Agricultural and Forest Meteorology 248 (2018) 228–239

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

Research Paper

Impact of CO_2 storage flux sampling uncertainty on net ecosystem exchange measured by eddy covariance

Giacomo Nicolini[a,*], Marc Aubinet[b], Christian Feigenwinter[c], Bernard Heinesch[b], Anders Lindroth[d], Ossénatou Mamadou[b], Uta Moderow[e], Meelis Mölder[f], Leonardo Montagnani[g,h], Corinna Rebmann[i], Dario Papale[a,j]

[a] Department for Innovation in Biological Agro-food and Forest systems, University of Tuscia, Viterbo, Italy
[b] Department of Biosystem Engineering, Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium
[c] Department of Environmental Sciences, University of Basel, Basel, Switzerland
[d] Department of Earth and Ecosystem Sciences, Lund University, Lund, Sweden
[e] Department of Hydrosciences, Technische Universität Dresden, Dresden, Germany
[f] Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden
[g] Faculty of Science and Technology, Free University of Bolzano, Bolzano, Italy
[h] Forest Services, Autonomous Province of Bolzano, Bolzano, Italy
[i] Helmholtz Centre for Environmental Research UFZ, Germany
[j] CMCC Euro Mediterranean Centre on Climate Change, IAFES Division, Viterbo, Italy

ARTICLE INFO

Keywords:
Carbon balance
Forest ecosystems
Profile measurement
ADVEX
ICOS

ABSTRACT

Complying with several assumption and simplifications, most of the carbon budget studies based on eddy covariance (EC) measurements quantify the net ecosystem exchange (NEE) by summing the flux obtained by EC (*FC*) and the storage flux (*SC*). *SC* is the rate of change of a scalar, CO_2 molar fraction in this case, within the control volume underneath the EC measurement level. It is given by the difference in the quasi-instantaneous profiles of concentration at the beginning and end of the EC averaging period, divided by the averaging period. The approaches used to estimate *SC* largely vary, from measurements based on a single sampling point usually located at the EC measurement height, to measurements based on profile sampling. Generally a single profile is used, although multiple profiles can be positioned within the control volume. Measurement accuracy reasonably increases with the spatial sampling intensity, however limited resources often prevent more elaborated measurement systems. In this study we use the experimental dataset collected during the ADVEX campaign in which turbulent and non-turbulent fluxes were measured in three forest sites by the simultaneous use of five towers/profiles. Our main objectives are to evaluate both the uncertainty of *SC* that derives from an insufficient sampling of CO_2 variability, and its impact on concurrent NEE estimates.Results show that different measurement methods may produce substantially different SC flux estimates which in some cases involve a significant underestimation of the actual SC at a half-hourly time scales. A proper measuring system, that uses a single vertical profile of which the CO_2 sampled at 3 points (the two closest to the ground and the one at the lower fringe of the canopy layer) is averaged with CO_2 sampled at a certain distance and at the same height, improves the horizontal representativeness and reduces this (proportional) bias to 2–10% in such ecosystems. While the effect of this error is minor on long term NEE estimates, it can produce significant uncertainty on half-hourly NEE fluxes.

1. Introduction

The estimation of the net ecosystem exchange (NEE) by the eddy covariance (EC) technique is based on simplifications of the mass balance equation, and on its integration over a control volume that extends horizontally on a representative surface and vertically from the soil level to the measurement height (Finnigan et al., 2003; Foken et al., 2012). After Reynolds averaging, integrating over the control volume, ignoring horizontal turbulent flux divergence and the horizontal variation of the vertical flux, the source/sink strength of a scalar c integrated over the height z of the control volume is given by (Aubinet et al., 2005; Feigenwinter et al., 2010a, 2008):

* Corresponding author.
E-mail address: g.nicolini@unitus.it (G. Nicolini).

http://dx.doi.org/10.1016/j.agrformet.2017.09.025
Received 23 June 2017; Received in revised form 23 September 2017; Accepted 30 September 2017
Available online 07 October 2017

$$NEE = \frac{1}{V_m}\overline{w'c'}\Big|_z + \int_0^z \frac{1}{V_m}\frac{\partial \overline{c}(z)}{\partial t}dz + \int_0^z \frac{1}{V_m}\overline{w}(z)\frac{\partial \overline{c}(z)}{\partial z}dz + \int_0^z \frac{1}{V_m}\left(\overline{u}(z)\frac{\partial \overline{c}(z)}{\partial x} + \overline{v}(z)\frac{\partial \overline{c}(z)}{\partial y}\right)dz \quad (1)$$

where NEE denotes the biological source/sink of CO_2, V_m is the molar volume of dry air ($m^3\ mol^{-1}$), c the CO_2 molar fraction ($\mu mol\ mol^{-1}$), t the time (s), u, v, w ($m\ s^{-1}$) are the wind velocity components in x, y and z directions respectively. Overbars refer to the Reynolds averaging operator. The first term on the right hand side (RHS) of Eq. (1) is the turbulent vertical flux (*FC*, $\mu mol\ m^2\ s^{-1}$) measured by the EC system at the reference height z (m). The second term on the RHS refers to the rate of change in storage of CO_2 (*SC*, $\mu mol\ m^2\ s^{-1}$), usually estimated from vertical profile measurements. The third and fourth terms denote the non-turbulent vertical and horizontal advection fluxes, respectively. In the prevalence of carbon budget studies, also those involved in continental and global monitoring networks like ICOS, Ameriflux or FLUXNET, the advection terms are rarely quantified directly either because of their assumed minor importance or because of the critical difficulties in measuring them with the required accuracy (Aubinet et al., 2010; Heinesch et al., 2007; Moderow et al., 2011; Vickers et al., 2012). More often their contribution is partially, and indirectly, taken into account by applying specific corrections (e.g. the friction velocity filter, McHugh et al., 2017). Especially in tall vegetation ecosystems, the storage term represents an important part of the mass balance equation. Although both terms (storage change and advection) are related and involved in the so called night-flux problem (Aubinet et al., 2010), advection will not be considered in this paper, referring the readers to specific literature (e.g. Feigenwinter et al., 2008; Kang et al., 2017; Montagnani et al., 2009; Yi et al., 2000 and aforementioned references). Here, the focus is on the storage term because it commonly represents the only non-turbulent term quantified and used to estimate the NEE. Thus, after this simplification Eq. (1) reduces to

$$NEE = FC + SC \quad (2)$$

The storage term *SC* reflects the temporal dynamics of CO_2 in the air volume below the *FC* measurement height, not influenced by turbulence. Positive *SC* values are due to an accumulation of CO_2 within the control volume while negative values mean a depletion. In practice, *SC* is the rate of change of CO_2 given by the difference in the instantaneous profiles of concentration at the beginning and end of the EC averaging period, divided by the averaging period itself (Finnigan, 2006). While cumulating over daily to yearly periods leads to a nullification of *SC*, it can be significant at short time intervals such as half-hours or hours, especially around sunrise and sunset. During night, when atmospheric stratification is stable and turbulence is suppressed, *SC* becomes important, equalling or even exceeding *FC*. It follows that, for a closer quantification of the true NEE, *SC* cannot be ignored (Papale et al., 2006).

The typical approach to compute *SC* is based on a single tower concentration profile assuming horizontal homogeneity (one-dimensional integration), although ideally, it should be derived from concentrations averaged in the whole volume (three-dimensional integration). Under non-ideal conditions, as in cases of heterogeneity in the source/sink distribution or of tall canopies forests, the error caused by this assumption can be large (Pattey et al., 2002; van Gorsel et al., 2009). Despite this evidence, it is not unusual that *SC* computation is further simplified based solely on the temporal changes of the concentrations measured at the tower top and assuming a constant CO_2 concentration in the air column underneath (also known as one-point, tower-top or discrete approach). The error associated with this additional assumption is proportional to the degree of decoupling between the CO_2 measurement height (generally the EC system height) and the below canopy air space. Some studies attested the comparability between profile and one-point *SC* estimates (Greco and Baldocchi, 1996; Knohl et al., 2003; Lee et al., 1999; Priante-Filho et al., 2004) while others reported general underestimates (Gu et al., 2012; Iwata et al., 2005). The choice of one sampling design over another is essentially due to technical (and cost-related) aspects. By reducing the sampling intensity, the accuracy of *SC* is also reduced, according to the spatial variability of the source/sink distribution. On the other hand, an overly complex setup could be expensive, difficult to implement and manage, and possibly not needed. A functional distribution of sampling points (SPs) is thus crucial in order to avoid errors in the estimates. Some studies assayed the effect of profile vertical configuration on *SC* and NEE estimates (Bjorkegren et al., 2015; Gu et al., 2012; Wang et al., 2016; Yang et al., 2007, 1999) based on a single vertical profile analysis. For example Gu et al. (2012) found that the CO_2 storage based on the tower-top measurement was underestimated by up to 34% with respect to the one based on their 8 level profile. Yang et al. (2007) reported that a profile system with 4 sampling levels or fewer, even if optimally distributed, is not adequate for CO_2 storage measurements in a forest with a complex vertical structure because its mean error is on the same order of magnitude as the nighttime NEE ($1.0–5.8\ \mu mol\ m^{-2}\ s^{-1}$).

In this paper we extend the analysis to a three-dimensional space with the objective to 1) quantify the error in *SC* measurements due to an insufficient sampling of the spatial variability of CO_2 concentration, 2) identify and evaluate an efficient measurement set-up, and 3) quantify the impact of *SC* sampling error on consequent NEE estimate.

The data used in this study are a part of the dataset of ADVEX, the CarboEurope-Integrated Project (CE-IP) advection campaigns (Feigenwinter et al., 2008). It is worth to note that we focused on the spatial sampling error only. Other main sources of error in *SC* measurements, as the temporal sampling error (Finnigan, 2006), are not considered (details on this can be found in e.g. Cescatti et al., 2016; Marcolla et al., 2014; Siebicke et al., 2011; Wang et al., 2016; Yang et al., 2007).

2. Materials and methods

2.1. The dataset: ADVEX

The ADVEX project aimed at providing a possible methodology to accurately quantify advective fluxes in EC measurements. Three analogous experiments were performed at three different forest sites across Europe in 2005 and 2006. The sites were part of the CE-IP and are characterized by different orography (Fig. 1): Norunda (NO in the following) in Sweden is on a basically flat surface, Renon (RE) in Italy, is located on a notable alpine slope (11°), Wetzstein (WS) in Germany is located on a hill ridge (Feigenwinter et al., 2008). Their mean altitude is 45 m, 1735 m and 782 m respectively, however the average temperature was comparable at the three sites. Also the species composition was similar and dominated by spruce (*Picea abies* L. Karst.), in association with *Pinus sylvestris* L. at NO and with *Pinus cembra* L. at RE. The average leaf area index was 4.5 at NO (Lagergren et al., 2005), 5.1 at RE (Marcolla et al., 2005), 7.0 at WS (Rebmann et al., 2010). The mean canopy height was about 25 m at NO and RE and 22 m at WS. NEE was rather different as in the ADVEX experiment year NO was on average a net source of CO_2 ($52\ g\ C\ m^{-2}\ y^{-1}$), RE was a strong sink ($-721\ g\ C\ m^{-2}\ y^{-1}$) and WS was a moderate sink ($-96\ g\ C\ m^{-2}\ y^{-1}$, Rebmann et al., 2010).

The experimental set-ups were similar, composed by one main tower (M) surrounded by four shorter towers (A–D, named satellites from now on) forming a quadrangle (Fig. 1). The main instrumentation on each tower is summarized in Table 1.

A comprehensive description of sites' characteristics, data collection, instruments and setup used during ADVEX can be found in the related bibliography (Aubinet et al., 2010; Feigenwinter et al., 2010a, 2010b, 2008; Moderow et al., 2011; Montagnani et al., 2010) and in the supplementary material of this article.

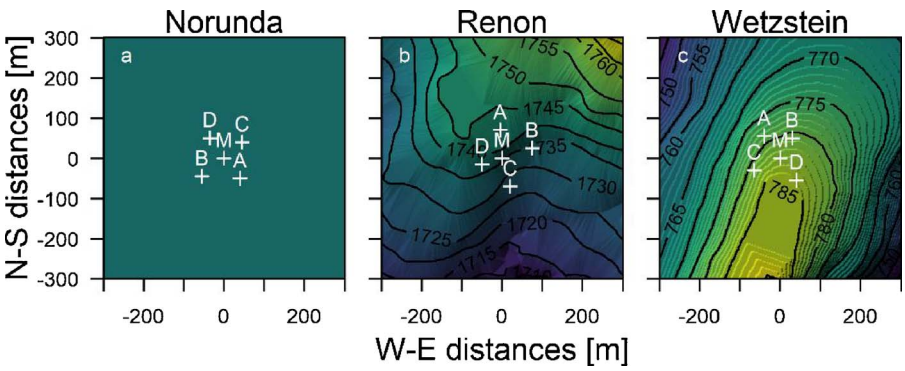

Fig. 1. Digital elevation model maps of the ADVEX experimental sites, Norunda (a), Renon (b) and Wetzstein (c). Difference in altitude (above sea level) between contour lines is 5 m. White crosses and letters represent main (M) and satellite towers (A, B, C, D) used in the experiment. The orography at Norunda is basically flat, with an average altitude of 45 m a.s.l.

Table 1
Main measurements and instrumentation during the ADVEX campaign.

	Norunda (NO), Sweden	Renon (RE), Italy	Wetzstein (WS), Germany
Experiment period	JUL 07, 2006 – SEP 19, 2006	JUL 21, 2005 – SEP 16, 2005	APR 12, 2006 – JUN 20, 2006
Main towers			
EC system[1]	LI7500[a]/USA – 1[b]	LI7500, LI6262[a]/Gill HSC	LI6262[a]/Gill R3[c]
EC height (m)	33	32	33
air pressure	PTA – 427[d]	CS – 105[e]	CS – 105[e]
profile system[2]	LI6262[a]/AB[f]	LI6262[a]/ MP103A	LI6262, LI7000[a]/ FW3[g]
profile heights[3] (m)	1.5, 8.5, 13.5, 19, 24.5, 28, 31.7, 36.9	1, 2, 4, 8, 16, 32	0.1, 0.3, 1, 2, 5, 9, 15, 23.8, 30
Satellite towers			
profile system[2]	LI6262[a]/FW3[g]	LI6262[a]/FW3[g]	LI6262/FW3[g]
profile heights[3] (m)	1.5, 6, 12, 30	1.5, 6, 12, 30	1.5, 4.4, 8.8, 24

[1] Gas analyser/sonic anemometer.
[2] Gas analyser/thermocouple.
[3] Sampling levels eight from the soil. On the main tower of Norunda there were additional levels at 43.8, 58.5, 73, 87.5, 100.6 m that were not considered in this work.
[a] LiCor, Lincoln, NE, US.
[b] METEK GmbH, Germany.
[c] Gill Instruments, UK.
[d] Vaisala Inc., Helsinki, Finland.
[e] Campbell SCi., Logan, US.
[f] Thermocouple, In Situ, Ockelbo, Sweden.
[g] Chromel–Constantan (Type E) thermocouple. Campbell Sci., Logan, US.

2.2. Analysis rationale

2.2.1. Calculation of fluxes

SC fluxes (μmol m^{-2} s^{-1}) from individual towers were calculated using a discrete form of the second term on the RHS of Eq. (1) that can be written as

$$SC \cong \overline{\rho_d}(z) \sum_{i=1}^{i=n} \frac{\Delta \bar{c}(z_i)}{\Delta t} \Delta z_i \quad (3)$$

in which the index *i* increases up to the total number of profile levels *n* (n = 8, 6, 9 for main towers in the case of NO, RE and WS, respectively and *n* = 4 in the case of satellite towers), ρ_d is air molar density calculated as a function of air moisture, temperature and pressure, $\Delta\bar{c}(z_i)$ is the CO_2 molar fraction difference at height z_i over the time period Δt (i.e. 1800*s*), Δz_i is the vertical extent of the corresponding *i*-th air layer. Several *SC* computations were used for the sampling error assessment, calculated according to different sampling designs by using the whole SPs of original profiles, series of lower-resolution profiles (LRP) obtained by excluding a certain number of SPs from the vertical integration, and their combinations. Then, four of them were considered in the assessment of *SC* in relation to turbulent fluxes, namely:

- *target SC* (SC_{TAR}), the storage flux considered as the closest representation of the true *SC* resulting from the spatial integration (averaging) of the individual *SC* measured at each tower (M, A, B, C, D) (Feigenwinter et al., 2008);
- *full profile SC* (SC_{FUL}), the storage flux resulting from the classical one-dimensional approach (vertical integration of the whole SPs of a profile);
- *one-point SC* (SC_{ONE}), the storage flux computed based solely on the temporal changes of [CO_2] measured at the tower top (highest profile level);
- *optimal SC* (SC_{OPT}), the storage flux resulting from the setup showing the best trade-off among measuring system complexity and performance.

It is important to clarify that we assumed that the storage flux measured using all the available data (SC_{TAR}) is the best possible estimation, and by far the most complete of all the storage measurements done at the sites. Though, this is not the real storage flux, which is practically impossible to measure because it would require the sampling of the rate of change of [CO_2] in the whole control volume. In the specific cases considered in this analysis, given the distances between the main and satellites towers (70 m maximum), all the *SC* fluxes were likely measured within the typical flux footprint climatology area at the sites, being the 4 satellites placed in proximity of the footprint picks, and this assumption can be accepted.

Turbulent fluxes (*FC*, μmol m^{-2} s^{-1}), were measured at the main towers at each site, details can be found e.g. in Moderow et al. (2009). NEE (μmol m^{-2} s^{-1}) was calculated by Eq. (2). The few gaps in the data set were not filled with the aim of basing the analysis on measured data only.

2.2.2. Storage flux uncertainty quantification and statistical analysis

The uncertainty in *SC* estimates has been evaluated considering both the vertical resolution of individual profiles and the spatial variability of [CO_2] (both vertical and horizontal). To this aim we produced series of *SC* based on a number of LRP as depicted in Fig. 2. For satellite towers, as they all originally have 4 sampling levels, 5 LRPs were considered, in addition to the full and one-point profile, using 3 and 2 SPs. For the main towers, LRPs were extracted by permutation of the original SPs, then, only the 5 most statistically-performing LRP within those with the same SPs number were considered. This, in addition to full and one-point profiles, resulted in a total of 32, 22 and 37 profiles for NO, RE and WS respectively (Fig. 2). Then, according to the specific case, we compared the actual *SC* with either SC_{FUL} or SC_{TAR} profile.

To evaluate the consistency between different *SC* we adopted some statistical goodness-of-fit measures (Zambrano-Bigiarini, 2017). Three of them were directly used for the evaluation while others were used as complementary information (and reported in the supplementary material). The used metrics are the slope of the regression line times the coefficient of determination (*bR2*), the Nash Sutcliffe efficiency (*NSE*)

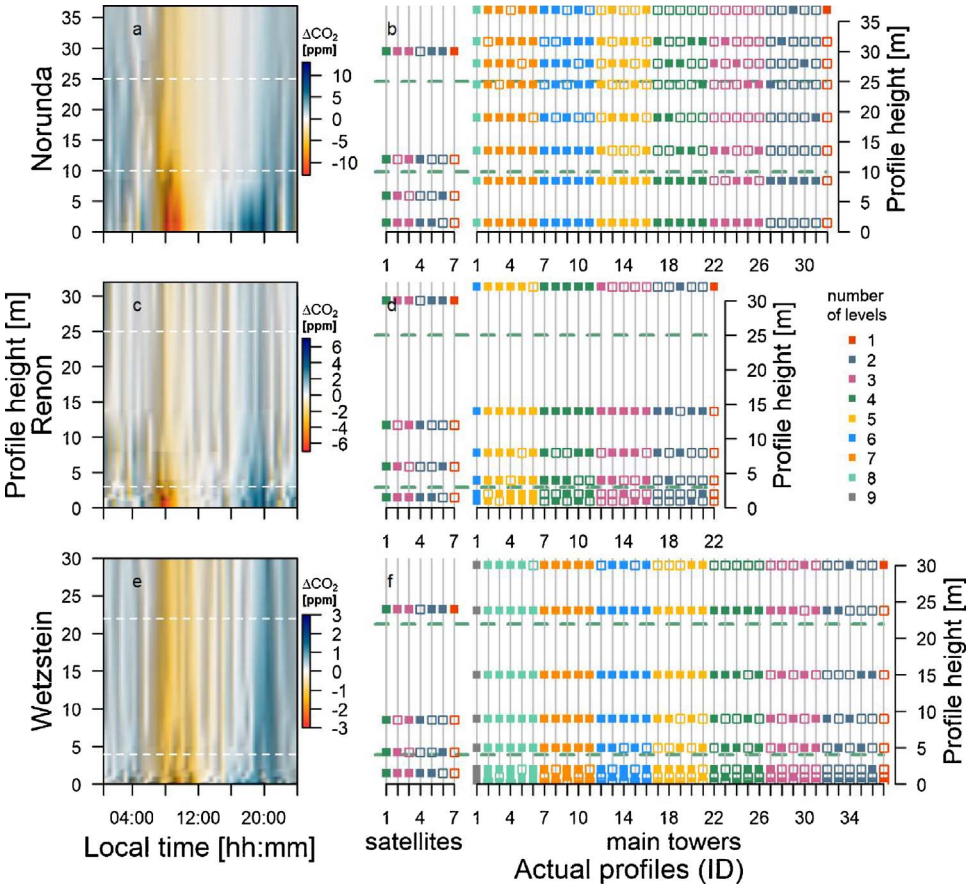

Fig. 2. Left panels: average daily trend of CO_2 concentration time derivative over the whole experimental periods for Norunda (a), Renon (c) and Wetzstein (e). *Right panels*: schematic representation of the considered low-resolution profiles (LRP) for satellites and main towers at Norunda (b), Renon (d) and Wetzstein (f). Full and empty squares highlight the used and excluded sampling point (SP) for each LRP respectively. Colours gather profiles with the same number of sampling points. For satellite towers LRPs are identical among sites as original profiles on these towers have the same number of SP. For main towers LRP are site-specific and were clustered in groups of 5 LRP with an equal number of SPs (yet differently displaced). The dotted lines (white on the left panels, light-green on the right panels) approximate the canopy space. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and the mean absolute error (*MAE*). *bR2* is a relative (dimensionless) measure obtained after Krause and Boyle (2005) by

$$bR^2 = \begin{cases} |b|R^2 for \;\; b \le 1 \\ |b|^{-1}R^2 for \;\; b > 1 \end{cases} \quad (4)$$

where b is the slope and R^2 is the coefficient of determination of the linear regression between the compared series (forcing the intercept to zero). This statistical combination allows accounting for both under- and over-estimations (described by b) as well as the spread of the series (described by R^2). *bR2* has its optimal value at 1. *NSE* is a dimensionless statistical metric (Nash and Sutcliffe, 1970, also known as model efficiency) that determines the relative magnitude of the residual variance compared to the observed data variance (here considered to be SC_{FUL} or SC_{TAR} according to the specific case). It is obtained by dividing the mean squared error (MSE) by the variance of the observations and subtracting that ratio from unity:

$$NSE = 1 - \frac{\sum_{t=1}^{n}(y_t - x_t)^2}{\sum_{t=1}^{n}(x_t - \mu_x)^2} = 1 - \frac{MSE}{\sigma_x^2} \quad (5)$$

where n is the total number of (time-related) data points, y_t is the actual *SC* value at time-step t, x_t is the reference *SC* value (SC_{FUL} or SC_{TAR}) at time-step t, and μ_x and σ_x are the mean and standard deviation of the series. *NSE* ranges from $-\infty$ to 1, the closer to 1 the more the two series are similar (NSE < 0 indicates that using the actual *SC* is no better than using its mean). According to Gupta et al. (2009), we decomposed the NSE values into separate components to stress the causes of possible series difference. These components represent the relative contribution of correlation (named $f1 = 2{\cdot}\alpha{\cdot}r$, with $\alpha = \sigma_y/\sigma_x$, the ratio of actual and reference *SC* and r their linear correlation coefficient), variability ($f2 = \alpha^2$) and normalized constant bias ($f3 = \beta_n^2$, with $\beta_n = (\mu_y{-}\mu_x)/\sigma_x$). To evaluate the error in variable units we used the MAE given by

$$MAE = N^{-1}\sum_{i=1}^{N}|x_t - y_t| \quad (6)$$

which describes the absolute uncertainty in terms of average deviations and where x_t and y_t are the reference (SC_{FUL} or SC_{TAR}) and the actual *SC* values at time-step t, respectively. Note that SC_{FUL} or SC_{TAR} were alternatively considered as reference flux according to which source of uncertainty we were interested in. As example, looking at the effect of the profile vertical configuration and density we used the storage flux from the respective full profile configuration (SC_{FUL}) as reference. Looking at the effect of using one single profile only, we used the *SC* flux that has been considered as the best approximation of the real flux (SC_{TAR}).

2.2.3. Storage flux uncertainty impact on EC data

After differences among *SC* obtained by the various LRP and combinations were assessed, we continued our analysis based on three *SC* typologies only: the SC_{FUL} and SC_{ONE} as they represent the two most widely used setups, and the SC_{OPT} as identified after the evaluation of the measurement system performance in relation to its complexity. We assessed the *SC* uncertainty in relation to *FC* and its impact on NEE estimates.

3. Results

3.1. Profile configuration impact on storage flux estimates

We compared individual LRP *SC* versus the respective SC_{FUL}, here taken as reference, with the aim of evaluating this effect on the most common measurement approach, i.e. the single profile. This mono-dimensional analysis was applied to each tower independently (both main and satellites towers) though, given the consistency of the

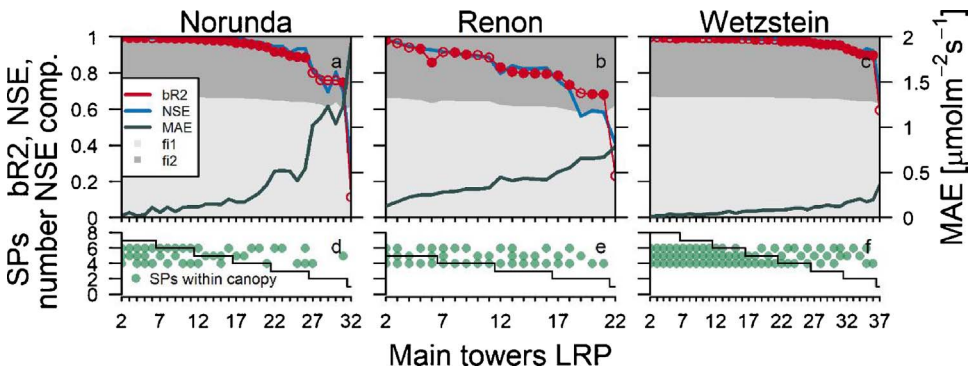

Fig. 3. Upper panels: performance statistics (bR2, NSE, the relative contribution of its components and MAE) of LRP against full profile storage fluxes of main towers. For bR2, full circles highlight slope > 1 and empty circles slope < 1. The NSE normalized bias component (fi3) is not visible due to its low relative contribution. Lower panels: a representation of individual LRP features is reported. Solid line shows the actual number of sampling points (SPs) and green circles the number and relative position of SPs within the canopy layer. The LRP ID resemble that of Fig. 2. Statistics are reported for the 3 sites, Norunda (a, d), Renon (b, e) and Wetzstein (c, f). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

obtained results, only the former has been reported (see supplementary material for the latter). The selected performance statistics are plotted against the respective LRP in Fig. 3, ordered according to Fig. 2. As expected, the performance decreases along with the number of considered levels. However, they are not linearly related as performance depends on the actual SP distribution. For NO an example is given considering two LRP having the same SPs distribution below and above the canopy layer, though with 3 and 1 SP respectively within it (LRP ID 4 and 14, with a total of 7 and 5 SPs over 8, see also Fig. 2). While bR2 and NSE remain around 1.0 indicating a good reproduction of the target flux data variance, MAE increases from 0.02 to 0.15 μmol $m^{-2} s^{-1}$ and the proportional bias from 0.4% to 1.25% respectively. Rearranging the same number of SPs (5) by moving one SP inside the canopy entails a reduction of MAE to 0.1 μmol $m^{-2} s^{-1}$ and a bias reduction to 0.7% (LRP ID 12).

Considering a half of the original SPs and placing one SP in the middle of the canopy layer (LRP ID 17) the MAE of resulting *SC* is 0.18 μmol $m^{-2} s^{-1}$, yet if the canopy point is moved to the upper canopy level MAE doubles (LRP ID 21). At NO estimates start to be critically affected when half of the SPs are considered: proportional biases are in the range of ± 7% and MAE of 0.2–2.0 μmol $m^{-2} s^{-1}$. At RE both bR2 and NSE start to markedly decrease as soon as even just one SP is omitted. At this site exemplary results are obtained by considering three LRP with 3 and 1 SPs within, 1 below and 1 above the canopy (LRP ID 2 and 10 with a total of 5 and 4 SPs respectively). MAE is 0.13 and 0.32 μmol $m^{-2} s^{-1}$ respectively while biases goes to 0.17% to −0.63%. However in this case, keeping the same SPs number (4) and moving one of the below canopy points into the canopy layer causes a worsening of the performances with a bias of −0.74%. Also at RE *SC* estimates start to be critically affected by profile density when half of the SPs are considered, but here overestimations reach 10% and MAE are in the range of 0.45–0.65 μmol $m^{-2} s^{-1}$, 20% to 35% of the typical nighttime fluxes at this site. At WS the effect of SPs density and positioning seem less evident. Considering again as example three LRP with 3, 2 and 1 SPs within the canopy layer, 2 below and 1 above it (ID 13, 19 and 25 with 6, 5 and 4 SPs over a total of 9), NSE and bR2 remain around unity, yet MAE doubles being 0.06, 0.08 and 0.12 μmol $m^{-2} s^{-1}$ respectively (3–6% of the typical nighttime fluxes). For this site the performances of LRPs start to worsen after more than half of the SPs are excluded from the *SC* computation.

Each statistic showed, at all sites, the unsuitability of computing *SC* with one SP only (LRP ID 32, 22 and 37 for the three sites respectively). The very low values of bR2 that in the best case of WS is near 0.6, reflect both a lack of correlation and a systematic underestimation which reaches the 62%, 47% and 17% of SC_{FUL} at NO, RE and WS respectively. MAE reaches values of 1.9, 0.8 and 0.4 μmol $m^{-2} s^{-1}$ in the main towers of NO, RE and WS respectively, and similar values were observed for satellites. By considering the NSE components, it is noticeable that the degree of agreement between SC_{FUL} and LRP *SC* is driven by the contribution of correlation (*f1* in Fig. 3) that on average is responsible for about the 60% of the statistic. The remaining 40% is almost entirely due to the effect of variability (*f2*). However, the latter component starts to become more and more important as soon as less SPs are considered. The constant bias contribution (*f3*) turns out to be negligible in all case. Taking the components of bR2 separately (see supplement material) shows that the slopes start to critically depart from 1 when 50% of SPs are used while correlation (R^2) remains high in general.

3.2. CO_2 spatial variability impact on storage flux estimates

We first assessed the performance of each individual full tower profile *SC*. To avoid spurious effects given by the different configuration of main and satellite tower profiles, we made a further profile manipulation. For the main towers we considered as full profile the one obtained by a subset of original SPs to match the number and displacement of the SPs of the concurrent satellites. Then, at each site, the target flux has been obtained by averaging the *SC* from the main tower (obtained as described above) and those from the full profile of the 4 satellites. By this analysis we assessed the error that potentially affects the measurement when a single profile (the most common approach) is taken as representative of the whole EC footprint. Even though this

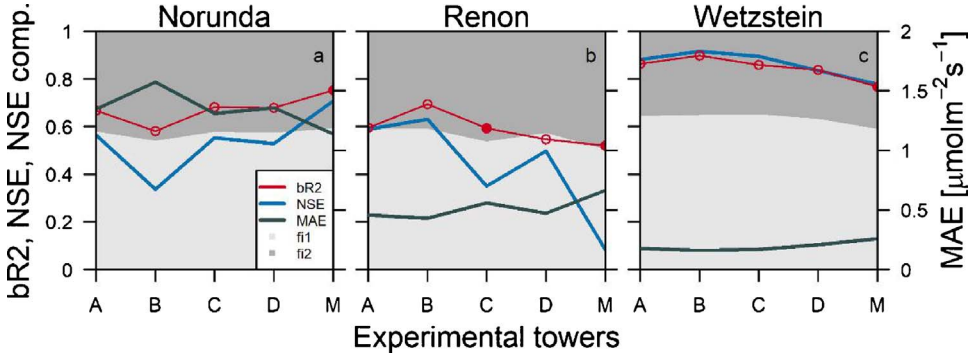

Fig. 4. Performance statistics (bR2, NSE, the relative contribution of its components and MAE) for the comparison of individual profile storage fluxes of main (M, subsampled to match the satellites configuration) and satellite towers (A, B, C, D) against the storage flux considered as reference (i.e. their mean). For bR2, full circles highlight slope > 1 and empty circles slope < 1. Statistics are reported for the 3 sites, Norunda (a), Renon (b) and Wetzstein (c).

analysis depends more on site-specific conditions, still interesting results can be obtained by analysing individual cases. In Fig. 4 bR2, NSE and MAE statistics of regressions are plotted for each tower. In general, relative indices (bR2 and NSE) are rather low while MAE is high.

As the three sites are characterized by different orography and vegetation density, results can be interpreted based on these differences. At NO, the soil surface is flat thus the variability of statistics likely depends on other causes (uneven local sources, wind patterns, or within canopy leaf area density profiles). The *SC* measured on M was the most similar to SC_{TAR} (here the reference *SC*) with a bR2 of 0.75, a NSE of 0.71 and a MAE of 1.14 μmol $m^{-2} s^{-1}$. On the contrary *SC* on satellite B showed the worst performance with bR2 decreasing to 0.58, NSE to 0.34 and MAE rising up to 1.57 μmol $m^{-2} s^{-1}$. In this site bR2 values are driven by its correlation component as slopes vary within −2% and −5%, indicating that important under- or overestimations do not verify independently from the location where *SC* is measured. At RE, orography conditions may play an important role as the station is located on a mountain slope (see Fig. 1). In addition, the vegetation distribution is rather uneven compared to the other sites. Both characteristics are reflected on the statistics variability. bR2 highlighted a general poor agreement between individual SC_{FUL} and SC_{TAR}, with values between 0.5 and 0.6. From its components a generally higher dispersion of values as (indicated by low R^2) and differences in flux magnitudes ranging from −3.2% and −11.3% for satellites B and D and to −10.76% and 11.9% for satellites A and C are noticeable. Is interesting to link these results with the actual location of towers. The former satellites (B and D) are positioned about at the same elevation as M (perpendicularly with respect to the main slope) while the latter (A and C) are along the slope. NSE and MAE values highlight the worst agreement for the M and C towers, showing that *SC* measured at these points neither reproduces *SC* temporal dynamics (NSE of 0.1 and 0.35) or magnitude (MAE of 0.66 and 0.56 μmol $m^{-2} s^{-1}$). At WS elevation changes are less pronounced with respect to RE and the vegetation is homogeneously distributed within the experimental area. Performance statistics are higher and consistent for each tower. This may suggest that *SC* variability, magnitude and dynamics are essentially reproduced by individual profiles independently from the sampling location. Anyway, the *SC* that diverged most from the reference fluxes was the one measured on M. The regression slope revealed an overestimation of about 11.5%, compared to underestimations observed for the other towers that range between −2 and −4%. The information deduced from NSE components was similar to that obtained from the profile density analysis, i.e. about a 60% of the statistic is due to the contribution of correlation and the remaining 40% is due to variability (*f1* and *f2* in Fig. 4).

The analyses above were based on individual profiles *SC* (LRP *SC* against SC_{FUL} in the first, SC_{FUL} against SC_{TAR} in the second). We then explored the effect of considering multiple profiles with a certain number of SPs in the control volume (along the satellites) in addition to the ones on the main tower. As the number of all the possible combinations of profiles and SPs was massive and difficult to evaluate, we selected a set of profile combinations as reported in Fig. 5 (and all are reported in the supplementary material). In practice, after a full statistical survey, we selected three profile typologies for M, i.e. full profile (ID = 1 in Figs. 2 and 5 for all the sites), the best performing profile among those with the 50% of total SPs (which consistently among Figs. 2, 3, 5 are identified by LRP ID 17 at NO with SPs at 1.5, 8.5, 19.0 and 31.7 m; LRP ID 12 at RE with SPs at 4.0, 14.0 and 32.0 m; LRP ID 22 at WS with SPs at 2.0, 9.0, 15.0 and 23.8 m) and the one-point profiles (which consistently among Figs. 2, 3, 5 are identified by LRP ID 32 at NO with the SP at 36.9 m; LRP ID 22 at RE with the SP at 32.0 m; LRP ID 37 at WS with the SP at 30.0 m). Then we used them to calculate *SC* in combination with 1–4 (all) satellite towers according to all the LRPs reported in Fig. 2.

Within each group, the amplitude of statistics' variability reflects the performances of the actual satellites LRP. For example at NO, considering the group that combines the semi-profile of M with 4 satellites (17 + 4 s in Fig. 5) the best statistics were clearly obtained by considering the full profile along satellites (bR2 = 0.99, NSE = 0.99, MAE = 0.03 μmol $m^{-2} s^{-1}$ and bias around = 0.5%). Though by considering only 2 out of 4 points on the satellites, the performances sensibly worsen (bR2 = 0.7–0.9, NSE = 0.6–0.9, MAE = 0.6–1.3 μmol $m^{-2} s^{-1}$ and bias up to15%). Two contrasting cases are especially interesting because of their potential application. The first is given by the setup that considers SC_{FUL} on M and 2 satellites (1 + 2 s in Fig. 5). When the 2 satellites were taken with their full profile, the resulting *SC* showed an average (N = 6) bR2 and NSE above 0.9 at each site, a MAE of 0.5–0.6, 0.18–0.22, less than 0.1 μmol $m^{-2} s^{-1}$ and a proportional bias of less than 1%, −4–6% and 1–2% at NO, RE and WS respectively. The second case is given by the setup that considers SC_{ONE} on M and 4 satellites (32 + 4s, 22 + 4s, 37 + 4 s in Fig. 5 for NO, RE and WS respectively). When the latter were taken with their full profile, the resulting *SC* showed still acceptable performances as bR2 was of 0.85 for NO and RE and 0.94 for WS, NSE above 0.9 at each site, and MAE about 0.38, 0.16 and 0.1 μmol $m^{-2} s^{-1}$ indicating that differences in flux dynamics, variances and average values were only small. However, fluxes were affected by a large bias that reaches −12%, −11% and −4% at NO, RE and WS, respectively.

To further assess the effect of the three-dimensional sampling on *SC* estimates by using a simpler measurement setup, we finally considered a singular profile (the one on M) after having averaged the $[CO_2]$ measured at a certain number of SPs with the simultaneous $[CO_2]$ measured at the same level along the satellite towers. As main and satellite towers had different SP distributions, such averaging was performed using linearly interpolated $[CO_2]$ on satellites. In Fig. 6 performance statistics are reported for a set of averaging strategies identified because of their practical feasibility. In general, all metrics were rather good, with MAE in the range of 0.25–0.75 μmol $m^{-2} s^{-1}$ at NO, 0.17–0.57 μmol $m^{-2} s^{-1}$ at RE and 0.1–0.24 μmol $m^{-2} s^{-1}$ at WS. The configuration named "ALL", which implies the use of all the SPs in both M and satellites, predictably performed best, yet acceptable values were also obtained by simpler setups. As example, the configuration that implies the mixing of the two lowermost SPs and one within the canopy layer (named 2L + C in Fig. 6) showed NSE values of 0.95, 0.55 and 0.85, bR2 of 0.93, 0.68 and 0.81 and MAE of 0.41, 0.44 and 0.21 μmol $m^{-2} s^{-1}$ for NO, RE and WS respectively. The proportional bias that characterized the *SC* estimates from this setup was always positive and limited to 3% at NO, 2% at RE (overall range is 1–7% and from 1 to 6% respectively) and to 10% at WS (ranging from 10 to 15%).

3.3. Optimal set-up selection

Based on the trade-off between performance and measuring system complexity (see the supplementary material for further detail) we identified an optimal setup (named 2L + C in Fig. 6 and optimal, SC_{OPT}, in the following). This setup requires the use of a vertical profile of which the $[CO_2]$ sampled at three SPs, the two lowermost and one in the canopy, are averaged using the corresponding levels in the four satellites (simulating a sampling in multiple locations with the mixing of the sampled air). Note that such a setup does not involve additional sampling levels but just a manipulation of the original ones. The resulting *SC* estimate has been considered in the following analyses together with SC_{FUL} and SC_{ONE} (estimated on M), and compared against the reference fluxes SC_{TAR}. Fig. 7 shows site-specific average *SC* diurnal cycles as obtained by these four different estimation schemes (SC_{FUL}, SC_{ONE}, SC_{OPT}, SC_{TAR}), their average half-hour difference with respect to SC_{TAR}, and concurrent *FC* diurnal cycles. While a discussion on *SC* intensity is beyond the scope of this analysis, it is worth to note the different magnitude of fluxes among sites, especially with respect to *FC*, as it influences the results obtained in the subsequent analyses (a companion figure is reported in the supplement, with scaled y-axes to better appraise *SC* and *SC* error, and with *FC* and *NEE* daily cycle for comparison of fluxes). It is noticeable that the diurnal trend is in general

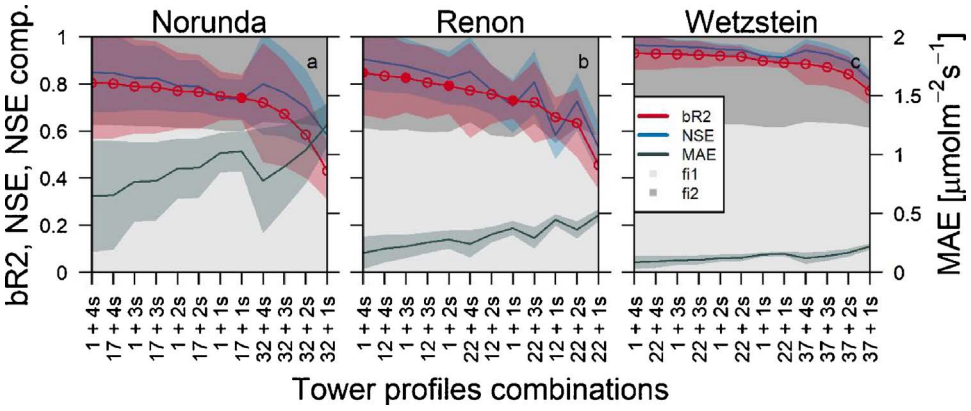

Fig. 5. Performance statistics (bR2, NSE, the relative contribution of its components and MAE) of different combinations of main tower profiles and groups of satellite towers, against target storage fluxes (SC_{TAR}). For the main towers, 3 profile typologies were considered: 1) full profile, with all originally available sampling points (identified by ID 1 in the x-axis); 2) the best performing among those with 50% of levels (ID 17, 12 and 22 in the x-axis for NO, RE and WS respectively); 3) one-point, using only one level at the tower top (ID 32, 22 and 37 in the x-axis for NO, RE and WS respectively). For satellite towers groups were made considering the same number of towers, independently from their profile configuration. They are identified by the number of actually used towers followed by 's' (e.g. 3 s means that 3 satellite towers were considered). For bR2, full circles highlight slope > 1 and empty circles slope < 1. Statistics are reported as mean and standard deviation within each group, for the 3 sites, Norunda (a), Renon (b) and Wetzstein (c).

followed by the different estimations with negative values denoting a depletion of the storage whereas positive values denoting that CO_2 is added to the storage. Though, SC_{ONE} shows the widest departure from SC_{TAR} during the morning and evening peaks, with underestimations of the order of actual fluxes. At WS this effect is less evident but yet present. SC_{FUL} also shows a certain degree of differences with SC_{TAR}, to some extent higher than SC_{OPT}.

3.4. Storage error impact on NEE

We then focused on *SC* setup-related errors in relation to *FC* and its impact on consequent NEE estimates. Only a light filtering has been made on half-hourly values, by removing a few exceptionally large values considered as spikes (± 60 µmol m^{-2} s^{-1} for NO and ± 40 µmol m^{-2} s^{-1} for RE and WS) and no attempt has been made to fill gaps in the datasets.

In Fig. 8 the distribution of the *SC* absolute error (SC_{FUL}, SC_{ONE}, SC_{OPT} respect to SC_{TAR}), expressed as percentage of concurrent *FC* values is shown and reported for daytime and nighttime conditions as median within flux bins of 4 µmol m^{-2} s^{-1}. At NO the measurement error of SC_{ONE} spans from 45% to 87% of the *FC* within ± 4 µmol m^{-2} s^{-1} and from 13% to 35% of the *FC* between ± 4 and ± 8 µmol m^{-2} s^{-1}. By considering SC_{OPT} the error can be reduced by 75% and 50% for the two *FC* classes respectively. At this site using a single full profile (SC_{FUL}) may still causes errors of the order of 20–40% of the most represented *FC* for both day and night time. At RE the relative error made by estimating *SC* from the top-most SP only is even bigger. During daytime the most represented flux classes are between 0 and −12 µmol m^{-2} s^{-1} (for a cumulative contribution of 60% of fluxes) and the *SC* error spans from 20 to 97% of concurrent *FC*. During nighttime the most represented *FC* class is the 0–4 µmol m^{-2} s^{-1} (60% of fluxes) and the error is around 80% of *FC*. In this case, measuring *SC* with an optimized setup may halve this error. At WS the advantage of using SC_{OPT} instead of SC_{ONE} is evident for night fluxes only. In fact, while during the day the errors are equivalent for the most represented *FC* classes, the error affecting the 0–4 µmol m^{-2} s^{-1} class during night (70% of fluxes) can be reduced from 22% to 6%.

To put into perspective the error on *SC* measurement and its consequence on NEE estimates, we compared the selected *SC* and concurrent NEE fluxes against respective TAR values (Fig. 9). Normal-probability contours of comparisons are plotted by using 25, 50, 75 and 95% confidence levels. The departure from the 1:1 ratio is generally evident for SC_{ONE} and SC_{FUL}, but when projected to NEE, differences were strongly smoothed out. However, while relative metrics improved, MAE are conserved. Hence, also for NEE, lower fluxes are the ones prone to major uncertainties. This effect is again more evident at NO with respect to RE and WS, suggesting that *SC* error only marginally affects the actual NEE estimates in the latter two sites.

Lastly, we performed an *ad-hoc* analysis of the NEE differences (by mean of the Bland-Altman method, Altman and Bland, 1983) to better assess the effective level of disagreement between estimates. In Fig. 10, the differences between two actual estimates are plotted against their mean. Two main inferences can be made: first, although the average differences are not exactly zero (as it should be in case of a fully stochastic variability of differences), they do not represent an effective bias (constant in this case) because their confidence interval crosses the zero line; second, the wide range of agreement, i.e. the range within which the 95% of differences are included, observed for NEE_{ONE} (panels on the left) is sensibly reduced for NEE_{FUL} and even more for NEE_{OPT} (central and right panels respectively). Similarly to the results obtained

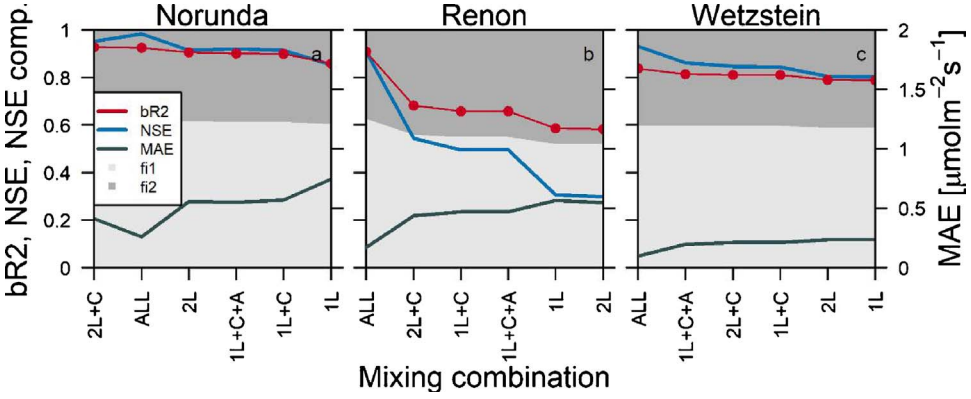

Fig. 6. Performance statistics (bR2, NSE, the relative contribution of its components and MAE) for the comparison of storage fluxes obtained by different combinations of individual sampling level [CO_2] mixing against target storage (SC_{TAR}) fluxes. Setup acronyms indicate which SP level were used for averaging: ALL, all the sampling points in M and their respective on satellites; 1L, only the lowermost level (1.5 m at NO, 1 m at RE, 0.1 m at WS); 2L, the two lowermost levels (1.5 and 8.5 m at NO, 1 and 2 at RE, 0.1 and 0.3 m at WS); 1L + C, the lowermost level and one SP in the lower part of the canopy space (1.5 and 13.5 m at NO, 1 and 8 m at RE, and 0.1 and 5 m at WS); 2L + C, the two lowermost levels and one in the lower part of the canopy space (1.5, 8.5, and13.5 m at NO, 1, 2, 8 m at RE, 0.1, 0.3, 5.0 m at WS); 1L + C + A, the lowermost level, one in the canopy and one just above the canopy top (1.5, 13.5 and 28.0 m at NO, 1, 8, 32 m at RE, 0.1, 5.0, 23.8 at WS). For bR2, full circles highlight slope > 1 and empty circles slope < 1. Statistics are reported for the 3 sites, Norunda (a), Renon (b) and Wetzstein (c).

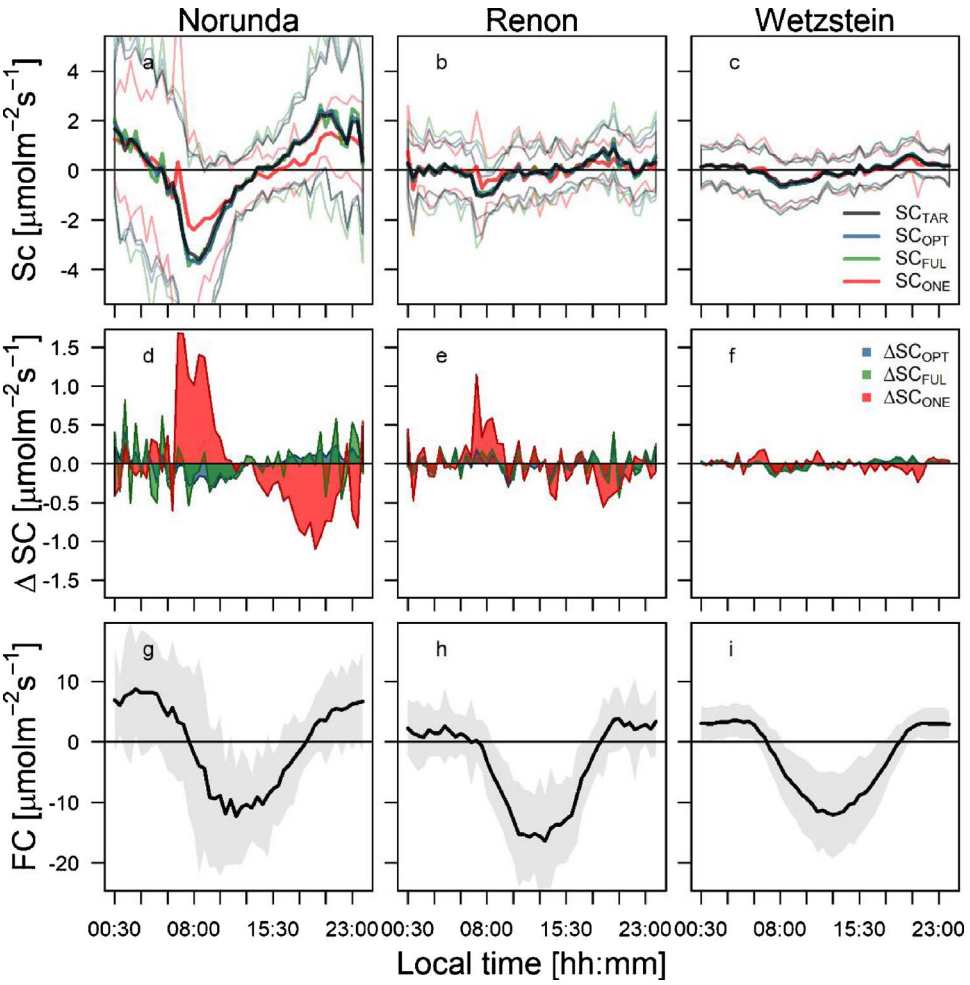

Fig. 7. Upper panels: mean diurnal cycles of selected storage fluxes (SC_{FUL}, SC_{ONE}, SC_{OPT}, SC_{TAR}) over the whole experimental periods. Values are reported as mean (thick lines) and standard deviation (thin lines) for each half-hour. Middle panels: SC absolute errors (residuals with respect to SC_{TAR}) averaged within each half-hour. Lower panels: mean diurnal cycles of EC fluxes (FC) reported as mean (thick lines) and standard deviation (shadow) for each half-hour. Data for Norunda (a, d, g), Renon (b, e, h) and Wetzstein (c, f, i).

above (Fig. 8 and 9) the magnitude of these differences would not be significant for high fluxes, yet large for the lower ones. For example, the contribution to the NEE uncertainty that derives from considering SC_{ONE} would be about ± 5.5, ± 2.2 and ± 1.0 µmol m^{-2} s^{-1} at NO, RE and WS respectively, reduced to ± 1.4, ± 1.3 and ± 0.6 µmol m^{-2} s^{-1} by considering SC_{OPT}. In general, with respect to measuring *SC* by the one-point approach, using the whole profile entails a reduction of the storage-related NEE uncertainty of about 37%, 22% and 30%, while using an optimized setup, such errors are reduced by 75%, 41% and 40% for NO, RE and WS respectively.

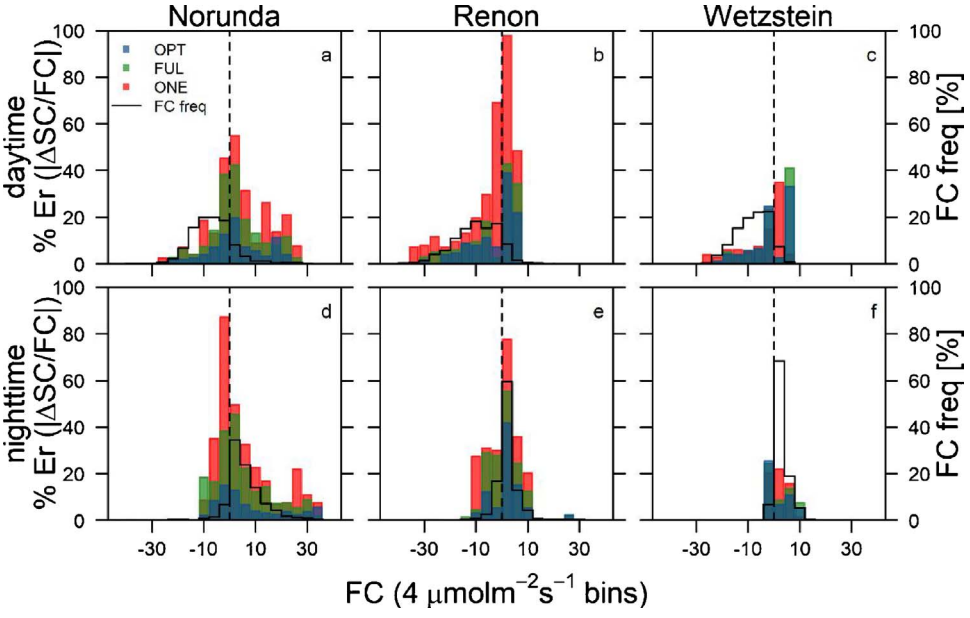

Fig. 8. Distribution of absolute errors of SC fluxes (SC_{FUL}, SC_{ONE}, SC_{OPT} respect to SC_{TAR}) expressed as percentage of concurrent EC fluxes (FC) during daytime (upper panels) and nighttime (lower panels) for Norunda (a, d), Renon (b, e) and Wetzstein (c, f). Daytime was determined by the condition PPFD > 5 µmol photons m^{-2} s^{-1}. Errors are reported as median within flux bins of 4 µmol m^{-2} s^{-1}. FC and SC fluxes < 2 µmol m^{-2} s^{-1} were not used for error estimation and classes with less than 5 observations were discarded. FC frequency distribution (black line) is reported on the secondary y-axis.

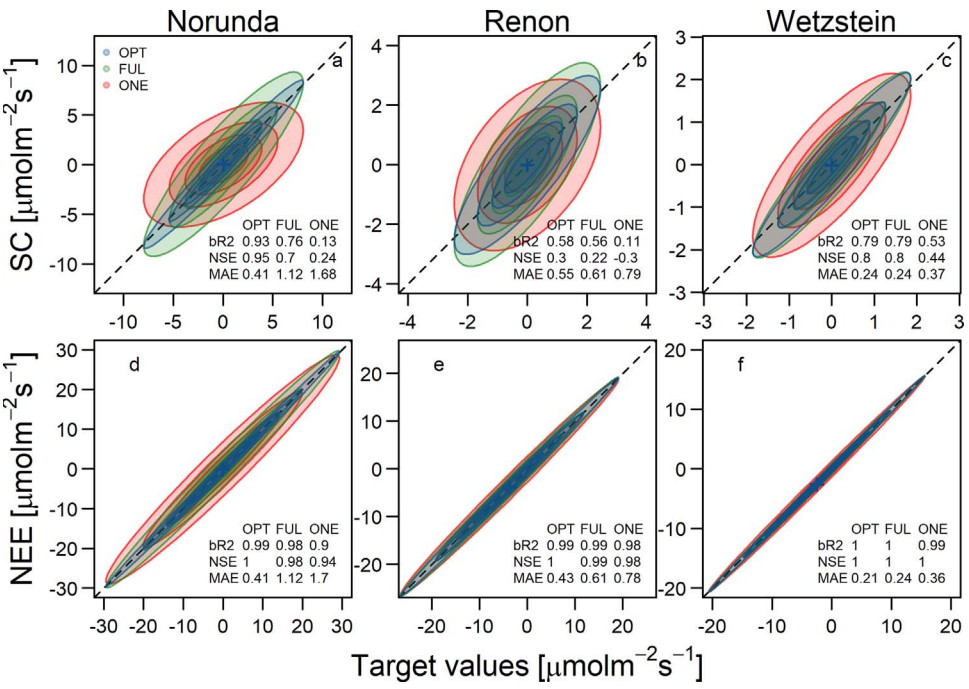

Fig. 9. Selected storage fluxes (SC_{FUL}, SC_{ONE}, SC_{OPT}, upper panels) and consequent NEE estimates (lower panels) compared to respective target values. Normal-probability contours represent the 25, 50, 75 and 95% confidence level, the dashed line represents the 1:1. Performance statistics (bR2, NSE and MAE) are reported in each plot corner. Results are displayed for Norunda (a, d), Renon (b, e) and Wetzstein (c, f).

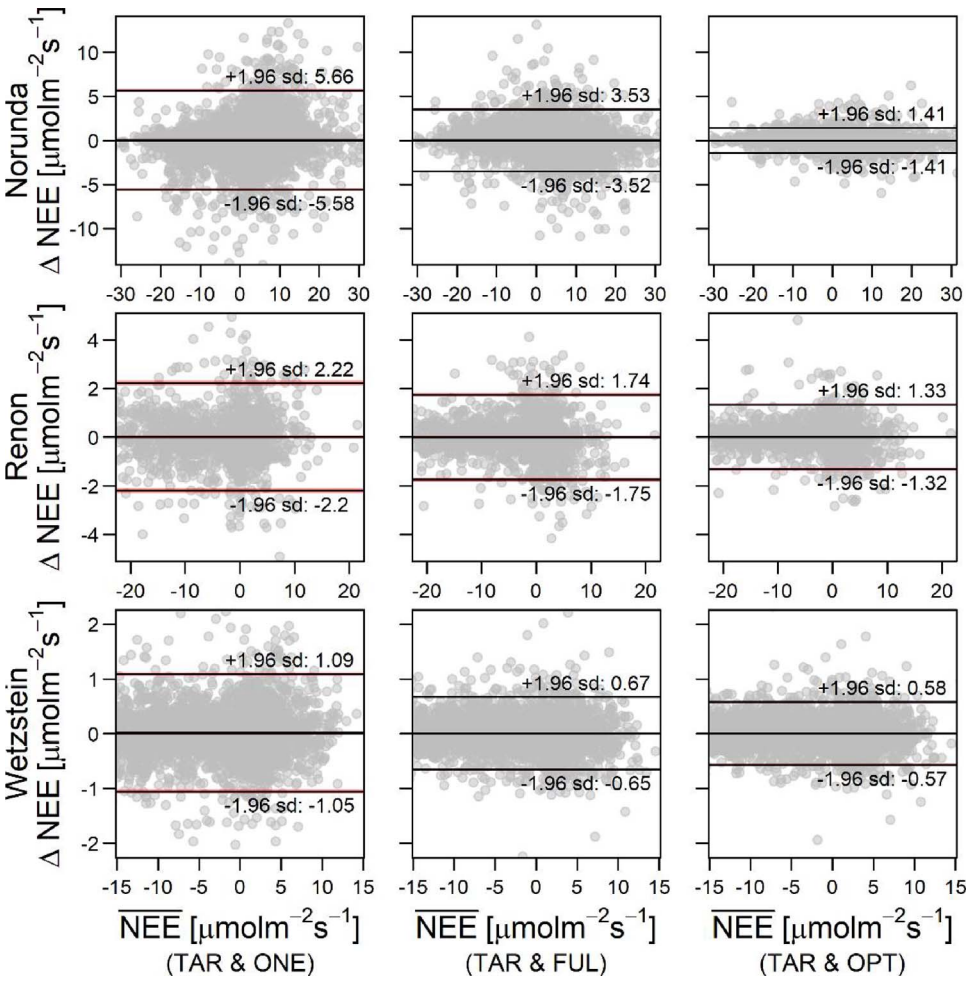

Fig. 10. Bland and Altman analysis for half-hourly NEE estimated by using the selected storage fluxes (differences between 'target' and 'one-point', 'full' and 'optimal' NEE against their means). The limits of agreement extend from −1.96 σ to +1.96 σ while the bias is represented by the gap between the mean (the line close to zero) and zero. Red shaded areas around mean and agreement limits lines show their confidence interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

In this paper we assessed the error in *SC* estimates due to a limited spatial resolution sampling of CO_2 variability, selected an effective measurement setup and finally, quantified the impact of such errors on consequent NEE estimates. Used data were collected during the ADVEX experiment, by means of a setup that was chosen and optimized to detect advection and not storage fluxes. Still, its design was suitable to assess the importance of storage fluxes, being by far more complete of what is generally done at eddy covariance sites. To evaluate the performance of the considered measurement setups, we used some effective dimensionless statistical measures named bR2 and the NSE, and the mean absolute error (MAE) as a measure of the uncertainty in absolute (and physical) terms. Although we considered the selected statistics as the most efficient in describing the relations between fluxes, they are not flawless and should be interpreted in conjunction with other measures of error. In the supplementary material of this article, additional statistics are reported and can be assessed to get further insights (reported results are not affected anyway).

The analysis of vertical profile configuration and density showed that performances decrease with the number of considered levels, though the decrease is not linearly related to the SP number as it depends on their distribution, mostly in relation to the canopy layer. This has been reported also by Wang et al. (2016) and Yang et al. (2007) and was observed at all the three analysed sites. This is probably due to, on one side, the important contribution of autotrophic respiration and photosynthesis in the canopy, and on the other, the structure of the canopy that promote the CO_2 accumulation. Given the same number of SPs a proper sampling of the canopy layer reduces both the estimate bias and uncertainty. However, at sites where a significant understorey vegetation is present (not the case for the sites analysed here) it may be important to properly sample this layer too.

We reported only the measuring performance of the top 5 LPR for a given number of sampling points (e.g. Fig. 3). This means that any different distributions of points cause a sensible worsening of the statistics. However, also the positioning of SPs inside the canopy layer depends on site-specific conditions. For example, differently from Norunda, at Renon and Wetzstein the lower canopy edge seems to play a major role as the point that samples in it is maintained even when a few sampling points are considered (see Fig. 3). All statistics highlighted, in every case, the divergence between *SC* measured with full profile and with the topmost SP only. The systematic underprediction reached 62%, 47% and 17% at NO, RE and WS respectively, reflecting the site-specific level of decoupling among the EC level and the ecosystem beneath. Also the resulting MAE values were high, equalling the 20–40% of the typical nighttime *SC* fluxes at the sites. This confirms the importance of measuring *SC* with, at least, one vertical profile in forest ecosystems, and that the use of the discrete approach based on a single measurement point at the eddy covariance system level introduces large errors and uncertainty.

The analysis of the effect of horizontal spatial sampling highlights the importance of the location when a single profile is used. In case of flat surfaces, as it was the case of Norunda, the error resulting from the use of individual tower profiles instead of a set of spatially distributed profiles was characterized by an increase of the uncertainty (MAE reached 1.6 $\mu mol\ m^{-2}\ s^{-1}$) rather than a real flux bias that was limited to 5%. In case of more complex orography as in Renon and Wetzstein, where advective fluxes are expected, we observed a different situation. At Renon individual towers *SC* poorly reproduce the target *SC* temporal dynamics (highlighted by bR2 and NSE) or magnitudes (MAE up to 0.7 $\mu mol\ m^{-2}\ s^{-1}$) and resulting biases critically ranged from −12% to 14%. In this site, the actual position of the profile has a great impact on the flux estimate. This is in line with a recent and detailed model-based study on local circulation at this site (Xu et al., 2017). At Wetzstein, individual tower estimates were more consistent. However, while relative (dimensionless) metrics remain at or above 0.8 in all cases, proportional biases ranged from −2% to 12% and MAE reached 0.25 $\mu mol\ m^{-2}\ s^{-1}$. These are not trivial values if we consider the low *SC* flux magnitude at this site.

The analysis of the *SC* fluxes derived from different combinations of towers and profile highlighted that *i.* the deployment of multiple towers does not automatically lead to improvements of *SC* estimates as it depends on actual tower profile designs and their location, and *ii.* despite certain multiple tower setups showed good performances (e.g. such as those involving the use of dense profiles) their implementation is however over-demanding, becoming unrealistic in the majority of experimental sites. To follow-up on this, we assessed the performance of possible setups that allow for the best sampling of $[CO_2]$ spatial gradient, being feasible at the same time. A very good approximation of the target *SC* flux was observed at Norunda and Wetzstein, while at Renon SC performance statistics were good but in a minor extent. One of the closest approximation of the target flux was provided by a relatively simple setup, with a vertical profile where 3 levels (the two closest to the ground and the one at the lower fringe of the canopy layer) have inlets for the sampling in different locations. This result highlighted the importance of the spatial variability of CO_2 accumulation below the canopy, that should be sampled to get a more accurate estimation of the storage flux. Such a sampling approach resemble the one proposed by Marcolla et al. (2005) and the one currently adopted in the ICOS network.

The analysis of the *SC* error in relation with the concurrent estimate of the turbulent flux showed that the error has the biggest impact on low fluxes (i.e. up to $|8|\ \mu mol\ m^{-2}\ s^{-1}$), which represent the majority of nighttime and early-morning EC fluxes. The importance of such an error is evident also when compared to *FC* over short period just after sunset when, as also reported by Galvagno et al. (2017), the turbulent flux is dominant and the advection terms are still negligible. The error magnitude can be sensibly reduced by using the proposed optimized setup.

Results suggest that *SC* error may only marginally affect the actual NEE estimates in the analysed sites. As a consequence, for budget studies (over temporal scales higher than daily sums) the *SC* measurement error could actually be acceptable because of its tendency of becoming negligible compared to NEE fluxes. However, when diurnal dynamics have to be determined, or gap-filling procedures to be applied (since they are based on single half-hourly values), errors in individual NEE may be important. This was also recently reported by Desai et al. (2015) for methane fluxes. As confirmed by the comparison of different NEE obtained by the selected *SC* fluxes (Fig. 10), NEE was not affected by a constant bias, however, the wide ranges of agreement (considered as an estimate uncertainty) may be not acceptable, especially for lower fluxes. Measuring *SC* with a proper setup (SC_{OPT}), allows to reduce the uncertainty on individual NEE up to about 75% with respect to the one-point approach, and up to about 60% with respect to using a single full profile. In this case, the advantage of optimally measuring *SC* was more evident for Norunda then for Renon or Wetzstein. In consideration of the inverse relationship between storage and advection, this result was somehow expected because the role of *SC* becomes more important at sites where advection is weak (advection drains the CO_2 accumulated in the canopy causing low *SC* fluxes). In fact in the ADVEX experiment Norunda was chosen as a "blank" site for advection, while the two latter sites were chosen exactly for the possible presence of advection. Therefore, the effect of a non-optimal *SC* for most of the sites would probably be in between these two extreme situations.

5. Conclusions

From our analysis we can conclude that: *i.* it is possible to improve the measure of the storage term by optimizing the sampling design integrating the vertical profile with a set of horizontal CO_2 sampling points so as to improve the spatial representativeness; *ii.* along the vertical profile, the canopy layer has an high importance followed by

the layers close to the ground. Provided that additional points would always improve the quality of the *SC* measurement, when possible it is better to increase the horizontal sampling rather than adding extra vertical points; *iii.* when the error in storage measurement is put in relation with the resultant NEE fluxes, the effect of *SC* errors is, as expected, bigger when *FC* is low and a proper *SC* measurement can reduce the error by 75% *iv*. the uncertainty resulting from an improper/insufficient *SC* sampling has a strong effect on single half hours while it is less important when budgets are of concern, *v*. the degree of importance of the storage term and its uncertainty is site-specific, with advection playing a key role: sites with presence of advection have generally lower storage fluxes and smaller effect of its uncertainty.

The analysis was based on data collected in three different sites, yet results were consistent in terms of main error sources and measuring system optimal design, suggesting that this approach can be used also in other similar sites. More data and experiments could help to further investigate this issue, in particular, for different green-house gases (such as CH_4 and N_2O) currently measured with the eddy covariance technique for which the dynamic of the below-canopy accumulation could be sensibly different.

Acknowledgements

The data from the ADVEX experiment have been collected with the support of the EU Project CarboEurope-IP and the authors thank all the technicians and scientists involved in the experiments. Dario Papale and Giacomo Nicolini thank the EU for supporting the ENVRIplus project funded by the Horizon 2020 Research and Innovation Programme under grant agreement 654182 and the RINGO project funded under the same program under grant agreement 730944.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.agrformet.2017.09.025

References

Altman, D.G., Bland, J.M., 1983. Measurement in medicine: the analysis of method comparison studies. J. Roy. Stat. Soc. D-Sta. 32, 307–317. http://dx.doi.org/10.2307/2987937.

Aubinet, M., Berbigier, P., Bernhofer, C., Cescatti, A., Feigenwinter, C., Granier, A., Grünwald, T., Havrankova, K., Heinesch, B., Longdoz, B., Marcolla, B., Montagnani, L., Sedlak, P., 2005. Comparing CO2 storage and advection conditions at night at different carboeuroflux sites. Bound. Layer Meteorol. 116, 63–93. http://dx.doi.org/10.1007/s10546-004-7091-8.

Aubinet, M., Feigenwinter, C., Heinesch, B., Bernhofer, C., Canepa, E., Lindroth, A., Montagnani, L., Rebmann, C., Sedlak, P., Van Gorsel, E., 2010. Direct advection measurements do not help to solve the night-time CO2 closure problem: Evidence from three different forests. Agric. For. Meteorol. 150, 655–664. http://dx.doi.org/10.1016/j.agrformet.2010.01.016.

Bjorkegren, A.B., Grimmond, C.S.B., Kotthaus, S., Malamud, B.D., 2015. CO2 emission estimation in the urban environment: measurement of the CO2 storage term. Atmos. Environ. 122, 775–790. http://dx.doi.org/10.1016/j.atmosenv.2015.10.012.

Cescatti, A., Marcolla, B., Goded, I., Gruening, C., 2016. Optimal use of buffer volumes for the measurement of atmospheric gas concentration in multi-point systems. Atmos. Meas. Tech. 9, 4665–4672. http://dx.doi.org/10.5194/amt-9-4665-2016.

Desai, A.R., Xu, K., Tian, H., Weishampel, P., Thom, J., Baumann, D., Andrews, A.E., Cook, B.D., King, J.Y., Kolka, R., 2015. Landscape-level terrestrial methane flux observed from a very tall tower. Agric. For. Meteorol. 201, 61–75. http://dx.doi.org/10.1016/j.agrformet.2014.10.017.

Feigenwinter, C., Bernhofer, C., Eichelmann, U., Heinesch, B., Hertel, M., Janous, D., Kolle, O., Lagergren, F., Lindroth, A., Minerbi, S., Moderow, U., Mölder, M., Montagnani, L., Queck, R., Rebmann, C., Vestin, P., Yernaux, M., Zeri, M., Ziegler, W., Aubinet, M., 2008. Comparison of horizontal and vertical advective CO2 fluxes at three forest sites. Agric. For. Meteorol. 148, 12–24. http://dx.doi.org/10.1016/j.agrformet.2007.08.013.

Feigenwinter, C., Mölder, M., Lindroth, A., Aubinet, M., 2010a. Spatiotemporal evolution of CO2 concentration, temperature, and wind field during stable nights at the Norunda forest site. Agric. For. Meteorol. 150, 692–701. http://dx.doi.org/10.1016/j.agrformet.2009.08.005.

Feigenwinter, C., Montagnani, L., Aubinet, M., 2010b. Plot-scale vertical and horizontal transport of CO2 modified by a persistent slope wind system in and above an alpine forest. Agric. For. Meteorol. 150, 665–673. http://dx.doi.org/10.1016/j.agrformet.2009.05.009.

Finnigan, J.J., Clement, R., Malhi, Y., Leuning, R., Cleugh, H.A., 2003. Re-evaluation of long-term flux measurement techniques. Part I: averaging and coordinate rotation. Bound. Layer Meteorol. 107, 1–48. http://dx.doi.org/10.1023/A:1021554900225.

Finnigan, J.J., 2006. The storage term in eddy flux calculations. Agric. For. Meteorol. 136, 108–113. http://dx.doi.org/10.1016/j.agrformet.2004.12.010.

Foken, T., Aubinet, M., Leuning, R., 2012. 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. http://dx.doi.org/10.1007/978-94-007-2351-1.

Galvagno, M., Wohlfahrt, G., Cremonese, E., Filippa, G., Migliavacca, M., Mora, U., Gorsel Van, E., 2017. Contribution of advection to nighttime ecosystem respiration at a mountain grassland in complex terrain. Agric. For. Meteorol. 237–238, 270–281. http://dx.doi.org/10.1016/j.agrformet.2017.02.018.

Greco, S., Baldocchi, D.D., 1996. Seasonal variations of CO2 and water vapour exchange rates over a temperate deciduous forest. Glob. Change Biol. 2, 183–197. http://dx.doi.org/10.1111/j. 1365–2486.1996. tb00071.x.

Gu, L., Massman, W.J., Leuning, R., Pallardy, S.G., Meyers, T., Hanson, P.J., Riggs, J.S., Hosman, K.P., Yang, B., 2012. The fundamental equation of eddy covariance and its application in flux measurements. Agric. For. Meteorol. 152, 135–148. http://dx.doi.org/10.1016/j.agrformet.2011.09.014.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377, 80–91. http://dx.doi.org/10.1016/j.jhydrol.2009.08.003.

Heinesch, B., Yernaux, M., Aubinet, M., 2007. Some methodological questions concerning advection measurements: a case study. Bound. Layer Meteorol. 122, 457–478. http://dx.doi.org/10.1007/s10546-006-9102-4.

Iwata, H., Malhi, Y., von Randow, C., 2005. Gap-filling measurements of carbon dioxide storage in tropical rainforest canopy airspace. Agric. For. Meteorol. 132, 305–314. http://dx.doi.org/10.1016/j.agrformet.2005.08.005.

Kang, M., Ruddell, B.L., Cho, C., Chun, J., Kim, J., 2017. Identifying CO2 advection on a hill slope using information flow. Agric. For. Meteorol. 232, 265–278. http://dx.doi.org/10.1016/j.agrformet.2016.08.003.

Knohl, A., Schulze, E.-D., Kolle, O., Buchmann, N., 2003. Large carbon uptake by an unmanaged 250-year-old deciduous forest in Central Germany. Agric. For. Meteorol. 118, 151–167. http://dx.doi.org/10.1016/S0168-1923(03)00115-1.

Krause, P., Boyle, D.P., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97. http://dx.doi.org/10.5194/adgeo-5-89-2005.

Lagergren, F., Eklundh, L., Grelle, A., Lundblad, M., Meelis, M., Lankreijter, H., Lindroth, A., 2005. Net primary production and light use efficiency in a mixed coniferous forest in Sweden. Plant Cell Environ. 28, 412–423.

Lee, X.H., Fuentes, J.D., Staebler, R.M., Neumann, H.H., 1999. Long-term observation of the atmospheric exchange of CO2 with a temperate deciduous forest in southern Ontario, Canada. J. Geophys. Res. 104, 15975–15984. http://dx.doi.org/10.1029/1999JD900227.

Marcolla, B., Cescatti, A., Montagnani, L., Manca, G., Kerschbaumer, G., Minerbi, S., 2005. Importance of advection in the atmospheric CO2 exchanges of an alpine forest. Agric. For. Meteorol. 130, 193–206. http://dx.doi.org/10.1016/j.agrformet.2005.03.006.

Marcolla, B., Cobbe, I., Minerbi, S., Montagnani, L., Cescatti, A., 2014. Methods and uncertainties in the experimental assessment of horizontal advection. Agric. For. Meteorol. 198–199, 62–71. http://dx.doi.org/10.1016/j.agrformet.2014.08.002.

McHugh, I.D., Beringer, J., Cunningham, S.C., Baker, P.J., Cavagnaro, T.R., Mac Nally, R., Thompson, R.M., 2017. Interactions between nocturnal turbulent flux, storage and advection at an ideal eucalypt woodland site. Biogeosciences 14, 3027–3050. http://dx.doi.org/10.5194/bg-2016-184.

Moderow, U., Aubinet, M., Feigenwinter, C., Kolle, O., Lindroth, A., Mölder, M., Montagnani, L., Rebmann, C., Bernhofer, C., 2009. Available energy and energy balance closure at four coniferous forest sites across Europe. Theor. Appl. Climatol. 98, 397–412. http://dx.doi.org/10.1007/s00704-009-0175-0.

Moderow, U., Feigenwinter, C., Bernhofer, C., 2011. Non-turbulent fluxes of carbon dioxide and sensible heat-a comparison of three forested sites. Agric. For. Meteorol. 151, 692–708. http://dx.doi.org/10.1016/j.agrformet.2011.01.014.

Montagnani, L., Manca, G., Canepa, E., Georgieva, E., Acosta, M., Feigenwinter, C., Janous, D., Kerschbaumer, G., Lindroth, A., Minach, L., Minerbi, S., Mölder, M., Pavelka, M., Seufert, G., Zeri, M., Ziegler, W., 2009. A new mass conservation approach to the study of CO2 advection in an alpine forest. J. Geophys. Res. Atmos. 114, 1–25. http://dx.doi.org/10.1029/2008JD010650.

Montagnani, L., Manca, G., Canepa, E., Georgieva, E., 2010. Assessing the method-specific differences in quantification of CO2 advection at three forest sites during the ADVEX campaign. Agric. For. Meteorol. 150, 702–711. http://dx.doi.org/10.1016/j.agrformet.2010.01.013.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I – a discussion of principles. J. Hydrol. 10, 282–290.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., Yakir, D., 2006. Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: algorithms and uncertainty estimation. Biogeosciences 3, 571–583. http://dx.doi.org/10.5194/bg-3-571-2006.

Pattey, E., Strachan, I., Desjardins, R.L., Massheder, J.M., 2002. Measuring nighttime CO2 flux over terrestrial ecosystems using eddy covariance and nocturnal boundary layer methods. Agric. For. Meteorol. 113, 145–158. http://dx.doi.org/10.1016/S0168-1923(02)00106-5.

Priante-Filho, N., Vourlitis, G.L., Hayashi, M.M.S., Nogueira, J.D.S., Campelo, J.H., Nunes, P.C., Souza, L.S.E., Couto, E.G., Hoeger, W., Raiter, F., Trienweiler, J.L.,

Miranda, E.J., Priante, P.C., Fritzen, C.L., Lacerda, M., Pereira, L.C., Biudes, M.S., Suli, G.S., Shiraiwa, S., Paulo, S.R., Do Silveira, M., 2004. Comparison of the mass and energy exchange of a pasture and a mature transitional tropical forest of the southern Amazon Basin during a seasonal transition. Glob. Change Biol. 10, 863–876. http://dx.doi.org/10.1111/j.1529-8817.2003.00775.x.

Rebmann, C., Zeri, M., Lasslop, G., Mund, M., Kolle, O., Schulze, E.-D., Feigenwinter, C., 2010. Treatment and assessment of the CO2-exchange at a complex forest site in Thuringia, Germany. Agric. For. Meteorol. 150, 684–691. http://dx.doi.org/10.1016/j.agrformet.2009.11.001.

Siebicke, L., Steinfeld, G., Foken, T., 2011. CO2 gradient measurements using a parallel multi-analyzer setup. Atmos. Meas. Tech. 4, 409–423. http://dx.doi.org/10.5194/amt-4-409-2011.

van Gorsel, E., Delpierre, N., Leuning, R., Black, A., Munger, J.W., Wofsy, S., Aubinet, M., Feigenwinter, C., Beringer, J., Bonal, D., Chen, B., Chen, J., Clement, R., Davis, K.J., Desai, A.R., Dragoni, D., Etzold, S., Grünwald, T., Gu, L., Heinesch, B., Hutyra, L.R., Jans, W.W.P., Kutsch, W., Law, B.E., Leclerc, M.Y., Mammarella, I., Montagnani, L., Noormets, A., Rebmann, C., Wharton, S., 2009. Estimating nocturnal ecosystem respiration from the vertical turbulent flux and change in storage of CO2. Agric. For. Meteorol. 149, 1919–1930. http://dx.doi.org/10.1016/j.agrformet.2009.06.020.

Vickers, D., Irvine, J., Martin, J.G., Law, B.E., 2012. Nocturnal subcanopy flow regimes and missing carbon dioxide. Agric. For. Meteorol. 152, 101–108. http://dx.doi.org/10.1016/j.agrformet.2011.09.004.

Wang, X., Wang, C., Guo, Q., Wang, J., 2016. Improving the CO2 storage measurements with a single profile system in a tall-dense-canopy temperate forest. Agric. For. Meteorol. 228–229, 327–338. http://dx.doi.org/10.1016/j.agrformet.2016.07.020.

Xu, X., Yi, C., Montagnani, L., Kutter, E., 2017. Numerical study of the interplay between thermo-topographic slope flow and synoptic flow on canopy transport processes. Agric. For. Meteorol. http://dx.doi.org/10.1016/j.agrformet.2017.03.004.

Yang, P.C., Black, T.A., Neumann, H.H., Novak, M.D., Blanken, P.D., 1999. Spatial and temporal variability of CO2 concentration and flux in a boreal aspen forest. J. Geophys. Res. 104, 27653–27661.

Yang, B., Hanson, P.J., Riggs, J.S., Pallardy, S.G., Heuer, M., Hosman, K.P., Meyers, T.P., Wullschleger, S.D., Gu, L.H., 2007. Biases of CO2 storage in eddy flux measurements in a forest pertinent to vertical configurations of a profile system and CO2 density averaging. J. Geophys. Res. Atmos. 112, 1–15. http://dx.doi.org/10.1029/2006JD008243.

Yi, C., Davis, K.J., Bakwin, P.S., Berger, B.W., Marr, L.C., 2000. Influence of advection on measurements of the net ecosystem-atmosphere exchange of CO2 from very tall tower. J. Geophys. Res. 105, 9991–9999.

Zambrano-Bigiarini, M., hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series, R package version 0.3-10, 2017, http://hzambran.github.io/hydroGOF/. DOI: 100.5281/zenodo.840087.