New Model for Vertical Distribution and Variation of Atmospheric Water Vapor – A Case Study for China

Kefei Zhang (profkzhang@cumt.edu.cn)
China University of Mining and Technology - Xuzhou Campus: China University of Mining and Technology https://orcid.org/0000-0001-9376-1148

Moufeng Wan
China University of Mining and Technology - Xuzhou Campus: China University of Mining and Technology

Suqin Wu
China University of Mining and Technology - Xuzhou Campus: China University of Mining and Technology

Zhen Shen
China University of Mining and Technology - Xuzhou Campus: China University of Mining and Technology

Dantong Zhu
China University of Mining and Technology - Xuzhou Campus: China University of Mining and Technology

Peng Sun
China University of Mining and Technology - Xuzhou Campus: China University of Mining and Technology

Longjiang Li
China University of Mining and Technology - Xuzhou Campus: China University of Mining and Technology

---

Research Article

Keywords: water vapor density, water vapor spatio-temporal model, water vapor vertical distribution

Posted Date: April 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1497870/v1

New Model for Vertical Distribution and Variation of Atmospheric Water Vapor – A Case Study for China

Moufeng Wan,[a,b] Kefei Zhang,[a,b] Suqin Wu,[a,b] Zhen Shen,[a,b] Dantong Zhu,[a,b] Peng Sun,[a,b] Longjiang Li,[a,b]

[a] *Jiangsu Key Laboratory of Resources and Environmental Information Engineering, China University of Mining and Technology, Xuzhou 221116, China*

[b] *School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China*

Corresponding author: Kefei Zhang, profkzhang@cumt.edu.cn

Abstract

For better modeling the variations in the vertical distribution of water vapor, in this study, a new function for the vertical variation in water vapor was derived, named $lapse_{RPWV}$. From the analyses of $lapse_{RPWV}$ time-series, it was found that its vertical distribution is strongly correlated with the relative magnitude of total precipitable water vapor (*TPWV*). This study proposed a method that used six data ranges of *TPWV* to determine the relative magnitude of *TPWV*. For the periodic variations in the classified $lapse_{RPWV}$ time-series in each of six *TPWV* ranges, a spatio–temporal $lapse_{RPWV}$ model was developed for each range. The new models were validated by comparing their predictions against the references from sounding data at 12 radiosonde stations in China, and their performances were also compared with that of the commonly used water vapor scale height (*H*) model. Results showed that, first, the number of stations that had reduced annual RMSE of *H* values in *TPWV* ranges from 1 to 6 accounted for 92%, 92%, 67%, 83%, 100%, and 100% of the total stations, respectively. Second, the proportions of the height range that had reduced annual RMSE of water vapor density (*WVD*) in all height ranges within all *TPWV* ranges were above 75% at the 12 stations. Last, considering all *TPWV* ranges as a whole, in each of 10 height ranges, the annual RMSEs of *WVD* of all the stations reduced at least 11%, 20%, 43%, 48%, 40%, 38%, 32%, 35%, 32%, and 28%, respectively.

Keywords: water vapor density; water vapor spatio-temporal model; water vapor vertical distribution

1 Introduction

Water vapor (WV) is the main greenhouse gas and also an important part of the earth's atmosphere (Chahine 1992). From global climate to local meteorology, it has a strong influence on climate at various spatio-temporal scales (Bevis et al. 1992; Zhao et al. 2012; Liu et al. 2013). WV mainly concentrates in the lower atmosphere and accounts for about 99% of the total WV content in the troposphere, with about half the total WV content in the altitude below 2 km (Viswanadham 1981). Although WV content in the atmosphere has a small portion (about 0 – 4%), it is the most active and variable component in the atmosphere, and it is also one of the meteorological parameters that are most difficult to be characterized (Rocken et al. 1997). WV concentration in different heights within the lower atmosphere over a site may vary by order of magnitude (Jacob 2001). The variation in the spatial distribution of WV, especially in the vertical distribution (Bevis et al. 1992), plays a vital role in the vertical stability of the atmosphere and the structural evolution of an atmospheric storm system (Jacob 2001).

According to the state equation of WV, the atmospheric water vapor density ρ_w (WVD, unit: g/m^3) at each layer of the atmosphere can be obtained by a function of the WV partial pressure e (unit: hPa) and temperature T (unit: K) of the same layer, expressed as (Reber and Swope 1972; Tomasi 1981)

$$\rho_w = \frac{e}{RT} \tag{1}$$

where R is the gas constant of water vapor ratio, and R = 0.4615 (unit: $J/(K \cdot g)$).

In the troposphere, generally speaking, the WV concentration is the largest at the ground surface level, there exists a correlation relationship of power-law between the atmospheric *WVD*s near the ground surface and the upper troposphere along the vertical direction over the same site. The relationship of power-law correlation can be expressed by the following exponential function, in which the trend of the decrease in *WVD* with the increase in altitude is assumed to be a uniform lapse rate (Reitan 1963)

$$\rho_{w_h} = \rho_{w_s} \exp(-\beta(h - h_s)) \tag{2}$$

where ρ_{w_s} and ρ_{w_h} (unit: g/m^3) are the WVDs at the ground surface height (h_s) and the height h exceeds h_s (unit: km), respectively, along the vertical direction in the troposphere over the site; β (unit: km^{-1}) is a constant for the height range from h_s to h. If β is known, the *WVD* at height *h* can be obtained from ρ_{w_s} and equation (2).

The common precipitable water vapor (PWV, whose unit was in millimeter in this study) is the depth to which liquid water would stand if all the WV in a vertical column of air of unit cross-sectional area was condensed. Let $PWV_{h_1}^{h_2}$ denote the partial WV in the height range from h_1 to h_2 along the vertical direction over a site, it can be calculated by the integral of *WVD* in the range

$$PWV_{h_1}^{h_2} = \frac{1}{\rho_v} \int_{h_1}^{h_2} \rho_w dh \tag{3}$$

where ρ_v is the water density ($\rho_v = 1$ g/cm^3); dh is the increment step (unit: km, the same as that of h). It is worth mentioning that if h_1 is at the surface and h_2 approaches the top of the troposphere, then the PWV in equation (3) is the common total PWV ($TPWV$) of the troposphere along the vertical direction.

Substituting equation (2) into (3) and let $h_1 = h_s$, the following can be derived (Reitan 1963)

$$PWV_{h_s}^{h} = \frac{\rho_{w_s}}{\rho_v}\left(\frac{1 - \exp(-\beta(h - h_s))}{\beta}\right) \quad (4)$$

When h in equation (4) approaches the tropopause, the WVD at this altitude is close to zero, and β can be approximated by the inverse of H, where H is the so-called atmospheric water vapor scale height (unit: km), i.e. the equivalent height under the assumption that atmospheric WV is uniformly distributed in the entire vertical range of the troposphere, and it has a physical interpretation of the depth through which the WVD reduces to 1/e of its value at the base of the troposphere (Byers 1957; Tomasi 1977). It is an important parameter in terms of its control on the radiative balance and convective stability of the atmosphere (Weaver and Ramanathan 1995). Equation (4) can be then simplified to the following formula (Tomasi 1977, 1981) (note: due to $\rho_v = 1$ g/cm^3, it is not showed hereafter for simplicity)

$$TPWV = \rho_{w_s} H \quad (5)$$

where $TPWV$ stands for total PWV of the site. The use of $TPWV$ here is for distinguishing it from a partial PWV within the troposphere expressed by equation (3).

Since H can be utilized to obtain $TPWV$ by multiplying with the surface WVD of the same site, some scholars studied the relationship between H and $TPWV$ in land and ocean regions(Tomasi 1984; Bobak and Ruf 1996; Otárola et al. 2010). One of the possible ways to obtain H over a site is to use the ratio of WVDs measured at two sites that are closed to each other in the horizontal dimension but have a significant difference in altitude (Reber and Swope 1972; Ruf and Beus 1997). Whereas, the dynamic nature of the atmosphere means that WV is very active in the troposphere, especially in the lower layers, and its amount varies with time and height. Simply taking a constant value for H or a periodic function that only contains the time variable to model the temporal variation in H over a site, is not reasonable, because the H value at the same site varies not only with time but also with height (Byers 1957; Reitan 1963; John et al. 2005; Kennett and Toumi 2005; Otarola et al. 2011; Zhang et al. 2015). Borger et al (Borger et al. 2020)developed an empirical parameterization for H and obtained a substantial improvement using the parameterization compared to the use of a prescribed constant WV profile.

In the relationship between sea surface temperature and column WV over tropical and subtropical oceans, H is taken as an index of vertical moisture gradient between the boundary layer and the free troposphere(Kanemaru and Masunaga 2013). Given the global temperature dependence of H, Kennett and Toumi (Kennett and Toumi 2005) examined the variation of H within the column to reflect changes in atmospheric moisture lifetime. As well, in the construction of tropospheric models, an exponential decrease function containing H, which itself is modeled as a seasonal function, is used for estimated ZWD (zenith wet delay) of GNSS (Global Navigation Satellite System) signal passing through troposphere (Ruffini et al. 1999; Schüler 2014). Moreover, to obtain a unique and stable WV estimate, in the process of constructing the tomographic models based on WV retrieved from GNSS measurements, the H value is often used as a vertical constraint which is an exponential function of WVD or wet refractive index in the estimation system of the tomographic models (Flores et al. 2000, 2001; Guo et al. 2016). In most applications, a constant H value selected from the range 1–3 km (based on the statistical distribution of

H) is typically used in the exponential function (Elósegui et al. 1998; Perler et al. 2011; Ding et al. 2018).

The vertical distribution of WV also correlates with the water vapor state caused by some meteorological factors e.g. temperature and WV pressure of the site (Jacob 2001). For accurately modeling the temporal variation trend in the vertical direction under different water vapor states in the troposphere, in this study, a function for the vertical variation in WV in the troposphere was derived based on the ratios of the lapse rate of atmospheric partial WV at any heights to the $TPWV$ of the site, named $lapse_{RPWV}$. Based on our analyses of the spatial and temporal characteristics of $lapse_{RPWV}$ time-series, it was found that the vertical distribution of $lapse_{RPWV}$ strongly correlates not only with the relative magnitude of its corresponding $TPWV$, but also the temporal periodicity. A method is studied to classify $TPWV$ into different data ranges for determining the relative magnitude of $TPWV$, and a new temporal model is developed for the fitting of the vertical $lapse_{RPWV}$ distribution according to the periodic variations in the classified time series of $lapse_{RPWV}$.

This paper is organized as follows. Section 2 introduces the methodology for the derivation of a function of the ratio of the lapse rate of partial WV at any height range to the $TPWV$ over a site; Section 3 describes data selection and division for quantification and unification of the correlation between the vertical distribution of $lapse_{RPWV}$ and $TPWV$, and the construction of a temporal $lapse_{RPWV}$ model. Section 4 presents test results, and Section 5 gives concluding remarks.

2 Methodology

2.1 Derivation of formula for the vertical distribution of water vapor

In practical applications, it is common that equation (2) is replaced by $\rho_{w_h} = \rho_{w_s}\exp(-(h-h_s)/H)$, and equation (5) is used to obtain the empirical value of H using measured $TPWV$ and surface WVD. However, this is not reasonable because equation (5) is derived from equation (4) under the condition that the height is close to the tropopause (rather than any other height). This implies that β in equation (2) can be replaced by $1/H$ only in the case that the height is close to the tropopause, rather than any other height below the tropopause.

WV is very active in the lower troposphere and its vertical distribution may not follow an exponential decrease trend all the time, e.g., WVD in the upper troposphere can be even greater than the lower troposphere sometimes. Therefore, it is necessary to analyze $WVDs$ at different heights in the troposphere, instead of directly using equation (2) or H. For obtaining a more accurate functional relationship between WVDs at two heights in the troposphere, a function of the ratio of the lapse rate of partial WV at any height range to the $TPWV$ was introduced in this study (termed $lapse_{RPWV}$), which reflects the variation of WV along the vertical direction, i.e., the vertical distribution of WV. The derivation of its formula is as follows.

In the troposphere, the WV content generally decreases with altitude, and the decline rate of the WV content with altitude is the so-called lapse rate, which by definition is the negative of the change in WV content with altitude. Using $PWV_{h_i}^{h_{trop}}$ to represent the WV content from any altitude h_i to the tropopause, the variation range of the $PWV_{h_i}^{h_{trop}}$ from the ground surface to the tropopause is from $TPWV$ to zero. Therefore, the lapse rate of PWV at h_i, denoted by ${lapse_{PWV}}_i$, can be expressed by

$$lapse_{PWV\,i} = \frac{PWV_{h_i}^{h_{trop}} - TPWV}{h_i - h_s} \tag{6}$$

where $lapse_{PWV\,i}$ is in a unit of mm/km; h_{trop} is the height of the tropopause.

Let $RPWV_i$ (unit: %) be the ratio of $PWV_{h_i}^{h_{trop}}$ to $TPWV$

$$RPWV_i = \frac{PWV_{h_i}^{h_{trop}}}{TPWV} \tag{7}$$

Extend equations (6) and (7) to any two heights h_i and h_j ($h_j > h_i$) in the troposphere, the following can be obtained

$$lapse_{PWV\,ij} = \frac{PWV_{h_i}^{h_j}}{h_i - h_j} \tag{8}$$

$$RPWV_{ij} = \frac{PWV_{h_i}^{h_j}}{TPWV} \tag{9}$$

From equations (8) and (9), the following can be derived

$$\frac{lapse_{PWV\,ij}}{TPWV} = \frac{RPWV_{ij}}{h_i - h_j} \tag{10}$$

Note that the sign of $h_i - h_j$ in the above equations is already negative for a descent trend, and the absolute value of equation (8) is the mean WVD (i.e., $\rho_{w_{ij}}$) in the height range between h_i and h_j.

Let $lapse_{RPWV_{ij}}$ (unit: 1/km) denote the left-hand term of equation (10), then

$$lapse_{RPWV_{ij}} = \frac{RPWV_{ij}}{\Delta h_{ij}} \tag{11}$$

$$lapse_{RPWV_{ij}} = -\frac{\rho_{w_{ij}}}{TPWV} \tag{12}$$

where $lapse_{RPWV_{ij}}$ is the ratio of PWV_{hi}^{hj} in a unit height (or thickness) between h_i and h_j to $TPWV$.

When h_i and h_j are very close (i.e. a very thin layer), WVD_i and WVD_j are very close as well. Replace the layer in equation (12) with its corresponding mid-height $h = \frac{h_i + h_j}{2}$, then equation (12) can be expressed as

$$lapse_{RPWV_h} = -\frac{\rho_{w_h}}{TPWV} \tag{13}$$

Let h_l and h_k be any two heights over the site within the troposphere, the following relationship can be derived

$$\rho_{w_l} = \rho_{w_k} \cdot \frac{lapse_{RPWV_l}}{lapse_{RPWV_k}} \tag{14}$$

where ρ_{w_l}, ρ_{w_k}, $lapse_{RPWV_l}$ and $lapse_{RPWV_k}$ are the *WVDs* and the rate of the variation in *PWV* at h_l and h_k, respectively.

In addition, it is worth mentioning that according to equations (5) and (13) the water vapor scale height *H* can be obtained from $lapse_{RPWV_h}$ at the surface height, i.e., $lapse_{RPWV_s}$, as below

$$H = -\frac{1}{lapse_{RPWV_s}} \tag{15}$$

2.2 Temporal model

Considering the characteristics of the annual and semi-annual variation of $lapse_{RPWV}$ time-series over a site, the trigonometric periodic function below was used to fit $lapse_{RPWV}$ for the site in this study

$$f(doy) = a_0 + \sum_{i=1}^{n} (a_i\, cos(i \cdot doy \cdot w) + b_i\, sin(i \cdot doy \cdot w)) \tag{16}$$

where i is the order of the trigonometric periodic function; a_0, a_i and b_i are the coefficients to be solved; *doy* is the day of year; *w* is the angular frequency.

3 Spatio–temporal characteristic of $lapse_{RPWV}$ and modeling

3.1 Data and data processing

Sounding data from 12 radiosonde stations located in the longitude range 100 ºE−125 ºE and latitude range 20 ºN−45 ºN in China over the 12-year from 2008 to 2019 were downloaded from the Integrated Global Radiosonde Archive (IGRA) (at https://www.ncdc.noaa.gov/data-access/weather-balloon/integrated-global-radiosonde-archive). The distribution of the 12 stations, which are in three climate zones, respectively, is shown in Fig. 1. The reasons for the selection of these stations are that they had at least 10 continuous sounding layers containing the 10 standard pressure levels (1000, 925, 850, 700, 500, 400, 300, 250, 200). The temporal resolution of the sounding data was 12 hours (observed at 00:00 and 12:00 UTC). The vertical sounding profiles contained various meteorological measurements including pressure, temperature, WV pressure, relative humidity, etc. at each sounding layer. The following procedures were carried out for the sounding data from each of the stations. First, according to equations (1), (3), and (5), the values of *WVD* at each of the sounding height layers, and *TPWV* and *H* were calculated, and then according to equations (7) and (11), $lapse_{RPWV}$ at the mid-height of two adjacent heights $h_l = (h_{i-1} + h_i)/2$ (i is the index of the sounding height layer (h_{layer})) was calculated. After this step was performed for all the sounding height layers across 12 -year period and for all the 12 stations on a *doy*-by-*doy* basis, the dataset containing doy, h_{doy}, $\rho_{w_{doy}}$, $TPWV_{doy}$, H_{doy}, $h_{l_{doy}}$ and $lapse_{RPWV_{doy}}$ was obtained for the investigation in the next sections.

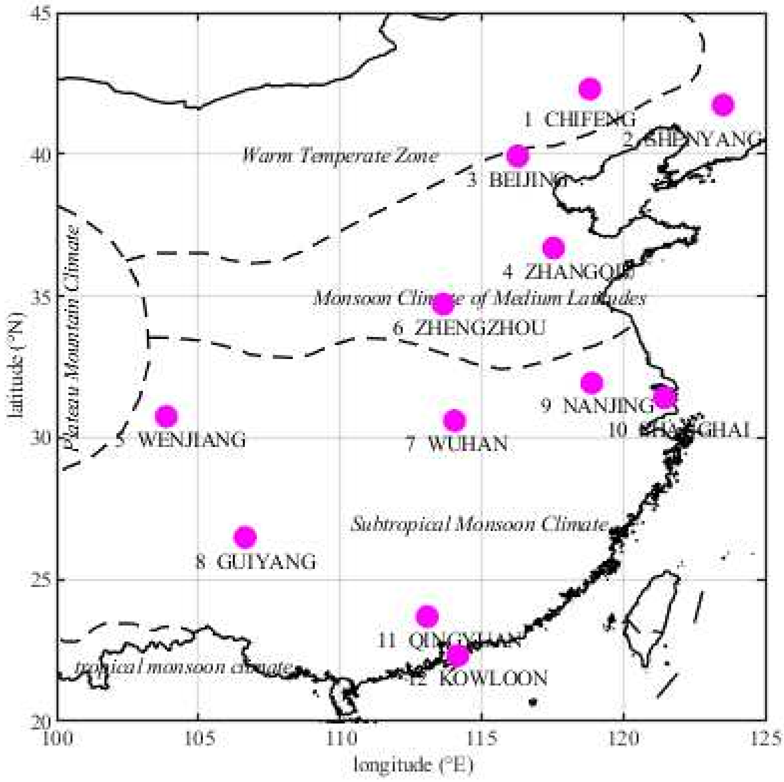


Fig. 1 Distribution of the 12 IGRA stations in China selected for this research (the dashed lines are division for different climate zones).

3.2 Spatio-temporal variation of $lapse_{RPWV}$

The spatial-temporal variation in $lapse_{RPWV_{doy}}$ over the 10-year from 2008 to 2017 over each of the 12 stations are shown in Fig. 2, where the doy, $h_{l_{doy}}$ and $lapse_{RPWV_{doy}}$ values were obtained from in the previous section, and the color bar indicates the value of $TPWV_{doy}$. We can see that the characteristics of the distributions of $lapse_{RPWV}$ over all the stations are very similar in both time and spatial domains, while the fluctuations in $lapse_{RPWV}$ over these stations in different climate zones, are different.

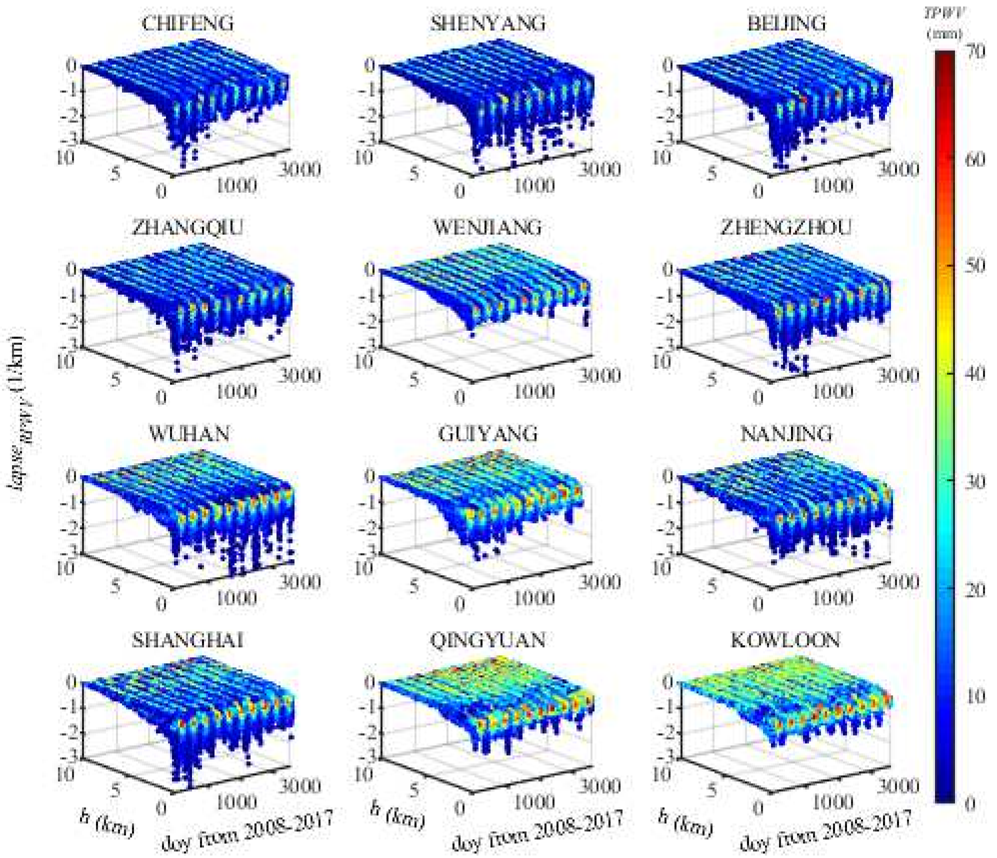


Fig. 2 Vertical distribution of $lapse_{RPWV_{doy}}$ in the 10-year from 2008 to 2017 over each of the 12 stations.

More specifically, at the CHIFENG (in a warm zone), SHENYANG, BEIJING, and ZHANGQIU stations (in monsoon climate of mid-latitudes), the vertical distribution of their $lapse_{RPWV_{doy}}$ at the same time and the same altitude are relatively similar, the fluctuation amplitudes of the vertical distribution of $lapse_{RPWV_{doy}}$ are relatively large than the other stations. The same results are from the WENJIANG, GUIYANG, QINGYUAN and KOWLOON stations (in subtropical monsoon climate), but their fluctuation amplitudes are relatively small compared to the previous 4 sites. While at the ZHENGZHOU (in monsoon climate of mid-latitudes), WUHAN, NANJING, and SHANGHAI stations (in subtropical monsoon climate), their $lapse_{RPWV_{doy}}$ are similar because they are at the junction of monsoon climate of medium latitudes and subtropical monsoon climate, and the fluctuation amplitudes of $lapse_{RPWV_{doy}}$ in a relatively uniform range – between the amplitudes of the above mentioned two regions. Overall, the $lapse_{RPWV}$ at the stations that are in the same climate zone have similarity.

At each station, the fluctuation amplitudes of the $lapse_{RPWV_{doy}}$ time-series were strong in winter but weak in summer, and its $TPWV_{doy}$ in winter was smaller than the other seasons, which implies notable temporal periodic variation in $lapse_{RPWV_{doy}}$, the same as $TPWV_{doy}$. In the vertical domain, the fluctuation amplitudes of $lapse_{RPWV_{doy}}$ in the lower troposphere, especially at 1−3 km altitudes, were larger than the upper troposphere at each station; and the $lapse_{RPWV_{doy}}$ increased with the increase in height and tended to be stable at the highest altitude, with a value close to 0.

3.2.1 Relationship between $lapse_{RPWV}$ and $TPWV$

To further study the relationship between $lapse_{RPWV}$ and *TPWV* over each of the 12 sites during the 10-year 2008 – 2017, the $lapse_{RPWV_{doy}}$ time-series of each site were sorted by their corresponding $TPWV_{doy}$ values in the same *doy*, and results are shown in Fig. 3, where the x-axis represents the value of *TPWV*, the y-axis represents the value of height and the z-axis represents the value of $lapse_{RPWV}$. Subfigure (a) shows the characteristics of the relationship between $lapse_{RPWV}$ and *TPWV* at six selected stations and subfigure (b) shows the results of the KOWLOON station in different-*doy*s in the four seasons, as an example of the 12 stations. Note that the selected CHIFENG and ZHANGQIU stations represent the stations located in the warm temperate zone and monsoon climate of medium latitudes, respectively; WENJIANG and KOWLOON represent the stations located in subtropical monsoon climate, and ZHENGZHOU and NANJING represent the stations at the junction of monsoon climate of medium latitudes and subtropical monsoon climate.

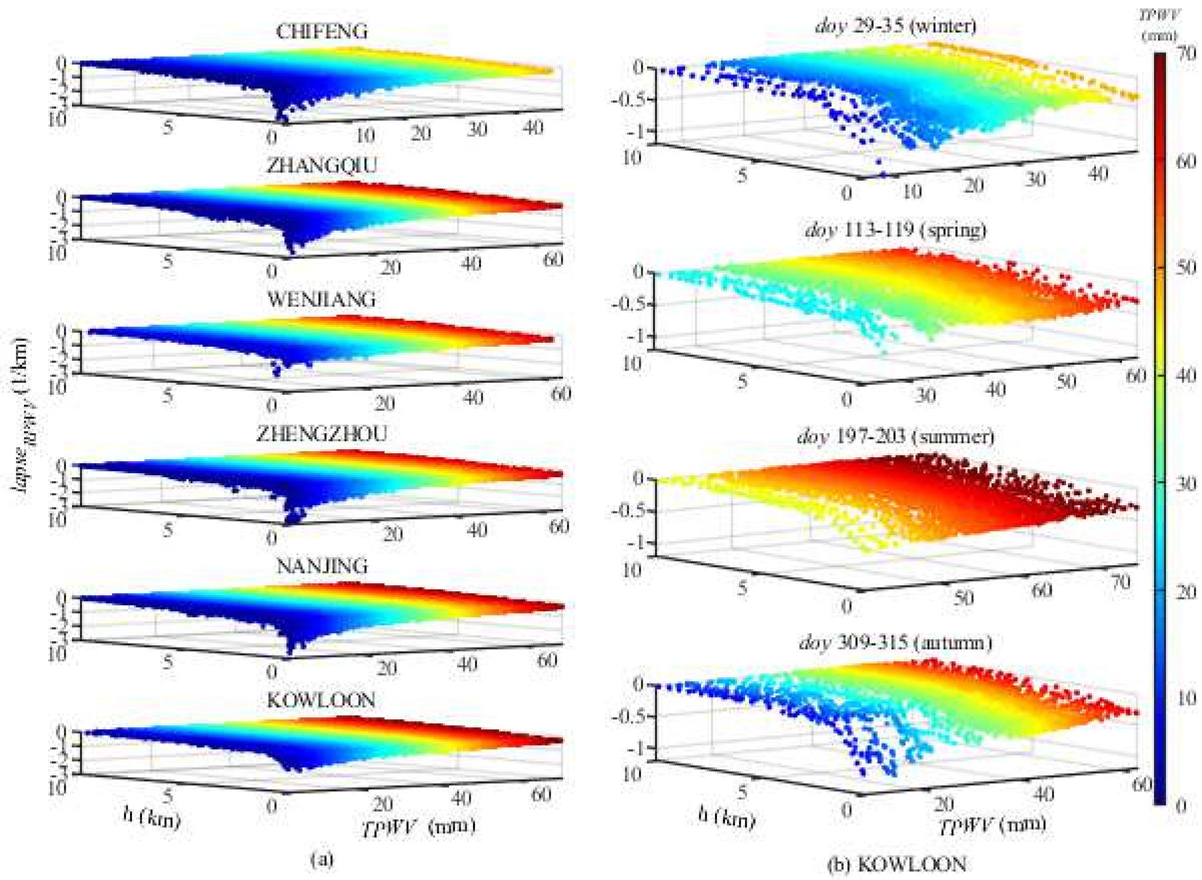


Fig. 3 Distribution of 10-year $lapse_{RPWV_{doy}}$ sorted by their corresponding $TPWV_{doy}$ values at each of the selected six stations (a) and results over KOWLOON in different *doy*s in the four seasons (b).

It can be seen from Fig. 3 (a) as a whole that the $TPWV_{doy}$ values over different stations show different fluctuation amplitudes of its corresponding vertical distribution of $lapse_{RPWV_{doy}}$. As the $TPWV_{doy}$ value gradually increases from small to large, the fluctuation amplitude of its corresponding vertical distribution of $lapse_{RPWV_{doy}}$ decreases and tends to be stable when $TPWV_{doy}$ closes to its maximum, and vice versa. For example, in each subfigure of Fig. 3 (a), the fluctuation amplitudes of the vertical distribution of $lapse_{RPWV_{doy}}$ in the *TPWV* range 0–20 mm, especially in small $TPWV_{doy}$ value, are larger than that in the *TPWV* range on the far right of the x-axis. Fig. 3 (b), which is the result at a *doy* time scale, shows that, although the $TPWV_{doy}$ values in different *doy*s in the four seasons fall into different ranges, the vertical distribution of $lapse_{RPWV_{doy}}$ in each of the four subfigures (or season) still tends to be stable with the increase of its corresponding $TPWV_{doy}$. This trend is consistent with the results of Fig. 3 (a) but on a different time scale. The results indicate that within a certain period the vertical distribution of $lapse_{RPWV}$ strongly correlates with the magnitude of its corresponding *TPWV* value relative to the *TPWV* values of the period, i.e., the vertical distribution of $lapse_{RPWV}$ is independent of its corresponding *TPWV* value itself. For example, in a selected period at the same site, there are 10 groups of data including 10 *TPWV* values and their corresponding 10 vertical distributions of $lapse_{RPWV}$. If the 10 *TPWV* values are sorted by their magnitude, and their corresponding vertical distributions of $lapse_{RPWV}$ are also sorted by the magnitude of the *TPWV* values, then the newly sorted 10 groups of *TPWV* values can be used to determine their corresponding vertical distributions of $lapse_{RPWV}$. The magnitude of one of the 10 *TPWV* values relative to the other nine *TPWV* values is often called the relative magnitude of this *TPWV*, which is denoted by *Rel-TPWV* hereafter in this paper, merely for convenience.

Comparing the four subfigures in Fig. 3 (b), one can also find that in different periods (or seasons), even though the *Rel-TPWV* values may be the same, i.e., under the same *Rel-TPWV* condition, but the fluctuation amplitudes of their corresponding vertical distributions of $lapse_{RPWV}$ are different and vary with time (those unlisted *doy*s had the same characteristic). For example, in the same position range on the x-axis from the top to the bottom subfigures, selecting the *TPWV* ranges of 10–20, 30–40, 40–50, and 10–20 mm, respectively, the fluctuation amplitudes in the third subfigure (*doy* 197–203, summer) are smaller than that of the other three subfigures. The same characteristic was also found at the other stations (which are not shown in this figure).

3.2.2 Criterion of partition for time, height, and *TPWV* intervals

The following conclusions can be drawn from the above $lapse_{RPWV}$ results over all stations. First, along the vertical direction, the $lapse_{RPWV}$ value increases with the increase in height and tends to become 0 from a negative value. Second, in the same time session, the vertical distribution of $lapse_{RPWV}$ is related to the *Rel-TPWV* of its corresponding *TPWV*. Last, under the same *Rel-TPWV* conditions, the fluctuation amplitude of the vertical distribution of $lapse_{RPWV}$ varies with time.

To analyze the temporal and spatial characteristics of $lapse_{RPWV}$, the aforementioned 10-year $lapse_{RPWV}$ time-series over each of the 12 stations selected were divided into several sections both in temporal and vertical domains. In the temporal domain, the time series were partitioned by the time scale of *doy*. In the vertical domain, considering both the density of the sounding data of the station and the characteristics of the fluctuation amplitude of $lapse_{RPWV}$ along the vertical direction – the fluctuation amplitude below 6 km is much larger than that above 6 km, and at below 3 km, the fluctuation amplitude is the largest. Hence, two different height intervals – a 0.5 km interval for below 6 km and a 1 km interval for above 6 km were adopted. More specifically, the vertical profile of the $lapse_{RPWV}$ time-series was portioned into several height ranges (denoted by h_r): $h_r =$

$[h_s,\ h_s+0.5\ km,\ \cdots, h_{layer}+1\ km, \cdots]$, $h_{layer} \geq 6\ km$, where h_s is the height of the station.

For further analyzing the relationship between the vertical distribution of $lapse_{RPWV}$ and its corresponding *Rel-TPWV* value of a station in a long period, the *Rel-TPWV* of *TPWV* of the site needs to be determined using the following procedure proposed in this study.

Step 1, according to the periodic characteristics of *TPWV* time series (Liu et al. 2015), a periodic function fitting the sample (i.e. measurements) of the 10-year $TPWV_{doy}$ time series was obtained. As an example, the magenta fitting curve is shown in Fig. 4 (a) for the KOWLOON station. Then the fitting function was applied to predict the *TPWV* value for each *doy*, which is denoted by $TPWV_{f_{doy}}$, and the discrepancies (residuals) between the $TPWV_{doy}$ and $TPWV_{f_{doy}}$ were calculated.

Step 2, the set of $TPWV_{doy}$ residuals for each *doy* in the duration of the 10 years were used to calculate the standard deviation of $TPWV_{doy}$ for the *doy*, then according to the periodic character of the time series of $TPWV_{doy}$ standard deviations from *doy* 1 to 366, a fitting periodic model was obtained and applied to predict the $TPWV_{doy}$ standard deviation for each *doy*, denoted by $\sigma_{f_{doy}}$. Fig. 4 (b) shows the time series of the $TPWV_{doy}$ standard deviations at each of the selected six stations, together with their fitting function (the magenta fitting curve).

It is worth mentioning that, to avoid the influence of *TPWV* in extreme weather conditions on the standard deviation of $TPWV_{doy}$, based on the statistical theory that the probability of a value being within a ± 2 standard deviations range is greater than or equal to 0.95 (percentile), any $TPWV_{doy}$ that had the absolute residual value greater than 2 standard deviations was regarded as an outline, thus to be excluded from the sample data for the calculation of the standard deviation. After a simple recursive process for outline removal was completed, the final $TPWV_{doy}$ standard deviation was obtained.

Step 3, both $TPWV_{f_{doy}}$ and $\sigma_{f_{doy}}$ on each *doy* were used to obtain the following five numerical boundaries for the *doy*: $TPWV_{f_{doy}} - 2\sigma_{f_{doy}}$, $TPWV_{f_{doy}} - \sigma_{f_{doy}}$, $TPWV_{f_{doy}}$, $TPWV_{f_{doy}} + \sigma_{f_{doy}}$, and $TPWV_{f_{doy}} + 2\sigma_{f_{doy}}$, and see the resultant five curves from *doy* 1 to 366 in Fig. 4 (a).

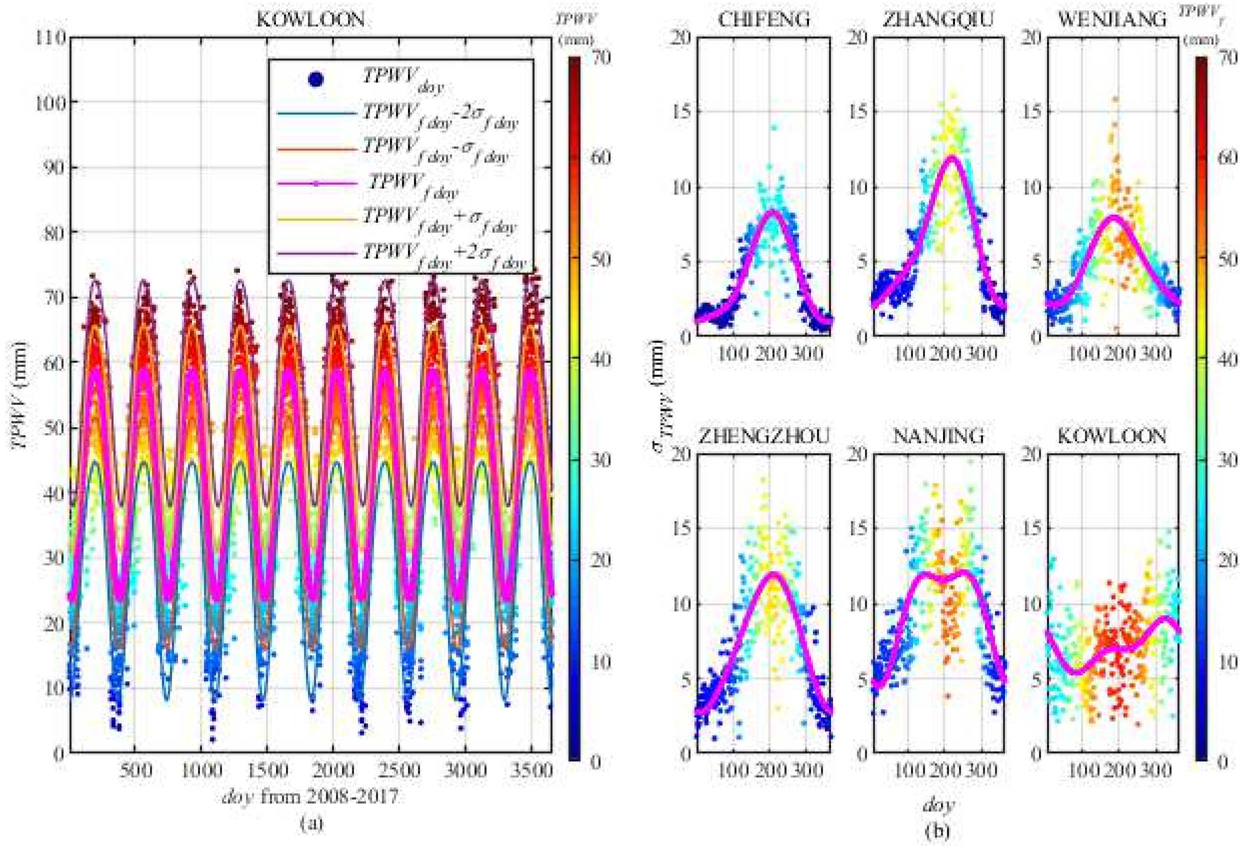


Fig. 4 (a) Time series of 10-year $TPWV_{doy}$ and its fitting function at KOWLOON; (b) standard deviation of $TPWV_{doy}$ in 366 *doy*s and its fitting function at each of the selected six stations.

Step 4, the above five numerical boundaries form six *TPWV* ranges (denoted by $TPWV_r$, *r*=1, 2, …, 6) for each *doy*: [less than $TPWV_{f_{doy}} - 2\sigma_{f_{doy}}$, $TPWV_{f_{doy}} - 2\sigma_{f_{doy}}$], [$TPWV_{f_{doy}} - 2\sigma_{f_{doy}}$, $TPWV_{f_{doy}} - \sigma_{f_{doy}}$], [$TPWV_{f_{doy}} - \sigma_{f_{doy}}$, $TPWV_{f_{doy}}$], [$TPWV_{f_{doy}}$, $TPWV_{f_{doy}} + \sigma_{f_{doy}}$], [$TPWV_{f_{doy}} + \sigma_{f_{doy}}$, $TPWV_{f_{doy}} + 2\sigma_{f_{doy}}$] and [$TPWV_{f_{doy}} + 2\sigma_{f_{doy}}$, greater than $TPWV_{f_{doy}} + 2\sigma_{f_{doy}}$]. The five numerical boundaries for all 366 *doy*s form six $TPWV_r$ curves.

The six $TPWV_r$ were taken as the quantified and unified *Rel-TPWV*, e.g., if $TPWV_{doy}$ values fall in the same $TPWV_r$, they are considered to have the same *Rel-TPWV* value. It is noted that in the special case that $TPWV_{f_{doy}} - 2\sigma_{f_{doy}}$ less than 0, $TPWV_r$ 1 does not exist and $TPWV_{f_{doy}} - 2\sigma_{f_{doy}}$ in $TPWV_r$ 2 is replaced by 0, as a result, only five $TPWV_r$ will be used in this case. Reflecting the results from the section above that is to say the vertical distribution of $lapse_{RPWV}$ is related to the *Rel-TPWV* of its corresponding *TPWV* in the same time session, i.e., if the *Rel-TPWV* is larger than others, the fluctuation amplitude of its corresponding $lapse_{RPWV}$ is smaller than the others, and also the vertical distribution is more stable, and vice versa. The *TPWV* values corresponding to each of the six $TPWV_r$ (from 1 to 6) are then considered as the following six states of WV: maximal disturbance, sub-disturbance, normal, normal, sub-saturated, and saturated, respectively; and their corresponding vertical distributions of $lapse_{RPWV}$ are considered as the following six vertical distributions of WV: maximal disturbance, sub-disturbance, normal, normal, sub-saturated, and saturated, respectively.

According to the above partition criterion for time, height, and *TPWV*, the 10-year sample data of

$lapse_{RPWV_{doy}}$ were first partitioned by the time partition criterion, then they were grouped by the $TPWV_r$ according to their corresponding $TPWV_{doy}$; finally, they were grouped by the h_r according to their heights. The resultant new groups of $lapse_{RPWV_{doy}}$ sample data will be used to study the characteristics of the temporal variations of $lapse_{RPWV}$ and model for each group.

3.2.3 Temporal and spatial characteristics of $lapse_{RPWV}$

For the spectrum analysis of each group of the 10-year $lapse_{RPWV_{doy}}$ time-series at each of the 12 stations obtained in the previous section, the Fourier transform method was used, and results at the KOWLOON station (as an example) are shown in Fig. 5 for the characteristics of the temporal variation of $lapse_{RPWV}$ of the station. The magenta curve in each subfigure of Fig. 5 (b) is the fit curve of the $lapse_{RPWV_{doy}}$ time-series has shown in the subfigure.

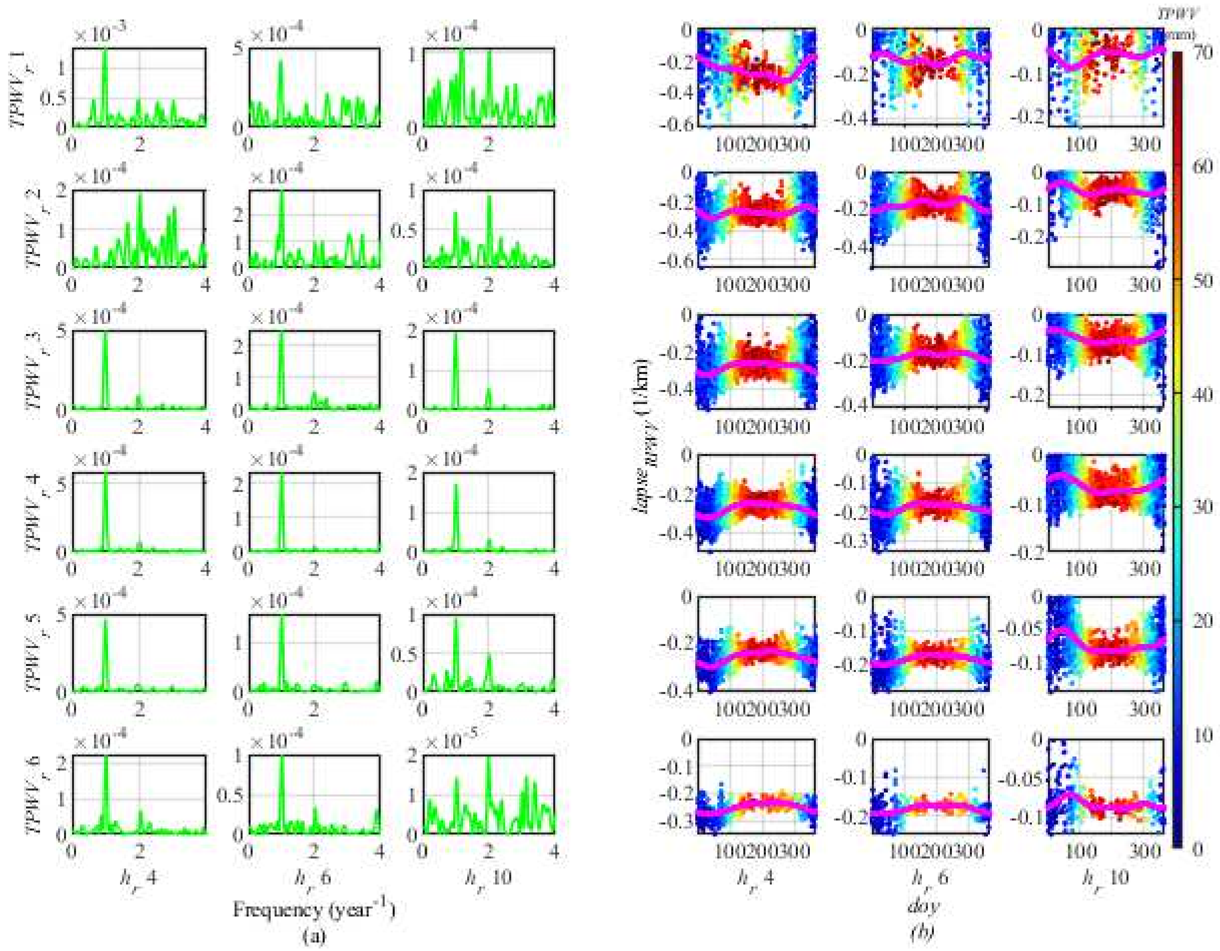


Fig. 5 (a) Power-period of 10-year $lapse_{RPWV_{doy}}$ time series at KOWLOON; (b) 10-year $lapse_{RPWV_{doy}}$ and their fitting periodic function in selected h_r within six $TPWV_r$ at KOWLOON. Note that six rows in both (a) and (b) represent the six $TPWV_r$.

As can be seen from Fig. 5 (a), the $lapse_{RPWV_{doy}}$ time-series in each of the selected h_r within six $TPWV_r$ generally present a notable periodic pattern, with significant annual and semi-annual cycles; in different h_r within different $TPWV_r$ show different periodic characteristics are shown. In h_r 6 within all $TPWV_r$, the annual cycle is significant, while the semi-annual cycle is weak, the opposite is true in h_r 10 within $TPWV_r$ 1, 2, and 6, while in some others the semi-annual cycle is equally significant with the annual cycle. Both in h_r 4 within $TPWV_r$ 2 and h_r 10 within $TPWV_r$ 1, the $lapse_{RPWV_{doy}}$ time-series show periods even less than a semi-annual cycle, e.g., a 4-month cycle. Comprehensive consideration of the main annual and semi-annual periodic characteristics of the $lapse_{RPWV_{doy}}$ time-series, the trig function of equation (16) taking the third order was adopted as its fitting function. It should be noted that the 4-month periodic characteristic may be insignificant or even not presented in a $lapse_{RPWV_{doy}}$ time-series. In this case, the coefficient of the 4 months term in equation (16) will be very small or even close to 0. Consequently, the trig function only contains the annual and semi-annual periodic terms. A comparison between the six rows in the same columns of the subfigures in Fig. 5 (b) from $TPWV_r$ 1 to 6, it can be observed that the fluctuation amplitude of $lapse_{RPWV_{doy}}$ also gradually decreases with the variation of $TPWV_r$ from 1 to 6 in each h_r, and then tends to be stable, e.g., with the fluctuation amplitude ranges of [–0.52, 0] 1/km to [–0.27, –0.23] 1/km, [–0.41, 0] 1/km to [–0.22, –0.18] 1/km, and [–0.20, 0] 1/km to [–0.12, 0] 1/km in h_r 4, 6, and 10, respectively. From the comparison between the same rows but in different columns, the fluctuation amplitude of $lapse_{RPWV_{doy}}$ in each of the six $TPWV_r$ is the largest in winter and smallest in summer, while it decreases with the increase in altitude, which is mainly due to a similar exponential decrease in water vapor with the increasing altitude.

Results at the other stations were also investigated and the same temporal variation pattern was observed, although the fluctuation amplitude of $lapse_{RPWV_{doy}}$ in different groups and different stations were different. The spatio-temporal modeling for each of the groups of the 10-year $lapse_{RPWV_{doy}}$ data will be carried out in the next section.

3.3 Construction of spatio-temporal $lapse_{RPWV}$ model

According to the classification of WV vertical distribution in section 3.2.2 that the vertical distribution of $lapse_{RPWV}$ in the $TPWV_r$ is considered as a type of WV vertical distribution, the true vertical distribution of $lapse_{RPWV}$ can be generalized as

$$lapse_{RPWV} = lapse_{RPWV_r} + disturbance_r \tag{17}$$

where $lapse_{RPWV_r}$ and $disturbance_r$ are the $lapse_{RPWV}$ and its disturbance term in the corresponding $TPWV_r$, respectively; *r* is from 1 to 6.

Based on the annual and semi-annual periodic characteristics of the $lapse_{RPWV_{doy}}$ time-series in each h_r within each $TPWV_r$, equation (16) (taking the third-order) was adopted as the fitting model for each of the groups of the 10-year $lapse_{RPWV_{doy}}$ sample data

$$\begin{aligned} lapse_{RPWV_{h,r}} = a_{0,h,r} + a_{1,h,r} \cos(doy \cdot w) + b_{1,h,r} \sin(doy \cdot w) + a_{2,h,r} \cos(doy \cdot 2w) + \\ b_{2,h,r} \sin(doy \cdot 2w) + a_{3,h,r} \cos(doy \cdot 3w) + b_{3,h,r} \sin(doy \cdot 3w) \end{aligned} \tag{18}$$

where $a_{0,h,r}$, $a_{1,h,r}$, $b_{1,h,r}$, $a_{2,h,r}$, $b_{2,h,r}$, $a_{3,h,r}$ and $b_{3,h,r}$ are the coefficients of the periodic terms; the subscripts *h* and r are the indexes of h_r and $TPWV_r$, respectively; $w = 2\pi/365.25$. The coefficients for each group of the $lapse_{RPWV_{doy}}$ sample data for the KOWLOON station (as an example) are partially selected and shown in Table 1.

Table 1 Coefficients of the $lapse_{RPWV}$ model for the selected h_r, i.e., h_r 1 and 5, within the selected $TPWV_r$, i.e., $TPWV_r$ 2 and 6, at KOWLOON station.

		a_0	a_1	b_1	a_2	b_2	a_3	b_3
…								
	h_r 1	-0.461	-0.018	0.018	-0.022	-0.027	0.012	-0.008
$TPWV_r$ 2	…							
	h_r 5	-0.227	-0.011	-0.008	-0.011	-0.015	-0.005	-0.001
	…							
…								
	h_r 1	-0.318	0.007	-0.002	0.010	0.003	0.001	0.003
$TPWV_r$ 6	…							
	h_r 5	-0.213	-0.014	-0.006	-0.002	-0.001	-0.001	0.005
	…							

3.4 Residuals of $lapse_{RPWV}$ model

The $lapse_{RPWV}$ residuals as the difference between the model-predicted and measured values as is shown in the sections above are shown in Fig. 6. We can see that, there are no pieces of very high or very low residuals in each of six $TPWV_r$ at all six stations. Most of the residuals in each subfigure distribute around 0 in an approximately symmetrical pattern, with approximately constant and uniform diffusion across the left and right, meaning that the residuals are random, hence the fitting performance of the new model is reasonable.

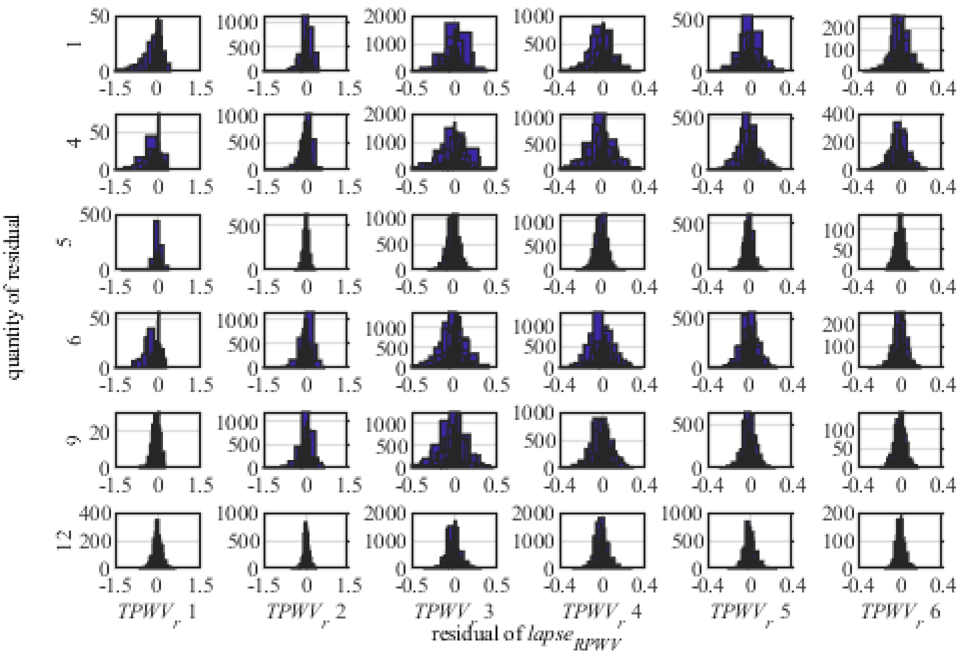


Fig. 6 Histogram of the residual distribution of $lapse_{RPWV}$ models in six $TPWV_r$ at each of the selected stations (with the serial numbers of 1, 4, 5, 6, 9, and 12, instead of using the name of the stations for convenience).

4 Evaluation of $lapse_{RPWV}$ model

To validate the aforementioned constructed $lapse_{RPWV}$ models, the H and ρ_w values for each *doy* in 2018 and 2019 predicted by the models (named $H_{m_{doy}}$ and $\rho_{w_{m_{doy}}}$, respectively) were compared against the H_{doy} and $\rho_{w_{doy}}$ reference values in the same two years using the procedure introduced in Section 3.1. The model results were also compared with that two values predicted by the commonly used *H* model, named $H_{H_{doy}}$ and $\rho_{w_{H_{doy}}}$, respectively. Since there are no global *H* models are available at present, according to the periodic characteristics of *H* (Otarola et al. 2011; Zhang et al. 2015), instead, the coefficients of the periodic model of *H* over each station were obtained from the H_{doy} time series in the 10 years from 2008 to 2017 by equation (16) (the second-order was adopted). The validation process is as follows.

(1) Validation of model-predicted $H_{m_{doy}}$ and $H_{H_{doy}}$.

According to the *Rel-TPWV* of $TPWV_{doy}$ relative to $TPWV_r$ on the same *doy* at a station site, the periodic function coefficients in each h_r within the corresponding $TPWV_r$ were selected, and the *doy* in 2018 and 2019 were used as the input of equation (18) to calculate the $lapse_{RPWV_{h,r}}$ value for the *doy*. Then $H_{m_{doy}}$ was calculated by the $lapse_{RPWV_{hs,r}}$ at the surface using equation (15). $H_{H_{doy}}$ on the same *doy* in 2018 and 2019 resulting from the *H* model were obtained using the coefficients of the periodic *H* model of the same station.

(2) Validation of model-predicted $\rho_{w_{m_{doy}}}$ and $\rho_{w_{H_{doy}}}$ in two cases below.

Case 1: $\rho_{w_{m_{doy}}}$ and $\rho_{w_{H_{doy}}}$ resulting from two models and $TPWV_{doy}$.

For the same station site, $\rho_{w_{H_{doy}}}$ at the surface and $\rho_{w_{m_{doy}}}$ in each h_r were calculated using the following two equations

$$\rho_{w_{H}}{}^{s}_{doy} = \frac{TPWV_{doy}}{H_{H_{doy}}} \tag{19}$$

$$\rho_{w_{m}}{}^{h_r}_{doy} = -TPWV_{doy} \cdot lapse_{RPWV_{h,r}} \tag{20}$$

Then, for the ith sounding height layer, $\rho_{w_{H}}{}^{i}_{doy}$ and $\rho_{w_{m}}{}^{i}_{doy}$ at height h^i_{doy} were calculated by equation (2) and interpolation between (h_{r1}, $\rho_{w_{m}}{}^{h_{r1}}_{doy}$) and ($h_{r2}$, $\rho_{w_{m}}{}^{h_{r2}}_{doy}$) (where $h_{r1} < h^i_{doy} < h_{r2}$), respectively.

Case 2: $\rho_{w_{m_{doy}}}$ and $\rho_{w_{H_{doy}}}$ resulting from two models and *WVD* at a specific altitude.

For the same station, $\rho_{w_{m}}{}^{j}_{doy}$ and $\rho_{w_{H}}{}^{j}_{doy}$ at height h^j_{doy} were calculated from $\rho_{w}{}^{i}_{doy}$ at the height h^i_{doy} (i and j are the indexes of the sounding layer from 1 to the last, respectively, and $i \neq j$) using the following equations

$$\rho_{w_{m}}{}^{j}_{doy} = \rho_{w}{}^{i}_{doy} \cdot \frac{lapse^j_{RPWV_m}}{lapse^i_{RPWV_m}} \tag{21}$$

$$\rho_{w_{H}}{}^{j}_{doy} = \rho_{w}{}^{i}_{doy} \exp(-(h^j_{doy} - h^i_{doy})/H_{H_{doy}}) \tag{22}$$

where $lapse^i_{RPWV_m}$ and $lapse^j_{RPWV_m}$ were obtained by interpolating (h_{r1}, $\rho_{w_{m}}{}^{h_{r1}}_{doy}$) and ($h_{r2}$, $\rho_{w_{m}}{}^{h_{r2}}_{doy}$), ($h_{r3}$, $\rho_{w_{m}}{}^{h_{r3}}_{doy}$) and ($h_{r4}$, $\rho_{w_{m}}{}^{h_{r4}}_{doy}$), respectively, where $h_{r1} < h^i_{doy} < h_{r2}$, $h_{r3} < h^i_{doy} < h_{r4}$, for an example.

(3) Using the above procedure to obtain $H_{m_{doy}}$, $\rho_{w_{m_{doy}}}$, $H_{H_{doy}}$ and $\rho_{w_{H_{doy}}}$ for all *doys* in 2018 and 2019 for each of the 12 stations, then the statistics including annual bias and root mean square error (RMSE) of the differences between the model-predicted results and reference values were calculated for the models' performance indicators. The formulas of the bias and RMSE are

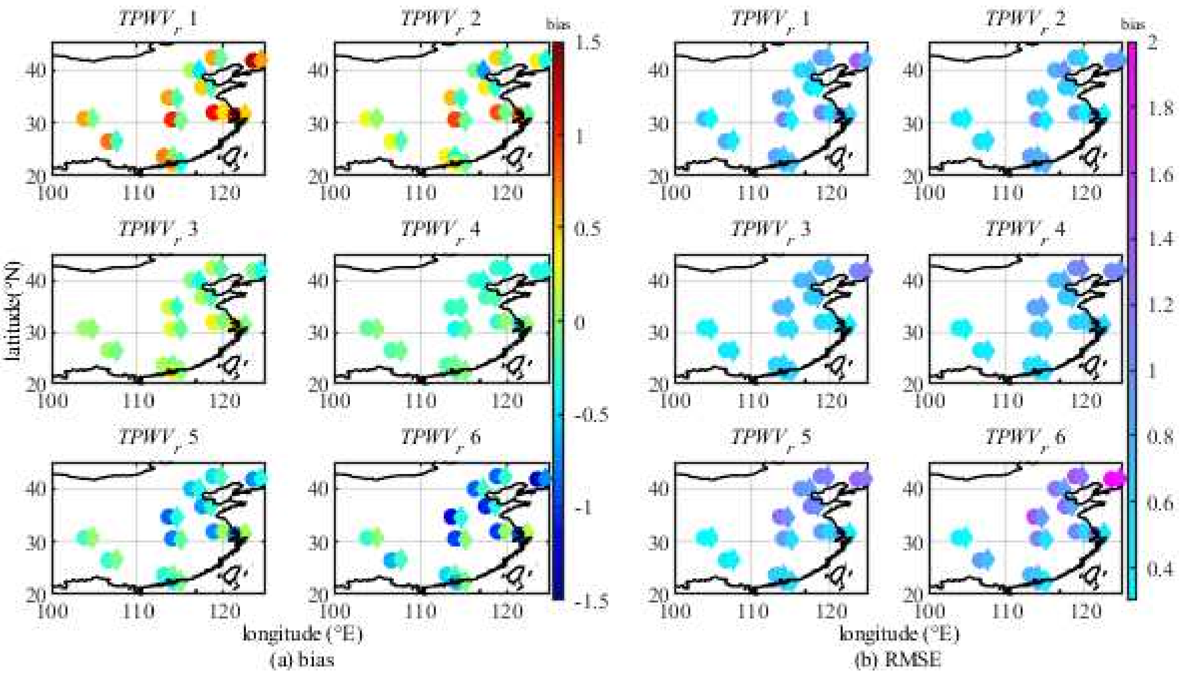
$$\text{bias} = \frac{1}{n}\sum_{i=1}^{n}\left(Value_{m_i} - Value_{r_i}\right) \tag{23}$$

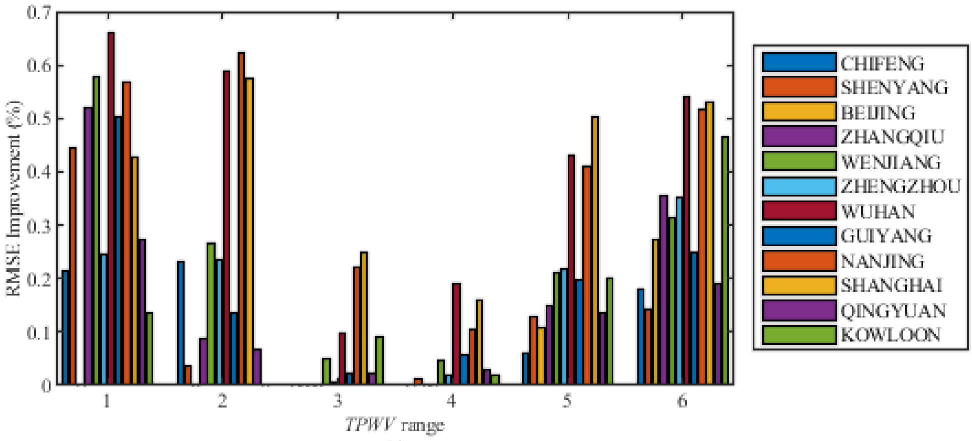
$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Value_{m_i} - Value_{r_i}\right)^2} \tag{24}$$

where i is the index of the sample data; the subscriptions m and r denote model and reference, respectively; n is the number of the samples contained in the statistics.

4.1 Accuracy of *H* predicted by two models

The annual biases and RMSEs of $H_{m_{doy}}$ and $H_{H_{doy}}$ in 2018 and 2019 at each of the 12 stations are shown in Fig. 7 (a) and (b), where the circular and diamond points denote $H_{H_{doy}}$ and $H_{m_{doy}}$, respectively. One can see that the annual biases of $H_{H_{doy}}$ of each station showed gradually increasing variation from negative to positive with the variation of $TPWV_r$ from 1 to 6, passing through the value of 0 with $TPWV_r$ from 3 to 4. While the biases of $H_{m_{doy}}$ are slightly greater than 0 at the CHIFENG, BEIJING, ZHANGQIU, and ZHENGZHOU stations, and that of other stations are around 0.





(c)

Fig. 7 Annual biases (a) and RMSEs (b) of H_{doy} in 2018 and 2019 at the 12 stations predicted from the two models; (c) Improvement ratio of annual RMSE of $H_{m_{doy}}$ relative to that of $H_{H_{doy}}$ in six $TPWV_r$ at each of the 12 stations.

Fig. 7 (b) shows the following (1) in $TPWV_r$ 1 and 2, the annual RMSEs of $H_{m_{doy}}$ were all smaller than that of $H_{H_{doy}}$ at all 12 stations except BEIJING, and the number of stations that had reduced annual RMSE accounted for 92% of the 12 stations. More specifically, in $TPWV_r$ 1 and at KOWLOON, the RMSE of $H_{m_{doy}}$ reduced about 14%, which was the minimum among all the stations. The maximum improvement was about 67% at WUHAN. In $TPWV_r$ 2, the two exceptional values were about 5% at SHENYANG and 63% at NANJING. (2) In $TPWV_r$ 3 and 4, the annual RMSEs of $H_{H_{doy}}$ were most less than that in the other $TPWV_r$ at all the stations, and they were most close to that of $H_{m_{doy}}$. Compared against the RMSEs of $H_{H_{doy}}$, the number of the stations that had reduced annual RMSE of $H_{m_{doy}}$ accounted for 67% and 83% of the 12 stations in the above two $TPWV_r$, respectively. (3) In $TPWV_r$ 5 and 6, the RMSEs of $H_{m_{doy}}$ were all smaller than that of $H_{H_{doy}}$ at all stations. The minimum and maximum improvements were about 8% at CHIFENG and 51% at SHANGHAI in $TPWV_r$ 5, respectively. The two exceptional values were about 15% at SHENYANG and 55% at WUHAN in $TPWV_r$ 6, respectively. In addition, the reduction in the annual RMSE of $H_{m_{doy}}$ in $TPWV_r$ from 1 to 6 at each station also can be observed from Fig. 7 (c).

4.2 Accuracy of ρ_w predicted by two models

For each of the 12 stations, $\rho_{w_{m_{doy}}}$ and $\rho_{w_{H_{doy}}}$ in 2018 and 2019 were obtained from the two models and $TPWV_{doy}$ and $\rho_{w_{doy}}$, respectively. Then the results were divided into 10 groups starting from the station height along the vertical direction with a 1 km interval. Their corresponding reference values obtained from the sounding data in the same 10 height ranges were used to evaluate the performance of the two models in each height interval. The bias and RMSE results are shown in the next two sections.

4.2.1 Accuracy of ρ_w resulting from models and *TPWV*

In Fig. 8, considering both the bias and RMSE, it can be seen that, at stations 1, 5, and 12, the annual biases of $\rho_{w_{m_{doy}}}$ are close to 0 in all 10 height ranges within all $TPWV_r$, and that of $\rho_{w_{H_{doy}}}$ all increase from a small negative value to 0 with the increase in height within all six $TPWV_r$. And the annual RMSEs of $\rho_{w_{H_{doy}}}$ at these three stations are at least about 2 times that of $\rho_{w_{m_{doy}}}$. The bias and RMSE values of the H model at station 5 in the height range from 1 to 5 are about from –4 to –2.5 g/m^3, and from 1 to 4.5 g/m^3, respectively. This performance of the H model is poor. At stations 4, 6, and 9 and in the same low height ranges, e.g., from 1 to 5, their absolute bias and RMSE values vary with the variation of $TPWV_r$ from 1 to 6, e.g., the two values in $TPWV_r$ 3 and 4 are less than that in the other $TPWV_r$, which reflects that the H model is only suitable to the normal water vapor state. Note that since WV content in a low height layer is much larger than that in a high layer, this section mainly focuses on the results in low height layers. Based on both the bias and RMSE of $\rho_{w_{m_{doy}}}$ and $\rho_{w_{H_{doy}}}$ resulting from the two models and $TPWV_{doy}$, the new model is obviously superior to the H model in all height ranges within all six $TPWV_r$ at all stations (the same results were also found from the other six unlisted stations).

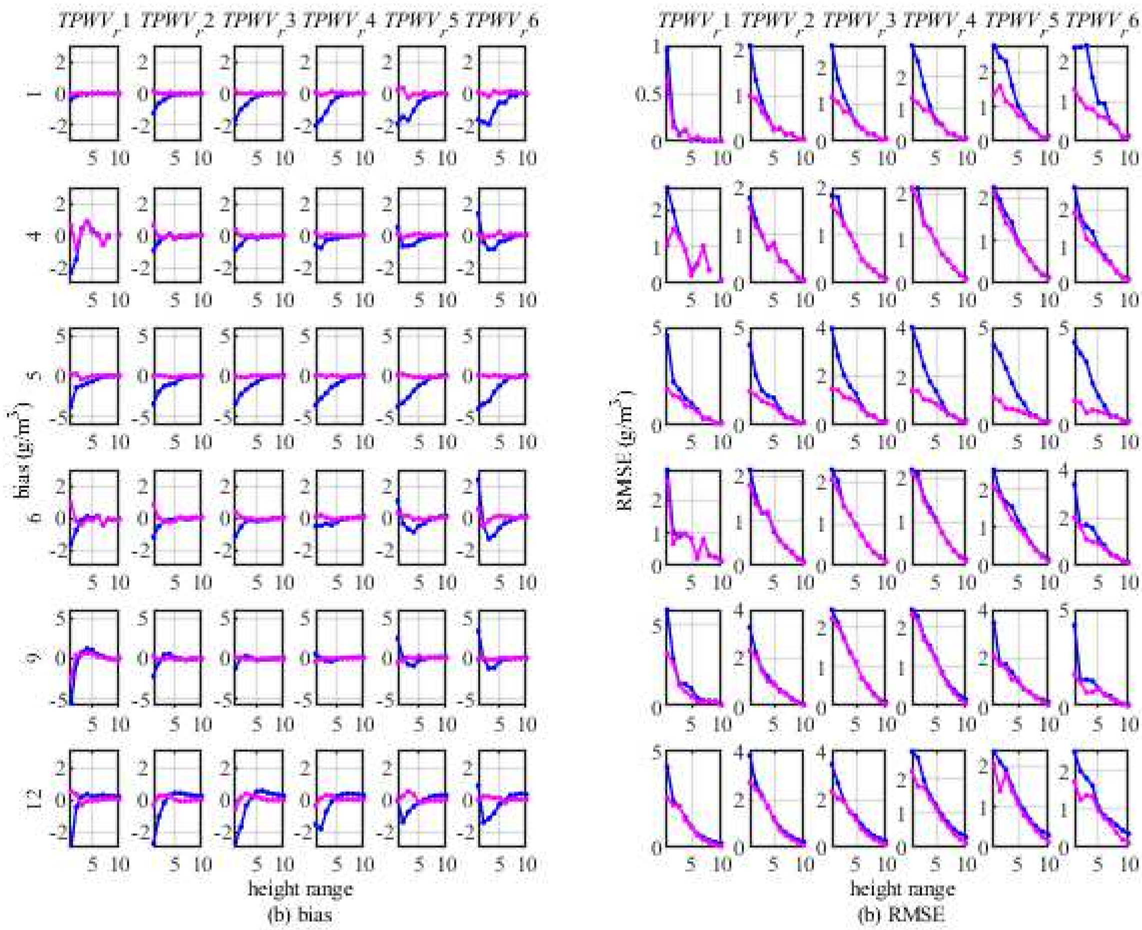


Fig. 8 Annual biases (a) and RMSEs (b) of $\rho_{w_{m_{doy}}}$ and $\rho_{w_{H_{doy}}}$ resulting from two models and $TPWV_{doy}$ in each of 10 height ranges within six $TPWV_r$ at selected six stations.

Due to the difficulty to find the difference between the RMSEs of $\rho_{w_{m_{doy}}}$ and $\rho_{w_{H_{doy}}}$ in higher height ranges in Fig. 8 (b), we counted the improvement ratios of the annual RMSE of $\rho_{w_{m_{doy}}}$ relative to that of $\rho_{w_{H_{doy}}}$ in the 10 height ranges within all $TPWV_r$ at the 12 stations, and results are shown in Fig. 9. It can be seen that the proportions of the number of the height ranges that had reduced the annual RMSE of $\rho_{w_{m_{doy}}}$ to all 10 height ranges within all six $TPWV_r$ (i.e. a total of 60 height ranges) are from 75% to 80% at the CHIFENG, BEIJING, ZHENGZHOU, and ZHANGQIU stations; and the results at the other stations were all over 90%.

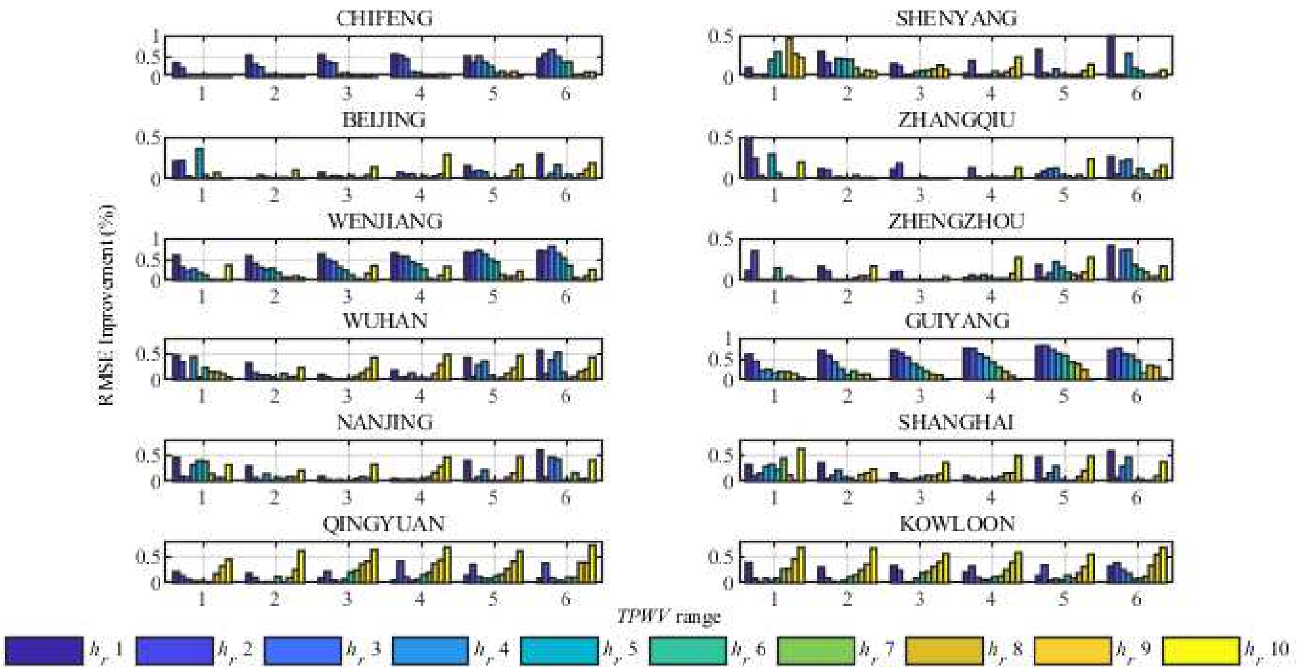
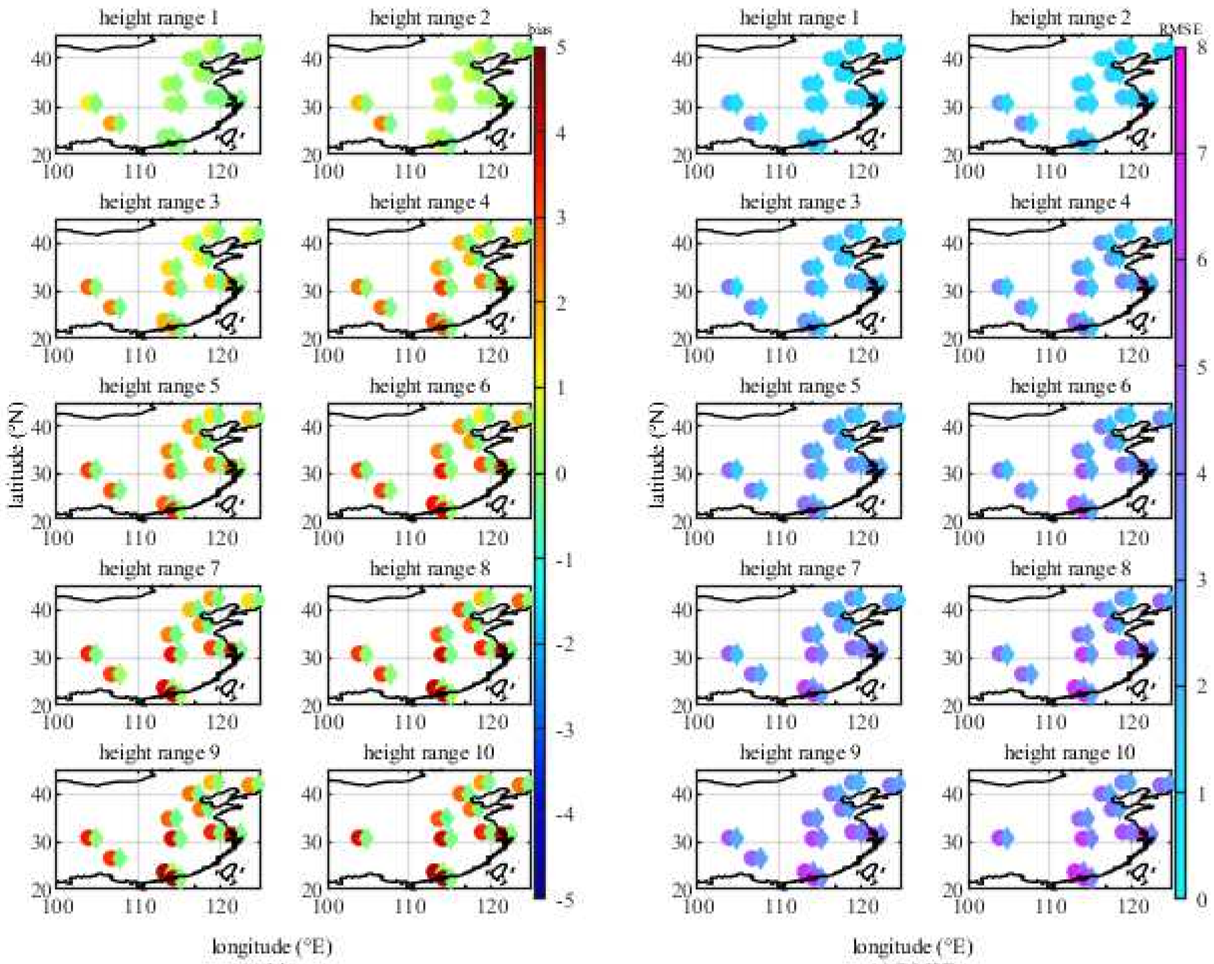


Fig. 9 Improvement ratio of annual RMSE of $\rho_{w_{m_{doy}}}$ resulting from new model and $TPWV_{doy}$ relative to that of $\rho_{w_{H_{doy}}}$ in each of 10 height ranges within each of six $TPWV_r$ at each of the 12 stations.

4.2.2 Accuracy of ρ_w resulting from models and *WVD* at a specific altitude

Fig. 10 (a) shows that the annual biases of $\rho_{w_{H_{doy}}}$ at all the 12 stations increase gradually with the increase in the height of $\rho_{w_{doy}}$, and in the height range 2 or above, they are all above 0, which indicates that the *H* model only performs well within the height range 1 and all the positive bias values mean systematical overestimation of the model. In contrast, the biases of $\rho_{w_{m_{doy}}}$ in the 10 height ranges all float around the 0 value. In Fig. 10 (b), the annual RMSEs of the two models at the 12 stations all show a tendency of larger RMSE in higher height ranges, i.e., the RMSE increases with the increase in the height of $\rho_{w_{doy}}$, with different amounts of increase. Therein, the minimum increment in the RMSE of $\rho_{w_{H_{doy}}}$ is 1.5 g/m^3 (from 3.5 to 5.0 g/m^3), and is at GUIYANG, where the RMSE of $\rho_{w_{m_{doy}}}$ is from 1.1 to 3.1 g/m^3; the maximum increment is 5.4 g/m^3 (from 1.7 to 7.1 g/m^3) and is at KOWLOON, where the RMSE of $\rho_{w_{m_{doy}}}$ is from 1.4 to3.9 g/m^3. Noticeably, the RMSE of $\rho_{w_{m_{doy}}}$ is less than that of $\rho_{w_{H_{doy}}}$ in all 10 height ranges. Compared with the $\rho_{w_{H_{doy}}}$ result, the RMSEs of $\rho_{w_{m_{doy}}}$ at all stations in each of the 10 height ranges reduce at least 11%, 20%, 43%, 48%, 40%, 38%, 32%, 35%, 32% and 28%, see Fig. 10 (c).

From the above results, it can be concluded that *WVD* at a lower height has resulted from the model and *WVD* at a higher height, while the small variation in the latter *WVD* (the variation in *WVD* at a higher height is small relative to *WVD* at the lower heights, but is large relative to *WVD* at the higher heights) has a great impact on the former. In addition, the poor result of the *H* model is also caused by the fact that the model is based on the exponential decline trend along with the vertical range from the surface *WVD* to the tropopause. Therefore, the new model is superior to the *H* model because it reflects the relationship between *WVD*s at different heights. And when the new model is applied, the *WVD* at a lower height should be selected as far as possible to calculate the *WVD* at a higher height through the model.



(a) bias (a) RMSE

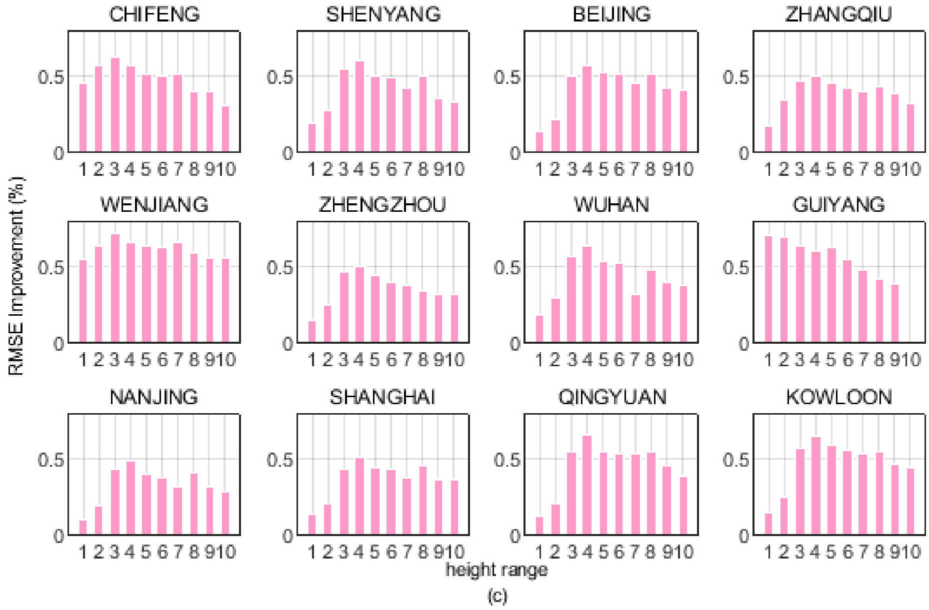


Fig. 10 Annual biases (a) and RMSEs (b) of $\rho_{w_{m_{doy}}}$ and $\rho_{w_{H_{doy}}}$ resulting from the two models and $\rho_{w_{doy}}$ from height ranges 1 to 10 at each of the 12 stations; (c) Improvement ratio of annual RMSE of $\rho_{w_{m_{doy}}}$ relative to that of $\rho_{w_{H_{doy}}}$.

5 Conclusion

The atmospheric water vapor in the troposphere varies with time and spatial location. Its vertical distribution presents temporal periodic features and also correlates with its state caused by some meteorological factors e.g. temperature and water vapor pressure. Water vapor scale height H is a common parameter reflecting the characteristics of the vertical distribution of atmospheric water vapor in the troposphere. The vertical distribution of water vapor may not be accurately exponential, since it is also correlative to the water vapor state over the site. In contrast, the traditional method of considering H value, to be either a constant or a periodic function is not a good choice.

For better modeling the temporal variation trend in the vertical distribution of water vapor under different water vapor states in the troposphere, in this study, a new function for the vertical variation in water vapor was derived by the ratio of the lapse rate of partial water vapor at any height range to the total precipitable water vapor (*TPWV)* over the site, named $lapse_{RPWV}$. From the analyses of the $lapse_{RPWV}$ time-series obtained from the 10-year from 2008 to 2017 sounding data over each of selected 12 radiosonde stations in China, it was found that the vertical distribution of the $lapse_{RPWV}$ not only strongly correlated with the relative magnitude of its corresponding *TPWV* at the same time and the same site, but also the $lapse_{RPWV}$ time-series show a periodic variation pattern in the temporal domain. For quantifying and unifying the standard of the relative magnitude of *TPWV*, this study proposed a method that was based on the periodic functions of *TPWV* and its standard deviation obtained from the 10-year data to construct six data ranges of *TPWV*. The vertical distributions of $lapse_{RPWV}$ corresponding to each of the six *TPWV* ranges (from 1 to 6) were considered as six vertical distributions of water vapor: maximal disturbance, sub-disturbance, normal, normal, sub-saturated, and saturated, respectively. Their corresponding *TPWV* ranges are also considered as the same six water vapor states.

From the investigation of the characteristics of the spatial and temporal variation of the processed 10-year $lapse_{RPWV}$ time-series in different height ranges within each of six $TPWV$ ranges, annual and semi-annual variation cycles were found in all the height ranges within all six $TPWV$ ranges at all the stations. Thus, a trigonometric periodic function was adopted for the new model fitting the 10-year $lapse_{RPWV}$ sample data of each height range within each of six $TPWV$ ranges at each station. The new fitting model was validated by comparing its prediction against the reference obtained from sounding data in 2018 and 2019 (out-of-sample data). Its results were also compared with that of the common H model. Results showed that the new $lapse_{RPWV}$ model significantly outperformed the H model in the following aspects. First, the annual accuracies of H values resulting from the new model were improved over the H model to various degrees, and the number of the stations that had reduced annual RMSE of H values in the $TPWV$ ranges from 1 to 6 accounted for 92%, 92%, 67%, 83%, 100%, and 100%, respectively, of the total stations. Secondly, the proportions of the number of the height ranges that had reduced annual RMSE of water vapor density (WVD) obtained from the new model and $TPWV$ in all height ranges within all $TPWV$ ranges were from 75% to 80% at the CHIFENG, BEIJING, ZHENGZHOU and ZHANGQIU stations; and the results at the other stations were all over 90%. Thirdly, considering all six $TPWV$ ranges and all stations as a whole, the annual RMSEs of WVD obtained from the new model and WVDs in the height ranges from 1 to 10 reduced at least 11%, 20%, 43%, 48%, 40%, 38%, 32%, 35%, 32%, and 28%, respectively. All results suggest that, under different water vapor states, the new $lapse_{RPWV}$ fitting model reflects not only the temporal relationship between the surface WVD and the $TPWV$ of a site but also the vertical distributions of WVDs of different heights well.

Acknowledgments

We acknowledge the National Oceanic and Atmospheric Administration (NOAA) for the provision of the Integrated Global Radiosonde Archive (IGRA) datasets.

References

Bevis M, Businger S, Herring TA, et al (1992) GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system. Journal of Geophysical Research 97:15787. https://doi.org/10.1029/92JD01517

Bobak JP, Ruf CS (1996) Prediction of water vapor scale height from integrated water vapor measurements. In: International Geoscience and Remote Sensing Symposium (IGARSS)

Borger C, Beirle S, Dörner S, et al (2020) Total column water vapour retrieval from S-5P/TROPOMI in the visible blue spectral range. Atmospheric Measurement Techniques 13:2751–2783. https://doi.org/10.5194/amt-13-2751-2020

Byers HR (1957) Significance of Different Vertical Distributions of Water Vapor in Arid and Humid Regions. Journal of the Meteorological Society of Japan Ser II 35A:330–335. https://doi.org/10.2151/jmsj1923.35A.0_330

Chahine MT (1992) The hydrological cycle and its influence on climate. Nature 359:373–380. https://doi.org/10.1038/359373a0

Ding N, Zhang S, Wu S, et al (2018) A new approach for GNSS tomography from a few GNSS stations

Elósegui P, Ruis A, Davis JL, et al (1998) An experiment for estimation of the spatial and temporal variations of water vapor using GPS data. Physics and Chemistry of the Earth 23:125–130. https://doi.org/10.1016/S0079-1946(97)00254-1

Flores A, de Arellano JVG, Gradinarsky LP, Rius A (2001) Tomography of the lower troposphere using a small dense network of GPS receivers. IEEE Transactions on Geoscience and Remote Sensing 39:439–447. https://doi.org/10.1109/36.905252

Flores A, Ruffini G, Rius A (2000) 4D tropospheric tomography using GPS slant wet delays. Annales Geophysicae 18:223–234. https://doi.org/10.1007/s00585-000-0223-7

Guo J, Yang F, Shi J, Xu C (2016) An Optimal Weighting Method of Global Positioning System (GPS) Troposphere Tomography. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9:5880–5887. https://doi.org/10.1109/JSTARS.2016.2546316

Jacob D (2001) The role of water vapour in the atmosphere. A short overview from a climate modeller's point of view. Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy 26:523–527. https://doi.org/10.1016/S1464-1895(01)00094-1

John T, Garg SC, Maini HK, et al (2005) Design of a simple low cost tethersonde data acquisition system for meteorological measurements. Review of Scientific Instruments 76:84501. https://doi.org/10.1063/1.1988029

Kanemaru K, Masunaga H (2013) A Satellite Study of the Relationship between Sea Surface Temperature and Column Water Vapor over Tropical and Subtropical Oceans. Journal of Climate 26:4204–4218. https://doi.org/10.1175/JCLI-D-12-00307.1

Kennett EJ, Toumi R (2005) Temperature dependence of atmospheric moisture lifetime. Geophysical Research Letters 32:n/a-n/a. https://doi.org/10.1029/2005GL023936

Liu Z, Chen B, Chan ST, et al (2015) Analysis and modelling of water vapour and temperature changes in Hong Kong using a 40-year radiosonde record: 1973-2012. International Journal of Climatology 35:462–474. https://doi.org/10.1002/joc.4001

Liu Z, Wong MS, Nichol J, Chan PW (2013) A multi-sensor study of water vapour from radiosonde, MODIS and AERONET: A case study of Hong Kong. International Journal of Climatology 33:109–120. https://doi.org/10.1002/joc.3412

Otárola A, Travouillon T, Schöck M, et al (2010) Thirty Meter Telescope Site Testing X: Precipitable Water Vapor. Publications of the Astronomical Society of the Pacific 122:470–484. https://doi.org/10.1086/651582

Otarola AC, Querel R, Kerber F (2011) Precipitable Water Vapor: Considerations on the water vapor scale height, dry bias of the radiosonde humidity sensors, and spatial and temporal variability of the humidity field

Perler D, Geiger A, Hurter F (2011) 4D GPS water vapor tomography: new parameterized approaches. Journal of Geodesy 85:539–550. https://doi.org/10.1007/s00190-011-0454-2

Reber EE, Swope JR (1972) On the Correlation of the Total Precipitable Water in a Vertical Column and Absolute Humidity at the Surface. Journal of Applied Meteorology 11:1322–1325. https://doi.org/10.1175/1520-0450(1972)011<1322:OTCOTT>2.0.CO;2

Reitan CH (1963) Surface Dew Point and Water Vapor Aloft. Journal of Applied Meteorology 2:776–779. https://doi.org/10.1175/1520-0450(1963)002<0776:SDPAWV>2.0.CO;2

Rocken C, van Hove T, Ware R (1997) Near real-time GPS sensing of atmospheric water vapor. Geophysical Research Letters 24:3221–3224. https://doi.org/10.1029/97GL03312

Ruf CS, Beus SE (1997) Retrieval of tropospheric water vapor scale height from horizontal turbulence structure. IEEE Transactions on Geoscience and Remote Sensing 35:203–211. https://doi.org/10.1109/36.563258

Ruffini G, Rius A, Cucurull L, Flores A (1999) Analysis of Water Vapor spatio-temporal structure over the Madrid Area using GPS data

Schüler T (2014) The TropGrid2 standard tropospheric correction model. GPS Solutions 18:123–131. https://doi.org/10.1007/s10291-013-0316-x

Tomasi C (1981) Determination of the total precipitable water by varying the intercept in Reitan's relationship. Journal of Applied Meteorology 20:1058–1069

Tomasi C (1977) Precipitable Water Vapor in Atmospheres Characterized by Temperature Inversions. Journal of Applied Meteorology 16:237–243. https://doi.org/10.1175/1520-0450(1977)016<0237:PWVIAC>2.0.CO;2

Tomasi C (1984) Vertical Distribution Features of Atmospheric Water Vapor in the Mediterranean, Red Sea, and Indian Ocean. Journal of Geophysical Research 89:2563–2566

Viswanadham Y (1981) The Relationship between Total Precipitable Water and Surface Dew Point. Journal of Applied Meteorology 20:3–8. https://doi.org/10.1175/1520-0450(1981)020<0003:TRBTPW>2.0.CO;2

Weaver CP, Ramanathan V (1995) Deductions from a simple climate model: factors governing surface temperature and atmospheric thermal structure. Journal of Geophysical Research 100:11585. https://doi.org/10.1029/95jd00770

Zhang B, Yao Y, Xu C (2015) Global empirical model for estimating water vapor scale height. Cehui Xuebao/Acta Geodaetica et Cartographica Sinica 44:. https://doi.org/10.11947/j.AGCS.2015.20140664

Zhao T, Dai A, Wang J (2012) Trends in tropospheric humidity from 1970 to 2008 over china from a homogenized radiosonde dataset. Journal of Climate 25:. https://doi.org/10.1175/JCLI-D-11-00557.1

Statements & Declarations

Funding

This work was supported by the State Key Program of the National Natural Science Foundation of China [grant number 41730109] and the Programme of Introducing Talents of Discipline to Universities [grant number B20046].

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Moufeng Wan. The first draft of the manuscript was written by Moufeng Wan and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability

The radiosonde data can be downloaded from: https://www.ncdc.noaa.gov/data-acces s/weather-balloon/integrated-global-radiosonde-archive.