TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Modeling of the weighted mean temperature based on the random forest machine learning 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), Wien 1040, Austria

[2] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China

EGU 2023 - European Geosciences Union General Assembly
Session G1.3 - New Developments in Mathematical Methods in Geodesy, with A Focus on Machine Learning
April 24-28, Wien, Austria

Introduction

Weighted mean temperature (T_m) is an important parameter in the atmospheric water vapor retrieval with the GNSS technique and the determination of a priori values for zenith non-hydrostatic delays (ZWD) based on the formula by Asken and Nordius (1987)

$$\mathrm{PW} = \mathrm{ZWD} \cdot \Pi_{(T_m)} = (\mathrm{ZTD} - \mathrm{ZHD}) \cdot \Pi_{(T_m)}$$

$$\Pi_{(T_m)} = \frac{10^6}{\rho_w R_v [\frac{k_3}{T_m} + (k_2 - \frac{R_d}{R_v} k_1)]}$$

$$T_m = \frac{\int \frac{e}{T} dh}{\int \frac{e}{T^2} dh} \approx \frac{\sum_{k=1}^{N} \frac{e_k}{T_k} \Delta h_k}{\sum_{k=1}^{N} \frac{e_k}{T_k^2} \Delta h_k}$$

(1) Numerical integration — requires temperature and humidity vertical profiles

(2) Global emprícial models such as GPT3 (Landskron and Böhm 2018)` — cannot well represent daily and complex variations

(3) T_m-T_s linear regression models Bevis formula (Bevis et al. 1994) — does not consider spatiotemporal variations and have poor accuracy in oceanic and polar regions

T_m modeling with surface meteorological parameters through the random forest machine learning

Dataset

Radiosonde data and GPS radio occultation measurements

Modeling: 4 years (**2016-2019**) of global atmospheric temperature and humidity profiles

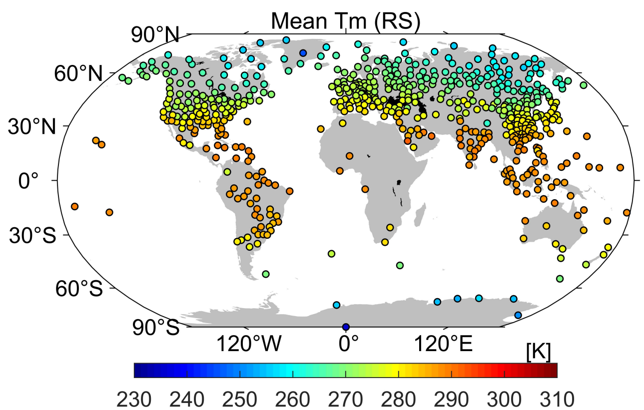
IGRA **radiosonde (RS)** data at **594** sites

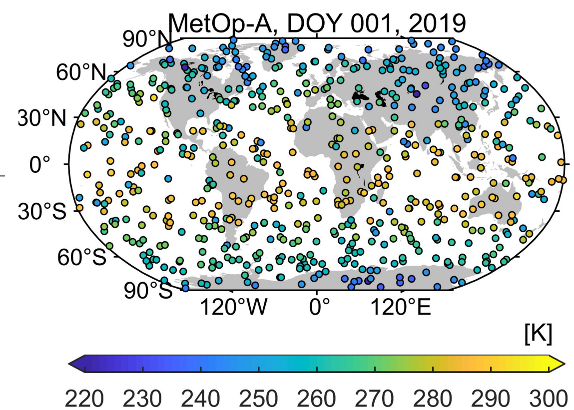
CDAAC **GPS radio occultation (RO)** 'wetPrf' product of **Metop-A and MetOp-B** satellites

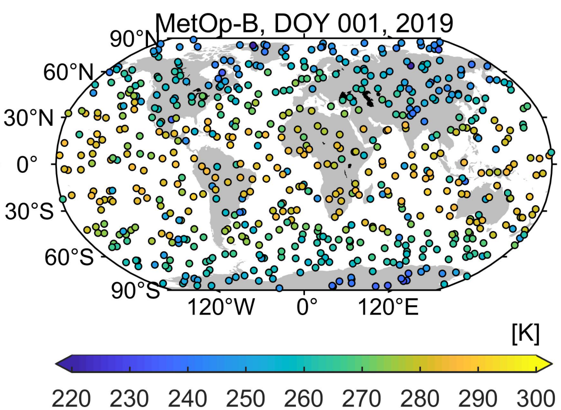
Validation: Global atmospheric temperature and humidity profiles for the year 2020

RS data at **659** sites

GPS RO 'wetPrf' products of **MetOp (-A, -B, and -C)** satellites





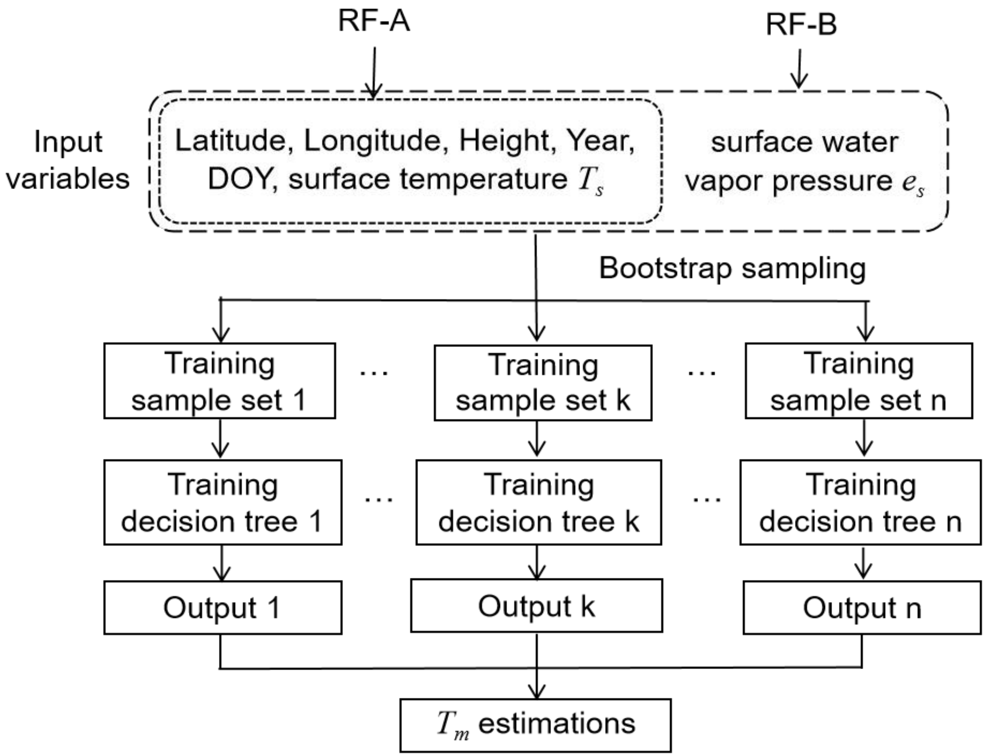


Methodology

T_m modeling based on the RF regression analysis

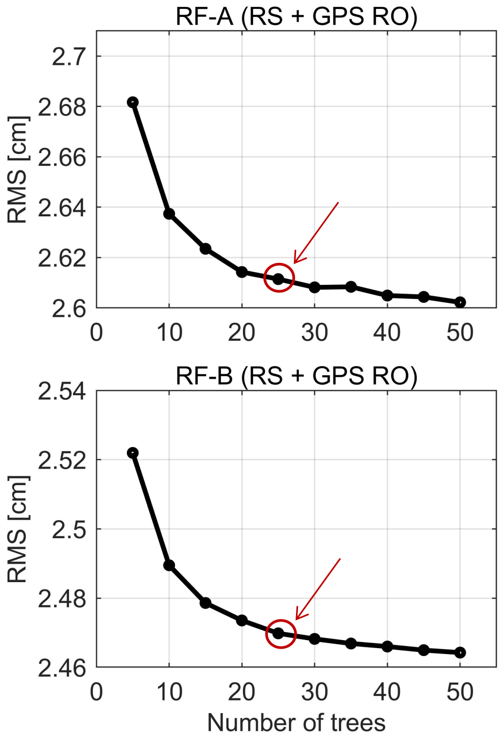
GTm_RF (RF-A) $T_m = f_A(\mathrm{Year}, \mathrm{DOY}, \mathrm{lat}, \mathrm{lon}, \mathrm{h_{ell}}, T_s)$

GTm_RF (RF-B) $T_m = f_B(\mathrm{Year}, \mathrm{DOY}, \mathrm{lat}, \mathrm{lon}, \mathrm{h_{ell}}, T_s, e_s)$



Determination of optimal number of decision trees

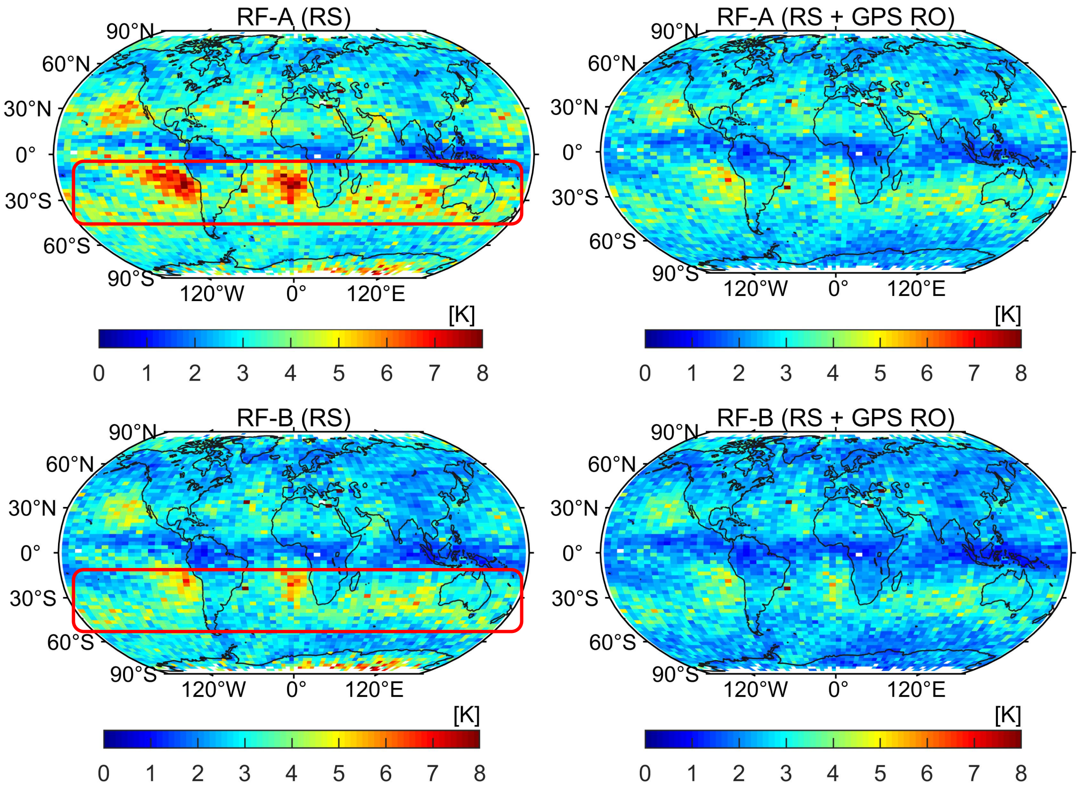
10-fold cross-validation



Comparisons of T_m modeling with different datasets

(a) T_m modeling with RS measurements

(b) T_m modeling with the integration of RS and GPS RO data



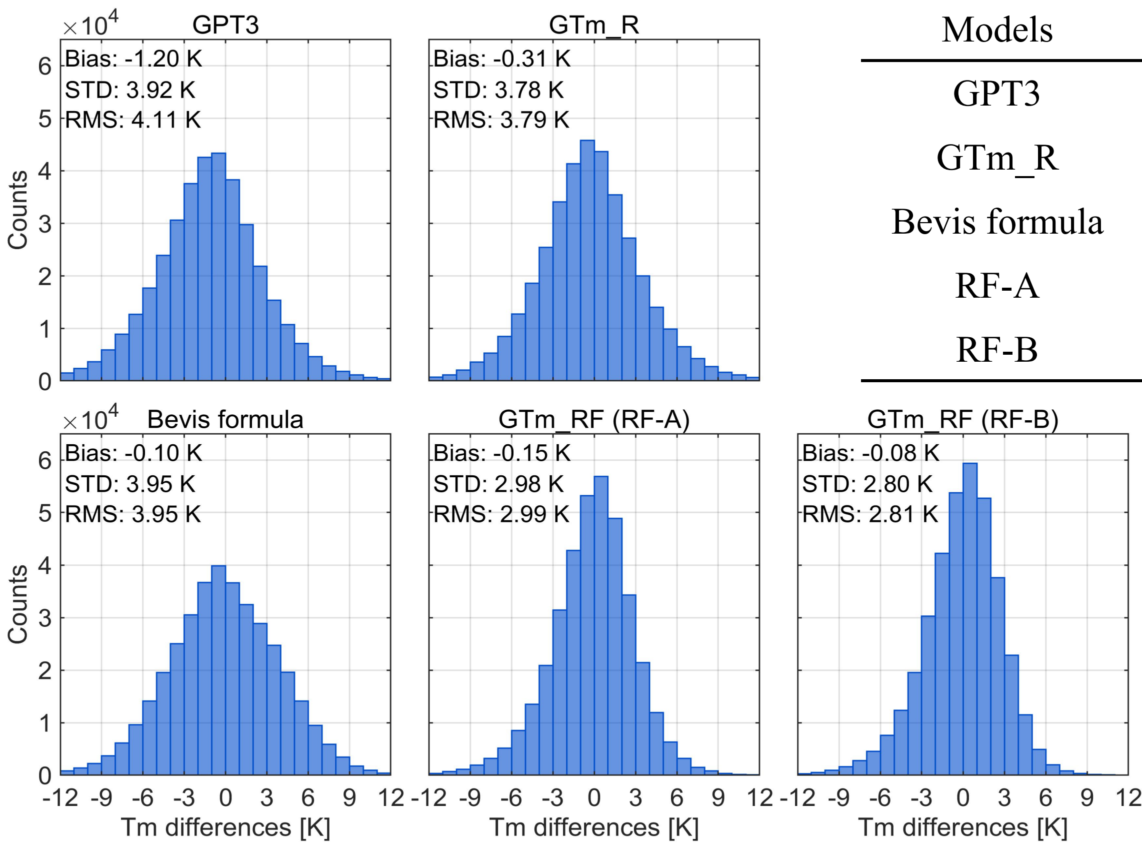
Positive biases of **2-8 K** for RF-A (RS) and **2-6 K** for RF-B (RS)

Positive biases within **1.5 K** for RF-A and RF-B (RS+GPS RO)

RS+GPS RO can improve the modeling accuracy in **oceanic and polar regions** and better model time-varying features of T_m **in the middle and upper troposphere**

Results

Validation with RS data



Models	Functional formualtion
GPT3	f (MJD, lat, lon, hell)
GTm_R	f (DOY, lat, lon, hell)
Bevis formula	f (T_s)
RF-A	f (Year, DOY, lat, lon, hell, T_s)
RF-B	f (Year, DOY, lat, lon, hell, T_s, e_s)

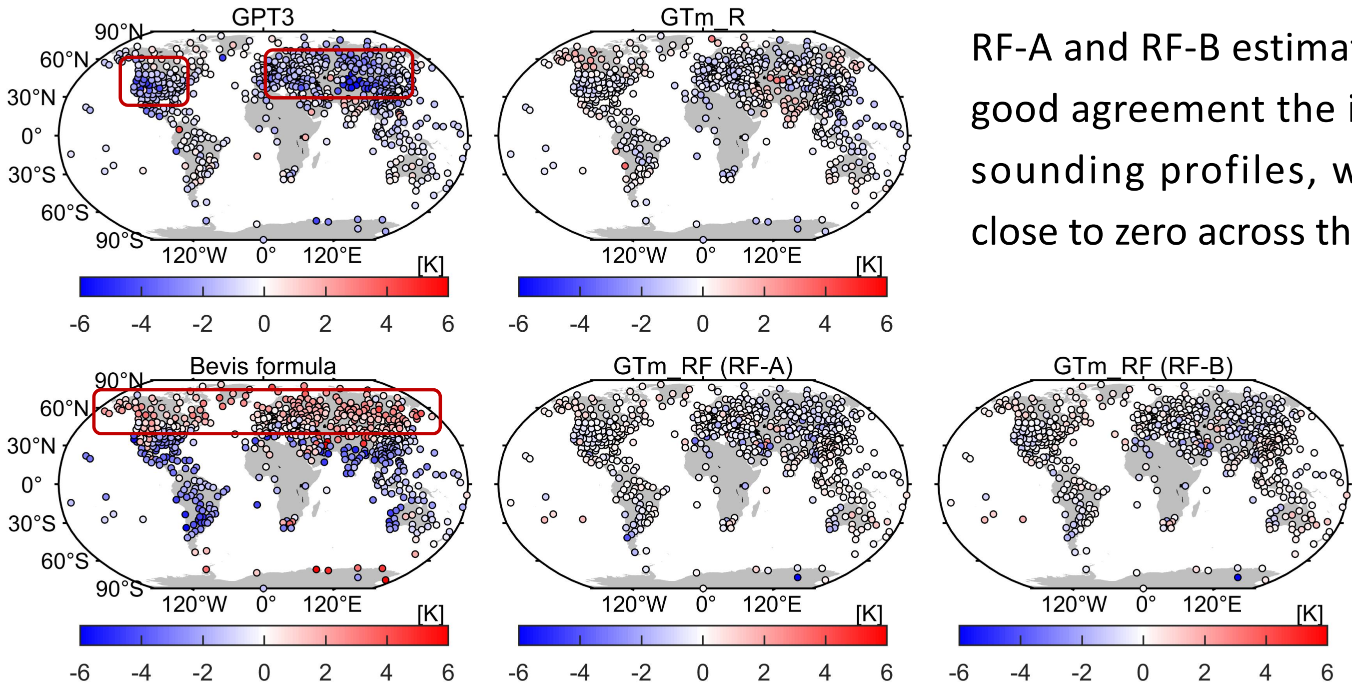
RF-A and **RF-B** models achieve overall agreements of **3.0 K** and **2.8 K** in comparison with **RS data**, whose accuracy improved by **21.1 %** and **25.8 %** than the GTm_R model

Results

Validation with RS data

GPT3 exhibits negative biases with absolute magnitudes of **3-6 K** due to the lack of vertical correction

Bevis formula has positive and negative biases with the latitude-dependent distribution for Tm

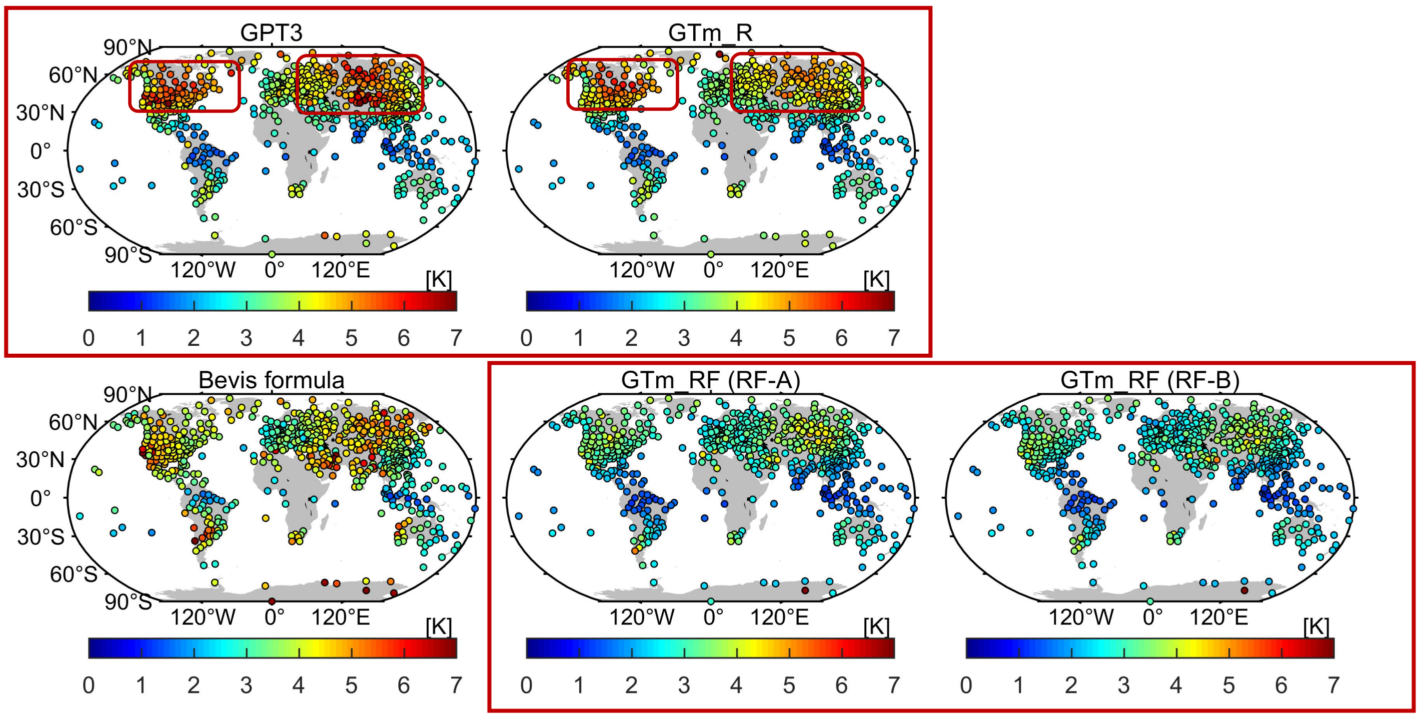


RF-A and RF-B estimations are in good agreement the integrals of sounding profiles, with biases close to zero across the globe

Results

Validation with RS data

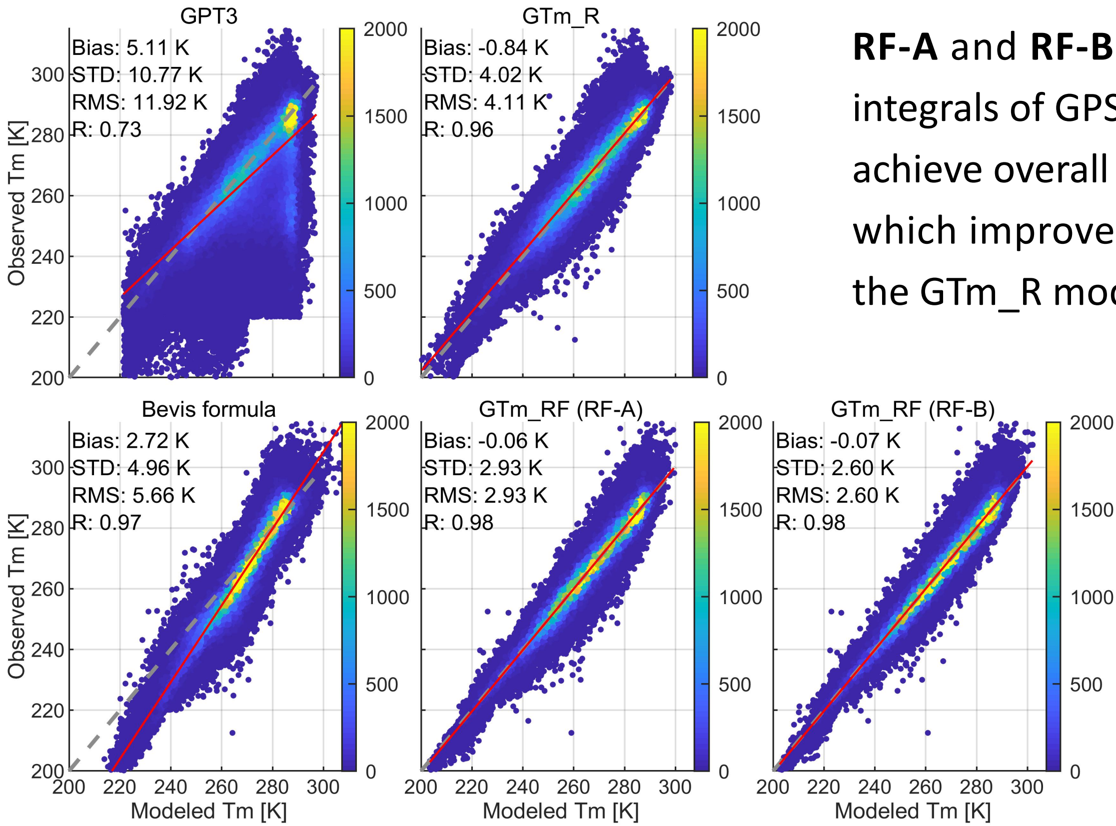
Considering surface meteorological parameters, T_m modeling has improved significantly in the middle and high latitudes of the northern hemisphere



Results

Validation with GPS RO measurements

GPS RO 'wetPrf' atmospheric profiles of MetOp-A/-B/-C satellites for the year 2020

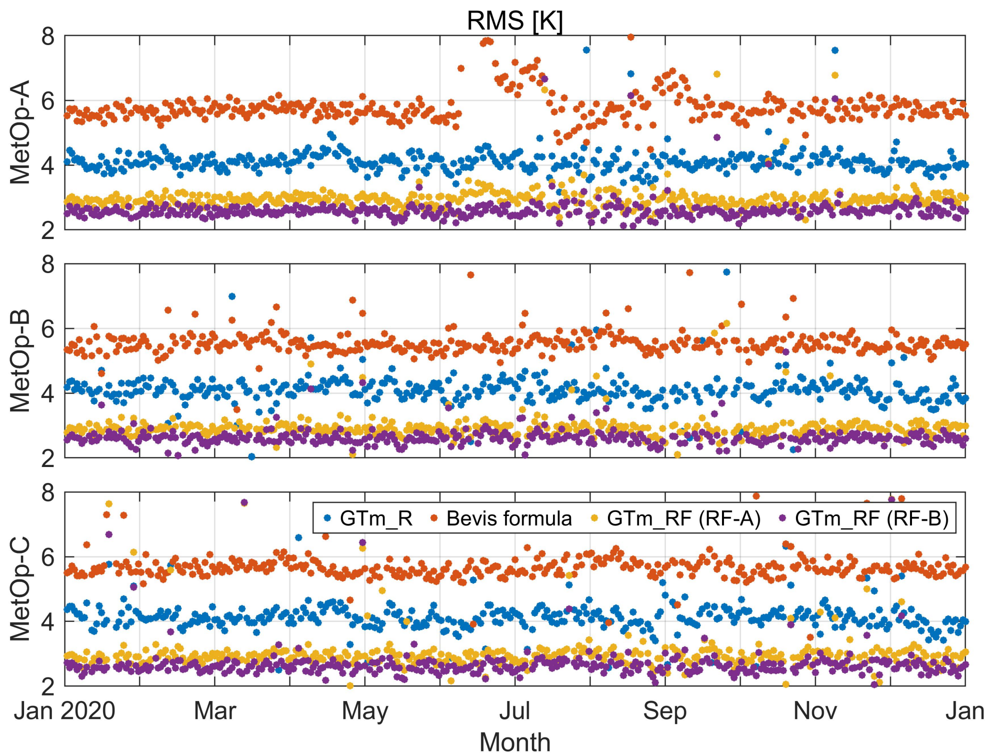


RF-A and **RF-B** models agree well with the integrals of GPS RO atmospheric profiles and achieve overall accuracies of **2.9 K** and **2.6 K**, which improved by **28.7 %** and **36.7 %** than the GTm_R model, respectively.

Results

Validation with GPS RO measurements

GPS RO 'wetPrf' atmospheric profiles of MetOp-A/-B/-C satellites for the year 2020



RF-A and **RF-B** achieve **relatively stable daily accuracy** throughout the year 2020, with RMS errors smaller than 3.0 K and 2.6 K, respectively. They perform better in reproducing time-varying T_m features.

Summary

T_m modeling with the integration of **RS** and **GPS RO** atmospheric profiles can improve accuracy in **oceanic and polar regions** and better account for time-varying T_m features at different heights in the troposphere

The RF-based T_m model with surface meteorological parameters (**RF-B**) obtains overall accuracy of **2.8 K** in comparison with **RS data** and **2.6 K** in contrast to **GPS RO** data, whose accuracy improves by **25.8 %** and **36.7 %** than empirical GTm_R model

Compared with empirical models, the accuarcy of **RF-based models (RF-A and RF-B)** has significantly improved in the middle and high latitudes of the northern hemisphere by considering **the correlation** with **surface meteorological parameters**

Thank you for your attention!

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