Analysis of the Weighted Mean Temperature of China Based on Sounding and ECMWF Reanalysis Data

WANG Xiaoying[1] (王晓英), SONG Lianchun[2*] (宋连春), and CAO Yunchang[3] (曹云昌)

1 *Institute of Remote Sensing, Nanjing University of Information Science & Technology, Nanjing* 210044

2 *National Climate Center, China Meteorological Administration, Beijing* 100081

3 *Meteorological Observation Center, China Meteorological Administration, Beijing* 100081

(Received August 31, 2011; in final form March 21, 2012)

ABSTRACT

Weighted mean temperature ($T_{\rm m}$) is one of the most important conversion parameters for calculating precipitable water vapor by the signal path wet delay in ground-based GPS meteorology. This paper first discusses the $T_{\rm m}$ regression models for Hong Kong (HK) and the associated error statistics relative to the true values of $T_{\rm m}$ from the numerical method. The results show that there is little difference in precision between annual and seasonal $T_{\rm m}$ regression models for HK. The Bevis $T_{\rm m}$-$T_{\rm s}$ (surface temperature) regression model is more suitable for northeastern China and the Qinghai-Tibetan Plateau than the local models. For areas lack of historical sounding data, the Kriging interpolation method and the ECMWF reanalysis product ERA-interim were employed to set up local $T_{\rm m}$-$T_{\rm s}$ models. The results indicate that the $T_{\rm m}$ derived by the ERA-interim data coincides well with that by the sounding data, and the Kriging interpolation method can successfully obtain the coefficients of local $T_{\rm m}$-$T_{\rm s}$ models, suggesting that these two approaches may serve as effective ways in the acquisition and localization of $T_{\rm m}$.

Key words: weighted mean temperature, ERA-interim, surface temperature, ground-based GPS, sounding data

Citation: Wang Xiaoying, Song Lianchun, and Cao Yunchang, 2012: Analysis of weighted mean temperature of China based on sounding and ECMWF reanalysis data. *Acta Meteor. Sinica*, **26**(5), 642–652. doi: 10.1007/s13351-012-0508-2.

1. Introduction

Water vapor is an active composition of the atmosphere, with extremely uneven distributions and large temporal and spatial variations in the atmosphere. The phase change of water vapor is directly related to precipitation and plays an important role in the atmosphere energy transfer, weather development, radiation budget of the earth-atmosphere system, and global climate change (Bevis et al., 1992; Rocken et al., 1995; Bi et al., 2006; Falconer et al., 2009; Perler et al., 2011). Low temporal and spatial resolutions of routine atmosphere observations restrict our knowledge on the spatiotemporal variation features of water vapor. As a result, it leads to low precision of the initial humidity field in numerical weather prediction (NWP) models as well as low accuracy of NWP results (Gutman et al., 2004; Song, 2004; Cao et al., 2005, 2006; Bastin et al., 2007). Ground-based GPS technology provides a new mean for atmospheric water vapor observation, which can be made all the time under all weather conditions at high precision. It supplements the traditional atmosphere observation methods (Song, 2004; Lutz, 2009).

When using path wet delay (PWD) to obtain precipitable water vapor (PWV) in ground-based GPS technology, there are three important computation formulas as follows.

$$\mathrm{PWV} = \Pi \times \mathrm{PWD}, \tag{1}$$

$$\Pi = \frac{10^6}{\rho_{\rm v} R_{\rm v}(k_2' + k_3/T_{\rm m})}, \tag{2}$$

$$T_{\rm m} = \frac{\displaystyle\int (e/T)\mathrm{d}z}{\displaystyle\int (e/T^2)\mathrm{d}z}, \tag{3}$$

Supported by the National Basic Research and Development (973) Program of China (2012CB955903 and 2010CB428401) and China Meteorological Administration Special Public Welfare Research Fund (GYHY201306011).

*Corresponding author: songlc@cma.gov.cn.

where Π is conversion coefficient, ρ_v is density of liquid water equal to 1.0 g cm^{-3}, R_v is gas constant for water vapor equal to 0.4613 J g^{-1} K^{-1}, and k_2' and k_3 are test constants of atmosphere refractive index equal to k_2'=22.1±2.2 K hPa^{-1} and k_3=(3.739±0.012)×10^5 K^2 hPa^{-1}. As indicated by Eq. (3), T_m is the average temperature of the atmosphere weighted by the water vapor pressure, called the weighted mean temperature.

To evaluate the impact of T_m on the PWV, the first-order partial derivative is derived as follows:

$$\partial(\mathrm{PWV})/\partial T_\mathrm{m} = \frac{10^6 k_3}{\rho_\mathrm{v} R_\mathrm{v}(k_2' T_\mathrm{m} + k_3)^2}\mathrm{PWD}. \qquad (4)$$

Typical values of zenith path hydrostatic delay (PHD) are 2.30 m and 0.00–0.40 m for PWD (Walpersdorf et al., 2001). Assuming the variables of PWD and T_m are constants, substitute the extreme values (e.g., PWD = 400 mm, T_m = 285 K) of all parameters into Eq. (4), the result is as follows:

$$\mathrm{d(PWV)/d}T_\mathrm{m} = 0.2243. \qquad (5)$$

Equation (5) shows that the T_m error of 1 K can cause a PWV error of 0.2243 mm in the case of a PWD value of 400 mm and T_m of 285 K. Therefore, accurate calculation of T_m is important. In ground-based GPS technology, there are several ways to acquire T_m: (1) considering T_m as a constant; (2) computing T_m by use of a numerical integration method based on Eq. (3) using radiosonde data or numeral prediction products; (3) computing T_m from T_s (T_s denotes surface temperature) based on specific local T_m-T_s regression models. In the above methods, considering T_m as a constant may lead to low precision of T_m and PWV; the numerical integration method has the highest precision and T_m thus derived is often taken as the "true" observation, but it is difficult to apply this method to real-time acquisition of T_m since the radiosonde data or numeral prediction products are not available in the real-time mode. Computing T_m from T_s based on the T_m-T_s regression models is simple and easy with acceptable precision in the current ground-based GPS technology.

Most studies use statistical regression models to obtain T_m. Bevis et al. (1992) made an analysis of 8718 radiosonde profiles spanning approximately a 2-yr interval for the stations in the United States with the latitude range of 27°–65°N and a height range of 0–1.6 km, and yielded a linear regression T_m=70.2+0.72T_s. The RMS deviation from this regression is 4.74 K, which is a relative error of less than 2%. The Bevis regression model has been applied to many studies of T_m in China (Li et al., 1999; Liu et al., 2006; Yue et al., 2008; Li et al., 2009; Yu and Liu, 2009; Wang et al., 2011a, b). However, it is found that the errors of the Bevis empirical model distribute unevenly in China and are generally more than 4 K with the extreme value of 8 K in some areas (Yu et al., 2011). The impact of the Bevis empirical model errors on the PWV can reach the millimeter level under an extreme weather condition with rich water vapor.

Previous studies focused on the setup of local T_m-T_s regression models for a specific region based on sounding data (Li et al., 1999; Liu et al., 2006; Yue et al., 2008; Li et al., 2009; Wang et al., 2011a). Yu and Liu (2009) and Yu et al. (2011) discussed the error statistics of the Bevis model over China and gave the dependence of the errors on the elevation of station. Wang et al. (2011b) investigated the correlation of the coefficients a and b of the local T_m model with longitude, latitude, and station elevation. Nonetheless, some aspects such as the spatiotemporal distribution characteristics of T_m and the associated retrieval error comparison between the Bevis model and local models over entire China are rarely researched. Yu and Liu (2009) presented a method adding a correction relevant to elevation in the Bevis model to obtain T_m over areas without historical sounding data.

Reanalysis of multi-decadal series of past observations has been widely utilized for the studies of atmospheric and oceanic processes and predictability. Since reanalysis data are produced using sophisticated and advanced data assimilation systems developed for NWP, they are more suitable than operational analysis for use in studies of long-term variability of climate. Reanalysis products are used increasingly in many fields that require an observational record of the state of either the atmosphere or its underlying land and ocean surfaces. So far, ECMWF has produced three reanalysis products including ERA-15, ERA-40, and ERA-interim, among which ERA-interim is a global reanalysis of the data-rich period since 1989. The ERA-interim data assimilation system uses a 2006 release of the Integrated Forecasting System (IFS Cy31r2), which contains many improvements in both the forecasting model and the analysis methodology, compared to those for ERA-40. The ERA-interim reanalysis was put into operation in March 2009, and has been running in near real-time to support climate monitoring (http://www.ecmwf.int/research/era/do/get/Reanalysis_ECMWF, retrieved on 19 December 2011).

This study adopts the numerical integration method and the sounding data of 87 international exchange sounding stations in China to obtain the spatiotemporal distribution characteristics of $T_{\rm m}$ over China. Meanwhile, we compare the results of $T_{\rm m}$ calculated from the Bevis model with those from the local regression models based on historical sounding data, and then evaluate the performance of the Bevis model and the local $T_{\rm m}$-$T_{\rm s}$ regression models. When setting up regression models for areas without sounding data, the Kriging spatial interpolation method is used to obtain the coefficients of the local $T_{\rm m}$-$T_{\rm s}$ regression models and the results are assessed in comparison with the numerical method results using the sounding data. In addition, the $T_{\rm m}$-$T_{\rm s}$ regression model results based on the ERA-interim reanalysis product are also compared with those based on the sounding data of Beijing and Hong Kong (HK), and the results indicate that the ERA-interim reanalysis provide another data source for local $T_{\rm m}$ calculation in areas without sounding data.

2. Data and method

In this paper, we use local $T_{\rm m}$-$T_{\rm s}$ regression models based on sounding data or the ERA-interim reanalysis product to derive $T_{\rm m}$, and use the "true" observation calculated from the sounding data with the numerical integration formula to validate the regression results. The data of 87 international sounding stations in China for 2003–2010 are downloaded from the University of Wyoming website http://weather.uwyo.edu/upperair/sounding.html (retrieved on 19 December 2011). The ERA-interim data are downloaded from http://data-portal.ecmwf.int/data/d/interim_daily (retrieved on 19 December 2011).

The numerical integration method is a discretion form of Eq. (3) and is described below. $T_{\rm m}$ can be retrieved from both sounding data and reanalysis products as follows (Li et al., 1999; Liu et al., 2006; Yue et al., 2008; Li et al., 2009; Yu and Liu, 2009; Wang et al., 2011a, b).

$$T_{\rm m}=\frac{\sum\limits_{i=1}^{n}\overline{\left(\frac{e_i}{T_i}\right)}\Delta h_i}{\sum\limits_{i=1}^{n}\overline{\left(\frac{e_i}{T_i^2}\right)}\Delta h_i}, \tag{6}$$

where $\overline{\left(\frac{e_i}{T_i}\right)}=\left(\frac{\frac{e_i}{T_i}+\frac{e_{i-1}}{T_{i-1}}}{2}\right)$, $\overline{\left(\frac{e_i}{T_i^2}\right)}=\left(\frac{\frac{e_i}{T_i^2}+\frac{e_{i-1}}{T_{i-1}^2}}{2}\right)$, Δh_i is the thickness of ith layer atmosphere, e_i and T_i are upper bound water vapor pressure and temperature of ith layer atmosphere respectively, and e_{i-1} and T_{i-1} are lower bound water vapor pressure and temperature of ith layer atmosphere respectively.

For sounding data, the water vapor pressure is equal to saturated vapor pressure of dew point temperature. For reanalysis product ERA-interim, the temperature, geopotential height (H), and relative humidity (RH) fields have a resolution of 1.5° latitude ×1.5° longitude and 37 vertical levels. The surface temperature ($T_{\rm s}$) and surface dew point temperature fields at the same horizontal resolution are also used. All variables are available 4 times daily at 0000, 0600, 1200, and 1800 UTC. The water vapor pressure (e) is obtained from RH and saturation water vapor pressure (E) as follows:

$$\begin{aligned}
&t=T-273.15,\\
&E=6.1078\times 10(7.5\times t/(237.3+t)),\ t>0,\\
&E=6.1078\times 10(9.5\times t/(265.5+t)),\ t\leqslant 0,\\
&e={\rm RH}\times E.
\end{aligned} \tag{7}$$

When Eq. (6) is used to calculate $T_{\rm m}$, all data at the levels below the station's geopotential height value are discarded. The surface water vapor pressure is calculated from the surface dew point temperature.

When comparing the $T_{\rm m}$ values obtained from sounding with that obtained from ERA-interim based on Eq. (6), interpolations are conducted as follows. In the vertical direction, the level at which the station is located is determined according to the station's actual geopotential height. Linear interpolation is used to obtain the surface temperature of the station based on the reanalysis data of two adjacent vertical grids, and exponential interpolation is used to obtain its water vapor pressure. In the horizontal direction, the temperature and water vapor pressure at each level above the station are obtained using bi-linear interpolation based on data of the four adjacent grids at the same level.

The T_m-T_s regression model is obtained using Matlab based on historical T_m and T_s sample pairs as follows. First, T_m is calculated by Eq. (6) twice a day for sounding data and four times a day for ERA-interim reanalysis product. Note that T_s can be directly obtained from the sounding data or ERA-interim reanalysis product. Then, the T_m-T_s regression model based on the total samples in a certain period of time can be set up with the Matlab software. For example, during a one-year period, there are about 730 samples for sounding data and 1460 samples for ERA-interim reanalysis product. The precision of the T_m-T_s regression model is evaluated by comparing T_m calculated by Eq. (6) with T_m calculated by the T_m-T_s regression model.

3. Regression of T_m at HK by seasons based on the sounding data

Based on the sounding data of HK (WMO station ID 45004) in 2005, T_m of this station was calculated by Eq. (6). Figure 1 shows the T_m annual change in 2005. T_m varies significantly by season between 275 and 295 K with the highest value in summer and lowest in winter. This gives us a rough idea of how T_m changes with season. It also raises a question as to if it is better to establish seasonal regression models according to seasonal sounding data rather than to obtain general/annual regression models according to consecutive multi-year sounding data time series, i.e., if seasonality of the regression should be considered and could be used to facilitate the regression accuracy?

In this section, the precision of the annual and seasonal T_m-T_s regression models for the HK station based on the sounding data are evaluated. There are 59 missing records for a total of 4325 valid times for the period 2003–2008 for HK sounding station. The HK_annual T_m-T_s model in Table 1 is set up based on all the sounding data from 2003 to 2008. We classify the total sounding data of 4325 entries from 2003 to 2008 by season, and about 1080 (T_m and T_s) samples are obtained for every season. Based on that, four seasonal T_m-T_s empirical models are set up, called HK_season (spring, summer, autumn, and winter), as shown in Table 1. Table 1 indicates that the a and b coefficients of spring and autumn models are close to that of the annual model, while the a and b coefficients of summer and winter models differ largely from that of the annual model.

We use the sounding data of HK station in 2009 to validate the local seasonal T_m-T_s empirical models. The values of T_m in 2009 are calculated using Eq. (6) and taken as real values. With regard to the seasonal regression models in Table 1, there are a total of 730 (T_m and T_s) samples in 2009 with averagely 185 samples in every season. The model values of T_m for the four seasons in 2009 are calculated from the seasonal T_m-T_s empirical models, respectively, and their statistic results are shown in Table 2. Although the a and b coefficients of summer and winter models differ largely from that of the annual model, there is almost no difference when replacing any of the seasonal models with the annual one to calculate T_m. On the contrary, the Bevis model has a larger bias than local seasonal models but almost the same RMS.

Why do the a and b coefficients of the summer and winter models differ largely from that of the annual model, but the T_m values differ little from that calculated from their linear combination? The linear trends of T_m from the four seasonal models and the annual model are shown in Fig. 2. The summer temperature in HK varies from 299.15 to 306.15 K (yellow vertical range) with the mean value of 301.15 K, while the winter temperature in HK varies from 287.15 to 293.15 K (lavender vertical range) with the mean value of 290.15 K. Figure 2 shows that the trend line of T_m from the summer model is very close to that from the

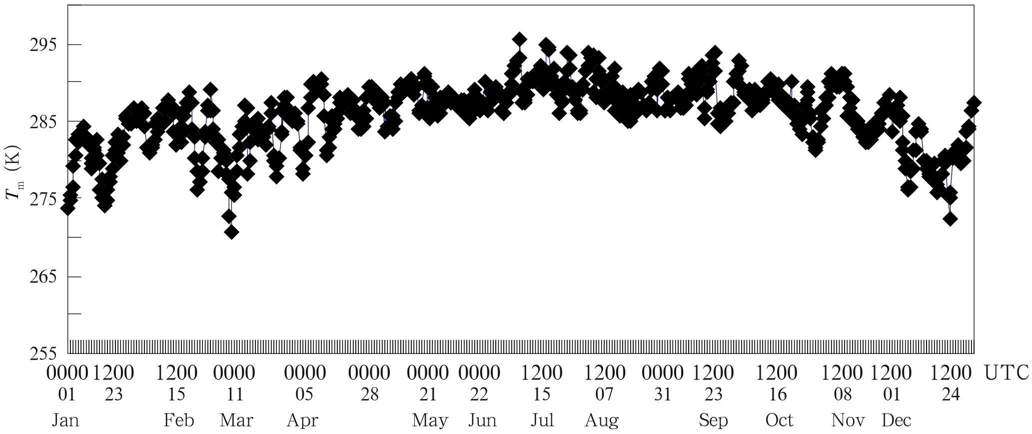

Fig. 1. Annual change of T_m obtained by using the numerical integration method and the 0000 and 1200 UTC sounding data of HK in 2005.

Table 1. Statistics of the seasonal T_m-T_s empirical models for HK

Model type	Sample time	Samples	Regression model	Bias (K)	RMS (K)
HK_year	Annual	4325	T_m=115.55+0.58T_s	1.49	1.88
HK_season (spring)	Spring (Mar–May)	1091	T_m=116.39+0.57T_s	1.59	1.99
HK_season (summer)	Summer (Jun–Aug)	1070	T_m=–7.55+0.99T_s	1.27	1.67
HK_season (autumn)	Autumn (Sep–Nov)	1089	T_m=124.36+0.55T_s	1.49	1.83
HK_season (winter)	Winter (Dec–Feb)	1075	T_m=50.88+0.80T_s	1.49	1.88

Table 2. Statistics of the seasonal T_m-T_s empirical models in comparison with the Bevis model for HK

Model name	Bias (K)				RMS (K)			
	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter
Bevis	3.06	2.31	3.25	4.38	2.21	1.49	1.88	1.95
HK_year	1.96	1.51	1.43	1.72	2.08	1.56	1.68	2.17
HK_season	1.81	1.61	1.91	2.30	2.08	1.57	1.68	2.18

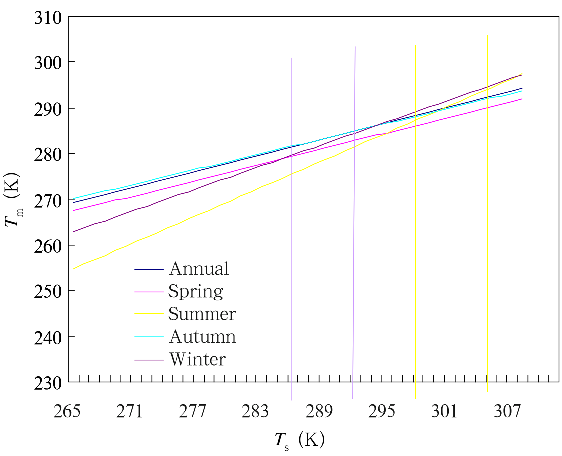

Fig. 2. Trend lines of the seasonal and annual regression models for HK.

annual model within the temperature range of 299.15–306.15 K, so is that with the winter model and annual model for the temperature range of 287.15–293.15 K. Thus, T_m calculated from the specific seasonal model and annual model is almost the same.

4. Spatiotemporal distribution of T_m over China derived from the sounding data

4.1 T_m distribution by the numerical integration method based on the sounding data

The average monthly T_m in January, April, July, and October is calculated based on the sounding data from 2003 to 2009 by Eq. (6) to reveal the spatial and temporal distribution characteristics of T_m over China. The results are shown in Fig. 3.

Figure 3 shows that the average monthly T_m over southeastern China is the highest throughout the year, whereas it is the lowest over northeastern China in

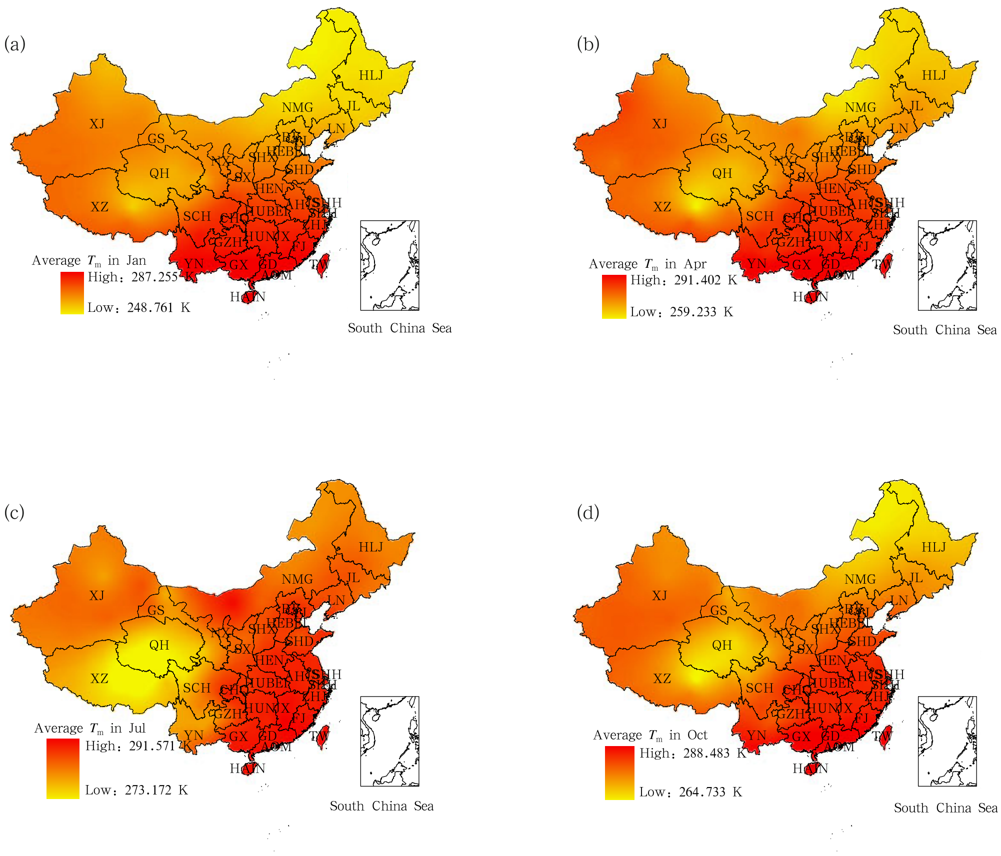

Fig. 3. Spatial distributions of the average monthly T_{m} (K) calculated by using the sounding data and Eq. (6) over China in (a) January, (b) April, (c) July, and (d) October.

January and over the Qinghai-Tibetan Plateau in July.

The general characteristics of spatiotemporal distributions of the average monthly surface temperature T_{s} in China are consistent with T_{m}. The correlation coefficient between T_{m} and T_{s} is 0.848 (Wang et al., 2011b), which shows the feasibility of estimating T_{m} by T_{s}.

Previous studies (e.g., Li et al., 1999; Liu et al., 2006; Yue et al., 2008; Li et al., 2009; Yu and Liu, 2009; Wang et al., 2011a, b) never investigated the overall spatiotemporal distribution of T_{m} over China. Figure 3 shows that the T_{m} distributions may have something to do with the climate pattern, population density, and topography. Southeastern China is under the influence of the subtropical monsoon climate with high population density, and T_{m} is relatively high in this region throughout the year. Northeastern China is dominated by the temperate monsoon climate, and T_{m} is comparatively low all the year except in summer. T_{m} over the Qinghai-Tibetan Plateau is significantly low compared with other areas in summer because of its unique topography.

4.2 T_{m} distribution by the regression method based on the sounding data

There are 120 sounding stations in China, among which 87 are international exchange stations. The data of the 87 stations from 2009 to 2010 are examined. We use the radiosonde data of 2009 to obtain the local T_{m}-T_{s} regression model for each station, then use the sounding data of 2010 to evaluate the local model and compare with the Bevis model.

The results shown in Table 3 and Fig. 4 reveal that not all the local T_{m}-T_{s} models are more precise than the Bevis model; this is inconsistent with the conclusion of previous studies (Li et al., 1999; Liu et al., 2006; Yue et al., 2008; Li et al., 2009; Yu and Liu, 2009; Wang et al., 2011a, b). In Table 3, dT_m means the difference between the real value of T_m in 2010 by Eq. (6) using sounding data and the model value of T_m by the local T_m-T_s model or the Bevis model using the radiosonde data of 2009, where a_l is the a coefficient of the T_m-T_s model and same is for b_l. The Bias_l, RMSe_l, Max_l, and Num_l are respectively the mean bias, standard deviation, maximum value of dT_m, and the number of samples with dT_m greater than five degrees based on the local model. Similarly, the Bias_B, RMSe_B, Max_B, and Num_B are the corresponding statistical variables for the Bevis model. The number of samples is about 730 in 2010.

The results indicate that there are 22 sounding stations for which the local T_m-T_s models are not better than the Bevis model. Table 3 only lists the statistical information of 9 stations. Table 3 and Fig. 4 show that the Bevis model is fit for areas of northeastern China and the Qinghai-Tibetan Plateau while it has a large deviation in southeastern and northwestern China, where the local T_m-T_s model works better.

5. Regression of T_m at two stations using the Kriging method based on the sounding data

The Kriging interpolation method provides an unbiased optimal estimation for regional variables in a limited area based on variation function spatial analysis. It takes into account the spatial correlation between the data points and is suitable for interpolation

Table 3. Comparison of the local T_m-T_s model with the Bevis model

Station number	a_l	b_l	Bias_l	RMSe_l	Max_l	Bias_B	RMSe_B	Max_B	Num_l	Num_B
51828	109.93	0.57	3.71	4.44	12.72	3.50	4.55	10.68	118	47
52323	98.07	0.61	3.72	4.59	12.01	3.64	4.58	8.64	132	55
52533	107.36	0.58	3.91	4.84	14.44	4.06	4.95	11.14	144	77
52681	95.62	0.63	4.03	4.83	12.26	3.88	4.79	12.76	145	110
52818	129.96	0.49	4.18	4.92	17.38	4.64	4.91	9.25	179	27
52836	80.33	0.67	3.77	4.62	15.45	4.03	4.50	11.14	122	31
54218	53.91	0.77	3.82	4.73	13.81	4.05	4.94	10.97	124	72
55299	67.01	0.72	3.92	4.84	9.1	4.87	4.84	4.81	72	0
56029	69.61	0.71	3.57	4.57	10.64	4.15	4.58	7.52	65	8

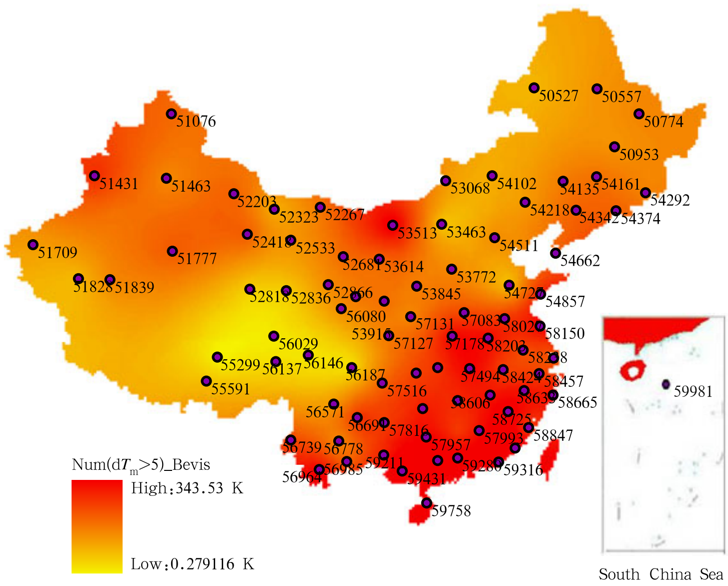

Fig. 4. Distribution of the number of samples with the T_m difference more than 5 K calculated from the Bevis model and the local T_m-T_s model based on the sounding data in China.

of spatial data. Details for Kriging interpolation can be seen in Zeng and Huang (2007).

The average distance of adjacent sounding stations is about hundreds of kilometers. For southeastern and northwestern China, the local T_m-T_s models are difficult to set up due to the absence of sounding data. The Kriging spatial interpolation method is used to obtain the a and b coefficients of the local T_m-T_s models for areas without sounding data. This method is evaluated at two sounding stations, i.e., Pizhou (WMO station ID 57972) and Kuche (WMO station ID 51644). Pizhou station locates in southeastern China where a dense network of sounding stations exists, whereas Kuche station locates in northwestern China where there is a lack of sounding stations. Selecting the two stations for test can validate if the Kriging interpolation method is able to obtain local T_m-T_s models for areas with both dense and coarse distributions of sounding stations at acceptable precision. For areas such as southeastern and northwestern China where the Bevis model fails to perform well, the local T_m-T_s models should be set up.

The evaluation for Pizhou station is conducted as follows: 1) use the Kriging method to generate a and b isosurfaces based on the a and b coefficients of the rest of the 86 sounding stations; 2) obtain the interpolated values of the a and b coefficients for this station, and the Kriging T_m-T_s model for this station is then obtained; 3) use the observed sounding data of 2010 to derive a real local T_m-T_s model, compare with the Kriging model derived in step 2, and calculate their differences. The same process as described above is applied to Kuzhe station. The results are shown in Table 4.

In Table 4, a_k is the a coefficient of the T_m-T_s model based on the Kriging interpolation method and same holds for b_k. The Bias_k, RMSe_k, Max_k, and Num_k are respectively the mean bias, standard deviation, maximum value of dT_m, and the number of samples with dT_m greater than five degrees based on the Kriging method. The a_l, b_l, Bias_l, RMSe_l, Max_l, and Num_l are the same as in Table 3.

Table 4 shows that the Kriging spatial interpolation method is feasible in setting up the appropriate local T_m-T_s model for areas with no historical sounding data such as southeastern and northwestern China, where the Bevis model performs not well.

Table 4. Comparison of the local T_m-T_s models and the Bevis model for Pizhou and Kuche stations

Station	a_l	b_l	a_k	b_k	Bias_l	RMSe_l	Max_l	Num_l	Bias_k	RMSe_k	Max_k	Num_k
Kuche	100.04	0.62	106.03	0.59	3.11	3.88	10.06	89	3.04	3.71	10.47	75
Pizhou	106.455	0.61	98.75	0.64	1.92	2.37	7.5	16	2.08	2.42	7.1	9

6. Evaluation of T_m at Beijing and HK calculated by the numerical and regression methods from two data sources

6.1 T_m values calculated by the numerical method from both data sources

We used ERA-interim and sounding data of 2003 for Beijing and 2007 for HK to derive T_m values by Eq. (6). The results are shown in Table 5.

There are 1460 samples for the ERA-interim data in both 2003 and 2007, from which 730 samples match the time of the sounding data for 2003 and 719 samples for 2007. Table 5 shows that there is little difference between the T_m values calculated by Eq. (6) from these two different datasets.

Table 5. Comparison of T_m computed by Eq. (6) from the ERA-interim data and from the sounding data for Beijing and HK

	Sample year (samples)	Bias (K)	RMS (K)
Beijing	2003 (730)	0.0014	1.3861
HK	2007 (719)	0.0006	0.5841

Statistical analysis in Table 5 is performed in the following way:

Difference: $d_i = T_{\mathrm{m(ERA\text{-}interim)}} - T_{\mathrm{m(sounding)}}$;

Systematic error: Bias=$\dfrac{\sum d_i}{n}$, where n is number of samples;

Accuracy: RMS=$\sqrt{\dfrac{\sum d_i^2}{n-1}}$. (8)

6.2 T_m values by the regression models from the ERA-interim data

Once the local T_m-T_s empirical formula has been set up, T_m can be calculated by T_s with high precision, and it can then be used to obtain PWV using PWD in ground-based GPS. We set up a T_m-T_s empirical formula using the ERA-interim data of 2003 for Beijing and 2007 for HK. The accuracy of this empirical formula is evaluated by the ERA-interim data of 2004 for Beijing and 2008 for HK. The statistical results are listed in Table 6.

Table 6. Statistics of the T_m-T_s regression models for Beijing and HK based on the ERA-interim data

Station	Beijing	HK
Year (sample number)	2003 (1460)	2007 (1460)
Regression RMS (K)	3.44	1.92
T_m-T_s model	T_m=0.80T_s+48.20	T_m=0.50T_s+137.90
Year (sample number) to evaluate the accuracy of the T_m-T_s model	2004 (1462)	2008 (1462)
Accuracy evaluation Bias	0.40	0.09
Accuracy evaluation RMS	3.99	1.96

How the statistical analysis in Table 6 was performed is detailed in Table 7 for both Beijing and HK.

6.3 T_m values by regression models from both data sources

To further understand the regressed T_m values from the ERA-interim data, the corresponding sounding data were also processed. Comparisons of T_m by the ERA-interim and sounding data are shown in Table 8. Table 8 shows that the local T_m-T_s empirical model $(T_m\text{-}T_s)_{\text{EAR-interim}}$ based on the EAR-interim reanalysis data coincides well with that based on the sounding data.

To sum up, the above comparisons demonstrate a good agreement between ERA-interim reanalysis product and sounding data when calculating T_m based on the numerical integration method and when setting up T_m-T_s regression models in Beijing and HK. This is easy to understand since radiosonde data are the key and main data source for the assimilation of reanalysis

Table 7. Statistical method used to obtain Table 6

Difference 2003	$d_{i2003}=T_m(\text{regression})-T_m(\text{Eq. (6) using 2003 data})$
Regression RMS	$\text{RMS}=\sqrt{\frac{\sum d_{i2003}^2}{n-1}}$, n: samples in 2003
Difference 2003–2004	$d_{i0304}=T_m(\text{regression})-T_m(\text{Eq. (6) using 2004 data})$
Accuracy evaluation Bias	$\text{Bias}=\frac{\sum d_{i0304}}{n}$, n: samples in 2004
Accuracy evaluation RMS	$\text{RMS}=\sqrt{\frac{\sum d_{i0304}^2}{n-1}}$

Table 8. Comparisons of the T_m-T_s empirical formulas derived from ERA-interim data and sounding data

Station	Beijing		HK	
Data source	ERA-interim	Sounding data	ERA-interim	Sounding data
Year (sample number)	2003 (1460)	2003 (730)	2007 (1460)	2007 (719)
Regression RMS	3.44	3.73	1.92	1.85
T_m-T_s model	T_m=0.80T_s+48.20	T_m=0.83T_s+40.53	T_m= 0.50T_s+137.90	T_m=0.55T_s+122.42
Year (sample number) to evaluate the accuracy of the model	2004 (1462)	2004 (731)	2008 (1462)	2008 (726)
Accuracy evaluation Bias	0.40	0.94	0.09	0.15
Accuracy evaluation RMS	3.99	4.90	1.96	2.02

products. For areas with rich sounding data, the reanalysis product in general has good quality too. It is also proved that the ECMWF reanalysis product ERA-interim can be applied to the acquisition of $T_{\rm m}$ for Beijing and HK areas.

7. Conclusions

This paper first compares the difference between the annual and seasonal $T_{\rm m}$-$T_{\rm s}$ regression models for HK in detail and then examines the applicability of the Bevis $T_{\rm m}$-$T_{\rm s}$ regression model over China. For areas lack of historical sounding data, the Kriging interpolation method and the ECMWF reanalysis product ERA-interim are used to set up local $T_{\rm m}$-$T_{\rm s}$ models. We draw the following conclusions:

(1) The annual and seasonal $T_{\rm m}$ models have considerable accuracy, $T_{\rm m}$ can be calculated precisely by annual $T_{\rm m}$ model for all seasons, and it is not necessary to set up local seasonal $T_{\rm m}$ models for HK sounding station.

(2) The Bevis model is fit for areas of northeastern China and the Qinghai-Tibetan Plateau, but it produces a large deviation of $T_{\rm m}$ in southeastern and northwestern China, where the local $T_{\rm m}$-$T_{\rm s}$ model should be set up.

(3) The Kriging spatial interpolation method can be used to set up the local $T_{\rm m}$-$T_{\rm s}$ model for areas with no historical sounding data such as southeastern and northwestern China, where the Bevis model performs not well.

(4) Acquisition of the local $T_{\rm m}$ based on ERA-interim reanalysis data coincides well with that based on the sounding data for Beijing and HK, which might provide a substitute available data source to set up local $T_{\rm m}$-$T_{\rm s}$ models for areas lack of sounding data.

For areas lack of historical sounding data, the Kriging interpolation method and the ECMWF reanalysis product ERA-interim were both employed to set up the local $T_{\rm m}$-$T_{\rm s}$ model. The ECMWF reanalysis product has wide data coverage and long data span on acceptable grid resolutions, whereas for the Kriging method in Section 5, there are only 120 sounding stations in China. Future research may focus on the validation of the ECMWF reanalysis product ERA-interim in setting up local $T_{\rm m}$-$T_{\rm s}$ models in comparison with the Kriging method. If the result is encouraging, the ECMWF reanalysis product can be used to set up the local $T_{\rm m}$-$T_{\rm s}$ models on the grid scale. Meanwhile, the Kriging method may also be used to interpolated the a and b coefficients of the local $T_{\rm m}$-$T_{\rm s}$ models for more homogeneously and rich distributed grids.

It should be pointed out that because the quality of the reanalysis product strongly depends on the quality and richness of the sounding data, for areas with no or few sounding data, the quality of the reanalysis product may not be good. It is difficult to say whether the reanalysis product is fit for $T_{\rm m}$ calculation for areas without sounding data. Nevertheless, it will be the direction of further research about whether the reanalysis product might serve as an effective and rich data source for the acquisition and localization of $T_{\rm m}$ for areas with no sounding data.

Acknowledgments. The ECMWF ERA-interim data used in this study are provided by the ECMWF, and sounding data are provided by the University of Wyoming.

REFERENCES

Bastin, S., C. Champollion, O. Bock, et al., 2007: Diurnal cycle of water vapor as documented by a dense GPS network in a coastal area during ESCOMPTE IOP2. *Bull. Amer. Meteor. Soc.*, **46**, 167–182.

Bevis, M., S. Businger, T. A. Herring, et al., 1992: GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system. *J. Geophys. Res.*, **97**(D14), 15787–15801.

Bi Yanmeng, Mao Jietai, Liu Xiaoyang, et al., 2006: Use ground-based GPS to sense slant path integrated water vapor. *Chinese J. Geophys.*, **49**(2), 335–342.

Cao Yunchang, Fang Zongyi, and Xia Qing, 2005: Preliminary analysis on the correlation between GPS sensed integrated water vapor and regional precipitation. *J. Appl. Meteor. Sci.*, **16**(1), 54–59. (in Chinese)

——, Chen Yongqi, Li Binghua, et al., 2006: The method of using ground-based GPS to obtain atmospheric water vapor profile. *Meteor. Sci. Tech.*, **34**(3), 241–245. (in Chinese)

Gutman, S. I., S. R. Sahm, S. G. Benjamin, et al., 2004: Rapid retrieval and assimilation of ground based

GPS precipitable water observations at the NOAA Forecast Systems Laboratory: Impact on weather forecasts. *J. Meteor. Soc. Japan*, **82**, 351–360.

Falconer, R. H., D. Cobby, P. Smyth, et al., 2009: Pluvial flooding: New approaches in flood warning, mapping and risk management. *J. Flood Risk Manage*, **2**, 198–208.

Li Guocui, Li Guoping, Du Chenghua, et al., 2009: Research on weighting mean temperature model during water vapor inversion using ground-based GPS in North China area. *J. Nanjing Inst. Meteor.*, **32**(1), 81–86. (in Chinese)

Li Jianguo, Mao Jietai, and Li Chengcai, 1999: The principle of using GPS to sense water vapour distribution and regression analysis on weighing mean temperature in eastern China. *Acta Meteor. Sinica*, **57**(3), 283–292. (in Chinese)

Liu Xuchun, Wang Yanqiu, Zhang Zhenglu, et al., 2006: The comparison analysis on the obtaining method of weighting mean temperature when using ground-based GPS to sense atmospheric integrated water vapor. *Journal of Beijing University of Civil Engineering and Architecture*, **22**(2), 38–40. (in Chinese)

Lutz, S., 2009: High-resolution GPS tomography in view of hydrological hazard assessment. PH. D. dissertation, ETH ZURICH, i-vi.

Perler, D., G. Alain, and H. Fabian, 2011: 4D GPS water vapor tomography: New parameterized approaches. *J. Geodesy.*, **85**(8), 539–550.

Rocken, C., van H. T., J. Johnson, et al., 1995: GPS/STORM-GPS sensing of atmospheric water vapor for meteorology. *J. Atmos. Ocean. Technol.*, **12**, 468–478.

Song Shuli, 2004: Monitor water vapor three-dimensional distribution using ground-based GPS network and its usage in meteorology. Ph. D. dissertation, Shanghai Astron. Obs., 1–14. (in Chinese)

Walpersdorf, A., E. Calais, J. Haase, et al., 2001: Atmospheric gradients estimated by GPS compared to a high resolution numerical weather prediction (NWP) model. *Physics and Chemistry of the Earth*, **26**(3), 147–152.

Wang Xiaoying, Song Lianchun, Dai Ziqiang, et al., 2011a: Feature analysis of weighted mean temperature $T_{\rm m}$ in Hong Kong. *Journal of Nanjing University of Information Science & Technology (Natural Science Edition)*, **1**, 47–52.

——, Dai Ziqiang, Cao Yunchang, et al., 2011b: Weighted mean temperature statistical analysis in ground-based GPS in China. *Geomatics and Information Science of Wuhan University*, **36**(4), 412–416. (in Chinese)

Yu Shengjie and Liu Lintao, 2009: Verification and analysis on water vapor weighting mean temperature regression formula. *Geomatics and Information Science of Wuhan University*, **34**(6), 741–744. (in Chinese)

——, Wan Rong, and Fu Zhikang, 2011: Research on localization of weighted mean temperature in GPS water vapor sounding. Conference on New Data Usage in Heavy Rain Sounding. Hubei, 8–10 February, Institute of Heavy Rain, China Meteorological Administration, Wuhan, 15–20. (in Chinese)

Yue Yingchun, Chen Chunming, and Yu Yan, 2008: Research on the usage of GPS technology to sense atmospheric water vapor in the South Pole area. *Sci. Survey Map.*, **33**(5), 81–84. (in Chinese)

Zeng Huaien and Huang Shenxiang, 2007: Research on spatial data interpolation based on Kriging interpolation. *Engineering of Surveying and Mapping*, **16**(5), 5–10. (in Chinese)