The properties of broad absorption line outflows based on a large sample of quasars

Zhicheng He$^{1,2,3,*}$ Tinggui Wang$^{1,2,†}$ Guilin Liu$^{1,2}$ Huiyuan Wang$^{1,2}$ Weihao Bian$^{4}$ Kirill Tchernyshyov$^{3}$ Guobin Mou$^{1,2,5}$ Youhua Xu$^{6}$, Hongyan Zhou$^{1,2}$, Richard Green$^{7}$ and Jun Xu$^{1,2}$

$^{1}$CAS Key Laboratory for Research in Galaxies and Cosmology, Department of Astronomy, University of Science and Technology of China, Hefei, Anhui 230026, China
$^{2}$School of Astronomy and Space Science, University of Science and Technology of China, Hefei 230026, China
$^{3}$Department of Physics & Astronomy, Johns Hopkins University, Bloomberg Center, 3400 N. Charles St., Baltimore, MD 21218, USA
$^{4}$Department of Physics and Institute of Theoretical Physics, Nanjing Normal University, Nanjing 210023, China
$^{5}$School of Physics and Technology, Wuhan University, Wuhan 430072, China
$^{6}$CAS Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Beijing 100012, China
$^{7}$Steward Observatory, University of Arizona, Tucson, AZ, 85721-0065, USA

E-mail: zcho@ustc.edu.cn
E-mail: twang@ustc.edu.cn
E-mail: glliu@ustc.edu.cn

BAL outflows can be estimated from the distance ($R$) of the absorption lines. The typical distance in this sample is tens of parsec, which is beyond the theoretical predicted location ($R$ of the distribution of $R$) between the outflowing gas and the galaxy center, the total hydrogen column density $N_{H}$ and the fraction $\Omega$ of the solid angle subtended by the outflowing gas. Because the ionization parameter $U_{H}$ of the plasma is inversely proportional to the product of hydrogen number density $n_{H}$ and $R^{2}$, the equation $U_{H} \propto 1/(n_{H}R^{2})$ can be obtained by measuring $U_{H}$ and $n_{H}$. In general, $n_{H}$ can be determined from the absorption lines of the excited states of ions (e.g., Fe II*, Si II*, S IV*), but this method is hindered by line blending and is therefore only applicable to quasars with relatively narrow absorption lines. During the last decade or so, outflow distances have been measured for only a dozen of individual quasars using this method$^{11,23}$, while the distributions of primary properties of BAL outflows remain in unknown. In this work, we present a novel method to determine $R$ by examining the variability of BAL troughs.

BAL troughs vary on timescales ranging from several days to years$^{6,23}$. There are two possibilities of BAL variability, i.e., the tangential movement of the absorbing gas and changes in the ionizing radiation incident on the gas. In the latter case, the variability timescales set a constraint on the ionization or recombination timescales, which depend solely on the incident ionizing continuum and gas density. A series of statistical investigations on BAL variability have been performed using the large multi-epoch spectroscopic dataset of Sloan Digital Sky Surveys (SDSS) DR10 and DR12 in ref$^{24,25}$. Ref$^{26}$ found that the majority of BAL variability is driven by variation of the ionizing continuum, and ref$^{27}$ further pinned down this fraction to be at least 80% of BALs. On the basis of the above works, we focus on BAL variability driven by ionizing continuum variation and use it to derive the primary physical properties of BAL outflows.

The ionization state of a gaseous outflow demands for a period of time (the recombination timescale, $t_{rec}$) to respond to changes in the ionizing continuum. The gas ionization is connected to the average intensity of ionizing continuum over $t_{rec}$, while the change in the average intensity of ionizing continuum between a pair of observations decreases when the $t_{rec}$ increases (see Supplementary Figure 1 in Methods for details). In principle, an absorption line should only vary from one observation to another if the $t_{rec}$ is shorter than the timescale on which the ionizing continuum varies and the time interval $\Delta T$ between the observations$^{28}$.

The underlying distribution of the recombination timescale $t_{rec}$ (hereafter abbreviated as $t_{r}$) of the outflow gas, i.e., $f(t_{r})$, determines the fraction $F(\Delta T)$ of variable BAL troughs that can be detected from a given large BAL quasar sample. We denote the probability of detecting the variability of a BAL with recombination timescale $t_{r}$ at $\Delta T$ as $p(t_{r}, \Delta T)$. Considering the ideal case, the BAL variability can (not) be detected when the recombination timescale $t_{r}$ is shorter (longer) than the observational time interval $\Delta T$. In this case, $p(t_{r}, \Delta T)$ is a step function, i.e., $p = 1$ for $t_{r} \leq \Delta T$ and $p = 0$ for $t_{r} > \Delta T$. Hence, we
can describe the ideal fraction curve as follows:

$$F_{\text{ideal}}(\Delta T) = \int_0^{\infty} p(t, \Delta T) f(t) \, dt = \int_0^{\Delta T} f(t) \, dt.$$  (1)

In reality, $p(t, \Delta T)$ is not a standard step function, and $F(\Delta T)$ depends not only on $\Delta T$ but also on the detection threshold. Given a certain detection threshold, we can write the actual fraction curve as the ideal fraction curve multiplied by a correction factor, i.e., $F(\Delta T) = K(\Delta T) F_{\text{ideal}}(\Delta T)$. Note that because $p(t, \Delta T)$ is not a standard step function, the actual $F(\Delta T)$ may deviate from $F_{\text{ideal}}(\Delta T)$, even in the unrealistic case of complete detection of all variability.

Measuring $F(\Delta T)$ is sufficient for deriving the underlying distribution $f(t)$. If $K(\Delta T)$ is constant, the underlying recombination timescale distribution $f(t)$ can be readily obtained by taking the derivative of $F(\Delta T)$ with respect to $\Delta T$. If $K(\Delta T)$ is not constant, the derivative of $F(\Delta T)$ will deviate from $f(t)$. However, in the Methods, we have performed a simulation showing that this deviation to be likely negligible for our sample.

$F(\Delta T)$ can be practically measured using the following method. Assuming that we have a sample of $N$ multiply-observed BAL quasars which have already been sorted according to the rest time interval between each pair of observations, we divide these $N$ quasars into $B$ bins, each of which contains approximately the same number of objects, $\Delta T_i$ is the mean time interval between the pair of observations for all quasars contained in the $i$-th bin. The fraction $F_i \equiv F(\Delta T_i)$ is measured to be $F_i = k_i/N_i$, where $k_i$ is the number of quasars with variable BAL, $N_i$ is the number of quasars contained in the same bin, and we have dropped $\Delta T_i$ for simplicity. Assuming that the detections of BAL variability are mutually independent in the $i$-th bin, the probability of detecting BAL variability follows a binomial distribution. The standard deviation of $k_i$ is $\sigma_{k_i} = \sqrt{k_i/(N_i-k_i)}$. As a result, one can use $F_i = \sqrt{F_i(1-F_i)/N_i}$ as an estimate of the measurement error of $F_i$. Due to the incomplete independence of spectral pairs, the actual uncertainty of the variable BAL fraction is larger than that estimated from a binomial distribution.

The $F(\log_{10} \Delta T)$ curve measured from the SDSS sample (see Supplementary Figure 2) is presented in panel a of Figure 1. The fits of C IV BAL troughs and the identification of C IV/BAL trough variabilities are shown in Methods and Supplementary Figure 3. The distributions of the parameter of the C IV BAL troughs are shown in Supplementary Figure 4. We assume that the logarithmic recombination time distribution is a Gaussian function $G(t, \sigma_t, \mu_t)$, where $\mu_t$ and $\sigma_t$ are the mean and standard deviation of the Gaussian distribution, respectively. Thus, the cumulative distribution function (CDF) of the Gaussian distribution is used to model the fraction curve,

$$F(t) = p_0 \left[ 1 + \text{erf} \left( \frac{t - \mu_t}{\sqrt{2} \sigma_t} \right) \right],$$  (2)

where $t \equiv \log_{10} \Delta T$ is the logarithmic time interval between each pair of observations, $\text{erf}(t) = 1/\sqrt{\pi} \int_{-t}^{t} e^{-x^2} \, dx$ is the error function, and $p_1 = t_t, p_2 = t_s$, i.e., the mean and standard deviation of the Gaussian distribution. The reduced $\chi^2$ of the best-fit model is 1.18. The best-fit mean and standard deviation of the Gaussian distribution are $t_s = 0.36 \pm 0.14$ and $t_t = 1.01 \pm 0.22$. The recombination timescale distribution is shown in panel b of Figure 1. Taking the above recombination timescale distribution as the input Gaussian, and employing the photoionization model (Supplementary Figure 5), we conduct a simulation test (see Methods) and generate the recovered Gaussians. The input and recovered Gaussians (Supplementary Figure 6) are consistent within 1σ uncertainty.

Ref.22 found that C IV, Si IV, and N V respond negatively to an increasing ionization parameter and then constrained the ionization parameter $\log_{10} U$ of most BAL outflows to be greater than 0 using photoionization simulations (see Figure 11 in ref.22). In view of this, we assume $\log_{10} U = 0$ for all the objects in our sample and perform our subsequent calculations accordingly, though we also report results based on other values of the ionization parameter for reference.

The recombination timescale $t_c$ of the C IV line is related to the electron density $n_e$ and the recombination rate $\alpha$ (see Methods for details). According to the measured $t_c$ distribution, the mean and standard deviation of the electron density distribution are $n_e = 10^{0.19 \pm 0.02}, 10^{0.79 \pm 0.02}, 10^{0.79 \pm 0.02}$ and $10^{0.79 \pm 0.02}$ cm$^{-3}$ at $\log_{10} U = -2, -1, 0$ and 1, respectively.

The outflow distance $R$ can be determined as long as the ionization state and density are known (see Methods for details). As shown in Figure 2, the mean and standard deviation of the $R$ distributions are $10^{1.70 \pm 0.54}, 10^{0.91 \pm 0.54}, 10^{4.11 \pm 0.54}$ and $10^{6.02 \pm 0.54}$ pc at $\log_{10} U = -2, -1, 0$ and 1, respectively. Our result that the typical outflow radius is tens of pc indicates that the BAL outflow locations is outside the theoretically predicted trough forming region (0.01 $\sim$ 1.0 pc) for accretion disc line-driven winds [42] but are smaller than the scales of most outflows that are derived using the excited state absorption lines [43].

The mass-flow rate $M_{\text{out}}$ and kinetic luminosity $\dot{E}_k$ are the key parameters to quantify the powerfulness of the feedback effect. As shown in Figure 2, the mean and standard deviation of the distribution of $M_{\text{out}}$ (see Methods and Supplementary Figure 7, 8 for details) are $10^{-0.09 \pm 0.70}, 10^{-0.27 \pm 0.70}, 10^{-1.61 \pm 0.70}, 10^{2.86 \pm 0.70}$ and $10^{-2.02 \pm 0.70}$ at $\log_{10} U = -2, -1, 0$ and 1, respectively. At $\log_{10} U = 0$, the typical value of $M_{\text{out}}$ is of ten to one hundred $M_\odot$ yr$^{-1}$.

The ratio of the mass-flow rate to accretion rate $M_{\text{out}}/M_{\text{acc}}$ is a proxy for exploring the relationship between the accretion system on small scales and the BAL outflows on relatively large scales. The accretion rate is given by $M_{\text{acc}} = L_{\text{bol}}/\eta c^2$, where $\eta = 0.1$ is the energy conversion efficiency. The mean and standard deviation of the distribution of $M_{\text{out}}/M_{\text{acc}}$ are $10^{-0.84 \pm 0.70}, 10^{-1.02 \pm 0.70}, 10^{0.86 \pm 0.70}$ and $10^{2.01 \pm 0.70}$ at $\log_{10} U = -2, -1, 0$ and 1, respectively. At $\log_{10} U = 0$, the typical value of $M_{\text{out}}/M_{\text{acc}}$ is a few times of $M_{\text{acc}}$. It is worth mentioning that ref.23 performed a series of two-dimensional radiation-hydrodynamical simulations of line-driven disc winds for black holes with masses in the range $M_{\text{BH}} = 10^9-10^{12} M_\odot$ and Eddington ratios in the range $\varepsilon = 0.1 - 0.5$. Their simulations predict that $M_{\text{out}}$ can become comparable to $M_{\text{acc}}$ when the Eddington ratio $\varepsilon$ is at least 0.3.

As shown in Figure 2, the mean and standard deviation of the kinetic-to-bolometric luminosity ratio $\dot{E}_k/L_{\text{bol}}$ (see Methods and Supplementary Figure 7, 8 for details) are $10^{-2.82 \pm 0.70}, 10^{-3.01 \pm 0.70}, 10^{-1.11 \pm 0.70}$ and $10^{1.92 \pm 0.70}$ at $\log_{10} U = -2, -1, 0$ and 1, respectively. At $\log_{10} U = 0$, the typical value of $\dot{E}_k/L_{\text{bol}}$ is a few percent. The threshold of $\dot{E}_k/L_{\text{bol}}$ for effective AGN feedback is still under debate. According to ref.24 a quasar outflow can effectively suppress star formation in the host galaxy by directly expelling the ISM when $\dot{E}_k/L_{\text{bol}}$ is at $10^{-1.3}$, i.e., 5%. Meanwhile, ref.25 proposed a two-stage feedback model. In their model, dense clouds expand when the outflow passes by, which increases the clouds’ cross-section and makes them more susceptible to radiative momentum driving and ionization heating by the quasar. For this case, an $\dot{E}_k/L_{\text{bol}}$ value of $10^{-2.3}$ (i.e., 0.5%) is able to produce enough feedback to suppress star formation in the host galaxy. In either case, a large fraction of the BAL outflows in our study appear powerful enough to regulate the growth of the SMBHs and their host galaxies.

Correspondence and request for materials should be addressed to Z.-C. H or T.-G. W or G.-L. L. (e-mail: zcho@ustc.edu.cn, twang@ustc.edu.cn, gliu@ustc.edu.cn).

Acknowledgements We acknowledge the financial support by the Strategic Priority Research Program "The Emergence of Cosmological Structures" of
Figure 1 | The fraction curve of BAL variabilities and the inferred recombination timescales distribution in the SDSS sample. a, the cumulative distribution function of a Gaussian distribution is used to model the fraction curve of BAL variabilities. The vertical error bars mark the 1σ uncertainty of the fraction curve and the horizontal one marks the width of each time bin. b, the inferred Gaussian distribution of the recombination timescales of BAL outflow gas. The mean and standard deviation of the recombination timescales are $t_{rec} = 10^{0.36 \pm 2.01}$ days.

1. Scannapieco, E. & Oh, S. P. Quasar feedback: the missing link in structure formation. *Astrophys. J.* 608, 62 (2004).
2. Murray, N., Quataert, E. & Thompson, T. A. On the maximum luminosity of galaxies and their central black holes: feedback from momentum-driven winds. *Astrophys. J.* 618, 569 (2005).
3. Cioffi, L., Ostriker, J. P. & Proga, D. Feedback from central black holes in elliptical galaxies. i. models with either radiative or mechanical feedback but not both. *Astrophys. J.* 699, 89 (2009).
4. Ostriker, J. P., Choi, E., Cioffi, L., Novak, G. S. & Proga, D. Momentum driving: which physical processes dominate active galactic nucleus feedback? *Astrophys. J.* 722, 648 (2010).
5. Hopkins, P. F. & Elvis, M. Quasar feedback: more bang for your buck. *Mon. Not. R. Astron. Soc.* 401, 7–14 (2009).
6. Di Matteo, T., Springel, V. & Hernquist, L. Energy input from quasars regulates the growth and activity of black holes and their host galaxies. *Nature* 433, 604 (2005).
7. Arav, N. et al. Measuring column densities in quasar outflows: Vlt observations of qso 2359–1241. *Astrophys. J.* 681, 954 (2008).
8. Moe, M., Arav, N., Bautista, M. A. & Korista, K. T. Quasar outflow contribution to agn feedback: Observations of qso sdss j0838+ 2955. *Astrophys. J.* 706, 525 (2009).
9. Hamann, F. et al. A high-velocity narrow absorption line outflow in the quasar [212329. 46- 005052.9]. *Mon. Not. R. Astron. Soc.* 410, 1957– 1974 (2011).
10. Borguet, B. C., Edmonds, D., Arav, N., Dunn, J. & Kriss, G. A. A 10 kpc scale seyfert galaxy outflow: Hst/cos observations of iras f22456–5125. *Astrophys. J.* 751, 107 (2012).
11. Arav, N. et al. Anatomy of the agn in ngc 5548-ii. the spatial, temporal, and physical nature of the outflow from hst/cos observations. *Astron. Astrophys.* 577, A37 (2015).
12. Chamberlain, C., Arav, N. & Benn, C. Strong candidate for agn feedback: Vlt/x-shooter observations of balqso sdss j0831+ 0354. *Mon. Not. R. Astron. Soc.* 450, 1085–1093 (2015).
13. Arav, N. et al. Evidence that 50% of balqso outflows are situated at least 100 pc from the central source. *Astrophys. J.* 857, 60 (2018).
14. Xu, X., Arav, N., Miller, T. & Benn, C. Vlt/x-shooter survey of bal quasars: Large distance scale and agn feedback. *arXiv preprint arXiv:1805.01544* (2018).
The distribution at the ionizing parameter

\[ \log_{10} \frac{L_{\text{bol}}}{L_{\text{ion}}} \] for different ionizing parameters. The error bars mark the standard deviation of the distributions. The distance distribution is \[ R = 10^{0.21\pm0.04} \text{ pc} \], the mass flow rate distribution is \[ M_{\text{out}} = 10^{4.61\pm0.70} \text{ M}_\odot \text{ yr}^{-1} \], the mass flow rate to accretion rate distribution is \[ M_{\text{out}}/M_{\text{acc}} = 10^{0.46\pm0.70} \], and the ratio of the kinetic-to-bolometric luminosity ratio distribution is \[ \frac{E_k}{L_{\text{bol}}} = 10^{-1.11\pm0.70} \].

Figure 2 | Distributions of the properties of the BAL outflows at different ionizing parameters. The distributions at the ionizing parameter \( \log_{10} U = 0 \) (red points) are the final results (marked with the mean and standard deviation values in the right-hand panels). The error bars mark the standard deviation of the distributions. A, the distance distribution is \( R = 10^{0.21\pm0.04} \text{ pc} \). B, the mass flow rate distribution is \( M_{\text{out}} = 10^{4.61\pm0.70} \text{ M}_\odot \text{ yr}^{-1} \). C, the mass flow rate to accretion rate distribution is \( M_{\text{out}}/M_{\text{acc}} = 10^{0.46\pm0.70} \). D, the ratio of the kinetic-to-bolometric luminosity ratio distribution is \( \frac{E_k}{L_{\text{bol}}} = 10^{-1.11\pm0.70} \).

METHODS

The change in the average intensity of ionizing continuum. For a typical high luminosity quasar, we assume that \( M_{\text{BH}} = 10^9 \text{ M}_\odot \), \( M_i = -26 \); the relaxation timescale \( \tau = 200 \text{ days} \), and the value of the structure function at infinity \( SF_{\infty} = 0.2 \text{ mag} \). As shown in Supplementary Figure 1, we generate a light curve at 1500\AA{} from the damped random walk (DRW) mode using the Python package astroML. The time interval between a pair of observations is assumed to be 30 days. The change in the average intensity of ionizing continuum (averaged over a recombination timescale \( t_r \)) before the observation) over the 100 days is obviously smaller than that of 10 days. In general, the change in the average intensity of ionizing continuum between a pair of observations decreases when the \( t_r \) of gas is increased (panel b of Supplementary Figure 1). In principle, the change in the average intensity of ionizing continuum can be ignored when the \( t_r \) is longer than the time interval between a pair of observations.

BAL Quasar Sample. We merge the BAL quasar catalog of SDSS data release 7 (DR7) and that of DR14. Then we compare this catalog with SDSS data release 14 (DR14) and select quasars with multiple spectroscopic observations. To investigate the variability of the C IV Ni 1549 BAL trough, we adopt a redshift cut \( 1.9 < z < 4.7 \). To ensure detection of major absorption lines, we only keep quasars with at least one spectrum with signal to noise (S/N) at SDSS \( q \) band greater than 10. After these cuts, we obtain a sample of 1728 BAL quasars with spectra taken at two or more epochs. For each quasar with \( m (m > 2) \) spectra, there are \( C_m^2 = m(m - 1)/2 \) spectral pairs. Then, there are total 9772 spectral pairs of 1728 BAL quasars in our sample. The amplitude of continuum variation is defined as

\[ \frac{\Delta L}{L} = 2 \frac{L_2 - L_1}{L_1 + L_2}, \]
where $L_1$ and $L_2$ are the fluxes measured at the first and second observations, respectively. The continuum (1500 Å) variation amplitudes $|\Delta L/L|$ versus the rest time interval $\log_{10} \Delta T$ are plotted in gray points in panel a of Supplementary Figure 2. The distribution of $\log_{10} \Delta T$ are shown in panel b of Supplementary Figure 2. The distribution of $|\Delta L/L|$ are shown in panel c of Supplementary Figure 2. The fiber positions of BOSS quasar targets were purposefully offset in order to optimize the throughput of light at 4000 Å, while the standard stars used for flux calibration are positioned for 5400 Å. This results in a large uncertainty in the flux calibration of quasar spectra in the BOSS survey.

In DR14, Re[23] have re-reduced BOSS spectra and improved the flux calibration by adding new atmospheric distortion corrections at the per-exposure level[24]. We cut the $|\Delta L/L|$ at 10% which is greater than the amplitude of spectrophotometric uncertainties of 6% level[24]. The $|\Delta L/L|$ distributions of our sample and the DRW model[22,23] of a typical quasar ($M_{\text{BH}} = 10^9 M_\odot$, $M_1 = -26$) are consistent at $|\Delta L/L| \geq 10\%$ (Kolmogorov-Smirnov test: $r = 0.02$, $p = 0.23$). These two distributions with $|\Delta L/L| < 10\%$ are significantly different ($r = 0.08$, $p = 1.6 \times 10^{-14}$).

As shown in black points of panel a of Supplementary Figure 2, to keep the $|\Delta L/L|$ flat in different time bins, we only select those spectral pairs of $10\% < |\Delta L/L| < 30\%$. Among this sample, there are 3686 spectra pairs from 915 quasars (black points). Through the identification of variable absorption lines, 1572 pairs (red points) of spectra from 432 BAL quasars are detected to have C IV BAL variations.

As shown in panel a of Figure 1 and panel d, e, f of Supplementary Figure 2, in order to measure the fraction $F(\Delta T)$, the selected 3686 pairs are sorted according to the rest time interval between the observations and divided among 38 bins. Each bin contains 100 spectra pairs except the first three bins. The first bin has 44 spectral pairs of $\log_{10} \Delta T < 0$. The second bin has 68 spectral pairs of $0 < \log_{10} \Delta T < 0.5$. The third bin has 74 spectral pairs. We have marked the widths (standard deviation) of all the bins in panel a of Figure 1 and panel d, e, f of Supplementary Figure 2.

The S/N of spectra may affect the detectability of BAL variability. As shown in the right panel of Supplementary Figure 2, the $S/N$ of the spectra at SDSS g band is nearly constant in all time intervals. The detectability of absorption line variability should therefore be approximately the same in the different time bins. In addition, there are two factors that may affect the BAL variability: the amplitudes of continuum variations $|\Delta L/L|$ and the basic physical parameters of the host quasars e.g., the bolometric luminosity $L_{\text{bol}}$ and the central SMBH mass $M_{\text{BH}}$[23]. The BAL variability increases with the amplitude of continuum variations while the timescale of continuum variations may increase with $L_{\text{bol}}$ and $M_{\text{BH}}$. The amplitudes of continuum variations $|\Delta L/L|$ and the monochromatic luminosities at 1500 Å are almost constant for all time bins. Thus, the influence of these two factors can also be ignored.

**Fitting C IV BAL trough.** To reliably characterize the continuum and delineate it from the C IV, N V BAL troughs, we use the unabsorbed quasar template[23] derived from SDSS Data Release 7 (DR7) to fit the spectra. Following ref[23,24], we use a double power-law function (Equation 1 in ref[23]) as the scale factor to scale these templates.

Dividing the spectra by the continuum, we obtain the normalized spectrum and then mark the contiguous deficient pixels as the possible intrinsic absorption lines region of C IV in the normalized spectrum. The marked region with a width of $\Delta \ln \lambda \geq 10^{-5}$ (greater than 300 km s$^{-1}$ in velocity) and statistically significant that 5σ will be screened as the intrinsic moderate to broad absorption line. Finally, we exclude the false ones (due to an improper fit in most cases) by the visual inspection. The distributions of the weighted centroid velocity[23] and width of C IV BAL are shown in panel a and b of Supplementary Figure 4.

The equivalent width (EW) of the C IV BAL troughs is calculated as follows:

$$\text{EW} = \int [1 - F_{\text{obs}}(\lambda)/F_{\text{cont}}(\lambda)] d\lambda.$$

The integration is done for the identified absorption line region. The averaged EW for each object of the 915 quasars are shown in Supplementary Figure 5.

The normalized residual flux in the trough of a partially obscured absorber is

$$I(\nu) = 1 - C(\nu) + C(\nu)e^{-\tau(\nu)},$$

where $C(\nu)$ and $\tau(\nu)$ are the covering factor and the optical depth of the ion at velocity $\nu$, respectively. The oscillator strengths of the blue and red components for the resonance doublet C IV 1548.2, 1550.8 Å are $f_{\text{blue}} = 0.19$ and $f_{\text{red}} = 0.095$, respectively. This renders an optical depth ratio $\tau_{\text{blue}}/\tau_{\text{red}}(\lambda_{\text{blue}}/f_{\text{blue}})/(\lambda_{\text{red}}/f_{\text{red}})$ close to 2. Thus, we will use the doublet components to fit the BAL troughs. The covering fraction $C$ has been found different at different velocities[23]. However, for simplicity, we only consider a constant covering factor for the whole BAL trough and allow the optical depth to vary with velocity. Note that our result of the C IV column densities is a conservative estimation. According to the partial covering model, we obtain a set of equations of $\tau(\nu_i)$ as follows:

$$\begin{align*}
I_{\nu_1} &= 1 - C + Ce^{-\tau(\nu_1)}, \\
I_{\nu_2} &= 1 - C + Ce^{-\tau(\nu_2)}, \\
I_{\nu_3} &= [1 - C + Ce^{-\tau(\nu_3)}][1 - C + Ce^{-2\tau(\nu_1)}], \\
I_{\nu_4} &= [1 - C + Ce^{-\tau(\nu_4)}][1 - C + Ce^{-2\tau(\nu_2)}], \\
I_{\nu_5} &= [1 - C + Ce^{-\tau(\nu_5)}][1 - C + Ce^{-2\tau(\nu_3)}], \\
I_{\nu_6} &= 1 - C + Ce^{-2\tau(\nu_4)},
\end{align*}$$

(5)

where $\tau$ is the optical depth of red component. There are $n$ equations in total with $n-k+1$ unknown variables, where $k = 5$ for a bin of 0.5 Å in wavelength. Since there are more constraints than unknown variables, the equation set has no exact solution. We therefore use the least-squares method to find a set of $\{\tau(\nu_i)\}$ that best fits these equations. To account for the noise in the flux and the uncertainty of continuum, the low-limit of the optical depth $\tau$ is set to $-0.1$ (corresponding to a normalized flux $I = 1.1$). For most of the troughs, the best fit results give a reduced $\chi^2$ around 1. An example fit is shown in panel a of Supplementary Figure 3. Note that the optical depth for the case of saturated absorption must be underestimated.

After the troughs are fitted, the C IV column densities are obtained by integrating the optical depth over the troughs:

$$\sum\text{N}_{\text{ion}} = \frac{3.7679 \times 10^{14} \text{cm}^{-2}}{\lambda f} \int \tau(\nu) d\nu,$$

(6)

where $\lambda$ and $f$ are the transition’s wavelength and oscillator strength, respectively, and the velocity $\nu$ is measured in km s$^{-1}$. The C IV column densities versus the BAL EW for the 915 quasars are shown in Supplementary Figure 5.

**Identification of the variable region of C IV BAL trough.** As described in ref[23,24], in order to identify the variable region of C IV BAL between a pair of spectra, we first need to exclude the influences of the continuum and emission line. To account for the potential variations of the continuum shape, we select the higher S/N spectrum of the pair of spectra as a template to match the other spectra by rescaling it using the double power-law function (Equation 1 in ref[23]). To account for variations of the emission line, we add/subtract a Gaussian to/from the rescaled spectrum. An example of the fit is displayed in panel b of Supplementary Figure 3. Compared with the unabsorbed quasar template matching, the rescaled template matching produces a better fit outside the absorption line region in most cases. As a result, we will measure the absorption line variability from the difference spectrum.

We take three steps to identify the variable absorption line components from the difference spectrum. Firstly, we search for the contiguous negative and positive pixels and mark all pixels where the difference is greater than 3σ. We screen the adjacent marked pixels and connect them to form a variable region. Secondly, we expand such regions into neighboring pixels which have the same sign but are less than 3σ significant level. Finally, we merge the neighboring regions which have the same variable sign and a separation of less than four pixels. The confidence with which a region of BAL is assigned to be variable is defined as: $N_a = \sum |\Delta \text{flux}|/\sqrt{\sum \sigma^2}$, where the flux uncertainties ($\sigma = \sqrt{\sigma^2_{\text{flux1}} + \sigma^2_{\text{flux2}}}$) of the two spectra includes the possible systematic uncertainties due to rescaling. We perform the identification of variable absorption in the wavelength coverage from 1410 Å to 1500 Å (corresponding velocity $2.7 \times 10^4$ km s$^{-1}$ to 0). 1572
pairs of spectra in 432 BAL quasars are detected (at 3σ detection threshold) to have C IV BAL variations. As discussed in ref[22], the coordinated variability in the troughs of different velocities are also found in our sample. 464 out of 1572 spectral pairs are detected two or more varied troughs. In the 464 spectral pairs, the varied troughs in 75%(347/464) spectral pairs are coordinated. This result also suggests that the BAL variabilities are likely due to clouds at different velocities responding to the same changes in ionizing flux.

The distributions of the weighted centroid velocity and width of C IV BAL for the 432 BAL quasars are plotted as orange lines in panel a and b of Supplementary Figure 4. There is no significant difference of the weighted centroid velocity distributions between the 915 quasars (black) and 432 quasars with varied BAL (Kolmogorov-Smirnov test: r = 0.06, p = 0.25). There is a weak difference of the BAL width distributions between the 915 quasars (black) and 432 quasars with varied BAL (r = 0.09, p = 0.01). The distributions of varied BAL region widths for 1572 spectral pairs are shown in panel c of Supplementary Figure 4. The number of occurrences of BAL absorption (black) and variable region (orange line) at different velocity bin are shown in panel d of Supplementary Figure 4. The purple one is the percentage of BAL variability, i.e., the ratio of the orange one to the black one. The percentages of BAL variability increase with the velocities.

\[ \Delta L / L = \text{log}_{10} \Delta T \]

The percentages of BAL variability increase with the velocities.

Panel a of Supplementary Figure 6 shows the fractions of the BAL variability, i.e., the ratio of the orange line to the black one. The percentages of BAL variability increase with the velocities from 0 to 347 km s\(^{-1}\) which is similar to ref[22]. The percentage of BAL variability is roughly constant (even a slight drop) from 2 × 10\(^{4}\) km s\(^{-1}\) to 2.7 × 10\(^{4}\) km s\(^{-1}\) which is similar to ref[22]. The variable regions are found across a wide range of velocities (mainly greater than 5.0 × 10\(^{4}\) km s\(^{-1}\)), suggesting that our results of the outflow distributions are the comprehensive statistical study for the outflow with different velocities.

Simulation test. To validate our method, we measure \( F(\Delta T) \) from a mock sample of 10\(^{3}\) quasar light curves generated by the DRW model, then estimate \( f(t_c) \) and compare it with the input distribution of recombination timescales \( t_c \). Each light curve has 10\(^{3}\) evenly spaced points, with the time interval between two neighboring points corresponding to 0.1 day.

In the mock sample simulation, we use the 1500 Å flux as an indicator of the ionizing flux, because the highly concordant variations of the absorption lines and UV continuum strongly suggest that the changes of the flux in the observed 1500Å rest-frame UV correlates with the changes of the ionizing flux (see ref[22]).

We use the Cloudy [30] (version c13.03) to simulate the BAL EW variation responding to the variation of ionizing flux. Assuming that the distribution of optical depth \( \tau_C \) in the C IV BAL trough is Gaussian and the covering factor \( C=1 \), we calculate the BAL EW curve as a function of the column density \( N_{\text{CIV}} \), at different velocities \( \sigma_C \). The percent change of BAL variability is roughly constant (even a slight drop) from 2 × 10\(^{4}\) km s\(^{-1}\) to 2.7 × 10\(^{4}\) km s\(^{-1}\) which is similar to ref[22].

The ionization parameter \( \chi \) of the absorber when the Si IV lines is in our calculations. Using the photoionization simulations, we can obtain the ionization parameter \( \chi \) for most BAL outflows. We therefore adopt the ionization parameter \( \log_{10} \chi = 0 \) for all the objects.

Hydrogen column density \( N_H \). The ionization parameter \( \chi \) of all the outflows can be estimated from the input Gaussian distribution (see Eq. 2) to model the fraction curves.

\[ \log_{10} \chi = \log_{10} \chi + 0.36 - 0.14 \text{ and } \tau_C = 1.01 - 0.22 \]

We use the Cloudy [30] to simulate the BAL EW variation responding to the variation of ionizing flux. Assuming that the distribution of optical depth \( \tau_C \) in the C IV BAL trough is Gaussian and the covering factor \( C=1 \), we calculate the BAL EW curve as a function of the column density \( N_{\text{CIV}} \), at different velocities \( \sigma_C \). The percent change of BAL variability is roughly constant (even a slight drop) from 2 × 10\(^{4}\) km s\(^{-1}\) to 2.7 × 10\(^{4}\) km s\(^{-1}\) which is similar to ref[22].

The ionization parameter \( \chi \) of the absorber when the Si IV lines is in our calculations. Using the photoionization simulations, we can obtain the ionization parameter \( \chi \) for most BAL outflows. We therefore adopt the ionization parameter \( \log_{10} \chi = 0 \) for all the objects.
where $\Delta L/L$ is the amplitude of change in the incident ionizing flux. Using the Chianti atomic database version 8.0 at a nominal temperature of $2 \times 10^4$ K, we take the recombination rates $\alpha_{\text{CV}} = 5.3 \times 10^{-12}$ cm$^3$ s$^{-1}$ (from C to C IV) and $\alpha_{\text{CIII}} = 2.1 \times 10^{-11}$ cm$^3$ s$^{-1}$ (from C IV to C III).

At the ionization parameters $\log_{10} U = -2, -1, 0$ and 1, the ratio of number densities of C V to C IV are $n_{\text{C V}} / n_{\text{CIV}} \approx 0.3, 5, 100$ and 1000, respectively.

**Calculation of the outflow properties.** The outflow distance $R$ can be determined as long as the ionization state and density are known. The ionization parameter is defined as follows:

$$U = \frac{Q_H}{4\pi R^2 n_H c}, \quad (8)$$

where $Q_H = \int^{\infty}_{0.5} \pi L_{\text{bol}}(\epsilon) \epsilon^{-\alpha} d\epsilon$ is the source emission rate of hydrogen-ionizing photons, $c$ is the speed of light, and $n_H \approx 0.83n_e$ is the hydrogen number density. Assuming that the distributions of $Q_H$ and $n_H$ are independent of each other, the distribution of $R$ is $\sqrt{Q_H}/(4\pi n_H c)$ is the product distribution of the above two distributions.

If the outflow is in the form of a thin partial shell ($\Delta R/R \ll 1$), $\dot{M}_{\text{out}}$ and $\dot{E}_k$ can be given by

$$\dot{M}_{\text{out}} = 4\pi \Omega C \mu m_p n_H v, \quad (9)$$

$$\dot{E}_k = 2\pi \Omega C \mu m_p n_H v^3, \quad (10)$$

where $\mu = 1.4$ is the mean atomic mass per proton, $m_p$ is the mass of the proton, and $v$ is the radial velocity of the outflow. Here we adopt the weighted centroid velocity of the C IV BAL trough, i.e., the mean of the velocities where each data point is weighted by its distance from the normalized continuum level. $\Omega$ and $C$ are the global covering factor and individual covering factor, respectively. Following the usual statistical approach for C IV BALs, the global covering factor is set to $\Omega = 0.2$.

Combining Eq. 8 and Eq. 9, the expression for $\dot{M}_{\text{out}}$ is

$$\dot{M}_{\text{out}} = 4\pi \Omega C \mu m_p n_H v \sqrt{\frac{Q_H}{4\pi U n_H c}}. \quad (11)$$

The $\dot{M}_{\text{out}}$ distribution can be deduced from the distributions of $CQ_H^{1/2} n_H v$ (see Methods for details) and $n_H$.

The accretion rate is $\dot{M}_{\text{acc}} = \dot{L}_{\text{bol}}/\eta c^2$, where $\eta = 0.1$ is the energy conversion efficiency. From Eq. 11, the $\dot{M}_{\text{out}}/\dot{M}_{\text{acc}}$ can be written as:

$$\frac{\dot{M}_{\text{out}}}{\dot{M}_{\text{acc}}} = \frac{4\pi \Omega C \mu m_p n_H v}{\dot{L}_{\text{bol}}} \sqrt{\frac{Q_H c^3}{4\pi U n_H}}. \quad (12)$$

The bolometric luminosities for the three quasar SED types are as follows: $L_{\text{bol}} = 4.2\lambda_{1500} L_{1500}$ (UV-soft), $L_{\text{bol}} = 6.6\lambda_{1500} L_{1500}$ (MF87), $L_{\text{bol}} = 4.1\lambda_{1500} L_{1500}$ (HE 0238).

From Eq. 10, the kinetic-to-bolometric luminosity ratio can be written as

$$\frac{\dot{E}_k}{\dot{L}_{\text{bol}}} = \frac{2\pi \Omega C \mu m_p n_H v^3}{\dot{L}_{\text{bol}}} \sqrt{\frac{Q_H}{4\pi U n_H c}}. \quad (13)$$

The $\dot{E}_k/\dot{L}_{\text{bol}}$ distribution can be derived from the distributions of $CQ_H^{1/2} n_H v^3 L_{\text{bol}}^{-1}$ and $n_H$.

We use the skewed Gaussian functions,

$$N(x) = G(p_1, p_2, p_3) \{1 + \text{erf} [p_4(x - p_1)]\} \quad (14)$$

to model the distribution of $Q_H$, $CQ_H^{1/2} n_H v$, $CQ_H^{1/2} n_H v L_{\text{bol}}^{-1}$ and $CQ_H^{1/2} n_H v^3 L_{\text{bol}}^{-1}$, where the $p_1$ is the mean of Gauss, $p_2$ is the standard deviation of Gauss, $p_3$ is the amplitude of Gauss and $p_4$ is the coefficient of skewness.

**Data availability.** The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request.
31. MacLeod, C. L. et al. Modeling the time variability of sdss stripe 82 quasars as a damped random walk. *Astrophys. J.* 721, 1014 (2010).
32. Kelly, B. C., Bechtold, J. & Siemiginowska, A. Are the variations in quasar optical flux driven by thermal fluctuations? *Astrophys. J.* 698, 895 (2009).
33. Kozłowski, S. et al. Quantifying quasar variability as part of a general approach to classifying continuously varying sources. *Astrophys. J.* 708, 927 (2009).
34. Guo, H., Wang, J., Cai, Z. & Sun, M. How far is quasar uv/optical variability from a damped random walk at low frequency? *Astrophys. J.* 847, 132 (2017).
35. Shen, Y. et al. A catalog of quasar properties from Sloan digital sky survey data release 7. *Astrophys. J.* Suppl. Ser. 194, 45 (2011).
36. Páris, I. et al. The Sloan digital sky survey quasar catalog: twelfth data release. *Astron. Astrophys.* 597, A79 (2017).
37. Dawson, K. S. et al. The baryon oscillation spectroscopic survey of sdss-iii. *Astron. J.* 145, 10 (2013).
38. Abolfathi, B. et al. The fourteenth data release of the Sloan digital sky survey: first spectroscopic data from the extended baryon oscillation spectroscopic survey and from the second phase of the Apache Point Observatory galactic evolution experiment. *Astrophys. J.* Suppl. Ser. 235, 42 (2018).
39. Margala, D. et al. Improved spectrophotometric calibration of the sdss-iii boss quasar sample. *Astrophys. J.* 831, 157 (2016).
40. Shen, Y. et al. The Sloan digital sky survey reverberation mapping project: technical overview. *Astrophys. J.* Suppl. Ser. 216, 4 (2014).
41. Arav, N., Korista, K. T., De Kool, M., Junkkarinen, V. T. & Begelman, M. C. Hubble space telescope observations of the broad absorption line quasar pg 0946+301. *Astrophys. J.* 516, 27 (1999).
42. Hall, P. B. et al. Unusual broad absorption line quasars from the Sloan digital sky survey. *Astrophys. J.* Suppl. Ser. 141, 267 (2002).
43. Arav, N., Borguet, B., Chamberlain, C., Edmonds, D. & Danforth, C. Quasar outflows and AGN feedback in the extreme UV: HST/COS observations of he 0238−1904. *Mon. Not. R. Astron. Soc.* 436, 3286−3305 (2013).
44. Leighly, K. M., Terndrup, D. M., Gallagher, S. C., Richards, G. T. & Dietrich, M. The z=0.54 lobar quasar sdss j085053.12+445122.5: I. spectral synthesis analysis reveals a massive outflow. arXiv:1808.02441 (2018).
45. Savage, B. D. & Sembach, K. R. The analysis of apparent optical depth profiles for interstellar absorption lines. *Astrophys. J.* 379, 245−259 (1991).
46. Rogerson, J. A. et al. Emergence and variability of broad absorption line quasar outflows. *Astrophys. J.* 862, 22 (2018).
47. Ferland, G. et al. The 2013 release of cloudy. *Revista mexicana de astronomía y astrofísica* 49, 137−163 (2013).
48. Mushotzky, R. F., Edelson, R., Baumgartner, W. & Gandhi, P. Kepler observations of rapid optical variability in active galactic nuclei. *Astrophys. J.* L743, L12 (2011).
49. Kasliwal, M. P., Vogtsey, M. S. & Richards, G. T. Are the variability properties of the kepler agn light curves consistent with a damped random walk? *Mon. Not. R. Astron. Soc.* 451, 4328−4345 (2015).
50. Steenbrugge, K., Feovˇcık, M., Kaastra, J., Costantini, E. & Verbunt, F. High-resolution x-ray spectroscopy of the low and high states of the Seyfert 1 galaxy ngc 4051 with Chandra LETGS. *Astron. Astrophys.* 578, 107−119 (2009).
51. Dietmers, R. et al. Multiwavelength campaign on mrk 509-iii. the 600 ks rgs spectrum: unravelling the inner region of an agn. *Astron. Astrophys.* 534, A38 (2011).
52. Mathews, W. G. & Ferland, G. J. What heats the hot phase in active nuclei? *Astrophys. J.* 323, 456−467 (1987).
53. Dunn, J. P. et al. The quasar outflow contribution to agn feedback: VLT measurements of sdss j0318-0600. *Astrophys. J.* 709, 611 (2010).
54. Arav, N. et al. Chemical abundances in an agn environment: X-ray/uv campaign on the markarian 279 outflow. *Astrophys. J.* 658, 829 (2007).
55. Borguet, B. C., Edmonds, D., Arav, N., Benn, C. & Chamberlain, C. Bal phosphorus abundance and evidence for immense ionic column densities in quasar outflows: VLT/x-shooter observations of quasar sdss J1512+1119. *Astrophys. J.* 758, 69 (2012).
56. Borguet, B. C., Arav, N., Edmonds, D., Chamberlain, C. & Benn, C. Major contributor to agn feedback: VLT x-shooter observations of s iv balqso outflows. *Astrophys. J.* 762, 49 (2013).
57. Arav, N. et al. Multiwavelength campaign on mrk 509-x. lower limit on the distance of the absorber from hst and stis spectroscopy. *Astron. Astrophys.* 544, A33 (2012).
58. Del Zanna, G., Dere, K., Young, P., Landi, E. & Mason, H. Chianti—an atomic database for emission lines. version 8. *Astron. Astrophys.* 582, A56 (2015).

**Supplementary Information**
Supplementary Figure 1 | The change in the average intensity of the ionizing continuum. a, the gray line is an example of the light curve at 1500Å, generated from the DRW model. The black points mark the fluxes at two observations with 30 days time interval. The blue/orange line is the average intensity of ionizing continuum over a recombination timescale \( t_{\text{rec}} \), 10/100 days before the two observations. The change in the average intensity of ionizing continuum over the 100 days is obviously smaller than that of 10 days. b, the change in the average intensity of ionizing continuum between the two observations decreases when the \( t_{\text{rec}} \) of gas is increased.
Supplementary Figure 2 | Properties of the BAL quasar sample. a, the continuum (1500Å) variation amplitudes |ΔL/L| versus the rest time interval log_{10} ΔT: 9772 spectral pairs of 1728 BAL quasars in the whole sample (gray points); 3686 spectra pairs from 915 quasars with 10% < |ΔL/L| < 30% (black points); 1572 pairs of spectra from 432 BAL quasars with C IV BAL variations (orange points). The gray vertical line marks the amplitude of spectrophotometric uncertainties of 6% level. The red horizontal line marks the level of |ΔL/L| = 10%. b, the distributions of the rest time interval log_{10} ΔT. The colors are the same with a. c, the distributions of the amplitudes of continuum variations |ΔL/L| for the whole sample (gray line) and the DRW model (blue line) of a typical quasar (M_{BH} = 10^9 M_☉, M_i = −26) are consistent at |ΔL/L| ≥ 10% (Kolmogorov-Smirnov test: r = 0.02, p = 0.23). These two distributions with |ΔL/L| < 10% are significantly different (r = 0.08, p = 1.6 × 10^{-14}).
Supplementary Figure 2 | Properties of the BAL quasar sample. d, the mean and standard deviation of the S/N of SDSS g band of the spectra for 3686 pairs from the 915 quasars at different time bins. e, the mean and standard deviation of the amplitudes of continuum variations $|\Delta L/L|$ at 1500Å. f, the mean and standard deviation of the monochromatic luminosities at 1500Å. All these quantities are nearly constants in all time intervals.
Supplementary Figure 3 | Fit of the C\textsuperscript{IV} BAL troughs and identification of variable absorption Lines. a, an example of the C\textsuperscript{IV}1548.2,1550.8\,\AA
doublet blended BAL trough. The gray line is the observational spectrum normalized by the continuum. The horizontal dashed line marks the level of the continuum. The red/blue line is the best fitted red/blue component of the doublets, and the thick black one is the product of the doublets. b, an example of matching the reference spectrum to another spectrum (in black) by multiplying the reference spectrum with a double power-law described in the text (the orange line). The blue curve represents the one with additional Gaussians to account for the change of the emission line equivalent width. The black horizontal line represents the varied region of C\textsuperscript{IV} BAL. The residuals of fits (solid line) and the combined spectrum uncertainties (dashed line) are plotted in the lower panel.
Supplementary Figure 4 | Distributions of the parameters of the C IV BAL troughs. The black line is the 915 quasar with $10\% < |\Delta L/L| < 30\%$ and the orange one is for the 432 quasar with varied BALs. a, the distributions of the weighted centroid velocities. There is no significant difference between the black and orange (Kolmogorov-Smirnov test: $r = 0.06$, $p = 0.25$). b, the distributions of BAL trough widths. There is a weak difference between the black and orange ($r = 0.09$, $p = 0.01$). c, the distributions of varied BAL region widths for 1572 spectral pairs. d, the number of occurrences of BAL absorption (black) and variable region (orange) at different velocity bin. The purple one is the percentage of BAL variability, i.e., the ratio of the orange one to the black one.
Supplementary Figure 5 | Simulation of the BAL EW response to the variation of ionizing flux. a, the black dots are the measured C IV BAL EW and column density \( N_{\text{CIV}} \) for the 915 quasars. The typical 1σ error bar of \( \log_{10} \) BAL EW is 0.014 dex. The typical 1σ error bar of \( \log_{10} N_{\text{CIV}} \) is 0.03 dex. The color lines are the BAL EW curves as a function of the \( N_{\text{CIV}} \) at different widths \( \sigma_v \). b, the response of \( N_{\text{CIV}} \) and BAL EW to the variations of ionization parameters \( \log_{10} U \) using Cloudy at the hydrogen column density \( \log_{10} N_{\text{H}} = 21 \text{ cm}^{-2} \) and 22 cm\(^{-2}\). The BAL EW will respond to the \( \log_{10} U \) along each color line at different widths \( \sigma_v \).
Supplementary Figure 6 | Simulation result. a, the black points with error bars are the $F(\log_{10} \Delta T)$ with 1σ uncertainty measured from the mock sample. We use the CDF of a Gaussian distribution (see Eq. 2) to model the fraction curves. The orange and purple lines are the best fitted models for $\log_{10} N_H = 21$ cm$^{-2}$ and $\log_{10} N_H = 22$ cm$^{-2}$, respectively. b, the input (black thick line) and recovered (orange for $\log_{10} N_H = 21$ cm$^{-2}$ and purple for $\log_{10} N_H = 22$ cm$^{-2}$) Gausses. The mean and standard deviation of the input Gaussian distribution are $t_c = 0.36 \pm 0.14$ and $t_\sigma = 1.01 \pm 0.22$. The best fit parameters of the recovered Gauss for $\log_{10} N_H = 21$ cm$^{-2}$ are: $t_c = 0.42 \pm 0.03$ and $t_\sigma = 0.98 \pm 0.06$. For $\log_{10} N_H = 22$ cm$^{-2}$, the best fit parameters are: $t_c = 0.43 \pm 0.03$ and $t_\sigma = 0.97 \pm 0.06$. The parameters of the recovered Gausses are consistent with the parameters of the input Gauss within 1σ uncertainty.

Supplementary Figure 7 | The hydrogen column density $N_H$ at the ionization parameter $\log_{10} U = 0$. The hydrogen column density $N_H$ of all the outflows can be estimated from the C IV column density $N_{CIV}$. The orange, black and purple lines represent the UV-Soft, MF87 and HE0238 SEDs, respectively.
Supplementary Figure 8 | Supplementary distributions at $\log_{10} U = 0$. We use the skewed Gaussian functions to model the distributions: a. $Q_H$; b. $CQ_H^{1/2} N_H$; c. $CQ_H^{1/2} N_{H_{2}} L_{bol}^{-1}$; d. $CQ_H^{1/2} N_{H_{2}} V^{3} L_{bol}^{-1}$. The dashed lines are the best fit results. The orange, black and purple lines represent the UV-Soft, MF87 and HE0238 SEDs, respectively.