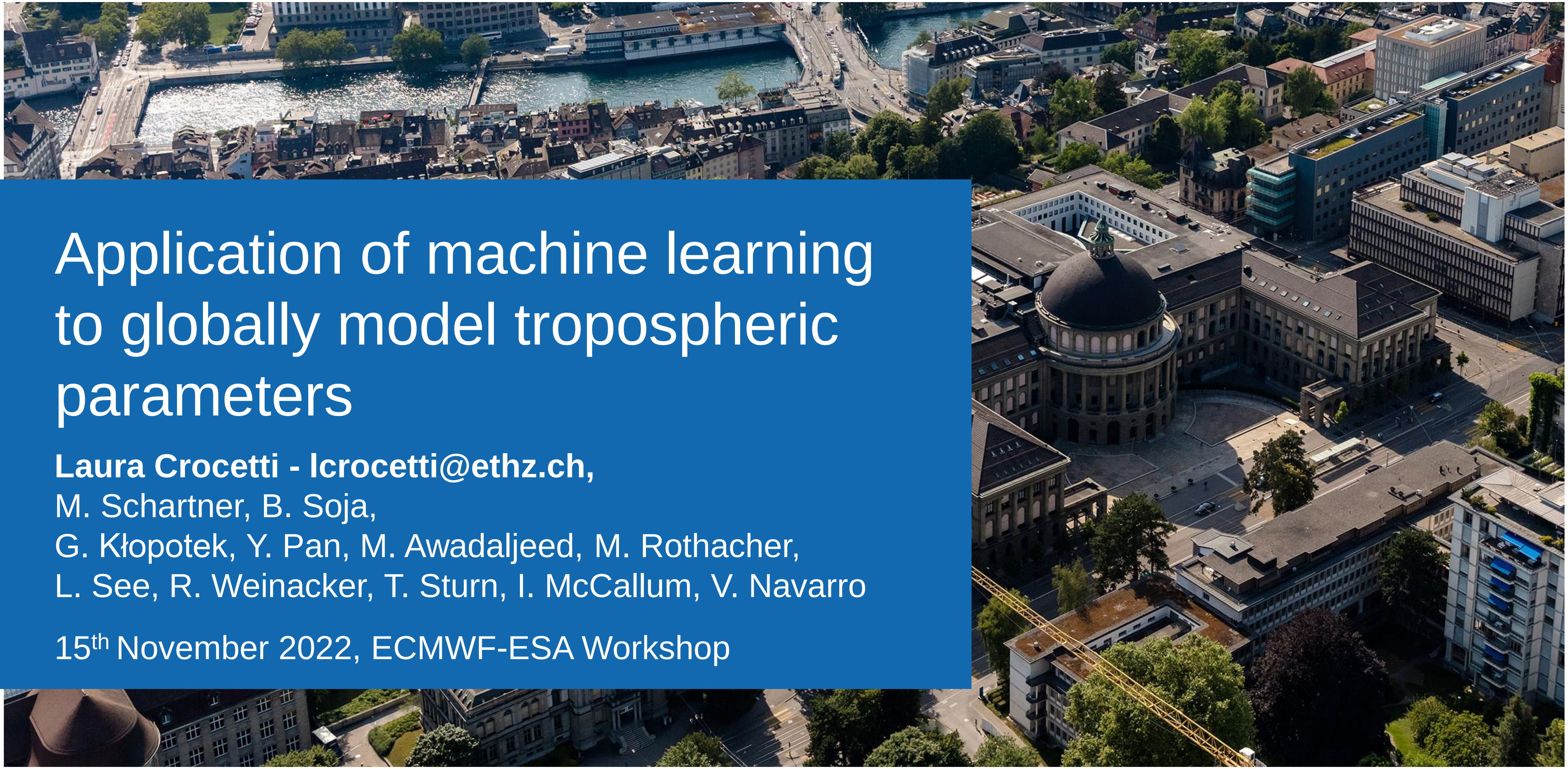


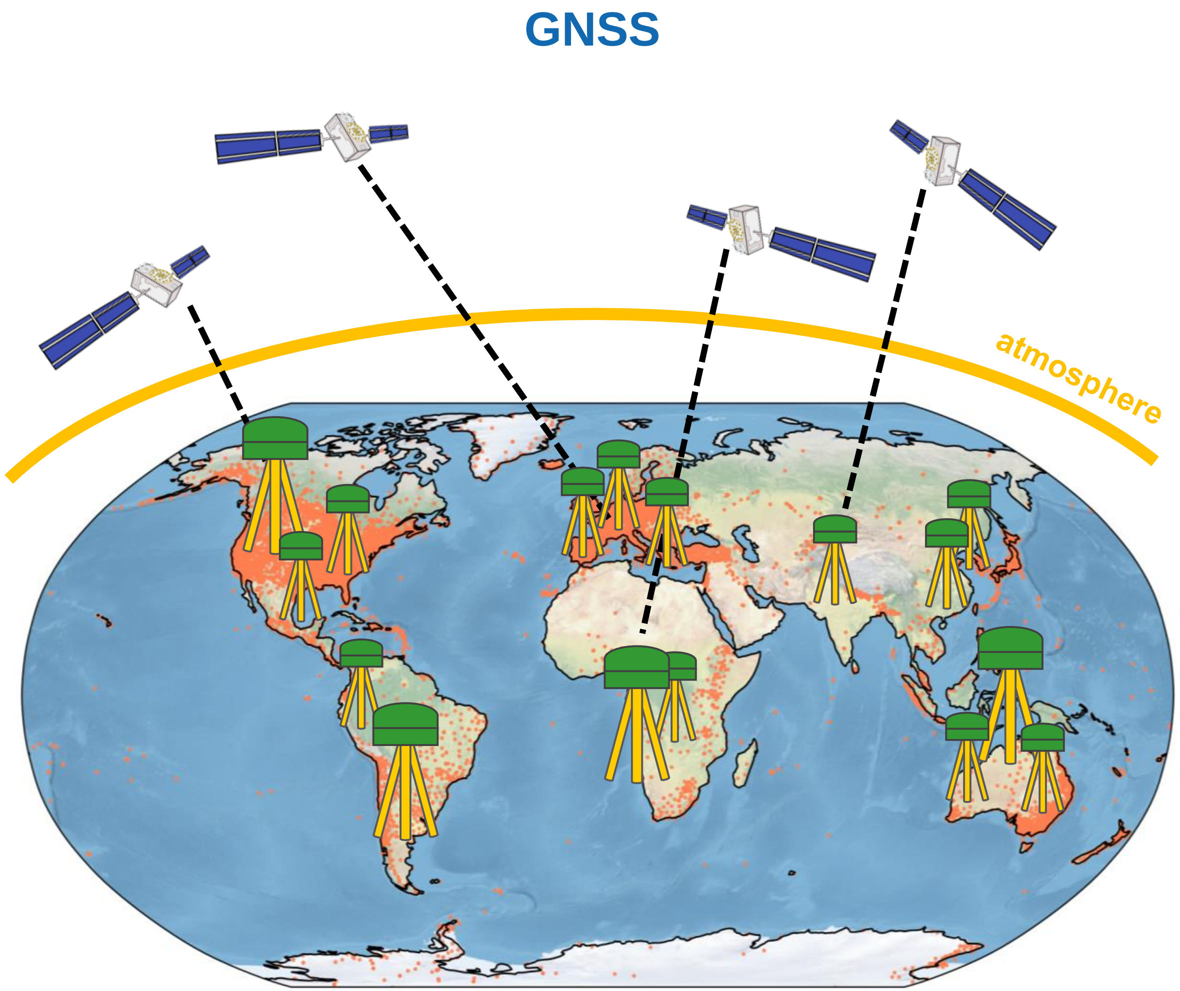
Application of machine learning to globally model tropospheric parameters

Laura Crocetti - lcrocetti@ethz.ch,
M. Schartner, B. Soja,
G. Kłopotek, Y. Pan, M. Awadaljeed, M. Rothacher,
L. See, R. Weinacker, T. Sturn, I. McCallum, V. Navarro

15th November 2022, ECMWF-ESA Workshop

Motivation

- Global Navigation Satellite System (GNSS)
  → not only used for positioning and navigation, but also for **atmospheric research**
- GNSS signals traverse the atmosphere
  → atmospheric properties can be retrieved
- Zenith hydrostatic delay
- **Zenith wet delay (ZWD)**
  → largely depends on water vapour
  → key parameter due to its relation with weather systems and climate change

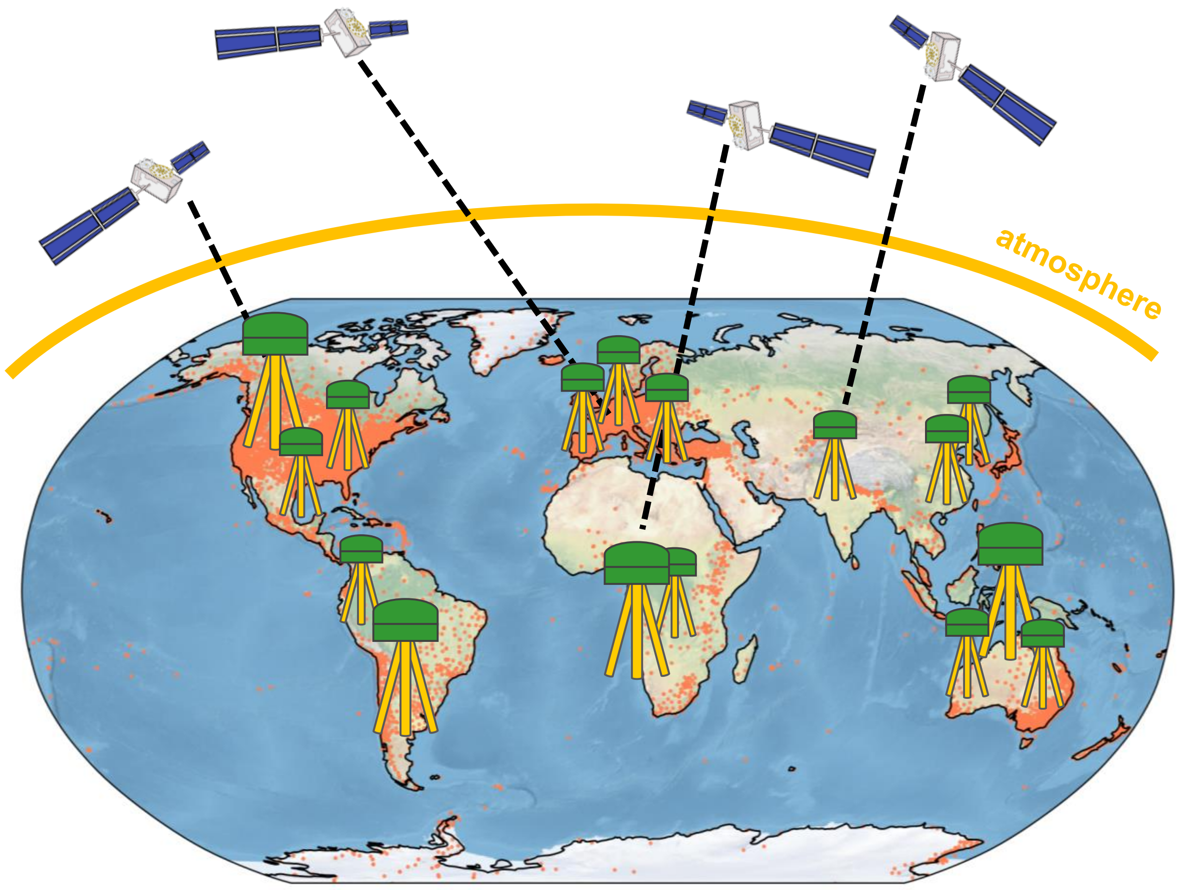


Motivation

- Global Navigation Satellite System (GNSS)
  → not only used for positioning and navigation, but also for **atmospheric research**
- GNSS signals traverse the atmosphere
  → atmospheric properties can be retrieved
- Zenith hydrostatic delay
- **Zenith wet delay (ZWD)**
  → largely depends on water vapour
  → key parameter due to its relation with weather systems and climate change

Aim: model **zenith wet delay** globally based on meteorological data using machine learning (ML)

GNSS

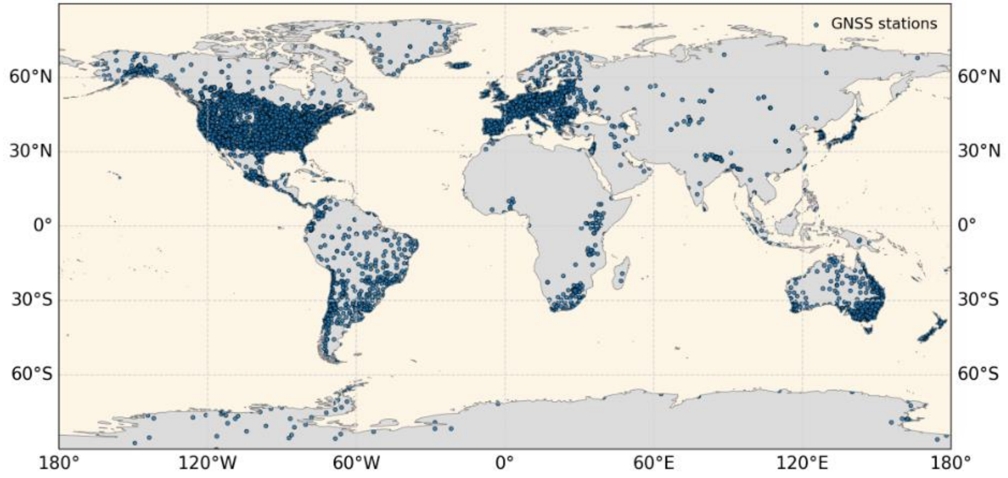


Data

TARGET

Zenith wet delay (ZWD)

- Source: Nevada Geodetic Laboratory (NGL)
  - more than 20.000 GNSS stations available
- Temporal resolution: 5 min → **hourly resolution**
- Spatial resolution: stations distributed globally
- Time span: all stations covering the **year 2019**



FEATURES

Meteorological data

- Source: ECMWF ERA5
- Temporal resolution: hourly
- Spatial resolution: 0.25°
- Time span: **year 2019**
- Several variables:
  - Specific humidity
  - Relative humidity
  - Temperature
  - Surface pressure
  - Total precipitation
  - Geopotential
  - Wind speed

Data

TARGET

Zenith wet delay (ZWD)

- Source: Nevada Geodetic Laboratory (NGL)
  - more than 20.000 GNSS stations available
- Temporal resolution: 5 min → **hourly resolution**
- Spatial resolution: stations distributed globally
- Time span: all stations covering the **year 2019**

added six pressure levels = (1000, 900, 800, 650, 450, 300) [hPa]

FEATURES

Meteorological data

- Source: ECMWF ERA5
- Temporal resolution: hourly
- Spatial resolution: 0.25°
- Time span: **year 2019**
- Several variables:
  - Specific humidity
  - Relative humidity
  - Temperature
  - Surface pressure
  - Total precipitation
  - Geopotential
  - Wind speed

Data

TARGET

Zenith wet delay (ZWD)

- Source: Nevada Geodetic Laboratory (NGL)
  - more than 20.000 GNSS stations available
- Temporal resolution: 5 min → **hourly resolution**
- Spatial resolution: stations distributed globally
- Time span: all stations covering the **year 2019**

added six pressure levels = (1000, 900, 800, 650, 450, 300) [hPa]

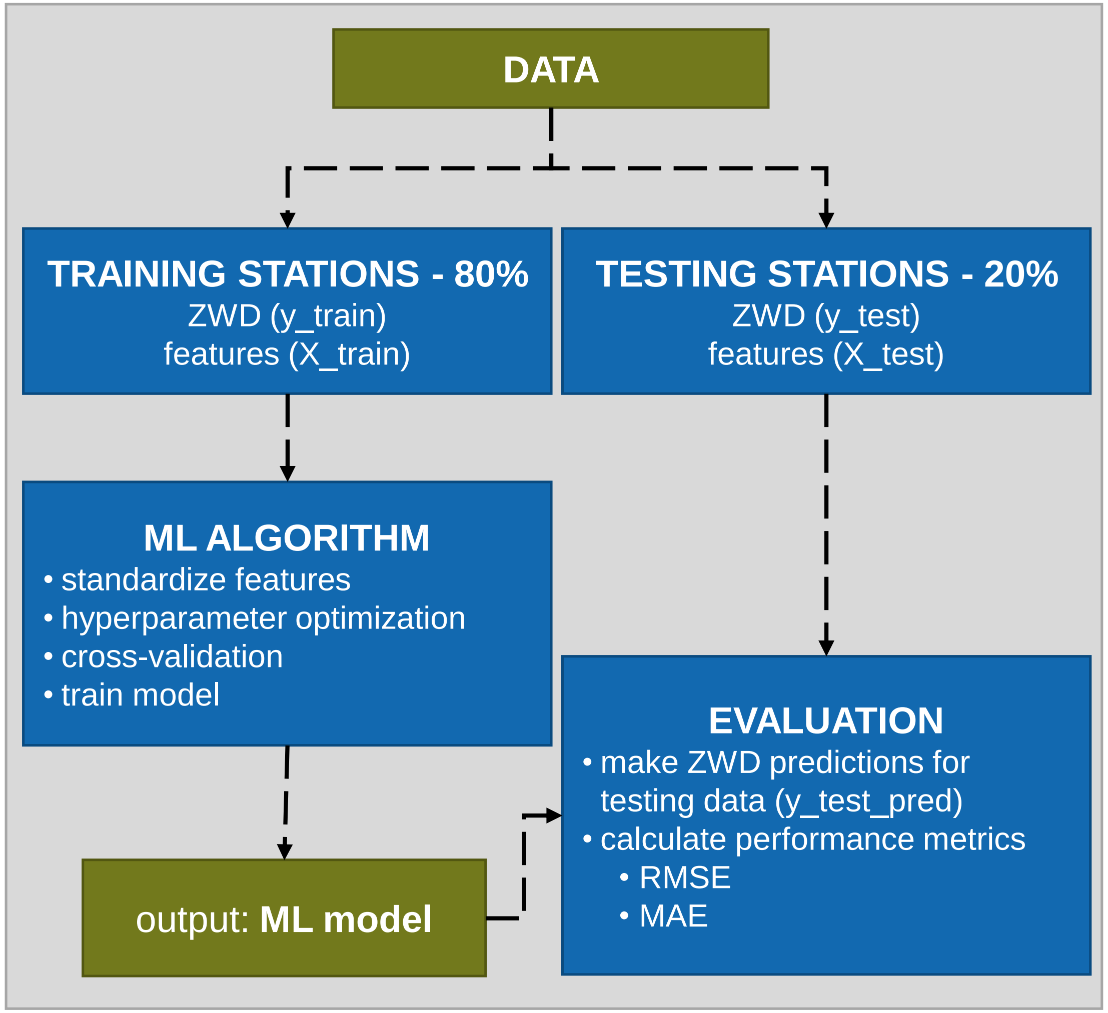
FEATURES

Meteorological data

- Source: ECMWF ERA5
- Temporal resolution: hourly
- Spatial resolution: 0.25°
- Time span: **year 2019**
- Several variables:
  - Specific humidity
  - Relative humidity
  - Temperature
  - Surface pressure
  - Total precipitation
  - Geopotential
  - Wind speed

- Latitude
- Longitude
- Height
- Time

Setup


DATA
TRAINING STATIONS - 80%
ZWD (y_train)
features (X_train)
TESTING STATIONS - 20%
ZWD (y_test)
features (X_test)
ML ALGORITHM
• standardize features
• hyperparameter optimization
• cross-validation
• train model
output: ML model
EVALUATION
• make ZWD predictions for testing data (y_test_pred)
• calculate performance metrics
• RMSE
• MAE

Setup

DATA

TRAINING STATIONS - 80%
ZWD (y_train)
features (X_train)

TESTING STATIONS - 20%
ZWD (y_test)
features (X_test)

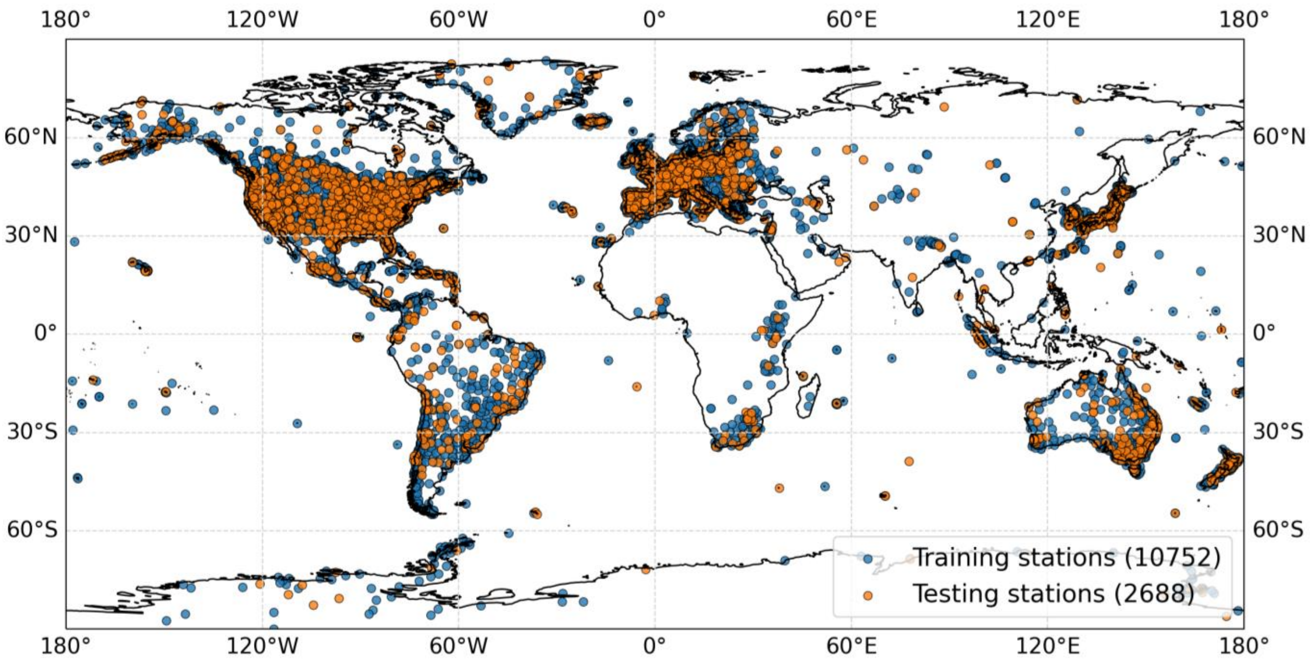
ML ALGORITHM
- standardize features
- hyperparameter optimization
- cross-validation
- train model

output: **ML model**

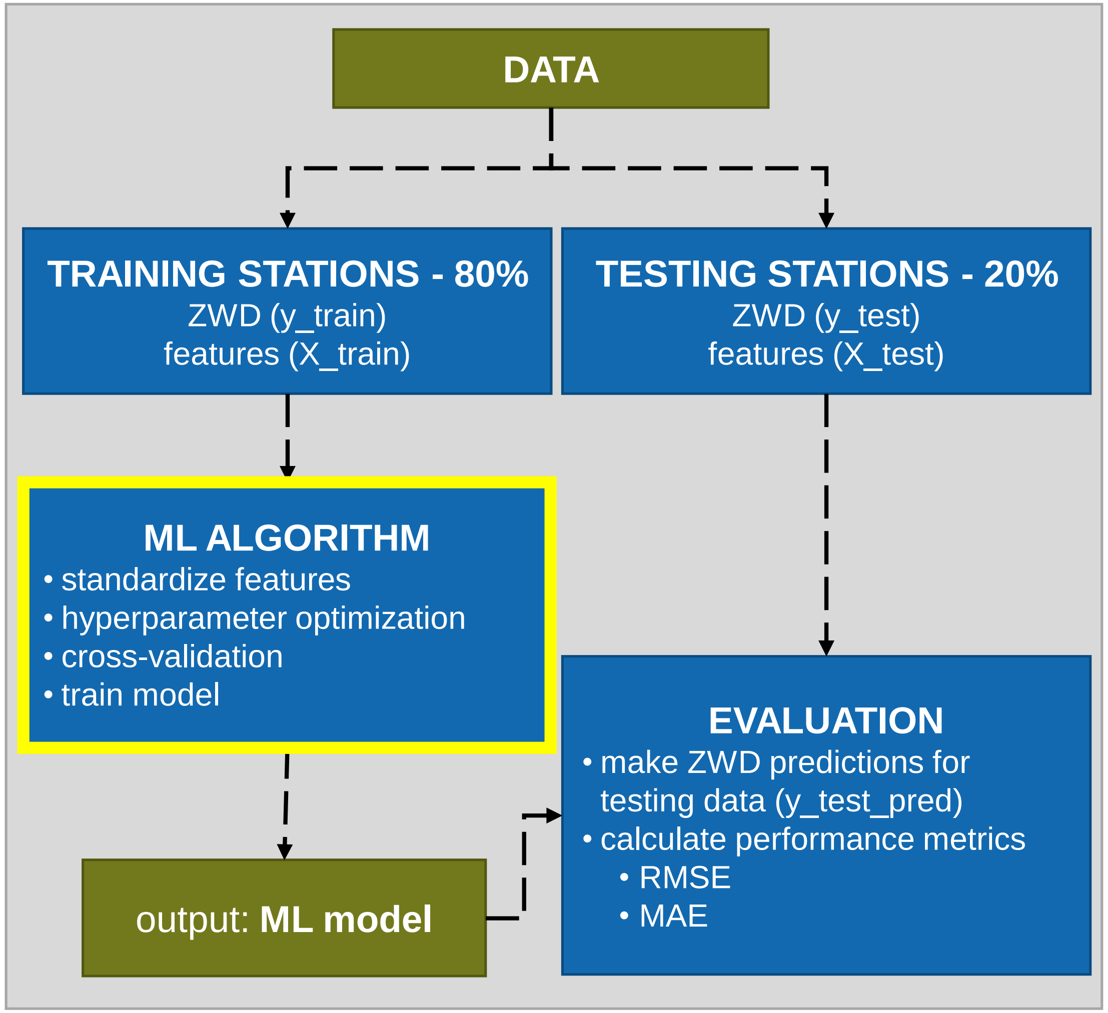
EVALUATION
- make ZWD predictions for testing data (y_test_pred)
- calculate performance metrics
  - RMSE
  - MAE

Global model for the year 2019

Distribution of training and test stations for all available stations (2019)

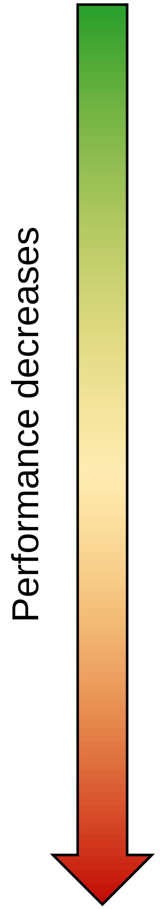


Setup



ML algorithms:

- **XGBoost**
- Random Forest
- HistGBoost
- Multilayer Perceptron
- Ridge Regression
- Stochastic Gradient Decent
- ElasticNet Regression
- Lasso Regression
- Linear Support Vector Machine
- AdaBoost



Setup

DATA

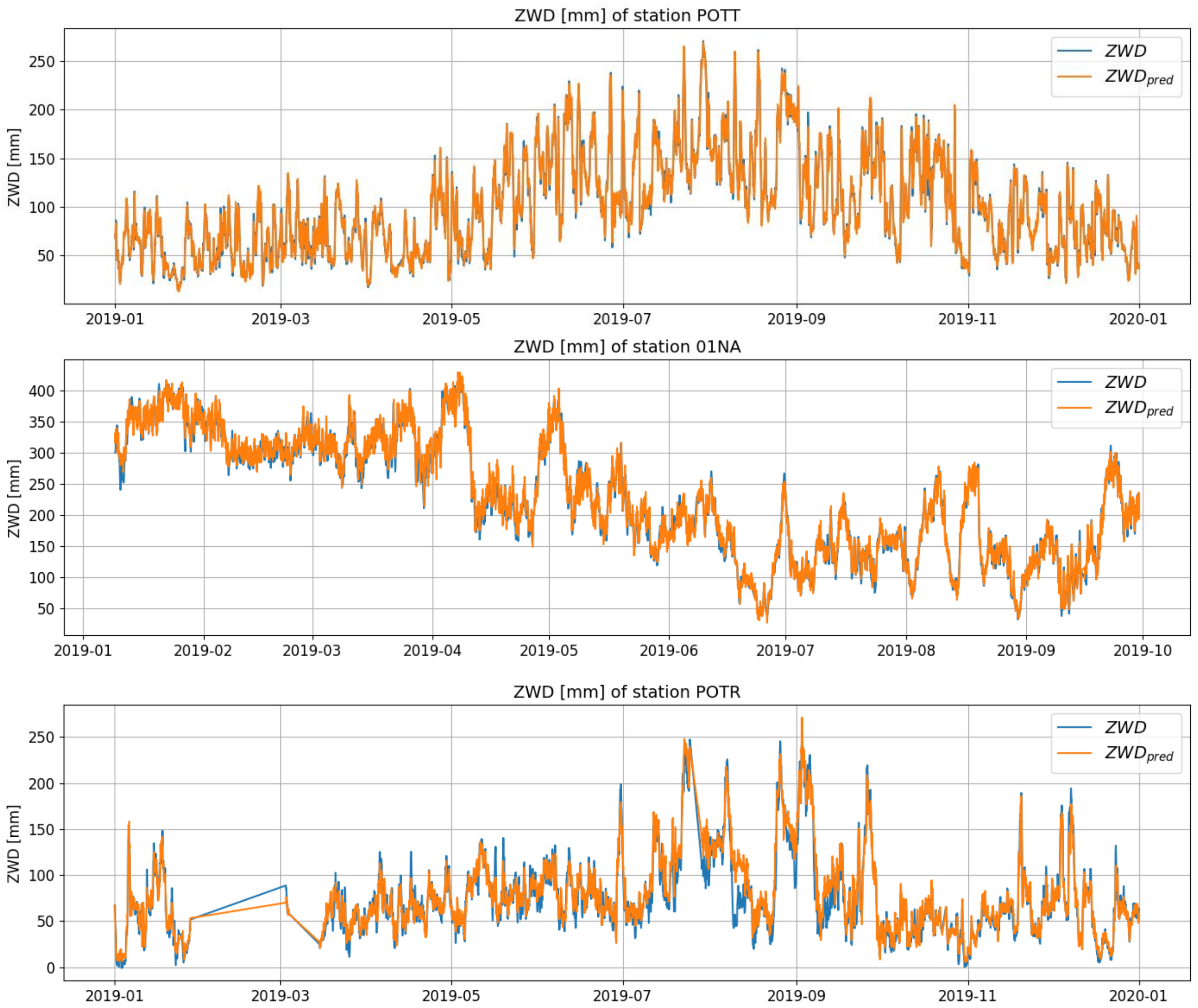
TRAINING STATIONS - 80%
ZWD (y_train)
features (X_train)

TESTING STATIONS - 20%
ZWD (y_test)
features (X_test)

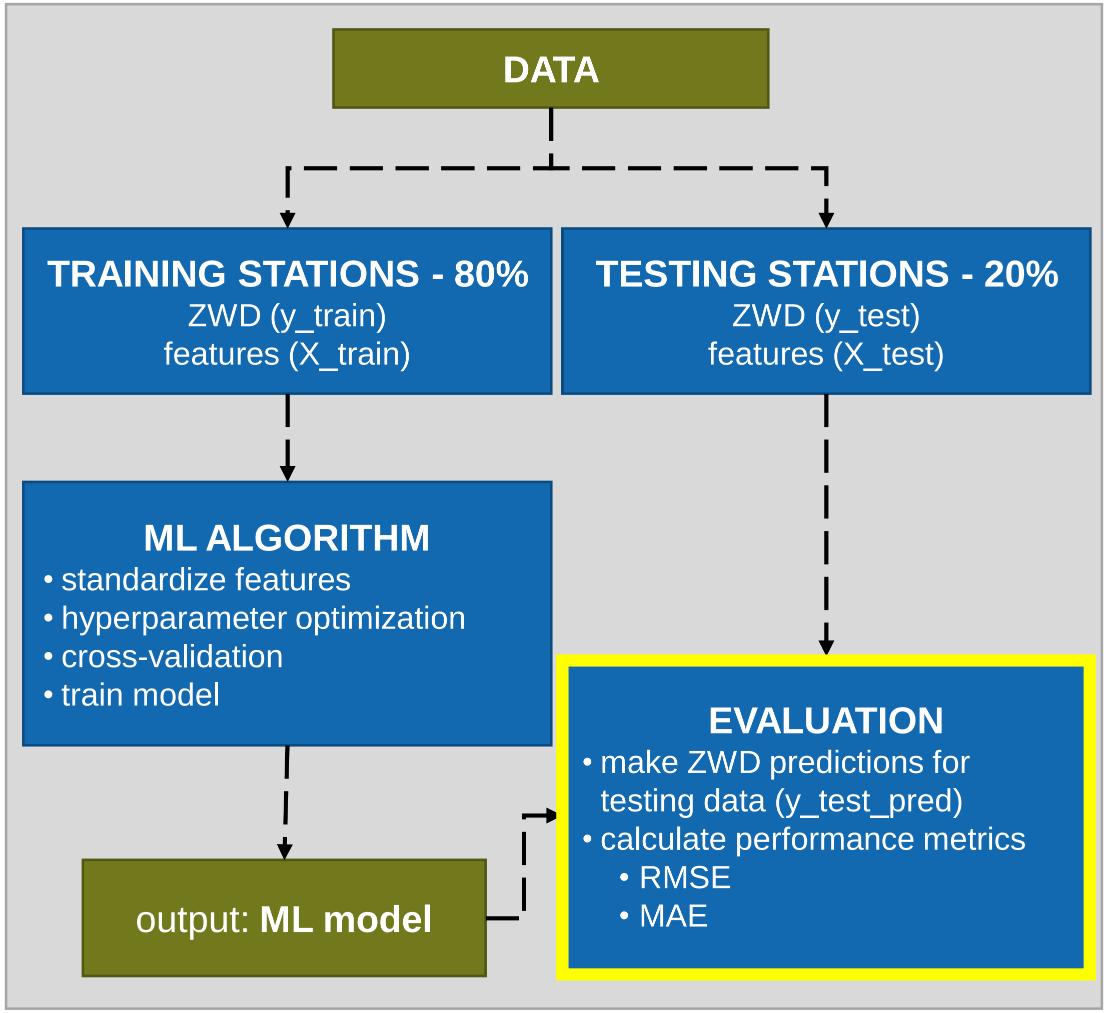
ML ALGORITHM
- standardize features
- hyperparameter optimization
- cross-validation
- train model

output: **ML model**

EVALUATION
- make ZWD predictions for testing data (y_test_pred)
- calculate performance metrics
  - RMSE
  - MAE



Setup

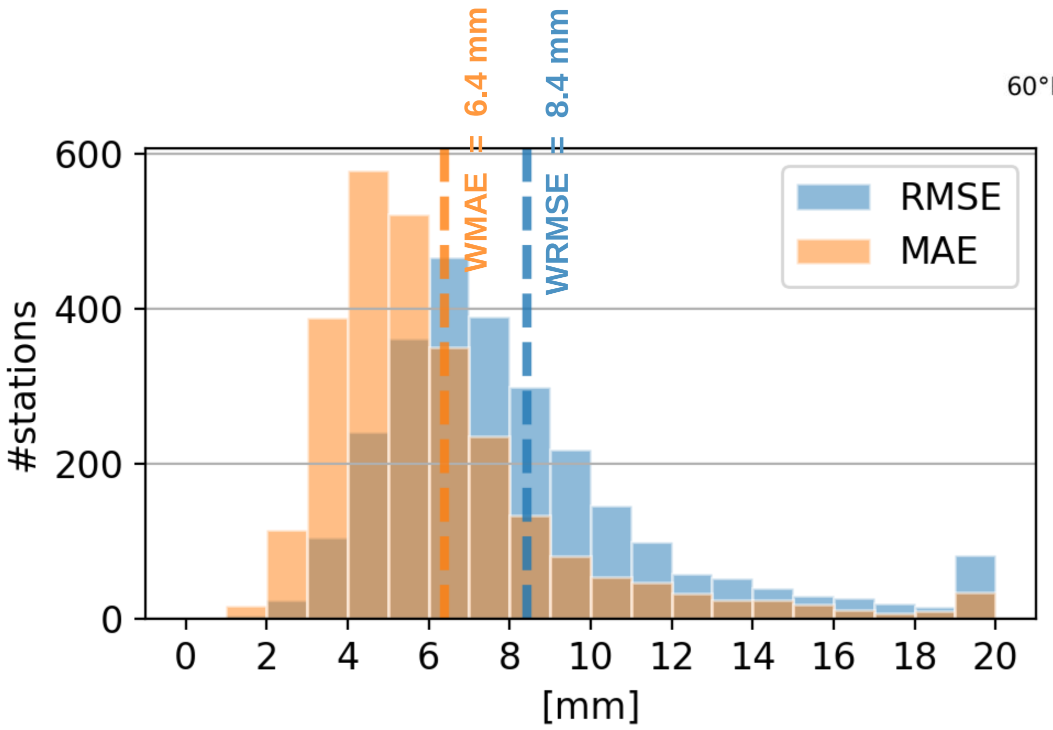


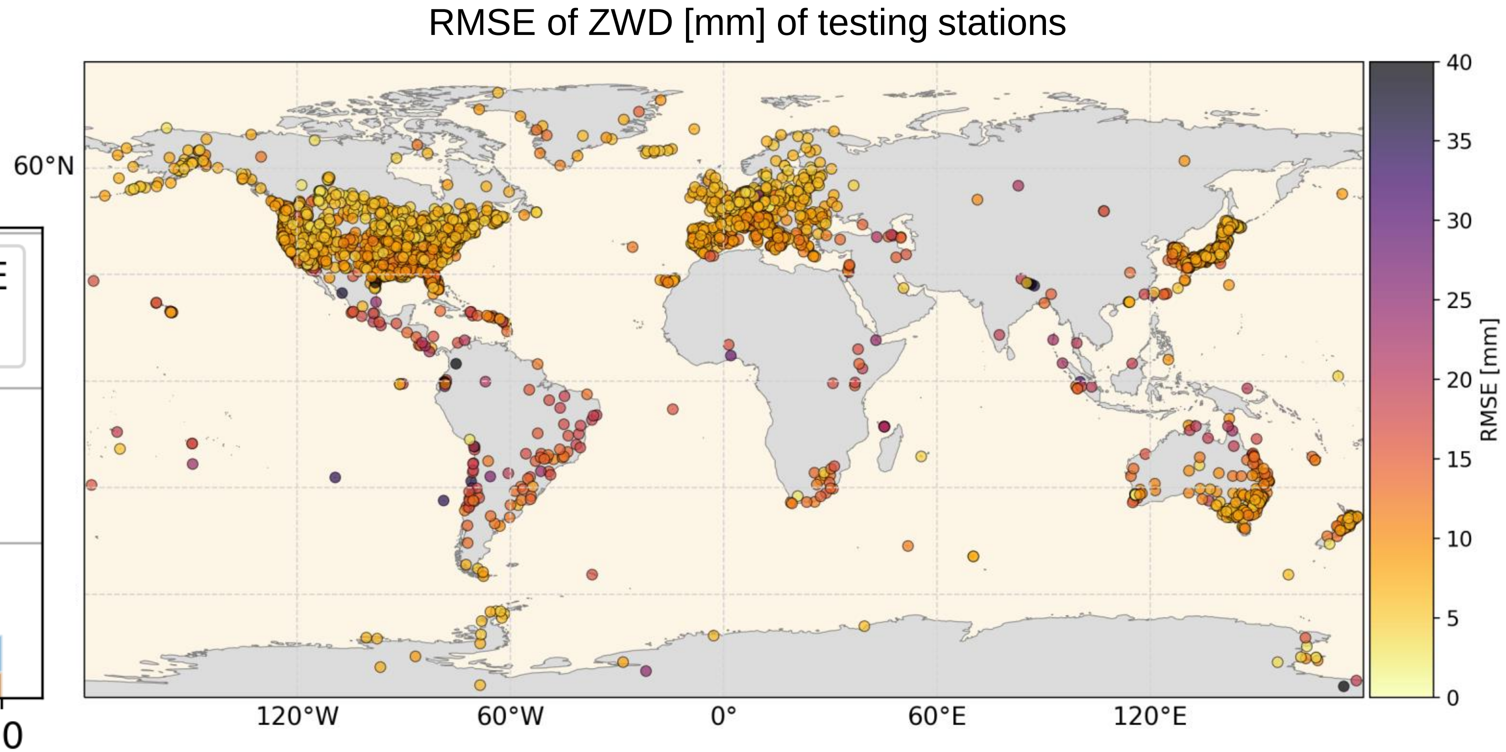
- Target: ZWD
- Features: latitude, longitude, height, time, specific humidity on six pressure levels (1000, 900, 800, 650, 450, 300 [hPa])
- ML algorithm: XGBoost
- Train model based on 10752 training stations for the year 2019
- Make ZWD predictions for 2688 testing stations for the year 2019

→ ZWD predictions for different stations for the same time period

→ No ZWD forecasting

Performance of individual test stations

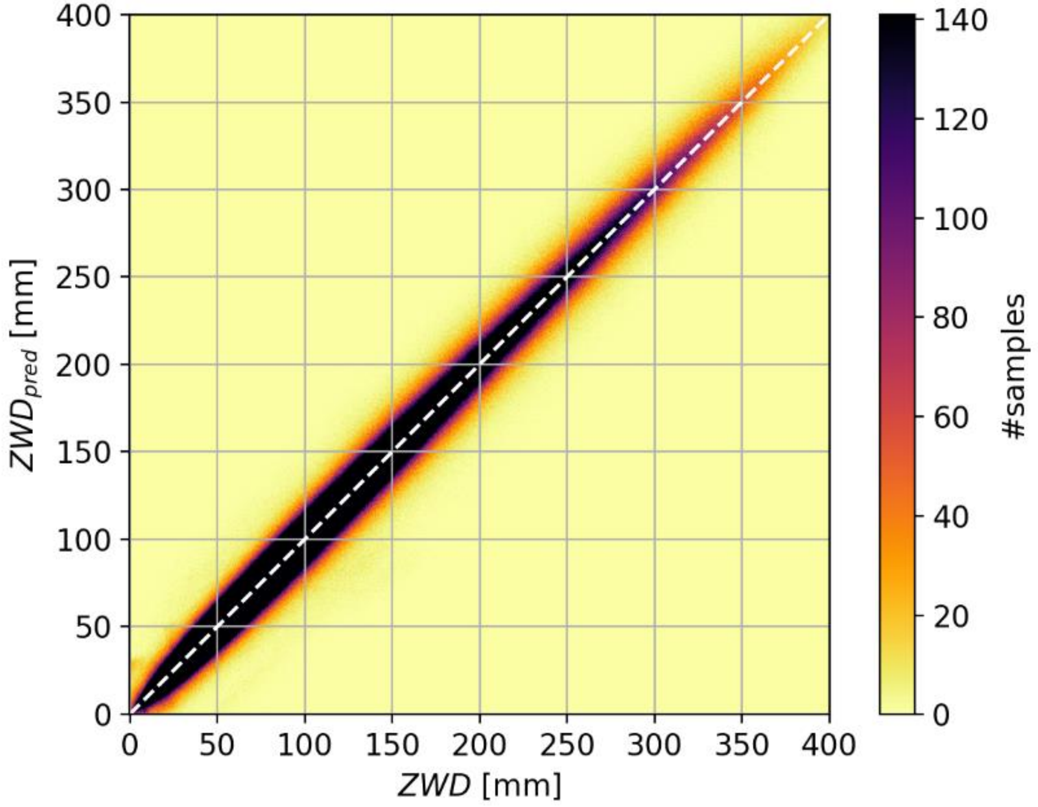




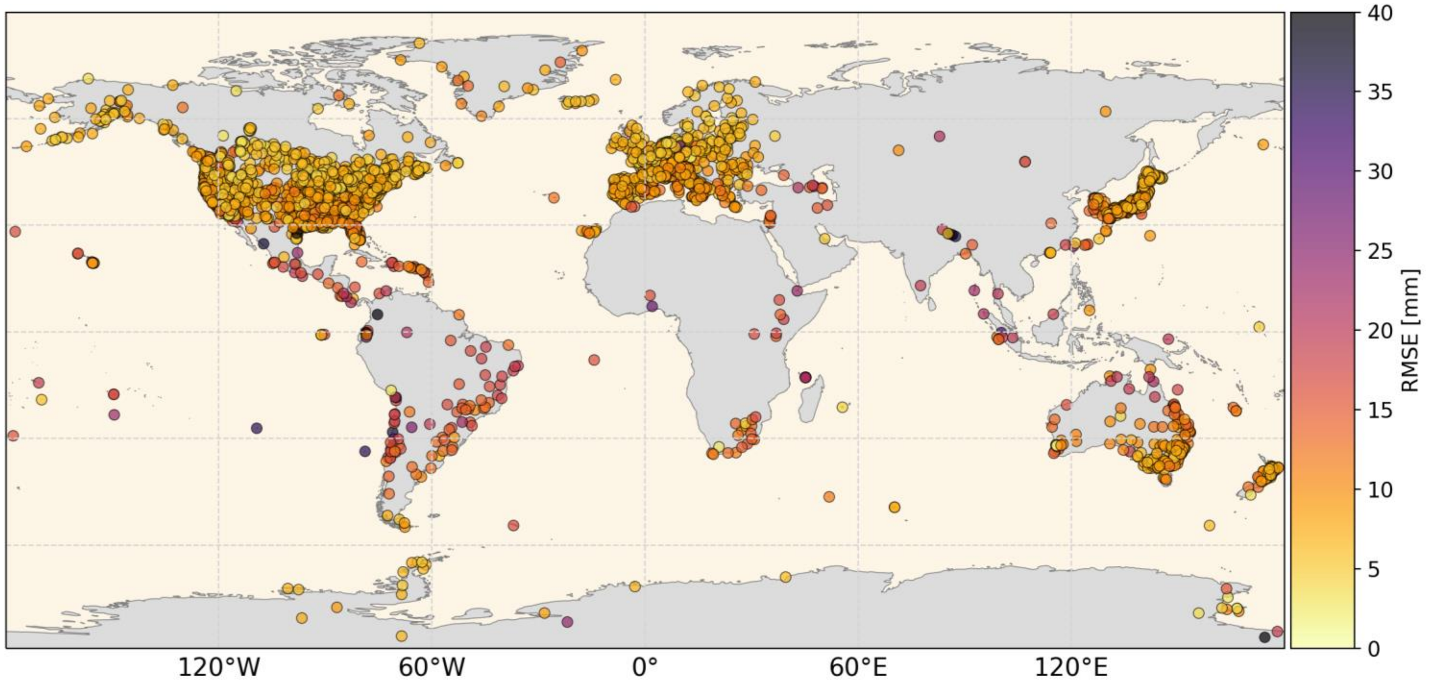
- Global average: 8.4 mm → very good performance; most stations have small errors
- Better performance in areas with dense GNSS station network and many stations

Performance of individual test stations

Comparison of predicted ZWD values to reference values at the test stations



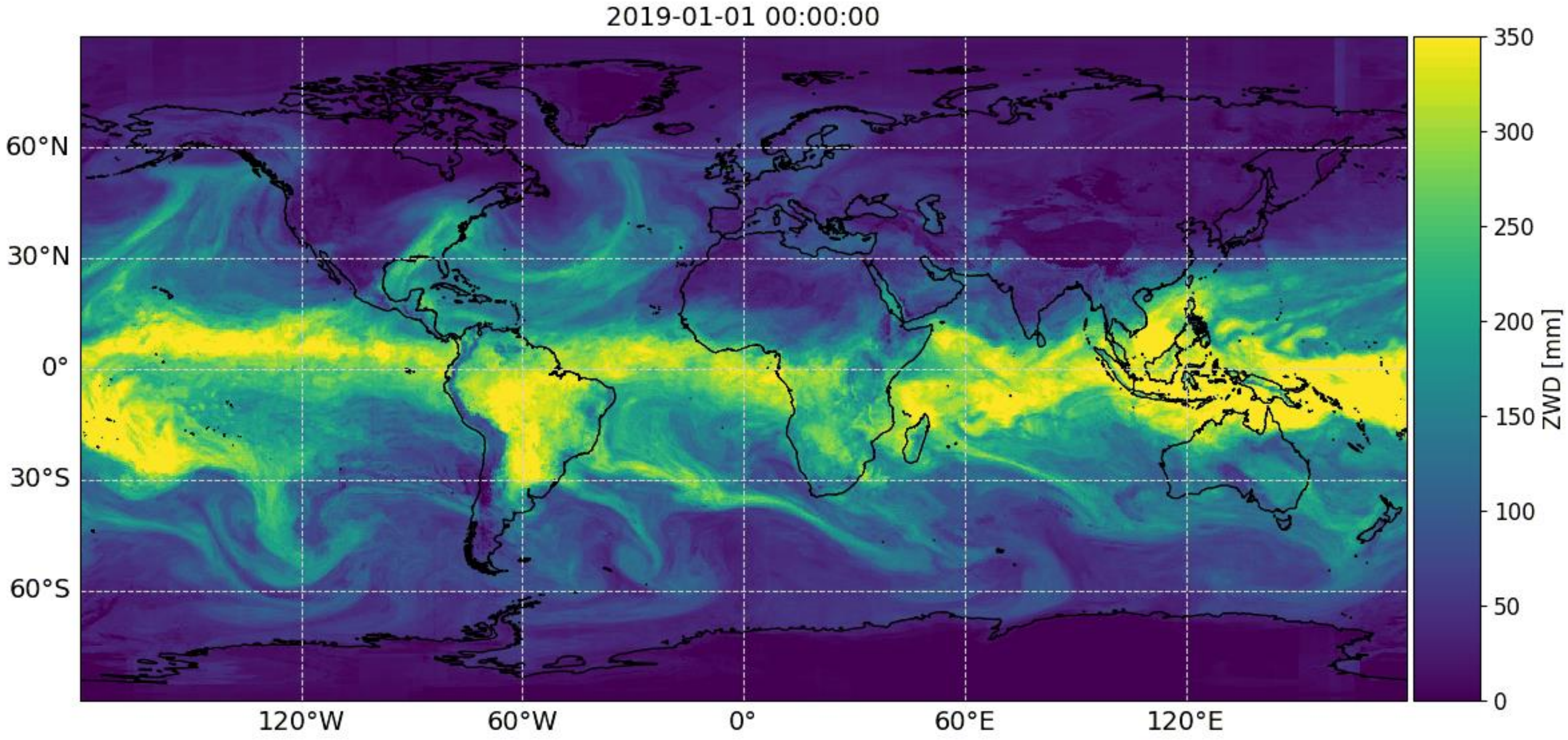
RMSE of ZWD [mm] of testing stations



- ZWD values cluster tightly along the identity line
- Model does not systematically over- or underpredict

ZWD predictions for 2019

- ML model can be applied at any location on Earth



External validation

Inter-comparison with two independent methods to estimate ZWD:

- (1) vertical integration of ERA5 data
- (2) vertical integration of radiosonde observation

External validation

Inter-comparison with two independent methods to estimate ZWD:

- (1) vertical integration of ERA5 data
- (2) vertical integration of radiosonde observation

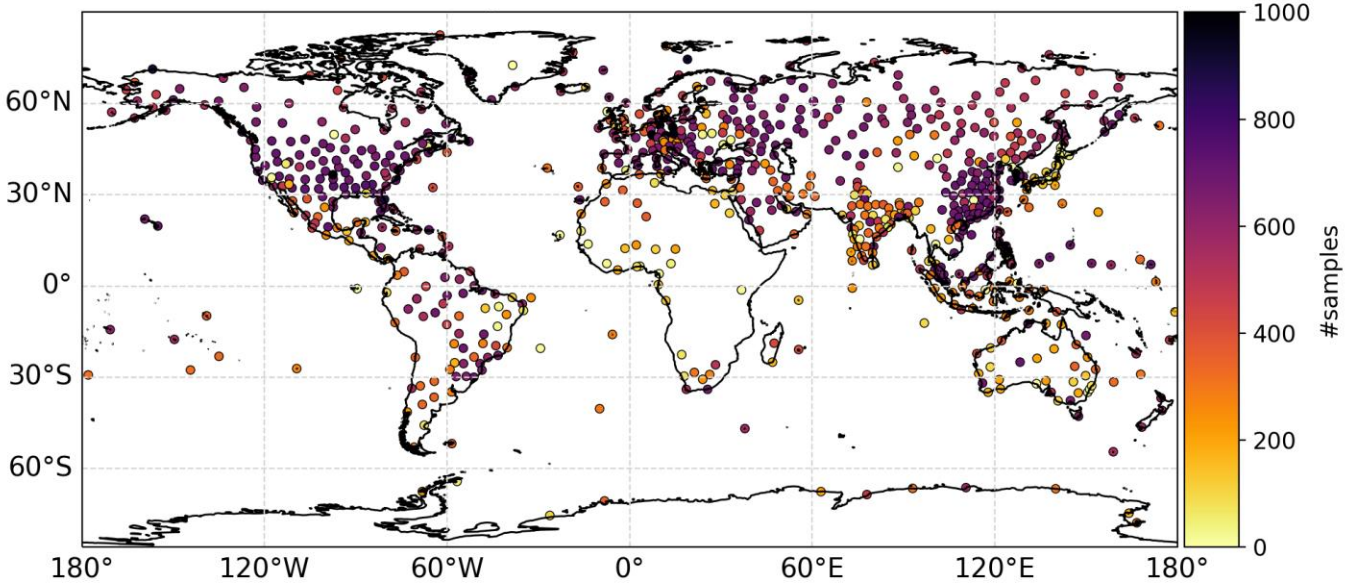
→ calculated ERA5-based ZWDs at the same locations as the 2688 test stations

External validation

Inter-comparison with two independent methods to estimate ZWD:

- (1) vertical integration of ERA5 data
- (2) vertical integration of radiosonde observation

Overview of radiosonde stations for the year 2019



- source: Integrated Global Radiosonde Archive (IGRA)
- 790 radiosonde stations available for 2019
- Geographic locations do not coincide with the GNSS stations
  - Radius of 20 km → 116 station pairs

→ calculated ML-based ZWD predictions at 116 locations of radiosonde stations

External validation

Inter-comparison with two independent methods to estimate ZWD:

- (1) vertical integration of ERA5 data
- (2) vertical integration of radiosonde observation

WRMSE [mm]	NGL (reference)	ML	ERA5	radiosonde
NGL (reference)				
ML	8.4			
ERA5	11.1	9.3		
radiosonde	14.7	15.2	14.1	

- ML model reproduces NGL better than a direct integration of ERA5
- higher compatibility of the proposed model with NGL and ERA5 was to be expected → NGL was reference, ERA5 based on same meteorological data

Global vs. specialised models

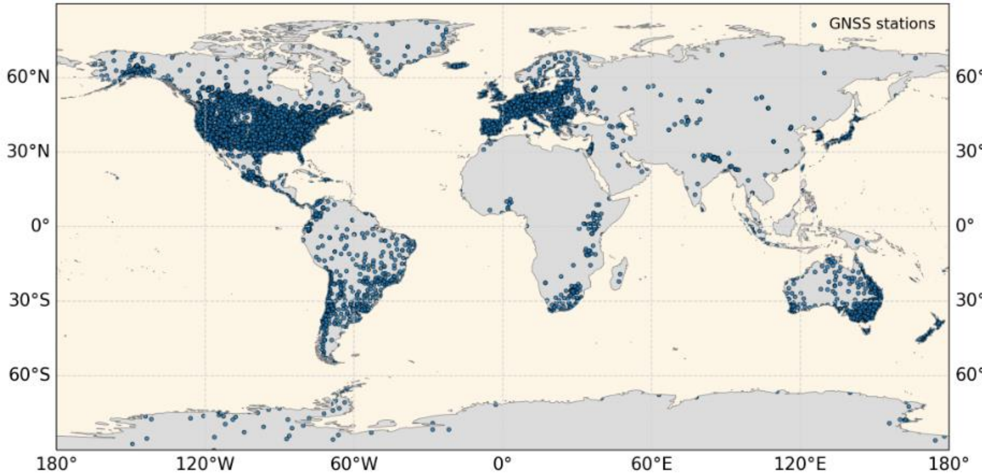
- Regional models
- Monthly models

Regional models

- Train regional models
- Apply global model to regional stations

	WRMSE [mm]		#stations
	regional	global	
Europe	7.0	7.4	3161
North America	7.4	7.8	6703
Australia	9.0	9.3	890
Asia	9.2	9.5	1780
Africa	13.6	13.6	212
South America	14.5	14.6	541
global average		**8.4**	**13440**

- Local models perform slightly better in regions with enough stations
- Performance of global model is comparable to performance of regional models
- Best performance in Europe and North America → highest number and highest density of GNSS stations



Monthly models

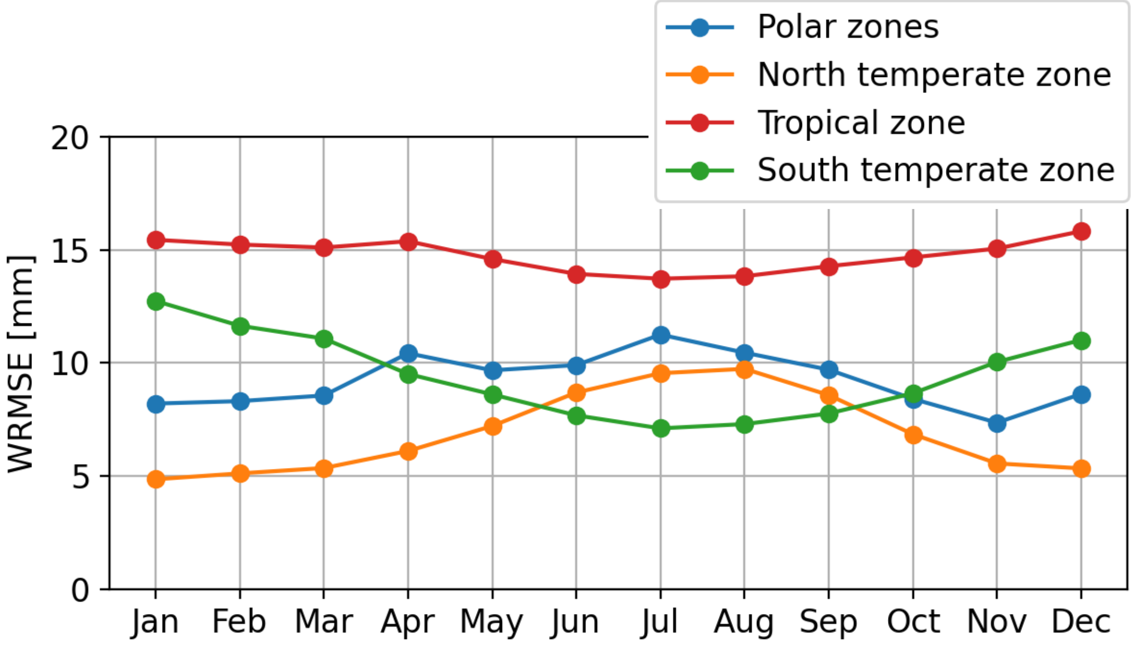
- Train monthly models
- Apply global model to monthly data

- Local models achieve slightly better performance in all months
- Performance of global model is comparable to performance of monthly models
- Best performance in “winter” months
- Global model can capture the seasonal variations of ZWD well

	WRMSE [mm]	
	monthly	global
January	6.2	6.6
February	6.3	6.8
March	6.5	6.9
April	7.0	7.5
May	7.8	8.3
June	8.9	9.5
July	9.6	10.2
August	9.8	10.4
September	8.9	9.4
October	7.5	8.1
November	6.6	7.1
December	6.6	7.0
all months		**8.4**

Monthly models

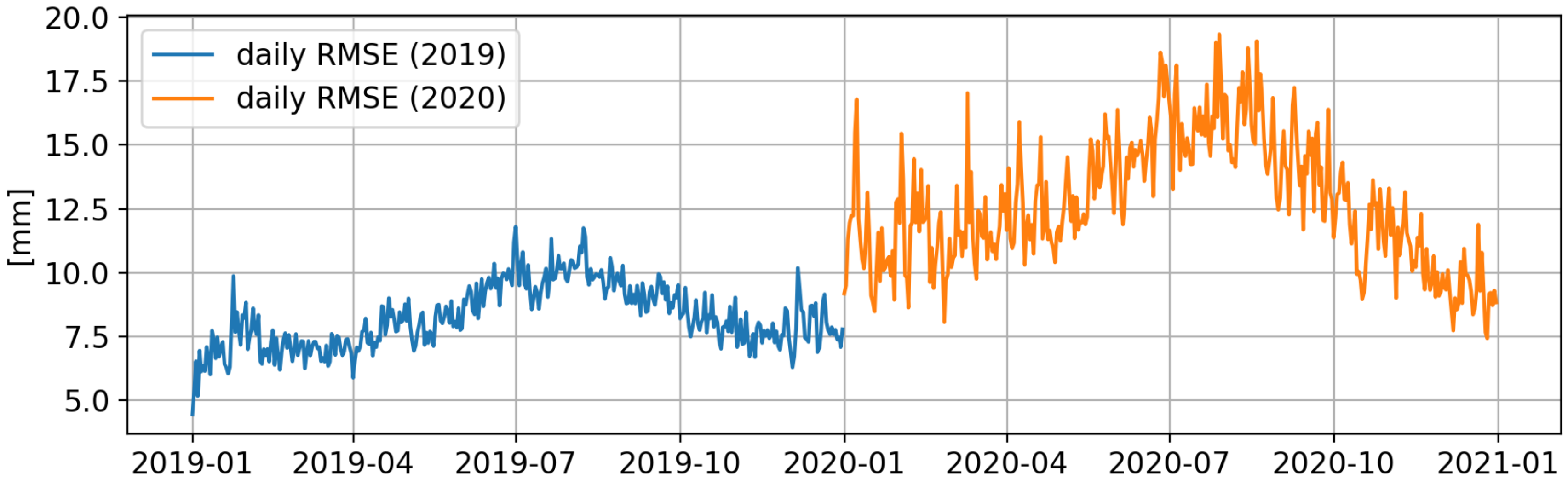
Performance of monthly models evaluated at different climate zones



	WRMSE [mm]	
	monthly	**global**
January	6.2	6.6
February	6.3	6.8
March	6.5	6.9
April	7.0	7.5
May	7.8	8.3
June	8.9	9.5
July	9.6	10.2
August	9.8	10.4
September	8.9	9.4
October	7.5	8.1
November	6.6	7.1
December	6.6	7.0
all months		**8.4**

Temporal prediction

- Apply model trained on data of 2019 to data of 2020



- Significant over-fitting of the current model to the conditions of 2019

Outlook

- Train model on **multiple years**
- Increasing data volume → neural networks and deep learning
- Improve temporal predictions
- Aim for a model that can be used for **near real-time applications**
- **Forecast data** will be utilized

Summary

- **Global ML-based ZWD model based on meteorological data**
- ML algorithm: **XGBoost**
- Final model achieves an WRMSE of **8.4 mm**
- Validation of ML-based ZWD predictions
  - with **ERA5**-based ZWDs
  - with **radiosonde**-based ZWDs
- Global model achieves similar performance compared to **regional and monthly models**
- **Seasonal behaviour** can be seen in monthly models
- Temporal prediction is possible but performance decreases
- Future work is planned

Laura Crocetti
lcrocetti@ethz.ch

ETH Zurich

Space Geodesy Group

Institute of Geodesy and Photogrammetry

Zurich, Switzerland

www.space.igp.ethz.ch