A Global-Gridded Dataset for Π based off ERA5 reanalysis

N. Z. Wong[1], L. Feng[2]and E. M. Hill[1,2]

[1]Asian School of the Environment, Nanyang Technological University
[2]Earth Observatory of Singapore, Nanyang Technological University

Key Points:

- A new T_m dataset based on the latest ERA5 reanalysis is now available.
- The GPT2w model is unable to account for intraseasonal variability
- Using surface temperature to estimate Π overemphasises diurnal variability over land

Corresponding author: Nathanael Wong, `nathanaelwong@fas.harvard.edu`

Abstract

We used the recently released ERA5 reanalysis data to create a global, gridded dataset for the constant of proportionality Π (Π_{RE5}) which converts Global Navigation Satellite Systems (GNSS) zenith wet delay signals into precipitable water vapour based on the work of Askne and Nordius (1987). A comparison of Π_{RE5} against the Π derived from ERA-Interim reanalysis (Π_{REI}) shows that Π_{RE5} has more pronounced diurnal variability, and that Π_{REI} slightly overestimates Π in regions of high topography. Comparison with other datasets derived from models such as GGOS Atmosphere (Π_{RGA}) or GPT2w (Π_{EG2}) also highlight deficiencies in these models at different spatial-temporal scales. Comparison of Π_{RE5} with values of Π derived from the linear approximation of T_m based on T_s (Bevis et al., 1992) (e.g. Π_{EBB}, Π_{EBM}) gives rise to significant bias in both the mean value and variability of Π. Lastly, we also perform the calculation of T_m using pressure coordinates instead of vertical coordinates (Π_{REP}), and find that the error in both the climatological mean and variance that results from this transformation is $< 0.01\%$. Since reanalysis data is given in pressure coordinates as opposed to vertical coordinates, using pressure coordinates in the calculation of T_m provides a much simpler method to calculate Π as opposed to integrating in vertical coordinates to find T_m.

Plain Language Summary

Text

1 Introduction

Atmospheric water vapour plays an extremely important role in the Earth's climate system, affecting the global energy balance and hydrological cycle. However, water vapour has high spatial and temporal variability, which makes it difficult to monitor accurately using traditional techniques such as radiosonde and satellites such as MODIS with remote-sensing instrumentation, which are expensive and have limited temporal resolution. These techniques also may suffer from biases over the years due to instrument changes, and there is often a trade-off between the cost-effectiveness of monitoring water vapour at high temporal and spatial resolution. This therefore hinders our representation of the global hydrological cycle in climate models.

However, the usage of GNSS instrumentation to study weather and climate can help to fill in this gap, as continuous GNSS (cGNSS) stations are be able to record data on the scale of minutes which allows for hundreds of measurements a day. The increasing density of cGNSS stations around the globe therefore allows for us to circumvent the problems faced in measuring water vapour using more traditional methods. Monitoring water vapour using cGPS (continuous GPS) stations is not a new technique, and was proposed by Bevis et al. (1992) even before the GPS constellation became fully operational in 1994. Since then, many studies have been conducted in Europe, the United States and East Asia using GPS to monitor precipitable water vapour. With the introduction of other cGNSS constellations such as GLONASS, BeiDou and Galileo, there may even be enough information to conduct tomographic studies of water vapour in the future.

GNSS technology relies on the transmission of radio wave signals from satellite to receiver in order to obtain precise positions. There are many different sources of error that need to be accounted for during this process. In space, the distortion of these signals is considered to be negligible, but when these signals travel through the atmosphere, two potential sources of error arise due to refraction of the radio waves the ionospheric delay, and the tropospheric delay. The delay due to the tropospheric component (tropospheric path delay) in the neutral atmosphere (troposphere and stratosphere) is not so easily resolved, and can be further divided into the hydrostatic delay and the wet de-

lay. With mapping functions, these components are mapped onto the zenith direction, to the zenith hydrostatic delay (ZHD) and the zenith wet delay (ZWD).

From the zenith wet delay, scientists are able to measure the precipitable water vapour in the atmosphere, which is the equivalent height if all the water vapour along the path was condensed into a column. Askne and Nordius (1987) were the first to establish the relationship between zenith wet delay and precipitable water using the dimensionless constant of proportionality Π as follows:

$$\mathrm{PWV} = \Pi \cdot \mathrm{ZWD}; \Pi = \frac{10^6}{\rho R_v \left(k_2' + \frac{k_3}{T_m}\right)} \tag{1}$$

$$T_m = \frac{\int_0^\infty \frac{e(z)}{T(z)}\,\mathrm{d}z}{\int_0^\infty \frac{e(z)}{T^2(z)}\,\mathrm{d}z} = \frac{\sum_i \frac{e_i}{T_i}\Delta h_i}{\sum_i \frac{e_i}{T_i^2}\Delta h_i} \tag{2}$$

where ρ is the density of water, R_v is the specific gas constant for water vapour k_2' and k_3 are refractivity constants, and T_m is the water-vapour-weighted column air temperature. T_m in turn is calculated through integrating vapour pressure e and temperature T from the surface to the top of the atmosphere. Therefore, assuming that all constants are accurate, it can be seen that the major sources of error in using GNSS to measure precipitable water vapour are errors in the measurement of ZWD and T_m. In recognition of this, there are many papers that have studied the uncertainties in the measurements of T_m through comparison of global gridded empirical models such as GPT2w and GTm-III with radiosonde data. These studies aim to come up with easier methods to estimate T_m and therefore Π without requiring intensive in-situ measurements that are sparse in many regions.

Ultimately though, many of these globally-gridded empirical models can trace their roots back to reanalysis datasets such as ERA-Interim. Although the best way to derive T_m is to use in-situ measurements of water vapour pressure and temperature from vertical profiles, such profiles are generally derived through radiosonde. The density of radiosonde measurements is sparse in many regions around the globe and practically nonexistent over oceans. Reanalysis datasets aim to optimally combine these measurements with models in order to provide a globally coherent, gridded dataset that, while not truly being observations themselves, can be taken to be close approximations. Therefore, J. Wang et al. (2005) championed to use of reanalysis data (e.g. ERA-40) to estimate water vapour pressure and temperature at different pressure heights to obtain T_m. X. Wang et al. (2016) used ERA-Interim reanalysis data to calculate T_m and found that the difference between reanalysis T_m and the "ground-truth" calculated using in-situ radiosonde data was 0.5% on average.

With the release of the latest ERA5 reanalysis dataset, meteorological output is available every hour at the 0.25° spatial resolution, which is significantly higher than ERA-Interim. We aim to take advantage of this wealth of data in order to produce globally gridded T_m and Π datasets at much higher resolution than what is currently available from empirical models, or even from GGOS Atmospheres, which was a globally-gridded T_m dataset based off ERA-Interim data. This paper compares the spatial and temporal mean and variability of the gridded dataset of Π that we derived from processing of ERA5 output data to other available datasets, both reanalysis and empirical. In Section 2, we elaborate on the methodology we used to calculate Π, and the other datasets and models that we used in comparison to our own. In Section 3, we explore the spatial and temporal mean and variability of our dataset, and compare our results when integrating in both vertical and pressure coordinates, while in Section 4, we compare our results with that from other datasets. Our conclusions are found in Section 5.

2 Methodology

In this section, we describe the methodologies that are used to calculate the water-vapour-weighted column-mean temperature T_m, and the constant of proportionality Π. A summary of these methods can be found in Table 1.

Name	Classification	Output:	Data Source / Input	Section	Reference
RE5	reanalysis	T_m	ERA5 $(T, T_s, T_d, q, z, \Phi)$	2.1	Hersbach and Dee (2016)
REP			ERA5 (T, T_s, T_d, q, p_s)	2.1.1	
REI			ERA-Interim (T, T_s, T_d, q, p_s)	2.1.1	Dee et al. (2011)
RGA			GGOS Atmosphere	2.1.2	Böhm and Schuh (2013)
EBB	empirical	T_m	ERA5 T_s; (a, b) Bevis et al. (1992)	2.2.1	Bevis et al. (1992)
EBM			ERA5 T_s; (a, b) Manandhar et al. (2017)		Manandhar et al. (2017)
EG2		Π	day-of-year; (lon,lat)	2.2.2	Böhm et al. (2015)
EMN			day-of-year; (lat,height)		Manandhar et al. (2017)

Table 1. A summary of the methods used to calculate T_m, and therefore Π.

We aim to investigate the diurnal $\Delta_d(\cdot)$, intraseasonal $\Delta_i(\cdot)$, seasonal $\Delta_s(\cdot)$ and interannual $\Delta_a(\cdot)$ mean-weighted variability (Eqn. 3) of Π for each of the different methods. Based on Eqn. 4 (X. Wang et al., 2016), we see that $\Delta\Pi \approx \Delta T_m$. We also compare the temporal variability at different scales across different datasets to evaluate the strengths and weaknesses of each methodology. Π_{RE5}, which denotes the Π dataset derived from the ERA5 reanalysis data in vertical coordinates, is taken to be the ground-truth against which all other datasets are compared against.

$$\Delta(\cdot) = \frac{\delta(\cdot)}{\overline{(\cdot)}}, \text{ where } \overline{(\cdot)} \text{ denotes the time-averaged mean of } (\cdot) \tag{3}$$

$$\Delta\mathrm{PWV} = \Delta\Pi + \Delta\mathrm{ZWD} \approx \Delta T_m + \Delta\mathrm{ZWD} \tag{4}$$

2.1 Calculating T_m from reanalysis data

In this study, we use both ERA5 (Hersbach & Dee, 2016) and ERA-Interim (Dee et al., 2011) reanalysis data at 1.0° resolution in both longitude and latitude. As the data output is given in pressure coordinates, we convert the pressure coordinates to vertical coordinates by finding the geopotential height $\Phi(p)$ at each pressure level p and dividing by the gravitational constant to obtain the vertical height. Thus, the following variables are required to calculate T_m in vertical coordinates:

- 3D Datasets: geopotential Φ, air temperature T, specific humidity q
- Surface Datasets: surface geopotential z, surface and dewpoint temperatures T_s, T_d

The vapour pressure $e(z)$ in Eqn. **??** for T_m can be calculated from either specific humidity q (Eqn. 5) or dewpoint temperature T_d (Eqn. 6).

$$e = p \cdot \frac{q}{q + (1-q)\varepsilon} \tag{5}$$

$$e = e_0 \exp\left[\frac{L}{R_v}\left(\frac{1}{T_0} - \frac{1}{T_d}\right)\right] \tag{6}$$

where p is the pressure, $\varepsilon = R_d/R_v$ is the ratio of the specific gas constants of the dry atmosphere and water vapour, e_0 and L are the saturation vapor pressure and enthalpy of vaporization at a reference temperature T_0.

We rewrite Eqn. 2 into the form of a discretized summation. For every timestep, the T_m for each gridpoint and vertical level is found by performing the summation of Eqn. 2 from the top of atmosphere to the height of each vertical level. We therefore obtain T_m for 37 different vertical levels. Since the lowest pressure level output in reanalysis models is at 1000 hPa, we note that the surface is often below this pressure coordinate. The conditions of the lower boundary are therefore established by the following method:

- If the surface geopotential z is higher than the geopotential at pressure level 1000 hPa, then
  - The 38th level taken to be at 1012.35 hPa.
  - The height of the 38th level is found using the hydrostatic balance, with the temperature given by surface temperature T_s.
- If the surface geopotential z is lower than the geopotential at pressure level 1000 hPa, then
  - The 38th level taken to be at the surface, and the vertical height is the surface orographic height.

The T_m at the surface is found using 1-D spline interpolation of degree $k = 1$ (i.e. linear interpolation) to the surface height z using the Fortran DIERCKX library (Dierckx, 1995) wrapped in Julia (`https://github.com/kbarbary/Dierckx.jl`).

2.1.1 Integration using pressure coordinates

As mentioned in previous sections, reanalysis output is given in pressure coordinates rather than vertical coordinates. Due to hydrostatic balance in the atmosphere and the relatively slow nature of vertical transport as opposed to horizontal flow, it can be assume that pressure is monotonically increasing downwards. Therefore, instead of retrieving the geopotential height of each pressure level at every different space-time coordinates and integrating in vertical coordinates, we integrate in pressure coordinates that are fixed in time and space. This reduces the amount of data needed to be downloaded and extracted in order to calculate T_m and therefore saves time and computational cost by about a quarter.

Assuming that R and g are constant throughout the atmosphere, the transformation of coordinates from vertical height to pressure can be performed using Eqn. 7-8:

$$\begin{aligned} \frac{\mathrm{d}p}{\mathrm{d}z} &= -\rho g \quad \text{(By hydrostatic balance)} \\ \therefore \mathrm{d}p &= -\rho g \,\mathrm{d}z = -\frac{p}{RT} g \,\mathrm{d}z \\ \therefore \mathrm{d}z &= -\frac{T}{p}\frac{R}{g}\,\mathrm{d}p \end{aligned} \tag{7}$$

$$\therefore T_m = \frac{\sum_i \frac{e_i}{T_i}\Delta z_i}{\sum_i \frac{e_i}{T_i^2}\Delta z_i} = \frac{-\sum_i \frac{e_i}{T_i}\frac{T_i}{p_i}\frac{R}{g}\Delta p_i}{-\sum_i \frac{e_i}{T_i^2}\frac{T_i}{p_i}\frac{R}{g}\Delta p_i} = \frac{\sum_i \frac{e_i}{p_i}\Delta p_i}{\sum_i \frac{e_i}{p_i T_i}\Delta p_i} \tag{8}$$

The following variables are required to calculate T_m in pressure coordinates:

- 3D Datasets: air temperature T, specific humidity q
- Surface Datasets: surface temperature, dewpoint and pressure T_s, T_d and p_s

As with integration in vertical coordinates, we note that since the lowest pressure level output in reanalysis models is at 1000 hPa, the surface pressure may often exceed 1000 hPa, especially over the ocean. In a manner similar to that when using vertical coordinates, the conditions of the lower boundary are therefore established by the following method:

- If the surface pressure z is lower than 1000 hPa, then
  - The 38th level taken to be at 1012.35 hPa.
- If the surface geopotential z is higher than 1000 hPa, then
  - The 38th level taken to be the surface pressure.

And T_m at the surface is found using 1-D spline interpolation of degree $k = 1$ (i.e. linear interpolation) to the surface pressure p_s using the Fortran DIERCKX library (Dierckx, 1995) wrapped in Julia (`https://github.com/kbarbary/Dierckx.jl`).

In section 3.2, we will compare the results between Π_{RE5} and Π_{REP}, where the latter denotes the Π dataset derived from ERA5 reanalysis output and integrated in pressure coordinates.

2.1.2 The GGOS Atmospheres Model

The GGOS Atmosphere model (Böhm & Schuh, 2013) is also similarly based off reanalysis data, specifically ERA-Interim data. The T_m output from the GGOS Atmosphere model has been used to create empirical models such as GTm-III (Yao et al., 2014) and has itself been used as ground truth in various studies (e.g. Lan et al. (2016); Liu et al. (2015)). We evaluate our method above against the GGOS Atmosphere T_m values from 1979-2018 by comparing the RGA dataset against the REI dataset, as REI also uses ERA-Interim data to calculate T_m. However, we are unable to do a direct comparison of our methodology with GGOS Atmosphere, because the method that Böhm and Schuh (2013) used to calculate T_m was not explicitly given.

The resolution of the RGA dataset is 2.5° longitude by 2.0° latitude. Therefore, to re-grid the GGOS Atmosphere output to the required 1.0° resolution we used 2-D cubic spline interpolation of degree $k = 3$, again using Julia-wrapped Fortran DIERCKX library (Dierckx, 1995) (`https://github.com/kbarbary/Dierckx.jl`). This may result in relatively large oscillations in regions where the topography changes sharply.

2.2 Approximating T_m using empirical relationships

Because vertical profiles of the atmosphere are sparse compared to surface measurements, several empirical models over the years have been developed in order to estimate either T_m or Π directly based on more easily retrieved meteorological variables such as surface temperature T_s, or even simply as a function of location and time.

2.2.1 Empirical models based on surface temperature

Bevis et al. (1992) was the first to use a linear method to estimate T_m from surface temperature T_s. From over 8000 radiosonde profiles in the United States, he derived the following linear relationship between T_m and T_s:

$$T_m = a + b \cdot T_s \tag{9}$$

where $a = 70.2$ and $b = 0.72$. However, it has long been recognised that these coefficients are highly dependant on location and season, and therefore there have been studies done to estimate a and b on regional scales for better estimates of T_m. Manandhar et al. (2017) derived estimates for a and b based on three different latitude categories: tropical, subtropical and temperature, which we also used to see if there was any significant difference when adjusting the linear relationship based on region. As it has been previously found that reanalysis models are able to model surface temperature with a high degree of accuracy, we use ERA5 surface temperature data here to estimate T_m via the linear method of Eqn. 9 using the coefficients of both Bevis et al. (1992) and Manandhar et al. (2017), to obtain the Π datasets Π_{EBB} and Π_{EBM} respectively.

2.2.2 Empirical models based on location and time

Unlike the methods described in Section 2.1 and 2.2.1 that are used to find T_m, empirical models such as GPT2w (Böhm et al., 2015), GTm-III and Manandhar et al. (2017) (hereafter referred to as MN2017) do not require any meteorological or climatological variables as input to calculate T_m. Instead, these models require as input only time/day-of-year and (x, y, z) positions. These empirical models are of course based on T_m values either directly calculated from reanalysis data (e.g. GPT2w), or from other models and approximations (e.g. GTm-III, MN2017).

In our study, we look at the GPT2w and MN2017 models and compare the mean values and variability of T_m in these models compared to T_m values calculated directly from reanalysis. As GPT2w is ultimately derived from ERA-Interim reanalysis data, it would be more appropriate to compare Π_{EG2} (the Π dataset derived from GPT2w) to Π_{REI} instead of Π_{RE5} in order to determine the veracity of the GPT2w model creation. However, we compare Π_{EMN} (the Π dataset derived from MN2017) to Π_{RE5}. We note that the temporal variability for both the GPT2w and MN2017 models are only valid for scales longer than the daily scale, and that they do not account for interannaul variabiliy. Therefore, we only investigate the intraseasonal and seasonal variability $\Delta_i(\cdot)$ and $\Delta_s(\cdot)$ for these two datasets.

3 Results

3.1 The spatial and temporal variability of Π_{RE5}

We plot the spatial and temporal variability of Π_{RE5} in Fig. 1. From the mean value μ_Π given in Fig. 1a, we note that Π is generally higher over the ocean as compared to over land, with the maximum mean values being found in the tropical Pacific and the Arabian Sea, and is lowest in Antarctica. We also note that Π is generally lower over regions of high topography, such as the Tibetian Plateau and Andes mountain range.

Of more interest is the temporal variability of Π at different scales. We note that the interannual variability $\Delta_a\Pi$ is small compared to the other modes of variability, with a minimum over the equator and increasing towards the poles (Fig. 1b). What is more interesting is that seasonal $\Delta_s\Pi$ and intraseasonal $\Delta_i\Pi$ variability are not highest over the pole in the northern hemisphere, but rather over eastern Sibera and Hudson Bay (Figs. 1d,e). This means that the magnitude of both $\Delta_s\Pi$ and $\Delta_i\Pi$ are tied to the variability of the polar vortices and the middle-troposphere, as opposed to a response based purely on surface temperature T_s. We deduce this based on the fact that both $\Delta_s\Pi$ and $\Delta_i\Pi$ are more significant over the ocean than would be otherwise expected from T_s alone.

Another empirical piece of evidence that the middle-troposphere is important in explaining the temporal variability of Π_{RE5} as opposed to surface temperature, lies in the fact that the diurnal variability $\Delta_d\Pi$ is greatest not over land, but rather predominantly over the regions where the mid-latitude storm tracks are found (Fig. 1c). In fact, in contrast to would be expected if the diurnal variability of Π was mostly due to T_s, the greatest variability in $\Delta_d\Pi$ is found over the ocean where storm tracks are most dense, rather than over land. We note that intraseasonal variability $\Delta_i\Pi$ also shows mild intensification over storm track regions, though this is partially overshadowed by variability close to the polar vortices (Fig. 1d).

We also investigate the relative importance of the different scales of temporal variability as a proportion of the total mean-weighted variability $\Delta_t\Pi$:

$$\Delta_t\Pi = \Delta_d\Pi + \Delta_i\Pi + \Delta_s\Pi + \Delta_a\Pi + \Delta_r\Pi \tag{10}$$

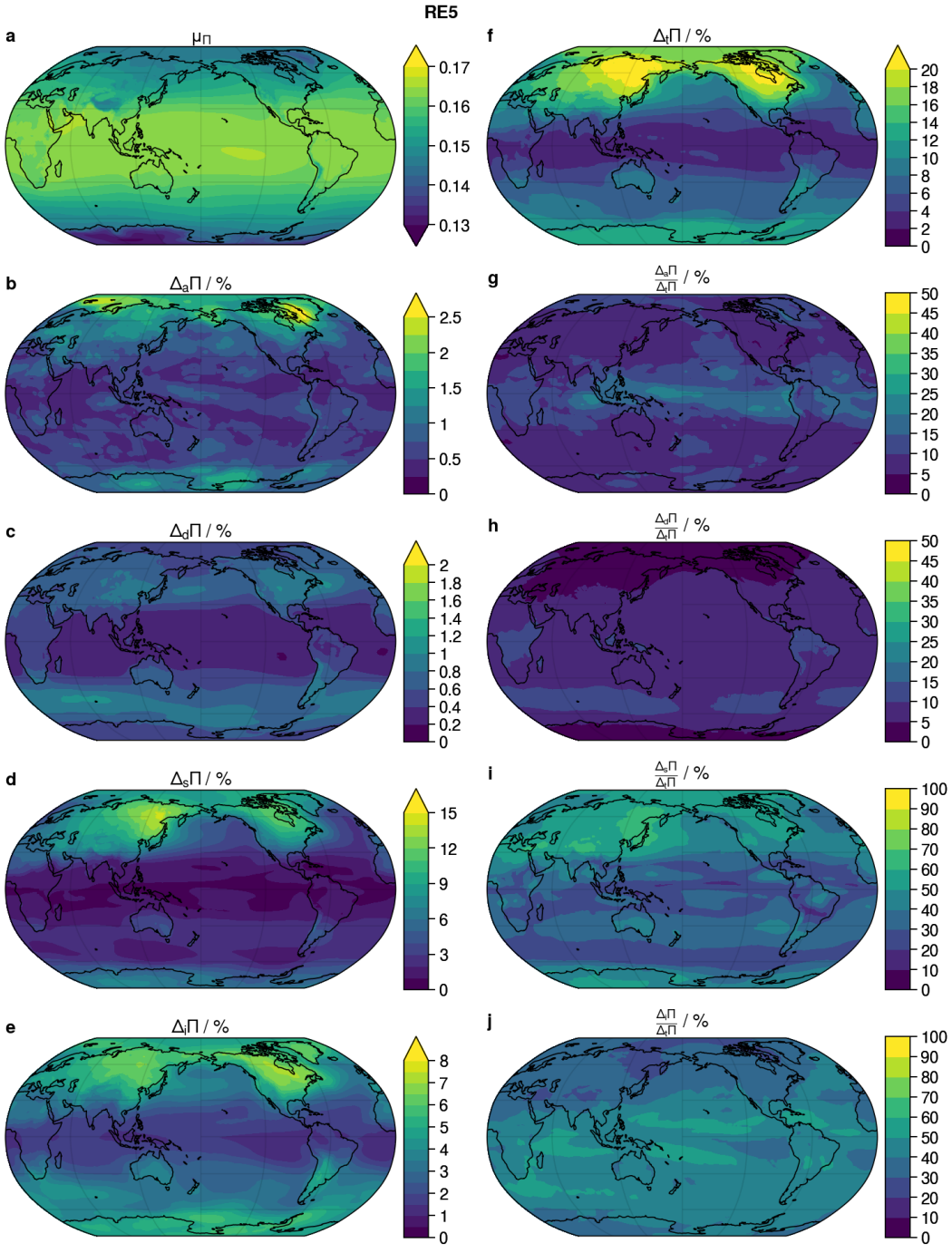

Figure 1. The spatial distribution of the (a) mean value μ_Π, (b) interannual variability $\Delta_a\Pi$, (c) diurnal variability $\Delta_d\Pi$, (d) seasonal variability $\Delta_s\Pi$ and (e) intraseasonal variability $\Delta_i\Pi$ for the dataset Π_{RE5}

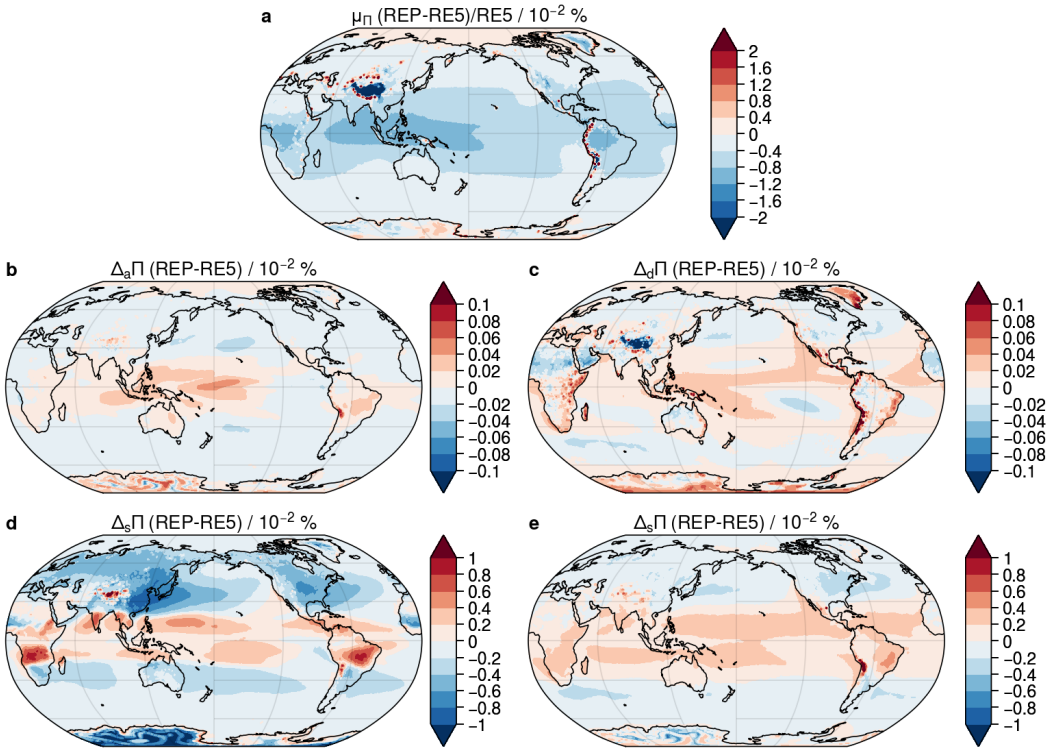

Figure 2. The spatial distribution of the difference in (a) mean value μ_Π, (b) interannual variability $\Delta_a\Pi$, (c) diurnal variability $\Delta_d\Pi$, (d) seasonal variability $\Delta_s\Pi$ and (e) intraseasonal variability $\Delta_i\Pi$ between the datasets Π_{RE5} and Π_{REP}

where $\Delta_r\Pi$ is the residual variability (which can be either positive or negative). When $|\Delta_r\Pi| \gg 0$, this indicates that the different modes of temporal variability are not independent of each other.

We see from Fig. 1h that the diurnal variability $\Delta_d\Pi$ does not play a large role in accounting for the variability of Π_{RE5}. Although it can be seen from Fig. 1h that $\Delta_d\Pi >$ 10% over a significant portion of land in the tropics, $\Delta_t\Pi < 10\%$ as well (Fig. 1f), and so overall $\Delta_d\Pi$ is not significant compared to μ_Π. In fact, in the tropics, interannual variability $\Delta_a\Pi$ may actually have a larger impact than $\Delta_d\Pi$ as the total variability $\Delta_t\Pi$ drops significantly with latitude. This drop in $\Delta_t\Pi$ as one moves closer to the equator is dominated by the change in seasonal variability $\Delta_s\Pi$. Indeed, we see that the importance of $\Delta_s\Pi$ does decrease as one moves to the equator, but still accounts for $\sim$ 30-50% of the total variability (Fig. 1i). As $\Delta_s\Pi$ decreases towards the equator, $\Delta_i\Pi$ therefore correspondingly becomes an increasingly important component of temporal variability, which is reflective of the unstable atmosphere of the tropics.

3.2 Integrating in Vertical vs. Pressure coordinates

In Section 2.1.1, we showed that we could integrate Eqn.1 in pressure coordinates by doing the coordinate transformations in Eqns. 7, 8, and in so doing reduce the amount of reanalysis data required to be downloaded and extracted. This saves a significant amount of overhead time and computational cost, but we wish to see if the resulting dataset Π_{REP} is close enough to Π_{RE5} that this method can be used to create the eventual gridded dataset.

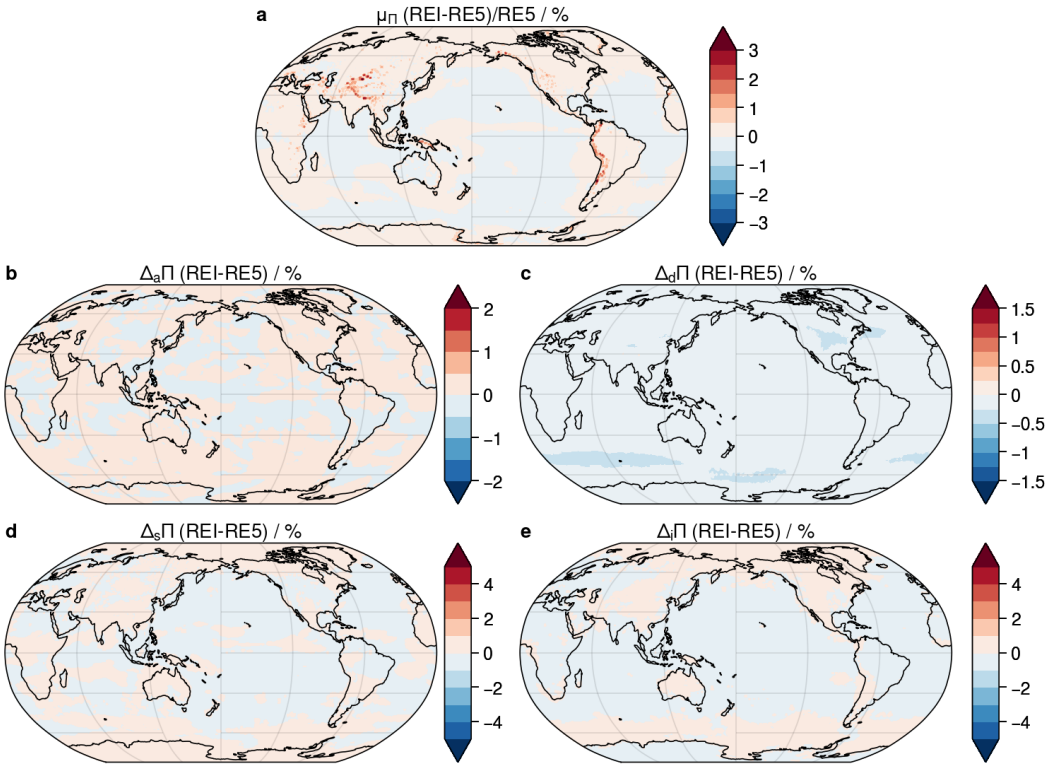

Figure 3. The spatial distribution of the difference in (a) mean value μ_Π, (b) interannual variability $\Delta_a\Pi$, (c) diurnal variability $\Delta_d\Pi$, (d) seasonal variability $\Delta_s\Pi$ and (e) intraseasonal variability $\Delta_i\Pi$ between the datasets Π_{RE5} and Π_{REP}

We therefore plot the spatial distribution of the difference in mean value μ_Π and mean-weighted temporal variability at all scales, between Π_{REP} and Π_{RE5}.

We see that the differences between Π_{RE5} and Π_{REP} are extremely small, on the order of $\sim \mathcal{O}(0.01\%)$ of μ_Π. The differences in mean value are, once again, generally largest in regions of high topography. However, apart from these regions, the differences between the two datasets are also most pronounced (if still small enough in magnitude to be considered negligible) in the tropical regions. The exception is seasonal variability $\Delta_s\Pi$, where the differences between the two datasets are found in the subtropical regions. Nonetheless, once again the differences between the two datasets is so small that they might as well be considered negligible, and therefore we can say that for all practical purposes that $\Pi_{\text{RE5}} \approx \Pi_{\text{REP}}$.

4 Comparison against other models

In this section, we compare our base dataset Π_{RE5} against the other datasets listed in Table 1.

4.1 Comparison against Π derived from ERA-Interim reanalysis

ERA5 is the fifth-generation ECMWF reanalysis product, meant to succeed ERA-Interim, which ceased operation in August 2019. We therefore wish to see if transitioning from ERA-Interim to ERA5 will cause any noticeable differences in the spatial-temporal variability of Π. One noticeable difference is that ERA5 reanalysis output is hourly, while

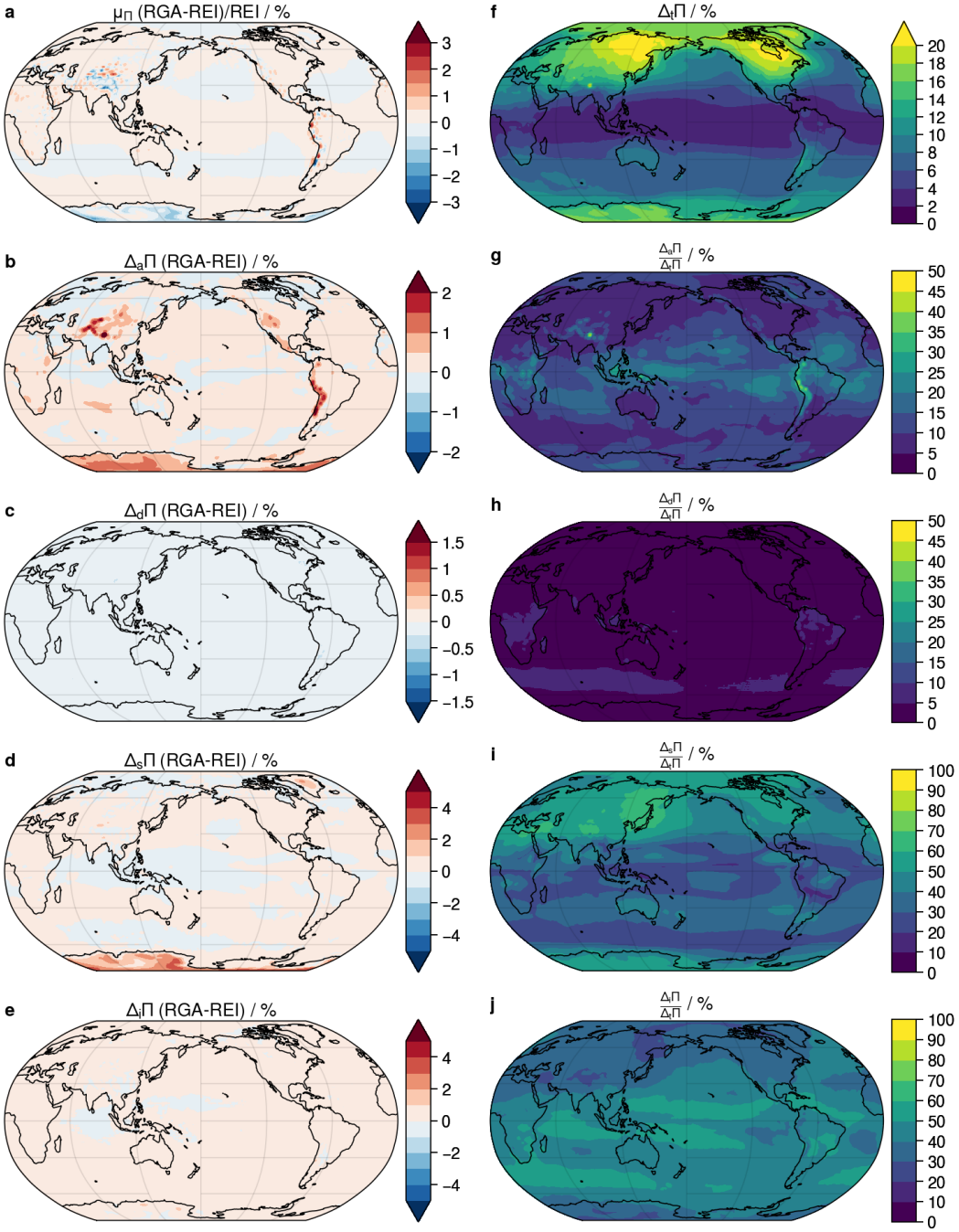

Figure 4. The spatial distribution of the difference in (a) mean value μ_Π, (b) interannual variability $\Delta_a\Pi$, (c) diurnal variability $\Delta_d\Pi$, (d) seasonal variability $\Delta_s\Pi$ and (e) intraseasonal variability $\Delta_i\Pi$ between the datasets Π_{RE5} and Π_{REP}

ERA-Interim output is 6-hourly. This means that ERA5 may be able to better represent the diurnal cycle of variability than ERA-Interim. For the purposes of comparison, we retrieved 1.0° spatial-resolution data for both datasets, and therefore we are unable to investigate the impact that increased spatial resolution has on ERA5 output compared to ERA-Interim at the smaller scale.

Nonetheless, from Fig. 3 we see that the differences between Π_{REI} and Π_{RE5} are in actuality very small compared to the differences between Π_{RE5} and the other datasets that we will compare against later. The most major differences are found in μ_Π and $\Delta_d\Pi$. We see that μ_Π tends to be larger in Π_{REI} compared to Π_{RE5} in regions where topography is more complex, and this is likely due to differences in the orographic heights between ERA5 and ERA-Interim. Meanwhile, $\Delta_d\Pi$ is shown to be smaller in Π_{REI} as compared to Π_{RE5}, and this is due to the higher temporal resolution of ERA5 that allows for a better representation of the maximum and minimum of the diurnal cycle as compared to 6-hourly output from ERA-Interim.

4.2 Comparison against Π derived from the GGOS Atmosphere Model

The usage of cubic spline interpolation to regrid the GGOS Atmosphere Model T_m output from 2.5° longitude by 2.0° latitude to a 1.0° resolution used in this study can indeed result in unwanted oscillations. However, when we compare Π_{RGA} against Π_{REI} (which is more appropriate as GGOS Atmospehre T_m output is based off ERA-Interim data) shows some minor oscillations in μ_Π in regions of complex topography and is especially prominent over the Tibetian Plateau and Andes mountains (Fig. **??**). The magnitude of the difference in μ_Π between the two datasets is on the order of $\sim \mathcal{O}(1\%)$ (Fig. **??**a), which indicates that the GGOS Atmosphere model output is, on a time-averaged basis, very close to that of our own derivation of Π from reanalysis data.

However, interannual variability $\Delta_a\Pi$ is greater for Π_{RGA} compared to Π_{REI}. Although $\Delta_a\Pi$ is itself very small, at $\sim \mathcal{O}(2{-}3\%)$ of μ_Π, it is worth noting that $\Delta_a\Pi_{\mathrm{RGA}}$ is 50% greater than $\Delta_a\Pi_{\mathrm{REI}}$ over regions of high topography. Combined with the differences in μ_Π between Π_{RGA} and Π_{REI}, we conclude that Π_{RGA} is less reliable over regions of sharp topography as compared to Π_{REI}. This is especially true for the tropical regions, where the total mean-weighted variability $\Delta_t\Pi$ is already relatively small, so any increase in $\Delta_a\Pi$ can have a significant impact on its relative importance. Indeed, we see that along the Andes mountains, the Tibetian plateau and the Maritime Continent where topography is steep, $\Delta_a\Pi$ accounts for around 30% of the total variability $\Delta_t\Pi$ (Fig. **??**).

Lastly, we also note that $\Delta_s\Pi$ sems to be greater over the polar ice caps in Greenland and Antartica. However, since we are unable to retrieve the exact method used to calculate T_m in the GGOS Atmosphere model, we cannot offer anything more than conjecture as to the differences.

4.3 Comparison against Π derived from Linear Estimation Methods

The results from Section 3.1-3.3.2 imply that using a linear model to derive T_m from T_s as in the form of Eqn. 9 results in significant differences in both the mean-state and in temporal variability. We compare our base dataset Π_{RE5} with two different linear models, Π_{EBB} and Π_{EBM}. Π_{EBB} is derived using the original equations from Bevis et al. (1992), while Π_{EBM} updates the original coefficients with those derived by Manandhar et al. (2017) as shown in Table 2.

It is made clear from Fig. 5 that changing the coefficients of a and b from Bevis et al. (1992) to Manandhar et al. (2017) in Eqn. 9 does little to change the biases observed in Π_{EBB} (Fig. 5a-e) compared to Π_{RE5}. In fact, we see that the mean climatology μ_Π of Π_{EBM} actually shows greater biases overall as compared to Π_{EBB}, especially

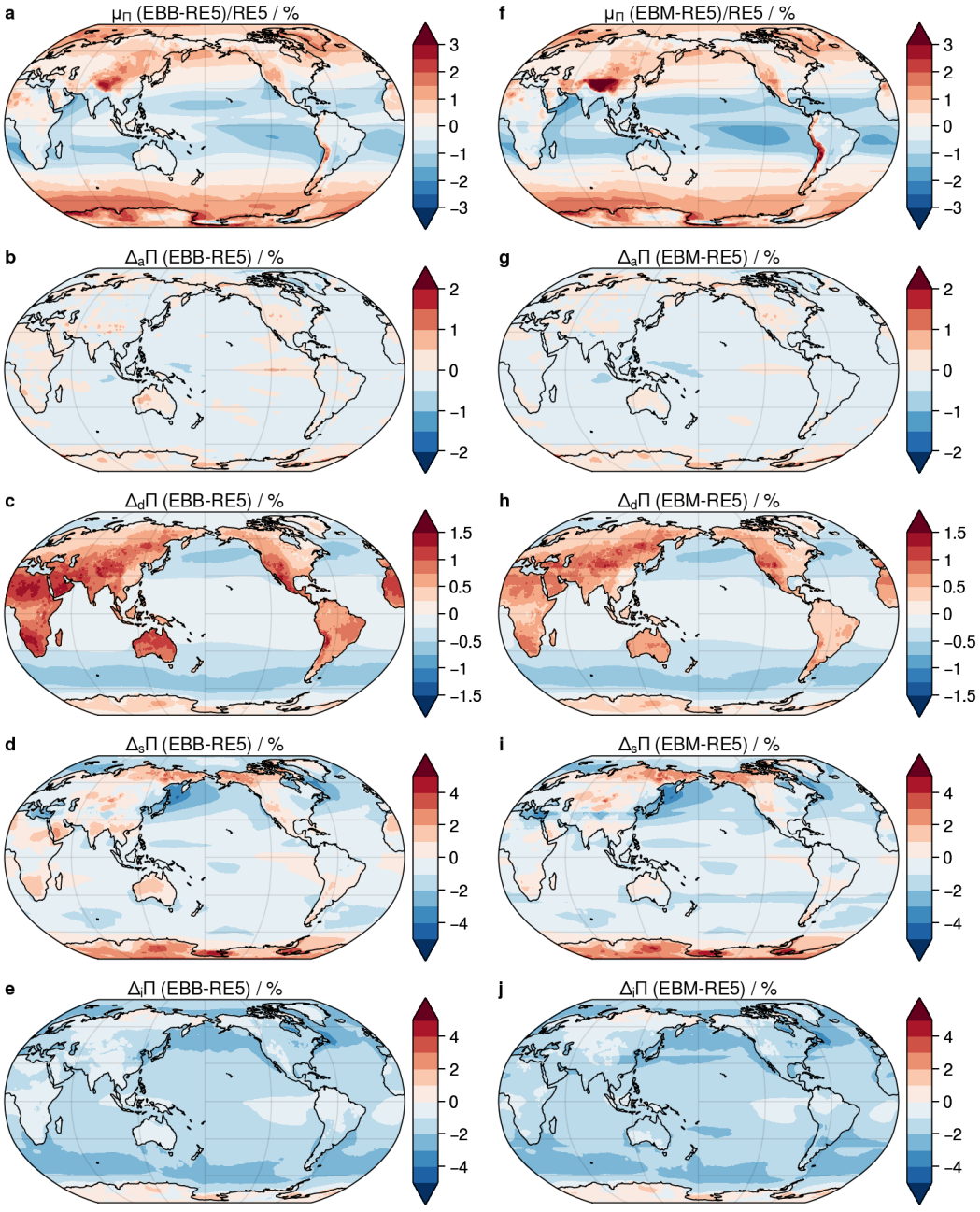

Figure 5. The spatial distribution of the difference in (a) mean value μ_Π, (b) interannual variability $\Delta_a\Pi$, (c) diurnal variability $\Delta_d\Pi$, (d) seasonal variability $\Delta_s\Pi$ and (e) intraseasonal variability $\Delta_i\Pi$ between the datasets Π_{RE5} and Π_{REP}

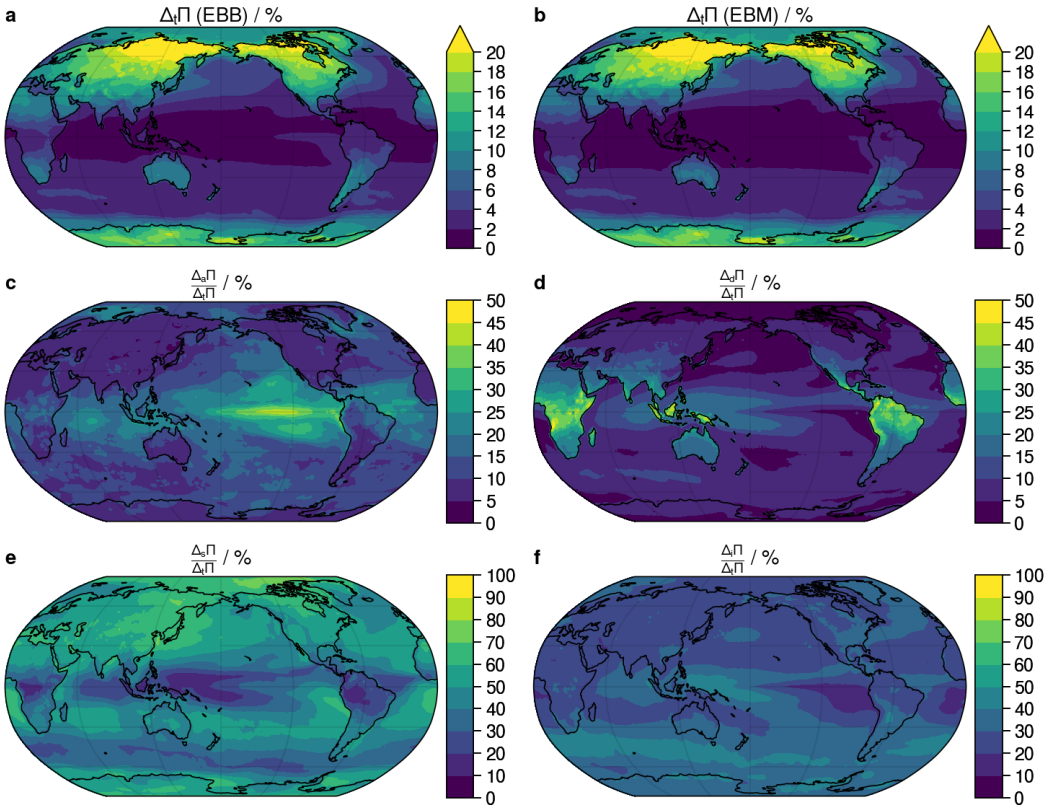

Figure 6. The spatial distribution of the difference in (a) mean value μ_Π, (b) interannual variability $\Delta_a\Pi$, (c) diurnal variability $\Delta_d\Pi$, (d) seasonal variability $\Delta_s\Pi$ and (e) intraseasonal variability $\Delta_i\Pi$ between the datasets Π_{RE5} and Π_{REP}

oer the ocean (where low biases become more prominent) or over mountains (where high biases are stronger).

ID	Latitude (ϕ) / °	a	b
EBB	$\lvert\phi\rvert \leq 90°$	70.2	0.72
EBM	$\lvert\phi\rvert \leq 23°$	129.13	0.52
	$23° < \lvert\phi\rvert < 36°$	106.36	0.60
	$\lvert\phi\rvert \geq 36°$	67.12	0.73

Table 2. A summary of the methods used to calculate T_m, and therefore Π.

The only clear advantage that Π_{EBM} has over Π_{EBB} is that the high-bias in $\Delta_d\Pi_{\mathrm{EBM}}$ over land is much less pronounced than for $\Delta_d\Pi_{\mathrm{EBB}}$. Overestimating the effect of the diurnal variability of Π may not be important in the mid-latitudes or polar regions where seasonal variability $\Delta_s\Pi$ overwhelmingly dominates the nature of variability in Π. However. in low-latitude regions where there is almost no seasonal variability (and from Figs. 5d,e,i,j it can clearly be seen that the linear relationship underestimates of both $\Delta_s\Pi$ and $\Delta_i\Pi$), overestimating that diurnal variability by even $\sim\mathcal{O}(1\%)$ relative to μ_Π can cause large chages to $\Delta_t\Pi$ (see Fig. 6d). We note that since the relationship between T_m

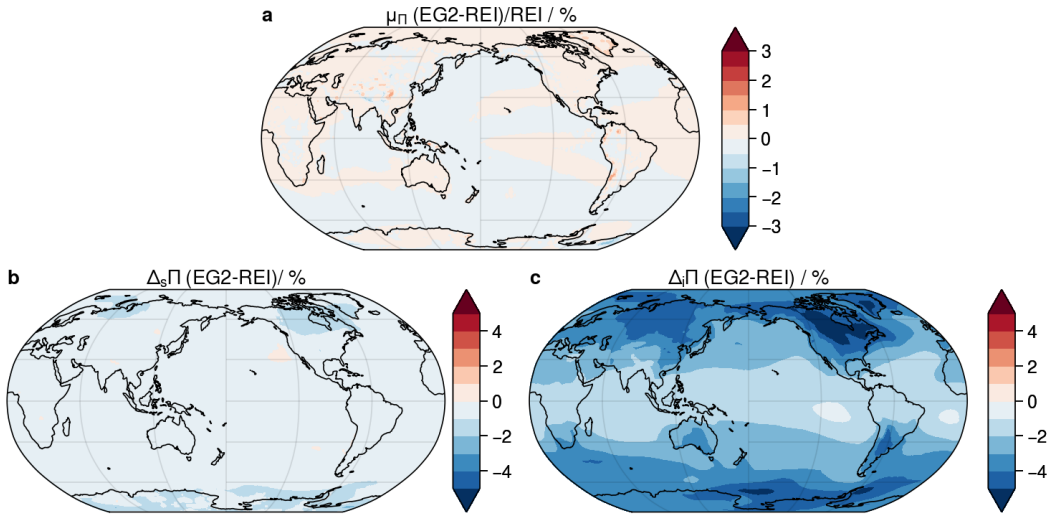

Figure 7. The spatial distribution of the difference in (a) mean value μ_Π, (b) interannual variability $\Delta_a\Pi$, (c) diurnal variability $\Delta_d\Pi$, (d) seasonal variability $\Delta_s\Pi$ and (e) intraseasonal variability $\Delta_i\Pi$ between the datasets Π_{RE5} and Π_{EG2}

and T_s is linear in Bevis et al. (1992), the importance of the different temporal modes in proportion to the total variability $\Delta_t\Pi$ is the same for Π_{EBB} and Π_{EBM}.

4.4 Comparison against Π derived from GPT2w

The GPT2w model (Böhm et al., 2015) is one of the most widely used models in GNSS Meteorology. It is based on 10 years of monthly-mean data from ERA-Interim at 1° horizontal resolution to calculate not just T_m, but also estimates of zenith-wet-delay and waver vapour pressure. Therefore it makes more sense to compare with Π_{EG2} (Π derived from the GPT2w model by converting from T_m using Askne and Nordius (1987)) with Π_{REI} instead of Π_{RE5}. Since GPT2w used only monthly mean data, it is meant to be used on the daily-seasonal scale as it is unable to account for the presence of the diurnal cycle. Furthermore, as it is an empirical model, interannual variability is not a relevant detail here. Our results are plotted in Fig. 7.

As would be expected, the differences in the yearly climatology μ_Π between Π_{EG2} and Π_{REI} are minimal except in regions of high topography. The difference in seasonal variability $\Delta_s\Pi$ (EG2−REI) is also relatively small at $< \mathcal{O}(1\%)$ compared to the actual seasonal variability of Π_{REI} ($\Delta_s\Pi_{\mathrm{REI}} < \mathcal{O}(15\%)$ in the subpolar regions). However, we see that day-to-day variability (given by intraseasonal variability $\Delta_i\Pi$) is much smaller when derived from the GPT2w model as opposed to when it is calculated directly from reanalysis data. This is due to the fact that, as mentioned above, the GPT2w model was derived from monthly-mean reanalysis data instead of the full reanalysis dataset and therefore is unable to adjust for daily variability.

From Π_{RE5} (Fig. 1), we see that $\Delta_i\Pi$ in the mid-latitudes and subpolar regions accounts for up to 30−40% of the total variability. The loss of the majority of $\Delta_i\Pi$ therefore means that a large portion of the variability in Π is lost. This means that using the GPT2w model to obtain values of Π may result in a measure of noticeable random error as compared to the true value derived from reanalysis data.

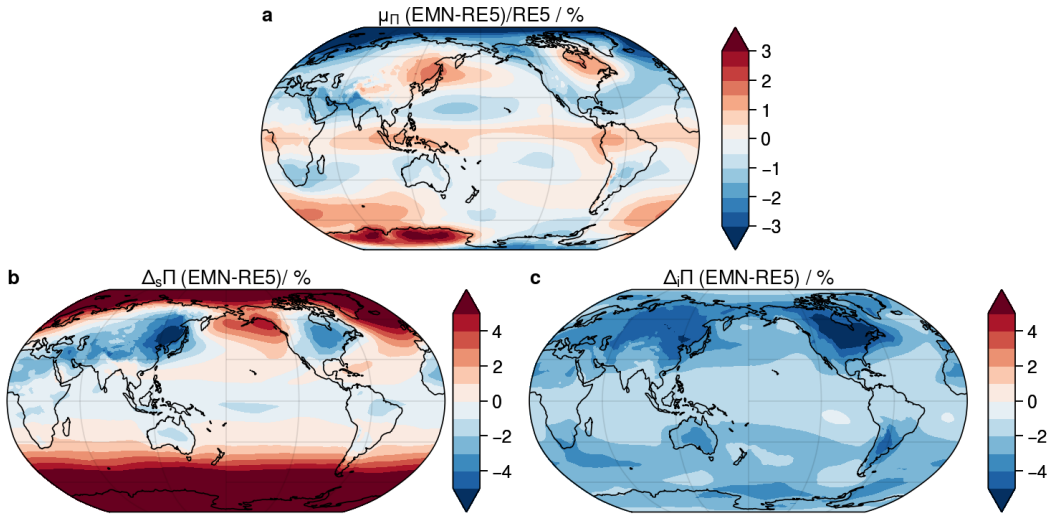

Figure 8. The spatial distribution of the difference in (a) mean value μ_Π, (b) interannual variability $\Delta_a\Pi$, (c) diurnal variability $\Delta_d\Pi$, (d) seasonal variability $\Delta_s\Pi$ and (e) intraseasonal variability $\Delta_i\Pi$ between the datasets Π_{RE5} and Π_{EG2}

4.5 Comparison against Π derived from PI2017

MN2017 is an empirical model for Π created by Manandhar et al. (2017) based on latitude, day-of-year and altitude. However, from Fig. 1a we see that Π_{RE5} has significant zonal asymmetry, and the zonally-symmetric nature of Π_{EMN} already causes significant biases in the climatological mean μ_Π, often in an oscillating manner as one moves around a given latitude-circle (Fig. 8a). Furthermore, because of this zonal=asymmerty in the actual values of Π derived from reanalysis data, we see that the maximum mean-weighted seasonal variability $\Delta_s\Pi$ of Π_{EMN} occurs not in the subpolar regions like in Π_{RE5}, but at the poles themselves. This causes a very noticeable high-bias in $\Delta_s\Pi$ for Π_{RMN} compared to Π_{RE5} (Fig. 8b), especially in the southern hemisphere.

We also note that MN2017 is an empirical model that accounts only for seasonal variability, as can be seen by Eqn. **??**, where the dependence of Π_{EMN} with time is iven by a cosine function of day-of-year. This means that intraseasonal variability $\Delta_i\Pi$ is considered negligible, which can be seen by Fig. 8c, where the difference in intraseasonal variability between Π_{RE5} and Π_{EMN} are given by almost the entire magnitude of $\Delta_i\Pi$ of Π_{RE5}.

5 Conclusion

We have derived a globally-gridded dataset for both the water-vapour-weighted mean column temperature T_m and the constant of proportionality Π between zenith wet delay (ZWD) and precipitable water vapour (PWV) using the latest generation of the ECMWF reanalysis, ERA5. The main benefit of using ERA5 over ERA-Interim is due to higher native spatial and temporal resolution of reanalysis output, which allows for better representation of complex topography and the diurnal cycle, as seen by comparing between the resulting datasets Π_{RE5} and Π_{REI} (Fig. 3). Seasonal and intraseasonal variability are very similar between the two, however.

Downloading reanalysis datasets simply to calculate Π may seem cumbersome to some, but we have also explored several different empirical models previously published

that aim to reduce or even eliminate the amount of data required, such as the linear estimation of Bevis et al. (1992) and the GPT2w model Böhm et al. (2015). We have explored the strengths and weaknesses of Π derived from these models with our own dataset, and have found that the linear Bevis method overemphasizes the diurnal cycle of variability of Π, while GPT2w is unable to account for day-to-day variability on a scale shorter than a seasonal/monthly timescale.

In light of this, we therefore have created a dataset for T_m and Π that we shall host, derived from the full 0.25° spatial and hourly temporal resolution of the ERA5 reanalysis dataset. We hope that such a dataset will become useful for the GNSS community, which strives to increase the accuracy and precision of GNSS positioning datasets.

Acknowledgments

Enter acknowledgments, including your data availability statement, here.

References

Askne, J., & Nordius, H. (1987, 5). Estimation of tropospheric delay for microwaves from surface weather data. *Radio Science*, *22*(3), 379–386. Retrieved from http://doi.wiley.com/10.1029/RS022i003p00379 doi: 10.1029/RS022i003p00379

Bevis, M., Businger, S., Herring, T. A., Rocken, C., Anthes, R. A., & Ware, R. H. (1992). GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system. *Journal of Geophysical Research*, *97*(D14), 15787–15801. Retrieved from http://doi.wiley.com/10.1029/92JD01517 doi: 10.1029/92JD01517

Böhm, J., Möller, G., Schindelegger, M., Pain, G., & Weber, R. (2015, 7). Development of an improved empirical model for slant delays in the troposphere (GPT2w). *GPS Solutions*, *19*(3), 433–441. Retrieved from http://link.springer.com/10.1007/s10291-014-0403-7 doi: 10.1007/s10291-014-0403-7

Böhm, J., & Schuh, H. (Eds.). (2013). *Atmospheric Effects in Space Geodesy*. Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://link.springer.com/10.1007/978-3-642-36932-2 doi: 10.1007/978-3-642-36932-2

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011, 4). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. Retrieved from http://doi.wiley.com/10.1002/qj.828 doi: 10.1002/qj.828

Dierckx, P. (1995). *Curve and surface fitting with splines*. Oxford University Press.

Hersbach, H., & Dee, D. (2016). ERA5 reanalysis is in production. *ECMWF Newsletter*(147), 7. Retrieved from https://confluence.ecmwf.int/pages/viewpage.action?pageId=74764925

Lan, Z., Zhang, B., & Geng, Y. (2016, 3). Establishment and analysis of global gridded Tm Ts relationship model. *Geodesy and Geodynamics*, *7*(2), 101–107. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S1674984716300015 doi: 10.1016/j.geog.2016.02.001

Liu, L., Li, J., Chen, X., & Cai, C. (2015, 12). Precision analysis on the weighted mean temperature of the atmosphere grid data offered by GGOS Atmosphere in Xinjiang. In G. Zhou & C. Kang (Eds.), (p. 98083I). Retrieved from http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2207379 doi: 10.1117/12.2207379

Manandhar, S., Lee, Y. H., Meng, Y. S., & Ong, J. T. (2017, 11). A Simplified Model for the Retrieval of Precipitable Water Vapor From GPS Signal.

IEEE Transactions on Geoscience and Remote Sensing, *55*(11), 6245–6253. Retrieved from http://ieeexplore.ieee.org/document/7994650/ doi: 10.1109/TGRS.2017.2723625

Wang, J., Zhang, L., & Dai, A. (2005). Global estimates of water-vapor-weighted mean temperature of the atmosphere for GPS applications. *Journal of Geophysical Research*, *110*(D21), D21101. Retrieved from http://doi.wiley.com/10.1029/2005JD006215 doi: 10.1029/2005JD006215

Wang, X., Zhang, K., Wu, S., Fan, S., & Cheng, Y. (2016, 1). Water vapor-weighted mean temperature and its impact on the determination of precipitable water vapor and its linear trend. *Journal of Geophysical Research: Atmospheres*, *121*(2), 833–852. Retrieved from http://doi.wiley.com/10.1002/2015JD024181 doi: 10.1002/2015JD024181

Yao, Y., Xu, C., Zhang, B., & Cao, N. (2014, 4). GTm-III: a new global empirical model for mapping zenith wet delays onto precipitable water vapour. *Geophysical Journal International*, *197*(1), 202–212. Retrieved from http://academic.oup.com/gji/article/197/1/202/690874/GTmIII-a-new-global-empirical-model-for-mapping doi: 10.1093/gji/ggu008