Development of a global model for zenith wet delays based on the random forest approach

Qinzheng Li[1,2], Johannes Böhm[1], Linguo Yuan[2], Robert Weber[1]

[1]Department of Geodesy and Geoinformation, Vienna University of Technology (TU Wien), Austria

[2]Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, China

EGU 2022 - European Geosciences Union General Assembly
Session G1.3 - Data Science and Machine Learning in Geodesy
May 23-27, 2022, Wien, Austria

Introduction

- Tropospheric delay is a major error source in the analysis of space geodetic observations

  High correlation between **residual delay** and other parameters of interest such as **station height** and receiver clock bias

  Incorrect modeling of tropospheric (wet) delay →
  - affects precision and accuracy of position solution
  - causes mis-interpretation of geophysical signals such as crustal deformation, terrestrial water storage variations

- Performance of tropospheric delay modeling is significantly limited due to the **high dynamics** of the **wet portion** in the neutral atmosphere

- Several positioning and navigation tasks such as **real-time applications** do not have the benefit of post-processing analyses, necessitating the availability of **accurate a-priori estimates for ZWD**.

Introduction

Ray-tracing through NWMs ⇨ Require the knowledge of **vertical profiles of temperature and humidity**

Saastamoinen (1972) model ⇨ Unable to account for spatial and temporal variations in the relationship

$$\mathrm{ZWD}=\frac{0.002277\cdot\left(\frac{1255}{T}+0.05\right)\cdot e_s}{1-0.00266\cos(2\varphi)-2.8\cdot10^{-7}h_s}$$

Global empirical troposphere models such as GPT3 ⇨ Unable to capture **daily** and **complex variations** of meteorological parameters

How can we **improve ZWD prediction accuracy** if meteorological data are available at some stations?

RF-based ZWD model

$$\mathrm{ZWD}_{\mathrm{RF}}=\begin{cases} f_{\mathrm{blind}}(\mathrm{DOY},\mathrm{lat},\mathrm{lon},\mathrm{h}_{\mathrm{ell}}) & \text{Without meteorological data} \\ f(\mathrm{DOY},\mathrm{lat},\mathrm{lon},\mathrm{h}_{\mathrm{ell}},e_s,T_s) & \text{Meteorological data are available} \end{cases}$$

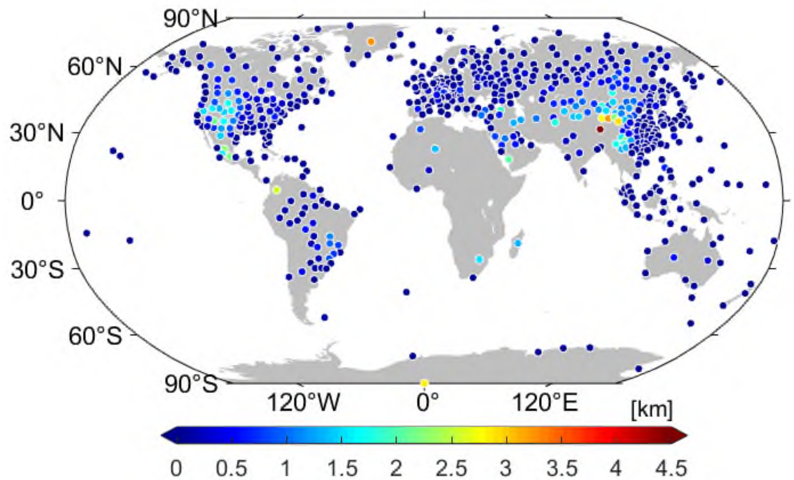
Dataset and Methodology

Radiosonde Measurements

Meteorological variables at discrete isobaric levels (usually twice per day)

Zenith wet delay (ZWD) are determined by the numerical integration of nonhydrostatic refractivity from the site's height to top sounding profiles

$$\mathrm{ZWD} = 10^{-6}\int_{h_s}^{\mathrm{top}} N_{nh}dh = 10^{-6}\int_{h_s}^{\mathrm{top}} [(k_2 - \frac{R_d}{R_w}k_1)\frac{p_w}{T} + k_3\frac{p_w}{T^2}]\cdot Z_w^{-1}dh$$



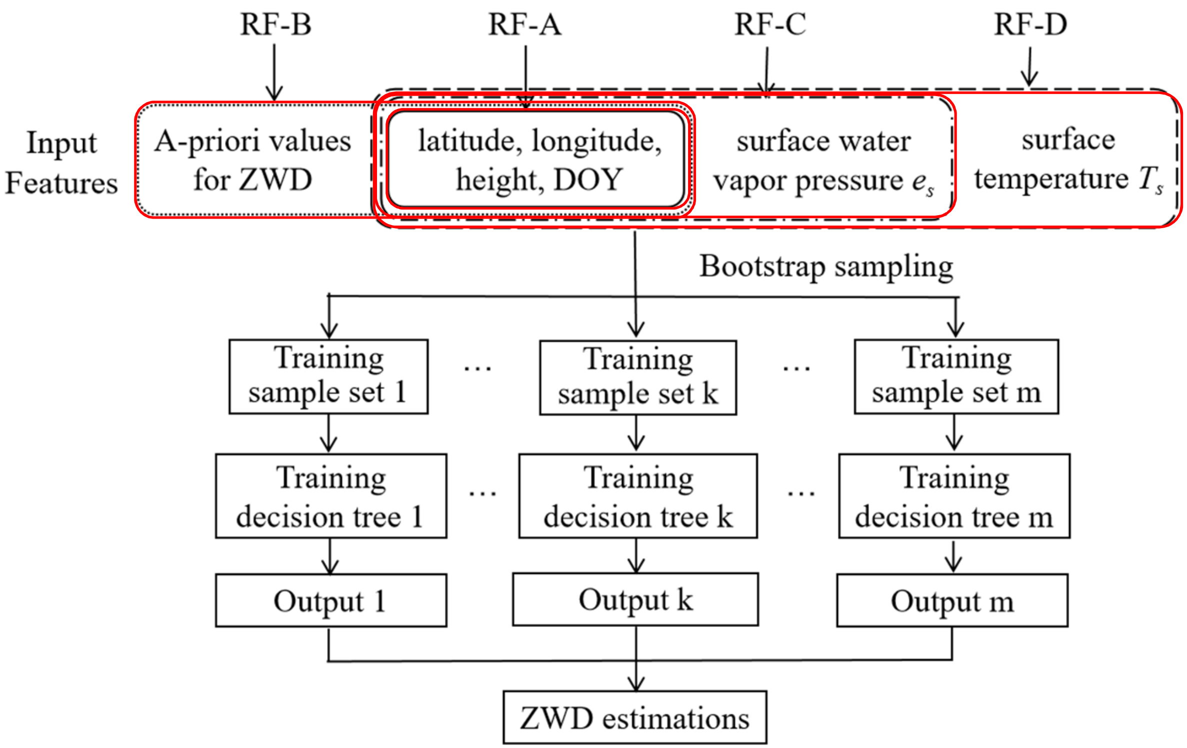
Modeling: 10 years (2010-2019) of radiosonde data at 586 sites

Validation: Sounding profiles for the year 2020

Dataset and Methodology

ZWD modeling based on the RF regression

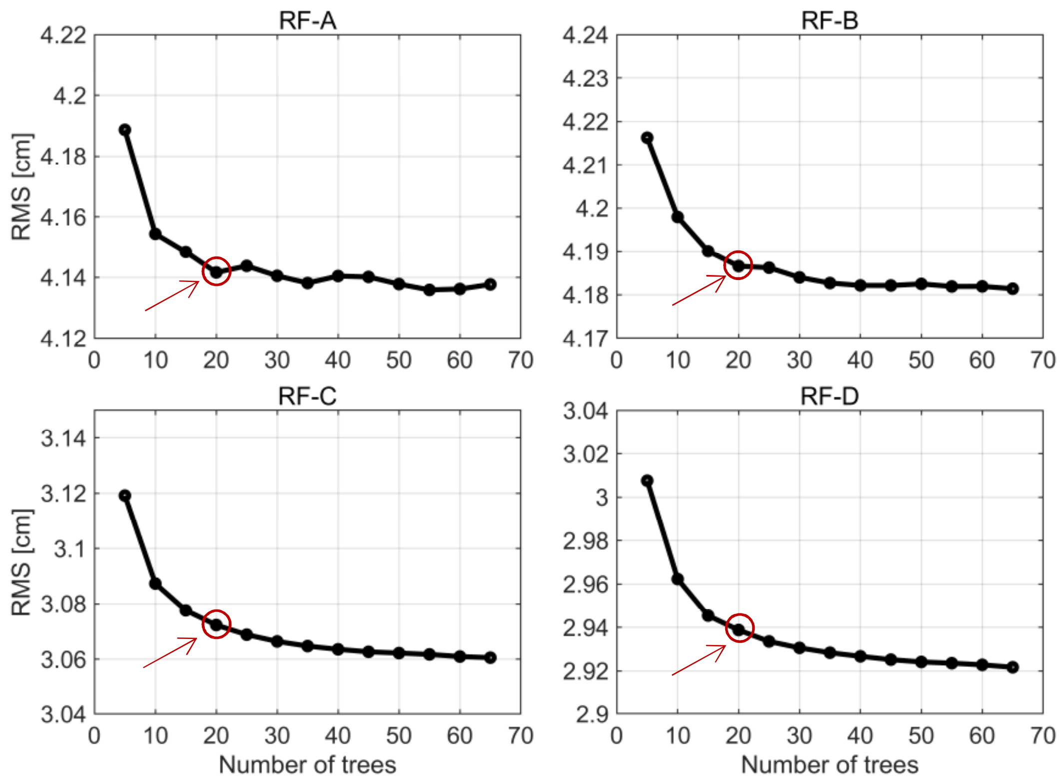
Functional formulation (RF-A): $ZWD_{RF-A} = f_a(DOY, lat, lon, h_{ell})$



Dataset and Methodology

Optional number of trees for a random forest system

10-fold **cross validation** RMS of ZWD differences between modeled and 'observed' values



ZWD modeling accuracy tends to be stable when the **number of trees** is larger than **20**

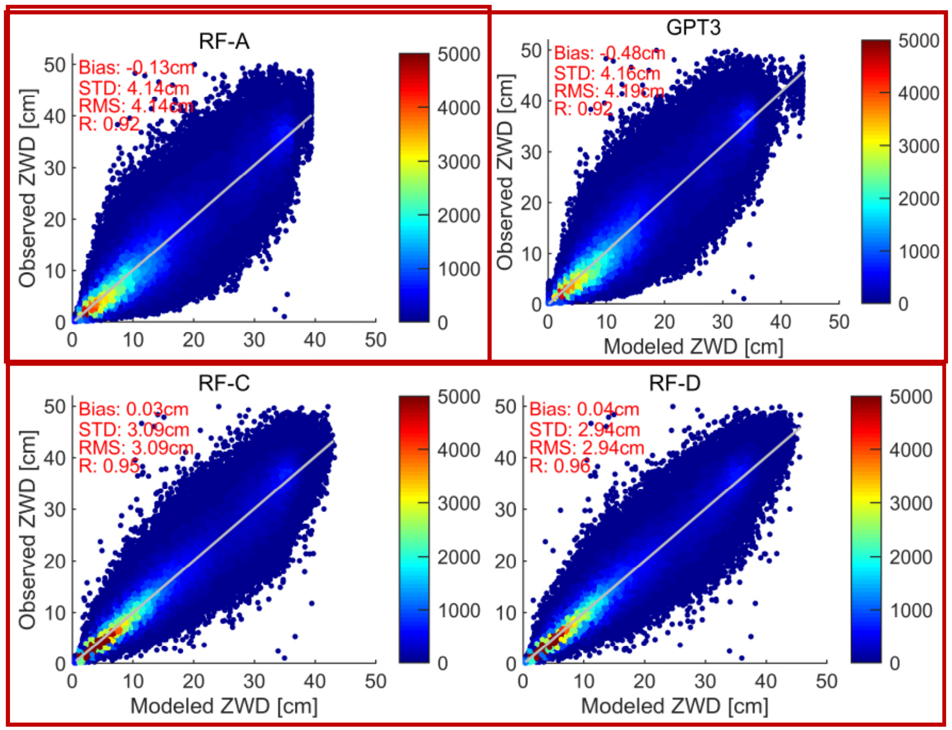
Ensemble learning can help the RF-based ZWD model to achieve **better generalization ability**

Results

Overall performance of the RF-based ZWD models

RF-A achieves a bias of -0.1 cm and RMS error of 4.1 cm, which is comparable to GPT3

ZWD predicting accuracy is significantly improved by adding surface meteorological parameters in the functional formulation , especially for $\mathbf{e_s}$



$$\mathrm{ZWD}_{\text{RF-A}} = f_a(\mathrm{DOY}, \mathrm{lat}, \mathrm{lon}, \mathrm{h}_{\mathrm{ell}})$$

$$\mathrm{ZWD}_{\text{RF-B}} = f_b(\mathrm{DOY}, \mathrm{lat}, \mathrm{lon}, \mathrm{h}_{\mathrm{ell}}, \mathrm{ZWD}_{\mathrm{priori}})$$

$$\mathrm{ZWD}_{\text{RF-C}} = f_c(\mathrm{DOY}, \mathrm{lat}, \mathrm{lon}, \mathrm{h}_{\mathrm{ell}}, e_s)$$

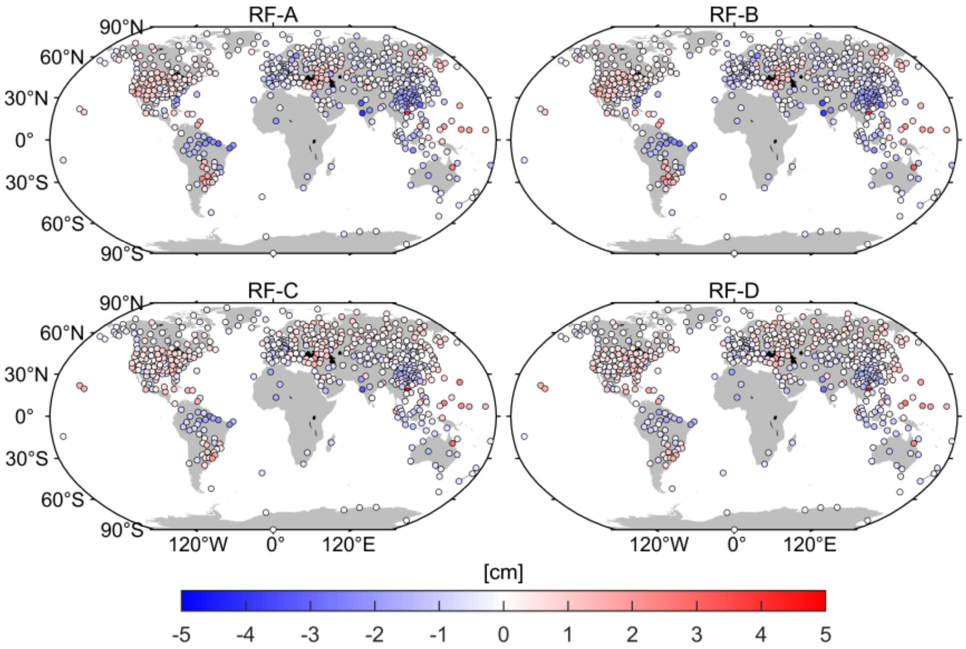
$$\mathrm{ZWD}_{\text{RF-D}} = f_d(\mathrm{DOY}, \mathrm{lat}, \mathrm{lon}, \mathrm{h}_{\mathrm{ell}}, e_s, T_s)$$

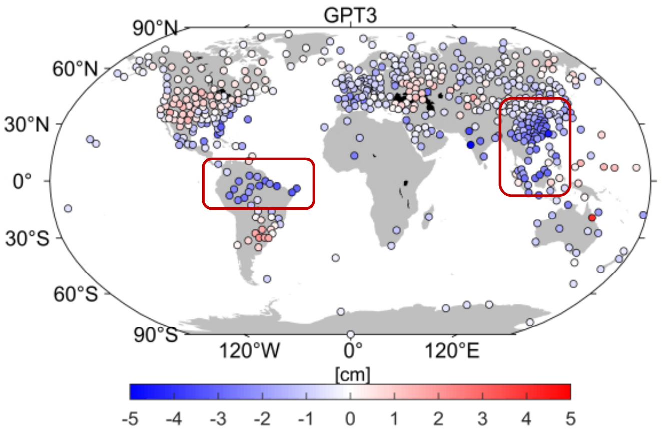
Model	Bias [cm]	STD [cm]	RMS [cm]	R
GPT3	-0.48	4.16	4.19	0.92
RF-A	-0.13	4.14	4.14	0.92
RF-B	-0.12	4.16	4.16	0.92
RF-C	0.03	3.09	3.09	0.95
RF-D	0.04	2.94	2.94	0.96

Results

Spatial features of model accuracy

RF-based ZWD models mitigate **negative biases** in the regions with **monsoon climate** and **tropical rainforest climate** types, in particular with the introduction of **surface meteorological information**

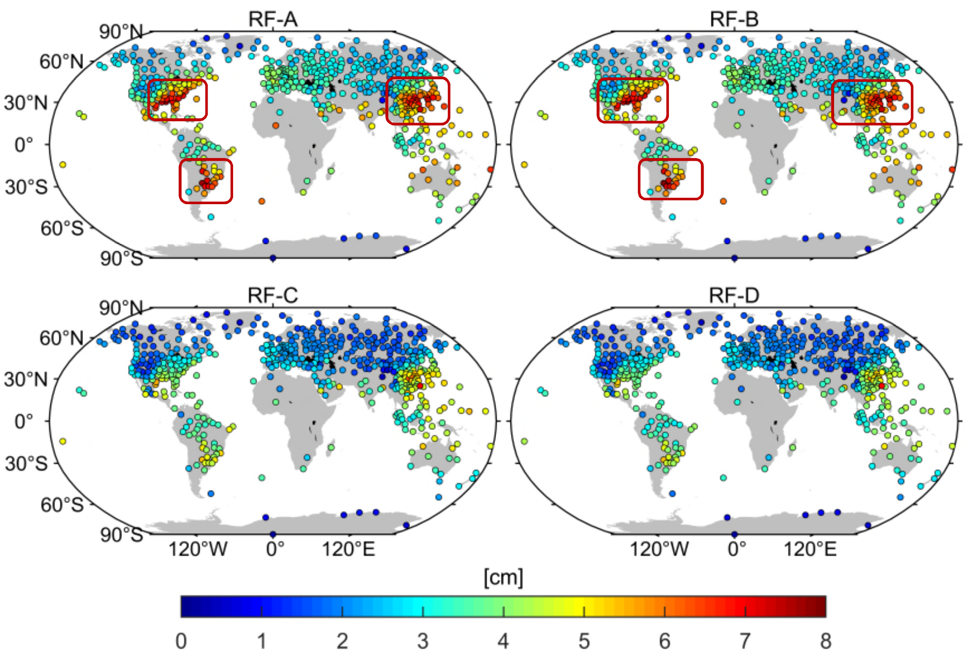




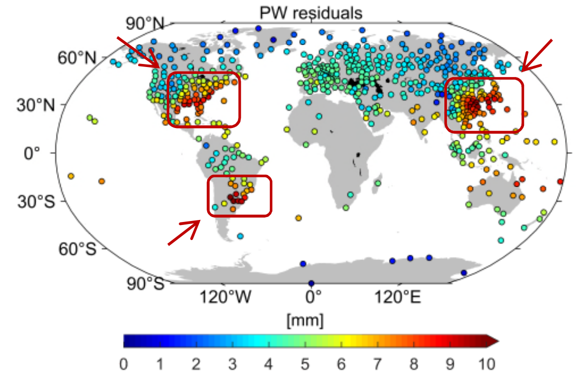
Spatial features of model accuracy

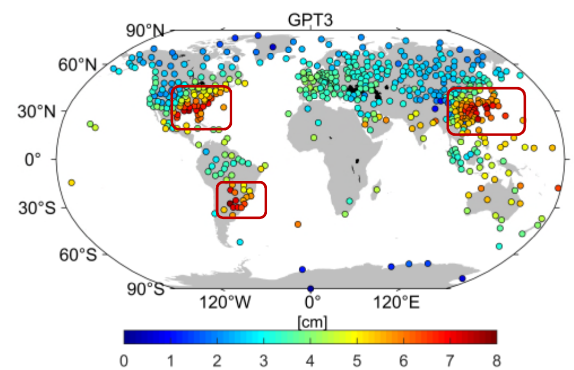
Complex variations of the **moisture in the lower atmosphere** across the globe by considering the **surface meteorological parameters** in the modeling

CWD prediction accuracy for RF-C and RF-D is clearly improved for the regions with **monsoon climate** type



Mean values and **seasonal components** in PW time series were removed





Summary

- The RF-based ZWD model with meteorological parameterization (**RF-D**) can achieve an overall accuracy of **2.9 cm** and a bias close to zero

- ZWD modeling accuracy is **significantly improved** by the introduction of **surface water vapor pressure** into functional formulation in comparsion with the **surface temperature**

- The RF-based ZWD models clearly mitigate **negative biases** in the regions with **monsoon climate** and **tropical rainforest climate** type

- Compared with the GPT3, ZWD modeling accuracy can be improved across the globe by considering **surface meteorological parameters**, especially in the regions with monsoon climate type

Thank you!

The subroutine of GZWD-RF model can be freely accessed from the VMF Data Server

https://vmf.geo.tuwien.ac.at/codes/gzwd_rf/

Qinzheng Li

e-mail: qinzheng.li@geo.tuwien.ac.at

Department of Geodesy and Geoinformation
Vienna University of Technology
Wiedner Hauptstraße 8
1040 Wien, Austria

Faculty of Geosciences and Environmental Engineering
Southwest Jiaotong University
Xi'an Road 999
Chengdu 611756, China