Received: 7 May 2019 | Accepted: 12 September 2019

DOI: 10.1111/gcb.14845

PRIMARY RESEARCH ARTICLE

Global Change Biology WILEY

Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis

Yeonuk Kim[1] | Mark S. Johnson[1,2] | Sara H. Knox[3] | T. Andrew Black[4] | Higo J. Dalmagro[5] | Minseok Kang[6] | Joon Kim[6,7,8] | Dennis Baldocchi[9]

[1]Institute for Resources Environment and Sustainability, University of British Columbia, Vancouver, BC, Canada

[2]Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, BC, Canada

[3]Department of Geography, University of British Columbia, Vancouver, BC, Canada

[4]Faculty of Land and Food Systems, University of British Columbia, Vancouver, BC, Canada

[5]Environmental Sciences Graduate Program, University of Cuiabá, Cuiabá, Brazil

[6]National Center for AgroMeteorology, Seoul, South Korea

[7]Department of Landscape Architecture & Rural Systems Engineering, Seoul National University, Seoul, South Korea

[8]Interdisciplinary Program in Agricultural & Forest Meteorology, Seoul National University, Seoul, South Korea

[9]Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA

Correspondence
Mark S. Johnson, Institute for Resources Environment and Sustainability, University of British Columbia, Vancouver, BC, Canada.
Email: mark.johnson@ubc.ca

Funding information
Brazilian Ministry of Science and Technology; California Department of Water Resources; British Columbia Knowledge Development Fund; Canada Foundation for Innovation; Korea government (MSIT), Grant/Award Number: 2018R1C1B6002917; Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RGPIN-2014-05065; National Research Foundation of Korea; US Department of Energy, Office of Science and Biological and Environmental Research; Brazilian National Institute for Science and Technology in Wetlands (INCT-INAU)

Abstract

Methane flux (FCH_4) measurements using the eddy covariance technique have increased over the past decade. FCH_4 measurements commonly include data gaps, as is the case with CO_2 and energy fluxes. However, gap-filling FCH_4 data are more challenging than other fluxes due to its unique characteristics including multidriver dependency, variabilities across multiple timescales, nonstationarity, spatial heterogeneity of flux footprints, and lagged influence of biophysical drivers. Some researchers have applied a marginal distribution sampling (MDS) algorithm, a standard gap-filling method for other fluxes, to FCH_4 datasets, and others have applied artificial neural networks (ANN) to resolve the challenging characteristics of FCH_4. However, there is still no consensus regarding FCH_4 gap-filling methods due to limited comparative research. We are not aware of the applications of machine learning (ML) algorithms beyond ANN to FCH_4 datasets. Here, we compare the performance of MDS and three ML algorithms (ANN, random forest [RF], and support vector machine [SVM]) using multiple combinations of ancillary variables. In addition, we applied principal component analysis (PCA) as an input to the algorithms to address multidriver dependency of FCH_4 and reduce the internal complexity of the algorithmic structures. We applied this approach to five benchmark FCH_4 datasets from both natural and managed systems located in temperate and tropical wetlands and rice paddies. Results indicate that PCA improved the performance of MDS compared to traditional inputs. ML algorithms performed better when using all available biophysical variables compared to using PCA-derived inputs. Overall, RF was found to outperform other techniques for all sites. We found gap-filling uncertainty is much larger than measurement uncertainty in accumulated CH_4 budget. Therefore, the approach used for FCH_4 gap filling can have important implications for characterizing annual ecosystem-scale methane budgets, the accuracy of which is important for evaluating natural and managed systems and their interactions with global change processes.

KEYWORDS
artificial neural network, comparison of gap-filling techniques, eddy covariance, machine learning, marginal distribution sampling, methane flux, random forest, support vector machine

1 | INTRODUCTION

The increase in atmospheric methane (CH_4) concentration is a global concern. Quantitating and predicting global sources and sinks of atmospheric CH_4 is crucial under changing climate, but uncertainties are higher than those for carbon dioxide (CO_2), particularly in relation to emission from wetlands and inland waters (Saunois et al., 2016). There is a need for direct and continuous ground-based measurements of methane flux (FCH_4) between land and atmosphere with the eddy covariance technique.

The eddy covariance technique has been widely used for quantitating carbon, water, and energy fluxes between land and atmosphere (Baldocchi, Falge, Gu, et al., 2001). In the past decade, flux towers which incorporate FCH_4 measurements have increased following the development of robust and low-power CH_4 gas analyzers (Baldocchi, 2014; McDermitt et al., 2011; Morin, 2019). As is the case with other trace gases, FCH_4 data based on the eddy covariance method commonly include data gaps due to unsuitable atmospheric conditions and system failures, with FCH_4 often having more gaps than CO_2 and H_2O. However, there is still no consensus on best practices for gap-filling FCH_4 (Dengel et al., 2013; Nemitz et al., 2018).

Robust FCH_4 gap-filling approaches are crucial since a small bias in gap-filled FCH_4 can cause a large difference in net greenhouse gas emissions when summing up fluxes in CO_2 equivalent (CO_2eq) terms. The radiative forcing of CH_4 is 34 times that of CO_2 on a 100-year basis when including climate–carbon feedbacks (Myhre et al., 2013). Furthermore, considering the sustained-flux global warming potential suggested for systems that have periods of sustained CH_4 emissions like wetlands, the radiative forcing of CH_4 is 45 CO_2eq on a 100-year basis (Neubauer & Megonigal, 2015).

There are a number of unique characteristics of FCH_4 which makes it more difficult to model it than other fluxes for gap-filling purposes. Firstly, FCH_4 depends nonlinearly on multiple drivers (Sturtevant et al., 2016), whereas the variations in CO_2 and energy fluxes can be explained by a few biophysical drivers, particularly by radiation. FCH_4 varies in relation to multiple drivers such as soil temperature, water table height (WTH), air pressure, other gas fluxes, and even friction velocity (*u*.; e.g., Brown, Humphreys, Moore, Roulet, & Lafleur, 2014; GaŽovic, Kutzbach, Schreiber, Wille, & Wilmking, 2010; Hatala, Detto, & Baldocchi, 2012; Herbst, Friborg, Ringgaard, & Soegaard, 2011; Kim, Verma, & Billesbach, 1999).

Secondly, most of the variance in CO_2 and energy fluxes stems from variability at the diel and seasonal timescales; in other words, diel and seasonal timescales of the power spectra are the dominant contributor to the total variance of the fluxes (e.g., Baldocchi, Falge, & Wilson, 2001; Stoy et al., 2005). For FCH_4, a larger part of the variance can occur at hourly or weekly–monthly timescales due to sporadic CH_4 emissions and the relationship between water table fluctuation (e.g., Chamberlain et al., 2018; Nemitz et al., 2018; Sturtevant et al., 2016). Some sites show a distinct diel pattern, but this is not always the case. Consistent diurnal and seasonal cycles are an important aspect for some gap-filling methods (e.g., Falge et al., 2001) and FCH_4 is challenging to model due to the lack of a diel and seasonal patterns in many cases.

Thirdly, interannual variability in CH_4 budgets is often not controlled by regularly measured biophysical drivers, but are controlled by slow changes in soil variables such as carbon content, salinity, and iron concentration (e.g., Chamberlain et al., 2018; Yuan et al., 2015). A FCH_4 gap-filling method based on an empirical regression such as machine learning (ML) algorithms requiring a long dataset for training may not reveal nonstationary trends in FCH_4 due to the slow changes in soil properties.

Furthermore, the impacts of biophysical drivers on FCH_4 are often lagged (Sturtevant et al., 2016). One of most critical examples is the impacts of drainage on FCH_4 in irrigated (flooded) rice fields. Midseason drainage is a common rice farming practice in Asia (Kim et al., 2016; Nishimura, Sawamoto, Akiyama, Sudo, & Yagi, 2004). Drainage facilitates methane oxidation, and thus FCH_4 in the remainder of the growing season is reduced after the week-long drainage period. Thus, drainage affects not only FCH_4 during the drainage period but also has a lagged effect persisting for a few weeks after the drainage period or even until the end of the growing season (Fumoto, Yanagihara, Saito, & Yagi, 2010; Kim et al., 2016). It is unclear that current gap-filling approaches can capture such a lagged effect (e.g., Reichstein et al., 2019).

It is challenging to select an appropriate gap-filling method for FCH_4 due to its characteristics and drivers. In some cases, using standard gap-filling methods for CO_2, water vapor, and energy fluxes have generated reasonable results, but this is not always the case. For instance, some researchers have applied a marginal distribution sampling (MDS) algorithm (Reichstein et al., 2005) for FCH_4 gap filling, which is a standard gap-filling method for CO_2 and H_2O fluxes (e.g., Alberto et al., 2014; Dalmagro et al., 2019; Housen et al., 2015; Tang et al., 2018). MDS has been widely used in the Fluxnet community by virtue of a robust online algorithm and an R package (Wutzler et al., 2018). MDS uses three key variables: global radiation, air temperature, and vapor pressure deficit. However, unlike CO_2 and energy fluxes, there is no direct causal relationship between FCH_4 and those variables. For this reason, recent studies have sought to replace inputs to MDS with different biophysical variables which have more direct relationships with FCH_4, such as soil temperature and WTH (e.g., Deventer et al., 2019; Kang et al., 2018; Taylor, Celis, Ledman, Bracho, & Schuur, 2018). However, this approach cannot fully resolve the multidriver dependency of FCH_4 because MDS was designed to use a small number of variables to reduce overfitting and to follow a perspective of parsimony. Another limitation of MDS for FCH_4 is that it cannot effectively predict missing data when gaps exceed 12 days (Moffat et al., 2007).

Because of such limitations for applying MDS to FCH_4, researchers have tested and applied ML algorithms like artificial neural networks (ANN) to gap-fill FCH_4 (e.g., Dengel et al., 2013; Knox et al., 2015). However, there is no consensus at present regarding standard input selection for ML algorithms (Nemitz et al., 2018). Including a large number of biophysical variables as inputs to ML algorithms can increase the complexity of the internal structures, and overfitting becomes more of an issue. Conversely, selection of too few variables or

suboptimal variables can decrease the quality of the resulting gap-filled FCH_4 since variables which include discriminative information can be disregarded. To avoid such issues, a few techniques have been used to guide selection of appropriate explanatory variables (i.e., feature selection; e.g., Dengel et al., 2013; Knox et al., 2015). However, feature extraction techniques such as principal component analysis (PCA) have rarely been tested with gap-filling algorithms.

Furthermore, previous studies for flux gap filling have mostly applied ANN only and rarely tested other ML algorithms (e.g., Dengel et al., 2013; Knox et al., 2015; Moffat et al., 2007; Papale & Valentini, 2003). Other studies have assessed flux upscaling methods which applied and compared various ML techniques (e.g., Tramontana et al., 2016; Xu et al., 2018), but it has been limited with respect to gap filling.

In this study, we compare the performance of MDS and three commonly used ML algorithms with different combinations of ancillary variables. ANN, support vector machine (SVM), and random forest (RF) were selected for this analysis. In addition, we applied PCA to ancillary biophysical variables as an input to the MDS and ML algorithms to resolve multidriver dependency of FCH_4 by reducing the number of inputs to the algorithms. We applied this approach to five benchmark FCH_4 datasets from temperate and tropical wetlands and rice paddies. The objectives of this research were to (a) evaluate a range of current and potential gap-filling methods for FCH_4; (b) explore the impact of bias introduced by the various gap-filling approaches on annual net FCH_4 estimates; and (c) assess characteristics that limit the utility of specific approaches, in particular cases, in order to provide guidance to the FCH_4 research community. Further details on the procedure can be found in Figure S1.

2 | MATERIALS AND METHODS

2.1 | Study sites

The five benchmark sites represent FCH_4 across various latitudes which are described in Table 1. Each site represents unique circumstances, which were valuable in testing the performance of a range of gap-filling approaches in relation to the challenging characteristics of FCH_4 described above. Three of these sites are wetland sites (abbreviation: WET) and the other two are irrigated rice paddy sites (abbreviation: RP). The number of site years included in this study is 12 years and 5 months. All sites used an LI-7700 open-path CH_4 analyzer (LI-COR Biosciences) and used different combinations of sonic anemometers and H_2O/CO_2 analyzers for the measurement of water vapor and CO_2 fluxes. We did not include sites operating closed-path CH_4 analyzers in this study since the open-path analyzer has been widely used by virtue of low-power requirement. It should be noted that there might be a slightly different performance of gap-filling algorithms for FCH_4 measured by the closed-path systems due to different random uncertainties between the two systems (Deventer et al., 2019; Peltola et al., 2014).

Corrected and filtered half-hourly flux data by each principal investigator (PI) were further filtered based on site-specific u_* thresholds. Following Nemitz et al. (2018), the same u_* thresholds used for CO_2 flux filtering were used for FCH_4 filtering. A moving point test was used to estimate the u_* thresholds, which is based on relating nighttime CO_2 fluxes to u_* (Gu et al., 2005; Papale et al., 2006). Detailed information on the instruments and quality controls can be found in the references in Table 1. In the following sections, we highlight one important characteristic for each site to emphasize how site-specific characteristics can influence gap-filling performance.

2.2 | WET.BB: A low u_* site

The Burns Bog wetland (WET.BB) is a temperate and permanent wetland near Vancouver, Canada (Ameriflux site CA-DBB; 49.12°N, 122.98°W). One of the important characteristics of WET.BB is that the u_* threshold value for this site was about 0.05–0.08 m/s (D'Acunha, Morillas, Black, Christen, & Johnson, 2019; Lee et al., 2017), which is lower than the typical range for wetlands (Barr et al., 2013; Falge et al., 2001). The average nighttime u_* in this site is also lower than the typical range. This means nighttime turbulence conditions at this site are not ideal. Previous research has reported that

TABLE 1 Site descriptions

Site code	Location	Lat., Lon.	Type	Study period	FCH_4 data coverage (%)		u_* Threshold (m/s)	Reference
					Day	Night		
WET.BB	Burns Bog, BC, Canada	49.12°N, 122.98°W	Rewetted peatland	April 2016 to August 2018	48	15	0.06	Lee et al. (2017)
WET.TW	Twitchell Island, CA, USA	38.10°N, 121.63°W	Alluvium wetland	January 2014 to December 2017	66	43	0.21	Chamberlain et al. (2018)
WET.PT	Pantanal, MT, Brazil	16.48°S, 56.41°W	Tropical wetland	March 2015 to November 2016	82	41	0.11	Dalmagro et al. (2018)
RP.HPK	Heanam, South Korea	34.47°N, 127.48°E	Rice paddy	June 2016 to June 2018	45	19	0.20	Lee, Kang, Kang, and Kim (2018)
RP.CRK	Cheorwon, South Korea	38.20°N, 127.25 °E	Rice paddy	May 2016 to October 2018	41	10	0.20	Huang et al. (2018)

Abbreviation: FCH_4, methane flux.

u_* and FCH_4 can have a physically based causal relationship by influencing the air–water gas exchange rate or the extent of the flux footprint (Herbst et al., 2011; Read et al., 2012). At the same time, low u_* can cause systematic measurement bias of FCH_4 due to a decrease in diffusivity and an increase in advection loss (Nemitz et al., 2018). In other words, the relationship between u_* and FCH_4 includes both actual and spurious relationships, but often the influence of spurious relationships can be reduced by introducing an u_* filter. However, u_* filtering can also remove valid measurements of FCH_4 that were repressed by low turbulence, thus causing an overestimation of FCH_4 (Herbst et al., 2011; Nemitz et al., 2018). At the WET.BB site, CH_4 gradients from water to air were higher than typical peatland pools, and open water evasion flux significantly contributed to the total FCH_4 (D'Acunha, 2019). This observation implies a low value of u_* could limit the transport of CH_4 to the atmosphere. Therefore, it is not clear that a threshold value for u_* at this site can selectively eliminate spurious effects without filtering physically repressed FCH_4 under low turbulence conditions.

2.3 | WET.TW: A nonstationary site

The Twitchell Island (WET.TW) site (Ameriflux site US-Tw4; 38.10°N, 121.63°W) is also a restored and permanent wetland in California, USA. WET.TW was restored in 2014, so the measurements have covered the whole history of wetland after the restoration (Chamberlain et al., 2018). FCH_4 at this site has been increasing during the measurement period, with a significant increase observed in 2017 (the FCH_4 time series can be found in Figure S2). This increase has not been fully explained by ancillary biophysical variables, but seemed to result from changes in soil properties (Chamberlain et al., 2018). If a gap-filling model cannot capture the nonstationarity and the increasing trend at this site, the gap-filling model can result in an underestimate of FCH_4 in later years (e.g., 2017) or an overestimate in earlier years (e.g., 2014). This site is, therefore, a suitable case to test the performance of gap-filling algorithms under nonstationary conditions.

2.4 | WET.PT: A site with unique power spectrum due to tropical climate

The WET.PT site is located in the northern Pantanal wetland near Poconé, Brazil (Ameriflux site BR-Npw; 16.50°S, 56.41°W), and has high seasonality with a pronounced dry period as well as an annual inundation period that ranges from 2 to 4 months (Dalmagro et al., 2018). WET.PT is thus a seasonal wetland with the WTH near or above the ground level only during wet season. CH_4 emissions occur almost exclusively during the wet season, with inundation acting like a switch (Dalmagro et al., 2019). In this site, soil respiration and FCH_4 are largely controlled by WTH and soil water content from hourly to weekly timescales (e.g., Johnson et al., 2013). Also, the seasonal variability of most biophysical variables in this site is much lower than that in temperate sites due to consistent warm conditions. The variance of FCH_4 at the hourly scale is the largest contributor to the total variance of FCH_4 at the WET.PT site, which was about 40%. High-frequency variability (e.g., hourly variability) is often hard to model (Knox et al., 2019). Therefore, this site is suitable to test the performance of gap-filling methods for sites that exhibit high-frequency fluctuations that can be challenging to predict.

2.5 | RP.HPK versus RP.CRK: Sites with delayed impacts of rice paddy drainage on FCH_4

The Heanam rice paddy (RP.HPK) and the Cheorwon rice paddy (RP.CRK) are both located in South Korea (AsiaFlux sites HPK and CRK). As is generally practiced in rice fields, the two rice paddies are flooded during the growing season, resulting in up to 4–5 months of inundation. Similar to WET.PT, CH_4 emissions occur exclusively following inundation due to the flood irrigation. The two rice paddy sites both included a midseason drainage event during the rice growing season. However, FCH_4 from the two sites responded differently to the drainage event. The impact of drainage on FCH_4 due to methane oxidation at RP.HPK was persistent, with FCH_4 not returning to the high rates observed prior to the drainage period (Figure S2). On the other hand, FCH_4 quickly returned to high rates shortly after the drainage period at the RP.CRK site (Figure S2). Therefore, the comparison of a gap-filling model with and without the use of lagged WTH variables in these two sites can reveal the usefulness of applying lags.

2.6 | Gap scenarios

In order to evaluate gap-filling algorithms, 10% of FCH_4 data was randomly removed and then gap filled. For this, we added short, medium, and long lengths of artificial gaps following Moffat et al. (2007). Here, the short-gap scenario is defined as a single half-hour data gap, the medium-gap scenario used data gaps of 32 hr (~1.5 days), and the long-gap scenario introduced artificial gaps of 12 days. Furthermore, we added an extremely long-gap scenario defined as 60 days (2 months) since FCH_4 measurements often include larger gaps, particularly during winter periods. Each gap scenario was permutated five times, resulting in a total of 20 FCH_4 time series with artificial gaps generated for each site. The random artificial gap generation and all of other analyses in this study were done using the R statistical language (R Core Team, 2018).

2.7 | PCA and input variables

Principal component analysis orthogonally rotates axes of the original variables to the directions of maximum variance to generate a new coordinate system which is a linear transform of the original dataset. The newly generated variables in the new coordinate system, which are referred to as principal components (PCs), are arranged in order of variance and are linearly uncorrelated with each other. Therefore, the first few PCs explain most of variance present in the original variables (Chen, Chu, & Li, 2012).

Principal component analysis is widely used with ML algorithms in many disciplines to resolve multicollinearity and to reduce the number of variables used as inputs for the ML algorithms (e.g., Howley, Madden,

O'Connell, & Ryder, 2006). In the field of micrometeorology, Chen et al. (2012) used PCA with a ML algorithm for latent heat flux gap filling. In this study, we applied PCA during preprocessing for the ML algorithms, as well as to prepare alternative inputs to the MDS algorithm.

Biophysical variables and other fluxes were quality controlled and gap filled by each PI (details in the references in Table 1). Gap-free biophysical variables and other fluxes related to FCH_4 were used for PCA after normalizing (Table 2). We selected the variables following Knox et al. (2016), and added LWOUT and H to consider surface temperature as a driver (e.g., Sachs, Giebels, Boike, & Kutzbach, 2010). Often, including the partitioned CO_2 fluxes instead of NEE can improve the performance of a gap-filling algorithm due to the high correlation between gross primary production and FCH_4 but we did not partition NEE into gross primary production and ecosystem respiration for use as input variables. This is because we tried to minimize uncertainty introduced by partitioning algorithms (e.g., Keenan et al., 2019). Also, various combinations of ancillary variables would be valuable to test (e.g., without flux variables vs. all available inputs), but it was beyond the scope of the current study.

2.8 | Spectral analysis

Analyzing the structure of a time series in the frequency domain is important to evaluate gap-filling approaches since the spectrum of frequencies can help identify how well a model represents dynamics in various timescales (Baldocchi, Falge, & Wilson, 2001). We computed the power spectra of PCs, and gap-filled FCH_4 to interpret results. For flux time series, wavelet transforms have been widely applied in recent research since it is suitable for nonstationary data (Stoy et al., 2005). However, we applied the Fourier transform to estimate the power spectrum rather than using a wavelet transform. The Fourier transform can compute high-resolution powers at higher frequencies (i.e., hourly scale, which is an important scale for FCH_4). Also, fluxes and ancillary variables in this analysis were near stationary except for gap-filled FCH_4 at WET.TW and RP.HPK, so the stationary assumption in the Fourier transform was generally satisfied. We used an algorithm which employs sine multitapers that can estimate a power spectrum in near-stationary conditions (Barbour & Parker, 2014). We integrated the power spectrum over frequency for the four timescales in Table 2 and estimated the percentage of total variance at each time scale (i.e., hourly, diel, weekly–monthly, and seasonal). The spectral analysis is discussed in detail in the Supporting Information (Figures S3–S6; Table S1).

2.9 | Gap filling using MDS

The MDS algorithm applies a look-up table method first, and then applies the mean diurnal curve method if necessary (Reichstein et al., 2005; Wutzler et al., 2018). In the look-up table method, a missing flux data point is replaced as the average of flux values within a certain time window under similar meteorological conditions. If the look-up table fails to gap fill due to missing ancillary data, the mean diurnal curve method is applied to fill the gap, which replaces the gap as the average flux value for the same time of day of the adjacent days.

TABLE 2 List of drivers, models, and timescales for parameters related to methane fluxes with abbreviations used in the text and figures

Biophysical variables and associated fluxes (mechanistic relationship)	
NEE	Net ecosystem exchange of CO_2 (carbon substrates and respiration)
LE	Latent heat flux (plant-mediated CH_4 transport related to transpiration)
H	Sensible heat flux (redundant with temperature measurements)
Rg	Global radiation (redundant with LE, NEE)
LWOUT	Outgoing long wave radiation (enzyme kinetics)
AirT	Air temperature (enzyme kinetics)
SoilT	Soil temperature (enzyme kinetics)
rH	Relative humidity (redundant with LE)
VPD	Vapor pressure deficit (redundant with LE and air temperature)
u_*	Friction velocity (CH_4 transport related to turbulence)
Pa	Air pressure (bubbles and ebullition)
WTH	Water table height (anaerobic condition or barrier for the CH_4 transport)
Gap-filling algorithm (inputs)	
MDS_{-RAV}	Marginal distribution sampling (conventional inputs: Rg, AirT, VPD)
MDS_{-SNW}	Marginal distribution sampling (alternative inputs: SoilT, NEE, WTH)
MDS_{-PC3}	Marginal distribution sampling (inputs: the first three principal components)
MDS_{-PC5}	Marginal distribution sampling (inputs: the first five principal components)
ANN_{-All}	Artificial neural network (inputs: all measured ancillary variables)
ANN_{-PCA}	Artificial neural network (inputs: the first seven principal components)
RF_{-All}	Random forest (inputs: all measured ancillary variables)
RF_{-PCA}	Random forest (inputs: the first seven principal components)
SVM_{-All}	Support vector machine (inputs: all measured ancillary variables)
SVM_{-PCA}	Support vector machine (inputs: the first seven principal components)
Time scale definition	
Hourly	30 min to 23 hr
Diel	23–25 hr
Weekly–monthly	25 hr to 2 months
Seasonal	2 months to years

Here, FCH_4 was gap filled using MDS with conventional inputs described above. We refer to this approach as $MDS_{\_RAV}$ (i.e. MDS with Rg, AirT, and VPD). We also evaluated the performance of the MDS algorithm using an alternative combination of inputs (i.e., SoilT, NEE, and WTH, referred to here as $MDS_{\_SNW}$); this combination of ancillary data can be more strongly related to FCH_4 fluxes than the conventional inputs (e.g., Kang et al., 2018; Taylor et al., 2018). However, $MDS_{\_RAV}$ and $MDS_{\_SNW}$ cannot resolve the multi-driver dependency of FCH_4. In this regard, we used outputs from PCA as inputs to the MDS algorithm since the first few PCs retain most of the variances of the whole drivers. This approach, referred to here as $MDS_{\_PC3}$ uses the first three PCs. Furthermore, we modified the MDS algorithm to increase the number of inputs to use the first five PCs as inputs to MDS (i.e., $MDS_{\_PC5}$). Detailed information on the algorithm modification can be found in Figure S7. Also, detailed margin values which determine the bin size for the look-up tables can be found in Table S2. The MDS algorithm was employed from "ReddyProc" R package (Wutzler et al., 2018).

2.10 | Gap filling using ML algorithms

Artificial neural networks, RF, and SVM were applied for FCH_4 gap filling. We tested two combinations of inputs for each algorithm to test the usefulness of PCA as inputs for ML algorithms used for FCH_4 gap filling. First, all related biophysical variables in Table 2 were used as inputs to the ML algorithms, and then three fuzzy variables (i.e., day of year, and cosine and sine functions with a year-long wavelength) were also added following Knox et al. (2016). We named these approaches as $ANN_{\_All}$, $RF_{\_All}$, and $SVM_{\_All}$. The other combination of inputs tested for the three ML algorithms was the first seven PCs, where the PCA for ML incorporated the three fuzzy variables (i.e., $ANN_{\_PCA}$, $RF_{\_PCA}$, and $SVM_{\_PCA}$).

One of the major challenges of using ML algorithms is that it is difficult to decide suitable values for required parameters and network architectures. For instance, the ANN algorithm requires a user to decide the number of hidden layers and the number of nodes in each layer (e.g., Dengel et al., 2013). To overcome this challenge, we used a threefold cross-validation repeated three times for each site and for each input combination using the "Caret" R package (Kuhn et al., 2019). The parameters include the number of nodes in each hidden layer for ANN, the number of variables randomly sampled (i.e., mtry) for RF, and inverse kernel width (i.e., sigma), as well as the cost of constraint violation (i.e., C) for SVM. The results from parameter tuning can be found in Table S3. In the following section, we briefly described the characteristics of each ML algorithm we applied.

2.11 | Artificial neural networks

The ANN algorithm is the most popular ML algorithm for flux gap filling (Dengel et al., 2013; Knox et al., 2015; Moffat et al., 2007; Papale & Valentini, 2003). ANN is composed of a few layers which include several nodes. Nodes in different layers are connected by weights which are estimated during the training. We employed two hidden layers following Knox et al. (2016) considering the similarity of input decisions with the research. We used "neuralnet" R package (Günther & Fritsch, 2010), the resilient backpropagation algorithm, following Dengel et al. (2013). We trained the neural network 1,000–10,000 times, which is comparable with Jammet, Crill, Dengel, and Friborg (2015). The training and prediction of FCH_4 was repeated several times. The median of 10 independent predictions was used to gap-fill FCH_4 for datasets including the artificial gaps, and the median of 20 independent predictions was used for gap filling of the original FCH_4 datasets which did not include artificial gaps, and included real gaps only.

2.12 | Random forest

Random forest algorithm introduced by Breiman (2001) is an ensemble method of regression tree. In flux upscaling studies, the accuracy of RF is better than or comparable with other ML algorithms (Bodesheim, Jung, Gans, Mahecha, & Reichstein, 2018; Tramontana, Ichii, Camps-Valls, Tomelleri, & Papale, 2015; Xu et al., 2018), but no previous studies have tested RF for the purpose of gap filling. The algorithm creates bootstrapped datasets and then generates independent regression trees using randomly sampled variables at each split node. Then, RF aggregates the prediction results of the individual trees. We created 400 regression trees for each case using the "randomForest" R package (Liaw & Wiener, 2002). One of the advantages of RF is that the algorithm can estimate "variable importance" by permuting a predictor variable and then determining by how much the prediction error increases.

2.13 | Support vector machine

Support vector machine for regression analysis (or support vector regression: SVR) was developed by Cortes and Vapnik (1995). SVM algorithm projects the input vectors into the higher dimensional feature space with a predefined kernel function to change a nonlinear problem into a linear regression problem. Also, SVM introduces slack variables to overcome nonseparability due to noise or extreme values in the dataset. In this study, we used the radial basic kernel function based on previous SVM studies for flux upscaling (e.g., Ichii et al., 2017; Xu et al., 2018). The "kernlab" R package was used in this study (Karatzoglou, Smola, Hornik, & Zeileis, 2004).

2.14 | Performance measures

We used Taylor diagrams to evaluate the performance of gap-filling approaches (Taylor, 2001) which can concisely summarize the root mean square error (RMSE), the correlation coefficient (*R*), and the ratio of the standard deviation (*SD*) of a model to that of the observation. The Taylor diagrams were generated using the "openair" R package (Carslaw & Ropkins, 2012). Taylor diagrams are discussed in more detail in the Supporting Information file. To investigate the results further, correlation matrices between residual and biophysical drivers were estimated using "corrplot" R package (Wei et al., 2017).

For bias evaluation, we used different strategies for half-hourly data and annual sums. For half-hourly FCH_4 data, bias was defined as the average amount by which the observation is greater or less than the modeled value (i.e., the residual). As for annual net FCH_4, it is impossible to define true values since the original FCH_4 data include actual data gaps. Therefore, we compared gap-filled annual sums with the ensemble median for each site. Also, we investigated how the power spectrum distribution explains the bias introduced by each model when determining annual sums.

2.15 | Random uncertainty

A gap-filling algorithm is often used to estimate random flux uncertainty since the model error has been believed to approach the measurement uncertainty of the eddy covariance method (Knox et al., 2019; Moffat et al., 2007; Richardson et al., 2008). The random error can be estimated using the residual of a gap-filling model (i.e., model residual method). We compared the model residual of the best approach with the random uncertainty estimated following Finkelstein and Sims (2001) in a few sites. The mean absolute error (MAE) was also used to compare model performance with the random uncertainty. Then, we quantified random uncertainty of cumulative fluxes using the model residual method by applying Monte Carlo simulation following Knox et al. (2019).

3 | RESULTS

3.1 | Performance of gap-filling algorithms

The overall performance of MDS for FCH_4 gap filling improved for all gap scenarios when using PCs as inputs instead of Rg, AirT, and VPD (Table 3; Figure 1 left). For the short-gap scenario, the RMSE of $MDS_{\_PC3}$ was 9% lower than that of $MDS_{\_RAV}$ on average of all sites. Performances of $MDS_{\_PC3}$ and $MDS_{\_PC5}$ were almost identical, which implies that increasing the number of PCs as inputs to MDS did not improve overall performance (Table 3). As for $MDS_{\_SNW}$, only the RP.HPK site showed improved performance over $MDS_{\_RAV}$ (Figure 1 right). The performance of MDS was lowest at WET.PT, and there was no evidence of performance improvement when including other inputs to MDS (Figure 1 blue arrows). The poor performance of MDS for WET.PT may have been caused by the high hourly scale FCH_4 variability, which is not well represented by the MDS method, which we describe further in the next section. The performances of MDS for the extremely long gap scenario were excluded in Figure 1 to maintain legibility, but are summarized in Table 3.

Unlike MDS, performances of ML algorithms generally decreased when PCA was used as input parameters (Figure 2). The arrows indicate a reduction in model accuracy, particularly for the ANN algorithm. For instance, the coefficient of determination (R^2) for $ANN_{\_All}$ is 0.1 higher than that of $ANN_{\_PCA}$ in WET.PT site for the long-gap scenario. These results indicate that PCA can be useful for deriving input parameters for MDS, but not for ML algorithms.

The first seven PCs, the inputs of ML algorithms, retained 95%–97% of the variance present in the ancillary datasets, but the seven PCs lost a relatively large portion of hourly and weekly–monthly variance up to 13% (details in Supporting Information). The loss of variance at hourly and weekly–monthly scales resulting from the application of PCA may reduce the performance of ML algorithms since the variability of FCH_4 at these timescales contributes notably to the total variance. Also, the order of PCs was not always identical with the importance of variables for training (an example in Figure S8), which implies that the disregarded PCs may include discriminative information. PCA is generally useful when the number of explanatory variables is too large to be handled by a predictive model (such as MDS in this study). The three selected ML algorithms handle the suite of explanatory variables (12–14). If one uses dozens of variables with the ML algorithm, the ML performance could improve with PCA, but this is not typical for FCH_4 gap filling.

We then compared different gap-filling algorithms using the best performing combinations of input parameters: $MDS_{\_PC3}$, $ANN_{\_All}$, $RF_{\_All}$, and $SVM_{\_All}$ (Figure 3). $RF_{\_All}$ consistently outperformed the other methods for all sites and all gap scenarios, with R^2 value ranging from .82 to .97 (Figure 3; Table 3). $ANN_{\_All}$ also showed consistently good performance for short-, medium-, and long-gap scenarios (R^2 = .72–.94), but it was inferior to $RF_{\_All}$ in all cases. Performances of $SVM_{\_All}$ were comparable with $ANN_{\_All}$ for the wetland sites, but were inferior for the rice paddy sites. Performances of $MDS_{\_PC3}$ were comparable with ML algorithms for the short-gap scenario, but the performances of $MDS_{\_PC3}$ decreased as gap lengths increased, which was also observed by Moffat et al. (2007) for NEE gap filling.

With extremely long gaps, the performance of MDS was much lower than for the ML techniques (Figure 3 open symbols). Also, performances of ANN and SVM were significantly lower than those for shorter gaps. In the WET.TW and RP.HPK sites, the performance reductions in ANN and SVM appear to be related to high interannual variability for these sites. Although there were small performance reductions for RF with extremely long gaps, the RF model provided reasonable results (R^2 = .85–.95) for all sites.

3.2 | Measurement uncertainty versus absolute error of RF

We compared the estimated random error with the absolute value of model residual of RF, the highest performance model, in the two rice paddy sites (Figure 4). The absolute error of $RF_{\_All}$ should be closest to the random error as the model performance approaches the noise limit, which was observed in the RP.CRK site (Figure 4 left). The mean value of random error was 20 nmol m^{-2} s^{-1}, and the MAE of $RF_{\_All}$ was 23 nmol m^{-2} s^{-1}. For the RP.HPK site, there was room to improve the gap-filling model since the model residual was relatively large compared to the random error (Figure 4 right). The MAE of $RF_{\_All}$ was 10 nmol m^{-2} s^{-1} which is 140% of the random error in this site (7 nmol m^{-2} s^{-1}).

TABLE 3 Coefficients of determination (R^2) for a range of gap-filling algorithms using standard input parameters and those derived from principal component analysis for short- (*s*), medium- (*m*), long- (*l*), and extremely long (el)-gap scenarios

Site	Gap scenario	MDS _RAV	MDS _SNW	MDS _PC3	MDS _PC5	ANN _All	ANN _PCA	RF _All	RF _PCA	SVM _All	SVM _PCA
WET.BB	*s*	.61 ± .03	.66 ± .03	.70 ± .03	.70 ± .04	.80 ± .03	.77 ± .02	**.92** ± .03	**.91** ± .01	.78 ± .02	.75 ± .01
	m	.58 ± .05	.61 ± .05	.62 ± .05	.63 ± .05	.81 ± .05	.79 ± .06	**.91** ± .03	**.93** ± .03	.79 ± .05	.77 ± .05
	l	.30 ± .22	.36 ± .22	.34 ± .20	.32 ± .20	.76 ± .19	.74 ± .19	.88 ± .06	.89 ± .08	.73 ± .19	.72 ± .19
	el	.07 ± .07	.07 ± .1	.06 ± .07	.07 ± .06	.64 ± .14	.56 ± .14	.85 ± .05	.76 ± .04	.56 ± .12	.50 ± .12
WET.TW	*s*	.71 ± .01	.76 ± .01	.76 ± .01	.77 ± .01	.79 ± .01	.74 ± .01	**.92** ± .01	**.93** ± .01	.77 ± .01	.74 ± .01
	m	.71 ± .04	.72 ± .03	.73 ± .03	.72 ± .03	.78 ± .02	.76 ± .02	**.92** ± .00	**.93** ± .01	.78 ± .02	.75 ± .02
	l	.69 ± .13	.69 ± .16	.69 ± .12	.67 ± .11	.82 ± .09	.77 ± .10	**.92** ± .03	**.93** ± .03	.82 ± .09	.78 ± .10
	el	.34 ± .15	.30 ± .16	.34 ± .13	.33 ± .13	.65 ± .09	.56 ± .10	.87 ± .03	.86 ± .03	.64 ± .09	.56 ± .10
WET.PT	*s*	.54 ± .05	.49 ± .04	.48 ± .04	.49 ± .05	.72 ± .03	.66 ± .04	.82 ± .06	.89 ± .05	.67 ± .05	.61 ± .05
	m	.54 ± .05	.49 ± .06	.52 ± .05	.52 ± .05	.77 ± .04	.71 ± .04	**.91** ± .01	**.90** ± .01	.72 ± .05	.67 ± .05
	l	.37 ± .09	.24 ± .13	.36 ± .10	.35 ± .09	.73 ± .07	.62 ± .06	.88 ± .05	.88 ± .06	.62 ± .09	.55 ± .08
	el	.15 ± .12	.14 ± .14	.16 ± .10	.15 ± .10	.56 ± .11	.50 ± .09	.88 ± .03	.85 ± .03	.50 ± .10	.46 ± .09
RP.HPK	*s*	.75 ± .01	.81 ± .04	.81 ± .03	.82 ± .03	**.92** ± .01	.89 ± .00	**.97** ± .00	**.96** ± .01	.69 ± .03	.66 ± .03
	m	.59 ± .05	.70 ± .05	.64 ± .05	.64 ± .04	**.94** ± .02	**.91** ± .02	**.97** ± .01	**.96** ± .01	.63 ± .08	.63 ± .09
	l	.54 ± .10	.64 ± .17	.59 ± .18	.57 ± .18	.87 ± .06	.83 ± .08	**.97** ± .01	**.95** ± .01	.67 ± .06	.70 ± .08
	el	.11 ± .20	.10 ± .17	.11 ± .21	.11 ± .21	.52 ± .40	.44 ± .38	**.93** ± .2	**.90** ± .09	.33 ± .30	.35 ± .31
RP.CRK	*s*	.74 ± .02	.72 ± .01	.79 ± .02	.78 ± .01	.85 ± .01	.82 ± .01	**.93** ± .01	**.93** ± .01	.78 ± .01	.76 ± .01
	m	.64 ± .07	.61 ± .07	.65 ± .07	.64 ± .07	.84 ± .03	.81 ± .02	**.93** ± .00	**.93** ± .01	.75 ± .03	.71 ± .04
	l	.45 ± .14	.51 ± .10	.56 ± .13	.56 ± .13	.86 ± .02	.82 ± .03	**.95** ± .01	**.93** ± .01	.79 ± .04	.72 ± .07
	el	.29 ± .17	.33 ± .17	.43 ± .19	.43 ± .19	.87 ± .31	.83 ± .30	**.95** ± .09	**.94** ± .07	.78 ± .30	.72 ± .30

Note: Median and standard deviation (*SD*) of R^2 for the five scenarios of each gap length is reported (bold: over .90).

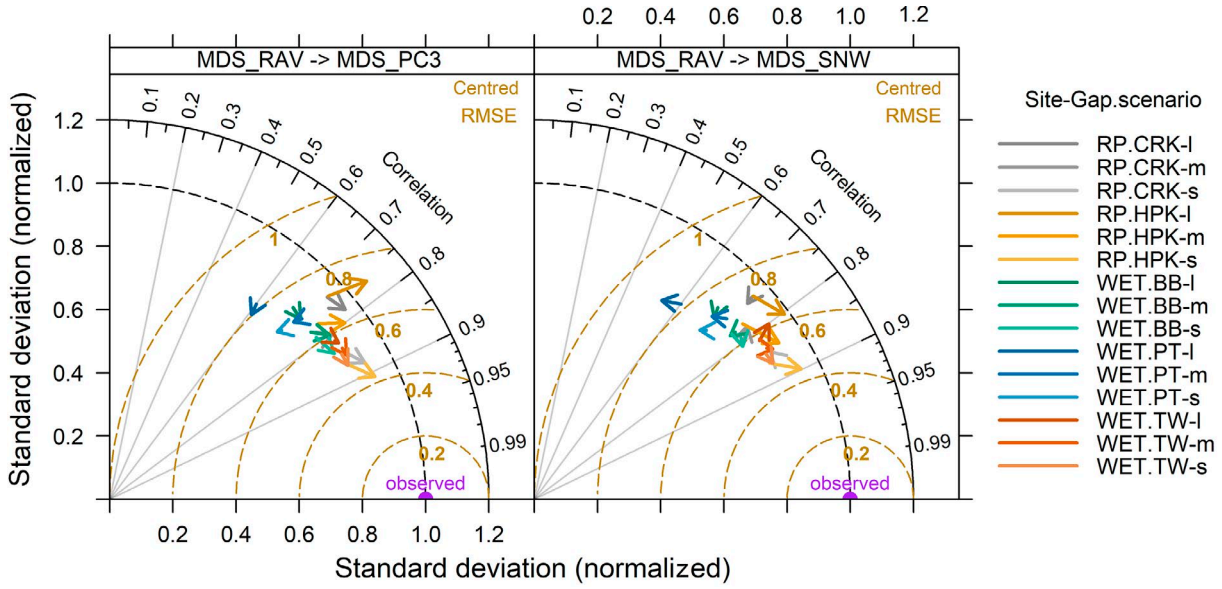

FIGURE 1 Taylor diagram showing marginal distribution sampling (MDS) performance changes when changing input variables. The starting points of the arrows represent root mean square error (RMSE), *r*, *SD* for $MDS_{\_RAV}$. The end points of the arrows represent those statistical values for $MDS_{\_PC3}$, and $MDS_{\_SNW}$ from left to right panels. If an arrow points toward observed values, it means model performance was improved. The color indicates each site, and the brightness indicates short- (*s*), medium- (*m*) and long (*l*)-gap scenarios [Colour figure can be viewed at wileyonlinelibrary.com]

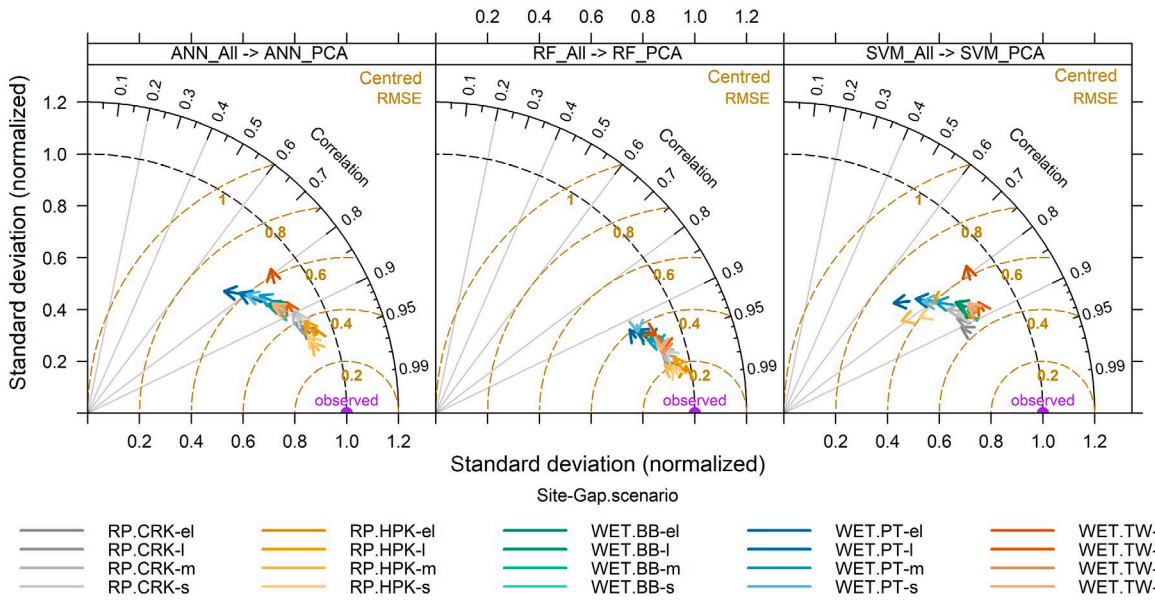

FIGURE 2 Taylor diagram showing change in machine learning algorithm performances when changing input variables. The starting points of the arrows represent root mean square error (RMSE), *r*, *SD* for $ANN_{\_All}$, $RF_{\_All}$, and $SVM_{\_All}$ from left to right panels. The end points of the arrows represent those statistical values for $ANN_{\_PCA}$, $RF_{\_PCA}$, and $SVM_{\_PCA}$. The color indicates each site, and the brightness indicates short- (*s*), medium- (*m*), long- (*l*), and extremely long (el)-gap scenarios [Colour figure can be viewed at wileyonlinelibrary.com]

3.3 | Dependency of the algorithm performance on timescale variability

Knox et al. (2019) found that performances of ANN for FCH_4 gap filling varied across various sites and the ANN's performances were low when the percentage of total FCH_4 variance at the hourly scale was high. In other words, a site showing high FCH_4 variability at the hourly scale is less predictable. In this research, we also found negative correlations between models' R^2 values and the percentage of total variance at the hourly scale for all gap-filling algorithms, but the slopes were different (Figure 5). For instance, the slopes of MDS were two to five times steeper than those of ML algorithms which means that performance of MDS is more sensitive to hourly variability. The performance of MDS was particularly lower than ML algorithms in

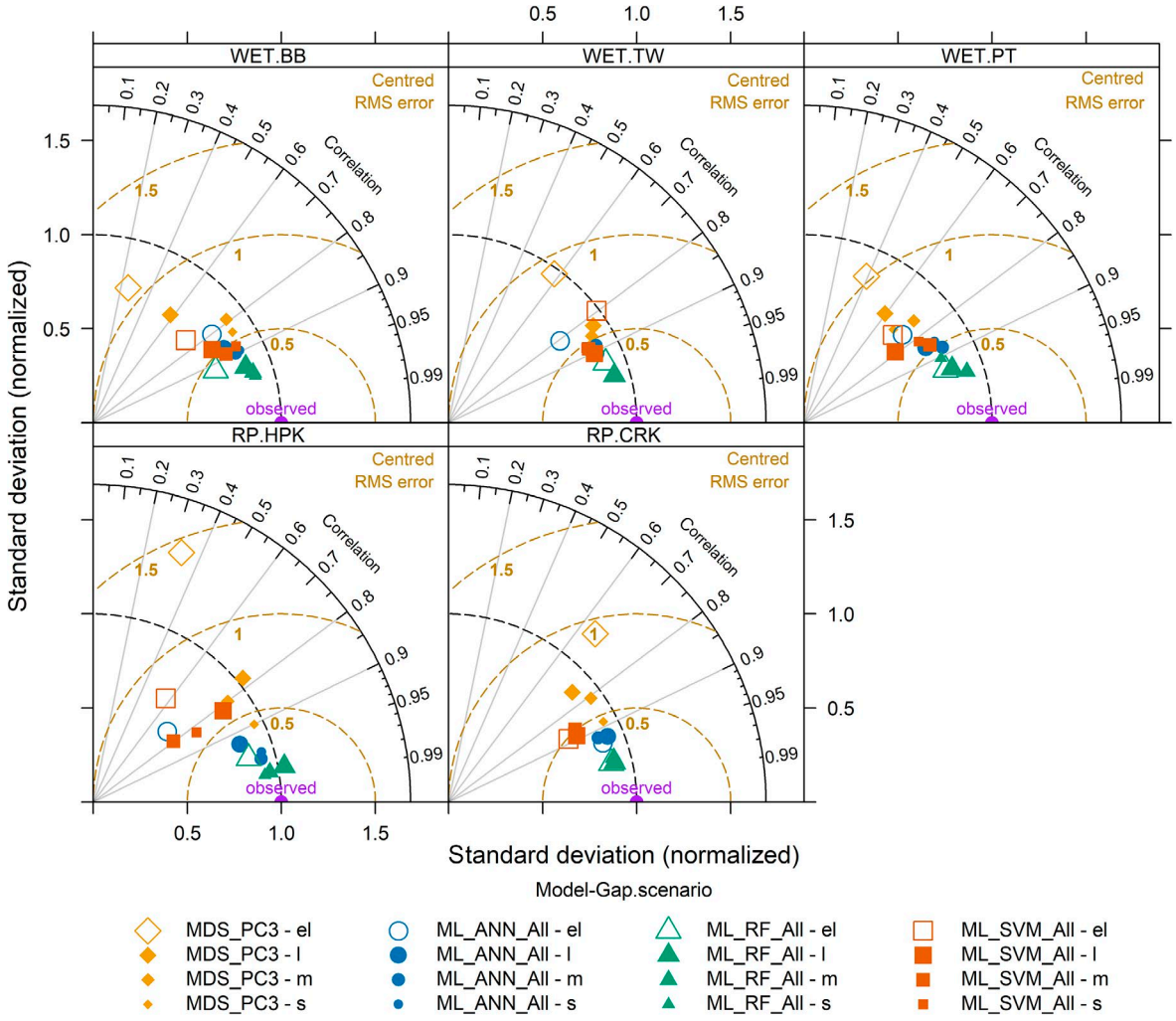

FIGURE 3 Taylor diagram comparing performances of algorithms for each site using the best performing combinations of input parameters. Different colors indicate different algorithms, and symbol sizes indicate lengths of gap. The extremely long gap (el) is highlighted with open symbols. The points represent median statistics of the five scenarios for each gap length [Colour figure can be viewed at wileyonlinelibrary.com]

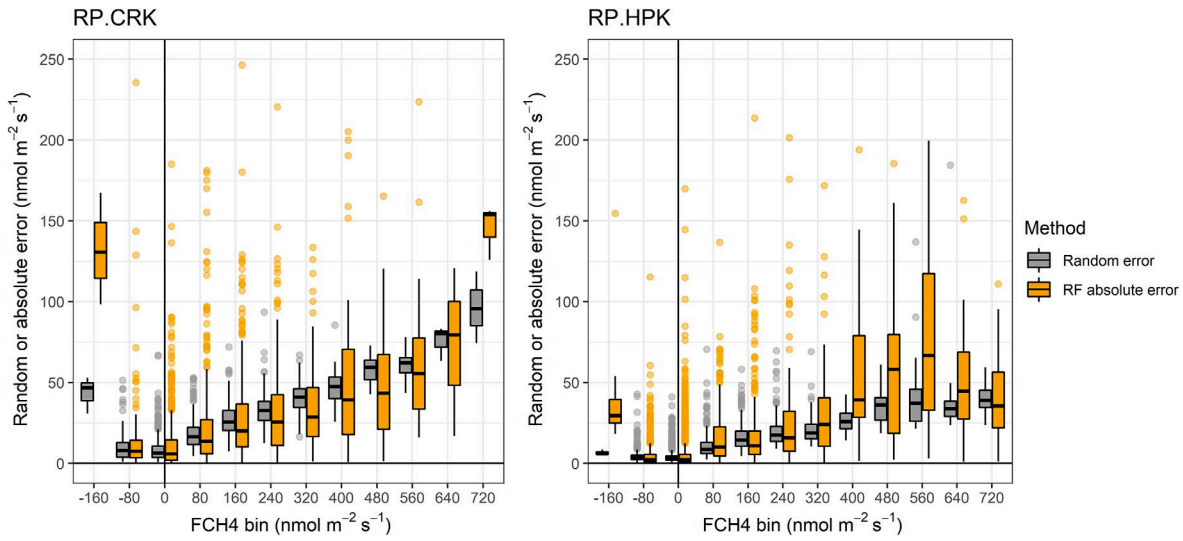

FIGURE 4 Methane flux (FCH_4) measurement and model uncertainties as a function of flux magnitude in the two rice sites. The yellow boxplot is the absolute error of $RF_{\_All}$, and the gray boxplot is estimated random error by Finkelstein and Sims (2001) [Colour figure can be viewed at wileyonlinelibrary.com]

WET.PT site where about 40% of FCH_4 variance was observed at the hourly scale.

3.4 | Bias of gap-filling algorithms

Overall, the half-hourly bias errors were not affected by usage of PCA but differed for each algorithm (Figure 6). SVM showed a consistently negative bias for all gap scenarios. Particularly, the negative bias of SVM was distinct for the two rice paddy sites (Figure 6 blue and vermillion colors). The bias errors of MDS increased for increasing gap lengths. The bias of ANN also increased for increasing gap lengths, but to a lesser degree. RF showed the best performance in terms of bias, and this algorithm was rarely affected by the length of gaps.

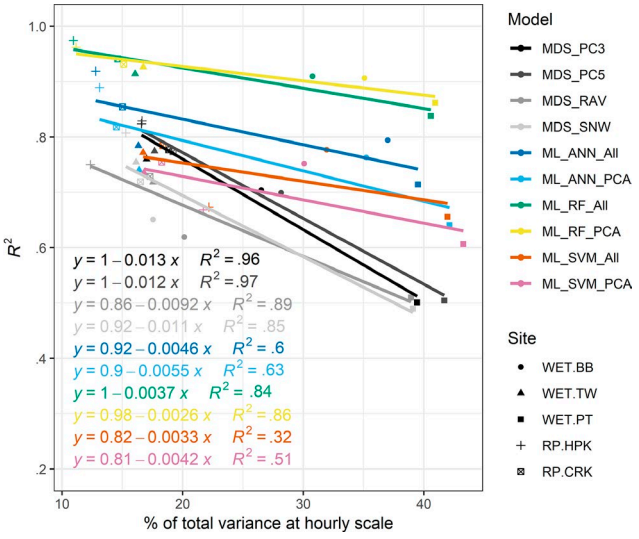

FIGURE 5 Relationship between coefficients of determination (R^2) for a range of gap-filling algorithms and high-frequency methane flux variability at each site (i.e., the percentage of total variance at the hourly scale). The short-gap scenario was used for this analysis [Colour figure can be viewed at wileyonlinelibrary.com]

We then examined the effect of bias of gap-filling approaches on annual methane budgets with the ensemble median for each site-year. The gap-filled annual sums in Table 4 were generated by the original datasets which did not include the artificial gaps, and the actual gaps in the original dataset were gap filled by each approach. The percentage of gap and accumulated random uncertainty (g CH_4-C m^{-2} $year^{-1}$) estimated by the model residual of $RF_{\_All}$ is also reported in Table 4. If the annual FCH_4 value was ±1.5 g CH_4-C m^{-2} $year^{-1}$ of the median value, the gap-filling model was classified as biased. Annual FCH_4 values more than ±2 g CH_4-C m^{-2} $year^{-1}$ away from the median value were classified as highly biased. A previous study on NEE gap filling suggested a median ±25 g CO_2-C m^{-2} $year^{-1}$ as a reliability criterion (Moffat et al., 2007), which is based on the accumulated NEE uncertainty (random + filled) estimated by Richardson and Hollinger (2005). Similarly, we defined the reliability criteria referring to the accumulated FCH_4 uncertainties estimated in Knox et al. (2019), who found that uncertainties in annual CH_4 estimates due to gap filling and random errors across 60 sites were on average ±2.1 g CH_4-C m^{-2} $year^{-1}$ at the 95% confidence interval.

Standard deviations of annual CH_4 budgets gap filled by each approach ranged from 0.56 to 2.62 g CH_4-C m^{-2} $year^{-1}$, while accumulated random uncertainties were within 0.13 g CH_4-C m^{-2} $year^{-1}$ (Table 4). This result implies that the total uncertainty of the gap-filled CH_4 annual budget would be mainly derived from the gap filling rather than the random error. The annual budgets generated by $MDS_{\_RAV}$, $MDS_{\_PC3}$, $MDS_{\_PC5}$, ANN, and RF were consistently within the reliable range (Table 4). Conversely, SVM or $MDS_{\_SNW}$ had several negative or positive biases, respectively. The negative bias of SVM was particularly enhanced in the two rice paddy sites.

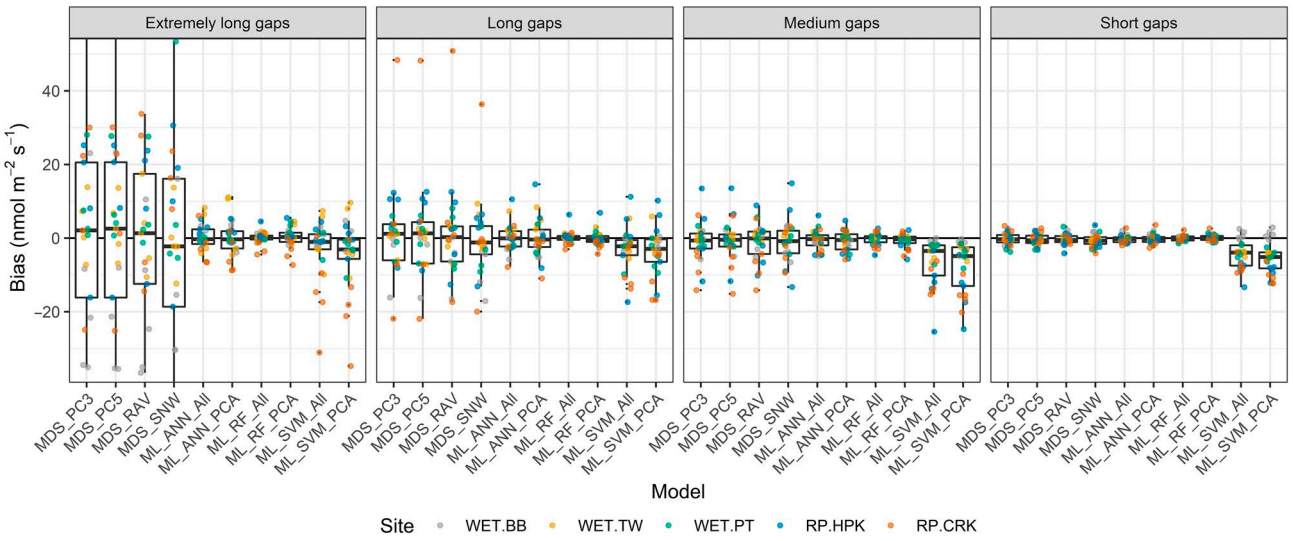

FIGURE 6 Half-hourly bias error. The box plots indicate each model's bias errors for the extremely long-, long-, medium-, and short-gap scenarios. The jitter points represent the five permutated scenarios with a different color for each site [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 Gap-filled annual sum of FCH_4 (g CH_4-C m^{-2} $year^{-1}$) using each approach [Colour figure can be viewed at wileyonlinelibrary.com]

Site-year (months)	% gap	Random error	$MDS_{\_RAV}$	$MDS_{\_SNW}$	$MDS_{\_PC3}$	$MDS_{\_PC5}$	$ANN_{\_All}$	$ANN_{\_PCA}$	$RF_{\_All}$	$RF_{\_PCA}$	$SVM_{\_All}$	$SVM_{\_PCA}$	Median	*SD*
WET.BB-16 (9)	72	0.04	12.8	13.8	12.7	12.6	12.7	12.8	13.0	12.3	13.5	14.0	12.8	0.56
WET.BB-17	72	0.05	**17.6**	**19.6**	**18.0**	**17.7**	**14.2**	14.7	14.9	14.5	15.7	16.1	15.9	1.83
WET.BB-18 (8)	60	0.04	12.8	12.8	12.5	12.4	10.8	10.8	11.1	11.3	11.6	11.6	11.6	0.79
WET.TW-14	44	0.06	17.0	16.7	16.4	16.2	17.3	17.2	17.1	17.2	16.5	16.3	16.9	0.42
WET.TW-15	45	0.07	27.4	27.8	26.9	26.8	26.7	26.2	27.1	26.7	26.2	25.4	26.8	0.67
WET.TW-16	46	0.08	33.8	34.1	33.6	33.6	33.8	33.5	33.8	34.0	33.1	32.6	33.7	0.45
WET.TW-17	47	0.1	48.7	**48.9**	48.5	48.5	47.2	46.5	47.5	47.2	**45.8**	**44.8**	47.3	1.36
WET.PT-15 (10)	40	0.12	11.4	11.8	11.6	10.7	11.2	10.0	11.2	10.6	10.6	9.6	10.9	0.70
WET.PT-16 (11)	38	0.12	16.6	17.3	17.1	15.8	16.7	15.2	16.1	15.6	15.9	14.7	16.0	0.83
RP.HPK-16 (6)	75	0.07	**23.5**	**23.3**	21.6	21.8	20.6	**20.0**	21.8	22.6	**16.0**	**16.5**	21.7	2.62
RP.HPK-17	73	0.07	20.0	19.9	18.8	18.9	20.3	20.0	21.3	**22.8**	**17.4**	**18.3**	20.0	1.54
RP.HPK-18 (6)	76	0.03	2.3	2.2	2.2	2.2	3.4	3.3	4.5	**5.1**	3.2	3.7	3.3	1.02
RP.CRK-16 (8)	80	0.1	21.1	**23.2**	20.7	20.8	20.3	20.4	22.1	22.0	**19.3**	19.4	20.8	1.22
RP.CRK-17	77	0.11	21.5	20.9	20.2	20.3	20.7	20.8	21.4	21.9	20.0	20.0	20.7	0.66
RP.CRK-18 (10)	75	0.13	24.0	24.7	23.4	23.4	23.6	23.3	23.7	23.6	**21.7**	**21.7**	23.5	0.94

Note: Bold values indicate biased annual FCH_4 values (more than ±1.5 g CH_4-C m^{-2} $year^{-1}$ from median values).
Underlined values indicate highly biased FCH_4 values (more than ±2.0 g CH_4-C m^{-2} $year^{-1}$ from median values).
Red color indicates positive bias and the blue color indicates negative bias from the ensemble median.
Abbreviations: ANN, artificial neural networks; MDS, marginal distribution sampling; RF, random forest; SVM, support vector machine.

4 | DISCUSSION

4.1 | Multidriver dependency of FCH_4

In order to investigate the degree to which the various gap-filling approaches incorporate multidriver dependency of FCH_4, the correlations between model residuals and biophysical drivers were explored. That is, if an approach properly explains multidriver dependency, there should be no significant correlation between the residual and the biophysical drivers in Figure 7.

The residuals of ANN and RF only had one to two significant relationships with the biophysical drivers, regardless of the usage of PCA (Figure 7). As for MDS, PCA reduced the number of significant relationships, which indicates an improvement over using original ancillary parameters as inputs to MDS. The residual of $MDS_{\_RAV}$ had seven significant relationships with the drivers, whereas only four to five relationships were significant for $MDS_{\_PC3}$ and $MDS_{\_PC5}$. Because PCA can compress most of the variance in the drivers, it could help resolve the multidriver dependency of FCH_4 compared to $MDS_{\_RAV}$ and $MDS_{\_SNW}$. This is expressed in the improved performance of MDS when using PCA. The residuals of SVM and $MDS_{\_SNW}$ have eight significant correlations and they showed high correlation coefficients with biophysical drivers (up to −0.1; Figure 7). These high correlations can be a reason for the bias in Table 4.

4.2 | Annual methane budget and timescales of FCH_4 variability

The different annual methane budgets generated by gap-filling approaches (Table 4) can be interpreted using the power spectral analysis. Gap-filled FCH_4 using different approaches showed different timescales of variability which influenced the annual methane budgets. We found that there were significant correlations between the annual methane budgets and the percentage of total FCH_4 variance at diel and seasonal scales across all gap-filling methods (Figure 8). The normalized annual methane budgets and the normalized FCH_4 variability at the diel scale had a negative correlation (R = −.83), whereas the annual budgets and the normalized FCH_4 variability at the seasonal scale had a positive correlation (R = .61). If WET.TW was disregarded, correlation coefficients of up to −.94 and .94 were observed for the diel and seasonal scales, respectively. This observation implies that the amplitude of diurnal and seasonal FCH_4 cycles generated by models is a key determinant of the annual methane budgets.

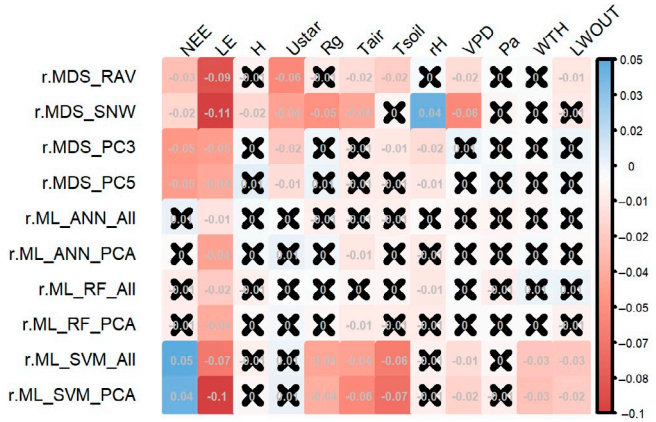

FIGURE 7 Correlation matrix showing the relationship between methane flux (FCH_4) residual (i.e., modeled FCH_4 − observed FCH_4) and biophysical drivers. Each row represents an approach and each column a driver. The color bar shows a correlation coefficient and the × sign indicates there is no significant (p < .05) correlation. The short-gap scenario was used [Colour figure can be viewed at wileyonlinelibrary.com]

The negative correlation between the FCH_4 variability at the diel scale and the annual methane budget may relate to the rate of occurrence of gaps in daytime versus nighttime periods. Because nighttime FCH_4 usually has more gaps to be filled, it is more affected by gap filling than daytime FCH_4. Thus, if a gap-filling model overestimates FCH_4, the difference between daytime and the nighttime FCH_4 can be reduced due to the high influence on nighttime values. This mechanism resulted in the negative correlations between FCH_4 variability at the diel scale and annual methane budgets (Figure 8). For instance, $MDS_{\_SNW}$ overestimated nighttime FCH_4 causing the lowest diel scale variability in general, and thus the positive bias of the annual methane budget (Table 4; Figure 8). For the WET.TW site, the data coverage for daytime and nighttime FCH_4 was similar (Table 1), and this may have resulted in a weak relationship between the diel power and the annual methane budget in the site (R = −.4).

The positive relationship between the annual methane budget and the percentage of total variance at the seasonal scale (Figure 8) is likely related to seasonal patterns in methane emissions. In periods of low FCH_4 due to cold temperatures (for WET.BB, WET.TW) or no inundation (for WET.PT, RP.HPK, and RP.CRK), there was no apparent difference between gap-filling approaches. Differences between approaches emerged for high FCH_4 periods. Therefore, the amplitude of seasonal scale cycles increased if a model overestimated FCH_4, and vice versa, resulting in a positive relationship between the FCH_4 variability at the seasonal scale and the annual methane budget (Figure 8). Unlike other sites, the WET.TW site showed the opposite trend. One possible explanation is that CH_4 emissions in the dormant period at WET.TW were higher than those at other sites (Figure S2), so gap-filling methods had an important role in the dormant season as well. Also, gap-filled FCH_4 for the WET.TW site showed high nonstationarity, hence the Fourier transform may not be suitable for this site (see Figure S3 right panel).

4.3 | u_* Threshold and CH_4 budget

We further tested gap-filling approaches with a higher u_* threshold value than the u_* threshold applied for the low nighttime u_* site WET.BB. The higher value of u_* threshold was generated by the bootstrapping method following Papale et al. (2006). We expected that the gap-filled annual methane budget using the

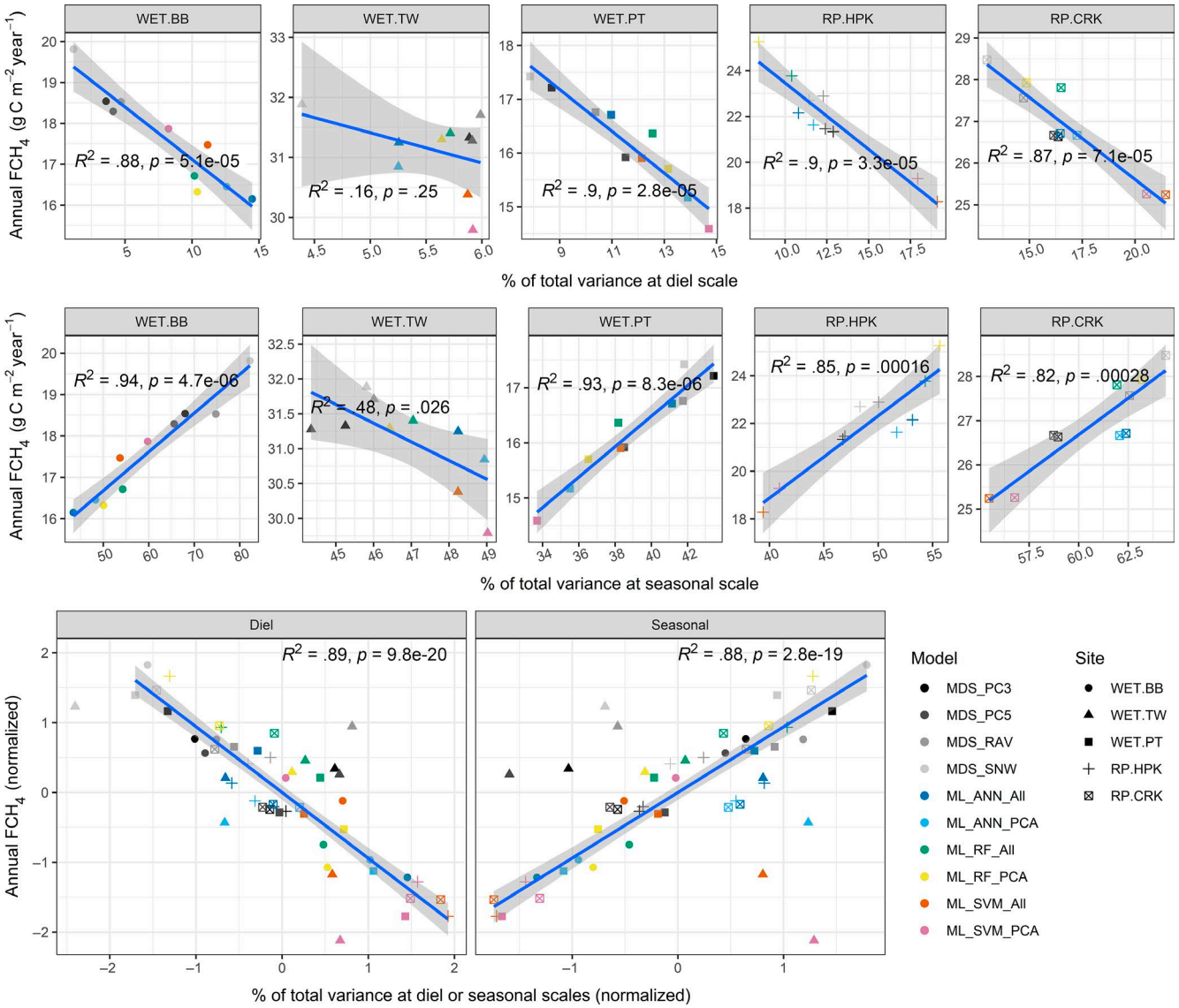

FIGURE 8 Relationships between the annual methane budget and percentage of total methane flux (FCH_4) variance at the diel and seasonal scales. The first and the second rows show the relationships for each site and each timescale. The last row shows the relationships for all sites with normalized data. Linear regression lines with confidence intervals are also shown. The WET.TW site was disregarded when generating the linear regression lines in the panels of the last row [Colour figure can be viewed at wileyonlinelibrary.com]

higher u_* threshold may be higher than that using the lower u_* threshold. This is because the higher u_* threshold can further filter low FCH_4 data related to not only measurement bias but also to actually repressed CH_4 at low turbulence condition (Nemitz et al., 2018). Overall, the annual methane budget gap filled by the MDS algorithm using the higher u_* threshold was about 1 g CH_4-C m^{-2} $year^{-1}$ higher than that using the lower u_* threshold in the WET.BB site (cf. D'Acunha et al., 2019). Surprisingly, however, annual budgets generated by all the ML algorithms were rarely affected by the u_* thresholds. This may due to the extrapolation capacity of the ML algorithms. To test the extrapolation capacity further, we trained a separate RF model removing high u_* periods (u_* values located in the fourth quartile (>75%) of the distribution) and tested its performance for gap filling the high u_* periods. RF well predicted FCH_4 for high u_* periods without any bias (data not shown). While further systematic examination will be necessary to establish the full role of u_* thresholds on FCH_4 data processing, our results suggest that the u_* threshold rarely affects the gap-filled annual methane budget using ML algorithms by virtue of their extrapolation capacity.

4.4 | Nonstationarity

Another feature of FCH_4 is nonstationarity, which may not be captured by ancillary biophysical observations. WET.TW is a representative nonstationary site with increases in annual methane emissions due to changes in soil properties (Chamberlain et al., 2018). Figure 9 shows the half-hourly bias error as a function of year for the WET.

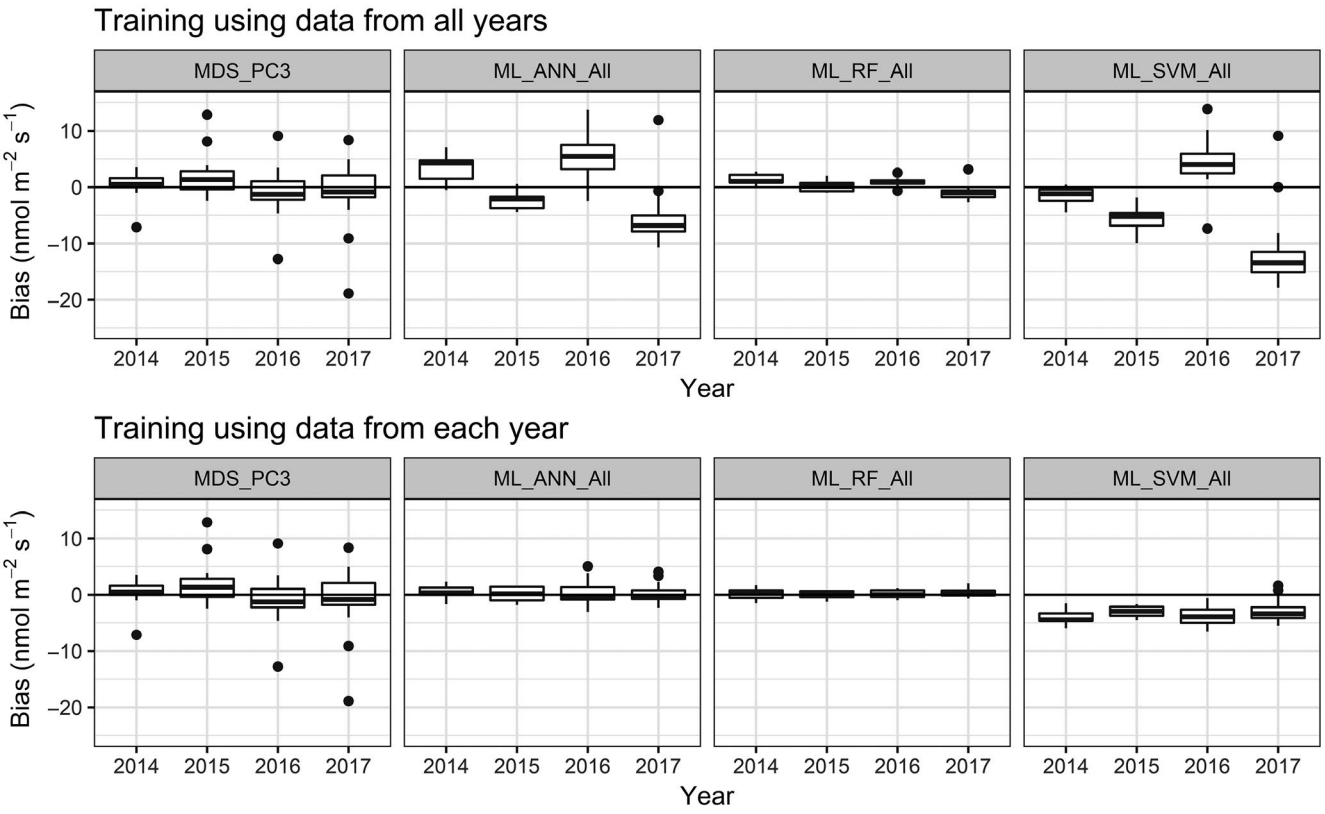

FIGURE 9 Half-hourly bias error for the WET.TW site as a function of year. Each panel represents each algorithm. The first row shows the results when machine learning (ML) algorithms were trained using data from all years. The second row is the results when they were trained separately for each year [Colour figure can be viewed at wileyonlinelibrary.com]

TW. If an algorithm cannot capture the increasing trend, there would be a negative bias in later years and/or a positive bias in earlier years. The bias of MDS did not change in each year by virtue of its moving window sampling strategy. On the other hand, ML algorithms showed a negative bias in the last year for models trained using data from the whole period (the first row in Figure 9). However, this bias problem was resolved when we trained the models separately for each year (the second row in Figure 9). It is noteworthy that RF was rarely affected by such nonstationarity even when the model was trained using the whole dataset.

4.5 | Delayed impacts of drainage

To test applicability of lagged variables, we applied the lagged WTH (1–4 weeks lag) to the RF model for the two rice paddy sites that experienced clearly lagged impacts of drainage. We only tested RF since its performance was the highest, and the algorithm can reveal the importance of specific variables.

In the RP.CRK site where the lagged impact of drainage on FCH_4 is thought to be less important (see the site description section), there was no performance improvement with the lagged variables (data not shown). In this site, the RF model was already close to the noise limit without the lagged variables, and thus there was little room for improvement (Figure 4 left). On the other hand, the performance of RF was improved with the lagged variables in RP.HPK site where the impact of drainage was persistent. The MAE of the RF model was 40% greater than the random error, but it was reduced to 28% of the random error after including the lagged variables. The absolute error of RF with the lagged variables approached the random error (Figure 10 right). The estimated variable importance also indicates the critical role of the lagged variables (Figure 10 left).

4.6 | Why RF?

In terms of both model accuracy and reliability from bias, the RF model showed the best performance across all sites in this analysis. The length of gaps rarely deteriorated the performance of the RF model. Also, RF was less affected by high FCH_4 variability at the hourly scale compared to the other approaches. In previous studies, RF was rarely applied for flux gap-filling purposes, but the results in this analysis suggest RF can be used as a standard FCH_4 gap-filling method. Then, why does the RF model show the best performance?

The best performance of RF for FCH_4 gap filling may relate to its training strategy which can handle highly nonlinear problems. RF can simultaneously incorporate continuous, discontinuous, and categorical variables, which include noise and gaps. Also, the training strategy of RF can avoid overfitting and the prediction result is stable. Furthermore, RF is straightforward to use compared to ANN (e.g.,

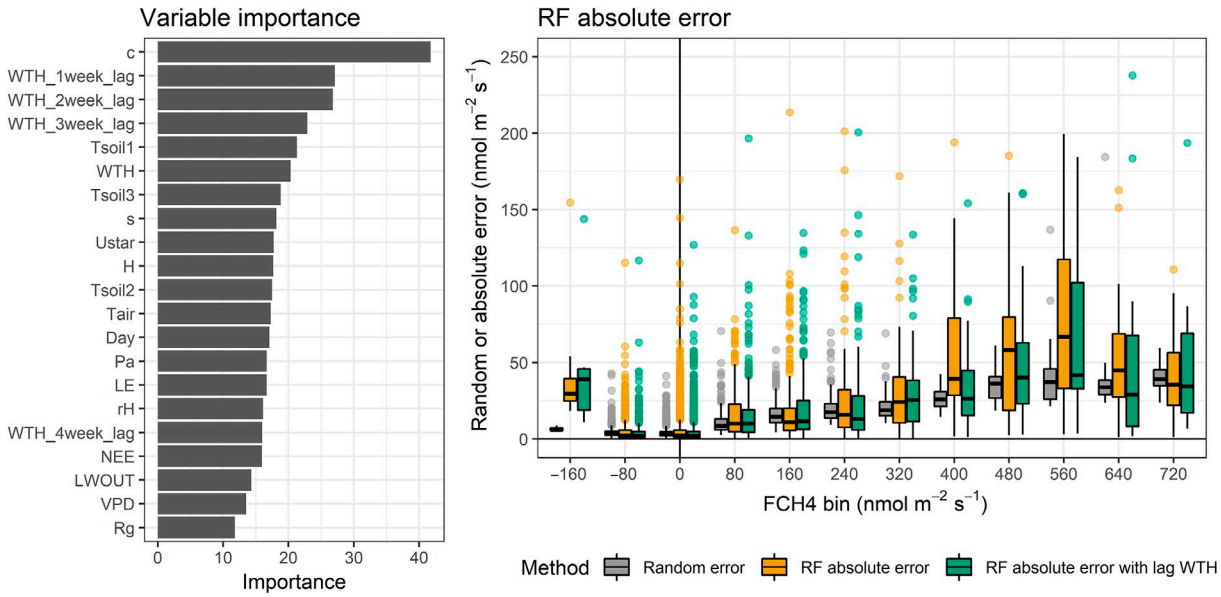

FIGURE 10 Variable importance for the random forest (RF) model with the water table height (WTH) lagged variables in the RP.HPK site (left) and methane flux (FCH_4) measurement and model uncertainties as a function of flux magnitude in the RP.HPK site (right). The variables c and s indicate cosine and sine functions, respectively, and other variables' abbreviations can be found in Table 2 [Colour figure can be viewed at wileyonlinelibrary.com]

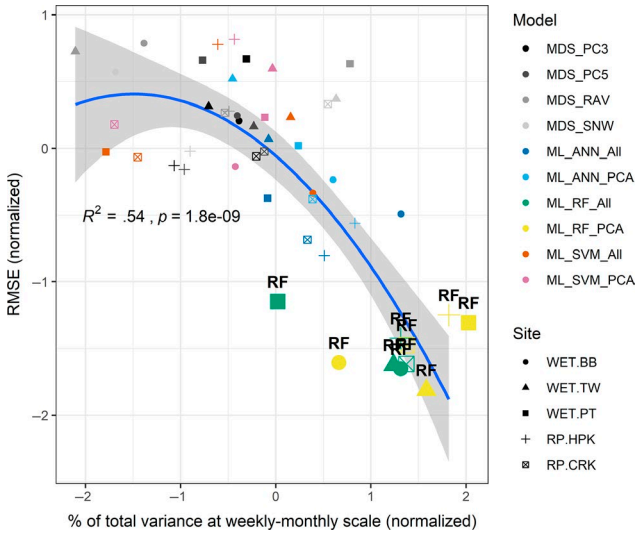

FIGURE 11 Relationships between root mean square error (RMSE) and the percentage of total methane flux (FCH_4) variance at weekly–monthly time scale. RMSE and the variance at the weekly–monthly scale were normalized. Random forest (RF) model is highlighted. The short-gap scenario was used for RMSE estimation. In the case of $MDS_{\_SNW}$ for the WET.PT site, a distinct outlier was disregarded in this figure (see Figure S9). The linear regression function is a second-order polynomial. Relationships for each site can be found in Figure S9 [Colour figure can be viewed at wileyonlinelibrary.com]

no need for stopping rules, no need for scaling, and less parameter requirements; cf. Tyralis, Papacharalampous, & Langousis, 2019).

We further explored data to find any unique feature of FCH_4 data gap filled by RF and found a consistent different characteristic compared with gap-filled FCH_4 by other approaches. Relative to other gap-filling approaches, FCH_4 gap filled by RF showed higher contributions of the weekly–monthly scale to the total variance (Figure 11). The flux variability at the weekly–monthly time scale is often nonlinearly and discontinuously related to environmental variables, particularly water table fluctuation (Johnson et al., 2013; Stoy et al., 2005; Sturtevant et al., 2016). Unlike other ML algorithms, RF may capture such nonlinear and discontinuous signals, which cause the higher variability of FCH_4 at this time scale.

4.7 | Summary and recommendations

In this study, we evaluated FCH_4 gap-filling performances of MDS and three ML algorithms with PCA. Based on this analysis, we highlight and recommend the following points to be considered for FCH_4 gap filling:

- RF showed the best overall performance for all sites, which was followed by ANN in terms of model accuracy, reliability, and bias. We do not recommend using PCA as an input for the ML algorithms tested (RF, ANN, and SVM).
- Absolute error of RF was very close to the random measurement uncertainty.
- MDS is a useful tool for FCH_4 gap filling if the lengths of gaps in the datasets are short enough and hourly variability of FCH_4 is not dominant in the total variance. When applying MDS, we recommend using PCA as input variables.
- When there are monthly scale gaps which should be filled, training the ML algorithms with multiyear data can provide reasonable results if FCH_4 remains stationary in interannual scales. RF showed the best performance for the monthly gap filling.

- The variability in gap-filled FCH_4 at the diel and seasonal timescales strongly influenced the resulting annual methane budgets, while the variability in weekly–monthly and hourly timescales was related to model accuracy.
- Gap-filling approaches tested in this study were generally able to address the multidriver dependency of FCH_4.
- The gap-filled annual methane budget using ML algorithms was less affected by the choice of u_* threshold compared to MDS.
- When using ML models, it is preferable to train the model using a shorter dataset such as a year rather than to train it with multiyear data in order to reduce the influence of nonstationarity of FCH_4.
- Including a lagged variable (e.g., WTH) as a predictor can improve the gap-filling performance at sites where an asynchronous impact is critical.

ACKNOWLEDGEMENTS

Research at Burns Bog (WET.BB) was funded through a NSERC Discovery Grant to Johnson with equipment obtained with funding from CFI (Canada Foundation for Innovation) and British Columbia Knowledge Development Fund (BCKDF) grants. The research tower was made possible by Metro Vancouver through a research agreement with the University of British Columbia (PI: A. Christen/A. Black). WET.BB researchers thank Zoran Nesic, Rick Ketler, Nick Lee, and Johannes Exler for their contributions to laboratory and fieldwork. Pantanal research efforts (WET.PT) were supported by the Brazilian National Institute for Science and Technology in Wetlands (INCT-INAU) and the Brazilian Ministry of Science and Technology. WET.PT researchers thank the Federal University of Mato Grosso for access to the field site and logistical support. Special thanks to the Xomanoflux group for field support. DDB was supported by the California Department of Water Resources, the US Department of Energy, Office of Science and Biological and Environmental Research. We thank our team of scientists in the Berkeley Biometeorology Lab for maintaining this site (Joe Verfaillie, Daphne Szutu, Sam Chamberlain, Elke Eichelmann, and Kyle Hemes). Research at Cheorwon and Haenam (RP.CRK and PR.HPK) were supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT; No. 2018R1C1B6002917). RP.CRK and PR.HPK researchers are grateful to the members of the Environmental Ecology Laboratory and the Complex Systems Science Laboratory at Seoul National University and the staff of the Measurement and Management department at National Center for AgroMeteorology, who installed and maintained the field sites. The eddy covariance data are available via Ameriflux (https://ameriflux.lbl.gov/), Asiaflux (http://www.asiaflux.net/), or on author request depending on sites. Example R scripts for the RF model and artificial gap generation are freely available via the online repository (https://github.com/yeonukkim/EC_FCH4_gapfilling). We thank Profs Dario Papale, Ankur Desai, and an anonymous reviewer for constructive comments on the earlier drafts.

ORCID

Yeonuk Kim https://orcid.org/0000-0003-2993-8687
Mark S. Johnson https://orcid.org/0000-0001-5070-7539
Sara H. Knox https://orcid.org/0000-0003-2255-5835
T. Andrew Black https://orcid.org/0000-0002-7494-9767
Higo J. Dalmagro https://orcid.org/0000-0002-2953-2575
Minseok Kang https://orcid.org/0000-0003-4901-4465
Joon Kim https://orcid.org/0000-0002-6381-8585
Dennis Baldocchi https://orcid.org/0000-0003-3496-4919

REFERENCES

Alberto, M. C. R., Wassmann, R., Buresh, R. J., Quilty, J. R., Correa, T. Q., Sandro, J. M., & Centeno, C. A. R. (2014). Measuring methane flux from irrigated rice fields by eddy covariance method using open-path gas analyzer. *Field Crops Research*, *160*, 12–21. https://doi.org/10.1016/j.fcr.2014.02.008

Baldocchi, D. (2014). Measuring fluxes of trace gases and energy between ecosystems and the atmosphere – The state and future of the eddy covariance method. *Global Change Biology*, *20*(12), 3600–3609. https://doi.org/10.1111/gcb.12649

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., … Evans, R. (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, *82*(11), 2415–2434. https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2

Baldocchi, D., Falge, E., & Wilson, K. (2001). A spectral analysis of biosphere–atmosphere trace gas flux densities and meteorological variables across hour to multi-year time scales. *Agricultural and Forest Meteorology*, *107*(1), 1–27. https://doi.org/10.1016/S0168-1923(00)00228-8

Barbour, A. J., & Parker, R. L. (2014). psd: Adaptive, sine multitaper power spectral density estimation for R. *Computers & Geosciences*, *63*, 1–8. https://doi.org/10.1016/j.cageo.2013.09.015

Barr, A. G., Richardson, A. D., Hollinger, D. Y., Papale, D., Arain, M. A., Black, T. A., … Schaeffer, K. (2013). Use of change-point detection for friction–velocity threshold evaluation in eddy-covariance studies. *Agricultural and Forest Meteorology*, *171–172*, 31–45. https://doi.org/10.1016/j.agrformet.2012.11.023

Bodesheim, P., Jung, M., Gans, F., Mahecha, M. D., & Reichstein, M. (2018). Upscaled diurnal cycles of land–atmosphere fluxes: A new global half-hourly data product. *Earth System Science Data*, *10*(3), 1327–1365. https://doi.org/10.5194/essd-10-1327-2018

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Brown, M. G., Humphreys, E. R., Moore, T. R., Roulet, N. T., & Lafleur, P. M. (2014). Evidence for a nonmonotonic relationship between ecosystem-scale peatland methane emissions and water table depth. *Journal of Geophysical Research: Biogeosciences*, *119*(5), 826–835. https://doi.org/10.1002/2013JG002576

Carslaw, D. C., & Ropkins, K. (2012). openair – An R package for air quality data analysis. *Environmental Modelling & Software*, *27–28*, 52–61. https://doi.org/10.1016/j.envsoft.2011.09.008

Chamberlain, S. D., Anthony, T. L., Silver, W. L., Eichelmann, E., Hemes, K. S., Oikawa, P. Y., … Baldocchi, D. D. (2018). Soil properties and sediment accretion modulate methane fluxes from restored wetlands. *Global Change Biology*, *24*(9), 4107–4121. https://doi.org/10.1111/gcb.14124

Chen, Y.-Y., Chu, C.-R., & Li, M.-H. (2012). A gap-filling model for eddy covariance latent heat flux: Estimating evapotranspiration of a subtropical seasonal evergreen broad-leaved forest as an example. *Journal of Hydrology*, *468–469*, 101–110. https://doi.org/10.1016/j.jhydrol.2012.08.026

Chu, H., Gottgens, J. F., Chen, J., Sun, G. E., Desai, A. R., Ouyang, Z., … Czajkowski, K. (2015). Climatic variability, hydrologic anomaly, and methane emission can turn productive freshwater marshes into net carbon sources. *Global Change Biology*, *21*(3), 1165–1181. https://doi.org/10.1111/gcb.12760

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/bf00994018

D'Acunha, B., Morillas, L., Black, T. A., Christen, A., & Johnson, M. S. (2019). Net ecosystem carbon balance of a peat bog undergoing restoration: Integrating CO_2 and CH_4 fluxes from eddy covariance and aquatic evasion with DOC drainage fluxes. *Journal of Geophysical Research: Biogeosciences*, *124*, 884–901. https://doi.org/10.1029/2019jg005123

Dalmagro, H. J., de Arruda, P. H., Vourlitis, G. L., Lathuillière, M. J., de S. Nogueira, J., Couto, E. G., & Johnson, M. S. (2019). Radiative forcing of methane fluxes offsets net carbon dioxide uptake for a tropical flooded forest. *Global Change Biology*, *25*(6), 1967–1981. https://doi.org/10.1111/gcb.14615

Dalmagro, H. J., Lathuillière, M. J., Hawthorne, I., Morais, D. D., Pinto, O. B. Jr, Couto, E. G., & Johnson, M. S. (2018). Carbon biogeochemistry of a flooded Pantanal forest over three annual flood cycles. *Biogeochemistry*, *139*(1), 1–18. https://doi.org/10.1007/s10533-018-0450-1

Dengel, S., Zona, D., Sachs, T., Aurela, M., Jammet, M., Parmentier, F., … Vesala, T. (2013). Testing the applicability of neural networks as a gap-filling method using CH_4 flux data from high latitude wetlands. *Biogeosciences*, *10*(12), 8185–8200. https://doi.org/10.5194/bg-10-8185-2013

Deventer, M. J., Griffis, T. J., Roman, D. T., Kolka, R. K., Wood, J. D., Erickson, M., … Millet, D. B. (2019). Error characterization of methane fluxes and budgets derived from a long-term comparison of open- and closed-path eddy covariance systems. *Agricultural and Forest Meteorology*, *278*, 107638. https://doi.org/10.1016/j.agrformet.2019.107638

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., … Wofsy, S. (2001). Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agricultural and Forest Meteorology*, *107*(1), 43–69. https://doi.org/10.1016/S0168-1923(00)00225-2

Finkelstein, P. L., & Sims, P. F. (2001). Sampling error in eddy correlation flux measurements. *Journal of Geophysical Research: Atmospheres*, *106*(D4), 3503–3509. https://doi.org/10.1029/2000jd900731

Fumoto, T., Yanagihara, T., Saito, T., & Yagi, K. (2010). Assessment of the methane mitigation potentials of alternative water regimes in rice fields using a process-based biogeochemistry model. *Global Change Biology*, *16*(6), 1847–1859. https://doi.org/10.1111/j.1365-2486.2009.02050.x

GaŽovic, M., Kutzbach, L., Schreiber, P., Wille, C., & Wilmking, M. (2010). Diurnal dynamics of CH_4 from a boreal peatland during snowmelt. *Tellus B: Chemical and Physical Meteorology*, *62*(3), 133–139. https://doi.org/10.1111/j.1600-0889.2010.00455.x

Gu, L., Falge, E. M., Boden, T., Baldocchi, D. D., Black, T. A., Saleska, S. R., … Xu, L. (2005). Objective threshold determination for nighttime eddy flux filtering. *Agricultural and Forest Meteorology*, *128*(3), 179–197. https://doi.org/10.1016/j.agrformet.2004.11.006

Günther, F., & Fritsch, S. (2010). neuralnet: Training of neural networks. *The R Journal*, *2*(1), 30–38. https://doi.org/10.32614/RJ-2010-006

Hatala, J. A., Detto, M., & Baldocchi, D. D. (2012). Gross ecosystem photosynthesis causes a diurnal pattern in methane emission from rice. *Geophysical Research Letters*, *39*(6), 1–5. https://doi.org/10.1029/2012GL051303

Herbst, M., Friborg, T., Ringgaard, R., & Soegaard, H. (2011). Interpreting the variations in atmospheric methane fluxes observed above a restored wetland. *Agricultural and Forest Meteorology*, *151*(7), 841–853. https://doi.org/10.1016/j.agrformet.2011.02.002

Howley, T., Madden, M. G., O'Connell, M.-L., & Ryder, A. G. (2006). The effect of principal component analysis on machine learning accuracy with high-dimensional spectral data. *Knowledge-Based Systems*, *19*(5), 363–370. https://doi.org/10.1016/j.knosys.2005.11.014

Huang, Y., Ryu, Y., Jiang, C., Kimm, H., Kim, S., Kang, M., & Shim, K. (2018). BESS-Rice: A remote sensing derived and biophysical process-based rice productivity simulation model. *Agricultural and Forest Meteorology*, *256–257*, 253–269. https://doi.org/10.1016/j.agrformet.2018.03.014

Ichii, K., Ueyama, M., Kondo, M., Saigusa, N., Kim, J., Alberto, M. C., … Zhao, F. (2017). New data-driven estimation of terrestrial CO_2 fluxes in Asia using a standardized database of eddy covariance measurements, remote sensing data, and support vector regression. *Journal of Geophysical Research: Biogeosciences*, *122*(4), 767–795. https://doi.org/10.1002/2016JG003640

Jammet, M., Crill, P., Dengel, S., & Friborg, T. (2015). Large methane emissions from a subarctic lake during spring thaw: Mechanisms and landscape significance. *Journal of Geophysical Research: Biogeosciences*, *120*(11), 2289–2305. https://doi.org/10.1002/2015JG003137

Johnson, M. S., Couto, E. G., Pinto, O. B. Jr, Milesi, J., Santos Amorim, R. S., Messias, I. A. M., & Biudes, M. S. (2013). Soil CO_2 dynamics in a tree island soil of the Pantanal: The role of soil water potential. *PLoS ONE*, *8*(6), e64874. https://doi.org/10.1371/journal.pone.0064874

Kang, M., Kim, J., Lee, S.-H., Kim, J., Chun, J. H., & Cho, S. (2018). Changes and improvements of the standardized eddy covariance data processing in KoFlux. *Korean Journal of Agricultural and Forest Meteorology*, *20*(1), 5–17. https://doi.org/10.5532/KJAFM.2018.20.1.5

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab-an S4 package for kernel methods in R. *Journal of Statistical Software*, *11*(9), 1–20. https://doi.org/10.18637/jss.v011.i09

Keenan, T. F., Migliavacca, M., Papale, D., Baldocchi, D., Reichstein, M., Torn, M., & Wutzler, T. (2019). Widespread inhibition of daytime ecosystem respiration. *Nature Ecology & Evolution*, *3*, 407–415. https://doi.org/10.1038/s41559-019-0809-2

Kim, J., Verma, S. B., & Billesbach, D. P. (1999). Seasonal variation in methane emission from a temperate Phragmites-dominated marsh: Effect of growth stage and plant-mediated transport. *Global Change Biology*, *5*(4), 433–440. https://doi.org/10.1046/j.1365-2486.1999.00237.x

Kim, Y., Talucder, M. S. A., Kang, M., Shim, K.-M., Kang, N., & Kim, J. (2016). Interannual variations in methane emission from an irrigated rice paddy caused by rainfalls during the aeration period. *Agriculture, Ecosystems & Environment*, *223*, 67–75. https://doi.org/10.1016/j.agee.2016.02.032

Knox, S. H., Jackson, R. B., Poulter, B., McNicol, G., Fluet-Chouinard, E., Zhang, Z., … Zona, D. (2019). FLUXNET-CH_4 synthesis activity: Objectives, observations, and future directions. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/bams-d-18-0268.1

Knox, S. H., Matthes, J. H., Sturtevant, C., Oikawa, P. Y., Verfaillie, J., & Baldocchi, D. (2016). Biophysical controls on interannual variability in ecosystem-scale CO_2 and CH_4 exchange in a California rice paddy. *Journal of Geophysical Research: Biogeosciences*, *121*(3), 978–1001. https://doi.org/10.1002/2015JG003247

Knox, S. H., Sturtevant, C., Matthes, J. H., Koteen, L., Verfaillie, J., & Baldocchi, D. (2015). Agricultural peatland restoration: Effects of land-use change on greenhouse gas (CO_2 and CH_4) fluxes in the

Sacramento-San Joaquin Delta. *Global Change Biology*, *21*(2), 750–765. https://doi.org/10.1111/gcb.12745

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., … Hunt, T. (2019). Caret: Classification and regression training. R package version 6.0-84. Retrieved from https://CRAN.Rproject.org/package=caret

Lee, S.-C., Christen, A., Black, A. T., Johnson, M. S., Jassal, R. S., Ketler, R., … Merkens, M. (2017). Annual greenhouse gas budget for a bog ecosystem undergoing restoration by rewetting. *Biogeosciences*, *14*(11), 2799–2814. https://doi.org/10.5194/bg-14-2799-2017

Lee, S.-H., Kang, M., Kang, N., & Kim, J. (2018). Haenam Paddy-field KoFlux (HPK) site with dry direct-seeding: Introduction. *Korean Journal of Agricultural and Forest Meteorology*, *20*(1), 18–33. https://doi.org/10.5532/KJAFM.2018.20.1.18

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

McDermitt, D., Burba, G., Xu, L., Anderson, T., Komissarov, A., Riensche, B., … Hastings, S. (2011). A new low-power, open-path instrument for measuring methane flux by eddy covariance. *Applied Physics B*, *102*(2), 391–405. https://doi.org/10.1007/s00340-010-4307-0

Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., … Stauch, V. J. (2007). Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agricultural and Forest Meteorology*, *147*(3), 209–232. https://doi.org/10.1016/j.agrformet.2007.08.011

Morin, T. H. (2019). Advances in the eddy covariance approach to CH_4 monitoring over two and a half decades. *Journal of Geophysical Research: Biogeosciences*, *124*(3), 453–460. https://doi.org/10.1029/2018jg004796

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., … Mendoza, B. (2013). Anthropogenic and natural radiative forcing. *Climate Change*, *423*, 658–740.

Nemitz, E., Mammarella, I., Ibrom, A., Aurela, M., Burba, G. G., Dengel, S., … Zahniser, M. (2018). Standardisation of eddy-covariance flux measurements of methane and nitrous oxide. *International Agrophysics*, *32*(4), 517–549. https://doi.org/10.1515/intag-2017-0042

Neubauer, S. C., & Megonigal, J. P. (2015). Moving beyond global warming potentials to quantify the climatic role of ecosystems. *Ecosystems*, *18*(6), 1000–1013. https://doi.org/10.1007/s10021-015-9879-4

Nishimura, S., Sawamoto, T., Akiyama, H., Sudo, S., & Yagi, K. (2004). Methane and nitrous oxide emissions from a paddy field with Japanese conventional water management and fertilizer application. *Global Biogeochemical Cycles*, *18*(2). https://doi.org/10.1029/2003GB002207

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., … Yakir, D. (2006). Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: Algorithms and uncertainty estimation. *Biogeosciences*, *3*(4), 571–583. https://doi.org/10.5194/bg-3-571-2006

Papale, D., & Valentini, R. (2003). A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Global Change Biology*, *9*(4), 525–535. https://doi.org/10.1046/j.1365-2486.2003.00609.x

Peltola, O., Hensen, A., Helfter, C., Belelli Marchesini, L., Bosveld, F. C., van den Bulk, W. C. M., … Mammarella, I. (2014). Evaluating the performance of commonly used gas analysers for methane eddy covariance flux measurements: The InGOS inter-comparison field experiment. *Biogeosciences*, *11*(12), 3163–3186. https://doi.org/10.5194/bg-11-3163-2014

R Core Team. (2018). *R: A language and environment for statistical computing*. Viena, Austria: R Foundation for Statistical Computing.

Read, J. S., Hamilton, D. P., Desai, A. R., Rose, K. C., MacIntyre, S., Lenters, J. D., … Wu, C. H. (2012). Lake-size dependency of wind shear and convection as controls on gas exchange. *Geophysical Research Letters*, *39*(9). https://doi.org/10.1029/2012gl051886

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., … Valentini, R. (2005). On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm. *Global Change Biology*, *11*(9), 1424–1439. https://doi.org/10.1111/j.1365-2486.2005.001002.x

Richardson, A. D., & Hollinger, D. Y. (2005). Statistical modeling of ecosystem respiration using eddy covariance data: Maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, applied to three simple models. *Agricultural and Forest Meteorology*, *131*(3), 191–208. https://doi.org/10.1016/j.agrformet.2005.05.008

Richardson, A. D., Mahecha, M. D., Falge, E., Kattge, J., Moffat, A. M., Papale, D., … Hollinger, D. Y. (2008). Statistical properties of random CO_2 flux measurement uncertainty inferred from model residuals. *Agricultural and Forest Meteorology*, *148*(1), 38–50. https://doi.org/10.1016/j.agrformet.2007.09.001

Sachs, T., Giebels, M., Boike, J., & Kutzbach, L. (2010). Environmental controls on CH_4 emission from polygonal tundra on the microsite scale in the Lena river delta, Siberia. *Global Change Biology*, *16*(11), 3096–3110. https://doi.org/10.1111/j.1365-2486.2010.02232.x

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., … Zhu, Q. (2016). The global methane budget 2000–2012. *Earth System Science Data*, *8*(2), 697–751. https://doi.org/10.5194/essd-8-697-2016

Stoy, P. C., Katul, G. G., Siqueira, M. B. S., Juang, J.-Y., McCarthy, H. R., Kim, H.-S., … Oren, R. (2005). Variability in net ecosystem exchange from hourly to inter-annual time scales at adjacent pine and hardwood forests: A wavelet analysis. *Tree Physiology*, *25*(7), 887–902. https://doi.org/10.1093/treephys/25.7.887

Sturtevant, C., Ruddell, B. L., Knox, S. H., Verfaillie, J., Matthes, J. H., Oikawa, P. Y., & Baldocchi, D. (2016). Identifying scale-emergent, nonlinear, asynchronous processes of wetland methane exchange. *Journal of Geophysical Research: Biogeosciences*, *121*(1), 188–204. https://doi.org/10.1002/2015JG003054

Tang, A. C. I., Stoy, P. C., Hirata, R., Musin, K. K., Aeries, E. B., Wenceslaus, J., & Melling, L. (2018). Eddy covariance measurements of methane flux at a tropical peat forest in Sarawak, Malaysian Borneo. *Geophysical Research Letters*, *45*(9), 4390–4399. https://doi.org/10.1029/2017GL076457

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192. https://doi.org/10.1029/2000JD900719

Taylor, M. A., Celis, G., Ledman, J. D., Bracho, R., & Schuur, E. A. G. (2018). Methane efflux measured by eddy covariance in Alaskan upland tundra undergoing permafrost degradation. *Journal of Geophysical Research: Biogeosciences*, *123*(9), 2695–2710. https://doi.org/10.1029/2018jg004444

Tramontana, G., Ichii, K., Camps-Valls, G., Tomelleri, E., & Papale, D. (2015). Uncertainty analysis of gross primary production upscaling using Random Forests, remote sensing and eddy covariance data. *Remote Sensing of Environment*, *168*, 360–373. https://doi.org/10.1016/j.rse.2015.07.015

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., … Papale, D. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, *13*(14), 4291–4313. https://doi.org/10.5194/bg-13-4291-2016

Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*, *11*(5), 910. https://doi.org/10.3390/w11050910

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). Package 'corrplot'. *Statistician*, *56*, 316–324.

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., … Reichstein, M. (2018). Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences*, *15*(16), 5015–5030. https://doi.org/10.5194/bg-15-5015-2018

Xu, T., Guo, Z., Liu, S., He, X., Meng, Y., Xu, Z., … Song, L. (2018). Evaluating different machine learning methods for upscaling evapotranspiration from flux towers to the regional scale. *Journal of Geophysical Research: Atmospheres*, *123*(16), 8674–8690. https://doi.org/10.1029/2018JD028447

Yuan, J., Ding, W., Liu, D., Kang, H., Freeman, C., Xiang, J., & Lin, Y. (2015). Exotic *Spartina alterniflora* invasion alters ecosystem–atmosphere exchange of CH_4 and N_2O and carbon sequestration in a coastal salt marsh in China. *Global Change Biology*, *21*(4), 1567–1580. https://doi.org/10.1111/gcb.12797

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Kim Y, Johnson MS, Knox SH, et al. 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*. 2020;26:1499–1518. https://doi.org/10.1111/gcb.14845