ETH zürich IIASA esa CAMALIOT

D4G POTSDAM JUNE 13-15, 2022
1ST WORKSHOP ON DATA SCIENCE FOR GNSS REMOTE SENSING

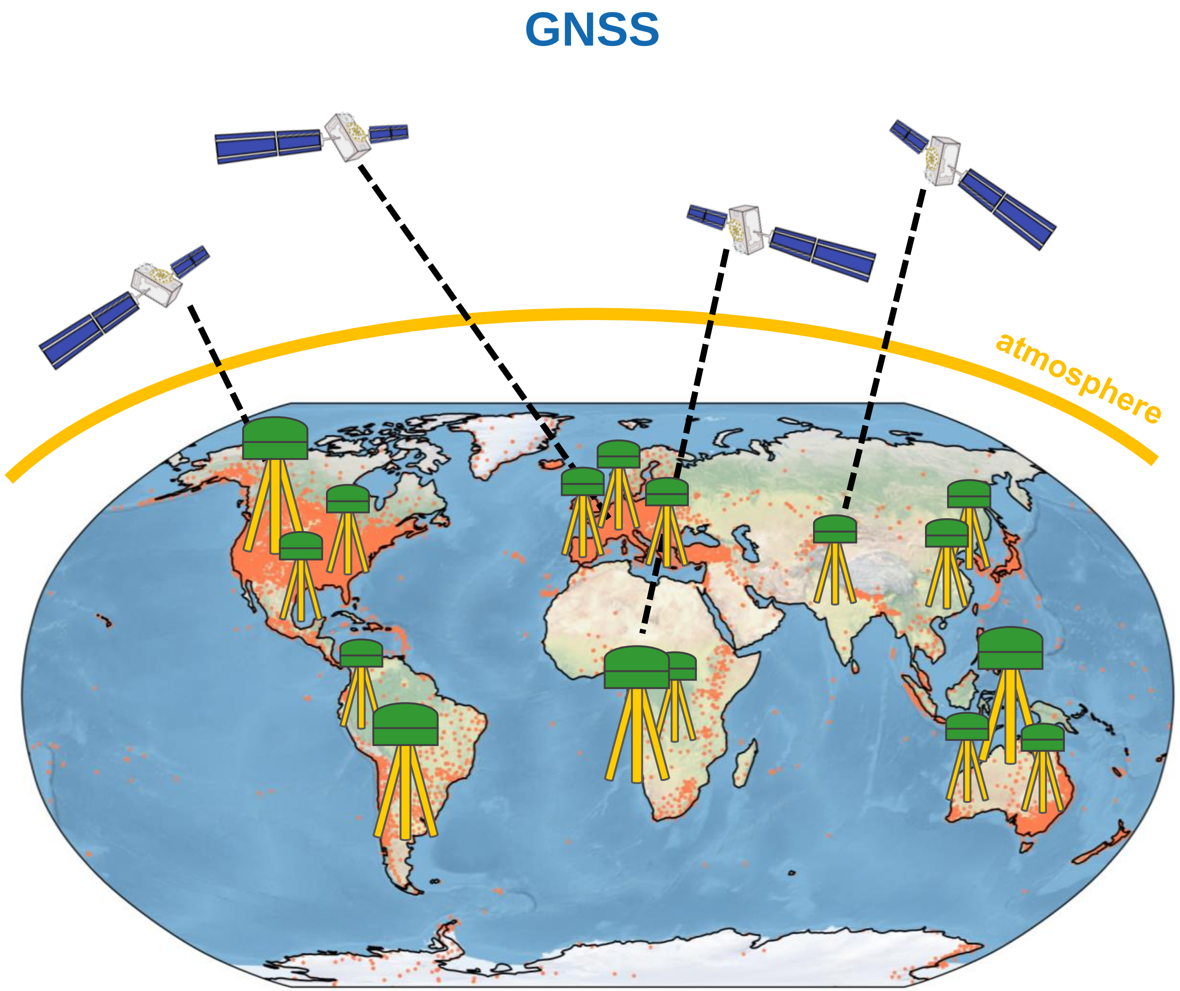
Machine learning algorithms for global modelling of Zenith Wet Delay based on GNSS measurements and meteorological data

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

14th June 2022, D4G22

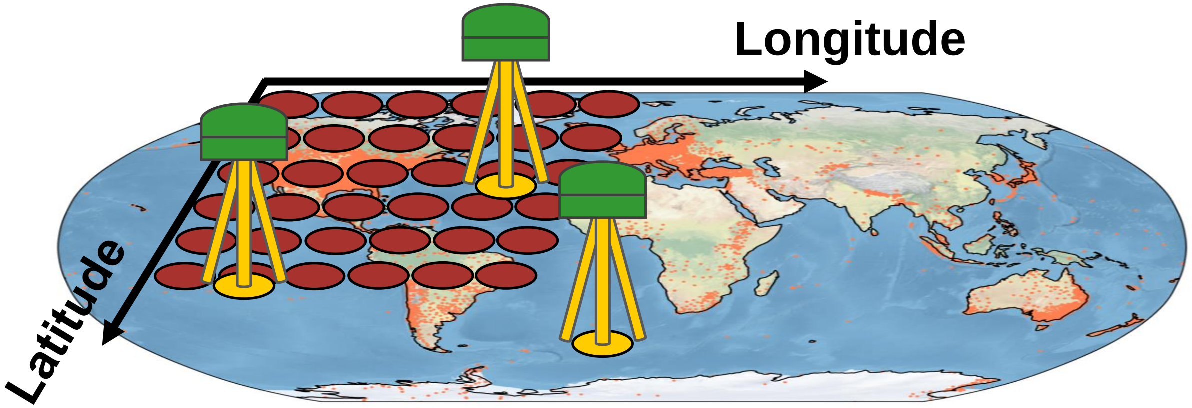
Motivation

- Global Navigation Satellite System (GNSS) – find also application in atmospheric research
- GNSS signals traverse the atmosphere
- Time delays → measurements of atmospheric properties



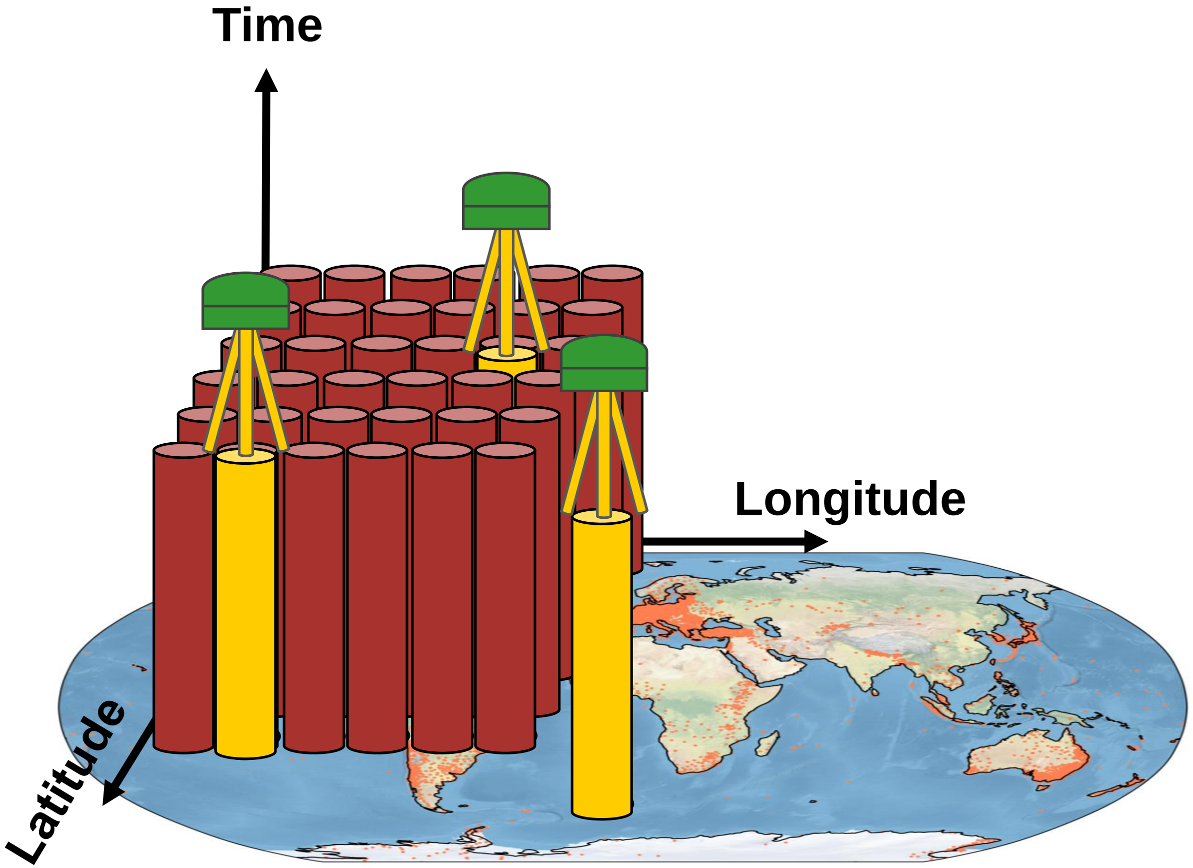
Motivation

- Global Navigation Satellite System (GNSS) – find also application in atmospheric research
- GNSS signals traverse the atmosphere
- Time delays → measurements of atmospheric properties
- Model **Zenith Wet Delay** globally based on meteorological data using Machine Learning (ML)



Motivation

- Global Navigation Satellite System (GNSS) – find also application in atmospheric research
- GNSS signals traverse the atmosphere
- Time delays → measurements of atmospheric properties
- Model **Zenith Wet Delay** globally based on meteorological data using Machine Learning (ML)



Data

ECMWF

Zenith Wet Delay (ZWD) TARGET

- Source: Nevada Geodetic Laboratory (NGL)
  - More than 19.000 GPS stations available
- Temporal resolution: 5 min → hourly resolution
- Spatial resolution: stations distributed globally
- Time span: year 2019

- Aim: predict ZWD in space (and time) using ML algorithms

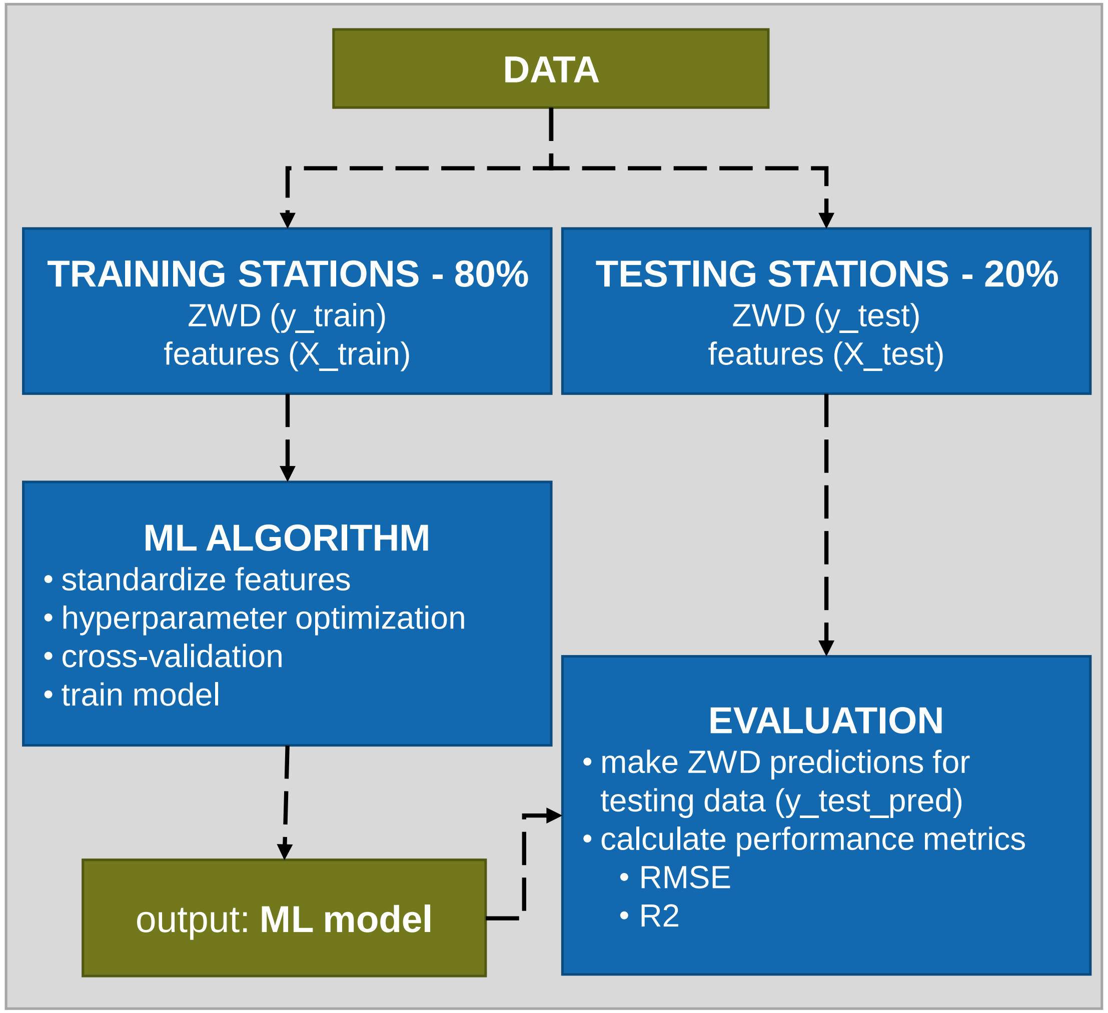
Meteorological data

- 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

FEATURES

- 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
EVALUATION
• make ZWD predictions for testing data (y_test_pred)
• calculate performance metrics
• RMSE
• R2
output: ML model

Setup

DATA

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

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

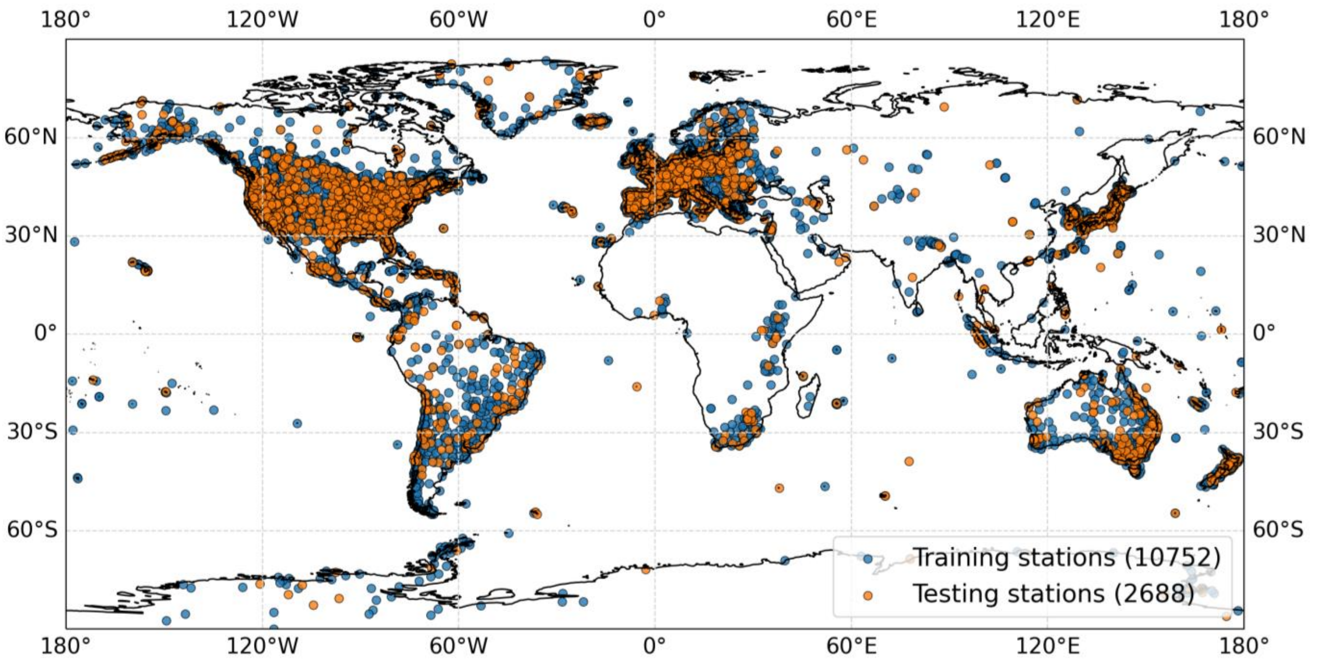
ML ALGORITHM
- standardize features
- hyperparameter optimization
- cross-validation
- train model

output: **ML model**

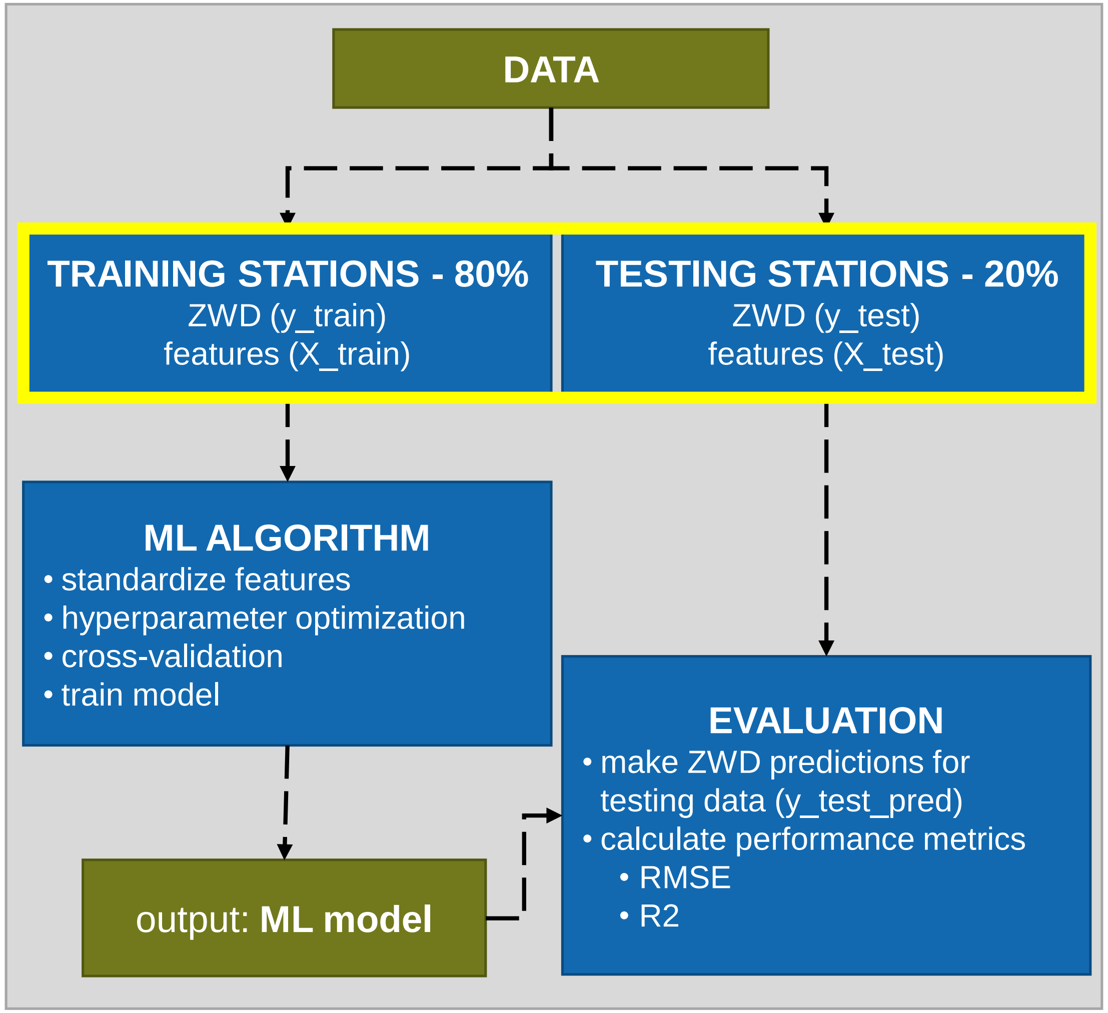
EVALUATION
- make ZWD predictions for testing data (y_test_pred)
- calculate performance metrics
  - RMSE
  - R2

Global model for the year 2019

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



Setup



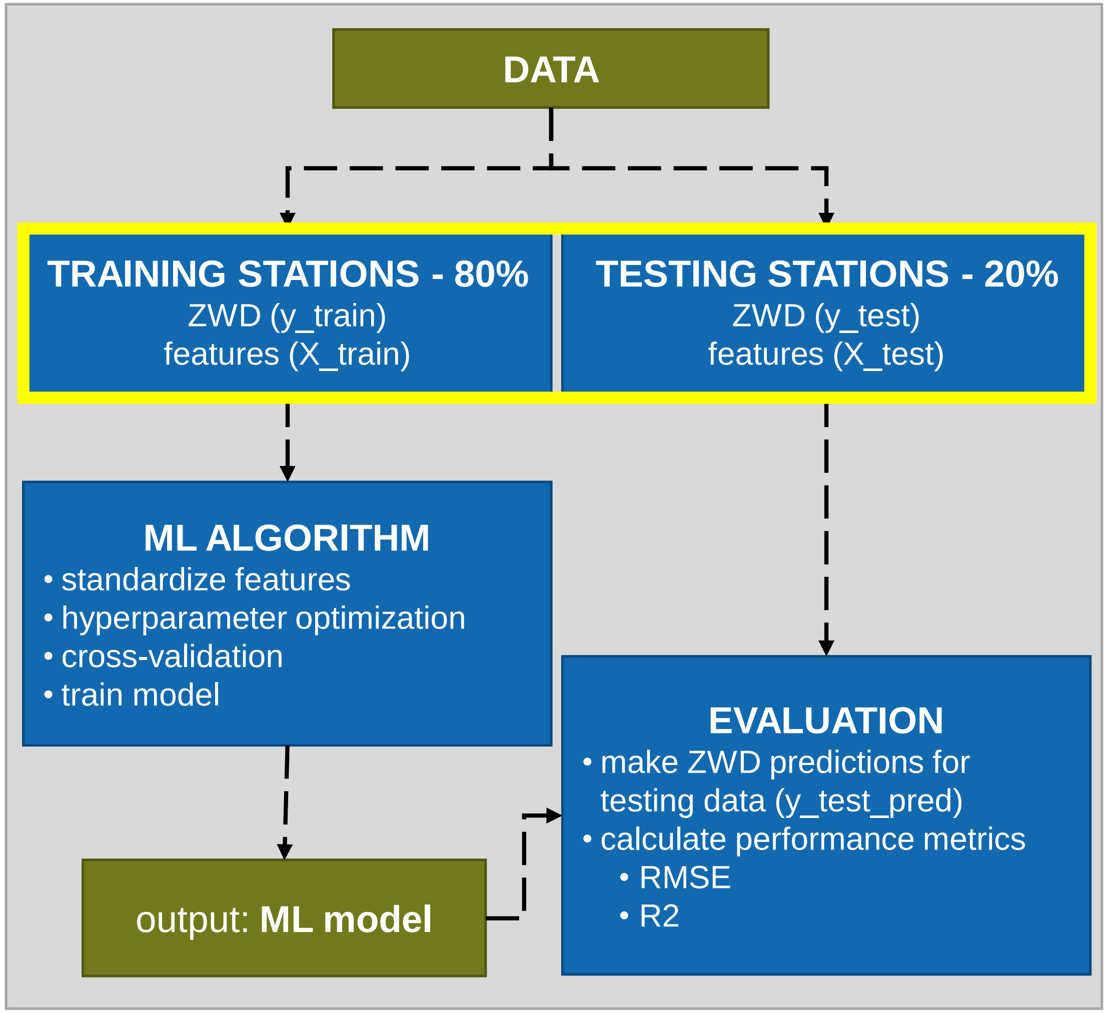
Target | **Features**

samples

	ZWD	lat	lon	time	doy	hour
2019-01-01 00:00:00/USAL	73.80000	40.33492	18.11149	1043136.00000	1.00000	0.00000
2019-01-01 01:00:00/USAL	72.40000	40.33492	18.11149	1043137.00000	1.04167	1.00000
2019-01-01 02:00:00/USAL	71.00000	40.33492	18.11149	1043138.00000	1.08333	2.00000
2019-01-01 03:00:00/USAL	68.70000	40.33492	18.11149	1043139.00000	1.12500	3.00000
2019-01-01 04:00:00/USAL	66.30000	40.33492	18.11149	1043140.00000	1.16667	4.00000
2019-01-01 05:00:00/USAL	66.50000	40.33492	18.11149	1043141.00000	1.20833	5.00000
2019-01-01 06:00:00/USAL	63.80000	40.33492	18.11149	1043142.00000	1.25000	6.00000
2019-01-01 07:00:00/USAL	60.50000	40.33492	18.11149	1043143.00000	1.29167	7.00000
2019-01-01 08:00:00/USAL	57.70000	40.33492	18.11149	1043144.00000	1.33333	8.00000
2019-01-01 09:00:00/USAL	56.70000	40.33492	18.11149	1043145.00000	1.37500	9.00000
2019-01-01 10:00:00/USAL	55.30000	40.33492	18.11149	1043146.00000	1.41667	10.00000
2019-01-01 11:00:00/USAL	52.00000	40.33492	18.11149	1043147.00000	1.45833	11.00000
2019-01-01 12:00:00/USAL	50.90000	40.33492	18.11149	1043148.00000	1.50000	12.00000
2019-01-01 13:00:00/USAL	49.30000	40.33492	18.11149	1043149.00000	1.54167	13.00000
2019-01-01 14:00:00/USAL	47.50000	40.33492	18.11149	1043150.00000	1.58333	14.00000
2019-01-01 15:00:00/USAL	48.10000	40.33492	18.11149	1043151.00000	1.62500	15.00000
2019-01-01 16:00:00/USAL	48.30000	40.33492	18.11149	1043152.00000	1.66667	16.00000
2019-01-01 17:00:00/USAL	49.20000	40.33492	18.11149	1043153.00000	1.70833	17.00000
2019-01-01 18:00:00/USAL	52.60000	40.33492	18.11149	1043154.00000	1.75000	18.00000
2019-01-01 19:00:00/USAL	56.10000	40.33492	18.11149	1043155.00000	1.79167	19.00000
2019-01-01 20:00:00/USAL	56.60000	40.33492	18.11149	1043156.00000	1.83333	20.00000
2019-01-01 21:00:00/USAL	56.80000	40.33492	18.11149	1043157.00000	1.87500	21.00000
2019-01-01 22:00:00/USAL	57.60000	40.33492	18.11149	1043158.00000	1.91667	22.00000

... | ...

Setup

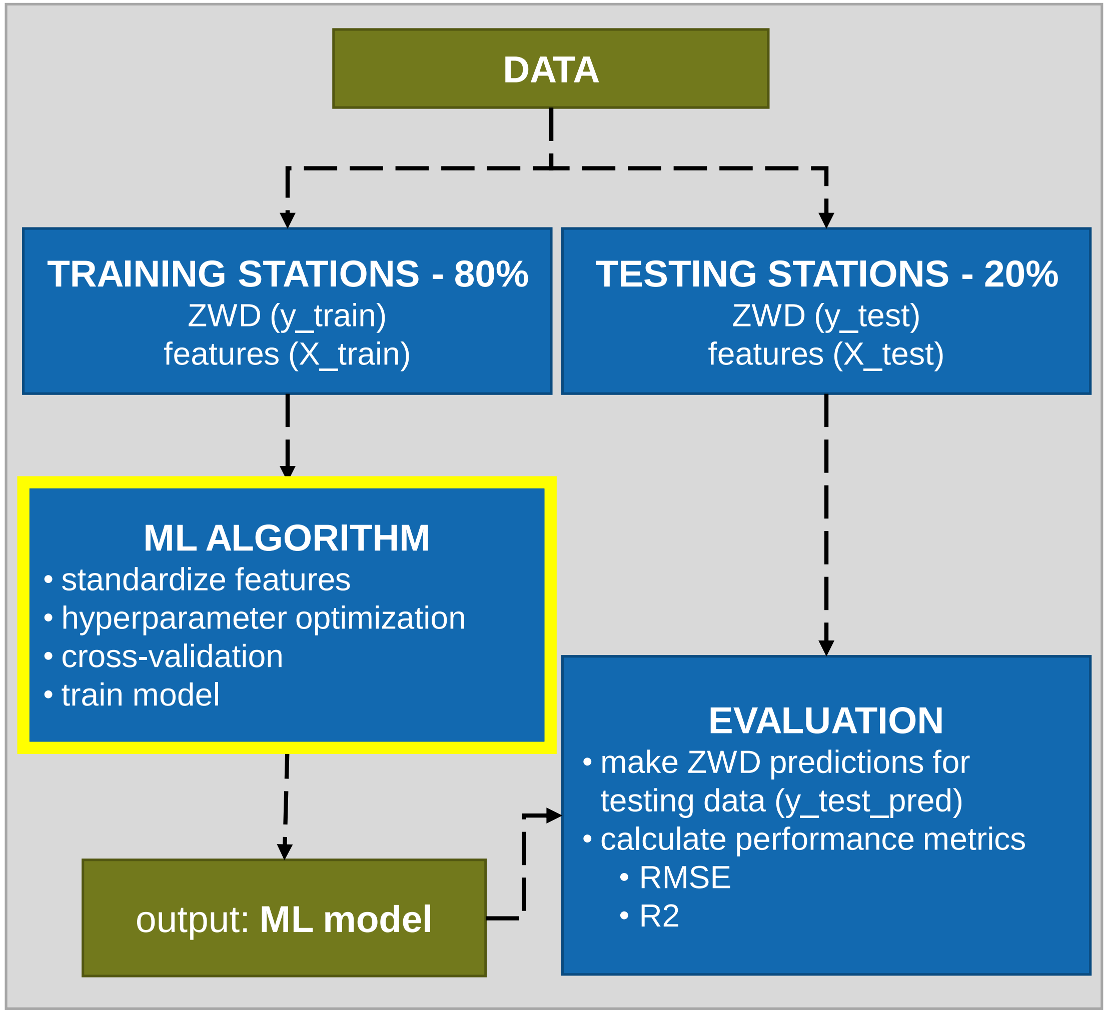


- Train model based on 10752 training stations for the year 2019
- Make predictions for 2688 testing stations for the year 2019

→ Predictions for different stations for the same time period

→ no predictions into the future

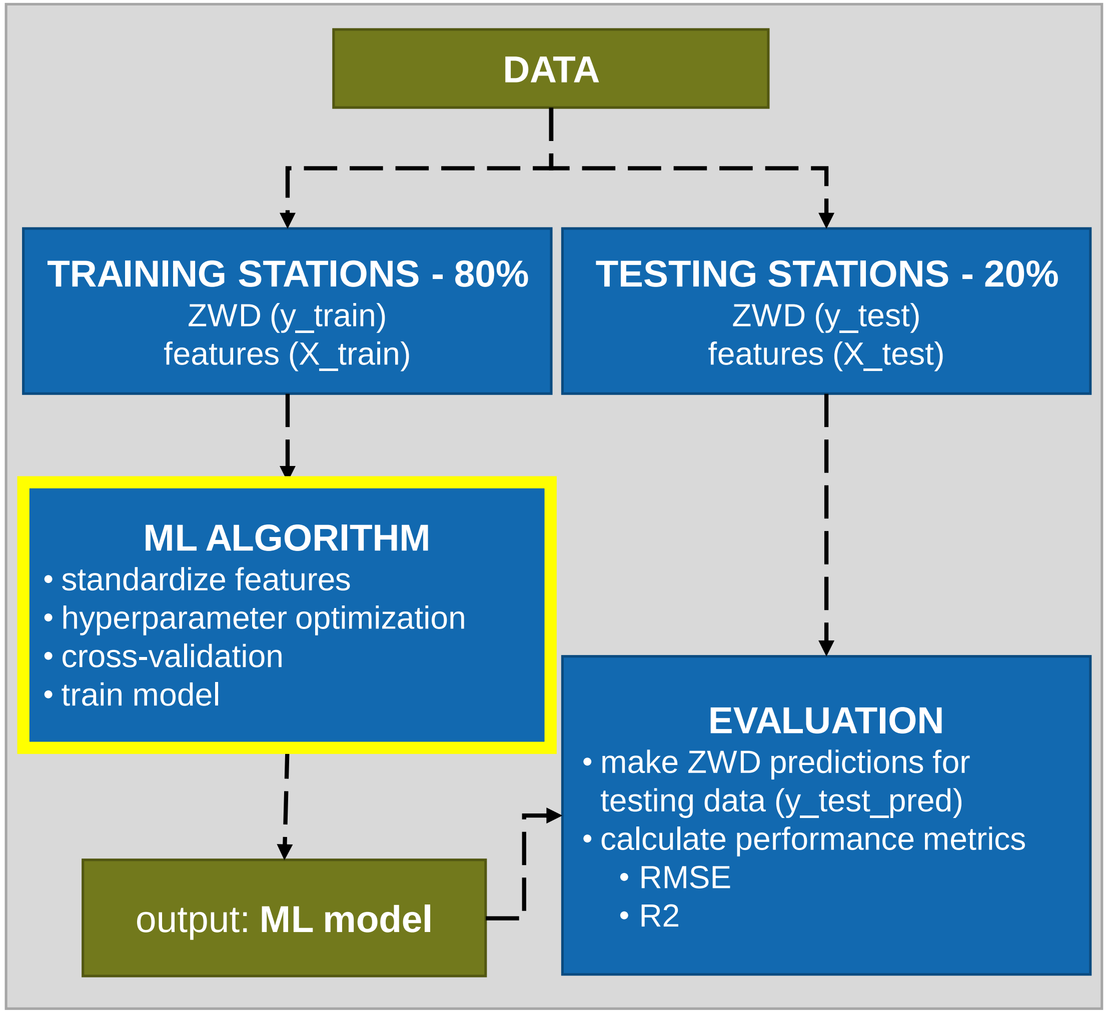
Setup



ML algorithms:

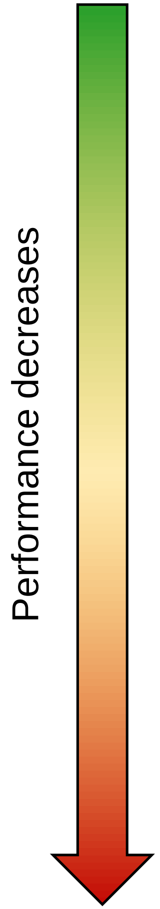
- **XGBoost**
- **Random Forest**

Setup

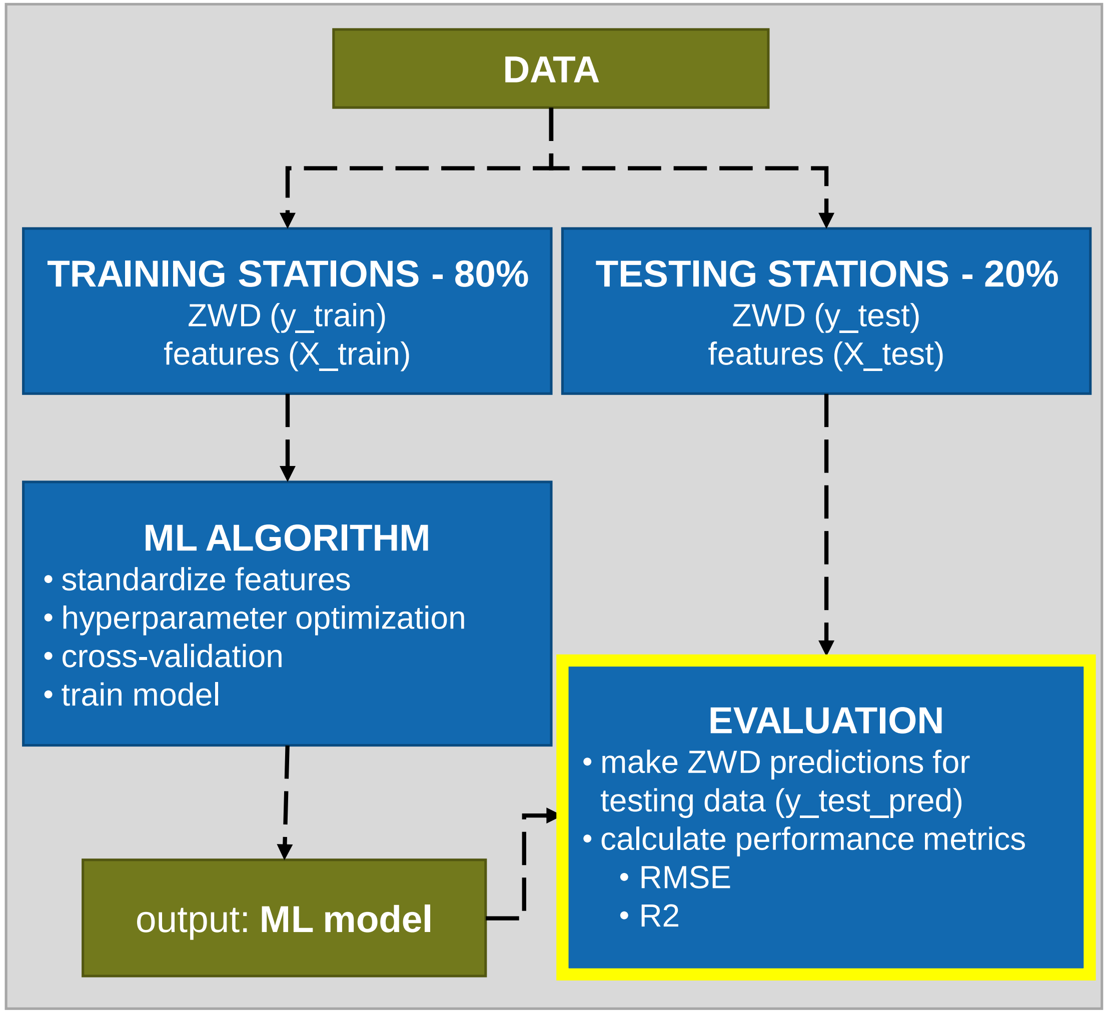


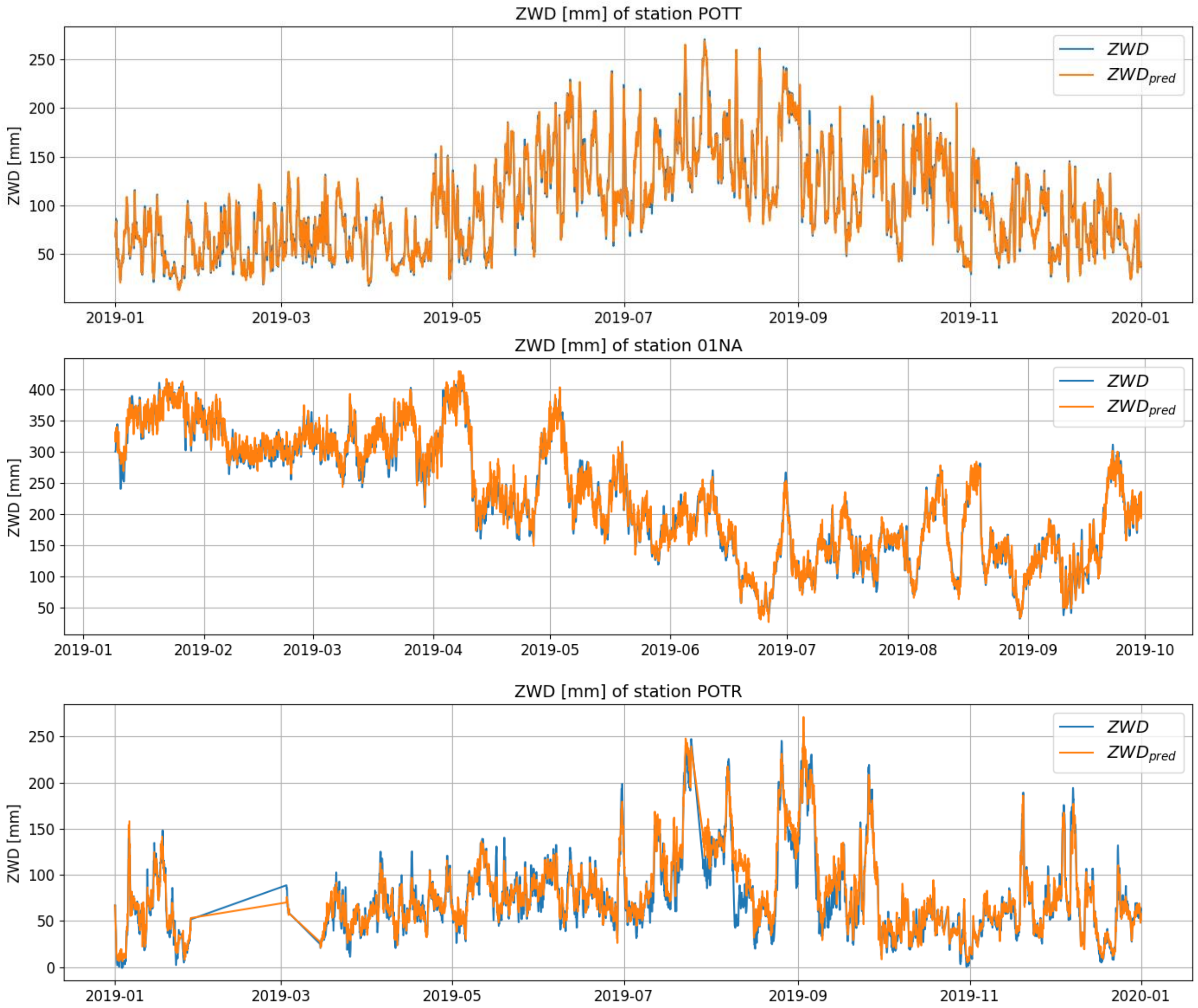
ML algorithms:

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

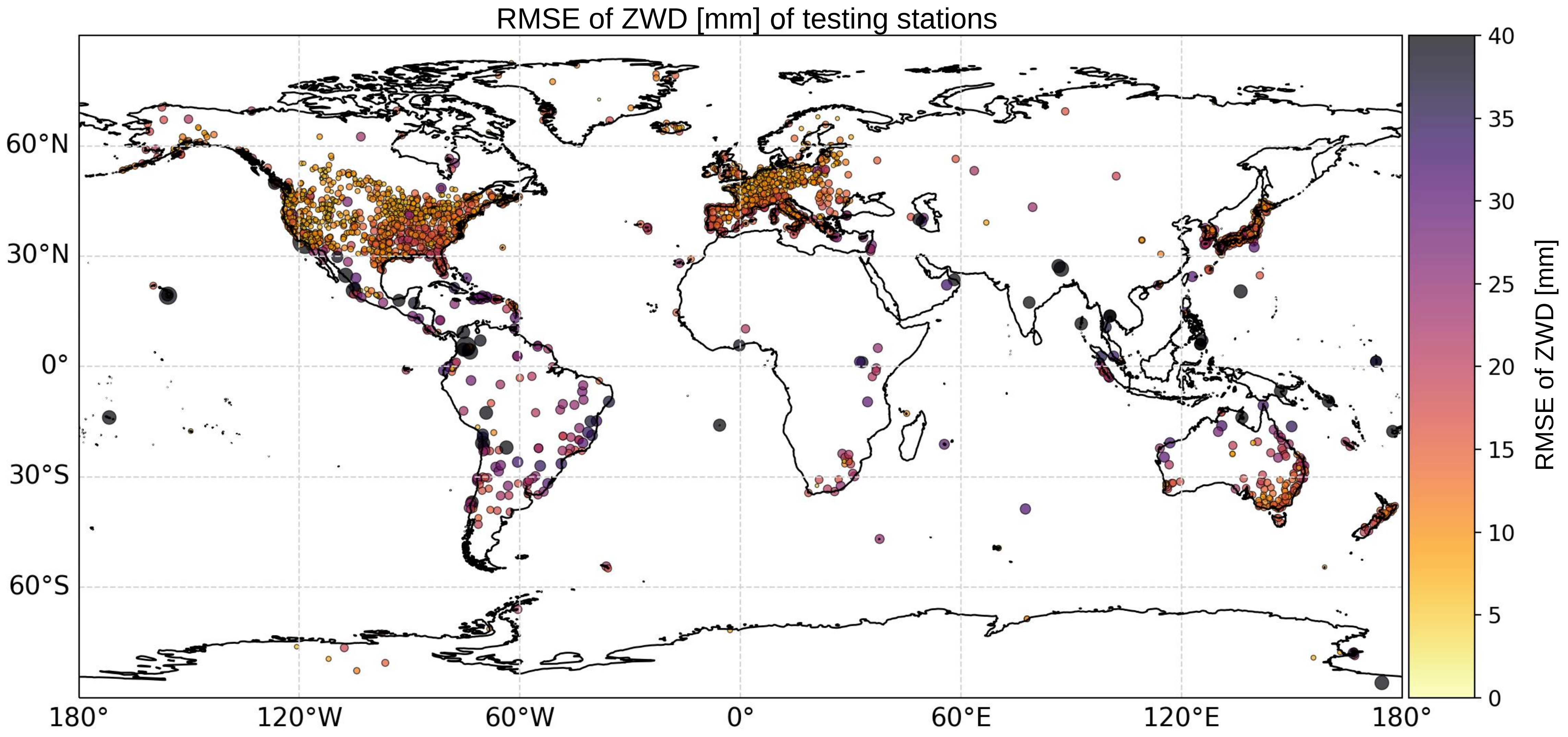


Setup

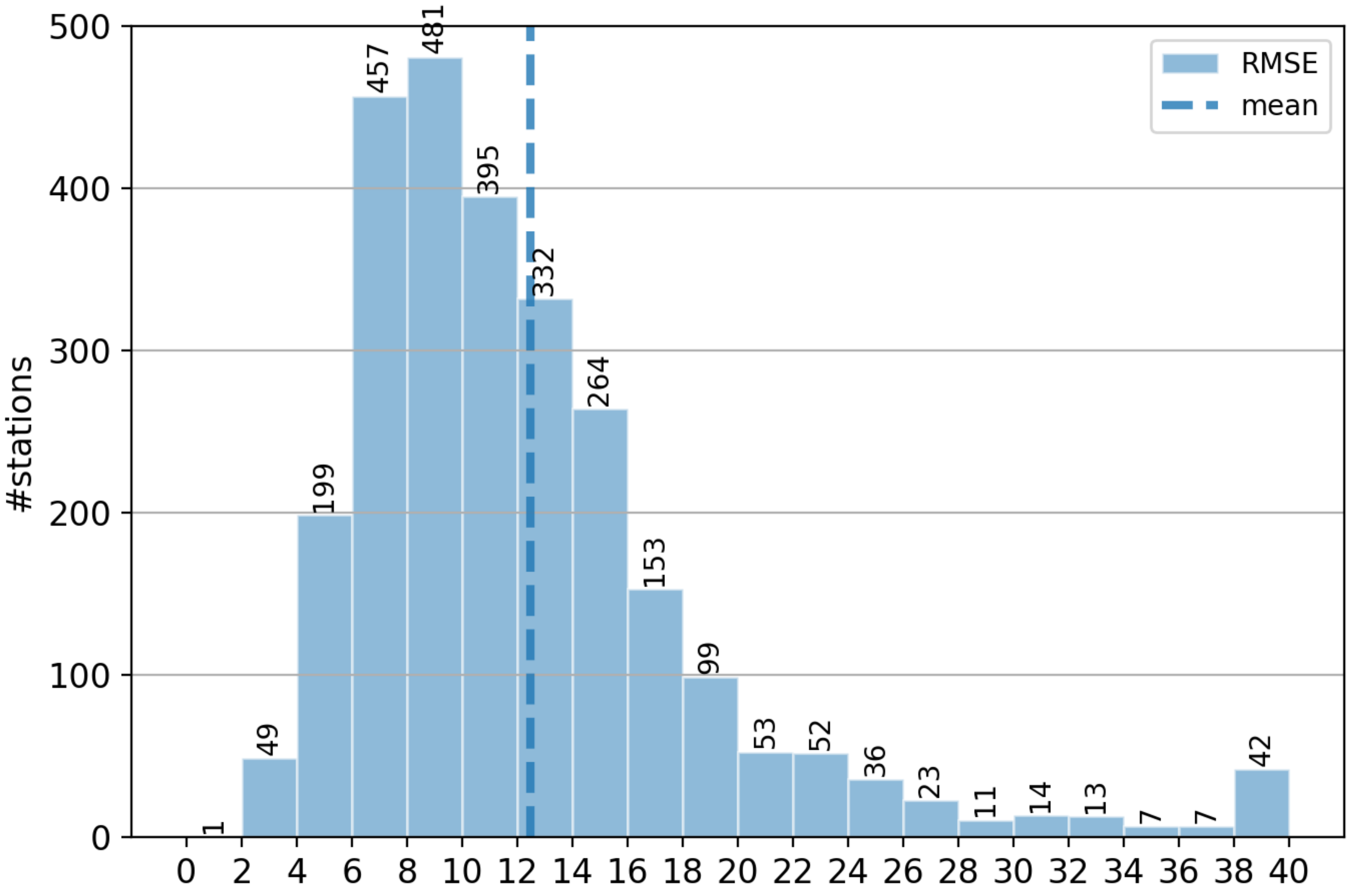

DATA
TRAINING STATIONS - 80%
ZWD (y_train)
features (X_train)
TESTING STATIONS - 20%
ZWD (y_test)
features (X_test)
ML ALGORITHM
• standardize features
• hyperparameter optimization
• cross-validation
• train model
output: ML model
EVALUATION
• make ZWD predictions for testing data (y_test_pred)
• calculate performance metrics
• RMSE
• R2


ZWD [mm] of station POTT
ZWD [mm] of station 01NA
ZWD [mm] of station POTR
ZWD [mm]
ZWD
ZWD_{pred}
2019-01
2019-02
2019-03
2019-04
2019-05
2019-06
2019-07
2019-08
2019-09
2019-10
2019-11
2020-01

Performance of individual test stations



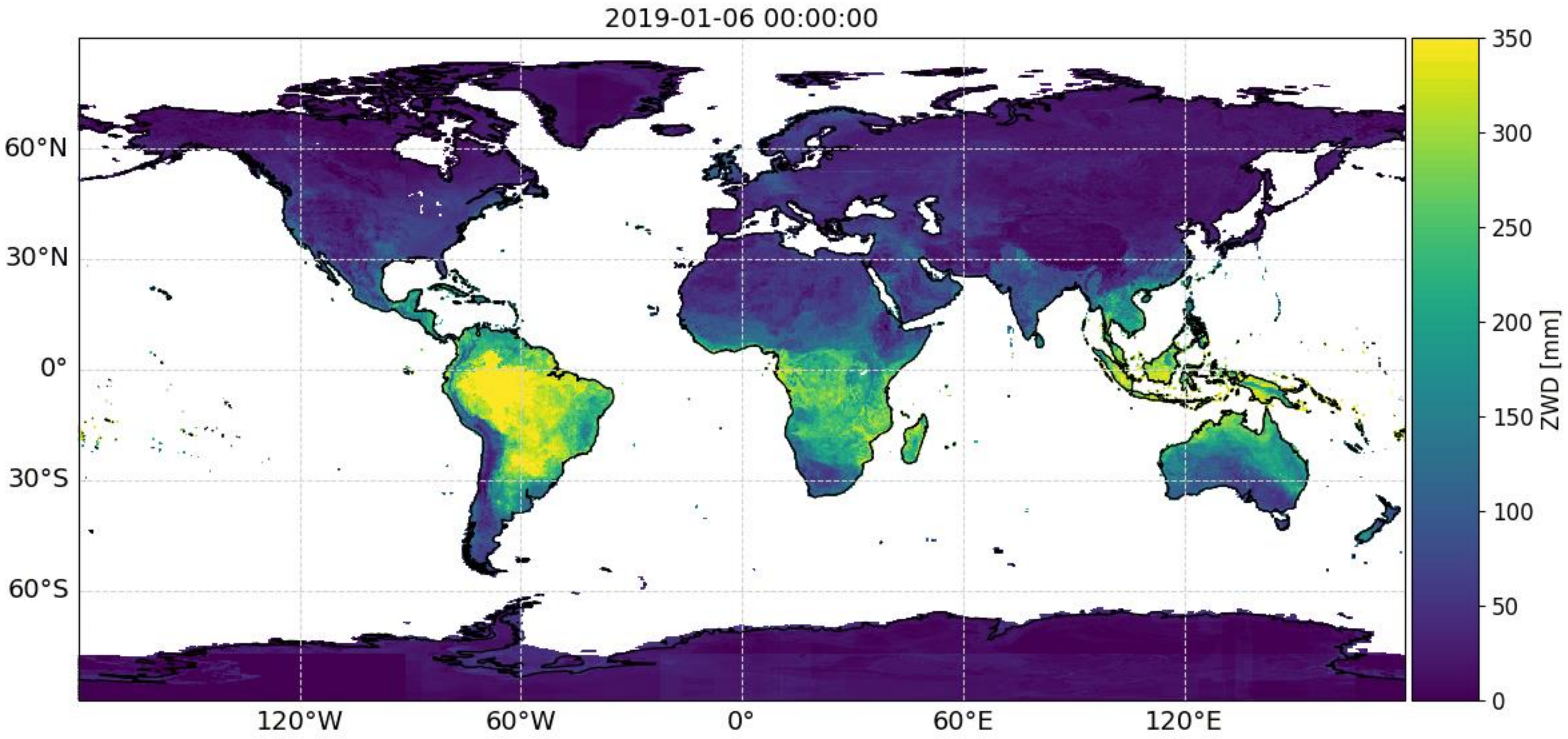
Performance of individual test stations



RMSE of ZWD [mm] of testing stations

mean = 12.4 mm

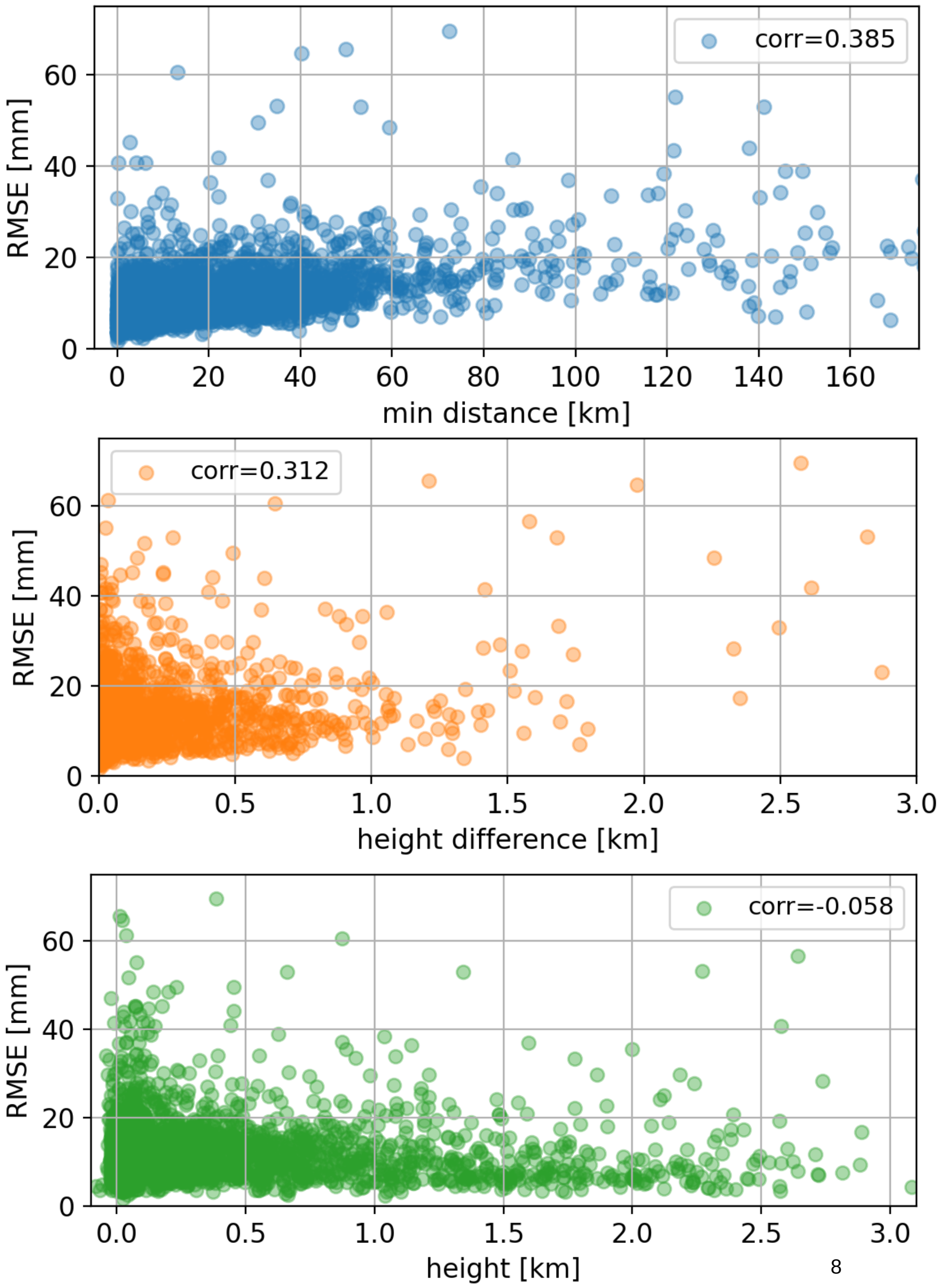
ZWD predictions for 2019



Understanding the performance

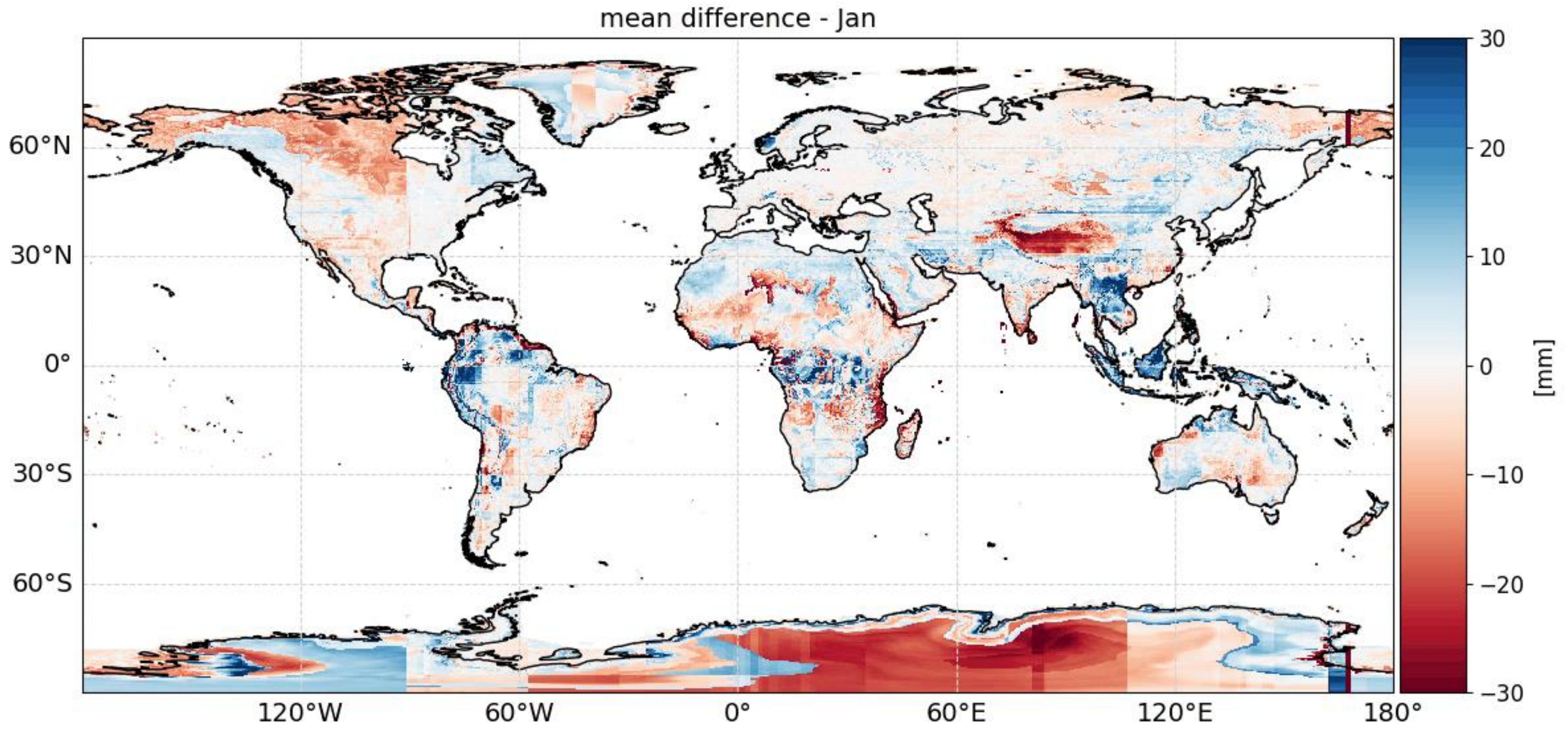
Comparison between RMSE and …

- minimum distance to next training station
  $\rightarrow \boldsymbol{corr = 0.34}$
- height difference to next training station
  $\rightarrow \boldsymbol{corr = 0.31}$
- height of GNSS station
  $\rightarrow \boldsymbol{corr = -0.06}$



Comparison between RF and XGBoost
Mean difference over each month for the year 2019

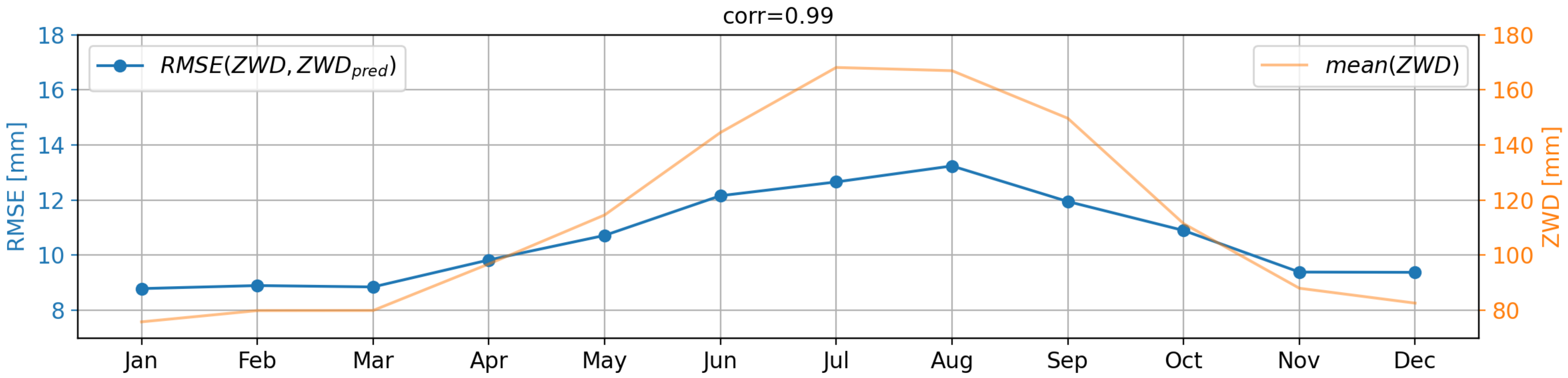
- Calculated ZWD predictions for year 2019, once with RF and once with XGBoost
- Took differences between RF predictions and XGBoost predictions
- Took mean over the differences for each month



Monthly models – performance for testing stations

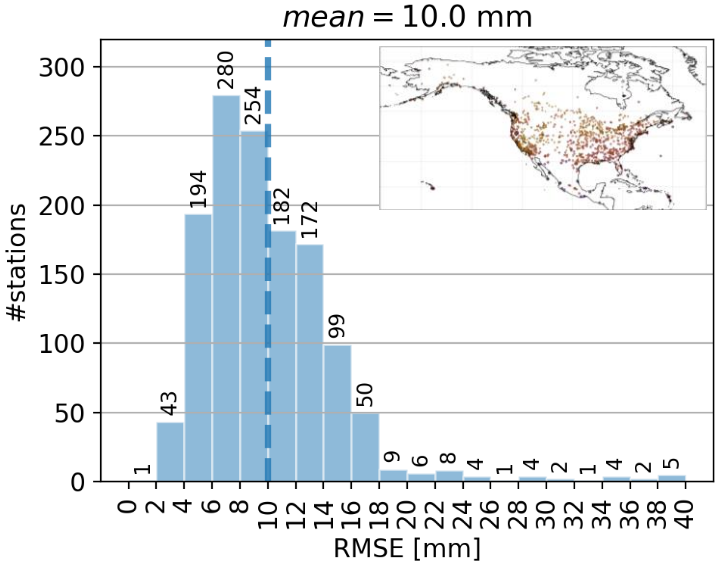
World 2019

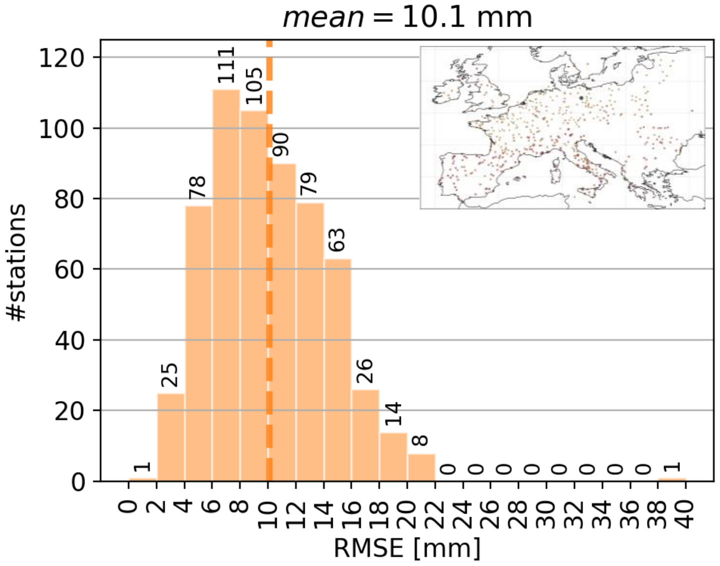
RMSE [mm]												
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	all months
8.8	8.9	8.9	9.8	10.7	12.1	12.7	13.2	12.0	10.9	9.4	9.4	**12.5**

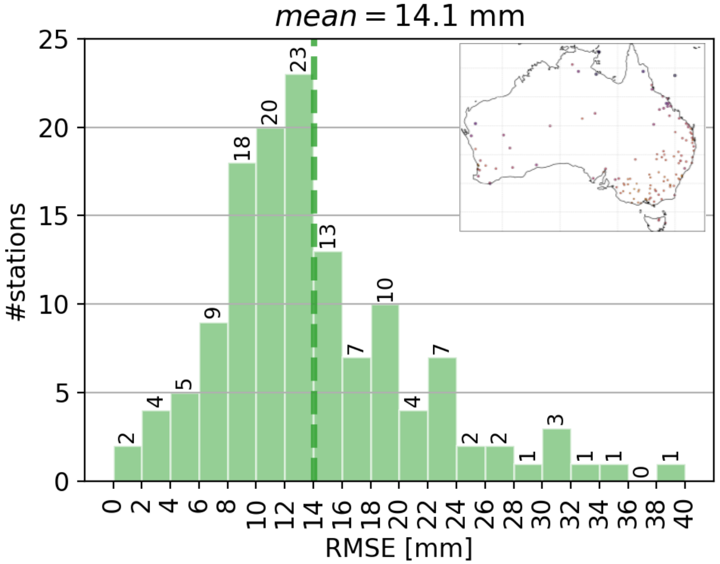


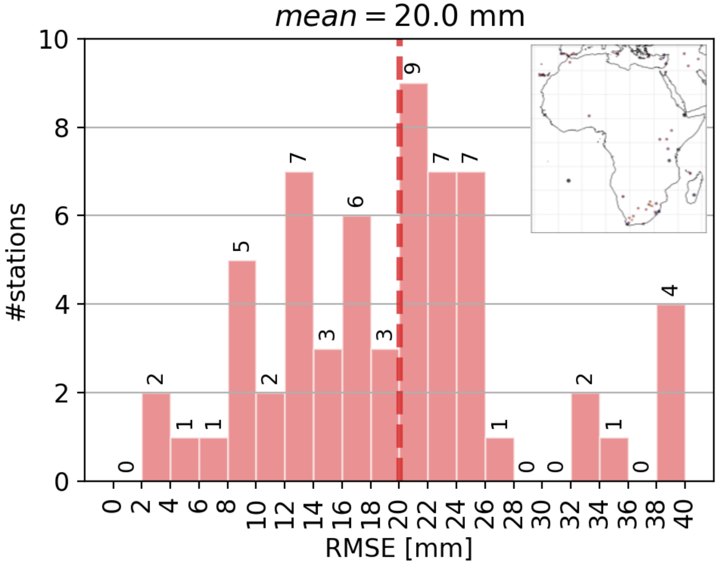
Regional models for the year 2019

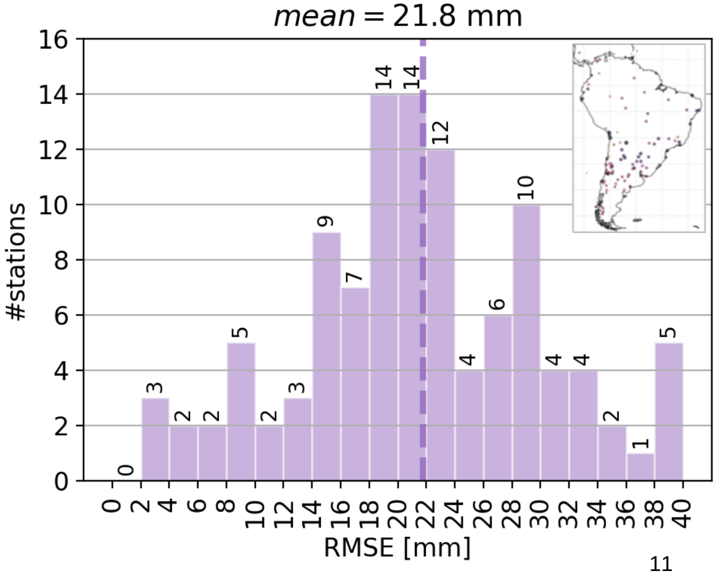
	RMSE [mm]	#stations
North America	10.0	6605
Europe	10.1	3004
WORLD	**12.5**	**13440**
Australia	14.1	665
Africa	20.0	302
South America	21.8	542







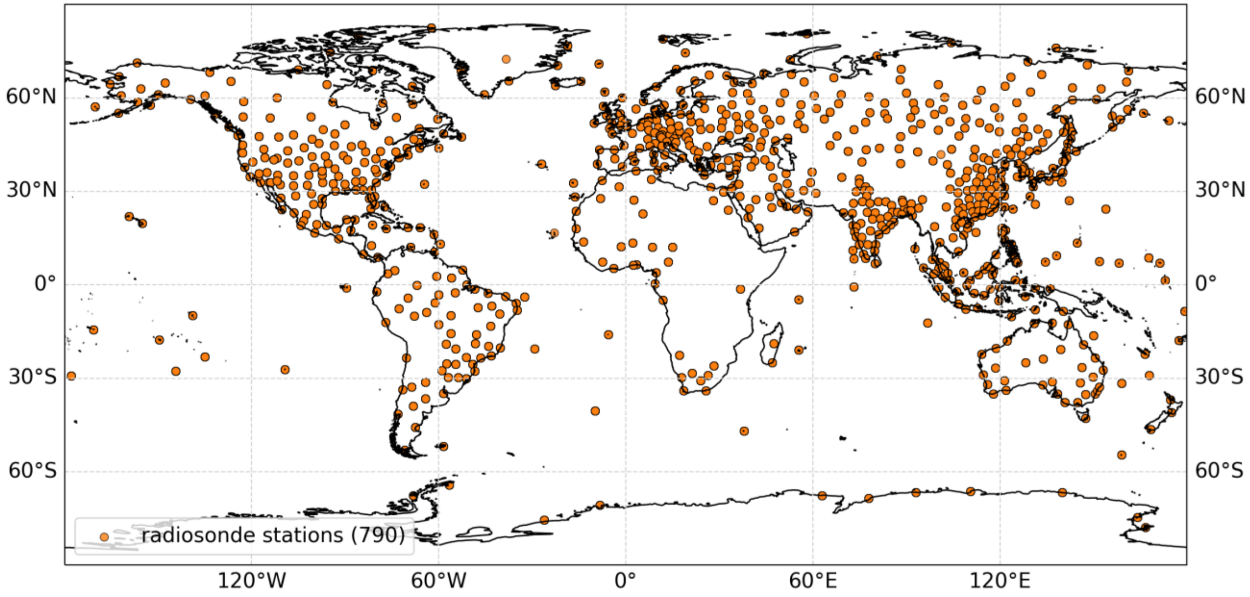




Validation of ZWD predictions with independent methods

- Use **radiosonde data** to calculate ZWD
- Several options for validation:
  - Compare ZWD at locations where radiosonde data are available
  - Compare ZWD at locations where additionally a GNSS station is available

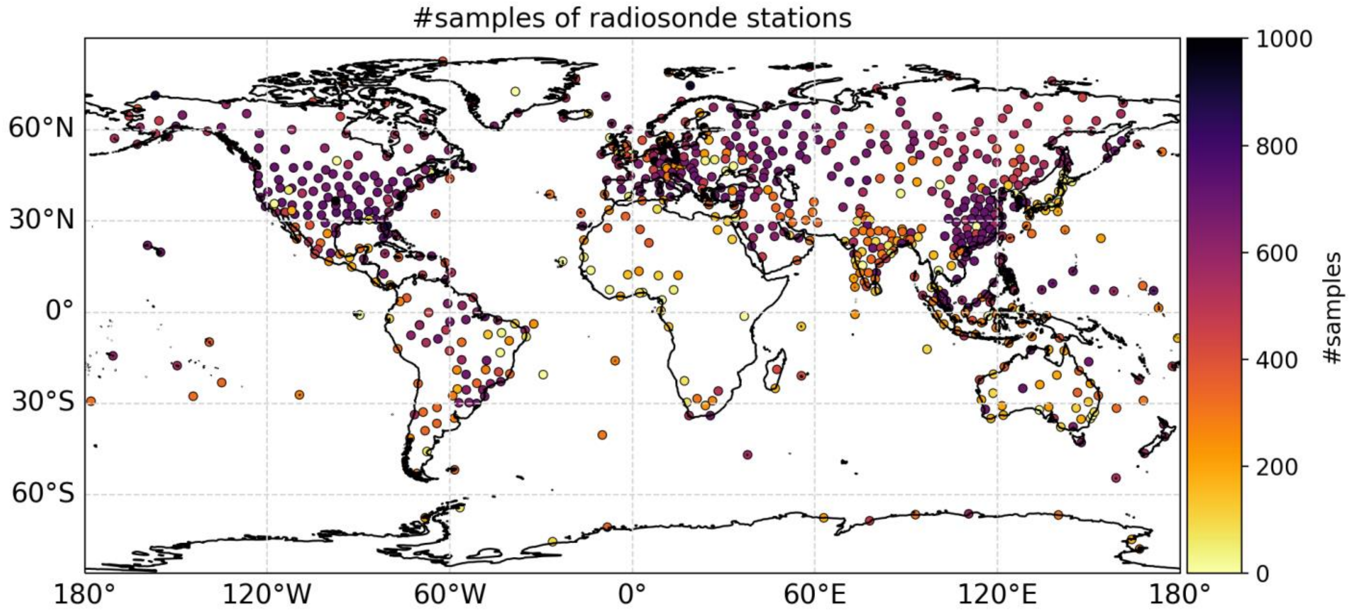
Overview of radiosonde stations for 2019

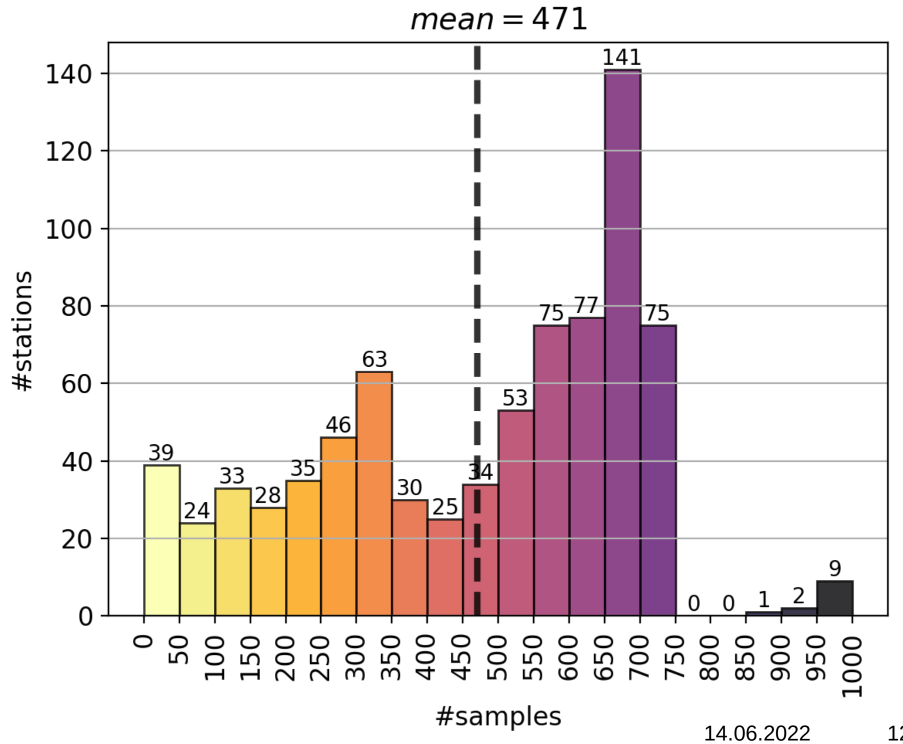


- 790 radiosonde stations available for 2019
- source:
  - Integrated Global Radiosonde Archive (IGRA)
  - https://www.ncei.noaa.gov/pub/data/igra/
- #samples differs for the stations

Validation of ZWD predictions with independent methods

- Use **radiosonde data** to calculate ZWD
- Several options for validation:
  - Compare ZWD at locations where radiosonde data are available
  - Compare ZWD at locations where additionally a GNSS station is available

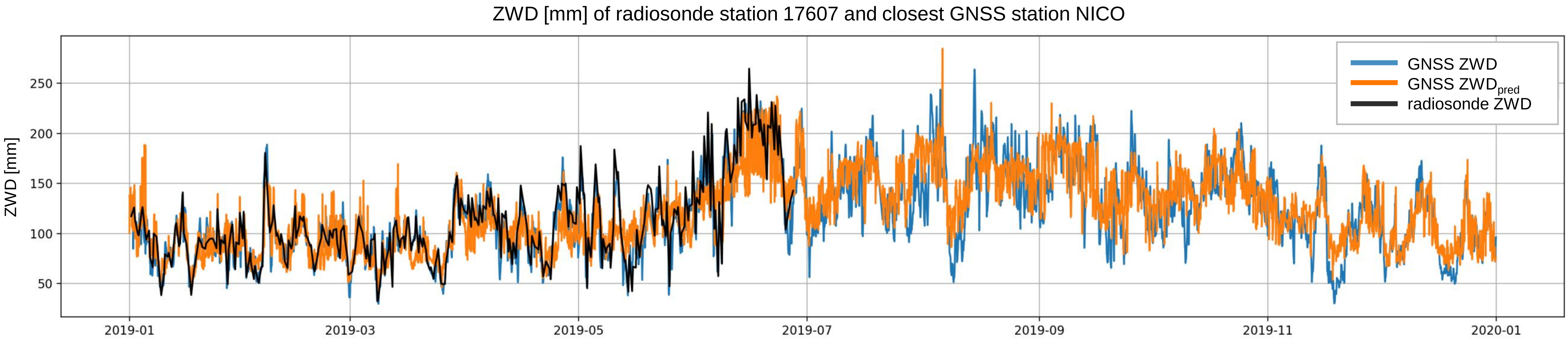




Validation of ZWD predictions with independent methods

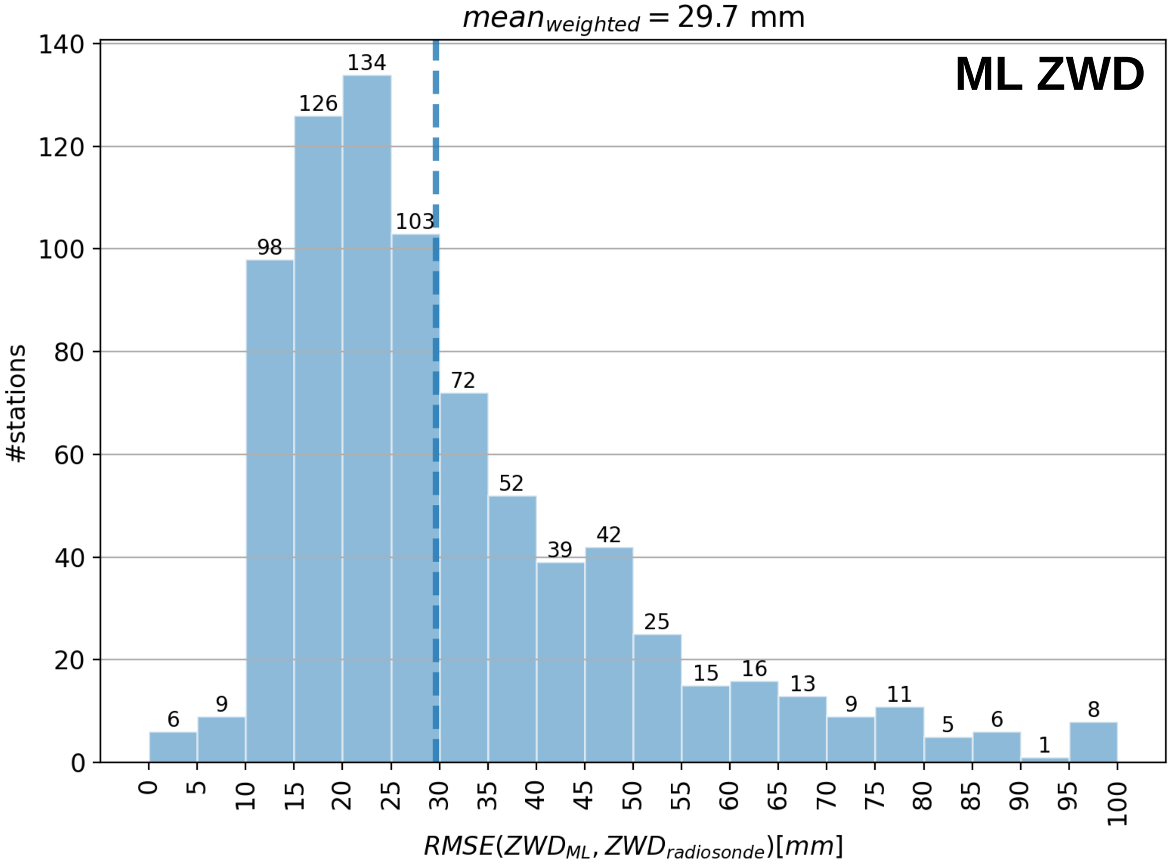
- Use **radiosonde data** to calculate ZWD
- Several options for validation:
  - Compare ZWD at locations where radiosonde data are available
  - Compare ZWD at locations where additionally a GNSS station is available

Example time series

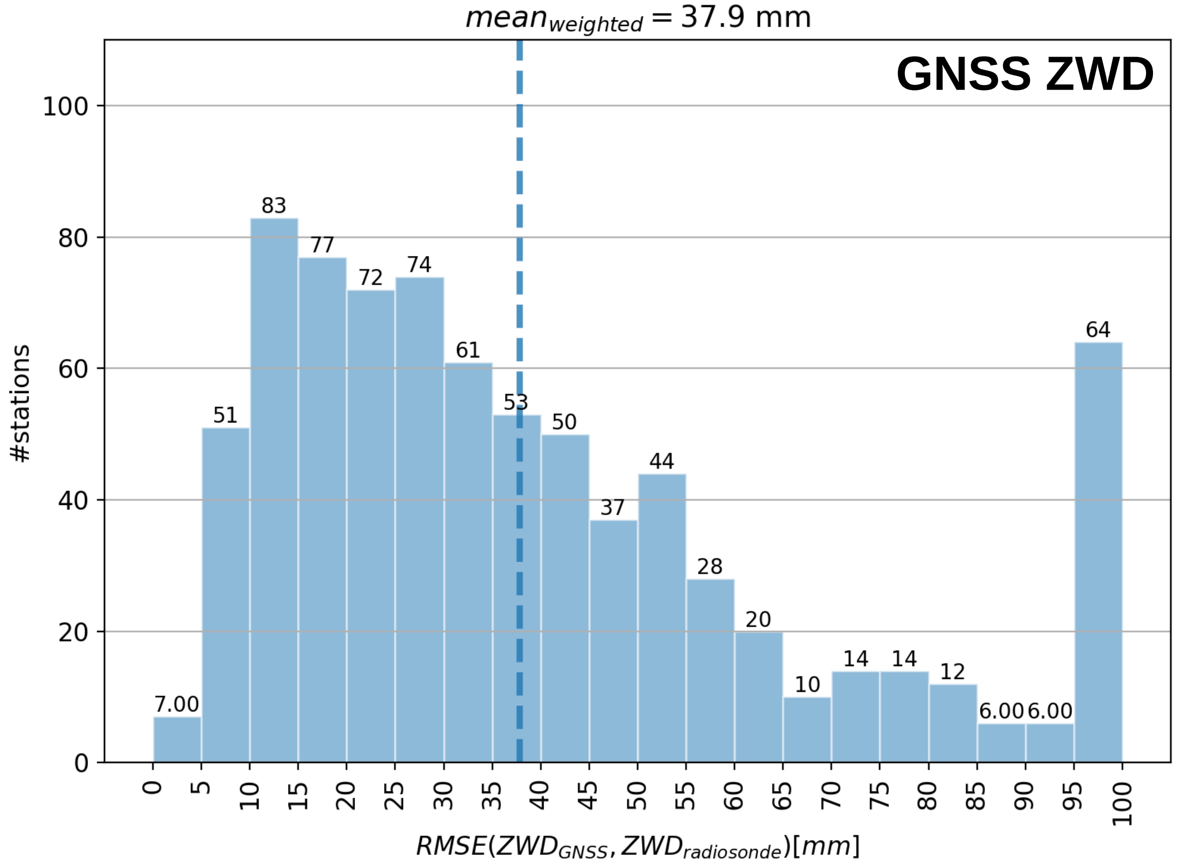


RMSE between ZWD of …

… radiosonde station and ML prediction



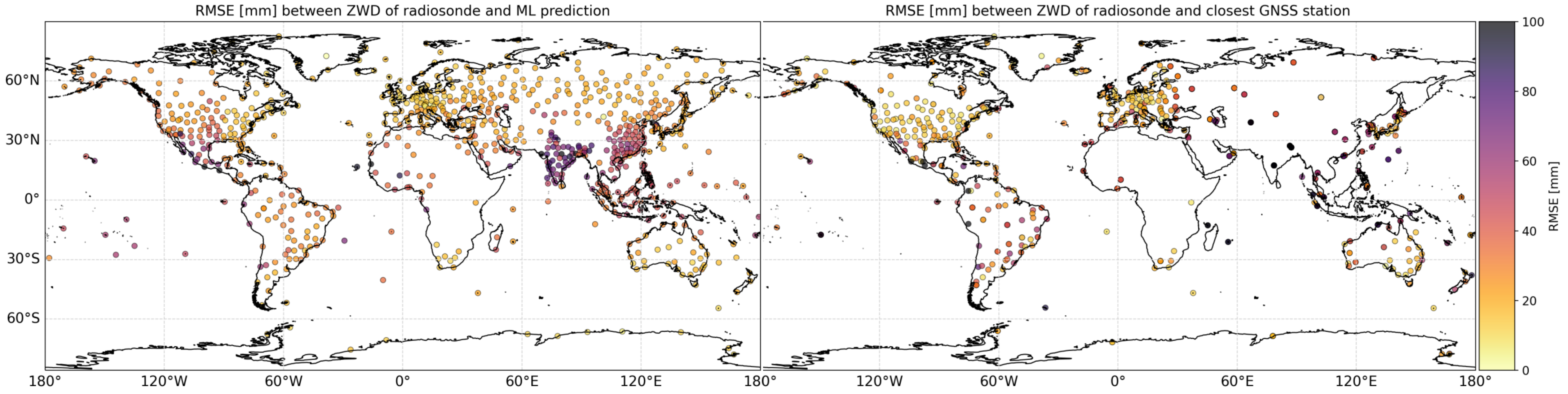
… radiosonde station and closest GNSS station



RMSE between ZWD of …

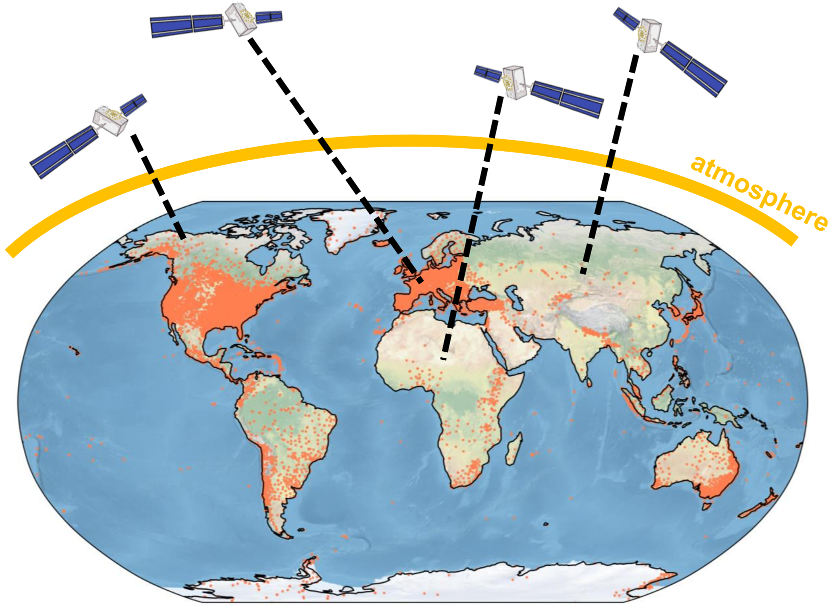
… radiosonde station and ML prediction

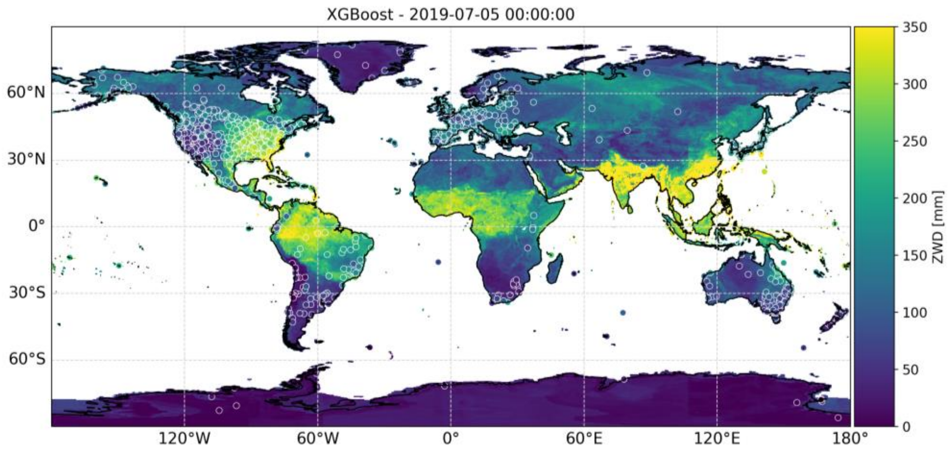
… radiosonde station and closest GNSS station



Summary

- Global ML-based ZWD model based on meteorological data
- Can make predictions of ZWD for every point on Earth (assuming meteorological data is available)
- Achieves an average RMSE of **12.4 mm**
- ML algorithms: **XGBoost** and Random Forest





ETH zürich

Laura Crocetti
lcrocetti@ethz.ch

ETH Zurich

Space Geodesy Group

Institute of Geodesy and Photogrammetry

Zurich, Switzerland

www.space.igp.ethz.ch