Atmos. Meas. Tech., 8, 4197–4213, 2015
www.atmos-meas-tech.net/8/4197/2015/
doi:10.5194/amt-8-4197-2015

Atmospheric Measurement Techniques
Open Access

Eddy-covariance data with low signal-to-noise ratio: time-lag determination, uncertainties and limit of detection

B. Langford[1], W. Acton[2], C. Ammann[3], A. Valach[2], and E. Nemitz[1]

[1]Centre for Ecology and Hydrology, Bush Estate, Penicuik, EH26 0QB, UK
[2]Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK
[3]Agroscope Research Station, Climate and Air Pollution Group, Zürich, Switzerland

Correspondence to: B. Langford (benngf@ceh.ac.uk)

Received: 28 January 2015 – Published in Atmos. Meas. Tech. Discuss.: 18 March 2015
Revised: 11 September 2015 – Accepted: 22 September 2015 – Published: 12 October 2015

Abstract. All eddy-covariance flux measurements are associated with random uncertainties which are a combination of sampling error due to natural variability in turbulence and sensor noise. The former is the principal error for systems where the signal-to-noise ratio of the analyser is high, as is usually the case when measuring fluxes of heat, CO_2 or H_2O. Where signal is limited, which is often the case for measurements of other trace gases and aerosols, instrument uncertainties dominate. Here, we are applying a consistent approach based on auto- and cross-covariance functions to quantify the total random flux error and the random error due to instrument noise separately. As with previous approaches, the random error quantification assumes that the time lag between wind and concentration measurement is known. However, if combined with commonly used automated methods that identify the individual time lag by looking for the maximum in the cross-covariance function of the two entities, analyser noise additionally leads to a systematic bias in the fluxes. Combining data sets from several analysers and using simulations, we show that the method of time-lag determination becomes increasingly important as the magnitude of the instrument error approaches that of the sampling error. The flux bias can be particularly significant for disjunct data, whereas using a prescribed time lag eliminates these effects (provided the time lag does not fluctuate unduly over time). We also demonstrate that when sampling at higher elevations, where low frequency turbulence dominates and covariance peaks are broader, both the probability and magnitude of bias are magnified. We show that the statistical significance of noisy flux data can be increased (limit of detection can be decreased) by appropriate averaging of individual fluxes, but only if systematic biases are avoided by using a prescribed time lag. Finally, we make recommendations for the analysis and reporting of data with low signal-to-noise and their associated errors.

1 Introduction

1.1 Motivation

Surface layer fluxes of gases such as carbon dioxide (CO_2) and methane (CH_4) are frequently determined using the eddy-covariance (EC) technique. This approach has allowed direct measurements of canopy-scale emission/deposition rates which are routinely incorporated into models of the carbon cycle and atmospheric chemistry. As with all measurements, the reported flux has an associated error, which should reflect both the systematic and random uncertainties of the measurement system. Systematic uncertainties arise e.g. from having an imperfect measurement system. For example, bandwidth limitations confine our ability to capture all the turbulent motions that contribute to the flux, and if uncorrected will introduce a bias. Another obvious systematic error is introduced by the uncertainty in the calibration standard. Identifying, minimising and correcting sources of systematic bias in flux measurements has been an active area of research (Businger, 1986; Lenschow and Raupach, 1991; Lenschow et al., 1994; Mann and Lenschow, 1994; Massman, 2000; Massman and Lee, 2002). In contrast, random errors do not bias the flux but reduce the overall confidence in an individual reported value. The main sources of random uncertainties in EC flux measurements are widely ac-

cepted as (i) the stochastic nature of turbulence sampling and (ii) instrument noise and the resolution at which samples are recorded. Numerous studies have focused on quantifying random uncertainties, ranging from rigorous theoretical investigations (Lenschow and Kristensen, 1985) to more practical approaches (Hollinger and Richardson, 2005). Usually these studies have addressed the problem from the perspective of an analytical system with fast-response time series and with good signal-to-noise ratio (SNR) e.g. fluxes of sensible heat, CO_2 or H_2O, because for these measurements, the uncertainty in the flux is typically dominated by natural turbulence variability. For example, Mauder et al. (2013) demonstrated that for fluxes of CO_2 and H_2O, the uncertainty associated with sensor noise was on the order of 1 % as opposed to stochastic errors which ranged between 20 and 30 %. The meaning of SNR is different for flux measurements than for concentration measurements. When measuring concentrations, the information (or "signal") is associated with the mean, while for flux measurements it is associated with the variability in the time series that reflects fluctuations induced by turbulence rather than by instrument noise. It is in the spirit of this latter definition that SNR is used throughout this paper.

Increasingly, eddy covariance is now being applied to measure fluxes of pollutants which are more difficult to measure precisely. Examples include measurements of volatile organic compounds (VOCs; Karl et al., 2002; Langford et al., 2010; Park et al., 2013), ozone (O_3; Coyle et al., 2009; Muller et al., 2009; Stella et al., 2013), nitric oxide (NO; Rummel et al., 2002), nitrogen dioxide (NO_2; Stella et al., 2013), nitrous oxide (N_2O; Eugster et al., 2007; Famulari et al., 2010; Jones et al., 2011) and aerosols (Nemitz et al., 2008; Ahlm et al., 2009; Farmer et al., 2011, 2013). Measuring these scalars at a rate sufficient to meet the requirements for eddy covariance (i.e. several Hz) often results in a low SNR and increases the overall uncertainty in the flux.

In addition, for many of these systems, co-location of anemometer and sensor is not possible. Closed-path sensors require inlet lines that create a time lag (τ) between the vertical wind velocities (w) and measured scalar concentrations (c). Correcting phase shifts between w and c by t data points (where $t = \tau \times$ sampling frequency) is a key step in the calculation of fluxes and is routinely done by assessing the cross-covariance function (i.e. the covariance as a function of time lag – see Eq. 4) between c and w which should reveal a maximum (in absolute terms), when the data are fully synchronised. Yet, when the random uncertainty is high, as is the case for many of these analysers, the cross-covariance becomes noisy, confounding the identification of a clear maximum. Through this data treatment, the low SNR in the concentration measurement, although a random error, may effectively introduce a systematic bias towards more extreme flux values. Recently, Taipale et al. (2010) reviewed the various options for determining time lags with reference to VOC fluxes which often have low SNR. Three commonly used approaches are the maximum (MAX), average (AVG) and prescribed (PRES) methods, which are all well suited for the automated post-processing of eddy-covariance data. Briefly, the PRES method involves using a constant time lag, predicted on the basis of the physical characteristics of the sampling system, i.e. sample flow rate and inner diameter and length of the inlet. The MAX method systematically searches for the absolute maximum value in the cross-covariance function between w and c within a predefined time window. Finally, the AVG method proposed by Taipale et al. (2010) applies a centred running mean to the cross-covariance function and then selects the flux from the unsmoothed cross-covariance function that corresponds to the maximum of the absolute running mean. The latter method was originally developed for use with VOC data measured by a proton transfer reaction mass spectrometer (PTR-MS), but is generically applicable to any data set with low SNR. There are currently no guidelines on the degree of smoothing that should be applied (i.e. the length of the running mean). In their study, Taipale et al. (2010) settle on a 5 s running mean, recognising that this is an arbitrary length.

With so many options available, it is clear that the calculated flux may differ depending on the chosen time-lag method. For example, in their study Taipale et al. (2010) confirm that using a prescribed time lag may result in a systematic underestimation of the flux, as the "true" time lag is likely to vary over time due to fluctuations in pumping speed but also due to the degree of absorption/desorption with the inlet wall and its effect on the effective transport time through the tube. Especially for the more water-soluble compounds this may change with humidity and the aerosol coating of the inlet. Similarly, systematically searching for a maximum within a noisy cross-covariance with multiple local maxima may well bias fluxes towards more extreme values (Laurila et al., 2012). The AVG method offers something of a compromise between the two approaches, but some systematic bias may still remain.

We hypothesise that the bias induced by using methods that search for a maximum in the cross-covariance is closely linked to the random error in the flux, which is in part a function of the SNR of the analytical instrumentation and in some cases may be greater than the systematic error induced from using a prescribed time lag. In order to address this hypothesis, an appropriate method is needed to quantify the random error in the flux and separate it into sampling and instrument error components.

1.2 Common approaches for quantifying random flux errors

Assuming the time lag is known, the random error (RE) of an eddy-covariance flux can be estimated in a variety of ways (Lenschow and Kristensen, 1985). One traditional method is mainly used to estimate the flux error due to the limited sampling of the stochastic nature of turbulence. It is based on the

variance of the instantaneous values of $w'c'$ and the integral timescale and is estimated as (Lumley and Pafanofsky, 1964; Wyngaard, 1973; Lenschow et al., 1994)

$$\mathrm{RE} = \left[\frac{2\sigma^2_{w'c'}\tau_{\mathrm{Fc}}}{L}\right]^{0.5}, \quad (1)$$

where L is the length of the averaging period in seconds, $\sigma^2_{w'c'}$ is the variance of the time series of instantaneous values of $w'c'$ over a typical averaging period (~ 30 min) and τ_{Fc} is the integral timescale, i.e. the average timescale over which correlation persists. The integral timescale can be directly estimated by integrating the area under the auto-covariance of $w'c'$ to the point of the first zero crossing. However, this gets more difficult for cases with high noise levels, as the auto-covariance becomes more scattered resulting in multiple zero crossings, including some artificially close to the zero time lag. This has the consequence that the integral timescale derived by this method becomes physically meaningless. In theory, additional sensor noise should increase the random error but in this case, the reduction in the estimate of τ_{Fc} has the knock-on effect of minimising the error estimate. Consequently, this approach appears unsuitable in situations where signal is limited. An alternative approach, for conditions of neutral stability, is to approximate the integral timescale by dividing the measurement height (z) above the zero plane displacement (d), by the mean wind speed ($\bar{u}$). Yet, Rannik et al. (2009) found this method to overestimate the integral scale by a factor of 2 and consequently the random flux error by a factor of 1.4.

Mahrt (1998) offered an alternative method that negates the use of the integral timescale by splitting the time series into sub-records and formulating the error as the standard error between sub-records (Mahrt, 1998; Rannik et al., 2009):

$$\mathrm{RE} = \left[\frac{\sigma^2_{\mathrm{Fc.sub}}}{k}\right]^{0.5}, \quad (2)$$

such that k is equal to the number of sub records and $\sigma^2_{\mathrm{Fc.sub}}$ is the variance of the k different fluxes calculated for the k different sub-records. This random error reflects a combination of the natural variability of the (genuine) atmospheric concentration and instrument noise, but since it calculates subsequent flux values, it appears likely that it is particularly sensitive to low frequency changes.

Another, quite widely used, method to determine the random uncertainty in eddy-covariance flux measurements was devised by Finkelstein and Sims (2001). Their approach is based on the variance of a covariance between two variables which are first auto- and cross-correlated. The random error is approximated through the integration of the auto-covariance and cross-covariance functions of the vertical wind velocity and scalar concentration as

$$\mathrm{RE} = \sqrt{\frac{1}{n}\left[\sum_{t=-m}^{m} f_{w'w'}(t) f_{c'c'}(t) + \sum_{t=-m}^{m} f_{w'c'}(t) f_{c'w'}(t)\right]}, \quad (3)$$

where $f_{w'c'}(t)$ is the cross-covariance function for a time lag of t data points defined as

$$f_{w'c'}(t) = \frac{1}{n-t}\sum_{i=1}^{n-t}(w_i - \bar{w})(c_{i+t} - \bar{c}) \quad (4)$$

and the auto-covariance terms are calculated analogously as the cross-covariance of an entity with itself. Here, n is the number of samples in the flux averaging period and m is the number of samples needed to capture at least the integral timescale (Rannik et al., 2015). Due to the difficulties of quantifying the integral timescale in noisy data (see above), for such time series, it would be advisable to choose a larger m. This mathematically rigorous estimate of the random flux error is implemented in the commonly used post-processing software EddyUH (https://www.atm.helsinki.fi/Eddy_Covariance/EddyUHsoftware.php) and EddyPro (http://www.licor.com/env/products/eddy_covariance/eddypro.html).

In cases where the time lag is unknown, random flux errors are often assessed based on the statistical properties of the cross-covariance function used to identify the time lag. This technique, first conceived by Wienhold et al. (1995) and developed further by Spirig et al. (2005), involves taking the standard deviation of the cross-covariance function at a distance far from the zero time lag (typically several times the integral timescale). In theory, the cross-covariance in this region reflects both random sensor noise and variability of the (genuine) atmospheric signal/concentration; thus, multiples of the standard deviation can yield a random flux error at a given confidence interval (e.g. $1.96 \times \sigma = 95$th; $3 \times \sigma = 99$th). This technique is described further in Sect. 2.2

While the cross-covariance method is widely used, few studies have attempted to go beyond this and isolate the effects of random sensor noise, mainly due to its negligible influence in many conventional eddy-covariance systems. Noticeable exceptions include the work of Shurpali et al. (1993) who proposed a new technique for estimating random instrument uncertainty which was popularised in the mid-nineties (Clement et al., 1995; Billesbach et al., 1998). In this approach the flux of a tracer is measured while sampling air with a constant mixing ratio e.g. directly sampling from a gas standard, and hence any observed flux, is purely a reflection of the random instrumental noise. This method has proved extremely robust, but has the obvious disadvantage of requiring routine data acquisition to stop while the random instrument uncertainty is assessed.

Billesbach (2011) proposed a more practical approach for quantifying random uncertainties from sensor noise. The so-called "random shuffle approach" involves random reshuffling of one of the variables in time and thereby removing any covariance between source/sink terms and transport, leaving only accidental correlations which can mostly be attributed to instrument noise. This is an intriguing option, yet, if we

consider a measured time series c, which is made up of some genuine signal (x) as well as instrumental noise (ε), then the effective amplitude of a time-shuffled time series is still composed of $x + \varepsilon$ and therefore the uncertainty is likely overestimated.

More recently, Mauder et al. (2013) approximated errors associated with random instrument noise by first calculating the signal-to-noise ratio of the concentration time series using an auto-covariance function and then using a basic error propagation to estimate the contribution of that noise to the uncertainty in the cross-covariance:

$$\mathrm{RE}_{\mathrm{noise}} = \sqrt{\frac{\left(\sigma_{\mathrm{C}}^{\mathrm{noise}}\right)^2 \sigma_w^2}{n}}. \quad (5)$$

In this approach, $\sigma_{\mathrm{C}}^{\mathrm{noise}}$ is the standard deviation of the instrument noise, derived using an auto-covariance function (see Sect. 2.1 for details), σ_w^2 is the variance of the vertical wind velocity and n is the number of data points in the flux averaging period. This method, also implemented by Peltola et al. (2014) and Rannik et al. (2015) is relatively simple to apply but as yet, its effectiveness has not been fully validated for use with eddy-covariance and disjunct eddy-covariance (DEC) data.

In this study we explore a further possibility for estimating the portion of the random error that is attributable to sensor noise by combining the ideas of Billesbach (2011) and Mauder et al. (2013), focusing in particular on the interplay between random instrument uncertainty, cross-covariance peak width and the systematic flux bias induced when determining the flux through the use of a cross-covariance function (Taipale et al., 2010; Laurila et al., 2012). In understanding this linkage, our aims are to (i) validate the use of Eq. (5) for use with EC and DEC data sets, (ii) outline an optimal strategy for calculating and reporting random errors, (iii) identify the optimal strategy for determining time lags for eddy-covariance data with low SNR and (iv) to draw conclusions on the validity of flux measurements made with low SNR.

2 Methods

2.1 Quantifying random white noise from analysers

Instrumental noise comes in both structured and unstructured forms. For example, the 50–60 Hz signal from a mains AC power supply might introduce a structured noise into a time series, and optical fringes often introduce periodic features in optical spectroscopic approaches. By contrast, uncorrelated white noise can result from minor fluctuations in the mechanics of instrument components, or fluctuations in temperature, pressure or humidity. Here we focus our attention on unstructured, white noise only, and define the SNR for a given time series c as

$$\mathrm{SNR} = \frac{\sigma_{x'}^2}{\sigma_{\varepsilon'}^2}, \quad (6)$$

where $\sigma_{x'}^2$ is the variance of the genuine signal of a measured time series, c ($c = x + \varepsilon$, where χ is genuine signal and ε is noise) and $\sigma_{\varepsilon'}^2$ is the variance of the noise. In order to establish the relative contributions of both signal and noise components of c we consider the auto-covariance to c' of the form

$$f_{c'c'}(t) = \frac{1}{n-t} \sum_{i=1}^{n-t} \left((x_i - \bar{x}) + \varepsilon_i\right) \left((x_{i+t} - \bar{x}) + \varepsilon_{i+t}\right), \quad (7)$$

where t is the number of data points associated with the time lag, n is the number of data points in the time series and the overbars denote averages. White noise can only contribute to the auto-covariance at $t = 0$ (Lenschow et al., 2000; Mauder et al., 2013) as it has no structure (i.e. the noise on an individual data point is uncorrelated to the noise of the adjacent data points). As the auto-covariance function moves away from zero, the contribution of instrument white noise is removed and thus the difference between $f_{c'c'}(0)$ and $f_{c'c'}(1)$ gives an estimate of random instrument noise. The underlying trend in the auto-covariance function of the genuine signal depends on the biophysical (source/sink strength) signature of the compound being measured, and the structure within the turbulence signal which is itself a function of atmospheric stability. The presence of a trend or "structure" in the auto-covariance is a sign of genuine signal in the data, whereas an auto-covariance with no underlying trend is the definition of white noise. Where a genuine signal is present, it is therefore necessary to extrapolate the positive auto-covariance function back to the zero point. This is typically done using only the first few points e.g. $f_{c'c'}(1–5)$. The noise can then be estimated as the difference between $f_{c'c'}(0)$ and $f_{c'c'}(1–5$ extrapolated) and is depicted in Fig. 1a which shows the auto-covariance applied to a time series of temperature measurements. The auto-covariance should decrease following a 2/3 power law (Wulfmeyer et al., 2010), but we find a linear extrapolation to be more appropriate. Lenschow et al. (2000) and Mauder et al. (2013) also adopted a linear extrapolation and suggest the deviation from the $2/3$ power law to result from the averaging effects of the analysers.

Using the auto-covariance function as opposed to the auto-correlation function (i.e. the normalised auto-covariance) means the calculated signal and noise are variances and retain their original units. Taking the square root gives the standard deviation of both the signal (σ_x) and noise components (σ_ε).

The auto-covariance is a convenient method when working in the time domain, but alternatives are available when analysing the data in the frequency domain. Figure 1b shows the variance spectrum for a time series of temperature measurements (T). The red line shows the spectra of unmodified

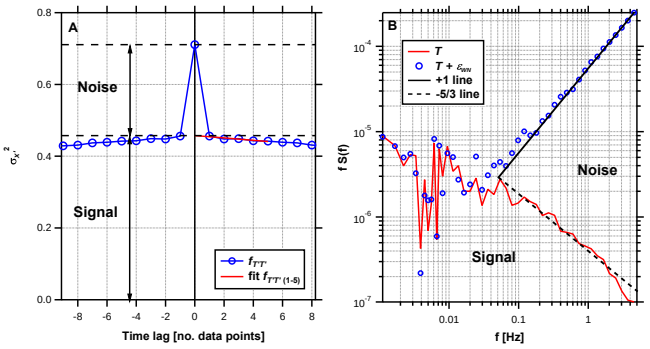

Figure 1. Illustration of the two methods for determination of analyser (temperature measurements, T) variance attributable to unstructured white noise, through **(a)** the use of an auto-covariance function in the time domain and **(b)** from the variance power spectra in the frequency domain. Panel **(a)** shows how the first few points (points 1 to 5 in this example) of the auto-covariance function can be used to extrapolate the contributions of signal and noise components at a lag of zero (red line).

temperature data from an ultrasonic anemometer and the blue line shows the same temperature data deteriorated through the addition of Gaussian white noise with a standard deviation of 1 K. In the frequency domain, on this plot, the fall-off towards higher frequencies in the inertial sub-range should follow a $-5/3$ slope (dashed line), while white noise follows a $+1$ slope (solid line). This enables the noise variance to be quantified as the area between solid and dashed lines in Fig. 1b. It should be noted that the auto-covariance method is unsuitable if the measured time series (c) is not used with the original time resolution, e.g. if it was first resampled to a different sampling frequency in order to match the sonic time resolution. Similarly, this technique does not apply to structured noise, because $f_{c'c'}(1)$ would still be affected by such noise. When using the frequency domain approach, structured noise may or may not show up as a departure from the $-5/3$ slope at high frequencies.

Throughout this paper we utilise the auto-covariance method in the time domain as it is readily applicable to both eddy-covariance and disjunct eddy-covariance data sets.

2.2 Quantifying random flux errors and the limit of detection

As discussed previously, the precision with which a flux can be measured is commonly approximated from the properties of the cross-covariance function between w' and c' ($f_{w'c'}(t)$).

For time lags much different from the true time lag (t), the standard deviation of $f_{w'c'}(t)$ provides a measure of the random error affecting the flux (Wienhold et al., 1995; Spirig et al., 2005):

$$\mathrm{RE}_{\sigma} = \sigma_{f_{w'c'}[-\Gamma,+\Gamma]}, \quad (8)$$

where Γ is a region of the cross-covariance function well away from the point of zero time lag. Typically, Γ defines two regions of the cross-covariance function, one covering the positive time shifts and the other covering the negative time shifts (e.g. $-\Gamma = -150$ to -180 s and $+\Gamma = 150$ to 180 s).

Multiplying this measure of the random error by α gives an estimate of the measurement precision at a given confidence interval ($\alpha = 1.96$ for the 95th percentile; $\alpha = 3$ for the 99th percentile) which can be used as the flux limit of detection (LoD) (i.e. $\mathrm{LoD}_{\sigma} = \alpha \times \mathrm{RE}_{\sigma}$). The flux detection limit does not only depend on the SNR of the concentration measurement, but also varies with wind speed and atmospheric stability. Therefore, for each new averaging period it is necessary to recalculate the LoD. Whilst this technique allows for the separation of a "genuine" flux signature from the general noise of the covariance, the determination of the standard deviation is often done using somewhat arbitrary boundaries (e.g. -150 to 180 s and $+150$ to 180 s, or defined as some multiple of the integral timescale; Spirig et al., 2005), and thus as the turbulence structure evolves throughout the day, these limits may become more or less appropriate. Any correlation between c' and w' within these bounds is either purely accidental and reflects the random noise in the time series or it is due to organised structures that persist over much longer time intervals suggesting that turbulence is not stationary or statistically not well covered in the measurement. Furthermore, a trend in scalar data can result in a cross-covariance which remains positive or negative over wide ranges rather than the expected fluctuation around zero. Figure 2a and b show the calculation of the LoD via the standard deviation (LoD_{σ}; blue dashed line) to sensible heat flux data from a forest site, for cases where the cross-covariance oscillates around zero and is offset from zero, respectively. In the latter case (Fig. 2b), many points of the cross-covariance function are above the limit of detection.

A modification of the LoD_{σ} approach is to calculate the random error based on the root mean squared deviation (RMSE) of $f_{w'c'}(t)$ from zero within the same specified region (greyed area – Fig. 2a and b), which reflects the variability in the cross-covariance function in these regions, but also its offset from zero as shown in Eq. (9).

$$\mathrm{RE}_{\mathrm{RMSE}} = \sqrt{\left(0.5\left(\left(\sigma_{f_{w'c'}[-\Gamma]}\right)^2 + \left(\overline{f_{w'c'}}_{[-\Gamma]}\right)^2 + \left(\sigma_{f_{w'c'}[+\Gamma]}\right)^2 + \left(\overline{f_{w'c'}}_{[+\Gamma]}\right)^2\right)\right)}, \quad (9)$$

where $\sigma_{f_{w'c'}}$ and $\overline{f_{w'c'}}$ represent the standard deviation and the average of the cross-covariance within a defined time window (Γ), respectively. Again, multiplying $\mathrm{RE}_{\mathrm{RMSE}}$ by α gives a limit of detection at a given confidence interval (i.e. $\mathrm{LoD}_{\mathrm{RMSE}} = \alpha \times \mathrm{RE}_{\mathrm{RMSE}}$).

Figure 2, panel c shows the LoD for sensible heat flux data calculated using both the $\mathrm{LoD}_{\mathrm{RMSE}}$ and LoD_{σ} methods. The two approaches agree very closely for periods where the cross-covariance fluctuates regularly around zero as in Fig. 2a, but where the covariance is predominantly of one sign, the LoD_{σ} approach derives significantly smaller lim-

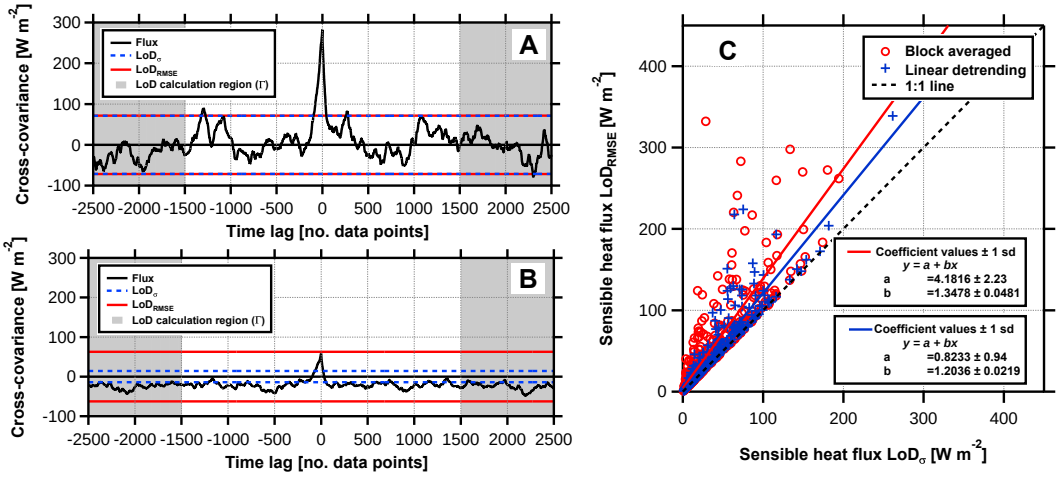

Figure 2. Cross-covariance functions for sensible heat fluxes (**a** and **b**) measured above a Douglas fir forest in Speuld, Netherlands, during two example measurement periods. Panel **(c)** shows the limit of detection for sensible heat fluxes calculated either by block averaging or linearly detrending the vertical wind velocity and temperature data. The limits of detection were calculated using both the standard deviation approach (LoD_σ) and the root mean square error approach ($\mathrm{LoD}_{\mathrm{RMSE}}$) approach. See text for further details.

its of detection, which, as shown in Fig. 2b, do not exceed the general noise of the cross-covariance function. While this does not prevent the identification of the true peak in the sensible heat flux in this example, this finding is particularly important for closed-path eddy-covariance systems in situations where the time lag between sensors is not known and signal quality is poor. In these instances, the identification of the genuine flux may be confounded by multiple regions of the cross-covariance exceeding the limit of detection and resulting in the reporting of erroneous fluxes (see Sect. 3.3.1 for further discussion on this issue). Importantly, if fluxes are calculated using linear detrending (i.e where a linear function is used to remove trends from the time series of w' and T' separately) as opposed to block averaging (where the mean of the 30 min averaging period is subtracted from w' and T', respectively), the effect is reduced somewhat but not completely removed. For this data set, 14 % of block averaged data would have been rejected using the $\mathrm{LoD}_{\mathrm{RMSE}}$ method as opposed to 4 % using the more traditional LoD_σ method. In contrast, when applying linear detrending to these data the percentage of data rejected fall to 6 and 3 % for the $\mathrm{LoD}_{\mathrm{RMSE}}$ and the LoD_σ methods, respectively. In light of these findings we utilise the $\mathrm{LoD}_{\mathrm{RMSE}}$ method for all calculations of the flux LoD in this study. Recommendations for the application of the $\mathrm{LoD}_{\mathrm{RMSE}}$ method to ozone eddy-covariance flux data are outlined in Nemitz et al. (2015).

2.3 Calculating the effect of instrument noise on the flux error

Analysis of the statistical properties of the cross-covariance function seems to offer a practical approach for approximating the total random error of the flux, because the variability of the cross-covariance function comprises both instrument noise and the variability of the (genuine) atmospheric concentration. Yet, as discussed above, isolating the instrumental component of the total random error remains a challenge. Here, we attempt to untangle the two errors using an approach similar to the "random shuffle" method of Billesbach (2011). Rather than shuffling the measured scalar time series to remove any covariance between c' and w', we generate a new time series of equal length comprised purely of Gaussian white noise, $\varepsilon_{\mathrm{WN}}$. The standard deviation of the white noise is set to match that of the instrument noise, σ_ε, which can be calculated using the auto-covariance method described in Sect. 2.1. The resulting time series shares the statistical properties of c', minus the contribution of the genuine analyser signal x' and therefore the contribution of instrument noise to the total random flux error can now be determined by applying the $\mathrm{RE}_{\mathrm{RMSE}}$ method to the cross-covariance of $f_{w'\varepsilon'_{\mathrm{WN}}}$

The four steps of this method are summarised as follows:

1. Perform an auto-covariance of c' to obtain the standard deviation of the instrument noise ε (e.g. σ_ε as described in Eq. 7).
2. Generate a time series of white noise ($\varepsilon_{\mathrm{WN}}$) with a standard deviation matching that of the instrument noise (e.g. $\sigma_{\varepsilon_{\mathrm{WN}}} = \sigma_\varepsilon$).
3. Calculate the cross-covariance $f_{w'\varepsilon'_{\mathrm{WN}}}$.
4. Apply the $\mathrm{RE}_{\mathrm{RMSE}}$ method to $f_{w'\varepsilon'_{\mathrm{WN}}[\Gamma]}$ to obtain the instrumental random error ($\mathrm{RE}_{\mathrm{noise}}$). Here, Γ represents a time window of 0–30 s.

In theory, this numerical exercise seeks to quantify the same error approximated by Mauder et al. (2013; Eq. 5), whilst

making no assumptions on the shape of the distribution of w'. Therefore we can use this approach to validate Eq. (5) for use with both EC and DEC data sets and assess its performance when applied to data sets with varying levels of SNR. The random error in the flux, attributable to instrument noise, is calculated using both these approaches and these results are compared in Sect. 3.1. As mentioned previously, the auto-covariance technique, also used by Mauder et al. (2013), is not sensitive to the effects of structured instrument noise. Consequently, the instrumental random error reported by both methods is likely underestimated. Nonetheless, the contribution from structured noise is included in our estimate of the total random error (i.e. by applying the $\mathrm{RE_{RMSE}}$ method to $f_{w'c'}$) and it is this parameter that is used to define the flux limit of detection (i.e. $\mathrm{LoD_{RMSE}} = \alpha \times \mathrm{RE_{RMSE}}$).

Although this proposed technique does not make any assumption about the distribution of w', it does make the assumption that the instrument noise follows a Gaussian distribution which is not always the case. For example, concentrations have often been found to be skewed towards larger values. In particular, tracers which show a high degree of variability at low mean concentrations may be log-normally distributed: concentrations are not constrained towards larger values, but cannot physically be negative, and therefore typically follow a log-normal distribution. The combined frequency distribution of w' and c' has been found to be more closely approximated by a Gram–Charlier equation than a Gaussian distribution (Milne et al., 2001). In addition, the statistical noise generated by instruments that derive concentrations from count events, e.g. counting particle number, ions (as the PTR-MS does for VOC fluxes or the aerosol mass spectrometer (AMS) for submicron aerosol chemical fluxes), follows a Poisson distribution. This Poisson distribution is again limited towards small values by zero, whilst it is not constrained towards larger values. This potentially introduces an asymmetry in the response, and it is not immediately clear that this cannot act on the flux in a different way than a normally distributed instrument noise.

With this in mind we performed several tests to determine if the covariance between the vertical wind velocity and a time series of white noise differs depending on the distribution of that noise (e.g. whether it is Gaussian, Poisson or log-normally distributed). For a single 30 min averaging period the covariance between w' and ε' was calculated iteratively, whereby the artificially generated noise ($\varepsilon_{\mathrm{WN}}$) was either Gaussian, Poisson or log-normally distributed as seen in Fig. 3a. For each of the 5000 iterations, a new time series of white noise was generated and the covariance was recalculated using a prescribed time lag. Figure 3b shows a distribution of the resulting white noise fluxes calculated using either Gaussian, Poisson or log-normally distributed noise. In each case, the fluxes are evenly spread about zero, which demonstrates that unstructured white noise creates a random uncertainty in the flux but does not induce a systematic bias, regardless of its distribution. In our calculation of the random

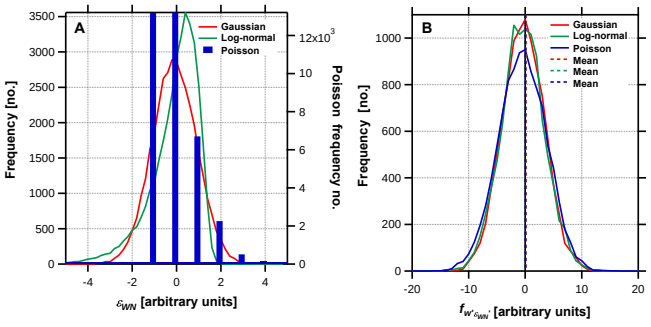

Figure 3. Panel **(a)** shows the frequency distributions of Gaussian, log-normal and Poisson noise ($\varepsilon_{\mathrm{WN}}$) of identical standard deviation. Panel **(b)** shows the frequency distributions of the flux calculated from the unstructured white noise over a period of 5000 iterations. The mean average flux for each noise distribution are marked with dashed lines, which are all close to zero and consistently confirm that no systematic bias is introduced to a flux measurement regardless of the type of noise used.

instrumental flux error, we have chosen to generate white noise with a Gaussian distribution. This simple test clearly demonstrates that our choice of noise distribution is actually unimportant and that the same result will be obtained, regardless of the distribution of the noise. Importantly, these findings provide assurances that flux biases are not created for eddy-covariance systems that induce a Poisson counting noise.

These findings confirm the theoretical considerations of Lenschow and Kristensen (1985), that, if the time lag is known, the presence of uncorrelated noise induces a random uncertainty in the flux but does not induce a systematic bias. Nonetheless, this conclusion does not consider the interplay of this noise with the determination of a time lag, which is vital for sensors that are spatially separated from the vertical wind velocity measurement, and its potential to introduce a bias that is a function of the random error.

2.4 Effect of instrument noise on time-lag determination

2.4.1 Signal-to-noise simulations

In order to investigate the influence of unstructured white noise from analysers and the method of time-lag determination on calculated fluxes, a series of simulations were performed using 31 days of sensible heat flux data (see the Supplement). Time lags were determined using the three main methods outlined above, MAX, AVG and PRES. For the AVG method, a further ten scenarios were implemented, whereby the running mean applied to the cross-covariance was increased from 0.5 to 5 s in 0.5 s intervals. In all scenarios the time lag was sought within a 10 s window which ranged from -5 to $+5$ s, with the true time lag obviously being 0 s. For the AVG method, the running mean was ap-

plied over a larger window to ensure the mean was properly centred for all data in the 10 s window. To accurately control the analyser noise level, the SNR of the temperature data was first quantified using the auto-covariance approach outlined above. The signal was then deteriorated by adding Gaussian white noise until a target signal-to-noise ratio was achieved to within a 1 % tolerance. Sensible heat fluxes were calculated using block averaging and a reference flux was determined by calculating the flux with zero lag from the unmanipulated temperature time series. This process was repeated 10 times for temperature data with a signal-to-noise ratio (SNR) ranging between 200 and 0.05. In addition, the above simulation was repeated three more times to assess the impact of adopting disjunct sampling protocols of 2.5, 5 and 7.5 s as is common for measurements of VOC fluxes by PTR-MS and aerosol fluxes by Q-AMS (quadrupole-AMS). The overall bias between simulated time series and the reference was determined by the gradient of the regression between 31 days of reference data versus those of the simulations.

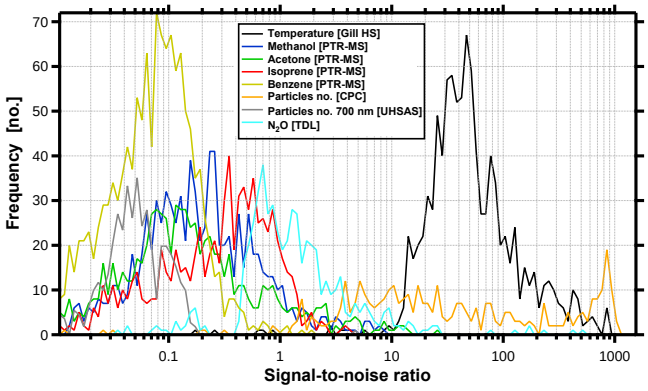

Figure 4. Signal-to-noise ratios of typical instruments used for the flux measurement of various trace gases and aerosols. Analysers include a sonic anemometer (temperature) and PTR-MS (isoprene, methanol and acetone) operated above a mixed oak forest at a height of 32 m and above a city (benzene). Particle number concentrations were measured by a CPC and UHSAS (single size bin) above a Douglas fir forest and N_2O measurements were made above a grassland field in 2003 using a first-generation tuneable diode laser.

2.4.2 Peak width simulation

The covariance between the genuine signals of c'(e.g $.x'$) and w' gives a peak with respect to time lag in the cross-covariance function. Superimposed on top of this peak are the contributions from the covariance of the error components of c' and w', e.g. $w'_\varepsilon c'_\varepsilon$. It stands to reason that the broader the peak, the greater the probability of detecting an extreme local maximum i.e. detecting a peak in $w'_\varepsilon c'_\varepsilon$ on top of the genuine covariance, $w'c'$. Consequently, the peak width of the genuine covariance becomes an important consideration when assessing the potential bias induced through the choice of time-lag determination. In order to assess the sensitivity of flux measurements to the peak width, we ran a second set of simulations on two identical artificially generated chirp signals (a signal that decreases in frequency over time). This type of signal, generated within LabVIEW (National Instruments, Austin, Texas, USA), was chosen as a convenient means of producing time series to represent the "genuine" signals of x' and w' which contained multiple frequencies. In a series of 12 iterations, Gaussian white noise (ε_{WN}) was added to x' until a target signal-to-noise ratio was met. In a second round of iterations the initial frequency of the chirp signal was decreased from 0.075 to 0.005 Hz in 0.005 Hz intervals. As the frequency was reduced, the full width at half the maximum (FWHM) of the covariance peak increased, resulting in a matrix of fluxes calculated using 12 signal-to-noise ratios (100–0.05) vs. 15 covariance peak widths, ranging between 0.9 and 12 s (FWHM). For each point in the matrix, the error relative to the flux, calculated from the unmodified chirp signals, was calculated.

2.5 Real world data

Simulating the effects of different time-lag determination methods and the effects of SNR is a useful exercise, but it is important to verify that these are representative of real world data. We assessed the performance of each lag method on example data from a variety of analysers operated in field experiments with varying levels of signal-to-noise, bearing in mind that the SNR for a given application will depend on concentrations and instrument operation. The analysers included an ultrasonic anemometer (Gill HS-50), a condensation particle counter (CPC, TSI Model 3776), an ultra-high-sensitivity aerosol spectrometer (UHSAS, PMS, Boulder, USA), a tuneable diode laser (TDL, Aerodyne Research Inc.) as well as disjunct data from a proton transfer reaction mass spectrometer (PTR-MS; Ionicon, Innsbruck, Austria). Figure 4 shows the frequency distribution of the SNR of 30 min averaging periods for each of the analysers.

A detailed description of each of the data sets used is supplied in the Supplement.

3 Results and discussion

3.1 Calculation of random flux errors

The procedure for calculating random errors using the auto-covariance approach was applied to EC fluxes of sensible heat (A) and DEC fluxes of isoprene (B) and acetone (C), and the results are shown in Fig. 5. The error bars denote the total random error obtained from the cross-covariance function (e.g. RE_{RMSE} applied to $f_{w'c'}$), and the central panels show how that error is divided between the random instru-

ment error (RE_{noise}) and natural variability of turbulence and the (genuine) scalar concentrations (RE_{var}). For the fluxes of sensible heat, the random instrumental error is very low, reflecting the excellent SNR of the sonic anemometer. In contrast, the instrumental error for fluxes of acetone is very large, likely due to the very low atmospheric concentrations of acetone. For isoprene, a clear diurnal cycle is visible which shows the relative contribution of the instrument error is largest at night when the isoprene signal is lowest and the signal-to-noise ratio is low. During the daytime, isoprene concentrations increase, improving the SNR of the analyser which sees the error due to natural variability of the (genuine) atmospheric concentration become the dominant source of uncertainty in the flux measurement. Overall, absolute errors are larger during the day, when turbulence is larger.

The lower panels show scatter plots of the random instrument error calculated using the numerically calculated Gaussian white noise flux versus the analytical approximation of Mauder et al. (2013; Eq. 5). The two methods give consistent results to within a few percent for both eddy-covariance and disjunct eddy-covariance data sets. Therefore, implementation of either method can enable operators to estimate the minimum detectable flux for their analyser under a given turbulence regime.

3.2 Bias effects of different time-lag determination methods

3.2.1 Dependence on signal-to-noise ratio

The simulations applied to sensible heat flux data reveal a distinct relationship between the signal-to-noise ratio of the raw temperature data and the relative flux bias for both the MAX and AVG lag determination methods. Figure 6a shows the results for 10 Hz eddy-covariance data. It is immediately apparent that methods that systematically search for a maximum (red trace) induce an average positive bias (towards more extreme emission or deposition) to the reported flux which increases linearly as the analyser signal deteriorates. For this data set, the relative bias can be as much as 18 %. Adopting the AVG method can significantly reduce this error provided the applied running mean is of a suitable length. However, selection of an inappropriate running mean may allow the bias to persist and can also become negative when overly long. The reason for the negative result lies in the fact that the shape of the peak in the covariance spectrum tends to be skewed, while the running average of the AVG peak fit is symmetrical. However, theory cannot currently explain the skewness which is therefore difficult to predict. By contrast, the use of a prescribed time lag (for the anemometer temperature data, the time lag is known to be zero), uncertainty increases as the signal is deteriorated more and more, but to a smaller degree, and its sign is random.

Figure 6b and c show the same set of simulations for fluxes calculated using the disjunct eddy-covariance method using sampling intervals of 2.5 s (panel b), 5 s (see Supplement) and 7.5 s (panel c), respectively. It is well understood that adopting a disjunct sampling approach reduces the statistical sample size and thus increases the random error (Lenschow et al., 1994; Rinne and Ammann, 2012). However, it is frequently assumed that the increased random error does not translate to a systematic bias in the measured fluxes, but our simulations show this not to be the case. The poor sampling statistics and high instrument noise combined with the MAX method for time-lag identification can potentially lead to 100 or even 200 % overestimation in the mean flux.

The method used to determine the time lag is a key factor in accurately resolving the flux as already demonstrated with the 10 Hz eddy-covariance data. Additional random uncertainty incurred from disjunct sampling amplifies the bias at signal-to-noise ratios less than 100, and in this instance resulting in relative errors of about 300 % at SNR = 0.01 where Δt is set to 7.5 s. Importantly, this offset appears avoidable if a prescribed time lag is used. When using a prescribed time lag, individual flux measurements are biased either high or low compared with the standard eddy-covariance approach that uses a larger number of data points, but when analysed collectively, the net error is close to zero. This implies that where the signal-to-noise ratio is very high it may be necessary to average over more data to reduce the relative random error and thus ensure no systematic bias is introduced.

These findings come with the caveat that in these simulations the prescribed time lag was a known quantity. When applied to real world data the adopted time lag must be a well-defined parameter that does not drift significantly over time. Failure to meet this requirement would undoubtedly result in a systematic underestimation of the flux, the magnitude of which would become a function of the cross-covariance peak width. This is discussed further in Sect. 3.2.2.

Of equal importance is the magnitude of the expected flux. Fluxes may be large even if the scalar mixing ratios are very noisy and thus the relative error is dependent on both the signal-to-noise ratio and the magnitude of the flux. Thus, although Fig. 6 describes the behaviour of the bias, the exact values depend on the magnitude of the fluxes and also the structure of the underlying turbulence data. However, these simulations do serve to highlight those aspects that make flux data more vulnerable to systematic errors.

3.2.2 Influence of peak width

Figure 7 shows the results of the peak width simulations on two artificially generated, multi-frequency signals. As well as reiterating the increase in the relative error associated with the analyser signal-to-noise ratio, this plot serves to demonstrate that the FWHM of the covariance peak is an equally important parameter. Broader covariance peaks, reflecting slower turbulence/larger eddies, result in a higher probability of an extreme maximum (i.e. the true cross-covariance between w' and c' plus random noise) being chosen and there-

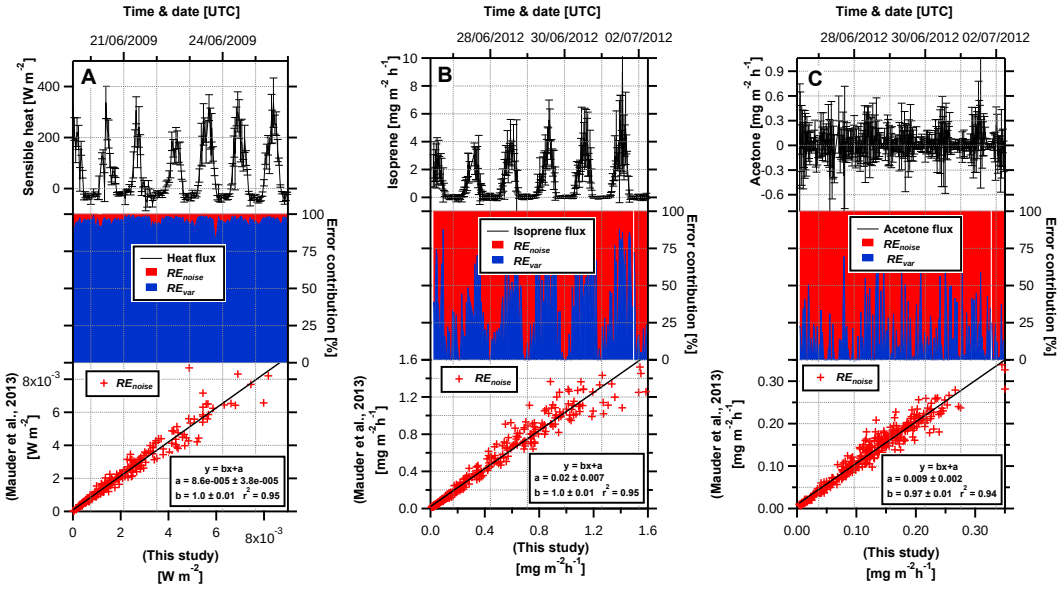

Figure 5. Sensible heat **(a)**, isoprene **(b)** and acetone **(c)** fluxes and accompanying errors. Upper panels show the measured fluxes with error bars denoting the total random error (RE). The central panels show how the total random error can be divided into errors associated with instrument noise (red, RE_{noise}) and the variability in turbulence at the genuine atmospheric concentration (blue, RE_{var}). The lower panels show scatter plots of the random instrument noise calculated using the analytical approximation of Mauder et al. (2013) and the numerical method outlined in this study.

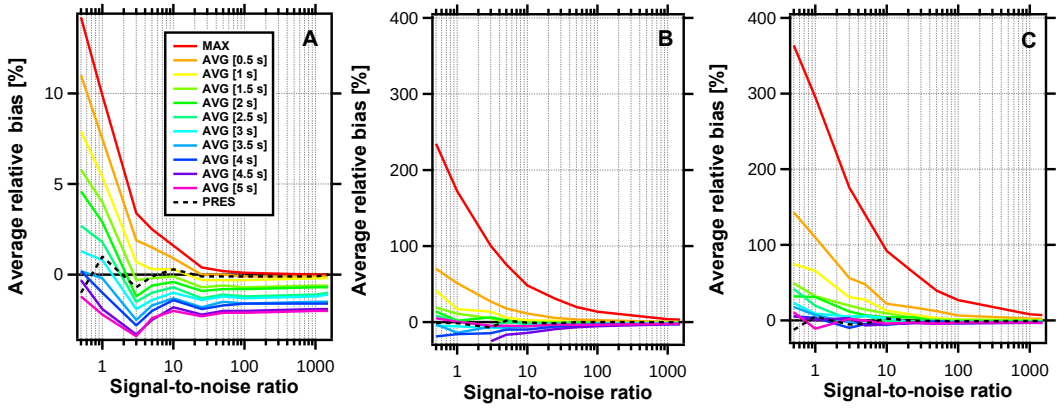

Figure 6. Panel **(a)** shows the average relative bias of a half-hourly flux as a function of the analyser signal-to-noise ratio for 31 days of 10 Hz eddy-covariance sensible heat flux data acquired at a height of 32 m. The signal-to-noise ratio of the temperature data was deteriorated to match predefined limits. The errors shown are relative to the sensible heat flux calculated using the unmodified temperature data and a constant (0 s) time lag. Panels **(b)** and **(c)** show the same plot for disjunct eddy-covariance data with sampling intervals of 2.5 and 7.5 sm respectively.

fore both the probability of overestimating the flux and the magnitude of the bias are closely linked to the peak width. The increase of mean eddy size with height means trace gas and aerosol flux measurements at higher elevations above ground are more at risk to systematic bias when the MAX or AVG methods are employed. Conversely, the probability of systematic underestimation of the flux through the use of a prescribed time lag at these measurement heights is somewhat reduced due to a greater tolerance afforded by the increased peak width. Massmann (2000) and Hörtnagl et al. (2010) recognised that further broadening of the covariance peak is possible through the attenuation of scalars through long inlet lines and also through an increase in the analyser integration time. Therefore, when working on tall towers above forests or urban canopies, one should be aware of the greater potential for systematic bias and should consider the use of a prescribed time lag which may provide the most representative (least biased) estimate of the flux. In

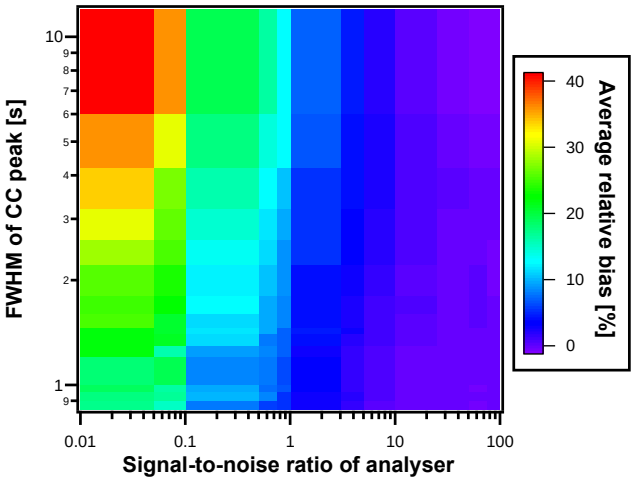

Figure 7. Image plot depicting the relationship between the average relative bias and both the signal-to-noise ratio of the analyser and the full width half maximum (FWHM) of the cross-covariance function peak for simulated eddy-covariance flux data calculated using the MAX time-lag method.

general terms, when sampling at lower heights such as above crops or grassland, the potential for bias is lessened, whereas the likelihood of underestimating the flux through the use of a prescribed time lag is increased.

When using a prescribed time lag, the attenuation of samples through long inlet lines, adsorption/desorption effects and fluctuations in pump flow rate are not typically considered. More often, the prescribed value is chosen purely on the basis of the inlet dimensions and a spot measurement of the flow rate, or through a single test where a pulse in concentration and wind speed is created near the anemometer/inlet. In these cases the potential for underestimating the flux is large. It is therefore good practice to initially search for the time lag using the AVG or MAX method and to plot the results as a histogram or time series. This may confirm that the time lag was indeed constant or it may reveal a clear peak or trend in time lags which can be used to set the prescribed value. For instruments that measure multiple species (e.g. mass spectrometers, optical spectrometers), it may be suitable to use the average time lag of a species that shows a clear flux (and thus clear time lags) as a proxy for the other compounds being measured. However, difference between gases in terms of solubility and therefore adsorption/desorption characteristics need to be considered. For example, it is well known that for closed-path sensor measurements of CO_2/H_2O a longer time lag is found for H_2O than for CO_2 (Ibrom et al., 2007). While most appropriate for instruments that measure multiple species simultaneously, this approach can also be applied to instruments that measure species sequentially (e.g. the quadrupole-based PTR-MS or AMS) as long as the scan cycle is accounted for when assigning time lags.

3.3 Real world flux measurements with low SNR

3.3.1 Mirroring

When measuring trace gas and aerosol fluxes, the fast sampling requirements of eddy covariance can result in low SNRs. Working in this region can see the random flux error equal or even exceed the magnitude of the flux, potentially introducing a bias as discussed above. In addition to these effects, where a maximum in the cross-covariance is still sought, the derived flux may switch between emission and deposition values of similar magnitude. This phenomenon, which we term "mirroring", is observed in the example flux data shown in Fig. 8 which were obtained using TDL, UHSAS and PTR-MS instruments and occurs because the random error in the flux is sufficiently large to span the zero line. It may be tempting to remove the negative (positive) fluxes on the basis of biophysical implausibility. For example, when measuring aerosol fluxes where only deposition fluxes are expected, it would be easy to dismiss positive fluxes as artefacts. Nevertheless, removal of these points is clearly incorrect and would introduce a positive bias to the reported average data (cf. Nemitz et al., 2002). Under such circumstances the reported flux is predominately driven by fluctuations in the amount of turbulence which evolves throughout the day to give a diurnal cycle which might be mistaken for a flux. Adopting the MAX or AVG methods exaggerates the mirroring by systematically choosing the furthest point away from zero which in the extreme case can result in the very unnatural flux distributions shown in Fig. 9. Adopting the AVG method with 5 s running mean limits this effect to a certain extent, but a noticible dip around zero remains. Importantly, the use of a prescribed time lag eliminates the splitting of data from either side of zero to give a much more natural looking flux distribution. For many compounds an assessment of the frequency distribution of flux data evaluated with the MAX method will highlight whether mirroring occurs and whether this approach is therefore not applicable. Care needs to be taken, however, when making this type of assessment on CO_2 fluxes, as the interplay between strong positive fluxes during the night and negative fluxes during the day could potentially result in a similar bimodal distribution. Figure 9 also nicely illustrates the flux bias introduced by using time-lag methods that systematically search for a maximum in the cross-covariance. In this instance the MAX method gives a mean flux 2.3 times larger than the PRES method.

In addition to the calculated fluxes, the red time traces in Fig. 8 show the Gaussian white noise flux ($f_{w'\varepsilon'_{WN}}$), i.e. the cross-covariance between an artificially generated white noise signal that shares the same standard deviation as the analyser noise. Here, both tracer and Gaussian noise fluxes have been calculated using the MAX method. For acetone, $f_{w'\varepsilon'_{WN}}$ is of a similar magnitude which indicates that in this example, the fluxes shown are almost entirely due to coincidental covariance between the vertical wind velocity and in-

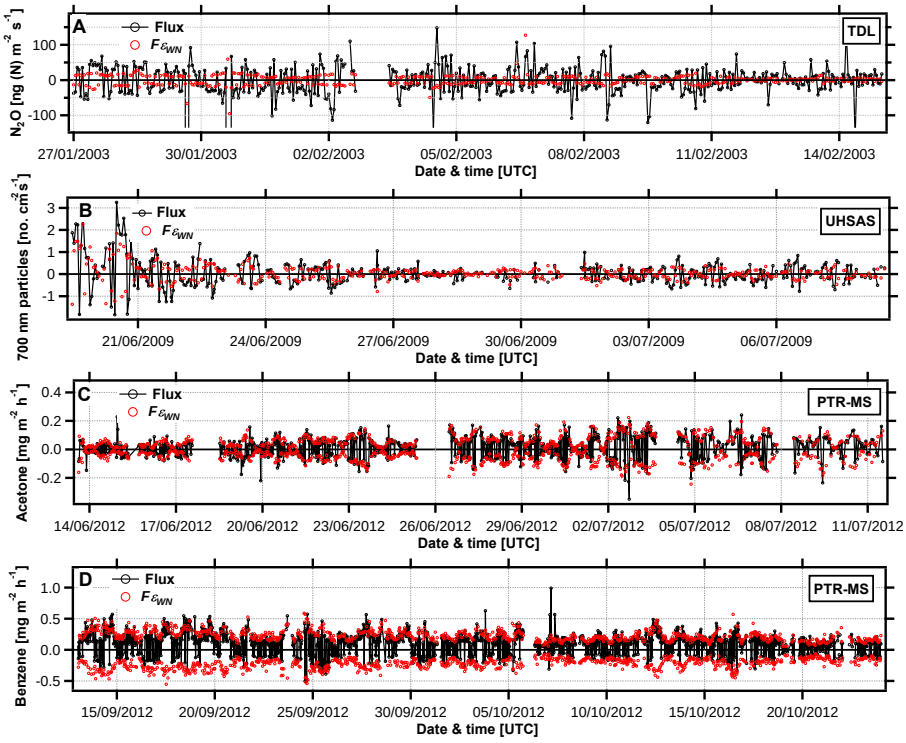

Figure 8. An example of "mirroring" in eddy-covariance data with low SNR processed with the MAX time-lag method. The data were obtained by TDL (panel **a**), UHSAS (data from a single size bin – panel **b**) and PTR-MS (panels **c** and **d**) instruments during four separate measurement campaigns. Red circles show a flux calculated from a cross-covariance between a time series of Gaussian white noise ($\varepsilon_{\mathrm{WN}}$) and the vertical wind velocity (w). The standard deviation of the white noise was set to match that of the instrument noise for each of the analysers and the flux was determined by searching for a maximum in the cross-covariance.

strument noise (i.e. the random error in the flux is completely dominated by instrumental noise). In contrast, the range of benzene and particle number fluxes both at least partially exceed the Gaussian white noise flux and show a sustained period of emission fluxes (e.g. 17 to 20 September) indicating the presence of a "genuine" flux which is, for certain periods, distinguishable from the random sensor noise flux. The remaining data would undoubtedly fall below conventional limits of detection and the individual 30 min flux measurements would ordinarily be rejected. Yet, the question remains whether any useful information on the net exchange can still be extracted from data such as these and is discussed in detail in Sect. 3.3.2. Finally, the TDL N_2O fluxes are consistently larger than the Gaussian white noise flux despite an apparent mirroring in the data. In this case it is likely that the instrument noise is comprised of both unstructured white noise and structured noise, perhaps from optical fringes which our method does not take into account. It should be noted, that we tried here to identify data series which showed the effects of limited SNR. All these instruments obviously can perform better in situations where fluxes are larger or where instrumentation parameters are further optimised.

3.3.2 Limit of detection for individual and averaged fluxes

For data where mirroring is observed, there are either no fluxes present or insufficient statistics to resolve them. If these data are to be utilised at a 30 min time resolution then they are of little use and should be rejected. In some cases, extending the averaging period may provide the additional statistical information required for resolving the flux, but it is also increasingly likely to violate the requirements for stationarity. Yet, in the literature, measured fluxes are seldom utilised at the resolution with which they are collected, but are more typically aggregated either to establish longer term budgets, by time of day or by a meteorological parameter such as light or temperature, in order to establish robust relationships for model parameterisations. Where data are averaged, presented and utilised in this way, the statistical significance of the average can be evaluated against the LoD of the ensemble average ($\overline{\mathrm{LoD}}$) calculated from the LoDs of the N individual data points that entered the average as

$$\overline{\mathrm{LoD}} = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\mathrm{LoD}^2}. \quad (10)$$

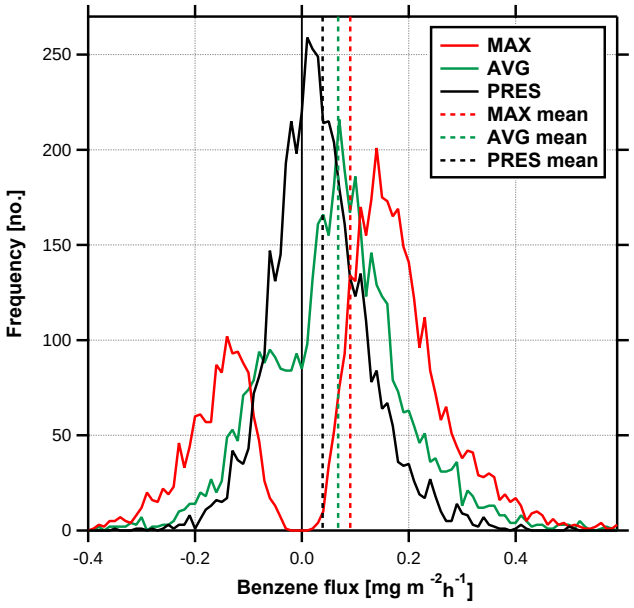

Figure 9. Distributions for benzene fluxes calculated using the MAX, AVG (5 s) and PRES time-lag methods. The benzene concentration data had an average signal-to-noise ratio of 0.09 and ranged between 0.007 and 0.24 at the 5th and 95th percentiles.

For this reason, data that fail averaging period specific LoD criteria, such as the time series shown in Fig. 8, should not be discounted out of hand as they may retain useful information on the net exchange when averaged, but they need to be reported together with an estimate of the random error. As the random errors are typically calculated on the raw data, any corrections applied to the calculated fluxes (e.g. low or high frequency loss corrections) should also be applied to the random error.

Figure 10a shows the averaged diurnal fluxes of the acetone time series shown in Fig. 8c for 1, 7, 14 and 21 day periods (Acton et al., 2015). The shaded areas represent the averaged $\mathrm{LoD}_{\mathrm{RMSE}}$ for the same period. Data falling within the shaded area cannot be resolved by the measurement system, but those points falling outside this area are statistically significantly different from zero at the 95th percent confidence interval. The fluxes were determined using both the MAX and PRES time-lag methods, with the latter based on the average isoprene time lag (plus the duty cycle offset). We have already demonstrated that using the MAX method highly biases the fluxes and consequently after 2 weeks of data are averaged, many of the fluxes appear to exceed the $\overline{\mathrm{LoD}}$. In contrast, those fluxes calculated with a prescribed time lag, which limits the bias, fail to exceed the $\overline{\mathrm{LoD}}$ even when averaged over a 3 week period and would be rejected. This serves as an important example of how the choice of time-lag determination can lead to the reporting of a flux which in essence is not statistically different from zero.

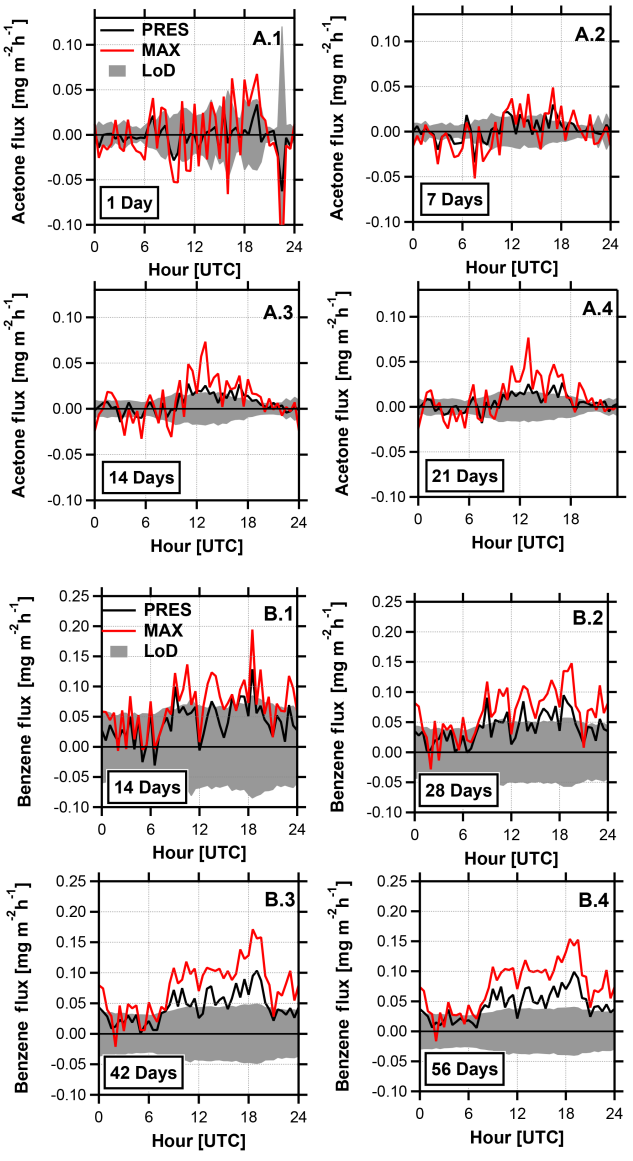

Figure 10. (a) Averaged diurnal profiles of acetone flux data obtained using a quadrupole PTR-MS. Increasing the number of averaged data points does not bring the PRES flux above the $\overline{\mathrm{LoD}}$ (greyed area indicates the $\overline{\mathrm{LoD}}$ at the 95th percentile), indicating no detectable flux. By contrast, the MAX fluxes, which are highly biased, show some periods above the $\overline{\mathrm{LoD}}$ which are an artefact of the MAX time-lag determination method. **(b)** Averaged diurnal profiles of benzene flux data obtained using a quadrupole PTR-MS. Increasing the number of data points averaged from 14 to 56 is sufficient to distinguish the flux calculated with a prescribed time lag from the LoD (greyed area indicates the LoD at the 95th percentile) and indicates a clear flux of benzene.

Figure 10b shows the same plot for the much longer time series of benzene (Valach et al., 2015). In this case we observe how the averaged fluxes eventually exceed the $\overline{\mathrm{LoD}}$ as the number of samples are increased. With this in mind, when targeting trace gas or aerosol fluxes with instrumenta-

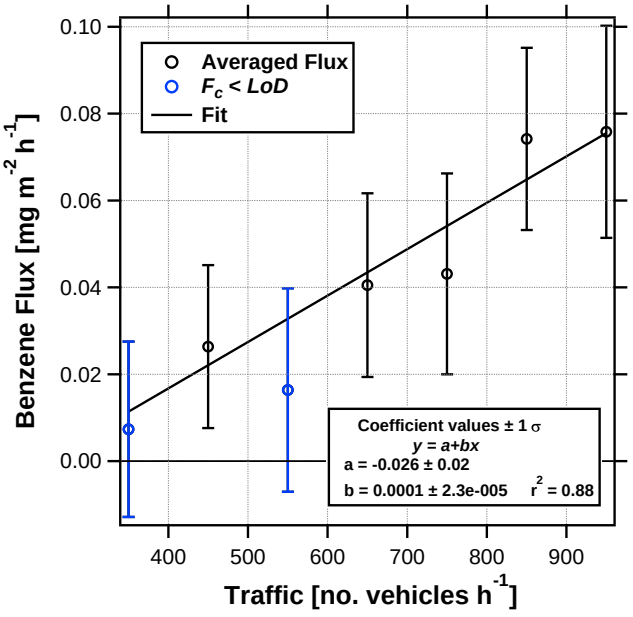

Figure 11. Benzene flux measurements calculated using a prescribed time lag and averaged as a function of traffic density. Error bars represent the ensemble LoD. In cases where the error bars intersect zero (blue points), the flux cannot be considered significantly different from zero.

tion limited in signal or where the expected fluxes are small, it may be prudent to attempt to measure for longer to ensure that statistically robust estimates on the net exchange are obtained. The improvement of LoDs for ensemble data does not only apply to temporal patterns. For example, Fig. 11 illustrates how benzene flux data can be averaged as a function of traffic density in order to parameterise emission rates from vehicles. Provided that the 30 min LoDs are averaged according to Eq. 10, flux data with limited SNR can be used in numerous ways that should go beyond the reporting of average diurnal emission rates. In this example, the ensemble LoD is shown as individual error bars and thus where those bars intersect the zero line, the flux is not significantly different from zero (e.g. the blue points). Rather than eliminating these data from the fit shown in Fig. 11 which might introduce a bias, our recommendation is to down weight these points according to their relative uncertainties.

4 Conclusions and recommendations

In this study we have carefully examined several key factors affecting the analysis of flux data with high noise level. Clearly, the effect of instrument noise on flux measurements has been studied before. Here we have developed a technique to quantify the uncertainty due to sensor white noise by first quantifying the amount of noise and then calculating a flux with this noise level. This numerical approach has been used to validate the approximations of Mauder et al. (2013) and shows consistent results when applied to both EC and DEC data sets. Both these methods can be easily implemented into eddy-covariance processing software and share the key advantage over the more traditional experimental approach of Shurpali et al. (1993) that measurements do not need to be interrupted for the assessment to take place. Nonetheless, it is important to reiterate that both of these approaches are not sensitive to the effects of structured noise, which cannot be quantified using the auto-covariance method (e.g. Eq. 7) upon which each of these approaches are based.

Most of the earlier analyses of random errors have been carried out under the assumption that the time lag between wind and concentration measurement is known. To our knowledge, the systematic bias introduced through the interplay between random sensor noise and the techniques used to determine the time lag has so far not been studied very systematically, although the problem has been highlighted in general terms in several publications and textbooks. Taipale et al. (2010) studied the effect of routines that are based on maximising the absolute value of the cross-covariance for disjunct data produced by PTR-MS and introduced the AVG approach to reduce this effect. We show here that, in general, the effectiveness of this approach depends on the length of the running mean chosen and the shape of the peak in the cross-covariance function.

Our work highlights the benefit of constraining the time lag of the air sampling and quantifying it as precisely as possible by external means when working with noisy sensors. In practical terms, this might mean controlling the inlet flow carefully and heating the inlet line to minimise adsorption/desorption effects, or deriving the time lag from a simultaneously measured compound with better SNR. Here we compile a list of general recommendations for the collection and processing of eddy-covariance data with limited SNR.

1. Where possible, log anemometer and scalar data to a single computer. This eliminates uncertainty in time lags due to clock drift and should restrict time lags to positive time shifts.

2. In-line flow meters should be used to monitor and record fluctuations in pumping speeds.

3. Use pressure controllers to limit fluctuations in sample flow rate (this may not always be possible when high flow rates are required to maintain turbulent flow and increases the required pumping capacity and therefore power consumption).

4. For water-soluble trace gases, heating of the entire inlet line should be considered to limit adsorption/desorption effects. Care needs to be taken not to generate aerosol evaporation artefacts for trace compounds that are distributed between the gas and aerosol phase according to a temperature dependent equilibrium.

5. Online monitoring of sample humidity may be necessary to account for adsorption desorption effects.

6. Use of the MAX method to generate an initial histogram or time series of time lags, followed by a second analysis using the PRES approach with a thus informed predefined time lag may be preferable to either estimating the time lag based on sampling flow rates alone or using the MAX method for final processing. This is because time lags estimated from the sampling flow do not consider the potential for a phase shift in the cross-covariance due to either signal attenuation or limited response of the instrumentation (Massman, 2000; Hörtnagl et al., 2010).

7. When reporting processed fluxes, results should be reported even if they are below the single-flux LoD, as long as they fulfil other quality control criteria. However, each individual flux value should be reported with its own quantification of the random uncertainty, so that uncertainties can be combined when fluxes are averaged.

We conclude that a significant number of fluxes (and derived values such as emission factors) reported in the literature are biased towards larger values (more distant from zero), because insufficient attention has been given to the way the time lag was estimated. By contrast, we demonstrate here the value of individual flux measurements, even if they individually fall below the LoD, for obtaining statistically significant longer-term averages, as long as care has been taken during data processing. In particular, fluxes below the LoD for the individual flux measurement should not be evaluated with the MAX method.

The Supplement related to this article is available online at doi:10.5194/amt-8-4197-2015-supplement.

Acknowledgements. We thank the members of the ECLAIRE flux community and three anonymous reviewers for their useful comments and recommendations. This work was funded by the EU FP7 grant ECLAIRE (no. 282910) and through the NERC grants ClearfLo (NE/H003169/1) and CLAIRE-UK (NE/I012036/1). W. J. Acton acknowledges financial support from BBSRC and Ionicon Analytik GmbH through the award of an industrial CASE studentship.

Edited by: S. Malinowski

References

Acton, W. J. F., Schallhart, S., Langford, B., Valach, A., Rantala, P., Fares, S., Carriero, G., Tillmann, R., Tomlinson, S. J., Dragosits, U., Gianelle, D., Hewitt, C. N., and Nemitz, E.: Comparison of three methods to derive canopy-scale flux measurements above a mixed oak and hornbeam forest in Northern Italy, Atmos. Chem. Phys. Discuss., submitted, 2015.

Ahlm, L., Nilsson, E. D., Krejci, R., Mårtensson, E. M., Vogt, M., and Artaxo, P.: Aerosol number fluxes over the Amazon rain forest during the wet season, Atmos. Chem. Phys., 9, 9381–9400, doi:10.5194/acp-9-9381-2009, 2009.

Billesbach, D. P.: Estimating uncertainties in individual eddy covariance flux measurements: A comparison of methods and a proposed new method, Agr. Forest Meteorol., 151, 394–405, 2011.

Billesbach, D. P., Kim, J., Clement, R. J., Verma, S. B., and Ullman, F. G.: An intercomparison of two tunable diode laser spectrometers used for eddy correlation measurements of methane flux in a prairie wetland, J. Atmos. Ocean. Technol., 15, 197–206, 1998.

Businger, J. A.: Evaluation of the accuracy with which dry deposition can be measured with current micrometeorological techniques, J. Climate Appl. Meteorol., 25, 1100–1124, 1986.

Clement, R. J., Verma, S. B., and Verry, E. S.: Relating chamber measurements to eddy-correlation measurements of methane flux, J. Geophys. Res.-Atmos., 100, 21047–21056, 1995.

Coyle, M., Nemitz, E., Storeton-West, R., Fowler, D., and Cape, J. N.: Measurements of ozone deposition to a potato canopy, Agr. Forest Meteorol., 149, 655–666, 2009.

Eugster, W., Zeyer, K., Zeeman, M., Michna, P., Zingg, A., Buchmann, N., and Emmenegger, L.: Methodical study of nitrous oxide eddy covariance measurements using quantum cascade laser spectrometery over a Swiss forest, Biogeosciences, 4, 927–939, doi:10.5194/bg-4-927-2007, 2007.

Famulari, D., Nemitz, E., Di Marco, C., Phillips, G. J., Thomas, R., House, E., and Fowler, D.: Eddy-covariance measurements of nitrous oxide fluxes above a city, Agr. Forest Meteorol., 150, 786–793, 2010.

Farmer, D. K., Kimmel, J. R., Phillips, G., Docherty, K. S., Worsnop, D. R., Sueper, D., Nemitz, E., and Jimenez, J. L.: Eddy covariance measurements with high-resolution time-of-flight aerosol mass spectrometry: a new approach to chemically resolved aerosol fluxes, Atmos. Meas. Tech., 4, 1275–1289, doi:10.5194/amt-4-1275-2011, 2011.

Farmer, D. K., Chen, Q., Kimmel, J. R., Docherty, K. S., Nemitz, E., Artaxo, P. A., Cappa, C. D., Martin, S. T., and Jimenez, J. L.: Chemically Resolved Particle Fluxes Over Tropical and Temperate Forests, Aerosol Sci. Technol., 47, 818–830, 2013.

Finkelstein, P. L. and Sims, P. F.: Sampling error in eddy correlation flux measurements, J. Geophys. Res., 106, 3503–3509, 2001.

Hollinger, D. Y. and Richardson, A. D.: Uncertainty in eddy covariance measurements and its application to physiological models, Tree Physiol., 25, 873–885, 2005.

Hörtnagl, L., Clement, R., Graus, M., Hammerle, A., Hansel, A., and Wohlfahrt, G.: Dealing with disjunct concentration measurements in eddy covariance applications: A comparison of available approaches, Atmos. Environ., 44, 2024–2032, 2010.

Ibrom, A., Dellwik, E., Larsen, S. E., and Pilegaard, K.: On the use of the Webb-Pearman-Leuning theory for closed-path eddy correlation measurements, Tellus B, 59, 937–946, 2007.

Jones, S. K., Famulari, D., Di Marco, C. F., Nemitz, E., Skiba, U. M., Rees, R. M., and Sutton, M. A.: Nitrous oxide emissions from managed grassland: a comparison of eddy covariance

and static chamber measurements, Atmos. Meas. Tech., 4, 2179–2194, doi:10.5194/amt-4-2179-2011, 2011.

Karl, T. G., Spirig, C., Rinne, J., Stroud, C., Prevost, P., Greenberg, J., Fall, R., and Guenther, A.: Virtual disjunct eddy covariance measurements of organic compound fluxes from a subalpine forest using proton transfer reaction mass spectrometry, Atmos. Chem. Phys., 2, 279–291, doi:10.5194/acp-2-279-2002, 2002.

Langford, B., Misztal, P. K., Nemitz, E., Davison, B., Helfter, C., Pugh, T. A. M., MacKenzie, A. R., Lim, S. F., and Hewitt, C. N.: Fluxes and concentrations of volatile organic compounds from a South-East Asian tropical rainforest, Atmos. Chem. Phys., 10, 8391–8412, doi:10.5194/acp-10-8391-2010, 2010.

Laurila, T. and Aurela Mand Tuovinen, J.: Eddy covariance measurements over wetlands, Eddy Covariance: A practical guide to measurements and data analysis, edited by: Aubinet, M., Vesala, T., and Papale D., Springer, Dordrecht, Heidelberg, London, New York, 14, 345–360, 2012.

Lenschow, D. and Kristensen, L.: Uncorrelated noise in turbulence measurements, J. Atmos. Oceanic Technol., 2, 68–81, 1985.

Lenschow, D. H. and Raupach, M. R.: The attenuation of fluctuations in scalar concentrations through sampling, J. Geophys. Res.-Atmos., 96, 15259–15268, 1991.

Lenschow, D. H., Mann, J., and Kristensen, L.: How long is long enough when measuring fluxes and other turbulence statistics, J. Atmos. Ocean. Technol., 11, 661–673, 1994.

Lenschow, D. H., Wulfmeyer, V., and Senff, C.: Measuring second-through fourth-order moments in noisy data, J. Atmos. Ocean. Technol., 17, 1330–1347, 2000.

Lumley, J. L. and Panofsky, H. A.: The structure of atmospheric turbulence. John Wiley & Sons, 239 pp., 1964.

Mahrt, L.: Flux sampling errors for aircraft and towers, J. Atmos. Ocean. Technol., 15, 416–429, 1998.

Mann, J. and Lenschow, D. H.: Errors in airborne flux measurements, J. Geophys. Res.-Atmos., 99, 14519–14526, 1994.

Massman, W. J.: A simple method for estimating frequency response corrections for eddy covariance systems, Agr. Forest Meteorol., 104, 185–198, 2000.

Massman, W. J. and Lee, X.: Eddy covariance flux corrections and uncertainties in long-term studies of carbon and energy exchanges, Agr. Forest Meteorol., 113, 121–144, 2002.

Mauder, M., Cuntz, M., Druee, C., Graf, A., Rebmann, C., Schmid, H. P., Schmidt, M., and Steinbrecher, R.: A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements, Agr. Forest Meteorol., 169, 122–135, 2013.

Milne, R., Mennim, A., and Hargreaves K.: The value of the coefficient in the relaxed eddy accumulation method in terms of fourth order moments, Bound. Lay. Meteorol., 101, 359–373, 2001.

Muller, J. B. A., Coyle, M., Fowler, D., Gallagher, M. W., Nemitz, E. G., and Percival, C. J.: Comparison of ozone fluxes over grassland by gradient and eddy covariance technique, Atmos. Sci. Lett., 10, 164–169, 2009.

Nemitz, E., Gallagher, M. W., Duyzer, J. H., and Fowler, D.: Micrometeorological measurements of particle deposition velocities to moorland vegetation, Q. J. Roy. Meteor. Soc., 128, 2281–2300, 2002.

Nemitz, E., Jimenez, J. L., Huffman, J. A., Ulbrich, I. M., Canagaratna, M. R., Worsnop, D. R., and Guenther, A. B.: An eddy-covariance system for the measurement of surface/atmosphere exchange fluxes of submicron aerosol chemical species – First application above an urban area, Aerosol Sci. Technol., 42, 636–657, 2008.

Nemitz, E., Coyle, M., Langford, B., Gerosa, G., Marzuoli, R., Stella, P., Benjamin Loubet, B., Potier, E., Joensuu, J., Altimir, N., Ammann, C., Vuolo, R., Pilegaard, K., and Weidinger, T.: Eddy-covariance flux measurements of ozone deposition: review and development of a common methodology, Atmos. Meas. Tech. Diss., in preparation, 2015.

Park, J.-H., Goldstein, A. H., Timkovsky, J., Fares, S., Weber, R., Karlik, J., and Holzinger, R.: Eddy covariance emission and deposition flux measurements using proton transfer reaction – time of flight – mass spectrometry (PTR-TOF-MS): comparison with PTR-MS measured vertical gradients and fluxes, Atmos. Chem. Phys., 13, 1439–1456, doi:10.5194/acp-13-1439-2013, 2013.

Peltola, O., Hensen, A., Helfter, C., Belelli Marchesini, L., Bosveld, F. C., van den Bulk, W. C. M., Elbers, J. A., Haapanala, S., Holst, J., Laurila, T., Lindroth, A., Nemitz, E., Röckmann, T., Vermeulen, A. T., and Mammarella, I.: Evaluating the performance of commonly used gas analysers for methane eddy covariance flux measurements: the InGOS inter-comparison field experiment, Biogeosciences, 11, 3163–3186, doi:10.5194/bg-11-3163-2014, 2014.

Rannik, Ü., Mammarella, I., Aalto, P., Keronen, P., Vesala, T., and Kulmala, M.: Long-term aerosol particle flux observations part I: Uncertainties and time-average statistics, Atmos. Environ., 43, 3431–3439, 2009.

Rannik, Ü., Haapanala, S., Shurpali, N. J., Mammarella, I., Lind, S., Hyvönen, N., Peltola, O., Zahniser, M., Martikainen, P. J., and Vesala, T.: Intercomparison of fast response commercial gas analysers for nitrous oxide flux measurements under field conditions, Biogeosciences, 12, 415–432, doi:10.5194/bg-12-415-2015, 2015.

Rinne, J. and Ammann, C.: Disjunct eddy covariance method, Eddy Covariance: A practical guide to measurements and data analysis, edidet by: Aubinet, M., Vesala, T., and Papale D., Springer, Dordrecht, Heidelberg, London, New York, 10, 291–306, 2012.

Rummel, U., Ammann, C., Gut, A., Meixner, F. X., and Andreae, M. O.: Eddy Covariance measurements of nitric oxide flux within an Amazonian rain forest, J. Geophys. Res., 107, LBA 17-1–LBA 17-9, doi:10.1029/2001JD000520, 2002.

Shurpali, N. J., Verma, S. B., Clement, R. J., and Billesbach, D. P.: Seasonal distribution of methane flux in a Minnesota peatland measured by eddy-correlation, J. Geophys. Res.-Atmos., 98, 20649–20655, 1993.

Spirig, C., Neftel, A., Ammann, C., Dommen, J., Grabmer, W., Thielmann, A., Schaub, A., Beauchamp, J., Wisthaler, A., and Hansel, A.: Eddy covariance flux measurements of biogenic VOCs during ECHO 2003 using proton transfer reaction mass spectrometry, Atmos. Chem. Phys., 5, 465–481, doi:10.5194/acp-5-465-2005, 2005.

Stella, P., Kortner, M., Ammann, C., Foken, T., Meixner, F. X., and Trebs, I.: Measurements of nitrogen oxides and ozone fluxes by eddy covariance at a meadow: evidence for an internal leaf resistance to NO_2, Biogeosciences, 10, 5997–6017, doi:10.5194/bg-10-5997-2013, 2013.

Taipale, R., Ruuskanen, T. M., and Rinne, J.: Lag time determination in DEC measurements with PTR-MS, Atmos. Meas. Tech., 3, 853–862, doi:10.5194/amt-3-853-2010, 2010.

Valach, A. C., Langford, B., Nemitz, E., MacKenzie, A. R., and Hewitt, C. N.: Seasonal and diurnal trends in concentrations and fluxes of volatile organic compounds in central London, Atmos. Chem. Phys., 15, 7777–7796, doi:10.5194/acp-15-7777-2015, 2015.

Wienhold, F. G., Welling, M., and Harris G. W.: Micrometeorological Measurement and Source Region Analysis of Nitrous-Oxide Fluxes from an Agricultural Soil, Atmos. Environ., 29, 2219–2227, doi:10.1016/1352-2310(95)00165-U, 1995.

Wulfmeyer, V., Pal, S., Turner, D. D., and Wagner, E.: Can water vapour Raman Lidar resolve profiles of turbulent variables in the convective boundary layer?, Bound.-Lay. Meteorol., 136, 253–284, 2010.

Wyngaard, J. C.: On surface-layer turbulence, in Workshop on Micrometeorology, edited by: Haugen, D. A., American Meteorology Society, Boston, 392 pp., 1973.