

1

2 **Diurnal Asymmetry of Desert Amplification and Its Possible**

3 **Connections to Planetary Boundary Layer Height:**

4 **A Case Study for the Arabian Peninsula**

5

6 Liming Zhou

7

8 *Department of Atmospheric and Environmental Sciences, SUNY at Albany, Albany, NY 12222,*

9 *USA*

10

11

12 Correspondence

13 Liming Zhou

14 Department of Atmospheric and Environmental Sciences

15 University at Albany, State University of New York

16 1400 Washington Avenue, Albany, NY 12222

17 Tel: (518) 442-4446; Fax: (518) 442-5825

18 Email: lzhou@albany.edu

19

20 Accepted for publication in

21 *Climate Dynamics*

22 January 4, 2021

23

Abstract. Recent studies using observations, reanalysis data and climate model simulations documented that 2m surface air temperature (T2m) has been amplified over the world's hottest and driest Sahara Desert and the Arabian Peninsula, referred to as desert amplification (DA). This study presents a comprehensive analysis of hourly surface observations, radiosonde measurements, and two latest state-of-the-art reanalysis products for the period 1979-2018 to examine the diurnal and vertical variations of DA and their connections with planetary boundary layer height (PBLH). It focuses on the Arabian Peninsula (AP), where observations are relatively abundant compared to the data scarce Sahara regions. Both observational and reanalysis data show that the diurnal cycle of surface warming rate depends, to some extent, inversely on the magnitude of climatological PBLH, and so DA has a distinct diurnal asymmetry with a stronger warming for a shallower PBLH. Results of upper air profiles reveal that DA is a bottom-heavy warming profile, which maximizes near the surface, decreases quickly with height, and is limited to the lower troposphere (>700 hPa) and surface. The major PBLH biases could explain, at least partially, some of the diurnal and vertical warming/cooling biases in the reanalyses. These results suggest that besides the surface radiative forcing, the PBLH may play an important role in modulating the diurnal and vertical structure of DA over the AP through heat redistributing via turbulent mixing.

Key words: Global warming, desert amplification, planetary boundary layer, diurnal asymmetry

1. Introduction

Deserts make up approximately 1/3 of the global land surface area (Zhou, 2016; Wei et al., 2017a). The Sahara and Arabian deserts, the world's two largest hot deserts, are formed in the subtropical subsiding branch of the Hadley cells and so generally associated with dry and cloudless weather conditions (Wu et al., 2009). The Sahara Desert and the Arabian Peninsula (SDAP) are among the driest and hottest regions on Earth and considered to be a hotspot in terms of climate change and impacts from regional to global scales through the influence of dust aerosols and atmospheric circulation (Knippertz and Todd, 2012; Vizzy and Cook, 2017; Thomas and Nigam, 2018). Observations and climate model simulations indicate adverse impacts of increasing warming and drought on fragile desert ecosystems in response to elevated greenhouse gas (GHG) concentrations (Huang et al., 2016; Thomas and Nigam, 2018).

By analyzing observational, reanalysis and projected land surface 2m air temperatures (referred to as T2m hereafter), several recent studies documented that T2m in mid- and low- latitudes has warmed the most over the SDAP. Zhou et al. (2015; 2016) examined the observational, reanalysis, and modeled T2m trends in 50°S-50°N by large-scale ecoregion for the period 1979-2012 and found dramatically increased warming rates with increasing surface aridity and the strongest warming over the driest and least vegetated SDAP. Cook & Vizzy (2015) evaluated annual mean T2m of three reanalyses and two observational gridded datasets for the period 1979-2012 and showed 2-4 times more warming over the Sahara than over the whole tropics. Evan et al (2015) examined *in-situ* observations of three stations and one atmospheric reanalysis for the months of July and August during the period of 1979–2012 and identified amplified

warming in summer over the Western Sahara Desert. Zhou (2016) further examined T2m changes in historical and projected simulations (1950-2100) from the Coupled Model Intercomparison Project phase 5 (CMIP5) and found strongest surface warming consistently and persistently seen over the SDAP during various 30-year periods after the 1980s, pointing to desert amplification (DA) in a warming climate. This work also showed that the magnitude of DA increased linearly with the global mean radiative forcing due to increasing GHGs. DA was reproduced by CMIP5 historical “all forcings” simulations, but was absent if only natural forcings were used, suggesting human influence (Zhou et al., 2015; 2016; Zhou, 2016). Using multiple satellite datasets, Wei et al. (2017a) indicated that DA was strongest at the surface, decreased with height, and mostly disappeared in the upper troposphere. The essential features of DA remained robust across all seasons, although the magnitude of DA was greater during summer months (Zhou et al., 2016; Vizzy and Cook, 2017; Wei et al., 2017a). These results suggest that DA is a fundamental feature of global warming patterns in mid- and low- latitudes and intensifies with increasing GHGs.

DA is conceptually similar to the well-known arctic amplification (AA), a subject of intensive research for several decades (e.g., Serreze and Barry, 2011). AA has been attributed to feedbacks associated with surface albedo, water vapor, cloud, and lapse rate, and to changes in atmospheric and oceanic heat transport (IPCC, 2007; 2013). In contrast, DA is an emerging new concept and its causes are largely unknown. Several recent studies suggested that DA may result mainly from large-scale greenhouse effects in a warming climate over the SDAP associated with increasing water vapor (Cook & Vizzy 2015; Zhou et al. 2016; Zhou 2016; Wei et al., 2017a; Evan et al., 2017). This attribution was proposed from the local surface energy budget perspective, based on

the results that the desert warming rate is well correlated spatially and temporally with enhanced downward longwave radiation (DLR) at the surface as a result of a warmer and moister atmosphere. It is suggested that DA may alter regional-scale climate and circulation over the deserts and surroundings and thus have significant environmental, societal, and economic consequences (Zhou 2016; Vizzy and Cook 2017). Hence, understanding major processes that control DA is essential for a complete assessment of climate change and impacts.

One distinct feature of hot desert climate is the pronounced diurnal cycle. The SDAP is characterized by extremely high temperatures during daytime and very low temperatures during nighttime, which creates the largest diurnal temperature range (DTR) in the world (Zhou et al. 2007; 2009; 2010). It is also marked by a large diurnal cycle in the atmospheric planetary boundary layer (PBL). Turbulent mixing in the PBL governs the vertical exchange of heat, moisture, momentum, and aerosols in the surface-atmosphere interface and thus strongly influences the atmospheric temperature, moisture, and wind (Stull, 1988). The PBL height (PBLH) represents the maximum height of the free atmosphere that is directly influenced by the Earth's surface and responds to surface impacts. On the diurnal time scale, the development of PBL typically consists of the deep convective boundary layer (CBL) during the day, the shallow stable boundary layer (SBL) at night, and their transition stages in the morning and evening periods. In general, PBLH depends proportionally on the intensity of surface heating over dry regions, and so the global PBL climatology shows the maximum PBLH in the SDAP, up to 3.5 km in summer months (Ao et al., 2012). Among various climate zones worldwide, the SDAP generally has the deepest and well-mixed PBL at daytime but the shallowest and most stably

stratified PBL at nighttime, with the strongest diurnal asymmetry in PBLH (Gamo, 1996; Messenger et al., 2010; Garcia-Carreras et al., 2015; Davy, 2018).

It is interesting to note that the spatial pattern of DA is coupled geographically well with that of the climatological DTR and PBLH over the SDAP, which are strongly connected with the unique diurnal features of desert land surface and PBL processes. Besides the positive radiative forcing at the surface, recent studies indicated that the PBLH modulates the T2m response to the surface forcing and is a stronger predictor of the diurnal asymmetry in surface warming (McNider et al. 2012; Davy and Esau, 2016; Davy et al., 2016). This implies that the strong diurnal cycle of PBLH over the SDAP may result in a diurnal asymmetry in DA. Also, the amplified surface warming associated with DA could modify the vertical warming profile or lapse rate over deserts via strong turbulent mixing in the PBL. However, recent detection and attribution of DA have been limited to the seasonal and annual features of daily mean T2m (Zhou et al., 2015; Cook & Vizzy 2015; Zhou et al. 2016; Zhou 2016; Wei et al., 2017a; Evan et al., 2017; Vizzy and Cook 2017), little attention is given to understand how DA varies diurnally and vertically. Hence, understanding the diurnal and vertical features of DA and their connections with PBLH is an important next step.

The diurnal variation is one of the most fundamental modes of variability of the global climate system and may function as a bridge between weather and climate (Yang and Slingo, 2001; Ruppert, 2016). Changes in surface temperatures such as daily maximum ($T_{\max}$), daily minimum ($T_{\min}$), and the DTR have been examined intensively to study climate change and variability (IPCC, 2007; 2013). Associated with global warming is a greater warming in $T_{\min}$ than $T_{\max}$, and

thus a substantial reduction in the DTR observed over many land areas since 1950 (Vose *et al.*, 2005; IPCC, 2007; 2013). The DTR calculated from $T_{\min}$ and $T_{\max}$ ($DTR = T_{\max} - T_{\min}$) are often analyzed to describe the diurnal cycle, but the full diurnal cycle is far more complex than the simple difference in the two numbers and needs analysis of high temporal resolution data (e.g., Vinnikov *et al.*, 2002; Davy *et al.*, 2017).

Despite its importance to Earth's climate, the Sahara has one of the sparsest networks of routine meteorological measurements of any landmass on Earth, with most measurements only available at the periphery of the desert (Marshall *et al.*, 2013). This data gap fundamentally limits our understanding of the Saharan climate because of insufficient observations available for data collection, assimilation or model validation (Garcia-Carreras *et al.*, 2015; Wei *et al.*, 2017a; 2017b). Also, weather and climate models have difficulties in realistically simulating the magnitude and diurnal evolution of PBLH and T2m, particularly over dry climates (Christensen *et al.*, 2008; McNider *et al.* 2012; Boberg and Christensen, 2012; Lewis and Karoly, 2013; Davy and Esau, 2016; Davy *et al.*, 2017; Wei *et al.*, 2017b; Davy, 2018). These data limitations and model deficiencies cast doubt on detecting and attributing DA. Further validation of DA using *in situ* observations is essential.

Surface and atmospheric observations in the Arabian Peninsula (AP) are relatively abundant compared to significant data gaps in the Sahara Desert. DA is a continental-scale warming pattern covering the entire SDAP and so the warming features in the AP are representative of DA. In addition, high temporal and spatial resolution reanalysis products have been released recently with improved quality. This provides a great opportunity to further understand and

validate DA using *in-situ* observations and advanced reanalysis data over the AP. The present paper analyzes high resolution data from surface and radiosonde observations and two widely used reanalysis data to examine the diurnal cycle and vertical structure of DA and its potential drivers over the AP. It focuses on the modern satellite data era for the period 1979-2018 to maximize spatial coverage of measurements that are assimilated into reanalysis products. This era also corresponds to the period when the observed DA signal is most significant.

The motivation for this study is threefold. The first goal is to search for more observational evidence of DA. The second goal is to examine the diurnal variation in surface warming associated with DA and potential drivers using hourly data, rather than the DTR as done previously. The third goal is to link the diurnal cycle of DA with that in the upper air. Climate change research has focused dominantly on T2m, including the detection and attribution of DA, but temperature changes are not limited to the Earth surface and can be extended into the free atmosphere (Brocard et al, 2013). The vertical structure of temperature changes can tell a whole story of climate change, and in this case, can improve our understanding of the diurnal coupling of surface and air temperatures over deserts.

2. Data and Methods

2.1. Study region

This study focuses on the AP and surrounding areas, where high-quality surface and radiosonde observations are available. It is worth noting that the spatial coverage of observational network

in the AP is sparser than that in North America and Europe but is relatively abundant for large-scale analysis in comparison to the other SDAP regions. The study domain, depicted as the rectangle box (16.5°N-32.5°N, 34.5°E-50°E) in Fig. 1, covers most of the AP countries and part of several neighboring states such as Egypt, Sudan, Iraq, Iran, and Israel. It includes the Arabian Desert, which occupies almost the entire AP, but excludes the relatively humid southern part of the AP affected by convective processes and clouds (Hassan et al., 2016; Patlakas et al., 2019).

The climate of the AP is extremely hot and dry, with infrequent low rainfall (Chowdhury and Al-Zahrani, 2013; Patlakas et al., 2019). The landscape consists of highlands in the western and southwestern regions, the vast arid and extra arid lands of the interior (Najd), the world's largest continuous bodies of sand deserts, and the Rub Al-Khali in the southeast (Patlakas et al., 2019). Despite sharing similar large-scale climate features, the AP also demonstrates some level of heterogeneity at local to regional scales due to variations in vegetation, topography, proximity to sea, and regional circulation patterns (Ahmed, 1997; Krishna, 2014). Among the AP countries, Saudi Arabia occupies ~4/5 of the AP and a large part of the Arabian Desert lies within the country. It is among the hottest countries with very low humidity in the world and average temperature ranging from 27°C-43°C in inland regions and 27-38°C in coastal regions (Krishna, 2014). For example, the highest (lowest) temperature of 52°C (-10°C) are recorded in two stations in Saudi Arabia (Almazroui et al. 2014). Saudi Arabia provides most of the station-based observations used in this study and is relatively well studied in terms of climate change compared to the Sahara Desert.

2.2. Observational and reanalysis data

2.2.1. Integrated global daily radiosonde data

Daily weather data for the atmosphere have been regularly obtained from radiosondes and pilot balloons dating back to 1905. The Integrated Global Radiosonde Archive Version 2 (IGRA2) from the U.S. National Climatic Data Center (NCDC) consists of quality-controlled sounding observations from various sources at >1500 global stations with varying periods of record (Durre et al., 2006). Measurements include atmospheric vertical profiles of temperature, humidity, wind and other variables at mandatory pressure levels. The IGRA2 has applied a comprehensive set of quality control procedures to the data to remove gross errors. So far, it is the largest and most complete dataset of quality-assured radiosonde observations freely available. Its temporal and spatial coverage is most complete over the U.S., Western Europe, Russia and Australia. The vertical resolution and extent of soundings improve significantly over time, with nearly three-quarters of all soundings reaching up to at least 100 hPa by 2003. However, the IGRA2 data have not been adjusted for inhomogeneities due to changes in instrumentation, observing practice, or station location.

It is well recognized that radiosonde data need bias correction for homogeneity before they can be used for trend analysis in climatic research (Thorne et al., 2011; Haimberger et al., 2012). One of the widely used homogeneity-adjusted radiosonde datasets based on the IGRA2 is the homogenization of global radiosonde temperatures with the Radiosonde Observation Correction Using Reanalyses (RAOBCORE) and the Radiosonde Innovation Composite Homogenization (RICH) (Haimberger, 2007; Haimberger et al., 2008). The RAOBCORE homogenization method

detected shifts in existing radiosonde observation time series and estimated the size of the shifts using background forecast time series from ERA-40 (1958-1978) and ERA-Interim (1979 onwards) as reference for break detection. However, the background forecasts may be influenced by biases in the radiosonde data, and by uncertainty from other observing systems, most notably satellites, and reanalysis models (Haimberger et al., 2012). To avoid this problem, the RICH homogenization method created reference series from neighboring radiosonde stations for breakpoint adjustment. It works well if the radiosonde network is not too sparse and only homogeneous pieces of the neighboring time series are used. Note that homogeneity adjustments were only made to radiosonde-based temperature measurements.

This study used the latest version v1.74 of the RICH dataset on 16 pressure levels for several reasons: (1) it has the longest data record with most stations compared to other homogeneity-adjusted radiosonde datasets over the AP; (2) it has homogeneity-adjusted measurements at both daytime and nighttime (00 and 12 Coordinated Time Universal or UTC); (3) it exhibits the closest match to the latest satellite observations in the tropics (Thorne et al., 2011). Note that the RAOBCORE adjusted data is assimilated into the ERA5, which is used in this study, while RICH adjusted data is not used for the ERA5.

Here the subdaily and daily temperatures from the RICH were downloaded for the period 1979-2018. Eight radiosonde stations over the study region were chosen following the data selection criteria (section 2.3.1). To estimate the climatology of PBLH for the same period, the original radiosonde soundings from the IGRA2 were also used (section 2.3.3).

2.2.2. Integrated global hourly surface observations

Hourly surface-based meteorological observations are available from the global-scale, quality-controlled integrated surface hourly dataset (DS3505) archived in the U.S. NCDC (<https://www7.ncdc.noaa.gov/CDO/cdopoemain.cmd>). The DS3505 consists of numerous global hourly and synoptic observations for regular weather parameters (e.g., temperature, dew point, wind, etc.) into a common format and data model, and thus provides a single collection of global hourly data with continuous updates. The primary data sources include the Automated Surface Observing System (ASOS), Automated Weather Observing System (AWOS), and various others from more than 35,000 stations worldwide for the period 1901-present. The data spatial and temporal coverage in the DS3505 is however not even. The best spatial coverage is evident in North America, Europe, Australia, and parts of Asia, but only a limited number of stations is in the Southern Hemisphere and the African Continent. Some stations have data as far back as 1901, but there is a substantial increase in data volume starting in the late 1970s. In terms of data continuity, some stations have over 50 years of continuous records, while others have “breaks” in the period of record (Smith et al., 2011).

Besides internal quality control (QC) procedures applied to several major input datasets to the DS3505, there have been continued incremental improvements in automated QC software since 2003 (Smith et al., 2011). The input data sources were first processed through automated and some manual QC and then additional QC software was developed and applied to the entire archive. The QC process included 54 QC algorithms checking for proper data format for each field, extreme values and limits, consistency between parameters, and continuity between

observations. The standardized and consistent QC procedure in the DS3505 was designed to eliminate obvious errors in the data, minimize overflagging of data, and ensure delivery of spatially variable, research-quality data. Detailed information regarding the QC process is described in Smith et al. (2011).

This study used the subhourly and hourly T2m from the DS3505 for the period 1979-2018. 21 surface stations over the AP were chosen following the data selection criteria (section 2.3.1).

2.2.3. High-resolution reanalysis products

This study used two of the latest state-of-the-art reanalysis products that provide hourly or 6-hourly analysis fields at relatively high spatial resolutions for the period 1979-2018.

ECMWF Reanalysis 5th Generation (ERA5) gives a numerical description of the recent climate, produced by combining vast amounts of historical observations into global estimates using advanced modelling and data assimilation systems (C3S, 2017). This climate reanalysis provides hourly estimates of many atmospheric, land and oceanic climate variables covering the period 1979 to present. The data cover the Earth on a 30 km grid and resolve the atmosphere using 137 levels from the surface up to a height of 80 km. The monthly mean data of analyzed meteorological fields: (1) hourly averaged temperature, humidity, wind speed, geopotential height on pressure levels, and (2) hourly averaged surface pressure and PBLH, at a spatial resolution of 0.5° longitude $\times 0.5^\circ$ latitude are used in this study. Hourly averaged T2m, surface humidity, and surface fluxes at the same spatial resolution are provided by the ERA5-Land (C3S,

2019). The ERA5-Land is a replay of the land component of the ERA5 climate reanalysis for the period 1981-present. It is produced to meet the needs of users for a more accurate surface dataset using the tiled ECMWF Scheme for Surface Exchanges over Land incorporating the land surface hydrology (H-TESEL) model. All available ERA5 and ERA5-land datasets are detailed at <https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset>.

The second Modern-Era Retrospective analysis for Research and Applications (MERRA-2) is a NASA atmospheric reanalysis that begins in 1980 with the enhanced use of satellite observations (Gelaro et al., 2017). The MERRA-2 data are provided on the same horizontal grid of 0.625° longitude $\times 0.5^{\circ}$ latitude at 42 pressure levels. The monthly mean data of analyzed meteorological fields: (1) 6-hourly (00, 06, 12 and 18 UTC) instantaneous temperature, humidity, wind speed, geopotential height on pressure levels, and (2) hourly averaged variables of T2m, surface humidity, surface pressure, PBLH, and surface fluxes, are examined in this study. Note that the hourly analyzed sounding data is not provided in the MERRA-2. A comprehensive list of available variables is detailed at <https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich785.pdf>.

2.3. Data processing and methods

The above four datasets have different temporal resolutions. The surface observations in the DS3505 have subhourly and hourly data; the radiosonde observations in the RICH contain subdaily and daily data; the two reanalysis products consist of monthly means of hourly averaged or 6-hourly instantaneous data. Since the signal of DA is limited to the surface and lower troposphere, the four datasets are processed into two types: (1) the hourly-averaged data near the

surface (e.g., T2m, PBLH, and surface fluxes) and (2) the sounding data at 00 and 12 UTC (e.g., air temperature and humidity). The latter only considers the 10 “mandatory reporting” pressure levels at and below 100 hPa shared by both radiosonde and reanalysis data: 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, and 100 hPa. Because the annual mean surface pressure over the study region is 948 hPa in the MERRA-2 and 952 hPa in the ERA5, the sounding results on 1000 hPa are excluded.

This work mainly analyzes temperature trends considering the arid nature of the AP climate as rainfall occurs only on a few days per year for most stations and humidity is persistently very low (Almazroui, 2014). Also, observed temperatures are of relatively high-quality, have much fewer missing data than other variables in the DS3505, and are homogeneity adjusted in the RICH. Only the annual mean data are examined to maximize the data coverage as seasonal mean data have a smaller sample size with a relatively higher ratio of missing data. Doing so will not bias the results with seasonality as DA is a large-scale warming pattern across all seasons (Vizy and Cook, 2017; Wei et al., 2017a). Because every variable analyzed here is an annual mean quantity, the term “annual mean” will be often omitted for brevity for the remainder of this paper.

2.3.1. Data processing for near surface and atmospheric air temperatures

The first task is to create annual mean anomalies for hourly T2m and for atmospheric temperatures at 00 and 12 UTC for the study period following the five steps detailed next. The first step is to create the daily hourly mean data. For each day, the subhourly and hourly data are

aggregated into hourly averaged data. This step is applied only to the surface station observations from the DS3505. The second step is to create the monthly mean data. For each month, the daily data are averaged to create the monthly mean. For every month, at least 10 days of data are required for the monthly averaging. Otherwise, missing data is assigned for that month. This step is applied to both the surface and radiosonde observations. The third step is to create the monthly mean anomalies by subtracting the long-term monthly mean (climatology) from the monthly mean data. The fourth step is to create the annual mean anomalies. For every year, the monthly mean anomalies are averaged to create the annual mean anomalies, and at least 6 months of data are required for the annual averaging. The fifth step is to create the long-term annual mean anomalies. At least 28 years (70%) of data for the study period 1979-2018 are required. The third to fifth steps are applied to all four datasets. The use of thresholds of 10 days per month (e.g., Li et al., 2020), 6 months per year (e.g., Wang and Wang, 2016), and 70% of the temporal coverage (e.g., Gertler and O’Gorman, 2019) is a reasonable compromise between the data length, completeness, and spatial coverage.

The second task is to identify radiosonde stations in the RICH and surface stations in the DS3505 from all available stations over the AP that having valid observations. There are 8 radiosonde stations and 21 surface stations in the study domain meeting the above data selection criteria. Most chosen stations are located within the airport premises, which are isolated from urban centers or industrial areas and comply with the World Meteorological Organization (WMO) standards (Almazroui et al., 2014). The name, WMO identifier number, and location for these stations are listed in Tables 1-2 and illustrated in Fig. 1a.

The third task is to estimate the linear trend of annual mean temperature anomaly time series over the study period using least squares fitting. A two-tailed student's t test is used to quantify the probability of whether the trend is statistically significant from zero. Before the trend analysis, a three-hour smoothing is applied to the surface hourly data for three reasons. First, the sample size among individual hours in the DS3505 varies largely for some stations due to irregular missing data and so the smoothing helps to reduce the sampling inhomogeneity. Second, the study domain spans 15.5° in longitude (34.5°E - 50°E) covering two time zones in local solar time, and so the smoothing helps to reduce the small phase difference in the diurnal cycle of data among different grids when spatial averaging is applied. Third, the reanalysis hourly averaged data consist of a continuous sequence of data averaged over the indicated interval and time stamped at 00:30 UTC, 01:30 UTC, ..., 23:30 UTC, while the reanalysis instantaneous data contain snapshots at synoptic times (00 and 12 UTC). The 3-hourly averaging helps to smooth out the timing difference among the data with different temporal resolutions. However, this three-hourly smoothing has limited impact if the data is smooth in the diurnal cycle. Trend analysis is performed at every station for the observations, every grid for the reanalysis, or for the spatially aggregated data.

2.3.2. Data processing for other variables

The above data processing calculates the climatology and linear trends in the annual mean data of (i) T2m at the hourly time scale and (ii) vertical temperature profiles at 00 and 12 UTC. The monthly means of daily $T_{\max}$, $T_{\min}$ and DTR were composited from the hourly T2m values and their annual mean trends are estimated accordingly. Similarly, the monthly means of hourly

averaged surface fluxes: sensible heat, latent heat, DLW, downward shortwave radiation (DSR), net shortwave and longwave radiation, and upward shortwave and longwave radiation, are processed and the linear trends of annual mean anomalies of these variables are also calculated for the MERRA-2 and ERA5. For simplicity, the trends of T2m, DLR and DSR, three frequently used variables, are referred to as $T2m_{trend}$ (C/decade), DLR_{trend} ($W/m^2/decade$), and DSR_{trend} ($W/m^2/decade$) hereafter, respectively.

2.3.3. Creating climatological PBLH

Both ERA5 and MERRA-2 provide the monthly mean of hourly averaged PBLH. The reanalysis PBLH is derived based on the bulk Richardson number in the ERA5 (C3S, 2017) following the conclusions of Seidel et al. (2012) and the total eddy diffusion coefficient of heat with a threshold value of $2 m^2s^{-1}$ in the MERRA-2 (Salmun et al., 2018). The monthly means of hourly averaged PBLH from the reanalysis are averaged to create the long-term climatology of PBLH, referred to as $PBLH_{climate}$.

The reanalysis-derived PBLH is a model-based estimate with large uncertainties and different PBLH estimation methods can produce substantially different values (section 3.5). To validate and intercompare the reanalysis PBLH, the bulk Richardson number (Ri) method (Vogelezang and Holtslag, 1996) is chosen to consistently diagnose the PBLH directly from the atmospheric soundings among different datasets. The Ri methods have proven to work reasonably well for both stable and convective boundary layers, and don't strongly depend on the sounding vertical

resolutions (e.g., Seidel et al., 2012; Zhang et al., 2013). The Ri is the ratio of turbulence associated with buoyancy to that associated with mechanical shear (Seidel et al., 2012):

$$Ri(z) = \frac{(g_0/\theta_{vs})(\theta_{vz}-\theta_{vs})(z-z_s)}{(u_z-u_s)^2+(v_z-v_s)^2}, \quad (1)$$

where g_0 is the acceleration of gravity, z is height, and s denotes the surface at the 2 m level, θ_v is virtual potential temperature, u and v are the zonal and meridional wind speed.

This study followed exactly the steps detailed in Seidel et al. (2012) to diagnose the PBLH for the radiosonde observations (IGRA2) and reanalysis datasets (ERA5 and MERRA-2). Due to the lack of observational winds at 2m, Ri is calculated by setting the surface winds to zero in the reanalyses for consistency with the radiosonde observations. With these assumptions, the PBLH is designed as the lowest level at which the bulk Ri reaches the critical value of 0.25. Scanning the Ri profile upward from the surface (at 2 m), the first level with $Ri \geq 0.25$ is identified, and linear interpolation between that level and the next lowest level provides an estimate of $z(Ri_{0.25})$. The PBLH estimated based on $z(Ri_{0.25})$ is referred to as $PBLH_{Ri_{0.25}}$. Note that all PBLH estimates in this study are measured in meters above ground level (AGL).

2.3.4. Multiple linear regression analysis

For the reanalysis data, $T2m_{trends}$ are found to be mostly related to DLR_{trend} and DSR_{trend} , which represent the surface longwave and shortwave radiative forcing, and $PBLH_{climate}$, which modulates the T2m response to the surface forcing. Next, these three variables are used as

independent variables in a multiple linear regression to quantify their relative contributions to the spatial and/or diurnal variation in $T2m_{trend}$ over the study domain:

$$T2m_{trend} = \beta_0 + \beta_1 * \frac{1}{PBLH_{climate}} + \beta_2 * DLR_{trend} + \beta_3 * DSR_{trend}, \quad (2)$$

where β_1 through β_3 are the partial regression coefficients estimated based on least squares fitting. Each coefficient represents the change in $T2m_{trend}$ to a one-unit change in the respective independent variable, holding all other variables constant. Its sign determines if the independent variable affects $T2m_{trend}$ positively or negatively. Statistical t and F tests can be performed to assess the statistical significance of each regression coefficient and the overall regression model, respectively. The adjusted R-squared ($R^2_{adjusted}$) measures the percentage of variance in $T2m_{trend}$ that can be explained by the three variables after adjusting the statistic based on the number of independent variables in the regression model.

2.3.5. Obtaining large-scale features

Besides the large-scale radiative forcing, $T2m_{trend}$ at individual stations is influenced by local factors and thus demonstrates some level of inter-station differences (section 2.1). In order to maximize large-scale warming patterns and minimize station-scale temperature variability, spatial averaging is applied at two spatial scales: (1) station mean and (2) regional mean. The former is simply an arithmetic mean of individual station data and is used for both the observational and reanalysis data. For the reanalysis, the station level data are obtained from the grid boxes where the chosen stations are located based on their geographic location (latitude and

longitude). The regional mean is applied only to the reanalysis fields using area-weighted averaging over the land grids within the rectangle study domain (16.5°N-32.5°N, 34.5°E-50°E) depicted in Fig.1.

The regional mean cannot be done for the observations because of limited stations available in the study domain and the non-uniform distribution of stations. As an alternative, an empirical orthogonal function (EOF) analysis is performed on the observed surface and radiosonde data to emulate their regional mean data that could be compared with the regional mean reanalysis data. EOF decomposes the data in terms of orthogonal basis functions and finds both spatial patterns (called EOF) and associated time series to extract the space–time modes of climate variability (Bjornsson and Venegas, 1997). The first EOF explains the greatest fraction of the total variance, the second for the largest part of the remaining variance, and so on. The EOF analysis helps to identify the most important modes of data variability, which describe the degree of coherence of spatial variation. A new temperature time series can be reconstructed based on the first EOF pattern to quantify the dominant large-scale warming patterns shared among all stations.

3. Results and discussion

3.1. Hourly T2m trends in surface observations

Figure 2 shows the diurnal cycle of $T2m_{trend}$ during 1979-2018, along with the diurnal cycle of climatological T2m, for the 21 surface stations in the DS3505. For each station, the $T2m_{trend}$ is positive at every hour and exhibits a strong diurnal pattern. Among the 504 (24 hours * 21

stations) trends, 501 are statistically significant ($p < 0.05$). The warming rate is generally in opposite phase with the climatological T2m value but with a lag of few hours. The largest/smallest warming is mostly seen around the transitions between day and night, with the largest warming in the early morning and the smallest warming in the late afternoon to the early evening. Among the 21 stations, the 24-hour averaged warming rate ranges from 0.39 to 0.93 °C/decade, and the diurnal range of the hourly warming rate is 0.20 to 0.77 °C/decade.

To focus on the large-scale warming features, Fig. 3a shows the diurnal cycle of station mean hourly $T2m_{trend}$. The hourly warming rate ranges 0.45-0.73 °C/decade, with a diurnal range of 0.28 °C/decade and a 24-hour average of 0.59 °C/decade. As stated previously, the leading EOF modes can capture the large-scale features shared among all individual stations. The reconstructed $T2m_{trend}$ from EOF1 (Fig. 3b), which can explain 79.6% of the total data variance from the original 21 stations, resembles Fig. 3a but in a slightly smaller magnitude. The hourly warming rate ranges 0.38-0.60 °C/decade, with a 24-hour average of 0.50 °C/decade. The climatology of T2m reaches the minimum of 18.7 °C at 03 UTC and the maximum of 30.0 °C at 12 UTC, while the warming trend maximizes at 05 UTC in the early morning and minimizes at 15 UTC around the sunsets. Like individual stations (Fig. 2), the diurnal cycle of station mean warming rate generally depends inversely on that of the climatological T2m value, but with a lag of few hours. Previous studies (e.g., Seidel et al., 2005) showed that the diurnal cycle of temperature peaks a few hours after local solar noon at the surface. Note that local solar time in Saudi Arabia = UTC + 3 hours.

Figure 4a shows interannual variations of station mean time series of T2m anomalies at 00 and 12 UTC from 1979 to 2018 averaged over the 21 surface stations in the DS3505. T2m at both UTC times exhibits a persistent and statistically significant ($p < 0.001$) upward trend and similar interannual variability. The warming trend is $0.60\text{ }^{\circ}\text{C/decade}$ at 00 UTC, and $0.57\text{ }^{\circ}\text{C/decade}$ at 12 UTC, indicating stronger warming rates at nighttime than daytime. Note that the $T_{2m\text{trend}}$ difference between 00 and 12 UTC is not large because both UTC times do not correspond to the times with the maximum/minimum warming rates.

There are no warming trends reported at hourly time scales over the AP in the literature. Despite some seasonal and spatial variations in magnitude, warming has been consistently observed across stations and seasons in Saudi Arabia (Rehman and Al-Hadhrami, 2012; Almazroui et al., 2012; 2014; Alghamdi and Moore, 2014; Athar, 2014; Krishna, 2014; Tarawneh and Chowdhury, 2018). Here several recent studies that reported daily mean, T_{max} , and T_{min} trends in Saudi Arabia are used to partially validate the results in Figs 2-3. Krishna (2014) analyzed annual mean T2m trends for the period 1984-2013 for 4 stations representing 4 different climatic zones of Saudi Arabia, and estimated significant warming trends of $0.52\text{--}0.69$, $0.31\text{--}0.62$, and $0.48\text{--}0.71\text{ }^{\circ}\text{C/decade}$ for the daily mean, T_{max} , and T_{min} . Alghamdi and Moore (2014) compared warming trends over the period 1985–2010 at two weather stations (urban vs. rural) in Riyadh city, Saudi Arabia. The rural station showed a warming trend of 0.69 and $0.83\text{ }^{\circ}\text{C/decade}$ for the T_{max} and T_{min} , respectively. The corresponding values for the urban station are 0.45 and $0.68\text{ }^{\circ}\text{C/decade}$. Tarawneh and Chowdhury (2018) calculated T2m changes during the period 1984-2013 for three stations representing the central, northern and southwest regions of Saudi Arabia, and documented the overall warming rates of $0.58\text{--}1.25\text{ }^{\circ}\text{C/decade}$ in summer and $0.43\text{--}0.66$

°C/decade in winter. To compare with these previous estimates, the annual mean $T_{\max}$, $T_{\min}$, and DTR trends for the 21 surface stations in the DS3503 were estimated. Their station mean trends (Table 3) are found to be 0.56, 0.72, and -0.16 °C/decade, respectively, and all are statistically significant ($p < 0.05$), indicating a stronger warming rate in $T_{\min}$ than $T_{\max}$ and thus a decline in DTR. Evidently, the T2m trends shown in Figs. 2-3 and Table 3 are in the range of recent observation-based estimates over the AP.

3.2. Temperature trends in radiosonde observations

Figure 5 shows the vertical profile of temperature trends from 850 to 100 hPa during the period 1979-2018 for the 8 radiosonde stations in the RICH. Note that the RICH data below 850 hPa are not analyzed due to poor quality as done by others (e.g., Thorne et al., 2011). Despite some differences, all stations show generally consistent results: cooling trends above 200 hPa and warming trends downward in the troposphere. Among the 144 trends (9 pressure levels * 2 UTC times * 8 stations), 77 are statistically significant at $p < 0.05$, and so are all the trends in the lowest three layers (500, 700, and 850 hPa). The largest warming trend is 0.65 °C/decade at 850 hPa and the biggest cooling rate is -0.34 °C/decade at 100 hPa. The trends at 00 and 12 UTC differ slightly in the free atmosphere but stronger warming is mostly seen at 00 UTC than 12 UTC in the lower troposphere. In general, there are two warming maxima in the profile, a strong one at 850 hPa and a very weaker one centered at 200-300 hPa.

Like the surface observations, the station mean results are examined to focus on the large-scale warming features. Fig 6a shows the vertical trend profile of station mean temperature averaged

over the 8 radiosonde stations in the RICH at 00 and 12 UTC for 100-850 hPa. The trend is negative above 200 hPa but is positive and increases downward in the entire troposphere, with a maximum of 0.43 °C/decade at 850 hPa. Among the 18 trends (9 pressure levels * 2 UTC times), 13 are statistically significant at $p < 0.05$, including the lowest three layers (500, 700, and 850 hPa). The corresponding mean $T_{2m_{trend}}$, for the 8 radiosonde stations, 0.68 °C/decade (00 UTC) and 0.48 °C/decade (12 UTC) from the DS3505, is also plotted in Fig. 6a. Figure 6b shows the reconstructed warming profile based on EOF analysis to capture the large-scale features shared among all individual radiosonde stations. The first EOF explains 54.3% (00 UTC) and 49.8% (12 UTC) of the total data variance from the original 8 stations. The vertical profile in Fig. 6b is very similar to that in Fig. 6a. Again, the station mean profiles exhibit two warming maxima (i.e., R-shaped): a very weaker top-heavy one centered at 250 hPa and a strong bottom-heavy one maximizing at the surface. The former represents the well-known warming profile peaking in the tropical upper troposphere associated with the positive water vapor feedback in a warming climate (Held and Soden, 2000; IPCC, 2007; 2013). The latter represents the signal of DA that maximizes at the surface and decreases with height as also documented by multiple satellite measurements (Wei et al., 2017a).

The warming trend is consistently larger at 00 UTC than 12 UTC in the lower troposphere and this day-night warming difference increases with pressure and maximizes at the surface (Fig. 6). This diurnal warming asymmetry is small between 500-700 hPa and become more evident for the layers > 700 hPa. As the station mean PBLH at 12 UTC over the AP is ~ 2.1 km (Table 4), it is reasonable to believe that the diurnal signal of DA is limited to the lower troposphere below 700 hPa. The free atmosphere in the tropics has relatively small diurnal variations because the

atmosphere is dynamically well mixed (Sherwood et al., 2005; Byrne and O’Gorman, 2016; 2018). For example, Seidel and Free (2009) analyzed the amplitude and phase of the climatological diurnal cycle of upper-air temperatures based on four-times-daily radiosonde data from 53 stations and found that the amplitude of the annual-average diurnal cycle (half the DTR) is largest (1 to 4 °C) at the surface, decreases with height quickly, and becomes very small (<0.4 °C) at 700 hPa and above. Brocard et al. (2013) analyzed the phase and amplitude of the diurnal temperature cycle based on 53 years of radiosonde measurements from a station in Switzerland and showed a strongly decreasing amplitude with height from about 3 °C at the surface to 0.2 °C at 700 hPa and above. These results suggest that the diurnal warming asymmetry may exist mainly in the lower troposphere (> 700 hPa) and at the surface.

There are no radiosonde-based diurnal warming trends reported previously in the AP. One key question is whether the warming profiles in Figs 5-6 are robust? Historically there were some disagreements on the tropospheric warming rates observed and modeled, particularly in the tropics. Thorne et al. (2011) comprehensively assessed recent homogenized radiosonde and satellite observing systems and model results and reported an overall agreement between modeled and observed throughout the tropospheric column because of considerable progresses made recently in treating uncertainties in both. Also, among the several radiosonde datasets used, the RICH showed the best match with satellite observations. The large-scale warming profile in the tropics (Fig. 8 in Thorne et al., 2011) highlighted the cooling in the stratosphere, the overall warming through the entire troposphere, and the largest warming in the upper troposphere centered at 200-300 hPa, consistent with the estimates in the middle and upper troposphere (Fig. 6) in this study. It differs largely in the lower troposphere from the bottom-heavy warming

profile over the AP, because it is averaged from the entire tropical regions dominated by oceans and the latter is from the moisture-limited deserts only. The bottom-heavy profile associated with DA was also documented by comparing the warming profile between rainforests and deserts using multiple satellite datasets (Wei et al. (2017a). Similar bottom-heavy warming profiles are also reported over Armenia with a dry highland continental climate (Gevorgyan, 2014).

3.3. Temperature trends in ERA5 and MERRA-2

It is essential that the reanalyses can capture some major T2m features observed before being used for analysis. Figure 4a shows interannual variations of station mean T2m anomalies at 00 and 12 UTC for 1979-2018 averaged over the 21 surface stations from the DS3505. The corresponding station mean T2m anomalies calculated from the ERA5-land and MERRA-2 are shown in Figs. 4b and 4c. Both reanalyses show warming trends that are statistically significant ($p < 0.0001$) at both UTC times. For the ERA5, the warming trend is $0.44\text{ }^{\circ}\text{C/decade}$ at 00 UTC and $0.41\text{ }^{\circ}\text{C/decade}$ at 12 UTC. The corresponding values for the MERRA-2 are $0.39\text{ }^{\circ}\text{C/decade}$ and $0.36\text{ }^{\circ}\text{C/decade}$, respectively. Evidently, the reanalyses capture well the major observed features of interannual variability and overall warming trends, but have a cooling bias, compared to the observed warming rate of $0.60\text{ }^{\circ}\text{C/decade}$ at 00 UTC and $0.57\text{ }^{\circ}\text{C/decade}$ at 12 UTC. As discussed previously, the surface warming rate difference (00 vs. 12 UTC) is not large because both UTC times do not correspond to the times with the maximum/minimum warming rate in T2m (Figs. 2-3).

The spatial patterns of T2m trend at 00 and 12 UTC from the ERA5-land (1981-2018) and MERRA-2 (1980-2018) over the SDAP and surrounding areas are shown in Fig. 7 to illustrate the large-scale warming patterns of DA and so the warming in the AP can be put in a proper context. Significant warming ($p < 0.05$) is widespread at both UTC times and the strongest trends at ~ 0.5 °C/decade are seen over a broad contiguous swath of land covering the entire Sahara and Arabian deserts. The warming is larger at 00 UTC than 12 UTC. These warming features are similar in the ERA5-land and MERRA-2. Again, the reanalyses have a systematic cooling bias in the warming rates but capture well the essential spatial features of DA (Zhou et al., 2015; 2016).

Figure 8 shows the diurnal cycle of $T2m_{trend}$ and climatological T2m from the ERA5-land (1981-2018) and MERRA-2 (1980-2018), at the station and regional mean levels over the AP. All hourly trends are statistically significant at $p < 0.05$. For the station mean T2m from the ERA5-land (Fig. 8a), the hourly warming trend ranges 0.41-0.45 °C/decade, with a diurnal range of 0.04 °C/decade and a 24-hour average of 0.43 °C/decade. The climatology of T2m has a minimum value of 17.4 °C at 02 UTC and a maximum value of 29.8 °C at 11 UTC, while the warming trend maximizes at 21 UTC and minimizes at 14 UTC. Similar diurnal features are seen for the regional mean T2m (Fig. 8c). For the station mean T2m from the MERRA-2 (Fig. 8b), the warming trend ranges 0.36-0.39 °C/decade, with a diurnal range of 0.03 °C/decade and a 24-hour average of 0.38 °C/decade; the climatology of T2m has a minimum value of 17.9 °C at 02 UTC and a maximum value of 31.4 °C at 11 UTC, while the warming trend maximizes at 00 UTC and minimizes at 08 UTC. The regional mean plot (Fig. 8d) shows similar features to the station mean plot except the minimum warming trend at 09 UTC.

Clearly, the reanalyses (Fig. 8) underestimate the observed warming rates (Fig. 3a), particularly the maximum warming rate and the diurnal asymmetry of warming. The observed climatology of T2m reaches the minimum of 18.7 °C at 03 UTC and the maximum of 30.0 °C at 12 UTC, while the warming trend maximizes at 05 UTC and minimizes at 15 UTC. The reanalysis climatology in T2m reproduces the observed diurnal range but differs by 1-2 hours in the minimum and maximum values than observed. Like the observations (Fig. 3), the reanalysis warming rate is generally in opposite phase with the climatological T2m values, indicating that the largest warming at nighttime and the smallest warming at daytime. However, the reanalyses differ by several hours in the maximum/minimum from the observations and from each other. These discrepancies are likely due to the differences in spatial resolution (point measurements versus coarse-resolution grid averaged data) and reanalysis deficiencies in modeling the surface radiative forcing, surface energy partitioning, and PBL mixing (section 3.5).

To compare with recent T2m trend estimates available over the AP (e.g., Alghamdi and Moore, 2014; Krishna, 2014; Tarawneh and Chowdhury, 2018), the annual mean climatology and trends of daily $T_{\max}$, $T_{\min}$, and DTR in the two reanalyses were calculated for the 21 surface stations in the AP (Table 3). Interestingly, the station mean reanalysis climatology is comparable to the observed values. The station mean trends for the $T_{\max}$, $T_{\min}$, and DTR are 0.42, 0.44, and -0.02 °C/decade, for the ERA5-land, and 0.37, 0.39, and -0.02 °C/decade for the MERRA-2, respectively. The corresponding observed trends are 0.56, 0.72, and -0.16 °C/decade, for the DS3505. Note that all the trends except these for the reanalysis DTR are statistically significant ($p < 0.05$). As discussed previously, the reanalyses capture the observed diurnal cycle of T2m warming and larger warming trends in $T_{\min}$ than $T_{\max}$, but largely underestimate the warming

rates, particularly at nighttime, and the magnitude of the diurnal asymmetry of warming. Also, the reanalysis DTR trend is substantially smaller than observed. These reanalysis biases are more pronounced in the MERRA-2 than the ERA5.

Figure 9 shows the vertical profile of temperature trend at 00 and 12 UTC from the ERA5 (1979-2018) and MERRA-2 (1980-2018), along with the corresponding T2m trends at the surface. The reanalyses show cooling in the stratosphere above 200 hPa and increasing warming through the entire troposphere, and two evident warming maxima in the vertical profile, one in the UT around 250 hPa, and the other near the surface. Overall, the reanalysis profile (Fig. 9) exhibits broadly similar vertical warming patterns (i.e., R-shaped) as observed (Fig. 6) over the AP. However, the observations (Fig. 6) show small warming differences between 00 and 12 UTC above 700 hPa, and stronger warming rate at 00 UTC than 12 UTC downward, while the reanalyses exhibit stronger warming at 12 UTC than 00 UTC between 500-850 hPa and a transition to the opposite below 900 hPa to the surface. Also, compared to the observations, the reanalysis generally has warming biases in the middle and upper troposphere and cooling biases in the lowest tropospheric layers, which are particularly evident in the MERRA-2.

3.4. Possible drivers of the diurnal and vertical warming features

The warming rate of T2m depends on surface radiative forcing and various response and feedback processes. At the global scale, T2m has generally increased more over drier regions and faster at nighttime than daytime in a warming climate, and this spatial and diurnal variability has been mostly explained by large-scale greenhouse effects and local to regional changes in cloud cover, precipitation, soil moisture, and vegetation (IPCC 2007; 2013; Zhou et al. 2007; 2009,

2010; Dirmeyer et al. 2013; Lindvall and Svensson, 2015; Davy and Esau, 2016; Wei et al., 2017b). For example, drier regions with less soil moisture and vegetation are associated with higher Bowen ratios and tend to experience larger warming rates due to less local evaporative cooling. These factors affect T2m differently by altering the land surface energy and hydrological balances over different regions, but their effects on surface warming are limited over the extremely dry and mostly cloud-free deserts. Land use change such as urbanization in the AP could affect the diurnal warming asymmetry. Urbanization generally creates urban heat island in most cities, but urban areas often exhibit cooler temperatures at day (urban cool island) and warmer temperatures at night than suburbs over deserts (Bounoua et al., 2009; Lazzarini et al., 2013; 2015). However, observational studies (Almazroui et al., 2013; Alghamdi and Moore, 2014) showed that urbanization lessened the warming rate in urban areas compared to surrounding rural areas but has not substantially contributed to the large-scale warming trends observed throughout Saudi Arabia.

Large-scale warming and moistening in response to increasing GHGs have been widely used to explain surface and atmospheric warming (IPCC 2007; 2013). Increased DLR in a warming climate associated with this global-scale greenhouse effect has been proposed as the primary surface radiative forcing for the DA over the SDAP (Zhou et al., 2015; 2016; Cook and Vizzy 2015; Zhou 2016; Evan et al., 2017; Wei et al., 2017a). In addition, the AP is one of significant dust sources in the world, and changes in DSR associated with dust aerosols can affect local temperature by modifying the radiative forcing via direct effects and feedback of desert dust (Islam and Almazroui, 2012). It is very likely that the radiative forcing associated with changes

in DLR and DSR may be the main drivers for the overall surface warming and its spatial and diurnal variations associated with the DA.

Besides the radiative forcing at the surface, the diurnal cycle of T2m warming is also tightly connected to the extent of turbulent mixing in the PBL, which is described by the PBLH (McNider et al., 2012; Dirmeyer et al. 2013; Davy and Esau 2014a, 2014b; 2016; Svensson and Lindvall, 2015; Wei et al., 2017b). As the lowest part of the atmosphere, the PBL controls and in turn, responds to the diurnal evolution of near-surface thermodynamic variables through turbulent exchanges of momentum, heat and moisture in the coupled land-atmosphere interface (Wei et al., 2017b). Recent studies highlighted that the PBLH modulates the T2m response to surface forcing and is a strong predictor of the strength of $T2m_{trend}$ (McNider et al. 2012; Dirmeyer et al. 2013; Davy and Esau 2014a, 2014b; 2016; Wei et al., 2017b). Davy et al. (2017) proposed a PBL-response mechanism that for a given forcing, the surface warming rate depends inversely on the PBLH and so is stronger for a shallower PBLH. It is interpreted as the forcing efficacy determined by the effective heat capacity of the atmosphere, which is defined by the PBLH. This mechanism helps to explain why the nighttime has warmed more rapidly than the daytime in observations across different regions. Its effect is expected to be most pronounced in regions such as the SDAP where there is a strong diurnal cycle in the PBLH, with an extremely shallow SBL forming at night. Hence it is possible that the diurnal and vertical warming features associated with the DA may be also tied to the diurnal evolution of PBLH over the AP.

Fig. 10 shows the station and regional mean diurnal cycle of climatological PBLH and T2m over the AP from the ERA5 and MERRA-2, respectively. In general, the diurnal phase of PBLH

follows closely that of T2m, but with a delay of ~1 hour, and the PBLH differs little between the station and regional mean results for each reanalysis. The PBL is deepest in the late afternoon at 12 UTC after the T2m reaches the daily maximum at 11 UTC. It is shallowest in the early morning at 03-04 UTC, 1-2 hours after the T2m reaches the daily minimum at 02 UTC. Combined with the relationship between $T2m_{trend}$ and climatological T2m shown in Figs 3 and 8, Fig. 10 supports generally the PBL-response mechanism that the diurnal cycle of surface warming rate depends inversely on the climatological PBLH and DA has a distinct diurnal asymmetry – the stronger warming rate for a shallower PBLH. To check this further, Fig. 11 shows the histogram of $T2m_{trend}$ as a function of $PBLH_{climate}$ from the ERA5-land (1981-2018) and MERRA-2 (1980-2018). The PBLH for all grids over the rectangle study domain are divided evenly into 10 bins based on a PBLH interval of 350 m in the ERA5 and 300 m in the MERRA2. Evidently, $T2m_{trend}$ decreases with increasing $PBLH_{climate}$, indicating an inverse relationship between $T2m_{trend}$ and $PBLH_{climate}$.

However, the $T2m_{trend}$ demonstrates some level of spatiotemporal heterogeneity and so do the $PBLH_{climate}$ and surface radiative forcing over the AP. It is difficult to establish the main drivers of $T2m_{trend}$ simply based on the station and regional mean diurnal cycle shown above, without considering the confounding impacts of radiative forcings and PBLH. Next, a multiple regression model (equation (2)) is used to explore the empirical relationships between the diurnal and spatial variance in $T2m_{trend}$ and three major contributing variables, DLR_{trend} , DSR_{trend} , and $PBLH_{climate}$ in the reanalyses. Note that a similar analysis cannot be done for the observations due to the lack of surface radiative forcing data. The magnitude of the standardized partial

regression coefficients β_1 through β_3 tell the relative contribution of the three variables to $T2m_{trend}$ and the sign of these coefficients tells the direction of their impacts on $T2m_{trend}$.

Table 4 lists the statistical results for the regression model by considering the diurnal and/or spatial variations in $T2m_{trend}$ for three cases: (i) the diurnal data over all the land grids within the study domain (i.e., the diurnal + spatial variation), (ii) the diurnal mean data over all the land grids within the study domain (i.e., the spatial variation), and (iii) the spatial mean diurnal data averaged over all the land grids within the study domain (i.e., the diurnal variation). The different cases are used to separate the contributions of the diurnal and spatial variations in the regression. Also, two types of PBLH are considered: (i) the PBLH provided by the reanalysis and (ii) the PBLH diagnosed from the reanalysis using the Ri method (i.e., $PBLH_{RI0.25}$). It is expected that $PBLH_{climate}$ matters most in the diurnal domain and DLR_{trend} and DSR_{trend} dominate in the spatial domain.

In the ERA5, when the reanalysis-derived PBLH is used, the regression coefficients for case (i) are 0.22, 0.85, and 0.22, for $1/PBLH_{climate}$, DLR_{trend} , and DSR_{trend} , respectively, and are all statistically significant ($p < 0.0001$), indicating that the radiative forcing of DLR has the dominant impact (positive) on $T2m_{trend}$, followed by $1/PBLH_{climate}$ (positive) and DSR_{trend} (positive). For case (ii), the radiative forcing of DLR has the dominant impact (positive) on $T2m_{trend}$, followed by DSR_{trend} (positive), while $PBLH_{climate}$ has a negligible effect. For case (iii), $1/PBLH_{climate}$ has the dominant impact (positive) on $T2m_{trend}$, followed by DLR_{trend} (positive), while DSR_{trend} is negligible. When the $PBLH_{RI0.25}$ is used, the regression results are nearly identical to these based on the reanalysis-derived PBLH. In the MERRA2, the results for all three cases agree

mostly with those in the ERA5, except that the impact of $PBLH_{climate}$ is weaker and the effect of DSR_{trend} is slightly stronger. The F test indicates that the overall regression model is statistically significant ($p < 0.0001$) for all cases in Table 4. The $R^2_{adjusted}$ value shows that 56%-99% of the data variance in $T2m_{trend}$ can be explained by the regression model. These statistical results suggest that the radiative forcing of DLR has the dominant effect on $T2m_{trend}$ in the spatial domain and $PBLH_{climate}$ could have the most control of the strength of the temperature response to the forcing in the diurnal domain.

The multiple regression results are generally consistent in the sign, magnitude, and significance of the three regression coefficients between the two reanalyses. However, the diurnal timing of maximum and minimum warming in $T2m$ differs between observed and reanalyzed. Also the reanalysis vertical profile shows some warming/cooling biases, and the $PBLH_{climate}$ has a less important role in explaining $T2m_{trend}$ in the MERRA-2 than in the ERA5. It is reasonable to assume that the free atmosphere (and the DLR forcing as well) in the tropics has relatively smaller diurnal variations than the PBLH because it is dynamically well mixed (Sherwood et al., 2005; Byrne and O’Gorman, 2016; 2018). Hence the reanalysis diurnal and vertical discrepancies may be tied more to the systematic biases in PBLH as described next.

3.5. Uncertainties in PBLH

A key question is whether the reanalysis PBLH diurnal cycle is reliable? So far, there are only two papers reported the diurnal cycle of PBLH over the AP in the literature. Abdel-Aal and Shonoda (2014) analyzed hourly meteorological data for the period 2009-2012 at Qurayyat

Province in Saudi Arabia and showed a strong seasonal variation in PBLH, ranging 1.4-1.9 km for the deepest CBL and 100-300 m for the shallowest SBL. Li (2019) used a high-resolution mesoscale model to simulate the diurnal and seasonal cycle of PBLH for three representative sites in Saudi Arabia and found a large seasonal variation of 0.6-3.8 km for the deepest CBL and 50-250 m for the shallowest SBL. However, these two studies also documented large diurnal, seasonal and spatial variations in the phase and magnitude of PBLH in the AP due to differences in local conditions. Hence, the estimated PBLHs from these two papers are inadequate to validate the reanalysis results in Fig. 10. One major feature in Fig. 10 is the systematic higher PBLH values in the MERRA-2 than the ERA5 at every hour in a range from a few hundred meters to over 1 km. The station-mean PBLH ranges 131-2211 m (790-2564 m), with a daily mean of 763 m (1554 m), in the ERA5 (MERRA-2). In particular, the MERRA-2 nocturnal PBLH is much larger than any estimates reported over the AP (Abdel-Aal and Shonoda, 2014; Li, 2019) and previous studies have documented that the MERRA-2 PBLH is biased high (e.g., Salmun et al., 2018; Ding et al., 2019).

It is difficult to judge the reliability of the reanalysis PBLH data because different methods are used to estimate the PBLH. Next, the PBLH is diagnosed using the same RI method for comparison among all three datasets. Table 5 lists the PBLH_Ri_{0.25} for the 8 radiosonde stations. In general, the reanalysis derived PBLH in the ERA5 is broadly comparable with PBLH_Ri_{0.25} at 00 and 12 UTC estimated from the IGRA2 and ERA5 profiles. This is expected as the RI method was also used in the ERA5 PBLH estimates although some adjustments were applied to this method (C3S, 2017). The reanalysis derived PBLH in the MERRA-2 are broadly comparable with PBLH_Ri_{0.25} at 12 UTC, but systematically higher than the PBLH_Ri_{0.25} at 00 UTC

estimated from the IGRA2 and MERRA-2 profiles. For example, the station mean PBLH_Ri_{0.25} at 12 UTC is 2090 m in IGRA2, 2238 m in ERA5 and 2665 m in MERRA-2, indicating a mean overestimate of 148 m in the ERA5 and by 575 m in the MERRA-2. The station mean PBLH_Ri_{0.25} at 00 UTC is 102 m in IGRA2, 23 m in ERA5 and 20 m in MERRA-2, which are generally consistent with the reanalysis-derived PBLH of 129 m in ERA5 but much lower than 849 m in the MERRA-2. The reanalysis PBLH_Ri_{0.25} at 00 UTC may be underestimated comparing to the IGRA2-based estimates due to the coarse vertical resolution of data used and the difficulties in estimating the nocturnal shallow SBL over the deserts (see more discussion later). Nevertheless, the reanalysis derived PBLH in the MERRA2 is systematically overestimated as also indicated by previous estimates over deserts (e.g., McGrath-Spangler et al., 2015; Wei et al., 2019b).

To examine this further, Fig. 12 shows the scatter plots of climatological PBLH between reanalysis-derived versus PBLH_Ri_{0.25} from the ERA5 (1979-2018) and MERRA-2 (1980-2018) by including all land grids within the study domain. For the ERA5, the correlation coefficient $R=0.97$ is statistically significant ($p<0.0001$) for a sample size of 21,144, indicating very good performance in the PBLH provided by the ERA5. For the MERRA-2, the corresponding $R=0.83$ is also statistically significant ($p<0.0001$) for a sample size of 2,884, indicating a good performance in the PBLH provided by the MERRA-2 as well. However, the MERRA-2 has much lower R , mainly due to significant overestimates of the PBLH at 00 UTC and 18 UTC than PBLH_Ri_{0.25}, indicating systematic positive biases in the MERRA-2 PBLH at nighttime.

The reanalysis PBLH is a model-based estimate and so is prone to biases due to model deficiencies. Modeling tests by McGrath-Spangler and Molod (2014) and McGrath-Spangler et al. (2015) showed large differences in the PBLH estimated by the two methods used in the ERA5 and MERRA-2, with maximum discrepancies in the nocturnal depth by as much as 1 km over northern Africa, which are similar to the results in Fig. 10. In numerical models, PBLH biases could manifest themselves as biases in surface warming and lapse rates in the troposphere (McGrath-Spangler et al., 2009; McGrath-Spangler and Denning, 2010; Svensson and Lindvall 2015; Wei et al., 2017b). Therefore, the major PBLH biases in the reanalysis could explain, at least partially, some of the afore-mentioned surface and atmospheric temperature trend biases in the magnitude and phase. Next, two examples are used to establish this possibility.

The first example is related to the reanalysis biases in the vertical warming profile (Fig. 9). The reanalyses generally have warming biases in the middle and upper troposphere and cooling biases in the lowest tropospheric layers, and a faster warming rate at daytime than nighttime at 500-900 hPa, particularly in the MERRA-2. It is well known that the Earth is mainly warmed bottom up, as most solar radiation is absorbed at the surface and this energy is transmitted through the rest of the atmosphere via PBL processes. Over the desert, sensible heat dominates and drives the PBL growth and there are strong correlations between PBLH and surface temperatures (Wei et al., 2017b). Differences in turbulent mixing result in different vertical redistributions of heat, which controls the vertical temperature profile or the lapse rate in the atmosphere. During the daytime, when the PBLH is higher, the vertical turbulent mixing is stronger and deeper and so more surface sensible heat can be transferred into upper atmospheric layers, leading to a cooler surface and warmer atmosphere. During the nighttime, when the

PBLH is higher, the downgradient turbulent diffusion is stronger and so more sensible heat will be transferred downward into lower atmospheric layers and the ground, leading to a warmer surface and a cooler atmosphere. The net effect of higher PBLH is to create a smaller DTR, a warmer (cooler) middle and upper troposphere (lower troposphere and land surface), and a warmer (cooler) daytime at daytime (nighttime) because the daytime impacts dominates. This appears to explain well the warming and cooling biases in the reanalysis, particularly the MERRA-2 because of its biased high PBLH. McGrath-Spangler and Molod (2014) indicated that the bulk Richardson number method better represented the PBLH over the Sahara than the other methods. This is also supported by the $PBLH_{RI0.25}$ estimated in Table 5, implying that the PBLH is more realistic in the ERA5 than in the MERRA-2 over the AP. As expected, the warming rates near the surface and in the atmosphere in the ERA5 are closer to those observed than the MERRA-2.

The second example is related to the reanalysis warming biases in T2m at nighttime when the reanalysis PBLH differs most. As shown previously, the reanalyses substantially underestimate nighttime warming and thus the diurnal warming asymmetry. For example, the station mean maximum warming rate is 0.73 °C/decade for the observations, 0.45 °C/decade for the ERA5, and 0.39 °C/decade for the MERRA-2. The biased high PBLH in the MERRA-2 will, according to the PBL-response mechanism, result in a biased low warming rate as shown in Fig. 8. Similar findings of overestimated PBLH under stable stratification and consequentially underestimated temperature trends are also reported in other observational and modeling studies (Seidel et al., 2012; McNider et al., 2012; Davy and Esau, 2016; Davy et al., 2017).

888 Accurate modeling of PBL processes is important in describing land-atmosphere interactions and
889 the diurnal and vertical temperatures. PBLH is one key measure of the strength of these
890 processes but lacks a unified definition and different PBLH estimation methods can produce
891 substantially different values, even for the same atmospheric profile (e.g., Seidel et al., 2010;
892 McGrath-Spangler and Molod, 2014). For example, McGrath-Spangler and Molod (2014)
893 compared seven PBLH estimation methods in the Goddard Earth Observing System (GEOS-5)
894 atmospheric general circulation model over land and identified the largest variations in the
895 nocturnal PBLH. McGrath-Spangler et al. (2015) further quantified the impacts of different
896 PBLH estimates within the GEOS-5 model on the turbulent length scale and the simulated
897 climate, and found that near-surface variables such as wind, temperature and humidity were
898 sensitive to the PBLH differences and such sensitivity was spatially and temporally
899 heterogeneous. Unfortunately, current numerical models have difficulties and large uncertainties
900 in representing key PBL processes, particularly in extreme and complex PBL conditions such as
901 the SDAP (Cuesta et al., 2009; Garcia-Carreras et al. 2013; Holtslag et al. 2013; Wei et al.,
902 2017b; Ao et al., 2017). For example, Garcia-Carreras et al. (2015) detailed a very complicated
903 picture of the vertical structure and diurnal evolution of the Saharan PBL using aircraft and
904 radiosonde measurements and a large-eddy simulation model; Gamo (1996) showed that the
905 thick CBL often has a weakly stable and nearly neutral stratification in the Sahara in the whole;
906 Flamant et al. (2007) found that the Saharan residual layer can be maintained for a whole day
907 sometimes. The subtle vertical structure of the Saharan PBL, particularly the small temperature
908 inversion and deep near-neutral residual layer, and its diurnal evolution, add further challenges
909 for PBL modeling (Cuesta et al., 2009; Couvreux et al. 2014). These complex PBL features are
910 identified in the Sahara, and possibly apply to the AP and other hot deserts as well. For example,

Ao et al. (2017) analyzed the diurnal variation of PBLH from two intense observation periods of experiments in summer of the Badain Jaran Desert and found that the deep CBL showed a diurnal variation of three- to five-layer structure in clear days and five-layer structure often around sunset or sunrise. Hence, the deficiencies and uncertainties in the reanalysis PBL processes can result in temperature biases and the timing differences in maximum and minimum warming.

The surface and atmospheric temperature changes in response to external forcings are a result of complex interactions among the atmosphere, PBL and land surface. Considering the complexity of turbulent mixing and the challenges in observing and modeling the PBL processes, it is very difficult to attribute the reanalysis biases in the fully coupled land-atmosphere system. For example, one major reanalysis bias discussed previously is the systematic underestimation of DTR and nighttime temperature trend in T2m. Although the ERA5 has more realistic PBLH than the MERRA-2, this systematic bias is smaller but still there. This is a long-standing issue in reanalysis and numerical models despite intensive attribution studies on this topic (e.g., Vose et al., 2005; Zhou et al., 2007; 2009; 2010; IPCC, 2007; 2013; Christensen et al., 2008; Boberg and Christensen, 2012; Lewis and Karoly, 2013; Wei et al., 2017a; 2017b; Du et al., 2018; Davy, 2018). In addition, other non-PBL processes may modulate the diurnal and vertical features of DA as well. For example, the reanalysis data represents a mean over the model grid-box on the order of hundreds of squared kilometers, while the station data come from point measurements by instruments over standard grass plots. Also, the surface elevation differs largely between the station sites and the model grids (Table 2). As the focus of the present study is the detection of DA in the context of AP, further attribution, however, is beyond the scope of this study.

4. Conclusions

This paper presents a comprehensive analysis of hourly surface observations, radiosonde temperature measurements, and two latest state-of-the-art reanalysis products (ERA5 and MERRA-2) for the period 1979-2018 to understand the diurnal cycle and vertical structure of DA over the AP. The diurnal cycle of T2m, PBLH, and surface fluxes and the atmospheric warming profiles from near surface (2m) to 100 hPa are analyzed. Observational and reanalysis data show consistently that DA is a bottom-heavy warming profile with a distinct diurnal asymmetry, which maximizes near the surface and decreases quickly with height and is limited to the lower troposphere and surface. The main findings are summarized as follows:

1. Observed and reanalysis data reveal consistent warming trends in T2m that are statistically significant for all stations and during every hour of day. The station mean surface warming rates in T2m are 0.59 °C/decade, with a strong diurnal range of 0.45-0.73 °C/decade for the observations. The corresponding values are 0.43 °C/decade, with a weak diurnal range of 0.41-0.45 °C/decade, for the ERA5, and 0.38 °C/decade, with a weaker diurnal range of 0.36-0.39 °C/decade for the MERRA-2. The reanalysis data capture well the overall warming and interannual variability but underestimate the warming rates and the diurnal asymmetry of warming.

2. Observed and reanalysis T2m data show the diurnal asymmetry of warming associated with PBLH over the AP. In general, the diurnal cycle of surface warming rate depends, to some

957 extent, inversely on the magnitude of climatological PBLH, which follows closely the diurnal
958 cycle of T2m, but differ by several hours in the diurnal phase. The surface observations show the
959 largest/smallest warming mostly around the transitions between day and night, while the
960 reanalysis data indicate the largest/smallest warming at nighttime/daytime.

961
962 3. Both surface and radiosonde observations indicate that DA is a bottom-heavy warming profile
963 limited to the lower troposphere and surface, and has a distinct diurnal asymmetry that
964 maximizes near the surface, decreases with height, and is mostly invisible above 700 hPa. The
965 reanalysis data capture this vertical warming profile but exhibit some biases.

966
967 4. The diurnal and vertical warming features could be, at least partially explained by a PBL-
968 response mechanism (Davy et al., 2017): for a given forcing, the surface warming rate depends
969 inversely on the PBLH and so is stronger for a shallower PBLH. The diurnal phase and the
970 magnitude of PBLH over the AP generally determines the diurnal warming asymmetry and its
971 vertical structure.

972
973 5. The major PBLH biases in the reanalysis could cause, at least partially, some of the surface
974 and atmospheric temperature trend biases in the magnitude and phase. In particular, the PBLH
975 estimated from the MERRA-2 is systematically higher than that from the ERA5, which could
976 help to explain some the warming biases in the middle and upper troposphere and cooling biases
977 in the lowest tropospheric layers.

979 These results suggest that besides the surface radiative forcing, the PBL turbulent mixing may
980 play an important role in modulating the diurnal and vertical structure of DA over the AP. The
981 reported warming trends and results of PBLH are in good agreement with theory and previous
982 findings in the literature. While other factors may have also asymmetrically affect the diurnal
983 temperature trends, this study concentrates on the PBLH as the PBL-response mechanism is
984 expected to maximize over the SDAP with the world's deepest daytime PBL and shallowest
985 nocturnal PBL (Davy et al., 2017). The role of PBL is to amplify the diurnal surface warming
986 stronger for a shallower PBLH over the deserts through heat redistributing via turbulent mixing.

987
988 To the best of my knowledge, this work is the very first comprehensive study to examine the
989 diurnal and vertical variations of warming trends over the AP and establish their relationships
990 with the PBLH. It highlights the importance and need for accurate descriptions of the PBL
991 processes with respect to the turbulent mixing in order to better characterize the temperature
992 diurnal cycle changes in reanalysis products and numerical models (Wei et al., 2017b). However,
993 the interactive mechanisms between near-surface temperatures and PBL processes are very
994 complex and it has been very challenging to establish cause and effect for a fully coupled land-
995 atmosphere system. The present work links the diurnal cycle warming rate to PBLH in a
996 statistical framework but there are large uncertainties in observational and reanalysis PBLH
997 estimates. Further attribution studies are needed to confirm its findings by conducting carefully
998 defined modeling sensitivity experiments.

999
1000 The reported findings have important implications as DA may accelerate over the arid and semi-
1001 arid regions in the context of global warming and has the strongest impacts on the SDAP (Zhou,

2016). Climate change is an important factor for sustainable water resource management and is an essential component for strategic water resource management in arid and semi-arid countries (Tarawneh and Chowdhury, 2018). The long, hot and dry weather, along with extreme temperatures, impose a significant strain on water resources as warming enhances evaporation from open reservoirs and domestic water demand. The climate of the AP is extremely arid, with high temperature variability, low annual rainfall, no natural perennial flow and limited groundwater reserves (Attada et al., 2017). As a result, the AP is extremely sensitive to climate fluctuations and is also highly vulnerable to climate change impacts (Almazroui et al., 2013; Attada et al., 2017). Understanding and predicting the AP climate can be beneficial for practical purposes in many different sectors, including water resources, agriculture, power generation, biodiversity, tourism, ecosystems, migration and food security (Almazroui et al., 2012; 2014).

Acknowledgements. L.Z. was supported by National Science Foundation (NSF AGS-1952745 and AGS-1535426).

References

- Abdel-Aal, M. and Shonoda, E., 2014. Environmental assessment for Qurayyat province in northern Saudi Arabia, Part I- determination of atmospheric stability & boundary layers. *Journal of Selçuk University Natural and Applied Science* (Online ISSN: 2147-3781), 777-787.
- Ahmed, B.Y.M., 1997. Climate classification of Saudi Arabia: An application of factor-cluster analysis. *GeoJournal*, 41.1, 69-84.
- Alghamdi, A.S., and Moore, T.W., 2014. Analysis and comparison of trends in extreme temperature indices in Riyadh City, Kingdom of Saudi Arabia, 1985–2010, *J. Climatol.*, 2014(1), 1–10.
- Almazroui, M, Islam, M.N., Jones, P.D., 2013. Urbanization effects on the air temperature rise in Saudi Arabia. *Clim. Change* 120: 109– 122, DOI: 10.1007/s10584-013-0796-2.
- Almazroui, M., Islam, M.N., Athar, H., Jones, P.D., Rahman, M.A., 2012. Recent climate change in the Arabian Peninsula: annual rainfall and temperature analysis of Saudi Arabia for 1978–2009. *Int. J. Climatol.* **32**: 953–966. DOI: 10.1002/joc.3446.
- Angell, K.J., 1999. Comparison of surface and tropospheric temperature trends estimated from a 63-station radiosonde network, 1958–1998. *Geophys. Res. Lett.* 26: 2761– 2764.
- Ao, C.O., Waliser, D.E., Chan, S.K., Li, J.-L., Tian, B., Xie, F., and Mannucci, A.J., 2012. Planetary boundary layer heights from GPS radio occultation refractivity and humidity profiles, *J. Geophys. Res.*, 117, D16117, doi:10.1029/2012JD017598.

1037 Ao, Y., Li, J., Li, Z., Lyu, S., Jiang, C., and Wang, M., 2017. Relation between the atmospheric
 1038 boundary layer and impact factors under severe surface thermal conditions. *Advances in*
 1039 *Meteorology*, vol. 2017, *Gale Academic OneFile*. <https://doi.org/10.1155/2017/8352461>.
 1040 Athar, H., 2014. Trends in observed extreme climate indices in Saudi Arabia during 1979–2008,
 1041 *Int. J. Climatol.* 34 (5), 1561–1574.
 1042 Attada, R., Dasari, H.P., Chowdary, J.S., Yadav, R.K., Omar, K., Ibrahim, H., 2018. Surface air
 1043 temperature variability over the Arabian Peninsula and its links to circulation patterns. *Int J*
 1044 *Climatol.* 2019. 39:445–464.
 1045 Bjornsson, H., and Venegas, S. A., 1997. A manual for EOF and SVD analyses of climatic data,
 1046 Feb. 1997, 52 pages. CCGCR Report No. 97-1.
 1047 Boberg, F. & Christensen, J. H., 2012. Overestimation of Mediterranean summer temperature
 1048 projections due to model deficiencies. *Nat. Clim. Change* 2, 433–436,
 1049 doi:10.1038/NCLIMATE1454 (2012).
 1050 Bounoua, L., Safia, A., Masek, J., Peters-Lidard, C., Imhoff, M., 2009. Impact of urban growth
 1051 on surface climate: a case study in Oran, Algeria. *J Appl Meteorol* 48(2):217–231.
 1052 Brocard, E., P. Jeannet, M. Begert, G. Levrat, R. Philipona, G. Romanens, and S. C. Scherrer,
 1053 2013. Upper air temperature trends above Switzerland 1959–2011, *J. Geophys. Res. Atmos.*,
 1054 118, 4303–4317, doi:10.1002/jgrd.50438.
 1055 Byrne, M. P. and P.A. O’Gorman, 2018. Trends in continental temperature and humidity directly
 1056 linked to ocean warming. *Proc. Nat. Acad. Sci.* doi: 10.1073/pnas.1722312115.
 1057 Byrne, M. P. and P.A. O’Gorman, 2016. Understanding decreases in land relative humidity with
 1058 global warming: Conceptual model and GCM simulations. *J. Climate*, 29:9045–9061. doi:
 1059 10.1175/JCLI-D-16-0351.1.

1060 Chowdhury, S., Al-Zahrani, M., 2013. Implications of climate change on water resources in
 1061 Saudi Arabia. *Arab. J. Sci. Eng.*, 38, 1959–1971.

1062 Christensen, J. H., Boberg, F., Christensen, O. B. & Lucas-Picher, P., 2008. On the need for bias
 1063 correction of regional climate change projections of temperature and precipitation. *Geophys.*
 1064 *Res. Lett.* **35**, L20709, doi:10.1029/2008GL035694.

1065 Copernicus Climate Change Service (C3S), 2017. ERA5: Fifth generation of ECMWF
 1066 atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate
 1067 Data Store (CDS), date of access. <https://cds.climate.copernicus.eu/cdsapp#!/home>

1068 Copernicus Climate Change Service (C3S), 2019. C3S ERA5-Land reanalysis. Copernicus
 1069 Climate Change Service, date of access. <https://cds.climate.copernicus.eu/cdsapp#!/home>

1070 Cuesta, J., Marsham, J.H., Parker, D.J., and Flamant, C., 2009. Dynamical mechanisms
 1071 controlling the vertical redistribution of dust and the thermodynamic structure of the west
 1072 Saharan atmospheric boundary layer during summer, *Atmos. Sci. Lett.*, 10, 34– 42,
 1073 doi:10.1002/asl.207.

1074 Davy, R., Esau, I., 2014a: Surface air temperature variability in global climate models. *Atmos Sci*
 1075 *Lett* 15:13–20. doi:10.1002/asl2.456.

1076 Davy, R., Esau, I., 2014b: Global climate models' bias in surface temperature trends and
 1077 variability. *Environ Res Lett* 9:114024. doi:10.1088/1748-9326/9/11/114024.

1078 Davy, R., 2018. The climatology of the atmospheric boundary layer in contemporary global
 1079 climate models. *J. Climate*, **31**, 9151–9173, <https://doi.org/10.1175/JCLI-D-17-0498.1>

1080 Davy, R., and Esau, I., 2016. Differences in the efficacy of climate forcings explained by
 1081 variations in atmospheric boundary layer depth. *Nat Commun* **7**, 11690.
 1082 <https://doi.org/10.1038/ncomms11690>.

1083 Davy, R., Esau, I., Chernokulsky, A., Outten, S., and Zilitinkevich, S., 2017. Diurnal asymmetry
 1084 to the observed global warming. *Int. J. Climatol.* doi:10.1002/joc.4688.

1085 Ding, F., Iredell, L. F., Theobald, M., Shen, S., Ostrenga, D., Wei, J. C., Meyer, D. J., Planetary
 1086 Boundary Layer Height from AIRS, MERRA-2, and GPS Radio Occultation Data Products
 1087 at NASA GES DISC, and Insights from Their Profiles Intercomparison. American
 1088 Geophysical Union, Fall Meeting 2019, abstract #A11T-2835, December 2019.

1089 Dirmeyer, P.A., Y. Jin, B. Singh, and X. Yan, 2013. Trends in land-atmosphere interactions from
 1090 CMIP5 simulations, *J. Hydrometeorol.*, 829–849, doi:10.1175/JHM-D-12-0107.1.

1091 Du, J., K. Wang, J. Wang, S. Jiang, and C. Zhou, 2018. Diurnal cycle of surface air temperature
 1092 within china in current reanalyses: Evaluation and diagnostics. *J. Climate*, **31**, 4585–4603,
 1093 <https://doi.org/10.1175/JCLI-D-17-0773.1>.

1094 Durre, I., Vose, R.S., Wuertz, D.B., 2006. Overview of the Integrated Global Radiosonde
 1095 Archive. *J. Climate* 19: 53– 68. <https://doi.org/10.1175/JCLI3594.1>.

1096 Flamant, C., Chaboureaud, J.-P. Parker, D. J., Taylor, C.M., Cammas, J.-P. Bock, O., Timouk, F.,
 1097 Pelon, J., 2007. Airborne observations of the impact of a convective system on the planetary
 1098 boundary layer thermodynamics and aerosol distribution in the inter-tropical discontinuity
 1099 region of the West African Monsoon, *Quarterly Journal of the Royal Meteorological Society*,
 1100 vol. 133, no. 626 A, pp. 1175–1189.

1101 Free, M., Seidel, D.J., Angel, J.K., Lanzante, J., Durre, I., and Peterson, T.C., 2005. Radiosonde
 1102 atmospheric temperature products for assessing climate (RATPAC): A new dataset of large-
 1103 area anomaly time series. *J. Geophys. Res.*, 110, D22101,
 1104 <http://dx.doi.org/10.1029/2005JD006169>.

1105 Gamo, M., 1996. Thickness of the dry convection and large-scale subsidence above deserts.
 1106 *Boundary-Layer Meteorology* **79**(3): 265–278.

1107 Garcia-Carreras, L., D.J. Parker, J.H. Marsham, P.D. Rosenberg, I.M. Brooks, A.P. Lock, F.
 1108 Marengo, J.B. McQuaid, and M. Hobby, 2015. The turbulent structure and diurnal growth of
 1109 the Saharan atmospheric boundary layer. *J. Atmos. Sci.*, **72**, 693–713,
 1110 <https://doi.org/10.1175/JAS-D-13-0384.1>

1111 Gelaro, R., and Coauthors, 2017. The Modern-Era Retrospective Analysis for Research and
 1112 Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454,
 1113 <https://doi.org/10.1175/JCLI-D-16-0758.1>.

1114 Gertler, C., and O’Gorman, P., 2019 Changing available energy for extratropical cyclones and
 1115 associated convection in Northern Hemisphere summer *Proc. Natl. Acad. Sci.* **116**. 4105-
 1116 4110.

1117 Gevorgyan, A., Surface and tropospheric temperature trends in Armenia, *Int. J. Climatol.* **34**:
 1118 3559–3573 (2014). <https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/joc.3928>.

1119 Haimberger, L., 2007. Homogenization of radiosonde temperature time series using innovation
 1120 statistics. *J. Climate* **20**, 1377-1403.

1121 Haimberger, L., C. Tavalato and S. Sperka, 2008. Towards elimination of the warm bias in
 1122 historic radiosonde temperature records - some new results from a comprehensive
 1123 intercomparison of upper air data, *J. Climate* **21**, 4587-4606.

1124 Haimberger, L., C. Tavalato and S. Sperka, 2012. Homogenization of the global radiosonde
 1125 dataset through combined comparison with reanalysis background series and neighboring
 1126 stations. *J. Climate* **25**, 8108–8131.

1127 Hassan, I., Ghumman, A. R., Hashmi, H. N., 2016. Global warming and temperature changes for
 1128 Saudi Arabia, *Journal of Biodiversity and Environmental Sciences*, 6663(81):2222-3045.
 1129 Held, I. M., and Soden, B. J., 2000. Water vapor feedback and global warming, *Annu. Rev.*
 1130 *Energy Environ.*, 25, 441–475, doi: 10.1146/annurev.energy.25.1.441.
 1131 Holtslag, A.A.M., Svensson, G., Baas, P., Basu, S., Beare, B., Beljaars, A.C.M., Bosveld, F.C.,
 1132 Cuxart, J., Lindvall, J., Steeneveld, G.J., Tjernström, M., Van De Wiel, B.J.H., 2013. Stable
 1133 atmospheric boundary layers and diurnal cycles—challenges for weather and climate models.
 1134 *Bull Am Meteorol Soc.* doi:10.1175/BAMS-D-11-00187.1
 1135 Holworth, G. C., 1964: Estimates of mean maximum mixing depths in the contiguous United
 1136 States *Mon. Weather Rev.* **92** 235–242.
 1137 IPCC, 2007. Climate change 2007: the physical science basis, the contribution of Working
 1138 Group I to the Fourth Assessment Report of the IPCC. Cambridge University Press,
 1139 Cambridge. ISBN: 978-0-521-88009-1 (2007).
 1140 IPCC, 2013. Climate change 2013: the physical science basis, the contribution of Working
 1141 Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.
 1142 Cambridge University Press, Cambridge. ISBN: 978-1-107-05799-1.
 1143 Islam, M.N., Almazroui, M. Direct effects and feedback of desert dust on the climate of the
 1144 Arabian Peninsula during the wet season: a regional climate model study. *Clim Dyn* **39**,
 1145 2239–2250 (2012). <https://doi.org/10.1007/s00382-012-1293-4>.
 1146 Knippertz, P., and Todd, M.C., 2012. Mineral dust aerosols over the Sahara: Meteorological
 1147 controls on emission and transport and implications for modeling, *Rev. Geophys.*, 50,
 1148 RG1007, doi:10.1029/2011RG000362.

1149 Krishna, L.V., 2014. Long-term temperature trends in four different climatic zones of Saudi
1150 Arabia, *International Journal of Applied Science and Technology*, 4 (5); 233-242.

1151 Lazzarini, M., Molini, A., Marpu, P.R., Ouarda, T.B.M.J., and Ghedira, H., 2015. Urban climate
1152 modifications in hot desert cities: The role of land cover, local climate, and seasonality,
1153 *Geophys. Res. Lett.*, 42, 9980–9989, doi:10.1002/2015GL066534.

1154 Lazzarini, M.; Marpu, P.R.; Ghedira, H., 2013. Temperature-land cover interactions: The
1155 inversion of urban heat island phenomenon in desert city areas. *Remote Sens. Environ.*, 130,
1156 136–152.

1157 Lewis, S., and Karoly, D., 2013. Evaluation of Historical Diurnal Temperature Range Trends in
1158 CMIP5 Models. *J. Climate*, 26(22), 9077-9089. Retrieved April 30, 2020, from
1159 www.jstor.org/stable/26192971

1160 Li, J., Chu, Y., Li, X., and Dong, Y., 2020 Long-term trends of global maximum atmospheric
1161 mixed layer heights derived from radiosonde measurements *Geophys. Res. Lett.* **15** 034054.

1162 Li, W., 2012. Study of diurnal cycle variability of planetary boundary layer characteristics over
1163 the Red Sea and Arabian Peninsula, in Earth Science Program. King Abdullah University of
1164 Science and Technology: Thuwal, Kingdom of Saudi Arabia.

1165 Lindvall, J., Svensson, G., 2015. The diurnal temperature range in the CMIP5 models. *Clim Dyn*
1166 44, 405–421 (2015). <https://doi.org/10.1007/s00382-014-2144-2>.

1167 Marsham, J.H., et al., 2013. Meteorology and dust in the central Sahara: Observations from
1168 Fennec supersite-1 during the June 2011 intensive observation period, *J. Geophys. Res.*
1169 *Atmos.*, 118, 4069–4089, doi:10.1002/jgrd.50211.

1170 McGrath-Spangler, E.L. and Molod, A., 2014. Comparison of GEOS-5 AGCM planetary
 1171 boundary layer depths computed with various definitions, *Atmos. Chem. Phys.*, 14, 6717–
 1172 6727, doi:10.5194/acp-14-6717-2014.

1173 McGrath-Spangler, E.L., A. Molod, L.E. Ott, and Pawson, S., 2015. Impact of planetary
 1174 boundary layer turbulence on model climate and tracer transport, *Atmos. Chem. Phys.*, 15,
 1175 7269– 7286, doi:10.5194/acp-15-7269-2015.

1176 McNider, R. T., G. J. Steeneveld, A. A. M. Holtslag, R. A. Pielke Sr., S. Mackaro, A. Pour-
 1177 Biazar, J. Walters, U. Nair, and J. Christy, 2012: Response and sensitivity of the nocturnal
 1178 boundary layer over land to added longwave radiative forcing. *J. Geophys. Res.* 117 D14106.

1179 Messenger, C., D. Parker, O. Reitebuch, A. Agusti-Panareda, C. M. Taylor and J. Cuesta, 2010.
 1180 Structure and dynamics of the Saharan atmospheric boundary layer during the West African
 1181 monsoon onset: Observations and analyses from the research flights of 14 and 17 July 2006,
 1182 *Q. J. Roy. Meteor. Soc.*, 107-124.

1183 Patlakas, P., Stathopoulos, C., Flocas, H., Kalogeri, C. and Kallos, G., 2019. Regional climatic
 1184 features of the Arabian Peninsula. *Atmosphere*, 10, 220.

1185 Ruppert, J. H., 2016, Diurnal timescale feedbacks in the tropical cumulus regime, *J. Adv. Model.*
 1186 *Earth Syst.*, 8, 1483– 1500, doi:[10.1002/2016MS000713](https://doi.org/10.1002/2016MS000713).

1187 Salmun, H., Molod, A., Collow, A., 2018. The use of MERRA-2 near surface meteorology to
 1188 understand the behavior of observed planetary boundary layer heights over the US Great
 1189 Plains. American Geophysical Union, Fall Meeting 2018, abstract #A13M-2638.

1190 Seidel, D. J., Y. Zhang, A. Beljaars, J.-C. Golaz, A. R. Jacobson, and B. Medeiros (2012),
 1191 Climatology of the planetary boundary layer over the continental United States and Europe,
 1192 *J. Geophys. Res.*, 117, D17106, doi:10.1029/2012JD018143.

1193 Sherwood, S. C., J. R. Lanzante, and C. L. Meyer, 2005. Radiosonde daytime biases and late
 1194 20th century warming, *Science*, 309, 1556– 1559, doi:10.1126/science.1115640.
 1195 Smith, A., Lott, N., and Vose, R., 2011. The integrated surface database: Recent developments
 1196 and partnerships. *Bull. Amer. Meteor. Soc.*, 92, 704–708.
 1197 Svensson, G., and Lindvall, J., 2015: Evaluation of near-surface variables and the vertical
 1198 structure of the boundary layer in CMIP5 models. *J. Climate*, 28, 5233–5253,
 1199 <https://doi.org/10.1175/JCLI-D-14-00596.1>
 1200 Tarawneh, Q.Y., Chowdhury, S., 2018. Trends of climate change in Saudi Arabia: Implications
 1201 on water resources. *Climate*, 6, 8.
 1202 Thomas, N., Nigam, S., 2017. Twentieth-century climate change over Africa: Seasonal
 1203 hydroclimate trends and Sahara Desert expansion, *J. Clim.*, 31, 3349-3370, 10.1175/JCLI-D-
 1204 17-0187.1
 1205 Thorne, P. W., J. R. Lanzante, T. C. Peterson, D. J. Seidel, and K. P. Shine, 2011. Tropospheric
 1206 temperature trends: History of an ongoing controversy, *WIREs Clim. Change*, 2(1), 66– 88,
 1207 doi:10.1002/wcc.80.
 1208 Vinnikov, K., Robok, A., and Basist, A., 2002. Diurnal and seasonal cycles of trends of surface
 1209 air temperature, *J. Geophys. Res.*, 107(D22), 4641, doi:10.1029/2001JD002007.
 1210 Vizzy, E. K. and Cook, K. H., 2017. Seasonality of the observed amplified Sahara warming trend
 1211 and implications for Sahel rainfall, *J. Clim.*, doi: 10.1175/JCLI-D-16-0687.1.
 1212 Vogelesang, D. H. P., and A. A. M. Holtslag (1996), Evaluation and model impacts of
 1213 alternative boundary-layer height formulation, *Boundary Layer Meteorol.*, 81, 245–269,
 1214 doi:10.1007/BF02430331.

1215 Vose, R. S., D. R. Easterling, and B. Gleason (2005), Maximum and minimum temperature
 1216 trends for the globe: An update through 2004, *Geophys. Res. Lett.*, 32, L23822,
 1217 doi:10.1029/2005GL024379.

1218 Wang, X., and K. Wang, 2016: Homogenized variability of radiosonde-derived atmospheric
 1219 boundary layer height over the global land surface from 1973 to 2014. *J. Climate*, **29**, 6893–
 1220 6908, <https://doi.org/10.1175/JCLI-D-15-0766.1>.

1221 Wei, N., Zhou, L., and Dai, Y., 2017a. Observational evidence for desert amplification using
 1222 multiple satellite datasets, *Scientific Reports* 7, 2043, doi:10.1038/s41598-017-02064-w.

1223 Wei, N., Zhou, L., and Dai, Y., 2017b. Evaluation of simulated climatological diurnal
 1224 temperature range in cmip5 models from the perspective of planetary boundary layer
 1225 turbulent mixing, *Climate Dynamics*, 49(1), 1-22, DOI 10.1007/s00382-016-3323-0.

1226 Wu, G.X., Liu, Y., Zhu, X., Li, W., Ren, R., Duan, A., and Liang, X., 2009. Multi-scale forcing
 1227 and the formation of subtropical desert and monsoon, *Ann. Geophys.*, 27, 3631–3644.

1228 Yang, G.Y., and J. Slingo, (2001), The diurnal cycle in the tropics. *Mon. Weather Rev.*, 129,
 1229 784–801.

1230 Zhou, L., 2016. Desert amplification in a warming climate, *Scientific Reports* 6, 31065; doi:
 1231 10.1038/srep31065.

1232 Zhou, L., Chen, H. and Dai, Y., 2015. Stronger warming amplification over drier ecoregions
 1233 observed since 1979, *Environ. Res. Lett.* 10, 064012, doi:10.1088/1748-9326/10/6/064012.

1234 Zhou, L., Chen, H., Hua, W., Dai, Y., and Wei, N., 2016. Mechanisms for stronger warming over
 1235 drier ecoregions observed since 1979, *Climate Dynamics*, DOI 10.1007/s00382-016-3007-9.

1236 Zhou, L., Dai, A., Dai, Y., Vose, R.S, Zou, C-Z., Tian, Y., and Chen, H., 2009: Spatial
 1237 dependence of diurnal temperature range trends on precipitation from 1950 to 2004, *Clim.*
 1238 *Dyn.*, 32:429–440, DOI 10.1007/s00382-008-0387-5.

1239 Zhou, L., Dickinson, R.E., Tian, Y., and Vose, R.S., 2007: Impact of vegetation removal and soil
 1240 aridation on diurnal temperature range in a semiarid region – Application to the Sahel, *Proc.*
 1241 *Natl. Acad. Sci. USA*, 104 (46), 17937-17942.

1242 Zhou, L., R. E. Dickinson, A. Dai, and P. Dirmeyer, 2010: Detection and attribution of
 1243 anthropogenic forcing to diurnal temperature range changes from 1950 to 1999: Comparing
 1244 multi-model simulations with observations, *Clim. Dyn.*, 35:1289-1307, DOI:
 1245 10.1007/s00382-009-0644-2.

Table 1. Surface weather stations used in this study from the DS3505 dataset

Station Name	WMO code	Latitude (°N)	Longitude (°E)	Elevation (m)
Ben Gurion Intl Airport, Israel	401800	32.011	34.887	41
Kuwait Intl Airport, Kuwait	405820	29.227	47.969	63
ARAR, Saudi Arabia	403570	30.907	41.138	553
Al Jouf, Saudi Arabia	403610	29.785	40.100	689
Hail, Saudi Arabia	403940	27.438	41.686	1015
Gassim, Saudi Arabia	404050	26.300	43.767	648
Al Ahsa, Saudi Arabia	404200	25.285	49.485	179
King Khaled Intl Airport, Saudi Arabia	404370	24.958	46.699	625
King Abdulaziz Intl Airport, Saudi Arabia	410240	21.680	39.157	15
Al Baha, Saudi Arabia	410550	20.296	41.634	1672
Bisha, Saudi Arabia	410840	19.984	42.621	1185
ABHA, Saudi Arabia	411120	18.240	42.657	2090
King Khaled Ab, Saudi Arabia	411140	18.297	42.804	2066
King Abdullah Bin Abdulaziz, Saudi Arabia	411400	16.901	42.586	6
Turaif, Saudi Arabia	403560	31.693	38.731	854
Qaisumah, Saudi Arabia	403730	28.335	46.125	358
Tabuk, Saudi Arabia	403750	28.365	36.619	778
Prince Mohammad Bin Abdulaziz, Saudi Arabia	404300	24.553	39.705	656
Riyadh Ab, Saudi Arabia	404380	24.71	46.725	635
Taif, Saudi Arabia	410360	21.483	40.544	1478
Nejran, Saudi Arabia	411280	17.611	44.419	1214

Table 2. Radiosonde stations used in this study from the RICH dataset

Station Name	WMO code	Latitude (°N)	Longitude (°E)	Elevation* (m)
Bet Dagan, Israel	ISM00040179	32.000	34.817	35 (414, 281)
Kuwait Intl Airport, Kuwait	KUM00040582	29.243	47.971	56 (397, 174)
Al-Qaisumah, Saudi Arabia	SAM00040373	28.317	46.133	358 (753, 339)
Tabuk, Saudi Arabia	SAM00040375	28.383	36.600	778 (584, 875)
Hail, Saudi Arabia	SAM00040394	27.433	41.683	1015 (943, 969)
Al-Madinah, Saudi Arabia	SAM00040430	24.550	39.700	654 (936, 846)
King Khaled Intl Airport, Saudi Arabia	SAM00040437	24.933	46.717	614 (693, 637)
Jeddah, Saudi Arabia	SAM00041024	21.700	39.183	15 (905, 593)

Note: *Elevation in parenthesis refers to the corresponding grid-averaged elevation in the ERA5 (first column) and MERRA2 (second column) reanalysis.

Table 3. Station means of the climatology (°C) and trends (°C/decade) of daily maximum T2m, minimum T2m, and DTR averaged over the 21 surface stations (Fig. 1a) from the DS3505 (1979-2018), ERA5-land (1981-2018), and MERRA-2 (1980-2018)

dataset	T _{max}		T _{min}		DTR	
	climatology	trend	climatology	trend	climatology	trend
DS3505	30.1	0.56	18.6	0.72	11.5	-0.16
ERA5-land	30.1	0.42	17.0	0.44	13.0	-0.02
MERRA-2	31.6	0.37	17.5	0.39	14.1	-0.02

Note: T_{max} – maximum T2m, T_{min} – minimum T2m, DTR – diurnal temperature rang of T2m. Trends (°C/decade) in bold and italics are statistically signifcant at the 5% level.

Table 4. Statistical results for the multiple linear regression in equation (2)

Scale	PBLH type	$T2m_{trend} = \beta_0 + \beta_1 * 1/PBLH_{clim} + \beta_2 * DLR_{trend} + \beta_3 * DSR_{trend}$					
		N (N*)	β_1	β_2	β_3	R ² _{adjusted}	F _{pval}
ERA5							
spatial + diurnal	PBLH	21144 (3299)	0.22	0.85	0.22	0.58	<0.01
	PBLH RI _{0.25}	21144 (3863)	0.22	0.81	0.21	0.58	<0.01
spatial	PBLH	881	0.02	1.15	0.52	0.73	<0.01
	PBLH RI _{0.25}	881	0.11	1.10	0.48	0.74	<0.01
diurnal	PBLH	24	2.53	1.70	0.06	0.95	<0.01
	PBLH RI _{0.25}	24	2.05	1.24	0.03	0.89	<0.01
MERRA-2							
spatial + diurnal	PBLH	17304 (2690)	0.00	0.74	0.26	0.56	<0.01
	PBLH RI _{0.25}	2884	0.05	0.72	0.29	0.60	<0.01
spatial	PBLH	721	0.06	0.72	0.37	0.66	<0.01
	PBLH RI _{0.25}	721	-0.02	0.75	0.37	0.68	<0.01
diurnal	PBLH	24	0.75	0.06	0.30	0.99	<0.01
	PBLH RI _{0.25}	4	-	-	-	-	-

Note: PBLH and PBLH_RI_{0.25} are defined in Table 5. N is the sample size of the regression for the ERA5-land and MERRA-2 data over the study domain, consisting of spatial (grid box) and diurnal (hourly or 6-hourly) temporal scales. N* is the effetcive degree of freedom after considering the spatial correlation between grid boxes. F_pval denotes the statistical significance (p value) of the F-test. The partial regression coefficients, β_1 , β_2 , and β_3 , are standardized and so can be compared to determine the relative contribution to T2m_{trend}. The values in bold and italics are statistically signifcant at p=0.01 and p=0.05, respectively.

1268 **Table 5.** PBLH (m) for the 8 radiosonde stations from the IRGA2, ERA5 and MERRA-2
1269 datasets

WMO code	IGRA2	ERA5		MERRA-2	
	PBLH_RI _{0.25}	PBLH	PBLH_RI _{0.25}	PBLH	PBLH_RI _{0.25}
00 UTC					
ISM00040179	84.3	106.7	20.1	432.5	21.5
KUM00040582	127.4	167.4	36.6	844.5	31.8
SAM00040373	100.6	134.5	21.7	1142.1	22.2
SAM00040375	87.5	127.5	16.1	563.5	9.5
SAM00040394	88.6	137.9	22.3	1252.3	12.1
SAM00040430	122.7	110.0	17.6	889.1	18.5
SAM00040437	104.8	130.0	20.1	1330.7	19.6
SAM00041024	101.7	116.7	26.5	335.6	22.4
Station mean	102.2	128.8	22.6	848.8	19.7
12 UTC					
ISM00040179	1199.3	1137.7	1075.8	1575.7	1570.7
KUM00040582	1784.4	2033.2	1885.0	2263.4	2147.5
SAM00040373	2275.5	2096.3	2226.9	2393.7	2380.4
SAM00040375	2131.4	2115.7	2395.7	2642.9	2870.0
SAM00040394	2793.8	2626.2	2809.3	2978.7	3022.2
SAM00040430	2774.0	2847.4	3136.5	3223.7	3600.8
SAM00040437	2520.9	2506.1	2720.2	2772.4	2851.0
SAM00041024	1240.3	1313.6	1650.5	2524.9	2877.5
Station mean	2089.9	2084.5	2237.5	2546.9	2665.0

1270 **Note:** PBLH refers to the PBLH provided by the reanalysis. PBLH_RI_{0.25} refers to the PBLH diagnosed
1271 directly from the atmospheric profiles using the bulk Richardson (Ri) number method (Eq. (1)), with the
1272 threshold value of Z(Ri=0.25).

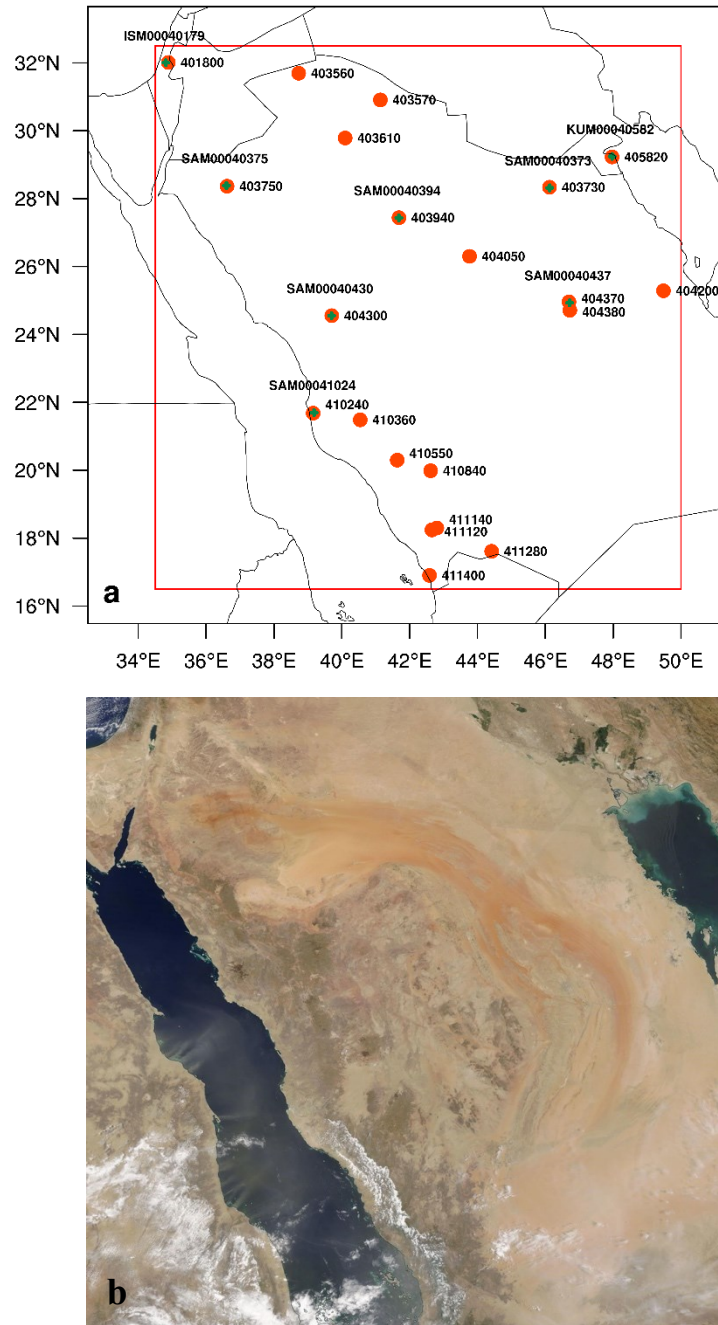


Fig. 1. (a) Surface (in red) and radiosonde (in green) stations used over the Arabian Peninsula (AP). The AMO identifier # for every radiosonde (11-digital) and surface (6-digital) station is shown. The geographic location and elevation for each station are listed in Tables 1 and 2. The rectangle box (16.5°N-32.5°N, 34.5°E-50.0°E) depicts the land area over which the regional mean is averaged for the reanalysis data. (b) The corresponding true-color satellite image on August 6, 2020 covering the same area as (a) obtained from the NASA EOSDIS worldview website: <https://worldview.earthdata.nasa.gov/>.

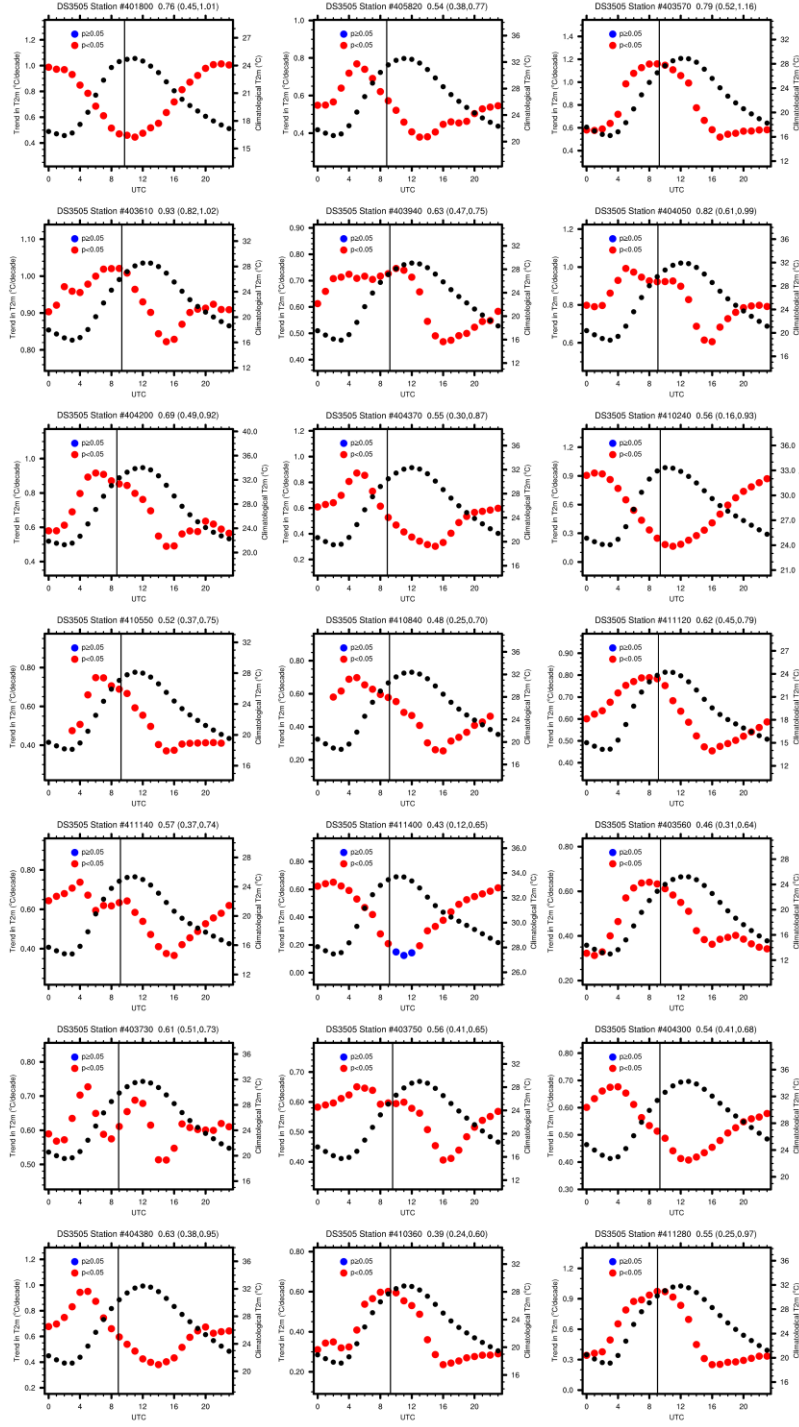


Fig. 2. The diurnal cycle of linear trend (°C/decade) of T2m (left y-axis, in red) for the period 1979-2018, along with the diurnal cycle of climatological T2m (°C, right y-axis, in black), for the 21 surface stations labeled in Fig. 1a. The T2m trends are statistically significant at $p \leq 0.05$ (in red). The daily mean, as well as the maximum and minimum (in parentheses) of the hourly T2m trend, are listed on the top of each panel. The vertical line in each panel indicates the local solar noon.

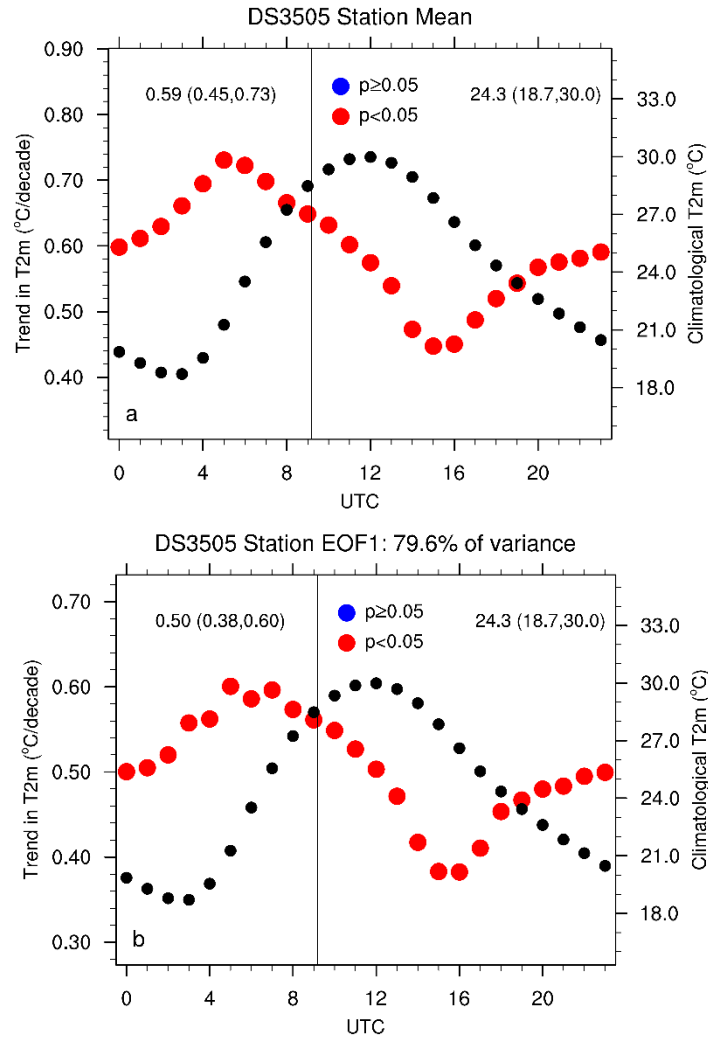


Fig 3. Same as Fig. 2 but for the station mean: (a) original T2m, and (b) reconstructed T2m using the first EOF, which explains 79.6% of the total data variance. The vertical line in each panel indicates the station mean local solar noon.

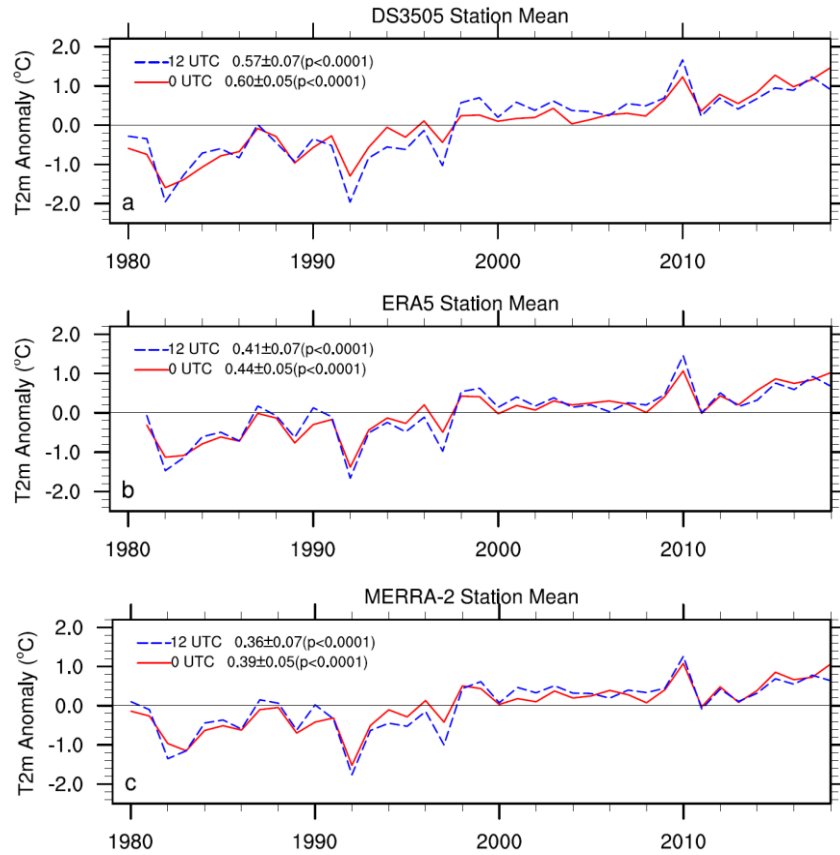


Fig 4. Interannual variations in station mean T2m anomalies (°C) at 00 and 12 UTC averaged over the 21 surface stations labeled in Fig. 1a from: (a) DS3505 (1979-2018), (b) ERA5-land (1981-2018), and (c) MERRA-2 (1980-2018). Linear trend (°C/decade) plus one standard deviation, along with its significance level (p value), are shown.

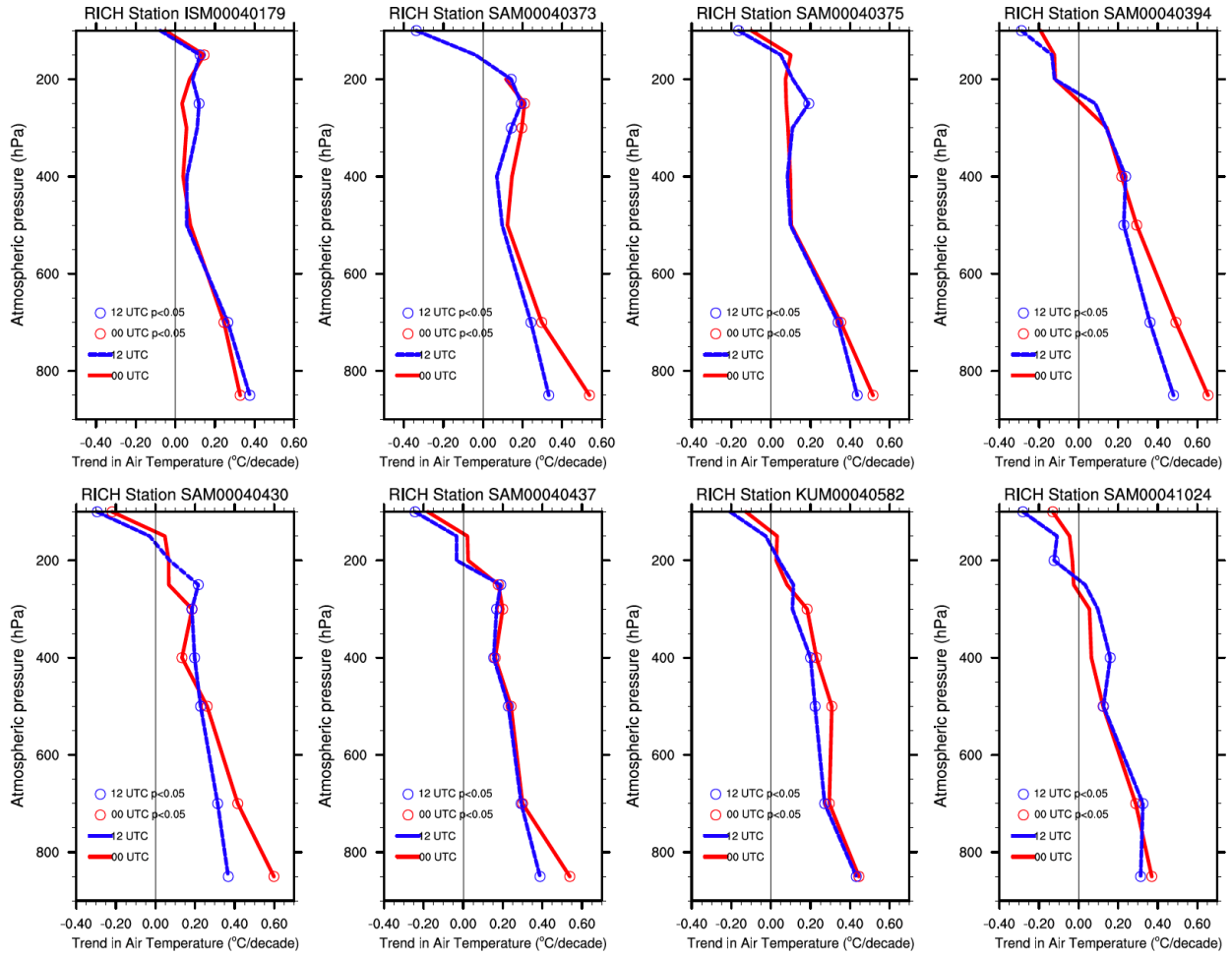


Fig 5. The vertical profile of linear trend ($^{\circ}\text{C}/\text{decade}$) of atmospheric air temperature at 00 (in red) and 12 UTC (in blue) during the period 1979-2018 for the 8 radiosonde stations labeled in Fig. 1a in the RICH dataset. The trends are statistically significant at $p < 0.05$ (in circle).

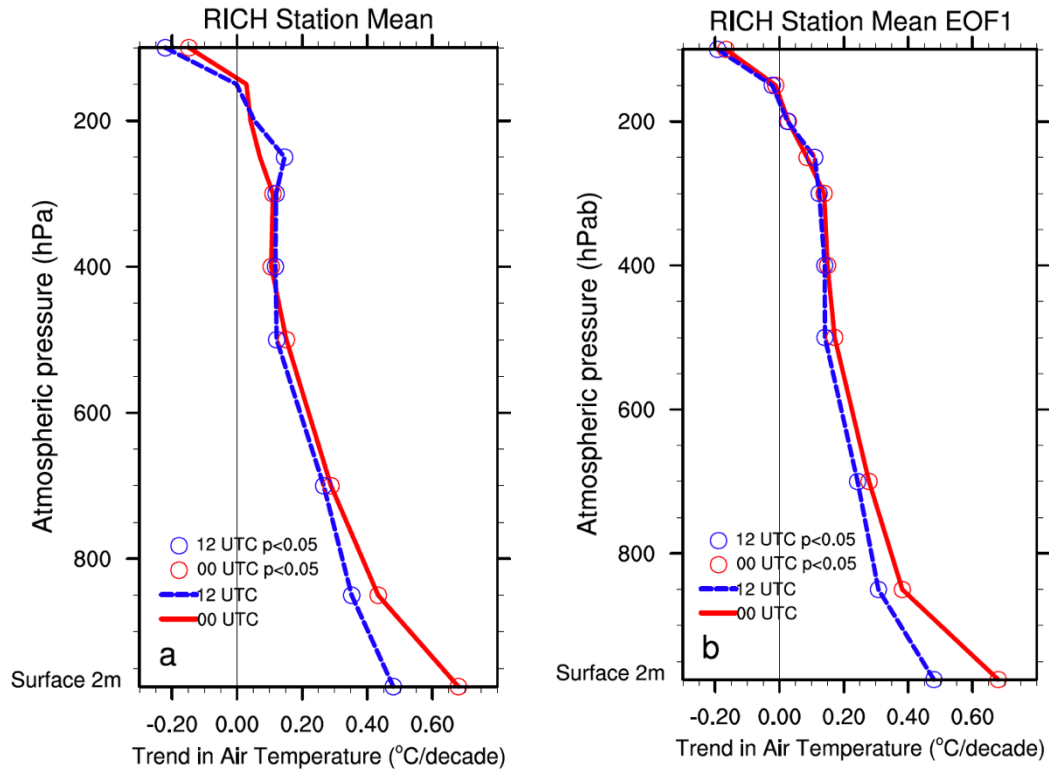


Fig 6. Same as Fig. 5 but for the station mean: (a) original temperature, and (b) reconstructed temperature using the first EOF, which explains 54.3% (00 UTC) and 49.8% (12 UTC) of the data variance. The corresponding station mean $T2m_{trend}$ at 00 and 12 UTC from the DS3505 are also shown at the bottom of the plot.

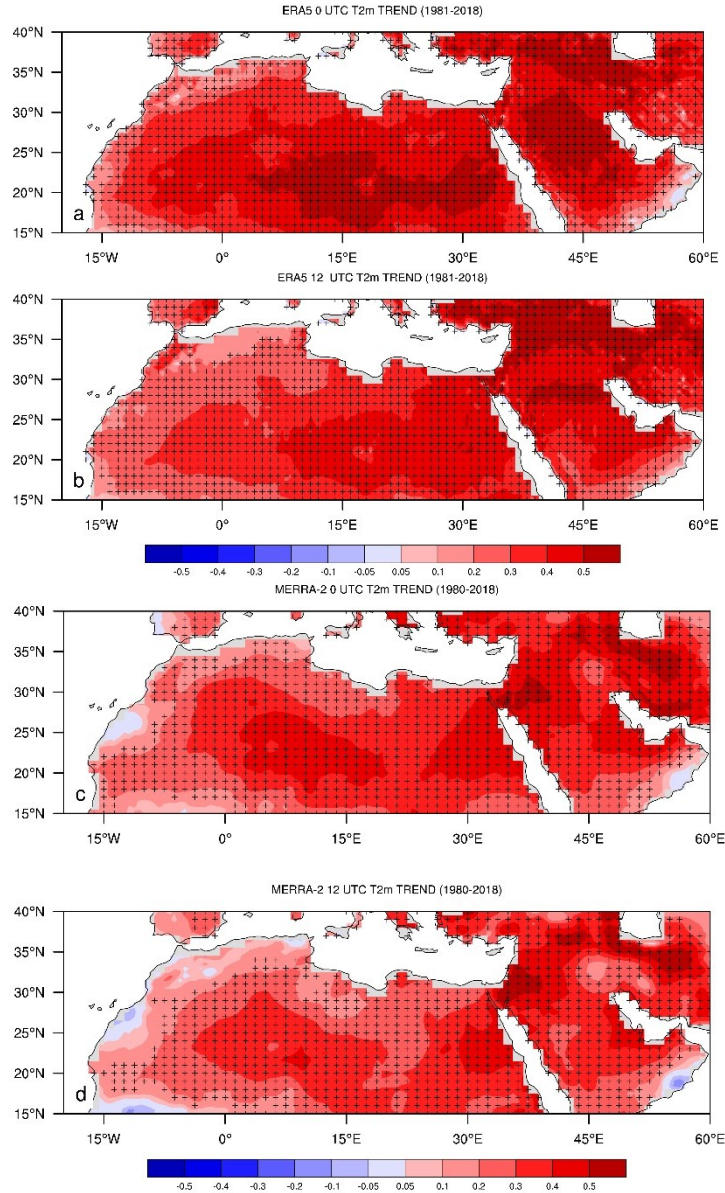


Fig. 7. Spatial patterns of linear trend (°C /decade) of T2m from ERA5-land (1981-2018) and MERRA-2 (1980-2018) over North Africa and the Arabian Peninsula: (a, c) at 00 UTC, and (b, d) at 12 UTC. Stippling indicates regions where the trend is statistically significant at $p < 0.05$.

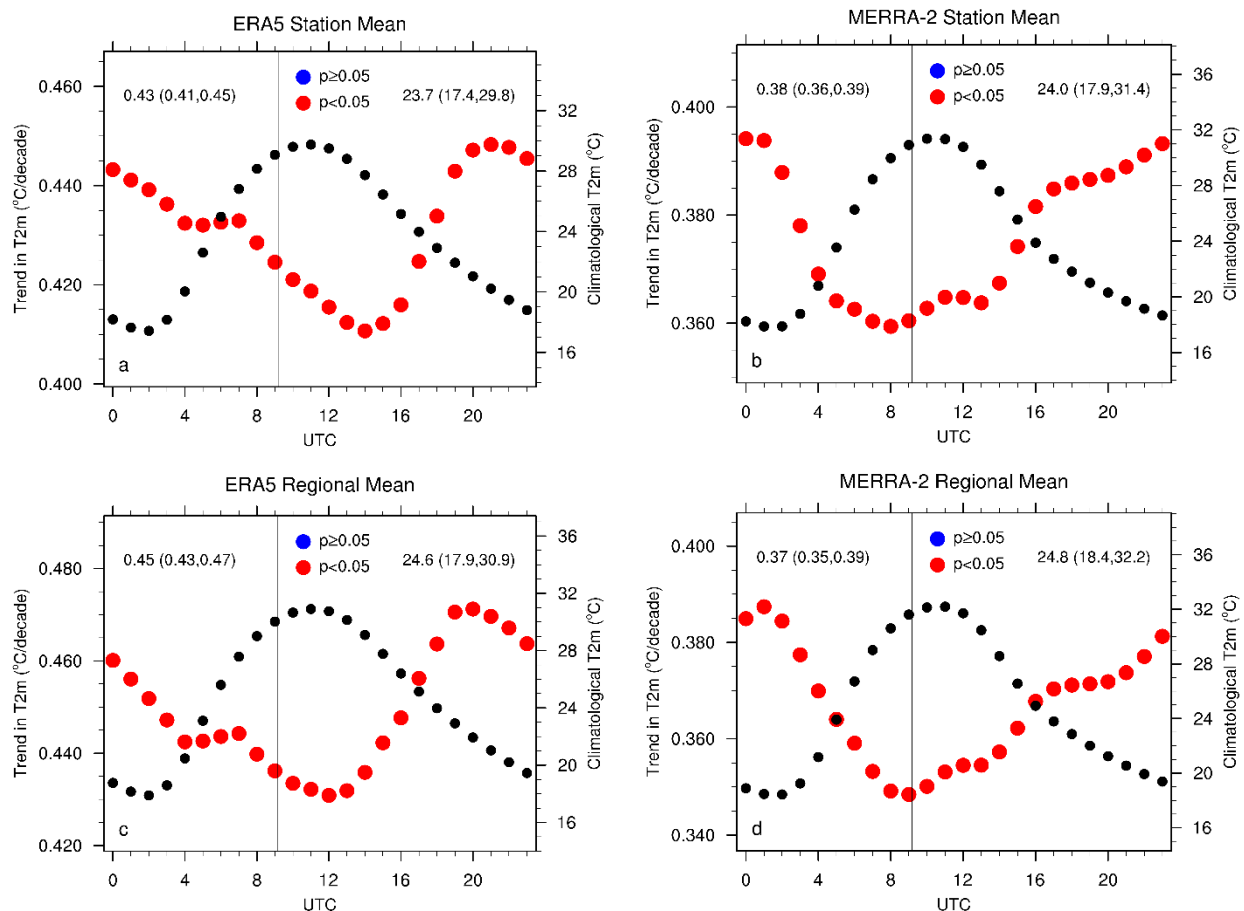


Fig. 8. The diurnal cycle of linear trend (°C/decade) of T2m (left y-axis), along with the diurnal cycle of climatological T2m (°C, right y-axis, in black), from the ERA5-land (1981-2018) and MERRA-2 (1980-2018): (a, b) station mean averaged over the 21 surface stations labeled in Fig. 1a, and (c, d) regional mean averaged over the rectangle domain depicted in Fig. 1a. The T2m trends are statistically significant at $p < 0.05$ (in red). The daily mean, along with the maximum and minimum (in parentheses), of the 24-hourly values for the T2m trend (left) and the T2m climatology (right) are listed in each panel. The vertical line in each panel indicates the station or regional mean local solar noon.

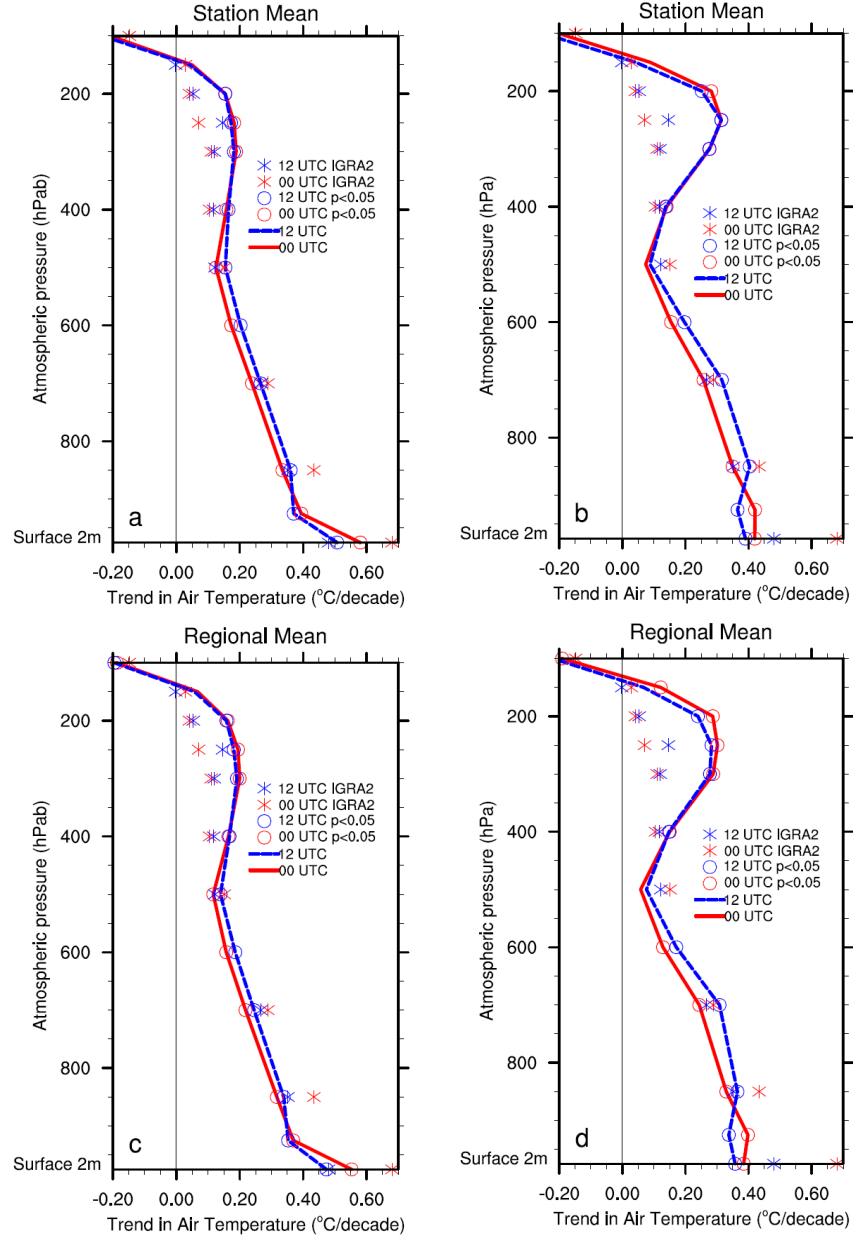


Fig. 9. The vertical profile of linear trend ($^{\circ}\text{C} / \text{decade}$) of atmospheric air temperature and surface T2m at 00 (in red) and 12 UTC (in blue) from the ERA5 (1979-2018) and MERRA-2 (1980-2018): (a, b) station mean for the 8 radiosonde stations labeled in Fig. 1a, and (c, d) regional mean for the rectangle domain depicted in Fig.1a. The trends are statistically significant at $p < 0.05$ (in circle). The corresponding station mean T2m trends ($^{\circ}\text{C}/\text{decade}$) are also plotted. For comparison purpose, the station mean profiles observed in Fig. 6a are also shown.

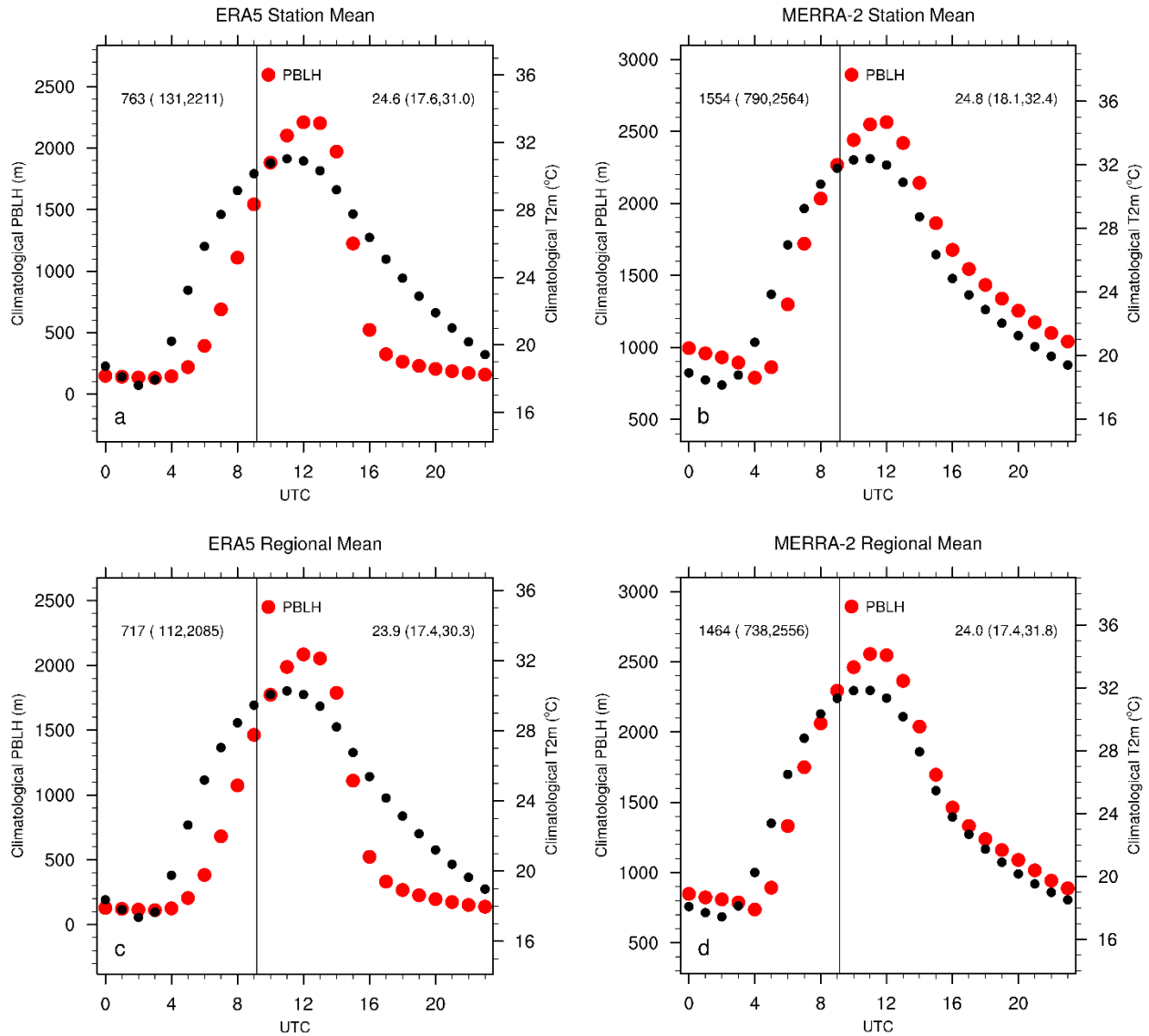


Fig. 10. The diurnal cycle of climatological PBLH (m, left y-axis, in red) and T2m (°C, right y-axis, in black) from the ERA5-land (1981-2018) and MERRA-2 (1980-2018): (a, b) station mean for the 21 surface stations labeled in Fig. 1a, and (c, d) regional mean for the rectangle domain depicted in Fig.1a. The daily mean, along with the maximum and minimum (in parentheses), of the 24-hourly values for the PBLH (left) and the T2m climatology (right) are listed in each panel. The vertical line in each panel indicates the station or regional mean local solar noon.

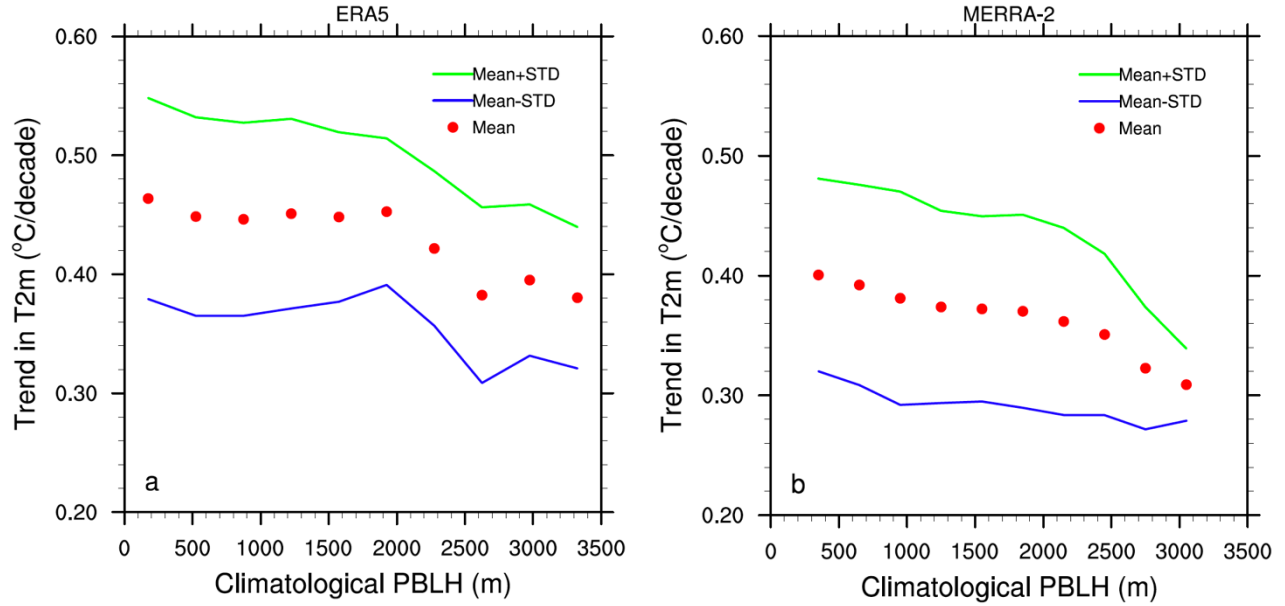


Fig. 11. The histogram of hourly T2m trend ($^{\circ}\text{C}/\text{decade}$, y-axis) plus one standard deviation (STD) as a function of climatological hourly PBLH (m, x-axis) from (a) ERA5-land (1981-2018) and (b) MERRA-2 (1980-2018). All the PBLHs over the land grids with the rectangle domain depicted in Fig. 1a are divided evenly into 10 bins based on a PBLH interval of 350 m in ERA5 and 300 m in MERRA2.

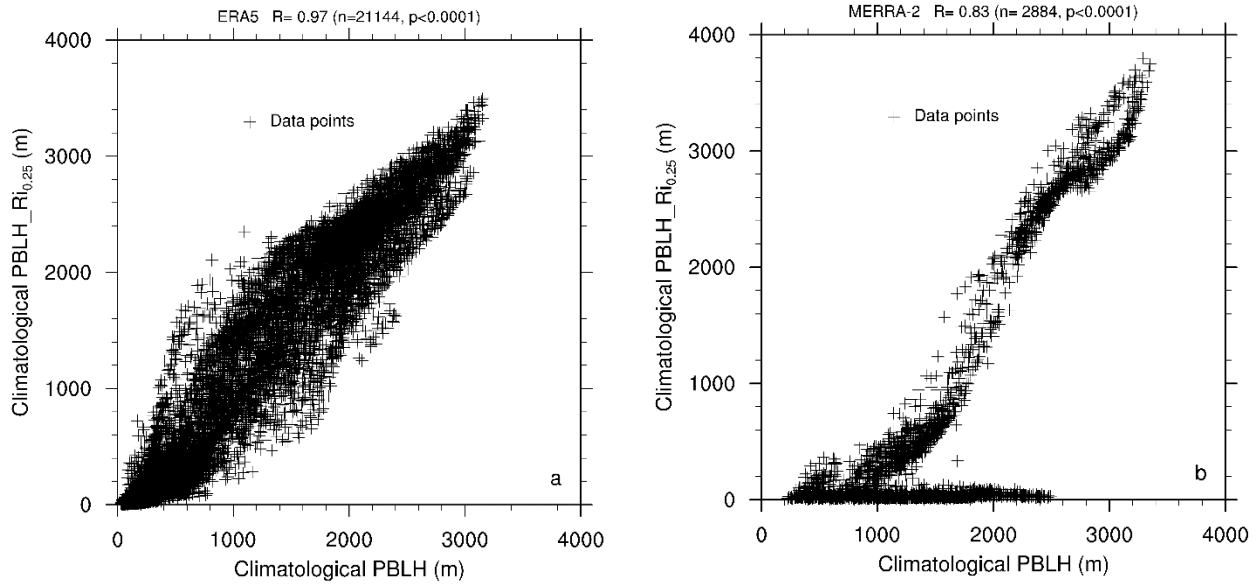


Fig. 12. Scatter plots of climatological hourly PBLH (m) derived by the reanalysis (x-axis) and diagnosed using the Richard number method, referred to as PBLH_Ri_{0.25} (m, y-axis), from the ERA5 (1979-2018) and MERRA-2 (1980-2018) over all land grids within the rectangle domain depicted in Fig.1a. The correlation coefficient R, its statistical significance (p value) and sample size (n) are listed on the top in each panel.