Finescale Clusterization Intermittency of Turbulence in the Atmospheric Boundary Layer

LEI LIU AND FEI HU

State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

(Manuscript received 30 September 2019, in final form 24 April 2020)

ABSTRACT

The intermittency of atmospheric turbulence plays an important role in the understanding of particle dispersal in the atmospheric boundary layer and in the statistical simulation of high-frequency wind speed in various applications. There are two kinds of intermittency, namely, the magnitude intermittency (MI) related to non-Gaussianity and the less studied clusterization intermittency (CI) related to long-term correlation. In this paper, we use a 20 Hz ultrasonic dataset lasting for 1 month to study CI of turbulent velocity fluctuations at different scales. Basing on the analysis of return-time distribution of telegraphic approximation series, we propose to use the shape parameter of the Weibull distribution to measure CI. Observations of this parameter show that contrary to MI, CI tends to weaken as the scale increases. Besides, significant diurnal variations, showing that CI tends to strengthen during the daytime (under unstable conditions) and weaken during the nighttime (under stable conditions), are found at different observation heights. In the convective boundary layer, the mixed-layer similarity is found to scale the CI exponent better than the Monin–Obukhov similarity. At night, CI is found to vary less with height in the regime with large mean wind speeds than in the regime with small mean wind speeds, according to the hockey-stick theory.

1. Introduction

In the study of atmospheric turbulence, two kinds of intermittency are commonly discussed: the finescale intermittency (also called the inertial subrange intermittency, the internal intermittency, or the small-scale intermittency) and the global intermittency (also called the energy-containing intermittency) (Mahrt 1999). The scale of the global intermittency is normally from several minutes to hours, and the scale of the finescale intermittency is much smaller. For high-Reynolds-number atmospheric turbulence, the smallest scale of the inertial subrange is even much smaller than the resolution of the ultrasonic anemometer, which is commonly used to measure turbulent velocity in the atmospheric boundary layer (Kit et al. 2017). In this study, we focus on the finescale intermittency in the inertial subrange.

The finescale intermittency of turbulent velocity fluctuations was found to be prevalent in the atmospheric boundary layer (Schertzer et al. 1997; Boettcher et al. 2003; Vindel and Yagüe 2011; Wichert et al. 2012; Liu et al. 2019). To improve the modeling of turbulent flows (Friedrich and Peinke 1997; Baile et al. 2011; Calif and Schmitt 2012) and obtain a better understanding of particle dispersal such as air pollutants and seeds (Wei et al. 2018; Duman et al. 2016) and the propagation of light and sound wave (Sreenivasan and Antonia 1997) in the atmospheric boundary layer, a better knowledge of the finescale intermittency of atmospheric turbulence is still needed.

The intermittency of turbulent velocity fluctuations at different spatial (time) scales can be exhibited by the statistics of velocity increments with different spatial (time) lags (Frisch 1995). For the atmospheric turbulence, because of widely used single-point measurements, the statistics of velocity increments with time lags is mostly studied. The probability density function of velocity increments is found to change from the non-Gaussian long-tailed distribution at smaller lags to the Gaussian-like distribution at larger lags, which means that the turbulent velocity fluctuations are more likely intermittent at smaller scales (Castaing et al. 1990; Noullez et al. 2000; Boettcher et al. 2003; Liu et al. 2010; Liu and Hu 2013; Liu et al. 2019). Thus, the quantities quantifying deviation from the Gaussian distribution, such as the flatness (Frisch 1995; Tabeling et al. 1996; Bos et al. 2007), the average variation rate of cumulants (Malécot et al. 2000), and the Kullback–Leibler divergence...
can be considered as measures of turbulence intermittency. Besides, the \( p \)-th order moment of velocity increments, that is, the structure function, is generally found to scale with lag (Frisch 1995),

\[
|u(t + \tau) - u(t)|^p \sim \tau^\xi_p,
\]

where \( u(t) \) is the velocity at time \( t \), \( \tau \) is the time lag, usually interpreted as the time scale of turbulence eddies in turbulence theories, and \( \xi_p \), called the exponent of the structure function, is a function of \( p \). According to the K41 theory of turbulence (Kolmogorov 1941), the exponent of the structure function \( \xi_p \) is a linear function of \( p \) for turbulence with the Gaussian distribution. Thus, the nonlinearity of \( \xi_p \) is also widely used to measure turbulence intermittency (Schmitt et al. 1994; Lauren et al. 1999; Vindel and Yagüe 2011; Wei et al. 2017): the more nonlinear \( \xi_p \) is, the more intermittent turbulence becomes.

In fact, except the intermittency related to non-Gaussianity, there is another kind of intermittency related to clusterization. One can image that if extremes (defining by absolute values greater than a threshold) in a time series are moved together, the new series could be more intermittent than the original one, even if the series is Gaussian distributed. This kind of intermittency is called the clusterization intermittency, and the one related to non-Gaussianity is called the magnitude intermittency (Bershadskii et al. 2004). A famous example of clusterization intermittency is the fractional Brownian noise, which shows intermittency but is Gaussian distributed (Mandelbrot and Ness 1968; Ding and Yang 1995). Few studies have focused on the clusterization intermittency of turbulent velocity fluctuations in the atmospheric boundary layer. The analysis of vertical wind velocities shows that the boundary layer nonstationarity can affect statistics of clusterization intermittency (Li and Fu 2013). It was found that the duration between clusters of nonstationary records is nearly power-law distributed, while the duration of stationary records is stretched exponential distributed. A series of systematic studies revealed that the clusterization intermittency is related to thermal stratification and roughness (Cava and Katul 2009; Cava et al. 2012). Recently, the interaction between clusterization intermittency and submesomotions in the stable boundary layer has gained much attention (Vercauteren and Klein 2015; Mortarini et al. 2018; Cava et al. 2019). Although above works found interesting results, many questions need further investigation. We want to know how the clusterization intermittency of atmospheric turbulence varies with turbulence scales, and how this kind of intermittency evolves with time, height, and surface-layer stability. The study could inform parameterization developments in the future study.

In this paper, we try to answer these questions by analyzing velocity increments with different time lags in the surface layer. There are two main differences between ours and previous studies. First, we use velocity increments, not wavelet analysis, proper orthogonal decomposition or other methods used in previous studies, to resolve turbulence scales. The use of velocity increments follows the convention in the study of magnitude intermittency. By using velocity increments, the difference between clusterization intermittency and magnitude intermittency with scale can be compared without considering systematic deviation among different scale-resolved methods, and the interpretation of clusterization intermittency by correcting or extending magnitude intermittency theories becomes possible. Second, our work stresses the connection between long-term statistical correlation and clusterization intermittency (see section 4), and basing on the connection, we propose a new measure to quantify clusterization intermittency (see section 5) and use it to analyze the characteristics of clusterization intermittency (see section 6).

2. Data

Data were collected in September 2009 at a site located in a steppe to the northeast of Xilinhaote (44.124°N, 116.297°E), in Inner Mongolia, China. The terrain, covered by grass about a few tens of centimeters high, is nearly horizontally uniform and flat, and the effect of heterogeneous terrain on clusterization intermittency is not considered in our analysis. Turbulence data, including high-frequency velocity and temperature, were collected by four ultrasonic anemometers (Campbell CSAT-3, 20Hz) deployed at 10-, 30-, 50-, and 70-m levels on a 100-m tower. The details about the experiments are available in Lyu et al. (2018).

The quality control algorithms, proposed by Vickers and Mahrt (1997), are used to find spikes, data with the amplitude resolution problem, dropouts, data violating absolute limits, data with unphysical high-order moments, and discontinuities. Records with unphysical high-order moments or discontinuities are discarded in the analysis. If the percentage of empty values or data violating absolute limits is greater than 0.5%, the record is not used in the analysis. If the percentage of spikes is greater than 1%, if the percentage of data with the amplitude resolution problem is greater than 70%, or if the dropouts near mean and extreme are greater than 6% and 10%, respectively, the record is also not used in the analysis. Besides, records seriously contaminated by white noises in the high-frequency range are also discarded. After the quality control, about 93.2% of 10-m, 89.7% of 30-m, 93.0% of 50-m, and 89.5% of 70-m
records remain. The problematic data in the remaining records are deleted and interpolated by the cubic spline method. The fraction of problematic data is 0.06%, 0.14%, and 0.13% for horizontal velocity, vertical velocity, and temperature, respectively.

To compute the Obukhov length and the friction velocity, we transform the instrument reference frame to the streamline reference by using double rotation (Kaimal and Finnigan 1994). In this paper, we mainly analyze the wind speed, that is,

\[ u = \sqrt{u_1^2 + u_2^2 + u_3^2}, \]  

where \( u_1, u_2, \) and \( u_3 \) are velocity components in the longitudinal, lateral, and vertical directions, respectively. Records are chronologically grouped into many samples each lasting 15 min, and the average period is also set to 15 min, unless otherwise noted. The double rotation of reference frame and the analysis method in sections 4 and 5 is applied on each 15-min sample. The choice of 15-min averaging time is based on the analysis of Bou-Zeid et al. (2010), where the sensitivity of turbulence statistics to averaging time is tested and no significant effect on results with averaging time from 15 min to 1 h is found.

The wind speed increments are used to resolve turbulence scales. The time lag of wind speed increment, representing characteristic time scale of turbulent eddies, is from 0.05 to 204.8 s in this study. The averaged spectra in the unstable and stable situations are shown in Fig. 1. The spectrum in the inertial subrange is approximately proportional to \( f^{-5/3} \) (Kolmogorov 1941). The frequency of \( 1/\tau_{\text{max}} \) is marked by a broken vertical line in Fig. 1, where \( \tau_{\text{max}} = 204.8 \) s is the largest time lag of wind speed increments in this study.

The deviation from \( f^{-5/3} \) between \( 1/\tau_{\text{max}} \) and about 1/50 Hz in the stable spectrum would be related to the contamination of submesomotions. The results calculated with the largest time lag, 50 s (not shown), have been compared with those with 204.8 s, and no significant difference is found. Thus, the choice of \( \tau_{\text{max}} = 204.8 \) s does not significantly impact the conclusions in the stable situations.

3. Interpretation of clusterization intermittency

In this section, we use time series models to interpret clusterization intermittency. Figure 2 shows the time series models of magnitude and clusterization intermittency, and their comparisons with wind speed time series. The details about these models are referred to Mantegna and Stanley (1999). The Lévy flight is a model of magnitude intermittency. A series of the Lévy flight with Cauchy distributed increments is shown in Fig. 2a and the adjacent increments of this series are shown in Fig. 2b. The non-Gaussian heavy-tailed distribution of increments makes large values more likely to occur than Gaussian white noises (see the red line in Fig. 2d). The Lévy flight increment series thus seems data with spikes (Fig. 2b). The fractional Brownian motion is a model of clusterization intermittency. A series of the fractional Brownian motion with a Hurst exponent of 0.9 is shown in Fig. 2c and the adjacent increments of this series are shown in Fig. 2d. Although the fractional Brownian increments are Gaussian distributed, they seem to be more intermittent than the Gaussian white noises because large values in the former tend to appear in clusters.

To illustrate the analysis, we choose two 15-min samples respectively observed in the daytime unstable boundary layer and the nocturnal stable boundary layer during the same day. The two samples are labeled by case A (unstable, see Fig. 2e) and case B (stable, see Fig. 2g); their flow characteristics are listed in Table 1. Similar to the fractional Brownian motion increments, large values of wind speed increments tend to appear in clusters, especially for the case in the unstable boundary layer (see Figs. 2f,h). In fact, atmospheric turbulence is a mixture of clusterization intermittency and magnitude intermittency. The magnitude intermittency has been extensively discussed. In this paper, we mainly focus on clusterization intermittency and propose a method to extract and quantify clusterization intermittency.

4. Long-term correlation and clusterization intermittency

From the view of the random process, Bunde et al. (2005) found that the clusterization intermittency is
intimately related to the long-term correlation in many numerical and natural signals. The long-term correlation means that the autocorrelation function, measuring correlation between two different times in a time series, decreases much slower than the exponential-like drop-off. In this section, we analyze the autocorrelation function of turbulent wind speed fluctuations at different scales and discuss the relationship between the long-term correlation and clusterization intermittency.

**a. Autocorrelation function**

The multifractal random walk (MRW), proposed by Bacry et al. (2001), is widely used to describe the time series with small-scale multiscaling, non-Gaussianity,
intermittency, and long-term correlation. Recently, the turbulent velocity fluctuations in the atmospheric boundary layer were found to be well described by the MRW generated by a lognormal cascade process (hereafter the lognormal MRW) (Liu et al. 2019). We use this model to discuss the autocorrelation function of turbulent velocity fluctuations. The details about the universal MRW theory are referred to Muzy and Bacry (2002) and Bacry and Muzy (2003).

Assuming that the velocity increments follow a lognormal MWR with a Hurst exponent of 0.5, the $p$th-order autocorrelation function of absolute increments at scale $\tau$ is

$$C_p(t, \tau) = \langle \delta u_x(t)^p \delta u_x(0)^p \rangle \sim \left( \frac{\tau}{T} \right)^{2\xi_p \mu} \left( \frac{1}{T} \right)^{-\mu \xi_p/4},$$

which holds when $\tau \ll t \ll T$ and $\mu > 0$ (Bacry et al. 2001). The symbol $\delta u_x(t)$ denotes the velocity increment between time $t + \tau$ and $t$, and its $p$th-order absolute moments scale with $\tau$:

$$\langle \delta u_x(t)^p \rangle \sim \tau^{\xi_p},$$

where $\xi_p$ is the exponent of structure function with a Hurst exponent of 0.5, that is,

$$\xi_p = \frac{p}{2} \left( 1 + \frac{\mu}{2} \right) - \frac{\mu p^2}{8}.$$  

The parameter $\mu$ in Eq. (3) is the magnitude intermittency exponent, which defines whether the increments are Gaussian ($\mu = 0$) or non-Gaussian ($\mu > 0$).

Generally, Liu et al. (2019) have verified the Hurst exponent in the lognormal MRW is apparently less than 0.5 for turbulent wind speeds in the atmospheric boundary layer. Thus, Eq. (3) cannot be used to describe the autocorrelation properties of turbulent wind speed data. Mathematically, it is difficult to derive the $p$th autocorrelation function of the lognormal MRW with an arbitrary Hurst exponent. If we replace the structure function exponent $\xi_p$ with the Hurst exponent $H = 1/2$ in Eq. (3) by an exponent $\xi_{p,H}$ with an arbitrary Hurst exponent $H$, and replace the magnitude intermittency exponent $\mu$ by an arbitrary parameter $\gamma$, an assumption of autocorrelation function of turbulent wind speeds is obtained:

$$C_p(t, \tau) \sim \left( \frac{\tau}{T} \right)^{2\xi_{p,H} \mu} \left( \frac{1}{T} \right)^{-\gamma \mu^2},$$

where $\tau \ll t \ll T$, $\gamma > 0$, and the Hurst exponent $H \in (0, 1)$. It has been strictly proved that

$$\xi_{p,H} = p(1 - m) - \frac{p^2 H^2}{2},$$

with an arbitrary parameter $m \in R$ (Muzy and Bacry 2002; Bacry and Muzy 2003). The direct verification of Eq. (6) is not easy. Due to limited data, the autocorrelation function would be scattered at large $t$, where Eq. (6) is assumed to hold. We thus verify it indirectly. According to Eq. (6),

$$G_p(t) = \frac{C_p(t, \tau)}{(t/T)^2\xi_{p,H}} \sim \left( \frac{1}{T} \right)^{-\gamma \mu^2},$$

Then, the function $G_p(t)$ is independent of $\tau$, and

$$G_p \sim \left( \frac{1}{T} \right)^{-\gamma (p-1)(p+1)\mu^2} \sim G_{p-1}^{\xi_{p-1}},$$

where

$$\xi_p = \left( \frac{p}{p - 1} \right)^2.$$  

The verification of Eq. (6) is thus replaced by the verification of Eq. (9).

Figure 3 shows the verification of Eq. (9) for case A and case B. We compute $C_p(t, \tau)$ at scales $\tau = 0.05, 10, 50$, and $100$ s and find that for a fixed $p$, if $C_p(t, \tau)$ is divided by $(t/T)^{2\xi_{p,H}}$, all data will collapse into a single curve in the plot of $G_p$ versus $G_{p-1}$. We then fit the data at $\tau = 0.05$ s with a linear function $\log G_p = \xi_p \log G_{p-1} + B$ in the log–log plot, where $B$ is the only fitting parameter, and find that Eq. (10) shows a good fit to the data. We have verified all samples and found that $G_p$ is independent of scale $\tau$ for the most samples, and more than 75% coefficients of determination are greater than $0.9$ for $p = 2, 3, \ldots, 10$, as shown in Fig. 3. Thus, Eq. (9) is true for most samples, and Eq. (6) is indeed a reasonable assumption for the autocorrelation function of turbulent velocity fluctuations in the atmospheric boundary layer. The power exponent of autocorrelation function in Eq. (6) reflects how slowly correlation decays with time in a time series. In the next section, the analysis of this exponent reveals that long-term correlations exist in the time series of wind speed increments and is related to

<table>
<thead>
<tr>
<th>Case A</th>
<th>Case B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean wind speed (m s$^{-1}$)</td>
<td>3.68</td>
</tr>
<tr>
<td>Mean temperature (°C)</td>
<td>14.66</td>
</tr>
<tr>
<td>Friction velocity (m s$^{-1}$)</td>
<td>0.18</td>
</tr>
<tr>
<td>Stability z/L</td>
<td>-14.02</td>
</tr>
<tr>
<td>Observation height (m)</td>
<td>30</td>
</tr>
<tr>
<td>Observation time (LST)</td>
<td>1200</td>
</tr>
</tbody>
</table>

Table 1. Flow characteristics of two samples shown in Figs. 2e and 2g. The surface-layer stability is measured by $z/L$, where $z$ is the observation height and $L$ is the Obukhov length.
clusterization intermittency (CI), which inspires us to propose a method to quantify clusterization intermittency in section 5.

b. Fluctuation analysis

In this section, we discuss the parameter $g$ in Eq. (6). The first-order autocorrelation function $C_1(t, \tau)$ shows the correlations in the time series of absolute increments. We have shown that $C_1(t, \tau) \sim \tau^{-g}$ in section 4a, where the parameter $g$ is independent of scale $t$. The parameter $g$ is called the correlation exponent, because it defines the correlation type. If $g$ lies in $(0, 1)$, $\int_0^\infty C_1(t, \tau) \, dt = \infty$, and the time series of absolute increments are thus long-term correlated (Beran 1994), while for the independent or short-term correlated absolute values, $\int_0^\infty C_1(t, \tau) \, dt < \infty$. Because of limited data, the autocorrelation function is usually scattered at large $t$. Thus, it is difficult to verify whether $g$ lies in $(0, 1)$ by analyzing autocorrelation function. Many methods, although controversy exists, have been proposed to estimate $g$ (Beran 1994; Taqqu et al. 1995; Franzke et al. 2012). The fluctuation analysis, proposed by Peng et al. (1992), is the one widely used in the literature.

The fluctuation analysis is described as follows. The profile of a discrete time series $x_i (i = 1, 2, \ldots, N)$ is defined by

$$Y_n = \sum_{i=1}^n x_i,$$  \hspace{1cm} (11)

where $n = 1, 2, \ldots, N$. Then, the profile increment with lag $s$ is obtained by $\delta Y_s = Y_{n+s} - Y_n$. If the standard deviation of $\delta Y_s$ called the fluctuation function, scales with large values of $s$, that is,

$$F(s) = (\delta Y^2 - \bar{Y}^2)^{1/2} \sim s^\alpha,$$  \hspace{1cm} (12)

where $\alpha$ is called the fluctuation exponent, one can prove that (Kantelhardt et al. 2001)

$$\alpha = 1 - \frac{\gamma}{2},$$  \hspace{1cm} (13)

where $\gamma$ is the correlation exponent of the time series $x_i$. Equation (13) shows that if the time series is long-term correlated, $\gamma \in (0, 1)$, and thus $\alpha \in (0.5, 1)$. If the time series is independent, $\alpha = 0.5$.

The fluctuation analysis is used to analyze the absolute velocity increments of case A and case B (see circles in Figs. 4a,b). Because the correlation exponent is independent of scale $t$, we just choose $t = 0.05$ s in the analysis. Results show that the fluctuation exponent is significantly greater than 0.5 but less than 1 for each case. Similar results are also found for most of the other samples (see “raw” in Fig. 4c). We then randomly shuffle the data and perform the same fluctuation analysis for the shuffled data. The randomly shuffling removes clusterization intermittency from data, but does not affect the underlying distribution and thus the magnitude intermittency. For case A and case B, the fluctuation exponents of the shuffled data are found to be approximately 0.5 (see squares in Figs. 4a,b). The analysis has been done for the whole samples, and we find that the fluctuation exponents of shuffled data are also around 0.5, similar to the fluctuation exponents of independent Gaussian samples (see Fig. 4c). We thus conclude that the absolute velocity increments in the atmospheric boundary layer are generally long-term correlated, and the clusterization intermittency is intimately related to the long-term correlation. If the clusterization intermittency disappears, the correlation in the time series changes from the long-term type to the short-term or the independent type.

The magnitude intermittency exponent $\mu$, obtained by fitting data to Eq. (7), is plotted versus the correlation

Unauthenticated | Downloaded 06/21/22 09:22 AM UTC
exponent $\gamma$, obtained by the fluctuation analysis. Results show that deviating from the prediction of Eq. (3), $\gamma$ is almost independent of $\mu$ (see Fig. 4d). This indicates that the clusterization intermittency is unrelated to the magnitude intermittency. They are two kinds of intermittency, not just two facets of the same intermittency. It means that clusterization and magnitude intermittency could be caused by two kinds of physical mechanisms. Distinguishing clusterization and magnitude intermittency is a first step to understand the underlying intermittency mechanisms.

5. Clusterization intermittency exponent

In this section, we discuss the clusterization intermittency exponent that can measure this kind of intermittency. The zero-crossing statistics of the telegraphic approximation of a time series is commonly used to design clusterization intermittency exponent (Bershadskii et al. 2004; Sreenivasan and Bershadskii 2006; Cava et al. 2012; Li and Fu 2013; Cava et al. 2019). We briefly review this method and then discuss its flaws (section 5b). Basing on above analysis of long-term correlation, we propose a new, better exponent (section 5c).

a. Telegraphic approximation

The magnitude and clusterization intermittency always coexist in a time series. Before defining measures, one should separate the two kinds of intermittency. This can be simply done by the telegraphic approximation (TA) of a time series $x(t)$ (Bershadskii et al. 2004):

$$\text{TA}(x) = \frac{1}{2} \left[ \frac{x(t) - h}{|x(t) - h|} + 1 \right].$$

(14)

By definition, the TA transforms the values of $x(t)$ greater than $h$ to 1 and the others to 0 (see Fig. 5). In practice, the threshold $h$ should be set carefully. If $h$ is set to be very large, the number of events greater than
the threshold are very small and TA(x) is sparse; if \( h \) is set to be very small, TA(x) is occupied by a few long events. To avoid poor statistics in both cases, the threshold should be set to some intermediate value. The choice of threshold is further discussed in section 5c.

b. Zero-crossing density

The zero crossing is a time point where TA(x) switches from 1 to 0 or from 0 to 1 (see Fig. 5). The zero-crossing density \( n_t \) is computed by counting the number of zero crossings in a running time window of size \( l \) and dividing this number by the total number of data in the window. The standard deviation of zero-crossing density \( n_t \) is usually assumed to be a power-law function of window size \( l \); that is,

\[
(n_t - \mu_t)^{1/2} \sim l^{-d},
\]  

(15)

where \( \mu_t \) is the mean of \( n_t \) and \( d \) is a measure of clusterization intermittency. The smaller than 1/2 the \( d \) is, the more clusterization intermittency becomes.

There are some flaws in the use of zero-crossing density to measure clusterization intermittency. First, in this method, \( d \) is assumed to equal 1/2 for any time series without clusterization intermittency. This assumption may be not generally true. The mathematical derivation of the zero-crossing statistics is very difficult. As far as we known, \( d \) is strictly proved to equal 1/2 just for the stationary Gaussian process with weak correlation (Kratz 2006), taken as an example without clusterization intermittency. Whether this conclusion is true for the non-Gaussian process is still an open problem. Because velocity increments at small scales are generally non-Gaussian, it seems unreasonable to measure clusterization intermittency by qualifying deviation of \( d \) from 1/2. Second, for the Gaussian process with long-range dependence, it has been proved that Eq. (15) is indeed true for very large window size (Slud 1994). However, it also indicates that it is difficult
obtain an accurate estimate of $d$, because the statistics is poor with large window size. Therefore, we propose a better measure, namely the return-time distribution, to quantify clusterization intermittency.

c. Return-time distribution

The return time, also called the waiting time or the interpulse period, is the duration between events with values greater than a threshold (see Fig. 5). For the statistically independent time series, the return time is strictly proved to be exponential distributed (Karlin and Taylor 1975; Liu et al. 2014):

$$f(\Omega) = \exp(-\Omega),$$

(16)

where $\Omega$ is the normalized return time, and $f(\Omega)$ is the probability density function of $\Omega$. The normalized return time $\Omega$ is defined by $r/\bar{r}$, where $r$ is the return time, and $\bar{r}$ is its mean. Unlike the zero-crossing statistics, the exponential distribution is true for any statistically independent random process, whether it is Gaussian or non-Gaussian. Thus, the exponential distribution can be reasonably considered as a benchmark to measure clusterization intermittency.

We here return to the problem of threshold choice. For a randomly shuffled time series, having no correlations and clusterization intermittency (see section 5b), the return-time distribution must be the exponential distribution. However, an unsuitable choice of threshold could cause poor statistics, and thus, the exponential distribution would not be observed, even for the...
independent time series. We randomly shuffle the absolute wind speed time series, choose a threshold, and make the telegraphic approximation of this series, and then calculate its return-time distribution. If the measured distribution can be well fitted by the exponential distribution, the threshold is considered to be suitable. The procedure of threshold choice is illustrated in Fig. 6, where the coefficients of determination $R^2$ between the measured distribution and the exponential distribution at different scales are shown. As discussed in section 5a, $R^2$ declines for both very small and very large thresholds. Because sample size decreases with increasing scale, $R^2$ declines earlier at larger scales when the threshold is very large. A suitable threshold can be chosen in the intermediate plateau range, where $R^2 > 0.95$. For case A, the threshold can be set between 1.5 and 4 times of mean absolute velocity increment for all scales. For case B, the threshold can be set between 2 and 3 times of mean absolute velocity increment for all scales.

We have analyzed all samples and find that if the threshold is set between 2 and 3 times of mean absolute velocity increment for all scales, more than 90% samples have $R^2 > 0.95$. In the following analysis, the threshold is set to 2.5 times of mean absolute velocity increments. With this threshold, about 95.8% samples have $R^2 > 0.95$.

Whether there is a universal and analytical return-time distribution for the long-term correlated time series is still an open problem. Models have been proposed in various applications. Here, we introduce some commonly used models. All of them include the exponential distribution as a specific case for the statistically independent time series.

(i) The stretched exponential distribution (Bunde et al. 2005; Altmann and Kantz 2005):

$$f(\Omega) = a_\kappa \exp[-(b_\kappa \Omega)^\kappa], \quad (17)$$

where

$$a_\kappa = b_\kappa \frac{\kappa}{\Gamma(1/\kappa)}, \quad b_\kappa = \frac{2^{2/\kappa} \Gamma\left(\frac{2 + \kappa}{2\kappa}\right)}{2 \sqrt{\pi}}, \quad (18)$$

for $0 < \kappa \leq 1$. The tail distribution of the stretched exponential distribution is

$$P(\Omega > \omega) = \int_\omega^\infty f(\Omega) \ d\Omega = \frac{a_\kappa}{b_\kappa \kappa} \Gamma\left(\frac{1}{\kappa} \omega b_\kappa \right). \quad (19)$$

If $\kappa = 1$, the stretched exponential distribution is just the exponential distribution for the statistically independent time series. By numerical analysis, Eichner et al. (2007) found that the stretched exponential distribution can only fit large return time, while small return time is well fitted by the power-law distribution. A similar result is also found for atmospheric...
turbulence (Schmitt 2007; Cava et al. 2012). Thus, the Weibull distribution, that is, the stretched-exponentially truncated power-law distribution, can also be considered as a model of return-time distribution.

(ii) The Weibull distribution (Laherrère and Sornette 1998):
\[
f(\Omega) = k\Omega^{-k} \exp\left[-\left(\Omega/\lambda\right)^k\right],
\]
where \( \lambda = 1/\Gamma(1 + 1/k) \) for \( 0 < k \leq 1 \). The tail distribution of the Weibull distribution is
\[
P(\Omega > \omega) = \int_\omega^\infty f(\Omega) d\Omega = \exp\left[-\left(\omega/\lambda\right)^k\right].
\]
If \( k = 1 \), the Weibull distribution is just the exponential distribution for statistically independent time series.

(iii) The \( q \)-exponential distribution (Ludescher et al. 2011; Manshour et al. 2016; Carbone et al. 2018):
\[
f(\Omega) = (2 - q)\beta[1 - (1 - q)\beta x]^{1/(1-q)},
\]
where \( \beta = 1/(3 - 2q) \) for \( 1 \leq q < 3/2 \). The tail distribution of the \( q \)-exponential distribution is
\[
P(\Omega > \omega) = \int_\omega^\infty f(\Omega) d\Omega = \left[1 - \left(\frac{1}{q'} - 1\right)\beta \omega\right]^{1/(1-q')},
\]
where \( q' = 1/(2 - q) \). The \( q \)-exponential distribution reduces to the exponential distribution as \( q \to 1 \).

6. Characteristics of clusterization intermittency

In this section, we present the characteristics of clusterization intermittency of turbulent velocity fluctuations in the atmospheric boundary layer. As discussed in the above section, the clusterization intermittency is measured by the clusterization intermittency exponent \( \kappa \), defined by the shape parameter of the
Weibull distribution. The smaller $k$ means more significant clusterization intermittency.

a. Diurnal variation

For case A and case B, the exponent $k$ is plotted as a function of scale $\tau$ shown in Fig. 9a, where it is found that $k$ decreases with increasing scale, and at small scales, $k$ decreases logarithmically with scale. We have performed the similar analysis on all samples and found that the clusterization intermittency exponent has a decreasing trend with increasing scale (see Fig. 9b). This indicates that the clusterization intermittency of turbulent velocities is normally stronger at larger scales, contrary to the magnitude intermittency that is normally
stronger at smaller scales. To estimate the overall clusterization intermittency for a sample, we average the clusterization intermittency exponent \( k \) over scales between \( t_1 \) and \( t_2 \):

\[
\bar{k} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} k(\tau) \, d\tau,
\]

where \( t_1 \) and \( t_2 \) are respectively set to the minimum (0.05 s) and maximum scales (204.8 s) shown in Fig. 9.

The time variation of \( \bar{k} \) at 10, 30, 50, and 70 m is shown in Fig. 10. From this figure, significant diurnal variation of clusterization intermittency is found at different heights. Normally, \( \bar{k} \) is smaller in daytime and larger in nighttime, which indicates that clusterization intermittency is more significant in daytime than in nighttime. The buoyancy is usually help for turbulence generation in daytime but usually depress turbulence generation in nighttime. The turbulence kinetic energy (TKE) is typically larger in daytime (Stull 1988). According to the K41 theory, the more energy turbulence extracts from large-scale eddies, the more energy finescale eddies gain in the cascade process before energy dissipates (Kolmogorov 1941). The gained energy could entail the presence of finescale coherent structures, such as vortex tubes and strain sheets (Wilczek 2015; Dong et al. 2020). Thus, clusterization intermittency manifesting itself in finescale coherent structures is typically larger in daytime than in nighttime. The diurnal variation indicates that clusterization intermittency is related to boundary layer stability. Here, we compute the variation of \( \bar{k} \) with the stability parameter \( z/L \), where \( z \) is the observation height and \( L \) is the Obukhov length. The results are shown in Fig. 12, where it is found that \( \bar{k} \) is normally smaller in the unstable boundary layer than in the stable boundary layer, and the clusterization intermittency is thus more significant under unstable conditions. A similar result is also found for the intermittent heat flux (Katul et al. 1994). In the unstable boundary layer, a dropping trend of \( \bar{k} \) is observed from neutral to unstable condition and then \( \bar{k} \) is almost constant. In the stable boundary layer, data are very scattered, and we cannot find an evident trend of \( \bar{k} \).

The variations of \( \bar{k} \) with height are analyzed over each 15-min epoch. The standard deviation of \( \bar{k} \) over the four observational heights is computed by

\[
\text{SD}[\bar{k}] = \sqrt{\frac{1}{3} \sum_{n=1}^{4} (\bar{k}_n - \langle \bar{k} \rangle_z)^2}.
\]
where $k_{zn}$ is the scale-averaged clusterization intermittency exponent at height $z_n$, and $\langle \cdot \rangle_z$ denotes the arithmetic mean over the four observational heights. It is found that 75% 15-min samples at daytime have $\text{SD}[k] < 0.03$ (Fig. 16). The slight variation of $k$ with height could be related to nonlocal turbulence generated by bulk buoyancy-driven instability in the very unstable boundary layer or the bulk wind shear in the nearly neutral boundary layer. Figure 13 shows two cases in the very unstable and the nearly neutral situations, respectively. In the very unstable case, ramp structures in the temperature time series (see $T$ in Fig. 13b) and large updrafts and downdrafts in the vertical velocity time series (see $u_3$ in Fig. 13b) are frequently observed, when thermals pass the observational tower (Garratt 1992). In the nearly neutral case, there are no evident ramp structures in the temperature time series (see $T$ in Fig. 13a) and updrafts and downdrafts appear less frequently in the vertical velocity time series (see $u_3$ in Fig. 13a). For the two cases, the clusterization intermittency is evident at all heights (see $d u_3$ in Fig. 13), and $k$ is found to vary slightly along height (see the black and green lines in Fig. 18). The values of $\text{SD}[k]$ in both cases are smaller than 0.02.

The fact that $k$ varies slightly with height indicates that observation height is not a very relevant parameter in the similarity analysis of clusterization intermittency in cases with nonlocal flow instabilities. In the convective surface layer, the Monin–Obukhov similarity is usually replaced by the mixing-layer similarity, where the observational height $z$ is replaced by the boundary layer height $h$ (Panofsky et al. 1977; Holtslag and Nieuwstadt 1986). We analyze the data with $z/L < -1$ and find that the dependence of clusterization intermittency exponent with $h/L$ is less scattered than that with $z/L$ (Fig. 14). The boundary layer height is estimated by the wavelength of maximum energy in the vertical velocity spectrum (Kaimal et al. 1982; Liu et al. 2011). The data in Fig. 14 are fitted with a power function by using the nonlinear least squares method. The fitting results show that

$$ k\left(\frac{h}{L}\right) = 0.17 \left(\frac{0.17}{-h/L}\right)^{0.10} + 0.19, \quad (26) $$

and

$$ k\left(\frac{z}{L}\right) = 0.54 \left(\frac{0.54}{-z/L}\right)^{5.43} + 0.28. \quad (27) $$

The coefficients of determination of the fittings with Eqs. (26) and (27) are 0.21 and 0.006, respectively.

**c. The hockey-stick theory**

The turbulence characteristics in the nocturnal stable boundary layer are usually analyzed by separating between two regimes, typically denoted as weakly stable and very stable (Mahrt 2014). One approach to delineate regimes is based on the hockey-stick theory (Sun et al. 2012). The theory states that the two regimes could be distinguished by the relationship between mean wind speed $\bar{u}$ and square root of turbulence kinetic energy $V_{\text{TKE}}$. The very stable regime where $V_{\text{TKE}}$ is weak and increases slightly with $u$ is called regime 1 and the weakly stable regime where $V_{\text{TKE}}$ increases rapidly with $u$ after $u$ exceeds a threshold value is called regime 2.

![Fig. 16. A boxplot of standard deviation of scale-averaged clusterization intermittency $k$ over height in daytime and in regime 1 and regime 2 in nighttime.](image-url)
to 0600 LST is shown in Fig. 15, and the threshold velocity is listed in Table 2. The two regimes can be clearly observed at all heights, and the threshold wind speed increases with height, as found in Sun et al. (2012).

The variation of $\overline{k}$ with height is analyzed in the two regimes. It is found that $\overline{k}$ varies more slightly in regime 2 than in regime 1, where 75% SD[$\overline{k}$] in regime 1 are larger than that in regime 2, and very large SD[$\overline{k}$] are more likely to appear in regime 1 (Fig. 16). The results coincide with the hockey-stick theory. According to this theory, turbulence in regime 1 is mainly generated by local wind shear and occurs frequently at some specific heights, while turbulence in regime 2 is mainly generated by bulk wind shear and occurs at most of observational heights (Sun et al. 2012).

Two cases in regime 1 and regime 2 are shown in Figs. 17a and 17b, respectively. In the case of regime 2, one can see that clusterization intermittency is evident at the four heights (see $\delta u$ in Fig. 17b), and $\overline{k}$ varies slightly with heights (see the blue line in Fig. 18). In the case of regime 1, clusterization intermittency is evident at the height of 10 and 30 m (see $\delta u$ in Fig. 17a). This could be related to the occurrence of jet-like wind shear below 30 m (see $u$ in Fig. 17), as found in (Cava et al. 2019).

At higher levels, the time series look more like Gaussian white noises. In this case, $\overline{k}$ is smaller at 10 and 30 m and becomes larger at higher heights (see the red line in Fig. 18). This example shows the possibility of the link between the jet-like event and clusterization intermittency. A better understanding of this link and the height dependence of clusterization intermittency would be obtained in the future research.

7. Conclusions

In this paper, we analyzed the multiscale clusterization intermittency of atmospheric turbulence by using a 20 Hz ultrasonic dataset lasting for one month. The turbulent velocity fluctuations at different scales, defined by the absolute velocity increments at different time lags, is found to be long-term correlated in statistics. By the fluctuation analysis, the clusterization intermittency, which refers to the fact that large velocity increments tend to appear in clusters, is found to be related to the long-term correlation. This finding reveals that clusterization intermittency is different to magnitude intermittency: the latter is related to non-Gaussianity,
while the former is not related to non-Gaussianity. Even the non-Gaussian random process, for example, the fractional Brownian motion (Mandelbrot and Ness 1968; Ding and Yang 1995), may have clusterization intermittency. Thus, the magnitude intermittency exponent cannot be used to measure clusterization intermittency.

Basing on the return-time statistics, we propose a new exponent to measure clusterization intermittency. The return-time statistics include return-time distributions of the telegraphic approximation of absolute velocity increments and the corresponding randomly shuffled increments respectively at different scales. Comparing the return-time distribution of shuffled increments with the exponential distribution, that is strictly proved to be true for any statistically independent time series, one can set an optimal threshold in the telegraphic approximation. For the return-time distribution of raw increments, we found that it can be well fitted by the Weibull distribution. The shape parameter of the Weibull distribution describes the occurrence probability of long return times, and thus can be used to measure clusterization intermittency. The observations of this parameter for all samples indicate the following: 1) Contrary to magnitude intermittency, clusterization intermittency weakens as scale decreases; 2) diurnal variation of clusterization intermittency is evident at different heights. Normally, the intermittency strengthens at daytime and weakens at nighttime; 3) clusterization intermittency is normally stronger in the unstable boundary layer than in the stable boundary layer. In the very unstable boundary layer, clusterization intermittency scales better by the mixed-layer similarity, where the boundary layer height replaces the observational height in the Monin–Obukhov similarity; 4) according to the hockey-stick theory, data in the nocturnal stable boundary layer are divided into two regimes. Clusterization intermittency varies more slightly with height in the regime with large mean wind speed than in the regime with small mean wind speed. The results coincide with the hockey-stick theory, where turbulence in the regime with large mean wind speed is mainly generated by bulk wind shear, and turbulence in the regime with small mean wind speed is mainly generated by local wind shear.

Our work indicates that finescale clusterization intermittency is intimately related to the boundary layer instability, which transfers energy from large-scale motions to finescale motions in the cascade process. In the future, we will combine data analysis and numerical model to build up some possible connections between large-scale motions and clusterization intermittency. This will be helpful to understand the nature of clusterization intermittency and design new parameterizations of intermittent turbulence to improve numerical model (i.e., large-eddy simulation) of boundary layer motions.

Acknowledgments. This work is supported by the National Nature Science Foundation of China (41675012) and the National Key Research and Development Program of China (2016YFC0208802 and 2017YFC0209605).

REFERENCES
Blender, R., K. Fraedrich, and F. Sienz, 2008: Extreme event return times in long-term memory processes near 1/f,


—, —, X.-L. Cheng, and L.-L. Song, 2010: Probability density functions of velocity increments in the atmospheric boundary...


