Technical note: Uncertainties in eddy covariance CO_2 fluxes in a semi-arid sagebrush ecosystem caused by gap-filling approaches

Jingyu Yao[1], Zhongming Gao[2], Jianping Huang[1], Heping Liu[2], and Guoyin Wang[3]

[1]Key Laboratory for Semi-Arid Climate Change of the Ministry of Education, College of Atmospheric Sciences, Lanzhou
University, Lanzhou, China

[2]Laboratory for Atmospheric Research, Department of Civil and Environmental Engineering, Washington State University,
Pullman, Washington, USA

[3]Department of Atmospheric and Oceanic Sciences & Institute of Atmospheric Sciences, Fudan University, Shanghai, China

Correspondence to: Jianping Huang (hjp@lzu.edu.cn)

Abstract. Gap-filling eddy covariance CO_2 fluxes is challenging at dryland sites due to small CO_2 fluxes. Here, four machine learning (ML) algorithms including artificial neural network (ANN), k-nearest neighbours (KNN), random forest (RF), and support vector machine (SVM) are employed and evaluated for gap-filling CO_2 fluxes over a semi-arid sagebrush ecosystem with different lengths of artificial gaps. The ANN and RF algorithms outperform the KNN and SVM in filling gaps ranging from hours to days, with the RF being more time efficient than the ANN. Performances of the ANN and RF are largely degraded for extremely long gaps of two months. In addition, our results suggest that there is no need to fill the daytime and nighttime NEE gaps separately when using the ANN and RF. With the ANN and RF, the gap-filling induced uncertainties in the annual NEE at this site are estimated to be within 16 g C m^{-2}, whereas the uncertainties by the KNN and SVM can be as large as 27 g C m^{-2}. To better fill extremely long gaps of a few months, we test a two-layer gap-filling framework based on the RF. With this framework, the model performance is improved significantly, especially for the nighttime data. Therefore, this approach provides an alternative in filling extremely long gaps to characterize annual carbon budgets and interannual variability in dryland ecosystems.

1 Introduction

The eddy covariance (EC) technique has been widely applied for monitoring energy and water fluxes as well as net ecosystem exchanges of carbon dioxide (NEE) and other trace gases between lands and the atmosphere (Baldocchi, 2003; Oncley et al., 2007; Stoy et al., 2013). However, due to multiple factors including power outages, instrument malfunctions and maintenance, and data quality checks, there exist gaps with approximately 20-60% of half-hourly data points annually at many long-term EC sites (Dragoni et al., 2007; Falge et al., 2001; Ma et al., 2007; Missik et al., 2019; Moffat et al., 2007;

Pastorello et al., 2020; Soloway et al., 2017; Wutzler et al., 2018). An average gap fraction of 30% in an annual dataset leads
30 to an uncertainty of ±25 g C m^{-2} year^{-1} for the annual NEE at forest sites (Moffat et al., 2007), while some EC sites report much greater uncertainties (Soloway et al., 2017). Therefore, gap-filling usually accounts for one large source of uncertainties in the annual NEE (Soloway et al., 2017), together with other sources of uncertainties such as measurement errors and bias related to non-closure of the surface energy balance (Gao et al., 2019; Wilson et al., 2002).

Robust NEE gap-filling approaches are critical for quantifying the annual and interannual variability of carbon budgets
35 (Falge et al., 2001; Moffat et al., 2007; Pastorello et al., 2020; Richardson and Hollinger, 2007; Soloway et al., 2017; Wutzler et al., 2018). Previous studies have developed and evaluated a number of NEE gap-filling approaches including non-linear regressions (NLR), look-up tables (e.g., marginal distribution sampling (MDS)), machine learning (ML) algorithms (e.g., artificial neural networks), and process-based models (Falge et al., 2001; Huang and Hsieh, 2020; Moffat et al., 2007; Wutzler et al., 2018). NLR fills NEE gaps based on regression analyses between NEE and meteorological variables such as
40 temperature (e.g., air or soil temperature) and light (e.g., photosynthetically active radiation), whereas MDS is based on look-up tables for similar meteorological conditions (i.e., global radiation, air temperature, and vapor pressure deficit) (Falge et al., 2001; Moffat et al., 2007; Reichstein et al., 2005). MDS has been widely used in gap-filling by virtue of an easy-to-use R package (Wutzler et al., 2018), although it cannot effectively fill the gaps of longer than 12 days (Moffat et al., 2007). ML-based methods are trained by presenting them with numerous meteorological variables as inputs and NEE as output data,
45 which have the potential to fill long gaps (Kim et al., 2020; Moffat et al., 2007). Artificial neural network (ANN) was applied to fill daytime and nighttime gaps separately (Knox et al., 2016; Moffat et al., 2007; Papale and Valentini, 2003). Other ML algorithms such as random forest (RF), k-nearest neighbours (KNN), and support vector machine (SVM) have recently been assessed for flux gap-filling (Huang and Hsieh, 2020; Kim et al., 2020) and flux upscaling (Xu et al., 2018). However, the performance of these ML-based algorithms has not been evaluated in filling gaps in EC fluxes for dryland
50 ecosystems with low NEE. The motivation of this gap-filling practice was driven by the fact that dryland ecosystems are very sensitive to water availability, functioning as carbon sinks in wet years and carbon sources in dry years (Biederman et al., 2017; Scott et al., 2015), and bias in gap-filled NEE may alter conclusions in sources or sinks of dryland ecosystems in case of relatively long gaps for eddy covariance data. In addition, different ML algorithms have distinctive internal structures that account for the underlying dependencies of outputs (i.e., NEE) on the inputs (i.e., meteorological variables) in different
55 ways, and uncertainties associated with different ML-based methods can also be assessed.

In this study, we evaluate the performance of four commonly used ML algorithms (ANN, KNN, RF, and SVM) in filling the gaps in the NEE data collected at an EC site over a semi-arid sagebrush ecosystem in the central Washington, USA, from 2016 to 2019, and assess the uncertainties in the annual NEE introduced by gap-filling methods.

2 Materials and methods

60   **2.1 Site description**

The eddy covariance flux tower is located in the Hanford Area in the U.S. State of Washington (AmeriFlux site: US-Hn1; 46°24′32″ N, 119°16′30″ W), and it started to collect data in December 2015 (Gao et al., 2019, 2020a, 2020b; Missik et al., 2019). This semi-arid site is predominantly covered by scattered shrubs and short grasses. Shrub species include *Artemisia tridentata* and *Chrysothamnus viscidiflorus*, and grasses include invasive weedy species (i.e., *Bromus tectorum and Salsola*

65   *kali*) and native grasses (i.e., *Poa secunda, Pseudoroegneria spicata, and Stipa comate*) (Missik et al., 2019). The long-term (1986-2015) mean annual precipitation was 197 mm, most of which occurred late in the fall and early in winter (Missik et al., 2019). The soil texture in the top layer of 30 cm is loamy sand with small rocks and gravel interspersed (Gao et al., 2017; Missik et al., 2019). In this study, the 4-year data from 2016 to 2019 are analyzed.

2.2 Eddy covariance and meteorological measurements

70   The EC system included a three-dimensional sonic anemometer (CSAT3, Campbell Scientific, Inc.) and an open path gas analyzer (LI-7500A, LI-COR, Inc.). In addition, a variety of microclimate data were measured, including four-component radiation, air temperature and relative humidity, wind speed and direction, precipitation, soil heat flux, and soil temperature and volumetric water content (Gao et al., 2017, 2019, 2020a; Missik et al., 2019). These data were sampled at a rate of 1 Hz and stored as 30-min averages. Further, 15-min meteorological data from two weather stations close to the tower site were

75   obtained from the Washington State University AgWeatherNet (AWN; https://weather.wsu.edu/). The two AWN stations are located within 8 km from the tower. The 15-min data were averaged to half-hourly values to fill gaps in the tower meteorological data of solar radiation (R_g), air temperature (T_{air}) and relative humidity (RH), vapor pressure deficit (VPD), wind speed (WS), precipitation (P), and soil temperature (T_{soil}). Thus, these half-hourly meteorological data for the study period are gap-free.

80   **2.3 EC data processing, quality control, and gap identification**

Raw 10 Hz EC data were processed using the EddyPro® software (version 7.06, LI-COR Biosciences, USA) to calculate the 30-min average fluxes of CO_2 (NEE) and latent (LE) and sensible (H) heat. The data were despiked and filtered for physically impossible values and abnormal diagnostic values of the sonic anemometer and the gas analyzer. The double rotation method was applied to the sonic anemometer data. Block averaging was used to determine the turbulent fluctuations

85   for each 30-min intervals. The fluxes were corrected for the effects of high- and low-pass filtering (Massman, 2000, 2001; Moncrieff et al., 2004) and air density fluctuations (Webb et al., 1980), respectively. The corrected fluxes were quality checked according to Mauder and Foken (2004). After quality checking, the "REddyProc" R package (Wutzler et al., 2018)

was used to determine the friction velocity (u_*) threshold and NEE data with low turbulence conditions were removed from the dataset.

90   For simplicity, a score of 2 was assigned to gaps due to field operations (e.g., instrument maintenance), electrical and/or instrument failures, a score of 1 to gaps due to low data quality (i.e., quality control and u_* filtering), and a score of 0 to flux data with good quality. Only the data with the score of 0 were used to train/test the gap-filling algorithms.

2.4 Machine learning algorithms

Four ML algorithms including the ANN, KNN, RF, and SVM were employed and evaluated for filling NEE gaps. In the
95   following sections, we briefly describe the characteristics and implementation of each ML algorithm. The required parameters in each algorithm are optimized using the "caret" R package (Kuhn et al., 2020) with a 10-fold cross-validation repeated ten times.

2.4.1 Artificial neural network (ANN)

The ANN algorithm has been successfully applied for filling NEE gaps in various ecosystems (Baldocchi & Sturtevant,
100   2015; Knox et al., 2016; Moffat et al., 2007; Papale & Valentini, 2003; Tramontana et al., 2016). In this study, we employed the "neuralnet" R package (Günther and Fritsch, 2010), the resilient backpropagation algorithm that has proven to be capable of gap-filling flux data (Dengel et al., 2013; Jammet et al., 2015; Kim et al., 2020; Knox et al., 2016). The required parameters for the ANN algorithm include the number of hidden layers and the number of nodes in each layer. Here, based on the parameter tuning results, we use two hidden layers with 12 and 10 nodes in the first and second hidden layers. We
105   train the neural network 1000 times, and the mean prediction results of the top 20 runs based on their training and testing R^2 values are used to fill NEE gaps (Baldocchi & Sturtevant, 2015; Knox et al., 2016).

2.4.2 K-nearest neighbours (KNN)

The KNN algorithm (Fix and Hodges, 1951) is a non-parametric ML approach and has been used in many applications. For example, Chen et al. (2012) applied the KNN algorithm for filling latent heat flux gaps, which fills the data gaps based on a
110   certain attribution of k neighbours in the feature space. In this study, we use the "caret" R package (Kuhn et al., 2020) to build the KNN where a suitable k value needs to be determined. Here, the optimized k value is 9.

2.4.3 Random forest (RF)

The RF algorithm (Breiman, 2001) has been applied for upscaling flux data to regional (Xu et al., 2018) and global (Jung et al., 2017; Zeng et al., 2020) scales and recently for gap-filling flux data (Huang and Hsieh, 2020; Kim et al., 2020). The RF
115   algorithm uses bootstrap aggregation and feature randomness when generating each individual tree to try to create many independent decision trees that operate as an ensemble of the prediction results. In this study, we create 500 regression trees

for each case using the "randomForest" R package (Liaw and Wiener, 2002), in which the tuning parameter is the number of randomly selected predictors (i.e., mtry equals 7).

2.4.4 Support vector machine (SVM)

120 The SVM algorithm (Cortes and Vapnik, 1995) has also been applied for gap-filling (Huang and Hsieh, 2020; Kim et al., 2020) and upscaling (Xu et al., 2018) flux data. The SVM algorithm can convert nonlinear regressions into linear regressions by projecting the original finite-dimensional space into the much higher-dimensional space with a predefined kernel function. In this study, we use the radial basic kernel function and the "kernlab" R package (Karatzoglou et al., 2004) where the tuning parameters include the inverse kernel width (i.e., sigma = 0.13) and the cost regularization parameter (i.e., C = 125 27).

2.5 Input variables

Besides the above mentioned meteorological variables, the input variables for the ML algorithms also include the normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) from the Moderate Resolution Imaging Spectroradiometer (MODIS), and three fuzzy variables (i.e., decimal day of year, and sine and cosine functions to represent 130 seasonal changes) following Moffat et al. (2007). We obtained the NDVI and EVI data around the flux tower location from the MOD13Q1 version 6 data product (https://lpdaac.usgs.gov/products/mod13q1v006/) at 16-day temporal and 250-m spatial resolutions (Didan, 2015). The 16-day NDVI and EVI data were resampled to 30 minutes using cubic spline interpolation.

2.6 Artificial gap scenarios and performance evaluation

135 In order to evaluate the gap-filling algorithms, artificial gaps with different lengths were randomly generated in the original flux data, accounting for 10% of the total data length. Following Moffat et al. (2007) and Kim et al. (2020), we considered four artificial gap lengths: one hour (1-H), one day (1-D), one week (1-W), and two months (2-M). Each gap length scenario was permuted 10 times, resulting in 40 distinct artificial gap scenarios plus the real gap scenario to be filled.

For each artificial gap scenario, we trained the models separately with only daytime or nighttime data, and with both daytime 140 and nighttime data. The performance of gap-filling algorithms was evaluated by comparing the estimated values (e_i) with the measured values (m_i) for the artificial gaps. Five commonly used performance metrics are used including the coefficient of determination (R^2), the absolute root mean square error (ARMSE) and relative root mean square error (RRMSE), mean absolute error (MAE), and the bias error (BE):

$$R^2 = \frac{\left(\sum_{i=1}^{n}(e_i - \bar{e})(m_i - \bar{m})\right)^2}{\sum_{i=1}^{n}(e_i - \bar{e})^2 \sum_{i=1}^{n}(m_i - \bar{m})^2} \qquad (1)$$

145 $\quad$ $$\text{ARMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - m_i)^2} \qquad\qquad (2)$$

$$\text{RRMSE} = \sqrt{\frac{\sum_{i=1}^{n}(e_i - m_i)^2}{\sum_{i=1}^{n}(m_i)^2}} \qquad\qquad (3)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|e_i - m_i| \qquad\qquad (4)$$

$$\text{BE} = \frac{1}{n}\sum_{i=1}^{n}(e_i - m_i) \qquad\qquad (5)$$

where the overbar denotes the mean value.

150

3 Results

3.1 NEE data gap evaluation

At the US-Hn1 site, different lengths of gaps were found in NEE data during the four years from 2016 to 2019 (Fig. 1). Gaps with short to medium lengths were usually caused by low data quality (i.e., gap score of 1), whereas gaps with medium to

155 $\quad$ extremely long lengths were mostly due to electrical and/or instruments failures (i.e., gap score of 2). Data gaps with a score of 1 frequently occurred in nighttime because u_* filtering mainly removes nighttime data. On average, gaps with scores of 1 and 2 accounted for about 28.4% and 30.2% of half-hour NEE data, respectively (Table 1). There were more extremely long gaps in 2018 and 2019 than 2016 and 2017 due to power failures in winter and early spring. Therefore, it is worth to examining the performances of gap-filling algorithms in filling different lengths of gaps. Overall, about 43.7% of NEE data

160 $\quad$ were available to calibrate and validate the gap-filling methods. In addition, for the data gaps with a score of 1, meteorological data from the flux tower were still available for NEE gap-filling; whereas for the data gaps with a score of 2, meteorological data from the flux tower also had gaps, and the data obtained from the two nearby AWN stations are thus used for gap filling.

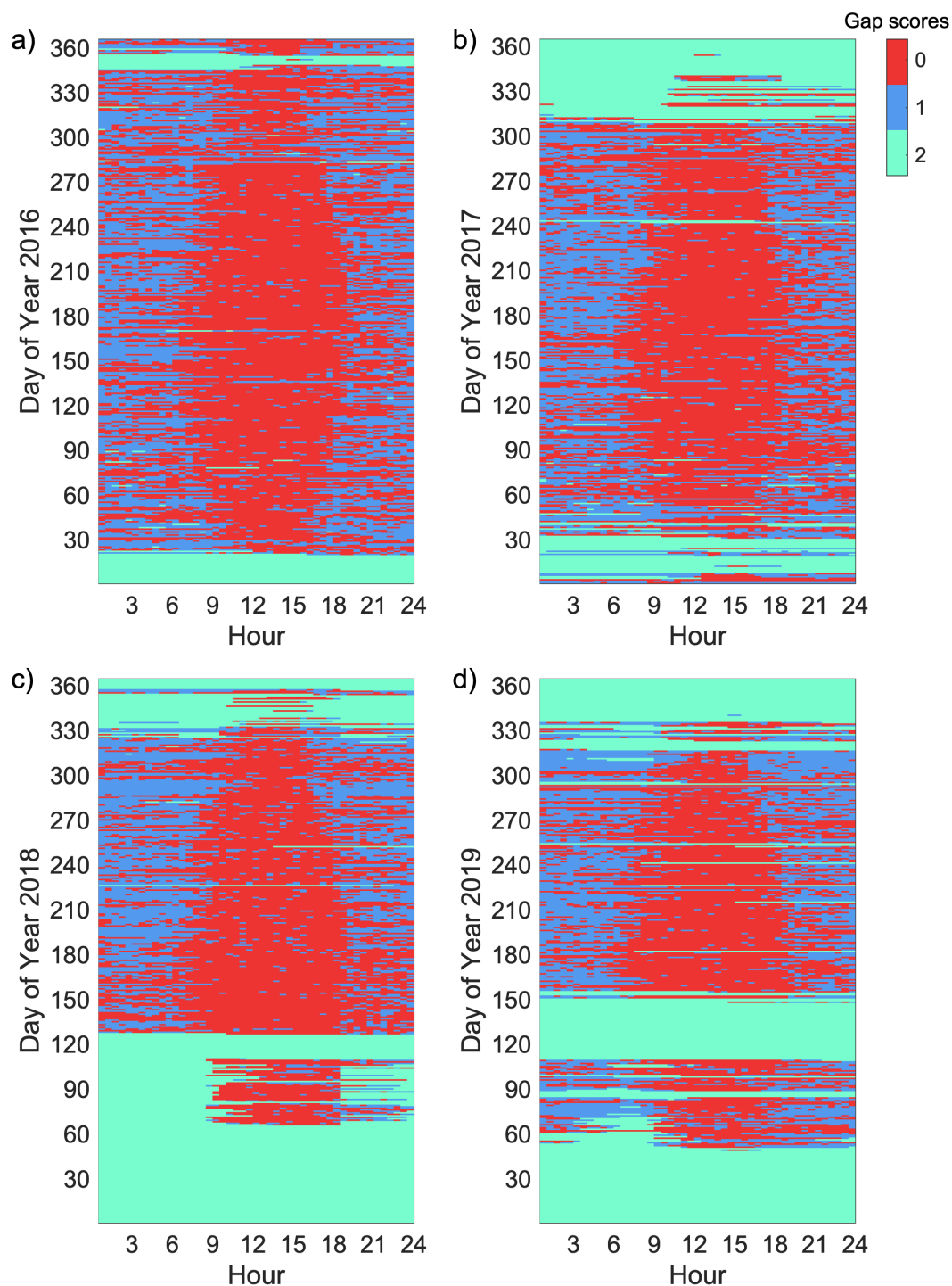



Figure 1 Distribution of NEE data gaps by day and hour from 2016 to 2019. NEE data gaps were classified as non-gaps (gap score: 0), gaps due to low data quality (gap score: 1), and gaps due to electrical and/or instrument failures (gap score: 2).

Table 1 Percentage of NEE data with different gap scores (0/1/2) for daytime and nighttime from 2016 to 2019. NEE gaps are classified as non-gaps (gap score: 0), gaps due to low data quality (gap score: 1), and gaps due to electrical and/or instrument failures (gap score: 2).

	Daytime	Nighttime	All
2016	36.1% / 9.6% / 2.7%	19.9% / 26.1% / 5.6%	56.0% / 35.7% / 8.3%
2017	31.5% / 9.8% / 7.1%	15.4% / 22.3% / 13.9%	46.9% / 32.1% / 21.0%
2018	27.8% / 6.8% / 14.1%	10.9% / 16.5% / 23.9%	38.7% / 23.3% / 38.0%
2019	23.4% / 7.1% / 17.9%	10.0% / 18.1% / 23.5%	33.4 % / 25.2 % / 41.4%
Overall	29.7% / 8.4% / 10.5%	14.0% / 20.7% / 16.7%	43.7% / 29.1% / 27.2%

3.2 Performance of ML gap-filling algorithms with different gap lengths

We first train and evaluate the ML algorithms without separating the daytime and nighttime data (i.e., using all the data). The overall performance of each algorithm degrades as the gap length increases, while the RF slightly outperforms the other algorithms for all the gap scenarios (Fig. 2 and Table 2). For the gap length of one hour (1-H), all the four ML algorithms have the highest R^2 and the lowest ARMSE, RRMSE, and MAE. As the gap length increases, R^2 for all ML algorithms decreases and ARMSE (RRMSE and MAE) increases. For 1-H, the RF has R^2, ARMSE, and RRMSE of 0.77±0.02, 0.66±0.03 µmol m^{-2} s^{-1}, and 0.47±0.02, respectively, whereas for the gap length of two months (2-M), the R^2 decreases to 0.60±0.22, and ARMSE and RRMSE increase to 0.91±0.07 µmol m^{-2} s^{-1} and 0.62±0.17, respectively. The magnitudes of BE for all ML algorithms also increase with the increasing gap lengths. The performance of these ML algorithms in NEE gap-filling at this semi-arid sagebrush site is comparable to that at some grassland sites (Huang & Hsieh, 2020), but lower than that at forest and cropland sites (Huang & Hsieh, 2020; Moffat et al., 2007).

Following Moffat et al. (2007), we also perform the algorithm training and evaluation separately for daytime and nighttime data. For the daytime data, the performance of each algorithm is similar to that for all the data, whereas the algorithm performance for the nighttime data is degraded with R^2 of 0.1–0.2 and RRMSE of 0.7–0.8, similar to the results for some forest sites in Moffat et al. (2007). The poor performance of the gap-filling algorithms for the nighttime data is primarily attributed to the shortage of available nighttime data for training the models (Fig. 1 and Table 1). However, the change of BE with the increased gap length at night is relatively small compared to that for the daytime data, especially for the ANN and RF. In addition, BE for long gaps (e.g., 2-M) has opposite signs for the daytime and nighttime data, resulting in a smaller BE for all the data.

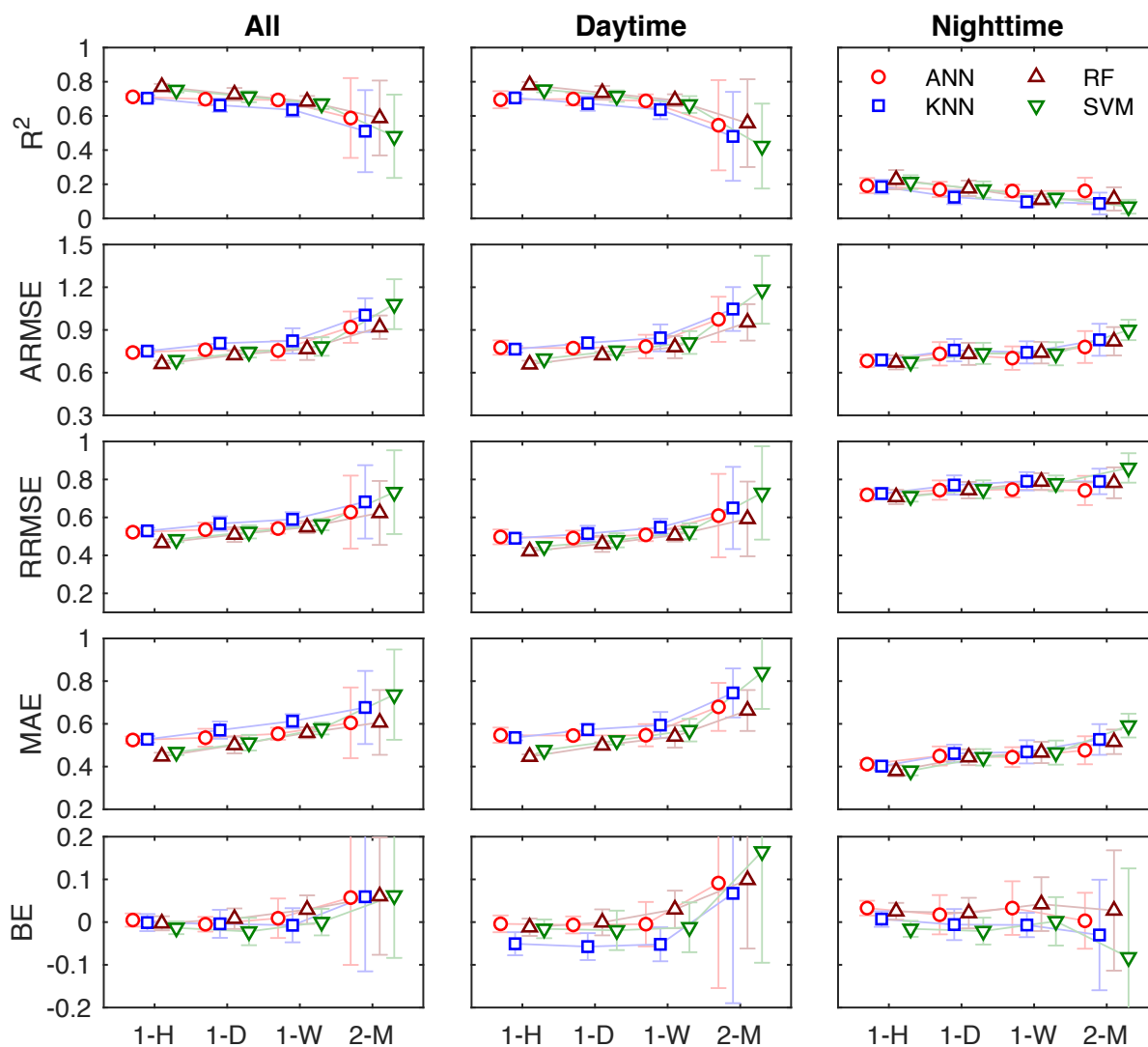



Figure 2 Performance of NEE gap-filling algorithms for the four gap length scenarios (i.e., one-hour (1-H), one-day (1-D), one-week (1-W), and two-month (2-M), respectively).

Table 2 Performance metrics of the four ML gap-filling algorithms for the four gap length scenarios. The ML algorithms are trained and evaluated without separating daytime and nighttime data.

	Gap length scenarios	R^2	ARMSE ($\mu mol\ m^{-2}\ s^{-1}$)	RRMSE	MAE ($\mu mol\ m^{-2}\ s^{-1}$)	BE ($\mu mol\ m^{-2}\ s^{-1}$)
ANN	One-hour	0.71±0.02	0.74±0.02	0.52±0.02	0.50±0.01	-0.005±0.016
	One-day	0.70±0.02	0.76±0.04	0.54±0.04	0.51±0.02	0.005±0.017
	One-week	0.69±0.03	0.76±0.07	0.54±0.03	0.51±0.04	-0.009±0.046
	Two-month	0.59±0.23	0.92±0.11	0.63±0.19	0.62±0.07	-0.057±0.158
KNN	One-hour	0.71±0.02	0.75±0.02	0.53±0.02	0.50±0.01	0.001±0.020
	One-day	0.66±0.04	0.81±0.04	0.57±0.04	0.55±0.02	0.004±0.033
	One-week	0.63±0.05	0.83±0.09	0.59±0.04	0.57±0.06	0.007±0.040
	Two-month	0.51±0.24	1.01±0.12	0.68±0.19	0.69±0.08	-0.059±0.175
RF	One-hour	0.77±0.02	0.66±0.03	0.47±0.02	0.43±0.01	0.003±0.017
	One-day	0.73±0.04	0.72±0.03	0.51±0.04	0.48±0.02	-0.008±0.024
	One-week	0.68±0.03	0.76±0.07	0.55±0.03	0.52±0.05	-0.029±0.033
	Two-month	0.60±0.22	0.91±0.07	0.62±0.17	0.61±0.05	-0.061±0.137
SVM	One-hour	0.75±0.02	0.69±0.02	0.48±0.02	0.45±0.01	0.014±0.015
	One-day	0.71±0.03	0.74±0.04	0.52±0.03	0.49±0.02	0.022±0.032
	One-week	0.67±0.03	0.78±0.06	0.56±0.03	0.53±0.04	-0.000±0.031
	Two-month	0.47±0.25	1.10±0.20	0.73±0.22	0.77±0.16	-0.062±0.146

3.3 Comparison of probability density functions (PDFs) between measured and estimated NEE

Figure 3 shows the comparison of the probability density functions (PDFs) between the measured and estimated NEE of the gap-filling algorithms. The estimated NEE by the gap-filling algorithms has similar shape of PDFs, which is also quite similar to that of the measured NEE with some difference in amplitudes. At this site, the measured half-hourly NEE ranges from -6 to 4 $\mu mol\ m^{-2}\ s^{-1}$, while the estimated NEE varies from approximately -5 to 2 $\mu mol\ m^{-2}\ s^{-1}$. For all the data, the PDFs of the estimated NEE present two peaks at around -0.2 and 0.2 $\mu mol\ m^{-2}\ s^{-1}$, whereas the PDF of the measured NEE only has one peak at around 0.2 $\mu mol\ m^{-2}\ s^{-1}$. That means that the PDFs of the estimated NEE have higher amplitude than the measured NEE in the range of -0.6 to 0.0 $\mu mol\ m^{-2}\ s^{-1}$. However, in the range of -2.0 to -0.8 $\mu mol\ m^{-2}\ s^{-1}$, the amplitude of the PDFs of the estimated NEE is lower than that of the measured NEE. The PDFs of the estimated NEE for the daytime data show one peak at around -0.2 $\mu mol\ m^{-2}\ s^{-1}$ with their shapes similar to those for all the data in the range with the negative NEE values; for the nighttime data, the estimated NEE has a narrower shape of PDF than that of the measured NEE. These

results suggest that the gap-filling algorithms underestimate the magnitudes of NEE in the range of -2.0 to 0.0 μmol m^{-2} s^{-1} and the magnitudes of the peak values of NEE.

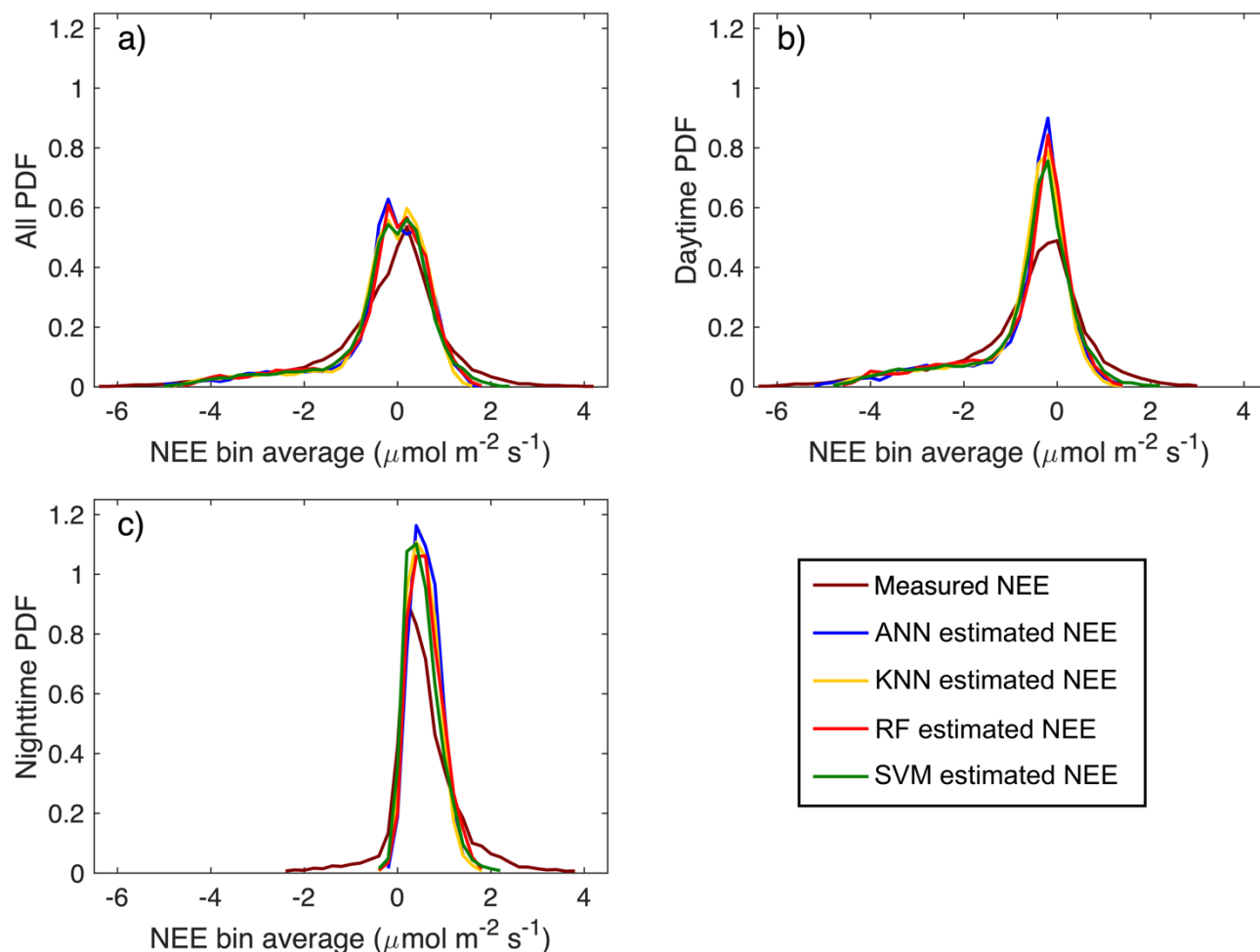



Figure 3 Comparison of probability density functions (PDFs) for the measured and estimated NEE of the gap-filling algorithms using daytime and nighttime data together and separately. The bin width is 0.2 μmol m^{-1} s^{-1} and bins with less than 10 data points are excluded.

3.4 Comparison of measurement error and absolute error of ML algorithms

Figure 4 shows the absolute errors as a function of bin-averaged NEE values for the gap-filling algorithms with the estimated random measurement errors (Finkelstein and Sims, 2001) as a reference. For all the data, the binned absolute errors for the ML algorithms are quite close to each other, and they are also close to the random errors in the range of negative NEE values but slightly deviated from the random errors in the range of positive NEE values. Note that the large deviations from the random errors at the edges are most likely due to the small number of data points as illustrated in Fig. 3. The binned absolute

errors for the daytime data are close to the random errors in the NEE range of about -4.0 to 1.0 µmol m^{-2} s^{-1}; whereas the binned absolute errors for the nighttime data are consistently higher than the random errors. For all the data, the mean value of the random errors is 0.56 µmol m^{-2} s^{-1}, while the MAE of the ML algorithms is 0.55, 0.59, 0.53, and 0.58 µmol m^{-2} s^{-1} for

225 the ANN, KNN, RF, and SVM, respectively. For the daytime data, the MAE of the ML algorithms (0.59, 0.62, 0.55, and 0.62 µmol m^{-2} s^{-1}) is all smaller than the mean value of the random error (0.66 µmol m^{-2} s^{-1}); whereas for the nighttime data, the MAE of the ML algorithms (0.46, 0.48, 0.47, and 0.49 µmol m^{-2} s^{-1}) is all larger than the mean value of the random error (0.32 µmol m^{-2} s^{-1}). Overall, the RF and ANN have better performance in filling NEE gaps at this semi-arid site, especially for the daytime data, although the RF is more time efficient than the ANN. In addition, all the four ML algorithms have low

230 performance in gap-filling the nighttime data, though the BE for the nighttime data is relatively small compared to the daytime data mostly due to the low magnitudes of the nighttime NEE.

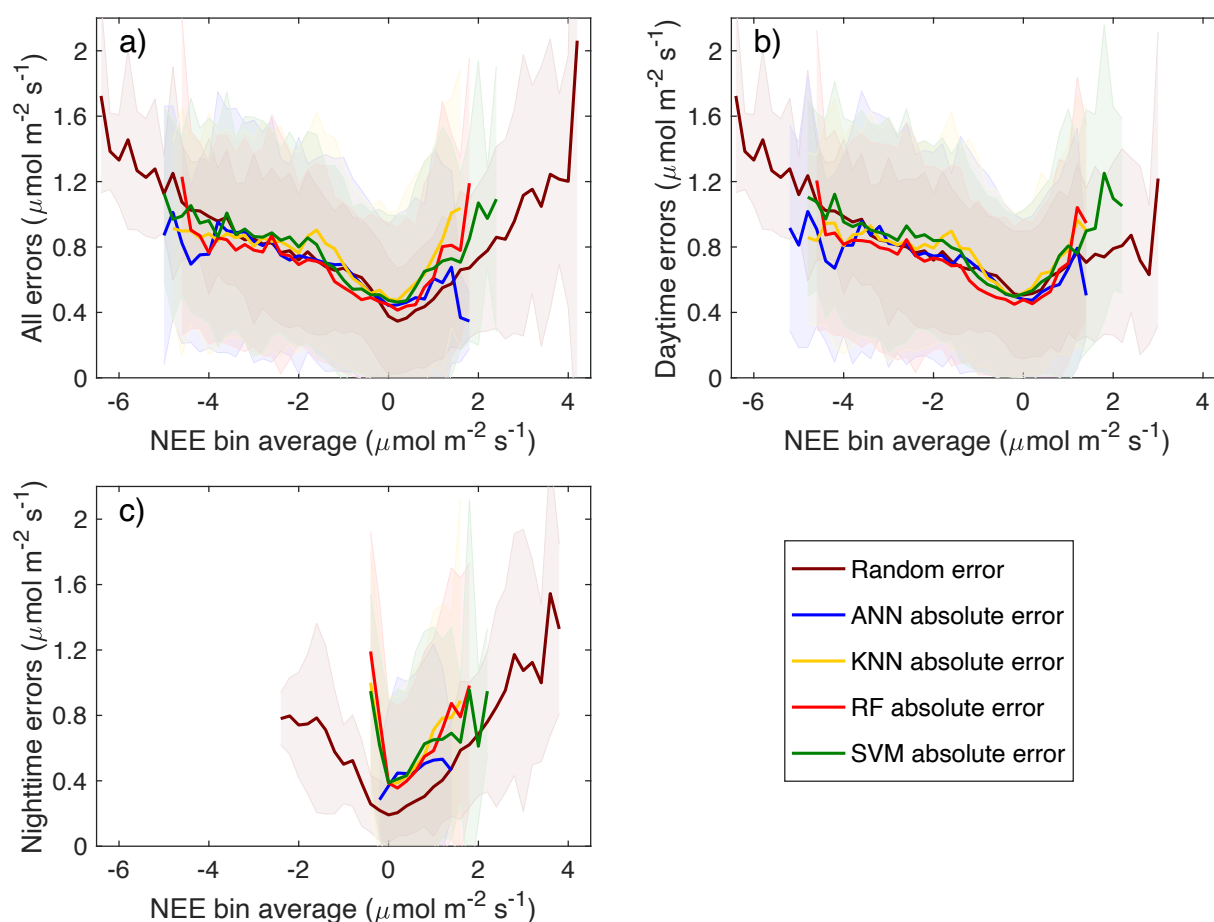



Figure 4 Comparison of NEE measurement and model uncertainties of the gap-filling algorithms using the data and the daytime and nighttime data. The random measurement error is estimated by Finkelstein and Sims (2001). The bin width is 0.2 µmol m^{-2} s^{-1} and bins with less than 10 data points are excluded.

4 Discussion

4.1 Uncertainties in carbon budgets caused by gap-filling

We now examine uncertainties in carbon budgets caused by gap filling with different training dataset and different methods. With the 40 distinct artificial gap scenarios, we obtain 40 gap-filled NEE time series for each method. By replacing the artificial gaps with the observed data, these time series datasets allow for an evaluation of the model consistency and reliability in filling the actual gaps. Figure 5 compares the monthly accumulative NEE of the gap-filled data during 2016-2019. The subscript A denotes that the daytime and nighttime NEE data are gap-filled together using the ML algorithms trained with all the data, and DN denotes that the daytime and nighttime data are gap-filled separately using the trained ML algorithms and then combined together to determine the monthly accumulative NEE. The error bar denotes one standard deviation of the monthly accumulative NEE of the 40 gap-filled time series. Here the gap-filled NEE using the MDS method is also included for comparison.

For months with gaps less than 7 days (Fig. 1; e.g., February to October in 2016), all the four ML algorithms are quite consistent and reliable with the mean standard deviations of the monthly accumulative NEE ranging from 0.4 (ANN) to 1.0 (SVM) g C m^{-2}. Both the RF and KNN have mean standard deviations of approximately 0.7 g C m^{-2} during these months. For months with long gaps (i.e., > 7 days), the mean standard deviations of the monthly accumulative NEE are 1.2, 1.3, 1.5, and 3.3 g C m^{-2} for the ANN, KNN, RF, and SVM, respectively. From this perspective, the ANN is the most reliable method in gap-filling because the predicted values are averages of the best 20 runs (Section 2.4.1). In other words, most of the ML algorithms are quite consistent in filling the gaps, and the differences in the monthly accumulative NEE caused by changes in training dataset are less than 1.5 g C m^{-2} except for the SVM.

The uncertainties in the monthly accumulative NEE as a result of the differences in the methods are now assessed. The difference among the methods ranges from 0.2 to 1.3 g C m^{-2} for months with gaps less than 7 days, and changes from 0.8 to 10.2 g C m^{-2} for months with long gaps, which is 0.8 to 4.8 g C m^{-2} without including the SVM. For the ANN and RF, the differences in the monthly accumulative NEE between A and DN are usually quite small for all the months, whereas the KNN presents opposite signs for A and DN in months with long gaps, which means that the KNN is unable to handle the daytime and nighttime data together. Overall, for months with short to medium gaps (i.e., < 7 days), there is no significant difference in the monthly accumulative NEE among the methods including the MDS method. For months with long gaps

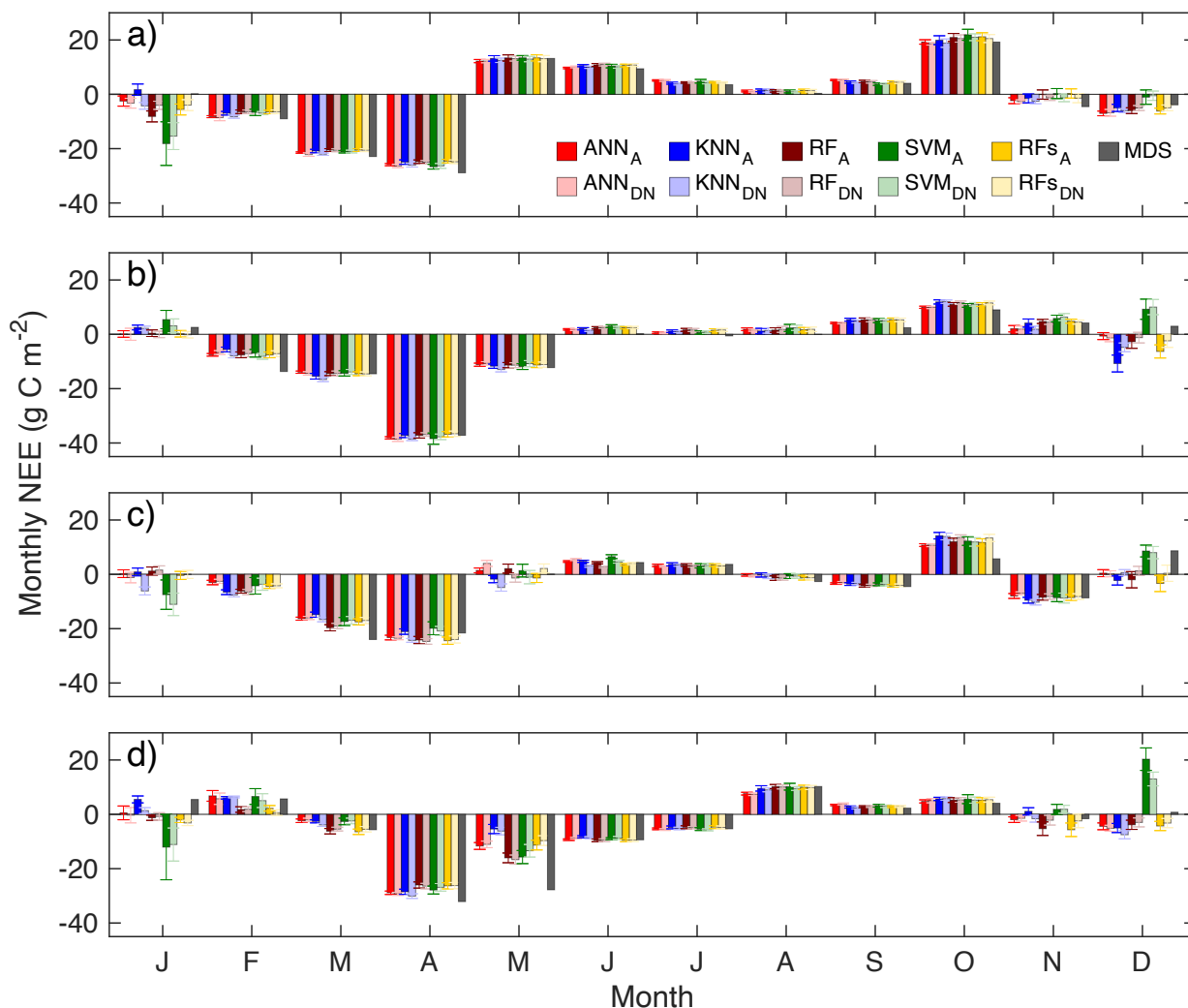
(i.e., > 7 days), the MDS method usually fails, and the ANN and RF have the best performance, and they have the potential to handle the daytime and nighttime data gap-filling together.

265

Figure 5 Monthly total net ecosystem exchange (NEE, g C m^{-2}) at the US-Hn1 in (a-d) 2016-2019. The subscript A denotes that the daytime and nighttime NEE data are gap-filled together using the ML algorithms trained with all the data, and DN denotes that the daytime and nighttime data are gap-filled separately and then combined to determine the monthly total values. RFs denotes the proposed two-layer RF based gap-filling framework, and MDS is the marginal distribution sampling algorithm. The error bar denotes one standard deviation

270 of the monthly total NEE of the 40 gap-filled time series.

The uncertainties in the annual total NEE are estimated by comparing the annual NEE obtained by each ML algorithm with their ensemble means (Fig. 6). Obviously, both the KNN and SVM are largely apart from the ensemble means, whereas the ANN and RF are relatively close to the ensemble means except for the RF in 2019. The mean standard deviations of the 40

trials for the annual NEE by ANN range from 2.5 to 4.3 g C m^{-2}, the mean standard deviations by the RF vary from 3.4 to 7.1
g C m^{-2}, and the mean standard deviations by the KNN and SVM vary from 5.1 to 13.9 g C m^{-2}. The differences between the
ANN and RF are within ±8.4 g C m^{-2}; and the differences between A and DN are less than 1.7 and 4.5 g C m^{-2} for the ANN
and RF, respectively. The overall uncertainties in the annual NEE caused by the ANN and RF are usually less than 15.5 g C
m^{-2}, while the uncertainties can be as large as 27.2 g C m^{-2} if including the KNN and SVM. Therefore, it is recommended to
use the ensemble mean of the ANN and RF as the best estimate of the annual NEE at the semi-arid sagebrush site. The
annual mean NEE by the ANN and RF is -15.4±4.7, -50.0±9.1, -31.4±7.2, and -40.3±8.7 g C m^{-2} for 2016, 2017, 2018, and
2019, respectively. In addition, the annual total NEE by the MDS is about 5.6~15.6 g C m^{-2} larger than that by the ANN and
RF.

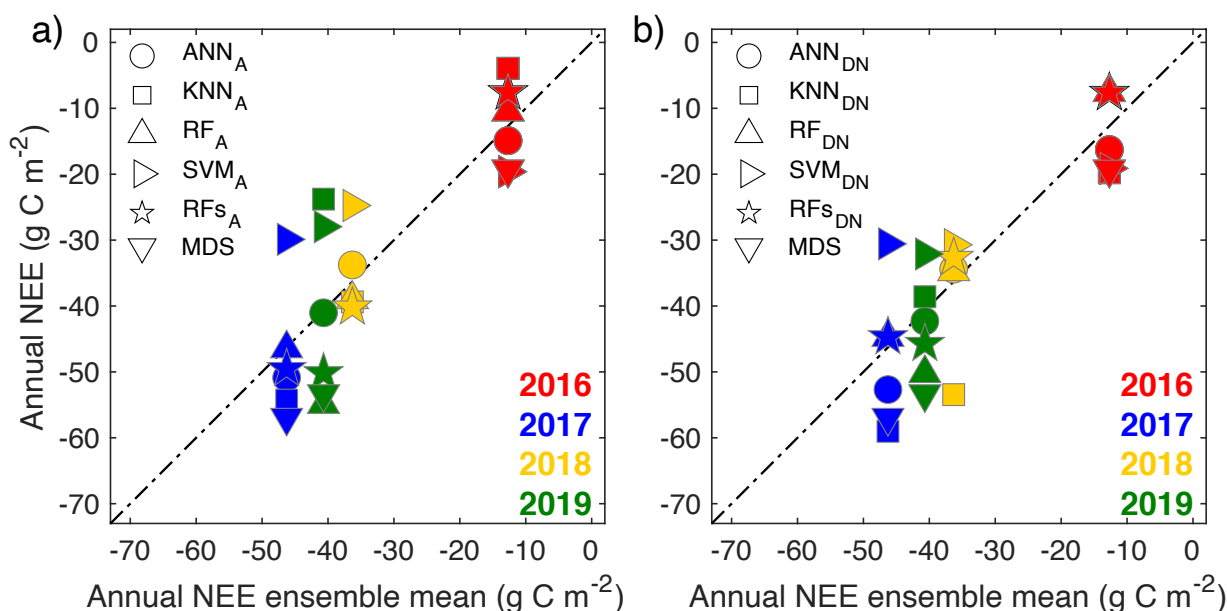



Figure 6 Comparison of the annual mean NEE from the ML algorithms and their ensemble mean. The subscript A denotes that daytime
and nighttime NEE data are gap-filled together using the ML algorithms trained with all the data, and DN denotes that the daytime and
nighttime data are gap-filled separately and then combined to determine the monthly total values.

4.2 A two-layer RF based gap-filling framework for extremely long gaps

We then propose and test a two-layer RF-based gap-filling framework (RFs) because 1) the performance of the ML
algorithms is only reliable for short to medium gaps; 2) there is no obvious difference in monthly and annual NEE between
training the model using all the data or the separated daytime and nighttime data; and 3) the RF is more time efficient than
the ANN. The procedures include: 1) train the RF model using the half-hourly data and fill the NEE gaps shorter than 7 days;
2) calculate daily means of the input variables and the partially filled NEE data; and 3) train the RF model using the daily

data and fill the gaps in the daily NEE data. Following the same procedures, the performance of this framework in filling the gaps in the daytime and nighttime daily means are also accessed. Using a 10-fold cross-validation repeated ten times, the R^2 of the second layer RF model is 0.78, 0.85, and 0.77 for all the data, the daytime data, and the nighttime data, respectively. The mean BE for this framework is 0.04 g C m^{-2} day^{-1}, which is slightly smaller than that of the original RF with the half-hourly NEE data (0.06 g C m^{-2} day^{-1}).

As shown in Figs. 5 and 6, the monthly and annual NEE by the RFs are quite close to those by the RF. For A (i.e., trained by all the data), the difference in the annual mean NEE ranges from 0.0 to 2.2 g C m^{-2}, whereas for DN (i.e., trained by the separated daytime and nighttime data), the difference varies from 0.8 to 3.3 g C m^{-2}. This test suggests that it is not necessary to fill all the gaps in the half-hourly NEE data if the focus is on assessing the uncertainties in cumulative annual mean NEE and interannual variability. In addition, the performance of the different methods is quite consistent when filling short to medium gaps (e.g., < 7 days), and thus using the ensemble mean of different methods as the input of the second RF layer has the potential to lower the uncertainties in the gap-filled data.

5 Conclusions

The performance of the four ML algorithms in filling the NEE data gaps is evaluated at a semi-arid sagebrush ecosystem site. Due to the relatively small range of NEE variations, the overall performance of these gap-filling algorithms at this site is lower than that at other forest sites, but comparable to that at other grassland sites. The RF algorithm outperforms the other algorithms in terms of the overall performance. It is not necessary to train the model separately for daytime and nighttime data when using the ANN and RF algorithms. The uncertainties in the monthly and annual NEE due to the gap-filling approaches are evaluated by the standard deviations of monthly NEE of multiple trials for each method and also accessed by the difference in the monthly NEE by the methods. With the ANN and RF, the uncertainties in annual NEE are usually within 16 g C m^{-2} at this semi-arid sagebrush site. Extremely long gaps in half-hourly NEE data due to power failures cannot be confidently filled by either of the methods because of the high uncertainties in R^2 and RRMSE, and thus we propose and test a two-layer RF based gap-filling framework. With this framework, the improvement in model performance is significant, especially for the nighttime data. Therefore, it is recommended that the two-layer RF based framework should be used if there are extremely long gaps existed in the NEE dataset and if there is a need to investigate its annual and interannual variability.

Code availability

The code used in this study are available from J.Y. and J.H. upon request.

Data availability

325   The data used in this study are deposited in a public domain repository at https://doi.org/10.6084/m9.figshare.14747952.

Author contribution

J.Y., Z.G., and H.L. contributed equally to this work. J.H designed the study with substantial input from all coauthors. Z.G. and H.L. conducted the field work and obtained and processed the EC data. J.Y. performed the gap-filling and analyzed the

330   results. J.Y and Z.G. drafted the manuscript. J.H., H.L., and G.W. contributed to the result analysis and interpretation. All authors commented on and approved the final manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

335   We acknowledge support by the Second Tibetan Plateau Scientific Expedition and Research Program (STEP), Grant no. 2019QZKK0602, the National Natural Science Foundation of China under grants 41521004, 41991231, and 41975075, the Foundation of Key Laboratory for Semi-Arid Climate Change of the Ministry of Education in Lanzhou University, the China 111 Project (no. B13045). We also acknowledge support by the U.S. Department of Energy (DOE) Office of Biological and Environmental Research (BER) as part of BER's Subsurface Biogeochemistry Research Program (SBR) at the Pacific

340   Northwest National Laboratory (PNNL). PNNL is operated by Battelle Memorial Institute for the U.S. DOE under Contract Number DE-AC05-76RL01830. H. Liu acknowledges support by National Science Foundation (NSF-AGS-1419614 and NSF-AGS-1853050).

References

345   Baldocchi, D. and Sturtevant, C.: Does day and night sampling reduce spurious correlation between canopy photosynthesis and ecosystem respiration?, Agric. For. Meteorol., 207, 117–126, doi:10.1016/j.agrformet.2015.03.010, 2015.

Baldocchi, D. D.: Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future, Glob. Chang. Biol., 9(4), 479–492, doi:10.1046/j.1365-2486.2003.00629.x, 2003.

Biederman, J. A., Scott, R. L., Bell, T. W., Bowling, D. R., Dore, S., Garatuza-Payan, J., Kolb, T. E., Krishnan, P.,
350   Krofcheck, D. J., Litvak, M. E., Maurer, G. E., Meyers, T. P., OECHEL, W. C., Papuga, S. A., Ponce-Campos, G. E., Rodriguez, J. C., Smith, W. K., Vargas, R., Watts, C. J., Yepez, E. A., Goulden, M. L., Garatuza Payan, J., Kolb, T. E., Krishnan, P., Krofcheck, D. J., Litvak, M. E., Maurer, G. E., Meyers, T. P., OECHEL, W. C., Papuga, S. A., Ponce-Campos, G. E., Rodriguez, J. C., Smith, W. K., Vargas, R., Watts, C. J., Yepez, E. A. and Goulden, M. L.: CO2 exchange and evapotranspiration across dryland ecosystems of southwestern North America, Glob. Chang. Biol., 23(10), 4204–4221,

355   doi:10.1111/gcb.13686, 2017.

Breiman, L.: Random forests, Mach. Learn., 45(1), 5–32, doi:10.1023/A:1010933404324, 2001.

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

360   Cortes, C. and Vapnik, V.: Support-vector networks, Mach. Learn., 20(3), 273–297, doi:10.1007/bf00994018, 1995.

Dengel, S., Zona, D., Sachs, T., Aurela, M., Jammet, M., Parmentier, F. J. W., Oechel, W. and Vesala, T.: Testing the applicability of neural networks as a gap-filling method using CH4 flux data from high latitude wetlands, Biogeosciences, 10(12), 8185–8200, doi:10.5194/bg-10-8185-2013, 2013.

Didan, K.: MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006, , NASA EOSDI,
365   doi:10.5067/MODIS/MOD13Q1.006, 2015.

Dragoni, D., Schmid, H. P., Grimmond, C. S. B. and Loescher, H. W.: Uncertainty of annual net ecosystem productivity estimated using eddy covariance flux measurements, J. Geophys. Res. Atmos., 112(17), 17102, doi:10.1029/2006JD008149, 2007.

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R.,
370   Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N. O., Katul, G., Keronen, P., Kowalski, A., Lai, C. T., Law, B. E., Meyers, T., Moncrieff, J., Moors, E., Munger, J. W., Pilegaard, K., Rannik, Ü., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K. and Wofsy, S.: Gap filling strategies for defensible annual sums of net ecosystem exchange, Agric. For. Meteorol., 107(1), 43–69, doi:10.1016/S0168-1923(00)00225-2, 2001.

Finkelstein, P. L. and Sims, P. F.: Sampling error in eddy correlation flux measurements, J. Geophys. Res. Atmos., 106(D4),
375   3503–3509, doi:10.1029/2000JD900731, 2001.

Fix, E. and Hodges, J. L.: Discriminatory analysis. Nonparametric discrimination; consistency properties, Randolph Field, Texas., 1951.

Gao, Z., Russell, E. S., Missik, J. E. C., Huang, M., Chen, X., Strickland, C. E., Clayton, R., Arntzen, E., Ma, Y. and Liu, H.: A novel approach to evaluate soil heat flux calculation: An analytical review of nine methods, J. Geophys. Res., 122(13),
380   6934–6949, doi:10.1002/2017JD027160, 2017.

Gao, Z., Liu, H., Missik, J. E. C., Yao, J., Huang, M., Chen, X., Arntzen, E. and McFarland, D. P.: Mechanistic links between underestimated CO2 fluxes and non-closure of the surface energy balance in a semi-arid sagebrush ecosystem, Environ. Res. Lett., 14(4), doi:10.1088/1748-9326/ab082d, 2019.

Gao, Z., Liu, H., Chen, X., Huang, M., Missik, J. E. C. C., Yao, J., Arntzen, E. and Mcfarland, D. P.: Enlarged non-closure
385   of surface energy balance with increasing atmospheric instabilities linked to changes in coherent structures, J. Geophys. Res. Atmos., 125(14), doi:10.1029/2020jd032889, 2020a.

Gao, Z., Liu, H., Arntzen, E., Mcfarland, D. P., Chen, X. and Huang, M.: Uncertainties in Turbulent Statistics and Fluxes of CO2 Associated With Density Effect Corrections, Geophys. Res. Lett., 47(15), doi:10.1029/2020GL088859, 2020b.

Günther, F. and Fritsch, S.: Neuralnet: Training of neural networks, R J., 2(1), 30–38, doi:10.32614/rj-2010-006, 2010.

390   Huang, I.-H. and Hsieh, C.-I.: Gap-Filling of Surface Fluxes Using Machine Learning Algorithms in Various Ecosystems, Water, 12(12), 3415, doi:10.3390/w12123415, 2020.

Jammet, M., Crill, P., Dengel, S. and Friborg, T.: Large methane emissions from a subarctic lake during spring thaw: Mechanisms and landscape significance, J. Geophys. Res. Biogeosciences, 120(11), 2289–2305, doi:10.1002/2015JG003137, 2015.

395   Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rödenbeck, C., Tramontana,

G., Viovy, N., Wang, Y. P., Weber, U., Zaehle, S. and Zeng, N.: Compensatory water effects link yearly global land CO2 sink changes to temperature, Nature, 541(7638), 516–520, doi:10.1038/nature20780, 2017.

Karatzoglou, A., Smola, A., Hornik, K. and Zeileis, A.: kernlab - An S4 Package for Kernel Methods in R, J. Stat. Softw., 400   11(9), 1–20, doi:10.18637/jss.v011.i09, 2004.

Kim, Y., Johnson, M. S., Knox, S. H., Black, T. A., Dalmagro, H. J., Kang, M., Kim, J. and Baldocchi, D.: Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis, Glob. Chang. Biol., 26(3), 1499–1518, doi:10.1111/gcb.14845, 2020.

Knox, S. H., Matthes, J. H., Sturtevant, C., Oikawa, P. Y., Verfaillie, J. and Baldocchi, D.: Biophysical controls on 405   interannual variability in ecosystem-scale CO2 and CH4 exchange in a California rice paddy, J. Geophys. Res. Biogeosciences, 121(3), 978–1001, doi:10.1002/2015JG003247, 2016.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C. and Hunt, T.: caret: Classification and Regression Training, R package version 6.0-86, Retrieved from: https://cran.r-project.org/package=caret, 2020.

410   Liaw, A. and Wiener, M.: Classification and Regression by randomForest, R News, 2(3), 18–22, 2002.

Ma, S., Baldocchi, D. D., Xu, L. and Hehn, T.: Inter-annual variability in carbon dioxide exchange of an oak/grass savanna and open grassland in California, Agric. For. Meteorol., 147(3–4), 157–171, doi:10.1016/j.agrformet.2007.07.008, 2007.

Massman, W. J.: A simple method for estimating frequency response corrections for eddy covariance systems, Agric. For. Meteorol., 104(3), 185–198, doi:10.1016/S0168-1923(00)00164-7, 2000.

415   Massman, W. J.: Reply to comment by Rannik on "A simple method for estimating frequency response corrections for eddy covariance systems," Agric. For. Meteorol., 107(3), 247–251, doi:10.1016/S0168-1923(00)00237-9, 2001.

Mauder, M. and Foken, T.: Documentation and Instruction Manual of the Eddy Covariance Software Package TK2, Abt. Mikrometeorologie, Arbeitsergebnisse, 26, Print: ISSN 1614-8916, 42 pp, 2004.

Missik, J. E. C., Liu, H., Gao, Z., Huang, M., Chen, X., Arntzen, E., Mcfarland, D. P., Ren, H., Titzler, P. S., Thomle, J. N. 420   and Goldman, A.: Groundwater-River Water Exchange Enhances Growing Season Evapotranspiration and Carbon Uptake in a Semiarid Riparian Ecosystem, J. Geophys. Res. Biogeosciences, 124(1), 99–114, doi:10.1029/2018JG004666, 2019.

Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., Beckstein, C., Braswell, B. H., Churkina, G., Desai, A. R., Falge, E., Gove, J. H., Heimann, M., Hui, D., Jarvis, A. J., Kattge, J., Noormets, A. and Stauch, V. J.: Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes, Agric. For. Meteorol., 425   147(3–4), 209–232, doi:10.1016/j.agrformet.2007.08.011, 2007.

Moncrieff, J., Clement, R., Finnigan, J. and Meyers, T.: Averaging, Detrending, and Filtering of Eddy Covariance Time Series, in Handbook of Micrometeorology, edited by X. Lee, W. Massman, and B. Law, pp. 7–31, Springer, Dordrecht, Dordrecht., 2004.

Oncley, S., Foken, T., Vogt, R., Kohsiek, W., DeBruin, H. A. R., Bernhofer, C., Christen, A., Gorsel, E., Grantz, D., 430   Feigenwinter, C., Lehner, I., Liebethal, C., Liu, H., Mauder, M., Pitacco, A., Ribeiro, L. and Weidinger, T.: The Energy Balance Experiment EBEX-2000. Part I: overview and energy balance, Boundary-Layer Meteorol., 123(1), 1–28 [online] Available from: http://dx.doi.org/10.1007/s10546-007-9161-1, 2007.

Papale, D. and Valentini, R.: A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization, Glob. Chang. Biol., 9(4), 525–535, doi:10.1046/j.1365-2486.2003.00609.x, 2003.

435   Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y. W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Ribeca, A., van Ingen, C., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardö, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J. M., Bowling, D. R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N.,

440 Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K. J., De Cinti, B., de Grandcourt, A., De Ligne, A., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di Bella, C. M., di Tommasi, P., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, 445 A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Sci. data, 7(1), 225, doi:10.1038/s41597-020-0534-3, 2020.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., 450 Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J. M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D. and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm, Glob. Chang. Biol., 11(9), 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x, 2005.

Richardson, A. D. and Hollinger, D. Y.: A method to estimate the additional uncertainty in gap-filled NEE resulting from 455 long gaps in the CO2 flux record, Agric. For. Meteorol., 147(3–4), 199–208, doi:10.1016/j.agrformet.2007.06.004, 2007.

Scott, R. L., Biederman, J. A., Hamerlynck, E. P., Barron Gafford, G. A. and Barron-Gafford, G. A.: The carbon balance pivot point of southwestern U.S. semiarid ecosystems: Insights from the 21st century drought, J. Geophys. Res. Biogeosciences, 120(12), 2612–2624, doi:10.1002/2015JG003181, 2015.

Soloway, A. D., Amiro, B. D., Dunn, A. L. and Wofsy, S. C.: Carbon neutral or a sink? Uncertainty caused by gap-filling 460 long-term flux measurements for an old-growth boreal black spruce forest, Agric. For. Meteorol., 233, 110–121, doi:10.1016/j.agrformet.2016.11.005, 2017.

Stoy, P. C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Arain, M. A., Arneth, A., Aurela, M., Bernhofer, C., Cescatti, A., Dellwik, E., Duce, P., Gianelle, D., van Gorsel, E., Kiely, G., Knohl, A., Margolis, H., McCaughey, H., Merbold, L., Montagnani, L., Papale, D., Reichstein, M., Saunders, M., Serrano-Ortiz, P., Sottocornola, M., Spano, D., 465 Vaccari, F. and Varlagin, A.: A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity, Agric. For. Meteorol., 171–172, 137–152, doi:10.1016/j.agrformet.2012.11.004, 2013.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S. and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences, 13(14), 4291–4313, doi:10.5194/bg-13-470 4291-2016, 2016.

Webb, E. K., Pearman, G. I. and Leuning, R.: Correction of flux measurements for density effects due to heat and water vapour transfer, Q. J. R. Meteorol. Soc., 106(447), 85–100, doi:10.1002/qj.49710644707, 1980.

Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P., Bernhofer, C., Ceulemans, R., Dolman, H., Field, C., Grelle, A., Ibrom, A., Law, B. E., Kowalski, A., Meyers, T., Moncrieff, J., Monson, R., Oechel, W., Tenhunen, J., 475 Valentini, R. and Verma, S.: Energy balance closure at FLUXNET sites, Agric. For. Meteorol., 113(1–4), 223–243, doi:10.1016/S0168-1923(02)00109-0, 2002.

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

480 Xu, T., Guo, Z., Liu, S., He, X., Meng, Y., Xu, Z., Xia, Y., Xiao, J., Zhang, Y., Ma, Y. and Song, L.: Evaluating Different Machine Learning Methods for Upscaling Evapotranspiration from Flux Towers to the Regional Scale, J. Geophys. Res. Atmos., 123(16), 8674–8690, doi:10.1029/2018JD028447, 2018.

Zeng, J., Matsunaga, T., Tan, Z. H., Saigusa, N., Shirai, T., Tang, Y., Peng, S. and Fukuda, Y.: Global terrestrial carbon fluxes of 1999–2019 estimated by upscaling eddy covariance data with a random forest, Sci. Data, 7(1), 1–11,

485   doi:10.1038/s41597-020-00653-5, 2020.