



**University of
Zurich** ^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2023

High-resolution tropospheric refractivity fields by combining machine learning and collocation methods to correct earth observation data

Shehaj, Endrit ; Miotti, Luca ; Geiger, Alain ; D'Aronco, Stefano ; Wegner, Jan Dirk ; Moeller, Gregor ; Soja, Benedikt ; Rothacher, Markus

DOI: <https://doi.org/10.1016/j.actaastro.2022.10.007>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-254959>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:




Shehaj, Endrit; Miotti, Luca; Geiger, Alain; D'Aronco, Stefano; Wegner, Jan Dirk; Moeller, Gregor; Soja, Benedikt; Rothacher, Markus (2023). High-resolution tropospheric refractivity fields by combining machine learning and collocation methods to correct earth observation data. *Acta Astronautica*, 204:591-598.

DOI: <https://doi.org/10.1016/j.actaastro.2022.10.007>

High-resolution tropospheric refractivity fields by combining machine learning and collocation methods to correct earth observation data

Journal Article

Author(s):

Shehaj, Endrit; Miotti, Luca; Geiger, Alain; D'Aronco, Stefano; [Wegner, Jan Dirk](#) ; [Möller, Gregor](#) ; [Soja, Benedikt](#) ; Rothacher, Markus

Publication date:

2023-03

Permanent link:

<https://doi.org/10.3929/ethz-b-000590523>

Rights / license:

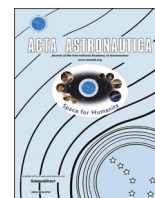
[Creative Commons Attribution 4.0 International](#)

Originally published in:

Acta Astronautica 204, <https://doi.org/10.1016/j.actaastro.2022.10.007>

Funding acknowledgement:

168952 - High-Resolution Atmospheric Water Vapor Fields by Spaceborne Geodetic Sensing, Tomographic Fusion, and Atmospheric Modeling (SNF)



High-resolution tropospheric refractivity fields by combining machine learning and collocation methods to correct earth observation data

Endrit Shehaj^{a,*}, Luca Miotti^a, Alain Geiger^a, Stefano D'Aronco^a, Jan D. Wegner^{a,b}, Gregor Moeller^a, Benedikt Soja^a, Markus Rothacher^a

^a Institute of Geodesy and Photogrammetry, ETH Zürich, Robert-Gnehm-Weg 15, 8093, Zürich, Switzerland

^b Institute for Computational Science, University of Zürich, Winterthurerstrasse 190, 8057, Zürich, 8006, Zürich, Switzerland

ARTICLE INFO

Keywords:

GNSS
Troposphere
Machine learning
Collocation
Zenith delay
Meteorological parameters

ABSTRACT

Signals used for Earth observation, when travelling through the atmosphere, are sensitive to refractivity; especially high spatio-temporal variations of water vapor are difficult to model and correct. Remaining unmodeled tropospheric delays deteriorate the positioning solution and therefore limit the accuracy of sensing and navigation applications. These delays are usually computed with empirical models based on ground meteorological parameters (pressure, temperature and water vapor partial pressure). However, existing models are not accurate enough for high-precision applications such as GNSS, where in consequence the so-called zenith total delay (ZTD) has to be estimated together with other unknown parameters (coordinates etc.).

For decades, the Institute of Geodesy and Photogrammetry at ETH Zurich has been studying collocation methods for the modeling of tropospheric delays using meteorological parameters, successfully interpolating pointwise or integral atmospheric observations. Meanwhile, machine learning has become a widely used and valuable alternative, when big datasets are available for the training process. Indeed, we have already successfully predicted ZTDs based on meteorological parameters with an accuracy of 1–2 cm for locations (GNSS stations) already used in the training phase. However, difficulties arise to predict delays at new locations.

In this work, we take a step forward in investigating the combination of machine learning algorithms and physical models used in a collocation approach to derive atmospheric delay fields at a very high resolution. Thus, without processing any GNSS data we can predict tropospheric delay fields everywhere in the area of investigation. In this paper, we firstly describe the designed architecture of the neural network, secondly, the combination of least-squares collocation and artificial neural networks for high-resolution prediction of tropospheric delays. We benefit from the complementary characteristics of these algorithms. While machine learning is capable of successfully predicting the variation of time series for given points, empirical models based on collocation are well suited for describing spatial variations within the area of investigation. Finally, we report the achieved performance for the entire territory of Switzerland (1–2 cm in terms of RMS), showing that the synergic combination of these algorithms can overcome the individual drawbacks of each method and provide more accurate delay estimates than either method individually.

Datasets of 11 years, covering the territory of Switzerland, consisting of GNSS ZTDs from 72 permanent AGNES/COGEAR (swisstopo, ETHZ) stations and meteorological data from MeteoSwiss were used for this research.

1. Introduction

Atmospheric water vapor is a very important gas in meteorology, climatology and hydrology. It is a crucial element of the hydrological cycle, as well as directly connected to natural disasters (floods, draughts, glacier melting, etc.) and global climate change [1]. However, its high

spatio-temporal variations make it very difficult to be modeled or predicted [2]. Meanwhile, microwave satellite signals, e.g. signals from Global Navigation Satellite Systems (GNSS) travel through the troposphere before they can be received at the user site. Due to high fluctuations of water vapor these signals are delayed and/or distorted, affecting the outcome of the technique. The experienced delay is a direct

* Corresponding author.

E-mail address: endrit.shehaj@geod.baug.ethz.ch (E. Shehaj).

<https://doi.org/10.1016/j.actaastro.2022.10.007>

Received 22 June 2022; Received in revised form 29 September 2022; Accepted 4 October 2022

Available online 11 October 2022

0094-5765/© 2022 The Authors. Published by Elsevier Ltd on behalf of IAA. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

function of the refractivity (and thus the meteorological parameters) along the signal path. The simplest approach to deal with this nuisance in GNSS signal processing, is to model it using empirical formulations based on ground meteorological data [3,4]. Although these empirical models may be accurate enough for low-cost GNSS receivers, higher accuracy is needed in case of geodetic applications. Clearly, the main problem of these models is the description of high variations of water vapor. Furthermore, in case of GNSS stations without close-by meteorological sensors, interpolating the meteorological parameters affects furthermore the results. Thus, scientists and engineers have routinely estimated these tropospheric effects during the processing of GNSS measurements. These has led to time series of tropospheric parameters with high-temporal resolution over several decades containing information about the water vapor, with the most typical one being the zenith total delay (ZTD) [5,6]. However, the spatial resolution may not be considered high, depending on the network.

Machine learning has become very popular in different fields of science, where the complex mathematical models can be trained appropriately from a large amount of data [7]. Indeed, in our previous works [8,9], we have successfully shown that the trained machine learning models can predict zenith delay time series based on meteorological parameters. Initially a simple algorithm (such as random forest) was deployed and then the results were further improved with a multi-layer perceptron (MLP) algorithm. Our trained models could predict time series for trained locations at 1–2 cm accuracy. However, the prediction at new locations was not very successful, experiencing a bias increasing with the distance to the closest trained location. Other works [10], uses Long Short Term Memory (LSTM) algorithms to predict future zenith wet delays based only on past time series, whilst [11] utilizes a Gaussian Process (GP) to predict tropospheric corrections in Interferometric Synthetic Aperture Radar (InSAR) based on zenith delays. Furthermore, some researchers use machine learning to fuse tropospheric delays in geodetic signals in numerical weather prediction models [12].

At the Institute of Geodesy and Photogrammetry, the least squares collocation software COMEDIE (Collocation of Meteorological Data for Interpretation and Estimation of Tropospheric Path Delays) has been developed and used for decades to interpolate/extrapolate atmospheric parameters produced by several meteorological sources (such as numerical weather prediction models, GNSS, InSAR, etc.) [13,14], [15,16]. The main feature of COMEDIE is the spatial interpolation of meteorological parameters; for instance, ZTDs can be interpolated with sub-cm level accuracy [15]. This is the case for the Swiss GNSS networks (shown later), with a relatively good height distribution of the network stations.

In this work, we take a step forward and combine the two methods previously utilized in our lab. On one side, we use machine learning algorithms to predict tropospheric delays from meteorological observations at stations already trained in the network. On the other side, we collocate the predicted delays and then interpolate them at any point in the area of investigation. Our goal is to benefit from the different spatio-temporal error distributions of machine learning and collocation.

1.1. Tropospheric delay in GNSS

There exists a direct relation between the delay microwave signals experience because of the troposphere and the meteorological parameters, as follows:

$$\Delta\rho_{TROPPO} = 10^{-6} \int_{Slant Path}^0 N(s) ds \tag{1}$$

$$N = k_1 \frac{P_{tot} - P_{wet}}{T} + k_2 \frac{P_{wet}}{T} + k_3 \frac{P_{wet}}{T^2} \tag{2}$$

where the tropospheric delay is the integral of the refractivity N along the slant path, which is directly described by the total pressure, water vapor partial pressure and the temperature (P_{tot} , P_{wet} and T). In Ref. [17] the coefficients, k_1 , k_2 and k_3 are respectively $77.6890 \text{ KhPa}^{-1}$, $71.2952 \text{ KhPa}^{-1}$ and $375,463 \text{ K}^2\text{hPa}^{-1}$, which are also the values we have used in our works.

Empirical models based on ground meteorological parameters are used to model tropospheric effects in GNSS signals (typical for low-cost receivers), [3,4]. For instance, the most typical one is the Saastamoinen model:

$$ZTD = 0.002277f(\varphi, h_{ellip}) \left[P_{tot} + \left(\frac{1255}{T} + 0.05 \right) P_{wet} \right] \tag{3}$$

where f is a function of the geographical latitude φ , and the ellipsoidal height h_{ellip} .

The ZTD is the most basic tropospheric parameter, that the GNSS community uses, which is estimated in the GNSS adjustment, where individual slant delays from each visible satellite are combined into one parameter.

1.2. GNSS and meteorological networks

In this work, the dataset used is the same as that of our previous works [8,9].

Fig. 1 shows the GNSS and the meteorological networks in Switzerland in blue and orange, respectively. The GNSS network includes three sub-networks: the AGNES network (operated by swisstopo), the COGEAR network (operated by the Mathematical and Physical Geodesy group (MPG) of ETH Zurich) and the NAGRA network which is the densification in the northeast part of Switzerland [18]. The meteorological network is the SwissMetNet network, operated by MeteoSwiss [19].

The dataset that we use covers the time period from 2008 to 2018, where swisstopo and MeteoSwiss have removed the outliers beforehand. An example of the time series is provided in Fig. 2, where the GNSS ZTD time series is plotted for station ERDE and the respective meteorological parameters for its closest meteo station SIO. Initially, it is important to point out that there are data missing in the time series, as shown in the GNSS ZTDs. Furthermore, from the four time series, a correlation can be seen between the periodicity of ZTDs and that of temperature and water vapor partial pressure, while the correlation with total pressure is less visible.

Notice that in this paper we focus on ZTDs, while it may be argued that modeling wet delays may be more reasonable, since the hydrostatic part in the ZTDs varies very little. One reason of our choice is the fact that for practical applications (such as tropospheric corrections to low-cost receivers such as smartphones) providing total tropospheric

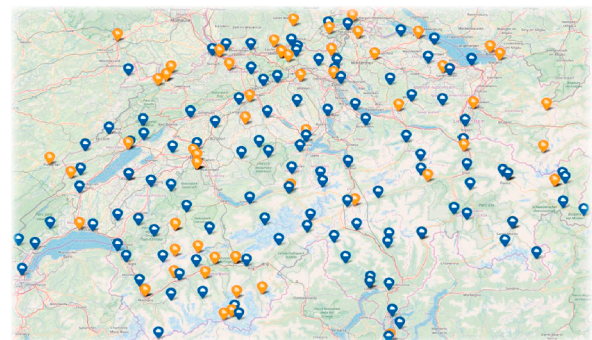


Fig. 1. AGNES/COGEAR/NAGRA GNSS (orange) network and SwissMetNet (blue) meteorological network, same as in Ref. [8]. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

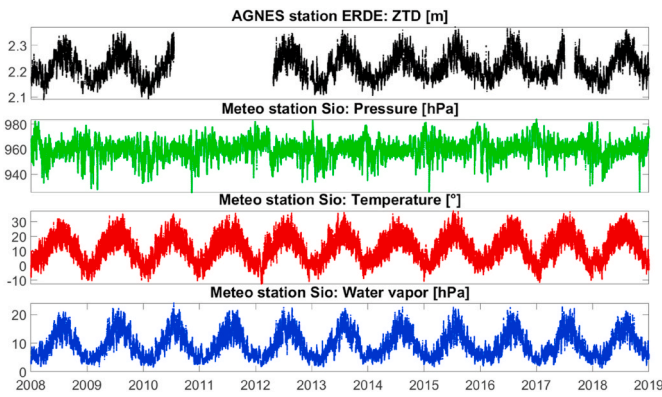


Fig. 2. ZTD time series for the station ERDE and meteorological parameters for the nearby station SIO, same as [8].

delays is more appropriate. Furthermore, since the total delay is a sum of the deterministic hydrostatic delay and a variable wet delay estimated during GNSS processing, the accuracy of the estimated total delay is the same as that of the wet delay. Therefore, we would expect ML to perform similarly for these two variables. In fact, we have performed some tests only using wet delays, and we obtained comparable results to the ones we display here.

More information regarding the dataset utilized here is provided in Ref. [9].

2. State-of-the-art

In our previous works [8,9], we have applied machine learning to predict GNSS time series of ZTDs from meteorological parameters. For comparison purposes, we also developed another approach, based on the collocation of meteorological parameters and the Saastamoinen model. Furthermore, collocation of ZTDs in the software COEMDIE has been performed for decades [15,20,21]. These methods, which are state-of-the-art for our laboratory, as well as for this work, are shown in Sections 2.1 to 2.3, and the respective results are summarized in Section 2.4. Further information can be found in the references above.

2.1. Machine learning

The machine learning algorithm used here is MLP, which is an artificial neural network. All neurons of one layer are fully connected to all neurons of the next layer whilst the first and the last layers are the input and output values, respectively, as displayed in Fig. 3.

Each neuron is a function of the input from the neurons in the previous layer, where each input is weighted and also a bias is added to the neuron itself. All the biases and weights of each neuron are the parameters of the neural network [22]. To introduce possible nonlinearities, the Rectified Linear Unit (ReLU) activation function is used, described in Ref. [22]. The network is trained by searching for a local minimum in the loss function, which is in our case the mean squared

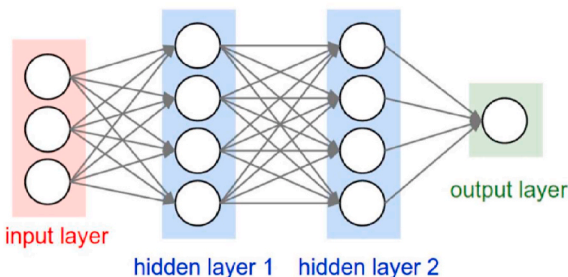


Fig. 3. Fully connected neural network (source: <http://cs231n.github.io/>).

error (mse) between the label and the prediction [23]. Table 1 summarizes the formulas of the neuron, the activation and loss functions, as well as the chosen parameters of the neural network.

In this work, we use neural networks to predict zenith delays, however in Refs. [8,9] we have used simpler algorithms such as random forest. We also deployed gradient boosting for internal comparisons and achieved similar results as with the random forest. Therefore, in this work we only display the results with neural network and use the neural network prediction to combine with collocation. MLP are universal approximators, meaning there exists a network that is the best fit to relate the inputs and outputs, [24]. In addition, from common knowledge the MLP algorithm mostly outperforms simpler machine learning algorithms, such as random forest, decision trees or gradient boosting. There may be more advanced machine learning algorithms, such as deep learning ones, which may perform better for our application than MLP, however they are more complex and have more parameters to be tuned. In this work, it is not our ultimate goal to find the best machine learning algorithm to predict ZTDs. Our objective is to show the combination of machine learning and collocation exploiting their different characteristics for time series prediction and spatial interpolation.

2.2. Saastamoinen model based on collocation

To compute the Saastamoinen model, initially we collocate the meteorological parameters (pressure, temperature and water vapor pressure) individually. Collocation accounts for a so-called ‘signal’ part, which is correlated noise of the residuals. The measurement l is described by the functional part f , the signal s and a Gaussian noise ϵ , as follows [15]:

$$l = f + s + \epsilon \tag{4}$$

For instance, the functional part of pressure measurements, is described as follows [13]:

$$P(x, y, h, t) = (P_0 + a(x - x_0) + b(y - y_0) + c(t - t_0)) \bullet e^{-\frac{h-h_0}{Hscale}} \tag{5}$$

where the parameters in red, estimated in the adjustment process, are P_0 at reference position, the gradients a , b and c in East, North and time, and the scale height $Hscale$. Therefore, the pressure can be computed at any point inside the network by introducing the coordinates and time x , y , h , t . As for x_0 , y_0 , h_0 and t_0 , they are the coordinates and time of a reference location, chosen usually as the mean of all measurements.

A covariance matrix, which is determined empirically and contains information about the correlations of the measurements, describes the

Table 1 Neural network description and chosen parameters.

Neuron	$neuron = f(\sum_{i=1}^n input_i * weight_i + bias)$		
Activation function	$f(x) = \max(0, x)$		
Loss function	$mse = \frac{1}{n} \sum_{i=1}^n (pred_i - label_i)^2$		
Layers	Input	Neurons	Parameters
	Hidden 1	27	-
	Hidden 2	512	14,336
	Hidden 3	128	65,664
	Hidden 4	128	16,512
	Hidden 5	128	16,512
	Output	128	16,512
Hyperparameters	Learning rate	Batch size	Epochs
	0.0001	1000	250
			129
Feature	Inputs		Output
	1	Pressure, temperature and water vapor partial pressure at 4 closest stations	ZTD
	2	Coordinates of the meteo stations	
	3	Coordinates of the GNSS stations	
Dataset	Training	Validation	Testing
	2008–2016	2017	2018

signal. More information is provided in Refs. [13–15].

By introducing the estimated parameters and the desired coordinates in Eq. (5), pressure can be calculated at any point inside the network. Thus, the collocation method is used for interpolation. Similarly, the collocation and interpolation of temperature and water vapor partial pressure is performed, [13].

Finally, the Saastamoinen model can be computed based on the interpolated parameters, using Eq. (3).

2.3. Collocation of ZTDs

In case of ZTDs, the functional part is [15]:

$$ZTD(x, y, h, t) = [ZTD_0 + a_{ZTD}(x - x_0) + b_{ZTD}(y - y_0) + c_{ZTD}(t - t_0)] \bullet e^{-\frac{h-h_0}{H_{ZTD}}} \tag{6}$$

where, ZTD_0 is the ZTD at the reference position (x_0, y_0, h_0, t_0) , a_{ZTD} , b_{ZTD} and, c_{ZTD} are the gradient parameters in the East and North coordinate and time, and H_{ZTD} is the scale height to be estimated.

As for the signal part, it is a function of a modeled covariance C_{ss} , $s(C_{ss}, x, t)$, where C_{ss} is defined as:

$$C_{ss}(i, j) = \frac{\sigma_0^2}{1 + \left[\left(\frac{x_i - x_j}{\Delta x_0} \right)^2 + \left(\frac{y_i - y_j}{\Delta y_0} \right)^2 + \left(\frac{h_i - h_j}{\Delta h_0} \right)^2 + \left(\frac{t_i - t_j}{\Delta t_0} \right)^2 \right]} \bullet e^{-\frac{h_i + h_j}{2h_0}} \tag{7}$$

with x, y, h, t being the space and time coordinates of two measured points i and j . $\Delta x_0, \Delta y_0, \Delta h_0, \Delta t_0, h_0$ are the correlation lengths in space and time and the scale height modifying the correlation lengths as a function of height. More information can be found in Ref. [20].

Furthermore, the refractivity can be formulated by simply deriving the ZTD in the zenith direction [20]:

$$N(x, y, h, t) = -\frac{\partial ZTD(x, y, h, t)}{\partial h} \tag{8}$$

$$N(x, y, h, t) = \frac{1}{H_{ZTD}} [ZTD_0 + a_{ZTD}(x - x_0) + b_{ZTD}(y - y_0) + c_{ZTD}(t - t_0)] \bullet e^{-\frac{h-h_0}{H_{ZTD}}} \tag{9}$$

Also the covariance matrices that relate the pathdelays with refractivity, describing the signal part, are shown in Ref. [20]; they are obtained by deriving the covariance of the delays with respect to height.

Therefore, using collocation fields of zenith delays and 3D fields of refractivity can be produced.

2.4. State-of-the-art results

In this section, we summarize the results obtained by our state-of-the-art methods. Whilst the results of the machine learning and Saastamoinen approach based on collocation have been displayed before in Refs. [8,9], the collocation results for this dataset have not been published before.

2.4.1. Machine learning and saastamoinen approach

In Fig. 4 an example of time series of ZTD is shown, predicted with

ML and the Saastamoinen model, for station WEHO; this station is simply chosen for illustration. In Fig. 5 the RMS for the year 2018 are displayed for each station, for both approaches (the Saastamoinen approach is represented by rectangles and the ML by circles). For every station the ML has a lower RMS compared to the Saastamoinen approach based on collocation. Furthermore, the ML predicted delays do not exceed an RMS of 2 cm, while this happens quite often for the Saastamoinen approach. A deeper analysis of these results is shown in Refs. [8, 9]. Since in this work we would like to put the main focus on section 3 where the combination is shown, here we do not make a deeper analysis of these results.

We must point out that we also used the GPT3 model to model the ZTDs, where the Askne and Nordius model is used to compute the wet part, supposedly more accurate. However, from our results, the Saastamoinen model resulted in the most accurate and therefore, in this section we only present those results.

2.4.2. Machine learning performance at untrained locations

The performance of ML is quite good at already trained locations, since we use several years of time series training the model. However, to show the performance at untrained stations, we removed the station from the training process and then used the ML model to predict time series of that station. An example (for station HGGL) is shown in Fig. 6, where we can clearly see a bias between the actual and the predicted delays. Furthermore, we did the same for three different stations, which have different distances to the closest trained GNSS station. From the results in Fig. 7, it is obvious that the further the untrained location is away from the trained station, the bigger the bias in the residuals (computed as difference of predicted and actual delays). Indeed, for station ZIMJ (which has a twin station nearby), there is no bias visible, and the residuals of station HGGL, which is the furthest away from the closest trained GNSS station, experience the largest bias.

The bias in the ML-only solution is explained by the limited data available for training the ML algorithm. Our GNSS network consists of 72 locations, therefore only these locations together with their closest 4 meteorological stations have been trained. For these locations, in total 9 years of data is available, which leads to a good performance of the ML algorithm for the trained locations. However, the algorithm fails to generalize to a certain extent when applied to new locations. This emphasizes the very different performance of time and space prediction, which is also a motivation behind this paper, to find an alternative to solve the spatial interpolation by combination with collocation.

2.4.3. Collocation

For the year 2018, we performed a collocation of the ZTDs for each GNSS station. The ZTDs of the other stations were used to estimate the collocation parameters. Finally, the parameters are used to compute the ZTD for the station excluded from the collocation process. Therefore, 71 stations enter the collocation and 1 station is excluded, and a cross validation between the collocated ZTDs and the actual ones estimated by GNSS processing is performed. This section is simply to show the good performance for spatial interpolation of our collocation method. Fig. 8 shows the residuals for 5 stations between the collocated and the actual ZTDs. These 5 stations are only chosen for illustration purposes. The residuals of the time series are much lower than 1 cm for most of the

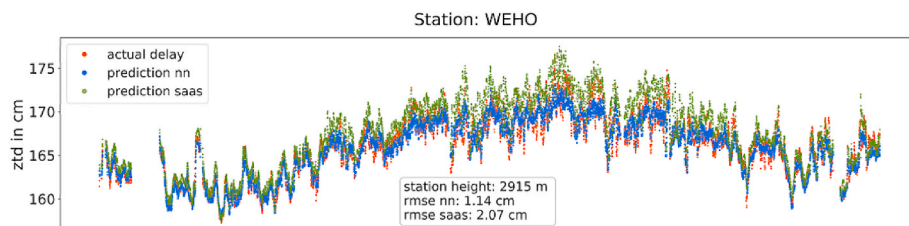


Fig. 4. Actual estimated delay and the respective neural network and Saastamoinen predictions (same as [8]).

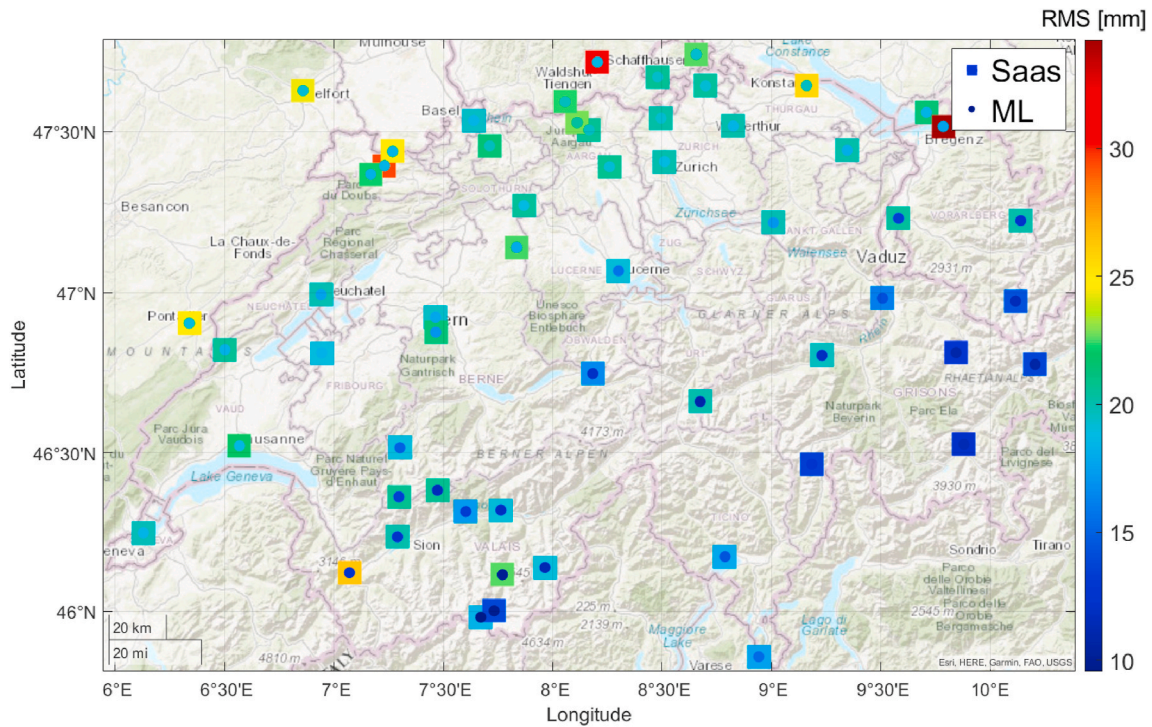


Fig. 5. Root mean square errors for the neural network and the Saastamoinen predictions for every station of the GNSS network.

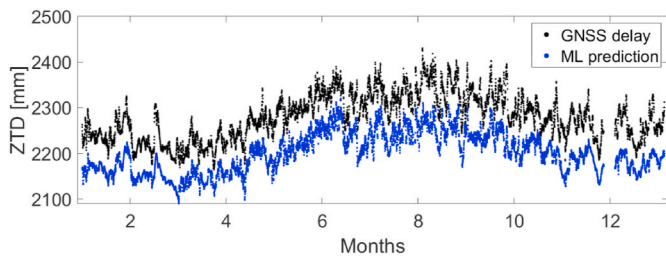


Fig. 6. Actual estimated delay and the respective neural network prediction at untrained station HGGL.

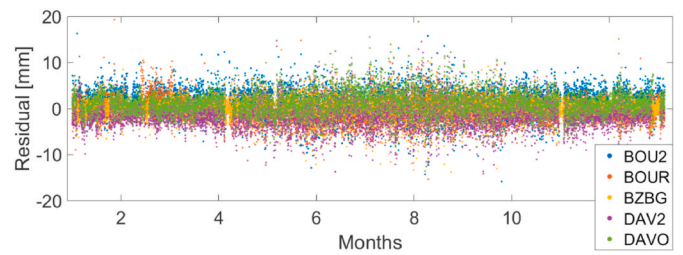


Fig. 8. ZTD residuals of the collocation for different GNSS stations during 2018.

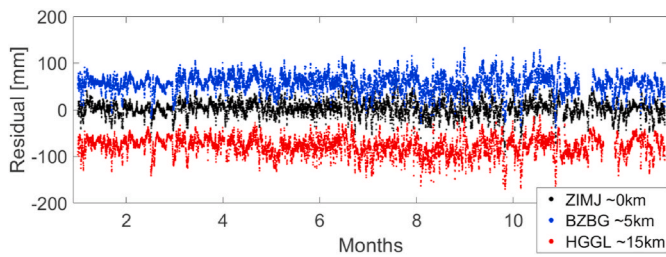


Fig. 7. Residuals of ML predictions for three untrained stations at different distances from the closest trained station.

epochs. Furthermore, Fig. 9 displays the RMS, standard deviation and mean of all the time series for the year 2018 for all stations. The statistics are clearly below 1 cm (or even 5 mm) for most of the stations (the statistics of only few stations are around 1 cm).

With further tuning of the model parameters in Section 2.3 it may be possible to improve the statistics of stations at about 1 cm; however, this is out of the scope of this paper where the main goal is to illustrate the spatial performance of collocation for our time series.

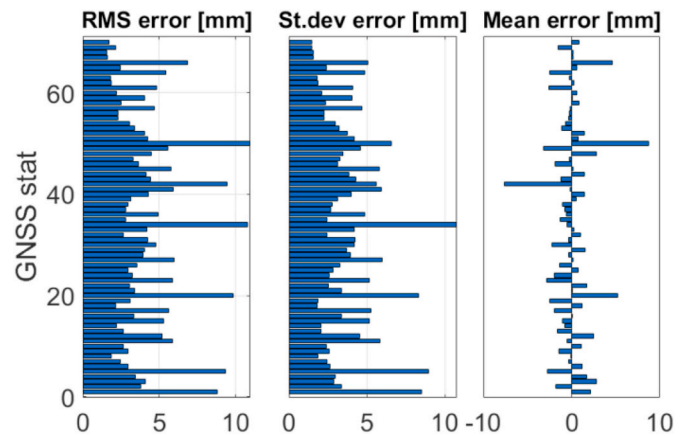


Fig. 9. Statistics of collocation of ZTDs for all GNSS stations during 2018.

3. Proposed combination

As mentioned before, the goal of this work is to benefit from the characteristics of machine learning and collocation. In the next Section 3.1 the combination architecture is displayed and in Section 3.2 the

results achieved with this combination approach are shown.

3.1. Combination architecture

Fig. 10, shows the workflow of computing ZTDs by using both, ML and collocation. There are three parts in the workflow:

- Used measurements: for the combination we make use of exactly the same measurements (meteorological parameters and zenith total delays) as in the case where we only use ML, as shown in Section 2.
- Machine Learning: in this part we perform exactly the same steps as performed in Section 2 for the ML approach, where training and validation are performed with the years 2008–2017. Then, with the trained ML models and the meteorological parameters for 2018, we predict ZTDs at already trained locations.
- Collocation: the predicted ZTDs from the ML algorithm are used as input to the collocation process, to compute the state vector, as explained in Section 2.3. Therefore, the model parameters and the coordinates of any point inside the network, can be used to generate ZTDs (as well as refractivities), at any location (trained or untrained).

3.2. Combination results

Section 3.2.1 shows the results of our combination method for the three stations also displayed in Section 2.4.2, while Section 3.2.2 contains the results of the combination approach over all Switzerland.

3.2.1. Specific stations

In Fig. 11 in the top plots, the ZTD time series at untrained locations are displayed predicted by ML (in blue), and the ZTD time series of our combination approach (in red), as well as the reference ZTD in black, for the year 2018. In the lower subplots, the respective residuals (difference between the predicted delays, via ML alone or combination, and the actual delays) are shown. For both stations, we can clearly see that the time series of the combination are much closer to the actual delay, while for the lower subplots we can clearly see that the residuals of the combined method (in red) do not experience a visible bias as those that only leverage ML to obtain a prediction (in blue). This is clear for both stations, at ~5 km and at ~15 km away from the closest trained GNSS station. In Table 2, all the statistics for these three stations are displayed. We can confirm from these statistics that with our combination method the bias and the standard deviation are improved compared to the case when the stations are not trained. Furthermore, the combination

statistics is similar to the case when the stations are part of the training dataset.

We emphasize that the time series of these three stations are not used to train the ML network, but the rest of the stations (69 stations) train the NN. Therefore, we use the predicted time series of the 69 stations for year 2018 as input to the collocation for computing the ZTD for the three untrained locations. Fig. 10 gives an overview about the workflow.

3.2.2. Results all over Switzerland

Fig. 12 shows the performance of the combination approach for all stations in Switzerland. In this case, we iteratively removed each of the stations from the training in ML and then performed the workflow shown in Fig. 10 using the rest of the stations. Therefore, we used the collocation of the predicted delays to interpolate ZTDs to the untrained station, for year 2018.

From these results we can see that:

- the results are area and/or station-dependent, i.e. south of Switzerland the combination statistics are better compared to North of Switzerland. It must be pointed out that this was also the case when all stations were trained using the ML approach, as shown in Fig. 5.
- the range of the RMS (as well as its distribution over Switzerland) for the combination (Fig. 12) is similar to the range of the RMS (and its distribution over Switzerland) for the case where all stations were used to train the ML models (Fig. 5).

The main reason for this distribution of RMS is the fact that in the north of Switzerland the GNSS stations are at lower altitudes, thus experiencing more the effects of water vapor. Therefore, the GNSS estimated ZTDs used for training and testing, are less accurate due to the higher water vapor variations. To verify this finding, initially we checked the relative error, and the values vary between 0.4 and 0.65%, with a general similar distribution as the RMS in Fig. 12, i.e. with larger relative errors in the north and smaller ones in the southern part. Furthermore, we plotted the standard deviation of the ZWDs (not shown here) also provided by swisstopo and we noticed that in the north the ZWDs vary more for the year 2018, while in the south there is lower variations. We point out that there are some stations which do not obey this rule, however most of the stations confirm this result.

Other possible reasons may be related to the GNSS station equipment quality (for instance the GNSS antenna), or GNSS processing. Indeed, we found a case of twin stations (few m away) which time series had biases larger than expected. We opted not to use those observation in our

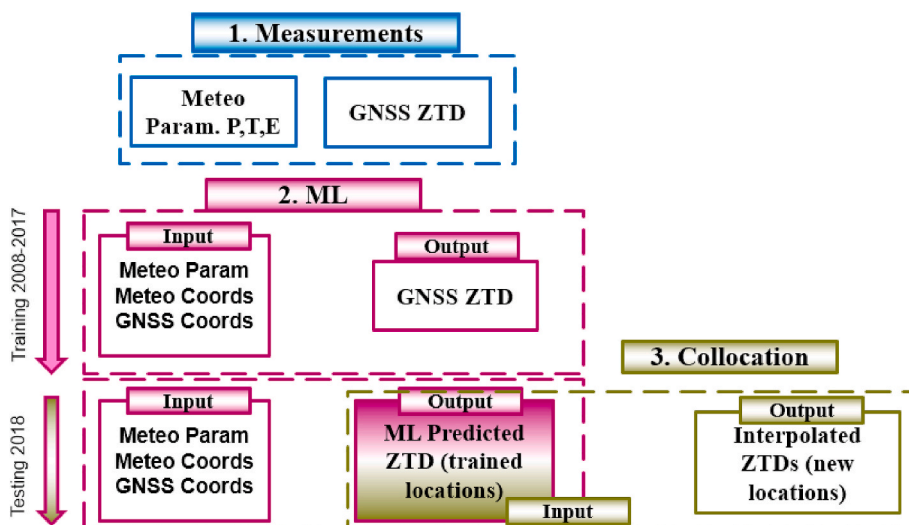


Fig. 10. Workflow of the proposed combination approach.

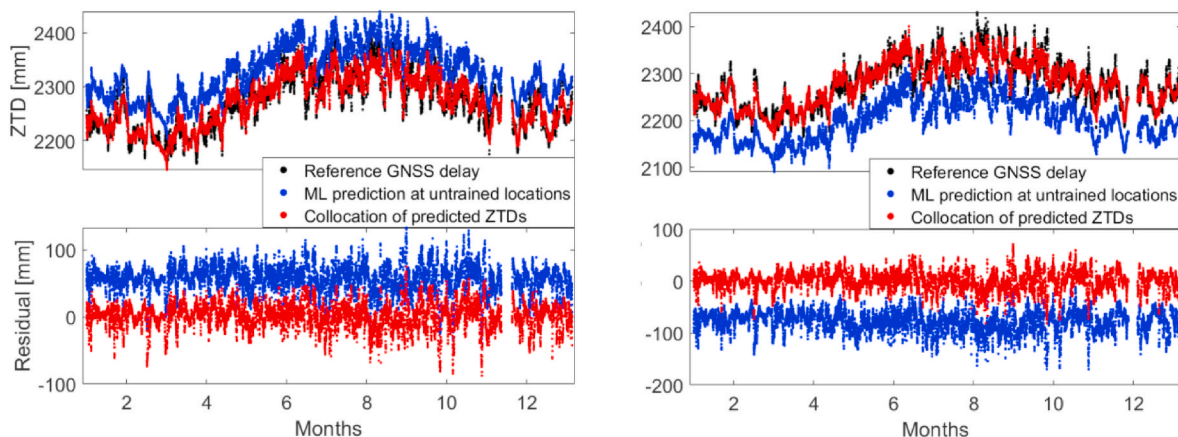


Fig. 11. Top plots: the actual estimated delay (black) and the respective neural network (blue) and combination (red) predictions for two untrained stations (BZBG (left) and HGGL (right), respectively ~5 km and ~15 km away from the closest trained station). Lower plots: the respective residuals of predictions of neural network and from the combined approach. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

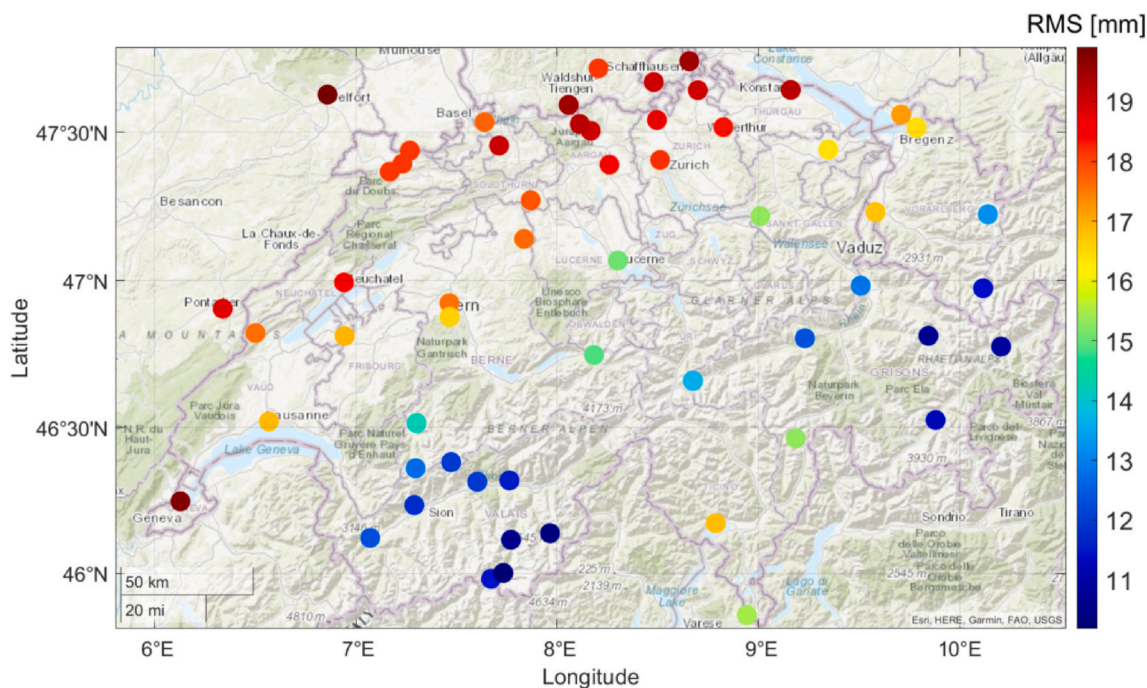


Fig. 12. Root mean square errors for the combination predictions, for all the GNSS network, for the year 2018.

evaluation, since this investigation was out of our scope.

4. Conclusions and discussion

In this work, we presented a combination approach where the different spatio-temporal characteristics of the error distribution of machine learning and collocation are exploited. In this context, machine learning is used to predict zenith delays based on meteorological

parameters, while collocation is used to spatially interpolate the zenith total delays predicted by ML.

The datasets utilized are meteorological and GNSS time series for 2008–2018 from the SwissMetNet and AGNES/COGEAR/NAGRA networks in Switzerland.

Initially, a review of the ML and collocation methods, as well as their performance for our dataset were provided. Afterwards, the combination workflow and the respective results were displayed. For the

Table 2
Statistics for the three investigated stations.

Unit [mm]	ZIMJ			BZBG			HGGL		
	RMS	STD	BIAS	RMS	STD	BIAS	RMS	STD	BIAS
SAAS	22.2	19.4	10.8	20.6	20.6	1.5	20.7	20.4	3.1
ML station included	17.2	17.0	2.5	19.5	19.5	0.6	19.2	19.1	1.8
ML station excluded	16.9	16.8	2.1	60.3	20.2	56.8	81.3	20.9	-78.6
Comb. ML&Coll	16.9	16.7	2.7	19.2	19.2	0.5	18.1	18.2	0.01

validation of our method, we used the GNSS time series themselves, which were removed from the training process and considered as a new untrained location. From the results displayed, we can conclude that our combined approach reaches similar results as in the case where these time series are used to train the ML network. Indeed, this was the case for the entire network in Switzerland, where the reported accuracy is in the range of 1–2 cm (depending on the station) in terms of RMS.

Therefore, we can produce tropospheric delays (and thus refractivity fields) with very high temporal and spatial resolution simultaneously, without processing any GNSS data. These outputs can be beneficial for low-cost GNSS receivers (to model their tropospheric effects), as well as to cross validate the estimated delays of permanent stations. Furthermore, they can be used in precise navigation to enhance Precise Point Positioning (PPP) methods, or to correct the tropospheric effects in other remote sensing techniques such as InSAR. Finally, their contribution to meteorology, climatology and hydrology (although at this point out of the scope of this work) may be considered, especially for locations, where the density of meteorological stations is low.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Swiss National Science Foundation (SNSF) for financing this work (project 200021E-168952) and swisstopo and MeteoSwiss for providing the GNSS measurements and meteorological data.

References

- [1] World Meteorological Organization, «public.wmo.int.», 31, Aug 2021 [Online]. Available: <https://public.wmo.int/en/media/press-release/weather-related-disasters-increase-over-past-50-years-causing-more-damage-fewer>. (Accessed 24 September 2021). Zugriff am.
- [2] A.C. Bernardes Parracho, Study of Trends and Variability of Atmospheric Water Vapour with Climate Models and Observations from Global Gns Network, PhD Thesis, Université Pierre et Marie Curie, 2018.
- [3] J. Saastamoinen, Contributions to the theory of atmospheric refraction: part II. Refraction corrections in satellite geodesy, *J. Geodes.* 107 (1973) 13–34.
- [4] H.S. Hopfield, Tropospheric effect on electromagnetically measured range: prediction from surface weather data, *Radio Sci.* 6 (3) (1971) 357–367.
- [5] M. Bevis, S. Businger, T.A. Herring, C. Rocken, R.A. Anthes, R.H. Ware, GPS meteorology: remote sensing of atmospheric water vapor using the Global Positioning System, *J. Geophys. Res.* 97 (1992) 15787–15801.
- [6] J.L. Davis, M.L. Cosmo, G. Elgered, Using the global positioning system to study the atmosphere of the earth: overview and prospects,» in GPS trends in precise terrestrial, airborne, and spaceborne applications, in: IAG Symposium vol. 115, Springer, Berlin, 1996, pp. 233–242.
- [7] A.E. Hassanien, *Machine Learning Paradigms: Theory and Application*, Springer, 2018.
- [8] L. Miotti, E. Shehaj, A. Geiger, D'Aronco, J.D. Wegner, G. Moeller, M. Rothacher, Tropospheric delays derived from ground meteorological parameters: comparison between machine learning and empirical model approaches, in: European Navigation Conference, 2020.
- [9] L. Miotti, *Machine Learning Approaches for Estimating Tropospheric Zenith Path Delay*, Master thesis, ETH Zurich, Zurich, 2019.
- [10] C. Kitpracha, S. Modiri, M. Asgarimchr, R. Heinkelmann, H. Schuh, Machine Learning Based Prediction of Atmospheric Zenith Wet Delay: A Study Using GNSS Measurements in Wettzell and Co-located VLBI Observations, EGU2019-4127 vol. 21, 2019, 2019.
- [11] R. Shamshiri, M. Motagh, H. Nahavandchi, A Machine Learning-Based Regression Technique for Prediction of Tropospheric Phase Delay on Large-Scale Sentinel-1 InSAR Time-Series, EGU, 2019.
- [12] B. Zhang, Y. Yao, Precipitable water vapor fusion based on a generalized regression neural network, *J. Geodes.* 95 (36) (2021).
- [13] V. Eckert, M. Cocard, A. Geiger, COMEDIE:(Collocation of Meteorological Data for Interpretation and Estimation of Tropospheric Pathdelays) Teil I: Konzepte, Teil II: Resultate; Technical Report 194, ETH Zürich. Grauer Bericht, 1992.
- [14] V. Eckert, M. Cocard, A. Geiger, COMEDIE :(Collocation of Meteorological Data for Interpretation and Estimation of Tropospheric Pathdelays) Teil III: Software; Technical Report 195, ETH Zürich. Grauer Bericht, 1992.
- [15] M. Troller, GPS based Determination of the Integrated and Spatially Distributed Water Vapor in the Troposphere, Geodätisch-geophysikalische Arbeiten in der Schweiz, Swiss Geodetic Commission., 2004.
- [16] E. Shehaj, K. Wilgan, O. Frey, A. Geiger, A collocation framework to retrieve tropospheric delays from a combination of GNSS and InSAR, *Navigation* 67 (2020) 823–842.
- [17] J.M. Rueger, «Refractive Index Formulae for Radio Waves,» in Proceedings of the FIG XXII International Congress, 2002. Washington, D.C. USA.
- [18] Swisstopo [Online]. Available: <http://pnac.swisstopo.admin.ch/pages/en/agnes-status.html#>, 2019. last visited: June 2019.
- [19] MeteoSwiss, «Automatic Monitoring Network: SwissMetNet Project,», 2020 [Online]. Available: <https://www.meteoswiss.admin.ch/home/measurement-and-d-forecasting-systems/land-based-stations/automatisches-messnetz.html>. (Accessed 30 January 2020). Zugriff am.
- [20] F. Hurter, GNSS meteorology in spatially dense networks, Geodätisch-geophysikalische Arbeiten in der Schweiz, Swiss Geodetic Commission., 2014.
- [21] K. Wilgan, F. Hurter, A. Geiger, W. Rohm, J. Bosy, Tropospheric refractivity and zenith path delays from least-squares collocation of meteorological and GNSS data, *J. Geodes.* 91 (2017) 117–134.
- [22] C. Nwankpa, W. Ijomah, A. Gachagan, S. Marshall, Activation Functions: Comparison of Trends in Practice and Research for Deep Learning, 2018 *arXiv: 1811.03378 [cs.LG]*.
- [23] S.O. Haykin, *Neural Networks and Learning Machines*, Pearson, 2009.
- [24] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, *Neural Network.* 2 (1989) 359–366.