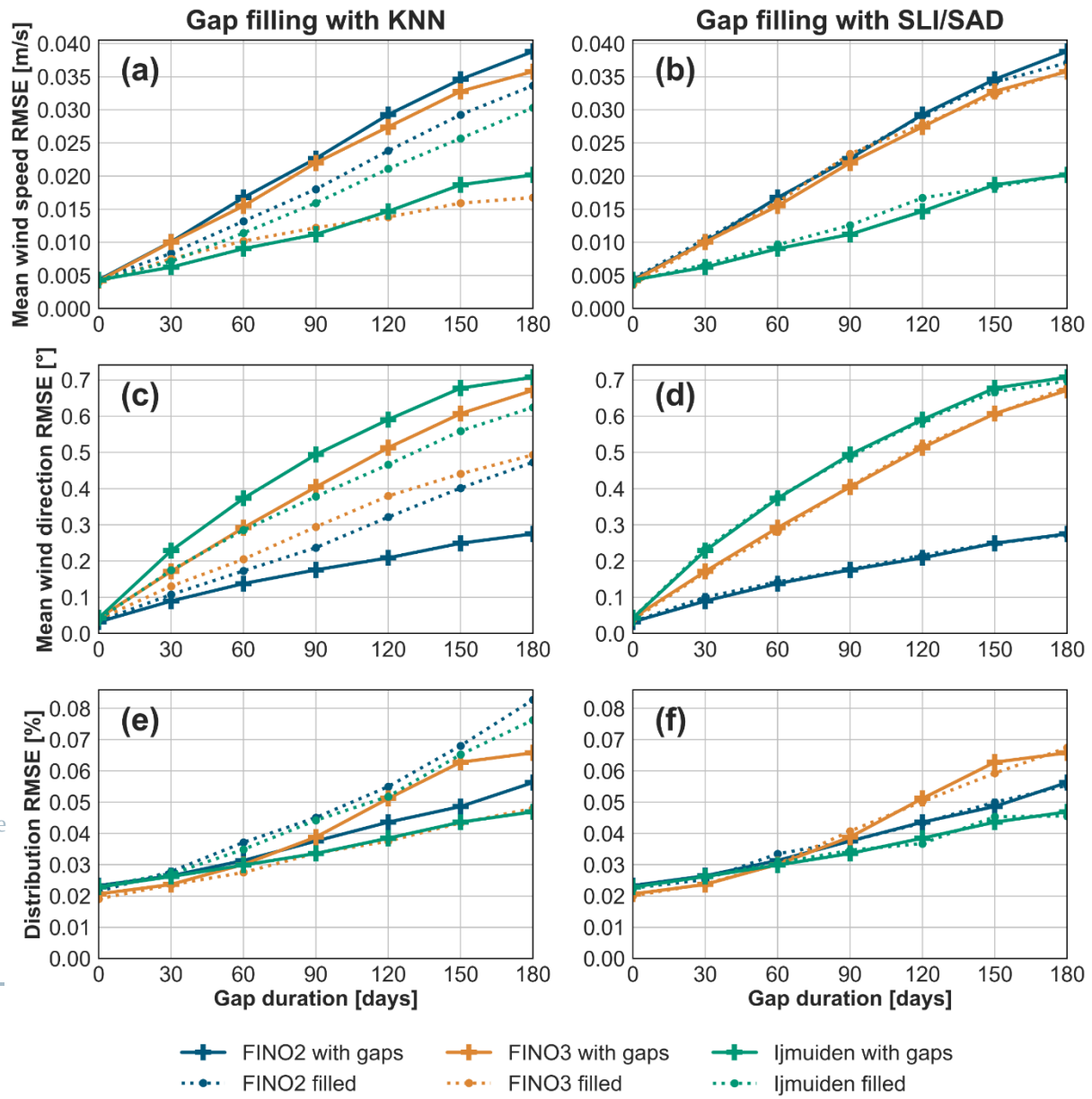
For coherent gaps



Fraunhofer IWES

Are long-term extrapolations affected by gap filling?

- No effect when filling with linear interpolation
- KNN- filling reduces gap effect in **"some"** cases
- **<u>Filling and extrapolating with the same reference data has no advantage</u>**



M. Jonietz et al., "Understanding the impact of data gaps on long-term offshore wind resource estimates," *Wind En. Sc*, 2023, 10.5194/wes-2023-127

References

- **M. Jonietz et al., “Understanding the impact of data gaps on long-term offshore wind resource estimates,” Wind En. Sc, 2023, https://doi.org/10.5194/wes-2023-127**

- S. Schwegmann et al., “Enabling Virtual Met Masts for wind energy applications through machine learning-methods,” En. And AI, 2023, https://doi.org/10.1016/j.egyai.2022.100209

- J. Gottschall et al., “Understanding the impact of data gaps on offshore wind resource estimates,” Wind En. Sc, 2021, https://doi.org/10.5194/wes-6-505-2021

© Fraunhofer IWES/Frank Bauer

Thank you for your attention

Fraunhofer IWES

Fraunhofer IWES

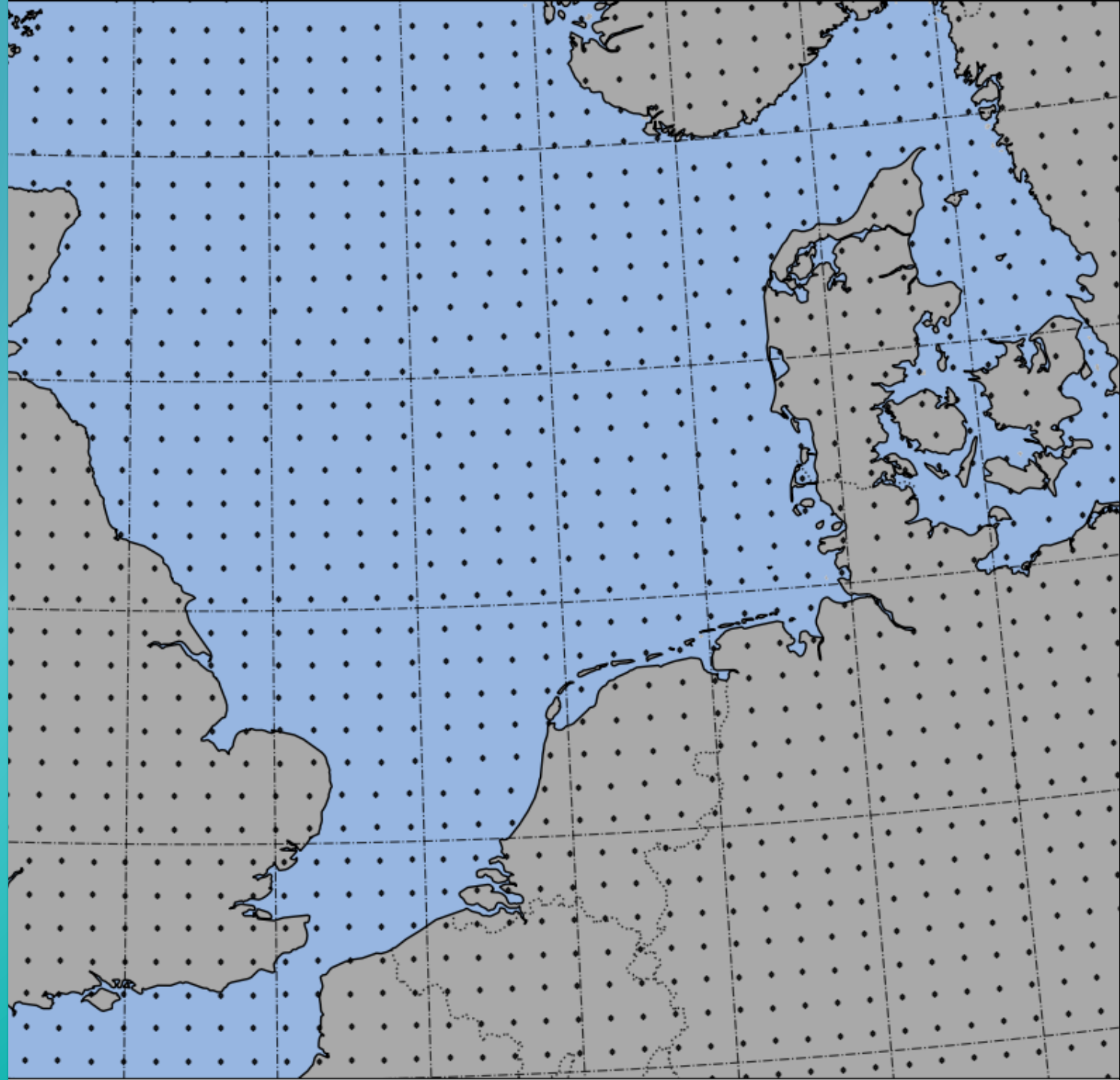
10.10.2024 / RAVE ML Workshop

Localized Wind Profile Predictions via a Machine Learning Approach

Farkhondeh (Hanie) Rouholahnejad, Julia Gottschall

Offshore wind profile estimation is key for site assessment.

How can machine learning mitigate the discrepancies caused by the large grid cells of ERA5?



- Dataset
- Method: Random forest-based models
- Results: Model performance
- Discussion: Model error interpretation
- Summary

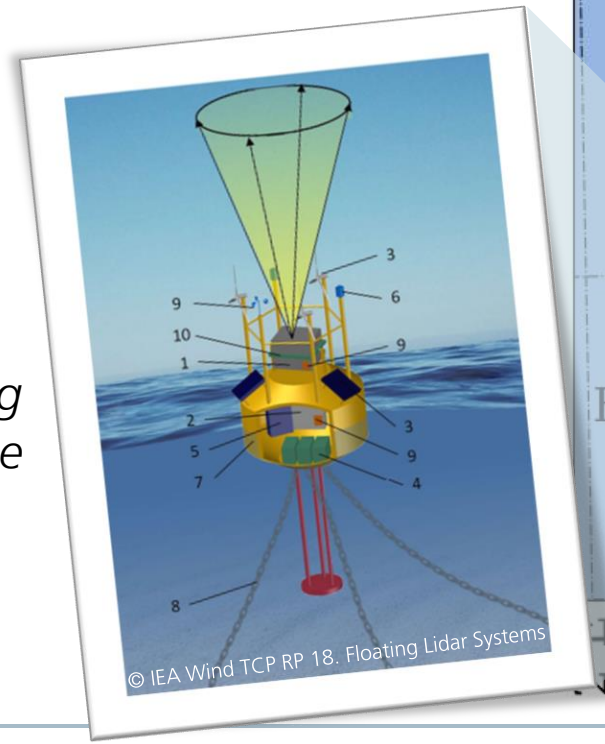
Data for Model Training and Validation

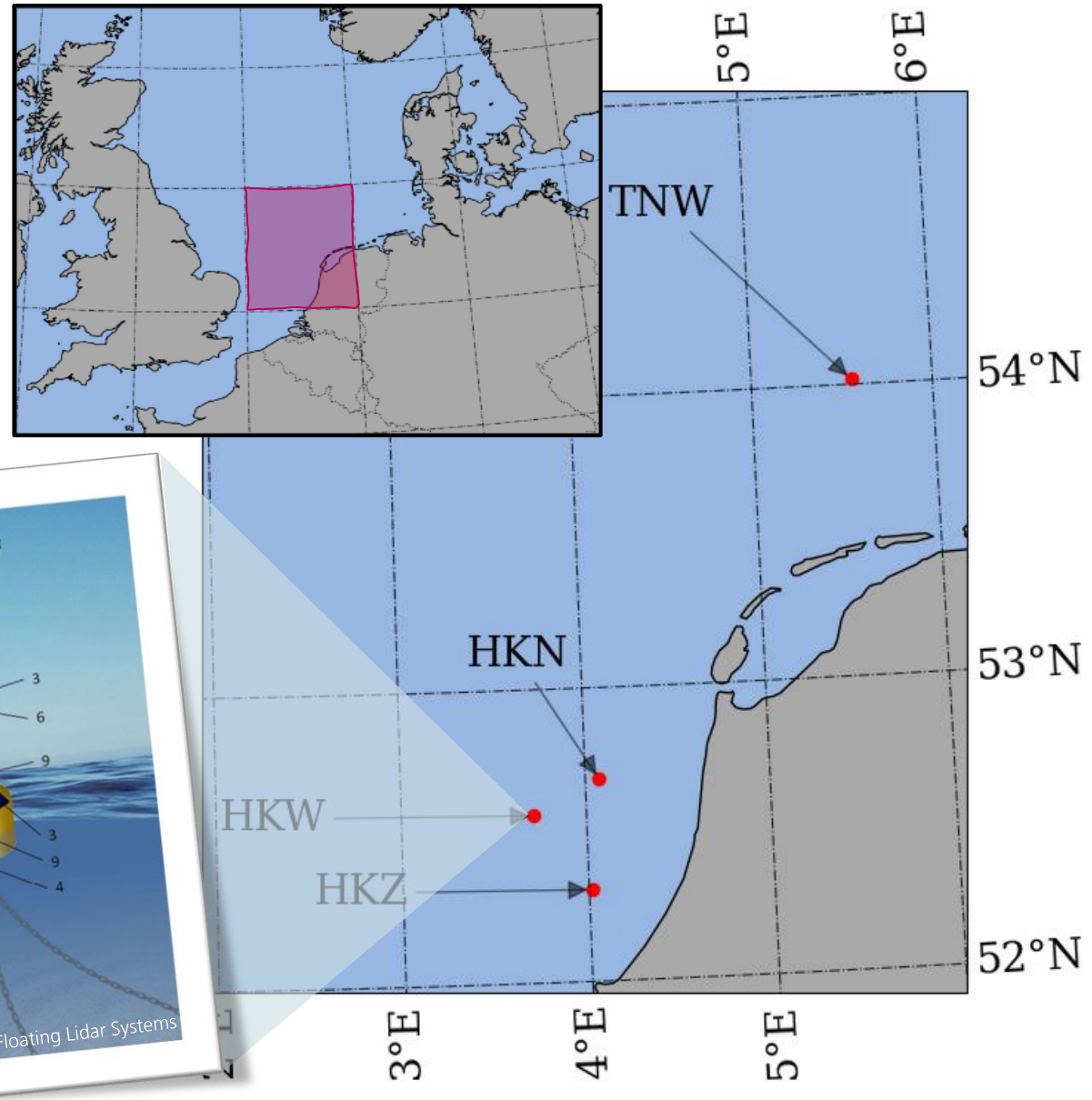
- Four locations within a 200 km wide region in the Dutch part of the North Sea
- Two-year floating lidar for each location.

⚠ WARNING

TERMINOLOGY ALERT!

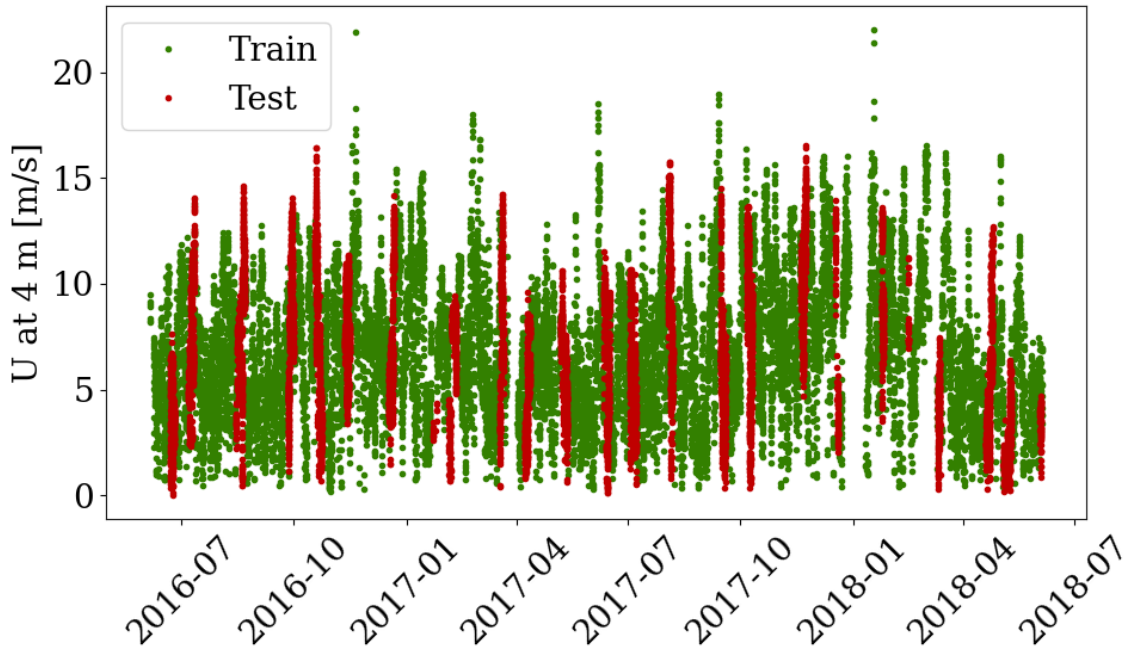
- ***Same site validation*** *is when training and validation subsets are from the same campaign.*
- ***Round robin validation*** *is otherwise.*



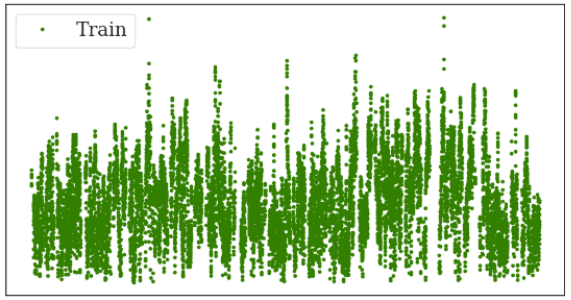


Fraunhofer IWES

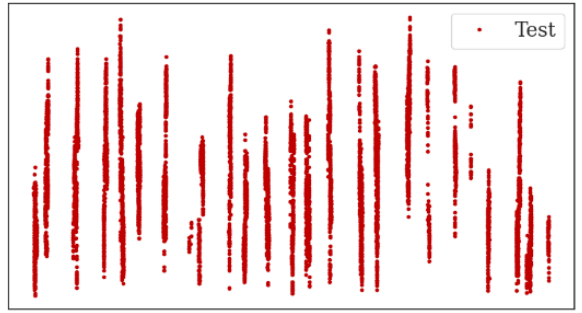
Test – Train Split



Train



Test

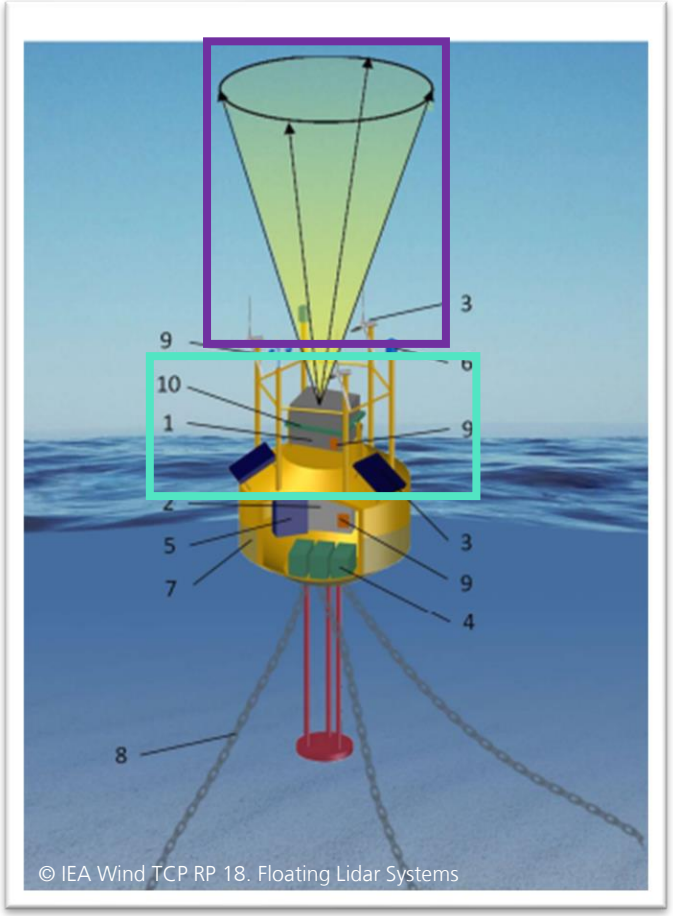


Fraunhofer IWES

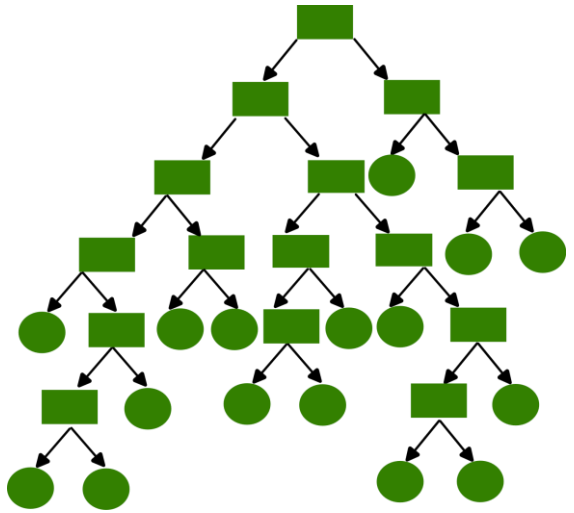
Model Input - Output

10 min Output

10 min Input



U_4m,
Pressure,
Temperature,
RH, SST



Wind Profile

Fraunhofer IWES

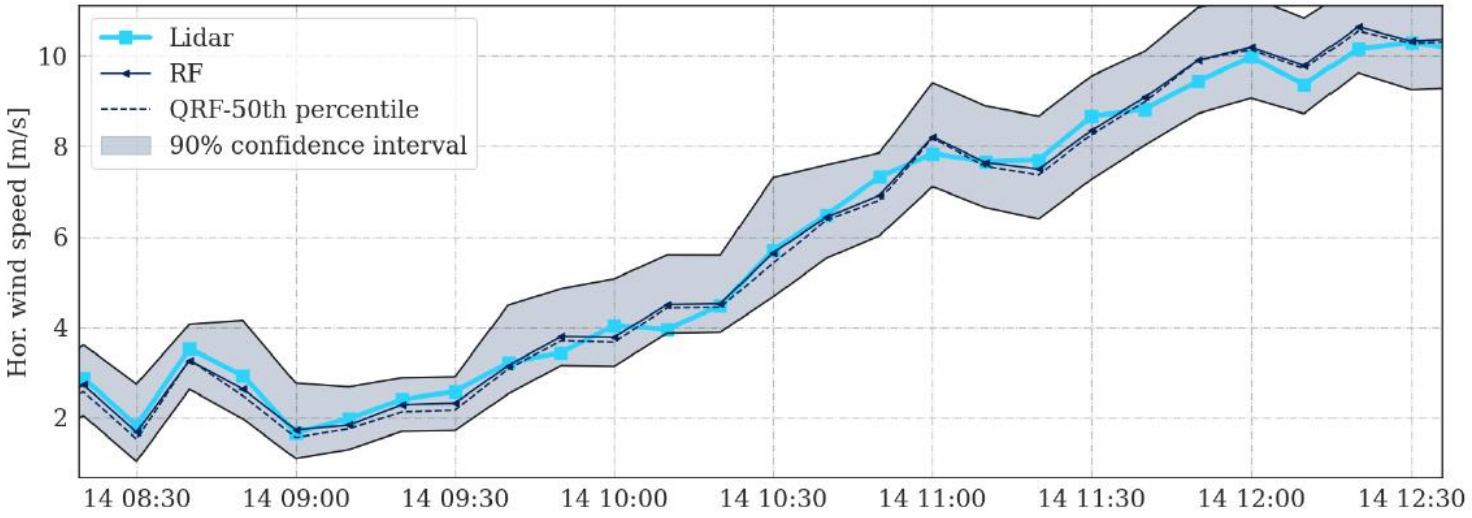
Random Forest and Quantile Regression

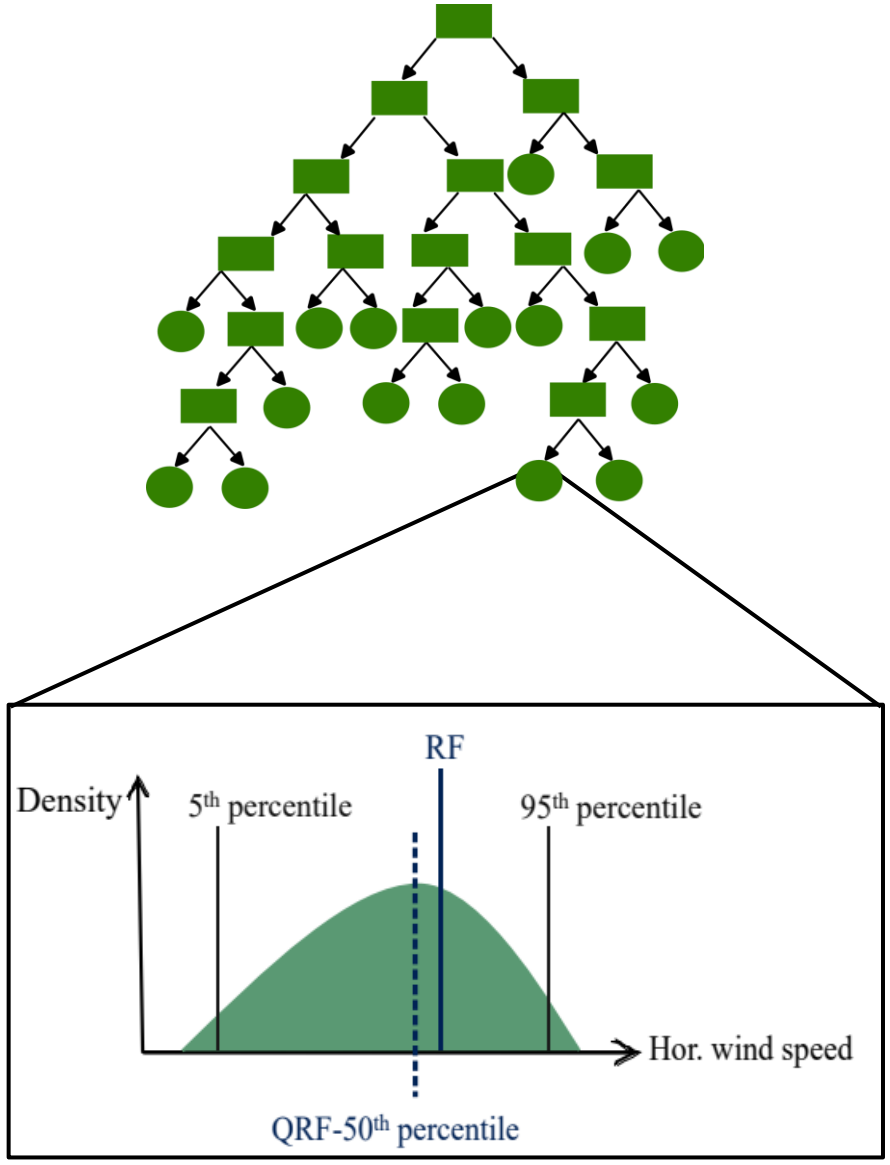
```
from sklearn.ensemble import RandomForestRegressor
```

Only the average of the samples in each leaf is stored

```
from sklearn_quantile import RandomForestQuantileRegressor
```

The sample distribution at each leaf is stored: any statistical parameter can be derived.

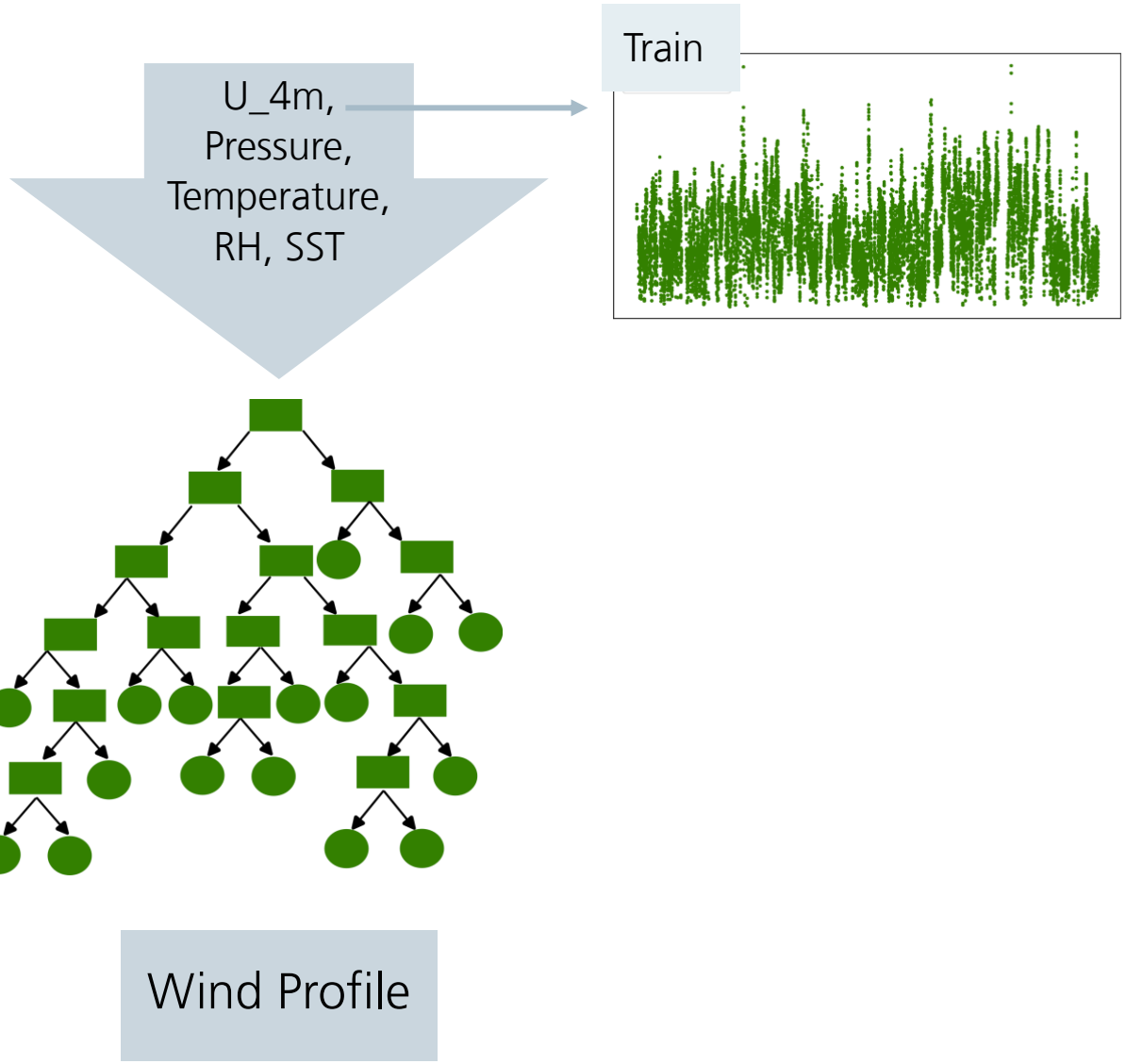




Fraunhofer IWES

Model Training

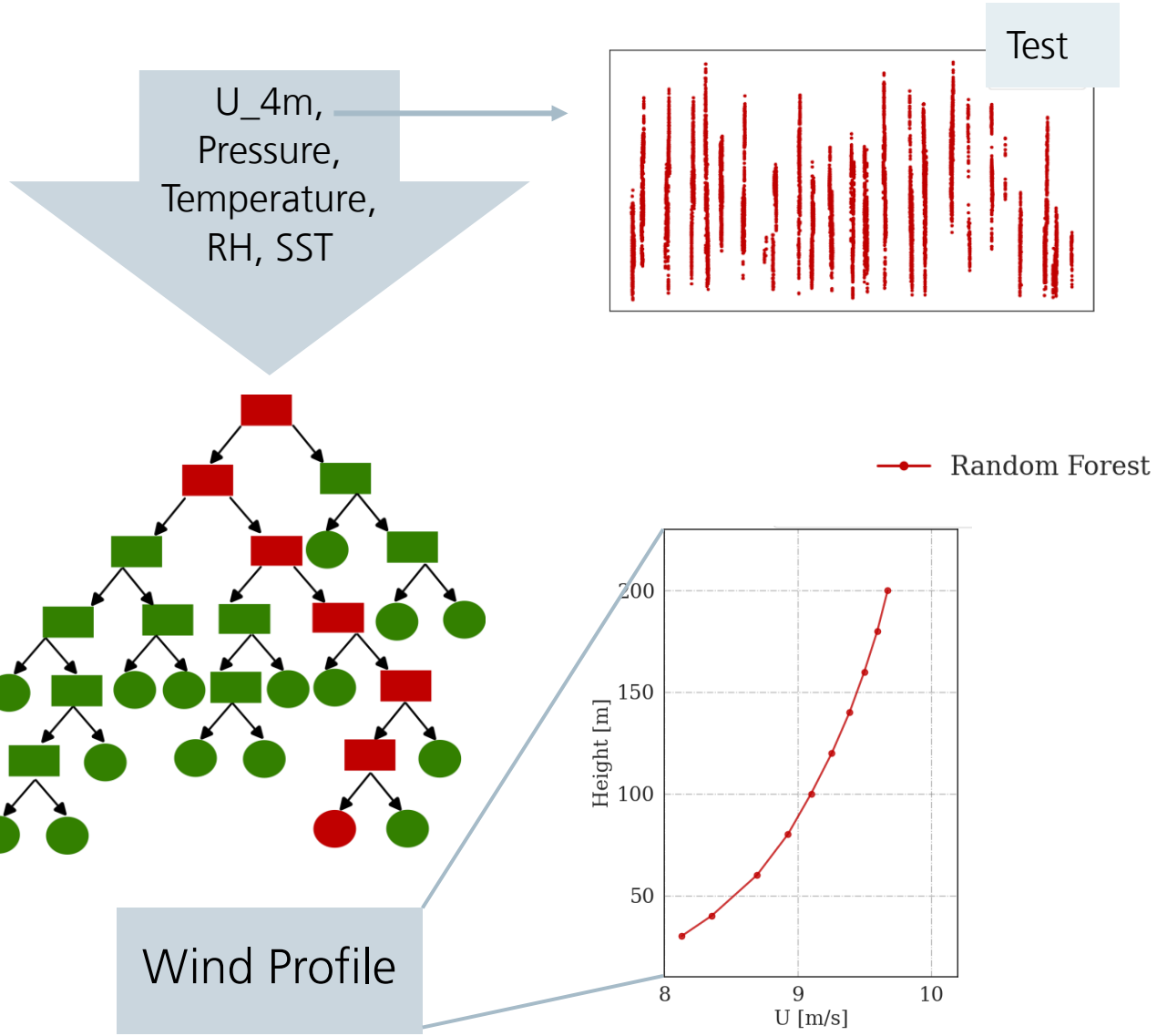
- Train the random forest model using 85% of the data.



Fraunhofer IWES

Model Testing

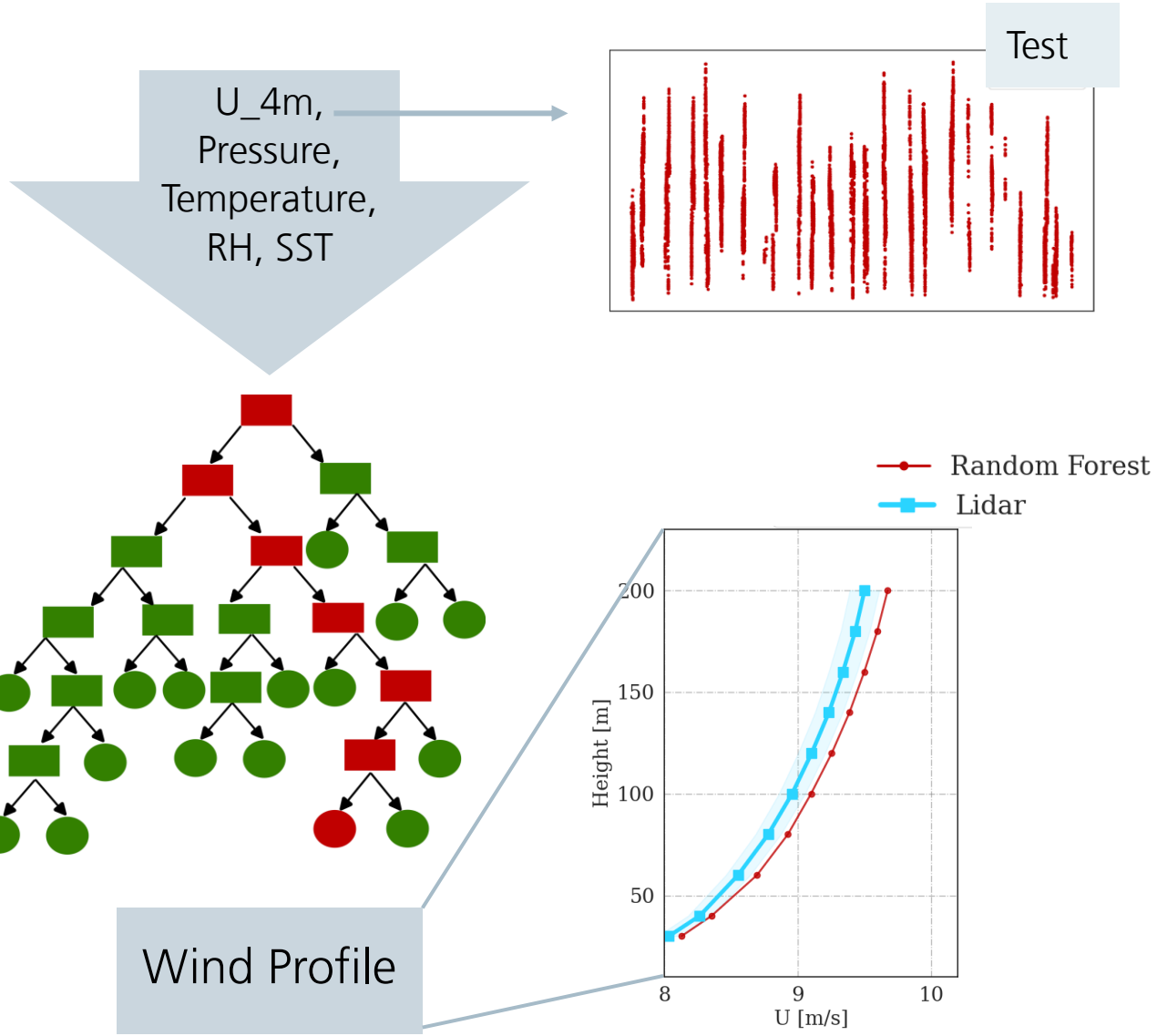
- Train the random forest model using 85% of the data.
- Validate the model on the remaining 15% based on same site and round robin approaches.



Fraunhofer IWES

Model Testing

- Train the random forest model using 85% of the data.
- Validate the model on the remaining 15% based on same site and round robin approaches.

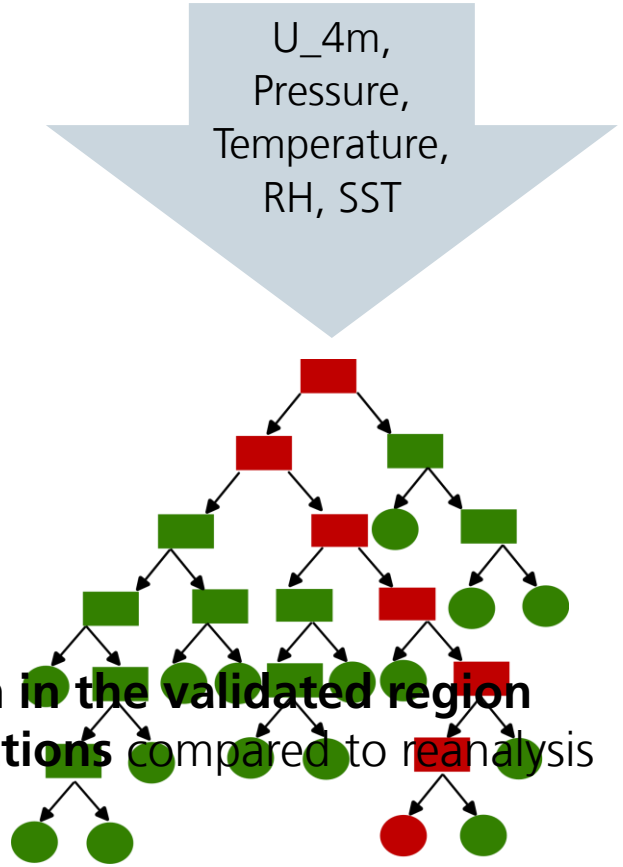


Fraunhofer IWES

Why to develop such a model?

contingent upon near-surface data availability

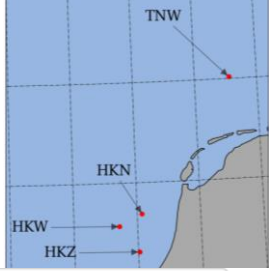
1. Fill the lidar **data gaps**
2. **Spatial extrapolation in the validated region**
3. More **localized predictions** compared to reanalysis datasets



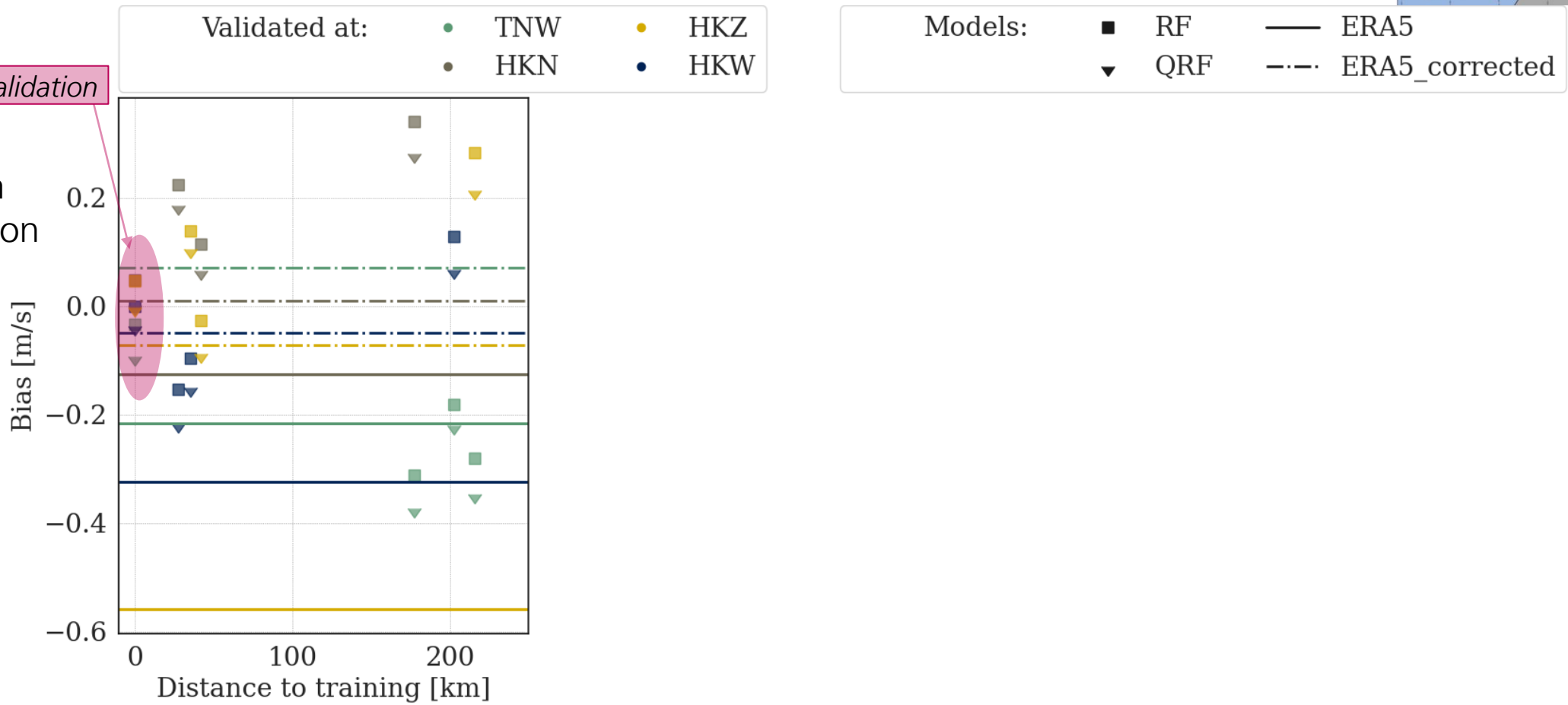
Fraunhofer IWES

How does the model perform?

Accuracy drops with distance to the training site.



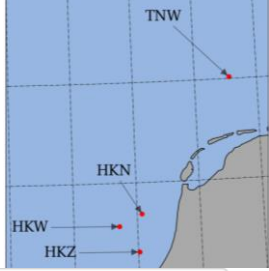
- The distance from the training location impacts the **bias**.



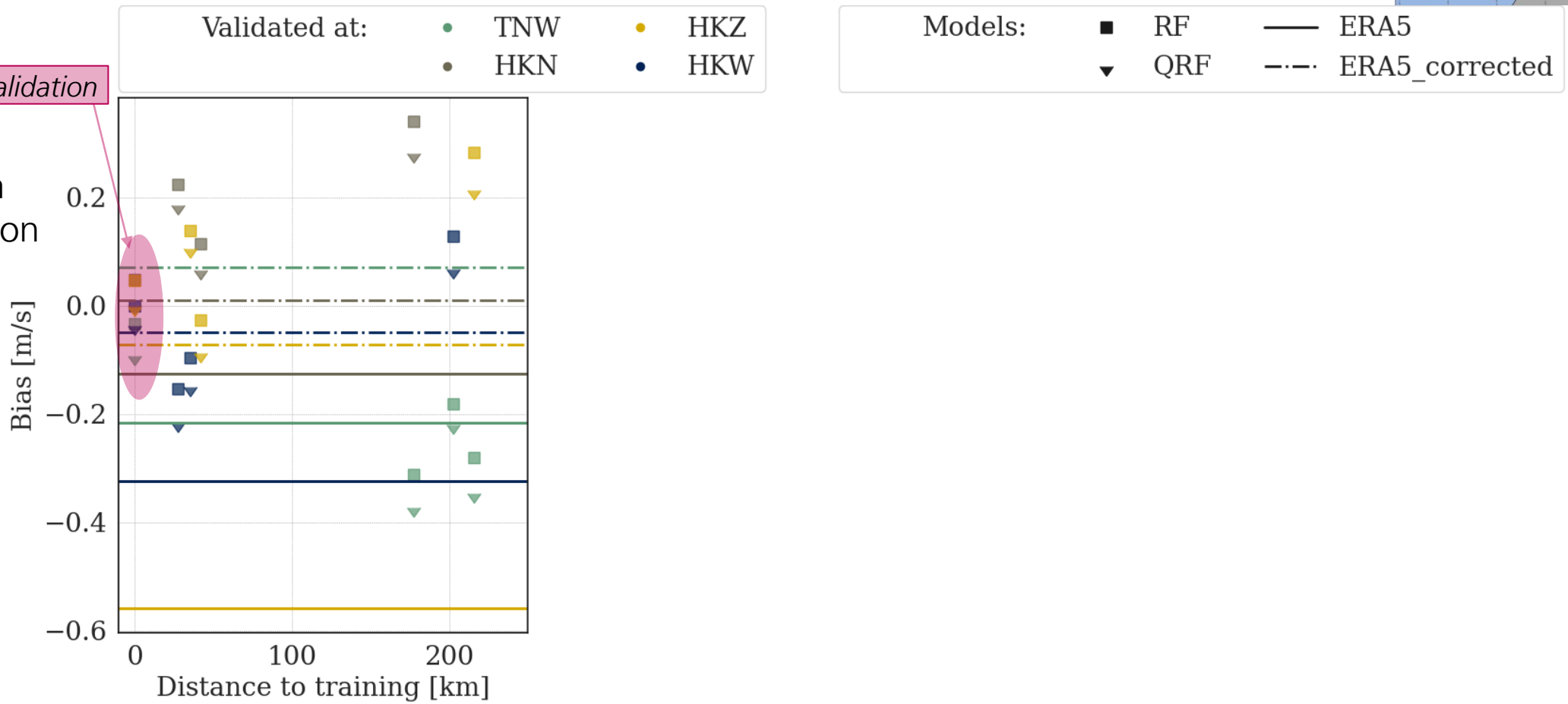
Error dependency on the distance to training site. Horizontal lines indicate ERA5 error metrics pre- and post-correction, via an MCP using the training subset to derive correlation parameters.

Fraunhofer IWES

Accuracy drops with distance to the training site.



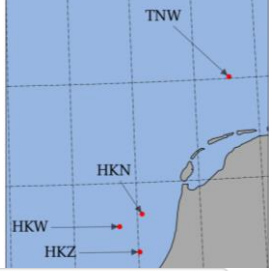
- The distance from the training location impacts the **bias**.



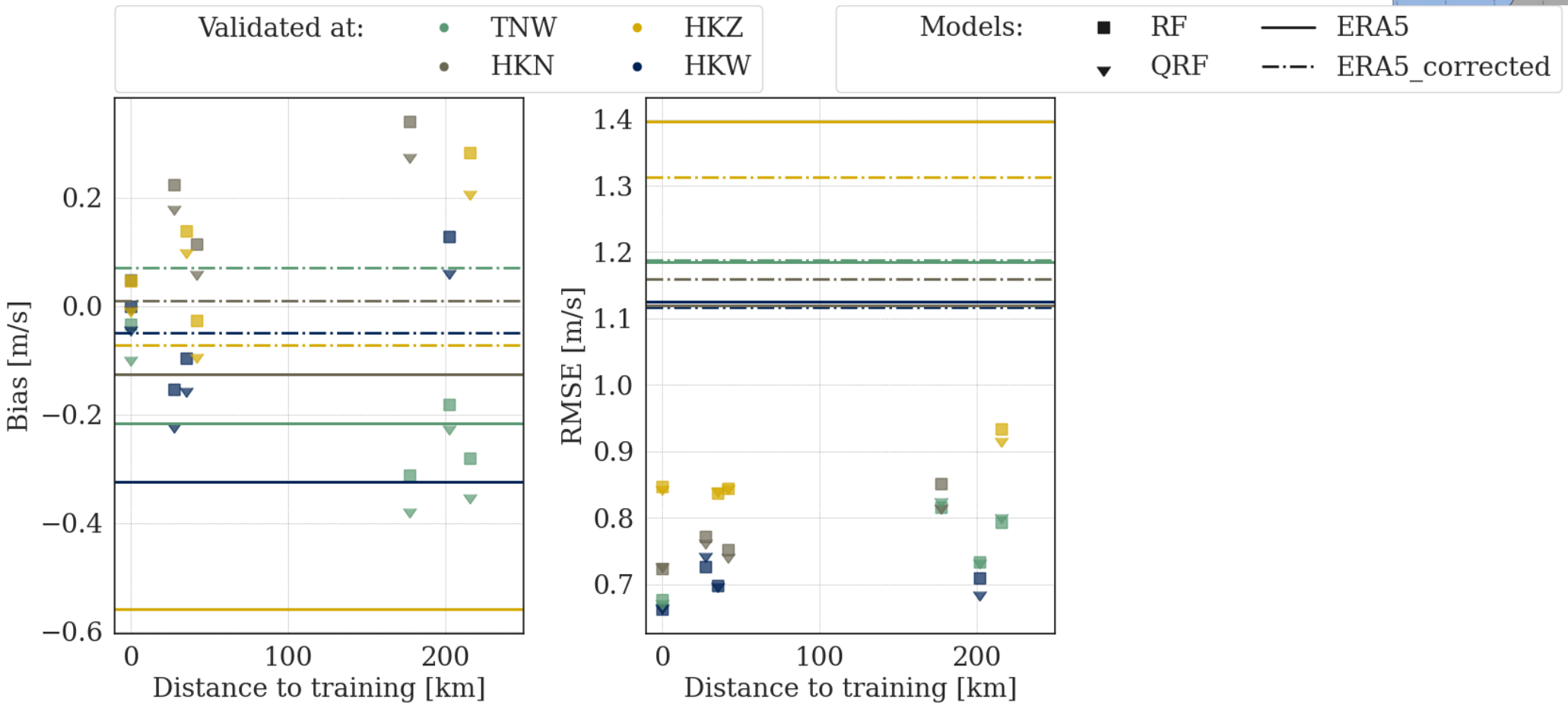
Error dependency on the distance to training site. Horizontal lines indicate ERA5 error metrics pre- and post-correction, via an MCP using the training subset to derive correlation parameters.

Fraunhofer IWES

Accuracy drops with distance to the training site.

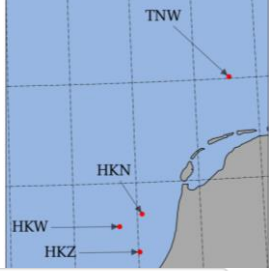


- The **RMSE** grows with distance to training site, but always lies below ERA5.

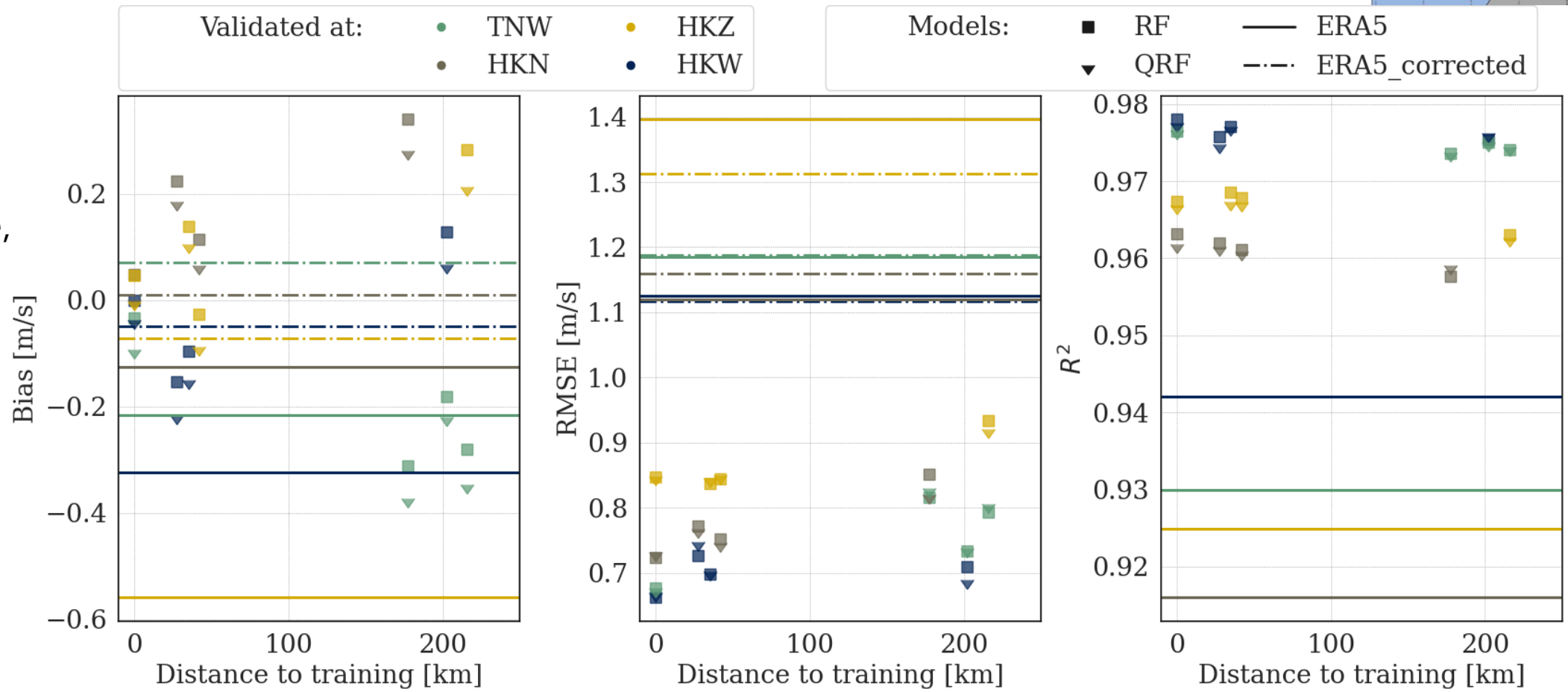


Error dependency on the distance to training site. Horizontal lines indicate ERA5 error metrics pre- and post-correction, via an MCP using the training subset to derive correlation parameters.

Accuracy drops with distance to the training site.



- **Regardless of the training site**, random forest models show better correlation with the observations.



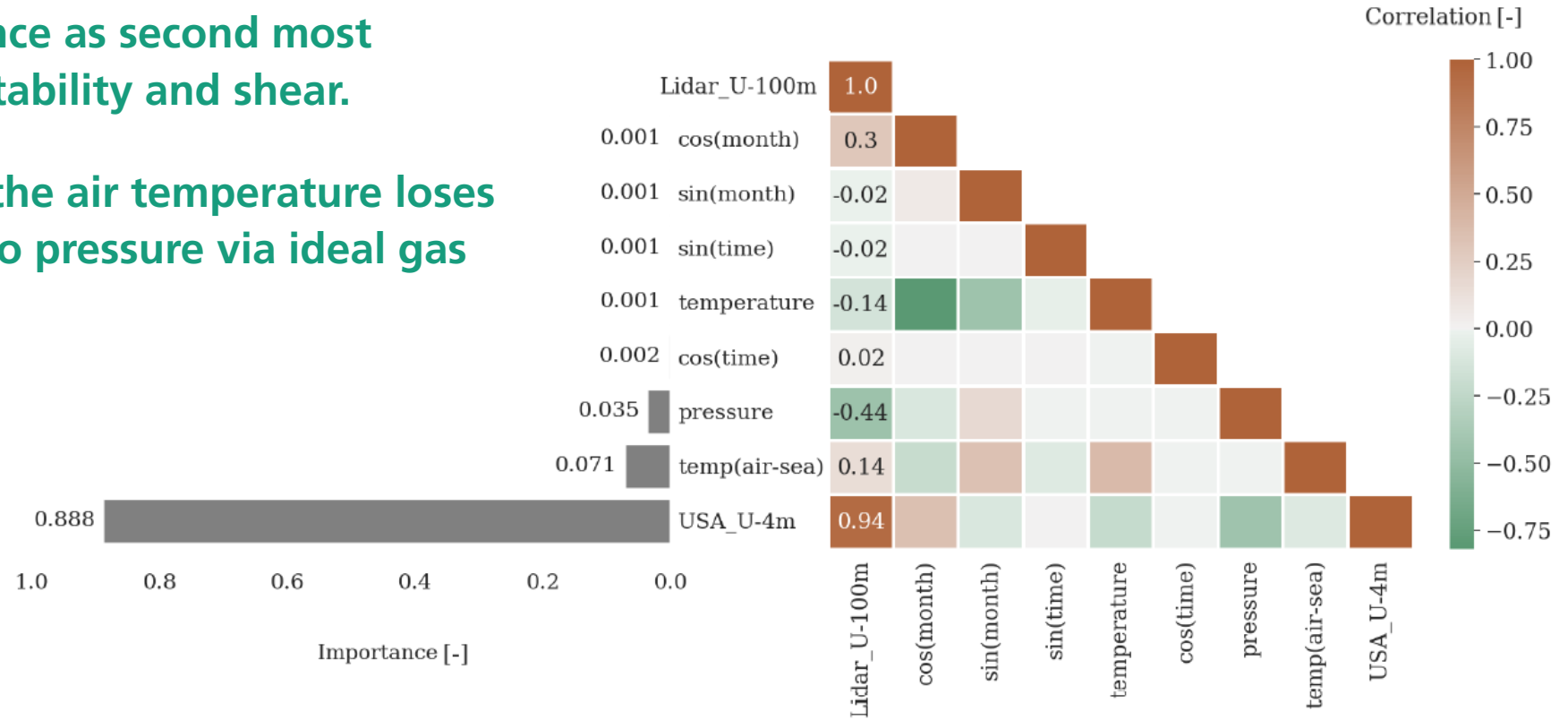
Error dependency on the distance to training site. Horizontal lines indicate ERA5 error metrics pre- and post-correction, via an MCP using the training subset to derive correlation parameters.

Fraunhofer IWES

Can the ML model capture the physics?

Feature importance

- Near surface wind speed is the most important feature.
- Air-sea temperature difference as second most important: key deriver for stability and shear.
- In presence of air pressure, the air temperature loses importance, as it is related to pressure via ideal gas law.

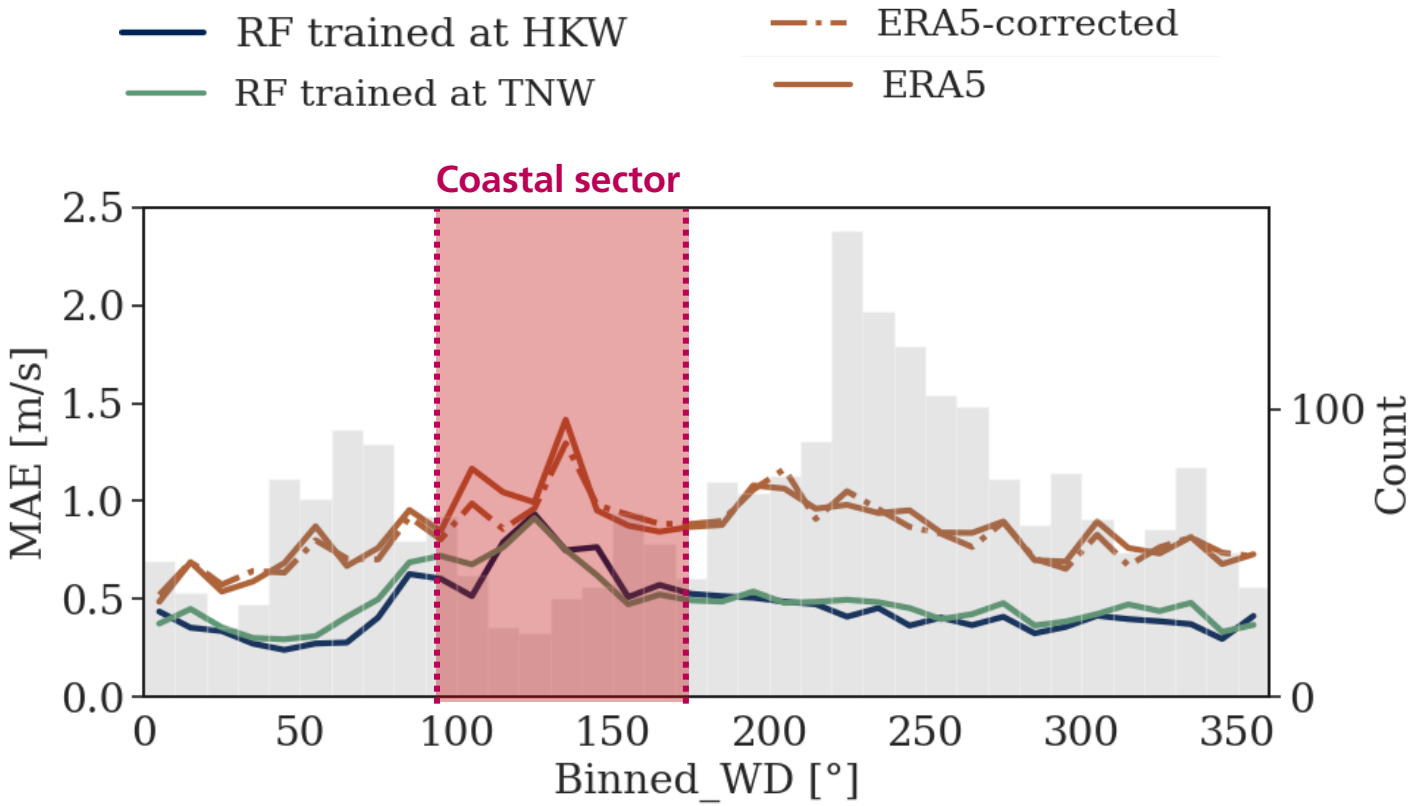


Fraunhofer IWES

What can the model error be attributed to?

The Mean Absolute Error peaks at the coastal wind sector.

- Both the RF model and ERA5 exhibit **higher Mean Absolute Error** (MAE) for the wind originating from the **coastal region.**
- Random forest is more accurate than of ERA5 in all wind sectors, including the coastal sector.

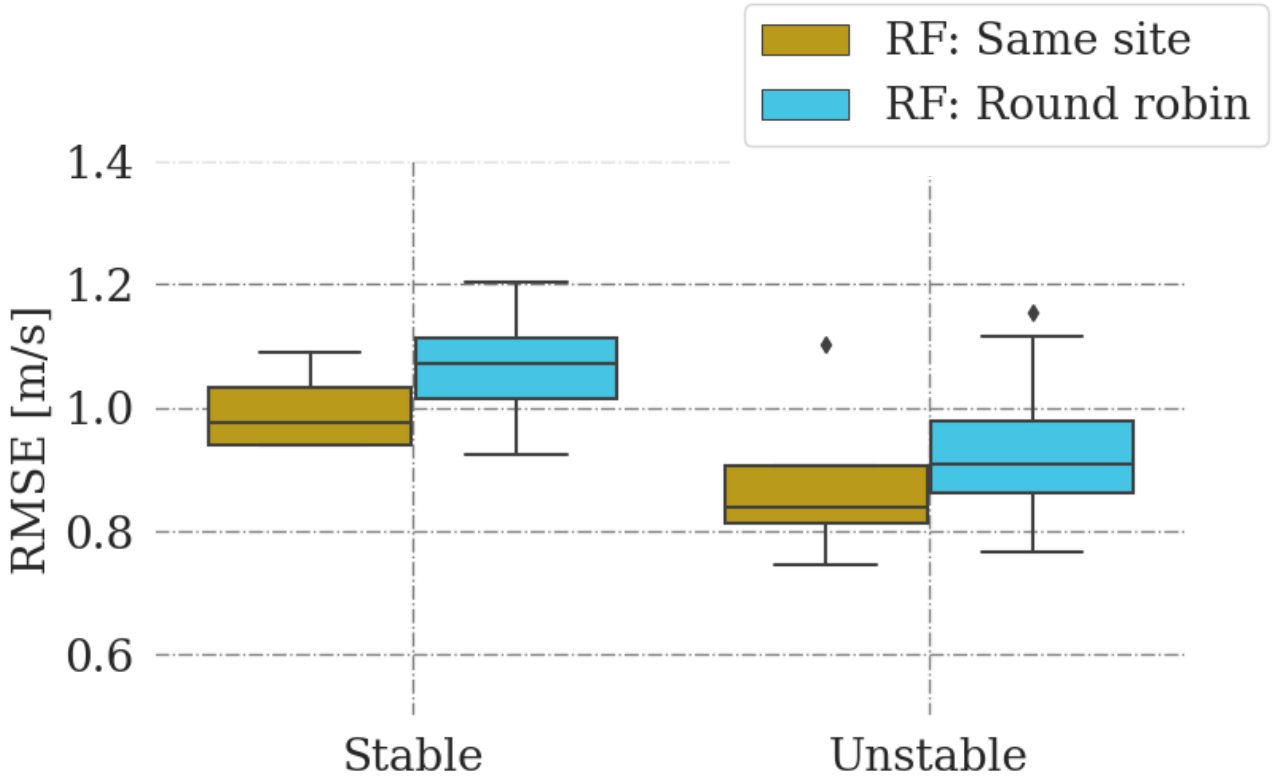


Error dependency on wind direction at HKW at 100 m. The bin counts are depicted in gray, with the coastal sector highlighted in red.

Fraunhofer IWES

The Random Forest model shows reduced accuracy for stable conditions.

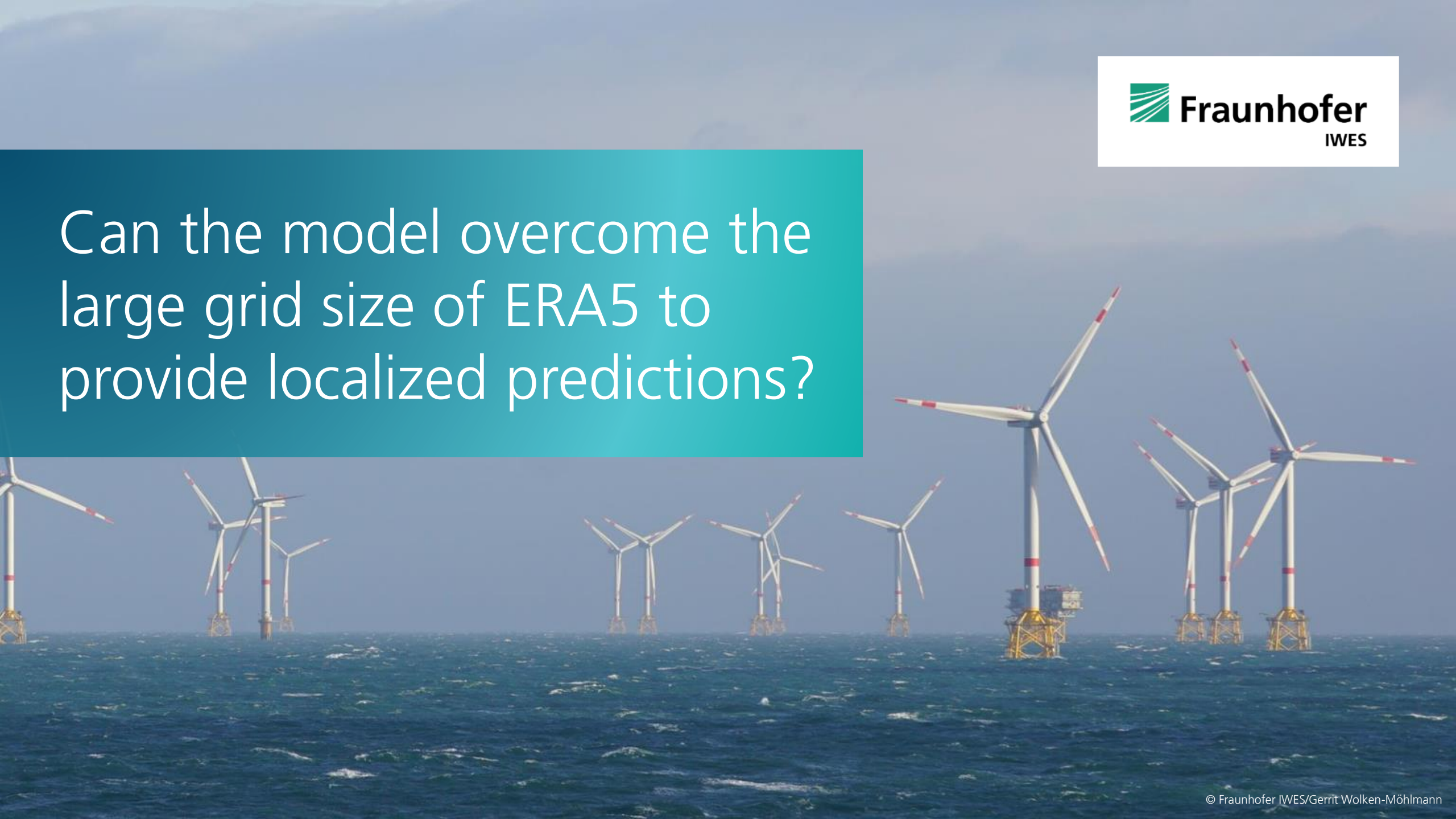
- Random forest-based models are less accurate when the wind aloft is **decoupled from the surface** (stable conditions).



Box plot depicting the average RMSE for stable and unstable conditions across all heights, with variations attributed to different locations.

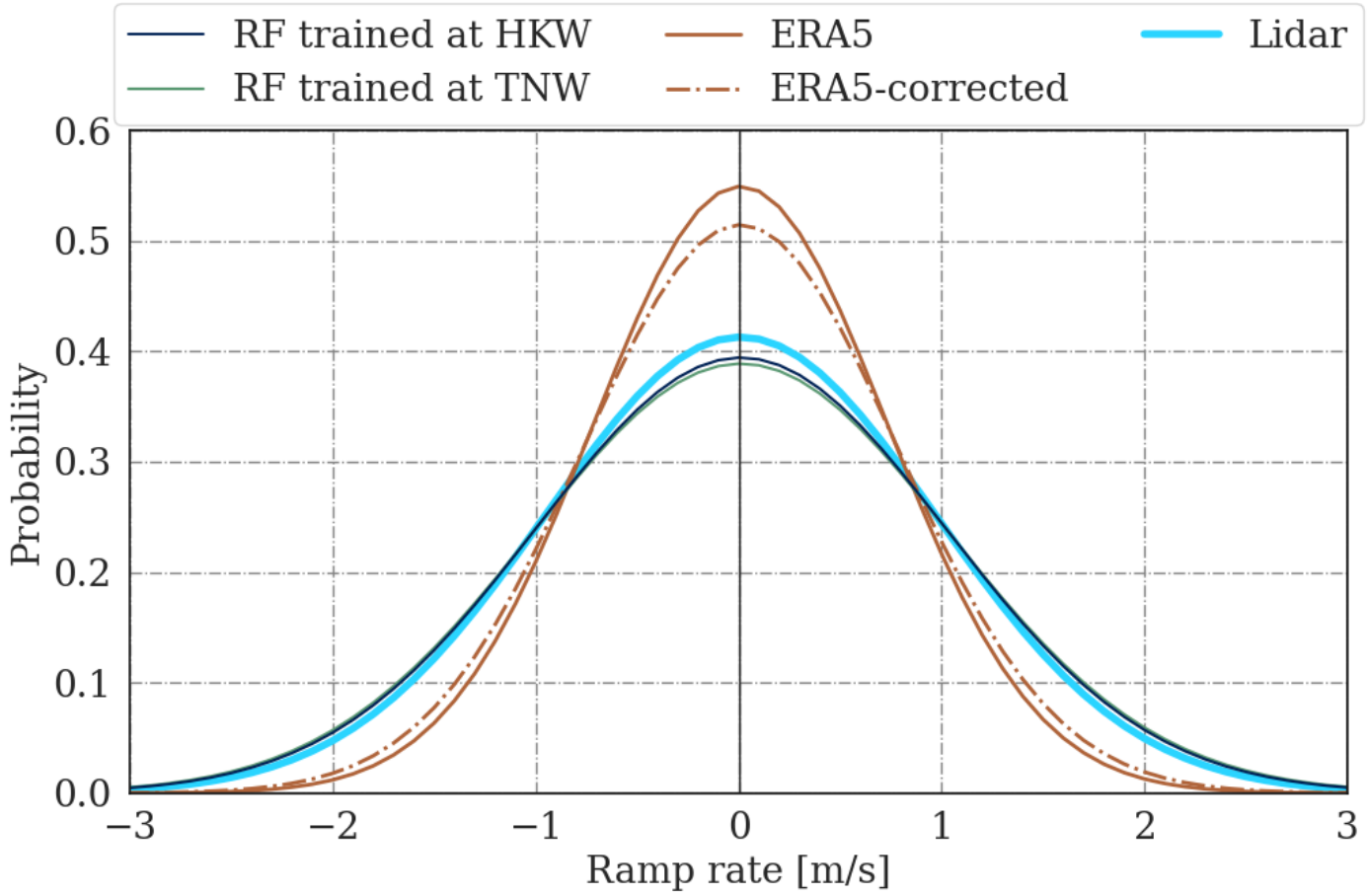
Fraunhofer IWES

Can the model overcome the large grid size of ERA5 to provide localized predictions?



The Random Forest model captures the wind speed variability more accurately than ERA5.

- ERA5 consistently **underestimates** wind speed variability, exhibiting a deviation of 22-30% from observed hourly ramp rates, which reduces to 16-27% after MCP correction.
- Random Forest predicts **more localized** wind profiles, demonstrating a deviation of 2-9%.

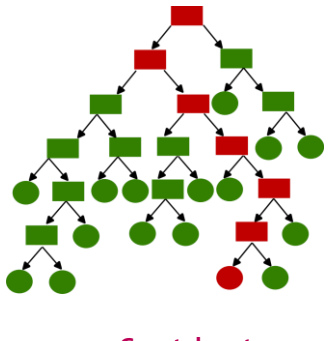


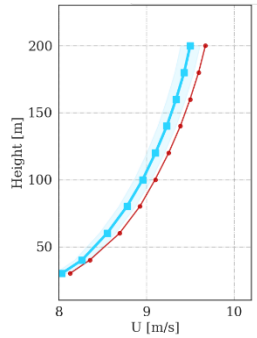
Absolute ramp rate distribution at HKW at 100 m for ERA5 before and after correction, and random forest for same site and round robin approaches.

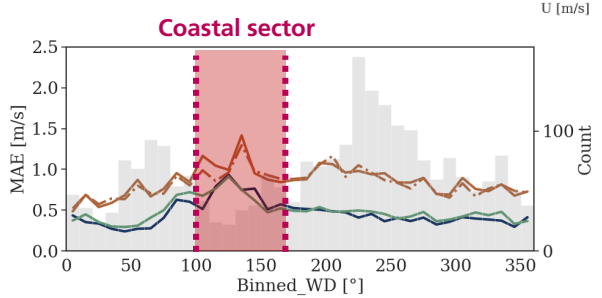
Fraunhofer IWES

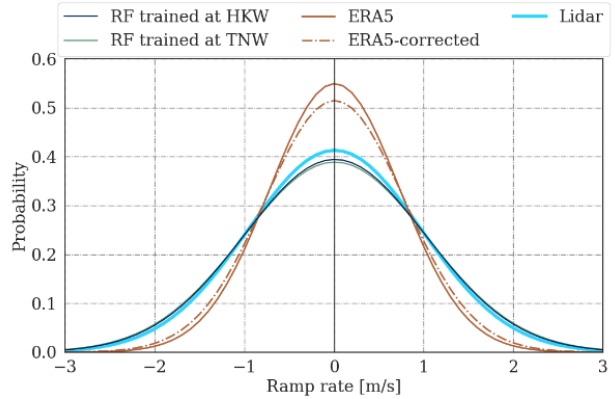
Summary and Conclusions

- A Random Forest model was validated in the North Sea in a 200 km wide region.
- Both the random forest model and ERA5 face challenges to model the wind originating from the **coastline**.
- The ERA5 underestimation of the **wind speed variability**, due to the large grid size, can be mitigated through the random forest model.
- Application: Lidar data gap filling vertical and horizontal extrapolation of wind profile
- Short coming: Shown results for free inflow – waked sectors are filtered out









Fraunhofer IWES

References

- **Rouholahnejad, F.**, Gottschall, J.: Characterization of Local Wind Profiles: A Random Forest Approach for Enhanced Wind Profile Extrapolation, Wind Energ. Sci. Discuss. [preprint], https://wes.copernicus.org/preprints/wes-2023-178, in review, 2023.
- Bodini, N. and Optis, M.: How accurate is a machine learning-based wind speed extrapolation under a round-robin approach?, Journal of Physics: Conference Series, 1618, 062 037, https://doi.org/10.1088/1742-6596/1618/6/062037, 2020a.
- Bodini, N. and Optis, M.: The importance of round-robin validation when assessing machine-learning-based vertical extrapolation of wind speeds, Wind Energy Science, 5, 489–501, https://doi.org/10.5194/wes-5-489-2020, 2020b.

© Fraunhofer IWES/Frank Bauer

Thank you for your time!

Fraunhofer IWES

Contact

Farkhondeh Rouholahnejad
Wind Data Analyst / Researcher
Fraunhofer Institute for Wind Energy Systems IWES
Am Seedeich 45
27572 Bremerhaven
Germany

Phone +49 471 14290-XXX
farkhondeh.rouholahnejad@iwes.fraunhofer.de

Selection of optional slides