Atmos. Chem. Phys. Discuss., 10, 16345–16384, 2010
www.atmos-chem-phys-discuss.net/10/16345/2010/
doi:10.5194/acpd-10-16345-2010
© Author(s) 2010. CC Attribution 3.0 License.

Atmospheric Chemistry and Physics Discussions

This discussion paper is/has been under review for the journal Atmospheric Chemistry and Physics (ACP). Please refer to the corresponding final paper in ACP if available.

Statistical dynamics of equatorial waves in tropical radiosonde wind data

T.-Y. Koh[1,2,3], **Y. S. Djamil**[1], **and C. K. Teo**[3]

[1]School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore
[2]Earth Observatory of Singapore, Nanyang Technological University, Singapore
[3]Temasek Laboratories, Nanyang Technological University, Singapore

Received: 1 May 2010 – Accepted: 16 June 2010 – Published: 1 July 2010

Correspondence to: T.-Y. Koh (kohty@ntu.edu.sg)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Weibull distributions were fitted to wind speed data from radiosonde stations in the global tropics. A statistical theory of equatorial waves was proposed to explain the shape parameter k obtained over Malay Peninsula and the wider Equatorial Monsoon Zone. The theory uses the $(-5/3)$-power law in quasi-2-D turbulence, classical Boltzmann statistics and the Central Limit Theorem. It provides a statistical dynamical basis for using empirical Weibull fits to derive wind speed thresholds for monitoring data quality. The regionally adapted thresholds retain more useful data than conventional ones defined from taking the regional mean plus three standard deviations. The new approach is shown to eliminate reports of atypically strong wind over Malay Peninsula which may have escaped detection in quality control of global datasets as the latter has assumed a larger spread of wind speed.

1 Introduction

Radiosonde observations provide arguably the most reliable long-term meteorological data, especially before the advent of satellites. They are used for routine weather analyses and forecasts, as well as validation of satellite retrievals (e.g., Divakarla et al., 2006; Stoffelen et al., 2005). Unprocessed radiosonde data contain many types of error (Gandin, 1988) and must pass through quality control (QC) before use. Because radiosonde data are collected all over the world under the auspices of World Meteorological Organization, QC methods are usually global in perspective and statistical in nature (e.g., Durre et al., 2006). The statistical methods are usually based on mathematics (e.g. by the use of standard deviation to detect outliers) rather than on dynamics (e.g. by examining properties emergent from an ensemble of physical systems). This may be because it is hard to generalize a single global statistical dynamics that is applicable to widely different climatic zones. Adopting the former "statistical mathematical" approach results in smaller regions with denser station network exerting greater influ-

ACPD

10, 16345–16384, 2010

ence in the formulation of QC criteria and thresholds than larger regions with more sparse network. The tropical landmasses in South America, Africa and Southeast Asia are good examples of the latter regions and the quality of radiosonde data from these regions requires some scrutiny even after QC.

In weather forecasting, modern data assimilation techniques incorporate additional QC based on the model first-guess fields and in-built error metrics. So erroneously large or small values in a data source that escape its own QC may be safely rooted out before assimilation. However, if the original QC of radiosonde data is excessively stringent for the lack of understanding of the possibility of extremal values, the data assimilation modules will not be able to recover the lost valid data. In that case, corrective relaxation of the original QC of radiosonde data will need to be accompanied by the extension of error metrics in data assimilation modules.

In the recent decades, there has been emerging interest in Southeast Asia by the international community studying the global atmosphere. Neale and Slingo (2003) pointed out that the diurnal cycle in the maritime continent is not well-captured by GCMs despite its importance to global circulation (Ramage, 1968). Zhu and Wang (1993) showed that strong interactions exist over Southeast Asia between the Asian-Australian monsoon and the intra-seasonal oscillations spanning the global tropics (Madden and Julian, 1971, 1994). There has been more research focused on this region's climate and weather, e.g. on El Nino impacts: Hendon (2003), Juneng and Tangang (2005); on Southeast Asian monsoon: Lau and Yang (1997), Chang et al. (2005); on tropical cyclones: Chang et al. (2003); on sea-breeze circulation: Hadi et al. (2002); Joseph et al. (2008). For the benefit of global and regional atmospheric research, besides gathering more data using non-conventional platforms in Southeast Asia (Koh and Teo, 2009), it is timely to re-examine the nature and quality of conventional radiosonde data from this region. This paper reports our investigations into the statistical dynamics of radiosonde wind data while on-going work on temperature and humidity data will be reported in future publications.

ACPD
10, 16345–16384, 2010

One may reasonably pose a general question: could the statistics of a set of wind data be understood from underlying *regional* atmospheric dynamics and thereby providing a basis for better quality monitoring? Unlike the global problem, a statistical dynamical approach is sound in principle here because the statistical properties of regional atmospheres are well determined by a few controlling factors from the region's climate (e.g. ambient stratification, humidity profile and prevailing wind pattern) and for the planetary boundary layer (PBL), from the surface characteristics (e.g. elevation, roughness, temperature, wetness). But there is a caveat: the underlying statistical dynamics must be revealed through data before QC; otherwise, data that could possibly reflect new physical understanding may have already been categorically rejected by existing QC methods based on statistical mathematics.

Literature on the statistical characterization of wind speed has mainly focused on the surface layer (e.g., Takle et al., 1978; LaBraga, 1994; Lun, 2000) and to a lesser extent, the PBL (e.g., Frank et al., 1997). Most of the literature employed the Weibull distribution (Wilks, 1995) to model wind speed. Justus et al. (1978) demonstrated that Weibull distribution fits surface wind better than the square-root-normal distribution used by Widger (1977). The Rayleigh distribution is another commonly used empirical fit for surface wind (Manwell et al., 2002) but this distribution is only a special case of the Weibull distribution with shape parameter k=2. The authors are unaware of any published characterization of tropospheric wind using Weibull distribution, but found Roney (2007) who fitted the Weibull distribution to lower stratospheric wind soundings. In all the reviewed literature, no quantitative explanation was attempted for why the Weibull distribution is a good fit to the wind data.

The objectives of this work are two-fold: (1) to elucidate the statistical dynamics of tropical wind by analyzing long-term records of raw radiosonde data from selected stations in Southeast Asia and extending the results to the wider tropics; (2) to assess the feasibility of using that statistical dynamical understanding for monitoring the quality of regional wind data. It is hoped that the results presented would motivate similar statistical dynamical studies in other tropical regions with sparse data coverage.

ACPD
10, 16345–16384, 2010

2 Data overview

Twice daily radiosonde observations were taken from the Department of Atmospheric Science, University of Wyoming (http://weather.uwyo.edu/upperair/sounding.html). Seven stations situated on the Malay Peninsula (MP) in Southeast Asia (Fig. 1) were used for the first part of this study. The peninsula spans a region of about 1200 km by 400 km oriented in the NW-SE direction. It is roughly the size of Great Britain or California, USA. It represents a conveniently sized region for which statistical homogeneity might be expected to underlie the prevailing mesoscale convective weather. The seven stations span the peninsula uniformly and together provide 35 years of data from 1973 to 2007 with some gaps interspersed in-between. Another 235 stations between 25° N and 25° S were used to test the extension of the findings from MP to the global tropics.

Wind speed at 00:00 UTC and 12:00 UTC on eleven mandatory pressure levels (1000, 925, 850, 700, 500, 400, 300, 250, 200, 150 and 100 mb) were used for all stations. For MP (Table 1), data was available for less than half the time at 12:00 UTC for Phuket and Songkhla, while data from Sepang and Phuket cover less than 20 years. Overall, the average number of stations on the peninsula reporting per day, out of seven total, lies in the range of 3.8 to 4.4 at 00:00 UTC and 2.6 to 3.1 at 12:00 UTC for all pressure levels excluding 925 mb (which has less data because it was only adopted as a mandatory level in the 1990s). Wind data were recorded to the nearest 1 knot and so where relevant, bin size in statistical analyses was specified in units of knots to avoid artificial clustering of data if otherwise specified in other units like $\mathrm{m\,s^{-1}}$.

3 Methodology

The frequency histogram for non-zero wind speed at 850 mb from all 7 stations on MP without quality control is shown in Fig. 2. Measurements of zero wind speed were ignored as they may actually denote calm condition or light wind speed which radiosonde records do not resolve. Equation (1) below shows the probability density function (PDF)

ACPD
10, 16345–16384, 2010

of the Weibull distribution that was empirically fitted to the wind speed data at each pressure level:

$$P(v;k,c)dv = \frac{k}{c}\left(\frac{v}{c}\right)^{k-1}\exp\left[-\left(\frac{v}{c}\right)^{k}\right]dv, \quad (1)$$

where v is the wind speed, c is the scale parameter and k is the shape parameter. Maximum Likelihood Estimation (MLE) was used to determine k and c (Wilks, 1995). The exponent k will be shown to be indicative of the underlying statistical dynamics.

The goodness of fit of the Weibull distributions to the histograms was tested with χ^2-statistics at 90% confidence level:

$$\chi^2 = \frac{n_{\text{tot}}}{\tau}\sum_{b=1}^{n_b}\frac{(p_{\text{obs},b}-p_{\text{fit},b})^2}{p_{\text{fit},b}},$$

where $p_{\text{obs},b}$ and $p_{\text{fit},b}$ are the observed and fitted probability of wind speed lying in bin b, n_b is the number of bins in the histogram and n_{tot} is the total number of data points. τ is a scale factor that compensates for the lack of independence among nearby data points in time and is taken as the critical number of days beyond which the lag auto-correlation of daily wind speed is not significant at 90% confidence level. χ^2 defined above is probably an upper bound on the true χ^2 because each station's measurement is not independent of the others. Thus, this χ^2-test is rather stringent, but it suffices for our purpose.

The root-mean-square (rms) velocity σ is given by

$$\sigma^2 \overset{\text{def}}{=} \int_0^\infty v^2 P(v)dv = c^2\int_0^\infty t^{2/k}\exp(-t)dt = c^2\Gamma\left(\frac{2}{k}+1\right), \quad (2)$$

where Γ is the gamma function. This implies that c is constrained by the climatological wind speed measured by σ for a given k. In this work, it found that $k\in[1.3,2.6]$ which implies $[\Gamma(2/k+1)]^{-1/2}\in[0.86,1.04]$. Thus in practice, $c\approx\sigma$ and c is a good indication of the climatological wind speed.

ACPD
10, 16345–16384, 2010

From the PDF, the expected number $n_{\text{fit}}(v, \delta v)$ of wind speed reports in the range (v, $v+\delta v$) was computed for the size of the dataset. The wind speed threshold v_{max} was defined such that $n_{\text{fit}} \leq 1$ for $v \geq v_{\text{max}}$. A value of $\delta v = 2$ knots was used because this was the finest resolution of wind speed for which the observed number n_{obs} approximates n_{fit}. (A large difference in n_{obs} was noted between odd- and even-valued v in knots, which may indicate that some stations actually measure in integral number of m s^{-1} but records in knots after applying the conversion $1\ \text{m s}^{-1} \approx 1.944$ knots.) By the Weibull distribution, wind speed records larger than v_{max} are unlikely to be reliable for the given dataset size. The principle of MLE ensures that estimates of k, c and hence v_{max} are not sensitive to the inclusion or exclusion of the rare extremal wind values.

4 Empirical results

Maximum-likelihood estimates of the scale and shape parameters as well as thresholds for wind speed at each level in MP are shown in Fig. 3. All Weibull fits are good according to the χ^2-test at 90% confidence level. Within the PBL (which for this paper is taken to be 850 mb and below), c increases with height but above the PBL, it is nearly invariant up to 500 mb with a value around 13 knots. In the upper troposphere, c increases sharply with height from 15.4±0.1 knots at 400 mb to reach a maximum of 42.2±0.2 knots at 150 mb.

k has the smallest value of 1.54±0.01 at 1000 mb and similar values around 5/3 in the PBL (925 mb and 850 mb) and at the tropopause (100 mb). The value of k increases upward from 1.71±0.01 at 700 mb to values somewhat bigger than 2 in the upper troposphere (300 mb to 150 mb).

The scaled threshold v_{max}/c shows the opposite vertical trend from k, which is expected from the threshold being pegged to $n_{\text{fit}}=1$: a larger k implies a stronger decay in the PDF at large v/c and hence a smaller scaled threshold. v_{max}/c has the largest value of 4.6 at 1000 mb and is around 4 in the PBL and at the tropopause. It decreases upward from 3.9 at 700 mb to values around 3 in the upper troposphere.

ACPD
10, 16345–16384, 2010

The same value of k=5/3 at the lower and upper boundaries of the troposphere despite different values of c and k≈2 in the upper troposphere characteristic of the Rayleigh distribution inspire confidence that the empirical fit does capture some underlying statistical dynamics of the equatorial troposphere, a topic to which we turn next.

5 Theoretical basis

The literature mentioned in Sect. 1 rarely justified the use of Weibull distribution beyond the fact that it does yield realistic fits to the observations. Moreover, most work dealt with surface wind for which the underlying assumptions may not be applicable to the troposphere or even the PBL. Therefore, the statistical dynamical underpinning of the Weibull distribution for near-equatorial wind must be sought anew from our understanding of atmospheric dynamics and statistical physics. The vertical profile of c is largely dictated by the climatology of planetary-scale Hadley and Walker circulation and the Asian-Australian monsoon. The vertical profile of k is the object of study in this section.

5.1 Equivalent shallow-water model

The theory of diurnal circulation (Rotunno, 1983) and the Matsuno-Gill model for trapped equatorial waves (Matsuno, 1966; Gill, 1980) and for Asian-Australian monsoon and Walker circulation (Gill, 1980) are useful for gaining fundamental understanding of the equatorial atmosphere. There are also intra-seasonal oscillations (ISO) (Madden and Julian, 1971, 1994) but these are less well understood. One attempt is found in Wang (2005) where a 1 1/2-layer model is proposed in which coupled Rossby-Kelvin wave modes reproduce realistic features of the observed equatorial ISO. For a more comprehensive review of ISO, see Waliser (2006). But underlying the above theories for equatorial atmosphere is the equivalent shallow-water model

ACPD
10, 16345–16384, 2010

which is adopted for the current work.

The horizontal wind vector $\boldsymbol{v}$ may be decomposed into different contributions, each associated with a different equivalent depth:

$$\boldsymbol{v} = \boldsymbol{v}_{\mathrm{WLK}} + \boldsymbol{v}_{\mathrm{AAM}} + \boldsymbol{v}_{\mathrm{ISO}} + \sum_{n} \boldsymbol{v}_{n} + \boldsymbol{v}_{\mathrm{DC}}, \tag{3}$$

where $\boldsymbol{v}_{\mathrm{WLK}}$, $\boldsymbol{v}_{\mathrm{AAM}}$, $\boldsymbol{v}_{\mathrm{ISO}}$ and $\boldsymbol{v}_{\mathrm{DC}}$ denote contributions due to the Walker circulation, Asian-Australian monsoon, ISO and diurnal cycle, respectively and the set $\{\boldsymbol{v}_{n}\}$ represents the contributions from equatorial waves of different vertical wave modes n and hence different equivalent depths (Wheeler and Kiladis, 1999). Over a long time such as 35 years, the set of values that each $\boldsymbol{v}_{X}$ takes (where X=WLK, AAM, ISO, n or DC) may be reproduced by the realizations of a random variable. Note that this is NOT saying that each contribution actually varies randomly in time.

5.2 Quasi two-dimensional turbulence

Due to strong vertical mixing throughout the PBL's shallow depth, the PBL may be considered as a shallow-water layer with the free troposphere above providing a non-rigid upper boundary condition. Likewise, gravity wave activity and associated turbulence from deep convective cloud tops are likely to create sufficient mixing for the tropopause region to be treated as a shallow internal boundary layer (IBL) with the free troposphere below and the lower stratosphere above providing non-rigid lower and upper boundary conditions, respectively.

The above concept of the PBL and IBL is different from the more conventional picture that they are boundary *surfaces* which experience the same internal waves as the troposphere and lower stratosphere. In the current picture, the PBL and IBL have small but finite thickness and respond to boundary conditions provided by the tropospheric interior and lower stratosphere. Because of the shallow depth, waves in the PBL or IBL exist in the fundamental mode, and so can be characterized by the layer's own

ACPD
10, 16345–16384, 2010

equivalent depth H[1]. Thus, we write

$$\boldsymbol{v} = \boldsymbol{v}_{\mathrm{WLK}} + \boldsymbol{v}_{\mathrm{AAM}} + \boldsymbol{v}_{\mathrm{ISO}} + \boldsymbol{v}_{H} + \boldsymbol{v}_{\mathrm{DC}}, \tag{4}$$

where $\boldsymbol{v}_H$ is the contribution from equatorial waves in the PBL or IBL.

Observed over long time scales, the PBL or IBL as a shallow-water flow exhibits quasi-two dimensional turbulence. The theoretical work of Weichman and Petrich (2001) showed that in turbulent shallow-water systems, coherent spatial structures form at large scales from energy inverse cascade from small scales.

The energy per unit mass E of a shallow-water system is:

$$E = \frac{1}{2}(\boldsymbol{v}\cdot\boldsymbol{v} + \phi),$$

where ϕ is geopotential at the shallow-water surface. In wave-number κ-space, the distribution of spectral energy density $\varepsilon = dE/d\kappa$ spans over an inertial range in which at steady state, there is no energy injection or dissipation so that E is conserved. The functional form of ε may be derived by dimensional analysis, following classic arguments advanced for Kolmogorov-type turbulence (Kolmogorov, 1941, 1962):

$$\left.\begin{aligned} [\varepsilon] &= L^3T^{-2} \\ [e] &= L^2T^{-3} \\ [\kappa] &= L^{-1} \end{aligned}\right\} \Rightarrow \varepsilon \sim e^{2/3}\kappa^{-5/3} \tag{5}$$

where L and T denote dimensions of length and time, e is the rate of energy injection per unit mass at small scales and is the only governing parameter in the inertial range.

Aircraft measurements in the Global Atmospheric Sampling Program (1975–1979) revealed a (−5/3)-power law for wavelengths between 6 km and 600 km (Gage and

[1]An analogy may be useful here: spectacles are often treated as infinitesimally thin lenses in optics; but they can also be understood as lenses with small but finite thickness where the light waves excited within the glass experience a refractive index distinct from that of the air on either side.

ACPD
10, 16345–16384, 2010

Nastrom, 1986). These measurements were between 9 km and 14 km altitude and mostly concentrated between 30° N and 65° N over Pacific Ocean, North America, Atlantic Ocean and Europe (Nastrom and Gage, 1985). Larsen et al. (1982) postulated that this scaling represents an inverse energy cascade from storm scales. Similar analyses with observations concentrated in the tropical troposphere have not been found in the literature. Nonetheless, we postulate that a (−5/3)-power law may exist at mesoscales (10 km to 10^3 km) in the tropical troposphere inclusive of the PBL and IBL.

The probability $P(\kappa)d\kappa$ of occurrence of a wave of wave-number in the interval $[\kappa, \kappa+d\kappa]$ above a station obeys classical Boltzmann statistics because the number of stations is a constant and total energy production balances total energy dissipation at steady state:

$$P(\kappa)d\kappa = \frac{1}{\varepsilon_0}\exp\left[-\frac{\varepsilon(\kappa)}{\varepsilon_0}\right]d\varepsilon\,, \tag{6}$$

where ε_0 is an appropriate energy scale. The above probability is normalized assuming that each ε-state corresponds to only one κ-state. This is reasonable on the equatorial β-plane as κ is the zonal wave-number while the trapped meridional modes are uniformly dense in κ-space.

Appendix A shows that within the postulated inertial range (10–10^3 km) sampled twice daily by radiosondes, only Yanai and Rossby waves are relevant. Since PBL wind anomalies are known to be isotropic (Mori, 1986; Ibarra, 1995; Koh and Ng, 2009) and the same is assumed for the IBL, Yanai waves cannot be dominant, leaving only isotropic Rossby waves to dominate the statistics. These waves obey the dispersion relation (cf. Eq. A7):

$$\omega \sim -\frac{\beta}{2\kappa}\,, \tag{7}$$

where κ is negative for westward phase propagation. For a wave disturbance, the velocity perturbation v depends on both the angular frequency ω and the amplitude δ

ACPD
10, 16345–16384, 2010

of the wave:

$$v \sim \omega\delta\,, \tag{8}$$

where ω and δ are independent of each other. Combining Eqs. (5)–(8), the probability $P(v)dv$ that a station records wind speed within $[v,\ v+dv]$ due to isotropic Rossby waves is

$$P(v)dv = \frac{1}{\varepsilon_0}\exp\left[-\left(\frac{v}{c}\right)^{5/3}\right]\frac{d\varepsilon}{dv}dv\,,$$

where the velocity scale c is defined such that

$$(\varepsilon/\varepsilon_0) = (v/c)^{5/3}\,. \tag{9}$$

Hence,

$$P(v)dv = \frac{5}{3c}\left(\frac{v}{c}\right)^{2/3}\exp\left[-\left(\frac{v}{c}\right)^{5/3}\right]dv\,. \tag{10}$$

Outside the inertial range, Eq. (5) and (7) do not apply but Eq. (6) is still valid. Assuming that $\varepsilon(\kappa)$ is continuous and monotonic decreasing for all scales (e.g. as reported by Gage and Nastrom, 1986), the cumulative probability F of a large-scale wave from outside the inertial range occurring above a station can be estimated from Eq. (6):

$$F(\kappa<\kappa_{\min}) = \int_0^{\kappa_{\min}} P(\kappa)d\kappa = \int_{\varepsilon_{\max}}^{\infty}\frac{1}{\varepsilon_0}\exp\left[-\frac{\varepsilon}{\varepsilon_0}\right]d\varepsilon$$

$$= \exp\left[-\frac{\varepsilon_{\max}}{\varepsilon_0}\right] = \exp\left[-\left(\frac{v_{\max}}{c}\right)^{5/3}\right],$$

where $\kappa_{\min}$ is the lower bound of the inertial range, $\varepsilon_{\max}$ is the corresponding highest spectral energy density in the inertial range and $v_{\max}$ is the largest wind speed for which

ACPD
10, 16345–16384, 2010

Eq. (10) holds. The radiosonde measurements in MP show that (v_{max}/c) is about 4 in PBL and IBL (Fig. 3). Hence, $F(\kappa<\kappa_{min})$ is only about 4×10^{-5}, showing a posteriori that the long waves outside the inertial range can be neglected when estimating the PDF of v. For the same reason, the contributions to the PDF of v by the Walker circulation, monsoons and ISO, being mediated by long waves could be ignored. (Note: there is no contradiction with Eq. (4) because the contributions to the *expectation* of $\boldsymbol{v}$ by the short waves largely cancel over time, allowing the less probable but non-cancelling contributions by the long waves to play a larger role.)

The contribution of the diurnal cycle on the PDF is weak probably because twice daily radiosondes only capture the fundamental frequency of the diurnal cycle but not the higher harmonics. Therefore, wind speed in the PBL and IBL should follow Eq. (10) which is the Weibull distribution with shape parameter $k=5/3$. This value of k is consistent with the values within error bars at 925 mb, 850 mb and 100 mb in Fig. 3b.

5.3 Approach to Gaussian statistics

In the free troposphere, as each random variable $\boldsymbol{v}_n$ in Eq. (1) is independent of the others, the Central Limit Theorem implies that the PDF of $\sum \boldsymbol{v}_n$ approaches Gaussian distribution as the total number N of vertical wave modes increases without bound. In the limit of large N, the mean of $\sum \boldsymbol{v}_n$ is the sum of all mean $\boldsymbol{v}_n$ which are zero, and the variance of $\sum \boldsymbol{v}_n$ is the sum of the variance of $\boldsymbol{v}_n$.

The contributions of the long waves to the PDF of v in the free troposphere are ignored for the same reasons as in the PBL and IBL. Therefore, wind velocity $\boldsymbol{v}$ follows the Gaussian distribution,

$$P\left(\boldsymbol{v}=\begin{bmatrix}v_x\\v_y\end{bmatrix}\right)dv_x dv_y \approx \frac{1}{\pi c^2}\exp\left[-\left(\frac{v_x}{c}\right)^2\right]\cdot\exp\left[-\left(\frac{v_y}{c}\right)^2\right]dv_x dv_y\,,$$

where c^2 is the variance of $\boldsymbol{v}$ and is twice the variance of v_x or v_y because the short waves are assumed to be isotropic. Integrating over all directions θ, the PDF for wind

ACPD
10, 16345–16384, 2010

speed v is the Rayleigh distribution:

$$P(v)dv = \int_0^{2\pi} P(\boldsymbol{v})v\,dv\,d\theta \approx \frac{2}{c^2} v \exp\left[-\left(\frac{v}{c}\right)^2\right] dv\,, \tag{11}$$

which is also Weibull distribution with shape parameter k=2. It must be emphasized that the derivation of Eq. (11) does not depend on the statistical dynamics of each contribution $\boldsymbol{v}_n$ except for the assumptions of isotropy and zero mean. Thus, the combined statistical dynamics of equatorial waves of many vertical wave-numbers explains why $k \approx 2$ is observed generally in the upper troposphere in Fig. 3b, without depending on the explanation for why k=5/3 in the PBL and IBL.

Next, further explanation is sought for the departure from Gaussian behavior evidenced by k<2 just above the PBL and the small deviations from k=2 in the upper troposphere.

Let σ_n^2 be the variance of $\boldsymbol{v}_n$ which is also the mean square velocity because mean $\boldsymbol{v}_n$ is zero. We define the following average variances and partial sums of $\boldsymbol{v}_n$ normalized to variance 1:

$$s_{N+1}^2 \overset{\text{def}}{=} \frac{1}{N+1}\sum_{n=1}^{N+1}\sigma_n^2 \quad ; \quad \boldsymbol{u}_{N+1} \overset{\text{def}}{=} \frac{1}{s_{N+1}\sqrt{N+1}}\sum_{n=1}^{N+1}\boldsymbol{v}_n\,, \tag{12}$$

$$s_{N,m}^2 \overset{\text{def}}{=} \frac{1}{N}\sum_{n=1,n\neq m}^{N+1}\sigma_n^2 \quad ; \quad \boldsymbol{u}_{N,m} \overset{\text{def}}{=} \frac{1}{s_{N,m}\sqrt{N}}\sum_{n=1,n\neq m}^{N+1}\boldsymbol{v}_n\,. \tag{13}$$

Theorem 2 of Arstein et al. (2004) is used to show that the average increase in Shannon's entropy of normalized partial sums of $\boldsymbol{v}_n$ is (cf. Appendix B):

$$\overline{\mathrm{Ent}(\boldsymbol{u}_{N+1}) - \mathrm{Ent}(\boldsymbol{u}_{N,m})} \geq \frac{1}{2}\ln\left(\widetilde{s_{N,m}^2}/\overline{s_{N,m}^2}\right) \tag{14}$$

where straight and curly overbars denote, respectively arithmetic and geometric mean over m=1 to N+1. The following statements can be made:

ACPD
10, 16345–16384, 2010

1. If mean square velocity σ_n^2 is invariant of the vertical mode n, $\widetilde{s^2_{N,m}} = \overline{s^2_{N,m}}$ and so Shannon's entropy is expected to increase monotonically with N as (Arstein, 2004)

$$\overline{\mathrm{Ent}(\boldsymbol{u}_{N+1}) - \mathrm{Ent}(\boldsymbol{u}_{N,m})} \geq 0\,. \tag{15}$$

2. If σ_n^2 varies with n, $\widetilde{s^2_{N,m}} < \overline{s^2_{N,m}}$ and $\ln\left(\widetilde{s^2_{N,m}} / \overline{s^2_{N,m}}\right) < 0$. The expected Shannon's entropy may decrease as N increases.

3. In all cases, the Central Limit Theorem requires that in the limit $N \to \infty$, $\boldsymbol{u}_N$ follows the Gaussian distribution which, among any distribution of variance 1, gives the maximum Shannon's entropy.

4. For large but finite N, the maximal reduction in Shannon's entropy can be simply estimated by supposing that $s_{N,m}$ takes one of two values, S_0 and S_1, with equal chance:

$$\frac{1}{2}\ln\left(\widetilde{s^2_{N,m}} / \overline{s^2_{N,m}}\right) \approx \frac{1}{2}\ln\left(\frac{2S_0S_1}{S_0^2 + S_1^2}\right) \leq 0\,. \tag{16}$$

Shannon's entropy associated with Weibull distribution of wind speed normalized to rms value of 1 is a function of the shape parameter k only (Appendix C):

$$\mathrm{Ent}(\boldsymbol{u}) = \ln\left(\frac{2\pi}{k}\right) - \ln\left[\Gamma\left(\frac{2}{k} + 1\right)\right] + \left(1 - \frac{2}{k}\right)\gamma + 1\,, \tag{17}$$

where $u = v/\sigma$ is the wind speed normalized by its rms value in Eq. (2) and γ is the Euler-Mascheroni constant. Using Eq. (17), Shannon's entropy corresponding to the k-values in Fig. 3b was computed and shown in Fig. 4. The theoretical bound to how

ACPD
10, 16345–16384, 2010

much entropy can be reduced from the maximal value assuming large but finite number of vertical wave modes was computed from Eq. (16) by using the estimates:

$$S_0 \approx \sigma(c_i, k_i)$$
$$S_1 \approx \sigma(c_{i-1}, k_{i-1}) \quad \text{for } i > 1 \tag{18}$$

where c_i and k_i are the values of c and k at pressure level i (the index i starts from 1 at 100 mb and increases downwards with pressure) and σ is the function given by Eq. (2). As Eq. (18) is only a rough estimate, the emphasis is on the vertical trend of the entropy reduction and not on the reduced entropy values per se.

With reference to Fig. 4, from 850 mb to 500 mb, Shannon's entropy increases monotonically towards the maximal value, $\ln\pi+1$, as k increases from 5/3 towards 2 (Fig. 3b). This is because there is little variation in the rms wind speed (which can be estimated from Fig. 3a) and the number of vertical wave modes N increases with increasing height. The latter assumption is consistent with diminishing vertical wavenumber of the fundamental mode and hence increasing spectral density of harmonics with increasing height. At 400 mb, the number of wave modes N is sufficiently large that Shannon's entropy comes close to the maximal value. At 300 mb and above, rapidly varying rms wind speed (Fig. 3a) leads to ever larger spread among the variance σ_n^2 of the wave modes. This keeps Shannon's entropy away from the maximal value and the corresponding k-value overshoots 2 (Fig. 3b). Note that Shannon's entropy is lowered the most where the rms wind speed changes the most rapidly from 200 mb to 150 mb.

In the PBL and IBL, Shannon's entropy is considerably smaller than the maximal entropy and is consistent with the dominance of only one wave mode. Near the surface at 1000 mb, the central mountain ridge on the peninsula may have an organizing influence on wind flow to lowering Shannon's entropy, depressing k below 5/3 (Fig. 3b).

ACPD
10, 16345–16384, 2010

6 Application to monitoring data quality

The preceding theory for Weibull distribution of near-equatorial wind speed supports the view that beyond the empirical threshold of validity of the distribution, v_{max}, wind speed data are likely to be dominated by noise and hence are suspect. It follows naturally to apply such thresholds to monitor the quality of the radiosonde data from MP. For demonstration purpose, data at three mandatory levels, 850 mb, 500 mb and 250 mb, were selected. Results showed that slightly more than half a percent of 278,711 available wind speed records at these three levels are suspect.

A common statistical threshold to reject outlying data is the mean plus three standard deviations. For a variable v, this threshold is denoted as v_{m3sd}. In Fig. 3(c), v_{max} is compared with v_{m3sd}, where the mean and standard deviation are computed from the MP data only. At all pressure levels, the threshold v_{max} is larger than the regional v_{m3sd}, implying that more useful regional data is retained in our statistical dynamical approach rather than the common statistical mathematical approach.

The MP data below our wind speed threshold v_{max} is compared with data from Integrated Global Radiosonde Archive (IGRA) (Durre et al., 2006) in Fig. 5. The finding is that a theoretically sound regional data-monitoring strategy can identify erroneously high wind speed that escapes detection in the QC of global datasets. This is possibly because global QC assumes a larger spread of wind values than is valid within a specific region like MP.

7 Extension to Equatorial Monsoon Zone

In this section, the preceding statistical dynamical theory for the Weibull distribution of radiosonde wind and the data monitoring strategy developed from it are tested for their relevance to other tropical regions.

242 stations across the global tropics (including MP) were first divided into three climatic zones according to the time-averaged zonal wind in the upper troposphere

ACPD
10, 16345–16384, 2010

(i.e. mandatory levels from 500 mb to 100 mb inclusive): (a) westerly zone: every level shows westerly mean wind; (b) mixed wind zone: both westerly and easterly mean wind are present; (c) easterly zone: every level shows easterly mean wind (Fig. 6). The existence of the mixed wind zone in the equatorial belt and its significance to cross-equatorial propagation of Rossby waves have been noted before (Webster and Holton, 1982). On a pressure level, each station is a point measurement and would under-sample the underlying dynamics. The k-value at each station would behave like a random variable itself with a probability distribution. Comparison of the vertical profiles of median k-values in Fig. 7 with Fig. 3b shows that the statistical dynamics in the westerly and mixed wind zones are probably different from that over MP, but the statistics in the easterly zone warrant further investigation.

The k-values in West Pacific (diamonds in Fig. 7c) are consistently larger than most other values in the easterly zone and bear closer resemblance to those in the mixed wind zone. The West Pacific stations are also the only ones in the easterly wind zone that do not lie within the monsoon region according to Fig. 1.2 of Ramage (1971). Therefore, the Equatorial Monsoon Zone (EMZ) is defined to encompass the stations in the easterly wind zone excluding the West Pacific stations.

Unlike MP, the EMZ spans half the globe across varying climatology of wind speed (Fig. 8a). So at each pressure level, wind speed v from each station must be normalized by its rms value σ estimated in Eq. (2) before the combined dataset of $u=v/\sigma$ can constitute a statistically homogeneous population that describable by a Weibull distribution of distinct shape and scale parameters (k_{EMZ}, c_{EMZ}). Because the combined dataset is the union of normalized subsets of rms value 1, σ_{EMZ} computed from k_{EMZ} and c_{EMZ} should be 1 for a large dataset if σ at each station correctly captures the climatological wind speed and the Weibull distribution is indeed appropriate for wind speed observed in the EMZ. From Fig. 8b, $\sigma_{\mathrm{EMZ}}\approx 1$ for all levels, verifying the above assumptions are good. The largest difference of σ_{EMZ} from one is only -0.055 and occurs at 1000 mb, possibly due to the complicating influence of local terrain and surface characteristics.

ACPD
10, 16345–16384, 2010

The characteristic profile of k_{EMZ} and the associated threshold for wind speed v_{max}^{EMZ} are shown in Fig. 8c and d, respectively. At each station, v_{max}^{EMZ} was estimated as the product of the local σ and the regional threshold u_{max} derived essentially from k_{EMZ}: the expected number $n_{fit}(u,\delta u)\leq 1$ for $u \geq u_{max}$ using a bin size δu of 2 knots divided by the mean σ at each level. Away from the surface, v_{max}^{EMZ} is larger than the v_{m3sd}^{EMZ} for about 90% or more of the stations, where v_{m3sd}^{EMZ} is the statistical mathematical threshold computed from the union set of all v measurements in EMZ for each level. As a threshold to flag off suspicious outlying radiosonde reports in EMZ, v_{max}^{EMZ} preserves more useful data than v_{m3sd}^{EMZ} because v_{max}^{EMZ} captures the region's statistical dynamics and is adapted to the local wind climatology. At 1000 mb, v_{m3sd}^{EMZ} is not a suitable reference for comparison as local v_{m3sd} should be used instead.

To understand the values of k_{EMZ}, the corresponding Shannon's entropy for u normalized to rms value of 1 is shown in Fig. 9. Unlike for MP, the spread of wind variance σ_n^2 of the wave modes could be estimated from the extensive spatial sampling of σ across EMZ (Fig. 8a). In Eq. (16), the theoretical bound to entropy reduction was computed using the following estimates:

$$S_0 \approx \sigma_{upp}\,,$$
$$S_1 \approx \sigma_{low}\,, \qquad (19)$$

where σ_{upp} and σ_{low} are the upper and lower quartiles of σ over all stations in the EMZ, respectively. The error bars for k in Fig. 8c and hence for entropy in Fig. 9 were estimated by generating another two sets of best-fit k_{EMZ} by separately removing the stations with the top and bottom 5 percentile of k-values (i.e. 2 stations) and computing the standard deviation among the three sets of best-fit k_{EMZ}.

Compared to the MP results in Fig. 4, the EMZ results in Fig. 9 show that the free troposphere (700 mb to 200 mb) is nearer to attaining maximal entropy because spatially distinct wave packets can exist for the same vertical mode, leading to a larger effective total number of modes. (Yanai wave packets have already been noted in the lower

ACPD
10, 16345–16384, 2010

stratosphere by Dunkerton, 1991. So it is plausible that wave packets exist in the troposphere where equatorial waves originate.) For the same reason, Shannon's entropy in the PBL is enhanced so that $k_{EMZ}>5/3$ at 850 mb and 925 mb (Fig. 8c). However, at these two levels just like at 150 mb, the large spread of σ across EMZ keeps Shannon's entropy away from the maximal value. The tropopause is generally observed to be at higher pressure than 100 mb over EMZ, especially above India during the summer monsoon (Seidel et al., 2001). Thus, stratospheric flow may exert some organizing influence to suppress Shannon's entropy at 100 mb leading to $k<5/3$. At 1000 mb, varied terrain and surface characteristics across EMZ may exert a disorganizing influence so that Shannon's entropy is enhanced leading to $k_{EMZ}>5/3$.

8 Summary

Empirical Weibull distributions of wind speed were derived by Maximum Likelihood Estimate for radiosonde data spanning more than 30 years from 7 stations in the Malay Peninsula (MP) and from 43 stations in the Equatorial Monsoon Zone (EMZ). The Weibull distribution is governed by two parameters: the shape parameter k is the key quantity explained in this paper; the rms wind speed σ (which determines the scale parameter c for a given k) is the result of planetary-scale climate dynamics. Wind in EMZ was non-dimensionalized by the local σ to remove the effect of geographic variation of climatology before empirical fitting of the Weibull distribution.

A statistical theory of equatorial waves was proposed to explain k based on the following key concepts.

1. Equatorial atmospheric dynamics is well represented by the equivalent shallow water model on the equatorial β-plane.
2. Equatorial waves exist in distinct vertical modes and horizontal wave-packets.
3. In the planetary boundary layer and the internal boundary layer at the tropopause,

ACPD
10, 16345–16384, 2010

- quasi-2-D turbulence produces an inertial range in the energy density spectrum with a characteristic (−5/3)-power law;
- Boltzmann statistics apply because the number of stations and total energy are constants;
- contribution by isotropic Rossby waves dominate the PDF of wind speed measured by radiosondes;
- thus, wind speed follows a Weibull distribution with k=5/3 for a small region like MP;
- for a large region like EMZ, distinction between wave-packets leads to $5/3<k<2$.

4. Within the tropospheric interior,
   - the simultaneous presence of many vertical modes causes Shannon's entropy to increase and the value of k to approach 2 compared to the PBL;
   - the spread of wind variance among wave modes at upper levels prevents Shannon's entropy from attaining the maximal value and causes k either to overshoot 2 (MP) or to remain below 2 (EMZ).
5. External organizing or disorganizing influences may suppress or elevate Shannon's entropy and cause k to deviate further or less from 2, respectively.

Best-fit Weibull distribution can be used to derive confidence thresholds for monitoring radiosonde wind speeds. The thresholds are generally larger than those obtained by taking the mean plus three standard deviations. More data is retained and data quality is improved because these thresholds are based on an understanding of the statistical dynamics of equatorial waves and they are adapted to the local climatology. Such an improved dataset from EMZ would ultimately benefit research and forecast.

ACPD
10, 16345–16384, 2010

Appendix A

Equatorial waves dynamics

The radiosonde samples are only twice daily and so gravity waves which have much higher frequencies are aliased into the lower frequencies in the dataset and cannot be properly investigated here. The following dispersion relations for the other equatorial waves are found in Gill (1982) (using different notation):

$$\omega_0 \stackrel{\text{def}}{=} \left(\beta\sqrt{gH}\right)^{1/2} \quad ; \quad \kappa_0 \stackrel{\text{def}}{=} \left(\frac{\beta}{\sqrt{gH}}\right)^{1/2},$$

$$\text{Rossby waves}: \frac{\omega}{\omega_0} = \frac{-\kappa/\kappa_0}{(\kappa/\kappa_0)^2 + 2m + 1}, \tag{A1}$$

$$\text{Yanai waves}: \frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} = \frac{\kappa}{\kappa_0}, \tag{A2}$$

$$\text{Kelvin waves}: \frac{\omega}{\omega_0} = \frac{\kappa}{\kappa_0}, \tag{A3}$$

where ω is the angular frequency and κ is the zonal wave-number (negative for westward phase propagation). β is the meridional gradient of planetary vorticity. m is a positive integer denoting the meridional mode of the Rossby wave of which the meridional velocity is proportional to the function:

$$Y(y) = \exp(-\kappa_0^2 y^2/2) H_m(\kappa_0 y), \tag{A4}$$

where $H_m(x)$ denotes Hermite polynomial of degree m. From Eq. (A1), the maximum frequency $\omega_{\max}$ for a Rossby wave of mode m and its corresponding wave-number κ_R are

$$\omega_{\max}(m) = \frac{\omega_0}{2\sqrt{2m+1}}, \tag{A5}$$

ACPD
10, 16345–16384, 2010

$$\kappa_R(m) = -\kappa_0\sqrt{2m+1}\,. \tag{A6}$$

The dispersion relations for the above equatorial waves are plotted in Fig. A1 for equivalent depths H from 20 m to 200 m. Note that κ_0 and ω_0 are not sensitive to the exact value of H chosen because they depend on the fourth root of H. (A ten-fold increase in H results in ω_0 increasing and κ_0 decreasing by less than 1.8 times.) With reference to Fig. A1, for period longer than 1 day and for wavelength in the postulated inertial range 10 to 10^3 km, only Yanai waves and Rossby waves are relevant.

There is much observational evidence that PBL winds are isotropic (Mori, 1986; Ibarra, 1995; Koh and Ng, 2009). For Yanai waves, Eq. (A2) shows that the meridional extent y_0 is $2\kappa_0^{-1}\sqrt{2\ln 10}$ such that $Y(\pm y_0/2)=0.1Y(0)$. This value is about 3400 km to 6000 km for H=20 m to 200 m, which is much longer than the zonal half-wavelengths within the inertial range (i.e. 5 km to 500 km). Such anisotropy makes it unlikely for Yanai waves to be important here.

As for Rossby waves of mode m, the meridional "wavelength" l_m corresponds to twice the distance between adjacent roots of Hermite polynomials (cf. Eq. A4). The asymptotic analysis of Calogero and Perelomov (1978) implies that near the equator for large m,

$$l_n \approx 2\pi(2m)^{-1/2}\kappa_0^{-1}\,.$$

The requirement of isotropy implies

$$\kappa = -2\pi/l_n \approx -\kappa_0\sqrt{2m} \approx \kappa_R\,,$$

where we have used Eq. (A6) for large m. Thus, isotropic Rossby waves have the highest frequency for a fixed meridional mode m for large m. (m<10 modes are not isotropic within the inertial range, see Fig. A1.) Combining Eqs. (A5) and (A6), the dispersion relation for isotropic Rossby waves is

$$\omega_{\max}(m) = -\frac{\beta}{2\kappa_R(m)} \tag{A7}$$

ACPD
10, 16345–16384, 2010

Appendix B

Proof of entropy increment

Theorem 2 of Arstein et al. (2004) states that the approach of $\sum \boldsymbol{v}_n$ to Gaussian statistics with increasing number of independent square-integrable random variables $\boldsymbol{v}_n$ obeys the inequality:

$$\mathrm{Ent}\left(\frac{1}{\sqrt{N+1}}\sum_{n=1}^{N+1}\boldsymbol{v}_n\right) \geq \frac{1}{N+1}\sum_{m=1}^{N+1}\mathrm{Ent}\left(\frac{1}{\sqrt{N}}\sum_{n=1,n\neq m}^{N+1}\boldsymbol{v}_n\right), \tag{B1}$$

where $\mathrm{Ent}(\boldsymbol{v})$ is Shannon's entropy which for a random two-dimensional vector variable $\boldsymbol{v}$ is

$$\mathrm{Ent}(\boldsymbol{v}) \stackrel{\mathrm{def}}{=} -\iint_{\text{all } \boldsymbol{v}} P(\boldsymbol{v})\ln P(\boldsymbol{v})\,d^2\boldsymbol{v}\,. \tag{B2}$$

Using definitions (12) and (13) and the identity $\mathrm{Ent}(s\boldsymbol{u})\equiv\ln s+\mathrm{Ent}(\boldsymbol{u})$, Eq. (B1) becomes

$$\ln s_{N+1} + \mathrm{Ent}(\boldsymbol{u}_{N+1}) \geq \overline{\ln s_{N,m}} + \overline{\mathrm{Ent}(\boldsymbol{u}_{N,m})},$$

where the straight overbar denotes arithmetic mean from m=1 to N+1. Further using the identities

$$s^2_{N+1} \equiv \overline{s^2_{N,m}}$$
$$\overline{\ln s^2_{N,m}} \equiv \ln \widetilde{s^2_{N,m}}$$

where the curly overbar denotes geometric mean from m=1 to N+1, the average increase in entropy is

$$\overline{\mathrm{Ent}(\boldsymbol{u}_{N+1}) - \mathrm{Ent}(\boldsymbol{u}_{N,m})} \geq \frac{1}{2}\ln\left(\widetilde{s^2_{N,m}}/\overline{s^2_{N,m}}\right). \tag{B3}$$

ACPD
10, 16345–16384, 2010

Appendix C

Shannon's entropy associated with Weibull distribution

First, wind speed v is normalized by its rms value σ to yield a non-dimensional variable $u=v/\sigma$ that follows Weibull distribution of variance 1:

$$
\begin{aligned}
P(u)du &= P(v)dv \\
&= k\sqrt{\Gamma\left(\frac{2}{k}+1\right)}\left[u\sqrt{\Gamma\left(\frac{2}{k}+1\right)}\right]^{k-1}\exp\left\{-\left[u\sqrt{\Gamma\left(\frac{2}{k}+1\right)}\right]^{k}\right\}du \\
&= -d[\xi(u)]\,.
\end{aligned}
$$

where Eq. (1) and (2) were used and $\xi(u)=\exp\left\{-\left[u\sqrt{\Gamma(2/k+1)}\right]^{k}\right\}$. From Eq. (B2), Shannon's entropy for u is

$$
\begin{aligned}
\mathrm{Ent}(u) &= -\int_0^\infty\left[\frac{P(u)}{2\pi u}\ln\frac{P(u)}{2\pi u}\right]u\,du\int_0^{2\pi}d\theta \\
&= -\int_0^\infty\ln\left[\frac{P(u)}{2\pi u}\right]P(u)du \\
&= -\int_0^1\ln\left[\frac{k}{2\pi}\Gamma\left(\frac{2}{k}+1\right)(-\ln\xi)^{(k-2)/k}\xi\right]d\xi \\
&= -\ln\left[\frac{k}{2\pi}\Gamma\left(\frac{2}{k}+1\right)\right]-\left(1-\frac{2}{k}\right)\int_0^1\ln(-\ln\xi)d\xi-\int_0^1\ln\xi\,d\xi \\
&= \ln\left(\frac{2\pi}{k}\right)-\ln\left[\Gamma\left(\frac{2}{k}+1\right)\right]+\left(1-\frac{2}{k}\right)\gamma+1\,.
\end{aligned}
$$

ACPD
10, 16345–16384, 2010

In the last step above, we made use of

$$\int_0^1 \ln(-\ln\xi)d\xi \equiv -\gamma\,,$$

$$\int_0^1 \ln\xi d\xi = [\xi\ln\xi]_0^1 - \int_0^1 d\xi = -1\,,$$

where the first integral is one expression for Euler-Mascheroni constant γ (Whittaker and Watson, 1996).

Acknowledgements. The authors would like to thank the Department of Atmospheric Science, University of Wyoming and National Oceanic and Atmospheric Administration, USA, for access to their radiosonde data archive. Some of the work in Sect. 7 was carried out by Evan Cheok, Victoria Junior College, Singapore, under the guidance of the first two authors.

References

Artstein, S., Ball, K. M., Barthe, F., and Naor, A.: Solution of Shannon's problem on the monotonicity of entropy, J. Am. Math. Soc., 17(4), 975–982, 2004.

Calogero, F. and Perelomov, A. M.: Asymptotic density of the zeros of hermite polynomials of diverging order, and related properties of certain singular integral operators, Lettere Al Nuovo Cimento, 23(18), 650–652, 1978.

Chang, C. P., Liu, C. H., and Kuo, H. C.: Typhoon Vamei: an equatorial tropical cyclone formation, Geophys. Res. Lett., 30(3), 1150, doi:10.1029/2002GL016365, 2003.

Chang, C. P., Wang, Z., McBride, J., and Liu, C. H.: Annual cycle of Southeast Asia – maritime continent rainfall and the asymmetric monsoon transition, J. Climate, 18, 287–301, 2005.

Divakarla, M. G., Barnet, C. D., Goldberg, M. D., McMillin, L. M., Maddy, E., Wolf, W., Zhou, L., and Liu, X.: Validation of atmospheric infrared sounder temperature and water vapor retrievals with matched radiosonde measurements and forecasts, J. Geophys. Res., 111, D09S15, doi:10.1029/2005JD006116, 2006.

Dunkerton, T.: Intensity variation and coherence of 3–6 day equatorial waves, Geophys. Res. Lett., 18(8), 1469–1472, 1991.

ACPD

10, 16345–16384, 2010

Durre, I., Vose, R. S., and Wuertz, D. B.: Overview of the integrated global radiosonde archive, J. Climate, 19, 53–68, 2006.
Frank, H. P. and Landberg, L.: Modelling the wind climate of Ireland, Boundary-Lay. Meteorol., 85, 359–378, 1997.
Gage, K. S. and Nastrom, G. D.: Theoretical interpretation of atmospheric wave-number spectra of wind and temperature observed by commercial aircraft during GASP, J. Atmos. Sci., 43(7), 729–740, 1986.
Gandin, L. S.: Complex quality control of meteorological observations, Mon. Weather Rev., 116, 1137–1156, 1988.
Gill, A. E.: Atmosphere-Ocean Dynamics, 4th edition, Academic Press, London, 662 pp., 1982.
Gill, A. E.: Some simple solutions for heat-induced tropical circulation, Quart. J. Roy. Meteor. Soc., 106, 447–462, 1980.
Hadi, T. W., Horinouchi, T., Tsuda, T., Hashiguchi, H., and Fukao, S.: Sea-breeze circulation over Jakarta, Indonesia: a climatology based on boundary layer radar observations, Mon. Weather Rev., 130, 2153–2166, 2002.
Hendon, H. H.: Indonesian rainfall variability: impacts of ENSO and local air-sea interaction, J. Climate, 16, 1795–1790, 2003.
Ibarra, J. I.: A new approach for the determination of horizontal wind direction fluctuations, J. Appl. Met. Clim., 34, 1942–1949, 1995.
Joseph, B., Bhatt, B. C., Koh, T. Y., and Chen, S.: Sea breeze simulation over Malay Peninsula over an intermonsoon period, J. Geophys. Res., 113, D20122, doi:10.1029/2008JD010319, 2008.
Juneng, L. and Tangang, F. T.: Evolution of ENSO-related rainfall anomalies in Southeast Asia region and its relationship with atmosphere-ocean variations in Indo-Pacific sector, Clim. Dynam., 25, 337–350, 2005.
Justus, C. G., Hargraves, W. R., Mikhail, A., and Graber, D.: Methods for estimating wind speed frequency distributions, J. Appl. Meteor., 17, 350–353, 1978.
Koh, T. Y. and Ng, J. S.: Improved diagnostics for NWP verification in the tropics, J. Geophys. Res., 114, D12102, doi:10.1029/2008JD011179, 2009.
Koh, T. Y. and Teo, C. K.: Towards a mesoscale observation network in Southeast Asia, Bull. Amer. Meteor. Soc., 90(4), 481–488, doi:10.1175/2008BAMS2561.1, 2009.
Kolmogorov, A. N.: The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers, C. R. Acad. Sci. USSR, 30, 301–305, 1941.

ACPD
10, 16345–16384, 2010

Kolmogorov, A. N.: A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number, J. Fluid Mech. 13, 82–85, 1962.
Labraga, J. C.: Extreme winds in the Pampa del Castillo Plateau, Patagonia, Argentina, with reference to wind farm settlement, J. Appl. Meteor., 33, 85–95, 1994.
Larsen, M. F., Kelley, M. C., and Gage, K. S.: Turbulence spectra in the upper troposphere and lower stratosphere between 2 hours and 40 days, J. Atmos. Sci., 39, 1035–1041, 1982.
Lau, K. M. and Yang, S.: Climatology and interannual variability of the Southeast Asian summer monsoon, Adv. Atmos. Sci., 14, 141–162, 1997.
Lun, I. Y. F. and Lam, J. C.: A study of Weibull parameters using long-term wind observations. Renew. Energ., 20, 145–153, 2000.
Madden, R. A. and Julian, P. R.: Description of a 40–50 day oscillation in the zonal wind in the tropical Pacific, J. Atmos. Sci., 28, 702–708, 1971.
Madden, R. A. and Julian, P. R.: Observations of the 40–50-day tropical oscillation – a review, Mon. Weather Rev., 122, 814–837, 1994.
Manwell, J. F., McGowan, J. G., and Rogers, A. L.: Wind characteristics and resources, in: Wind Energy Explained, Theory, Design and Application, Wiley, New York, 21–80, 2002.
Matsuno, T.: Quasi-geostrophic motions in the equatorial area, J. Meteor. Soc. of Japan, 44, 25–42, 1966.
Mori, Y.: Evaluation of several “single-pass” estimators of the mean and the standard deviation of wind direction, J. Clim. Appl. Met., 25, 1387–1397, 1986.
Nastrom, G. D. and Gage, K. S.: A climatology of atmospheric wavenumber spectra observed by commercial aircraft, J. Atmos. Sci., 42, 950–960, 1985.
Neale, R. B. and J. M. Slingo: The maritime continent and its role in the global climate: a GCM study, J. Climate, 16, 834–848, 2003.
Ramage, C. S.: Role of a tropical maritime continent in the atmospheric circulation, Mon. Weather Rev., 96, 365–370, 1968.
Ramage, C. S.: Monsoon Meteorology, Academic Press, New York, 296 pp., 1971.
Roney, J. A.: Statistical wind analysis for near-space applications, J. Atmos. Sol.-Terr. Phy., 69, 1485–1501, 2007.
Rotunno, R. A.: On the linear theory of the land and sea breeze, J. Atmos. Sci., 41, 1999–2009,1983.
Seidel, D. J., Ross, R. J., and Angell, J. K.: Climatological characteristics of the trop-

ACPD
10, 16345–16384, 2010

ical tropopause as revealed by radiosondes, J. Geophys. Res., 106(D8), 7857–7878, doi:10.1029/2000JD900837, 2001.
Stoffelen, A., Pailleux, J., Källén, E., Vaughan, J. M., Isaksen, L., Flamant, P., Wergen, W., Andersson, E., Schyberg, H., Culoma, A., Meynart, R., Endemann, M., and Ingmann, P.: The atmospheric dynamics mission for global wind measurement, Bull. Amer. Meteorol. Soc., 86, 73–87, 2005.
Takle, E. S. and Brown J. M.: Note on the use of Weibull statistics to characterize wind-speed data, J. Appl. Meteor., 17, 556–559, 1978.
Waliser, D. E.: Intraseasonal variability, in: Asian Monsoon, Springer-Praxis, Heidelberg, Germany, 203–257, 2006.
Wang, B.: Theory, in: Intraseasonal Variability in the Atmosphere-Ocean Climate System, Springer-Verlag, Heidelberg, Germany, 307–360, 2005.
Webster, P. J. and Holton, J. R.: Wave propagation through a zonally varying basic flow: the influences of mid-latitude forcing in the equatorial regions, J. Atmos. Sci., 39, 722–733, 1982.
Weichman, P. B. and Petrich, D. M.: Statistical equilibrium solutions of the shallow water equations, Phys. Rev. Lett., 86(9), 1761–1764, 2001.
Wheeler, M. and Kiladis, G. N.: Convectively coupled equatorial waves: analysis of clouds and temperature in the wavenumber-frequency domain, J. Atmos. Sci., 56, 374–399, 1999.
Whittaker, E. T. and Watson, G. N.: A Course in Modern Analysis, 4th edition, Cambridge University Press, England, 680 pp., 1996.
Widger Jr. W. K.: Estimations of wind speed frequency distributions using only the monthy average and fastest mile data, J. Appl. Meteor., 16, 244–247, 1977.
Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press, London, 467 pp., 1995.
Zhu, B. and Wang, B.: The 30–60 day convection see-saw between the tropical Indian and western Pacific Oceans, J. Atmos. Sci., 50, 184–19, 1993.

ACPD
10, 16345–16384, 2010

Table 1. Table 1 shows the period for which radiosonde wind data were available at 00:00 UTC and at 12:00 UTC (with the precise start and end dates in parentheses). (Note that data from Kuantan came in two periods.) Within each period, the proportion of days for which data were actually available lies within the shown percentage range for all pressure levels used in this study, except 925 mb which was instituted as a mandatory level only in the 1990s and so the lower percentage at this level is shown in parentheses. The total period spanned and the number of station reports per level are given in the last line. Data periods less than twenty years or data availability less than 50% (excluding 925 mb) are highlighted in bold.

	00:00 UTC		12:00 UTC	
Station	Date range	Percentage of data available	Date range	Percentage of data available
Phuket	**20 Sep 1988–31 Dec 2007**	51–64% (43%)	**31 Jul 1990–04 Oct 1994**	**24–46%** (4%)
Penang	1 Jan 1973–31 Dec 2007	77–83% (38%)	1 Jan 1973–31 Dec 2007	74–80% (36%)
Sepang	**16 Jul 1999–31 Dec 2007**	90–93% (93%)	**17 Jul 1999–31 Dec 2007**	86–92% (92%)
Changi	24 Aug 1980–31 Dec 2007	84–88% (54%)	19 Jul 1983–31 Dec 2007	51–55% (28%)
Kuantan	2 Jan 1973–9 Jan 2000 5 Feb 2005–30 Dec 2007	65–77% (26%)	25 Oct 1973–30 Nov 2000 4 Feb 2005–30 Dec 2007	59–71% (24%)
Kota Bharu	1 Jan 1973–30 Dec 2007	60–80% (36%)	1 Aug 1975–30 Dec 2007)	52–75% (36%)
Songkhla	1 Jan 1973–31 Dec 2007	67–80% (21%)	1 Jan 1973–24 Jun 1998	**37–49%** (1%)
	Date range	Number of station reports	Date range	Number of station reports
All stations	1 Jan 1973–31 Dec 2007	48 520–55 719 (26 369)	1 Jan 1973–31 Dec 2007	33 153–39 944 (16 960)

ACPD
10, 16345–16384, 2010

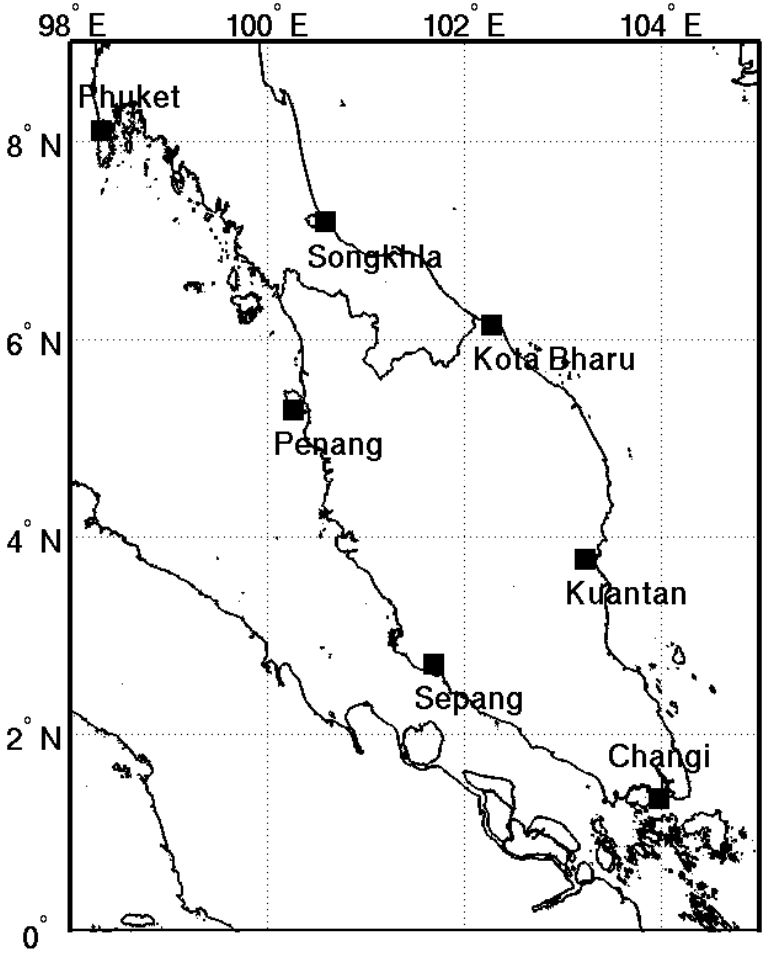

Fig. 1. Location of radiosonde stations on the Malay Peninsula (MP) used in this study.

ACPD
10, 16345–16384, 2010

Statistical dynamics of equatorial waves in tropical radiosonde wind data

T.-Y. Koh et al.

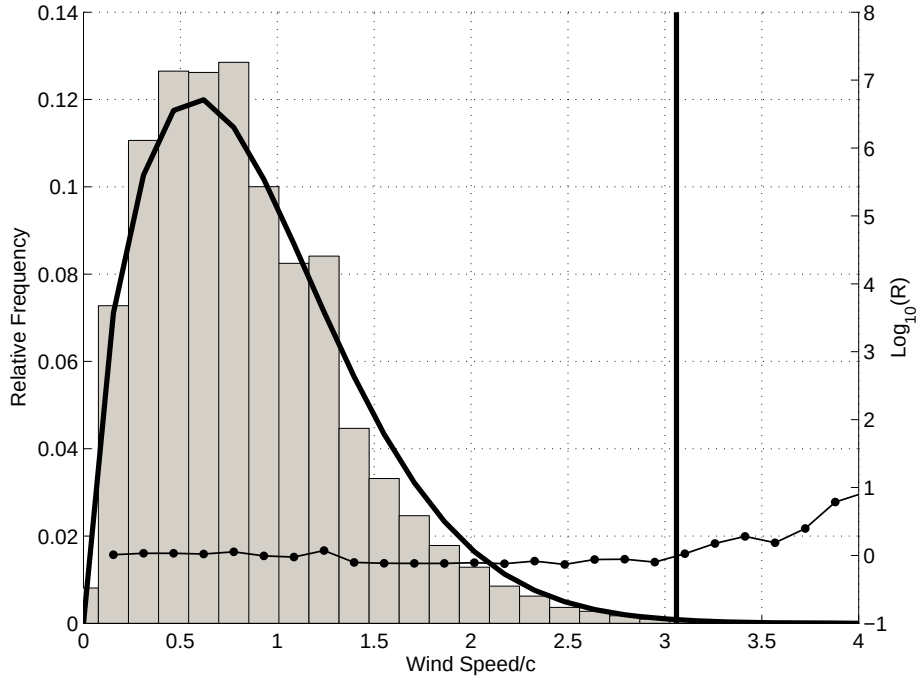

Fig. 2. Frequency histogram of the scaled wind speed at 850 mb for all 7 stations on MP from 1973 to 2007 using a bin size of 2 knots. The frequency is normalized over the total number of measurements. The thick curve is the empirically fitted Weibull distribution with shape parameter k=1.67 and scale parameter c=12.9 knots (or 6.62 m s^{-1}). For clarity, the ratio R of the observed to the fitted frequency is included as the thin line with dots using the logarithmic axis on the right. The bold vertical line denotes the threshold v_{max}=39.4 knots, beyond which wind speed data is flagged as erroneous.

ACPD
10, 16345–16384, 2010

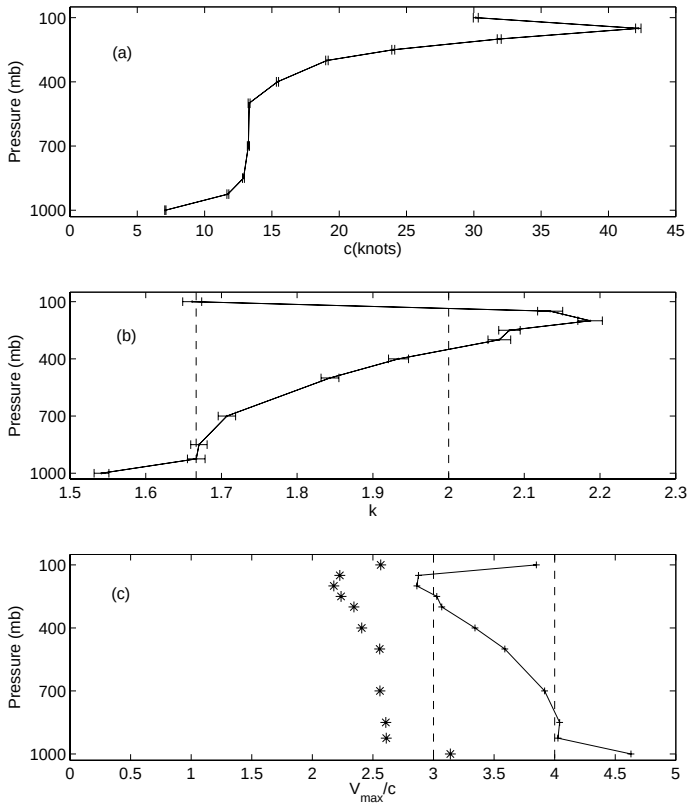

Fig. 3. Plots of empirically fitted attributes of the Weibull distribution for wind speed at different pressure levels in MP: **(a)** scale parameter c; **(b)** shape parameter k, **(c)** scaled threshold $v_{\max}/c$ for wind speed. Error bars for c and k are estimated by MLE at 95% confidence level. Vertical dashed lines correspond to k=5/3, 2 in (b) and $v_{\max}/c$=3, 4 in (c). Asterisks in (c) denote the threshold v_{m3sd} (mean plus three standard deviations) at each pressure level.

ACPD
10, 16345–16384, 2010

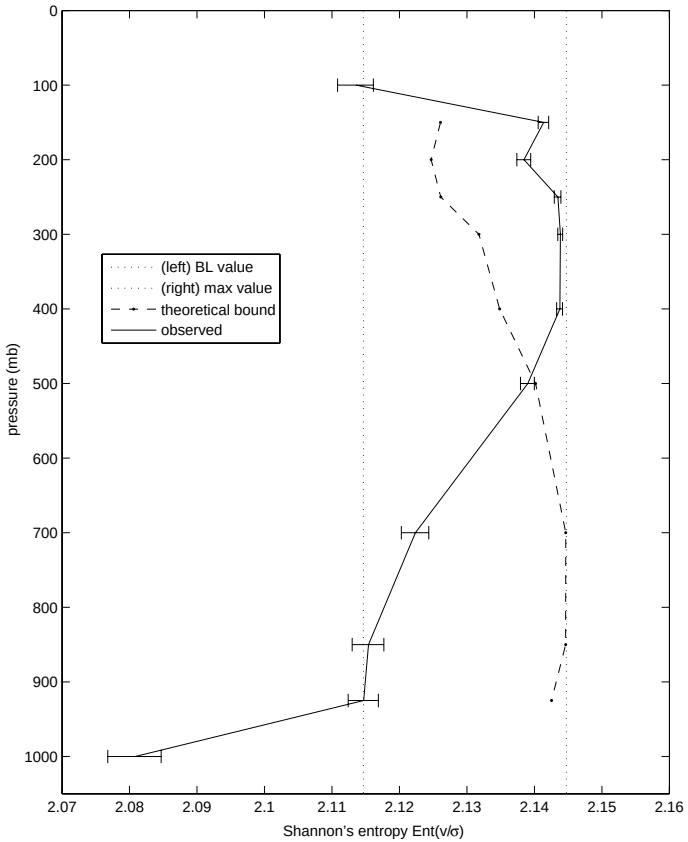

Fig. 4. This figure shows Shannon's entropy for the Weibull's distributions fitted to the radiosonde wind speed data from MP (solid line). The wind speed was first normalized by its rms value. The maximal entropy and the value associated with $k=5/3$ (dotted lines) are shown, as well as the theoretical bound (dashed line) for reduction from maximal entropy.

ACPD
10, 16345–16384, 2010

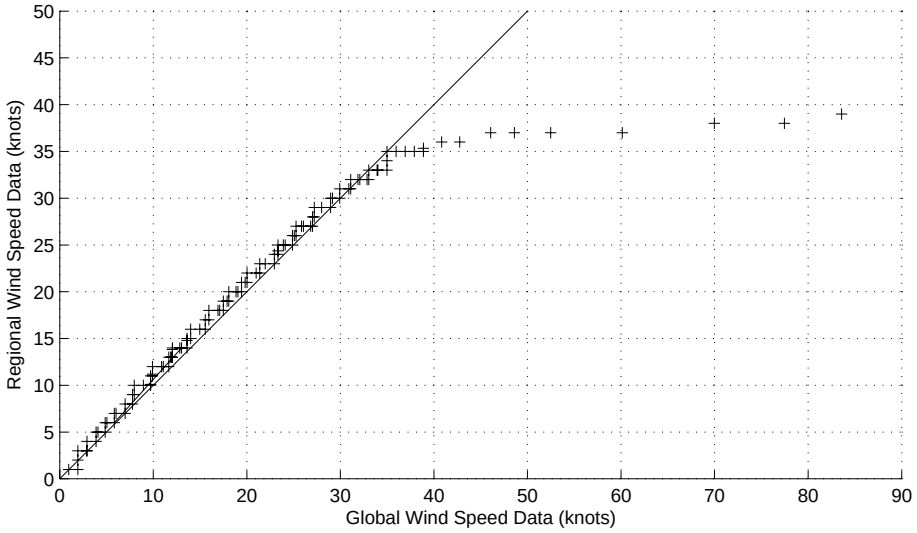

Fig. 5. The 850 mb-wind speed over MP from Wyoming data archive after screening with the threshold v_{max} plotted against the wind speed from IGRA database for the same stations and period. Each ordered pair denoted by a cross refers to the same 0.01%-quantile in both datasets. The straight line shows where the crosses should lie if both datasets had the same distribution.

ACPD
10, 16345–16384, 2010

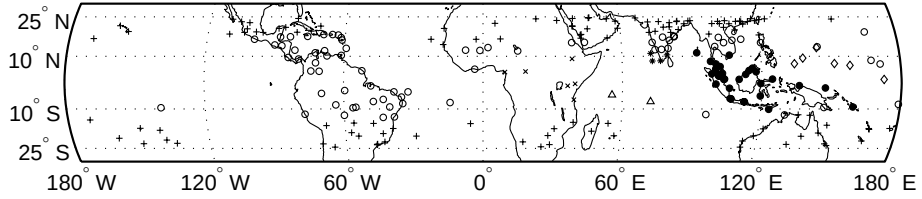

Fig. 6. All 242 tropical radiosonde stations used in the latter part of the study (including the seven stations in MP): "+" signs denote stations in the upper-level (500 mb to 100 mb) westerly zone; circles denote stations in the upper-level mixed wind zone; all other symbols denote stations in the upper-level easterly zone. Within the easterly zone, stations are denoted by their geographical regions (number of stations shown in brackets): "x" sign=Africa (6); asterisk=South Asia (4); dot=Southeast Asia (31); triangle =Indian Ocean (2); diamond= West Pacific (6).

ACPD
10, 16345–16384, 2010

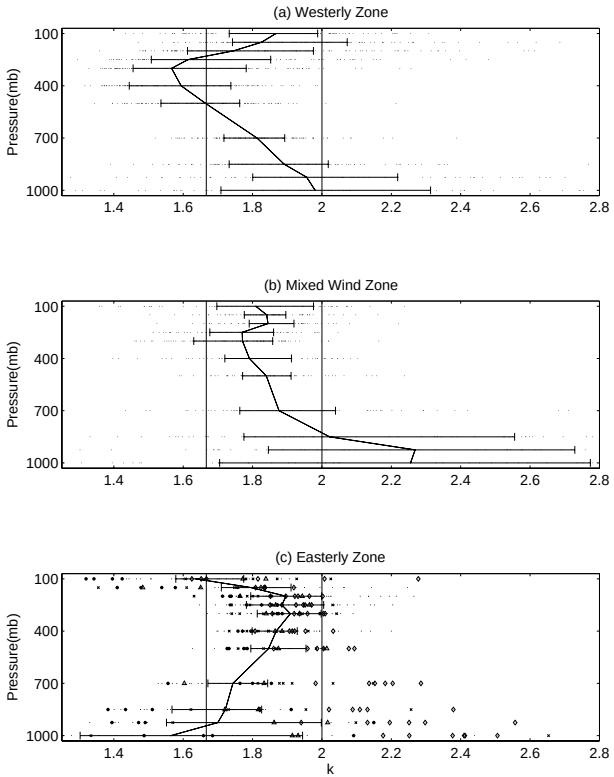

Fig. 7. Scatter plot showing the values of k across vertical levels at tropical stations in the three upper-level climatic wind zones. For each zone: median values at each level are connected to show a vertical profile; the delimited horizontal bars denote the inter-quartile range at each level. For the easterly zone, k-values of stations in Africa ("x" sign), South Asia (asterisk), Southeast Asia (dot), Indian Ocean (triangle) and West Pacific (diamond) are marked with the same symbols as in Fig. 6.

ACPD
10, 16345–16384, 2010

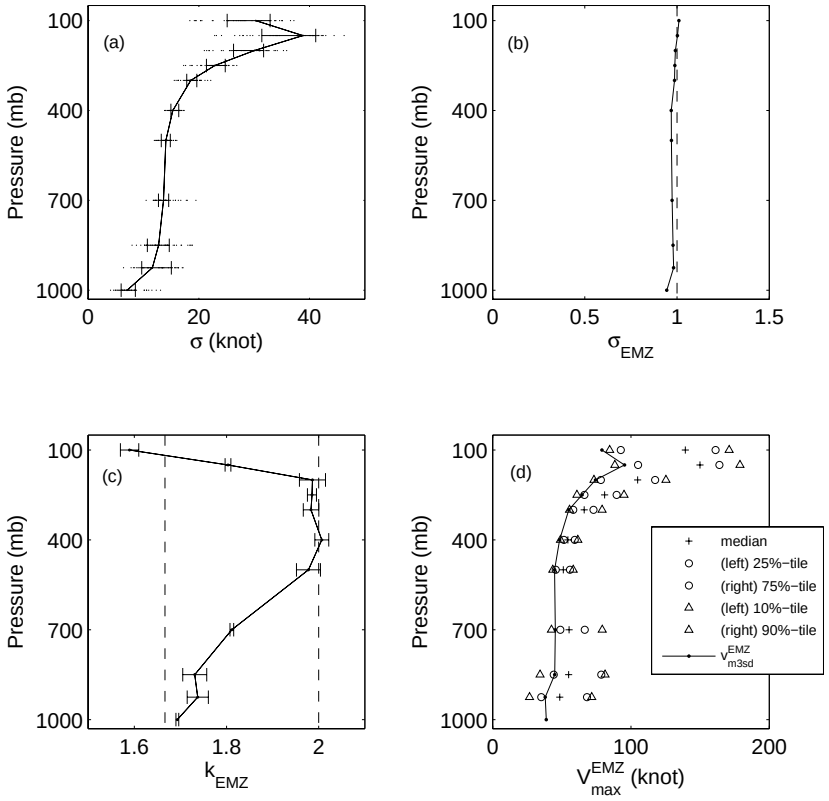

Fig. 8. **(a)** Scatter plot of rms wind speed σ computed using Eq. (2) at all 43 stations in the Equatorial Monsoon Zone (EMZ). The vertical profile connects the median values and delimited horizontal bars denote the inter-quartile ranges. **(b)** The rms value σ_{EMZ} of non-dimensional wind speed $u=v/\sigma$ computed from the fitted Weibull distribution of u in EMZ. **(c)** The shape parameter k_{EMZ} describing the fitted Weibull distribution of u in EMZ with error bars shown. **(d)** The spread of threshold wind speed (symbols) among the stations in EMZ compared to (line).

ACPD
10, 16345–16384, 2010

Statistical dynamics of equatorial waves in tropical radiosonde wind data

T.-Y. Koh et al.

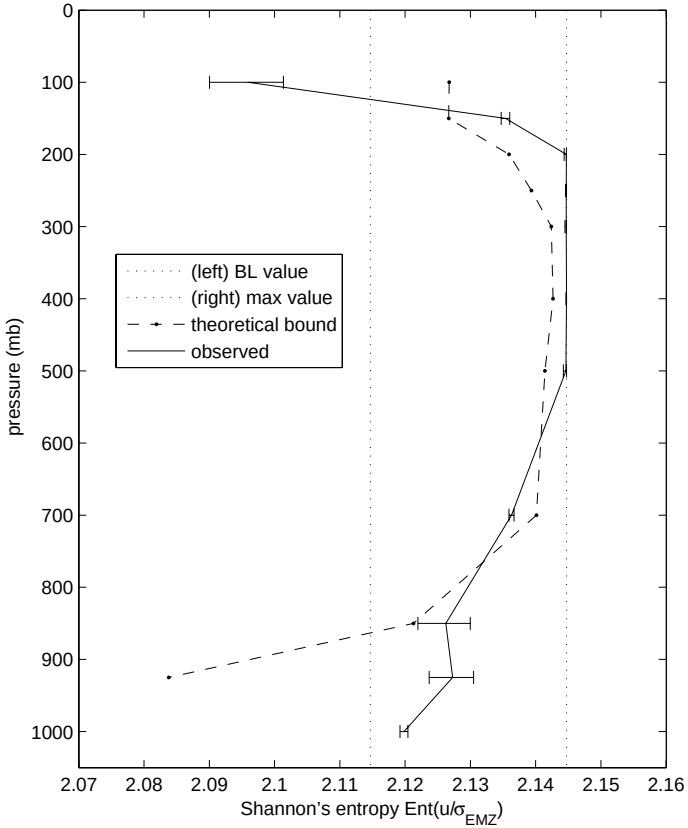

Fig. 9. This Figure shows Shannon's entropy for the Weibull's distributions fitted to the non-dimensional wind speed $u=v/\sigma$ from Equatorial Monsoon Zone (solid line). u was first normalized by its rms value σ_{EMZ} as the latter is close to but not exactly 1. The maximal entropy and the value associated with $k=5/3$ (dotted lines) are shown, as well as the theoretical bound (dashed line) for reduction from maximal entropy.

ACPD
10, 16345–16384, 2010

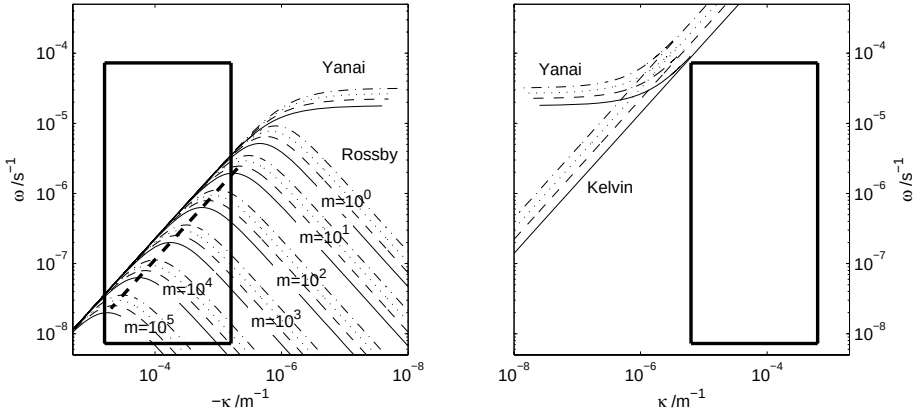

Fig. A1. Angular frequency ω versus zonal wave-number κ plotted on log-log scales for Rossby, Yanai and Kelvin waves for equivalent depths of 100 m (solid), 200 m (dashed), 500 m (dotted) and 1000 m (dash-dotted). The slanted bold dashed line shows the dispersion relation for isotropic Rossby waves for meridional mode number $m>10$. The left and right panels are for negative and positive wave-numbers, respectively. The boxes denote the range of frequency and wave-number corresponding to period of 1 to 10^4 days and wavelength of 10 to 10^3 km.

ACPD
10, 16345–16384, 2010