



Eddy Covariance flux errors due to random and systematic timing errors during data acquisition

Gerardo Fratini¹, Simone Sabbatini², Kevin Ediger¹, Giacomo Nicolini², Domenico Vitale², Dario Papale²

5 ¹LI-COR Biosciences Inc., Lincoln, 68504, Nebraska, USA

²Dipartimento per la Innovazione nei sistemi Biologici, Agroalimentari e Forestali - DIBAF, Viterbo, 01100, Italy

Correspondence to: Gerardo Fratini (gerardo.fratini@licor.com)

Abstract. Modern Eddy Covariance (EC) systems collect high-frequency data (10-20 Hz) via instruments' digital outputs. While this helps overcome the limits of traditional mixed analog/digital systems, it also introduces the new problem of guaranteeing data synchronicity when the clocks of the involved devices cannot themselves be synchronized, which is often the case with instruments providing data via serial or Ethernet connectivity. By means of a simulation study, we found that in most cases random timing errors can safely be neglected, as they don't impact fluxes to a large extent. On the contrary, realistic systematic timing errors potentially arising in asynchronous systems can lead to flux underestimations as large as 10%, by means of attenuation of high-frequency flux contributions. We characterize the transfer function of such filters as a function of the error amount and find cut-off frequencies as low as 1 Hz, which means that synchronization errors can dominate high-frequency attenuations in open- and enclosed-path EC systems. In most cases, such timing errors cannot be detected nor quantified a posteriori. Therefore, we suggest testing the ability of prospect EC data logging systems to assure required synchronicity and propose a procedure to implement such a test which relies on readily available equipment.

1 Introduction

20 Calculating atmosphere-biosphere exchange fluxes of a gaseous species with the Eddy Covariance (EC) method requires collecting *synchronous* data of the vertical wind component w and of the concentration c of the gas, at a suitable frequency (e.g. 10 or 20 Hz). After proper data treatment, the covariance of the two time series is calculated, from which the flux is derived (e.g. Foken et al. 2012). In this context, *synchronicity* means that w and c data for any given timestamp (i.e., the data that is multiplied together in the covariance) refers to values of the physical quantities at a given point in space and time.

25 However, despite any level of integration within an EC system, wind and concentration data are measured by different instruments. In addition, data collection is performed by means of a variety of more or less engineered data acquisition systems. Delays in the data flows, zero-hold issues, digital clock drifts, required separation of the measuring devices and artefacts in the data acquisition strategy can lead to poor synchronicity, i.e. to *misalignments* of the time series, that may result in significant flux errors of both random and systematic nature. Commercial solutions implementing sound engineering practices do exist for long-established EC of CO₂ and H₂O, which guarantee data synchronicity within specs that meet EC requirements. However, most solutions are not scalable to other species, in the sense that it is not possible to simply replace a gas analyser with another one and achieve the same synchronization performance. Furthermore, it is customary for many research groups to assemble EC systems "in-house", especially when addressing gas species that haven't been popular enough to grow strong commercial interest. In all such cases, it is important to verify that data misalignments are not being introduced by the data logging strategy, which can lead to biases in resulting fluxes.

35 In this paper, we discuss the potential types and sources of misalignment in digital EC systems and quantify their effects on resulting fluxes. In conjunction with site-specific characteristics such as the typical cospectral shapes, this information can help design an appropriate data collection system. We also propose a strategy for evaluating prospect EC data collection systems from the point of view of data synchronicity, before they are used in routine field activities.



1.1 Analog vs. digital EC systems

Traditionally, a combination of analog and digital transmission systems has been used to collect EC data (Eugster and Plüss, 2010). For example, analog signals from the gas analyzer were sent to an interface unit responsible of digitizing the data before merging it with anemometric data, itself coming from an A/D converter into the interface unit, typically from a sonic anemometer-thermometer (SAT). More recently, specular solutions, with the analog data from the SAT sent to an interface unit residing in the gas analyzer system, became available and were widely adopted. With both these approaches, the data was presented to the user for the flux calculation as single files with wind and gas timeseries merged and synchronized by the interface unit.

Analog data output allows the data to easily cross clock domains. The clock that is used to sample the original signal does not need to be synchronized to the clock that samples the analog output. This makes it very convenient to merge data from systems with unsynchronized clocks. However, collecting data in analog form has several limitations and risks. Firstly, the number of analog channels available either as outputs from the instruments or as inputs to the interface unit is typically limited to 4 or 6, which dramatically reduces the number of variables that can be collected. In fact, historic EC raw datasets are comprised of 6 or maximum 7 variables: the 3 wind components (u , v , w), the sonic temperature (T_s) and the concentration of the gases of interest (c , traditionally CO_2 and H_2O); more rarely, a diagnostic variable for the anemometric data was also collected. For this reason, critical information such as the full diagnostics of both instruments and their status (e.g. the temperature and pressure in the gas analyser cell or the signal strength) or the original raw measurement (speed of sound, raw data counts etc.) are not collected in analog systems, limiting the means for quality screening and limiting the possibility of future re-computation starting from the most fundamental measured data. Another problem with analog data collection is that signals are subject to degradation due to dissipation, electromagnetic noise and ageing of cables and connectors, which reduces the quality of collected data (Barnes, 1987). In addition, although raw measurements are analog in nature, they are typically immediately digitized (*native* digital format provided by the manufacturer) and then - in case of analog data collection - they are re-converted to analog, sent to the interface unit and there converted back to digital; these A/D-D/A conversions potentially degrade the signal adding noise and dampening high-frequency signal components (Eugster and Plüss, 2010).

In recent years, “fully digital” EC systems became available, where both data streams are collected in their respective *native* digital format, i.e. without A/D conversions others than those implemented by the manufacturer to provide digital outputs. Fully digital systems largely or completely overcome both problems with mixed analog/digital systems, with more robust, less corruptible data transmission protocols and the possibility of collecting all variables made available by the individual instruments (e.g. Eugster and Plüss, 2010). However, combining digital data streams from different instruments brings about new challenges, most notably with respect to data synchronization. While moving between clock domains is trivially simple in an analog system, it is much more challenging with digital data when the involved clocks can be completely asynchronous to each other. Un-synchronized data can affect the correlation between variables and as a consequence, systematic biases can be introduced in resulting fluxes. Therefore careful handling of digital data is essential during every step of the data acquisition.

1.2 The problem of clocks synchronization in digital systems

Digital clocks are electronic oscillator circuits that use the mechanical resonance of a vibrating crystal of piezoelectric material to create an electrical signal with a precise frequency, which is then used to keep track of time. The number and quality of clocks involved in an EC system vary with the data collection strategy and technology. In typical applications, the clocks involved are:

- The sampling clocks of the sensing instruments (in the typical EC system, the SAT and the gas analyzer), responsible for sampling data at the prescribed rate with sufficient precision and accuracy;



- For systems that transmit data serially (RS-232 or RS-485), the serial clock of the same sensing instrument, which may or may not be *correlated* to its sampling clock;
- The clocks of the logging device(s) (data logger, PC, etc.), responsible for attaching a timestamp to the data. If a single logging device is used this is usually also responsible for merging data streams from the different instruments; in case dedicated logging systems are used for different instruments, merging is performed in post-processing and the clocks of the different loggers must, therefore, be aligned sufficiently frequently (e.g. every second, using a GPS signal).

Typical open digital communication protocols used for EC instruments are serial (RS-232, RS-485) and packet-based data protocols (Ethernet). In devices that transmit data via serial communication, such as SATs, there are no means to synchronize the sampling clock of the device to that of the data logger. With such devices, the best that can be done is to assign a timestamp *after* transmission, based on the clock of the data logger (this last step should be performed carefully to avoid large inaccuracies due to serial port latencies, especially in PC-based systems). In addition, devices implementing serial communication have an asynchronous clock that drives those protocols (e.g. Dobkin et al., 2010). If this clock is *correlated* with the device's sampling clock, the receiving data logger can - at least in principle - reconstruct the sampling clock. However, in devices that do not correlate sampling and serial clocks (such as those that output data in a software thread), the system scheduler then determines when data is transmitted, thereby completely isolating the serial clock from the sampling clock and making it impossible for the data logger to reconstruct the sampling clock.

Packet-based data communications such as Ethernet even further isolate the sampling clock from the transmission clock. In devices using this protocol, it is therefore impossible to reconstruct a sampling clock. However, for Ethernet-based systems additional protocols are available, such as NTP or PTP, to actually synchronize all system clocks. The synchronized system clocks then allow the data to be correctly timestamped *before* transmission, eliminating any synchronization issue, provided that downstream software can align the various data streams based on their timestamps (e.g. Mahmood et al., 2014).

1.3 Types of timing errors

In typical EC data acquisition setups, time series collected by different instruments can show three distinct types of misalignments (Figure 1):

Time lags: these are *constant* offsets in otherwise perfectly aligned time series. They can be the result of constant electronic delays or of delays due to digital signal processing (DSP). More frequently, they result from a physical separation of the sampling volumes or from the delay due to the time needed for the passage of air in a sampling line.

Random timing errors (RTEs in the following) occur when the timestamps assigned to the data differ from the exact time dictated by the nominal sampling frequency, and such differences are *randomly distributed* so that, on average, the actual frequency is equal to the nominal one. In practice, in the EC context, it is more useful to consider the random differences in the timestamps assigned to data from one instrument with respect to that of the paired instrument. In fact, in the hypothetical case in which the two instruments would have the exact same sequence of random errors, that would not introduce any misalignment and hence no flux bias.

Systematic timing errors (STEs) occur when the timestamps assigned to the data differ from the exact time dictated by the nominal sampling frequency, and such differences are *systematic*, e.g. the actual time step is slightly longer or shorter than the nominal one for time spans of the order of the flux averaging interval. Again, in EC we are only concerned with systematic *relative* errors, for identical errors in the two concerned instruments would entail no misalignment and hence no flux bias.

Instances of each type of misalignment can, and typically will, be present at the same time to various degrees.



1.4 Sources of misalignment and their effects on time series

1.4.1 Spatial separation between sampling volumes

In a SAT, the *sampling volume* is the volume of air between the upper and lower sets of transducers. Similarly, in an open-path gas analyser such as the LI-7500 CO₂/H₂O analyzer and the LI-7700 CH₄ analyzer (LI-COR Biosciences Inc., Lincoln, NE, USA), the sampling volume is the volume of air between the upper and lower mirrors. In a closed- or enclosed-path gas analyzer such as the LI-7000, the LI-7200 (LI-COR Biosciences Inc.) and the EC155 (Campbell Scientific Inc., Logan, HT, USA), instead, the sampling volume can be identified with the volume of the intake device, e.g. a rain cup.

Even in the hypothetical situation of perfectly synchronized timestamps for wind and gas data, if the respective instruments' sampling volumes have to be spatially separated to avoid presently intractable flow distortion issues in the anemometer, as is notably the case with open-path setups (see, for example, Wyngaard, 1988; Frank et al., 2016; Horst et al., 2016; Grare et al., 2016 and Huq et al. 2017), the corresponding timeseries will be affected by misalignment, possibly to varying degrees. Indeed, assuming the validity of Taylor's hypothesis of frozen turbulence, wind and concentration data will be affected by a time-lag (the time air takes to travel between the two sampling volumes), which will be further modulated by wind intensity and direction. Additionally, modification of turbulence structure intervening while air parcels transit through the dislocated instrument volumes may introduce further uncertainty in flux estimates (Cheng et al., 2017). In case of co-located sensors (e.g. Hydra-IV, CEH; IRGASON, Campbell Scientific Inc.) this problem is not present but is replaced by the flow distortion issues mentioned above and not addressed in the present study.

1.4.2 Spatial separation between measuring volumes

In a SAT, the sampling volume coincides with the measuring volume, i.e. wind velocity is measured exactly where it is sampled. The same is true for open-path gas analyzers. However, closed-path and enclosed-path analyzers take the sampled air into a measuring cell via a sampling line that can be anywhere between 0.5 to 50 meters long, with its inlet usually placed very close to the SAT sampling volume. This implies a delay of the time series of gas concentrations with respect to the wind time series. Such delay can be more or less constant in time depending on the possibility of actively controlling the sampling line flowrate. In systems without flow controllers, the flowrate may vary significantly in response to power fluctuations or tube clogging and so would the corresponding time-lags.

1.4.3 Clock errors

Quartz crystal clocks universally used in electronic devices are subject to two main types of error: periodic jitter and frequency drift (McParland, 2017).

Period jitter

Period jitter in clock signals is the random error of the clock with respect to its nominal frequency. It is typically caused by thermal noise, power supply variations, loading conditions, device noise, and interference coupled from nearby circuits. Jitter is a source of RTE in time series.

Frequency drift

The oscillation frequency of a clock varies with temperature, leading to *drifts* of the measured time and hence to STEs. The drift of a clock can be expressed as the amount of time gained (or lost) as a result of the drift per unit of time, with suitable units being $\mu\text{s/s}$. For example, a drift of $-30 \mu\text{s/s}$ means that a clock accumulates $30 \mu\text{s}$ of delay per second, or about 2.6 seconds over the course of one day ($2.6 = 30 \cdot 10^{-6} \cdot (24 \cdot 60 \cdot 60)$). The dependence of crystal oscillation frequency on temperature varies, even dramatically, with the type and angle of crystal cut and can be modeled as quadratic (BT, CT, DT cuts) or cubic (AT cuts) (Hewlett Packard 1997). Figure 2 shows exemplary drift curves for different crystal cuts. Typically,



the nominal frequency (e.g. 32 kHz) is specified at 20 or 25 °C. Apart from that temperature, the frequency can vary for example according to (for a BT cut):

$$\frac{f-f_0}{f_0} = -\alpha(T - T_0)^2 \cdot 10^{-6} \quad (1)$$

5

where $f_0 = f(T_0 = 25\text{ °C})$ and typical values of α range 0.035-0.040.

Clocks in EC systems can be exposed to large variations of temperature (day-night, seasonal cycles). Because we are concerned with *relative* drifts, we are interested in *differences* in the temperatures experienced by the instruments' sampling/logging clock as well as with differences in their temperature responses. Clocks experiencing similar temperatures and with similar temperature responses, would minimize relative drift. On the contrary, clocks with opposite responses to temperature will result in relative drifts that are close to the sum of the individual drifts, e.g. in the case of AT clocks with different angles of rotation at relatively high temperatures (i.e. above 30 degC, see Figure 2)

It is also to be noted that temperature-compensated clocks do exist, which have accuracies of around $\pm 2\ \mu\text{s/s}$. As we will show later, such drifts can be safely neglected, as long as clocks are synced sufficiently often (e.g. once a day). For completeness, we note that clock drifts also occur due to the ageing of components. However, the absolute values of typical ageing rates ($< 1\ \mu\text{s/year}$) are of no concern in EC applications. Because STEs in EC systems are caused primarily or exclusively by clock drifts, in the rest of the paper we will use the terms STE and *drift* interchangeably.

1.4.4 Further sources of timing errors in digital asynchronous systems

Recently, Ethernet connectivity became available in commercial loggers and industrial PCs (e.g. SmartFlux 2 and 3 by LI-COR Biosciences Inc., CR3000 and CR6 by Campbell Scientific Inc.) and can be used for data acquisition in EC systems. The acquisition is usually done using TCP, a packet-based protocol specifically designed to preserve the accuracy of the data during transmission. TCP is however not designed to preserve the temporal aspect of the packets. The TCP receiving system must buffer up packets and signal the sender if an error occurs in any packet. This can cause packets to even arrive out of order, even though they are always delivered to the application in order. Therefore, TCP-based systems are subject to significant RTEs and, potentially, to STEs.

Serial communication devices typically use a first-in-first-out (FIFO) policy to buffer data, on both the sending and the receiving sides. The FIFO increases the efficiency and throughput by reducing the number of interrupts the CPU has to handle (e.g. Park et al., 2003). Without a FIFO buffer, the CPU has to interrupt for every data unit. With a FIFO, the CPU is interrupted only when a FIFO is full, or a programmed amount of data is ready. However, the FIFO becomes a problem on a system where it's desired to correlate the serial clock to the sampling clock. On such a system, the FIFO introduces timing jitter on both the transmitter and the receiver, hence inducing RTEs in the system.

1.5 Dealing with timing errors in EC practice

The fundamental difference between time-lags on one side and RTE/STE on the other side is that constant time-lags can, at least in principle, be addressed *a posteriori* during data processing. The topic of correctly estimating and compensating time-lags has long been discussed in the EC literature (Vickers & Mahrt 1997; Ibrom et al., 2007; Massman and Ibrom, 2008; Langford et al, 2015), and corresponding algorithms are available in EC processing software. We will therefore not further discuss time-lags in this paper.

Random and systematic timing errors, instead, are not identifiable and therefore it is not possible to correct them. However, their effect on flux estimates, as we will see, can become significant. For this reason, the only viable strategy to reduce flux biases is to design the data acquisition system in a way that prevents or minimizes the possibility of their occurrence.



The focus of this paper is, therefore, the quantification of flux underestimations as a function of RTE and STE, so as to derive quantitative specifications for a data acquisition system that minimizes EC flux losses. We further propose a simple scheme for evaluating existing data acquisition systems with respect to data synchronization.

2 Materials and Methods

5 2.1 Simulation design

In order to accurately quantify how time alignment errors affect flux estimates, we performed a simulation study. As a reference, we used the covariance estimated from high-frequency data of vertical wind speed (w) and sonic temperature (T_s) which are by definition perfectly synchronized since they are computed starting from the same raw data (the travelling time of sound signals between pairs of transducers in a SAT). We also assumed that high-frequency timeseries are provided at perfectly constant time steps of 0.1 (10 Hz) or 0.05 (20 Hz) seconds. Subsequently, we manipulated the vector of times at which the sonic temperature data was sampled in order to simulate realistic ranges of RTEs and STEs. Values of sonic temperature in correspondence to the new, simulated timestamps were estimated by linearly interpolating the closest data points in the original series (Figure 3).

Before calculating covariances, standard EC processing steps were applied such as spike removal (Vickers and Mahrt, 1997), tilt correction by double rotation method (Wilczak et al., 2001) and fluctuations estimation via block-averaging. Covariance estimates obtained with the new versions of T_s were then compared with the reference to quantify the effect of simulated timing errors. Flux biases would be almost identical to biases in covariances, weakly modulated by corrections intervening between covariance and flux computation, such as spectral corrections and consideration of WPL effects (e.g. Fratini et al. 2012). The simulation study was implemented in the source code of EddyPro v6.2.1 (LI-COR Biosciences Inc, Lincoln, NE; Fratini and Mauder, 2014).

For the present analysis, we simulated RTEs ranging ± 1 to ± 100 ms, that is, up to the same order of the sampling interval. As an example, with a simulated ± 10 ms RTE using 10 Hz data (nominal time step = 100 ms), simulated time steps for T_s varied randomly between 90 and 110 ms, with an average of 100 ms. We note that RTEs of 10-100 ms won't usually be caused by clock jitter, which is typically several orders of magnitude smaller but can easily be caused by acquisition systems based on serial or Ethernet communication not specifically designed to collect synchronous timeseries, as described above.

For STEs, we simulated relative drifts ranging 10 to 180 $\mu\text{s/s}$ (specifically 10, 30, 60, 90, 120, 150 and 180 $\mu\text{s/s}$). For example, to simulate a STE of 60 $\mu\text{s/s}$ we kept the original w timeseries (time step equal to 100 ms) and modified the time step of T_s to be 100.006 ms. This may seem a negligible difference, which however accumulates a difference of 108 ms between w and T_s within 30 minutes and manifests itself as a difference of one row in the length of the time series (i.e. 18000 value for w and 17999 for T_s @ 10 Hz). Similarly, systematic errors of 120 and 180 $\mu\text{s/s}$ would lead to 2 and 3 row differences, respectively.

Systematic timing errors such as frequency drifts essentially act as low-pass filters, which can be described by characterising their transfer function, provided that the drift is known, as in our simulation. Here, for each 30-min period and for each STE amount, we calculated an *in-situ* transfer function as the frequency-wise ratio of drifted and original w - T_s cospectra:

35

$$\text{TF}_{EXP}(\text{STE}|f_n) = \frac{\text{CO}(wT_s^{\text{STE}}|f_n)}{\text{CO}(w,T|f_n)}, \text{ for STE} = [10, \dots, 180] \mu\text{s/s} \quad (2)$$

where T_s^{STE} (K) is the simulated sonic temperature for each STE value, f_n (Hz) is the natural frequency and $\text{TF}_{EXP}(\text{STE}|f_n)$ is the *in-situ* transfer function for STE. We repeated this calculation for a number of cospectra ranging 1000-2000 (depending on data availability). The ensemble of all transfer functions so obtained and for each STE amount was then fitted

40



with the following function, that was found to reasonably approximate the data obtained for all drifts at all sites in the most relevant frequency range (0.01 ÷ 5 Hz):

$$TF(f_{co}|f_n) = (1 + \beta) \frac{1}{1 + \left(\frac{f_n}{f_{co}}\right)^\alpha} - \beta \quad (3)$$

5 where:

f_{co} (Hz) is the transfer function cut-off frequency and α and β are fitting parameters whose values were found to vary very little around $\alpha = 2.65$ and $\beta = 0.25$.

2.2 Datasets

We performed simulations on 4 datasets acquired from EC sites representative of various ecosystem types and climatic regimes and characterized by the different height of measurement and height of the canopy:

- *IT-Ro2*: a deciduous forest of Turkey Oak (*Quercus cerris* L.) in Italy. Eddy covariance measurements were carried out from 2002 to 2013 and the period used for the simulations was May 2013, when the canopy height was 15 m and the measurement height 18 m (Rey et al., 2002).
- *IT-Ro4*: located at about 1 km from IT-Ro2, this site is a rotation crop where EC measurements have been carried out from 2008 to 2014. Data used for the simulations include 43 days in 2012 when crimson clover (*Trifolium incarnatum* L.) was cultivated (maximum canopy height 60 cm, measurement height of 3.7 m)
- *DK-Sor*: evergreen forest near Sorø, DK. EC measurements are performed since 1997: during the period used for the simulation (the entire 2015) the forest was 25 m tall and the EC system placed at 60 m a.g.l. (Pilegaard et al., 2011).
- *IT-CA3*: fast-growing, short-rotation coppice (SRC) of poplar clones planted in 2010, located in Castel d'Asso, Viterbo, Italy. The EC tower was installed at the end of 2011, and measurements were taken until mid-2015. The period used for the simulation included 9 months over the period 2012-2015, with a canopy height ranging 0-5.3 m, and the measurement height between 3 and 5.5 m (Sabbatini et al., 2016).

2.3 Validation of the simulation design

Although the proposed simulation design enables the evaluation of resulting errors using readily available EC data, we note that interpolating data sampled at 10 or 20 Hz frequency can potentially introduce artefacts (due to the lack of information at higher frequencies) such as, for example, an undue reduction of the sonic temperature variance, which would result in artificial reduction of the $w-T_s$ covariance. In order to detect any such effects, we preliminarily implemented a validation procedure, making use of one week of sonic data from a Gill HS (Gill Instruments, Lymington, UK) collected at 100 Hz. The validation involved the following steps:

1. Subsampling at 10 Hz and simulating timing errors as described above, i.e. interpolating starting from the subsampled data.
2. Subsampling at 10 Hz and simulating timing errors by interpolating the original 100 Hz data.
3. Comparing $w-T_s$ covariances obtained in 1 and 2.

The timing errors simulated interpolating the original 100 Hz measurements (option 2 above) are much less prone to artefacts because interpolation occurs between data that are 0.01 seconds apart, an interval too short for any significant flux signal to occur. Using this procedure, we could verify that there is no detectable difference between results obtained with 100 Hz and 10 Hz data (not shown), which implies that the interpolation procedure is not introducing significant artefacts in the estimation of variances and covariances and therefore the simulation can be performed with virtually any historic EC dataset using the available code.



3 Results and discussion

Figure 4 compares covariances $w-T_s$ obtained with increasing amounts of RTE against the reference covariance obtained with the original, perfectly synced, timeseries. Reduction in covariance estimates is fairly negligible provided that RTE is of the same order of magnitude of the sampling interval or less. Largest discrepancies were observed for the IT-CA3 and IT-Ro4 sites with a covariance underestimation of 3% for RTE of amplitude 100 ms. As mentioned earlier, such large timing errors are never the result of electronic clock jitter and may instead be caused by a data transmission system not designed for time synchronization, such as TCP.

Conversely, flux biases induced by systematic timing errors are both more significant and more variable. Figure 5 shows that a STE of 60 $\mu\text{s/s}$ (1 row of difference in a 30 minutes file with data collected at 10 Hz) can lead to errors anywhere between 0 \div 4%, increasing to 1 \div 8% for a STE of 120 $\mu\text{s/s}$ (2 rows of difference) and to 1 \div 11% for a STE of 180 $\mu\text{s/s}$ (3 rows of difference).

Figure 6A shows an example of the transfer functions derived using the procedure described in Sect 2.1. The Figure refers to the site IT-CA3, but the filters obtained for the other sites had very similar characteristics, as illustrated in Fig. 6B using the mean cut-off frequencies computed for all sites at each STE amount: the tight $\pm 3.5\sigma$ range merely demonstrates that the low-pass filter properties of the STE are independent from the data used to derive it, and only vary with the error amount.

Nonetheless, in Figure 5 we showed how the same filter leads to very different flux underestimations at a different site. For example, a systematic error of 180 $\mu\text{s/s}$ led to a flux biases of 1% and 11% at IT-Ro2 and IT-Ro4, respectively. The reason is related to the distribution of the flux contributions across the frequency domain, as shown in Figure 7: the more the flux cospectrum is shifted towards higher frequencies, the more it will be dampened by the low-pass filter and the higher the resulting flux bias will be. In other words, systematic timing errors are a source of high-frequency spectral losses, not dissimilar to the ones traditionally considered in EC (Moncrieff et al. 1997; Massmann, 2000; Ibrom et al. 2007).

To put this new source of high-frequency losses in perspective, we note that for EC systems based on an enclosed-path gas analyzer (LI-7200), cut-off frequencies ranging 1.1 Hz (for less optimized) up to 7-8 Hz (for systems with optimized intake rain cup and heated sampling line) were reported in literature (e.g. Fratini et al. 2012; De Ligne et al., 2014; Metzger et al. 2016). Similar values are usually found in systems based on open-path setups. Significant STEs can thus easily become leading sources of flux biases in modern EC systems, with the additional complication that they are hard to detect, quantify and correct. In fact, once acquired and stored in files, it is generally not possible to detect whether a drift between data streams occurred, and hence to correct it in post-processing. Missing lines in one data stream could be either filled in by the data acquisition software (e.g. by means of the “last observation carried forward” technique) or could be compensated by dropping one line in the paired, longer series. In both cases, one would be unable to detect the problem, which is however obviously not solved by these solutions, meant only to build a complete rectangular dataset. On the contrary, a mismatch of one or two lines in the length of the time series is *not* necessarily the sign of an occurring STE, as it could also be the result of an imperfect timing in opening/closing a data stream, or some combination of the two factors. For this reasons, it is very difficult, if not practically impossible, to detect STEs, distinguish them from other timing errors or artefacts and, more importantly, to infer the type and amount of error that is being introduced in the covariances. It is therefore advised to evaluate the performance of a data acquisition system before it is put in operation.

If analog inputs are available in the EC instruments, a simple way to evaluate synchronicity in the data logging system is to connect a signal generator to both EC instruments (SAT and gas analyzer) and collect the data via the data logging system in the configuration that would be adopted in normal operation (Figure 8). The result is 2 replicates of the known signal data, whose timestamps will in general not be synchronized, in the sense that the same nominal timestamp will be attached to two different pieces of data. The two datasets can then be compared to calculate the phase difference between the clocks and hence assess RTEs and STEs. For example, a cross-correlation of the time series yields the phase difference. It's important to test the timeseries at different time intervals, e.g. $\frac{1}{2}$ hour for several days, in actual field conditions that undergo



significant temperature variations. The two instruments are synchronized if the cross-correlation yields the same result every time. If this constant phase offset (time-lag) is different from zero, this measurement quantifies signal delays in a system, which can be addressed either by optimizing the data logging system or by taking this offset into account while setting up the time-lag automatic computation in post-processing. A cross-correlation that changes over time, instead, is a strong indication of occurring STEs. It is very difficult to anticipate the evolution in time of the phase change as it depends on the clocks cuts, quality, temperature sensitivity. In general, we may expect a linear trend in the phase if temperatures don't vary strongly (see also later) or a trend modulated by a diurnal pattern if temperature plays a role.

If a signal generator is not available, but analog outputs from at least one EC instrument are available, a viable variant of this test consists in transmitting (at least) one of the analog outputs to the analog input of the companion instrument (e.g. w sent via analog output of the SAT to an analog input channel of the gas analyser). In this way the raw-data files contain 2 replicates of that variable, each collected according to the timing of the respective instrument: the timestamps of the digital version is logged according to timing of the sending instrument (SAT, in the example), while those of its analog version are logged according to the timing of the receiver (gas analyzer). The same cross-correlation analyses described above can then be performed. Note that, in both versions of the test, the way raw data is stored may have a strong impact on how to interpret the results. For instance, depending on the specifics of the data acquisition system, collecting a unique file with 3 days worth of data or collecting 30-min files for three days, can provide different results, e.g. because the act of closing a file and opening a new one can cause the data streams to be partially or completely "realigned".

To exemplify the test, we collected about 3 days of 20Hz wind data from a SAT (HS-100, Gill Instruments Ltd., Lymington, UK) both in native digital format (via a RS232 port, indicated with the subscript d in the following) and in analog format via the A/D of a LI-7550 Analyzer Interface Unit (LI-COR Biosciences Inc.) which was then collected via a second RS232 port (indicated with subscript a), using an industrial-grade PC running Windows XP. Thus, the data logging system under testing was "a Windows PC collecting EC instruments via RS232, which were setup to transmit data in streaming mode". The two data streams were completely independent to each other, and we attached timestamps to the records based on the OS clock as the data was made available from the serial port to the application collecting the data. We then merged the two datasets based on timestamps and split the resulting 3-day file into 30 minutes. Finally, we calculated time-lags between pairs of homologous variables (e.g. u_d vs u_a but results were identical for all anemometric variables), which are shown in Figure 9. The linearity of the data suggests that the system is affected by a fairly constant STE of about 50 $\mu\text{s/s}$, as quantified by the slope of the line.

Using the same setup, we further collected 2 days of data directly stored as 30-min files and again computed time-lags between homologous variables. The acquisition system was able to realign the two series at the beginning of each half hour resetting the time-lag between them to roughly zero. Nevertheless, calculating time-lags on overlapping 5 min periods, we found that *within* each half-hourly period the time-lags increased of 0.05 s (32% of the times) and of 0.1 s (68% of the times), which again indicates an STE ranging 30-60 $\mu\text{s/s}$ which, as shown above (Fig. 5), can lead to detectable flux biases.

Conclusions

Undoubtedly, modern EC systems should log high-frequency data in a native digital format, so as to collect all possible measurement, diagnostic and status information from each instrument and assure the creation of robust, self-documented datasets, which are essential to the long-term research goals of climate and greenhouse gases science. Nonetheless, given the needs of researchers to often assemble EC systems from instrumentation that is not necessarily designed to be integrated, extreme care and expertise must be put in the handling of different digital data streams, in a context where data synchronicity is essential. We have shown that failure to do so can result in significant biases for the resulting fluxes, which depend on the type of timing error (random or systematic) and its amplitude, as well as on the cospectral characteristics at the site. We have



also explained how such errors are virtually impossible to detect and quantify in historic timeseries. It is, therefore, necessary to avoid them upfront, via proper design and evaluation of the data logging system.

Deploying a simple testing setup that makes use of equipment usually available to the EC experimentalists, we demonstrated how, for example, a naive data collection performed asynchronously on a Windows XP industrial-PC leads to significant relative drifts among the timeseries, which is bound to generate flux underestimations. With minor *ad hoc* adjustments, the same testing setup can be used to evaluate any EC data logging system.

Finally, we recommend the scientific community to promote collaboration and synergy among manufacturers of EC equipment, for technological solutions that *guarantee* sufficient synchronicity do exist - such as Ethernet connectivity deploying the PTP protocol - but, in order to be utilized, they require all instrumentation to be compatible with those technologies, which is yet not still the case.

Acknowledgments

DP, GN and DV thanks the ENVRplus project funded by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement 654182 and the RINGO project funded under the same program under grant agreement 730944. SS thanks the COOP+ project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 654131.

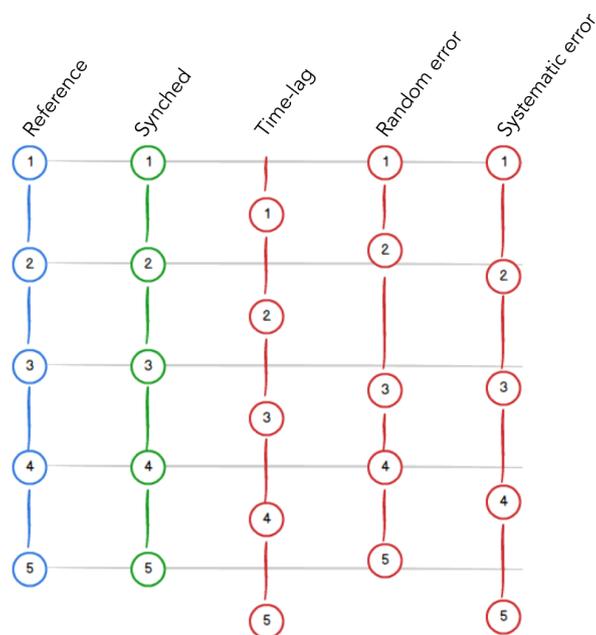
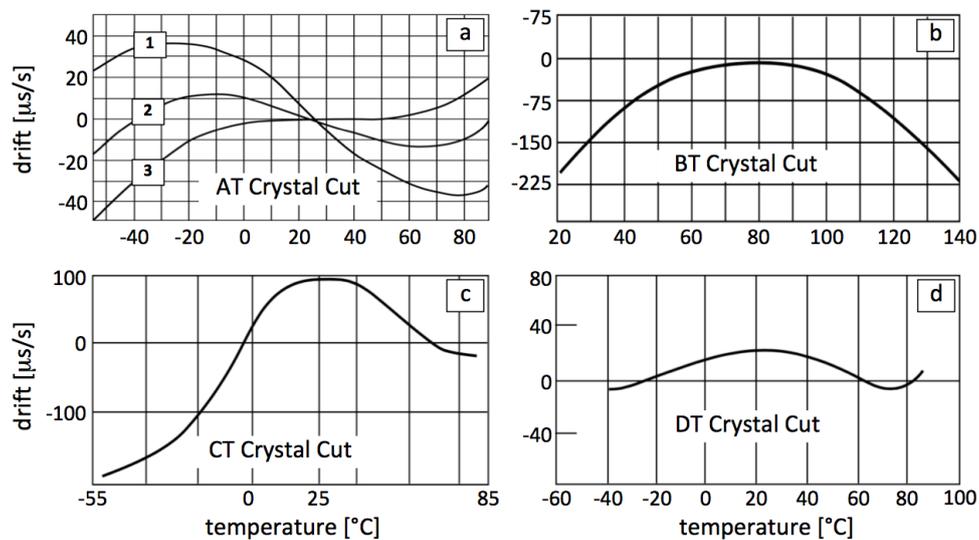
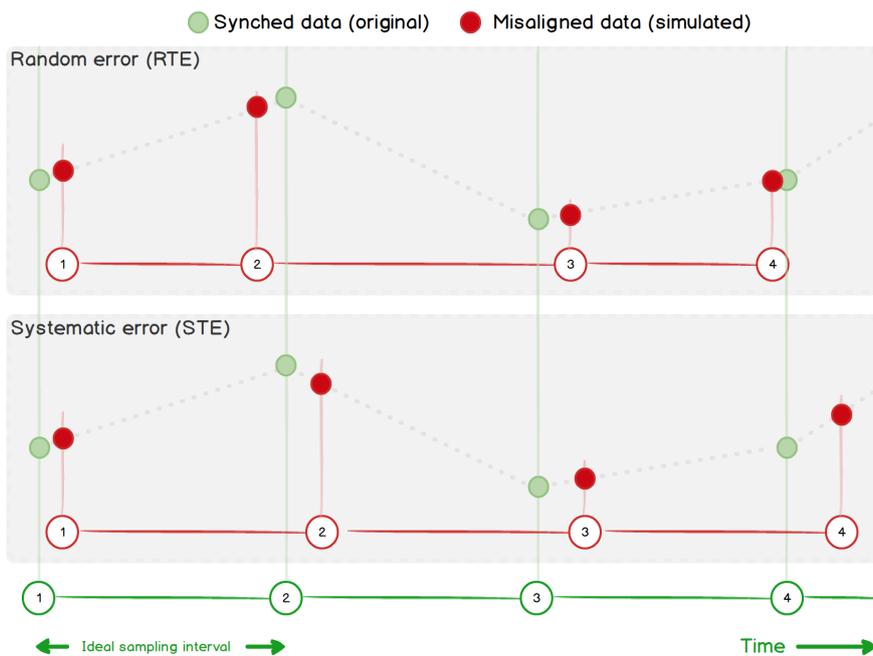


Figure 1: Schematic of the three types of misalignments common in EC data. Given a reference time series (e.g. w , blue), the ideal paired gas concentration time series is perfectly aligned (green). The three red time series exemplify (from left to the right): a constant offset (time-lag), random variations around the perfect alignment (random error) and a systematically larger time step (systematic error). Real data is typically affected by a mix of all error types in varying amounts.



5 Figure 2: Exemplary temperature dependence curves for clocks with AT (a), BT (b), CT (c) and DT (d) cuts. In (a) the 3 curves refer to different angles of rotations of the crystal. Reproduced and adjusted from Hewlett Packard (1997).



5 **Figure 3:** Sketch of the timing error simulation via linear interpolation. Green dots represent original T_s data points, equally spaced at the prescribed (nominal) time steps. In the upper panel, RTEs are simulated as time steps randomly varying around the nominal value. In the lower panel, STEs are simulated as a time step larger than the nominal one, whereby the difference between the assigned and correct timestamps always increase in time.

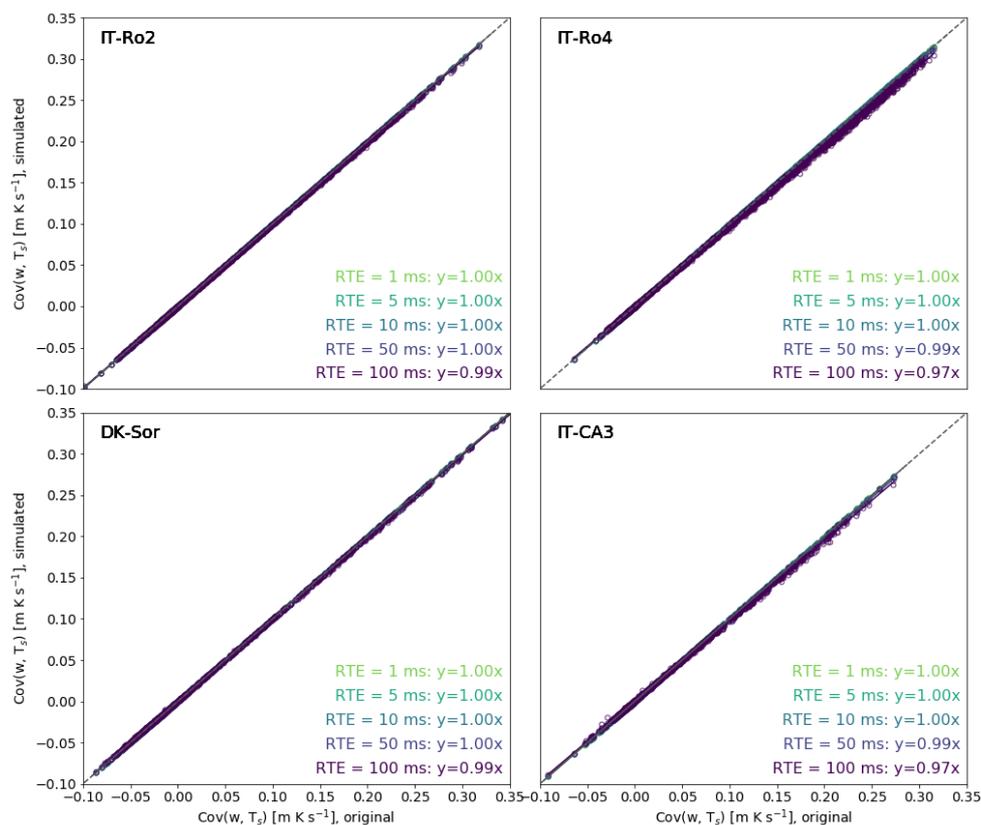


Figure 4: Simulation of RTEs. Covariances w - T_s obtained with a set of simulated random errors of different amplitudes (y-axes) are compared to the covariances w - T_s computed with the original time series (x-axes), for the four sites. All regressions had offset equal to zero and $r^2 > 0.99$.

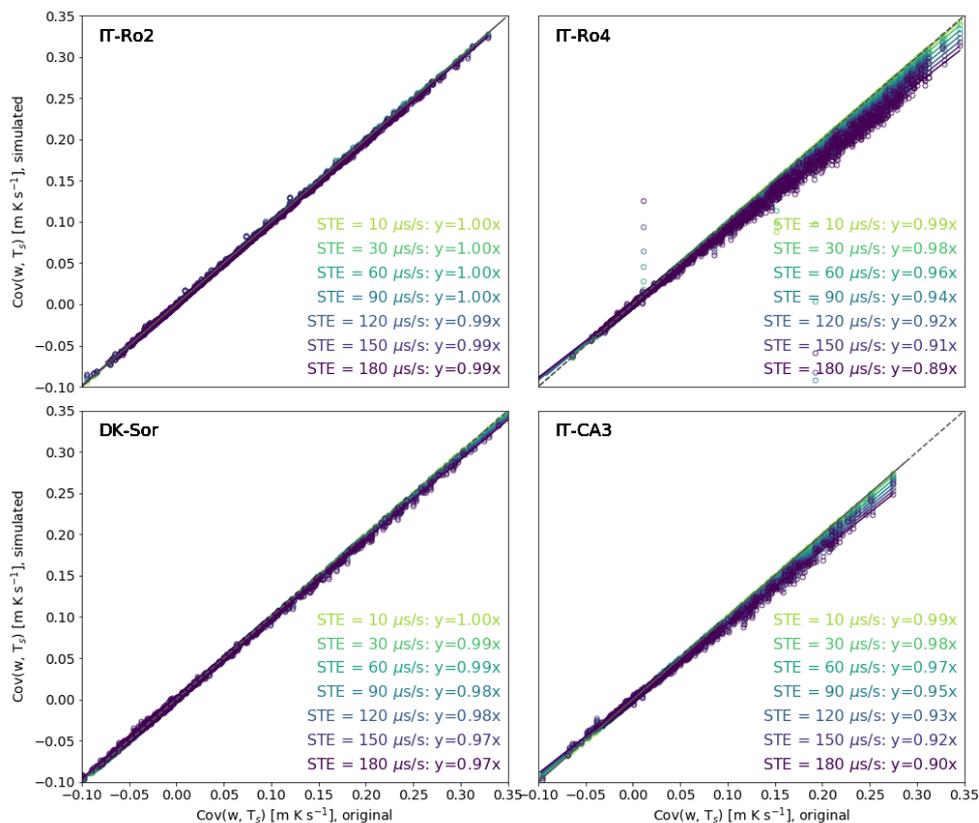
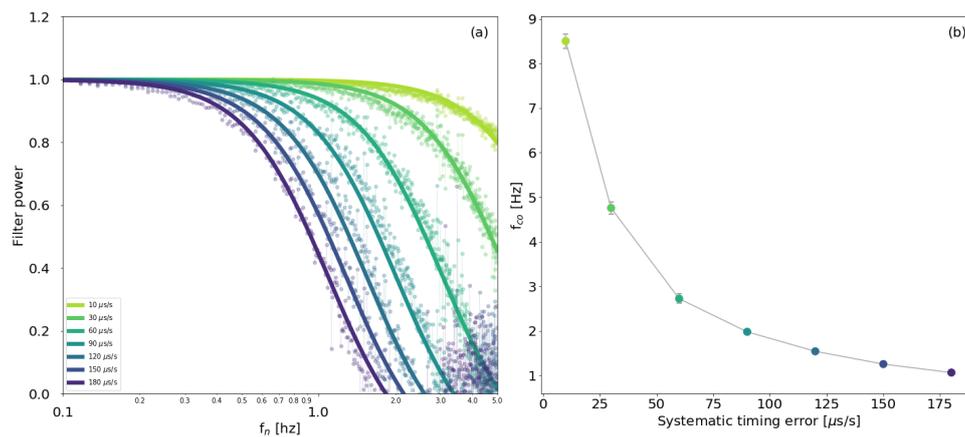
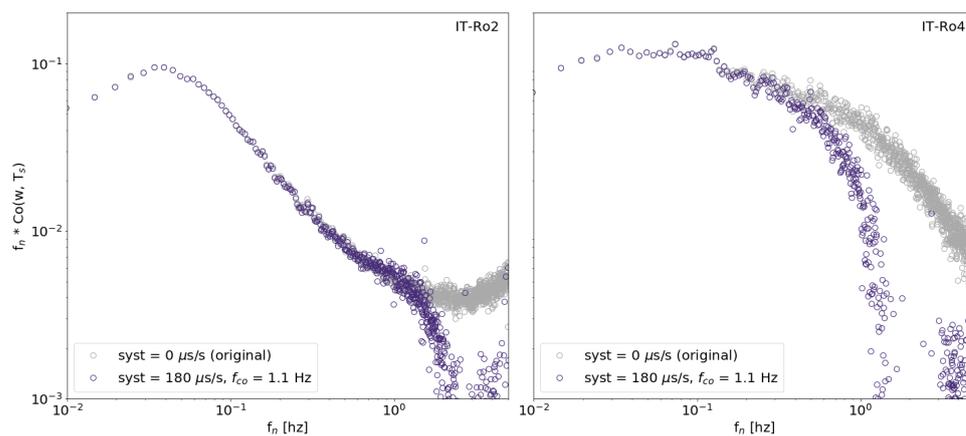


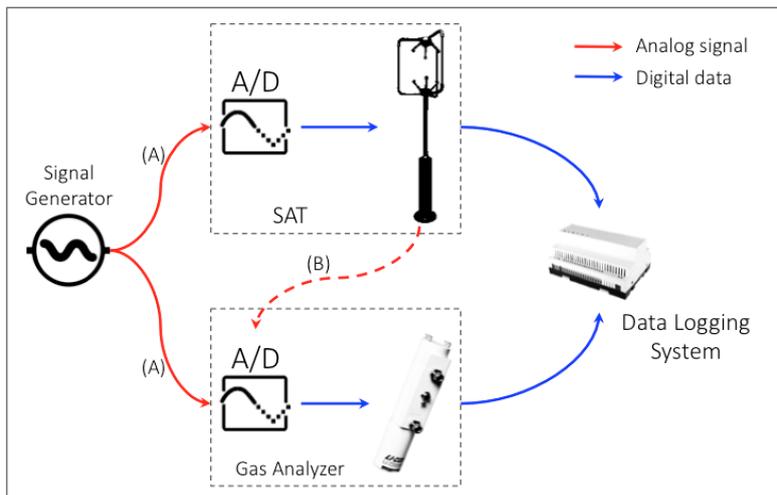
Figure 5: Simulation of STE. Covariances $w-T_s$ obtained with a set of simulated systematic errors of different amplitudes (y-axes) are compared to the covariances $w-T_s$ computed with the original time series (x-axes), for the four sites. All regressions had offset equal to zero and $r^2 > 0.98$.



5 Figure 6: Transfer function for the STE at different error amounts, as derived using Eqs. 2 and 3 (a). Mean values and 3.5σ ranges of the transfer function cut-off frequencies across the four sites, as a function of the error amount (b).



5 **Figure 7:** Example of how the same transfer function - here for a STE of 180 μ s/s - leads to very different flux biases via high-frequency spectral losses, for sites characterized by very different flux distributions across the frequency range.



5 **Figure 8:** Schematic of possible setups to evaluate the ability of a data logging system to synchronise EC data. In setup A the same (known) analog signal (solid red lines) is sent to the analog inputs of the EC instruments where it is digitized. In setup B analog wind data (dashed red line) is sent to the gas analyzer, where it is digitized. In both setups, the 2 digital data streams are then collected by the data logging system. The clocks involved and how timestamps are attached to data depend on the specifics of the system under consideration.

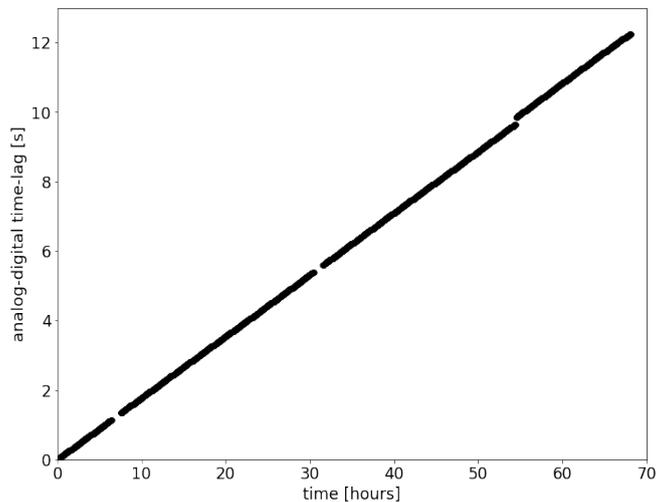


Figure 9: Evolution of time-lags between two replicates of the same variable (u wind component, but identical results were obtained with v and w), one collected with the SAT native digital format and one collected via analog outputs from the SAT. Data were collected in two files roughly 70 hours long and then split into 30-min chunks for the computation of time-lags.



References

- Barnes, J. R.: *Electronic System Design: Interference and Noise Control Techniques*, New Jersey: Prentice-Hall Inc., 1987.
- Cheng, Y., Sayde, C., Li, Q., Basara, J., Selker, J., Tanner, E., Gentine, P.: Failure of Taylor's hypothesis in the atmospheric surface layer and its correction for eddy-covariance measurements, *Geophysical Research Letters*, 44(9), 4287-4295, 2017.
- 5 De Ligne, A., Joly, L., Chopin, H., Cousin, J., Chauvin, N., Decarpentier, T., and Aubinet, M.: Laboratory and field experiments in order to dimension the IRGA gas sampling system, 1st ICOS Science Conference on Greenhouse Gases and Biogeochemical Cycles, Brussels, Belgium, 23–25 September 2014, 27, 2014.
- Dobkin, R., Moyal, M., Kolodny A., and Ginosar, R.: Asynchronous Current Mode Serial Communication, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 18 (7), 1107-1117, 2010.
- 10 Eugster, W., and Plüss, P.: A fault-tolerant eddy covariance system for measuring CH₄ fluxes, *Agric. For. Meteorol.*, 150, 841–851. doi:10.1016/j.agrformet.2009.12.008, 2010.
- Foken, T., Aubinet, M., and Leuning, R.: The Eddy Covariance Method, in: Aubinet, M., Vesala, T., Papale, D. (Eds.), *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*. Springer, Dordrecht, pp. 1–20. doi:10.1007/978-94-007-2351-1, 2012.
- 15 Frank, J. M., Massman, W. J. , Swiatek, E., Zimmerman, H.A., and Ewers, B.E.: All sonic anemometers need to correct for transducer and structural shadowing in their velocity measurements. *Journal of Atmospheric and Oceanic Technology*, 33(1): 149- 167, 2016.
- Fratini, G., Ibrom, A., Arriga, N., Burba, G., and Papale, D.: Relative humidity effects on water vapour fluxes measured with closed-path eddy-covariance systems with short sampling lines. *Agric. For. Meteorol.* 165, 53–63.
- 20 doi:10.1016/j.agrformet.2012.05.018, 2012.
- Fratini, G., and Mauder, M.: Towards a consistent eddy-covariance processing: an intercomparison of EddyPro and TK3. *Atmos. Meas. Tech. Discuss.* 7, 2107–2126. doi:10.5194/amtd-7-2107-2014, 2014.
- Grare, L., Lenain, L., and Melville, W.K.: The influence of wind direction on Campbell Scientific CSAT3 and Gill R3-50 sonic anemometer measurements. *Journal of Atmospheric and Oceanic Technology*, 33(11): 2477-2497, 2016
- 25 Horst, T.W., Vogt, R., and Oncley, S.P.: Measurements of flow distortion within the IRGASON integrated sonic anemometer and CO₂/H₂O gas analyzer. *Boundary-LayerMeteorology*, 160(1): 1-15, 2016.
- Huq, S., De Roo, F., Foken, T., and Mauder, M.: Evaluation of probe-induced flow distortion of Campbell CSAT3 sonic anemometers by numerical simulation. *Boundary-LayerMeteorology*, 165(1): 9-28, 2017.
- Ibrom, A., Dellwik, E., Flyvbjerg, H., Jensen, N.O., and Pilegaard, K.: Strong low-pass filtering effects on water vapour flux
- 30 measurements with closed-path eddy correlation systems. *Agric. For. Meteorol.* 147, 140–156. doi:10.1016/j.agrformet.2007.07.007, 2007.
- Langford, B., Acton, W., Ammann, C., Valach, A., and Nemitz, E.: Eddy-covariance data with low signal-to-noise ratio: time-lag determination, uncertainties and limit of detection. *Atmos. Meas. Tech. Discuss.* 8, 2913–2955. doi:10.5194/amtd-8-2913-2015, 2015.
- 35 Mahmood, A., Exel, R., and Sauter, T.: Delay and Jitter Characterization for Software-Based Clock Synchronization Over WLAN Using PTP. *IEEE Transactions on Industrial Informatics*, 10 (2), 1198-1206, 2014.
- Massman, W.J.: A simple method for estimating frequency response corrections for eddy covariance systems. *Agric. For. Meteorol.* 104, 185–198, 2000.
- Massman, W.J., and Ibrom, A.: Attenuation of concentration fluctuations of water vapor and other trace gases in turbulent
- 40 tube flow. *Atmos. Chem. Phys.* 8, 6245–6259. doi:10.5194/acp-8-6245-2008, 2008.
- McParland, D.: How to select the right oscillator for your Design: Four key questions to help ensure that you meet your design's requirements (2017) *Electronic Products*, 51 (4), 2017.



- Metzger, S., Burba, G.G., Burns, S.P., Blanken, P.D., Li, J., Luo, H., and Zulueta, R.C.: Optimization of an enclosed gas analyzer sampling system for measuring eddy covariance fluxes of H₂O and CO₂. *Atmos. Meas. Tech.*, 9, 1341–1359, 2016.
- Moncrieff, J.B., Massheder, J.M., de Bruin, H., Elbers, J., Friborg, T., Heusinkveld, B., Kabat, P., Scott, S., Soegaard, H., and Verhoef, A.: A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. *J. Hydrol.* 188–189, 589–611. doi:10.1016/S0022-1694(96)03194-0, 1997.
- 5 Park J., Mackay S., and Wright E.: *Practical Data Communications for Instrumentation and Control*. Newnes, Oxford. Pages 400, ISBN 9780750657976, 2003.
- Pilegaard, K., Ibrom, A., Courtney, M.S., Hummelshøj, P., and Jensen, N.O.: Increasing net CO₂ uptake by a Danish beech forest during the period from 1996 to 2009. *Agric. For. Meteorol.* 151, 934–946. doi:10.1016/j.agrformet.2011.02.013, 2011.
- 10 Rey, A., Pegoraro, E., Tedeschi, V., De Parri, I., Jarvis, P. G., and Valentini, R.: Annual variation in soil respiration and its components in a coppice oak forest in Central Italy. *Global Change Biology*, 8(9), 851–866, 2002.
- Sabbatini, S., Arriga, N., Bertolini, T., Castaldi, S., Chiti, T., Consalvo, C., Njakou Djomo, S., Gioli, B., Matteucci, G., and Papale, D.: Greenhouse gas balance of cropland conversion to bioenergy poplar short-rotation coppice. *Biogeosciences* 13, 95–113. doi:10.5194/bg-13-95-2016, 2016.
- 15 Smith, S.W.: *The scientist and engineer's guide to Digital Signal Processing*. California Technical Publishing, 2002.
- Vickers, D., and Mahrt, L.: Quality control and flux sampling problems for tower and aircraft data. *J. Atmos. Ocean. Technol.* 14, 512–526, 1997.
- Wilczak, J.M., Oncley, S.P., Stage, S.A.: Sonic anemometer tilt correction algorithms. *Boundary-Layer Meteorol.* 99, 127–150, 2001.
- 20 Wyngaard, J. C.: Flow-distortion effects on scalar flux measurements in the surface layer: Implications for sensor design. In Hicks, B.B. (Eds) *Topics in Micrometeorology. A Festschrift for Arch Dyer*. Springer, Dordrecht, 1998.