Zenith Troposphere Delay Prediction based on BP Neural Network and Least Squares Support Vector Machine

Poster-EGU2020-5417
Prof. Dr. Xu Tianhe
Shandong University at Weihai
thxu@sdu.edu.cn

Tianhe Xu[1] Song Li[2] Nan Jiang[1]

Shandong University, Institute of Space Science, 180 West Wenhua Road, Weihai, P.R. China, 264209
Changan University, School of Geological and Surveying Engineering, Xi'an, China,710054

Abstract

With the rapid development of artificial intelligence, machine learning has become an high-efficient tool applied in the fields of GNSS data analysis and processing, such as troposphere, ionosphere or satellite clock modeling and prediction. In this paper, zenith troposphere delay (ZTD) prediction algorithms based on BP neural network (BPNN) and least squares support vector machine (LSSVM) are proposed in the time and space domain. The main trend terms in ZTD time series are deducted by polynomial fitting, and the remaining residuals are reconstructed and modeled by BPNN and LSSVM algorithm respectively. The test results show that the performance of LSSVM is better than that of BPNN in term of prediction stability and accuracy by using ZTD products of International GNSS Service (IGS) of 20 stations in time domain. In order to further improve LSSVM prediction accuracy, a new strategy of training samples selection based on correlation analysis is proposed. The results show that using the proposed strategy, about 80% to 90% of the 1-hour prediction deviation of LSSVM can reach millimeter level depending on the season, and the percentage of the prediction deviation value less than 5 mm is about 60% to 70%, which is 5% to 20% higher than that of the classical random selection in different month. The mean values of RMSE in all 20 stations using the new strategy are 1-3mm smaller than those of the classical one. Then different prediction span from 1 to 12 hours is conducted to show the performance of the proposed method. Finally, the ZTD predictions based on BPNN and LSSVM in space domain are also verified and compared using GNSS CORS network data of Hong Kong, China.

New model LSSVM_ZTD

The core idea of new model is that historical ZTD are viewed as inputs to make informed estimates that are predictive in determining the direction of future trends, and ZTD of 24 hours before the starting moment of forecasting are considered as historical data, ZTD of few hours after starting moment need to be predicted. There are two sets of simulation experiments to determine two independent variables affecting precision of new model, ***strategy for selecting training samples and the time span of prediction.***

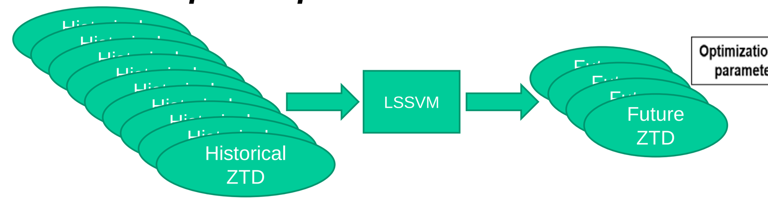

Fig 8.Topological diagram of the LSSVM_ZTD model

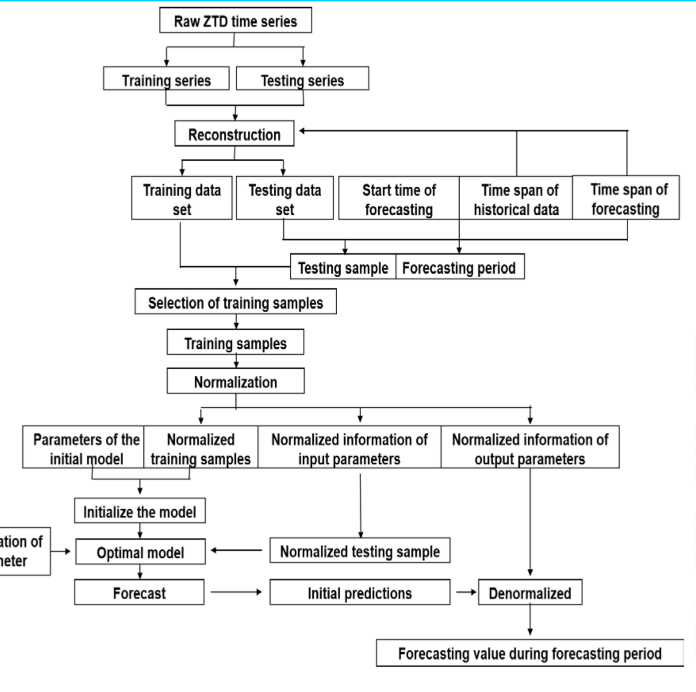

Fig 7. Flow chart of LSSVM_ZTD model

New model LSSVM_ZTD (strategy)

Strategy 1: randomly select ZTD in consecutive periods to form set $\{x_{train\_i}, y_{train\_i}\}$.
Strategy 2: choose training samples that are more relevant to the testing sample, which means that x_{train_i} with smaller euclidean distance from the testing sample x_{test_i}. Take consideration of the fact that the further out the forecast, the higher the chance that the estimate will be inaccurate, the time span ***T*** are fixed as **one hour** When determining strategies.

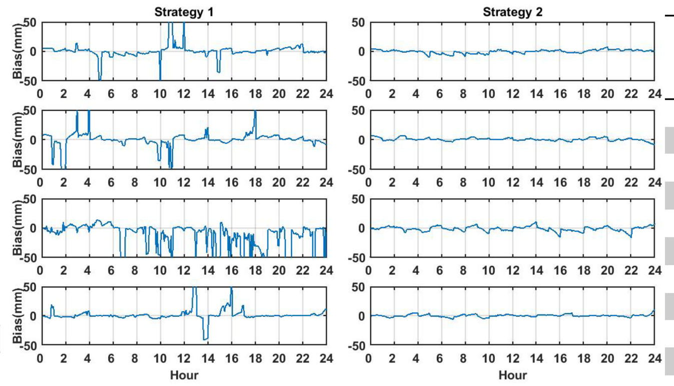

Fig 9. Bias of different strategies at FAIR station

Table 4. RMSE with different strategies at different stations

Station	strategy 1		strategy 2	
	Availability rate	RMSE (mm)	Availability rate	RMSE (mm)
GODE	29.17%	46.81	57.29%	4.23
USUD	58.33%	29.83	76.04%	4.74
DRAO	78.13%	7.72	88.54%	2.56
LHAZ	59.03%	15.91	92.01%	2.45
YELL	80.90%	9.41	93.40%	3.28
AIRA	68.75%	25.56	77.08%	4.06
FAIR	83.68%	13.11	97.22%	2.59
JOZE	67.01%	15.64	76.04%	4.23
MANA	65.63%	18.40	69.44%	3.83
ONSA	47.92%	22.50	63.89%	3.85
VILL	56.25%	16.80	78.47%	4.07
mean	63.16%	20.15	79.04%	3.63

Table 3. RMSE with different strategies at FAIR station

month	strategy 1	strategy 2
12-02	7.85 [1.41,37.91]	2.63 [0.52,6.47]
03-05	9.48 [0.51,37,51]	2.10 [0.31,6.05]
06-08	28.62 [1.90,113.59]	3.84 [0.87,9.59]
09-11	6.49 [0.39,45.79]	1.79 [0.38,4.93]
mean	13.11	2.59

(1) The accuracy of model built with strategy 2 is higher than that of model with strategy 1, among them, the average availability rate of eleven IGS stations increased by 15.88%, and the average RMSE decreased by 16.52mm.

Introduction of data and experimental results

The involved data are all from IGS_ZPD files provided by IGS, and the ZTD data at a 5-min time intervals are used as data source, the 11-year time series of ZTD observations are used to model and forecast **by BPNN and LSSVM.** ZTD from 2008 to 2019 are defined as the training data, ZTD of 2019 are assumed to be unknown data used for testing the precision of model built. These sites involved are ARTU, BJFS , CHAN, GANP, GODE, USUD, DRAO, LHAZ, KOUR, YELL, AIRA, FAIR, JOZE, ONSA, ZIM2, PERT, VILL, SANT, TLSE.

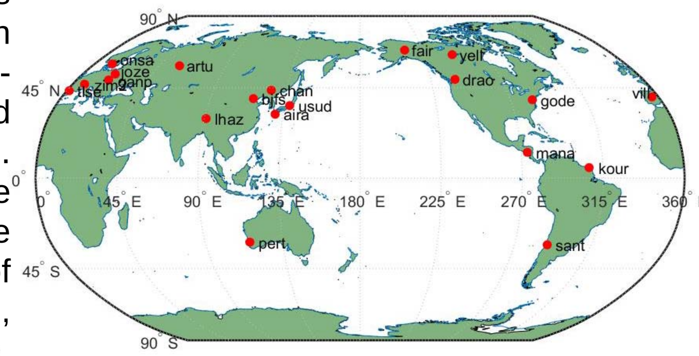

Fig 1. IGS stations distribution

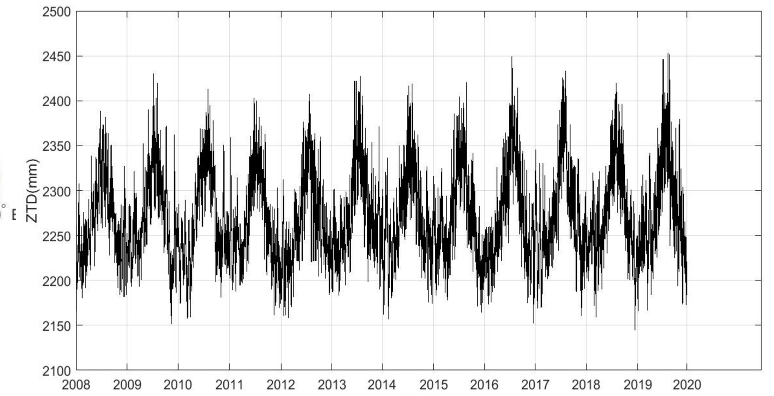

Fig 2. Raw ZTD time series at FAIR site

(1)The accuracy of ZTD provided by IGS can reach 4-5mm, with higher precision and large data volume.
(2)The ZTD time series has obvious annual and semiannual periods, and there are also complicated high-frequency signal that is hard illustrated by special formulas.
(3)Results of experiments at stations with data integrity over 90% are chosen for analysis, which reflect the relationship between ZTD and time parameters more comprehensively.
(4)Predictions of FAIR station with the highest data integrity are taken as an example to visualize the experimental results of model.

The ZTD of forecast period from IGS are regarded as the true value, and there are three indicators for evaluating the accuracy of model, 1) the difference between predictions of model and true value, called **bias**, is basic indexes; 2) the predicted value is considered as the available value if the bias of it is less than 5mm, by which all predicted value are determined whether it is available. **The availability rate**, the number of usable value divide the total number of predicted results, is secondary indicator; 3) **the root mean square error (RMSE)**, is core indicator. If there are many models of the same type, we will average RMSE of them to evaluate accuracy of this type model.

New model LSSVM_ZTD (time span)

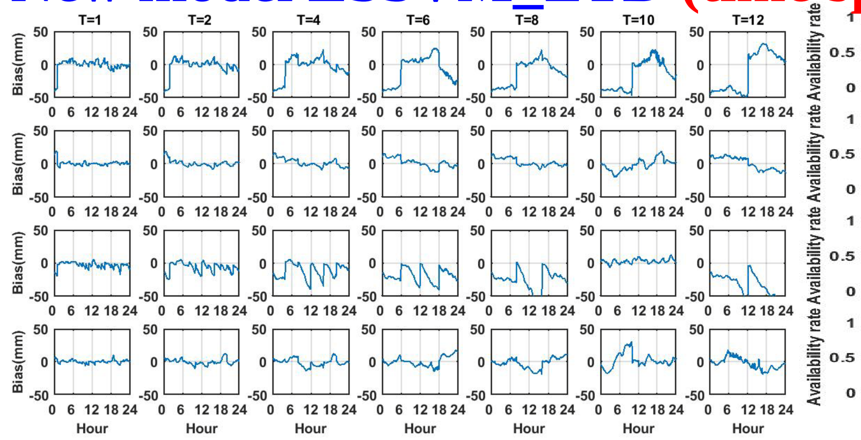

Fig 10 .Bias with different span at FAIR station

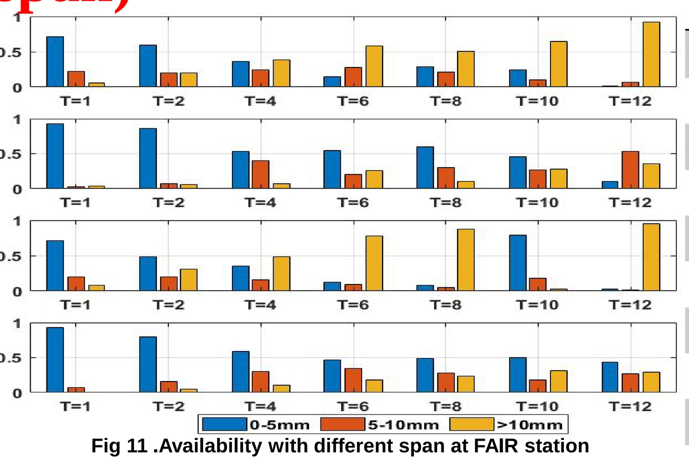

Fig 11 .Availability with different span at FAIR station

Table 5. RMSE with different span at FAIR station (mm)

Month	T=1	T=2	T=4	T=6	T=8	T=10	T=12
12-02	5.35 [0.73,38.96]	8.11 [2.11,39.17]	12.61 [1.61,37.60]	19.35 [5.20.36.36]	17.94 [6.70,36.36]	21.76 [9.75,39.79]	30.03 [18.72,41.34]
03-05	2.40 [0.23,16.13]	3.72 [1.47,13.77]	5.42 [2.43,11.02]	6.51 [2.90,11.17]	5.18 [2.80,9.38]	7.25 [2.62,10.93]	9.31 [9.20,9.41]
06-08	4.43 [0.93,16.35]	8.56 [1.69,21.00]	14.66 [3.48,23.95]	21.18 [16.66,26.40]	28.49 [18.90,41.97]	4.35 [3.01,5.60]	36.07 [30.12,42.03]
09-11	2.21 [0.60,6.32]	3.62 [1.42,9.11]	5.38 [2.28,9.11]	6.72 [3.65,11.15]	7.32 [3.67,11.92]	8.02 [3.80,16.26]	8.55 [6.79,10.31]
mean	3.60	6.00	9.52	13.44	14.73	10.34	20.99

Table 6. Availability and RMSE with different duration of different stations (mm)

Station	T=1		T=2		T=4		T=6		T=8		T=10		T=12	
	RMSE	Availability	RMSE	Availability	RMSE	Availability	RMSE	Availability	RMSE	Availability	RMSE	Availability	RMSE	Availability
GODE	3.87	79.70%	7.07	60.40%	12.65	34.90%	18.77	38.80%	23.34	23.60%	15.20	27.00%	35.18	13.50%
USUD	5.23	70.20%	9.50	52.00%	11.58	46.40%	15.99	39.30%	17.03	35.70%	16.99	22.40%	20.72	27.20%
DRAO	3.76	77.50%	6.20	60.80%	8.71	43.50%	8.62	47.10%	10.33	37.90%	8.77	47.70%	11.68	27.60%
LHAZ	2.77	88.40%	4.56	70.60%	6.57	55.00%	7.09	60.60%	8.97	44.50%	9.85	51.90%	10.74	39.70%
YELL	3.45	81.80%	5.83	60.90%	8.66	48.20%	10.86	38.80%	11.36	40.10%	19.88	21.50%	13.49	34.60%
AIRA	8.53	61.70%	14.67	36.30%	23.64	25.60%	30.20	17.10%	34.15	16.60%	13.93	33.20%	43.23	12.80%
FAIR	3.60	82.30%	6.00	68.50%	9.52	46.00%	13.44	32.10%	14.73	36.40%	10.34	49.90%	20.99	14.70%
JOZE	4.25	73.80%	6.55	63.50%	7.96	55.90%	12.80	53.90%	18.42	43.00%	13.98	37.80%	22.34	28.90%
MANA	4.21	74.70%	6.41	57.00%	10.04	44.10%	11.62	39.70%	17.58	30.30%	18.08	29.30%	16.90	26.60%
ONSA	2.86	87.00%	4.84	71.80%	8.11	49.80%	10.69	47.80%	14.61	37.60%	13.41	26.40%	19.91	22.70%
VILL	3.74	78.50%	6.47	61.10%	8.52	48.70%	15.20	20.90%	13.41	36.00%	10.71	33.20%	17.35	16.60%
mean	4.20	77.80%	7.10	60.30%	10.54	45.30%	14.12	39.60%	16.72	34.70%	13.74	34.60%	21.14	24.10%

according to the last experiment, the training samples more relevant to the testing sample are selected **(strategy 1)**.
(1) The forecasting accuracy of model decreases with the increase of forecasting span, when span is two hours, the average RMSE of eleven stations closes to 7mm, and the average availability rate is above 60%.

Classic model BP_ZTD

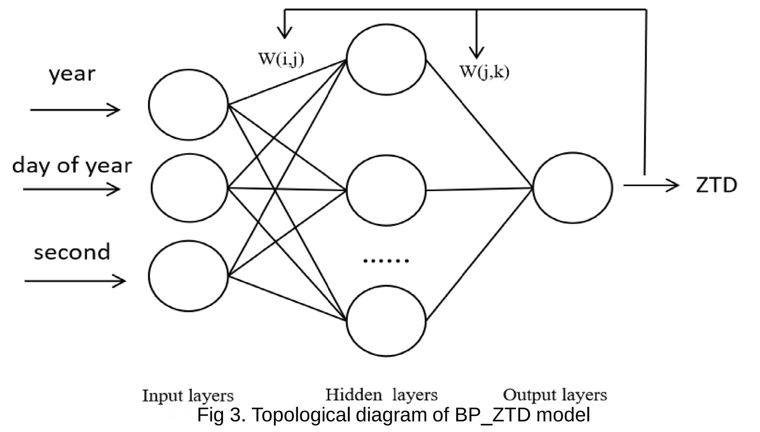

Fig 3. Topological diagram of BP_ZTD model

Table 1.Parameters of BPNN

parameter name	Value
Input_train	year,day of year,second(2008-2019)
Output_train	ZTD(2008-2019)
Input_test	year,day of year,second(2019-2020)
Output_test	ZTD(2019-2020)
hiddenum	7
net.divideFcn	'dividerand'
net.layers{1}.transferFcn	'tansig'
net.layers{2}.transferFcn	'purelin'
net.trainParam.epochs	100
net.trainParam.lr	0.1
net.trainParam.goal	0.001
net.trainParam.max_fail	8

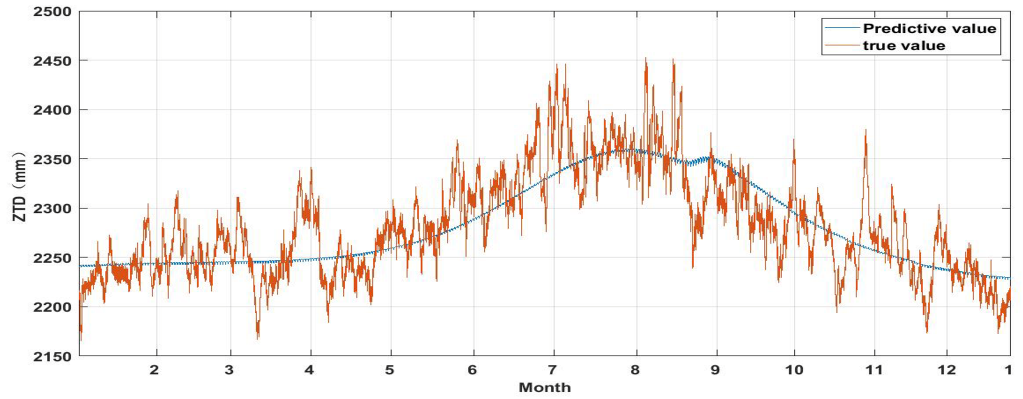

Fig 4. Predictions and true values at FAIR station

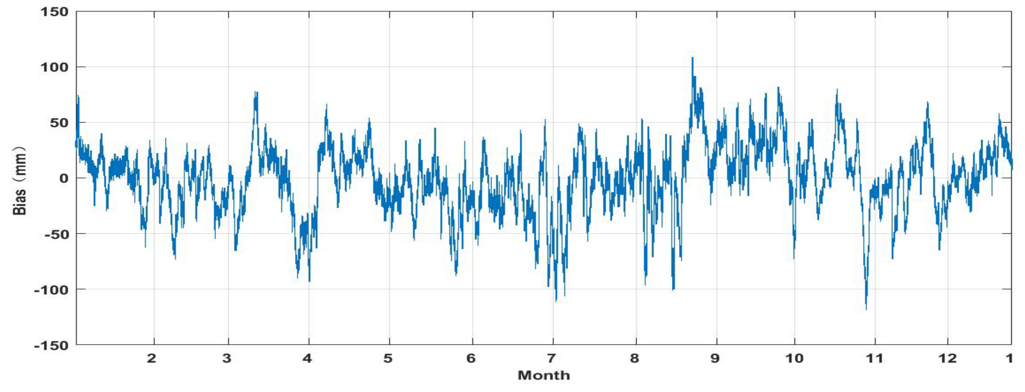

Fig 5. Bias time series at FAIR station

Table 3. Forecast results at various stations

Station	Data integrity	RMSE (mm)	Availability rate
GODE	95.07%	69.87	4.68%
USUD	96.33%	37.71	11.87%
DRAO	96.81%	27.32	15.03%
LHAZ	94.84%	17.89	23.24%
YELL	96.94%	24.68	16.56%
AIRA	96.01%	60.99	5.06%
FAIR	98.84%	31.35	13.10%
JOZE	94.93%	39.06	10.97%
MANA	90.08%	69.83	6.68%
ONSA	94.84%	38.95	10.87%
VILL	96.17%	33.01	10.46%
mean	95.53%	40.97	11.68%

(1) The classic model BP_ZTD can well predict ZTD annual period rather than high-frequency signals. It is failed to predict high-frequency signals of the ZTD time series.
(2) There are lower accuracy of BP_ZTD model, most of bias range from -50 to 50mm, and the maximum value of bias is more than 100 mm, the average RMSE of eleven sites is around 40mm, the maximum value of RMSE reaches 69 mm, the minimum value closes to 17mm, ZTD can be predicted with the centimeter accuracy by the BP_ZTD model and about 10 percent of ZTD predictions are viewed as the available..

New model LSSVM_ZTD(comparision of different models)

Based on the conclusions of the above two simulation studies, the new model LSSVM_ZTD with training samples having **high correlation with testing samples for modeling (strategy 2)** and **two-hour time span of forecasting are built,** by which we can predict ZTD of any sites at any consecutive two hours (in unit of whole hour) as long as we know the 24-hour historical ZTD immediately before the beginning moment of forecasting and the ZTD data of this station for the past few years. 30-day result are obtained through loop constructing many LSSVM_ZTD models for different testing samples.

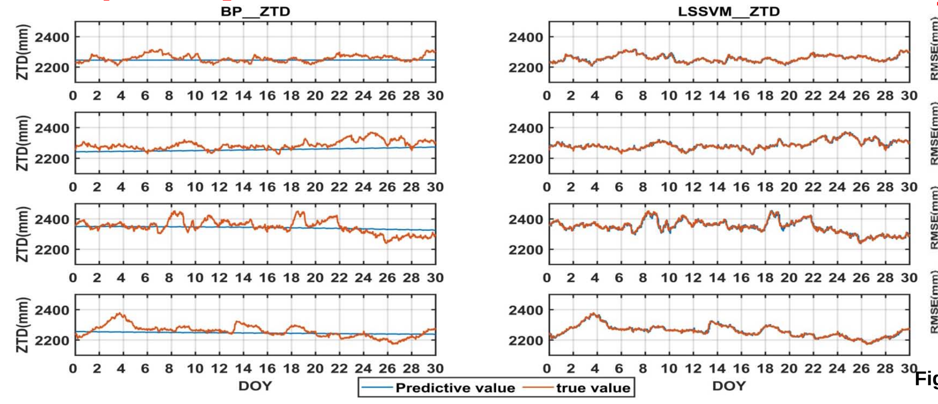

Fig 12 . Forecast results of BP ZTD and LSSVM ZTD at FAIR station

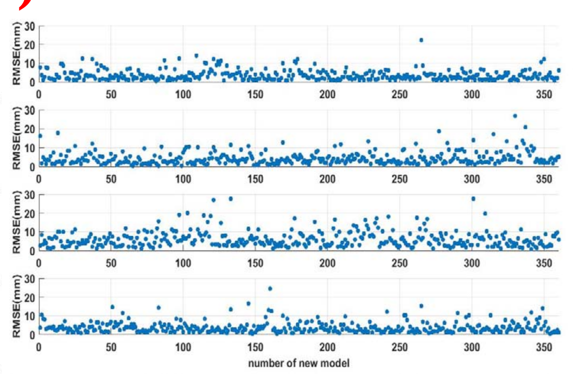

Fig 15. RMSE of different LSSVM_ZTD models at FAIR station

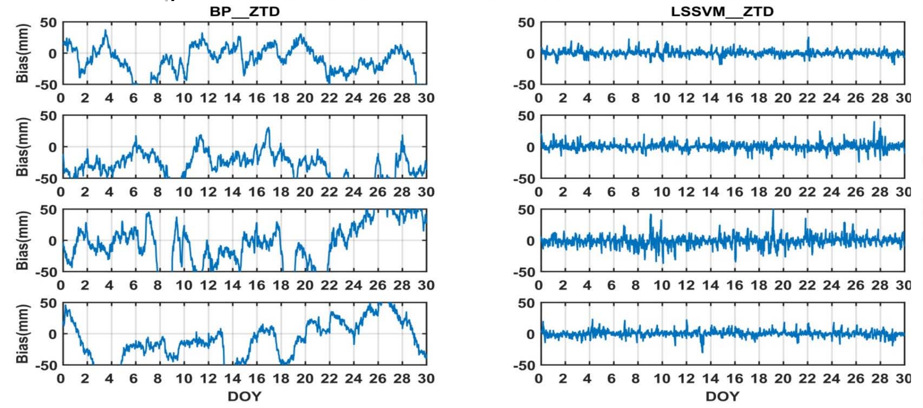

Fig 13 .Bias of BP_ZTD and LSSVM_ZTD at FAIR station

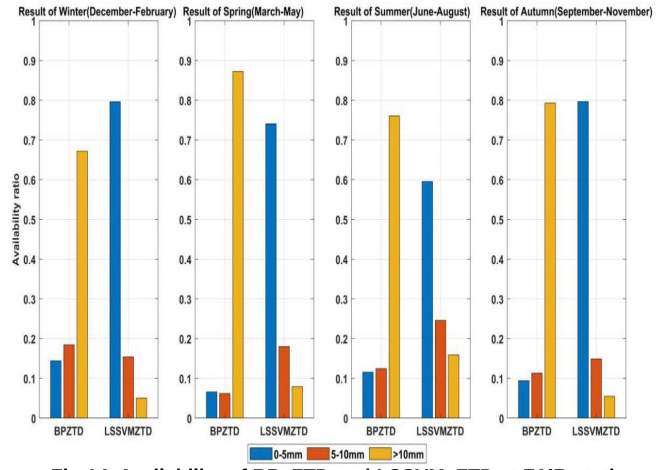

Fig 14. Availability of BP_ZTD and LSSVM_ZTD at FAIR station

Table 7. RMSE with different models at FAIR station (mm)

Month	BP_ZTD	LSSVM_ZTD
12-02	24.59	3.66 [0.62，22.40]
03-05	39.50	4.38 [0.45，26.69]
06-08	39.54	6.35 [0.92，37.93]
09-11	37.36	3.78 [0.59，24.39]
mean	35.25	4.54

Table 8. Result with different models at different stations

Station	BP_ZTD		LSSVM_ZTD	
	Availability	RMSE (mm)	Availability	RMSE (mm)
GODE	7.55%	57.4	55.67%	7.81
USUD	8.42%	35.88	57.40%	7.14
DRAO	14.20%	29.52	71.08%	4.75
LHAZ	20.65%	17.82	74.11%	4.33
YELL	18.02%	22.62	76.57%	4.20
AIRA	4.50%	61.45	50.46%	8.24
FAIR	10.49%	35.25	73.22%	4.54
JOZE	9.69%	34.81	59.75%	6.29
MANA	8.65%	37.91	53.22%	7.72
ONSA	14.99%	35.79	67.38%	5.54
VILL	10.29%	34.30	64.37%	5.86
mean	11.59%	36.61	63.93%	6.04

Table 9 Efficiency with different models at different stations

Station	BP_ZTD	LSSVM_ZTD
	Time (s)	Time (s)
GODE	73.48	105.71
USUD	73.69	106.05
DRAO	35.81	108.54
LHAZ	54.41	106.67
YELL	73.99	106.02
AIRA	72.18	103.11
FAIR	74.22	104.76
JOZE	71.02	110.37
MANA	69.89	108.13
ONSA	73.90	107.05
VILL	73.83	105.15
mean	67.86	106.51

(1)The prediction results of the new model LSSVM_ZTD are analyzed and compared with the that of the classic model BP_ZTD, the results show that the prediction accuracy of the new model LSSVM_ZTD is an order of magnitude higher than that of the classic model BP_ZTD, where the average RMSE of eleven stations is reduced by 30.57 mm, and the average availability rate is increased by 52.34%, the efficiency of LSSVM_ZTD model is slightly below that of BP_ZTD model, where it takes about 100 seconds to build a LSSVM_ZTD model.
(2)when modeling the ZTD time series changing drastically, such as summer, there is the poor accuracy of LSSVM_ZTD model.

Conclusions

(1)The prediction accuracy of classic model BP_ZTD is at the centimeter level.
(2)The prediction accuracy of the new model LSSVM_ZTD is an order of magnitude higher than that of the classic model BP_ZTD.
(3) LSSVM_ZTD with selecting the training samples by correlation analysis can further improve the prediction accuracy.

References

[1] Jan Douša, Michal Eliaš (2018) A two-stage tropospheric correction model combining data from GNSS and numerical weather model. GPS Solutions. https://doi.org/10.1007/s10291-018-0742-x.
[2] Ma Jianwu, Tao Tingye, Yin Weisong (2017). GPS tropospheric Delay Interpolation Model Based on RBF Neural Network [J] . Metal Mine(10): 33-35.
[3] Zou Lele(2019), Wind Power Prediction Based on Improved Particle Swarm Optimization Algorithm for Least Squares Support Vector Machine [D]. Xinjiang University.
[4] Xia Tan, Zhong Jia, Yadan Zhang (2018)Non-invasive continuous blood pressure measurement based on mean impact value method, BP neural network, and genetic algorithm. Technology and Health Care 26 (2018) S87–S101 S87 DOI 10.3233/THC-174568.
[5] V. Cherkassky, and Y. Ma(2004), Practical Selection of SVM Parameters and Noise Estimation for SVM regression, Neural Networks, 17(1), 2004, 113-126.

Acknowledgements

This work was supported by Natural Science Foundation of China (41874032) and the National Key Research and Development Program (2016YFB0501701)