Atmos. Chem. Phys., 11, 4177–4189, 2011
www.atmos-chem-phys.net/11/4177/2011/
doi:10.5194/acp-11-4177-2011

Atmospheric Chemistry and Physics

Statistical dynamics of tropical wind in radiosonde 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 – Published in Atmos. Chem. Phys. Discuss.: 1 July 2010
Revised: 21 March 2011 – Accepted: 28 April 2011 – Published: 6 May 2011

Abstract. Weibull distributions were fitted to wind speed data from radiosonde stations in the global tropics. A statistical theory of independent wind contributions was proposed to partially explain the shape parameter k obtained over Malay Peninsula and the wider Equatorial Monsoon Zone. This statistical dynamical underpinning provides some justification 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. New scientific questions are raised in the pursuit of statistical dynamical understanding of meteorological variables in the tropics.

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 statistical mechanics). 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 influence 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 data values that are too different from first guesses may be rooted out before assimilation. However, in the tropics, the quality of first-guess fields may sometimes be suspect because model performance is known to be poorer and less data is assimilated prior to making the first guess. Therefore, a QC methodology dependent only on the collected data itself and underpinned by statistical dynamical understanding may be useful, at least as an independent check of data quality before data assimilation and their associated QC checks.

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 general circulation models (GCM) 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 Niño impacts: Hendon (2003), Juneng and Tangang (2005); on Southeast Asian monsoon: Lau and Yang

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

(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.

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.

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 given 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 units like $\mathrm{m\,s^{-1}}$. A bin size of $\delta v = 2$ knots was used for the frequency histograms (e.g. Fig. 2) throughout this work because this was the finest resolution practically achievable. A large difference in the data frequency was noted between odd- and even-valued v in knots, which may indicate that some stations actually measure in integral number of $\mathrm{m\,s^{-1}}$ but record in knots after applying the conversion $1\,\mathrm{m\,s^{-1}} \approx 1.944$ knots.

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) 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

Table 1. Table 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 the column Date Range. (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 italics.

	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 1990–4 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)

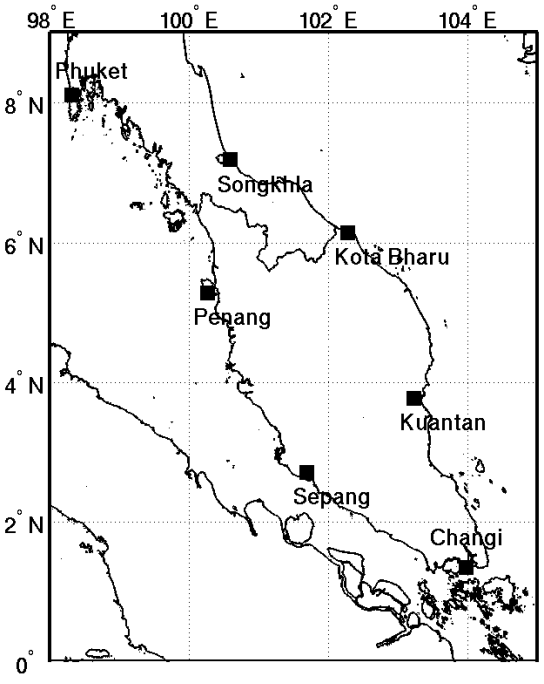

Fig. 1. Location of radiosonde stations on the Malay Peninsula (MP) used in this study.

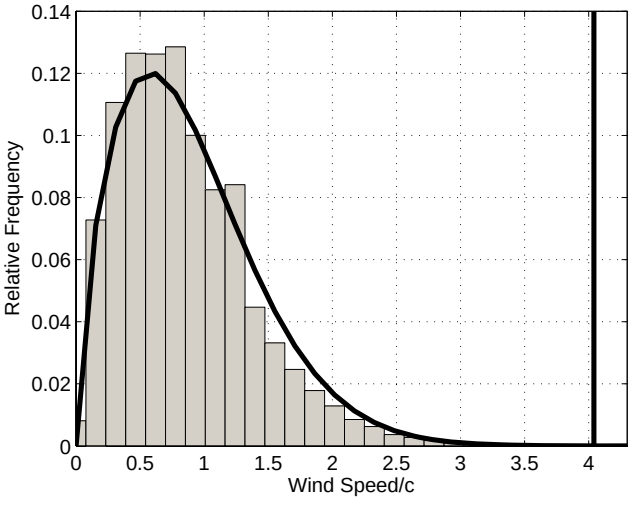

Fig. 2. Frequency histogram of the scaled wind speed at 850 mb for all 7 stations on MP from 1973 to 2007 using a bin size of 2 knots. The frequency is normalized over the total number of measurements. The thick curve is the empirically fitted Weibull distribution with shape parameter $k = 1.67$ and scale parameter $c = 12.9$ knots (or 6.62 m s^{-1}). The bold vertical line denotes the threshold $v_{\max}$= 52.1 knots, beyond which wind speed data is flagged as erroneous.

(MLE) was used to determine k and c (Wilks, 1995). The exponent k will be shown to be indicative of the underlying statistical dynamics.

There was a very weak dependence of k and c on the range of raw data $[0, v_{\text{fit}}]$ to which MLE was applied for sufficiently large v_{fit}. This dependence is inherently weak because log-likelihood for a set of n_{tot} data points is defined as:

$$\ln \ell \stackrel{\text{def}}{=} \sum_{i=1}^{n_{\text{tot}}} \ln P(v_i)$$
$$\Rightarrow \frac{\sum_{v_i > v_{\text{fit}}} \ln P(v_i)}{\sum_{v_0 < v_i < v_{\text{fit}}} \ln P(v_i)} < \frac{n(v_i > v_{\text{fit}})}{n(v_0 < v_i < v_{\text{fit}})} \ll 1,$$

for sufficiently large v_{fit}

where v_0 is such that $P(v)$ decreases monotonically for $v > v_0$ and n refers to the number of data points. The contribution of extreme values to the log-likelihood is evidently negligible and so k and c estimated by MLE are not sensitive to erroneously large values of the data. Nonetheless, in order to proceed in practice, the value of v_{fit} used for the Weibull fit was selected by minimizing the mean absolute difference between the fitted distribution and the frequency histogram over all available data. (Other criteria for choosing v_{fit} were tested with no significant difference in the results.)

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

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

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

$$\sigma^2 \stackrel{\text{def}}{=} \int_0^\infty v^2 P(v) dv = c^2 \int_0^\infty t^{2/k} \exp(-t) dt \qquad (2)$$
$$= c^2 \Gamma(\tfrac{2}{k} + 1)$$

where Γ is the gamma function (Arfken, 2000). This implies that c is constrained by the climatological wind speed measured by σ for a given k. In this work, it was 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.

From the PDF, the expected number $n_{\text{fit}}(v)$ of wind speed reports in the bin $[v,\ v+\delta v]$ of the frequency histogram was computed for the size of the dataset. The wind speed threshold $v_{\max}$ was defined such that $n_{\text{fit}}(v) \leq 1$ for $v \geq v_{\max}$. By the fitted Weibull distribution, wind speed records larger than $v_{\max}$ are unlikely to be reliable for the given dataset size. It was checked that $v_{\max}$ was smaller than v_{fit} at all levels, confirming the validity of the Weibull fit including values around $v_{\max}$.

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. The value of $k$increases upward from 1.67 ± 0.01 at 925 mb to values somewhat bigger than 2 in the upper troposphere (300 mb to 150 mb). Similar values of around 5/3 are noted at the tropopause level (100 mb) and in the PBL (925 mb and 850 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.

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 statistical dynamics. 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 Approach to Gaussian statistics

Suppose the horizontal wind vector $\boldsymbol{v}$ may be decomposed into numerous contributions associated with different tropical meteorological phenomena:

$$\boldsymbol{v} = \sum_n \boldsymbol{v}_n \qquad (3)$$

For instance, $\boldsymbol{v}_1$ can arise from diurnally excited gravity waves (Rotunno, 1983), $\boldsymbol{v}_2$ from equatorial waves (Wheeler and Kiladis, 1999), $\boldsymbol{v}_3$ from intra-seasonal oscillations (Madden and Julian, 1971, 1994; Waliser, 2006), $\boldsymbol{v}_4$ from Asian-Australian monsoon (Wang, 2006), $\boldsymbol{v}_5$ from Walker circulation (Katz, 2002) under inter-annual variations etc. One

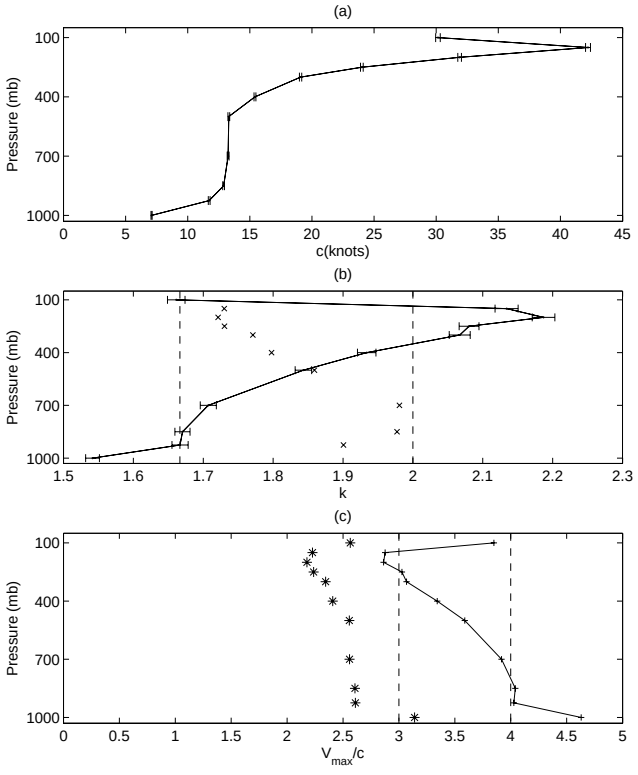

Fig. 3. Plots of empirically fitted attributes of the Weibull distribution for wind speed at different pressure levels in MP: **(a)** scale parameter c; **(b)** shape parameter k, **(c)** scaled threshold $v_{\max}/c$ for wind speed. Error bars for c and k are estimated by MLE at 95 % confidence level.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

might even split the contributions among sub-categories, distinguishing between: land-sea and mountain-valley diurnal circulations; Kelvin and Rossby waves of different equivalent depth; monsoon cold surges and westerly wind bursts; El Niño – Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD). But the detailed cause of each $\boldsymbol{v}_n$ is not important to the following argument as long as there are more than a few independent $\boldsymbol{v}_n$ contributing to $\boldsymbol{v}$. Over a long time such as 35 years, the set of values that each $\boldsymbol{v}_n$ takes may be reproduced by the realizations of a random variable with its own characteristic probability distribution. Note that this is NOT saying that each contribution actually varies randomly in time.

Assuming each random variable $\boldsymbol{v}_n$ in Eq. (3) 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 of random variables, N, increases without bound. Note that the mean of $\sum \boldsymbol{v}_n$ is the sum of mean $\boldsymbol{v}_n$, and the variance of $\sum \boldsymbol{v}_n$ is the sum of the variance of $\boldsymbol{v}_n$. (Appendix A shows that the Central Limit Theorem is still applicable even in the case where the set of $\boldsymbol{v}_n$ has members with non-zero covariance.)

Thus, in the limit of large N, wind velocity $\boldsymbol{v}$ follows the Gaussian distribution,

$$P(\boldsymbol{v} = \begin{bmatrix} v_x \\ v_y \end{bmatrix}) \, d^2\boldsymbol{v} = \frac{1}{\pi \, c^2} \exp\left[-\left(\frac{v_x - \bar{v}_x}{c} \right)^2 \right] \cdot \exp\left[-\left(\frac{v_y - \bar{v}_y}{c} \right)^2 \right] dv_x dv_y \tag{4}$$

where c^2 is the variance of $\boldsymbol{v}$ and is twice the variance of v_x or v_y because the wind velocity anomaly is assumed to be isotropic. The isotropic assumption is supported by the virtual absence, or at best, weakness of anisotropy from observations (e.g. Mori, 1986; Ibarra, 1995; Koh and Ng, 2009).

For zero mean wind, integrating over all directions θ, the PDF for wind speed v is the Rayleigh distribution:

$$P(v)dv = \int_0^{2\pi} P(\boldsymbol{v}) v \, dv \, d\theta \approx \frac{2v}{c^2} \exp\left[-\left(\frac{v}{c} \right)^2 \right] dv \tag{5}$$

which is also Weibull distribution with shape parameter k = 2. It must be emphasized that the derivation of Eq. (5) does not depend on the statistical dynamics of each contribution $\boldsymbol{v}_n$ (or $\boldsymbol{w}_n$) except for the empirically justifiable assumptions of isotropy. The assumption of zero mean wind will be examined later. Thus, the statistical dynamics of many independent contributions from tropical meteorological phenomena may explain why $k \approx 2$ is observed generally in the upper troposphere in Fig. 3b. In fact, Rayleigh distribution cannot be rejected by the χ^2-test at 90 % confidence level for levels between 500 mb and 150 mb inclusive. (Note that this does not mean that Rayleigh distribution is the best-fit distribution.)

5.1.1 Departure from Gaussian statistics

For the levels below 500 mb and at 100 mb, the χ^2-test rejects Rayleigh distribution at 90 % confidence level. Moreover, MLE fits of Weibull distribution did show $k < 2$ for those levels and small but significant deviations from $k = 2$ in the upper troposphere. Further theoretical understanding for such departure from Gaussian behavior is sought below by: (1) examination the effect of non-zero mean wind; (2) introducing Shannon's entropy as a measure of non-Gaussianity and explaining its variation in the lower and upper troposphere.

5.1.2 Non-zero mean wind

In the presence of non-zero mean wind, Appendix B shows that for isotropic Gaussian wind anomalies as in Eq. (4), the PDF for wind speed v is

$$P(v)\,dv = \underline{\exp\left(-\frac{v_{\mathrm{m}}^2}{c^2} \right) \cdot I_0\left(\frac{2v_{\mathrm{m}} v}{c^2} \right)} \cdot \frac{2v}{c^2} \exp\left(-\frac{v^2}{c^2} \right) dv \tag{6}$$

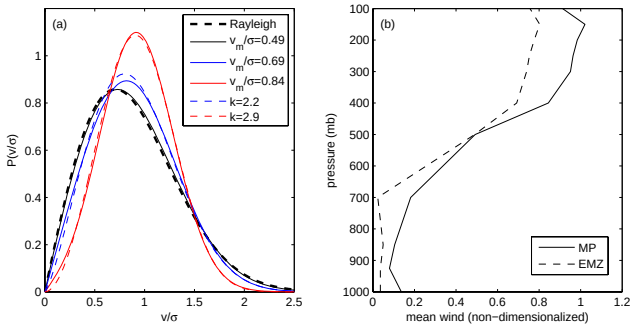

Fig. 4. (a) The Rayleigh distribution (bold dashed) is compared to the PDF of wind speed when the mean wind $v_{\rm m}$ is non-zero (solid) assuming Gaussian wind anomalies. Examples of Weibull distributions with $k > 2$ (dashed) are also shown for comparison. Wind speed v is normalized by its rms value σ for each distribution to facilitate comparison. **(b)** Normalized mean wind $v_{\rm m}/\sigma$ for the set of 7 stations on Malay Peninsula (MP) and mean normalized wind $u_{\rm m}$ for Equatorial Monsoon Zone (EMZ), where $u = v/\sigma$ for each station.

where $v_{\rm m}$ is the magnitude of the mean wind vector and I_0 is the modified Bessel function of the first kind. When $v_{\rm m}$ is zero, the underlined factor in Eq. (6) is unity, recovering Rayleigh's distribution. This factor comprises two terms: as $v_{\rm m}$ increases, the exponential function decreases while the modified Bessel function increases. The two tendencies tend to balance for small $v_{\rm m}$, but the latter wins out for large $v_{\rm m}$ and so the PDF departs increasingly from Rayleigh distribution.

The effect of non-zero mean wind $v_{\rm m}$ on the PDF in Eq. (6) was computed and is shown in Fig. 4a, where the PDFs are expressed in terms of wind speed normalized by its rms value σ to facilitate comparison. For $v_{\rm m}/\sigma < 0.5$, the effect can be neglected as $P(v/\sigma)$ approximates that for Rayleigh distribution. While for $v_{\rm m}/\sigma > 0.5$ the effect of non-zero mean wind is significant, the PDF can in practice be approximated by Weibull distributions with $k > 2$ (cf. dashed and solid lines in Fig. 4a). This explains why Weibull fits are still good even in the presence of large mean wind.

The magnitude of the mean wind $v_{\rm m}$ in MP is shown in Fig. 4b, normalized by the rms wind speed σ over the region. At 500 mb and below, $v_{\rm m}/\sigma \leq 0.49$, which implies the mean wind has negligible effect on $P(v)$ and is not the main cause of $k \neq 2$ at those levels. At 400 mb and above, $v_{\rm m}/\sigma \geq 0.84$. So, using Fig. 4a, Weibull fits to $P(v)$ would result in $k = 2.9$ or larger at those levels, if the wind anomalies were Gaussian. Thus, strong mean wind can explain why k tends to sometimes overshoot 2 at upper levels, but it is also clear that the wind anomalies are non-Gaussian because k is considerably less than 2.9 (Fig. 3b). There must be another cause for $k \neq 2$ in the upper troposphere that reduces the value of k.

5.1.3 Shannon's entropy

Shannon's entropy for a random two-dimensional vector variable $\boldsymbol{v}$ is defined as

$$\mathrm{Ent}[P(\boldsymbol{v})] \stackrel{\mathrm{def}}{=} -\int\int_{\mathrm{all}\boldsymbol{v}} P(\boldsymbol{v}) \ln P(\boldsymbol{v})\, d^2\boldsymbol{v} \qquad (7)$$

Among all probability density functions $P(\boldsymbol{v})$ of unit variance, Shannon's entropy is maximal for the Gaussian distribution only (Artstein et al., 2004). Thus, the Central Limit Theorem may be understood as an approach towards maximal Shannon's entropy. Small Shannon's entropy denotes strong departure from Gaussianity in the distribution.

Assuming that isotropy prevails, Appendix C shows that Shannon's entropy of Weibull distributions of unit variance is related to the shape parameter k,

$$\mathrm{Ent}[P(v/\sigma)] = \ln\left(\tfrac{2\pi}{k}\right) - \ln\left[\Gamma(\tfrac{2}{k}+1)\right] + (1-\tfrac{2}{k})\gamma + 1 \qquad (8)$$

where Γ is again the gamma function (Arfken, 2000) and γ is the Euler-Mascheroni constant (Whittaker and Watson, 1996). The expression shows that Shannon's entropy has maximum value of $E_{\max} = \ln\pi + 1$ at $k = 2$ and decreases monotonically for k larger or smaller than two (graph not shown). Thus, Shannon's entropy corresponding to the k-values at 500 mb and below in Fig. 3b was computed and shown in Fig. 5. Shannon's entropy at upper levels was not computed because the effect of strong mean wind on $P(v/\sigma)$ implies that Eq. (8) is not applicable to wind speed but to wind anomaly magnitude only.

From Eq. (8), to understand the variation of k is to understand the variation of Shannon's entropy. In Appendix D, we show that notwithstanding the Central Limit Theorem, when the variances are non-uniform among the velocity contributions $\boldsymbol{v}_n$ in Eq. (3), Shannon's entropy of $P(v/\sigma)$ can decrease by an amount as much as ΔE as the number of independent contributions N increases, i.e. the approach to Gaussianity is not monotonic in general. For large N, the theoretical lower bound for Shannon's entropy, $(E_{\max} - \Delta E)$, could be estimated roughly (see Appendix D for details).

The vertical trend in $(E_{\max} - \Delta E)$ in Fig. 5 shows that even for large N, wind anomalies can have the most departure from Gaussianity between 400 mb and 150 mb, which is consistent with our deduction at the end of Sect. 5.1.2. Thus, $k \neq 2$ in the upper troposphere may arise in part from the non-uniform variance among the wind contributions $\boldsymbol{v}_n$. For illustration, rough estimates of the theoretical lower bound for k for wind anomaly magnitude were computed from $(E_{\max} - \Delta E)$ by inverting Eq. (8) (crosses in Fig. 3b). The decreasing effect on k by non-uniform variance appears to compete with the increasing effect on k by strong mean wind in the upper troposphere, resulting in k close to and sometimes overshooting 2.

From 925 mb to 500 mb in Fig. 5, ΔE is negligible because the variance is roughly uniform among wind contributions $\boldsymbol{v}_n$. But Shannon's entropy is much less than $E_{\max}$.

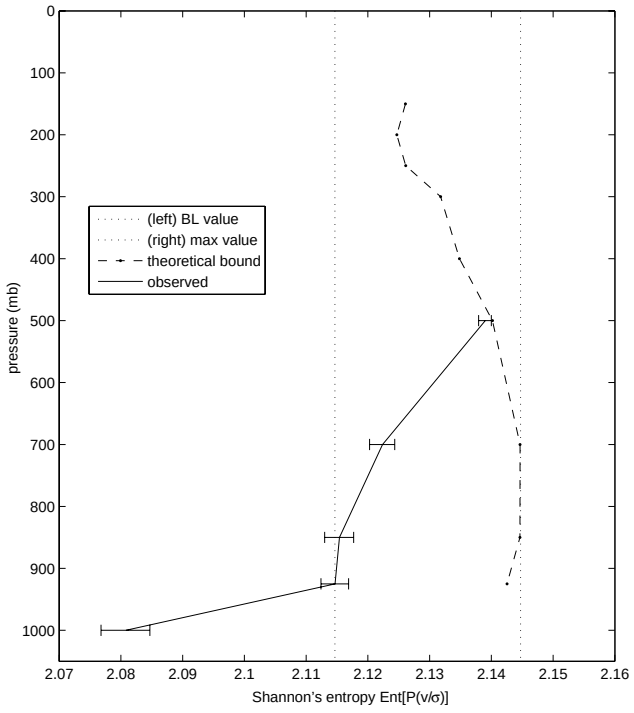

Fig. 5. Shannon's entropy for the Weibull's distributions fitted to the radiosonde wind speed data from MP at 500 mb and below (solid line). The wind speed was first normalized by its rms value. The maximal Shannon's entropy and the value associated with $k = 5/3$ (dotted lines) are shown. The theoretical bound (dashed line) for reduction from maximal Shannon's entropy is computed for wind anomaly using Eq. (8).

This means N is not large enough for the wind anomalies to approach Gaussianity. Artstein et al. (2004) proved that when the velocity contributions $\boldsymbol{v}_n$ have uniform variance, Shannon's entropy increases monotonically as the number of contributions N increases (see Appendix D). This is consistent with the trend in Shannon's entropy in the lower troposphere, suggesting an increase in the number of independent contributions N with height.

6 Application to monitoring data quality

The preceding understanding for Weibull distribution of 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 about half a percent of 278 711 available wind speed records at these three levels are suspect.

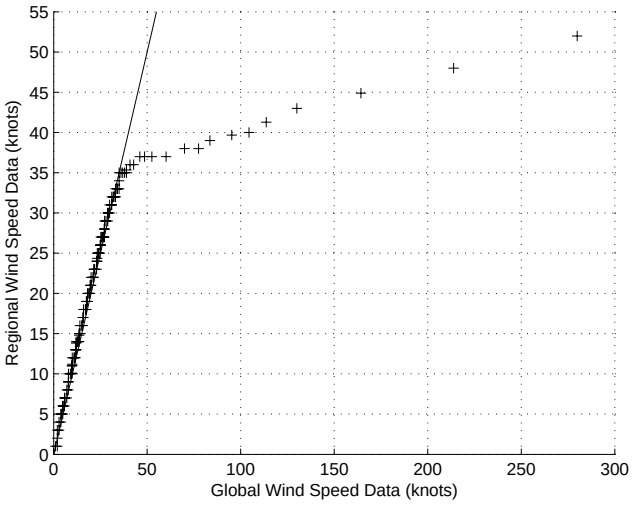

Fig. 6. The 850 mb-wind speed over MP from Wyoming data archive after screening with the threshold $v_{\max}$ (52.1 knots) plotted against the wind speed from IGRA database for the same stations and period. Each ordered pair denoted by a cross refers to the same 0.01 %-quantile in both datasets. The straight line shows where the crosses should lie if both datasets had the same distribution.

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. 3c, $v_{\max}$ 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. 6. 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. Similar large erroneous wind speed in Indonesia reported over the Global Telecommunication System (GTS) was also noted by Okamoto et al. (2003).

In modern data assimilation systems, such as used by European Centre for Medium-Range Weather Forecast (ECMWF), east-west and north-south wind components are analyzed separately and often assumed to follow Gaussian distributions of equal variance. The non-Gaussianity identified in the last section, especially in the lower troposphere where the mean wind is weak may be cause for reexamination of these assumptions. Moreover, in eliminating unrealistic wind speed (e.g. $v > v_{\max} = 52.1$ knots at 850 mb in MP, see Fig. 6), the proposed QC method would raise the

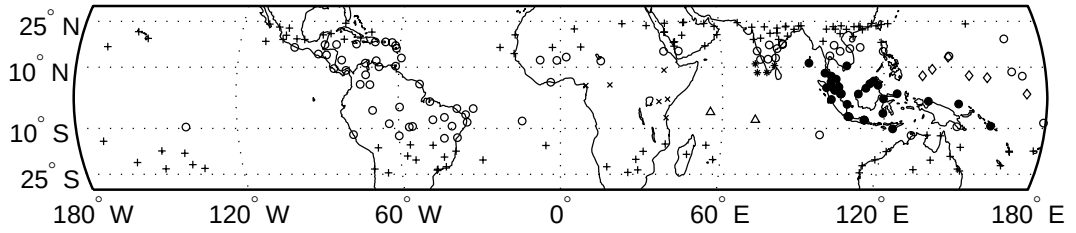

Fig. 7. All 242 tropical radiosonde stations used in the latter part of the study (including the seven stations in MP): "+" signs denote stations in the upper-level (500 mb to 100 mb) westerly zone; circles denote stations in the upper-level mixed wind zone; all other symbols denote stations in the upper-level 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).

quality of data assimilated and incrementally improve model analyses and re-analyses in the tropics. This would eventually contribute to the quality of model first-guess fields so that they could be used more reliably to check tropical observations at the time of assimilation.

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 (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. 7). 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. 8 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. 8c) 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 figure 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.

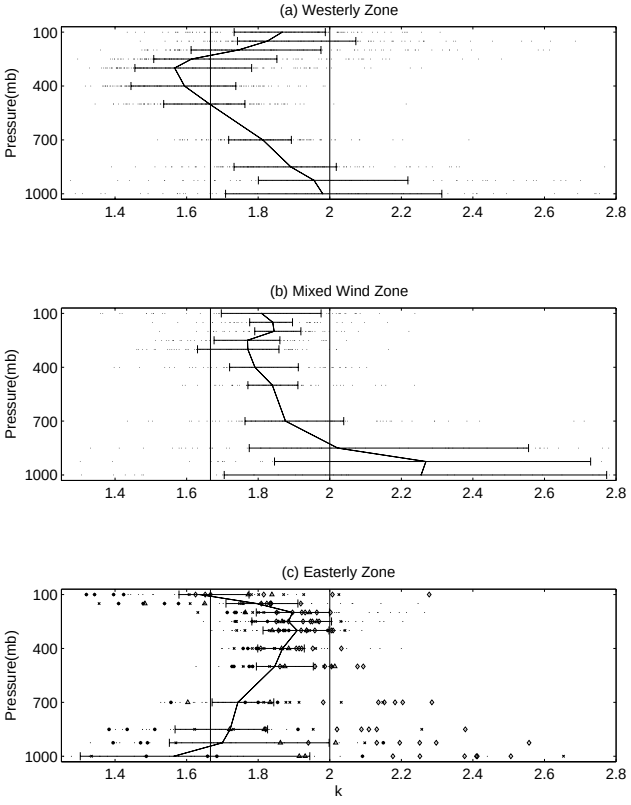

Fig. 8. Scatter plot showing the values of k across vertical levels at tropical stations in the three upper-level climatic wind zones. For each zone: median values at each level are connected to show a vertical profile; the delimited horizontal bars denote the inter-quartile range at each level. For 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. 7.

Unlike MP, EMZ spans half the globe across varying climatology of wind speed (Fig. 9a). 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

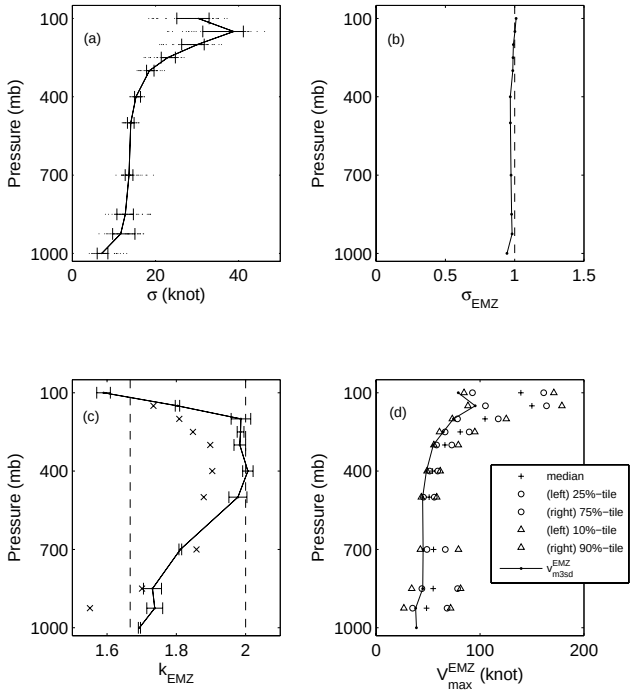

Fig. 9. (a) Scatter plot of rms wind speed σ computed using Eq. (2) at all 43 stations in the Equatorial Monsoon Zone (EMZ). The vertical profile connects the median values and delimited horizontal bars denote the inter-quartile ranges. **(b)** The rms value σ_{EMZ} of non-dimensional wind speed $u = v/\sigma$ computed from the fitted Weibull distribution of u in EMZ using Eq. (2). **(c)** The shape parameter k_{EMZ} describing the fitted Weibull distribution of u in EMZ with error bars shown. Crosses denote theoretical lower bound for k for wind anomaly magnitude. **(d)** The spread of threshold wind speed (symbols) among the stations in EMZ compared to (line).

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. From Fig. 9b, we see that $\sigma_{\mathrm{EMZ}} \approx 1$ for all levels indeed. 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.

It was not possible to carry out the χ^2-test for the Weibull fit for the wind speed in the EMZ because the spatial autocorrelation is hard to estimate reliably from an irregular station distribution. However, the fact that $\sigma_{\mathrm{EMZ}} \approx 1$ is indirect but clear evidence that the Weibull distribution is a good fit because otherwise, Eq. (2) would have yielded wrong values not only for σ_{EMZ} but also for σ at each station in the first place.

The characteristic profile of k_{EMZ} and the associated threshold for wind speed $v_{\mathrm{max}}^{\mathrm{EMZ}}$ are shown in Fig. 9c and d respectively. At each station, $v_{\mathrm{max}}^{\mathrm{EMZ}}$ was estimated as the product of the local σ and the regional threshold u_{max} derived essentially from k_{EMZ}: the expected number $n_{\mathrm{fit}}(u, \delta u) \leq 1$ for $u \geq u_{\mathrm{max}}$ using a bin size δu of 2 knots divided by the mean σ at each level. Away from the surface, $v_{\mathrm{max}}^{\mathrm{EMZ}}$ is larger than the $v_{\mathrm{m3sd}}^{\mathrm{EMZ}}$ for about 90 % or more of the stations, where $v_{\mathrm{m3sd}}^{\mathrm{EMZ}}$ 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, $v_{\mathrm{max}}^{\mathrm{EMZ}}$ preserves more useful data than $v_{\mathrm{m3sd}}^{\mathrm{EMZ}}$ because $v_{\mathrm{max}}^{\mathrm{EMZ}}$ captures the region's statistical dynamics and is adapted to the local wind climatology. At 1000 mb, $v_{\mathrm{m3sd}}^{\mathrm{EMZ}}$ 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. 10. As before, Shannon's entropy was not computed for 400 mb and above because of the effect of strong mean wind ($u_{\mathrm{m}} > 0.69$) on $P(u/\sigma_{\mathrm{EMZ}})$ (Fig. 4). The error bars for k in Fig. 9c and hence for Shannon's entropy in Fig. 10 were estimated by generating another two sets of best-fit k_{EMZ} by separately removing the stations with the top or bottom 5 percentile of k-values (i.e. top or bottom 2 stations) and computing the standard deviation among the three sets of best-fit k_{EMZ}. The theoretical lower bound for Shannon's entropy, $(E_{\mathrm{max}} - \Delta E)$, could also be estimated roughly for large N as before (see Appendix D for details).

Compared to the MP results in Fig. 5, the EMZ results in Fig. 10 show that the lower troposphere (1000 mb to 500 mb) is nearer to attaining maximal entropy because there is a larger number of independent velocity contributions N arising from spatial de-correlation within EMZ. Below 850 mb, it appears that the non-uniformity of variance among the contributions across EMZ may be keeping Shannon's entropy away from the maximal value (dashed line in Fig. 10). Above 500 mb, competing effects from non-uniform variance among velocity contributions (that decrease k) and large mean wind (that increase k) tend to balance leading to $k \approx 2$, although nearer the tropopause the former effect seems to dominate (crosses in Fig. 9c). Note that $k \approx 2$ does not imply Gaussianity (but the converse would be true).

8 Summary and discussion

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 investigated in this paper; the scale parameter c is determined by a given k and the rms wind speed σ which is the result of planetary-scale climate dynamics. Wind in EMZ was non-dimensionalized by the local σ to remove the effect of

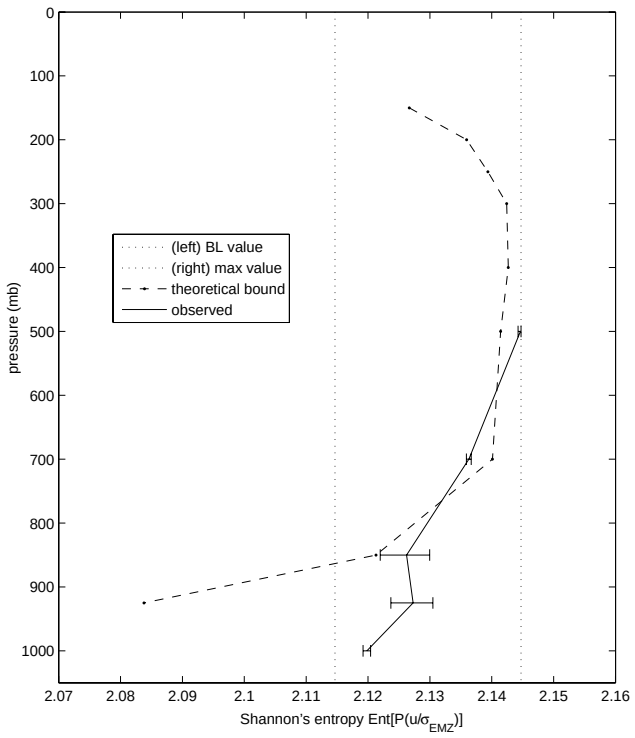

Fig. 10. Figure 10 shows Shannon's entropy at 500 mb and below for the Weibull's distributions fitted to the non-dimensional wind speed $u = v/\sigma$ from Equatorial Monsoon Zone (solid line). u was first normalized by its rms value σ_{EMZ} as the latter is close to but not exactly 1. The maximal entropy 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.

geographic variation of climatology before empirical fitting of the Weibull distribution.

A statistical theory of independent physical contributions to the observed wind was proposed to explain the observed k as follows.

1. The increase in the number of such contributions N causes Shannon's entropy to rise and the value of k to approach 2 from the lower to mid-troposphere.

2. In the upper troposphere, N is likely to be large. But the non-uniformity of variance among the velocity contributions prevents Shannon's entropy from attaining the maximal value and tends to decrease k, while strong mean wind tends to increase k. Thus, k has values close to 2 (EMZ) or sometimes may overshoot 2 (MP).

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 near-equatorial wind and they are adapted to the local climatology. Such an improved dataset from EMZ would ultimately benefit research and forecast.

The existence of non-Gaussianity in the troposphere appears to be a natural consequence of non-linear dynamical models. Sardeshmukh and Sura (2009) showed in an adiabatic GCM that skewness and excess kurtosis (which are identically zero for Gaussian distribution and hence represent non-Gaussian behaviour) are associated mainly with small-scale turbulent fluxes. Interestingly, they also showed that the statistical relation between skewness and excess kurtosis can be reproduced in linear stochastic models when additive (state-independent) and multiplicative (state-dependent) Gaussian white noises are correlated. This may hint at further investigation of the non-Gaussianity identified in this work with linear equatorial wave models.

The current work also raises specific questions: (1) What are the physical causes of the dominant velocity contributions? (2) Why does the number of independent velocity contributions N seem to increase with height? (3) Does the seemingly common value of $k = 5/3$ observed in the PBL (850 and 925 mb) and at the tropopause (100 mb) in MP reflect any statistical dynamics occurring at local scales at those levels or is it mere coincidence in this dataset? (4) How do we understand the profile of k outside of the EMZ? (5) What distributions do tropical temperature and humidity follow and what are their underlying statistical dynamics? These questions and others leave much room for exciting research into the statistical dynamics of regional atmospheres.

Appendix A

Non-zero covariance between velocity contributions

If a particular $\boldsymbol{v}_n$ has non-zero covariance with another $\boldsymbol{v}_{\mathrm{m}}$, the two velocity contributions would not be independent. However, it is easy to define two new variables as follows:

$$\boldsymbol{w}_n \stackrel{\mathrm{def}}{=} \boldsymbol{v}_n - \mathrm{cov}(\boldsymbol{v}_n, \boldsymbol{v}_{\mathrm{m}})[\mathrm{var}(\boldsymbol{v}_{\mathrm{m}})]^{-1}\boldsymbol{v}_{\mathrm{m}}$$
$$\boldsymbol{w}_{\mathrm{m}} \stackrel{\mathrm{def}}{=} \left\{1 + \mathrm{cov}(\boldsymbol{v}_n, \boldsymbol{v}_{\mathrm{m}})[\mathrm{var}(\boldsymbol{v}_{\mathrm{m}})]^{-1}\right\}\boldsymbol{v}_{\mathrm{m}}$$

where the variance and covariance for vectors are defined in e.g. Feller (1968). The contributions to $\boldsymbol{v}$ could be re-expressed as

$$\boldsymbol{v}_n + \boldsymbol{v}_{\mathrm{m}} \equiv \boldsymbol{w}_n + \boldsymbol{w}_{\mathrm{m}}$$

It is readily verified that the new random variables, $\boldsymbol{w}_n$ and $\boldsymbol{w}_{\mathrm{m}}$, have zero covariance. In this way, the Central Limit Theorem may be applied as before.

Appendix B

PDF for Gaussian wind anomaly under non-zero mean wind

From Eq. (4),

$$P(\boldsymbol{v})d^2\boldsymbol{v} = \frac{1}{\pi c^2}\exp\left[-\frac{(\boldsymbol{v}-\overline{\boldsymbol{v}}\cdot\boldsymbol{v}-\overline{\boldsymbol{v}})}{c^2}\right]d^2\boldsymbol{v}$$

$$= \frac{1}{\pi c^2}\exp\left(\frac{v_{\mathrm{m}}^2}{c^2}\right)\cdot\exp\left(\frac{2v_{\mathrm{m}}v}{c^2}\cos\theta\right)\cdot\exp\left(-\frac{v^2}{c^2}\right)v\,dvd\theta$$

where v_{m} is the magnitude of the mean wind and θ is the angle between the wind vector and the mean wind. Integrating over all angles, the PDF for wind speed v is

$$P(v)dv = \exp\left(-\frac{v_{\mathrm{m}}^2}{c^2}\right)\cdot I_0\left(\frac{2v_{\mathrm{m}}v}{c^2}\right)\cdot\frac{2v}{c^2}\exp\left(-\frac{v^2}{c^2}\right)dv \quad \text{(B1)}$$

where I_0 is the modified Bessel function of the first kind (Arfken et al., 2000):

$$I_0(\alpha) \equiv \tfrac{1}{2\pi}\textstyle\int_0^{2\pi}\exp(\alpha\cos\theta)d\theta$$

Appendix C

Shannon's entropy associated with Weibull distribution

We assume that the wind velocity $\boldsymbol{v}$ is isotropic and hence has zero mean. Equivalently, we can take $\boldsymbol{v}$ as an isotropic wind anomaly if the mean wind is non-zero. Let the variance of the distribution be σ^2. For the magnitude v obeying Weibull distribution, we define a non-dimensional variable $u = v/\sigma$ that follows Weibull distribution of unit variance:

$$P(u)\,du = P(v)\,dv$$

$$= k\sqrt{\Gamma(\tfrac{2}{k}+1)}\left[u\sqrt{\Gamma(\tfrac{2}{k}+1)}\right]^{k-1}\exp\left\{-\left[u\sqrt{\Gamma(\tfrac{2}{k}+1)}\right]^k\right\}du$$

$$= -d[\xi(u)]$$

where Eqs. (1 and 2) were used and $\xi(u) = \exp\left\{-\left[u\sqrt{\Gamma(\frac{2}{k}+1)}\right]^k\right\}$. From Eq. (7), Shannon's entropy for u is

$$\mathrm{Ent}[P(u)] = -\int_0^\infty\left[\frac{P(u)}{2\pi u}\ln\frac{P(u)}{2\pi u}\right]u\,du\int_0^{2\pi}d\theta \quad \text{(C1)}$$

$$= -\int_0^\infty\ln\left[\frac{P(u)}{2\pi u}\right]P(u)\,du$$

$$= -\int_0^1\ln\left[\tfrac{k}{2\pi}\Gamma(\tfrac{2}{k}+1)(-\ln\xi)^{(k-2)/k}\xi\right]d\xi$$

$$= -\ln\left[\tfrac{k}{2\pi}\Gamma(\tfrac{2}{k}+1)\right] - \left(1-\tfrac{2}{k}\right)\int_0^1\ln(-\ln\xi)d\xi - \int_0^1\ln\xi\,d\xi$$

$$= \ln\left(\tfrac{2\pi}{k}\right) - \ln\left[\Gamma(\tfrac{2}{k}+1)\right] + (1-\tfrac{2}{k})\gamma + 1$$

In the last step above, we made use of

$$\textstyle\int_0^1\ln(-\ln\xi)\,d\xi \equiv -\gamma$$
$$\textstyle\int_0^1\ln\xi\,d\xi = [\xi\ln\xi]_0^1 - \int_0^1 d\xi = -1$$

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

Appendix D

Derivation of entropy increment

Recalling Eq. (3), let σ_n^2 be the variance of $\boldsymbol{v}_n$. We define the following average variances and partial sums of $\boldsymbol{v}_n$ normalized to unit variance:

$$s_{N+1}^2 \overset{\text{def}}{=} \frac{1}{N+1}\sum_{n=1}^{N+1}\sigma_n^2 \quad ; \quad \boldsymbol{u}_{N+1} \overset{\text{def}}{=} \frac{1}{s_{N+1}\sqrt{N+1}}\sum_{n=1}^{N+1}\boldsymbol{v}_n \quad \text{(D1)}$$

$$s_{N,m}^2 \overset{\text{def}}{=} \frac{1}{N}\sum_{n=1,n\neq m}^{N+1}\sigma_n^2 \quad ; \quad \boldsymbol{u}_{N,m} \overset{\text{def}}{=} \frac{1}{s_{N,m}\sqrt{N}}\sum_{n=1,n\neq m}^{N+1}\boldsymbol{v}_n \quad \text{(D2)}$$

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

$$\mathrm{Ent}\left[P\left(\tfrac{1}{\sqrt{N+1}}\sum_{n=1}^{N+1}\boldsymbol{v}_n\right)\right] \geq \frac{1}{N+1}\sum_{m=1}^{N+1}\mathrm{Ent}\left[P\left(\tfrac{1}{\sqrt{N}}\sum_{n=1,n\neq m}^{N+1}\boldsymbol{v}_n\right)\right] \quad \text{(D3)}$$

Using definitions (D2) and (D3) and the identity $\mathrm{Ent}(s\boldsymbol{u}) \equiv \ln s + \mathrm{Ent}(\boldsymbol{u})$, Eq. (D3) becomes

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

where the straight overbar denotes arithmetic mean from $m=1$ to $N+1$. Using the identities,

$$\frac{s_{N+1}^2 \equiv \overline{s_{N,m}^2}}{\overline{\ln s_{N,m}^2} \equiv \ln\widetilde{s_{N,m}^2}}$$

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

$$\overline{\mathrm{Ent}[P(\boldsymbol{u}_{N+1})]} - \overline{\mathrm{Ent}[P(\boldsymbol{u}_{N,m})]} \geq \tfrac{1}{2}\ln\left(\widetilde{s_{N,m}^2}\Big/\overline{s_{N,m}^2}\right) \overset{\text{def}}{=} \Delta E \quad \text{(D4)}$$

ΔE is the minimal increment in Shannon's entropy that can be expected.

The following statements can be deduced from Eq. (D4):

1. If mean square velocity σ_n^2 is invariant of n, $\widetilde{s_{N,m}^2} = \overline{s_{N,m}^2}$. Thus, Shannon's entropy is expected to increase monotonically with N (Artstein, 2004).

$$\overline{\mathrm{Ent}[P(\boldsymbol{u}_{N+1})]} - \overline{\mathrm{Ent}[P(\boldsymbol{u}_{N,m})]} \geq 0 \quad \text{(D5)}$$

2. If σ_n^2 varies with n, it is straightforward to show that $\widetilde{s_{N,m}^2} < \overline{s_{N,m}^2}$ and $\ln\left(\widetilde{s_{N,m}^2} \Big/ \overline{s_{N,m}^2}\right) < 0$. Thus, Shannon's entropy may even be expected to decrease as N increases because $\Delta E < 0$.

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

By supposing that σ_n take with equal chance one of two values, S_0 and S_1, the largest possible reduction in Shannon's entropy ΔE can be simply estimated using Eq. (D4):

$$\Delta E \approx \tfrac{1}{2} \ln\left(\frac{2S_0S_1}{S_0^2+S_1^2}\right) \leq 0 \tag{D6}$$

For MP, we used the ratio of rms wind speed σ of neighbouring levels (Eq. 2) as proxy for the ratio of variances of velocity contributions:

$$\left(\frac{S_0}{S_1}\right)_i \approx \frac{\sigma(c_i,k_i)}{\sigma(c_{i-1},k_{i-1})} \text{ for } i > 1 \tag{D7}$$

where the level index i increases downwards with pressure.

For EMZ, unlike for MP, the variance of wind contributions could be estimated from the extensive spatial sampling of rms wind speed across EMZ (Fig. 9a). So we used the following estimate:

$$\left(\frac{S_0}{S_1}\right)_i \approx \left(\frac{\sigma_{\text{upp}}}{\sigma_{\text{low}}}\right)_i \tag{D8}$$

where σ_{upp} and σ_{low} are the upper and lower quartiles of the rms wind speed at each station in the EMZ respectively. As the choices in Eqs. (17 and 18) are only rough estimates, the emphasis is on the vertical trend of the theoretical bound ΔE and not on the values per se.

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. We are also grateful to the constructive comments by anonymous reviewers which have helped improved the quality of this manuscript.

Edited by: T. J. Dunkerton

References

Arfken, G. B., Weber, H. J., and Harris, F.: Mathematical Methods for Physicists, 5th Ed., Academic Press, p. 1112, 2000.

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.

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.

Durre, I., Vose, R. S., and Wuertz, D. B.: Overview of the Integrated Global Radiosonde Archive, J. Climate, 19, 53–68, 2006.

Feller, W.: An Introduction to Probability Theory and its Applications, 3rd Ed., Wiley, New York, 227–233, 1968.

Frank, H. P. and Landberg, L.: Modelling the wind climate of Ireland, Bound.-Lay. Meteorol., 85, 359–378, 1997.

Gandin, L. S.: Complex quality control of meteorological observations, Mon. Weather Rev., 116, 1137–1156, 1988.

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. Meteorol. 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.

Katz, R. W.: Sir Gilbert Walker and a Connection between El Niño and Statistics, Stat. Sci., 17, 97–117, 2002.

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, B. Amer. Meteorol. Soc., 90(4), 481–488, doi:10.1175/2008BAMS2561.1, 2009.

Labraga, J. C.: Extreme winds in the Pampa del Castillo plateau, Patagonia, Argentina, with reference to wind farm settlement, J. Appl. Meteorol., 33, 85–95, 1994.

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, Renewable Energy., 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 character-

istics and resources. In: Wind Energy Explained, Theory, Design and Application, Wiley, New York, 21–80, 2002.

Mori, Y.: Evaluation of several "single-pass" estimators of the mean and the standard deviation of wind direction, J. Clim. Appl. Meteorol., 25, 1387–1397, 1986.

Neale, R. B. and Slingo, J. M.: The maritime continent and its role in the global climate: a GCM study, J. Climate, 16, 834–848, 2003.

Okamoto, N., Yamanaka, M. D., Ogino S.-Y., Hashiguchi, H., Nishi, N., Sribimawati, T., and Numaguti, A.: Seasonal variations of tropospheric wind over Indonesia: comparison between collected operational rawinsonde data and NCEP reanalysis for 1992–1999, J. Meteorol. Soc. Japan, 81, 829–850, 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, 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.

Sardeshmukh, P. D. and Sura, P.: Reconciling non-Gaussian climate statistics with linear dynamics, J. Climate, 22, 1193–1207, doi:10.1175/2008JCLI2358.1, 2009.

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, B. Am. 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. Meteorol., 17, 556–559, 1978.

Wang, B.: The Asian Monsoon. Springer-Praxis, Heidelberg, Germany, p. 844, 2006.

Waliser, D. E.: Intraseasonal variability, in: The Asian Monsoon, Springer-Praxis, Heidelberg, Germany, 203–257, 2006.

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.

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 ed, 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. Meteorol., 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, 184–199, 1993.