

Stable gap-filling for longer eddy covariance data gaps: A globally validated machine-learning approach for carbon dioxide, water, and energy fluxes

Songyan Zhu^{a,*}, Robert Clement^a, Jon McCalmont^a, Christian A. Davies^b, Timothy Hill^a

^a College of Life and Environmental Science, University of Exeter, Streatham Campus, Rennes Drive, Exeter EX4 4RJ, UK

^b Shell International Exploration and Production Inc, Shell Technology Centre, Houston, TX 77082, United States of America

ARTICLE INFO

Keywords:

Global land ecosystems
Carbon exchange
Eddy covariance
Long gaps
Robust gap-filling

ABSTRACT

Continuous time-series of CO₂, water, and energy fluxes are useful for evaluating the impacts of climate-change and management on ecosystems. The eddy covariance (EC) technique can provide continuous, direct measurements of ecosystem fluxes, but to achieve this gaps in data must be filled. Research-standard methods of gap-filling fluxes have tended to focus on CO₂ fluxes in temperate forests and relatively short gaps of less than two weeks. A gap-filling method applicable to other fluxes and capable of filling longer gaps is needed.

To address this challenge, we propose a novel gap-filling approach, Random Forest Robust (RFR). RFR can accommodate a wide range of data gap sizes, multiple flux types (i.e. CO₂, water and energy fluxes). We configured RFR using either three (RFR₃) or ten (RFR₁₀) driving variables. RFR was tested globally on fluxes of CO₂, latent heat (LE), and sensible heat (H) from 94 suitable FLUXNET2015 sites by using artificial gaps (from 1 to 30 days in length) and benchmarked against the standard marginal distribution sampling (MDS) method.

In general, RFR improved on MDS's R² by 15% (RFR₃) and by 30% (RFR₁₀) and reduced uncertainty by 70%. RFR's improvements in R² for H and LE were more than twice the improvement observed for CO₂ fluxes. Unlike MDS, RFR performed well for longer gaps; for example, the R² of RFR methods in filling 30-day gaps dropped less than 4% relative to 1-day gaps, while the R² of MDS dropped by 21%.

Our results indicate that the RFR method can provide improved gap-filling of CO₂, H and LE flux timeseries. Such improved continuous flux measurements, with low bias, can enhance our understanding of the impacts of climate-change and management on ecosystems globally.

1. Introduction

To keep climate change to below 1.5 °C within reach (Wollenberg et al., 2016; Glanemann et al., 2020; Smith et al., 2021), Natural Climate Solutions (NCS) (Griscom et al., 2017) may be the most cost-effective approach immediately ready for large-scale deployment (Cohen-Shacham et al., 2019), because land ecosystems absorb approximately one third of anthropogenic C emission per year (Friedlingstein et al., 2020). NCS have already been implemented in 66% of countries (Chausson et al., 2020), but measuring and verifying the effectiveness of NCS remains challenging (Skinner and Dell, 2015; Smith et al., 2020; Bautista et al., 2021).

Eddy covariance (EC) has been suggested as part of the solution to the NCS measurement challenge e.g. inaccessible and hard-to-observe carbon pool changes (Baldocchi, 2020; Keith et al., 2021; Hemes et al., 2021). EC can monitor (ecosystem-scale) mass (CO₂, water, CH₄,

and N₂O.) and energy fluxes continuously (Aubinet et al., 2012; Hill et al., 2017; Baldocchi, 2020), with a broad convergence between EC and other carbon exchange quantification methods (Skinner and Dell, 2015; Campioli et al., 2016). Currently over 400 EC towers are contributing datasets to the global synthesis project FLUXNET (Baldocchi et al., 2001; Baldocchi, 2014; Pastorello et al., 2020).

However, data gaps hinder the application of EC flux time-series (Aubinet et al., 2012). Most EC data gaps occur as a result of instrument failure (e.g. power loss and sensor malfunction) (Papale et al., 2006), rejection of data during quality control (Mauder et al., 2008), and data loss through adverse environmental conditions (Falge et al., 2001). Gap-filling approaches for EC include the research-standard Marginal Distribution Sampling (MDS) (Reichstein et al., 2005; Pastorello et al., 2020), which fills gaps by considering the covariance of fluxes with meteorological drivers (global radiation, air temperature and vapour pressure deficit) and the temporal autocorrelation of the flux values

* Corresponding author.

E-mail address: sz394@exeter.ac.uk (S. Zhu).

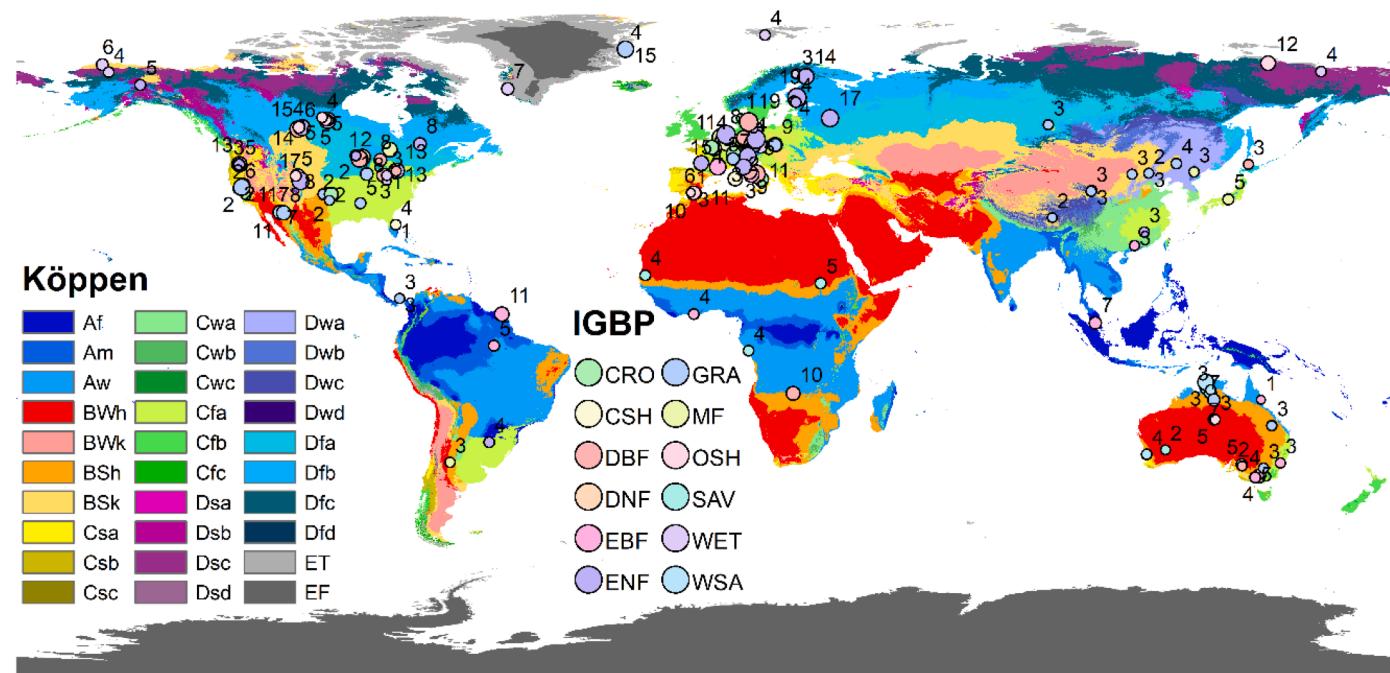


Fig. 1. FLUXNET2015 sites (dots) used for gap-filling. The underlying map represents the Koppen climate classifications. Dot colors represent the International Geosphere-Biosphere Programme (IGBP) land cover classification. Dot sizes represent the data length in years of sites (noted by the numbers aside).

(Reichstein et al., 2005), and other numerical methods (e.g. machine-learning) aiming for improving gap-filling performance (Vitale et al., 2019; Irvin et al., 2021).

Previous comparisons of gap-filling approaches have tended to focus on gaps in carbon fluxes of up to two weeks in temperate forests (Moffat et al., 2007) despite being routinely applied globally to carbon, water, and heat fluxes. Whilst MDS has been demonstrated as an effective gap-filling method for filling short gaps using a small driver set (Moffat et al., 2007), it was reported not designed for long temporal data gaps (Kang et al., 2019). Additional uncertainty in filled long NEE gaps (~three weeks) was reported (Richardson and Hollinger, 2007), but still no robust methods have been proposed for filling long gaps. Machine-learning (e.g. random forest) methods outperformed MDS in filling, e.g. methane flux, gaps, but they require 7–14 drivers (e.g. leaf area index) to fill gaps (Menzer et al., 2015; Kim et al., 2020). It remains unknown if more recent machine-learning methods can improve on MDS for the same driver sets and as machine learning can leverage information from a larger, expanded driver set.

In this paper, we present a gap-filling approach for NEE (CO_2 fluxes), H (sensible heat), and LE (latent energy), based on a new Random Forest-Robust (RFR) algorithm, that is designed to be effective for longer data gaps. RFR was implemented using two different driver sets to simulate good and poor driver availability: 1) the same three meteorological drivers as MDS (RFR_3) and 2) an expansion to ten drivers (RFR_{10}) to explore if additional gap-filling improvements can be seen by exploiting this wider range of drivers. We evaluated RFR_3 and RFR_{10} against MDS by using 94 globally distributed sites (806 site-years) from the FLUXNET database. Gap-filling and validation were carried out for artificial gaps much longer than previous validations (Moffat et al., 2007), with a combination of short (24 h), long (7-day), and very long (30-day) missing periods. Finally, we independently verified gap-filling performance by comparing the EBR (energy balance ratio) of measured data to the EBR of gap-filled data. To explore the limitations of approaches, gap-filling performance was examined for daytime and night-time periods and for different international Geosphere-Biosphere Programme (IGBP) ecosystems surface classifications.

2. Methodology

2.1. FLUXNET 2015 site selection

The FLUXNET 2015 dataset contains open access data (at half-hourly resolution) from 206 globally distributed sites, comprising quality-controlled ecosystem-scale NEE, H, and LE fluxes along with associated meteorological and biological variables (Pastorello et al., 2020). Whilst installed and maintained by different researchers, a uniform flux post-processing procedure was applied to all sites (Pastorello et al., 2017, 2020). We used half-hourly FLUXNET 2015 products: NEE_VUT_REF, NEE_VUT_REF_QC, H_F_MDS, H_F_MDS_QC, LE_F_MDS, and LE_F_MDS_QC (<https://fluxnet.org/data/fluxnet2015-dataset/fullset-data-product/>). Quality control flags (*_QC) were used to identify gap-filled fluxes already present in the datasets. Not all 206 sites were appropriate for validating gap-filling approaches (sites used and their background information are shown in Fig. 1 and Tables S1 and S2), 48 sites did not provide quality control information for H and LE and 86 did not have the required drivers to implement RFR_{10} . In addition, 12 sites did not contain enough original (non-gap-filled) data to accommodate the artificial gaps for validation. Due to these constraints, a sub-set of 94 sites were analyzed for gap-filling for the complete NEE, H, and LE.

2.1.1. Environmental gap filling drivers

We used pre-filled environmental drivers provided by the FLUXNET2015 database. Drivers for MDS and RFR_3 were downward short-wave radiation (SW_IN_F), vapour pressure deficit (VPD_F_MDS), and air temperature (TA_F_MDS). The additional seven drivers for the extended RFR_{10} were net radiation (NETRAD), wind speed (WS), wind direction (WD), soil heat flux (G_F_MDS), soil temperature (TS_F_MDS), relative humidity (RH), and soil water content (SWC_F_MDS).

2.1.2. Site characteristic descriptors

For each site, we extracted descriptors of geographical location, land-use classification, local meteorology, climate classification, and instrumental setup to provide comprehensive information on gap-filling performance analysis (Tables S1 and S2). Descriptors extracted from the FLUXNET site meta-data include continent, altitude, the International

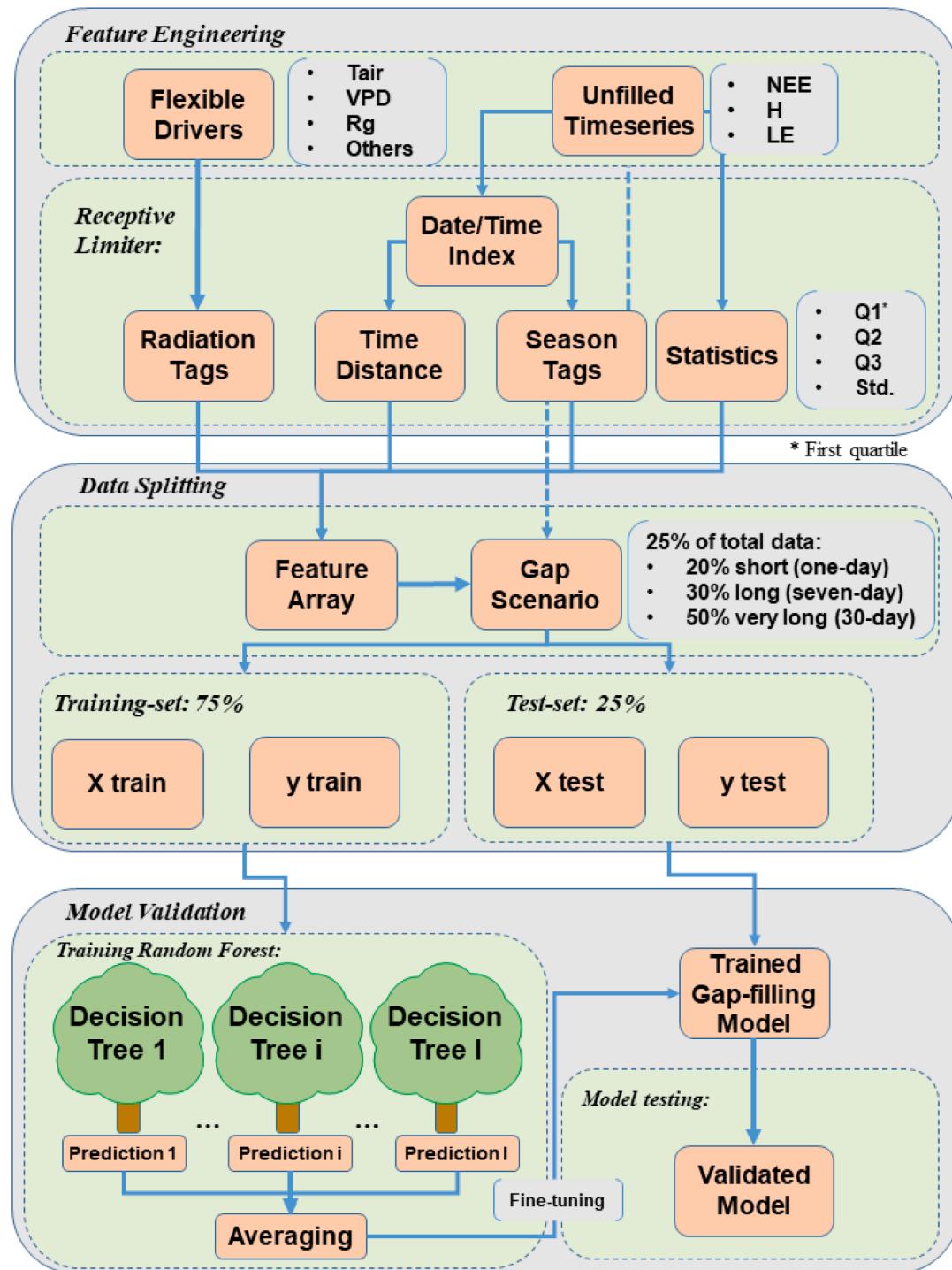


Fig. 2. Workflow of implementing RFR: Feature engineering (top grey panel), data splitting via gap scenario (middle grey panel), and model validation (bottom grey panel).

Geosphere-Biosphere Programme (IGBP), and Koppen's climate classification (Falge et al., 2017; Pastorello et al., 2020). From the FLUXNET2015 database we extracted mean annual temperature ($^{\circ}\text{C}$), precipitation (mm) and wind speed (m s^{-1}). Instrumental setup was classified by sensor type (i.e. open-path, closed-path, or both), instrument-to-canopy height ratio and data set duration. Information on site setup was determined by a literature search of the primary publications for each site.

2.2. Artificial gap scenario

Artificial gaps were generated within the datasets to be filled using three approaches; 25% of total half-hours were randomly removed comprised of three different gap lengths: short gaps (24 h, 20% of total gaps), long gaps (7-day, 30% of total gaps) and very-long gaps (30-day, 50% of total gaps). Differently located random gap scenarios were generated for each site. For each site, NEE, H, and LE shared the same gaps. Where the artificial gaps overlapped with existing 'real' gaps we required at least 50% original measured data be present, if this criterion

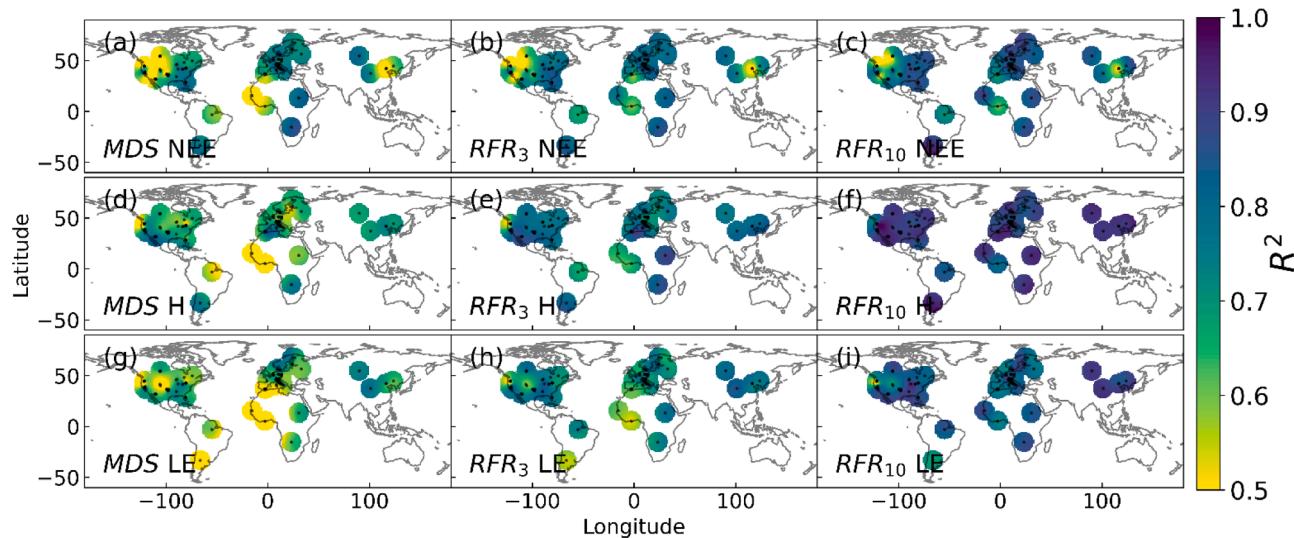


Fig. 3. R^2 map of comparing MDS, RFR₃, and RFR₁₀ filled gaps with measurements for NEE, H, and LE, respectively. It shows spatial distribution of the variance explained by gap-filling (R^2) at 94 FLUXNET2015 sites. (We also provide validation at 194 sites (1346 site-years) covering six continents, 11 IGBP classes, and 18 Koppen climate classes (Table S9)).

was not met, the artificial gap was discarded and randomly re-generated until it meets the ‘>50%’ criterion. Sites with insufficient original measured data to provide the required gap lengths were rejected from the analysis.

2.3. Gap-filling approaches

The benchmark MDS was implemented using the R package REdyProc (v. 1.2.2) (Wutzler et al., 2018), further details on the MDS approach can be found in (Reichstein et al., 2005). Our novel machine learning approach, Random Forest Robust (RFR), was developed using the ‘fluxlib’ package (<https://github.com/soonyenju/fluxlib>) in Python 3.6+, and is based on Random Forest implemented in Scikit-Learn (v. 0.24.1) (Pedregosa et al., 2011) with a new feature selector called ‘receptive limiter’ (details are given in section **Error! Reference source not found.**). Training of the RFR was performed for each site separately. Because our RFR approach contains two distinct driver sets, a total of three methods (MDS, RFR₃ and RFR₁₀) were validated at each site.

The RFR approach has been widely implemented in ecological applications (Breiman, 2001; Jung et al., 2009; Tramontana et al., 2015; Zeng et al., 2020). Our implementation of RFR comprises three steps: feature engineering, data splitting, and model validation (Fig. 2).

The ‘Receptive Limiter’ is the core of feature engineering, continuous data are extracted and binned into discrete categories, downward solar radiation is tagged as, for example, ‘weak’ ($<10 \text{ W m}^{-2}$), ‘medium’ ($10\text{--}100 \text{ W m}^{-2}$), or ‘strong’ ($>100 \text{ W m}^{-2}$). Time distance from the beginning of the time-series (in hours) is extracted as a feature to capture the potential ecosystem growing or degrading trends. Seasons (by the month of time-series) are tagged as ‘winter’ (DJF in North Hemisphere; JJA in South Hemisphere), ‘spring’ (MAM in North Hemisphere; SON in South Hemisphere), ‘summer’ (JJA in North Hemisphere; DJF in South Hemisphere), and ‘autumn’ (SON in North Hemisphere; MAM in South Hemisphere). Daily flux quartiles and standard deviations are extracted from quality-controlled flux time-series as RFR input features separately from NEE, H, and LE to preclude potential outliers in filled gaps. Features and fluxes are split into training and testing data (training-set and test-set). Training data is used to separately feed the RFR. Hyperparameters of RFR are automatically optimized using the GridSearchCV function of Scikit-Learn. The trained RFR models are subsequently validated against the test-set.

2.4. Evaluation indicators

Statistical comparisons between gap-filled and original measured values within the artificial gaps were carried out for NEE, H, and LE at each site using the coefficient of determination (R^2), slope of linear regression, Root Mean Squared Error (RMSE, g C (carbon) $\text{m}^{-2} \text{ d}^{-1}$ for NEE and W m^{-2} for H and LE), and bias (same units as RMSE).

The bias is defined as:

$$\text{bias} = \frac{\sum^F \text{ill.} - \sum^M \text{eas.}}{n}$$

Where *Fill*. denotes the filled gaps *Meas*. denotes the measured fluxes (of corresponding artificial gaps) *n* is the length of gaps measured as the number of half-hours

These descriptive statistics are also determined separately for daytime and night-time periods, where daytime is defined as periods above a threshold of $20 \text{ W m}^{-2} \text{ Rg}$ (Papale et al., 2006).

Welch’s T-test (Derrick et al., 2016) was used to determine gap-filling improvement by RFR₃ over MDS and by RFR₁₀ over RFR₃ separately within the 95% confidence interval.

We use the energy balance ratio (EBR) of the gap-filled periods as an independent measure of gap-filling bias in the energy fluxes (i.e. LE and H) (Foken et al., 2011; Perez-Priego et al., 2017). According to the following formula (Eshonkulov et al., 2019):

$$\text{EBR} = \frac{\sum^{(H+LE)}}{\sum^{(NETRAD-G)}}$$

Where EBR = energy balance ratio NETRAD= ground downward net radiation (W m^{-2}), derived from FLUXNET201 G = ground heat flux (W m^{-2}), derived from FLUXNET2015

3. Results

3.1. Gap-filling performance

In general, North America and Europe comprised the most sites and Europe was seen with the highest R^2 for NEE, H, and LE; while South America and Africa were seen with the lowest for H and LE (Fig. 3). Comparing NEE with H and LE, northwest North America and northeast Asia were seen with low R^2 ; but R^2 for NEE in South America and Africa were relatively higher. As regards to gap-filling approaches, RFR₃ was

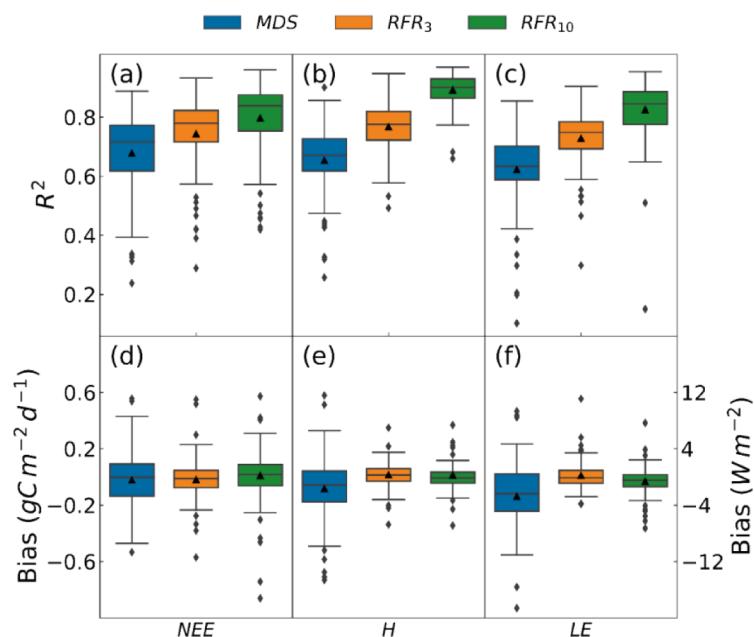


Fig. 4. R^2 and bias boxplots of MDS, RFR₃, and RFR₁₀ gap-filling for NEE (a, d), H (b, e), and LE (c, f), respectively. In this and following boxplots, bars show the third quartile, median and the first quartile as three bars on the boxes in descending order, while the black triangles indicate the mean.

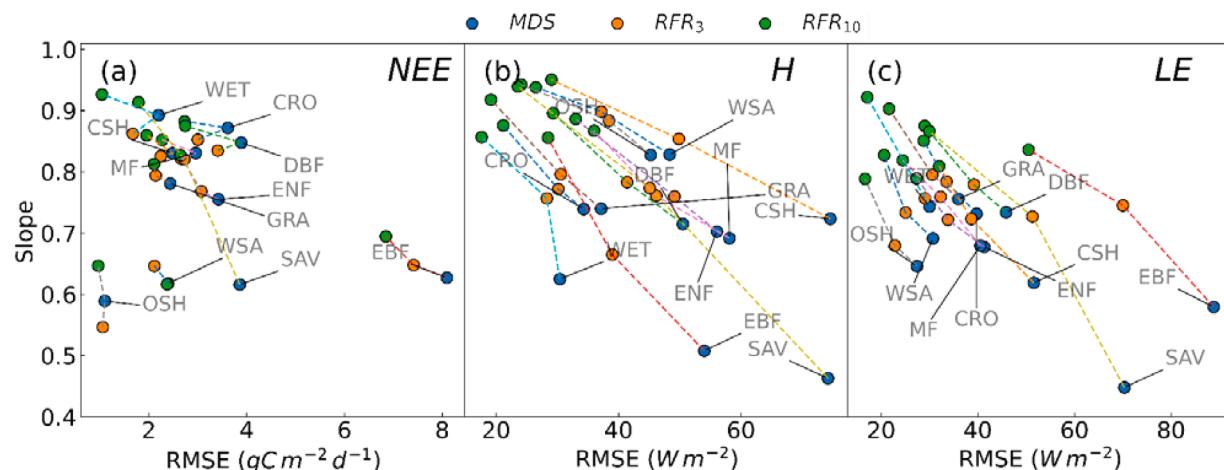


Fig. 5. Scatter plot of gap-filling RMSE against slope. Location of each dot represents the median of metrics for one gap-filling approach (blue for MDS, orange for RFR₃, and green for RFR₁₀) of one IGBP. Dots concentrating on the top-left corner reflect higher values of slope but smaller values of RMSE, vice versa. Dots for the same IGBP are collected by dashed lines (line colors differ by IGBP ecosystem classification). CRO: Croplands, CSH: Closed Shrublands, DBF: Deciduous Broadleaf Forests, EBF: Evergreen Broadleaf Forests, ENF: Evergreen Needleleaf Forests, GRA: Grasslands, MF: Mixed Forests, OSH: Open Shrublands, SAV: Savannas, WET: Permanent Wetlands, WSA: Woody Savannas.

seen with higher R^2 over MDS, and RFR₁₀ was seen with further higher R^2 , especially in South America and Africa.

RFR generally outperformed MDS gap filling for all fluxes with higher R^2 and narrower bias interquartile range (IQR) (Fig. 4). RFR₃ was out performed by RFR₁₀ for R^2 but not for bias, where RFR₃ had a marginally lower bias for LE and NEE (Fig. 4e,f).

Across all three fluxes (NEE, H, LE), median R^2 showed RFR₃ explaining 9%, 16%, and 18% more variance than MDS, respectively, and the RFR₁₀ explaining a further 8%, 16%, and 13%, respectively (Fig. 4a-c and Table S3). **Reference source not found.** More details can be found in Table S4.

Both RFR₃ and RFR₁₀ resulted in similar reductions in the IQR of biases over MDS, nearly 40% for NEE (Fig. 4d) and more than 70% for H and LE (Fig. 4e,f). All methods showed a similar median bias (across all sites) for NEE, ranging from -0.02 to $0.01 \text{ g C m}^{-2} \text{ d}^{-1}$ (Table S3).

Reference source not found.).

Similar pattern of the gap-performance was seen, with RFR₁₀ performing better than RFR₃ and both RFRs performing better than MDS in terms of slope and RMSE for all three fluxes (NEE, LE and H) and all ecosystems (Fig. 5). RFR₃ increased the slope by 5% over MDS, with RFR₁₀ nearly doubling this to 11%. Meanwhile, RFR₃ reduced the RMSE by 17% compared to MDS and RFR₁₀ reduced RMSE 21% compared to MDS (Table S4).

The improvements in gap-filling slope and RMSE brought by RFR methods were larger for H (Fig. 5b) and LE (Fig. 5c) than for NEE (Fig. 5a). The improvement of RFR₁₀ was particularly evident for H and LE in ecosystems that MDS (and even RFR₃) struggle with (e.g., SAV and EBF, Fig. 5b,c). Compared to MDS, the slope for RFR methods increased 3% for NEE, 16% for H, and 15% for LE; corresponding RMSE decreased 15% for NEE, 34% for H, and 26% for LE (Table S4).

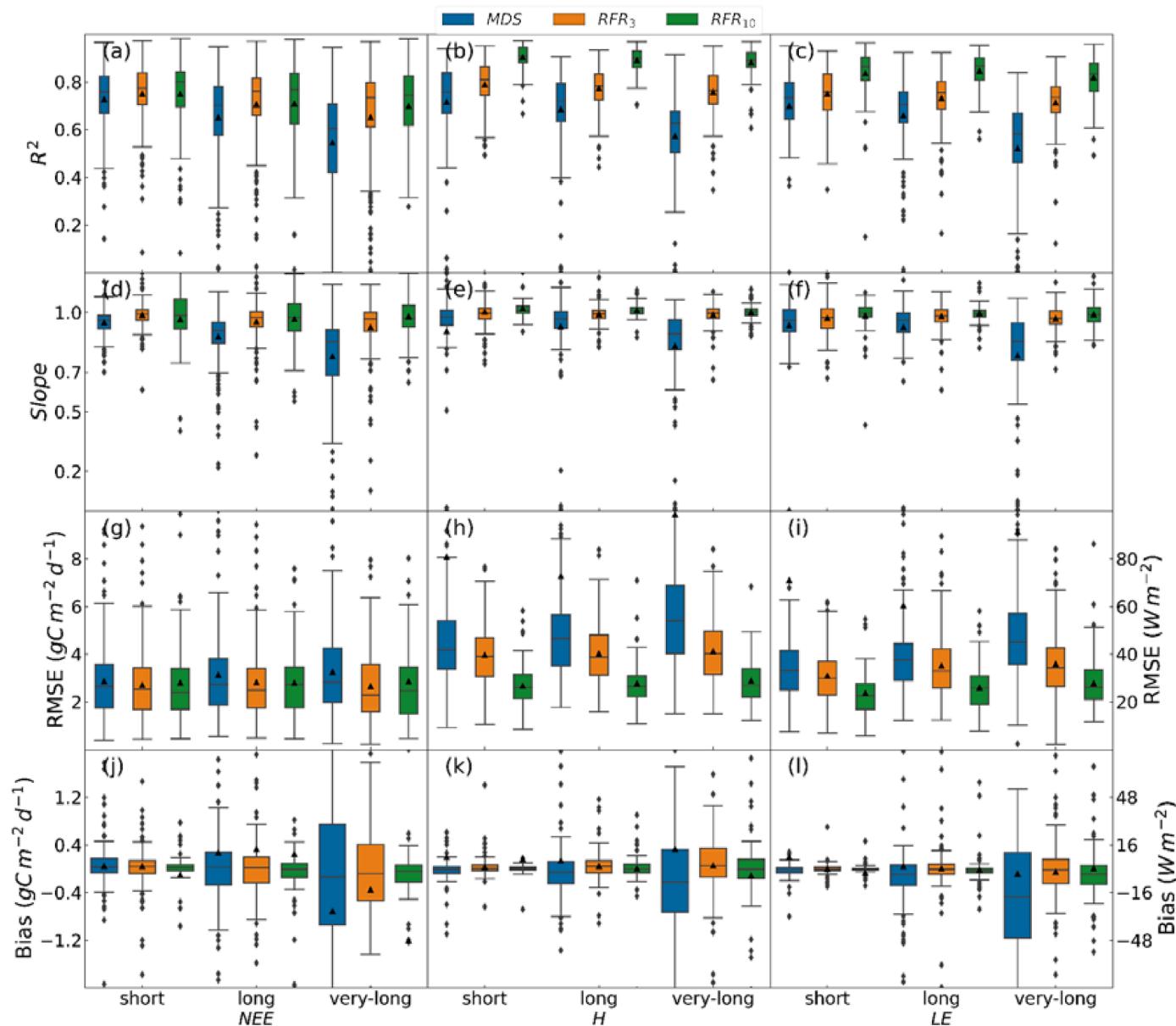


Fig. 6. Boxplots showing gap-filling performance of three methods in short, long, and very-long gaps of same sites from the combined artificial gap scenario. The figures shows the performance in terms of R^2 , slope, RMSE, and bias for NEE (a, d, g, and j), H (b, e, h, and k), and LE (c, f, I, and l), of R^2 (a–c), linear slope (d–f), RMSE (g–i), and bias (j–l).

3.2. Sensitivity of gap-filling to gap length

Considering gap length scenarios separately, the RFR methods showed greater resilience to longer gaps compared to MDS (Fig. 6). R^2 (Fig. 6a–c) and slope (Fig. 6d–f) of RFR₁₀ were higher than RFR₃ and further higher than MDS in short, long, and very-long gaps; while RMSE (Fig. 6g–i) and IQR of bias (Fig. 6j–l) of RFR₁₀ were smaller than RFR₃ and further smaller than MDS in short, long, and very-long gaps. More details can be found in Tables S3 and S7.

All four statistical measures of the RFR methods were less sensitive to gap-length than MDS (Fig. 6 and Table S3). For example, as gap length increased from short (1-day) to very-long (30-day), R^2 on average for the three fluxes decreased by 21% (MDS), 4% (RFR₃), and 4% (RFR₁₀); gap-filling uncertainty in terms of bias interquartile range increased by 44% (MDS), 42% (RFR₃), and 6% (RFR₁₀). In addition, RFR methods for H and LE showed higher accuracy in filling longer gaps than for MDS (e.g., higher mean R^2 and narrower R^2 IQR, Fig. 6a–c).

Using filled artificial gaps (H and LE) and their measured

counterparts, RFR methods (in particular RFR₁₀) exhibited energy balance ratios closer to those calculated for the corresponding original measurements than did MDS (Fig. 7). The averaged EBR was separately 80% (measured), 80% (RFR₁₀), 78% (RFR₃), and 73% (MDS). In regard to ecosystem types, overall EBR of croplands were smaller than other ecosystems. It was seen in all ecosystem types that RFR₁₀ EBR was closer to measured values than RFR₃ and even closer to the measured values than MDS, such discrepancy in EBR between MDS and RFR methods was the largest in SAV (Fig. 7).

3.3. Limitations of gap-filling approaches

Gap-filling performance, in terms of R^2 , in the daytime was much better than at night (Fig. 8). NEE, for example, median nighttime R^2 decreased compared to daytime by 80% (MDS), 70% (RFR₃) and 85% (RFR₁₀). It can be seen that the difference between daytime and nighttime gap-filling R^2 for H (Fig. 8d) and LE (Fig. 8f) was larger than for NEE (Fig. 8b). Bias in the daytime NEE was larger than at night (Fig. 8

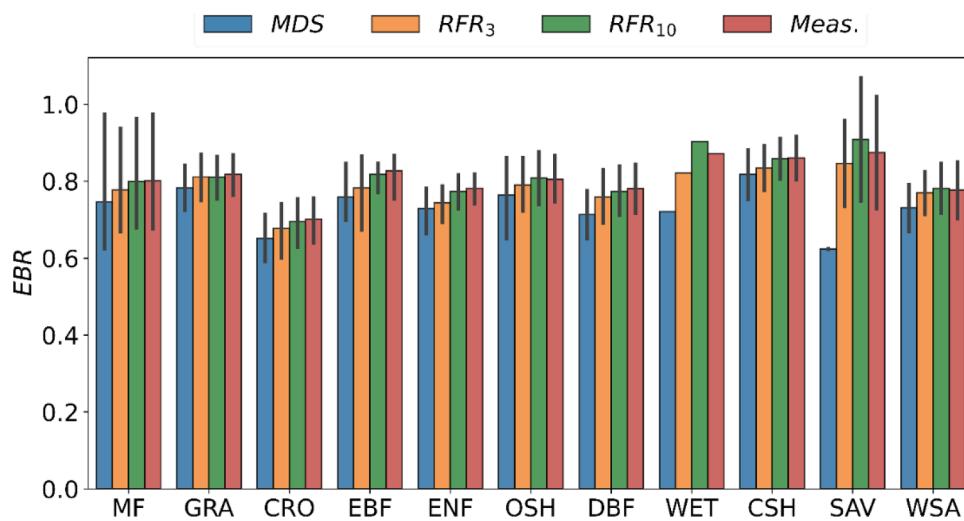


Fig. 7. Means (bars) and standard deviations (black vertical lines) of energy balance ratio (EBR) for filled artificial gaps and corresponding measurements (Meas.).

Error! Reference source not found.a, however no consistent pattern was observed for H ([Fig. 8 Error! Reference source not found.c](#)) and LE ([Fig. 8 Error! Reference source not found.e](#)). More details can be found in Tables S5 and S6.

Performance of the gap-filling routines varied by IGBP ecosystem landcover classification. Evergreen broadleaf forest (EBF) was seen with the lowest R^2 and large nocturnal bias for NEE ([Fig. 8 Error! Reference source not found.a and b](#)), H ([Fig. 8 Error! Reference source not found.d](#)), and LE ([Fig. 8 Error! Reference source not found.e](#)). Savannah (SAV) showed large nocturnal biases for all three fluxes.

4. Discussion

4.1. Global gap-filling performance and intercomparison between approaches in different landcover classifications

This work follows the earlier gap-filling study of NEE by [Moffat et al. \(2007\)](#), as well as H and LE ([Vitale et al., 2019](#)), long gap uncertainty study ([Richardson and Hollinger, 2007](#)), and recent studies regarding high-performance of machine-learning on methane gap-filling ([Kim et al., 2020; Irvin et al., 2021](#)). We updated and integrated previous analyzes by applying machine-learning approaches with modifications to fill very long gaps in NEE (CO_2), H, and LE fluxes and greatly extended the geographical range of test sites. Our results showed a consistent improvement in gap-filling using RFR compared to MDS for all the 94 global sites that were suitable for our complete analysis (See methods). This improvement was seen for all three fluxes (NEE, LE and H), with the greatest improvements for H and LE. For longer gaps, usually resulting from system failure ([Richardson and Hollinger, 2007](#)), the improvement on gap-filling by RFR could be considerable ([Fig. 6](#) and Table S3**Error! Reference source not found.**), which supports the recommendation for RFR given in [Kim et al. \(2020\)](#).

In agreement with previous studies, MDS showed satisfactory gap-filling performance in most cases ([Fig. 6](#) and Table S3**Error! Reference source not found.**) because individual gap-lengths are normally shorter than 1.5 days ([Moffat et al., 2007](#)). RFR methods improved the gap-filling accuracy (e.g. 15% R^2 increase by using RFR₃ and 30% R^2 increase by using RFR₁₀) while reducing uncertainties (e.g. interquartile range of bias decreased by 70%) for NEE, H, and LE globally ([Fig. 3](#) and Table S3**Error! Reference source not found.**) and statistically significantly (Table S10) for most of sites (Tables S4–S6). Such improvement can be attributed to the complex architecture of random forest and the “receptive-limiter” approach used in this study. The benefit of the receptor-limiter we used can be seen by comparing random forest gap-filling performance with and without the “receptive-limiter”

([Fig. S1](#)).

The improvement for H and LE on gap-filling by using RFR was much larger than for NEE compared with MDS ([Fig. 4](#)). Currently, studies of gap-filling focused on H or LE are fewer than NEE at a global scale ([Foltynová et al., 2020](#)), resulting in a knowledge gap around these energy fluxes. Reliable gap-filling methods (for H and LE) like RFR can help address this knowledge gap and will help to inform debates around the environmental impacts (positive or negative) of nature-based solutions and the mitigation of global climate change ([Stenzel et al., 2018](#)).

Using the extended driver set in RFR₁₀ showed advantages in gap-filling for R^2 , slope, and RMSE, but the uncertainty also increased in some circumstances. Where the focus was solely on annual sums – especially when only shorter gaps exist – RFR₃ produced the smallest range in biases. The advantages of using extended drivers (RFR₁₀) became more apparent under the more challenging gap scenarios (i.e. longer gaps and night-time).

Our analysis has shown a large variation in gap-filling performance for different ecosystems. RFR indeed improved gap-filling performance, but it still struggled with NEE, H, and LE for savannah (SAV), evergreen broadleaf forest (EBF), and open shrubland (OSH) ([Fig. 5](#)) and geographically in Africa, South America, and northwest North America ([Fig. 3](#)). The reason causing the poor gap-filling performance for ecosystems like EBF and ecosystems like SAV may be different. Inferred by the small RMSE and slope, the poor performance in SAV could be accounted by the weak flux signal there ([Fig. 5a](#)). In contrast, the RMSE was large while the slope was small for EBF ([Fig. 5a](#)), which indicates the fluxes there could be large. The poor gap-filling performance for EBF could be caused by the subtle seasonality, e.g. in Brazil, that does not correlate with photosynthetically active radiation ([Restrepo-Coupe et al., 2013](#)). Given the large improvement of using extended drivers, one possible solution in the future could be introducing other environmental drivers, like leaf area index and/or satellite-based vegetation index, as suggested by ([Kang et al., 2019](#)).

4.2. Gap-filling longer gaps and uncertainty analyzes

The performance of MDS reduced significantly for very-long gaps, whereas RFR continued to operate with similar statistical performance. Within our 94 selected sites (which are biased towards complete datasets) MDS failed to gap-fill 5.47% NEE half-hours from 19.50% sites, 0.30% H half-hours from 13.07% sites and 0.35% LE half-hours from 13.73 sites. Crucially for NCS, RFR did a better job at maintaining gap-filling performance for longer data gaps, for example, R^2 of MDS in filling very-long gaps decreased by >15%, but the decrease for RFR methods was less than 5% ([Fig. 6, Error! Reference source not found.](#)

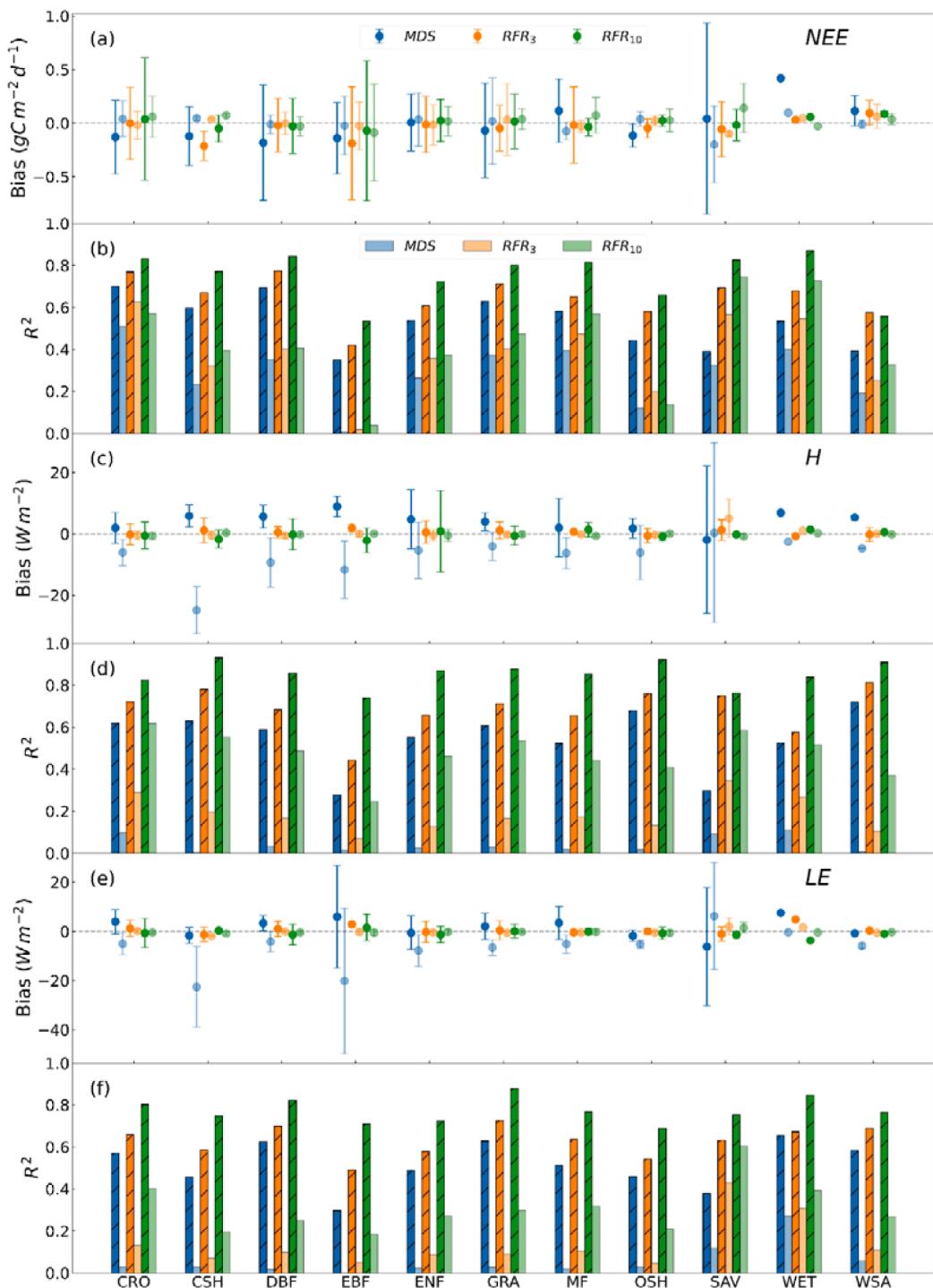


Fig. 8. Day and night Gap-filling median bias (with error bars) and R^2 grouped by IGBP for NEE (a and b), H (c and d), and LE (e and f). The solid dots and bars are for daytime gap-filling, while the lighter dots and bars are for night-time gap-filling.

Fig. S2, and Table S3Error! Reference source not found.).

Whilst both RFR methods outperformed MDS for long gaps, the performance of RFR₁₀ was significantly better than RFR₃ (Fig. 6). Where drivers are available RFR₁₀ should be considered over RFR₃ or MDS for sites with data gaps that exceed a few days in length. It is worth noting however that the average ratio of gap to data in the Fluxnet2015 (at the half hour resolution) is 67.53% (i.e. on average datasets are missing 67.53% of their total half hours) and that of this 67.53%, 97.1% are short gaps, 2.77% are long gaps and 0.13% are very long gaps. Similarly, the real gap ratio for H is 39.77%, and 98.60% are short gaps, only

1.20% are long gaps and 0.20% are very-long gaps; the real gap ratio for LE is 44.99%, and 98.87% are short gaps, only 0.99% are long gaps and 0.14% are very-long gaps. It might be suggested, however, that the data present in FLUXNET are likely to represent ‘best-case’ data with contributions from better-maintained sites, it is likely that gap scenarios may be more challenging at many other sites.

As an independent verification, the energy balance ratio (EBR) of 94 sites was 80% (using measured H and LE), 80% (using RFR₁₀ gap-filled H and LE), 78% (using RFR₃ gap-filled H and LE), and 73% (using MDS gap-filled H and LE); also suggesting the application of RFR methods can

be reliable in gap-filling energy fluxes. In this case, flux time-series gap-filled by using RFR methods can be beneficial to climate models and/to support satellite remote sensing validations.

4.3. Implications of gap-filling performance for cumulative fluxes

In terms of gap-filling uncertainty, the mean global carbon sequestration rate is approximate $17.5 \text{ g C m}^{-2} \text{ yr}^{-1}$ for terrestrial ecosystems (Levin, 2001; Griscom et al., 2017), and a week-long gap would result in an additional uncertainty of $30 \text{ g C m}^{-2} \text{ yr}^{-1}$ in the worst cases (Richardson and Hollinger, 2007). Our findings suggest lower overall uncertainties, the bias interquartile range across 94 sites equated to an annual bias of $84 \text{ g C m}^{-2} \text{ yr}^{-1}$ (MDS), $45 \text{ g C m}^{-2} \text{ yr}^{-1}$ (RFR₃), and $55 \text{ g C m}^{-2} \text{ yr}^{-1}$ (RFR₁₀) (Table S3Error! Reference source not found.), that is comparable to Richardson and Hollinger (2007). This reduction in NEE uncertainty by using RFR could be very valuable to near carbon neutral ecosystems (Soloway et al., 2017). RFR methods also reduced uncertainty for H and LE to $<2 \text{ W m}^{-2}$ from 5 W m^{-2} of MDS, and the improvement was good compared with $> 3 \text{ W}^{-2}$ at most sites (Vitale et al., 2019). This reduction in uncertainty seen using RFR could play an important role in accurately estimating global evapotranspiration. Therefore, RFR methods, especially the RFR₃, are suggested with great potential in remote NCS applications where longer gaps can occur more easily due to instrument failure. In remote areas, EC system maintenance in a regular and frequent manner becomes difficult, as NCS applications aim to be low-cost.

4.4. Limitations of this study

RFR performed reliably in our study scenarios of gap lengths up to one month, but we might expect performance to drop off substantially as gap lengths increase beyond this. We did not test longer gaps due to the reduction in the numbers of FLUXNET sites that could be included in this analysis but could usefully be the focus in a future study. Furthermore, as with other comparisons studies such Moffat et al. (2007), we did not consider non-randomly located gaps in this study, for example, gaps created due to regular maintenance schedules, or perhaps routine harvesting operations in agricultural systems. Devising data gap probabilities based on potential environmental and management challenges that were realistic across all 94 sites would be extremely challenging. However, we suggest that focused studies looking at gap-filling performance for non-random gaps could be an important focus for later studies.

The performance of gap-filling methods has been observed to be better during daytime than night-time Moffat et al. (2007). Whilst our present study, RFR₁₀ performed slightly better than RFR₃, and both improved on MDS, in capturing the diurnal patterns of NEE, the gap-filling performance at night remains poor compared to daytime (e.g. $R^2 < 0.6$ in many ecosystems). One reason is the low friction velocity at night, up-to 70% of data can be rejected at night due to stable atmospheric conditions etc.(Aubinet et al., 2012) and lower magnitude of nocturnal fluxes. In addition, gap-filling at night is challenging because the shortwave solar radiation (vital to gap-filling) vanishes (Reichstein et al., 2005).

5. Conclusion

In this study, a robust gap-filling approach (i.e. RFR) is proposed for filling long gaps in NEE, H, and LE fluxes. Validated against MDS globally with gap sizes ranging from 1 to 30 days, we found that RFR methods improve the gap-filling performance particularly for H and LE and extended drivers are beneficial to gap-filling performance (i.e. RFR₁₀ outperforms RFR₃). RFR₃ and RFR₁₀ separately improves gap-filling accuracy by 15% and 30% while reduces uncertainty by 70%. Unlike MDS, RFR methods maintain performance with gap-lengths up to one month. Compared with filling 1-day long gaps, the gap-filling performance (in terms of R^2) of filling 30-day long gaps degrades by 21% for

MDS and degrades by <4% for RFR methods. No obvious difference is found between RFR₃ and RFR₁₀ performance degradation. In addition, RFR methods, in particular the RFR₁₀ largely reduces the uncertainty in filling 30-day long gaps, its uncertainty is less than 1/3 of MDS. Three challenges are to be addressed in the future for better applying RFR gap-filling to eddy covariance for natural climate solutions: (1) the difficulties of gap-filling at night which is a lasting challenge to eddy covariance requires further research, (2) the still poor performance for certain ecosystems (i.e. evergreen broadleaf forest, savannah, and open shrubland) that might be addressed by introducing extra environmental drivers, (3) the question of gap-filling performance for even longer gaps and non-random gaps that will be considered in our future studies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. We would also like to provide the source code hosted at <https://github.com/soonyenju/fluxlib/tree/master/fluxlib> for public access.

Acknowledgments

The authors thank the FLUXNET and the research groups for providing the CC-BY-4.0 (Tier one) open-access eddy covariance data (https://fluxnet.org/login/?redirect_to=/data/download-data/). They also thank the ReddyProc (https://cran.r-project.org/web/packages/R_EddyProc/index.html) team and scikit-learn (<https://scikit-learn.org/stable/install.html>) team for the packages that help the implementation and validation for gap-filling approaches. Songyan Zhu would like to acknowledge a Shell funded PhD studentship and Timothy Hill acknowledge funding from a joint UK NERC-FAPESP grant no. NE/S000011/1 & FAPESP-19/07773-1.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.agrformet.2021.108777](https://doi.org/10.1016/j.agrformet.2021.108777).

Reference

- Aubinet, M., Vesala, T., Papale, D., 2012. Eddy covariance: a Practical Guide to Measurement and Data Analysis. Springer Science & Business Media.
- Baldocchi, D., 2014. Measuring fluxes of trace gases and energy between ecosystems and the atmosphere-the state and future of the eddy covariance method. *Glob. Chang. Biol.* 20, 3600–3609.
- Baldocchi, D., Falge, E., Gu, L., et al., 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Am. Meteorol. Soc.* 82, 2415–2434.
- Baldocchi, D.D., 2020. How eddy covariance flux measurements have contributed to our understanding of global change biology. *Glob. Chang. Biol.* 26, 242–260.
- Bautista, N., Marino, B.D., Munger, J.W., 2021. Science to commerce: a commercial-scale protocol for carbon trading applied to a 28-year record of forest carbon monitoring at the Harvard forest. *Land* 10, 163 (Basel).
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Campioli, M., Malhi, Y., Vicca, S., et al., 2016. Evaluating the convergence between eddy-covariance and biometric methods for assessing carbon budgets of forests. *Nat. Commun.* 7, 13717. <https://doi.org/10.1038/ncomms13717>.
- Chausson, A., Turner, B., Seddon, D., et al., 2020. Mapping the effectiveness of nature-based solutions for climate change adaptation. *Glob. Chang. Biol.* 26, 6134–6155. <https://doi.org/10.1111/gcb.15310>.
- Cohen-Shacham, E., Andrade, A., Dalton, J., et al., 2019. Core principles for successfully implementing and upscaling Nature-based Solutions. *Environ. Sci. Policy* 98, 20–29. <https://doi.org/10.1016/j.envsci.2019.04.014>.
- Derrick, B., Toher, D., White, P., 2016. Why Welch's test is type I error robust. *Quant. Methods Psychol.* 12 (1).
- Eshonkulov, R., Poyda, A., Ingwersen, J., et al., 2019. Evaluating multi-year, multi-site data on the energy balance closure of eddy-covariance flux measurements at cropland sites in southwestern Germany. *Biogeosciences* 16, 521–540. <https://doi.org/10.5194/bg-16-521-2019>.
- Falge, E., Baldocchi, D., Olson, R., et al., 2001. Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agric. For. Meteorol.* 107, 43–69.

- Falge, E., et al. "FLUXNET research network site characteristics, investigators, and bibliography, 2016." ORNL DAAC (2017).
- Foken, T., Aubinet, M., Finnigan, J.J., et al., 2011. Results of a panel discussion about the energy balance closure correction for trace gases. *Bull. Am. Meteorol. Soc.* 92, 13–18. <https://doi.org/10.1175/2011BAMS3130.1>.
- Foltynová, L., Fischer, M., McGloin, R.P., 2020. Recommendations for gap-filling eddy covariance latent heat flux measurements using marginal distribution sampling. *Theor. Appl. Climatol.* 139, 677–688. <https://doi.org/10.1007/s00704-019-02975-w>.
- Friedlingstein, P., O'Sullivan, M., Jones, M.W., et al., 2020. Global carbon budget 2020. *Earth Syst. Sci. Data* 12, 3269–3340. <https://doi.org/10.5194/essd-12-3269-2020>.
- Glanemann, N., Willner, S.N., Levermann, A., 2020. Paris climate agreement passes the cost-benefit test. *Nat. Commun.* 11, 1–11. <https://doi.org/10.1038/s41467-019-13961-1>.
- Griscom, B.W., Adams, J., Ellis, P.W., et al., 2017. Natural climate solutions. *Proc. Natl. Acad. Sci.* 114, 11645–11650.
- Hemes, K.S., Runkle, B.R.K., Novick, K.A., et al., 2021. An ecosystem-scale flux measurement strategy to assess natural climate solutions. *Environ. Sci. Technol.* 55, 3494–3504. <https://doi.org/10.1021/acs.est.0c06421>.
- Hill, T., Chocholiek, M., Clement, R., 2017. The case for increasing the statistical power of eddy covariance ecosystem studies: why, where and how? *Glob. Chang. Biol.* 23, 2154–2165.
- Irvin, J., Zhou, S., McNicol, G., et al., 2021. Gap-filling eddy covariance methane fluxes: comparison of machine learning model predictions and uncertainties at FLUXNET-CH₄ wetlands. *Agric. For. Meteorol.* (108528), 308–309. <https://doi.org/10.1016/j.agrformet.2021.108528>.
- Jung, M., Reichstein, M., Bondeau, A., 2009. Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences* 6, 2001–2013.
- Kang, M., Ichii, K., Kim, J., et al., 2019. New gap-filling strategies for long-period flux data gaps using a data-driven approach. *Atmosphere* 10, 568 (Basel).
- Keith, H., Vardon, M., Obst, C., et al., 2021. Evaluating nature-based solutions for climate mitigation and conservation requires comprehensive carbon accounting. *Sci. Total Environ.* 769, 144341. <https://doi.org/10.1016/j.scitotenv.2020.144341>.
- Kim, Y., Johnson, M.S., Knox, S.H., et al., 2020. Gap-filling approaches for eddy covariance methane fluxes: a comparison of three machine learning algorithms and a traditional method with principal component analysis. *Glob. Chang. Biol.* 26, 1499–1518.
- Levin, S., 2001. Encyclopedia of biodiversity. Academic Press, San Diego.
- Mauder, M., Foken, T., Clement, R., et al., 2008. Quality control of CarboEurope flux data-part 2: inter-comparison of eddy-covariance software. *Biogeosciences* 5, 451–462.
- Menzer, O., Meiring, W., Kyriakidis, P.C., McFadden, J.P., 2015. Annual sums of carbon dioxide exchange over a heterogeneous urban landscape through machine learning based gap-filling. *Atmos. Environ.* 101, 312–327. <https://doi.org/10.1016/j.atmosenv.2014.11.006>.
- Moffat, A.M., Papale, D., Reichstein, M., et al., 2007. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agric. For. Meteorol.* 147, 209–232.
- Papale, D., Reichstein, M., Aubinet, M., et al., 2006. Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences* 3, 571–583.
- Pastorello, G., Papale, D., Chu, H., et al., 2017. A new data set to keep a sharper eye on land-air exchanges. *Eos Trans. Am. Geophys. Union* 98, Online.
- Pastorello, G., Trott, C., Canfora, E., et al., 2020. The FLUXNET2015 dataset and the ONEflux processing pipeline for eddy covariance data. *Sci. Data* 7, 1–27.
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al., 2011. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Perez-Priego, O., El-Madany, T.S., Migliavaca, M., et al., 2017. Evaluation of eddy covariance latent heat fluxes with independent lysimeter and sapflow estimates in a Mediterranean Savannah ecosystem. *Agric. For. Meteorol.* 236, 87–99. <https://doi.org/10.1016/j.agrformet.2017.01.009>.
- Reichstein, M., Falge, E., Baldocchi, D., et al., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Glob. Chang. Biol.* 11, 1424–1439.
- Restrepo-Coupe, N., Da Rocha, H.R., Hutyra, L.R., et al., 2013. What drives the seasonality of photosynthesis across the Amazon basin? A cross-site analysis of eddy flux tower measurements from the Brasil flux network. *Agric. For. Meteorol.* 182, 128–144.
- Richardson, A.D., Hollinger, D.Y., 2007. A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps in the CO₂ flux record. *Agric. For. Meteorol.* 147, 199–208. <https://doi.org/10.1016/j.agrformet.2007.06.004>.
- Skinner, R.H., Dell, C.J., 2015. Comparing pasture C sequestration estimates from eddy covariance and soil cores. *Agric. Ecosyst. Environ.* 199, 52–57.
- Smith, P., Beaumont, L., Bernacchi, C.J., Byrne, M., Cheung, W., Conant, R.T., et al., 2021. Essential outcomes for COP26. *Glob. Chang. Biol.* <https://doi.org/10.1111/gcb.15926>.
- Smith, P., Soussana, J.F., Angers, D., et al., 2020. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. *Glob. Chang. Biol.* 26, 219–241.
- Soloway, A.D., Amiro, B.D., Dunn, A.L., Wofsy, S.C., 2017. Carbon neutral or a sink? Uncertainty caused by gap-filling long-term flux measurements for an old-growth boreal black spruce forest. *Agric. For. Meteorol.* 233, 110–121. <https://doi.org/10.1016/j.agrformet.2016.11.005>.
- Stenzel, F., Greve, P., Tramberend, S., 2018. Increase water stress more than climate change. *Nat. Commun.* 1–9. <https://doi.org/10.1038/s41467-021-21640-3>.
- Tramontana, G., Ichii, K., Camps-Valls, G., et al., 2015. Uncertainty analysis of gross primary production upscaling using random forests, remote sensing and eddy covariance data. *Remote Sens. Environ.* 168, 360–373.
- Vitale, D., Bilancia, M., Papale, D., 2019. A multiple imputation strategy for eddy covariance data. *J. Environ. Inform.* 34, 68–87.
- Wollenberg, E., Richards, M., Smith, P., et al., 2016. Reducing emissions from agriculture to meet the 2 °C target. *Glob. Chang. Biol.* 22, 3859–3864. <https://doi.org/10.1111/gcb.13340>.
- Wutzler, T., Lucas-Moffat, A., Migliavaca, M., et al., 2018. Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences* 15, 5015–5030.
- Zeng, J., Matsunaga, T., Tan, Z.-H., et al., 2020. Global terrestrial carbon fluxes of 1999–2019 estimated by upscaling eddy covariance data with a random forest. *Sci. Data* 7, 1–11.