

Dear Editor,

thank you for recommending the revised manuscript for publication. In this document we address the minor revisions suggested by Referee #3 and provide the manuscript with marked changes below.

*Wutzler and others demonstrate the REddyProc package. I feel that the manuscript will be an important contribution to the flux community given the number of users who will likely use the package.*

Thank you.

*Abstract was a bit qualitative, numbers would help. For example, how much lower and higher are the uncertainty estimates noted toward the end of the abstract?*

We updated the abstract:

Lower uncertainty estimates of both  $u^*$  and resulting gap-filled fluxes by 50% with the presented tool was achieved by ...

Higher estimates of uncertainty in day-time partitioning (about twice as high) resulted

*'allowed a boost in' is too wordy.*

We replaced by „greatly advanced“

*'and temperature' instead of 'or temperature' on line 7 page 2.*

We corrected as suggested.

*Figure 1 already includes a thumbprint plot for RE that seems a bit implausible. Is the red/orange horizontal line through the middle of the plot really true?*

The “reddish” band does not come from “gap filled” NEE estimates but the band coincides from mainly measured data where there is a very high NEE at nighttime (if you look the nighttime NEE immediately before the big gap looks quite high). Hence its in the original measurement data. The color scale is assigned so that the maximum value of the dataset gets the red color code.

We decided to not discuss this distracting detail in the paper.

*The legend of Fig. 2 is a bit elusive but can be figured out with careful study. Clarifying it would help.*

We reworded the figure caption as follows:

Concept of the  $u^*$ -filter: Night-time NEE at low  $u^*$  is biased towards lower NEE values compared to cases with higher  $u^*$ . Unbiased NEE should scatter around the same plateau because environmental conditions are similar. The  $u^*$  threshold (dashed line), i.e. the value below which this bias is considered significant, is estimated by a moving point method on  $u^*$  bins (crosses) across half-hourly records (circles). The example uses a subset data from DE-Tha.

*P 5 and elsewhere: please use 'van Gorsel'. Regarding section 2.1, will improved approaches following van Gorsel et al. (2007), sigma\_w-based approaches (Jocher et al., see earlier work by Acevedo: <https://www.sciencedirect.com/science/article/pii/S0168192308001962>), estimates of drainage (see Hayek et al., 10.1016/j.agrformet.2017.12.186), atmospheric stability approaches (Novick et al., 2004) or other non-ustar approaches be implemented? It's been my feeling that a large part of the flux community has moved beyond ustar.*

We apologize for not correctly citing the two-word surname and corrected it. Currently there are no plans to include more recent approaches. The issue tracker on the ReddyProc github site is a good place to suggest, argue for, and detail such a suggestion.

*2.1.1: I've long-wondered what is the best threshold for 'nighttime' especially given the characteristic decline in turbulence with the collapse of the daytime boundary layer may occur earlier. Is there a justification for  $R_g < 10 W m^{-2}$ ? (see also the last line of page 8).*

We selected this threshold and other similar values because we had not reason to deviate from the standard approaches, here DP06. We did not discuss this issue in the manuscript but give some thoughts here in the replies.

The night-time threshold for  $u^*$  estimation  $10 \text{ Wm}^{-2}$  as used by Papale et al 2006 was selected to our knowledge mainly to 1) guarantee a certain amount of nighttime data for the estimation of  $E_0$  and  $R_b$  (this we think is the most important reason); and 2) to avoid problems in the selection of nighttime data in sites with drift in the pyrometer (i.e. positive values at nighttime due to wrong calibrations or drift that were pretty common in the past, now we think that the quality control is stricter in this regard). Also for the point 2 the crosscheck with the potential radiation is important.

Further, the  $R_g$  measurement can be a little unclean and therefore small values below 10 are included, for example bright moon hours at night, slight offsets of the instrument, or a not fully dark sky due to some back scatter from low hanging clouds after sunset. For the partitioning, the algorithm additionally even filters for nighttime by calculating the potential radiation and only "true" nighttime values will be considered.

Layering might already set in before, i.e. at higher radiation values like  $30 \text{ W m}^{-2}$  and  $u^*$  estimation and filtering could include even more data points. We are not aware of any study on this and suspect a good threshold to be very site specific and maybe even season specific - for example thinking of hours with low sun angles in Northern sites. However, for this, REddyProc allows the PIs and users to experiment with their own  $R_g$  thresholds.

*Regarding 2.2.2, often alternate micrometeorological information is available. Is it possible in REddyProc to add multiple micrometeorological measurements of the same variable?*

We do not fully understand the request. E.g. „Using PAR when  $R_g$  is missing“ is currently not implemented. However, the user can flexible define different meteorological variables and thresholds with the Lookup-table approach. However, then it is not "the" MDS algorithm any more. There is also currently no feature to use a different temperature column aside from air temperature yet. Again, the github issue tracker would be a good place to propose and details such a feature.

*On page 10, BGC16 was used and defined before. Please define it at first instance.*

Section 3 is dedicated to describe the benchmark tools. When using acronym BGC16 before, its difficult to describe it sufficiently. Hence those usages refer to section 3 (e.g. Table 1 and p8110).

*‘Papale C-implementation’ on p. 13, just use DP06. A quick re-read for minor issues like this would improve the paper. Also, what is a ‘big scatter’ on page 13?*

We hope that a re-read captured those minor issues, as we corrected on page 13. We reworded „Big scatter“ by „although individual threshold estimates differed“

*Regarding the LRC fits, is a least squares cost function applied or least absolute deviation to avoid the influence of outliers (see Richardson 2005/6 papers). Some discussion of the cost function is necessary in my opinion.*

The fitting for least absolute deviation is appropriate if measurement error is Laplace distributed, which seems to be the case both for soil respiration and NEE data. However, when comparing fluxes at similar environmental conditions and hence, of similar magnitude, the distribution is better approximated by a Gaussian distribution, and the apparent Laplace shape is probably generated by a superposition of several Gaussian distributions, where standard deviation scales with flux magnitude (Lasslop 2008). Hence, in accordance with Lasslop 2010 we used the usual normal assumption in fitting the LRC curve.

We extended the section on fitting the LRC curve and repeated the cost equation from Lasslop et al. in supplement 1 (section 2 , step 3).

# Basic and extensible post-processing of eddy covariance flux data with REddyProc.

Thomas Wutzler<sup>1</sup>, Antje Lucas-Moffat<sup>2,3</sup>, Mirco Migliavacca<sup>1</sup>, Jürgen Knauer<sup>1</sup>, Kerstin Sickel<sup>1</sup>, Ladislav Šigut<sup>4</sup>, Olaf Menzer<sup>5</sup>, and Markus Reichstein<sup>1</sup>

<sup>1</sup>Max Planck Institute for Biogeochemistry, Hans-Knöll-Straße 10, 07745 Jena, Germany

<sup>2</sup>German Meteorological Service, Centre for Agrometeorological Research, Bundesallee 33, 38116 Braunschweig, Germany

<sup>3</sup>Also at: Thuenen Institute of Climate-Smart Agriculture, Bundesallee 65, 38116 Braunschweig, Germany

<sup>4</sup>Global Change Research Institute CAS, Bělidla 986/4a, CZ-60300 Brno, Czech Republic

<sup>5</sup>Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA

**Correspondence:** Thomas Wutzler  
(twutz@bgc-jena.mpg.de)

## Abstract.

With the eddy covariance (EC) technique, net fluxes of carbon dioxide (CO<sub>2</sub>) and other trace gases as well as water and energy fluxes can be measured at the ecosystem level. These flux measurements are a main source for understanding biosphere-atmosphere interactions and feedbacks by cross-site analysis, model-data integration, and up-scaling. The raw fluxes measured with the EC technique require an extensive and laborious data processing. While there are standard tools<sup>1</sup> available in open source environment for processing high-frequency (10 or 20 Hz) data into half-hourly quality checked fluxes, there is a need for more usable and extensible tools for the subsequent post-processing steps. We tackled this need by developing the REddyProc package in the cross-platform language R that provides standard CO<sub>2</sub>-focused post-processing routines for reading (half-)hourly data from different formats, estimating the  $u_*$  threshold, gap-filling, flux-partitioning, and visualizing the results. In addition to basic processing, the functions are extensible and allow easier integration in extended analysis than current tools. New features include cross-year processing and a better treatment of uncertainties. A comparison of REddyProc routines with other state-of-the-art tools resulted in no significant differences in monthly and annual fluxes across sites. Lower uncertainty estimates of both  $u_*$  and resulting gap-filled fluxes by 50% with the presented tool ~~was~~ were achieved by an improved treatment of seasons during the bootstrap analysis. Higher estimates of uncertainty in day-time partitioning (about twice as high) resulted from a better accounting for the uncertainty in estimates of temperature sensitivity of respiration. The provided routines can be easily installed, configured, ~~used, and integrated in further inquiry. Hence~~ and used. Hence, the eddy covariance community will benefit from using the REddyProc package, allowing easier integration of standard post-processing with extended analysis.

---

<sup>1</sup>[fluxnet.fluxdata.org/2017/10/10/toolbox-a-rolling-list-of-softwarepackages-for-flux-related-data-processing](https://fluxnet.fluxdata.org/2017/10/10/toolbox-a-rolling-list-of-softwarepackages-for-flux-related-data-processing)

## 1 Introduction

The availability of ecosystem level observations of net ecosystem exchange (NEE) of carbon dioxide ( $\text{CO}_2$ ) and other gases, latent heat (LE) and sensible heat (H) fluxes measured by the eddy covariance (EC) method (Aubinet et al., 2000) ~~allowed a boost in~~ greatly advanced ecosystem understanding at site to global scales (Baldocchi et al., 2017). The EC method provides half-hourly or hourly records of turbulent fluxes between an entire ecosystem and the atmosphere. These data are derived from high frequency measurements (10 or 20 Hz) of wind speed and direction together with measurements of air scalar characteristics such as  $\text{CO}_2$  and water vapor concentration, ~~or~~ and temperature. Methods to compute fluxes from high frequency measurements, methods for the quality check and quality assessment (QA/QC), and methods for the storage corrections have been consolidated in the last decades (Rebmann et al., 2012; Foken et al., 2012; Aubinet et al., 2012) and are available as open source software<sup>2</sup>.

Although measured continuously, the (half)hourly EC data contain gaps due to instrument malfunctioning or records which are not representative of the ecosystem because of micrometeorological conditions under which the assumptions of the EC technique are not met (details in e.g. Foken and Wichura, 1996; Foken et al., 2004; Göckede et al., 2004; Foken et al., 2012). Hence, (half-)hourly records are marked with different quality flags and need further extensive post-processing as described by Papale et al. (2006).

NEE from periods with low friction velocity ( $u_*$ ) (Aubinet et al., 2012) need to be detected and filtered out to avoid systematic biases in nighttime NEE (Papale et al., 2006). The screened flux time series with gaps need to be filled (Reichstein et al., 2005a) using the available flux data and meteorological measurements. An additional information can be obtained from NEE thanks to flux partitioning methods, that provide model estimates of gross primary production (GPP) and ecosystem respiration ( $R_{\text{eco}}$ ) (Reichstein et al., 2005a). These gross fluxes are important to understand land-atmosphere interactions.

All these post-processing steps need to be performed routinely for EC data. Hence, it is desirable to have automated and reproducible post-processing tools available that can be easily used, extended, and integrated into researcher's own workflow. For this purpose we compiled all routines for the important  $\text{CO}_2$ -focused post-processing steps in the `REddyProc` package for the free and cross-platform, ~~free~~-R language. The `REddyProc` package loads time series of quality checked and storage corrected fluxes and the basic set of meteorological variables and provides a software environment to perform  $u_*$  threshold detection and filtering, gap filling and partitioning. Furthermore, a series of other functionalities like data import routines and data visualization are provided.

The objectives of the paper are to 1) provide a reference that describes the methodology of the processing used in the `REddyProc` package, and 2) show that the obtained results do not differ systematically from results obtained with standard post-processing implemented in the FLUXNET community (based on Papale et al., 2006; Reichstein et al., 2005a; Lasslop et al., 2010; Pastorello et al., 2017).

---

<sup>2</sup> [fluxnet.fluxdata.org/2017/10/10/toolbox-a-rolling-list-of-softwarepackages-for-flux-related-data-processing](http://fluxnet.fluxdata.org/2017/10/10/toolbox-a-rolling-list-of-softwarepackages-for-flux-related-data-processing)

Table 1 explains used abbreviations. The first part of the paper (section 2) describes the post-processing methods. The second part (Section 3) presents the benchmarks the `REddyProc` implementation with standard post-processing tools. It details differences in the implementations and possible consequences in obtained results and aggregated fluxes. An appendix gives an overview of the package with general design, an example of the post-processing, and links resources that get readers started with post-processing their own data.

**Table 1.** Abbreviations used repeatedly in the paper.

Symbol	Description
EC	eddy covariance
CO <sub>2</sub>	carbon dioxide
NEE	net ecosystem exchange towards the atmosphere in $\mu\text{mol CO}_2 \text{ m}^{-2}\text{s}^{-1}$ (aggregated in $\text{gC m}^{-2}\text{yr}^{-1}$ )
GPP	gross primary productivity (same units as NEE)
R <sub>eeco</sub>	ecosystem respiration (same units as NEE)
H, LE	sensible and latent heat flux in $\text{Wm}^{-2}$
$u_*$	friction velocity in $\text{ms}^{-1}$
R <sub>g</sub>	shortwave incoming global radiation in $\text{Wm}^{-2}$
T <sub>air</sub>	air temperature in °C
T <sub>soil</sub>	soil temperature in °C
VPD	vapor pressure deficit in hPa
LUT	look-up table (section 2.2.1)
MDC	mean diurnal course (section 2.2.2)
MDS	marginal distribution sampling (section 2.2.3)
$E_0$	temperature sensitivity parameter in eq. 1
$R_{Ref}$	respiration at reference temperature parameter eq. 1
LRC	light response curve (section 2.3.2)
DP06	C-implementation of the $u_*$ threshold estimation by Dario Papale (section 3)
BGC16	2016-version of the online tool provided by the MPI-BGC, Jena, (section 3)

[2.2](#)

## 2 Methods of ~~Post-processing~~ post-processing

The post-processing relies on half-hourly or hourly measurements of NEE and ancillary meteorological data of  $u_*$ , global radiation ( $R_g$ ), air or soil temperature ( $T_{air}$ ,  $T_{soil}$ ), and vapor pressure deficit (VPD). The fluxes should be quality checked and, if applicable, storage corrected before their use in the package.

- 5 The post-processing follows a specific workflow:
1. determination and filtering of periods with low turbulent mixing ( $u_*$ -filtering),
  2. replacing missing data in the half-hourly/hourly records (gap-filling), and
  3. partitioning NEE into the gross fluxes GPP and  $R_{eco}$  (flux-partitioning).

Usage of the `REddyProc` package follows this data post-processing workflow (Fig.1). The following sections explain the steps  
10 in more detail.

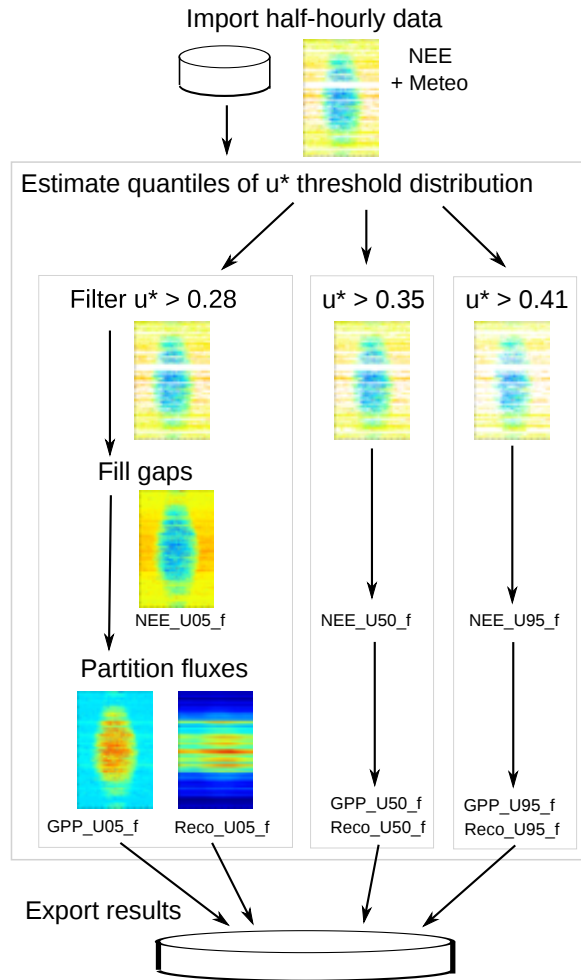
### 2.1 $u_*$ -filtering

Determining periods with low turbulent mixing is a critical step in the EC data post-processing. Standard steady state and integral turbulence characteristics tests in the initial processing exclude the most problematic records of sensible (H), latent (LE) heat and  $CO_2$  fluxes (Foken and Wichura, 1996). However, it is well known (summarized in Aubinet et al., 2012, chapter 5),  
15 that such a quality checking strategy is not sufficient, especially in the case of  $CO_2$ . Stable stratification that is present often during the night-time dampens turbulence and leads to an underestimation of the night-time NEE, i.e. the ecosystem respiration (van Gorsel et al., 2007). Massman and Lee (2002) proposed that unfavorable conditions could be detected by inspecting the relationship of night-time NEE versus  $u_*$ . Within similar time period and similar environmental conditions respiration should not be dependent on the  $u_*$ . At low  $u_*$  values, a negatively biased respiration is measured. A heuristic class of methods, which is  
20 widely accepted, assumes that a threshold of  $u_*$  can be established above that nighttime fluxes are considered valid. Hence, the  $u_*$  threshold is the minimum  $u_*$  above which respiration reaches a plateau (Fig. 2). This threshold is specific for each season of a site year. Uncertainties in the  $u_*$  threshold estimate represent one of the largest uncertainty components in the post-processing of NEE.

There are at least two methods of estimating the  $u_*$  threshold: the moving point method (Reichstein et al., 2005c; Papale et al.,  
25 2006), which is currently more routinely used, and the break-point detection method (Barr et al., 2013).

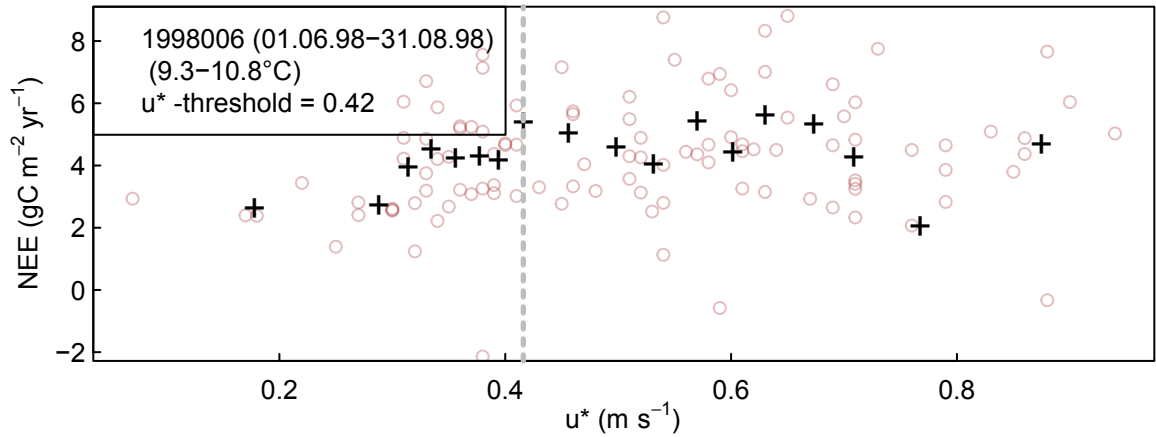
#### 2.1.1 Moving point method for $u_*$

The method of Papale et al. (2006) detects a plateau in the relationship of night-time NEE versus  $u_*$  among all records within temperature subset by a moving point test of records binned into different  $u_*$  bins.



**Figure 1.** The workflow starts with importing the half-hourly (or hourly) data, in this example the year 1998 of site DE-Tha. Next, a probability distribution of  $u_*$  threshold is estimated for each season. Gap-filling and flux-partitioning is performed for several quantiles of this distribution for an estimate of uncertainty. Finally the results are exported.

The nighttime data (default:  $R_g < 10 \text{ Wm}^{-2}$ ) is split into different times of year, here called seasons, to account for differing surface roughness. Then the data of each season is split into default six temperature subsets of equal size (according to quantiles). Within each temperature subset data is split into 20 about equally sized  $u_*$  bins. The default moving point method, called *Forward2*, determines the threshold based on these  $u_*$  bins. It checks for each bin if the mean NEE is higher than 0.95 times the mean of the following 10 bins. If this holds true also for the next bin, the mean  $u_*$  of the bin is reported as threshold. There are often subsets of data, where no clear threshold can be detected. Hence, there are quality criteria on whether the estimate of a given subset is used in subsequent aggregation. One quality criterion specifies that temperature and  $u_*$  should not be correlated within the temperature subset, another requires a minimum number of valid records within a subset. Next, the  $u_*$  estimates for



**Figure 2.** Concept of the  $u_*$ -filter for a subset of data from DE-Tha.: Night-time NEE at low  $u_*$  friction velocities, i.e. with low turbulence, is biased towards lower NEE values compared to cases with higher  $u_*$ , although other. Unbiased NEE should scatter around the same plateau because environmental conditions are similar. The  $u_*$  threshold (dashed line), i.e. the value below which this effect bias is considered significant, i.e.  $u_*$  threshold (dashed line), is estimated by a moving point method on  $u_*$  bins (crosses) across half-hourly records (circles). The example uses a subset data from DE-Tha.

different temperature classes and seasons (details in section 3.2.1) are aggregated to derive a robust  $u_*$  estimate. Within one season, the median is taken across the estimates of different temperature subsets. Within one year, the maximum is taken across the associated seasons.

Records during night-time with  $u_*$  smaller than the estimated threshold are flagged as invalid and are replaced in the subsequent gap-filling processing step. In addition, each half-hour after records with  $u_*$  smaller than the threshold is flagged to be invalid.

### 2.1.2 Breakpoint detection method for $u_*$

Alternatively, a breakpoint-detection can be applied to the unbinned data, that avoids the sensitivity of the Moving point method to the specifics of the binning schemes (Barr et al., 2013). REdDyProc provides this method by estimating the breakpoint based on unbinned records within the seasons/temperature subsets using the segmented R-package (Mugge, 2003, 2008). However, REdDyProc differs from Barr et al. (2013) by keeping the same aggregating scheme of seasonal/temperature estimates to annual thresholds as with the Moving point method.



### 2.1.3 Bootstrapping uncertainty of the $u_*$ threshold

Estimates of the  $u_*$  threshold are often sensitive to the specifics of the combination of methods and the data, e.g. the binning, minimum number of records within a season or temperature subset, criteria in aggregation, etc. Therefore, bootstrap (re-sampling with replacement) is applied to generate 200 artificial replicates of the dataset and for each replicate the threshold is estimated (Efron and Tibshirani, 1986; Davison and Hinkley, 1997). The 5th, 50th, and 95th percentile of the estimates are reported as a range of threshold estimates. The subsequent post-processing steps of gap-filling and partitioning are then repeated using those different thresholds to propagate the uncertainty of  $u_*$  threshold estimation to derived quantities such as annual NEE, GPP and  $R_{eco}$

## 2.2 Gap-filling methods

After quality checks and  $u_*$ -filtering, the dataset of half-hourly NEE fluxes may contain up to 50 percent gaps (sometimes this fraction is even higher, depending on the site conditions). For the benchmark datasets used in this manuscript, the percentage of gaps before  $u_*$ -filtering was on average 32% and after  $u_*$ -filtering 60% and 48% for upper and lower  $u_*$  threshold estimate, respectively. Filling of gaps in half-hourly NEE data is necessary to obtain complete time series for the calculation of daily averages or balances such as monthly or seasonal sums. The following three gap-filling methods are implemented in REddyProc .

### 2.2.1 Look-up tables

In the look-up table (LUT) approach, the fluxes are binned by the meteorological conditions within a certain time window. Within the chosen time window and respective bin, each meteorological variable deviates less than a fixed margin to ensure similar meteorological conditions. The missing value of the flux is then calculated as the average value of the binned records and its uncertainty estimated from their standard deviation.

The original LUT of Falge et al. (2001) consisted of fixed periods over a year, while in REddyProc the meteorological conditions are sampled with a moving window around the gap to be filled.

### 2.2.2 Mean diurnal course

NEE fluxes have a mean diurnal course (MDC) that follows the course of the sun with only respiration during nighttime and a combination of respiration and photosynthesis during daytime. This autocorrelation of the fluxes is exploited by taking the average value at the same time of day within a moving time window of adjacent days (Falge et al., 2001). In REddyProc , the same time of day includes also the fluxes of the adjacent hour ( $\pm 1$  hour).

Though the MDC method only showed a medium performance in the gap filling comparison for net carbon fluxes by (Moffat et al., 2007), it has the advantage that this approach can be used even if no meteorological information is available.

### 2.2.3 Marginal distribution sampling

5 The so-called marginal distribution sampling (MDS) by (Reichstein et al., 2005b) exploits the covariation of the fluxes with the meteorological variables and their temporal autocorrelation based on the two methods LUT and MDC described above.

The filling of each half-hourly NEE with the MDS algorithm depends on the availability of the meteorological data of  $R_g$ , Tair, and VPD: 1) If all three meteorological variables are available, LUT will be used with default margins of  $50 \text{ W m}^{-2}$ ,  $2.5 \text{ }^\circ\text{C}$ , and  $5.0 \text{ hPa}$ , respectively. 2) If Tair or VPD are missing, only the variable  $R_g$  will be used. 3) If no meteorology is available, the gaps are filled with MDC. Following a specific sampling procedure, the MDS algorithm increases the number of days in  
10 the vicinity of the gap until there are enough data points (at least two) for gap-filling. A more detailed description with a flow diagram is provided in Supplement 1.

The MDS algorithm is optimized for carbon dioxide and water fluxes and can also be used to estimate the uncertainty of the half-hourly fluxes. In the comparison of gap-filling methods by (Moffat et al., 2007) [Moffat et al. \(2007\)](#), the MDS algorithm performed well for different artificial gap scenarios ranging from single half-hours to several days. Due to its flexibility in  
15 dealing with missing meteorological input data and its fast and highly automated routines available as online tool (BGC16, section 3), the MDS gap-filling method has been widely used.

### 2.3 Flux-partitioning methods

The gross fluxes of GPP into the land system and  $R_{\text{eco}}$  out of the land system are the two opposing parts of NEE:  $\text{NEE} = R_{\text{eco}} - \text{GPP}$ . Availability of GPP and  $R_{\text{eco}}$  is pivotal as they are the two biggest terms of the carbon cycle (e.g. Chapin et al., 2006; Jung  
20 et al., 2011). Moreover, understanding their sensitivity to environmental drivers (e.g., radiation, temperature, and soil moisture) is important to interpret land-atmosphere interactions and to improve earth systems models (Reichstein et al., 2012). Therefore several methods were developed to partition NEE into these two components (Reichstein et al., 2005c; Lasslop et al., 2010; Moffat, 2012; Wehr and Saleska, 2015; Desai et al., 2008; Stoy et al., 2006).

The two most widely used methods are the so-called night-time partitioning and day-time partitioning (Reichstein et al., 2012).  
25 The night-time partitioning (Reichstein et al., 2005c) relies on the temperature response function of nighttime NEE fluxes that are representative of  $R_{\text{eco}}$ . It assumes that this relationship is applicable also to daytime data. The relationship is then used to predict  $R_{\text{eco}}$  from measured temperature and GPP is computed as a difference between  $R_{\text{eco}}$  and NEE. This method is currently the most widely used approach. Alternatively, the day-time partitioning (Lasslop et al., 2010) fits a model to observations of daytime NEE and global radiation, accounting for the effects of radiation and VPD on GPP as well as the effects of temperature  
30 on  $R_{\text{eco}}$ .

### 2.3.1 Nighttime flux-partitioning

The method of Reichstein et al. (2005c) estimates a temporally varying respiration-temperature relationship from night-time data. First night-time data is selected by a threshold of  $R_g < 10 \text{ Wm}^{-2}$ , which is congruent with the BGC online tool (BGC16, section 3), but differs from the  $20 \text{ Wm}^{-2}$  reported in Reichstein et al. (2005c). [Additionally, night-time data is constrained](#)  
5 [between computed sunset and sunrise.](#)

Next, temperature sensitivity,  $E_0$  of the Lloyd and Taylor (1994) relationship (1) is estimated by fitting the model to successive 15 day periods of night-time data, and the resulting  $E_0$  series is aggregated to an annual estimate.

$$R_{eco}(T) = R_{Ref} \exp \left[ E_0 \left( \frac{1}{T_{Ref} - T_0} - \frac{1}{T - T_0} \right) \right], \quad (1)$$

where  $T_0$  is kept constant at  $-46.02^\circ\text{C}$  (Lloyd and Taylor, 1994) and where reference temperature  $T_{Ref}$  is  $15^\circ\text{C}$ , which is  
10 congruent with the BGC online tool (BGC16, section 3) but differs from the  $10^\circ\text{C}$  reported in (Reichstein et al., 2005c). For robustness each fit is repeated on a trimmed data set excluding records with residuals outside the 5%-95% residual distribution. The annual aggregate is the mean across the 3 valid estimates having the lowest uncertainty in the fit. Single estimates of  $E_0$  are considered valid, if there were minimum of 6 records, temperature ranged across at least  $5^\circ\text{C}$ , and estimates were inside the range of 30 to 450 K.

15 Subsequently, the respiration at reference temperature,  $R_{Ref}$ , is re-estimated from night-time data using the annual  $E_0$  temperature sensitivity estimate for 7-day windows shifted consecutively for 4 days. The estimated value is then assigned to the central time-point of the 4 day period and linearly interpolated between periods. Hence, the obtained respiration-temperature relationship varies across time.

Finally,  $R_{eco}$  is estimated for both day- and night-time from the temporarily varying  $R_{eco}$  temperature relationship and daytime  
20 GPP is computed as  $R_{eco} - \text{NEE}$ .

### 2.3.2 Daytime flux-partitioning

The method of (Lasslop et al., 2010) models NEE using the common rectangular hyperbolic light-response curve (LRC) (Falge et al., 2001):

$$NEE = \frac{\alpha\beta R_g}{\alpha R_g + \beta} + \gamma, \quad (2)$$

25 where  $\alpha$  ( $\mu\text{mol CO}_2 \text{ J}^{-1}$ ) is the canopy light utilization efficiency and represents the initial slope of the light-response curve,  $\beta$  ( $\mu\text{mol CO}_2 \text{ m}^{-2}\text{s}^{-1}$ ) is the maximum  $\text{CO}_2$  uptake rate of the canopy at infinite  $R_g$ , and  $\gamma$  ( $\mu\text{mol CO}_2 \text{ m}^{-2}\text{s}^{-1}$ ) is a term

accounting for ecosystem respiration. The hyperbolic light-response curve is modified to account for the temperature dependency of respiration after Gilmanov et al. (2003) by setting respiration  $\gamma$  to the Lloyd and Taylor respiration model (Lloyd and Taylor, 1994) (1). Further, the constant parameter  $\beta$  in (2) is replaced by an exponential decreasing function (Körner, 1995) at higher VPD values (3).

$$5 \quad \beta = \begin{cases} \beta_0 \exp[-k(\text{VPD} - \text{VPD}_0)] & \text{if VPD} > 10 \text{ hPa} \\ \beta_0 & \text{otherwise} \end{cases}, \quad (3)$$

where the  $\text{VPD}_0$  threshold is 10 hPa in accordance with earlier findings at the leaf level (Körner, 1995) ignoring potential vegetation specific differences.

Parameter  $T_0$  in (1) was fixed as in the night-time partitioning (section 2.3.1). Parameter  $T_{Ref}$  was fixed in each window to the median temperature within the window. The other parameters ( $E_0, R_{Ref}, \alpha, \beta_0, k$ ) of the model are estimated by the following  
10 steps. 1) A time varying temperature sensitivity  $E_0$  is estimated from night-time data for a window shifted by two days. 2) The  $E_0$  estimates are smoothed across successive windows by fitting a Gaussian Process (Rasmussen and Williams, 2006; Menzer et al., 2013) using the `mlegp` R-package that also estimates uncertainty of the smoothed  $E_0$ . Next, a prior respiration,  $R_{Ref}$ , for reference temperature  $T_{Ref} = 15$  °C is re-estimated from night-time data for each window with smoothed  $E_0$ . 3) Parameters of the rectangular hyperbolic light-response curve ( $R_{Ref}, \alpha, \beta_0, k$ ) are fitted using only day-time data and the previously  
15 determined temperature sensitivity ( $E_0$ ) for each window. 4) Finally, for each NEE record, GPP and  $R_{eco}$  are estimated with the parameter set of the previous valid window and the parameters of the next valid window, and the two results are interpolated linearly by the time difference to the window centers. Supplement 1 reports necessary technical details to these steps.

Note, that contrary to the night-time based flux-partitioning, both GPP and  $R_{eco}$  are model predictions and do not add up exactly to observed NEE.

### 20 3 Benchmarking REddyProc post-processing steps

The post-processing steps implementations of REddyProc were benchmarked with the post-processing tools widely used in the FLUXNET processing. Specifically, REddyProc (version 0.8.1)  $u_*$ -filtering results were compared with results by a C-implementation from Dario Papale (Papale et al., 2006), here referred to as DP06. Results of REddyProc (version 1.1.2) gap-filling and flux-partitioning were compared with results obtained by the 2016-version web based tool provided by the  
25 Max Planck Institute for Biogeochemistry, Jena, best described in Reichstein et al. (2005a). The tool was accessed in 2016 (29.07.2016) and is hereafter referred as BGC16. Annually and monthly aggregated values, here, refer to the mean across all valid values in a month or a year, which can differ from real annual or monthly budgets in the presence of large gaps. The first section describes the dataset used for benchmarking for each processing step implemented in the package. Within each of the

following sections for the processing steps, subsections describe differences in the code, report the results of benchmarking, and discuss them. Supplement 1, additionally, provides more detailed results and statistics.

### 3.1 Dataset used for benchmarking

Data of twenty-five sites with open data policy of the LaThuile FLUXNET dataset<sup>3</sup> were used for benchmarking. The sites are located in different climate zones and belong to a variety of plant functional types (Table 2) to guarantee testing different conditions (i.e. presence of snow, management such as cuts and crop rotation, sites disturbed) and ecosystem types (e.g. deciduous vs evergreen forests, grasslands and croplands). For each site the following variables were used: NEE already filtered for quality flags (Foken and Wichura, 1996), despiked and  $u_*$  filtered (Papale et al., 2006), random error of NEE computed as described by Reichstein et al. (2005a), Tair, Tsoil, Rg, and VPD. Moreover, NEE time series before the  $u_*$ -filtering and the  $u_*$  data were downloaded from AMERIFLUX and the European Flux Database to test the  $u_*$  threshold estimation. Finally, time series of gap-filled NEE ( $NEE_f$ ), GPP partitioned with the night-time based method ( $GPP_{NT}$ ) (Reichstein et al., 2005a) were downloaded from the LaThuile dataset, while GPP partitioned with the daytime method ( $GPP_{DT}$ ) were computed with BGC16.

### 3.2 $u_*$ -filtering: Benchmarking with DP06

Estimation of  $u_*$  threshold by `REddyProc` using the default moving point method (section 2.1.1) was benchmarked to estimation based on Papale's DP06 C-implementation (Papale et al., 2006). The benchmark applied a bootstrap sample of size 60 and recorded lower, median, and upper quantiles of 10%, 50% and 90% instead of the default 5% and 95% based on a larger sample size to save computing time.

The different estimates of the  $u_*$  threshold have potential consequences for the inferred fluxes. To explore these consequences, we used the different [resulting](#) thresholds to mark gaps, gap-fill the data, and compute the annual NEE based on the gap-filled time series. NEE uncertainty was estimated by the difference between NEE based on lower quantile  $u_*$  and NEE based on upper quantile  $u_*$  estimate.

#### 3.2.1 Differences in code

The biggest difference of `REddyProc` compared to DP06 is that `REddyProc` by default employs seasons that can span across years. With the '*Within one year*' classification option, which is employed also by DP06, records of December are associated to the same season as January and February of the *same* year. With the default '*continuous*' classification, seasons start the same as in DP06 by default in March, June, September, and December. However, December is treated in the same season as January and February of the *next* year to avoid discontinuities at year boundaries. The annual  $u_*$  threshold is then applied according to those continuous seasons spanning year boundaries. For example, the processing of 2014 data would by

---

<sup>3</sup>www.fluxdata.org

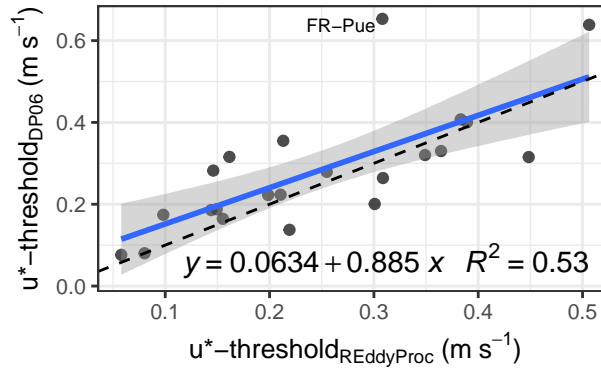
**Table 2.** Description of sites and times used for benchmarking `REddyProc`.

<sup>a</sup> Abbreviations for land cover type from International Geosphere-Biosphere Programme (IGBP) classification: CRO: cropland, DBF: deciduous broadleaf forest, EBF: evergreen broadleaf forest, ENF: evergreen needleleaf forest, GRA: grassland, OSH: open shrubland, WET: permanent wetland, WSA: woody savanna.

<sup>b</sup> Abbreviations for climate from Köppen-Geiger classification: Af: equatorial, rainforest; BSh: hot arid steppe; Cfa: humid, warm temperate, hot summer; Cfb: humid, warm temperate, warm summer; Csa: summer dry, warm temperate, hot summer; Dfb: cold, humid, warm summer; Dfc: cold, humid, cold summer.

Site	Year	Lat	Lon	Land cover <sup>a</sup>	Climate <sup>b</sup>
CA-NS7	2004	56.64	-99.95	OSH	Dfc
CA-TP3	2005	42.71	-80.35	ENF	Dfb
CH-Oe2	2004	47.29	7.73	CRO	Cfb
DE-Hai	2002	51.08	10.45	DBF	Cfb
DE-Tha	1998	50.96	13.57	ENF	Cfb
DK-Sor	2006	55.49	11.65	DBF	Cfb
ES-ES1	2000	39.35	-0.32	ENF	Csa
ES-VDA	2005	42.15	1.45	GRA	Cfb
FI-Hyy	1998	61.85	24.29	ENF	Dfc
FI-Kaa	2001	69.14	27.30	WET	Dfc
FR-Gri	2006	48.84	1.95	CRO	Cfb
FR-Hes	1998	48.67	7.06	DBF	Cfb
FR-Lq1	2006	45.64	2.74	GRA	Cfb
FR-Lq2	2006	45.64	2.74	GRA	Cfb
FR-Pue	2003	43.74	3.60	EBF	Csa
IE-Dri	2004	51.99	-8.75	GRA	Cfb
IL-Yat	2005	31.34	35.05	ENF	BSh
IT-Amp	2004	41.90	13.61	GRA	Cfa
IT-MBo	2005	46.02	11.05	GRA	Cfb
IT-SRo	2001	43.73	10.28	ENF	Csa
PT-Esp	2004	38.64	-8.60	EBF	Csa
RU-Cok	2004	70.62	147.88	OSH	Dfc
SE-Nor	1997	60.09	17.48	ENF	Dfb
US-Ton	2004	38.43	-120.97	WSA	Csa
VU-Coc	2002	-15.44	167.19	EBF	Af

default use data from Winter 2014 (starting in December 2013) to Autumn 2014 (ending in November 2014). `REddyProc` also allows more flexibility with the '*User specified*' classification into seasons as explained below.



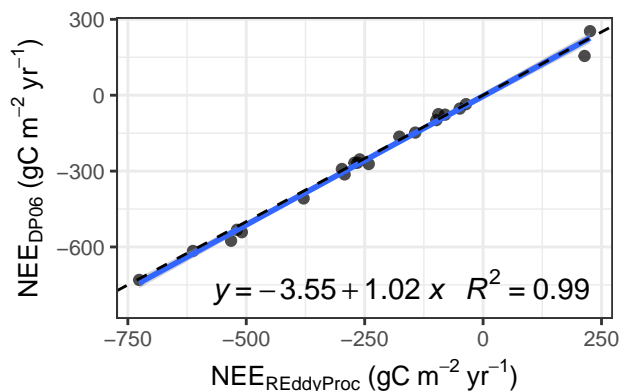
**Figure 3.** ~~Large scatter~~ Deviations between single sites but retained relationship between  $u_*$  threshold derived with DP06 versus  $u_*$  threshold derived with REddyProc across site-years as shown by a regression (solid line with shaded uncertainty bound) close to the 1:1 line (dashed).

There are further slight differences between REddyProc and DP06. Both methods bin in a way such that the number of records in each bin is similar. If there are numerically equal  $u_*$  values, they are sorted into the same bin, resulting in bins with unequal record numbers. In DP06 sometimes no records are sorted into the subsequent bins hampering the moving point detection. Contrary, the binning with REddyProc ensures that there are a minimum number of records in all bins. This often results in fewer bins ~~than without numerically equal  $u_*$  values~~. Moreover, differing from ~~the Papale C implementation~~ DP06, REddyProc employs several more quality criteria. First, when comparing the threshold bin to NEE in the following bins, it makes sure that there are least 3 bins to infer a plateau in NEE. Next, when aggregating the thresholds of different temperature classes to season, it ensures that a threshold was found in at least 20% of the temperature classes. For those seasons where no threshold could be determined, the annual estimate is used. When there are too few records within one year, a single season comprising all records is used for threshold estimation.

In difference to DP06, REddyProc re-samples data only within seasons instead across the entire year during the bootstrap, in order to protect periods of similar  $u_*$ -NEE relationship and to avoid seasonal biases in re-sampling.

### 3.2.2 Benchmark results

The general relationship in the estimation of the  $u_*$  threshold was retained between the two methods (Fig. 3), although ~~there was a big scatter~~ individual threshold estimates differed. The exceptionally high threshold value of  $> 0.6 \text{ m s}^{-1}$  for site FR-Pue was very probably an overestimate by DP06. However, one has to remember that each estimate has a high uncertainty, and the differences between the two methods were in the range of this uncertainty (suppl 1) The estimate of the uncertainty of the  $u_*$  thresholds with REddyProc was, however, only half of the uncertainty range estimated by DP06 (suppl 1). This increased precision was mainly due to the modified bootstrapping scheme, that respects the  $u_*$  seasons.



**Figure 4.** Strong correspondence in NEE based on  $u_*$  threshold estimated by REddyProc and NEE based on  $u_*$  threshold estimated by DP06 across site-years.

When propagating the differences in  $u_*$  to differences in annual NEE, there was no bias and ~~very-low~~ decreased scatter across sites between all the methods (Fig. 4), despite the differences in  $u_*$  threshold. The absolute differences in annual NEE between the methods were small (mostly  $< 20 gC m^{-2} yr^{-1}$ ), and mostly lower than half of the uncertainty range estimated from the bootstrap (suppl 1) REddyProc estimates  $u_*$  thresholds with roughly double the precision compared to DP06, due to  
 5 protecting seasons during bootstrap (suppl 1).

### 3.2.3 Discussion of $u_*$ threshold estimation

The agreement between NEE based on  $u_*$  estimates of REddyProc moving point implementation and current FLUXNET standard post-processing (DP06) (Fig. 4) indicates that the sensitivity of NEE to the  $u_*$  threshold estimate in the inferred ranges is low, which also explains the large uncertainty of the  $u_*$  threshold estimate. One reason for the missing effect could be site  
 10 selection of this study without many sites affected by advection, which show limited saturation of the NEE -  $u_*$  dependence. Since in such cases the filtering does not work properly anyway, it should not change the conclusions for NEE. Hence, we infer that  $u_*$  estimates of both DP06 and REddyProc are appropriate due to the negligible effect on NEE sums. The agreement implies that both methods can be interchanged in studies that are based on aggregated values, such as annual carbon budgets or for upscaling, without the need to reprocess data.

15 However, the increase in estimated precision, i.e. lower standard deviation, of the  $u_*$  threshold estimate also yields an increase in estimated precision of the annual NEE by 50% (suppl 1). This will lead to improved accuracy and usability of EC measurements and any downstream, post-processed data products in model-data integration studies.

While the default seasons and their aggregation are in line with previous approaches, REddyProc allows to use site specific knowledge to derive better threshold estimates. For example, if there is disturbance such as harvest, the  $u_*$  threshold is expected



to change and a different threshold should be applied for filtering before and after the disturbance. In this case the user can define a season change at harvest date and use season-specific threshold estimates instead of the annually aggregated estimate (section B7).

### 3.3 Gap-filling: Benchmark with BGC16

- 5 The gap-filling implementation of REddyProc was benchmarked with the BGC online tool (BGC16, section 3) that used pvWave-code from Reichstein et al. (2005c).

#### 3.3.1 Differences in code

Compared to the BGC16, the new implementation of the MDS algorithm in REddyProc was not limited to single years but filled the gaps with a window moving continuously over all years in the input data. This had the advantage of smoother gap-filling over the end of year **timestamp** and will especially be of interest for sites in which vegetation is not dormant during this time. This new feature led to different, probably more realistic gap-filled NEE values in the beginning and end of the year. There were also slight differences in the ~~window size between the old and new code~~ sequence of window sizes between REddyProc and BGC16. For MDC, ~~the old window size~~ window size with BGC16 had a few more intermediate day steps than ~~the new implementation~~ REddyProc, which affected longer gaps with missing meteorology. The default meteorological variables and margins for LUT (see Chapter 4.2.2 above) were the same in both implementations.

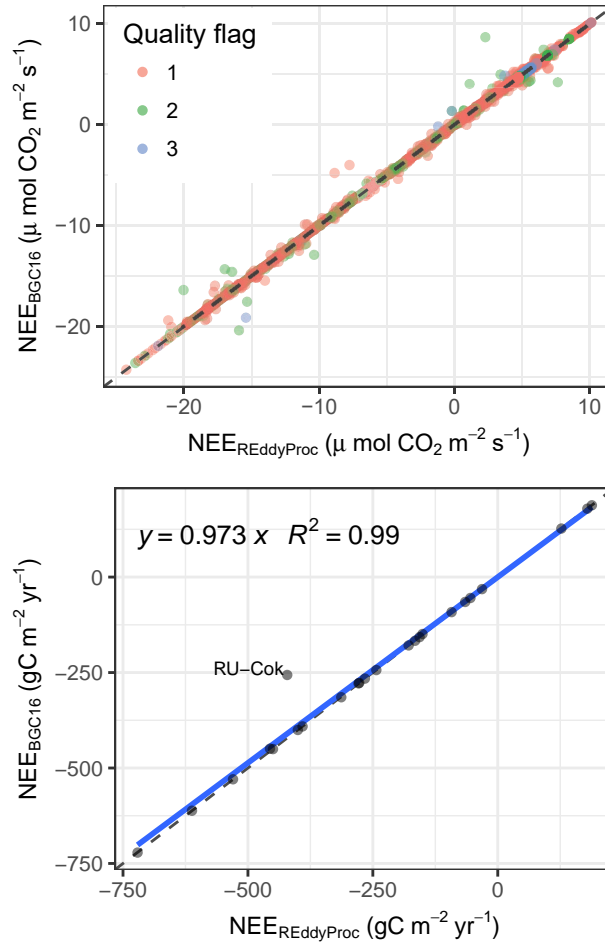
While REddyProc restricts gap-filling to interpolation of gaps, BGC16 also extrapolated missing records in periods without measurements.

#### 3.3.2 Benchmark results and discussion for gap-filling

In the benchmark, REddyProc gap-filling was run using the same measured NEE as input that passed the QA/QC routines and  $u_{*}$ -filtering. The annually aggregated values comprised both, filled gaps and originally valid records.

REddyProc gap-filling results agreed with the results of BGC16. Few discrepancies at half-hourly time scale were mostly at longer gaps due to usage of fewer window sizes, as shown for the DE-Tha case (Fig. 5 top). At annually aggregated time scale, the agreement between methods was strong ( $R^2 = 0.99$ ) (Fig. 5 bottom). The outlier of site RU-Cok is due to the availability of only a few months of data for the whole year. While REddyProc filled gaps in the time period with data available, BGC16 extrapolated also into the time before and after. The seasonal cycle was well reproduced at each site (suppl 1).

The good agreement between NEE based on gap-filling by REddyProc and gap-filling by BGC16 (Fig. 5) imply that both gap-filling tools can be used interchangeably without the need to reprocess data.



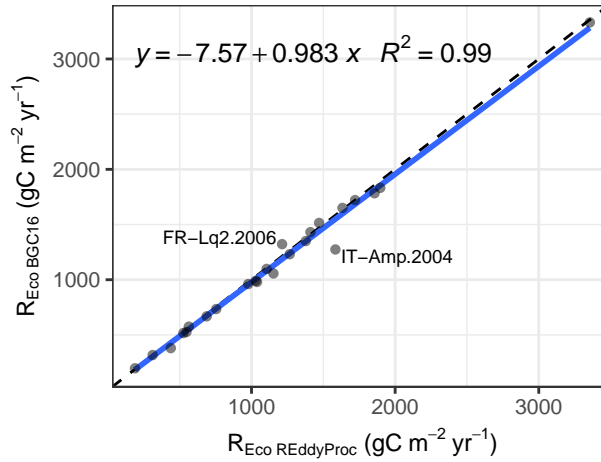
**Figure 5.** Predictions of NEE by `REddyProc` after gap-filling agree with BGC16 both at half-hourly values (top), shown for the DE-Tha 1998 use case, and annual means across sites (bottom). Larger quality flags are associated with larger window sizes.

### 3.4 Nighttime flux-partitioning: Benchmark BGC16

The night-time based flux-partitioning was benchmarked to [the BGC online tool \(BGC16, section 3\)](#), that [which](#) used `pvWave` code developed by Reichstein et al. (2005c).

#### 3.4.1 Differences in code

- The main features of the `REddyProc` implementation of the night-time based partitioning algorithm were very similar to BGC16, with using a reference temperature of 15°C and trimming the estimates of temperature sensitivity  $E_0$  before aggregating them (section 2.3.1). `REddyProc` differed from BGC16 in computing the potential radiation that is used in subsetting the



**Figure 6.** Predictions of annually aggregated ecosystem respiration,  $R_{eco}$  from `REddyProc` night-time partitioning agree with the predictions by BGC16.

night-time data to derive  $E_0$  and  $R_{Ref}$  (Reichstein et al., 2005). While `REddyProc` used the exact solar time for the calculation of the potential radiation, where the sun culminates exactly at noon, BGC16 used the local winter time which differs from the solar time depending on the location within the time zone.

### 3.4.2 Benchmark results and discussion for night-time flux-partitioning

- 5 Annual aggregated values of  $R_{eco}$  predicted by `REddyProc` were in very good agreement ( $R^2 = 0.99$ ; slope  $\approx 1$ ) with BGC16 as shown in Fig. 6 and in supplement 1.

In order to evaluate the effects of the differences introduced in the code described above, we also computed  $R_{eco}$  by prescribing in `REddyProc` either  $E_0$ , selection of night-time data, or both from BGC16 output. Results are reported in supplement 1 and showed that the most important factor affecting the  $R_{eco}$  computed with `REddyProc` was the different selection of night-time

- 10 data, though the differences were almost negligible at annual time scale.

The two implementations agree very well for most sites at annual time-scale. Because of no systematic deviations across sites, the spatial upscaling of fluxes should not be affected by `REddyProc` implementation. However, for some sites, such as IT-Amp, the quite large relative errors indicates problems related to selection of night-time data and problems due to large gaps in the dataset.

### 3.5 Daytime flux-partitioning: Benchmark with BGC16

The day-time flux-partitioning was benchmarked with results of the BGC online tool (BGC16, section 3), which is based on pvWave code developed by Lasslop et al. (2010) and used in the processing of the 2015 Fluxnet release (Pastorello et al., 2017).

#### 3.5.1 Differences in code

- 5 BGC16 differed from `REddyProc` (section 2.3.2) mainly in aspects of separation of nighttime data, estimation of temperature sensitivity from night-time data, uncertainty estimation, treatment of missing values, and optimization library code.

While for separating nighttime data `REddyProc` used exact solar time where sun culminates exactly at noon, BGC16 used the local winter time zone time.

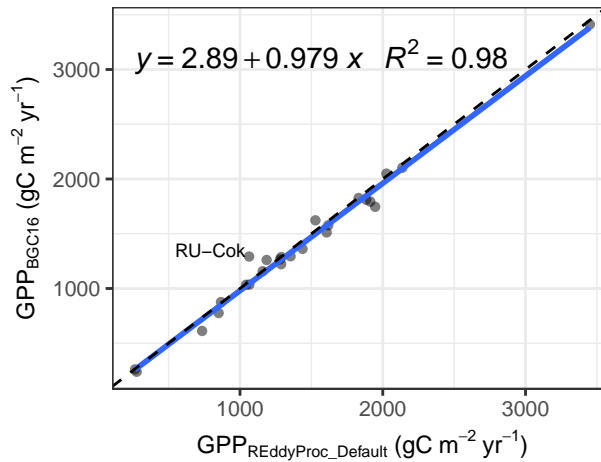
- For the estimation of temperature sensitivity  $E_0$  from night-time data BGC16 used a reference temperature of 15 °C, instead of  
10 the median temperature inside the window. Hence, it estimated stronger correlations between parameters for windows with a different temperature range. Moreover, it omitted smoothing of the estimated  $E_0$  across time, often leading to large fluctuations of the  $E_0$  estimates across few days (suppl 1), larger estimates of its uncertainty, and differences in subsequent estimation of LRC parameters.

- For uncertainty estimation, BGC16 relied on the curvature of the LRC fit's optimum instead of a bootstrap procedure. Hence, it  
15 could not take into account the uncertainty of  $E_0$  estimated from night-time data before the day-time LRC fit. Moreover, during interpolation of fluxes based on previous and next valid estimates, the distance weights differed. While `REddyProc` assigns the estimates to the time of the mean of valid record in a window, BGC16 assigned it to the start of the third day, also if there were only valid data for the first day in the window.

- For weighting the records in the LRC fit, BGC16 used the raw estimated NEE uncertainty of each record. It did not check  
20 for high leverage of spurious low NEE uncertainty estimates. Its estimates, therefore, were in some windows very strongly influenced by a few records, and failed if a NEE uncertainty estimate of zero was provided. Moreover, when there were missing values or values below zero in provided NEE uncertainty, it set all uncertainty to 1, while `REddyProc` fills the gaps by setting the missing uncertainty to the maximum of 20% of respective NEE but at least  $0.7 \mu\text{molCO}_2\text{m}^{-2}\text{s}^{-1}$ .

- Treatment of missing values was not considered by BGC16 and assumed to be handled prior to the processing. Hence, it did not  
25 handle missing VPD values and did not re-try the LRC fit without the VPD-effect in order to use also records with missing VPD. Moreover, as described above, when there were missing values of NEE uncertainty, weighting records in the LRC fit was omitted.

- For compatibility with BGC16, the above code-differences can be disabled in `REddyProc`. But differences in optimization  
library code and specifically the conditions of non-convergence on scattered data could not be eliminated, which led to differences  
30 in results as shown in the following section.



**Figure 7.** Prediction of annually aggregated GPP from REddyProc day-time partitioning agree with BGC16 across sites.

### 3.5.2 Benchmark results for day-time partitioning

Annually GPP predictions of both implementations showed no significant bias across the test sites (Fig. 7), although, there was some scatter ~~for the among~~ individual predictions. A similar scatter was observed when comparing the predictions of the default REddyProc options to the predictions with compatibility options. Most of the differences were caused by decreasing the unreasonable high influence of NEE records with small NEE uncertainty (suppl 1).

The largest differences in aggregated fluxes between implementations were due to the extrapolation of fitted parameters to periods where no parameter fits were obtained. In many of those cases, there were fits at the boundaries of these periods, whose validity was questionable. Whether those fits passed the quality check or not had a large influence on the extrapolation and hence on the aggregated values. For example, at RU-Cok parameter estimates for valid periods agreed between implementations. However, no valid parameters could be obtained for winter months. While REddyProc reported missing values, BGC16 reported GPP values based on summer parameterizations also for periods further away from summer, which in turn lead to higher annual GPP estimates.

Uncertainty estimates of gross fluxes ~~were larger about doubled~~ with REddyProc due to accounting for uncertainty in temperature sensitivity estimates from night-time data (suppl 1).

### 3.5.3 Discussion of day-time flux-partitioning

Agreement between aggregated fluxes predicted by the day-time method and absence of bias for the test sites (Fig. 7) suggest that the methods can be used interchangeably for upscaling, although differences in results of influential sites can potentially propagate to differences in upscaled estimates. REddyProc provides a quality flag for the results of the day-time partitioning,

that allows excluding less reliable data in upscaling studies. For the results associated with good quality flags, we set stronger trust in the `REddyProc`-based estimates.

The day-time flux-partitioning is quite sensitive to the details of LRC fit. Small changes in treatment of extreme or missing NEE uncertainty estimates or changes in pre-processing and treatment of missing values cause different estimates of LRC parameters and propagate to predicted fluxes of GPP and  $R_{\text{eco}}$ . Although, we put much effort in trying to reproduce the results of BGC16, we were not able to eliminate all differences, especially in the subtle details in the parameter optimization library codes. The differences in predicted half-hourly fluxes, however, average out across sites and across time (suppl. 1) making this issue less severe at larger scales.

The estimated uncertainties are even more sensitive. Both implementations occasionally produce unreasonably high outliers that affect the aggregated values. `REddyProc`, in general, estimates higher uncertainties of predicted fluxes, because it accounts for uncertainty in temperature sensitivity. Note, that the uncertainty introduced to annually aggregated fluxes due to flux-partitioning is smaller than uncertainty due to uncertain  $u_*$  threshold estimate. Hence, differences or difficulties in uncertainty estimation caused by flux-partitioning do affect conclusions of the overall uncertainty estimates to a lesser extent.

#### 4 Conclusions

The `REddyProc` software provides a set of tools for the  $\text{CO}_2$ -focussed post-processing of eddy covariance flux data including  $u_*$ -filtering, gap-filling and flux-partitioning and propagation of the uncertainty from the  $u_*$ -filtering to the gap-filled NEE and partitioned GPP and  $R_{\text{eco}}$ .

The freely available R-package enables researchers to integrate the flux data processing into their own, offline environment or work stream without the need of uploading data. This seamless integration allows to improve overall workflow, to speed up processing routines and to ultimately generate cleaner, reproducible scientific results.

The compatibility of the implemented methods to the available standard tools provides continuity of the data analysis, when adopting `REddyProc` for processing EC-data. `REddyProc` can closely reproduce results of the widely used BGC online tool (BGC16, section 3).

A number of enhancements provide more flexibility to the user in the processing of their data. For instance, the new processing allows to treat multi-year data without breaks at annual boundaries that can significantly affect sites in the southern hemisphere or sites characterized by vegetation activity in winter. Another new feature of `REddyProc` is the flexibility to define different seasons for the application of the  $u_*$ -filtering and gap-filling routines, which is important for sites with discontinuous surface cover associated with snow melt, dry seasons, or harvest.

Sensitivity of the results to subtle details of the implementation, however, call for caution when interpreting results. This is especially true for  $u_*$  threshold estimation and the day-time flux-partitioning, and especially for data with long gaps.

Continued integration of new methodological developments into the package will support research using EC data. We strive to provide new developments in a basic and extensible manner, while paying attention to compatibility with results of reference implementations.

In summary, research using (half-)hourly eddy covariance data can benefit from building blocks for standardized and extensible post-processing provided by `REddyProc`.

## Appendix A: The `REddyProc` package

The `REddyProc` processing tool is freely available in two options: a) online as a web-service<sup>4</sup> with a smaller range of user option, and b) as a package of the open-source R environment with a larger set of user options and with each of the steps and methods available independently.

10 The `REddyProc` package can be installed by typing at the R-terminal:

```
install.packages("REddyProc")  
library(REddyProc)  
?REddyProc
```

Alternatively, there is [a-an](#) automatically built docker image (Metzger et al., 2017) at docker hub allowing to run RStudio with `REddyProc` from a browser without any other installation besides docker. Regarding installation issues and docker images we refer the reader to the github project homepage<sup>5</sup>.

15 Some general principles and choices in the design of `REddyProc` that lead to trade-offs between robustness and flexibility are explained in supplement 1.

## Appendix B: Example application

This section reports an example R session using `REddyProc`. Code is shown in a shaded area and corresponding output with monospace font.

### 20 B1 Importing the half-hourly data

The workflow starts with importing the half-hourly data. The example reads a text file with data of the year 1998 from the DE-Tha site and converts the separate decimal columns year, day, and hour to a POSIX timestamp column. Next, it initializes the `sEddyProc` class.

---

<sup>4</sup> [www.bgc-jena.mpg.de/bgi/index.php/Services/REddyProcWeb](http://www.bgc-jena.mpg.de/bgi/index.php/Services/REddyProcWeb)

<sup>5</sup> [github.com/bgctw/REddyProc](https://github.com/bgctw/REddyProc)

```

#### load libraries used in this vignette
library(REddyProc)
library(dplyr)
#### Load data with 1 header and 1 unit row from (tab-delimited) text file
fileName <- getExamplePath('Example_DETha98.txt', isTryDownload = TRUE)
EddyData.F <- if (length(fileName)) fLoadTXTIntoDataframe(fileName) else
  # or use example dataset in RData format provided with REddyProc
  Example_DETha98
#### Add time stamp in POSIX time format
EddyDataWithPosix.F <- fConvertTimeToPosix(EddyData.F, 'YDH', Year.s = 'Year'
  , Day.s = 'DoY', Hour.s = 'Hour')
#### Initialize R5 reference class sEddyProc for post-processing of eddy data
#### with the variables needed for post-processing later
EddyProc.C <- sEddyProc$new('DE-Tha', EddyDataWithPosix.F,
  c('NEE', 'Rg', 'Tair', 'VPD', 'Ustar'))

```

## B2 Estimating the $u_*$ threshold distribution

The second step is the estimation of the distribution of  $u_*$  thresholds to identify periods of low friction velocity ( $u_*$ ), where NEE is biased low. Discarding periods with low  $u_*$  is one of the largest sources of uncertainty in aggregated fluxes. Hence, several quantiles of the distribution of the uncertain  $u_*$  threshold are estimated by a bootstrap.

- 5 The friction velocity,  $u_*$ , needs to be in a column named “Ustar” of the input dataset.

```

uStarTh <- EddyProc.C$sEstUstarThresholdDistribution(
  nSample = 100L, probs = c(0.05, 0.5, 0.95))
#filter(uStarTh, aggregationMode == "year")
select(uStarTh, -seasonYear)

```

##	aggregationMode	season	uStar	5%	50%	95%
## 1	single	<NA>	0.4162500	0.3658125	0.4414583	0.6197520
## 2	year	<NA>	0.4162500	0.3658125	0.4414583	0.6197520
## 3	season	1998001	0.4162500	0.3658125	0.4414583	0.6197520
10 ## 4	season	1998003	0.4162500	0.3198846	0.4037981	0.5615500
## 5	season	1998006	0.3520000	0.3192476	0.3858571	0.4393708
## 6	season	1998009	0.3369231	0.2195485	0.3906458	0.5235175
## 7	season	1998012	0.1740000	0.1829667	0.4189286	0.6180982



The output reports  $u_*$  estimates of 0.42 for the original data and 0.37, 0.44, 0.62 for lower, median, and upper quantile of the estimated distribution. The threshold can vary between periods of different surface roughness, e.g. before and after harvest. Therefore, there are estimates for different time periods of the year, called seasons, reported as different rows. These season-estimates can be aggregated to entire years or to a single value across years, reported by rows with corresponding aggregation mode.

The subsequent post processing steps will be repeated using the three quantiles of the  $u_*$  distribution. They require to specify a  $u_*$ -threshold for each season and a suffix to distinguish the outputs related to different thresholds.

For this example of an evergreen forest site, the same annually aggregated  $u_*$  threshold estimate will be chosen for each of the seasons within a year. In order to distinguish the automatically generated columns, the column names of the estimation results are written to variable `uStarSuffixes`.

```
uStarThAnnual <- usGetAnnualSeasonUStarMap(uStarTh)[-2]
uStarSuffixes <- colnames(uStarThAnnual)[-1]
print(uStarThAnnual)
```

```
##      season      U05      U50      U95
## 1 1998001 0.3658125 0.4414583 0.619752
## 2 1998003 0.3658125 0.4414583 0.619752
## 3 1998006 0.3658125 0.4414583 0.619752
## 4 1998009 0.3658125 0.4414583 0.619752
## 5 1998012 0.3658125 0.4414583 0.619752
```

### B3 Gap-filling

The second post-processing step is filling the gaps using information of the valid data. In this case, the same annual  $u_*$  threshold estimate is used for each season, as described above, and the uncertainty will be computed also for valid records (`FillAll`).

```
EddyProc.C$sMDSGapFillAfterUStarDistr('NEE',
  UstarThres.df = uStarThAnnual,
  UstarSuffix.V.s = uStarSuffixes,
  FillAll = TRUE
)
```

The screen output (not shown here) already shows that the  $u_*$ -filtering and gap-filling was repeated for each given estimate of the  $u_*$  threshold, i.e. column in `uStarThAnnual`, with marking 22% to 38% of the data as gap.

For each of the different  $u_*$  threshold estimates, a separate set of output columns of filled NEE and its uncertainty is generated, distinguished by the suffixes given with `uStarSuffixes`. Suffix “\_f” denotes the filled value and “\_fsd” the estimated standard deviation of its uncertainty.

```
grep("NEE.*_f$", names(EddyProc.C$sExportResults()), value = TRUE)
grep("NEE.*_fsd$", names(EddyProc.C$sExportResults()), value = TRUE)
```

```
## [1] "NEE_U05_f" "NEE_U50_f" "NEE_U95_f"
5 ## [1] "NEE_U05_fsd" "NEE_U50_fsd" "NEE_U95_fsd"
```

## B4 Partitioning net flux into GPP and $R_{eco}$

The third post-processing step is partitioning the net flux (NEE) into its gross components GPP and  $R_{eco}$ . The partitioning algorithm needs a precise criteria between night-time and day-time. Therefore, geographical coordinates and the time zone need to be provided to allow computing exact solar time of sunrise and sunset. Further, missing values in the used meteorological data need to be filled.

```
EddyProc.C$sSetLocationInfo(Lat_deg.n = 51.0, Long_deg.n = 13.6, TimeZone_h.n = 1)
EddyProc.C$sMDSGapFill('Tair', FillAll.b = FALSE)
EddyProc.C$sMDSGapFill('VPD', FillAll.b = FALSE)
```

Now we are ready to invoke the partitioning, here by the night-time approach, for each of the several filled NEE columns.

```
#variable uStarSuffixes was defined above at the end of uStar threshold estimation
resPart <- lapply(uStarSuffixes, function(suffix){
  EddyProc.C$sMRFluxPartition(Suffix.s = suffix)
})
```

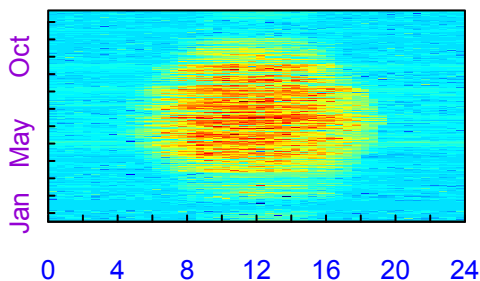
The results are stored in columns `Reco` and `GPP_f` modified by the respective  $u_*$  threshold suffix.

```
grep("GPP.*_f$|Reco", names(EddyProc.C$sExportResults()), value = TRUE)
```

```
## [1] "Reco_U05" "GPP_U05_f" "Reco_U50" "GPP_U50_f" "Reco_U95" "GPP_U95_f"
```

The visualizations of the results by a fingerprint plot gives a compact overview.

```
EddyProc.C$sPlotFingerprintY('GPP_U50_f', Year.i = 1998)
```



## B5 Estimating the uncertainty of aggregated results

First, the mean of the GPP across all the year is computed for each  $u_*$ -scenario and converted from  $\mu\text{mol CO}_2 \text{ m}^{-2}\text{s}^{-1}$  to  $\text{gC m}^{-2}\text{yr}^{-1}$ .

```
FilledEddyData.F <- EddyProc.C$sExportResults()
#suffix <- uStarSuffixes[2]
GPPAggCO2 <- sapply(uStarSuffixes, function(suffix) {
  GPPHalfHour <- FilledEddyData.F[[paste0("GPP_", suffix, "_f")]]
  mean(GPPHalfHour, na.rm = TRUE)
})
molarMass <- 12.011
GPPAgg <- GPPAggCO2 * 1e-6 * molarMass * 3600*24*365.25
print(GPPAgg)
```

```
5 ##          U05          U50          U95
## 1894.097 1956.090 1985.817
```

The difference between those aggregated values is a first estimate of uncertainty range in GPP due to uncertainty of the  $u_*$  threshold.

```
(max(GPPAgg) - min(GPPAgg)) / median(GPPAgg)
```

In this run of the example a relative error of about 4.7% is inferred.

- 10 For a better but more time consuming uncertainty estimate, specify a larger sample of  $u_*$  threshold values, for each repeat the post-processing, and compute statistics from the larger sample of resulting GPP columns. This can be achieved by specifying a larger sequence of quantiles when calling `sEstUstarThresholdDistribution` in section B2.

```
EddyProc.C$sEstUstarThresholdDistribution(
  nSample = 200, probs = seq(0.025, 0.975, length.out = 39) )
```

## B6 Storing the results in a csv-file

The results still reside inside the `sEddyProc` class. To export them to an R Data.frame, the newly generated columns need to be appended to the columns with the original input data. Then this data.frame is written to a text file in a temporary directory.

```
FilledEddyData.F <- EddyProc.C$sExportResults()
CombinedData.F <- cbind(EddyData.F, FilledEddyData.F)
fwriteDataframeToFile(CombinedData.F, 'DE-Tha-Results.txt', Dir.s = tempdir())
```

## B7 Specifying seasons where $u_*$ thresholds differs

- 5 With changing surface roughness, e.g. on harvest or leaf-fall, also the  $u_*$ -NEE relationship can change. Therefore the  $u_*$  threshold needs to be re-estimated at different times of the year, called seasons. The default uses continuous seasons, for details see section 3.2.1. In order to yield results corresponding to DP06, the user can specify `seasonFactor.v = usCreateSeasonFactorMonthWithinYear(EddyData.C$DATA$sDateTime, startMonth= c(3,6,9,12))` as an argument to routine `sEstUstarThreshold`. By default the annual aggregate of the season thresholds, i.e. maximum across seasons, is used to
- 10 identify unfavorable conditions, but the seasonal estimates can also be used instead.

Moreover, the users can specify also other user-defined seasons, e.g. when harvest dates are known (see package vignette `DEGebExample`). They can create a grouping by specifying exact starting days of the periods by function `usCreateSeasonFactorYdayYear`, or they can provide a column with the data that indicates e.g. the same group for two wet seasons. Each season is associated to the year corresponding to the center day between first and last day of the season.

- 15 With all methods, there is a required minimum number of 160 records within a season. If there are too few records, the data of the seasons within one year are combined and the  $u_*$  threshold for this seasons is set to the estimate obtained for the data of the entire year.

(Reichstein et al., 2005a)

*Acknowledgements.* This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks:

- 20 AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia and USCCC. The ERA-Interim reanalysis data are provided by ECMWF and processed by LSCE. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux and AsiaFlux offices.

The authors acknowledge Dario Papale, Gilberto Pastorello and Trevor F Keenan for the discussions on the benchmarking of REddyProc and PvWave code. M. M. and M. R. acknowledge the Alexander Von Humboldt foundation that funded part of this research activity with the Max Planck Research Preis to Markus Reichstein. M. M. acknowledge the MSCA-ITN project TRUSTEE.

L.Š. was supported by the Ministry of Education, Youth and Sports of the Czech Republic within the CzeCOS program, grant number 5 LM2015061, and within the National Sustainability Program I (NPU I), grant number LO1415.

## References

- Aubinet, M., Grelle, A., Ibrom, A., Rannik, U., Moncrieff, J., Foken, T., Kowalski, A. S., Martin, P. H., Berbigier, P., Bernhofer, C., Clement, R., Elbers, J., Granier, A., Grünwald, T., Morgenstern, K., Pilegaard, K., Rebmann, C., Snijders, W., Valentini, R., and Vesala, T.: Estimates of the annual net carbon and water exchange of European forests: The EUROFLUX methodology, *Advances in Ecological Research*, 30, 114–177, 2000.
- Aubinet, M., Vesala, T., and Papale, D., eds.: *Eddy Covariance*, Springer Netherlands, <https://doi.org/10.1007/978-94-007-2351-1>, 2012.
- Baldocchi, D., Chu, H., and Reichstein, M.: Inter-annual variability of net and gross ecosystem carbon fluxes: A review, *Agricultural and Forest Meteorology*, <https://doi.org/10.1016/j.agrformet.2017.05.015>, 2017.
- Barr, A., Richardson, A., Hollinger, D., Papale, D., Arain, M., Black, T., Bohrer, G., Dragoni, D., Fischer, M., Gu, L., et al.: Use of change-point detection for friction–velocity threshold evaluation in eddy-covariance studies, *Agricultural and forest meteorology*, 171, 31–45, <https://doi.org/10.1016/j.agrformet.2012.11.023>, 2013.
- Chapin, F. S., Woodwell, G. M., Randerson, J. T., Rastetter, E. B., Lovett, G. M., Baldocchi, D. D., Clark, D. A., Harmon, M. E., Schimel, D. S., Valentini, R., Wirth, C., Aber, J. D., Cole, J. J., Goulden, M. L., Harden, J. W., Heimann, M., Howarth, R. W., Matson, P. A., McGuire, A. D., Melillo, J. M., Mooney, H. A., Neff, J. C., Houghton, R. A., Pace, M. L., Ryan, M. G., Running, S. W., Sala, O. E., Schlesinger, W. H., and Schulze, E. D.: Reconciling carbon-cycle concepts, terminology, and methods, *Ecosystems*, 9, 1041–1050, 2006.
- Davison, A. and Hinkley, D.: *Bootstrap methods and their application*, Cambridge University Press, Cambridge, 1997.
- Desai, A. R., Richardson, A. D., Moffat, A. M., Kattge, J., Hollinger, D. Y., Barr, A., Falge, E., Noormets, A., Papale, D., Reichstein, M., and Stauch, V. J.: Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, *Agricultural and Forest Meteorology*, 148, 821 – 838, <https://doi.org/https://doi.org/10.1016/j.agrformet.2007.11.012>, 2008.
- Efron, B. and Tibshirani, R.: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statistical science*, pp. 54–75, 1986.
- Falge, E., Baldocchi, D., Olson, R. J., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N. O., Katul, G., Keronen, P., Kowalski, A., Ta Lai, C., Law, B. E., Meyers, T., Moncrieff, J., Moors, E., Munger, J. W., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: Gap filling strategies for defensible annual sums of net ecosystem exchange, *Agric. For. Meteorol.*, 107/1, 29–43, 2001.
- Foken, T. and Wichura, B.: Tools for quality assessment of surface-based flux measurements, *Agricultural and forest meteorology*, 78, 83–105, [https://doi.org/10.1016/0168-1923\(95\)02248-1](https://doi.org/10.1016/0168-1923(95)02248-1), 1996.
- Foken, T., Göckede, M., Mauder, M., Mahrt, L., Amiro, B., and Munger, W.: Post-field data quality control, in: *Handbook of micrometeorology*, pp. 181–208, Springer, 2004.
- Foken, T., Leuning, R., Oncley, S. R., Mauder, M., and Aubinet, M.: Corrections and Data Quality Control, pp. 85–131, Springer Netherlands, Dordrecht, [https://doi.org/10.1007/978-94-007-2351-1\\_4](https://doi.org/10.1007/978-94-007-2351-1_4), 2012.
- Gilmanov, T. G., Johnson, D. A., and Saliendra, N. Z.: Growing season CO<sub>2</sub> fluxes in a sagebrush-steppe ecosystem in Idaho: Bowen ratio/energy balance measurements and modeling, *Basic and Applied Ecology*, 4, 167 – 183, <https://doi.org/http://dx.doi.org/10.1078/1439-1791-00144>, 2003.
- Göckede, M., Rebmann, C., and Foken, T.: A combination of quality assessment tools for eddy covariance measurements with footprint modelling for the characterisation of complex sites, *Agricultural and Forest Meteorology*, 127, 175–188, <https://doi.org/10.1016/j.agrformet.2004.07.012>, 2004.

- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, *Journal of Geophysical Research: Biogeosciences*, 116, n/a–n/a, 2011.
- 5
- Körner, C.: *Leaf Diffusive Conductances in the Major Vegetation Types of the Globe*, pp. 463–490, Springer Berlin Heidelberg, Berlin, Heidelberg, [https://doi.org/10.1007/978-3-642-79354-7\\_22](https://doi.org/10.1007/978-3-642-79354-7_22), 1995.
- Lasslop, G., Reichstein, M., Papale, D., Richardson, A., Arneth, A., Barr, A., Stoy, P., and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, *Global Change Biology*, 16, 187–208, <https://doi.org/10.1111/j.1365-2486.2009.02041.x>, 2010.
- 10
- Lloyd, J. and Taylor, J.: On the temperature dependence of soil respiration, *Functional ecology*, 8, 315–323, 1994.
- Massman, W. J. and Lee, X.: Eddy covariance flux corrections and uncertainties in long-term studies of carbon and energy exchanges, *Agricultural and Forest Meteorology*, 113, 121–144, 2002.
- Menzer, O., Moffat, A. M., Meiring, W., Lasslop, G., Schukat-Talamazzini, E. G., and Reichstein, M.: Random errors in carbon and water vapor fluxes assessed with Gaussian Processes, *Agricultural and Forest Meteorology*, 178–179, 161–172, <https://doi.org/10.1016/j.agrformet.2013.04.024>, 2013.
- 15
- Metzger, S., Durden, D., Sturtevant, C., Luo, H., Pingintha-Durden, N., Sachs, T., Serafimovich, A., Hartmann, J., Li, J., Xu, K., and Desai, A. R.: eddy4R 0.2.0: a DevOps model for community-extensible processing and analysis of eddy-covariance data based on R, Git, Docker, and HDF5, *Geoscientific Model Development*, 10, 3189–3206, <https://doi.org/10.5194/gmd-10-3189-2017>, 2017.
- 20
- Moffat, A. M.: A new methodology to interpret high resolution measurements of net carbon fluxes between terrestrial ecosystems and the atmosphere, Ph.D. thesis, Friedrich-Schiller-Universität Jena, 2012.
- Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., Beckstein, C., Braswell, B. H., Churkina, G., Desai, A. R., et al.: Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes, *Agricultural and Forest Meteorology*, 147, 209–232, <https://doi.org/10.1016/j.agrformet.2007.08.011>, 2007.
- 25
- Muggeo, V.: Segmented: An R package to fit regression models with broken-line relationships, *R News*, 8 (1), 20: 25, 2008.
- Muggeo, V. M.: Estimating regression models with unknown break-points, *Statistics in medicine*, 22, 3055–3071, 2003.
- Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., and et al.: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, *Biogeosciences*, 3, 571 – 583, <https://doi.org/10.5194/bg-3-571-2006>, 2006.
- 30
- Pastorello, G., Papale, D., Chu, H., Trotta, C., Agarwal, D., Canfora, E., Baldocchi, D., and Torn, M.: The FLUXNET2015 dataset: The longest record of global carbon, water, and energy fluxes is updated, *Eos Trans. AGU*, <https://doi.org/10.1029/2017EO071597>, 2017.
- Rasmussen, C. E. and Williams, C. K. I.: *Gaussian Processes for Machine Learning*, The MIT Press, 2006.
- Rebmann, C., Kolle, O., Heinesch, B., Queck, R., Ibrom, A., and Aubinet, M.: *Data Acquisition and Flux Calculations*, pp. 59–83, Springer Netherlands, Dordrecht, [https://doi.org/10.1007/978-94-007-2351-1\\_3](https://doi.org/10.1007/978-94-007-2351-1_3), 2012.
- 35
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., et al.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global Change Biology*, 11, 1424–1439, <https://doi.org/10.1111/j.1365-2486.2005.001002.x>, 2005a.

- Reichstein, M., Katterer, T., Andren, O., Ciais, P., Schulze, E. D., Cramer, W., Papale, D., and Valentini, R.: Temperature sensitivity of decomposition in relation to soil organic matter pools: critique and outlook, *Biogeosciences*, 2, 317–321, <https://doi.org/10.5194/bg-2-317-2005>, 2005b.
- Reichstein, M., Subke, J. A., Angeli, A. C., and Tenhunen, J. D.: Does the temperature sensitivity of decomposition of soil organic matter depend upon water content, soil horizon, or incubation time?, *Global Change Biology*, 11, 1754–1767, 2005c.
- Reichstein, M., Stoy, P. C., Desai, A. R., Lasslop, G., and Richardson, A. D.: Partitioning of Net Fluxes, pp. 263–289, Springer Netherlands, Dordrecht, [https://doi.org/10.1007/978-94-007-2351-1\\_9](https://doi.org/10.1007/978-94-007-2351-1_9), 2012.
- Stoy, P. C., Katul, G. G., Siqueira, M. B., Juang, J.-Y., Novick, K. A., Uebelherr, J. M., and Oren, R.: An evaluation of models for partitioning eddy covariance-measured net ecosystem exchange into photosynthesis and respiration, *Agricultural and Forest Meteorology*, 141, 2 – 18, <https://doi.org/10.1016/j.agrformet.2006.09.001>, 2006.
- van Gorsel, E., Leuning, R., Cleugh, H. A., Keith, H., and Suni, T.: Nocturnal carbon efflux: reconciliation of eddy covariance and chamber measurements using an alternative to the u.-threshold filtering technique, *Tellus B: Chemical and Physical Meteorology*, 59, 397–403, 2007.
- Wehr, R. and Saleska, S.: An improved isotopic method for partitioning net ecosystem–atmosphere CO<sub>2</sub> exchange, *Agricultural and Forest Meteorology*, 214, 515–531, 2015.



# Supplement 1 to Basic and extensible post-processing of eddy covariance flux data with REddyProc.

Thomas Wutzler<sup>1</sup>, Antje Lucas-Moffat<sup>2,3</sup>, Mirco Migliavacca<sup>1</sup>, Jürgen Knauer<sup>1</sup>, Kerstin Sickel<sup>1</sup>, Ladislav Šigut<sup>4</sup>, Olaf Menzer<sup>5</sup>, and Markus Reichstein<sup>1</sup>

<sup>1</sup>Max Planck Institute for Biogeochemistry, Hans-Knöll-Straße 10, 07745 Jena, Germany

<sup>2</sup>German Meteorological Service, Centre for Agrometeorological Research, Bundesallee 33, 38116 Braunschweig, Germany

<sup>3</sup>Also at: Thuenen Institute of Climate-Smart Agriculture, Bundesallee 65, 38116 Braunschweig, Germany

<sup>4</sup>Global Change Research Institute CAS, Bělidla 986/4a, CZ-60300 Brno, Czech Republic

<sup>5</sup>Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA

**Correspondence:** Thomas Wutzler  
(twutz@bgc-jena.mpg.de)

## 1 Marginal Details of marginal distribution sampling (MDS) for gap-filling details

The marginal distribution ~~sampling Look-up-table (section 2.2.3)~~ combines the two methods ~~look-up-table (LUT) and Mean-mean~~ diurnal course (MDC) ~~methods depending on the availability of the meteorological data.~~ Depending on the availability of the meteorological data, ~~Three, three~~ different conditions are identified for each half-hourly NEE flux:

1. The data of the three meteorological variables (Rg, Tair, and VPD) are available.
2. Tair or VPD are missing, but Rg is available.
3. Also Rg is missing.

Case 1): The missing value is replaced by the average value under similar meteorological conditions with respect to Rg, Tair and VPD in a LUT approach. If no similar meteorological conditions (minimum of two half-hourly fluxes) are present within the starting time window of 7 days, the ~~windows window~~ size is increased to 14 days.

Case 2): The same LUT approach is taken, but similar meteorological conditions can only be defined via Rg within a time window of 7 days.

Case 3): The missing value is replaced with the mean diurnal course (MDC). The number of days start with one day, thus

a linear interpolation of available data at adjacent hours ( $\pm 1$  hour) at the same day. The number is then increased to  $\pm 1$  and  $\pm 2$  days.

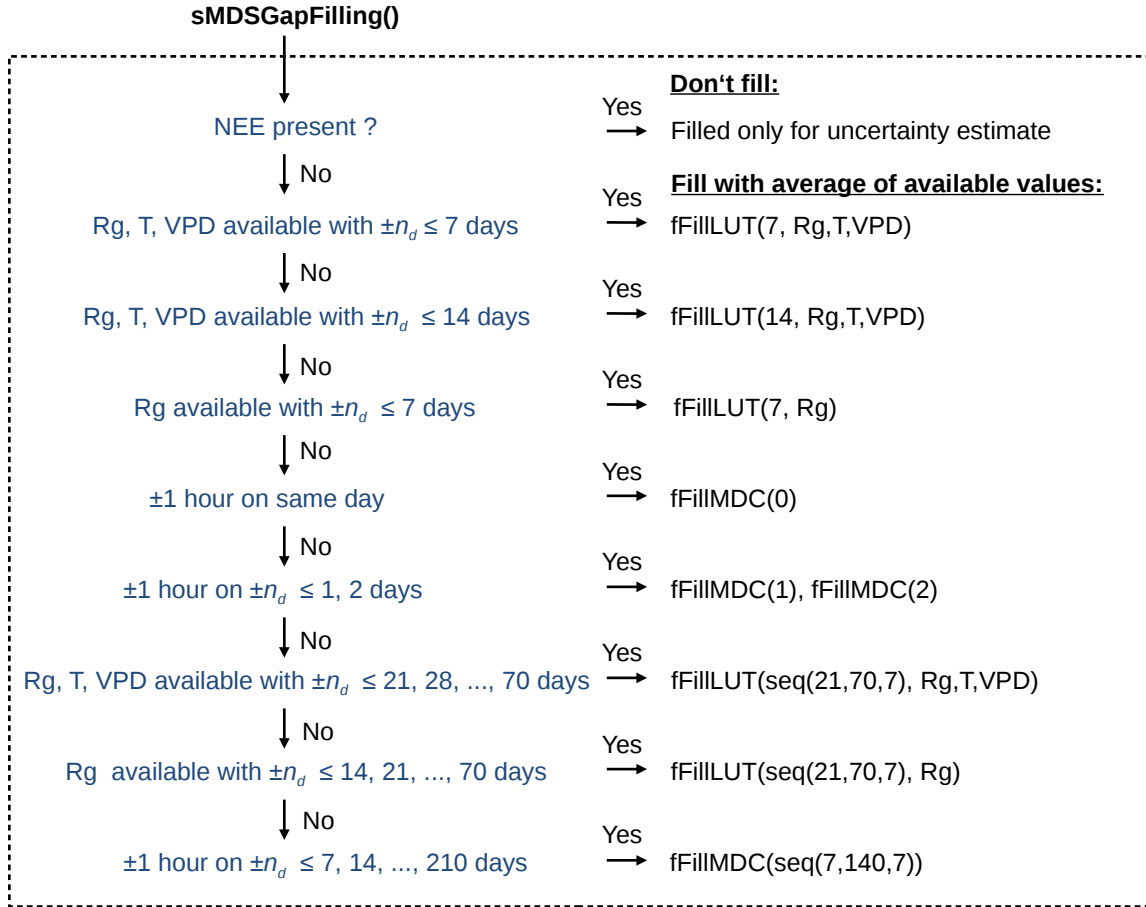
If after these steps the NEE values could not be filled, the procedure is repeated with increased window sizes until the value can be filled, see flow diagram in Fig. 1.

The ~~provided~~ quality flag depends on both, the number of meteorological conditions present,  $n_c$ , and the number of days in the window,  $n_d$ . When using LUT, ~~i.e. when  $n_c \geq 1$ , it the flag~~ equals 1 ~~if the  $n_d \leq 14$  when  $n_c$  is either 1 or 3 and  $n_d \leq 7$ .~~ It equals 3 if either  $n_c = 3$  and  ~~$n_d \geq 56$  or  $n_c < 3$  and  $n_d \geq 28$ .~~ ~~It equals 2 in the  $n_d > 28$ , or  $n_c = 1$  and  $n_d > 14$ .~~ In all other LUT cases, ~~it equals 2.~~ When using MDC, i.e.  $n_c = 0$ , the quality flag equals 1 if  ~~$n_d \leq 1$  it is the same day,~~ it equals 3 if  ~~$n_d > 5$   $n_d \geq 2$ ,~~ and it equals 2 in ~~other MDC all other~~ cases.

~~The three default variables and margins are Rg with  $\pm 50$   $Wm^{-2}$ , Tair with  $\pm 2.5$  °C, and VPD with  $\pm 5.0$  hPa. These can be also be specified by the user.~~

## 2 Details of daytime flux-partitioning

This section reports details of the steps ~~described in the post-processing methods section of for~~ daytime flux-partitioning ~~(section 2.3.2).~~



**Figure 1.** Flow diagram of the MDS gap-filling algorithm as implemented in REdDyProc. See table 1 in the [main](#) paper for abbreviations.

In step 1, parameter  $E_0$  is estimated for 12 day windows. Only records with temperature above  $-1$  °C are valid for estimation. Reference temperature  $T_{Ref}$  in eq. 1 of the paper is set to the median temperature of the window in order to decrease correlation between estimates of  $R_{Ref}$  and  $E_0$ . A missing estimate is reported for non-valid windows with too few valid records ( $\text{minNRecInDayWindow} = 10$ ), non-convergence of the fitting procedure, or an  $E_0$  estimate outside the bounds  $[50,400]$ . Missing estimates are filled during the smoothing in step 2.

In step 2, the Gaussian Process takes into account the uncertainty of the  $E_0$  fit from night-time in each window. However, if the correlation of  $E_0$  across subsequent windows is high, the uncertainty is reduced similar as with repeated measurements. The respiration  $R_{Ref}$  for windows where no fit could be obtained is set to the value from the previous valid window. Again, only records with temperature above  $-1$  °C are used.

In step 3, fitting of other parameters is done for each window centered at the same record as the windows of step 1.

By default the fit uses the same weak prior on parameters as BGC16 (Lasslop et al., 2010). The ~~prior locations are optimization minimizes a cost function using the "BFGS" method, which is a quasi-Newton method as published by Shanno (1970). The cost function assumes normally distributed model-data-residuals and normally distributed vague priors (eq. 1) (Lasslop et al., 2010, eq. 5).~~

$$c = \sum_i \frac{(NEE_{pred,i} - NEE_{obs,i})^2}{\sigma_{NEEadj,i}^2} + \sum_j \frac{(\theta_j - \theta_{prior,j})^2}{\sigma_{\theta,j}^2}, \quad (1)$$

where  $NEE_{pred,i}$  is computed by the LRC equation, and parameters  $\theta = (k, \alpha, R_{Ref}, \beta)$  have prior locations of  $k = 0.05$ ,  $\alpha = 0.1$ ,  $R_{Ref}$  = night-time estimate, and  $\beta$  = range of NEE values, i.e. the difference between 97% and 3% quantile. The prior uncertainty is  $sd_k = 50$ ,  $sd_\beta = 600$ ,  $sd_\alpha = 10$ , and  $sd_{R_{Ref}} = 80$ ,  $\sigma_k = 50$ ,  $\sigma_\beta = 600$ ,  $\sigma_\alpha = 10$ , and  $\sigma_{R_{Ref}} = 80$ .

In each daytime fit, the NEE records are weighted according to their variance (Lasslop et al., 2010, eq. 5). In order to avoid unreasonable leverage of records with a very low estimate of NEE uncertainty, the records of the above 0.7 quantile of weights are associated the weight of the 0.7 quantile. This The uncertainty of NEE by default is assigned a lower bound (eq. 1). This lower bound differs from Lasslop et al. (2010) and assigns low influence to records with high uncertainty, but avoids in order to avoid the problem of the high leverage high leverage of a few records with very low estimates of NEE-uncertainty.

There are certain

$$\sigma_{NEEadj,i}^2 = \max[\sigma_{NEE,i}^2, q_{0.3}(\sigma_{NEE}^2)], \quad (2)$$

where  $q_{0.3}(\sigma_{NEE}^2)$  is the 30% quantile of the vector of estimated standard deviations of NEE.

There are several quality criteria and fall-backs during the daytime fitting in order to obtain reasonable fits. If there are too few valid records ( $\text{minNRecInDayWindow} < 10$ ), or the fitting did not converge, or VPD parameter  $k < 0$ , then the fit is repeated without the VPD effect, because often there are records where VPD is missing but other variables are available. If fitting did not converge or parameter estimate of  $\alpha$  is larger than 0.22, the fit is repeated with  $\alpha$  fixed to the last valid value of  $\alpha$  from fits in previous windows. If there are still too few records or the fitting was not valid, a missing result is reported for the window. In addition a missing result is reported if estimated  $\alpha < 0$  or  $R_{Ref} < 0$  or  $\beta_0 < 0$  or  $\beta_0 > 250$ , or if  $\beta_0 > 100$  and at the same time estimated standard deviation  $sd_{\beta_0} \geq \beta_0$  (Table A1 in Lasslop et al., 2010). The Variance-Covariance matrix of the parameter uncertainty is estimated by bootstrapping the day-time fit. In each sample, the prescribed temperature sensitivity  $E_0$  is drawn from a normal distribution with standard deviation estimated in step 2. In this way also the uncertainty of the night-time fit propagates to the uncertainty of the day-time parameters and subsequently to the inferred gross fluxes.

In step 4, variance of the flux estimates are computed based on the Variance-Contrivance matrices obtained in step 3 with each parameter estimate (Lasslop et al., 2010, eq. 6). The two standard deviations based either on the previous and subsequent valid estimates of the Variance-Contrivance matrices are linearly interpolated with respect to the time difference to the estimates.

### 3 General design of REddyProc

There are some general principles and choices in the design of REddyProc that explain which may lead to some trade-offs.

Preventing accidental errors is one goal of the package. Therefore, there is extensive checking on formatting and data availability when importing the data and also throughout post-processing steps. Especially, the high-level routines will issue warnings or stop post-processing if they detect inconsistencies or lack of sufficient data, or changes in specification of critical standard parameters. These checks reinforce a sound standard post-processing also for non-expert users. On the other hand, these checks will render the standard routines not usable for datasets where not enough data are available. It is still possible to use REddyProc with sparse data for some purposes by using lower-level routines. But this requires experience in both, post-processing and R programming.

Smooth inter-annual processing is the next goal. The data is not partitioned into annual chunks for post-processing to avoid artificial discontinuities between years. The user can still specify different periods or seasons, e.g. when meteorological conditions change after harvest, but the boundaries do not need to align with years. REddyProc therefore works with the entire dataset in memory. Potentially this can lead to longer post-processing times when working with many years of EC data but still can be handled on a usual notebook.

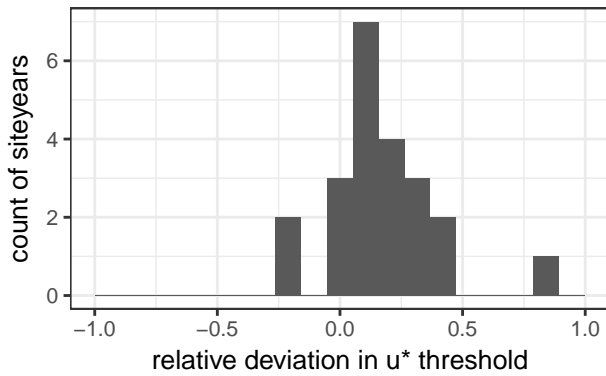
Memory efficient processing required some extended R programming. Specifically we used R5 classes<sup>1</sup> to avoid frequently copying the entire dataset in memory. Users should be aware that calling functions on the sEddyProc class not only provides a return value but also changes the data of the R5 class (Chapter 'OO field guide' in Wickham, 2014). Users who want to integrate the post-processing in their own codes are encouraged to learn about R5 classes, however, it is not a prerequisite.

Continuity with other tools ensures that switching to REddyProc does not introduce discontinuities with results obtained from other tools. Specifically, we reproduced the  $u_*$  threshold estimation of the C implementation by Dario Papale (DP06), and the gap-filling, night-time flux partitioning, and day-time flux-partitioning from BGC16, the 2016 version of the BGC online tool (Reichstein et al., 2005). In some details, the standard parameterization differs from BGC16, e.g. using seasons that span across calendar years in  $u_*$  threshold estimation, but it is usually possible to set compatible parameters with REddyProc.

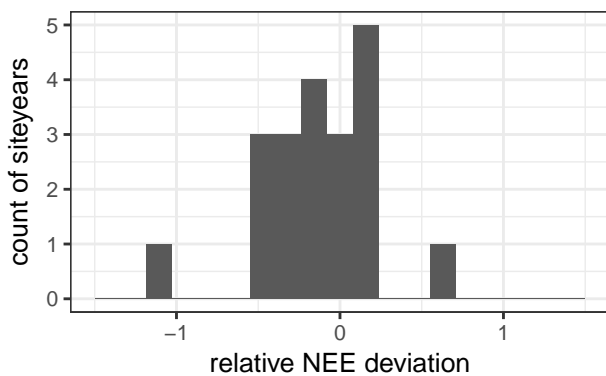
## 4 Additional benchmark statistics

This section provides figures, in addition to the main figures presented in the paper's benchmark section, that require the reader being comfortable with histograms and probability distribution functions.

<sup>1</sup> [www.rdocumentation.org/packages/methods/versions/3.4.3/topics/ReferenceClasses](http://www.rdocumentation.org/packages/methods/versions/3.4.3/topics/ReferenceClasses)



**Figure 2.** Histogram of differences between  $u_*$  threshold estimates of different methods (DP06 - REddyProc) normalized by the uncertainty range (90% quantile - 10% quantile estimated by DP06). Absolute values are smaller than one meaning that the difference between the methods is smaller than range of estimates by the DP06 method only. The clustering of values on the positive side suggests a bias towards larger values with DP06.

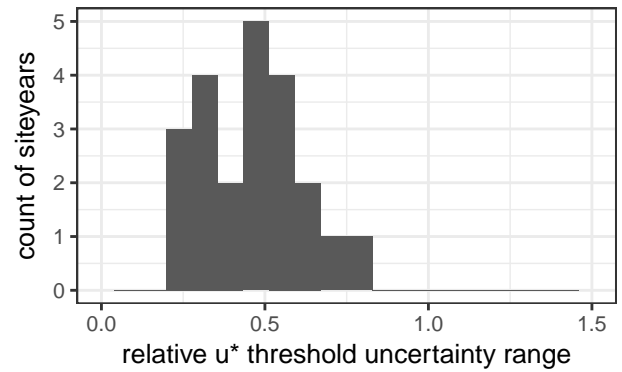


**Figure 3.** Histogram of differences between annual NEE based on  $u_*$  estimates of different methods (DP06 - REddyProc) normalized by the uncertainty range of NEE due to  $u_*$  (NEE based on 90%  $u_*$  quantile - NEE based on 10%  $u_*$  quantile estimated by DP06). Absolute values are mostly smaller than one, meaning that the difference in annual NEE between methods is smaller than the uncertainty due to uncertainty of  $u_*$  threshold from DP06 only.

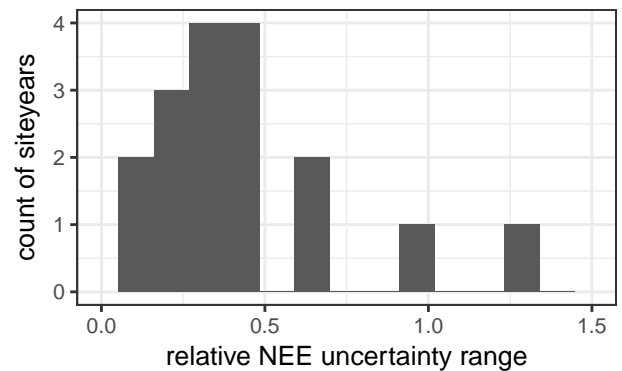
The histograms display how often a certain value occurs across all the site-months. A centering of a difference away from zero or for a ratio away from one denotes a bias.

#### 4.1 $u_*$ threshold estimation

There was a bias in  $u_*$  estimation compared to DP06 (Fig. 2). However, the bias and the differences were only a small fraction of the uncertainty of the estimate itself. Moreover, the bias did not propagate to estimated NEE based on different  $u_*$  values (Fig. 3).



**Figure 4.** Histogram of ratio (REddyProc / DP06) of uncertainty ranges of  $u_*$  (90% quantile - 10% quantile). The clustering around 0.5 shows that the estimated uncertainty with REddyProc is only about half the uncertainty estimated by DP06.



**Figure 5.** Histogram of ratio (REddyProc / DP06) of uncertainty ranges of annual NEE (NEE based on 90% quantile of  $u_*$  threshold - NEE based on 10% quantile of  $u_*$  threshold). The clustering below a value of one indicates that the lower uncertainty in  $u_*$  threshold with REddyProc also propagates to lower uncertainty in annual NEE.

There was a significant difference in the estimate of uncertainty of the  $u_*$  thresholds (Fig. 4). The uncertainty estimated by REddyProc was only half as high as the one estimated by DP06. It resulted from acknowledging the seasons of similar conditions also during bootstrap. The lower  $u_*$  threshold uncertainty propagated to the uncertainty estimates of NEE (Fig. 5).

#### 4.2 Gap-filling

The evaluation of annual aggregated NEE data obtained with REddyProc and BGC16 showed good agreement across sites.

Note, that the aggregated NEE data contained both, measured and gap-filled data for the purpose of evaluating the impact of the processing on the aggregated NEE. The determination

coefficient ( $R^2$ ) showed a very good agreement between the two methods both at annual and monthly time scale (Table 1). The relative mean absolute error (RMAE) is low: about -3 % for annual aggregation and -0.47 % for monthly aggregation. Both Modeling Efficiency (EF) and Mean Bias Error (MBE) showed a very low bias between the two products for both monthly and annual aggregations (Table 1).

**Table 1.** Gap-filling evaluation statistics of yearly and monthly cumulated and gap-filled NEE data obtained with REddyProc and BGC16. Statistic abbreviations are explained in the paper’s benchmarking section

	Yearly	Monthly
N	25	281
pearson	0.99	1.00
MBE	-0.02	0.00
RMBE	2.8%	-0.08%
MAE	0.02	0.00
RMAE	-3.0%	-0.5%
RMSE	0.09	0.01
$R^2$	0.98	1.00
EF	0.98	1.00

### 4.3 Night-time partitioning

The results show good agreement between  $R_{eco}$  estimated using REddyProc and BGC16 for  $R^2$  and modelling efficiency (Table 2). The relative RMSE is 6.56 and 13.45 % for yearly and monthly aggregation, respectively. The high RRMSE is due to few site years as reported in Table 3, and this is confirmed by the lower RMAE (3.87 % and 6.33 % for yearly and monthly, respectively), which is less sensitive to outliers.

The difference in  $R_{eco}$  related to the selection of night-time data are not negligible: the differences in RRMSE of 1.03% and negligible differences in

$$R^2$$

. Also, the use of  $E_0$  prescribed from BGC16 lead to negligible difference in  $R^2$  of about 0.02. Therefore, though very small, the selection of night-time data is the most important difference introduced by REddyProc.

### 4.4 Day-time partitioning

The time-variable estimate of temperature sensitivity of ecosystem respiration with day-time partitioning is a significant source of uncertainty for gross fluxes GPP and  $R_{eco}$ . REddyProc accounts for the previously unaccounted uncertainty for estimating uncertainty of these gross fluxes by a bootstrap (Fig. 6). The BGC16 estimate of annual uncertainty in Fig. 6 is a low estimate compared to a full quantification,

**Table 2.** Nighttime partitioning evaluation statistics across sites of annually and monthly aggregated ecosystem respiration ( $R_{eco}$ ) estimated with REddyProc and BGC16.

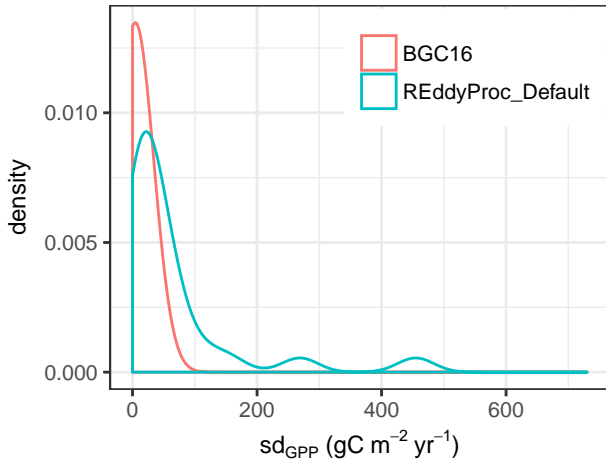
	Yearly	Monthly
N	25	297
pearson	0.99	0.99
MBE	25.4	2.15
RMBE	2.2%	2.2%
MAE	44.8	6.17
RMAE	3.8%	6.3%
RMSE	75.6	13.2
RRMSE	6.5%	13.5%
$R^2$	0.99	0.98
EF	0.99	0.97

	RMBE	RMAE	RRMSE	$R^2$	EF
CA-NS7	15.12	16.29	32.64	0.96	0.86
CA-TP3	-1.99	25.10	38.59	0.87	0.82
DE-Hai	6.00	6.48	10.40	0.99	0.96
DE-Tha	2.83	3.52	6.19	0.99	0.99
DK-Sor	3.50	4.45	7.34	1.00	0.99
ES-ES1	1.68	3.34	4.59	0.93	0.92
ES-VDA	3.96	9.98	15.93	0.95	0.93
FI-Hyy	2.72	3.05	4.47	1.00	1.00
FI-Kaa	-4.52	5.51	9.07	0.99	0.99
FR-Gri	3.96	4.14	6.35	1.00	0.99
FR-Hes	4.11	5.06	9.06	0.99	0.99
FR-Lq1	-2.89	5.76	7.24	0.99	0.99
FR-Lq2	-8.27	9.49	11.49	0.99	0.97
FR-Pue	0.65	1.27	1.55	1.00	1.00
IE-Dri	-0.03	2.41	3.24	1.00	1.00
IL-Yat	1.66	2.89	3.72	0.99	0.99
IT-Amp	24.34	29.38	45.33	0.91	0.48
IT-MBo	2.03	3.00	4.49	1.00	1.00
IT-SRo	-1.46	1.86	2.10	1.00	0.99
PT-Esp	9.06	12.08	19.94	0.62	0.45
RU-Cok	-0.76	28.54	38.98	0.27	-0.07
SE-Nor	-1.05	2.09	3.62	1.00	1.00
US-Ton	2.93	7.45	11.82	0.93	0.92
VU-Coc	0.73	2.64	3.83	0.93	0.93

**Table 3.** Nighttime partitioning evaluation statistics at sites level of the of ecosystem respiration ( $R_{eco}$ ) estimated with REddyProc and BGC16.

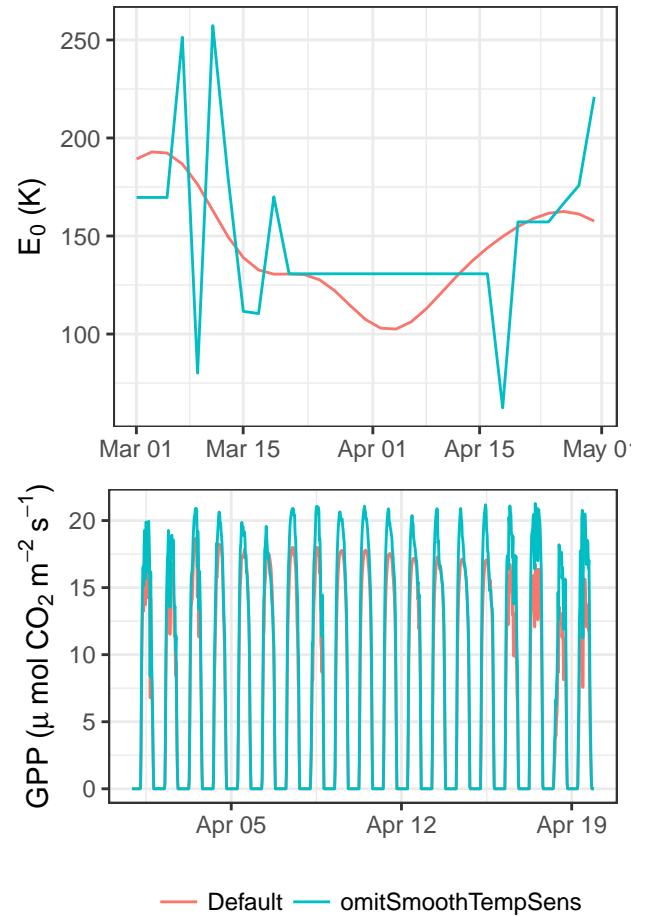
because it assumes no correlation between half-hourly errors. An improved quantification of correlations requires the full variance-covariance matrix of the LRC parameter fits (Lasslop et al., 2010; Menzer et al., 2013), which were not available for BGC16.

The introduced uncertainty is reduced by smoothing the temperature sensitivity estimates ( $E_0$ ) across several successive windows (Fig. 6-7 top) before estimating parameters of the LRC. This smoothing has also an effect on predicted half-hourly gross fluxes (Fig. 7 bottom).

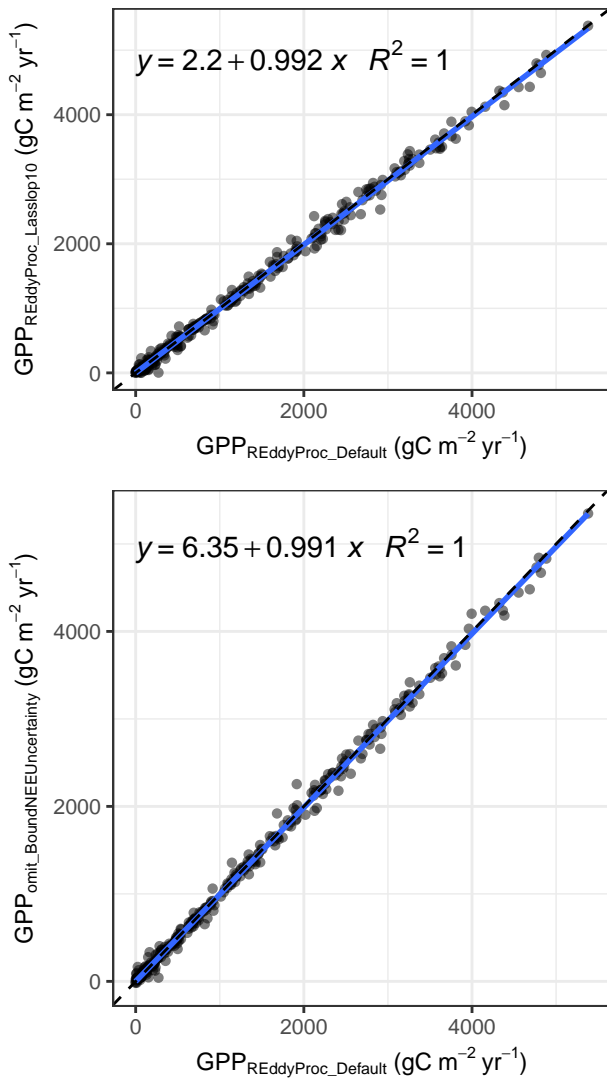


**Figure 6.** Density plot of estimated standard deviation of uncertainty of the annually aggregated GPP across sites due to uncertainty in parameters estimation during day-time based flux partitioning. Higher estimates with REddyProc are caused by taking into account the uncertainty in temperature sensitivity,  $E_0$ .

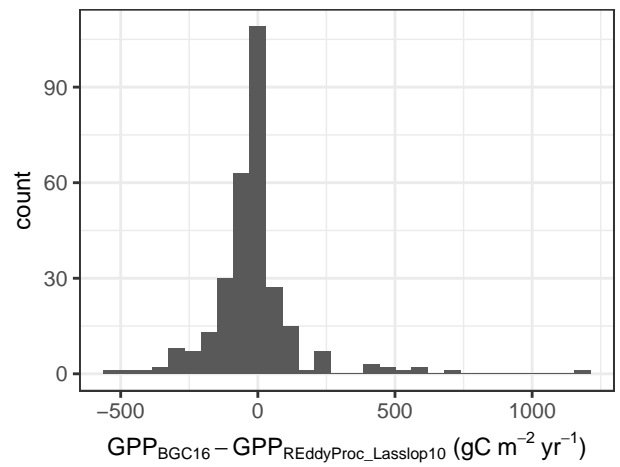
Results of daytime partitioning are sensitive to subtle details of the procedure. Hence, there is quite much scatter introduced by the differences of REddyProc processing with default options or with options that maximize compatibility with BGC16 (Fig. 8 top). One such a subtle options is to decrease or not to account for the unreasonably high leverage of some observations during the fit to the light-response curve by some records having a very small estimate of its uncertainty (Fig. 8 bottom). Due to the sensitivities of the day-time partitioning, there are still differences between REddyProc with compatibility options and BGC16 (Fig. 9).



**Figure 7.** Effects of smoothing (top) successive estimates of temperature sensitivity,  $E_0$ , on predicted GPP (bottom) for site PT-Esp.



**Figure 8.** Sensitivity of estimated monthly GPP fluxes to specific processing details results in scatter between GPP predictions based on different REddyProc options. Most of the differences between default options and compatibility options (top) are caused by differences in weighting different records during the fit (bottom.)



**Figure 9.** Histogram of difference between monthly GPP predictions of REddyProc with Lasslop10 compatibility options BGC16.

## References

- Lasslop, G., Reichstein, M., Papale, D., Richardson, A., Arneeth, A., Barr, A., Stoy, P., and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, *Global Change Biology*, 16, 187–208, <https://doi.org/10.1111/j.1365-2486.2009.02041.x>, 2010.
- Menzer, O., Moffat, A. M., Meiring, W., Lasslop, G., Schukat-Talamazzini, E. G., and Reichstein, M.: Random errors in carbon and water vapor fluxes assessed with Gaussian Processes, *Agricultural and Forest Meteorology*, 178-179, 161–172, <https://doi.org/10.1016/j.agrformet.2013.04.024>, 2013.
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., et al.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global Change Biology*, 11, 1424–1439, <https://doi.org/10.1111/j.1365-2486.2005.001002.x>, 2005.
- Shanno, D. F.: Conditioning of quasi-Newton methods for function minimization, *Mathematics of Computation*, 24, 647–647, <https://doi.org/10.1090/s0025-5718-1970-0274029-x>, 1970.
- Wickham, H.: *Advanced R*, Taylor & Francis Inc, 2014.