Investigation of near-global daytime boundary layer height using high-resolution radiosondes: First results and comparison with ERA5, MERRA-2, JRA-55, and NCEP-2 reanalyses

Jianping Guo[a], Jian Zhang[b*], Kun Yang[c], Hong Liao[d], Shaodong Zhang[e], Kaiming Huang[e],Yanmin Lv[a], Jia Shao[f], Tao Yu[b], Bing Tong[a], Jian Li[a], Tianning Su[g], Steve H.L. Yim[h,i], Ad Stoffelen[j], Panmao Zhai[a], and Xiaofeng Xu[k]

[a] State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China

[b] Hubei Subsurface Multi-scale Imaging Key Laboratory, Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China

[c] Department of Earth System Science, Tsinghua University, Beijing 100084, China

[d] Nanjing University of Information Science and Technology, Nanjing 210044, China

[e] School of Electronic Information, Wuhan University, Wuhan 430072, China

[f] College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

[g] Department of Atmospheric and Oceanic Sciences, University of Maryland, College Park, Maryland 20740, USA

[h] Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, Hong Kong, China

[i] Stanley Ho Big Data Decision Analytics Research Centre, The Chinese University of Hong Kong, Shatin, Hong Kong, China

[j] The Royal Netherlands Meteorological Institute (KNMI), 3730 AE De Bilt, The Netherlands

[k] China Meteorological Administration, Beijing 100081, China

*Correspondence to:

Dr. Jian Zhang (Email: zhangjian@cug.edu.cn)

Abstract

The planetary boundary layer (PBL) governs the vertical transport of mass, momentum and moisture between the surface and the free atmosphere, and thus the determination of PBL height (BLH) is recognized as crucial for air quality, weather and climate analysis. Although reanalysis products can provide important insight into the global view of BLH in a seamless way, the *in situ* observed BLH on a global scale remains poorly understood due to the lack of high-resolution (1-s or 2-s) radiosonde measurements. The present study attempts to establish a near-global BLH climatology at synoptic times (0000 and 1200 UTC) and in the daytime using high-resolution radiosonde measurements over 300 radiosonde sites worldwide for the period 2012 to 2019, which is then compared against the BLHs obtained from four reanalysis datasets, including ERA5, MERRA-2, JRA-55, and NCEP-2. The variations of daytime BLH exhibit large spatial and temporal dependence, and as a result the BLH maxima are generally discerned over the regions such as Western United States and Western China, in which the balloon launch times mostly correspond to the afternoon. The diurnal variations of BLH are revealed with a peak at 1700 local solar time (LST). The most promising reanalysis product is ERA5, which underestimates BLH by around 130 m as compared to radiosondes released during daytime. In addition, MERRA-2 is a well-established product and has an underestimation of around 160 m. JRA-55 and NCEP-2 might produce considerable additional uncertainties, with a much larger underestimation of up to 400 m. The largest bias in the reanalysis data appears over the Western United States and Western China and it might be attributed to the maximal BLH in the afternoon when the PBL has grown up. Statistical analyses further indicate that the biases of reanalysis BLH products are positively associated with orographic complexity, as well as the occurrence of static instability. To our best knowledge, this study presents the first near-global view of high-resolution radiosonde derived boundary layer height and provides a quantitative assessment of the four frequently used reanalysis products.

Keywords. Radiosonde; boundary layer height; reanalysis; sensible heat flux

1. Introduction

The planetary boundary layer (PBL) is where most of exchanges of heat, moisture, momentum and mass take place between the free atmosphere and ground surface (Stull, 1988; Liang and Liu, 2010). The spatial and temporal variability of PBL, through a variety of physical processes, has a profound influence on research fields such as air quality (Stull, 1988; Li *et al.,* 2017), convective storm (Oliveira *et al*., 2020) and global warming (Davy and Esau, 2016), among others. It is well known to be influenced by radiative cooling at night and by downward solar radiation reaching the ground surface at daytime, respectively, forming a stable boundary layer (SBL) and convective boundary layer (CBL), with a typical PBL depth (BLH) of less than 500 m and 1–3 km (Zhang *et al.,* 2020a), respectively. For climate models, most of the PBL processes occur at sub-grid scales and thus are either underrepresented or not fully represented (von Engeln and Teixeira, 2013). Meanwhile, there are many problems in elucidating the PBL processes using numerical model simulations (Martins *et al.,* 2010), even over the relatively homogeneous ocean (Belmonte and Stoffelen, 2019), which is likely due to the scarcity of fine-scale vertical observations of the atmosphere.

Over the oceans Belmonte and Stoffelen (2019) performed a climatological comparison between state-of-the-art reanalysis and scatterometer surface winds in the PBL, revealing mean and transient PBL model errors. Houchi *et al.* (2010), based on high-resolution radiosondes, verified the climatological wind profiles and found in particular a factor of 2–3 lower wind shear simulated by the European Centre for Medium-Range Weather Forecasts (ECMWF) model. Wind shear is recognized to be able to significantly modulate turbulent mixing of atmospheric pollutants (Zhang *et al.,* 2020b), and thus the inabilities of the model in this regard may have repercussions for air quality prediction.

The critical interaction between PBL turbulence and vertical structures of thermodynamic variables, as the heart of PBL physics, makes the determination of BLH a big challenge, due largely to the difficulty for those instruments with coarse vertical resolution in resolving the sharp gradients of temperature and water vapor at the top of the PBL, and estimating PBL-top entrainment and lateral entrainment (Teixeira *et al*., 2021). Thus, this highlights the importance of high-resolution vertical measurements of thermodynamic variables. The temporal and spatial variations in BLH have been extensively assessed in previous studies at a regional or national scale, such as the contiguous United States (Seidel *et al.,* 2012; Zhang *et al.,* 2020a), Europe (Palarz *et al*., 2018), Arctic and Antarctic (Zhang *et al*., 2011), which are mainly implemented by low-resolution radiosonde measurements, reanalysis or both. Fortunately, a few pioneering studies in characterizing BLH have adopted high-resolution measurements at a national scale over China (Guo *et al.* 2016; Zhang *et al*., 2018, Su *et al*., 2018) and United States (Seidel *et al*., 2010). Notable diurnal and seasonal cycles have been revealed (e.g., Guo *et al.,* 2016; Short *et al.,* 2019). Besides the regional results, several attempts have been made to provide global-scale retrievals of BLH using the Global Positioning System radio occultation (GPS RO) and Integrated Global Radiosonde Archive (IGRA) version 2 (Seidel *et al.,* 2010; Gu *et al.,* 2020; Ratnam and Basha, 2010), in which seasonal variations and maritime-continental contrasts of BLHs have been achieved. The measurements of GPS RO, at a vertical resolution of 100 m around the PBL top, are typically used to determine BLH by searching for the altitude with a sharp gradient in the refractivity profile (Basha *et al.,* 2018). However, such sharp gradient of refractivity might overestimate BLH compared to other methods that the community usually used, such as the parcel method (Seidel *et al.,* 2010). Compared with high-resolution soundings, IGRA is sparsely sampled in the vertical (about 10-30 layers below 500 hPa), which could result in large uncertainties in estimating BLH. Likewise, additional errors could be introduced in reanalysis products for their sparse vertical resolutions (about 6-42 layers below 500 hPa), which are equivalent to or bigger than IGRA. A large spread emerges in the explicit determination of BLH from a variety of instruments, in spite of that the BLH detection based on radiosonde is the most accepted methodology for deriving CBL and SBL (Seidel *et al.,* 2012; de Arruda Moreira *et al.,* 2018).

A wide range of reanalysis products, such as those from the fifth generation ECMWF atmospheric reanalysis of the global climate (ERA5), the National Aeronautics and Space Administration (NASA) Modern-Era Retrospective-analysis for Research and Applications version 2 (MERRA-2), Japanese 55-year Reanalysis (JRA-55), and the NCEP climate forecast system version 2 (NCEP-2), provide a rich ensemble of climate data products (Saha *et al.,* 2014; Hersbach *et al.,* 2020; Kobayashi *et al.,* 2015; Gelaro *et al.,* 2017), but are sensitive to both empirical parameterizations and the diagnostic method chosen, while verification by direct observations of BLH are sparse (Seibert *et al.,* 2000). Some inter-comparisons between instruments or model data, such as radiosonde, CALIOP, and ERA-interim reanalysis have been previously conducted, and a good consistency has been yielded in seasonal and spatial variation (e.g., Guo *et al.,* 2016; Zhang *et al.,* 2016). However, Basha *et al.* (2018) demonstrate that ERA-interim can underestimate BLH by around 900 m compared to GPS RO. This underestimation may be caused by different kinetic or thermodynamic assumptions use. For instance, ERA-interim is implemented with a bulk Richardson number method (Palm *et al.,* 2005), which is believed to be suitable for all atmospheric conditions (Anderson, 2009). It is worth highlighting that the state-of-art reanalysis could be one of the most promising data sources for obtaining the synoptic or climatological features of BLH.

Despite much progress made in developing the BLH products, there are still some unresolved issues in quantifying the variability of BLH from a global perspective. These issues include: the worldwide variation of BLH by high-resolution vertical soundings, the inter-comparisons among reanalysis datasets, and further evaluations with radiosonde observations, especially in the daytime based on the same retrieval algorithm. To this end, this study seeks to address the following scientific questions: (1) a climatological distribution of near-global BLH by using high-resolution radiosonde measurements; (2) inter-comparisons of ERA5, MERRA-2, JRA-55, and NCEP-2 with additional evaluation with radiosondes; and (3) investigate potential sources for the biases of BLH between observation and reanalysis. The rest of the paper is organized as follows. The descriptions of high-resolution radiosonde data, reanalysis products, and the bulk Richardson number method are given in Section 2. Section 3 presents the spatial distributions of BLH by radiosonde and reanalyses and their inter-comparisons. A brief conclusion and remarks are finally outlined in Section 4.

2. Data descriptions and BLH retrieval method

2.1 High-resolution radiosonde measurements

In 2018, IGRA provided atmospheric soundings at around 445 radiosonde sites across the globe, including pressure, temperature, humidity and wind vector. The number of pressure levels below 500 hPa is around 10-30. By comparison, for high-resolution radiosondes, the sampling rate is 1-s or 2-s, corresponding to a vertical resolution of approximately 5–10 meters throughout the atmosphere. The high-resolution radiosonde measurements used in the present study are obtained from 342 sites around the world, which are provided by several organizations, including the China Meteorological Administration (CMA), the National Oceanic and Atmospheric Administration (NOAA) of United States, the German Deutscher Wetterdienst (Climate Data Center), the Centre for Environmental Data Analysis (CEDA) of United Kingdom, the Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN), and University of Wyoming.

The CMA maintains the China Radiosonde Network (CRN), which contains 120 operational stations homogeneously distributed across mainland China with a vertical sampling rate of 1 second (5–8 m resolution), since 2011 (Guo *et al.,* 2016; 2019; Zhang *et al.,* 2016; 2018; Su *et al.,* 2020). The NOAA started the Radiosonde Replacement System (RRS) program in 2005, which involved 89 sites with a vertical resolution of 5 m (Zhang *et al.,* 2019). The German Deutscher Wetterdienst (Climate Data Center) has been sharing the radiosonde measurements at 14 sites with a sampling rate of 2 seconds since 2010. Moreover, the 10 m resolution soundings at 12 sites was provided by the CEDA, which began to share soundings since 1990, and 8 radiosonde sites were shared by GRUAN with a vertical resolution smaller than 10 m. An additional 93 sites came from the University of Wyoming, which started in 2017, with a sampling rate of 2-s or 1-s. In total, over 678,000 soundings at 342 stations are used here for the period of January 2012 to December 2019 in total of eight years, including 633,000 soundings at the regular release times of 0000 and 1200 UTC and 43,000 more irregular observations during intensive observation period (IOP).

Radiosonde measurements are taken twice per day following the World Meteorological Organisation (WMO) protocol for synoptic times at 0000 and 1200 UTC (Seibert *et al.,* 2000), except for special field campaign observations at specified stations or time ranges during IOPs. The protocol implies that stations at different longitudes sample the diurnal cycle differently. For instance, stations near 0°E (London) and 180°E (Samoa) sample at midnight and midday, while stations near 90°E (Bangladesh) and 90°W (Chicago) sample at dawn and dusk, with intermediate longitudes at linearly varying intermediate local solar times (LSTs) of day. For wintertime regions near 90°W and 90°E, the release times are insufficient for evaluating the BLH during daytime. Hence, the BLH estimates from regular radiosondes will vary with longitude and season (McGrath-Spangler and Denning, 2012). Generally, the principal PBL mechanism at night is associated with an SBL, which gradually transitions into CBL in the morning (Stull, 1988; Zhang *et al.,* 2018). The transition from SBL to CBL is generally quick and occurs swiftly after sunrise, but the reverse process can be slow in the late evening (Taylor *et al.,* 2014). Despite the dominance of CBL during the daytime, an SBL still occurs, especially in the event of overcast sky (Zhang *et al.,* 2018; 2020) and near strong divergence in moist convective downbursts (King *et al.,* 2017). To illustrate the daytime variation of BLH, we only selected the soundings that are launched 2 hours after sunrise and 2 hours before sunset. The sunrise and sunset times are gauged in a longitude bin size of 15 degrees and based on the latitude of station and the calendar day of the release. Using this definition, a total of 190,013 profiles including soundings launched at both synoptic times and during IOP, spanning January 2012 to December 2019, are used to obtain the BLHs in the daytime.

The spatial distribution of file number for each site is displayed in Figure S1, in which the sites with less than 10 matches are excluded.

2.2 ERA5, MERRA-2, JRA-55 and NCEP-2 reanalysis datasets

ERA5 is the successor of ERA-interim and has undergo a variety of improvements, including more recent parameterization schemes and data assimilation system, better spatial resolution, both horizontally and vertically (137 levels), and improved representation of evaporation balance, cyclones, soil moisture, and global precipitation (Hersbach *et al.,* 2020). The BLH is composited in the ERA5 product on a 1440×721 grids with 0.25° longitude and 0.25° latitude resolution. It is computed by the bulk Richardson number method, with a temporal resolution of 1 hour.

MERRA-2 is the latest atmospheric reanalysis of the modern satellite era produced by NASA's Global Modeling and Assimilation Office (GMAO). It includes aerosol data assimilation, improvements on ozone, and cryospheric processes (Gelaro *et al.,* 2017). In this product, the BLH is packaged and defined by identifying the lowest level at which the heat diffusivity drops below a threshold value (McGrath-Spangler and Denning, 2012). The formula for calculating BLH is as follows:

$$\mathrm{BLH(MERRA2\_packaged)} = 44308 \times (1 - \left(P_{PBLtop}/P_{Surface}\right)^{0.1903} \quad (1)$$

where $\mathrm{BLH(MERRA2\_packaged)}$ is in unit of meter, P_{PBLtop} the BLH (packaged parameter in MERRA-2, in unit Pa), and $P_{Surface}$ the surface pressure (in unit Pa). However, to preclude the uncertainty raised by different methods adopted, the BLH by MERRA-2 is extracted by bulk Richardson number method, by utilizing the parameters of horizontal wind, temperature, geopotential height, relative humidity (RH), and surface pressure as inputs. These input data are provided on a grid of 576×361 points with 0.625° longitude and 0.5° latitude resolution and has 42 pressure levels (about 16 layers below 500 hPa), with a temporal resolution of 3 h.

JRA-55 is the second Japanese global atmospheric reanalysis commissioned by the Japan Meteorological Agency (JMA) (Kobayashi *et al.,* 2015). Data contains 37 pressure levels between 1 hPa and 1000 hPa (16 layers below 500 hPa), provided on a grid of 288×145 points, with a horizontal spacing of $1.25°\times1.25°$ and a temporal resolution of 6 hours. The parameters, including geopotential height, temperature, horizontal wind, surface pressure, and RH, are used to assess BLH as before.

NCEP-2 has the coarsest model resolution than ERA5 (Rinke *et al.,* 2019), with a spatial resolution of 2.5° longitude and 2.5° latitude. The total level is 17 (6 layers below 500 hPa), which is substantially less than MERRA-2, JRA-55 or ERA5, and the temporal resolution is 6 hours. Similar parameters to JRA-55 are preserved to compute BLH. It is noteworthy that all model times include 0000 and 1200 UTC and hence collocate well with the synoptic radiosonde times.

2.3 Bulk Richardson number method

In the spirit of a like-for-like comparison, the BLHs derived from radiosonde and reanalysis data (MERRA-2, JRA-55, and NCEP-2) are calculated using the bulk Richardson number (BRN), which also serves as the built-in algorithm in ERA5 for BLH products. The BRN, an algorithm used to reflect how strongly buoyancy is coupled to the vertical momentum (Scotti, 2015), has been widely used for the climatological study of BLH from radiosonde measurements thanks to its applicability and reliability for all PBL regimes (Anderson 2009; Seidel *et al.,* 2012; Guo *et al.,* 2019). It determines the BLH by identifying the level at which the bulk Richardson number, represented by $Ri(z)$, reaches its critical value (Palm *et al*., 2005) and is formulated as:

$$\mathrm{Ri(z)} = \frac{\left(\frac{g}{\theta_{vs}}\right)(\theta_{vz}-\theta_{vs})z_{AG}}{(u_z-u_s)^2+(v_z-v_s)^2+(bu_*^2)} \tag{2}$$

where g is the gravitational acceleration, z_{AG} the height above ground level (AGL), θ_v the virtual potential temperature, u_* the surface friction velocity, and u and v the horizontal wind components and b a constant, which is usually set to zero due to the fact that friction velocity is much weaker compared with the horizontal wind (Seidel *et* *al.*, 2012). The subscripts of z and s denote the parameters at z height above ground and ground level, respectively.

It is known that $Ri(z)$ increases with increasing free flow stability (Zilitinkevich and Baklanov, 2002). Below a critical value of 0.25, the flow is dynamically unstable and likely cause turbulent motion. Nevertheless, since turbulence can also occur away from this critical value (Haack *et al.,* 2014), care must be taken in that the critical value might not be well defined, leading to uncertainty in estimating BLH. Meanwhile, the BLH estimates were found not to change very much by differing the input of critical values ($Ri = 0.2; 0.25; 0.3$) (Guo *et al.,* 2016). Therefore, for a given discrete Ri profile, here we identify the BLH as the interpolated height at which the $Ri(z)$ firstly crosses the critical value of 0.25 starting upward from the ground surface. Besides, it is well recognized that the vertical resolution of radiosonde measurement has large impact on the BLH estimated. For instance, BLHs are usually lower for a sparser vertical resolution (Seidel *et al*., 2012). Therefore, factors that cause uncertainty in estimating BLH by using the bulk Richardson method include, but not limited to, meteorological parameters, the surface friction, vertical resolution of data and the critical value of Ri.

2.4 Collocation procedure and a case study

In contrast to the reanalysis data, the longitude, and latitude distributions of high-resolution radiosonde are irregular. A precise comparison between reanalysis data and sounding is required for consistency in time, latitude, and longitude. The matching procedures implemented in this present study go as follows. (1) A latitudinal and longitudinal matching procedure is carried out by finding the geographical grid cell of the reanalysis product that contains the radiosonde station. (2) Time matching for ERA5 is to find the exact UTC time (hour) of the weather balloon launch. (3) For MERRA-2, NCEP-2, and JRA-55 datasets, the requirement is to limit the time difference with the weather balloon launch time to 1 hour.

A case at 0600 UTC 06 Jun 2016, Chongqing (29.6°N, 106.4°E, 541 m) is shown in Figure 1. In this case, BLH obtained by sounding is 1,337 m and is closest to that by

ERA5, which underestimates the height by 72 m. Compared with the radiosonde profile, MERRA-2 can capture the main vertical structures and the magnitude of wind speed (WS), RH, and temperature, but not the fine-scale vertical variations (Figure 1b). It also slightly undervalues the BLH by 125 m. The basic parameters outlined by NCEP-2, for instance, RH (5% larger than sounding), temperature (3°C less than sounding), and wind speed (4.5 m/s larger than sounding), all have notable differences with the sounding (Figure 1c). Eventually, The NCEP-2 derived BLH is considerably underestimated by 729 m. By and large, the profiles from JRA-55 are not as accurate as those from MERRA-2. More specifically, the wind speed at some heights, prominently above 2 km, is underestimated (Figure 1d); the mean RH is 4% less than that from the sounding. As a result, JRA-55 substantially underestimates BLH by 399 m. Based on this case, we can note that the performances of ERA5 and MERRA-2 are obviously better than those from JRA-55 and NCEP-2 in terms of the BLH. The remarkable underestimation by NCEP-2 can be attributed to the underestimations in near-surface virtual potential temperature (roughly 2.46 K less than sounding) and temperature. By comparison, the smaller BLH in JRA-55 could be attributed to the underestimated RH.

2.5 Normalized sensible/latent heat flux in the daytime

The sensible heat flux represents the level of energy that induces CBL growth (Wei *et al*., 2017), whereas the latent heat fluxes characterize the evaporation of moisture from the soil to the CBL, which feedbacks on the development of CBL and the formation of PBL cloud (Pal and Haeffelin, 2015). For a given amount of heat flux, small latent heat fluxes usually mean more energy being available for PBL growth (Chen *et al*., 2016). When less energy is constrained by the moist ground, more energy is available to heat the air. Moreover, the surface heat flux is closely associated with near-surface meteorological variables. For instance, a lower RH usually indicates a larger sensible heat flux and lower latent heat flux (Guo *et al*., 2019; Zhang *et al*., 2013). Suppose that the heat supplied to the air at the radiosonde balloon launch time is the area shaded under the heat flux curve (Fig.11.12 in Stull 1988), the normalized sensible heat flux in the daytime is defined by

$$\overline{Q_H} \propto \int_{T_{sunrise}}^{T_{launch}} Q_H \rho^{-1} c_p^{-1} dt \tag{3}$$

where $T_{sunrise}$ and T_{launch} are the sunrise time and radiosonde balloon launch time, Q_H the sensible heat flux, ρ the near-surface density and c_p equals 1004 J°C$^{-1}kg^{-1}$. The similar principle is applied to the calculation of normalized latent heat flux as well.

3. Results and discussion

3.1 Overview of BLHs at two synoptic times and over the day

The near-global mean BLHs at 0000 UTC from 2012 to 2019 by four reanalysis products are shown in Figure 2, in which the results obtained from radiosonde are overlaid by colored circles. The stations with sounding covering at least 2 continuous years are kept. The four reanalysis products yield an analogous result with respect to the spatial variation of BLHs, which are positively correlated with the sounding-derived BLH, with correlation coefficients of 0.90, 0.81, 0.47, 0.46 for ERA5, MERRA-2, NCEP-2, and JRA-55, respectively. It is evident that the BLHs from NCEP-2 over the continents of Africa, Asia, and South America are 300 m thicker than those of the other three products (Figure 2b). Furthermore, the BLH in Antarctic by ERA5 is notably 500 m lower than that by NCEP-2 and MERRA-2 (Figure 2a). Most of the mean BLHs by radiosonde are consistent with the reanalysis products, except that the values from all four reanalysis products over the Pacific Ocean and the contiguous U.S. are underestimated by about 300 m. Moreover, it is worth to note here that the BLHs by JRA-55 are considerably underestimated by around 1 km over these regimes. For 0000 UTC, the regions nearly from the east coast to the west coast of Pacific Ocean (UTC+8 to UTC+12, and UTC–12 to UTC–8) are covered by sunshine, and thus are filled with deeper PBL.

Comparable results at 1200 UTC are presented in Figure S2. Africa, the Middle East, and the west of India and China, corresponding to local noon and afternoon, have maximal BLHs of around 1.8 km. Moreover, it is noteworthy that the values from NCEP-2 and JRA-55 over these areas are visibly lower than those from ERA5 and MERRA-2, particularly over Africa and the Middle East, whereas these low values can barely be validated with soundings due to their sparse distribution. Over these areas, the BLHs are underestimated by reanalysis by about 200 m relative to the sounding results. Notably, BLHs from NCEP-2 over the continents of Africa are 1 km lower than those from ERA5 and MERRA-2. According to the results at 0000 and 1200 UTC, the comparisons between reanalysis products and soundings demonstrate that the BLHs are well resolved in the nighttime but are underestimated at daytime by reanalysis datasets.

For the near-global variation of BLH at a certain synoptic time, daytime and nighttime appear on the map simultaneously, but as a function of longitude, which is displayed in Figure 2. Thus, the variations at a fixed synoptic time on the map create a picture of the diurnal BLH variation. Given the dominance of CBL in the daytime, investigating the BLHs in the daytime is thus favorable for unravelling the underlying causes for the discrepancies existed in the BLHs from both radiosonde and reanalysis. Therefore, the following results show the variations of daytime BLH only, unless otherwise noted.

The climatological mean variations in the daytime BLH from the soundings and four reanalysis products are drawn in Figure 3. The period spans from January 2012 to December 2019 for most of the stations provided by China, the U.S., Germany, and the U.K. As implied by the results from soundings (Figure 3e), the deepest PBL is observed over the Tibetan Plateau (TP) and the northwest of China, the south of Africa, and the west of U.S, with values as high as 1.7 km. The possible reason for this phenomenon is that the weather balloons over these regions are basically launched in the early afternoon of boreal summer (June–July–August) when the maximal BLH is usually observed (Collaud Coen *et al.,* 2014; Guo *et al.,* 2016). The BLHs over the Pacific Ocean are noticeably large, with values of 1.3 km. The longitudinal variation of BLH is evident, likely due to LST variations of the soundings. Additionally, BLHs in the middle and low latitudes are larger than high latitudes, which is consistent with the findings in Gu *et al.* (2020).

By and large, the climatological results of BLH by radiosonde and four model products are comparable, indicating that both capture the spatial variations implied by the sounding LST times sampled. Among the model products, ERA5 shows the best prediction of BLH contrasted with radiosonde, with a correlation coefficient of 0.88 (Figure 3a). Furthermore, the results from MERRA-2 are positively correlated with those from the soundings, with a correlation coefficient of 0.66 (Figure 3b). The performances of JRA-55 and NCEP-2 are significantly poorer than those of ERA5 and MERRA-2, with correlation coefficients of 0.4 and 0.41, respectively (Figure 3c, d). The values of BLH over the west of U.S and the west of China are seriously underestimated by NCEP-2 and JRA-55 by around 800 m. Thus, we note that ERA5 and MERRA-2 are more robust in deriving the BLH, purely based on the climatological distribution of BLHs.

Figure 4 illustrates the diurnal variations in BLH at 0000 and 1200 UTC and during daytime. A notable diurnal variation can be noticed, with a minimum of 343 m at 0400 LST and a maximum of 1224 m at 17 LST (Figure 4a). The magnitude in BLH during daytime are essentially larger than that at 0000 and 1200 UTC and has a maximal value of 1926 m at 1700 LST (Figure 4b). It follows that most of soundings (about 78%) that are released at 0000 and 1200 UTC are excluded by the collocation procedure designed for collecting samples in the daytime. Note that the result during daytime will not significant change with/without IOP data.

3.2 Correlations with near-surface meteorological variables and surface heat flux

The PBL is the lowest part of the troposphere and evolves diurnally due to near-surface thermodynamic variables through turbulent exchanges of momentum, heat, and moisture (Pithan *et al*., 2015). Thus, the surface meteorological variables depend on the underlying land surface and its coupling with the PBL, and they could act as a good proxy for BLH under some specific circumstances (Zhang *et al*., 2013; Zhang *et al*., 2018). An analysis of the correlation between the BLHs by radiosondes and near-surface meteorological variables is presented in Figure 5. The variables include near-surface air temperature at 2 m AGL (T_{2m}), pressure (Ps), RH, and WS, which are extracted from the first level in sounding. The first level is assumed to be associated with the near-surface variables (Serreze *et al*., 1992; Wang and Wang 2016). We note that BLH, T_{2m}, RH and WS all have substantial diurnal and seasonal variability as partly expressed in Eq. (2).

Moderate positive (negative) correlation coefficients can be noticed between BLH and T_{2m} (RH), with mean values of 0.39/-0.51 (Figure 5a, c), implying that both T_{2m} and RH could be an adequate indicator for the temporal variation of BLH. Moreover, the correlations between BLH and WS are also positively notable, with a mean value of 0.24 (Figure 5d). By contrast, the correlation between Ps and BLH is negatively significant above most of the regions (Figure 5b).

The correlation analyses between BLH and normalized heat fluxes, which are assessed by ERA5 reanalysis products, are displayed in Figure 6. It is notable that positive/negative correlation coefficients usually exist in normalized sensible/latent heat flux, with a global mean of 0.29 and -0.31. This correlation is not high because BLH also depends on the radiative heating/cooling and the temperature profile in different stations (Yang *et al*., 2004).

For the climatological variation of BLH, the near surface variables such as T_{2m}, RH and WS, and the normalized sensible/latent heat flux could be a good indicator. Conversely, the development of BLH could also limit the magnitude of RH (McGrath-Spanglerm, 2016).

3.3 Comparisons with reanalysis products

The radiosonde stations are mainly dispersed over the U.S, China, Australia, Europe, the Pacifica Ocean, and the polar region, and only a few stations contribute over the rest of the world. The polar region contains a station with a latitude larger/lower than 67.7°N/°W. Therefore, six regions are specifically examined in terms of the bias between radiosonde and model product.

The BLH differences between ERA5 and radiosonde are shown in Figure 7, in which we specify the differences over the six above-mentioned regions. As observed in Figure 7e, the BLH over most of the stations is underestimated to a slight extent, with a near-global mean of 131.96 m. As expected, the most underestimated regions cover the west of U.S, and southern China (Figure 7e), with a difference of around 200 m. In addition, it is worth mentioning that the BLHs over the Pacific Ocean are overestimated in four seasons, with a bias of around 400 m (Figure 7h). Among the six classified regions, BLHs in Europe, East Asia, and polar are reliably determined by ERA5, with an average bias of around 50 m (Figure 7b, c, i). The bias seems to exhibit a seasonal dependence, and it is around 62 m larger in the warm seasons compared to cool seasons in both hemispheres. Regardless of the small bias, the newest model product, ERA5, properly estimates the BLH, especially above the regions of Europe, the eastern U.S, East Asia, and polar.

Similarly, the BLHs by MERRA-2 are underestimated, with a near-global mean bias of 166.35 m (Figure 8), which is slightly larger than that of ERA5 (131.96 m). This could indicate that the MERRA-2-derived BLH is more dispersed than ERA5. The spatial distribution of bias value is broadly identical to that of ERA5, except that the BLHs over Europe, Australia, and polar region are well estimated by MERRA-2, due to much smaller mean biases at 42.78 m, 52.98 m, and 66.20 m, respectively (Figure 8b, g, i).

In addition, the packaged BLH in MERRA-2 is also evaluated with radiosonde. BLH is as high as 3 km over the TP region at 0600 UTC (Figure S3), corresponding to an overestimation of 0.8 km over this region (Figure S4). Over the rest regions, BLH is slightly or moderately overestimated by around 50 m. However, The BLH difference among various methods could reach up to a kilometer or even more (Seidel *et al*., 2010), which is probably owing to the variety of kinetic or thermodynamic theories applied in different algorithms.

By comparison, the mean bias produced by JRA-55 is larger than those from ERA5 and MERRA-2, with a mean value of 351.49 m, as shown in Figure 9. The BLHs above most stations are underestimated by JRA-55, particularly for the sites over western China and western U.S, and the Pacific Ocean, with an underestimation of about 800 m. The most underestimated stations cluster at the latitude range of 40–45°N, with a mean difference of around 1 km (Figure 9f). Although the near-global mean of bias is significantly larger than ERA5 and MERRA-2, the estimations over Europe and the polar regions seem to be more in line with the observations , with mean values of 174.99 m and 93.84 m, respectively (Figure 9b, i).

The mean bias by NCEP-2 is larger than that by JRA-55, with a mean value of 420.86 m, as illustrated by Figure 10. The distribution results are similar to JRA-55, except for Europe and Australia, where the bias is about twice that of JRA-55.

In general, the comparison analysis of the daytime BLH results between soundings and four reanalysis datasets indicates that ERA5 reanalysis produces the BLH that is closest to the high-resolution soundings. Interestingly, MERRA-2 can provide a good spatial distribution of BLH. JRA-55 and NCEP-2 can only give a good prediction over some regions, most of which tends to produce a much larger BLH estimates compared to those from ERA5 and MERRA-2.

3.4 Potential sources for the bias between reanalysis products and radiosonde

The possible sources for the difference between radiosonde and reanalysis could be rather complicated. From the spatial pattern of BLH discrepancy results between radiosonde and reanalysis (Figures 7–10), we can notice that the regions with large differences tend to be observed over regions with high elevation, such as the TP in China and Rocky mountain in the U.S. These regions generally have much more complex orography. Coincidently, the soundings over the above-mentioned two regions are all obtained from afternoon, in which the PBL develops to the maximum (Figure 4).

As expected, highest biases generally are accompanied with peak BLHs, which has also been confirmed in our previous studies (cf. Figure 2c in Li *et al*., 2017). Therefore, the biases depend on the LST when the weather balloon is launched, which at least could not be ruled out.

In addition, the large differences primarily appear in the low and middle latitudes, where thermal convection frequently occurs. Therefore, it is reasonable to infer that static stability could exert an influence on the comparison results. Then, we will analyze the probable influences from terrain and static stability on BLH differences.

We evaluate the influence from the orographic complexity around the sounding station and calculate the standard derivation (STD) of elevation within 1°x1° grid, with the help of 30 arc second digital elevation model (DEM) dataset. Terrain is complex over the western China and western US where most of soundings are released in afternoon and large BLH biases are usually found. Therefore, for all soundings that are launched during the time period spanning from 1300 LST to 1800 LST we analyze the relationship between BLH biases and the standard derivation of the DEM (Figure 11). It follows that the influence from the orography appears instrumental, given the correlation coefficient varying from -0.84 to -0.95. Furthermore, the errors or uncertainties in ERA5 are less easily impacted by the orographic complexity given a much flatter fitted line (Figure 11a).

Based on the correlation between orographic complexity (manifested by the STD of the DEM) and the bias of a reanalysis relative to radiosonde measurements, it is likely that the performances of MERRA-2, JRA-55, and NCEP-2 might be restricted by the complex underlying terrains. One of the reasons could be because global reanalysis with coarse resolution that cannot resolve the sub-grid processes due to topography. However, ERA5 appears to be less dependent on terrain. In other words, the models used in ERA5 show sufficient capability and excellent performance in reproducing the atmospheres, particularly in the PBL over complex terrains.

Lower tropospheric stability (LTS) is an indicator to describe the thermodynamic state of the lower atmosphere and is defined by the differences in potential temperature at 700 hPa and 1000 hPa (Guo *et al*., 2016). Typically, the smaller the LTS, the more unstable the low troposphere. The mean LTS over each station is defined by the ensemble mean by four reanalysis datasets, and its spatial distribution is depicted in Figure 12. The lower troposphere over the western United States and western China is more unstable compared to the rest of the world, with LTS of around 6K (Figure 11a), which is likely associated with afternoon launch time of weather balloons. According to the correlation between the bias of BLH and the mean LTS, it is clear that the underestimation in BLH by JRA-55 and NCEP-2 products are negatively correlated with LTS, with correlation coefficients of 0.32 and 0.36 (Figure 12b).

Besides the LTS, the role of lifted index could be another influential factor. The lifted index is a predictor of latent instability (Galway, 1956), and it is defined as the temperature difference between the environment temperature and an air parcel lifted adiabatically at 500 hPa. The index is computed by the air temperature, RH, and pressure profiles from radiosondes. We calculate the percentage of negative lifted index above each station, which represents the occurrence rate of latent instability that exists in the daytime (Figure 12c). The stations with high probability of strong instability, denoted by $\mathrm{P(lifted\ index} < 0)$, are predominantly dispersed over the west U.S, the west and south of China, and the Pacific Ocean, reaching a percentage as high as around 70%. These stations are regularly overlapped with great biases in the reanalysis products as shown in Figures 7–10. According to the analysis, it is clear that all four reanalysis products are positively associated with $\mathrm{P\ (lifted\ index} < 0)$, with correlation coefficients ranging from -0.34 to -0.47 (Figure 12d). The positive (negative) correlation coefficients in lifted index suggests that the underestimation by reanalysis might be associated with the instability activity in the lower troposphere that has not been adequately represented or simulated by the models used in reanalyses. In light of the surface heating during the day and the growth of the PBL due to air ascent, it is also inferred that afternoon BLHs suffer the greatest errors if this is caused by inadequate air mixing within the free troposphere in models.

4. Conclusions and summary

A climatology of near-global BLH from high-resolution radiosonde measurements has been yielded for the daytime BLH. The high-resolution radiosonde data has a much finer spatial resolution of 5 m or 10 m, compared to that by IGRA, and can establish a finer and more precise structure of the PBL. In addition, direct comparisons among four well-established reanalysis model products have been conducted. The present study adopts over 300 sounding stations with high-resolution, spanning from 2012 to 2019, to investigate the climatological variation of near-global BLH in the daytime and evaluates four model products at the radiosonde sampling.

Notable spatial variation can be observed in the climatological mean of BLH at 0000 and 1200 UTC. In the afternoon, the regions over the Western United States and Western China have the largest BLHs with values as high as 1.7 km, whereas 0000 and 1200 UTC compare generally to earlier times of day (LST) in the rest of the world with hence lower BLH. In addition, BLHs in the middle and low latitudes are larger than those in high latitudes. The T_{2m} and RH, and the normalized sensible/latent heat flux are a good predictor for the spatio-temporal evolution of BLH. The most important result is we found that all the four reanalysis products generally underestimate the daytime BLH, with a near-global mean varying from around 132 m to 420 m. The largest bias in reanalysis appears over the Western United States and Western China, where the boundary layers grow vigorously in the afternoon. ERA5 and MERRA-2 definitely have better performance than JRA-55 and NCEP-2 in terms of the magnitude of BLH and a higher correlation coefficient with the soundings. The newest version of reanalysis, ERA5, has the smallest bias and the highest positive correlation relative to radiosondes. The underestimation by NCEP-2 and JRA-55 is robust over some regions, for instance, western China and western U.S, with differences even exceeding 800 m.

However, all products can obtain a precise estimate over some regions, for instance, Europe, the eastern U.S, and polar, probably due to morning LST soundings and smaller daytime PBL development. The BLH over the Pacific Ocean is underestimated in all seasons and by all products. The underestimation tends to have a seasonal dependence, i.e., the warm season has a larger underestimation. However, BLH is moderately overestimated by the packaged BLH parameter in MERRA-2, possibly due to different BLH-deriving methods used.

We investigated two possible sources contributing to the biases, including topography and static stability. The analysis shows that the DEM spread does have a negative correlation with the bias, suggesting that the reanalysis data cannot provide a reliable simulation result under complex terrain conditions. In addition, reanalysis BLH errors tends to be negatively correlated with the occurrence rate of unstable air, suggesting that the reanalyses do not accurately determine BLH when the lower troposphere is unstable.

Although this study suffers from the inhomogeneous distribution of the radiosonde sites, the climatological BLHs at the near-global scale can help us understand the variation characteristics of BLH in different regions and for different LST. For the first time, we present near-global BLH estimates from high-resolution radiosondes, and further conduct a comprehensive comparison of BLH products for four widely used reanalysis datasets using the BLHs derived from the soundings. The findings provide insights into the limitations of reanalysis data and, more importantly, are expected to greatly benefit future research works related to applications of different kinds of reanalysis data in the future.

Acknowledgements

This study is jointly supported by the National Key Research and Development Program of the Ministry of Science and Technology of China under grant

2017YFC1501401, the National Natural Science Foundation of China under grant 41771399, 41531070 and 41874177, and S&T Development Fund of CAMS (2021KJ008). The authors would like to acknowledge the National Meteorological Information Centre (NMIC) of CMA, NOAA, German Deutscher Wetterdienst (Climate Data Center), U.K Centre for Environmental Data Analysis (CEDA), GRUAN, and the University of Wyoming (http://data.cma.cn/en, ftp://ftp.ncdc.noaa.gov/pub/data/ua/data/1-sec/, https://cdc.dwd.de/portal/, https://catalogue.ceda.ac.uk/,ftp://ftp.ncdc.noaa.gov/pub/data/gruan/processing/level2/RS92-GDP/version-002/, http://weather.uwyo.edu) for providing the high-resolution sounding data. We would like to thank the ECWMF for ERA5 data (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form), GMAO for MERRA-2 (https://disc.gsfc.nasa.gov/datasets?keywords=MERRA-2&page=1), NCAR and Japan Meteorological Agency for JRA-55 (https://climatedataguide.ucar.edu/climate-data/jra-55), NOAA for NCEP-2 (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html). NASA for 30 arc second digital evaluation height (DEM) data (https://search.earthdata.nasa.gov/).

References

Anderson, P. S.: Measurement of Prandtl number as a function of Richardson number avoiding self-correlation, *Bound-Layer Meteorol.,* 131, 345–362, https://doi.org/10.1007/s10546-009-9376-4, 2009.

Ao, C. O., Waliser, D. E., Chan, S. K., Li, J.-L., Tian, B., Xie, F., and Mannucci, A. J: Planetary boundary layer heights from GPS radio occultation refractivity and humidity profiles, *J. Geophys. Res. Atmos.,* 117(D16), https://doi.org/10.1029/2012JD017598, 2012

Basha, G., and Ratnam, M. V.: Identification of atmospheric boundary layer height over a tropical station using high resolution radiosonde refractivity profiles:

Comparison with GPS radio occultation measurements, *J. Geophys. Res.-Atmos.,* 114, D16101, https://doi.org/10.1029/2008JD011692, 2009.

Basha, G., Kishore, P., Ratnam, M. V., Ravindra Babu, S., Velicogna, I., Jiang, J. H., and Ao, C. O.: Global climatology of planetary boundary layer top obtained from multi-satellite GPS RO observations, *Clim. Dynam.,* 52, 2385–2398. https://doi.org/10.1007/s00382-018-4269-1, 2018

Belmonte Rivas, M. and Stoffelen, A.: Characterizing ERA-Interim and ERA5 surface wind biases using ASCAT, *Ocean Sci.,* 15, 831–852, https://doi.org/10.5194/os-15-831-2019, 2019.

Chen, X., Škerlak, B., Rotach, M. W., Añel, J. A., Su, Z., Ma, Y., and Li, M.: Reasons for the extremely high-ranging planetary boundary layer over the western Tibetan Plateau in winter, *J. Atmos. Sci.,* 2021–2038, https://doi.org/10.1175/JAS-D-15-0148.1, 2016.

Collaud Coen, M., C. Praz, A. Haefele, D. Ruffieux, P. Kaufmann, and Calpini., B.: Determination and climatology of the planetary boundary layer height by in-situ and remote sensing methods as well as the COSMO model above the Swiss plateau, *Atmos. Chem. Phys.,* 14, 15,419–15,462, https://doi.org/10.5194/acp-14-13205-2014, 2014.

Davy, R., and I. Esau: Differences in the efficacy of climate forcings explained by variations in atmospheric boundary layer depth, *Nat. Commun.,* 7, 11690, https://doi.org/10.1038/ ncomms11690, 2016.

de Arruda Moreira, G., J. L. Guerrero-Rascado, J. A. BravoAranda, et al.: Study of the planetary boundary layer by microwave radiometer, elastic lidar and Doppler lidar estimations in Southern Iberian Peninsula, *Atmos. Res.,* 213, 185–195, https://doi.org/10.1016/j.atmosres.2018.06.007, 2018.

Galway, J. G.: The lifted index as a predictor of latent instability, *Bull. Am. Meteorol. Soc.,* 37, 528–529, 1956

Gelaro R, et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), *J. Climate*, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Gu, J., Zhang, Y. H., Yang, N., and Wang, R.: Diurnal variability of the planetary boundary layer height estimated from radiosonde data, *Earth Planet. Phys.,* 4(5), 479–492, http://doi.org/10.26464/epp2020042, 2020.

Guo, J., et al.: The climatology of planetary boundary layer height in China derived from radiosonde and reanalysis data, *Atmos. Chem. Phys.,* 16(20), 13309–13319. https://doi.org/10.5194/acp-16-13309-2016, 2016.

Guo, J., et al.: Shift in the temporal trend of boundary layer height trend in China using long-term (1979–2016) radiosonde data, *Geophys. Res. Lett.,* 46 (11): 6080-6089, doi: 10.1029/2019GL082666, 2019.

Guo, J., et al.: The climatology of lower tropospheric temperature inversions in China from radiosonde measurements: roles of black carbon, local meteorology, and large-scale subsidence, *J. Climate*, 9327–9350, https://doi.org/10.1175/JCLI-D-19-0278.1, 2020.

Haack, A., Gerding, M., and Lübken, F.-J.: Characteristics of stratospheric turbulent layers measured by LITOS and their relation to the Richardson number, *J. Geophys. Res.-Atmos.,* 119, 10,605–10,618. https://doi.org/10.1002/2013JD021008, 2014.

Hersbach, Hans, et al.: The ERA5 global reanalysis, *Q. J. Roy. Meteor. Soc.,* 146(730), 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Houchi, K., Stoffelen, A., Marseille, G. J., and De Kloe, J.: Comparison of wind and wind shear climatologies derived from high-resolution radiosondes and the ECMWF model, *J. Geophys. Res.-Atmos.,* 115, D22123, https://doi.org/10.1029/2009JD013196, 2010.

King, G. P., Portabella, M., Lin, W., Stoffelen, A.: Correlating extremes in wind and stress divergence with extremes in rain over the Tropical Atlantic, EUMETSAT Ocean and Sea Ice SAF Scientific Report OSI_AVS_15_02, Version 1.0, 2017.

Kobayashi, et al.: The JRA-55 reanalysis: General specifications and basic characteristics, *J. Meteor. Soc. Japan*, 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

Li, H., Y. Yang, X.-M. Hu, Z. Huang, G. Wang, B. Zhang, and Zhang, T.: Evaluation of retrieval methods of daytime convective boundary layer height based on lidar data, *J. Geophys. Res.-Atmos.,* 122, 4578–4593, https://doi.org/10.1002/2016JD025620, 2017.

Liu, S., and Liang, X.-Z.: Observed diurnal cycle climatology of planetary boundary layer height, *J. Climate*, 23(21), 5790–5809. https://doi.org/10.1175/2010JCLI3552.1, 2010

Martins, J. P. A., J. Teixeira, P. M. M. Soares, P. M. A. Miranda, B. H. Kahn, V. T. Dang, F. W. Irion, E. J. Fetzer, and Fishbein, E.: Infrared sounding of the trade-wind boundary layer: AIRS and the RICO experiment*, Geophys. Res. Lett*., 37, L24806, https://doi.org/10.1029/2010GL045902, 2010.

McGrath-Spangler, E. L.: The impact of a boundary layer height formulation on the GEOS-5 model climate, *J. Geophys. Res.-Atmos.,* 121, 3263–3275, https://doi.org/10.1002/2015JD024607, 2016.

McGrath-Spangler, E. L., and Denning, A. S.: Estimates of North American summertime planetary boundary layer depths derived from space-borne lidar, *J. Geophys. Res.-Atmos.*, 117, D15101, https://doi.org/10.1029/2012JD017615, 2012.

Oliveira, M. I. et al.: Planetary boundary layer evolution over the Amazon rainforest in episodes of deep moist convection at the Amazon Tall Tower Observatory, *Atmos. Chem. Phys.*, 20, 15–27, https://doi.org/10.5194/acp-20-15-2020, 2020.

Palarz, A., Celiński-Mysław, D., and Ustrnul, Z.: Temporal and spatial variability of surface-based inversions over Europe based on ERA-Interim reanalysis, *Int. J. Climatol.*, 38(1), 158–168, https://doi.org/10.1002/joc.5167, 2018.

Pal, S., and M. Haeffelin, M.: Forcing mechanisms governing diurnal, seasonal, and interannual variability in the boundary layer depths: Five years of continuous lidar observations over a suburban site near Paris, *J. Geophys. Res.-Atmos.,* 120, 11,936–11,956, https://doi.org/10.1002/2015JD023268, 2015.

Palm, S. P., A. Benedetti, and Spinhirne, J.: Validation of ECMWF global forecast model parameters using GLAS atmospheric channel measurements, *Geophys. Res. Lett.*, 32, L22S09, https://doi.org/10.1029/2005GL023535, 2005.

Pithan, F., Angevine, W., and Mauritsen, T.: Improving a global model from the boundary layer: total turbulent energy and the neutral limit Prandtl number, *J. Adv. Model. Earth. Syst.,* 7, 791–805, https://doi.org/10 .1002/2014MS000382, 2015.

Ratnam, M. V., Basha, G.: A robust method to determine global distribution of atmospheric boundary layer top from COSMIC GPS RO measurements, Atmos. Sci. Lett., 11, 216–222, https://doi.org/10.1002/asl.277, 2010.

Rinke, A., Segger, B., Crewell, S., Maturilli, M., Naakka, T., Nygård, T., Vihma, T., Alshawaf, F., et al.: Trends of vertically integrated water vapor over the arctic during 1979-2016: Consistent moistening all over?, *J. Climate*, 32(18), 6097–6116, https://doi.org/10.1175/JCLI-D-19-0092.1, 2019.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y. T., Chuang, H.Y., Iredell, M. and Ek, M.: The NCEP climate forecast system version 2, *J. Climate*, 27(6), 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1, 2014.

Scotti, A.: Biases in Thorpe-scale estimates of turbulence dissipation. Part II: energetics arguments and turbulence simulations, *J. Phy. Oceanog.,* 45(10), 2522–2543, https://doi.org/10.1175/JPO-D-14-0092.1, 2015.

Seibert, P., Beyrich, F., Gryning, S. E., Joffre, S., Rasmussen, A., and Tercier, P.: Review and inter-comparison of operational methods for the determination of the mixing height, *Atmos. Environ.,* 34, 1001–1027, https://doi.org/10.1016/S1352-2310(99)00349-0, 2000.

Seidel, D. J., Ao, C. O., and Li, K.: Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis, *J. Geophys. Res.-Atmos.,* 115(D16), https://doi.org/10.1029/2009JD013680, 2010.

Seidel, D. J., Zhang, Y., Beljaars, A., Golaz, J.-C., Jacobson, A. R., and Medeiros, B.: Climatology of the planetary boundary layer over the continental United States and Europe, *J. Geophys. Res.-Atmos.,* 117(D17), https://doi.org/10.1029/2012JD018143, 2012.

Serreze, M. C., J. A. Maslanik, M. C. Rehder, R. C. Schnell, J. D. Kahl, and E. L. Andreas, E. L.: Theoretical heights of buoyant convection above open leads in the winter Arctic pack ice cover*, J. Geophys. Res.-Atmos.*, 97, 9411–9422, 1992.

Short, E., Vincent, C. L., & Lane, T. P: Diurnal cycle of surface winds in the Maritime Continent observed through satellite scatterometry, *Mon. Weather. Rev.,* 147(6), 2023–2044, https://doi.org/10.1175/MWR-D-18-0433.1, 2019.

Stull, R. B.: An Introduction to Boundary Layer Meteorology. Kluwer Academic, 666 pp, Dordrecht, the Netherlands, 1988.

Su, T., Li, Z., and Kahn, R.: Relationships between the planetary boundary layer height and surface pollutants derived from lidar observations over China: regional pattern and influencing factors, *Atmos. Chem. Phys.,* 18, 15921–15935, https://doi.org/10.5194/acp-18-15921-2018, 2018.

Su, T., Li, Z., Zheng, Y., Luan, Q., and Guo, J.: Abnormally shallow boundary layer associated with severe air pollution during the COVID-19 lockdown in China, *Geophys. Res. Lett.,* 47(20), https://doi.org/10.1029/2020GL090041, 2020.

Taylor, A. C., Beare, R. J., and Thomson, D. J.: Simulating dispersion in the evening-transition boundary layer, *Bound-Layer Meteorol.,* 153, 389–407, https://doi.org/10.1007/s10546-014-9960-0, 2014.

Teixeira, J., J. R. Piepmeier, A. R. Nehrir, C. O. Ao, S. S. Chen, C. A. Clayson, A. M. Fridlind, M. Lebsock, W. McCarty, H. Salmun, J. A. Santanello, D. D. Turner, Z. Wang, and X. Zeng: Toward a global planetary boundary layer observing system: the NASA PBL incubation study team report. NASA PBL Incubation Study Team. 134 pp, 2021.

von Engeln, A., and Teixeira, J.: A planetary boundary layer height climatology derived from ECMWF reanalysis data, *J. Climate*, 26(17), 6575–6590, https://doi.org/10.1175/JCLI-D-12-00385.1, 2013.

Wang, X., and Wang, K.: Homogenized variability of radiosonde-derived atmospheric boundary layer height over the global land surface from 1973 to 2014. *J. Climate*, 29, 6893–6908, https://doi.org/10.1175/JCLI-D-15-0766.1, 2016.

Wei, N., Zhou, L., and Dai, Y.: Evaluation of simulated climatological diurnal temperature range in CMIP5 models from the perspective of planetary boundary layer turbulent mixing, *Clim. Dynam.,* 49, 1–22, https://doi.org/10.1007/s00382-016-3323-0, 2017.

Yang, K., T. Koike, H. Fujii, T. Tamura, X. Xu, L. Bian, and Zhou, M.: The Daytime Evolution of the Atmospheric Boundary Layer and Convection over the Tibetan Plateau: Observations and Simulations, J. Meteorol.Soc.Jpn., 82 (6), 1777-1792, 2004.

Zhang, Y., Sun, K., Gao, Z., Pan, Z., Shook, M. A., and Li, D.: Diurnal climatology of planetary boundary layer height over the contiguous United States derived from AMDAR and reanalysis data, *J. Geophys. Res.-Atmos.,* 125, https://doi.org/10.1029/2020JD032803, 2020a.

Zhang, Y., J. Guo, Y. Yang, Y. Wang, and S.H.L. Yim: Vertical wind shear modulates particulate matter pollutions: A perspective from Radar wind profiler observations in Beijing, China, *Remote Sens.,* 12(3), 546. https://doi.org/10.3390/rs12030546, 2020b.

Zhang, W., Guo, J., Miao, Y., Liu, H., Li, Z., and Zhai, P.: Planetary boundary layer height from CALIOP compared to radiosonde over China, *Atmos. Chem. Phys.,* 16, 9951–9963, https://doi.org/10.5194/acp-16-9951-2016, 2016.

Zhang, W., Guo, J., Miao, Y., Liu, H., Song, Y., Fang, Z., He, J., Lou, M., Yan, Y., Li, Y., and Zhai, P.: On the summertime planetary boundary layer with different thermodynamic stability in China: A radiosonde perspective, *J. Climate*, 31(4), 1451–1465, https://doi.org/10.1175/JCLI-D-17-0231.1, 2018.

Zhang, J., Zhang, S. D., Huang, C. M., Huang, K. M., Gong, Y., Gan, Q., and Zhang, Y. H.: Latitudinal and topographical variabilities of free atmospheric turbulence from high-resolution radiosonde data sets, *J. Geophys. Res.-Atmos.,* 124, 4283–4298, https://doi.org/10.1029/2018JD029982, 2019.

Zhang, Y., D. J. Seidel, J.-C. Golaz, C. Deser, and Tomas, R. A.: Climatological characteristics of Arctic and Antarctic surface-based inversions, *J. Climate*, 24, 5167–5186, https://doi.org/10.1175/2011JCLI4004.1, 2011.

Zhang, Y. H., Seidel, D. J., and Zhang, S. D.: Trends in planetary boundary layer height over Europe, J. Climate, 26(24), 10,071–10,076, https://doi.org/10.1175/JCLI-D-13-00108.1, 2013.

Zilitinkevich, S., and Baklanov, A.: Calculation of the height of the stable boundary layer in practical applications, *Bound-Layer Meteorol.,* 105(3), 389–409. https://doi.org/10.1023/A:1020376832738, 2002.

Figures:

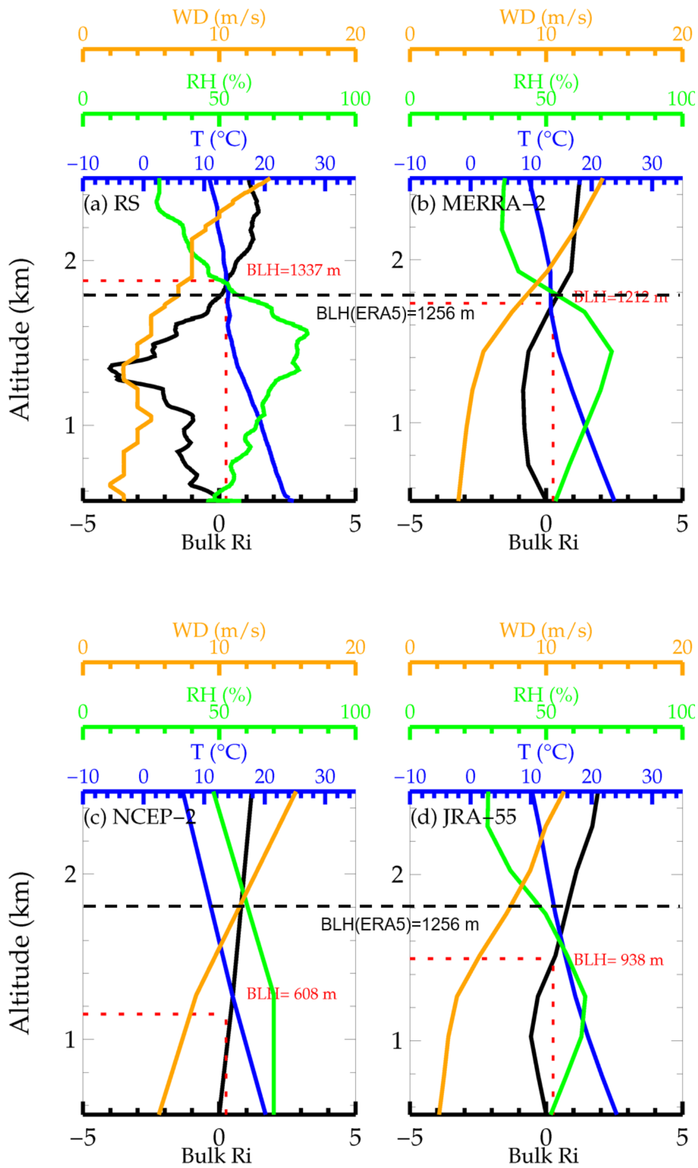

Figure 1. Profiles of basic atmomospheric parameters from the ground up to 2.5 km AGL, including wind speed (orange), bulk Ri (black), temperature (blue), and RH (green) at 0600 UTC (1400 LST) 06 Jun 2016 at Chongqing (29.6°N, 106.4°E, 541 m) from radiosonde (a), MERRA-2 (b), NCEP-2 (c), and JRA-55 (d) reanalysis datasets. The boundary layer height (BLH) in each subplot is marked as red dash lines and red texts, and the BLH for ERA5 is 1265 m in this case (black dash lines).

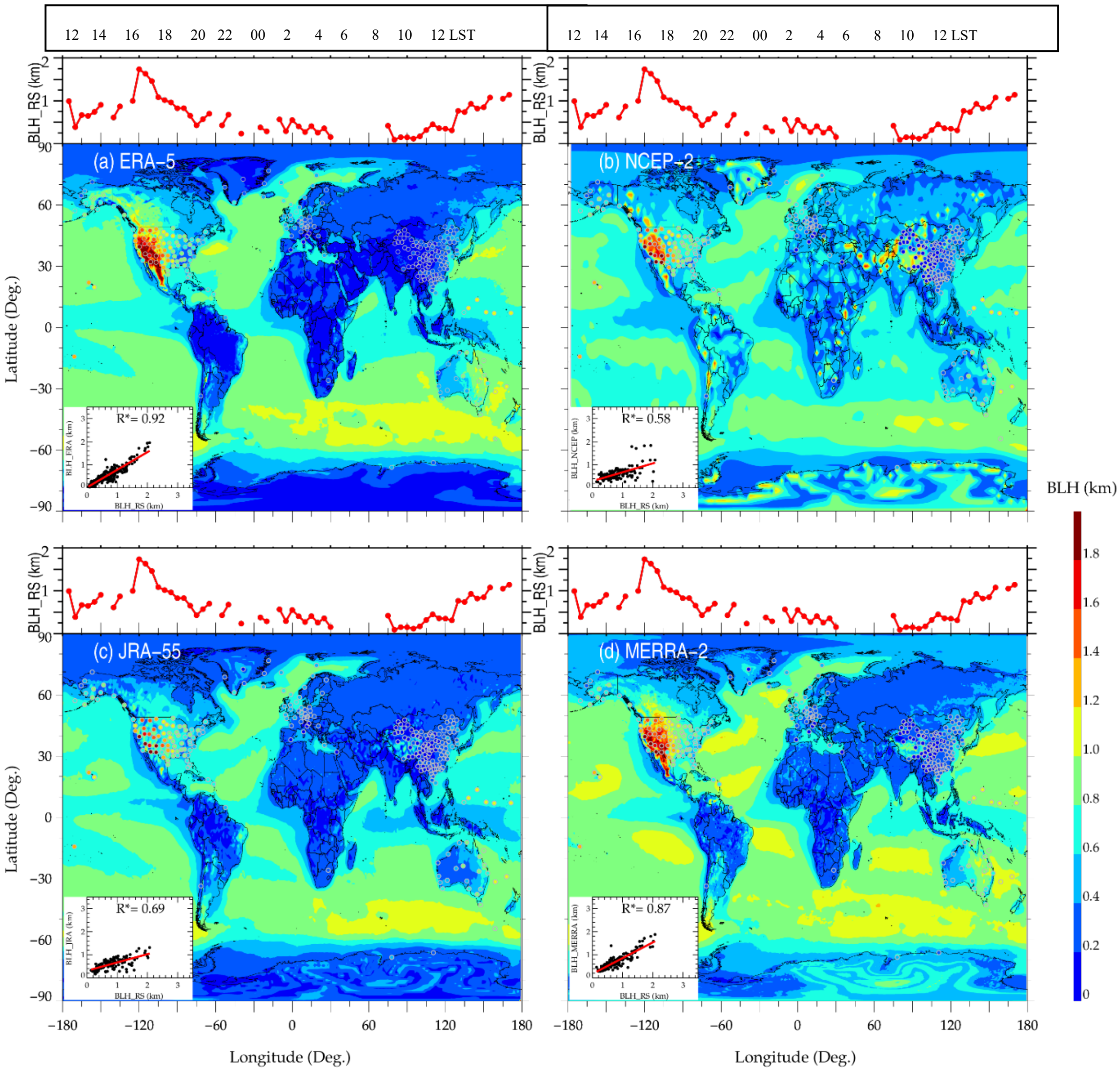

Figure 2. The mean BLH estimated from ERA5 (a), NCEP-2 (b), JRA-55 (c), and MERRA-2 (d) reanalysis data at 0000 UTC during years 2012 − 2019. The dots with gray marginal lines in each map denote the mean BLH derived by sondes at 0000 UTC, and the red dotted lines present the mean BLH derived by radiosonde on a grid with 5° longitude. Stations with less than 10 profiles are not included in the analysis. The 2D scatter plot in the left bottom corner of each panel illustrates the correlations between reanalysis-derived and sonde-derived BLHs at 0000 UTC, where the asterisk (*) superscripts indicate that the correlation coefficients are statistically significant ($p<0.05$) and the red lines denote the least-squares regression line.

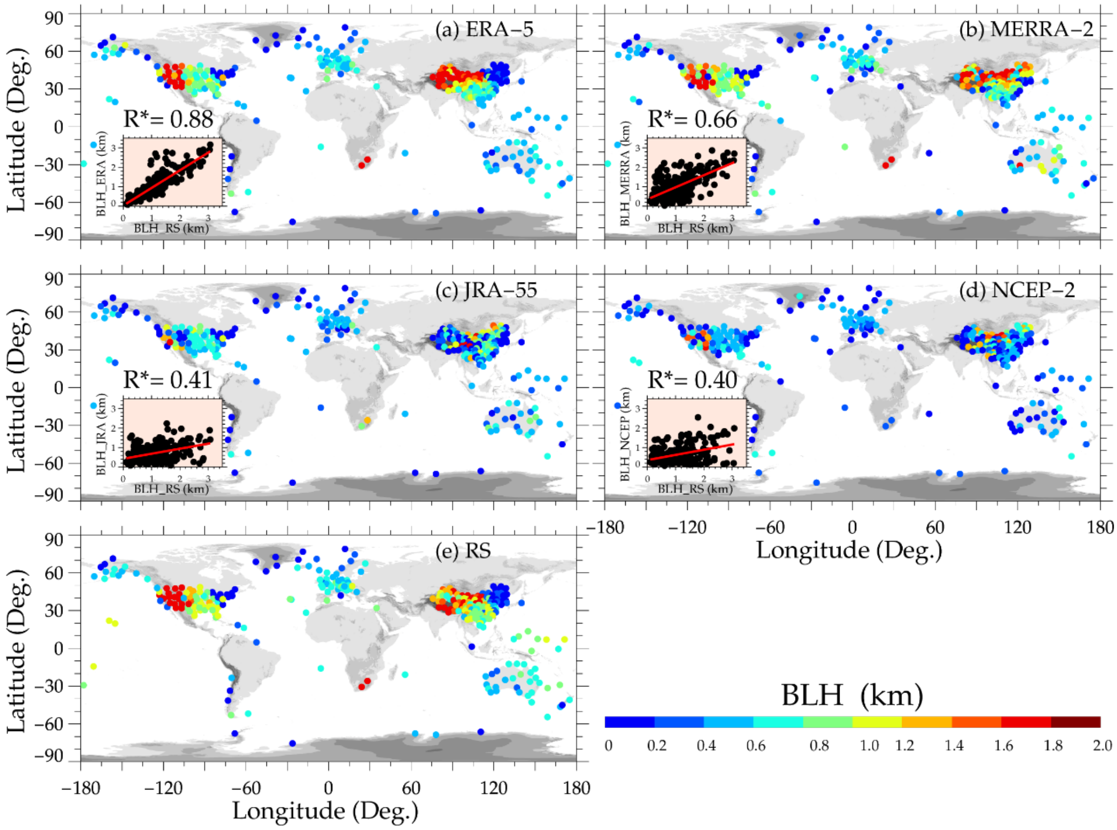

Figure 3. Spatial distributions of the mean BLHs determined at the near-global high-resolution radiosonde observational network locations during the daytime for the period 2012 to 2019, which is extracted from ERA5 (a), MERRA-2 (b), JRA-55 (c), NCEP-2 (d), and radiosonde measurements (e), respectively. Similar to Figure 2, the scatter plot illustrates the correlations between reanalysis-derived and sonde-determined BLHs in the daytime.

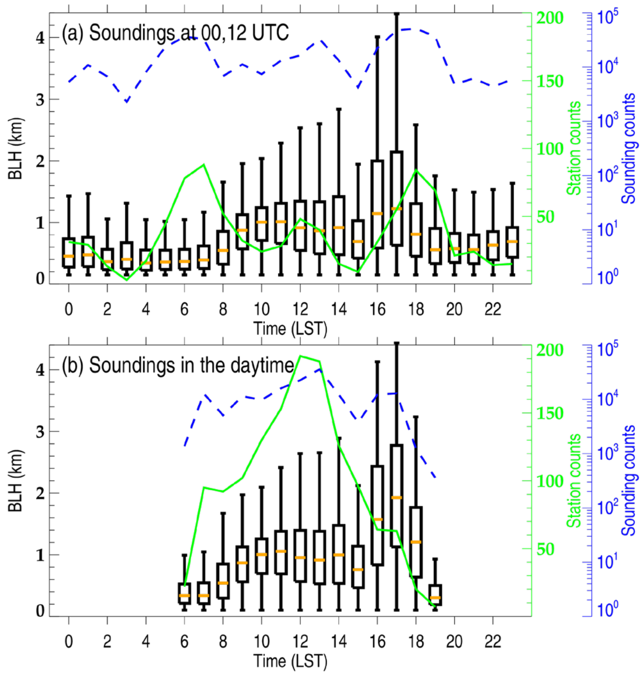

Figure 4. Box and whisker plots of diurnal variation (in LST, 24 hours) of BLH determined by all soundings operationally launched at 0000 and 1200 UTC (a) and by the soundings launched at both synoptic times and intensive observation times that are limited to the daytime alone (b). Solid green line and dotted blue line highlight the number of sonde station and total sounding for each hour of day, respectively.

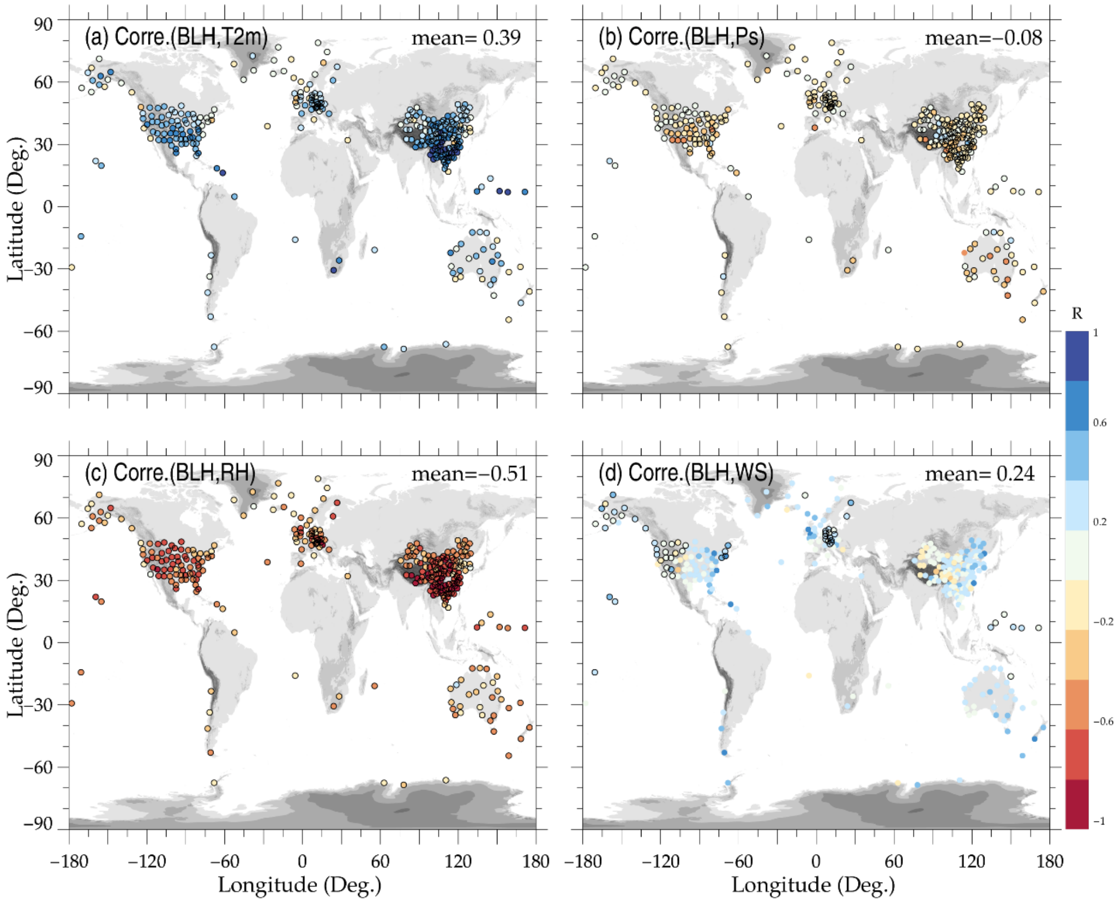

Figure 5. Correlations between the radiosonde-derived BLHs and near-surface air temperature at 2m AGL (T_{2m}; a), near-surface pressure (Ps; b), near-surface RH (c), and near-surface wind speed (WS; d). Dots outlined in black denote that the correlation coefficient values are statistically significant ($p<0.05$), and the mean correlations are texted in the upper right corner of each panel.

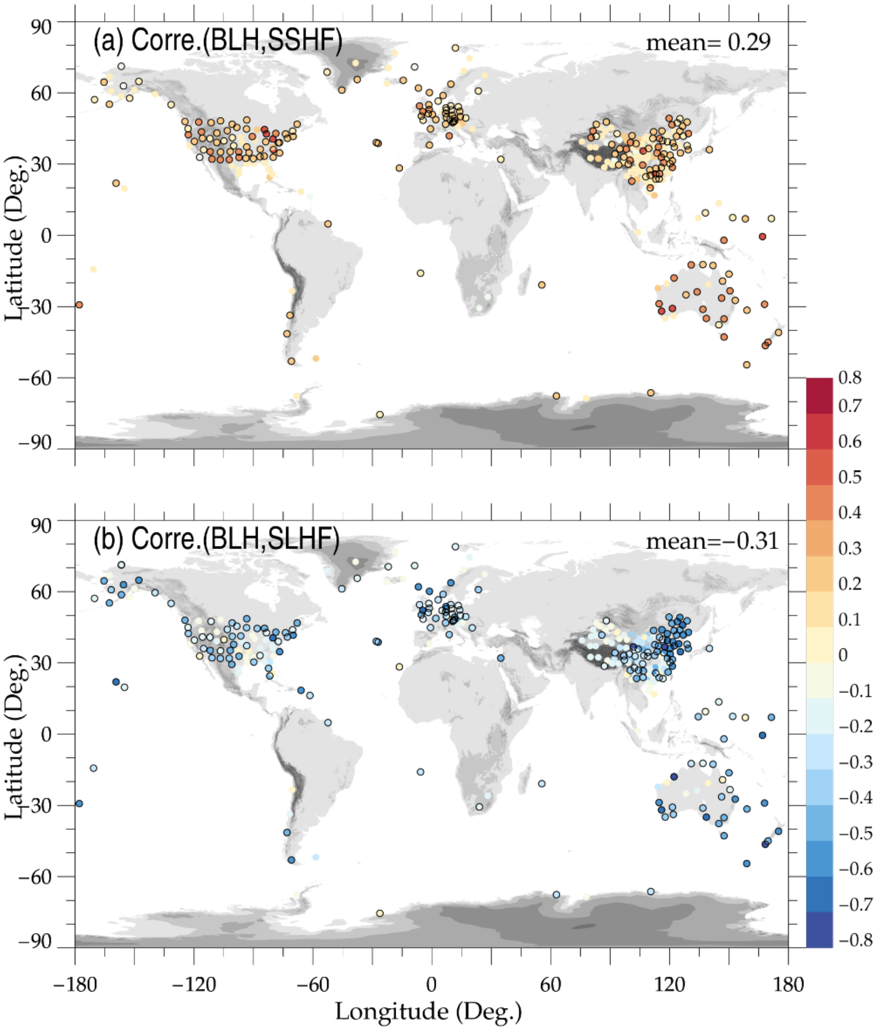

Figure 6. Similar as Figure 5, but for the correlations between BLHs versus normalized surface sensible (a) and latent heat fluxes (b).

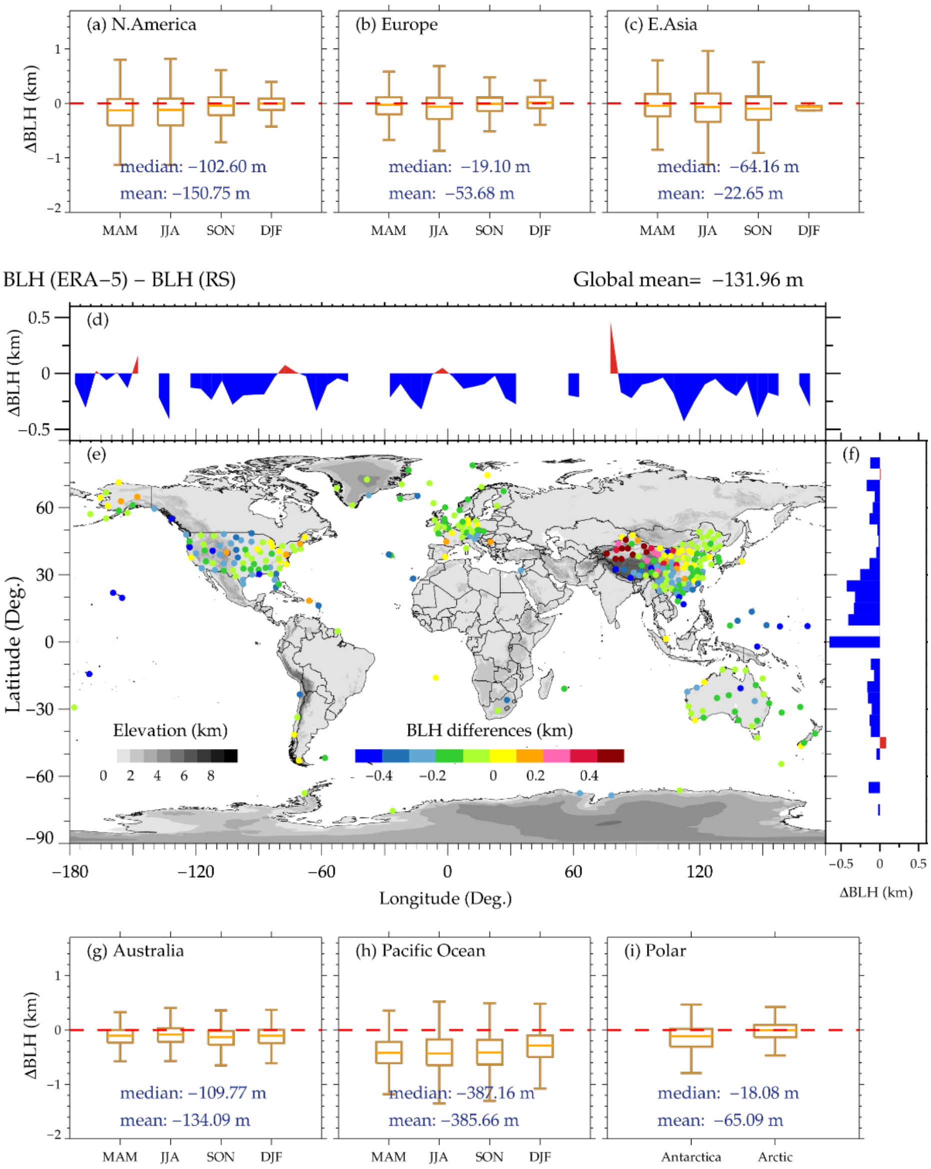

Figure 7. Statstical results of BLH differences between ERA5 and radiosonde. The spatial distribution of mean differences is highlighted in (e). Also shown are the distributions of mean BLH differences as a function of longitude (d) and latitude (f). The box and whisker plot of BLH differences over the six regions of interest (i.e., North America, Europe, East Asia, Australia, Pacific Ocean, Polar) over four seasons are displayed in (a-c), (g-i). The seasons are defined as follows: MAM, March–April–May; JJA, June–July–August; SON, September–October–November; DJF, December–January–February.

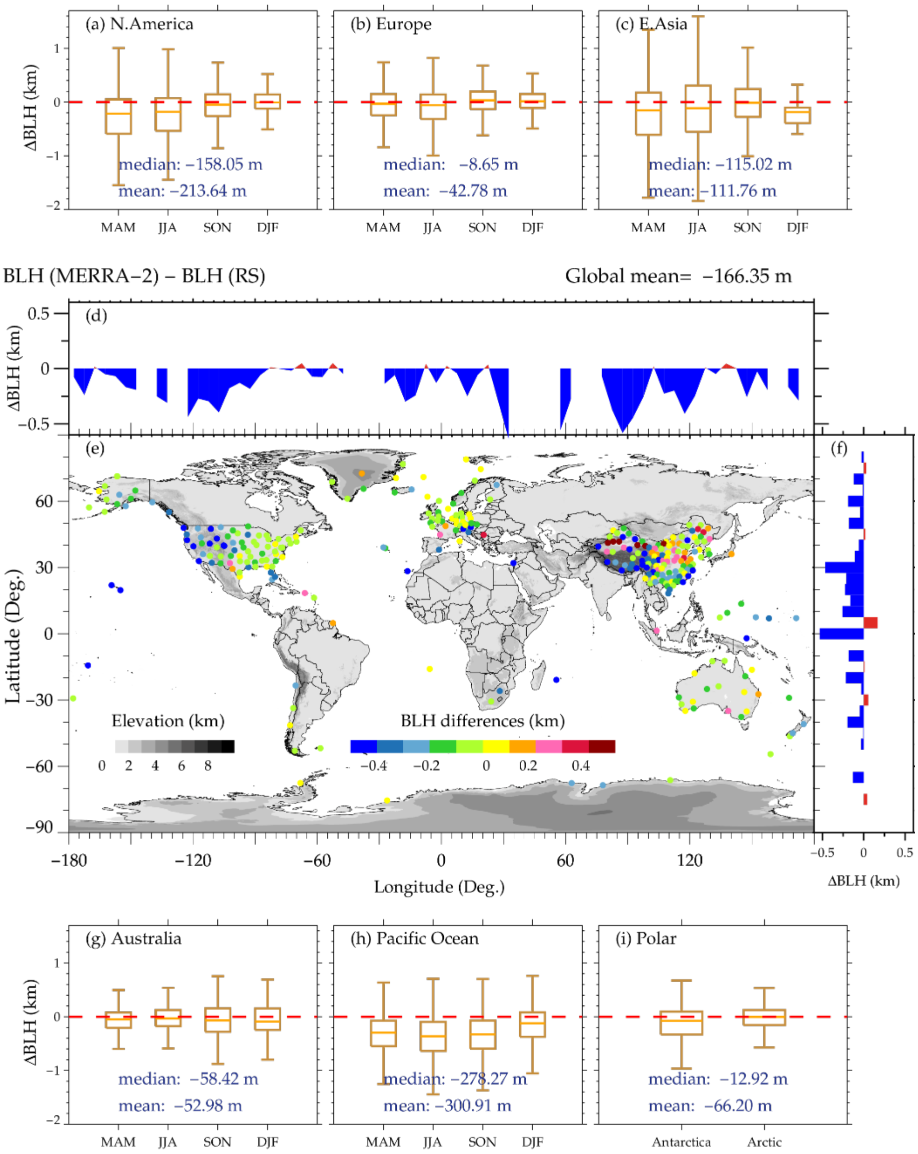

Figure 8. Similar as Figure 7, but for the differences between MERRA-2-derived BLHs and radiosonde-determined BLHs.

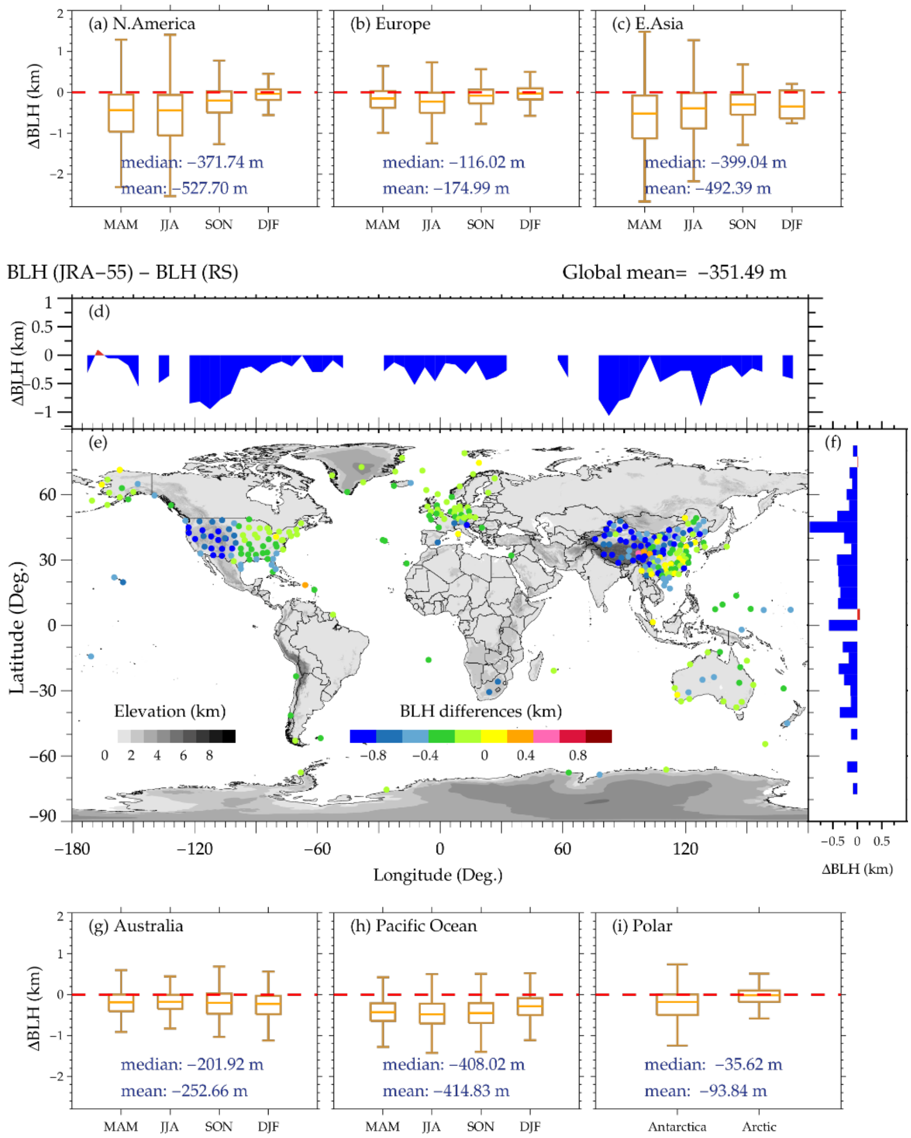

Figure 9. Similar as Figure 7, but for the differences between JRA-55-derived BLHs and radiosonde-determined BLHs.

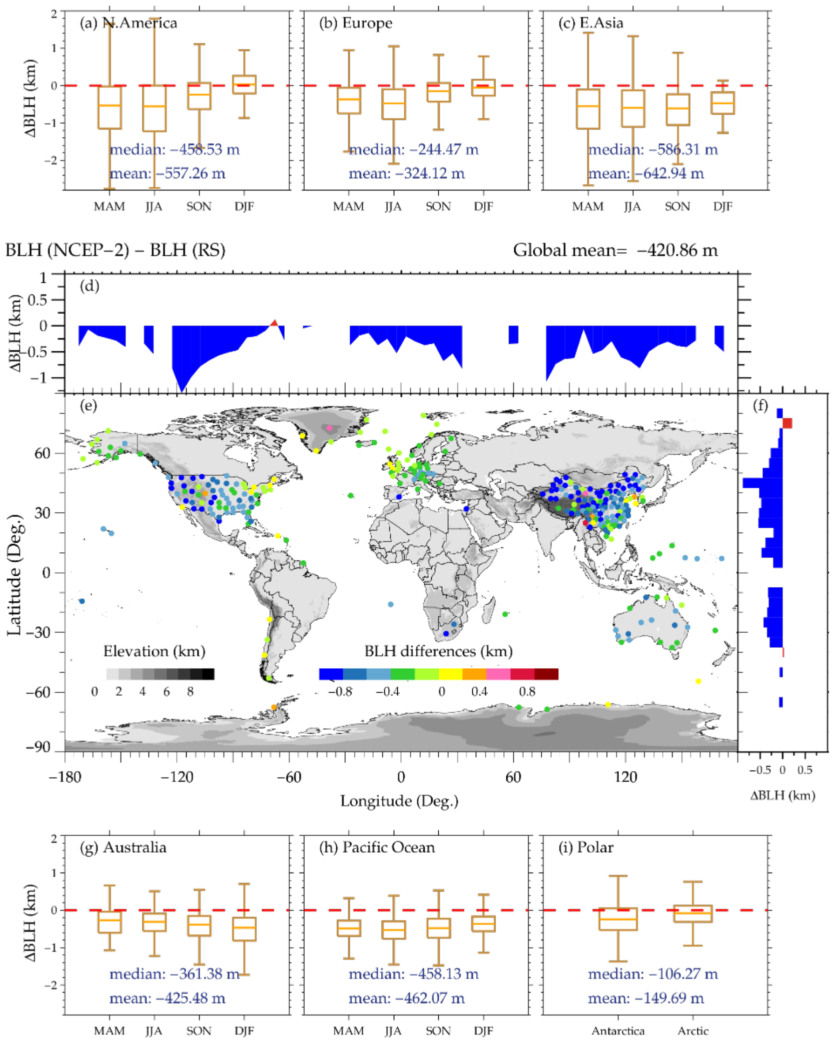

Figure 10. Similar as Figure 7, but for the differences between NCEP-2-derived BLHs and radiosonde-determined BLHs.

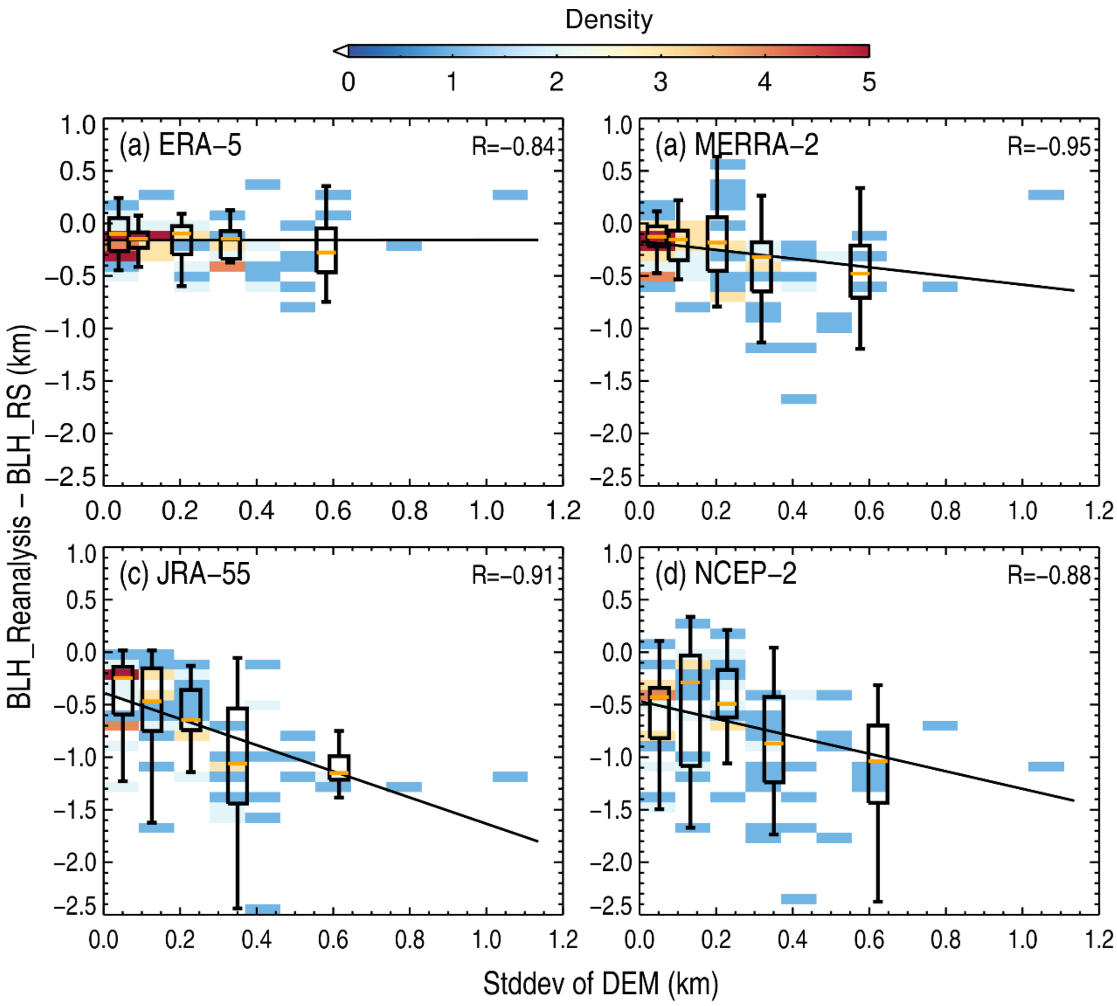

Figure 11. Density plots of the differences of BLHs between radiosonde and ERA5 (a), MERRA-2 (b), JRA-55 (c), and NCEP-2 (d) as a function of the standard derivation of the DEM, where the black lines denote the least-squares regression line. The box-and-whisker plots of the anomalies of BLH in five evenly intervals are overlaid in each panel, and the correlation coefficients are marked in the upper right corner of each panel. Note that all samples are collected from soundings that are launched in the afternoon, spanning from 1300 LST to 1800 LST.

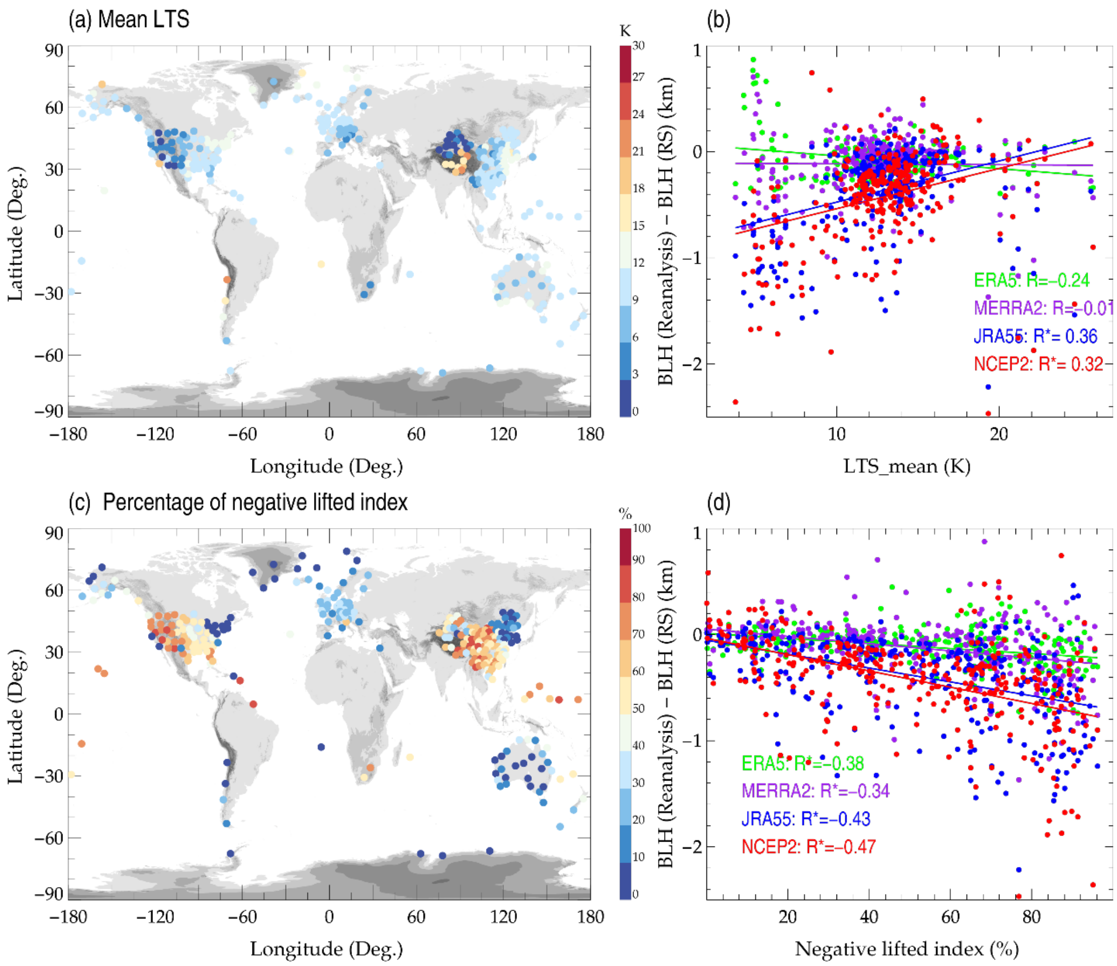

Figure 12. Spatial distribution of the ensemble means of lower tropospheric stability in the daytime (a). The scatter plots showing the difference of model- minus sounding-derived BLHs from four reanalysis datasets versus the anomalies of LTS as derived from four reanalysis relative to those from soundings (b). The variations in the percentage of negative lifted index (c), and the anomalies of BLH as a function of negative lifted index (d).