The Spatial Distribution of Atmospheric Water Vapor Based on Analytic Hierarchy Process and Genetic Algorithm

Fengjun Wei[1], Chunhua Liu[2], Rendong Guo[3*], Xin Li[4], Jilei Hu[5], Chuanxun Che[6]
School of Energy and Building Engineering, Shandong Huayu University of Technology, Dezhou, 253034, China[1, 2, 3, 6]
Research and Development Center of Building Energy Saving Engineering Technology, Shandong Huayu University of Technology, Dezhou, 253034, China[1, 2, 3]
School of Civil Engineering, Shenyang University, Shenyang, 110044, China[4]
Key Laboratory of Geological Hazards on Three Gorges Reservoir Area-Ministry of Education, China Three Gorges University, Yichang, 443002, China[5]

***Abstract*—The inversion of water vapor spatial distribution using ground-based global navigation satellite systems is a technique that utilizes the propagation delay of satellite signals in the atmosphere to retrieve atmospheric water vapor information. To further promote the accuracy of the information obtained by this method, a satellite system is designed to solve the spatial distribution of atmospheric water vapor based on chromatography technology and genetic algorithm. Firstly, the accuracy of the empirical air temperature and pressure model to calculate the zenith statics delay is analyzed. To optimize the global weighted average temperature model, a model that considers the decreasing rate of atmospheric weighted average temperature and a model based on the linear relationship between surface heat and weighted average temperature are proposed. The idea of removal interpolation restoration is introduced to achieve regional interpolation of atmospheric precipitable water. Finally, in response to the problem of multiple solutions in the current water vapor chromatography equation, a genetic algorithm based chromatography method is put forward to achieve the solution of atmospheric water vapor spatial distribution. The experimental analysis shows that the average root mean square error and average absolute error of the design method of the research institute are 1.78g/m3 and 1.41g/m3, respectively, which can realize the calculation of atmospheric water vapor density distribution with high accuracy.**

Keywords*—Global navigation satellite system; spatial distribution of water vapor; genetic algorithm; chromatography technology

I. Introduction

The atmosphere is a crucial carrier of the earth's gas and water cycle, providing oxygen for humans and animals, carbon dioxide for plant photosynthesis, and water for all life [1]. Through the greenhouse effect, the atmosphere maintains a suitable temperature on the Earth, and through the absorption of harmful rays by ozone, protects earth's organisms from harm. Water vapor is an important component of the atmosphere, mainly concentrated in the lower atmosphere. It has an important impact on the atmospheric greenhouse effect and circulation [2,3]. The atmosphere water vapor has complex temporal and spatial changes, and accurately measuring, modeling, and predicting the content and changes of water vapor is of great importance for studying the greenhouse effect and climate change. With the continuous expansion of Global Navigation Satellite System (GNSS) and its related fields, GNSS technology used to study the atmosphere water vapor spatial distribution (AWVSD) has made significant progress [4,5]. To achieve a more accurate solution of water vapor distribution, research is conducted on the estimation of atmospheric precipitable water volume (PWV) based on GNSS technology, and the idea of removal interpolation restoration is introduced to optimize the accuracy of the calculation. In terms of three-dimensional water vapor chromatography, a chromatography manner with genetic algorithm (GA) is proposed to address the problems in the chromatography equation.

The gap between research and existing national and international research: Currently, there are generally four main methods for calculating atmospheric water vapor content. The first method is direct calculation using measured sounding data. Sounding data are relatively old, objective and accurate, so they are often used to verify other calculation methods. However, due to the limited number of sounding stations, it is not possible to describe the spatiotemporal distribution of regional water vapor in detail, especially for areas with uneven terrain loads and stations. The second is to use remote sensing data to invert the atmospheric water vapor content, a new technology that has been developed and widely used in recent years, but whose measurement accuracy still needs to be improved. The third is to establish the relationship between the water vapor content and the surface meteorological elements for the calculation. The calculation is simple and the results are ideal, but this greatly increases the number of stations. Fourth, NCEP/NCAR reanalysis data is currently the most widely used method for calculating water vapor content. Gridded data can compensate for the shortcomings of insufficient stations in calculating the spatial distribution of water vapor, but the coarse grid affects the fine analysis and its applicability is not high enough. The method developed by the Research Institute is an improved method based on the use of remote sensing data to invert the atmospheric water vapor content, ensuring high accuracy in the spatial distribution of water vapor with fewer stations.

Motivation and potential benefits of the research: Over the past century, the global climate has undergone a significant change characterised by global warming, and the warming of the climate system has become an indisputable fact. And global warming has also led to more frequent natural disasters caused by extreme weather events, with extremely high losses for society. Extreme precipitation often leads to more severe flood disasters. Detecting water vapor in real time and controlling its spatiotemporal changes, as well as studying its interaction with precipitation, is of great research importance for precipitation forecasting. However, traditional water vapor detection methods have drawbacks such as high operational costs, large station requirements and insufficient accuracy. The use of GNSS to invert the spatial distribution of water vapor has good performance, but the accuracy needs to be improved. Research is underway to improve this technology to achieve high performance water vapor spatial distribution detection.

The main contributions and impacts of the research include:

1) Using different model expressions, different practical resolutions, and different spatial resolution modeling data, the impact characteristics of different modeling factors on the accuracy of empirical temperature and pressure models were studied. A temperature and pressure model based on segmented practical ideas was constructed, and the accuracy of calculating ZHD estimates for different ZHDs in measured meteorological parameters was evaluated within the context of de Qianqiu.

2) The weighted average temperature decline rates for different height intervals were obtained and analyzed, and a CPT2wh model was established. At 1 × Provide the average value, annual amplitude, and semi-annual amplitude of Tm decline rate on a grid point. The proposed model effectively improves the lack of elevation correction in Tm estimation and significantly improves the accuracy of Tm estimation for stations with large elevation differences.

3) A new method for detecting the spatial distribution of water vapor based on genetic algorithm and Analytic Hierarchy Process has been proposed. This method does not need to rely too much on the conditions of the Moon Lake, prior information and external meteorological data, and converts the inverse problem of the matrix into the function optimization problem, which effectively solves the ill conditioned problem of the equation.

4) A GNSS-PWV spatial interpolation method based on the idea of removal interpolation recovery is proposed. This method does not require the surface measured meteorological data of the station to be measured, nor does it require regression analysis of the observed data, which can effectively avoid the issue of elevation correction in PWV interpolation.

The research and work can address the shortcomings of traditional methods of spatial detection of water vapour and provide more accurate detection results. The data will be useful as a guide for government planning and regional responses to climate change.

The specific content of the study is divided into three parts:

The first part is the literature review section, which analyzes the current research status at home and abroad, and explores the key points that need to be overcome in the research.

The second part is the methodology section, which mainly introduces the technologies required for the AWVSD calculation method designed by the research institute, and makes improvements to address the limitations of related technologies.

The third part is the experimental analysis section, which mainly analyzes the performance of the research institute's design methods.

II. Related Works

The amount of water content in clouds is an important parameter for studying the impact of clouds on climate. Many scholars have used different methods to calculate and analyze the AWVSD characteristics. Shi et al. used the HYSPLIT platform to simulate the Lagrangian trajectory of air envelopment in East China during the summer monsoon. It investigated four different periods during its seasonal migration from south to north [6]. Huang et al. used radio temperature measurement data from 2012 to 2017 to establish an empirical model of atmospheric weighted average temperature (Tm) in Guilin, China. Then, they used observation data from 11 GNSS stations in Guilin to study the spatiotemporal characteristics of GNSS derived PWV under heavy rain from June to July 2017 [7]. Lee et al. found in their observation of the time process of H2O2 generation in condensate droplets that it was typically generated from droplets smaller than 10 microns [8]. Tan et al. obtained the land surface temperature (LST) of Dongting Lake (China) from Landsat 7 data, and discussed its relationship with land cover (LULC) type. The outcomes denoted that LST varied greatly among different LULC types, with higher LST in building areas and lower LST in other areas. Water bodies played a critical supervision role in reducing LST [9].

GNSS technology was gradually developing and becoming a focus of research for scholars. Pan et al. considered the advantages and disadvantages of remote sensing technology and Global Navigation Satellite System Interferometric Reflection (GNSS-IR) in trajectory recognition of water content (VWC), and proposed a point surface fusion method with GA combined with backpropagation neural network (GA-BP) for GNSS IR and MODIS data to raise the accuracy of VWC estimation [10]. Zheng et al. used the signal-to-noise ratio of navigation satellite signals to invert sea level based on observation data from MAYG on the east coast of Africa from 2017 to 2019 [11]. LÜ et al. used Differential Interferometric Synthetic Aperture Radar (InSAR) and pixel offset tracking to gain the line of sight displacement and near-field range displacement of the M6.9 earthquake in Menyuan, Qinghai from SAR images. At the same time, they obtained high-speed displacement waveforms of 16 GNSS stations through historically accurate point positioning schemes and inverted the seismic fracture process [12]. Lewen et al. conducted

monitoring of changes in the spatial environment of the line of sight based on global and regional GNSS reference station networks for inversion of tropospheric and ionospheric parameters [13].

GA is an efficient, parallel, and global search method. Jalali Z et al. used GAs and constant development manners to optimize the exterior walls of an office building. The SPEA-2 algorithm was utilized for multi-objective optimization [14]. Garud et al. extensively reviewed the applicability of artificial neural networks (ANN), fuzzy logic (FL), and GA, as well as their hybrid models, using AI for effectiveness prediction. Besides, some literature on predicting solar radiation using ANN, FL, GA, and their hybrid models has been summarized [15]. Shariati et al. collected a database of 1030 datasets to intelligently predict the strength of concrete containing slag and fly ash as partial substitutes for cement [16]. Abualigah et al. proposed an efficient task scheduling optimization method based on a multiverse optimizer and genetic algorithm (MVO-GA). MVO-GA was utilized to improve the effect of task transmission through cloud networks, provide appropriate transfer decisions, and rearrange transfer tasks with the efficiency weights collected in the cloud [17].

Based on the above literature analysis, there are various methods for calculating and analyzing the AWVSD, but the accuracy of the obtained AWVSD analysis methods still needs to be improved. To this end, the study utilizes GNSS technology to estimate atmospheric PWV, and then uses GA algorithm for water vapor chromatography. A solution model for AWVSD based on Analytic Hierarchy Process (AHP) and GA is obtained.

III. A Solution Model for AWVSD based on AHP and GA

Water vapor can effectively absorb solar longwave radiation and surface infrared radiation, thereby providing insulation for the entire Earth system. Its content in the atmosphere has complex spatial and temporal variations, constantly affecting weather characteristics and climate environments around the world. Therefore, the accurate measurement, modeling and prediction of water vapor levels and changes are of great importance for the study of the greenhouse effect and climate change. At the same time, an in-depth understanding of the spatiotemporal changes of water vapor plays an important role in improving the accuracy of weather forecasting and conducting disaster weather warnings. When electromagnetic waves pass through the Atmosphere of Earth, they are affected by the ionospheric delay and the flow delay, which will lead to the extension of the electromagnetic wave propagation practice and the bending of the propagation path. Therefore, monitoring and studying the distribution of water vapor in the atmosphere is of great importance for services such as radio communications, navigation, positioning and timing. In order to improve the accuracy of information obtained from current water vapor spatial distribution inversion techniques, research has been carried out on key technologies for obtaining atmospheric precipitable water and tomographic inversion of atmospheric water vapor density, with a focus on ground-based GNSS inversion. Firstly, the empirical air temperature and pressure model is analyzed to calculate the accuracy of the zenith Statics delay. In order to optimize the global weighted average temperature model, the model of worrying about the decline rate of the atmospheric weighted average temperature and the model based on the linear relationship between the surface heat and the weighted average temperature are proposed. And introduce the idea of removal interpolation restoration to achieve regional interpolation of atmospheric precipitable water. Finally, in response to the problem of multiple solutions in the current water vapor chromatography equation, a genetic algorithm based tomography method is proposed to achieve the solution of atmospheric water vapor spatial distribution.

1) GNSS-PWV spatial interpolation based on the idea of removal interpolation recovery: In the GNSS solution model, each satellite signal corresponds to a tropospheric delay, and directly estimating them will result in rank deficiency in the equation. The convective delay is modeled as the sum of zenith directional delay (ZTD) and the product of atmospheric horizontal gradient and their corresponding projection functions. The calculation method of ZTD is shown in equation (1).

$$\begin{aligned} STD &= m_h(ele)\times m_w(ele)\times ZWD \\ &+ m_\Delta(ele)\times\cot(ele)\times\left[G_{NS}\cos(azi)+G_{WE}\sin(azi)\right] \end{aligned} \quad (1)$$

In formula (1), STD means the total delay of tropospheric oblique path; m_h stands for the dry mapping function; m_w stands for the wet mapping function; m_Δ expresses the atmospheric horizontal gradient mapping function; G_{NS} and G_{WE} express the atmospheric horizontal gradient in the north-south and east-west directions respectively; ele and azi indicate the satellite altitude angle and azimuth angle respectively; ZWD denotes the zenith wet delay; ZHD means the zenith statics delay. Usually, atmospheric precipitable water vapor (PWV) and integrated water vapor (IWV) are utilized to characterize the information of atmospheric water vapor content. The conversion relationship between the two is shown in equation (2).

$$IWV = \rho_w \times PWV \quad (2)$$

In equation (2), ρ_w means the density of liquid water. ZWD and PWV's conversion relationship is shown in equation (3).

$$\begin{cases} PWV = \Pi \cdot ZWD \\ \Pi = \dfrac{10^6}{\rho_w \times R/m_w \times \left(k_2' + k_3/T_m\right)} \end{cases} \quad (3)$$

In equation (3), Π means the conversion factor; R denotes the general gas constant; m_w indicates the molar mass of wet air; k_2' and k_3 express the atmospheric refractive index constants; T_m means the atmospheric weighted average temperature. The process of inverting PWV through ground GNSS is shown in Fig. 1.

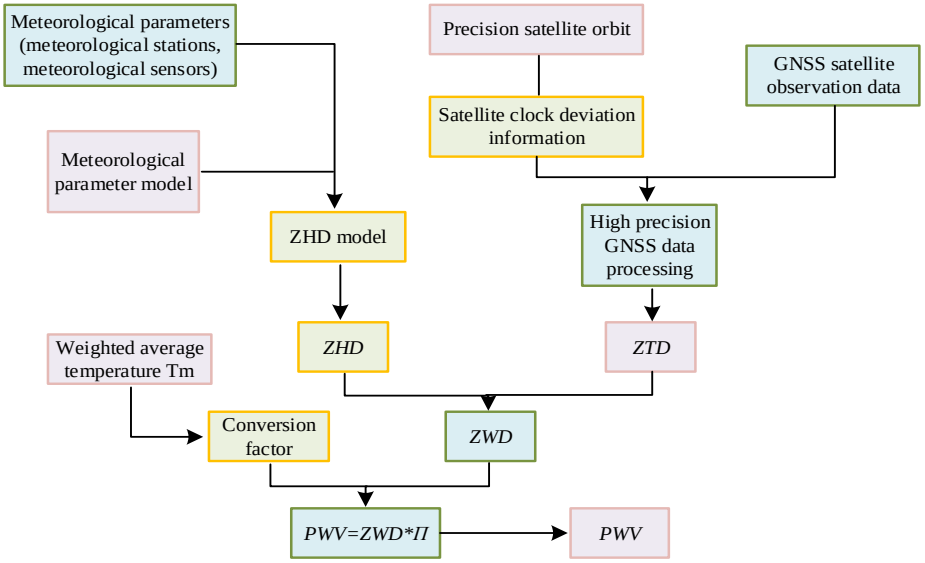


Fig. 1. Process of PWV inversion based on ground GNSS.

After obtaining the GNSS station ZWD, it is necessary to use conversion factors to obtain the atmospheric PWV. It is a function of the weighted Tm, and accurately obtaining Tm is the useful way to improving the accuracy of GNSS-PWV. To this end, the GPT2w model is applied to calculate Tm. When using GPT2w to calculate Tm, the study first selects the four model grid points closest to the desired station, and uses the model coefficients stored in external grid files to calculate the Tm estimates of the four grid points. Then, the interpolation algorithm is used to interpolate the Tm estimates of the four grid points to gain the Tm estimates of the station to be measured. However, during the research, it is found that the GPT2w model lacks elevation correction when calculating Tm, which limits the accuracy of the model in some cases, especially on the waiting points with significant elevation differences from the GPT2w grid points. For this purpose, the study adopts the index of virtual temperature, temperature decreasing rate, and water vapor pressure (WVP) decreasing factor to adjust the pressure, temperature, and WVP accordingly. At the same time, it applies the vertical Tm decline rate to the GPT2w model. Considering the limitations of the relationship between Tm and surface temperature (Ts) used globally, a GGTm Ts model is established using GGOSAtmosphere and ECMWF data. By providing high-precision Tm-Ts relationships on global grid points, more accurate Tm estimation can be achieved.

GNSS technology is used to gain PWV, which has become an important data source for meteorological departments. However, due to environmental and economic factors, GNSS stations are often scattered and have large distances, which limits the analysis of AWVSD in certain applications. Therefore, research is conducted on spatial interpolation of GNSS-PWV. It has a closed connection between water vapor and terrain, so it needs to consider the impact of terrain factors on PWV interpolation. To address the above issues, a method based on the idea of removal interpolation restoration is proposed for spatial interpolation of GNSS-PWV. The interpolation method is shown in Fig. 2.

Using the coordinates and time information of the measuring station, the GPT2w model can calculate meteorological parameters, and use these two parameters to estimate the ZWD of the measuring station. The calculation method is shown in equation (4).

$$ZWD = 10^{-6} \times \left(k_2' \times k_3 / T_m \right) \times \frac{R / m_d \times g_m}{(\lambda + 1)} \times e \quad (4)$$

In equation (4), m_d means the molar mass of dry air; g_m indicates the average gravity; e and λ denote the WVP and the decline rate of WVP respectively. After obtaining ZWD, it is converted into PWV using a conversion factor. The conversion method is shown in equation (5).

$$PWV = \frac{m_w \times g_m}{m_d \times p_w \times (\lambda + 1)} \times e \quad (5)$$

Using the GPT2w model, the PWV estimation at the corresponding time of the station calculated by using the above formulas (4) and (5) is denoted as GPT2w_PWV. The PWV true value obtained by using GNSS technology to high-precision process the observation data is denoted as GNSS_PWV. Firstly, it calculates the difference between the two types of PWVs based on PWV_residual. Then, the interpolation algorithm is used to interpolate the PWV_residual of the GNSS station to obtain the difference of the desired station. Finally, the obtained difference and its GPT2w_PWV are interpolated into the PWV of the station to be tested. This method achieves interpolation without the need for surface meteorological observation data and regression fitting processes, and considers the influence of elevation issues on PWV interpolation. The distribution and elevation of GNSS stations are shown in Fig. 3.

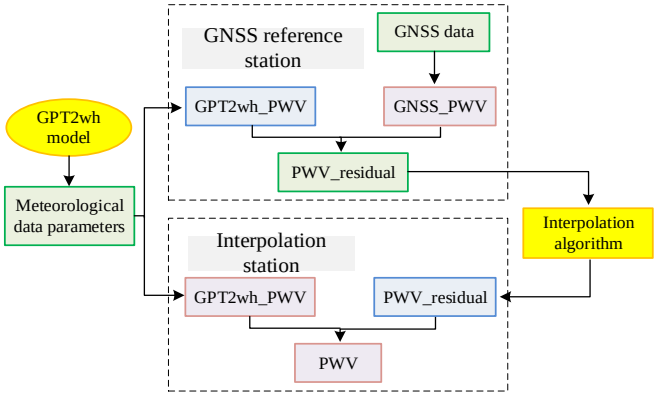


Fig. 2. Interpolation method process.

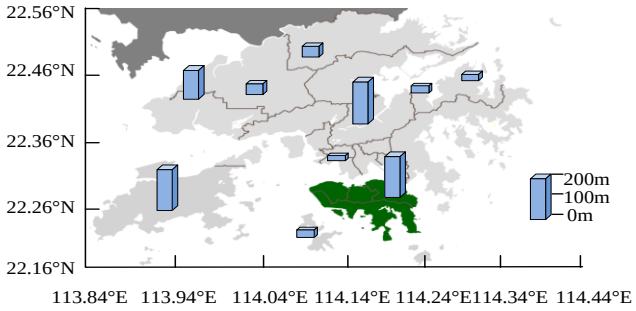


Fig. 3. Distribution and elevation of GNN stations.

Based on the above content, research has estimated PWV in the atmosphere using the GPT2w model and Tm-Ts model. However, GNSS stations are often scattered and have large distances, which limits the analysis of AWVSD in some applications. Research conducts the spatial interpolation of GNSS-PWV to improve the accuracy of existing PWV interpolation methods.

2) GA-based 3D water vapor AHP algorithm: Through GNSS observation data and meteorological information, precise PWV of the station can be obtained, while PWV is only the water vapor content in a unit low area column that runs through atmosphere. It is the average water vapor content in the zenith direction of multiple rays on the oblique path near the station, and cannot accurately show the variation in water vapor space, thereby limiting its application in meteorology such as weather forecasting. A GNSS water vapor tomography (WVT) manner is put forward, which uses optimization methods for equation solving to avoid matrix inversion, but relies heavily on constraint equations and external observation data. The parameters of the water vapor chromatography method based on genetic algorithm are: the population size is 200, the crossover probability is set to 0.8, the Choice function is Roulette, the crossover function is Intermediate, and the mutation function is adaptive feasibility.

Chromotography techniques (CT) refer to the method of inverting the spatial distribution of parameters using integrated observations from different positions and directions within the study area. In GNSS 3D WVT, the oblique path water vapor content (SWV) generated by the GNSS satellite signal passing through the tomography area is the observed measurement, and the WVD in each grid is the desired parameter. The WVT observation equation is established by calculating the intercept within each grid, as shown in equation (6).

$$SWV^q = \sum_{i=1}^{n} d_i^q \cdot x_i \qquad (6)$$

In equation (6), q indicates the satellite ray number for WVT; n means the total number of tomographic grids; d_i^q means the intercept of the q th satellite ray passing through the i th grid, and x_i refers to the WVD value within the i tomographic grid. The equation diagram is shown in Fig. 4.

Due to the fixed position of GNSS stations in the chromatography area, satellite signals will vary with their altitude and azimuth angles. When constructing the observation equation for water vapor chromatography, these SWVs passing through the side of the chromatography study area are defined as invalid satellite rays, and only the effective satellite rays passing through the top of the chromatography area are selected. It collects all available WWVs for water vapor chromatography and constructs a water vapor chromatography equation set using equation (7).

$$y_{m\times 1} = A_{m\times n} \cdot x_{n\times 1} \qquad (7)$$

In equation (7), y means the matrix containing all SWVs; m expresses the total number of SWVs that can be used for water vapor chromatography; A indicates the matrix

containing the intercept of satellite rays passing through the grid, where the number of never passing through the grid is 0. x means the matrix composed of all the parameters to be solved, and n means the total number of WVD in the grid to be solved. For some small-scale water vapor tomographic regions, the vertical layering of the tomographic grid is almost parallel, and the mutual exchange of water vapor densities in each layer of the grid may not affect the overall results, which leads to the multiplicity of observation equations. In addition, the limited number of GNSS stations in the tomographic region often leads to uneven and insufficient distribution of GNSS satellite signal rays in the tomographic region. Furthermore, there is a case where the coefficient matrix A is rank deficient. Therefore, in addition to the observation equation, the study also constructed horizontal, vertical and top-level constraints. The geometric diagrams of the first two constraint equations are shown in Fig. 5.

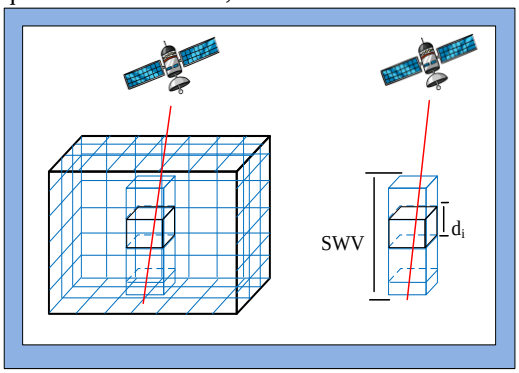
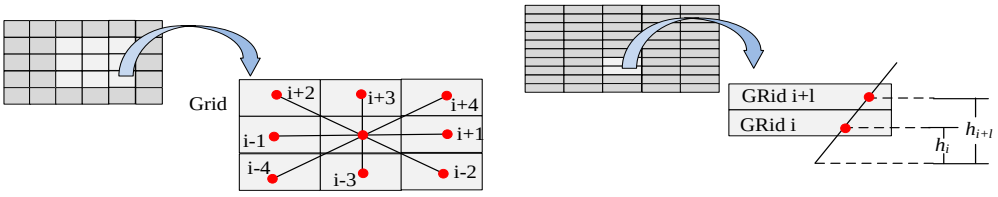


Fig. 4. Graphical representation of GNSS three-dimensional WVT observation equation.



(a)Horizontal constraint (b)Vertical constraint

Fig. 5. Geometric diagram of horizontal and vertical constraint equations.

Equation (8) is established by using distance dependent Gaussian weighting function.

$$w_1^i x_1 + w_2^i x_2 + \ldots + w_{i-1}^i x_{i-1} - x_i + w_{i+1}^i x_{i+1} + \ldots w_l^i x_l = 0 \qquad (8)$$

In equation (8), l indicates the total amount of grids in the same horizontal layer; x_l denotes the atmospheric WVD of the l th grid; w_l^i expresses the horizontal weight coefficient of the l th grid in the horizontal layer relative to the i th grid, and d is the distance between the center points of the corresponding two grids. The vertical constraint equation is shown in equation (9).

$$x_i = x_{i+1} \cdot e^{(h_{i+l} - h_i)/h} \qquad (9)$$

In equation (9), h is the height of the water vapor chromatography area. For top level constraints, research suggests that the top level region of the tomographic model has very little water vapor, so the WVD of all top level tomographic grids is directly constrained to be 0. Synthesizing three constraints and using the least square (LS) method can obtain the atmospheric WVD solution as shown in equation (10).

$$x = \left(A^T PA + B^T PB\right)^{-1} \cdot \left(A^T Py\right) \qquad (10)$$

In equation (10), B denotes the coefficient matrix of the constraint equation. For the 3D WVT method based on GA, it needs to primarily construct the tomography equation, and then the idea of optimization equation is used to solve it. The calculation method is shown in equation (11).

$$\min f(x) = (y - Ax)^T P(y - Ax), x \in R+ \qquad (11)$$

In equation (11), the x value that minimizes $f(x)$ is the result of water vapor chromatography. After the tomographic equation is obtained, a fitness function is constructed, and then some groups representing the approximate value of the grid WVD are randomly generated. Then, according to the fitness value of the grid WVD, the grid density estimation value of the subsequent generations is selected as the parent. After selection, a new grid WVD approximation solution is formed by using the group of parents mentioned above to calculate new offspring through crossover and mutation. It calculates the fitness value of the approximate value of the WVD of each group of grids. When the fitness value of a group meets the requirements or reaches the number of searches, the GA algorithm stops searching, or continues to iterate circularly. The specific process of GA based water vapor chromatography method is shown in Fig. 6.

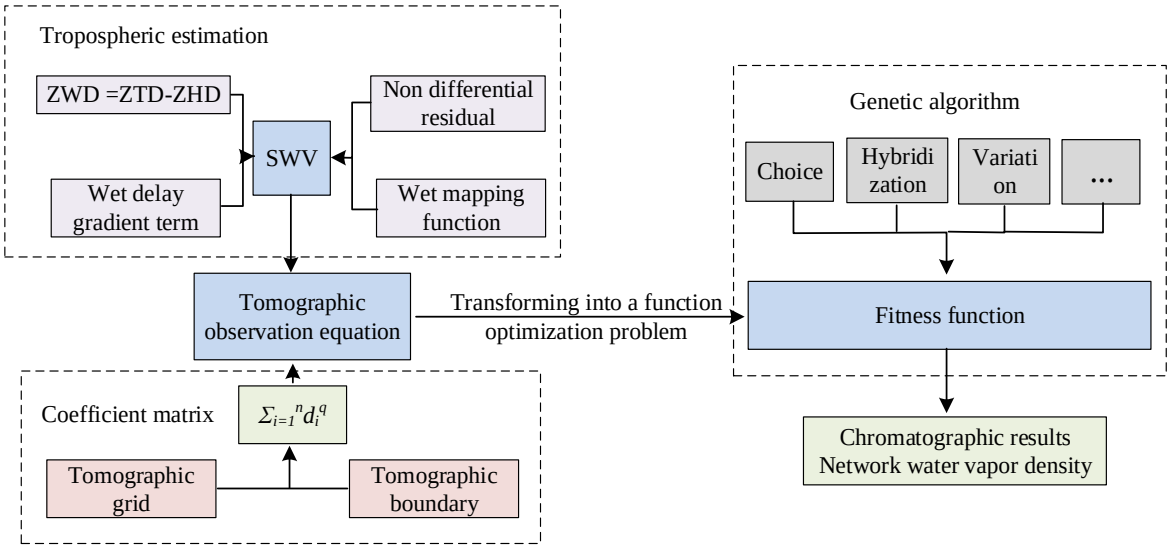


Fig. 6. Flow chart of water vapor chromatography method based on GA algorithm.

Based on the above operations, x that minimizes the value of the fitness function is obtained, that is, the best solution of equation (11), from which the distribution information of atmospheric WVD over the tomographic area is obtained.

IV. Results and Discussion

The purpose of the performance evaluation is to measure the good or bad results of the research work in order to further identify and optimize the performance differences between different models. In order to achieve a more accurate solution of water vapor distribution, research will be carried out on the estimation of atmospheric precipitable water volume (PWV) based on GNSS technology, and the idea of removal interpolation restoration will be introduced to optimize the accuracy of the calculation. In terms of three-dimensional water vapor chromatography, a chromatography method based on genetic algorithm is proposed to solve the problems in the chromatography equation. To test the detection performance of the water vapor spatial distribution detection method developed by the research institute, a series of experiments were designed to test the model improvement effect and detection accuracy. When GPT2w model was applied to work out Tm, the study used a decreasing rate to adjust and optimize it vertically. Here, the optimized model was named GPT2wh. To test the improvement effect of the model, the accuracy of GPT2wh was verified by using Tm calculated from sounding station data. The experiment selected 459 sounding stations containing meteorological profile information for more than half a year in 2020, and used the two improved models before and after to obtain the daily Tm values of the corresponding stations from 0 to 12 o'clock. Bias and RMSE were used as statistical variables for comparative analysis, as shown in Fig. 7.

In Fig. 7, for the GPT2wh model, the accuracy of most sounding stations was better than 5K, with a proportion of 88.45%; For the GPT2wh model, stations with an accuracy of 5K accounted for 76.45%. Compared to the GPT2w, the maximum improvement of RSME by the GPT2wh model was 7.35K, from 11.35K to 4.00K. The mean RMSE of the GPT2wh model was 3.83K, which was 0.33K less than that of the GPT2w, and the accuracy was improved by about 8%. The different colors in Fig. 7(a) and 7(b) could distinguish between different types of stations, with positive and negative Bias stations, respectively. This was mainly due to the height difference between the sounding station and its four model grid points, while the GPT2wh model included a Tm decay rate that could be vertically corrected for Tm, with significant reductions in positive and negative bias values. The mean deviation of the GPT2wh model was 0.32K and the mean deviation of the GPT2w model was 0.94K. Based on the contents of Fig. 7, the improved model gave a more accurate Tm value.

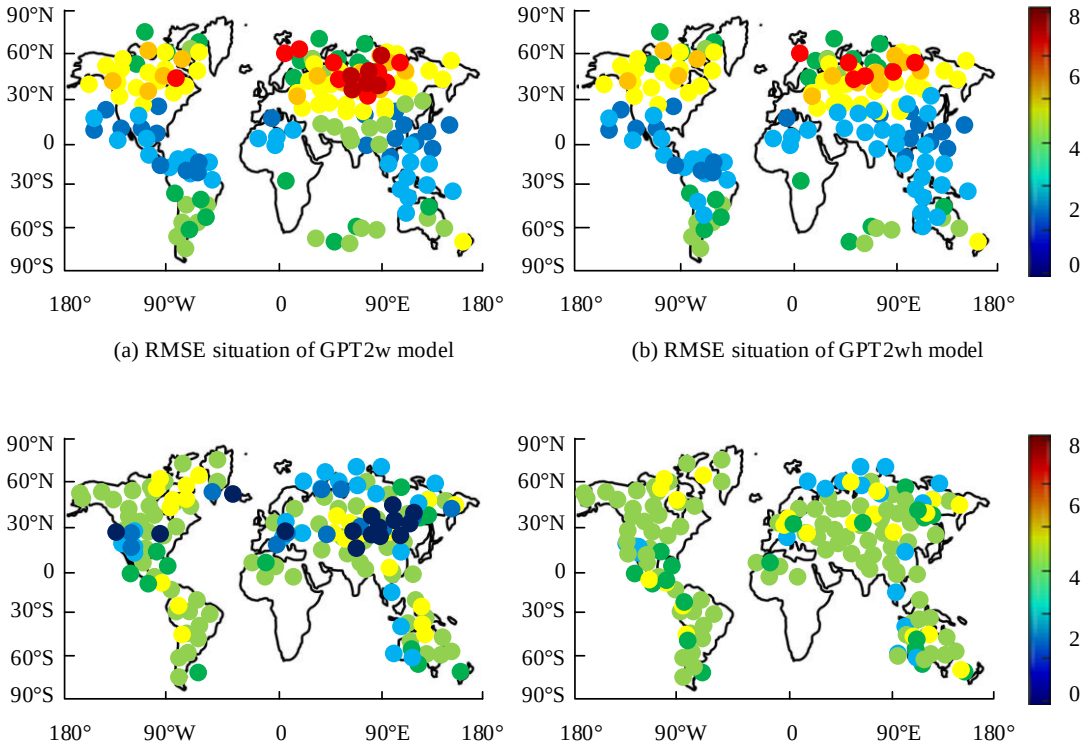
To further test the performance of the two models, statistical analysis was conducted on the height difference between each sounding station and its corresponding model nodes, and the trend curve of the two indicators changing with the height difference was plotted in Fig. 8.

In Fig. 8, the RMSE and Bias values of the GPT2w increased with the increase of height distinguish, while the GPT2wh could get stable RMSE and Bias in different height difference ranges. The average bias of the GPT2w model varies widely at different altitudes, from 0.32k to 10.34k; the average bias of the GPT2wh model is relatively stable and small. As the altitude difference increases, the average RMSE of the GPT2w model increases significantly, while the GPT2wh model can achieve smaller and smaller RMSE. It can be seen that the GPT2wh model has limited improvement in Tm when the height difference is small, while the improvement effect on Tm is more significant when the height difference is large. This indicated that the improved model could well promote the accuracy of height difference method station estimation Tm. And as the height difference increased, the promotion of the GPT2whbecame more significant.

To validate the effectiveness of the PWV spatial interpolation method proposed by the research institute, station cross validation and grid point PWV interpolation

using GNSS station PWV were used to compare the interpolation accuracy of the entire region. Firstly, different interpolation schemes were used for comparison: the first group of schemes did not take into account of the influence of elevation factors on PWV interpolation, and directly used IDW, Kriging, and TPS. The second group added the GPT2wh model on top of the first group, while the third group used the 3DRing and 3DTPS algorithms that considered elevation factors. Adding the GPT2wh model on top of the third group formed the fourth group of schemes. By comparing the PWV interpolation schemes mentioned above, the impact of elevation factors on the PWV interpolation results has been verified, and the effectiveness of using the removal interpolation restoration idea for PWV interpolation in different situations has been evaluated. The RMSE distribution of daily station cross validation for different schemes is shown in Fig. 9.

In Fig. 9, the first group of schemes showed the worst with a larger RMSE, while the third group considered the elevation factor and reduced the RMSE of the interpolation results. Based on the idea of removal interpolation recovery, after adding the GPT2wh model, the second group effectively reduced the RMSE value compared to the first group and the fourth group compared to the third group. Based on the content of Fig. 9, this idea could effectively improve the interpolation accuracy of PWV. In the results of Kriging and 3DKriging, the interpolation algorithm that adds elevation factors can improve the accuracy of PWV interpolation for most stations, but some stations do not show significant improvement, especially for stations located at 113.84° E and 22.26° N. Because when the HKNP station is the desired station, its elevation is no longer within the elevation range of the other 11 reference stations, and the elevation is relatively large. Even if the influence of elevation factors is taken into account in 3DDrilling, 11 reference stations cannot provide sufficient and reliable elevation reference information for the HKNP station, resulting in a small improvement in PWV interpolation accuracy. The second highest station is HKST. Although its elevation is significantly different from most of the survey stations, its elevation is included in the elevation range of 11 reference stations. Therefore, using Kriging for PWV interpolation accuracy crossover, while using 3DKriging can better improve the PWV accuracy of the HKST survey station. Through analysis, it can be concluded that the accuracy of 3DDrilling interpolation of PWV depends largely on the selection of reference stations, often requiring the elevation range of reference stations to cover as much as possible the elevations of all stations to be measured. The same problem exists for 3DTPS. The PWV interpolation method, which is based on the idea of removal interpolation restoration, can effectively solve the above problems. For the HKNP station, the RMSE of Kriging GPT2wh interpolated PWV is 1.58m, which is 4.45m better than the RMSE of Kriging; the RMSE of TPS-GPT2wh interpolated PWV is 1.43m, which is 2.97m better than the RMSE of TPS. Compared to PWV interpolation using 3DDrilling and 3DTPS, Kriging-GPT2wh and TPS-GPT2wh not only improve the accuracy of HKNP for special height stations, but also perform better for all other stations.



(a) RMSE situation of GPT2w model

(b) RMSE situation of GPT2wh model

(c) Bias situation of GPT2w model

(d) Bias situation of GPT2wh model

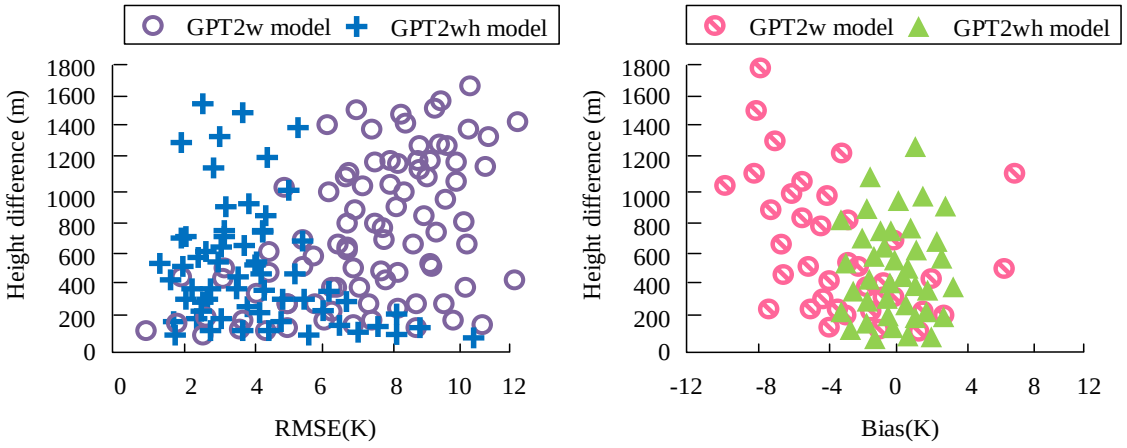
Fig. 7. Bias and RMSE scatter plots of the GPT2w model before and after improvement.

To further verify the accuracy of each interpolation method, the PWV of each grid point interpolated by ten methods was statistically compared with the PWV provided by ECMWF. The statistical information of grid points for different interpolation methods is listed in Table I.

In Table I, the grid point PWV verification results were similar to the station cross validation results. The accuracy of grid point verification results was better than that of station cross validation, regardless of MAE, RME, or CRE. This was because of the overall difference in PWV obtained from ECMWF grid data and GNSS data.

The LS method was the most commonly used solution method for 3D WVT, and a large amount of experiments demonstrated that this method could obtain high-precision atmospheric WVD. To assess the GA algorithm's accuracy based tomography method designed by the research institute, the results were compared with those of LS. In the experiment, linear regression and box chart were used to analyze the atmospheric water vapor density distribution of the two methods. The chart included experiments under all weather conditions, as shown in Fig. 10.

In Fig. 10 (a), the scatter distribution and regression line of atmospheric WVD showed a good linear relationship between the two chromatographic methods. The beginning point and slope of the regression equation were 0.534 and 0.9542, respectively. Fig. 10(b) showed the distribution of the difference in atmospheric WVD between the two chromatography methods, with Q1 and Q3 being -0.83g/m3 and 0.61g/m3, respectively, indicating that the proportion of the difference in atmospheric WVD calculated by the two chromatography methods within 1g/m3 exceeded 50%. The upper and lower limits of the box plot were 2.73g/m3 and -2.88g/m3, respectively. The outlier only accounted for 3.21%. From the content of Fig. 10, the atmospheric density outcomes obtained by the GA based tomography method were consistent with that obtained by the LS method, which proved that this method could achieve high precision solution of AWVSD. The RMSE and MAE obtained by comparing the genetic algorithm based tomography results with sounding data and ECMWF data were 1.45/1.24g/m3 and 1.35/1.01g/m3, respectively, while the RMSE/MAE obtained by comparing the least squares tomography results with sounding data and ECMWF data were 1.46/1.24g/m3 and 1.37/1.15g/m3, respectively. Overall, it can be seen that both the genetic algorithm based tomography and the least squares tomography methods can provide good atmospheric water vapor results compared to the reference values in both sunny and rainy experiments, and the statistical results of the former are better than those of the latter.



(a)RMSE changes with height difference (b)Bias changes with height difference

Fig. 8. Variation of RSME and bias for estimating tm values by different models with station altitude difference.

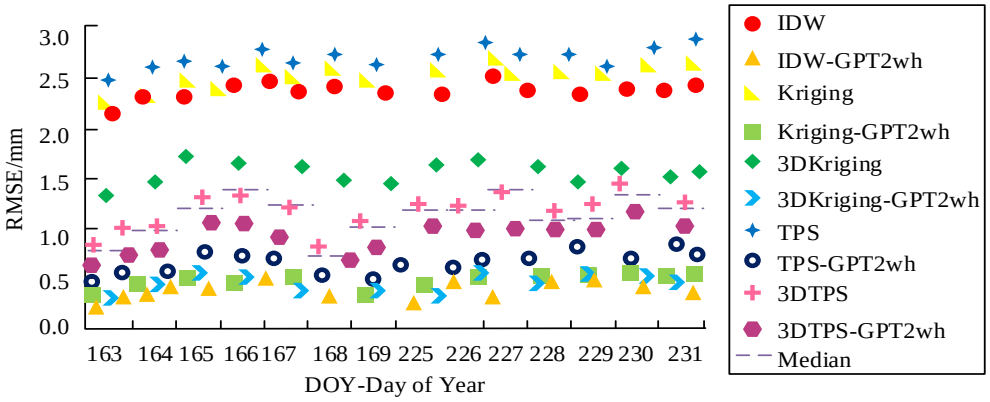
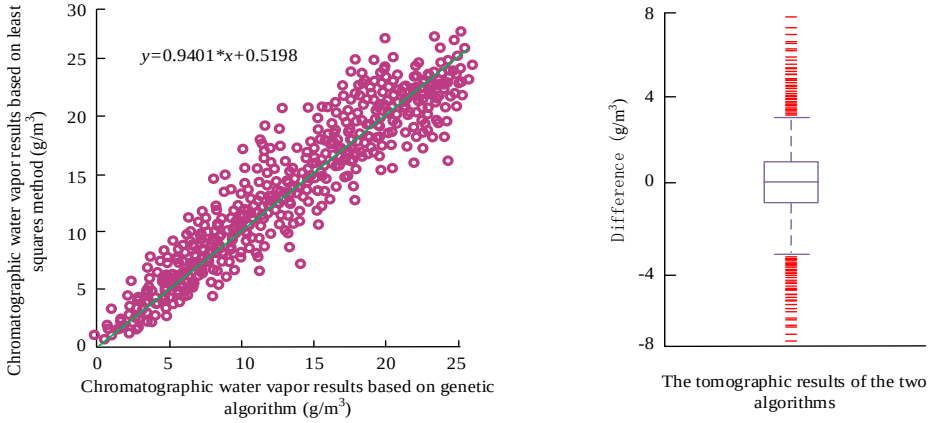


Fig. 9. RMSE distribution of various PWV interpolation methods in different year product days.

TABLE I. COMPARISON OF GRID POINT STATISTICAL INFORMATION OF DIFFERENT INTERPOLATION METHODS

Interpolation method	Sunny day(mm)			Rain(mm)		
	MAE	RMSE	CRE	MAE	RMSE	CRE
IDW	2.459	2.966	0.462	2.455	2.967	0.471
IDW-GPT2wh	1.640	2.045	0.193	1.643	2.045	0.196
Kriging	2.537	3.012	0.472	2.533	3.023	0.471
Kriging-GPT2wh	1.470	1.801	0.141	1.481	1.803	0.143
3DKriging	1.820	2.203	0.217	1.812	2.204	0.220
3DKriging-GPT2wh	1.587	1.896	0.160	1.592	1.892	0.162
TPS	2.873	3.324	0.692	2.882	3.312	0.693
TPS-GPT2wh	1.630	1.923	0.168	1.631	1.935	0.169
3DTPS	1.978	2.375	0.245	1.974	2.348	0.243
3DTPS-GPT2wh	1.699	2.004	0.182	1.698	2.003	0.181



(a)Linear regression diagram (b)Box diagram

Fig. 10. Linear regression and box plot of chromatographic results obtained from two chromatographic methods.

To further test the performance of the tomography method (Method 1) designed by the research institute, three-dimensional WVT experiments were conducted under different weather conditions using GNSS observation data offered by the Hong Kong Satellite Positioning Reference Station Network SatRef, and three-dimensional water vapor information was obtained under different weather conditions. Compare Method 1 with the currently popular centralized chromatography method, and the specific results were shown in Table II. The comparison methods included: GNSS WVT with consideration for boundary signals and vertical constraints (Method 2), a robust adaptive WVT (Method 3), WVT with fusion of ECMWF grid data (Method 4), and water vapor distribution tomography based on Kalman filtering (Method 5).

In Table II, regardless of whether it is sunny or rainy, the RMSE values of water vapor chromatography density for Method 1 were lower than those for the other four methods. During the entire chromatographic experiment, the average RMSE and MAE values of Method 1 were 1.78g/m3 and 1.41g/m3, respectively. According to the comprehensive table, Method 1 could obtain higher precision atmospheric WVD values. The meteorological profile data provided by radiosonde stations can be used to calculate water vapor density values, which are studied as reference values to evaluate the accuracy of water vapor tomography methods based on genetic algorithms. Due to the daily launch of radiosonde balloons at 0:00 and 12:00 UTC, the study chose to compare the water vapor chromatography results of the corresponding time periods under sunny and rainy conditions. During the rainy period from August 9, 2022 to August 15, 2022, the atmospheric water vapor density calculated from radiosonde data and the results of water vapor tomography based on genetic algorithms will vary with altitude. It is found that the atmospheric water vapor density decreases with increasing altitude; the atmospheric water vapor profile obtained by the genetic algorithm based tomography method is in agreement with the atmospheric water vapor profile obtained by radio sounding. In absolute terms, the agreement is better in the upper atmosphere. Considering the relative error, the corresponding values for tomographic grids with heights greater than 5km and less than 5km are 31% and 15%, respectively. This is because the water vapor density value in the upper atmosphere is relatively small, and small differences between sounding data and tomographic results can also lead to significant relative errors; the proportion of the atmospheric water vapor content within 5km of the Earth's surface exceeds 90%.

TABLE II. COMPARISON OF WATER VAPOR CHROMATOGRAPHY RESULTS OF VARIOUS METHODS

Project		Method 1	Method 2	Method 3	Method 4	Method 5
Sunny day	RMSE	1.84	2.01	1.98	2.22	2.31
	MAE	1.42	1.98	1.74	2.01	2.14
Rain	RMSE	1.71	2.03	1.96	2.19	2.27
	MAE	1.39	1.96	1.79	2.00	2.12
Average value	RMSE	1.78	2.02	1.97	2.21	2.29
	MAE	1.41	1.97	1.77	2.00	2.13

V. CONCLUSION

The amount of water content in clouds not only has a significant impact on the growth of cloud droplets and the formation and intensity of precipitation, but also serves as a predictive parameter for global climate data simulation and an important parameter for studying the impact of clouds on climate. To achieve a more accurate solution of AWVSD, research was conducted on estimating PWV based on GNSS technology, while introducing the idea of removal interpolation restoration to optimize the accuracy of the calculation. In three-dimensional water vapor chromatography, a chromatography method based on GA was proposed to address the problems in the chromatography equation. Through experimental analysis, the accuracy of most sounding stations in the GPT2wh model was better than 5K, with a proportion of 88.45%, which was more accurate than the Tm value obtained before improvement. The first group of schemes displayed the worst with a larger RMSE, while the third group considered the elevation factor and reduced the RMSE of the interpolation results. Based on the idea of removal interpolation recovery, after adding the GPT2wh model, the second group effectively reduced the RMSE value compared to the first group and the fourth group compared to the third group. This idea could effectively improve the interpolation accuracy of PWV. The average RMSE and MAE values for Method 1 were 1.78g/m3 and 1.41g/m3, respectively. The tomography method designed by the research institute could obtain higher precision atmospheric WVD values. The detection method for water vapor spatial distribution proposed by the research institute needs to be further improved and refined. Further research and work are needed after this proposal:

Step 1: Various empirical meteorological parameter models have been studied and constructed, but with the continuous enrichment and improvement of analytical data and spatial resolution of meteorological data, more dense meteorological grid parameter models need to be provided. In future research work, it is necessary to further consider how to reasonably sample global grid points, refine the positional relationship between the desired points and grid points, and improve the accuracy of spatial distribution research.

Step 2: A stable 3D water vapor chromatography algorithm and program have been developed, but the efficiency and automation level of the program still need to be improved. Subsequent research content needs to achieve automated batch processing of GNSS observation data, automated mapping, and comparative verification of sounding data.

Step 3: The study area selected by the Research Institute is the Hong Kong region, which has a large number of GNSS stations and a small geographical area. However, the research has not explored how to achieve water vapor spatial distribution detection in areas with large areas and sparse GNSS stations. In the future, regional water vapor spatial inversion can be achieved by exploring reasonable grid partitioning strategies, multi-source observation data fusion and other means.

ACKNOWLEDGMENT

The research is supported by: R&D center of building energy saving engineering technology of Shandong Huayu University of Technology, No.: 2019-03.

REFERENCES

[1] Guo R., Han D., Chen W., Dai L., Ji K., Xiong Q., Müller-Buschbaum P. Degradation mechanisms of perovskite solar cells under vacuum and one atmosphere of nitrogen. Nature Energy, 2021, 6(10): 977-986.

[2] Schumacher D L, Keune J, Dirmeyer P, Miralles D G. Drought self-propagation in drylands due to land–atmosphere feedbacks. Nature geoscience, 2022, 15(4): 262-268.

[3] Halldorsson V. National sport success and the emergent social atmosphere: The case of Iceland: International Review for the Sociology of Sport, 2021, 56(4):471-492.

[4] Xing G, Deng C, Di J, Zhu H, Yu C. High-temperature behaviour of V2AlC powders under nitrogen atmosphere. Ceramics International, 2022, 48(10): 14424-14431.

[5] Zheng N, Chai H, Chen L, Ma Y, Tian X. Snow depth retrieval by using robust estimation algorithm to perform multi-SNR and multi-system fusion in GNSS-IR. Advances in Space Research, 2023, 71(3): 1525-1542.

[6] Shi Y, Jiang Z, Liu Z, Li L. A Lagrangian analysis of water vapor sources and pathways for precipitation in East China in different stages of the East Asian summer monsoon. Journal of Climate, 2020, 33(3): 977-992.

[7] Huang L, Mo Z, Xie S, Liu L, Chen J, Kang C, Wang S. Spatiotemporal characteristics of GNSS-derived precipitable water vapor during heavy rainfall events in Guilin, China. Satellite Navigation, 2021, 2(1): 1-17.

[8] Lee J K, Han H S, Chaikasetsin S, Marron D P, Waymouth R M, Prinz F B, Zare R N. Condensing water vapor to droplets generates hydrogen peroxide. Proceedings of the National Academy of Sciences, 2020, 117(49): 30934-30941.

[9] Tan J, Yu D, Li Q, Tan X, Zhou W. Spatial relationship between land-use/land-cover change and land surface temperature in the Dongting Lake area, China. Scientific reports, 2020, 10(1): 1-9.

[10] Pan Y, Ren C, Liang Y, Zhang Z, Shi Y. Inversion of surface vegetation water content based on GNSS-IR and MODIS data fusion. Satellite Navigation, 2020, 1(1): 1-15.

[11] Zheng N, Chen P, Li Z. Accuracy analysis of ground-based GNSS-R sea level monitoring based on multi GNSS and multi SNR. Advances in Space Research, 2021, 68(4): 1789-1801.

[12] LÜ M Z, CHEN K J, CHAI H S, CENG J, ZHANG S, FANG L. Joint inversion of InSAR and high-rate GNSS displacement waveforms for the rupture process of the 2022 Qinghai Menyuan M6. 9 earthquake. Chinese Journal of Geophysics, 2022, 65(12): 4725-4738.

[13] Lewen Z, Jiaqian R, Yang D. Platform for GNSS real-lime space environment parameter inversion and its accuracy evaluation. Nanjing Xinxi Gongcheng Daxue Xuebao, 2021, 13(2): 204-210.

[14] Jalali Z, Noorzai E, Heidari S. Design and optimization of form and facade of an office building using the genetic algorithm. Science and Technology for the Built Environment, 2020, 26(2): 128-140.

[15] Garud K S, Jayaraj S, Lee M Y. A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models. International Journal of Energy Research, 2021, 45(1): 6-35.

[16] Shariati M, Mafipour M S, Mehrabi P, Ahmadi M, Wakil K, Trung N T, Toghroli A. Prediction of concrete strength in presence of furnace slag and fly ash using Hybrid ANN-GA (Artificial Neural Network-Genetic Algorithm). Smart Structures and Systems, An International Journal, 2020, 25(2): 183-195.

[17] Abualigah L, Alkhrabsheh M. Amended hybrid multi-verse optimizer with genetic algorithm for solving task scheduling problem in cloud computing. The Journal of Supercomputing, 2022, 78(1): 740-765.

[18] Yang Y, Song X. Research on face intelligent perception technology integrating deep learning under different illumination intensities. Journal of Computational and Cognitive Engineering, 2022, 1(1): 32-36.

[19] Guo Y, Mustafaoglu Z, Koundal D. Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms. Journal of Computational and Cognitive Engineering, 2023, 2(1): 5-9.

[20] Hidayat I, Ali M Z, Arshad A. Machine Learning-Based Intrusion Detection System: An Experimental Comparison. Journal of Computational and Cognitive Engineering, 2022, 2(2):88-97.