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 $\rightarrow$ 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 $\rightarrow$ measurements of atmospheric properties
- Model Zenith Wet Delay globally based on meteorological data using Machine Learning (ML)
Data

**Zenith Wet Delay (ZWD)**
- 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

**TARGET**
- 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

Laura Crocetti – lcrocetti@ethz.ch
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

Global model for the year 2019

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

Laura Crocetti – lcrocetti@ethz.ch
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

---

**Target**

<table>
<thead>
<tr>
<th>Date</th>
<th>Lat</th>
<th>Lon</th>
<th>Time</th>
<th>Day</th>
<th>Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019-01-01</td>
<td>40.33492</td>
<td>18.11149</td>
<td>01:36:00.0000</td>
<td>01:00:00</td>
<td>0.0000</td>
</tr>
<tr>
<td>2019-01-02</td>
<td>40.33492</td>
<td>18.11149</td>
<td>01:35:00.0000</td>
<td>01:00:00</td>
<td>0.0000</td>
</tr>
<tr>
<td>2019-01-03</td>
<td>40.33492</td>
<td>18.11149</td>
<td>01:34:00.0000</td>
<td>01:00:00</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**Features**

<table>
<thead>
<tr>
<th>Date</th>
<th>Lat</th>
<th>Lon</th>
<th>Time</th>
<th>Day</th>
<th>Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019-01-01</td>
<td>40.33492</td>
<td>18.11149</td>
<td>01:36:00.0000</td>
<td>01:00:00</td>
<td>0.0000</td>
</tr>
<tr>
<td>2019-01-02</td>
<td>40.33492</td>
<td>18.11149</td>
<td>01:35:00.0000</td>
<td>01:00:00</td>
<td>0.0000</td>
</tr>
<tr>
<td>2019-01-03</td>
<td>40.33492</td>
<td>18.11149</td>
<td>01:34:00.0000</td>
<td>01:00:00</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

---

Laura Crocetti – lcrocetti@ethz.ch
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 ALGORITHM**
- standardize features
- hyperparameter optimization
- cross-validation
- train model

**EVALUATION**
- make ZWD predictions for testing data ($y_{\text{test}}$, $\hat{y}_{\text{test}}$)
- calculate performance metrics
  - RMSE
  - R2

**TRAINING STATIONS** - 80%
- ZWD ($y_{\text{train}}$)
- features ($X_{\text{train}}$)

**TESTING STATIONS** - 20%
- ZWD ($y_{\text{test}}$)
- features ($X_{\text{test}}$)

**output:** ML model

**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

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

**output:** ML model

Laura Crocetti – lcrocetti@ethz.ch
Performance of individual test stations

RMSE of ZWD [mm] of testing stations
Performance of individual test stations

mean = 12.4 mm
ZWD predictions for 2019
Understanding the performance

Comparison between RMSE and ...

- minimum distance to next training station
  \[ corr = 0.34 \]

- height difference to next training station
  \[ corr = 0.31 \]

- height of GNSS station
  \[ corr = -0.06 \]
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

<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>all months</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE [mm]</td>
<td>8.8</td>
<td>8.9</td>
<td>8.9</td>
<td>9.8</td>
<td>10.7</td>
<td>12.1</td>
<td>12.7</td>
<td>13.2</td>
<td>12.0</td>
<td>10.9</td>
<td>9.4</td>
<td>9.4</td>
<td>12.5</td>
</tr>
</tbody>
</table>

corr=0.99

Laura Crocetti – lcrocetti@ethz.ch
### Regional models for the year 2019

<table>
<thead>
<tr>
<th>Region</th>
<th>RMSE [mm]</th>
<th>#stations</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>10.0</td>
<td>6605</td>
</tr>
<tr>
<td>Europe</td>
<td>10.1</td>
<td>3004</td>
</tr>
<tr>
<td>WORLD</td>
<td>12.5</td>
<td>13440</td>
</tr>
<tr>
<td>Australia</td>
<td>14.1</td>
<td>665</td>
</tr>
<tr>
<td>Africa</td>
<td>20.0</td>
<td>302</td>
</tr>
<tr>
<td>South America</td>
<td>21.8</td>
<td>542</td>
</tr>
</tbody>
</table>

**Mean RMSE by Region**

- **North America**: 10.0 mm
- **Europe**: 10.1 mm
- **WORLD**: 12.5 mm
- **Australia**: 14.1 mm
- **Africa**: 20.0 mm
- **South America**: 21.8 mm
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

[ZWD [mm] of radiosonde station 17607 and closest GNSS station NICO](chart)
RMSE between ZWD of ... radiosonde station and ML prediction

mean_{weighted} = 29.7 mm

ML ZWD

... radiosonde station and closest GNSS station

mean_{weighted} = 37.9 mm

GNSS ZWD
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
Laura Crocetti
lcrocetti@ethz.ch

ETH Zurich
Space Geodesy Group
Institute of Geodesy and Photogrammetry
Zurich, Switzerland
www.space.igp.ethz.ch