Agricultural and Forest Meteorology 332 (2023) 109365

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Gap-filling carbon dioxide, water, energy, and methane fluxes in challenging ecosystems: Comparing between methods, drivers, and gap-lengths

Songyan Zhu [a,*], Jon McCalmont [b], Laura M. Cardenas [c], Andrew M. Cunliffe [a], Louise Olde [c], Caroline Signori-Müller [a], Marcy E. Litvak [d], Timothy Hill [a]

[a] Department of Geography, Faculty of Environment, Science and Economy, University of Exeter, Streatham Campus, Rennes Drive, Exeter EX4 4RJ, United Kingdom
[b] King's College, School of Biological Sciences, University of Aberdeen, Aberdeen AB24 3FX, United Kingdom
[c] Rothamsted Research, Net Zero and Resilient Farming, North Wyke, Devon EX20 2SB, United Kingdom
[d] Department of Biology, University of New Mexico, Albuquerque, NM, United States

ARTICLE INFO

Keywords:
Eddy covariance
Gap-filling
Managed & low-flux ecosystems
ERA5 drivers

ABSTRACT

Eddy covariance serves as one the most effective techniques for long-term monitoring of ecosystem fluxes, however long-term data integrations rely on complete timeseries, meaning that any gaps due to missing data must be reliably filled. To date, many gap-filling approaches have been proposed and extensively evaluated for mature and/or less actively managed ecosystems. Random forest regression (RFR) has been shown to be stable and perform better in these systems than alternative approaches, particularly when filling longer gaps. However, the performance of RFR gap filling remains less certain in more challenging ecosystems, e.g., actively managed agri-ecosystems and following recent land-use change due to management disturbances, ecosystems with relatively low fluxes due to low signal to noise ratios, or for trace gases other than carbon dioxide (e.g., methane).

In an extension to earlier work on gap filling global carbon dioxide, water, and energy fluxes, we assess the RFR approach for gap filling methane fluxes globally. We then investigate a range of gap-filling methodologies for carbon dioxide, water, energy, and methane fluxes in challenging ecosystems, including European managed pastures, Southeast Asian converted peatlands, and North American drylands.

Our findings indicate that RFR is a competent alternative to existing research standard gap-filling algorithms. The marginal distribution sampling (MDS) is still suggested for filling short (< 12 days) gaps in carbon dioxide fluxes, but RFR is better for filling longer (> 30 days) gaps in carbon dioxide fluxes and also for gap filling other fluxes (e.g. sensible heat, latent energy and methane). In addition, using RFR with globally available reanalysis environmental drivers is effective when measured drivers are unavailable. Crucially, RFR was able to reliably fill cumulative fluxes for gaps > 3 moths and, unlike other common approaches, key environment-flux responses were preserved in the gap-filled data.

1. Introduction

The eddy covariance (EC) technique measures the net exchange of mass and energy between the land surface and the atmosphere, and eddy covariance observational networks (e.g., FLUXNET) have expanded monitoring efforts of carbon, water and energy cycles and helped standardise and distribute flux data, (Baldocchi 2020). In recent years, eddy covariance applications have been extended to measure fluxes of other greenhouse gases [e.g., methane] (Eugster and Plüss 2010; Saunois et al., 2016). However, just as with CO_2, the completeness of these flux time series is limited by instrumental failures and data quality issues that result in missing data 'gaps'.

Many 'gap-filling' approaches have been applied to model missing values based on the existing data (Reichstein et al., 2005; Moffat et al., 2007; Kim et al., 2020; Zhu et al., 2022). These gap-filling techniques range from process-based models, e.g., biosphere energy-transfer hydrology model (Knorr and Kattge 2005), to empirical models such as non-linear regression and artificial neural networks [ANN] (Braswell et al., 2005; Noormets et al., 2007) after the pioneering studies (Papale and Valentini 2003; Reichstein et al., 2005). Marginal distribution

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

https://doi.org/10.1016/j.agrformet.2023.109365
Received 6 September 2022; Received in revised form 16 December 2022; Accepted 7 February 2023
Available online 24 February 2023

sampling [MDS] (Reichstein et al., 2005; Moffat et al., 2007) and ANN (Delwiche et al., 2021; Mahabbati et al., 2021) are widely adopted as research-standard gap-filling approaches and remain the benchmark for comparison of novel approaches. Recently proposed machine learning-based methods – e.g., random forest regression (RFR) – exhibited close or even better gap-filling performance than MDS and ANN (Kim et al., 2020; Zhu et al., 2022). However, an understanding of the reliability of eddy covariance gap-filling algorithms is still incomplete. For example, Moffat et al. (2010) and Albert et al. (2017) separately analysed the net carbon flux responses to photosynthetic radiation and air temperature. It remains unknown if these gap-filling approaches can preserve other flux-environment responses – e.g., the carbon flux to water table depth response in McCalmont et al. (2021) – that are also crucial to investigating the biosphere-atmosphere interactions.

First, the flux of primary interest in most cases is carbon dioxide (CO_2), which has seen research effort concentrated more on this gas than other fluxes. For example, studies on filling methane flux gaps (> two months particularly) are still uncommon (Delwiche et al., 2021) as methane flux-measuring instruments only became more practical in the 2010s (McDermitt et al., 2011). Methane flux gap-filling is typically more challenging than CO_2 fluxes, due to high variability and responses to multiple environmental controls: soil temperature and seasonality were the most important drivers for wetlands (Irvin et al., 2021) but water table depth is also important in cases where water table fluctuations were substantial (Kim et al., 2020). Previous studies mainly focused on the local scale (Hommeltenberg et al., 2014; Morin et al., 2014) or a certain types of ecosystem (Dengel et al., 2013; Irvin et al., 2021). Considering the recently released global methane flux database (Delwiche et al., 2021), a global multi-ecosystem study is thereby possible and will benefit our understanding of methane flux-environment interactions.

Second, most early eddy covariance towers were installed in productive and/or less-disturbed natural ecosystems, and this sampling distribution poses challenges to the development of gap-filling (Moffat et al., 2007; Irvin et al., 2021; Zhu et al., 2022). Flux gap-filling for other types of ecosystems can be challenging, these include managed ecosystems and ecosystems with flux rates close to zero (Lucas-Moffat et al., 2018; McKenzie et al., 2021; Yao et al., 2021a). In managed ecosystems, for example, agricultural activities, can substantially alter flux temporal dynamics (McCalmont et al., 2021; Cardenas et al., 2022), quantifying the frequency and intensity of management activities can be challenging for training a machine learning model to gap-fill these timeseries. For low-flux ecosystems – e.g., drylands that comprise around 40% of the global land surface (Huang et al., 2016; Cunliffe et al., 2022) – the low signal-to-noise ratio makes gap filling challenging.

In addition, it is valuable to consider the possibility of gap-filling only with drivers derived from publicly available meteorology reanalyses datasets. The high financial cost of an eddy covariance system and the cost of redundant meteorological measurements are a significant limiting factor in the extension of flux monitoring networks (Hill et al., 2017), resulting in an incomplete picture of global scale ecosystem carbon cycling (Schimel et al., 2015). In this case, gap-filling with open meteorological reanalysis data could help promote the application of eddy covariance globally. Furthermore, the use of meteorological reanalysis data may help improve flux estimates in regions where redundancy in flux and meteorological measurements is not available or is very sparse (Xiao et al., 2012).

In this study we explore the effectiveness of gap-filling techniques for net ecosystem exchange (NEE, i.e., CO_2 flux), sensible heat (H), latent energy (LE), and methane flux (FCH4) in challenging ecosystems. To achieve this, we first globally evaluate gap-filling techniques for long FCH4 gaps. We then, for the first time, investigate impacts of machine learning algorithms, environmental drivers, and gap lengths on the gap-filling performance in seven challenging ecosystems, including three managed European grassland pasture sites, two Southeast Asian peatland conversion sites and two North American dryland sites. The aim of this study is to inter-compare and validate gap-filling approaches and determine factors that impact gap-filling performance in challenging ecosystems.

2. Methodology

2.1. Study designs

This study comprises two parts (A & B). In part A, we test our gap-filling algorithm at 77 sites of a global methane flux database (FLUXNET-CH4). Following the routines proposed in Zhu et al. (2022), we first tested the effectiveness of random forest (RFR) for filling FCH4 globally as RFR has been repeatedly suggested particularly for gap-filling long gaps (Kim et al., 2020; Irvin et al., 2021; Zhu et al., 2022). In part B, we evaluated gap-filling performance for NEE, H, LE, and/or FCH4 across machine learning algorithms to separate out the leading performance limitations in seven challenging ecosystems. The two parts will be referred to as '*Part A*' and '*Part B*' as study designs, analyses, and presentations in between are different.

2.2. Sites description

For this study we used eddy covariance measurements from 1) a global open-access FCH4 dataset (FLUXNET-CH4) for *Part A* and 2) sites we maintained in three challenging ecosystems for *Part B*.

2.2.1. FLUXNET-CH4 sites (Part A)

The FLUXNET-CH4 Version 1.0 Community Product released in 2021 is the first global FCH4 dataset (Delwiche et al., 2021). We used all the 77 open-access eddy covariance sites (across 204 site years) on a wide range of soil types [see Table S1 for details] (Delwiche et al., 2021). The mean and median gap ratio (gap half-hours / total half-hours) of the 77 sites are both 70%. In this study, incoming shortwave radiation (SW_IN_F), air temperature (TA_F), and vapour pressure deficit (VPD_F), provided the key three environmental drivers set ($driver_3$); in addition to these, incoming longwave radiation (LW_IN_F), precipitation (P), soil temperature (TS_1), friction velocity (USTAR), wind speed (WS_F), water table depth (WTD_F), and all other available drivers were added to form the extended multiple drivers set ($driver_m$) (https://fluxnet.org/data/fluxnet-ch4-community-product/data-variables/).

2.2.2. Challenging sites (Part B)

We also evaluated gap-filling techniques for three types of challenging ecosystems:

1) Three temperate grasslands as managed pasture (ROTH_HS, ROTH_PP, and ROTH_HSC) affected by grazing and other agricultural activities. These sites are from the Rothamsted North Wyke Farm Platform in the United Kingdom, (NWFP, established in 2010, see more at https://nwfp.rothamsted.ac.uk/) which provides a platform to research ecosystem responses to livestock grazing under different management practices in lowlands of southwest England. Rainfall in the area is averages around 1000 mm yr^{-1} with a mean air temperature of ca. 10 °C. The three pastures were typically grazed from April to October with cattle (ca. 4 ha^{-1}), lamb (ca. 17 ha^{-1}), and sheep (ca 10 ha^{-1}). ROTH_PP, the Permanent Pasture, is considered as a control; it retains the original sown species (predominantly perennial ryegrass, *Lolium perenne*) and has not been ploughed for the previous 10 years (Orr et al., 2016). ROTH_HS (i.e., High sugar grass) and ROTH_HSC (i.e., White clover/High sugar grass mix) were separately ploughed and re-seeded in 2013 (ROTH_HS) and 2014 (ROTH_HSC) with the *Lolium perenne* grass variety *AberMagic* and the combination of AberMagic and the white clover variety *AberHerald* (more details can be found in Orr et al. (2016) and Cardenas et al. (2022)).

Table 1
Site background information and gap ratio for NEE, H, LE, and FCH4. 10 Hz data were processed into half-hourly mean fluxes rates using EddyPro software (v6.2.2 LI-COR Environmental, Lincoln, Nebraska, USA).

	Managed pastures			Converted peatlands		Drylands	
Tower	ROTH_HS	ROTH_PP	ROTH_HSC	SAB	SEB	SEG	SES
Coordinates	50.77°N, 3.91°W	50.77°N, 3.91°W	50.77°N, 3.90°W	3.16°N, 113.42°E	3.17°N, 113.35°E	34.34°N, 106.74°W	34.36°N, 106.70°W
Data time	01/01/2017–04/06/2019 (31 months)	01/01/2017–10/07/2019 (31 months)	01/01/2017–04/07/2019 (31 months)	12/12/2016–31/01/2020 (38 months)	16/04/2017–10/02/2020 (34 months)	03/10/2018–31/12/2020 (34 months)	03/10/2018–31/12/2020 (34 months)
Plant type	High Sugar grass	Permanent Pasture	High sugar grass/white clover mix	Early converted peatland	Mature converted peatland	Bouteloua-dominated grassland	Larrea tridentata shrubland
Canopy height (m)	0.05	0.05	0.05	2.6	8	0.3	0.75
Measurement height (m)	1.59	1.57	1.59	6	20	3	2.5
Sonic anemometer	Windmaster Pro[1]	Windmaster Pro[1]	Windmaster Pro[1]	R3–50[1]	R3–50[1]	CSAT-3[2]	CSAT-3[2]
Infra-red gas analyser (IRGA)	LI-7200/(LI-7700 for FCH4)[3]	LI-7200/(LI-7700 for FCH4) [3]	LI-7200[3]	LI-7200/7550[3]	LI-7200/7550[3]	LI-7500[3]	LI-7200/LI-7500[3]
	Existing gap ratio						
NEE	69%	64%	63%	68%	80%	12%	11%
H	65%	64%	63%	70%	73%	7%	6%
LE	80%	75%	74%	70%	73%	11%	10%
FCH4	83%	83%	/	/	/	/	/

[1] Gill Instruments Ltd, Lymington, Hampshire, UK.
[2] Campbell Scientific, Logan, Utah, USA.
[3] LI-COR Environmental, Lincoln, Nebraska, USA.

Table 2
Driver sets for the seven sites at three challenging ecosystems. Sites within the same ecosystem use the same driver sets. SW is shortwave solar radiation (W m^{-2}), TA is air temperature (°C), VPD is vapour pressure deficit (kPa), PPFD is photosynthetic photon flux density (μmol m^{-2} s^{-1}), USTAR is friction velocity (m s^{-1}), WS is wind speed (m s^{-1}), NETRAD is net radiation (W m^{-2}), P is precipitation (mm), TS is soil temperature (°C), SWC is soil water content (m^3m^{-3}), and SHF is soil heat flux (W m^{-2}). The subscript 'era' indicates the corresponding drivers are re-analysed ones.

Managed pastures			*Converted peatlands*			*Drylands*		
$driver_3$	$driver_m$	$driver_{era}$	$driver_3$	$driver_m$	$driver_{era}$	$driver_3$	$driver_m$	$driver_{era}$
SW	SW	SW_{era}	SW	SW	SW_{era}	SW	SW	SW_{era}
TA	TA	TA_{era}	TA	TA	TA_{era}	TA	TA	TA_{era}
VPD	VPD	VPD_{era}	VPD	VPD	VPD_{era}	VPD	VPD	VPD_{era}
	PPFD			USTAR			PPFD	
	USTAR			WS			USTAR	
	WS			WTD			WS	
	NETRAD			/			P	
	P			/			NETRAD	
	TS			/			/	
	SWC			/			/	
	SHF			/			/	

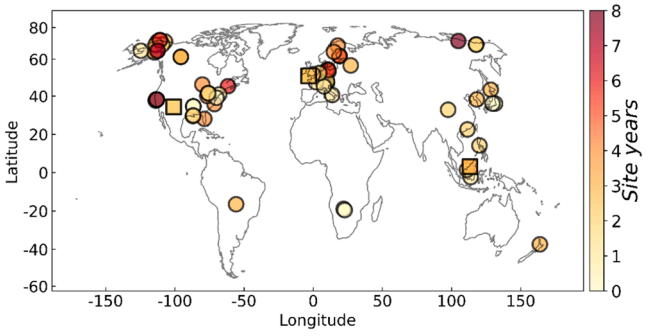

Fig. 1. Locations and site years of 77 FLUXNET-CH4 sites (circles) and 7 sites (squares) in the challenging ecosystems.

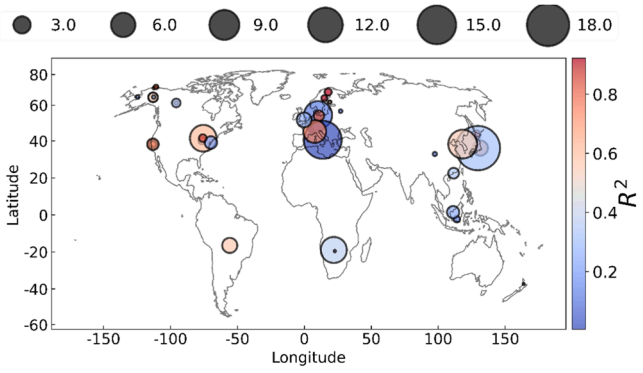

Fig. 2. Overview of gap-filling performance (R^2: circle colours, MBE: circle sizes) at the FLUXNET-CH4 sites. The performance measures are averages of gap-filling techniques and different artificial gap lengths. The unit for MBE is nmol CH_4 m^{-2} s^{-1}.

2) Two oil palm (*Elaeis guineensis*) plantations established into tropical peatland in Sarawak, Northern Malaysian Borneo, which provide datasets capturing both a developing plantation ecosystem and the mature phase under tropical conditions. The converted sites (Sabaju (SAB) and Sebungan (SEB)) were established into land

Table 3
Statistics of gap-filling performance for three approaches: MDS, RFR3, and RFRm. MDS and RFR3 use $driver_3$ set and RFRm uses $driver_m$ site. Q1 and Q3 are the first quartile and the third quartile, respectively. The unit for RMSE and MBE are nmol CH_4 m^{-2} s^{-1}.

	R^2	Slope	RMSE	MBE
	MDS			
Min	0.00	0.00	1.84	−5.14
Q1	0.08	0.13	15.95	−0.59
Median	0.34	0.43	33.78	0.05
Mean	0.37	0.42	60.78	0.99
Q3	0.66	0.66	63.17	0.92
Max	0.89	0.92	346.88	21.24
	RFR3			
Min	0.00	0.01	1.80	−14.40
Q1	0.16	0.25	13.12	−0.45
Median	0.44	0.47	28.73	0.05
Mean	0.47	0.48	52.29	0.59
Q3	0.77	0.79	52.82	1.21
Max	0.93	0.97	298.48	17.89
	RFRm			
Min	0.00	0.01	1.76	−13.05
Q1	0.30	0.31	12.63	−0.23
Median	0.52	0.50	27.86	0.13
Mean	0.53	0.53	48.54	0.54
Q3	0.79	0.79	52.18	0.71
Max	0.94	0.94	253.92	20.12

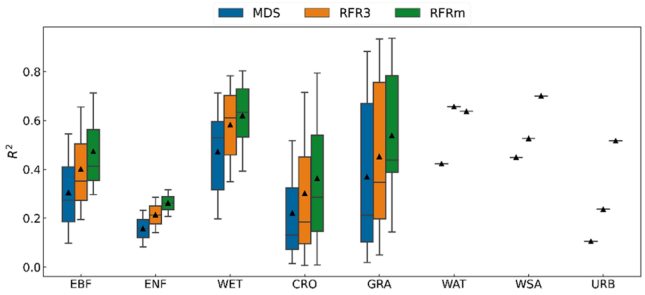

Fig. 3. Gap-filling R^2 boxplots for the three approaches grouped by the International Geosphere-Biosphere Programme classification (IGBP). Triangles in the boxes are mean values. CRO: Croplands, EBF: Evergreen Broadleaf Forests, ENF: Evergreen Needleleaf Forests, GRA: Grasslands, URB: Urban and Built-Up Lands, WAT: Water Bodies, WET: Permanent Wetlands, and WSA: Woody Savannas (https://fluxnet.org/data/badm-data-templates/igbp-classification/).

previously cleared of peat swamp forest with forest residues remaining on site, unburnt and compacted into rows, with soil drainage carried out through a regular grid system of drainage canals cut into the peat. The SAB dataset captures the early conversion period immediately following conversion (years 1–3), starting at bare soil with palms developing rapidly over the three years, while the SEB dataset covers the mature, cropping phase (years 10–12). Rainfall in the area is typically high at ca. 3000 mm yr^{-1} with a mean air temperature of ca. 26 °C. Full details of the study site area, experimental set up, data collection, processing and quality control can be found in McCalmont et al. (2021).

3) Two dryland sites, located in the Northern Chihuahuan Desert, New Mexico, USA, to evaluate the gap-filing performance under the low signal-to-noise conditions. These are AmeriFlux core sites (US-Seg and US-Ses, separately referred to as SEG and SES hereafter) (Anderson-Teixeira et al., 2011; Litvak 2016a, b; Boschetti et al., 2021). Rainfall in the area is ca. 230 mm yr^{-1} with a mean air temperature of ca. 15 °C. US-Seg experienced a severe wildfire in 2009 (https://ameriflux.lbl.gov/sites/siteinfo/US-Seg). Instrumentation and more background information of these seven sites can be found in Table 1. The data filtering and other data processing steps

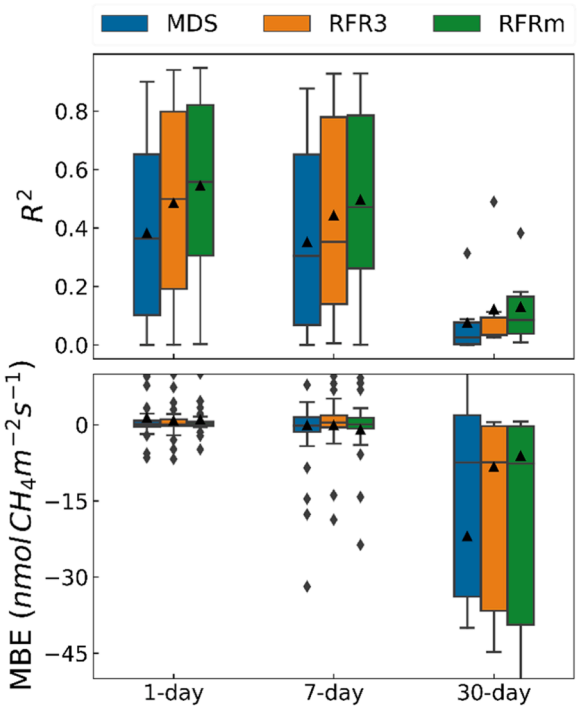

Fig. 4. Methane flux gap-filling performance in terms of R^2 and MBE for 1-day, 7-day, and 30-day long gaps. The horizontal lines and triangles within the boxes indicate medians and means, respectively. The lower and higher whiskers separately are first and third quartiles.

followed Reichstein et al. (2005), Papale et al. (2006), and referred to the FLUXNET processing standards Pastorello et al. (2020).

2.3. Gap-filling pipeline

2.3.1. Environmental drivers

Gap-filling approaches were driven by environmental variables, we investigate the influence of three different driver sets:

1) three typically measured key drivers ($driver_3$) – shortwave solar radiation (SW), air temperature (TA), and vapour pressure deficit (VPD)
2) three measured key drivers along with additional in-situ measured drivers ($driver_m$)
3) three modelled key drivers from re-analysed public records ($driver_{era}$)

In *Part A*, the driver sets for gap-filling FLUXNET-CH4 sites were introduced above. For the challenging sites in *Part B*, measured half-hourly drivers ($driver_3$ and $driver_m$) and reanalysis drivers ($driver_{era}$) derived from a publicly available database (European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) (Hersbach et al., 2018) are given in It is noteworthy that reanalysis data may not represent the exact tower-level meteorological conditions (Vuichard and Papale 2015; Lipson et al., 2022). We used driverera directly aims to test flux estimation when or where measurement data are unavailable.

In addition to environmental drivers, we also used auxiliary drivers (denoted as AUX) – i.e., hour, day of year, and year information for each half-hourly data point.

Table 2. Gaps in measured drivers were filled with ERA5 data following Vuichard and Papale (2015). ERA5 provides global hourly meteorology at 0.25° × 0.25° since 1979, and we used air and dew point temperature at 2 m above the ground and downward solar radiation at

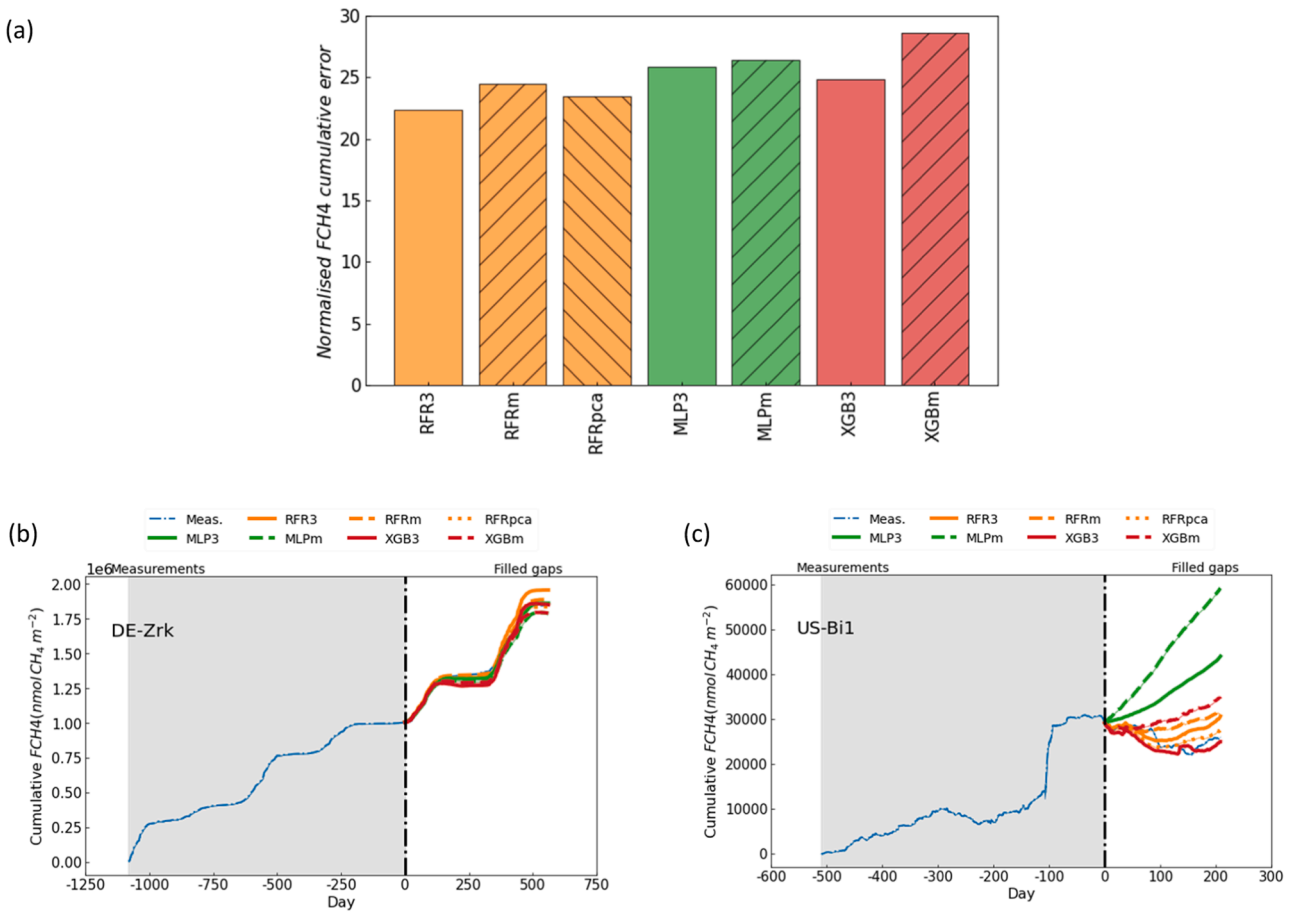

Fig. 5. Panel (a) shows the normalised FCH4 very-long gap (vl4) filling errors. The values are sums of normalised errors at FLUXNET-CH4 sites and are grouped by gap-filling approaches. The term 'error' here means the cumulative difference between filled artificial gaps and measurements in the vl4 scenario. The term 'normalised' means, at each site, the error was divided by the FCH4 sum. Panel (b) and (c) show the performance at two FLUXNET-CH4 sites DE-Zrk (c) and US-Bi1 (d) in the very-long gap (vl4) scenario where the first two-thirds of time series (i.e., the grey area left to the solid black vertical line) to train the gap-filling models while the last one-third of time series (i.e., the area right to the solid black vertical line) were used as the artificial gap to evaluate the gap-filling performance. The fluxes are presented in the cumulative manner to evaluate the aggregated errors. Gaps originally existed in measurements (i.e., blue lines) were removed beforehand to test the gap-filling performance.

the ground surface. ERA5 gridded time-series were interpolated into coordinates of the seven EC towers, respectively. It is noteworthy that reanalysis data may not represent the exact tower-level meteorological conditions (Vuichard and Papale 2015; Lipson et al., 2022). We used $driver_{era}$ directly aims to test flux estimation when or where measurement data are unavailable.

In addition to environmental drivers, we also used auxiliary drivers (denoted as AUX) – i.e., hour, day of year, and year information for each half-hourly data point.

2.3.2. *Artificial gap scenarios*

In *Part* A, gap-filling validation compared the filled artificial gaps with corresponding measured fluxes. The validation of RFR gap-filling for FCH4 took the same machine-learning algorithm implementation and artificial gap scenario as Zhu et al. (2022). In this scenario, 25% of half-hours were randomly removed from FCH4 time series to create artificial gaps with three gap-lengths: 20% were 24-hour long gaps, 30% were 7-day long gaps, and the last 50% were 30-day long gaps.

In *Part B*, to fully evaluate the gap-filling performance on various gap-lengths based on Moffat et al. (2007) and Zhu et al. (2022), we used ten gap scenarios:

a) Very-short gaps (vs) of single half-hour
b) Short gaps (s) of eight consecutive half-hours
c) Medium gaps (m) of 64 consecutive half-hours (≈ 1.5 days)
d) Long gaps (l) of 12 consecutive days
e) Mixed-length gaps (M1) of combining scenarios a to d
f) Very-long gaps (vl1) of 30 consecutive days
g) Very-long gaps (vl2) of 60 consecutive days (≈ 2 months)
h) Very-long gaps (vl3) of 90 consecutive days (≈ 3 months)
i) Very-long gaps (vl4) by making the whole last 1/3 time series as artificial gaps
j) Mixed-length gaps (M2) of combining 1-day, 7-day, and 30-day long gaps, see (Zhu et al., 2022).

Scenarios a – e, identical to Moffat et al. (2007), were used to represent typical length gaps caused by de-spiking, data quality control, or system failure. Scenarios f – j were used to assess the capability of gap-filling techniques to deal with very long gaps. Note that the non-artificial gaps – i.e., the 'real' missing half-hours in original measurements – were removed in evaluating gap-filling performance. All the evaluations and validations were carried out on artificial gaps.

2.3.3. *Gap-filling approaches*

Gap-filling techniques in this study include MDS and six machine-learning algorithms. In both *Part A*, we used MDS and the random forest (RFR) algorithms. In and *Part B*, we used all seven algorithms. The MDS method was implemented via the widely used REddyProc (v. 1.2.2) open source R package (Wutzler et al., 2018). The implementation of the six machine-learning algorithms followed the workflow in Zhu et al.

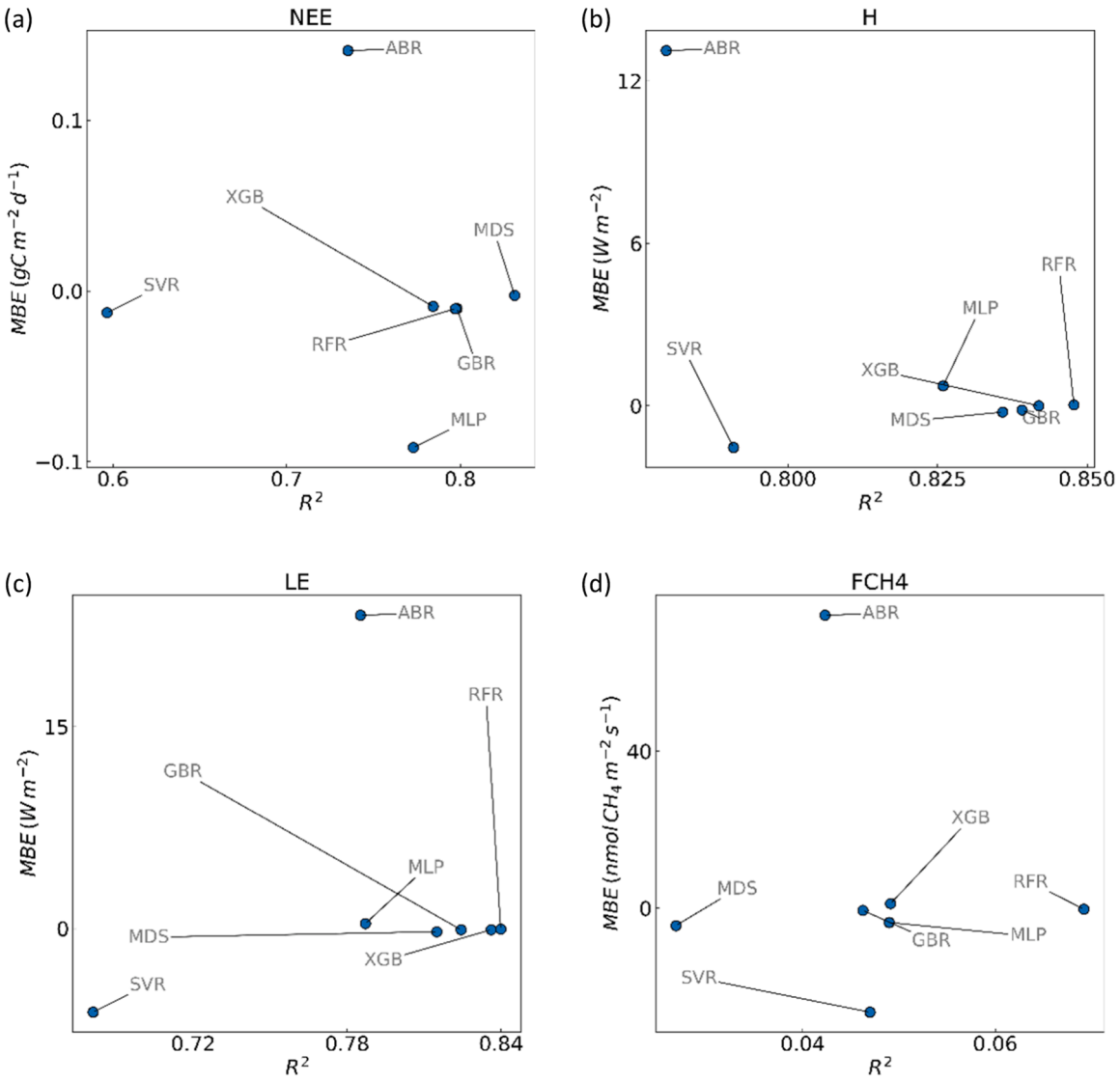

Fig. 6. Gap-filling R^2 against MBE grouped by approaches for NEE (a), H (b), LE (c), and FCH4 (d). The dots represent mean R^2 and MBE across ecosystems, driver sets, and gap-lengths a – e.

(2022). Full description of all the seven approaches can be found below. The effectiveness of each gap-filling approach was assessed by comparing gap-filled to original values using the coefficient of determination (R^2) and slope of the ordinary least squares regression, root mean squared error (RMSE) and the mean bias error (MBE). The statistical comparisons were calculated using the Python SciPy (V1.7.1) package (Virtanen et al., 2020). We applied the following methods, in addition to our statistical metrics, to further evaluate the gap-filling performance: 1) the cumulative errors to evaluate the accuracy in calculating annual sums particularly for the long gap length scenarios (vl1 – vl4) and 2) the '(permutation) feature importance', which suits all the algorithms here (Altmann et al., 2010), to measure the contribution of drivers to the machine learning algorithms. Furthermore, we also investigate the capability of reproducing already known relationships between fluxes and environmental variables – e.g., the 2nd order polynomial fit of ecosystem respiration to water table depth (WTD) (McCalmont et al., 2021).

2.3.3.1. Gap-filling workflow. The whole gap-filling workflow included approach validation and application. We first applied this workflow to all the FLUXNET-CH4 towers in *Part A*. Then in Part *B*, we applied it to the challenging towers to separately test the seven gap-filling approaches for the flux of interests (i.e., NEE, H, LE, or FCH4) in various artificial gap scenarios described in Section 2.3.2 – i.e., in every workflow implementation, we tested one approach in one scenario for one flux at one tower. Specifically, in the validation step, we randomly masked out flux measurements to create artificial gaps. As the artificial gaps may overlap with existing 'real' gaps, we applied a criterion which required at least 50% original measured data be present (Zhu et al., 2022). Otherwise, we would randomly recreate gaps unless the criterion was met. Then we filled the artificial gaps to compare with corresponding measurements. In the application step, we applied the validated approach to fill the 'real' gaps. See the algorithm paper for more technical details (Zhu et al., 2022).

2.3.3.2. Marginal distribution sampling (MDS). For MDS, a standard gap-filling approach, gaps were filled 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. Where only flux data are missing, but meteorological data are present, the missing flux value was filled with the mean value of fluxes under similar meteorological conditions within a seven-day window. If no meteorological data are available in the time window, the value was filled with a mean value from the same time of day (the mean diurnal

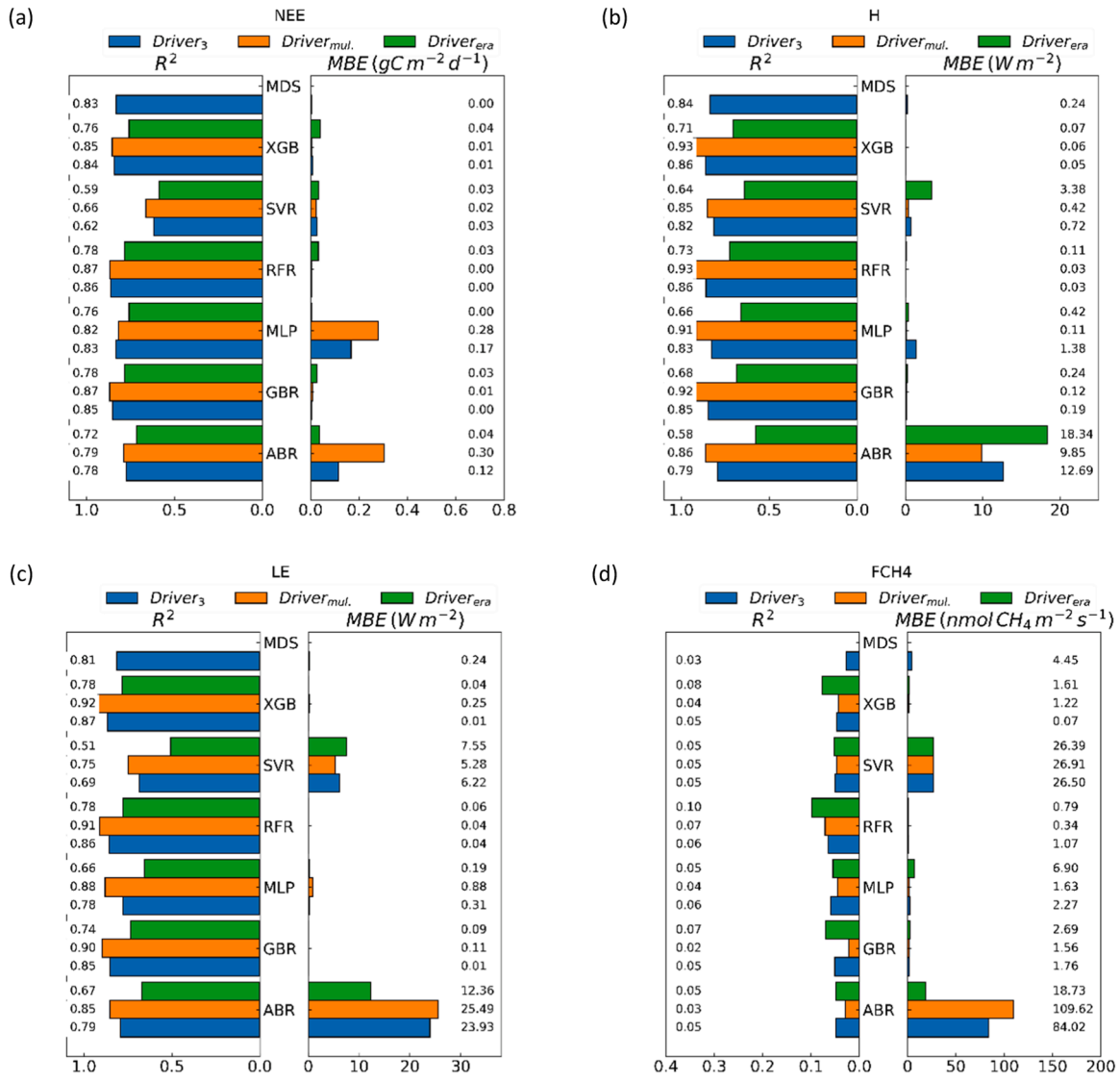

Fig. 7. Gap-filling R^2 and MBE grouped by driver sets for NEE (a), H (b), LE (c), and FCH4 (d) on average of ecosystems and gap-length scenarios.

course), initially ± 1 hour either side of the missing value or from increasingly widening windows (i.e., linear interpolation of data either side of the missing value). Standard deviations of these mean values used in gap-filling are recorded along with a categorical classification of the confidence level of the filled value based on approach and window size (Reichstein et al., 2005).

2.3.3.3. Machine-learning algorithms. Here, we tested three commonly used gap-filling algorithms [(1) – (3)] and three additional algorithms [(4) – (6)] to assess their potential for improving the gap-filling performance:

1) Multiple layer perceptron (MLP) (Hinton 1989);
2) Support vector regressor (SVR) (Platt 1999);
3) Random forest regressor (RFR) (Breiman 2001);
4) Xgboost (XGB) (Chen and Guestrin 2016);
5) Ada boost regressor (ABR) (Freund and Schapire 1997);
6) Gradient boosting regressor (GBR) (Friedman 2001).

Algorithm (1) is an artificial neural network (ANN) but we use the specific label, multiple layer perceptron (MLP) to avoid ambiguity because the term ANN now encapsulates various kinds of neural networks (Abiodun et al., 2018). It is also a standard gap-filling approach particularly for gaps longer than one month (Delwiche et al., 2021; Mahabbati et al., 2021). The SVR was also an established gap-filling algorithm, it converts non-linear regressions into higher-dimensional linear regression by a predefined kernel function (Khan et al., 2021; Yao et al., 2021b).

Decision tree-based algorithms, especially the RFR, were reported to be superior to the standard gap-filling approaches (Kim et al., 2020; Mahabbati et al., 2021; Zhu et al., 2022). Hence, we also used other mainstream decision tree-based algorithms [algorithm (4) – (6)] to further test the effectiveness of tree-based algorithms in gap-filling. In addition, information redundancy – i.e., the correlation between drivers – can be detrimental to the gap-filling performance (Kim et al., 2020). Therefore, we also adopted the RFR with principal component analysis (RFR_{pca}) to reduce redundant information as in Kim et al. (2020).

We used the Scikit-Learn package (v 0.23.1) (Pedregosa et al., 2011) within Python (v 3.6) to provide interfaces to all machine-learning algorithms except XGB which was provided independently (v 1.1.1, https://xgboost.readthedocs.io/en/latest/index.html) (Chen and Guestrin 2016). Hyperparameters of all the machine learning algorithms were set as default (see the links after algorithm names). Details of the six machine-learning algorithms can be found in the Supplementary materials.

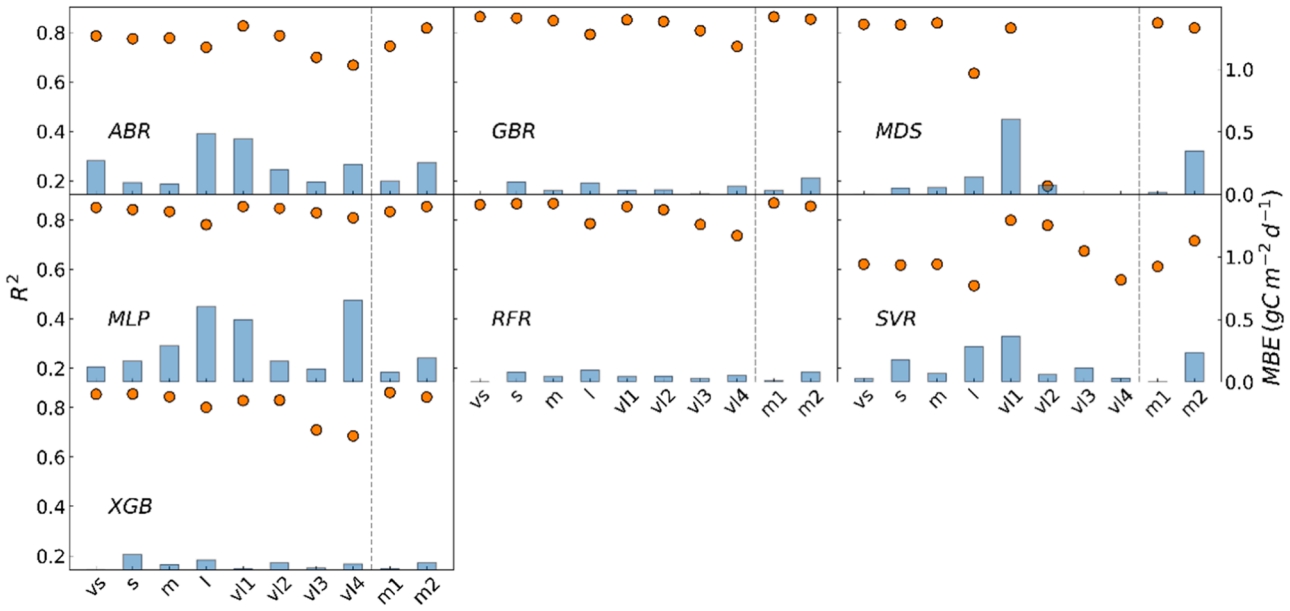

Fig. 8. NEE gap-filling R^2 (dots) and MBE (bars) for various gap-lengths.

3. Results

3.1. Gap-filling evaluation for FLUXNET-CH4 sites (Part A)

3.1.1. Global validation of RFR gap-filling for methane fluxes

The performance for filling long methane gaps exhibited large spatial variability but no regional patterns (Fig. 2). The gap-filling performance was the best in wet tundra and fen ecosystems in north Europe and northwest America as well as rice and bog ecosystems in northeast Asia and North America (indicated by higher R^2 and smaller bias, see https://fluxnet.org/data/fluxnet-ch4-community-product/ for the types and distribution of ecosystems). As for the methane flux gap-filling approaches per se (Figure S2), three-driver random forest (RFR3) performed better than marginal distribution sampling (MDS), as indicated by a 29% higher R^2 median; Furthermore multiple-driver random forest (RFRm) performed better than RFR3 by 19% (Table 3). Regarding the gap-filling error in terms of RMSE and uncertainty in terms of the interquartile range of bias, three-driver random forest was better than marginal distribution sample due to a 10% smaller RMSE median and a smaller uncertainty (1.51 vs. 1.66 nmol $CH_4\ m^{-2}\ s^{-1}$). Using multiple drivers further reduced the error of random forest gap-filling by 3% and with a smaller uncertainty of 0.94 nmol $CH_4 m^{-2}\ s^{-1}$ (Table 3).

Following the International Geosphere-Biosphere Programme (IGBP) classification, gap-filling R^2 of RFRm was higher than RFR3 and further higher than MDS in nearly all classes (Fig. 3), alongside with the opposite pattern of error – i.e., error of RFRm was smaller than RFR3 and further smaller than MDS (Figure S5). Details for gap-filling performance in terms of site classification in Delwiche et al. (2021) can be found in Figure S5 too. The R^2 distribution was in relation to types of ecosystems, fen and marsh ecosystems had higher R^2 than other site classes.

Considering both IGBP and site (Delwiche et al., 2021) classifications, taking gap-filling results of RFRm (Fig. 3b) as an example, higher R^2 values were observed in classes with higher fluxes. Meanwhile, classes with higher fluxes were also seen with relatively larger gap-filling errors, and these characteristic were seen for the other two gap-filling methods. Flux values and R^2 exhibited a positive second-order polynomial relationship (Figure S5) and this positive relationship was more obvious for the site classification (Fig. 3b). In other words, IGBP classes (of the same site class) with higher fluxes showed higher R^2. For example, in the bog class, evergreen broadleaf forests (EBF) had higher averaged flux value and R^2 than permanent wetlands (WET) and further higher than evergreen needleleaf forests (ENF). However, this pattern was not seen obviously for site classes.

When filling longer gaps, R^2 of all three gap-filling approaches decreased while the bias increased, particularly for filling the 30-day gaps (Fig. 4). For filling all three gap-lengths, multiple-driver random forest still performed better than three-driver random forest and further performed better than marginal distribution sampling. Comparing filling 1-day long gaps and 7-day long gaps, the gap-filling performance of all three approaches were relatively stable (R^2 medians decreased by less than 30% while the uncertainty difference was approximately 2 nmol $CH_4\ m^{-2}\ s^{-1}$). However, as gap-length increased further from 7-day to 30-day, the gap-filling performance of all three approaches declined greatly (R^2 medians dropped by nearly 90% while the uncertainty increased to nearly 40 nmol $CH_4\ m^{-2}\ s^{-1}$).

3.1.2. Intercomparisons between machine learning algorithms

As regards the cumulative gap-filling errors at global FLUXNET-CH4 sites (Fig. 5a), the random forest regressor (RFR) had smaller error compared to the other machine learning algorithms. Here, we show the results of Xgboost (XGB) as an example of other decision tree-based algorithms. The marginal distribution sampling was not employed because it cannot fill very long gaps (Zhu et al., 2022). Comparing the number of drivers, the cumulative error of using multiple drivers (e.g., RFRm) was higher than using the three essential drivers (RFR3). As an example, Fig. 5b and c show the typical gap-filling performance at two FLUXNET-CH4 sites – DE-Zrk with strong seasonality and US-Bi1 with low seasonality. Gap-filling approaches at the two sites exhibited contrasting performance. Cumulative fluxes filled by all approaches at site DE-Zrk were in good agreement with corresponding measurements (Fig. 5b). Whilst large disagreement was observed at site US-Bi1 (Fig. 5c). As regards performance difference between algorithms, the RFR and XGB estimated cumulative methane flux (FCH4) much closer to measurements (Fig. 5e) than the research-standard multiple layer perceptron (MLP) algorithm.

3.2. Gap-filling evaluation in challenging ecosystems (Part B)

3.2.1. Comparison between methods, drivers, and gap-lengths

Averaged across scenarios a – e, i.e., scenarios in Moffat et al. (2007), random forest (RFR) performance was best for most fluxes (indicated by

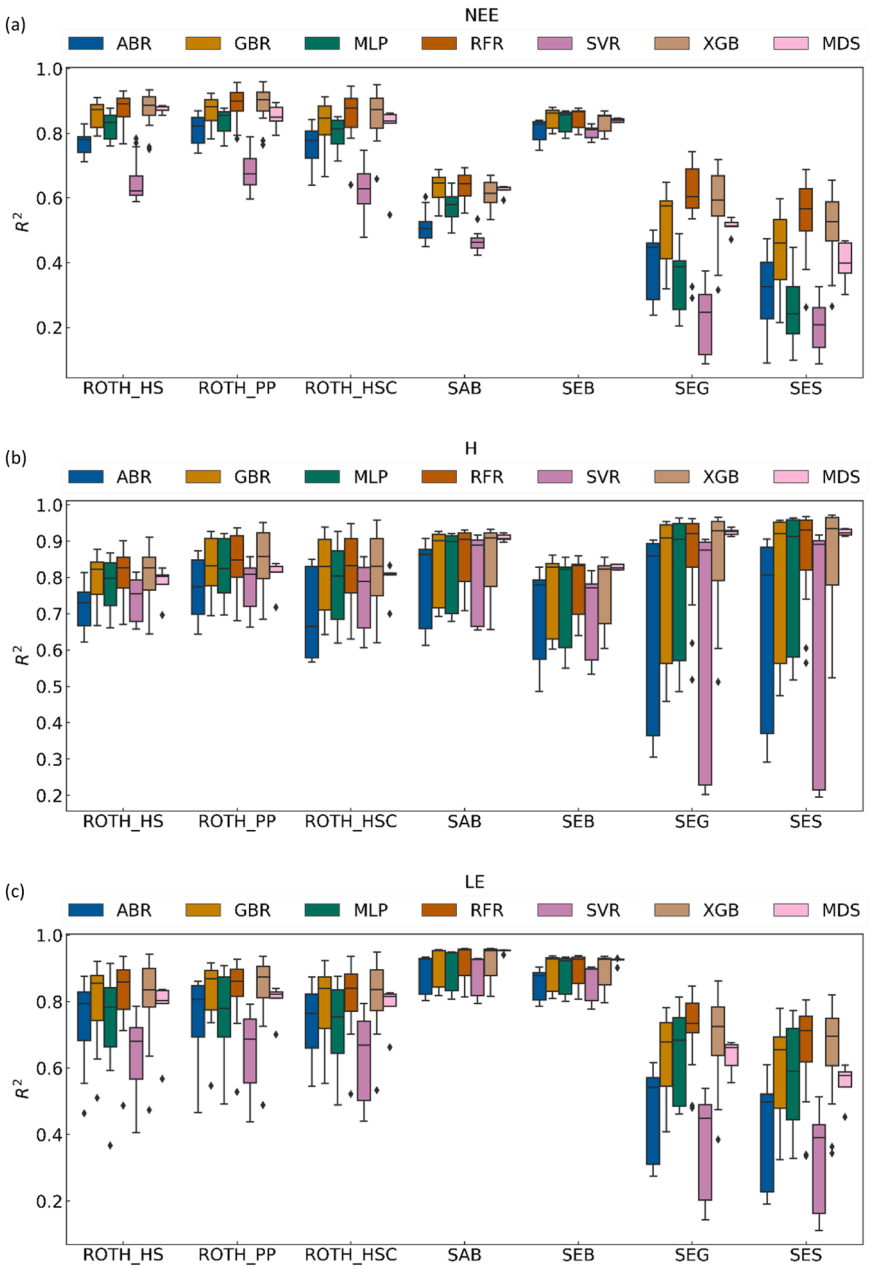

Fig. 9. Gap-filling performance in terms of R^2 of various approaches at seven towers in the three challenging ecosystems for NEE (a), H (b), and LE (c). This figure is based on the aritificial gap-lengths a – e.

higher R^2 and smaller bias, Fig. 6). For net ecosystem exchange (NEE), gap-filling performance of marginal distribution sampling (MDS) was the best due to the highest R^2 and a relatively small bias. Random forest regression (RFR) and gradient boosting regression (GBR) also exhibited relatively good gap-filling performance ($R^2 > 0.8$); the bias difference between approaches was smaller than 0.05 g C m^{-2} d^{-1} (Fig. 6a). In contrast, the gap-filling performance was the worst by using support vector regression (SVR, $R^2 < 0.6$) and Ada boost regression (ABR, bias > 0.1 g C m^{-2} d^{-1}). For sensible heat (H) and latent energy (LE), random forest regression (RFR), gradient boosting regression (GBR), and marginal distribution sampling (MDS) still showed the best gap-filling performance ($R^2 > 0.5$ and bias < 5 W m^{-2}); in this case, Xgboost (XGB) also exhibited an equivalent R^2 and bias (Fig. 6b and c). Again, support vector regression (SVR) and Ada boost regression (ABR) showed the worst performance. For methane fluxes (FCH4), random forest regression (RFR) showed relatively better performance but the gap-filling performance of all approaches were bad ($R^2 < 0.1$, Fig. 6d). Bias of Ada boost regression (ABR) exceeded 40 nmol CH_4 m^{-2} s^{-1} while it was smaller than 20 nmol CH_4 m^{-2} s^{-1} for all other approaches.

Generally, for net ecosystem exchange (NEE), sensible heat (H), and latent energy (LE), the gap-filling R^2 of using multiple drivers ($driver_m$, $m > 3$) was higher than using three drivers ($driver_3$) and modelled drivers ($driver_{era}$) for all the approaches (Fig. 7a –c). In contrast, for methane fluxes (FCH4), the R^2 for modelled drivers ($driver_{era}$) was the highest for most of the approaches (Fig. 7d). Gap-filling bias showed no uniform characteristics in comparisons between driver sets; for example, multiple layer perceptron (MLP) bias in filling net ecosystem exchange gaps for using modelled drivers ($driver_{era}$) was close to zero, but bias for using three drivers ($driver_3$) was 0.17 g C m^{-2} d^{-1} and bias for using multiple drivers ($driver_m$) was even larger, reaching 0.28 g C m^{-2} d^{-1} (Fig. 7a). However, bias of random forest regression (RFR) for using $driver_m$ was close or smaller than $driver_3$ and smaller than $driver_{era}$ for

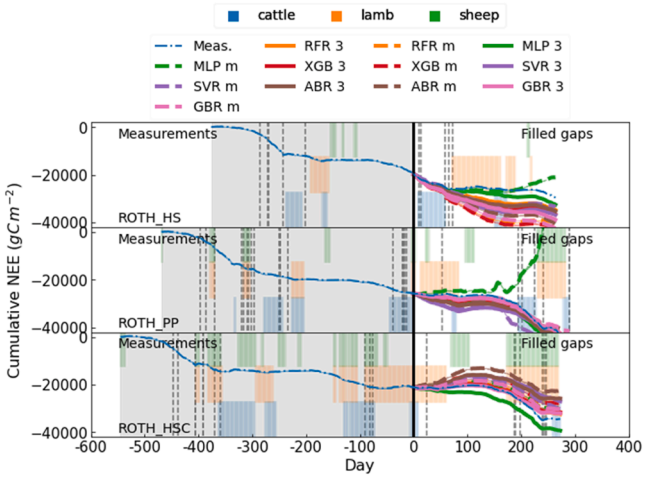

Fig. 10. Gap-filled fluxes against measurements in the very-long gap (vl4) scenario where the first two-thirds of time series (i.e., the grey area left to the solid black vertical line) to train the gap-filling models while the last one-third of time series (i.e., the area right to the solid black vertical line) were used as the artificial gap to evaluate the gap-filling performance. The fluxes are presented in the cumulative manner to evaluate the aggregated errors. The blue, orange, and green blocks represent the grazing period for cattle, lamb, and sheep, respectively. The grey dashed vertical lines are the occurrence of management activities.

net ecosystem exchange, sensible heat, and latent energy but not for methane (Fig. 7).

The gap-filling R^2 consistently decreased as the gap-length increased from very-short (vs) to long (l) gaps in the 'standard' artificial gap scenario (Moffat et al., 2007) for all the approaches (Fig. 8). Similar results for other fluxes can be found in Figure S4. In comparison between marginal distribution sampling (MDS), random forest regression (RFR) and gradient boosting regression (GBR), MDS R^2 decreased by the largest amount as gap-length increased from medium (m) to long (l), nearly 25%. In the meantime, MDS showed the largest and continuous absolute bias increase as the gap-length increased from very-short (vs) to long (s); but the bias variations of random forest and gradient boosting were smaller. For very-long gaps (v1 – v4), R^2 also decreased as the gap-length increased. R^2 decreased to a much lower ratio (~ 25%) for machine-learning approaches than for MDS. In particular, MDS failed to fill gaps when the gap-length reached 3-month (vl3). random forest regression (RFR), gradient boosting regression (GBR), and Xgboost (XGB) had relatively small absolute bias amongst gap lengths; and random forest bias variations were smaller compared to other approaches.

Gap-filling R^2 for the standard artificial gap scenario (Moffat et al., 2007) showed relatively obvious ecosystem-level patterns (Fig. 9). Gap-filling R^2 for the control permanent pasture (ROTH_PP) was higher and with narrower interquartile range (IQR) than managed pastures (ROTH_HS and ROTH_HSC). R^2 for the mature converted tropical peatland (SEB) was higher and with narrower IQR than the plantation establishment phase peatland (SAB). Dryland sites (SEG and SES) were seen with the lowest R^2 and/or the widest IQR compared with other ecosystems. These phenomena were particularly obvious for net ecosystem exchange (Fig. 9a).

3.2.2. Further evaluations

Filled gaps by most approaches were in line with the corresponding measurements (Fig. 10). For grazing events (i.e., chromatic blocks) that affected biomass amount and fluxes relatively slowly, no obvious increasing difference between filled gaps and measurements was seen during grazing periods when other management activities did not take place. These management activities include spraying herbicides and grass cutting (full details of management practices at the Rothamsted sites can be found here: https://nwfp.rothamsted.ac.uk/). In contrast, the upheaval in flux was mainly observed around management activities (e.g., around day 250 in ROTH_HSC).

Regarding machine learning algorithms, the research-standard multiple layer perceptron (MLP) exhibited very unstable cumulative flux compared with corresponding measurements. The Xgboost [XGB] (Fig. 10a), support vector regressor [SVR] (Fig. 10b), and Ada boost regressor [ABR] (Fig. 10c) showed relatively larger cumulative difference against measurements than other machine learning algorithms. In ROTH_HS and ROTH_HSC, the difference between filled gaps and measurements were larger for using multiple drivers ($driver_m$) than for using three essential drivers ($driver_3$).

In site SAB, a previous study observed 2nd order polynomial responses of night-time net ecosystem exchange (NEE) to water table depth (WTD). As shown in Fig. 11, the reproduced 2nd order polynomial fits of WTD against filled gaps using the multiple layer perceptron (MLP), support vector regressor (SVR), and Ada boost regressor (ABR) were barely in agreement with the fit for measurements. In contrast to Xgboost (XGB), for random forest regressor (RFR) and gradient boosting regressor (GBR), the reproduced fit of using multiple drivers ($driver_m$) was more in line with the fit for measurements than using three drivers ($driver_3$). Overall, the RFRm reproduced 2nd order polynomial fit was the closest to the fit for measurements. Flux responses to other environmental variables for all challenging sites can be seen in Figure S6.

The shortwave radiation (SW) contributed the most information (importance > 50%) for gap-filling net ecosystem exchange (NEE) and latent energy (LE) in managed pasture sites [ROTH_HS, ROTH_PP, ROTH_HSC] and Malaysian converted sites [SAB and SEB] (Fig. 12). SW and the net radiation (RN) were the dominant driver (importance close to 90%) for gap-filling sensible heat (H) at all the seven challenging sites. The SW and RN contributed nearly 50% for gap-filling LE at the dryland sites [SEG and SES]. However, no clearly dominant environmental drivers were seen for gap-filling methane flux (FCH4) at ROTH_HS & ROTH_PP and for gap-filling NEE at SEG & SES (Fig. 12). Wind speed and direction (WIND) contributed the most information for gap-filling FCH4 at ROTH_HS and ROTH_PP. For gap-filling NEE at SAB and SEB, the sum importance of the three essential drivers (i.e., solar radiation, vapour pressure deficit, and air temperature) was smaller than 40%.

4. Discussion

4.1. Global methane flux gap-filling feasibility and limitations

In in this study, filling long methane gaps with random forest at global FLUXNET-CH4 sites exhibited inferior performance. In contrast, random forest demonstrated great performance for filling short methane gaps in Kim et al. (2020) and for filling long net ecosystem exchange, sensible heat, and latent energy fluxes in Zhu et al. (2022). In comparison with MDS, the advantages of using random forest were sound (Fig. 3) and this is in line with Irvin et al. (2021). However, filling long methane flux gaps is still challenging. The gap-filling performance varied largely by sites (Figure S2). This discrepancy in R^2 between sites could relate to the strength of methane flux (Fig. 3) and variability in methane flux time series. For example, at sites BR-Npw and US-Tw4 (Figure S2) that were also in Kim et al. (2020), the gap-filling performance was good ($R^2 > 0.8$) and methane time series at both sites exhibited very strong seasonality. On the contrary, at sites with poor gap-filling performance (e.g., AT-Neu and CH-Dav where $R^2 < 0.05$, Figure S2), the temporal variations of methane time series were more irregular. In agreement with Irvin et al. (2021), the correlation between gap-filing performance and seasonality/periodicity in methane time series was broadly seen, sites with good gap-filling performance showed strong periodicity and vice versa (Figure S2). This suggested that filling methane gaps heavily depended on the periodicity of drivers to

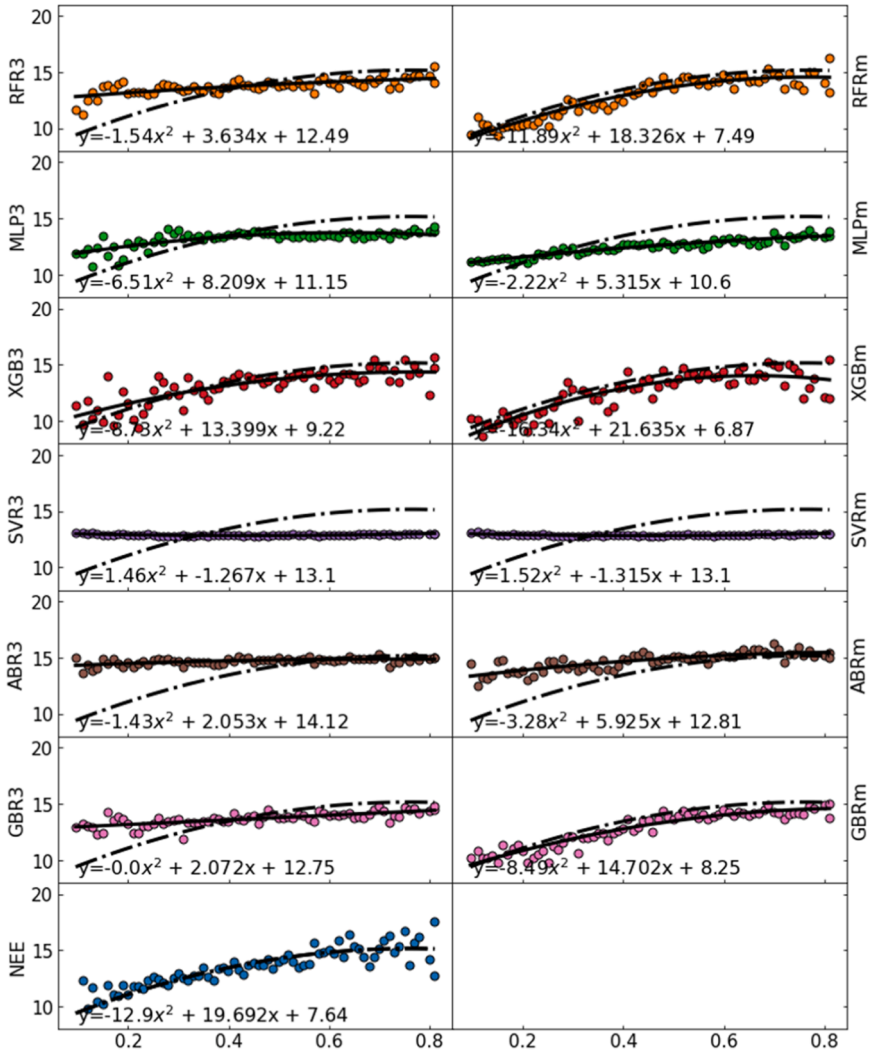

Fig. 11. Reproduced night-time NEE (y-axes) responses to water table depth (WTD, x-axes) below the surface binned to 0.01 m increments in site SAB. Data were from the very-long gap (vl4) scenario. The solid curves are the responses for measured fluxes while the dashed curves are for filled gaps using different machine learning algorithms and driver sets.

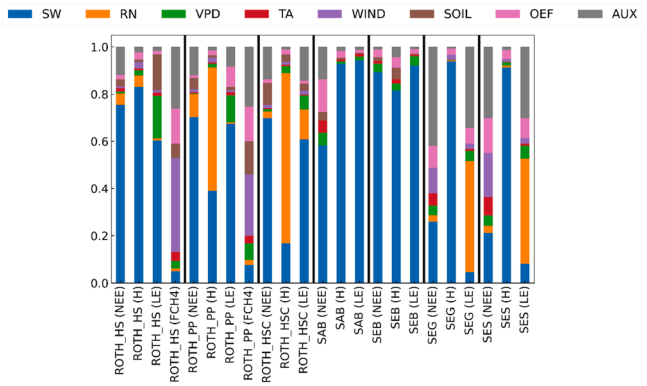

Fig. 12. Feature importance to gap-filling for fluxes in challenging ecosystems. The sum of drivers in every ecosystem equals to 100%. SW: shortwave radiation, RN: net radiation, TA: air temperature, WIND: wind speed and wind direction, SOIL: soil temperature, soil water content, and soil heat flux if possible; OEF: other environmental features like precipitation; AUX: auxiliary drivers like season and day of year.

reproduce the temporal dynamics in methane fluxes.

Where gap filling failed to replicate this periodicity may simply be due to a lack of data at specific sites [e.g., data for FI-Hyy only covered four months, see Table S1 and Figure S2] or local ecosystem type and climate [e.g., CH—Hgu where the dominant vegetation was alpine meadow and the methane time series showed no periodicity] (Delwiche et al., 2021). According to Fig. 5, despite the low R^2, the cumulative gap-filling error can be relatively small for machine learning algorithms excluding the artificial neural network (i.e., MLP). This suggests that filling very long methane flux gaps can be feasible if the goal is to estimate annual sums.

Further improvements in the gap-filling performance will benefit from understanding the ecosystems. As we tested both classic and state-of-art machine learning algorithms (Fig. 5), further technical advances may not enhance the gap-filling performance for methane fluxes. The dependence of gap-filling performance on methane periodicity is a significant challenge. Machine learning approaches can exploit the temporal structure information and achieved good gap-filling performance at sites with strong periodicity, therefore it infers that the dominant environmental drivers of methane fluxes are complex and may vary largely by ecosystem type (Figure S3). Hence, understanding the study ecosystem and identifying the dominant driver (Knox et al., 2021) can be very helpful to the challenging sites with poor gap-filling

performance for methane and other fluxes.

4.2. Selection of machine-learning algorithms and driver sets

In agreement with previous studies (Falge et al., 2001; Reichstein et al., 2005; Moffat et al., 2007), marginal distribution sampling (MDS) was effective in filling net ecosystem exchange (NEE) gaps ≤ 12 days (Fig. 6a). As gaps shorter than 12 days cover most typical gap-lengths caused by data quality control or short-term system failure, MDS is therefore still recommended to be the standard gap-filling approach (Pastorello et al., 2020). For longer gaps, support vector regression (SVR) and Ada boosting regression (ABR) approaches are not recommended due to their lower R^2 and larger absolute MBE (Fig. 6). In contrast, both random forest regression (RFR) and gradient boosting regression (GBR) showed similar performance (Fig. 6), but RFR may be preferred due to its relatively smaller bias and smaller bias variations with increasing gap-length (Fig. 8). Multiple layer perceptron (MLP) (i.e., shallow layered neural networks) was the most stable approach to gap-length (Fig. 8). The deep learning techniques that have been broadly applied in other environmental sciences (Reichstein et al., 2019; Zhu et al., 2021) might show potential in further improving the gap-filling performance for very long gaps.

For net ecosystem exchange (NEE), sensible heat (H), and latent energy (LE) fluxes, the selection of driver set affected gap-filling R^2 by ~10% (Fig. 7). In general, using multiple drivers (i.e., $driver_m$) improves the fraction of flux variance explained by the machine learning gap-filling model. In line with our findings presented in Zhu et al. (2022), we have shown that using three drivers ($driver_3$) can achieve comparable gap-filling performance to using $driver_m$ even in the more challenging ecosystems. Averaged R^2 of using modelled drivers ($driver_{era}$) was higher than 0.7 while the bias was less than 0.05 g C m^{-2} d^{-1} for NEE and less than 0.5 W m^{-2} for H and LE (Fig. 7). This suggests that reanalysis data can be effective in estimating fluxes when or in regions where measured flux and meteorological data are unavailable.

4.3. Shortcomings and advantages of machine-learning approaches for the challenging ecosystems

Machine-learning approaches may fail in gap-filling when the flux environmental driving mechanism was unclear (Fig. 12 and Figure S3) unless strong flux periodicity was present. This was particular the case for gap-filling methane when flux periodicity was extremely low in the time series. For example, methane was typically low at sites ROTH_HS and ROTH_PP, but in the summer of 2018, the ecosystem experienced short-term rapid methane increases which could be in relation to the presence of livestock in the field (Figure S1). Identifying the 'right' drivers was also crucial to gap-filling fluxes in the dryland sites SEG and SEG. Fluxes in dryland ecosystems are dependant on water availability (Barnes et al., 2021), but according to Table 2, we did not have the soil water supply data. This could be one major reason limiting the gap-filling performance in dryland sites. Besides, sudden biomass or flux changes were not captured (e.g., management activities in Fig. 10) by the machine learning algorithms. It was difficult to quantify the management activities as algorithm drivers (Orr et al., 2016) and this may cause the failed capture. Therefore, determining the drivers directly and explicitly correlating with flux variations in these challenging ecosystems may be the way to improve the gap-filling performance for these machine-learning approaches.

In contrast, machine learning algorithms can well extrapolate the impacts of slow biomass changes from the past into the future – e.g., grazing events at sites ROTH_HS, ROTH_PP, and ROTH_HSC (Fig. 10) and plant growth at site SAB (Figure S1) – even though such change information was not directly used as drivers. For example, random forest regression (RFR) successfully reproduced cumulative net ecosystem exchange (NEE) for both early-stage and mature converted peatlands (Figure S1b). This was assessed by removing the last 30% of the flux time series (scenario vl4) and gap-filling it. The reproduced sums matched well with the removed measurements for the mature ecosystem but showed slight overestimation for the early-stage ecosystem (Figure S1b).

In agreement with the literature (Kim et al., 2020; Mahabbati et al., 2021; Irvin et al., 2021; Zhu et al., 2022), there was no single algorithm that stood out clearly as the best in our range of options (Fig. 6), but the random forest (RFR) was found to be a very competitive alternative to the existing standard gap-filling algorithms. Decision tree-based algorithms, e.g., RFR, were shown to be more resilient to short-term disturbances, such as management events (Fig. 10), than the research-standard multiple layer perceptron (MLP). In comparison with other decision tree-based algorithms, RFR was also advantageous in reproducing the flux responses to environmental drivers [e.g., water table depth in SAB] (Fig. 11). Furthermore, the use of RFR approaches can improve the explained variance by 20% to 25% (Fig. 9) in dryland ecosystems where all gap-filling approaches struggled. Further eddy covariance towers and studies are in need in drylands because they cover ~40% of the global land area (Huang et al., 2016) but are under-sampled with eddy covariance towers (Boschetti et al., 2019). For filling very long gaps (vl4), the performance of RFR was particularly promising on timescales ranging from multi-hour to multi-day. Hence, RFR is recommended to use in future eddy covariance studies.

Spatiotemporal scaling remains a challenge – i.e., for how long or for how far can flux data be extrapolated in time and space dimensions. In the spatial dimension, extrapolating eddy covariance fluxes typically uses satellite remote sensing and gridded meteorology data, and this study field is referred to as flux upscaling (Jung et al., 2020). However, satellites cannot provide both the high spatial and high temporal resolution observations needed to directly compare with eddy covariance fluxes at half-hourly and tower-level scales. In the temporal dimension, machine learning algorithms exhibited promise in predicting fluxes in the coming year. Predictions of fluxes in the more distant future may be possible in the absence of environmental (e.g., vegetation species and/or temperature) changes outside the measurement range. However for both spatial and temporal scalability more work is required.

5. Conclusion

The accuracy of gap-filling techniques is critically important to the continuous flux measurements from eddy covariance. For the first time, we comprehensively evaluated what factors affect the gap-filling performance and by how much, particularly in challenging ecosystems. We have shown that while increasing the number *of in-situ* driver measurements improves gap filling performance, utilisation of publicly available regional datasets, when combined with machine learning techniques, particularly random forest regression (RFR), can still provide good results. RFR also showed superior performance when considering more challenging ecosystems with high levels of management interventions, in this case grazed pasture or tropical peatland converted to agriculture. While marginal distribution sampling (MDS) still performed well with gaps up to the medium range, for much longer gaps, RFR was a clear improvement. Gains in performance were also seen for RFR in gap filling methane datasets, but more limited and inconsistent at the site-specific level. Critically, the environment-flux responses emerged in RFR but not in MLP gap-filled data. The use of RFR for future gap-filling is thereby further recommended. Despite being a significant global climate impact, ecosystem scale datasets for methane flux are only very recently becoming available and much work remains to improve our understanding of ecosystem drivers, and the relationships between them, for this important greenhouse gas Fig. 1.

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.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank the FLUXNET-CH4 research groups for providing the CC-BY-4.0 (Tier one) open-access eddy covariance data (https://fluxnet.org/data/fluxnet-ch4-community-product/) and ERA5 (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5) for providing meteorology reanalysis data. They also thank the ReddyProc (https://cran.r-project.org/web/packages/R EddyProc/index.html) team, scikit-learn (https://scikit-learn.org/s table/install.html) team, and Xgboost team (https://xgboost.readthedocs.io/en/stable/python/python_api.html) for the packages that help the implementation and validation for gap-filling approaches. SZ and TH would like to acknowledge funding from Shell to support the PhD studentship. Rothamsted thanks BBSRC grants BBS/E/C/000I0320 and BBS/E/C/000J0100. The Eddy Covariance equipment deployed in this work was funded by CIEL (https://www.cielivestock.co.uk/) and the raw data is available on the Farm Platform Portal (https://nwfp.rothamsted.ac.uk/).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2023.109365.

References

Abiodun, O.I., Jantan, A., Omolara, A.E., et al., 2018. State-of-the-art in artificial neural network applications: a survey. Heliyon 4, e00938.
Albert, L.P., Keenan, T.F., Burns, S.P., et al., 2017. Climate controls over ecosystem metabolism: insights from a fifteen-year inductive artificial neural network synthesis for a subalpine forest. Oecologia 184, 25–41.
Altmann, A., Toloşi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. Bioinformatics 26, 1340–1347.
Anderson-Teixeira, K.J., Delong, J.P., Fox, A.M., et al., 2011. Differential responses of production and respiration to temperature and moisture drive the carbon balance across a climatic gradient in New Mexico. Glob. Change Biol. 17, 410–424.
Baldocchi, D.D., 2020. How eddy covariance flux measurements have contributed to our understanding of global change biology. Glob. Change Biol. 26, 242–260.
Barnes, M.L., Farella, M.M., Scott, R.L., et al., 2021. Improved dryland carbon flux predictions with explicit consideration of water-carbon coupling. Commun. Earth Environ. 2, 1–9.
Boschetti, F., Cunliffe, A., Clement, R., et al., 2019. Quantification of the spatial variability of CO2/H2O fluxes in dryland ecosystems using low-cost EC systems. Geophys. Res. Abstr. EGU2019–EGU7757. Vienna.
Boschetti, T.C .F, Cunliffe, A.M., Clement, R., Anderson, K., Sitch, S., Brazier, R.E., Hill, 2021. Half hourly fluxes of sensible heat, latent energy and carbon, observed by eight eddy covariance towers in the Northern Chihuahuan Desert. N. Am. 2018–2019.
Braswell, B.H., Sacks, W.J., Linder, E., Schimel, D.S., 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. Glob. Change Biol. 11, 335–355.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Cardenas, L., Olde, L., Loick, N., et al., 2022. CO2 fluxes from three different temperate grazed pastures using Eddy covariance measurements. Sci. Total Environ., 154819
Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM sigkdd international conference on knowledge discovery and data mining, pp. 785–794.
Cunliffe, A.M., Boschetti, F., Clement, R., et al., 2022. Strong correspondence in evapotranspiration and carbon dioxide fluxes between different eddy covariance systems enables quantification of landscape heterogeneity in dryland fluxes. J. Geophys. Res.: Biogeosci., e2021JG006240
Delwiche, K.B., Knox, S.H., Malhotra, A., et al., 2021. FLUXNET-CH 4: a global, multi-ecosystem dataset and analysis of methane seasonality from freshwater wetlands. Earth Syst. Sci. Data 13, 3607–3689.
Dengel, S., Zona, D., Sachs, T., et al., 2013. Testing the applicability of neural networks as a gap-filling method using CH4 flux data from high latitude wetlands. Biogeosciences.
Eugster, W., Plüss, P., 2010. A fault-tolerant eddy covariance system for measuring CH4 fluxes. Agric. For. Meteorol. 150, 841–851.
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.
Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55, 119–139. https://doi.org/10.1006/jcss.1997.1504.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232. https://doi.org/10.1214/aos/1013203451.
Hersbach, H., Bell, B., Berrisford, P., et al., 2018. ERA5 Hourly Data on Single Levels from 1979 to Present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS).
Hill, T., Chocholek, M., Clement, R., 2017. The case for increasing the statistical power of eddy covariance ecosystem studies: why, where and how? Glob. Change Biol. 23, 2154–2165.
Hinton, G.E., 1989. Connectionist learning procedures. Artif. Intell. 40, 185–234. https://doi.org/10.1016/0004-3702(89)90049-0.
Hommeltenberg, J., Mauder, M., Drösler, M., et al., 2014. Ecosystem scale methane fluxes in a natural temperate bog-pine forest in southern Germany. Agric. For. Meteorol. 198, 273–284.
Huang, J., Yu, H., Guan, X., et al., 2016. Accelerated dryland expansion under climate change. Nat. Clim. Change 6, 166–171.
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. Agricultural and Forest Meteorology 308–309:108528. https://doi.org/10.1016/j.agrformet.2021.108528.
Jung, M., Schwalm, C., Migliavacca, M., et al., 2020. Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. Biogeosciences 17, 1343–1365. https://doi.org/10.5194/bg-17-1343-2020.
Khan, M.S., Jeon, S.B., Jeong, M.-.H., 2021. Gap-filling eddy covariance latent heat flux: inter-comparison of four machine learning model predictions and uncertainties in forest ecosystem. Remote Sens. (Basel) 13, 4976.
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. Change Biol. 26, 1499–1518.
Knorr, W., Kattge, J., 2005. Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. Glob. Change Biol. 11, 1333–1351.
Knox, S.H., Bansal, S., McNicol, G., et al., 2021. Identifying dominant environmental predictors of freshwater wetland methane fluxes across diurnal to seasonal time scales. Glob. Change Biol. 27, 3582–3604.
Lipson, M., Grimmond, S., Best, M., et al., 2022. Harmonized gap-filled datasets from 20 urban flux tower sites. Earth Syst. Sci. Data Discuss. 1–29.
Litvak, M., 2016a. AmeriFlux US-Seg Sevilleta Grassland. Lawrence Berkeley National Lab.(LBNL), Berkeley, CAUnited States.
Litvak, M., 2016b. AmeriFlux US-Ses Sevilleta Shrubland. Lawrence Berkeley National Lab.(LBNL), Berkeley, CAUnited States.
Lucas-Moffat, A.M., Huth, V., Augustin, J., et al., 2018. Towards pairing plot and field scale measurements in managed ecosystems: using eddy covariance to cross-validate CO2 fluxes modeled from manual chamber campaigns. Agric. For. Meteorol. 256, 362–378.
Mahabbati, A., Beringer, J., Leopold, M., et al., 2021. A comparison of gap-filling algorithms for eddy covariance fluxes and their drivers. Geosci. Instrum., Methods Data Syst. 10, 123–140.
McCalmont, J., Kho, L.K., Teh, Y.A., et al., 2021. Short-and long-term carbon emissions from oil palm plantations converted from logged tropical peat swamp forest. Glob. Change Biol. 27, 2361–2376.
McDermitt, D., Burba, G., Xu, L., et al., 2011. A new low-power, open-path instrument for measuring methane flux by eddy covariance. Appl. Phys. B 102, 391–405.
McKenzie, S.M., Pisaric, M.F., Arain, M.A., 2021. Comparison of tree-ring growth and eddy covariance-based ecosystem productivities in three different-aged pine plantation forests. Trees 35, 583–595.
Moffat, A.M., Beckstein, C., Churkina, G., et al., 2010. Characterization of ecosystem responses to climatic controls using artificial neural networks. Glob. Change Biol. 16, 2737–2749.
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.
Morin, T., Bohrer, G., Frasson R d, M., et al., 2014. Environmental drivers of methane fluxes from an urban temperate wetland park. J. Geophys. Res.: Biogeosci. 119, 2188–2208.
Noormets, A., Chen, J., Crow, T.R., 2007. Age-dependent changes in ecosystem carbon fluxes in managed forests in northern Wisconsin. USA. Ecosyst. 10, 187–203.
Orr, R., Murray, P., Eyles, C., et al., 2016. The North Wyke Farm Platform: effect of temperate grassland farming systems on soil moisture contents, runoff and associated water quality dynamics. Eur. J. Soil Sci. 67, 374–385.
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.
Papale, D., Valentini, R., 2003. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. Glob. Change Biol. 9, 525–535.
Pastorello, G., Trotta, 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.
Platt, J., 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv. Large Margin Classifiers 10, 61–74.

Reichstein, M., Camps-Valls, G., Stevens, B., et al., 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

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. Change Biol. 11, 1424–1439.

Saunois, M., Jackson, R., Bousquet, P., et al., 2016. The growing role of methane in anthropogenic climate change. Environ. Res. Lett. 11, 120207.

Schimel, D., Pavlick, R., Fisher, J.B., et al., 2015. Observing terrestrial ecosystems and the carbon cycle from space. Glob. Change Biol. 21, 1762–1776. https://doi.org/10.1111/gcb.12822.

Virtanen, P., Gommers, R., Oliphant, T.E., et al., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272.

Vuichard, N., Papale, D., 2015. Filling the gaps in meteorological continuous data measured at FLUXNET sites with ERA-Interim reanalysis. Earth Syst. Sci. Data 7, 157–171.

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., et al., 2018. Basic and extensible post-processing of eddy covariance flux data with REddyProc. Biogeosciences 15, 5015–5030. https://doi.org/10.5194/bg-15-5015-2018.

Xiao, J., Chen, J., Davis, K.J., Reichstein, M., 2012. Advances in upscaling of eddy covariance measurements of carbon and water fluxes. J. Geophys. Res.: Biogeosci. 117.

Yao, J., Gao, Z., Huang, J., et al., 2021a. Uncertainties in eddy covariance CO 2 fluxes in a semiarid sagebrush ecosystem caused by gap-filling approaches. Atmos. Chem. Phys. 21, 15589–15603.

Yao, J., Gao, Z., Huang, J., et al., 2021b. Uncertainties in eddy covariance CO 2 fluxes in a semiarid sagebrush ecosystem caused by gap-filling approaches. Atmos. Chem. Phys. 21, 15589–15603.

Zhu, S., Clement, R., McCalmont, J., et al., 2022. Stable gap-filling for longer eddy covariance data gaps: a globally validated machine-learning approach for carbon dioxide, water, and energy fluxes. Agric. For. Meteorol. 314, 108777.

Zhu, S., Xu, J., Yu, C., et al., 2021. DecSolNet: a noise resistant missing information recovery framework for daily satellite NO2 columns. Atmos. Environ. 246, 118143 https://doi.org/10.1016/j.atmosenv.2020.118143.