# A comparison of gap-filling algorithms for eddy covariance fluxes and their drivers

**Atbin Mahabbati[1], Jason Beringer[1], Matthias Leopold[1], Ian McHugh[2], James Cleverly[3], Peter Isaac[4], Azizallah Izady[5]**

[1] School of Agriculture and Environment, The University of Western Australia, 35 Stirling Hwy, Crawley, Perth WA, 6009, Australia

[2] School of Ecosystem and Forest Sciences, The University of Melbourne, Richmond, VIC, 3121, Australia

[3] School of Life Sciences University of Technology Sydney Broadway NSW  2007

[4] OzFlux Central Node, TERN Ecosystem Processes, Melbourne, VIC 3159, Australia

[5] Water Research Center, Sultan Qaboos University, Muscat, Oman

*Correspondence to:* Atbin Mahabbati ([atbin.m@hotmail.com](mailto:atbin.m@hotmail.com))

## Abstract

The errors and uncertainties associated with gap-filling algorithms of water, carbon and energy fluxes data, have always been one of the main challenges of the global network of microclimatological tower sites that use eddy covariance (EC) technique. To address these concerns, and find more efficient gap-filling algorithms, we reviewed eight algorithms to estimate missing values of environmental drivers, and separately, nine algorithms for the three major fluxes typically found in EC time series. We then examined the algorithms' performance for different gap-filling scenarios utilising the data from five EC towers during 2013. This research's objectives were a) to evaluate the impact of the gap lengths on the performance of each algorithm; and b) to compare the performance of traditional and new gap-filling techniques for the EC data, for fluxes and separately for their corresponding meteorological drivers. The algorithms' performance was evaluated by generating nine gap windows with different lengths, ranging from a day to 365 days. In each scenario, a gap period was chosen randomly, and the data were removed from the dataset, accordingly. After running each scenario, a variety of statistical metrics were used to evaluate the algorithms' performance. The algorithms showed different levels of sensitivity to the gap lengths; The Prophet Forecast Model (FBP) revealed the most sensitivity, whilst the performance of artificial neural networks (ANNs), for instance, did not vary as much by changing the gap length. The algorithms' performance generally decreased with increasing the gap length, yet the differences were not significant for the windows smaller than 30 days. No significant difference between the algorithms were recognised for the meteorological and environmental drivers. However, the linear algorithms showed slight superiority over those of machine learning (ML), except the random forest algorithm (RF) estimating the ground heat flux (RMSEs of 28.91 and 33.92 for RF and

38 classic linear regression (CLR) respectively). However, for the major fluxes, ML algorithms and the
39 MDS showed superiority over the other algorithms. Even though ANNs, random forest (RF) and
40 extreme gradient boost (XGB) showed comparable performance in gap-filling of the major fluxes, RF
41 provided more consistent results with slightly less bias, as against the other ML algorithms. The results
42 indicated no single algorithm that outperforms in all situations, but the RF is a potential alternative
43 for the MDS and ANNs as regards flux gap-filling.

44

## 1.    Introduction

46        To address the global challenges of climatological and ecological changes, environmental
47 scientists and policymakers are demanding data that are continuous in time and space.  In addition,
48 there is a need for quantifying and reducing uncertainties in such data, including observations of
49 carbon, water and energy exchanges that are crucial components in national/international flux
50 networks and global earth observing systems.  Satellites partially fill this gap as they provide excellent
51 spatial coverage but have limited temporal resolution, and not measured at a point scale. As such,
52 high-quality long-term site observations of ecosystem process and fluxes are needed that are
53 continuous in time and space. The global eddy covariance (EC) flux tower network (FLUXNET),
54 consists of its regional counterparts (i.e. AmeriFlux, EUROFLUX, OzFlux, etc.) and was established in
55 the late 1990s to address the global demand for such information (Aubinet et al., 1999; Baldocchi et al.,
56 2001; Beringer et al., 2016a; Hollinger et al., 1999; Menzer et al., 2013; Tenhunen et al., 1998). Despite
57 EC data being frequently used to validate process modelling analyses, field surveys, and remote
58 sensing assessments (Hagen et al., 2006), there are some serious concerns regarding the technique's
59 challenges, e.g. data gaps and uncertainties. Hence, filling data gaps and reducing uncertainties
60 through better gap-filling techniques are highly needed.

61        Even though the EC is a common technique to measure fluxes of carbon, water and energy,
62 there are some challenges in providing robust, high-quality continuous observations. One of the
63 challenges regarding the technique, and therefore, the flux networks, is addressing data gaps and the
64 uncertainties associated with the gap-filling process, mainly when the gap windows are long (longer
65 than 12 consecutive days, as described by Moffat et al., (2007). These gaps happen quite often for a
66 variety of reasons, such as values out of range, spike detection or manual exclusion of date and time
67 ranges, instrument or power failure, herbivores, fire, eagles nests, lightning, researchers on leave, etc.
68 (Beringer et al., 2016b). Since the EC flux towers are often located in harsh climates, their data are more
69 susceptible to adverse weather (i.e. rain conditions), and they sometimes prevent quick access to sites
70 for repair and maintenance. As a result, this issue can, in turn, produce gaps which might be relatively
71 long (Isaac et al., 2017), and thus, problematic as follows. Firstly, loss of data is considered a threat to
72 scientific studies depending on the missing data quantity, pattern, mechanism and nature (Altman
73 and Bland, 2007; Molenberghs et al., 2014; Tannenbaum, 2010). That is because using an incomplete
74 dataset might lead to biased, invalid and unreliable results (Allison, 2000; Kang, 2013; Little, 2002).
75 Second, continuous gap-filled data are required to calculate the annual or monthly budgets of carbon
76 or water balance components (Hutley et al., 2005).

Other than the challenges caused by missing data, there are several sources of errors and uncertainties in the EC technique. Firstly, random error is associated with the stochastic nature of turbulence, associated sampling errors (incomplete sampling of large eddies, uncertainty in the calculated covariance between the vertical wind velocity and the scalar of interest), instrument errors, and footprint variability (Aubinet et al., 2012). For instance, Dragoni et al. (2007) analysed EC-based data of Morgan-Monroe State Forest for eight years (1999-2006) and assessed that instrument uncertainty was equal to 3% of the total annual NEE. Another primary source of uncertainty in EC measurements is systematic errors caused by methodological challenges and instrument calibration problems (e.g. sonic anemometer errors, spikes, gas analyser errors, etc.). Finally, one of the sources of uncertainties is data processing, especially data gap-filling (Isaac et al., 2017; Moffat et al., 2007; Richardson et al., 2012; Richardson and Hollinger, 2007).

There are several uncertainties pertaining to gap-filling of missing values, including measurement uncertainty (Richardson and Hollinger, 2007), lengths and timing of the gaps (Falge et al., 2001; Richardson and Hollinger, 2007) and the particular gap-filling algorithm that is used (Falge et al., 2001; Moffat et al., 2007). However, there are two dominant issues of long data gaps and the choice of a particular gap-filling algorithm (Aubinet et al., 2012). Firstly, long gaps can significantly increase the total amount of uncertainty as the ecosystem behaviour might change because of different agricultural periods or phenological phases (e.g. growing season, harvest period, bushfire, etc.). And thereby show different responses under similar meteorological conditions (Aubinet et al., 2012; Isaac et al., 2017; Richardson and Hollinger, 2007). Consequently, the period in which a long gap happens is important. For example, research undertaken by Richardson & Hollinger (2007) on data from a range of FLUXNET sites revealed that a week data gap during spring green-up in a forest led to a higher uncertainty over a three-week gap period during winter. Second, each gap-filling algorithm has its strengths and weaknesses; for instance, Moffat et al. (2007) compared 15 different commonly-used gap-filling algorithms. They found no significant difference between the performance of the algorithms with "good" reliability based on analysis of variance of RMSE. Besides, the overall gap-filling uncertainty was within $\pm 25$ g C m$^{-2}$ yr$^{-1}$ for most of the proper algorithms, whereas, the other algorithms generated higher uncertainties of up to $\pm 75$ g C m$^{-2}$ yr$^{-1}$, showing that the uncertainty provided by reliable methods can be considerably smaller. This result is similar to the findings of Richardson & Hollinger (2007) who found that for the datasets used in their study that uncertainties of up to $\pm 30$ g C m$^{-2}$ yr$^{-1}$ were from long gaps by appropriate algorithms. Considering that the data provided by EC tower networks are of use for research, government and policymakers, robust gap-filling is a need to quantify and reduce uncertainties in flux estimations.

Several methods have been typically used to fill data gaps in both fluxes and their meteorological drivers to manage the missing data problem. Due to computational constraints of complex algorithms, early works to impute EC data gaps used interpolation methods based mostly on linear regression or temporal autocorrelation (Falge et al., 2001; Lee et al., 1999). These approaches

116 were replaced quickly by more sophisticated methods such as non-linear regressions (Barr et al., 2004;
117 Falge et al., 2001; Moffat et al., 2007; Richardson et al., 2006); look-up tables (Falge et al., 2001; Law et
118 al., 2002; Zhao and Huang, 2015); artificial neural networks (ANNs) (Aubinet et al., 1999; Beringer et
119 al., 2016a; Cleverly et al., 2013; Hagen et al., 2006; Isaac et al., 2017; Kunwor et al., 2017; Moffat et al.,
120 2007; Papale and Valentini, 2003; Pilegaard et al., 2001; Staebler, 1999); mean diurnal variation (Falge
121 et al., 2001; Moffat et al., 2007; Zhao and Huang, 2015), multiple imputations (Hui et al., 2004; Moffat
122 et al., 2007), etc. Each of these methods has its pros and cons as follows: a) Interpolation methods such
123 as the Mean Diurnal Variation (MDV), do not need any drivers, yet, their accuracy is lower than other
124 approaches (Aubinet et al., 2012). Moreover, this method may provide biased results on extremely
125 clear or cloudy days (Falge et al., 2001). MDV is not recommended when a gap is longer than two
126 weeks, for it cannot consider the non-linear relations between the drivers and the flux, leading to a
127 high level of uncertainty (Falge et al., 2001). And b) The look-up table, especially its modified version,
128 Marginal Distribution Sampling (MDS), has provided performance close to ANNs, and are more
129 reliable and consistent than the other algorithms so far. Hence, MDS was chosen as one of the standard
130 gap-filling methods in EUROFLUX (Aubinet et al., 2012). Nevertheless, the performance of MDS in
131 gap-filling of extra long gaps is not well known (Kim et al., 2020). c) ANNs have commonly been used
132 to gap-fill EC fluxes since 2000 and because of their robust and consistent results are considered as a
133 standard gap-filling algorithm in several networks, e.g. ICOS, FLUXNET, OzFlux, etc. (Aubinet et al.,
134 2012; Beringer et al., 2017; Isaac et al., 2017). Despite their reliable performance, ANNs –and generally
135 all other ML algorithms- face some challenges. Over-fitting, for instance, is a big concern and can
136 happen when the number of degrees of freedom is high, while the training window is not long enough
137 respectively, or the quality of the training dataset is low. This challenge becomes acute when the gaps
138 happen while the ecosystem behaviour changes and shows different responses under similar
139 meteorological conditions. Furthermore, there is a desire to have the training windows short so that
140 the algorithm can track the ecosystem behaviour shift. Yet, this increases the risk of over-fitting
141 depending on the algorithm. In other words, the training window length should be neither too short
142 to cause over-fitting, nor too long to lead algorithms to ignore ecological condition changes. Besides,
143 long gaps are considered as one of the primary uncertainty sources of $CO_2$ flux in the FLUXNET
144 (Aubinet et al., 2012). As a result, studying the effects of the gap lengths, as well as the window length
145 whereby an algorithm is trained are both critical challenges associated with the environmental data
146 gap-filling.

147

148     Apart from the limitations and disadvantages of the mentioned algorithms, gap-filling of fluxes
149 (e.g. NEE) experiences some other challenges that make it necessary to find or develop new gap-filling
150 algorithms. That is because the current methods are not flexible enough to perform well in special
151 occasions or extreme values (Kunwor et al., 2017), and there is almost no room to optimise them to
152 improve their outcome (Moffat et al., 2007). Moreover, even using the best available algorithm, such
153 as ANNs, the model (gap-filling) uncertainty still accounts for a sizable proportion of the total
154 uncertainties, especially when the gaps are relatively long. Since the 2000s when MDS and ANNs were
155 chosen as the most reliable gap-filling methods for EC flux observations, many new ML and

optimisation algorithms have been developed and used in various scientific fields. Some of which have shown superiority over ANNs, either individually or as a part of a hybrid or ensemble model, e.g. (Gani et al., 2016). As a result, comparing the cutting-edge algorithms with the current standard ones can show whether there is any room to improve the gap-filling process within the field. According to the concerns mentioned above, this paper had two objectives. a) To find out the impact of different gap lengths on the performance of each algorithm. And b) to compare the performance of traditional with new gap-filling techniques, separately for fluxes and their meteorological drivers, particularly soil moisture, for this has always been a challenging variable to gap-fill due to biology and heterogeneity of soil parameters. To address these objectives, we utilised nine different algorithms (Extreme Gradient Boost (XGB), Random Forest Algorithm (RF), Artificial Neural Networks (ANNs), Marginal Distribution Sampling (MDS), Classic Linear Regression (CLR), Support Vector Regression (SVR), Elastic net regularisation (ELN), Panel Data (PD) and Prophet Forecast Model (FBP)) to fill the gaps of the major fluxes, and eight of them (excluding MDS) to fill the gaps of the environmental drivers. We then assessed their relative performance to evaluate potentially better ways to fill EC flux data. To test the approaches, we used five flux towers from the OzFlux network. To evaluate the performance of these algorithms, nine scenarios for gaps were planned – from a day to a whole year - and applied to the datasets, and different common performance metrics (e.g. RMSE, MBE, etc.), as well as visual graphs were used.

## 2.    Materials and methods

In order to address the first objective of this research, nine different gap lengths were superimposed to the datasets, i.e. 1, 5, 10, 20, 30, 60, 90, 180 and 365 days. To address the second objective, we chose nine different algorithms to fill the gaps, including a wide variety of different approaches, e.g. from a simple algorithm like CLR to the cutting-edge ML algorithms, such as XGB (MDS was not used to gap-fill the environmental drivers). The data used in this paper came from five EC towers of the OzFlux Network, i.e. Alice Springs Mulga, Calperum, Gingin, Howard Springs and Tumbarumba from 2012 to 2013, with a time resolution of 30 minutes, except for Tumbarumba (60 minutes). Additionally, data coming from three additional sources outside of the network were also used as ancillary data to help the algorithms fill environmental drivers' gaps.

### 2.1.  Data

The data used for this research came from the OzFlux, which is the regional Australian and New Zealand flux tower network that aims to provide a continental-scale national research facility to monitor and assess Australia's terrestrial biosphere and climate (Beringer et al., 2016a). As described in Isaac et al. (2017), all OzFlux towers continuously measure and record meteorological and flux variables at resolutions up to 10 Hz, and use a 30 min averaging period, with a few exceptions (data are available from (http://data.ozflux.org.au/portal). The network acquires additional data from the Australian Bureau of Meteorology (BoM), the European Centre for Medium-Range Weather Forecasting (ECMWF), and the Moderate Resolution Imaging Spectroradiometer (MODIS) on the TERRA and AQUA satellites for alternative data for gap-filling flux tower datasets (Isaac et al., 2017). As explained by Isaac et al. (2017), OzFlux uses the BoM automated weather station (AWS) datasets to gap-fill the meteorological data, the BoM weather forecasting model (ACCESS-R) for radiation and

soil data from 2011 onward, and MODIS MOD13Q1 for Normalised Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI). Moreover, the data provided by BIOS2, a physically-based model-data integration environment for tracking Australian carbon and water (Haverd et al., 2015), were also used as another ancillary source for varieties of environmental features. Current ACCESS-R and MODIS data are available from the BoM OPeNDAP (http://www.opendap.org/) server and TERN-AusCover data (http://www.auscover.org.au/), respectively.

The datasets used in this research came from five towers from the OzFlux Network between 2012 and 2013, each representative of a different climate and land cover of Australian ecological conditions; i.e. Alice Springs Mulga: Tropical and Subtropical Desert, Calperum: steppe, Gingin: Mediterranean, Howard Springs: Tropical Savanna, Tumbarumba: Oceanic (Table 1) (Beringer et al., 2016a). The datasets included 15 meteorological drivers as well as three major fluxes recorded (Table 2) based upon EC technique at a 30-minute temporal resolution, except for Tumbarumba, which was hourly. Additionally, relevant ancillary datasets for the mentioned towers were used to follow the OzFlux Network gap-filling protocol (Table 3). Each dataset was quality checked at three levels based on the OzFlux Network protocol described in (Isaac et al., 2017) and applied using PyFluxPro ver. 0.9.2. To address the underestimation of canopy respiration by EC measurements at night, we used the CPD method (Barr et al., 2013) to reject nightly records when the friction velocity fell below each site's threshold value. After dismissing the inappropriate measurements, overall coverage of 72-88 % and 21-48 % were achieved for diurnal and nocturnal records during 2013 (the year to which the artificial gaps were superimposed), respectively.

*Table 1. The information of the five towers that their data were used, including their name, location, dominant species and climate.*

| Site | Location | Species | Climate | Latitude, Longitude (degree) |
|---|---|---|---|---|
| Alice Springs Mulga [AU-ASM] | Pine Hill cattle station, near Alice Springs, Northern Territory | Semi-arid mulga (Acacia aneura) ecosystem | Tropical and Subtropical Desert Climate (Bwh) | -22.2828° N, 133.2493° E |
| Calperum [AU-Cpr] | Calperum Station, 25 km NW of Renmark, South Australia | Recovering Mallee woodland | Steppe Climate (Bsk) | -34.0027° N, 140.5877° E |
| Gingin [AU-Gin] | Swan Coastal Plain 70 km north of Perth, Western Australia | Coastal heath Banksia woodland | Mediterranean Climate (Csa) | -31.3764° N, 115.7139° E |
| Howard Springs [AU-How] | E of Darwin, NT | Tropical savanna (wet) | Tropical Savanna Climate (Aw) | -12.4943° N, 131.1523° E |
| Tumbarumba [AU-Tum] | Near Tumbarumba, NSW | Wet temperate sclerophyll eucalypt | Oceanic climate (Cfb) | -35.6566° N, 148.1517° E |

224    *Table 2. List of variables and their units used in this research, including the three main fluxes and their environmental drivers.*

| List of variables | Units |
| --- | --- |
| Drivers: | |
| Ah | Absolute Humidity (g m$^{-3}$) |
| Fa | Available energy (W m$^{-2}$) |
| Fg | Ground heat flux (W m$^{-2}$) |
| Fld | Downwelling long-wave radiation (W m$^{-2}$) |
| Flu | Upwelling long-wave radiation (W m$^{-2}$) |
| Fn | Net radiation (W m$^{-2}$) |
| Fsd | Downwelling short-wave radiation (W m$^{-2}$) |
| Fsu | Upwelling short-wave radiation (W m$^{-2}$) |
| ps | Surface pressure (kPa) |
| Sws | Soil water content (m m$^{-1}$) |
| Ta | Air temperature (C) |
| Ts | Soil temperature (C) |
| Ws | Wind speed (m s$^{-1}$) |
| Wd | Wind direction (deg) |
| Precip | Precipitation (mm) |
| q | Specific Humidity (kg kg$^{-1}$) |
| Fluxes: | |
| Fc (also NEE) | CO$_2$ flux (μmol m$^{-2}$ s$^{-1}$) |
| Fh (also H) | Sensible heat flux (W m$^{-2}$) |
| Fe (also LE) | Latent heat flux (W m$^{-2}$) |

225

226    The datasets whereby each environmental variable was gap-filled are shown in Table 3. For each of
227    these variables, the same variable of the ancillary source was used to fill the gaps. For instance, to gap-
228    fill Ah, the Ah records of AWS, ACCESS-R and BIOS2 were used. To gap-fill the missing values of
229    fluxes, i.e. Fc (NEE), Fh (H) and Fe (LE), eight drivers were used as follows: Ta, Ws, Sws, Fg, vapour
230    pressure deficit  (VPD) , Fn, q and Ts based on a combination of Random Forest (RF) feature selection
231    and testing out a series of feature combinations. Different Python Programming Language libraries
232    (ver. 3.6.4) were utilised for training and testing the algorithms, i.e. xgboost for XGB, fbprophet for
233    FBP statsmodels for PD and sklearn for the rest of algorithms.  Each algorithm was tuned individually
234    using grid search, and the number of nodes, layers, irritations, etc. were chosen accordingly.

235

236

237    *Table 3. The ancillary sources used to gap fill each environmental driver.*

| List of variables (y) | Ancillary Source |
| --- | --- |
| Drivers: | |
| Ah | AWS, ACCESS-R, BIOS2 |
| Fa | ACCESS-R, BIOS2 |
| Fg | ACCESS-R, BIOS2 |
| Fld | ACCESS-R, BIOS2 |
| Flu | ACCESS-R, BIOS2 |
| Fn | ACCESS-R, BIOS2 |
| Fsd | ACCESS-R, BIOS2 |
| Fsu | ACCESS-R, BIOS2 |
| ps | AWS, ACCESS-R |
| Sws | ACCESS-R, BIOS2 |

| | |
|---|---|
| Ta | AWS, ACCESS-R, BIOS2 |
| Ts | ACCESS-R, BIOS2 |
| Ws | AWS, ACCESS-R |
| Wd | AWS, ACCESS-R |
| Precip | AWS, ACCESS-R, BIOS2 |

## 2.2. Gap-filling algorithms

Eight imputation algorithms for estimating 15 environmental drivers and 9 algorithms for the 3 major fluxes were chosen to make the comparison. These algorithms were selected in such a way that a variety of approaches were tested, from the standard methods like ANNs and MDS, to the newer algorithms, which have rarely or never been used in the field, such as Extreme Gradient Boosting and panel data (Table 4).

Table 4. The name and the abbreviation of the gap-filling algorithms.

| Algorithm abbreviation | Full name |
|---|---|
| XGB | Extreme Gradient Boost |
| RF | Random Forest Algorithm |
| ANNs | Artificial Neural Networks |
| MDS | Marginal Distribution Sampling |
| SVR | Support Vector Regression |
| CLR | Classical Linear Regression |
| PD | Panel data |
| ELN | Elastic net regularisation |
| FBP | The Prophet Forecasting Model (Facebook Prophet) |

**Marginal Distribution Sampling (MDS)**

Reichstein Reichstein et al. (2005) introduced the MDS is an enhanced look-up table method, which considers both the covariation of fluxes with meteorological variables and the temporal auto-correlation of the fluxes (Aubinet et al., 2012). Alongside the ANNs, the MDS is considered one of the standard gap-filling methods for flux data amongst the FLUXNET, and is selected in this study to help the community have a clear idea of the performance of other algorithms. Unlike the other algorithms used in this research, we used Fsd, Ta and VPD as the input features for the MDS to be consistent with standard application of the MDS, and for using more than three or four drivers is not generally recommended (Aubinet et al., 2012). The PyFluxPro ver. 0.9.2 was used to apply the algorithm (modified code used for the gaps longer than 10 days).

**Artificial Neural Networks (ANNs)**

Rooted in the 1950s, artificial neural networks are ML methods inspired by biological neural networks and are classified as supervised learning methods (Dreyfus, 1990; Farley and Clark, 1954). ANNs work based on several connected units called nodes, which are used to mimic a neuron's functionality in an animal brain by sending and receiving signals to other nodes. The ANNs technique used in this paper was the Multi-layer Perceptron regressor, which optimises the squared-loss using stochastic gradient descent. Sklearn.neural_network.MLPRegressor was used to apply this method

in Python, and its hyperparameters were 800 and 500 for "hidden_layer_sizes" and "max_iter", respectively based on grid search. ANNs are one of the current standard approaches for gap-filling in FLUXNET and in this research were picked out as a performance reference for other algorithms.

**Classical Linear Regression (CLR)**

A classical linear regression is an equation developed to estimate the value of the dependent variable (y) based on independent values ($x_i$). In contrast, each $x_i$ has its specific coefficient and an overall intercept value. In this method, these coefficients are determined by minimising the squared residuals (errors) of estimated vs observed values, called least squares. A CLR algorithm can be formulated as follows (Freedman, 2009):

$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_i X_i + \varepsilon \qquad (1)$$

where y is the dependent variable, $\alpha$ is the interception, $X_i$s are independent variables, and $\beta_i$ is coefficient of $X_i$, and $\varepsilon$ is the error term. We chose this algorithm as a baseline to find out how better more complicated algorithms can estimate dependent variables comparatively.

**Random Forests (RF)**

Random forest, a supervised ML algorithm, used for both classification and regression, consists of multiple trees constructed systematically by pseudorandomly selecting subsets of components of the feature vector, that is, trees constructed in randomly chosen subspaces (Ho, 1998). The RF algorithm has been developed to overcome the over-fitting problem, a commonplace limitation of its preceding decision tree-based methods (Ho, 1995, 1998). Sklearn.ensemble.RandomForestRegressor was used to apply this method in Python, and the hyperparameters used were 5 and 1000 for "max_depth" and "n_estimators", respectively based on grid search.

**Support Vector Regression (SVR)**

As a non-linear method, support vector regression was developed based on Vanpik's concept of support vectors theory (Drucker et al., 1997). An SVR algorithm is trained by trying to solve the following problem:

minimise $\frac{1}{2}$ $\|w\|^2$

subject to $\begin{pmatrix} y_i - \langle w, x_i \rangle - b \leq \varepsilon, \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon, \end{pmatrix}$

where $x_i$ and $y_i$ are training sample and target value in a row. The inner product plus intercept $\langle w, x_i \rangle + b$ is the prediction for that sample, and $\varepsilon$ is a free parameter that serves as a threshold.

299 sklearn.svm.SVR was used to apply this method in Python, and the hyperparameters that used were
300 1 and 0.001 for "C" and "gamma", respectively based on grid search.

**Elastic net regularisation (ELN)**

302 The elastic net is a linear regularised regression method that exerts small amounts of bias by
303 adding two penalty components to the regressed line to decline the coefficients of independent
304 variables and thus, provides better long-term predictions. Given that these two penalty components
305 come from ridge regression and LASSO, the elastic net is considered as a hybrid model consists of
306 ridge and LASSO regressions, overcoming the limitations of both. The estimates from the ELN method
307 can be formulated as below (Zou and Hastie, 2005):

$$\hat{\beta}(elastic\ net) = \frac{\left(|\hat{\beta}(OLS)| - {\lambda_1}/{2}\right)}{1 + \lambda_2} sgn\{\hat{\beta}(OLS)\} \tag{2}$$

308

309 where $\hat{\beta}$ is the coefficient of each ELN independent variable, $\lambda_1$ and $\lambda_2$ are penalty coefficients of
310 LASSO and ridge regression respectively, $\hat{\beta}(OLS)$ is the coefficient of an independent variable
311 calculated based on ordinary least squares, and *sgn* stands for the sign function:

$$sgn(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases} \tag{3}$$

312

313 The ELN regression is good at addressing situations when the training datasets have small samples
314 or when there are correlations between parameters. sklearn.linear_model.ElasticNet was used to
315 apply this method in Python, and the hyperparameters used were as follows: {'alpha': 0.01,
316 'fit_intercept': True, 'max_iter': 5000, 'normalize': False} based on grid search.

317

**Panel data (PD)**

319 Panel data is a multidimensional statistical method, mainly used in econometrics to analyse
320 datasets, which involve time series of observations amongst individual cross-sections (Baltagi, 1995)
321 usually based on ordinary least squares (OLS) or generalised least squares (GLS). A two-way panel
322 data model consists of two extra components beyond a CLR as follows (Baltagi, 1995; Hsiao et al.,
323 2002; Wooldridge, 2008):

$$y_{it} = \alpha + \beta X_{it} + u_{it} \qquad i = 1, 2, ..., N\ ;\ \ t = 1, 2, ..., T \tag{4}$$

$$y_{it} = \alpha + \beta X_{it} + \mu_i + \lambda_t \tag{5}$$

324 where i and t denote the cross-section and time series dimension in a row, y is a dependent-variable
325 vector, X is an independent variable matrix, $\alpha$ is a scalar, $\beta$ is the coefficient of the independent-
326 variable matrix, $\mu_i$ is the unobservable individual-specific effect, and $\lambda_t$ is the unobservable time-

specific effect. Panel data abilities to provide a holistic analysis of different individuals, as well as determining the specific impact of every single time caused its superiority over CLR.   Since PD requires cross-sections to be applied, we used a cross-section tower for each of the main five tower as follows: Ti Tree East for Alice Springs Mulga, Whroo for Calperum, Great Western Woodlands for Gingin, Daly River for Howard Springs, and Cumberland Plain for Tumbarumba. The cross-section towers were chosen based on their distances (the closest ones with common years of data).

**Extreme Gradient Boost (XGB)**

Extreme gradient boost is a reinforced method of Gradient Boost introduced in 1999 that works based on parallel boosted decision trees and similar to RF can be used for a variety of data processing purposes including classification and regression (Friedman, 2002; Jerome H. Friedman, 2001; Ye et al., 2009). XGB method is resistive to over-fitting and provides a robust, portable and scalable algorithm for large-scale boosting decision-trees-based techniques. sklearn.ensemble.GradientBoostingRegressor was used to apply this method in Python, and its hyperparameters were chosen based on grid search as follows: {'learning_rate': 0.001, 'max_depth': 8, 'reg_alpha': 0.1, 'subsample': 0.5}.

**The Prophet Forecasting Model (FBP)**

The Prophet Forecasting Model, also known as "prophet", is a time series forecasting model developed by Facebook to manage the common features of business time series and designed to have intuitive parameters that can be adjusted without knowing the details of underlying model (Taylor and Letham, 2017). A decomposable time series model was used (Harvey and Peters, 1990) to develop this model, with three main components: trend, seasonality, and holidays as the equation below (Taylor and Letham, 2018):

$$y(t) = g(t) + s(t) + h(t) \tag{6}$$

where g(t) is the trend function, which models non-periodic changes, s(t) is a function to represent periodic changes, e.g. seasonality, and h(t) assesses the effects of potential anomalies which occur over one or more days, e.g. holidays.

## 2.3.  *The gap scenarios*

In order to find out the effect of gap size on the performance of our gap-filling algorithms, the data was removed randomly from nine different gap windows (i.e. 1, 5, 10, 20, 30, 60, 90, 180 and 365 consecutive days) during 2013. Afterwards, the data from 2012 to 2013 were used to train the algorithms (excluding the superimposed gaps). Finally, the trained algorithms were used to fill the artificial gaps superimposed to the datasets. The entire process permutated five times in each scenario to ensure the performance was not sensitive to the gap position (i.e seasonally). As such, 15 variables,

362 9 window lengths, 8 gap-filling methods (MDS excluded), and 5 permutations across 5 towers resulted
363 in 27,000 computations for the meteorological features. Similarly, 3 fluxes, 9 window lengths, 9 gap-
364 filling methods, and 5 permutations across 5 towers resulted in 6,075 computations for the major
365 fluxes, overall.

## 2.4. Statistical performance measures

367 Different statistical metrics were used to evaluate algorithms' performance and enable
368 comparison between measured values from the flux towers with each gap-filling algorithm prediction.
369 These metrics included the coefficient of determination (R-squared) to measure the square of the
370 coefficient of multiple correlations (Devore, 1991), the variance of measured and modelled values ($S^2$)
371 to indicate how well algorithms could follow the variations of the recorded data, the root mean square
372 error (RMSE), the mean bias error (MBE) to capture distribution and bias of residuals, variance ratio
373 (VR) to compare the variance of estimated values with those of measured, and the Index of Agreement
374 (IoAd) to compare the sum of the squared error to the potential error (Bennett et al., 2013).
375 Abbreviations and formulas of these metrics are illustrated as follows (Bennett et al., 2013):

$$R^2 = \frac{[\sum(p_i - \bar{p})(o_i - \bar{o})]^2}{\sum(p_i - \bar{p})^2 \sum(o_i - \bar{o})^2} \tag{7}$$

376

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{N - 1} \tag{8}$$

377

$$RMSE = \sqrt{\frac{\sum(p_i - o_i)^2}{N - 1}} \tag{9}$$

378

379

$$MBE = \frac{\sum o_i - p_i}{N - 1} \tag{10}$$

380

$$VR = \frac{\sigma_p^2}{\sigma_o^2} \tag{11}$$

381

$$IoAd = 1 - \frac{\sum_{i=1}^{n}(o_i - p_i)^2}{\sum_{i=1}^{n}(|p_i - \bar{o}| + |o_i - \bar{o}|)^2} \tag{12}$$

382

383 where $o_i$ and $p_i$ are individual measured and predicted values respectively, $\bar{o}$ and $\bar{p}$ are the means of
384 o and p, and $\sigma^2$ is the variance. $S^2$ is calculated separately for the observed and predicted values with
385 the respective values defined as x representing every observed or predicted value. All of these metrics
386 were calculated for each of the gap scenarios, and then the results of five permutations were

387 concatenated. Afterwards, the metrics were calculated to avoid Simpson's paradox or any relevant
388 averaging issue as described by Kock and Gaskins (2016).

## 3. Results

390

### 3.1. Fluxes

#### 3.1.1 CO$_2$ flux (Fc)

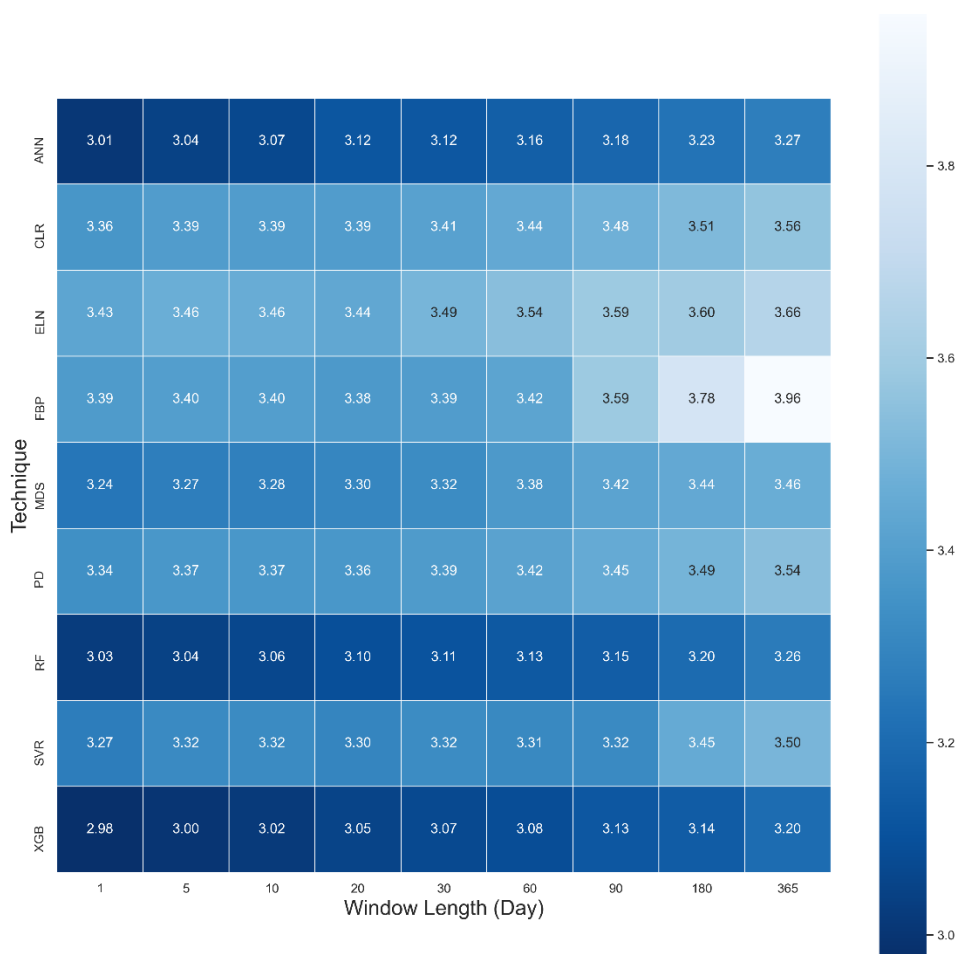
393    Even though factors such as ground heat flux (Fg) and net radiation (Fn) are fluxes, we dealt
394 with them as environmental drivers since they drive the three major turbulent fluxes. The metrics
395 used to evaluate the algorithms' performance (RMSE, R$^2$, MBE, IoAd and VR) (Table 5) illustrated that
396 overall, the performance of these algorithms, particularly the ML ones, was similar, closely followed
397 by the MDS. The XGB provided the lowest values of RMSE and one of the highest R$^2$, while the FBP
398 and ELN had the lowest and highest values of R$^2$ and RMSE, respectively. The algorithms, however,
399 showed different levels of sensitivity to the gap lengths, e.g. the CLR and PD showed smaller
400 sensitivity, while the FBP showed the most sensitivity (Figure 1).

401 *Table 5. The average amounts of performance metrics for each gap-filling algorithm regarding Fc, which includes all window*
402 *lengths and sites, ranked by RMSE using the Tukey's HSD test at the level of 5 per cent.*

| Algorithm | Mean RMSE | Mean R$^2$ | Mean MBE | Mean IoAd | Mean VR |
|---|---|---|---|---|---|
| XGB | 3.07 [a] | 0.59 | -0.43 | 0.90 | 0.66 |
| RF | 3.12 [a] | 0.58 | -0.37 | 0.91 | 0.71 |
| ANNs | 3.13 [a] | 0.56 | -0.33 | 0.90 | 0.69 |
| SVR | 3.34 [b] | 0.47 | -0.32 | 0.86 | 0.75 |
| MDS | 3.35 [b] | 0.51 | -0.41 | 0.85 | 0.70 |
| PD | 3.41 [b,c] | 0.48 | -0.35 | 0.81 | 0.54 |
| CLR | 3.44 [b,c] | 0.49 | -0.36 | 0.81 | 0.55 |
| ELN | 4.52 [c] | 0.43 | -0.37 | 0.73 | 0.39 |
| FBP | 4.15 [d] | 0.47 | -0.06 | 0.77 | 0.68 |

403

404 These outcomes were expected for the XGB as it uses a more regularised model formalisation to
405 control over-fitting (Chen and Guestrin, 2016) which, on paper, leads to better performance as against
406 its ML rivals. The relatively poor performance of FBP was also foreseen for unlike other algorithms,
407 FBP did not use any feature to estimate flux values, other than the previous time series of flux values.
408 However, the weaker performance of the ELN compared to CLR was unforeseen as by adding two
409 penalty components to the regression line, the ELN is supposed to improve the long-term prediction
410 compared to the traditional linear regression methods. Tukey's HSD (honestly significant difference)
411 test at the level of five per cent was applied to the results to determine whether the difference amongst
412 the algorithms was significant (Table 5). Where the null hypothesis was there is no significant
413 difference between the mean values of the RMSE. According to the results, there were significant
414 differences between certain algorithms, and the XGB, RF and ANNs were different from the rest,
415 showing that these three performed considerably better. Tukey's HSD test, however, did not reject the
416 second error probability between RF, XGB and ANNs meaning that the three algorithms were not
417 significantly different from each other. This result agrees with the results of Falge et al. (2001) and

418  Moffat et al. (2007) in the sense that ANNs are one of the best available gap-filling algorithms, and
419  there is no significant difference amongst the appropriate algorithms. However, the test showed that
420  the performance of the MDS was significantly different from the ANNs. It seems that the difference
421  has occurred because of the longer gaps (> 10 days) that had been absent from the previous studies.
422  Finally, it is worth mentioning that Tukey's HSD is well known as a conservative test. That being said,
423  despite no meaningful difference based on Tukey's HSD, XGB and RF might have performed better
424  than ANNs, as the superiority of RF in gap-filling of methane flux over the ANNs, SVR, and MDS has
425  recently been claimed by Kim et al. (2020).

426

427  *Figure 1. A heat map of mean RMSE values of Fc across all sites based on 9 algorithms and 9 window lengths in 2013.*

428

To address this paper's first objective, which was to find out the sensitivity of the gap-filling algorithms to the gap window length, we used the averaged RMSE, R2 and MBE for each gap size using the output of all algorithms for all sites (Table 6). The outcome illustrates that the longer the window length got, the larger the RMSE became. Yet, no such pattern was recognisable for the $R^2$ and MBE. As a result, generally, any consecutive gaps longer than 30 days seem to decline the algorithms' performance noticeably. A reason for this may be that longer windows do not let the algorithms accommodate seasonal changes and, therefore, different canopy physiological behaviour.

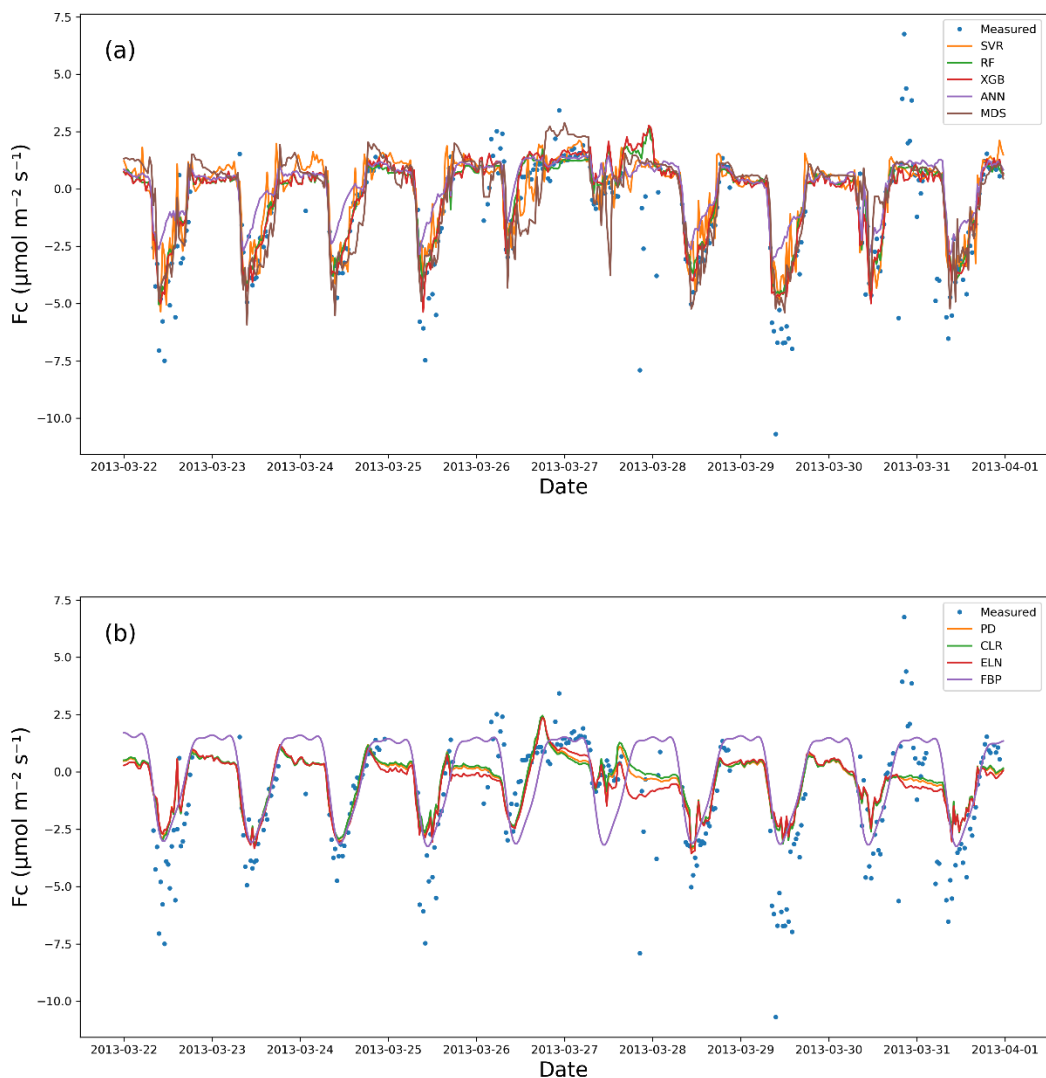
*Table 6. The average RMSE, $R^2$, and MBE for Fc gap-filling based on the window length including the outcome of all sites; the differences of RMSE values were tested using the Tukey's HSD test at the level of 5 per cent.*

| Window length | Mean RMSE | Mean $R^2$ | Mean MBE |
|---|---|---|---|
| 1-day | 3.23 [a] | 0.53 | -0.27 |
| 5-days | 3.25 [a] | 0.52 | -0.31 |
| 10-days | 3.26 [a] | 0.51 | -0.29 |
| 20-days | 3.27 [a] | 0.51 | -0.31 |
| 30-days | 3.29 [a] | 0.51 | -0.31 |
| 60-days | 3.32 [a] | 0.49 | -0.35 |
| 90-days | 3.37 [a] | 0.51 | -0.38 |
| 180-days | 3.43 [a] | 0.50 | -0.41 |
| 365-days | 3.49 [a] | 0.49 | -0.37 |

438

According to the MBE values (Table 5), mainly, all algorithms had negative MBE indicating an overestimation of the Fc values. This bias varied from tower to tower and depended on the window lengths. For instance, the MBE's absolute values were larger in Gingin and Tumbarumba, while considerably smaller (closer to zero) at Alice Springs Mulga and Calperum (Supplementary). The lower leaf area index of the two later sites, and thus their smaller amounts of photosynthesis are likely to be the reason for this. FBP, nonetheless, provided substantially lower mean bias (-0.06) compared to the other algorithms, which varied between -0.32 and -0.43.

Observations from the EC technique often include extremely low or high values after QC, especially at night, when some of the theoretical assumptions might be violated. One of the practical challenges associated with the EC technique is that it is often difficult to distinguish between the good data and the noise (Aubinet et al., 2012; Burba and Anderson, 2010). This problem seems to affect the outcomes of the gap-filling algorithms in this research, as none of them performed ideally in capturing the observed variance (Table 5**Error! Reference source not found.**). Even though RMSE, $R^2$ and IoAd showed the superiority of the XGB, RF and ANNs, the variance ratio between the estimated and measured values revealed different information (Table 5), which is recognisable in Figure 2. The variance ratios (VR) showed that SVR captured the extreme values of Fc better than the other algorithms, 0.75 on average. The other ML algorithms –plus the MDS- though, performed closely with

456    regard to capturing the extremes that matches both the expectations, and the performance metrics
457    (Table 5).

458

459    *Figure 2. Measured vs estimated values of Fc for Calperum based on a 10-day gap window (March 22 - March 31, 2013): (a) the*
460    *ML algorithm plus the MDS, and (b) the linear models plus FBP.*

461    The linear algorithms, CLR, PD, and ELN, performed worse concerning the VR compared to the ML
462    algorithms with the VR of Fc for Calperum (Figure 2**Error! Reference source not found.**) confirming
463    this. Based on the figure, as expected, the ELN performed the worst in capturing the fluctuations in
464    Fc (VR = 0.39), while the performance of the other algorithms, apart from the top five, was not
465    significantly better the exception of FBP. It is noteworthy that CLR, PD, and ELN frequently predicted
466    nocturnal photosynthesis. Overall, the results showed a significant difference between the top five
467    algorithms (XGB, RF, ANNs, SVR, and MDS) and remaining algorithms, particularly in capturing the

468 fluctuations and the min-max range of Fc. However, a comprehensive comparison shows a slight
469 superiority of the XGB and RF.

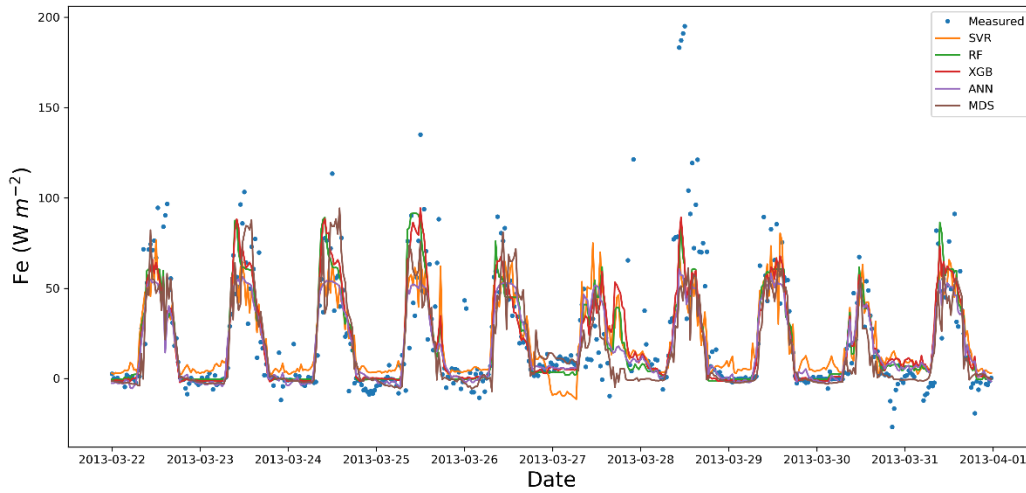
### 3.1.2 Latent heat flux (Fe)

471 The performance of algorithms for Fe was similar to that for Fc with respect to RMSE, MBE
472 and $R^2$, as shown in Table 7. This similarity was not surprising since these processes are partially
473 coupled via stomatal conductance (Scanlon and Kustas, 2010; Scanlon and Sahu, 2008). Again, the top
474 three ML algorithms performed better, with XGB and RF being statistically significant   as shown by
475 the Tukey's HSD (Table 7). The null hypothesis was not rejected while comparing FBP and SVR,
476 whereas the better performance of the other algorithms was confirmed.  As a result, the FBP and SVR
477 provided the most unsatisfactory results in estimating Fe, according to the average values of the
478 RMSE. No significant improvement in RMSE occurred when the gap lengths became shorter than 60
479 days, meaning that the algorithms' performance did not vary considerably from a 30-day to a one-day
480 window, especially for the top algorithms (XGB, RF, and ANNs). CLR and PD results were very
481 similar to those for Fc, showed lower RMSE and higher $R^2$ values as against ELN, but the ELN led to
482 a slightly lower MBE. The MBE values also showed moderately high values for the SVR, meaning that
483 there was an absolute bias in its outcome, which might be related to overfitting. The source of the bias
484 shown by the SVR algorithm (Figure 3), was because it could not capture the minimum values
485 appropriately, resulting in a considerable overestimation. A common issue in estimating Fe values,
486 which had affected all algorithms other than the FBP, was the inability to capture the negative  values.
487 In contrast to Fc results, the ANNs did not perform as well as the XGB and RF, which could be due to
488 not capturing the maximum values compared to its rivals.

489 *Table 7. The average metrics for Fe gap-filling based on the algorithms, ranked by RMSE using the Tukey's HSD test at the level*
490 *of 5 per cent.*

| Algorithm (Fe) | Mean RMSE | Mean R² | Mean MBE |
|---|---|---|---|
| XGB | 34.95 [a] | 0.74 | -3.48 |
| RF | 35.63 [a] | 0.74 | -3.33 |
| ANNs | 37.77 [a,b] | 0.67 | -3.94 |
| MDS | 41.74 [b,c] | 0.64 | -3.27 |
| PD | 43.28 [b,c] | 0.64 | -6.35 |
| CLR | 43.51 [c] | 0.64 | -6.66 |
| Eln | 44.34 [c] | 0.59 | -5.13 |
| SVR | 46.63 [c,d] | 0.59 | -20.45 |
| FBP | 50.53 [d] | 0.52 | 3.01 |

491

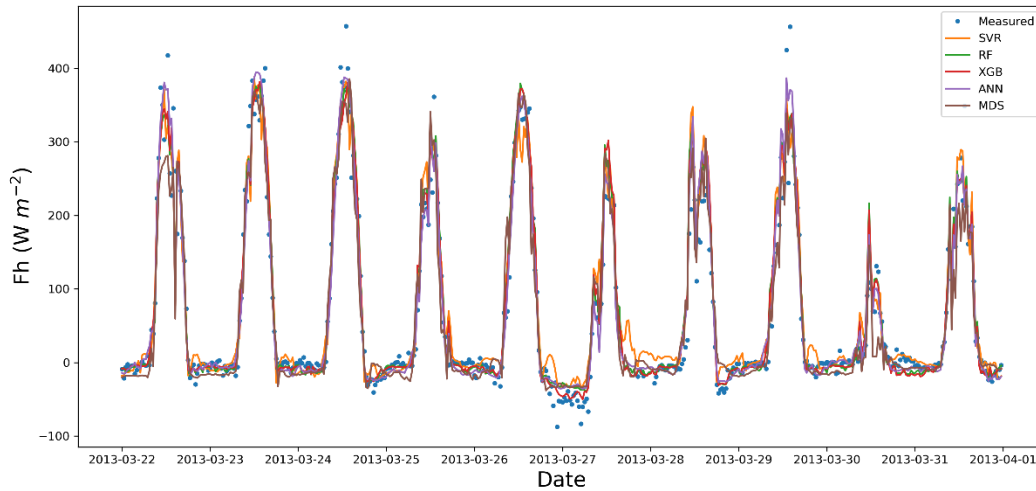
492

493

494 *Figure 3. Measured vs estimated values of Fe for Calperum based on a 10-day gap window (March 22 - March 31 2013).*

495     *3.1.3  Sensible heat flux (Fh)*

496    As with the other flux results, the metrics of RMSE, $R^2$ and MBE showed slight superiority for
497 XGB and RF, as well as the inferiority of the SVR and FBP over the other algorithms (Table 8).
498 Likewise, the SVR provided relatively large negative values of MBE, showing considerable
499 overestimation. The Tukey's HSD test of the average RMSE values confirmed that the performance of
500 the FBP was significantly different from the rest at the level of 5 per cent, making FBP the weakest
501 performer for Fh. On the other hand, although there was no significant difference amongst the XGB,
502 RF and ANNs, the first two were considerably superior over the other algorithms as regards the
503 Tukey's HSD test. Similarly to Fe, estimated values of Fh using SVR had a negative bias (Figure 4)
504 because it was not able to provide appropriate estimations of Fh minimum values. In contrast, the
505 ANNs performed the best in capturing the minimum values, while the other top algorithms
506 performed almost equally well. Despite the close performance in capturing the minimum values,
507 ANNs and MDS did not perform as well as XGB and RF in capturing the overall values, resulting in
508 an higher RMSE. Finally, like the other fluxes, the PD performed slightly better than the CLR and
509 ELN.

510 *Table 8. The average metrics for Fh gap-filling based on the algorithms, ranked by RMSE using the Tukey's HSD test at the level*
511 *of 5 per cent.*

| Algorithm (Fh) | Mean RMSE | Mean $R^2$ | Mean MBE |
|:---:|:---:|:---:|:---:|
| XGB | 37.23 [a] | 0.92 | -0.21 |
| RF | 37.55 [a] | 0.91 | -0.09 |
| ANNs | 40.13 [a,b] | 0.90 | -0.08 |
| MDS | 43.30 [b,c] | 0.88 | -9.51 |
| SVR | 43.80 [b,c] | 0.88 | 0.35 |
| PD | 44.96 [c] | 0.88 | 1.36 |
| CLR | 45.03 [c] | 0.88 | 1.64 |
| Eln | 45.19 [c] | 0.87 | 2.16 |
| FBP | 72.91 [d] | 0.73 | 1.07 |

512

513  *Figure 4. Measured vs estimated values of Fh for Calperum based on a 10-day gap window (March 22 - March 31 2013).*

514

### 3.2.    Meteorological and Environmental Drivers

516   Since meteorological and environmental drivers are needed to fill the gaps of the three
517   turbulent fluxes (Fc, Fe and Fh), the eight algorithms (excluding the MDS) were used to fill these
518   drivers' gaps. The metrics of $R^2$, RMSE, and MBE were calculated for all five towers and nine window
519   lengths (16 meteorological and environmental drivers). Overall, for most meteorological drivers, the
520   linear algorithms, especially the CLR and PD, performed slightly better than the ML algorithms such
521   as the XGB, RF, ANNs and SVR, except for Ah, Fg and Fn. This unexpected superiority can be
522   explained based on the two following reasons. Firstly, unlike the fluxes, the input and output features
523   were the same here, e.g. Ta for Ta, which led to solid correlations (e.g. up to 0.99 for atmospheric
524   pressure - ps) as well as strong linear relationships between the independent and dependent features.
525   These strong correlations helped the linear algorithms perform well and reduced ML algorithms'
526   ability to capture non-linear behaviour of complicated problems. Second, ML algorithms' slight
527   inferiority could be due to data noise where simple linear algorithms such as the CLR are usually
528   relatively less sensitive to the noise. Therefore, over-fitting is not an issue for them when the number
529   of observations is big enough (i.e. at least 10 to 20 observations per parameter (Harrell, 2014)). The
530   exceptions were Ah, Fn and Fg, for which values were estimated more accurately by the XGB, ANNs
531   and RF, especially Fg  where the RMSE of RF and CLR for Fg was 28.91 versus 33.92  respectively).
532   Tukey's HSD test for the mean RMSE values of Fg confirmed that the XGB, ANNs and RF significantly
533   better results , while, like all other fluxes and drivers, the FBP was the worst algorithm (Table 9). Yet,
534   according to the same test for the other drivers, there was no significant difference between the
535   algorithms, other than the FBP, which provided the most significant mean values of the RMSE (results
536   not shown). Importantly, though, none of the algorithms offered adequate estimations for soil
537   moisture (Sws), particularly in drier regions. This weak performance happened because Sws changes
538   dramatically during rainfall in a pulsed manner often from zero to saturation in short space of time,

539  whereas, the algorithms had been trained based on the datasets mostly reflecting non-rainy periods.
540  These datasets, consequently, could not fit the algorithms in a way that they could estimate Sws
541  accurately when precipitation occurs and the soil moisture increases dramatically. For instance, in a
542  wet region like Tumbarumba, where the soil faces rainy days frequently, the time series are much less
543  spikey. Thus, the overall performance was better in these regions than the drier ones (e.g. $R^2$ of 0.45
544  and 0.26 on average for Tumbarumba and Calperum, respectively). In addition, the dataset used to
545  gap-fill the soil moisture was a model derivation from gridded data or regional reanalysis and
546  therefore, may not close to reality. Another challenge of estimating soil moisture comes from the low
547  spatial coherence of soil moisture is that it can be extremely different just a couple of hundred metres
548  away, due to storms, topography, soil structure heterogeneity, etc. (Reichle et al., 2004; Sahoo et al.,
549  2008).

550

551  *Table 9. The average amounts of RMSE for Fg gap-filling based on the algorithms, using the Tukey's HSD test at the level of 5*
552  *per cent.*

| Algorithm (Fg) | Mean RMSE |
|---|---|
| RF [a] | 28.91 |
| XGB [a, b] | 29.19 |
| ANNs [b, c] | 29.58 |
| SVR [c] | 31.46 |
| CLR [d] | 33.92 |
| PD [d] | 33.93 |
| ELN [d] | 34.09 |
| FBP [e] | 39.10 |

553

## 4.  Discussion

555

556

557  Nine gap-filling algorithms were used in this study: Extreme Gradient Boost as XGB, Random
558  Forest Algorithm as RF, Artificial Neural Networks as ANNs, Marginal Distribution Sampling as
559  MDS, Support Vector Regression as SVR, Classical Linear Regression as CLR, panel data as PD, Elastic
560  net regularisation as ELN, and The Prophet Forecasting Model as FBP. All algorithms performed
561  similarly in estimating the meteorological and environmental drivers (turbulent fluxes included)
562  across all stations, except the FBP, which performed poorly for it did not use any ancillary data. The
563  best results were achieved for the 30-day gaps and shorter, while the worst results obtained for the
564  most extended windows, 180 and 365 days. Although most of the algorithms performed almost
565  equally well in estimating meteorological and environmental drivers, the linear algorithms (CLR, ELN
566  and PD) performed slightly better, though not significantly using Tukey's HSD test. The only apparent
567  exception was Fg, for which the RF provided more accurate and robust estimations. The ML
568  algorithms and MDS, on the other hand, showed their superiority over the linear algorithms while
569  estimating the main fluxes, Fc, Fe and Fh. For Fc, the XGB, RF and ANNs performed significantly

570 better than the FBP and all linear algorithms ( i.e. the CLR, PD and ELN, yet, followed closely by the
571 SVR and MDS). The superiority of the ML algorithms and their intimate performance agreed with the
572 results of previous researchers (Falge et al., 2001; Moffat et al., 2007), who showed the superiority of
573 non-linear algorithms and no significant difference amongst the top algorithms in estimating Fc.
574 Besides, the slight superiorities of XGB and RF over ANNs, our results confirm that RF performs better
575 for EC flux gap-filling, as noted by Kim et al. (2020) for methane.

576     The XGB was the most novel ML algorithm used in this research and based on the most
577 performance metrics provided comparatively robust results in estimating the fluxes. In estimating the
578 meteorological drivers though, the XGB did not show any superiority over the other algorithms,
579 especially the linear ones. Moreover, the XGB needed four to six times longer time to be trained and
580 tuned, making it a less feasible algorithm when time or the processing power are important factors or
581 several years of data are needed to be gap-filled. Hence, we do not recommend the XGB as an
582 alternative to the current standard algorithms. Nevertheless, because of its local superiorities, this
583 algorithm might be suitable to use in an ensemble model alongside the algorithms with different
584 weaknesses.

585     The RF was the best all-around algorithm amongst the nine algorithms used in this study,
586 providing the best consistent and robust estimates of the fluxes (similar to XGB) but also being less
587 complicated and performing faster than the XGB. The RF also provided the best results for Fg, where
588 the linear algorithms did not perform well. This superiority of RF over ANNs, MDS, and SVR has
589 been shown previously by Kim et al. (2020) for gap-filling of methane, showing that it is worth testing
590 the RF for other towers, and fluxes across the FLUXNET.

591     The ANNs estimated the fluxes better than the linear algorithms, notably for Fc, yet not as
592 robust as the XGB and RF in general. For Fc and Fh, the ANNs provided bias, mainly due to
593 overestimating minimum values when the window lengths were longer than 30 days. However, since
594 the superiority of the XGB and RF was not considerable, it is difficult at this point to suggest using
595 XGB or RF as better alternatives. That is because the utility of ANNs have been validated for a long
596 time in different locations and considered as one of the most reliable algorithms in the field for more
597 than a decade (Aubinet et al., 2012; Hagen et al., 2006; Kunwor et al., 2017; Moffat et al., 2007). In other
598 words, the superiority of RF, should be assessed in several future studies to convince the network to
599 suggest RF instead of ANNs, or identify it as another standard gap-filling method.  Furthermore, there
600 are a wide variety of different ANNs algorithms used in the field (Beringer et al., 2016b; Hagen et al.,
601 2006; Isaac et al., 2017; Kunwor et al., 2017; Moffat et al., 2007), and the minor superiority of RF and
602 XGB cannot be generalised without additional case studies. As such, we suggest other researchers to
603 use the RF, especially for Fh and Fc alongside the ANNs to find out which one performs better in the
604 challenging scenarios (e.g. when the gaps are long). Another option is to develop ensemble models to
605 improve the results over a single algorithm (Moffat et al., 2007). Ideally, a model with a higher level
606 of flexibility is required in the field (Hagen et al., 2006; Kunwor et al., 2017; Richardson and Hollinger,
607 2007). Finally, the ANNs, like the other ML algorithms, did not show a consistent superiority over the
608 linear algorithms regarding the environmental drivers. Therefore, we do not recommend using ML
609 algorithms in such scenarios, except for Fg, for which RF seems to be a better option.

610    The MDS performed close to, yet not as well as the XGB, RF, and ANNS in gap-filling the fluxes.
611    Its performance was close to the SVR, but was more reliable for Fe and Fh. It is worth mentioning that
612    this performance was achieved despite the MDS using fewer input features. Its performance, however,
613    was comparable with the ML algorithms, particularly when the gap lengths were relatively shorter
614    (equal to or smaller than 10 days). As such, we recommend using the MDS when the gaps are not long
615    or the available input features are limited, especially considering that the MDS performs significantly
616    faster than the ML algorithms, and is easier to use.

617    The SVR showed consistent inferiority over the other ML algorithms and did not fulfill our
618    expectations, neither for the meteorological drivers nor for the major fluxes. The only strength of the
619    SVR  was that it captured the extreme values better than any other algorithm. However, because of
620    the larger RMSE the mentioned advantage seems to be achieved suspiciously and might have
621    occurred due to over-fitting. This dubious performance shows the SVR is perhaps more vulnerable to
622    the over-fitting issues regarding these data types. Hence, we suggest the SVR not to be used in
623    environmental modelling related to the reviewed drivers and fluxes, whatsoever.

624    The CLR, the simplest algorithm used in this research, provided a comparatively acceptable
625    performance in estimating the meteorological drivers, except for Fg. This algorithm, however, did not
626    perform well in assessing the fluxes, especially Fc, mainly because of its inability to capture the
627    extreme values caused by the non-linear nature of Fc to its drivers. Overall, considering the CLR
628    simplicity, resource-saving and robust performance for drivers, this algorithm seems to be the most
629    suitable way to fill the gaps of meteorological parameters in similar scenarios, where the same
630    ancillary dataset are available.

631    The PD performed slightly better than the CLR, yet it did not show a significant superiority over
632    the other linear algorithms used in the research. This unforeseen weak performance can be explained
633    due to a couple of reasons. First, one of the assumptions of using the PD is that the cross-sections'
634    behaviour (here towers) is similarly under the similar conditions (the independent variables), and the
635    only thing leads to the difference is the specific characteristics of each individual cross-section.
636    Contrariwise, it seems that the five towers selected in this research violated this assumption due to
637    them being in widely different ecosystems. Based on the previous studies in which the PD performed
638    well (Izady et al., 2013, 2016; Mahabbati et al., 2017), it appears that a decent level of homogeneity is
639    vital for the PD to perform satisfactorily. As in all previous cases, the cross-sections ecosystem had
640    significant similarities, and the distance between them was smaller. Therefore, the characteristics of
641    cross-sections, such as radiation, climate, rainfall, etc. had considerably more remarkable similarity
642    and homogeneity compared with the towers used in this research. Finally, it is worth mentioning that
643    PD has been commonly used to analyse the time series with a time resolution of weekly or longer,
644    with some exceptions using daily time steps. In this research, the data resolution was half-hourly
645    instead, which dramatically increased the computational demands of the algorithm, led to days of
646    processing for a single run. This demand happened because the algorithm creates a dummy variable
647    for each time step and the relevant matrix of variables becomes too large to compute by a regular PC.
648    Considering the computational expense of this algorithm, we recommend other researches not to use
649    PD when the time resolution is shorter than daily. Despite the limitation, we still encourage further

650    use of PD whenever there is a decent homogeneity level amongst the cross-sections and the time
651    resolution is daily or longer.

652         As a hybrid linear model, the ELN did not show any superiority over the CLR, despite its
653    modifications to provide more accurate estimations. Even though ELN performed well in estimating
654    the drivers with slight supremacy on some occasions (e.g. Fld, the CLR is a more proper algorithm to
655    choose for gap-filling the drivers due to its simplicity and less calculation requirement).

656         The FBP was a unique algorithm used in this research, as it did not use any independent
657    variables to estimate the values of drivers and fluxes. The FBP performance was the least satisfactory
658    of all the algorithms. Therefore, FBP cannot be considered as a reliable alternative for current
659    algorithms to fill the gaps, especially longer ones.

660         Given that some of the environmental drivers that affect Fc are different during the day versus
661    night, separating the diurnal and nocturnal datasets to train the algorithms could improve the
662    outcome. Mainly because of the $u^*$ threshold filtering and other problems associated with the
663    nocturnal period, the portion of diurnal data is generally, by far, outweighs the nocturnal data portion,
664    which potentially leads to a bias in the algorithm. The same challenge is associated with soil moisture
665    estimation, as the behaviour of the system's behaviour on sunny days is utterly different from during
666    the rainy periods. Moreover, the system memory and the antecedent condition are undeniable features
667    associated with soil moisture (Ogle et al., 2015). Therefore, using models that can address these
668    considerations are more likely to improve the estimations.

669         Finally, it is noteworthy that some of the flux drivers used in this study as input features for
670    the gap-filling algorithms are not commonly used or might not globally be available. However,
671    considering that similar relative performance has been achieved in other researches for which
672    different sets of input features had been used (Kim et al., 2020), the relative performance of the
673    algorithms reviewed in this research should generally provide similar relative performance while
674    using different input features.

675    ## 5.    Conclusions

676         Eight different gap-filling algorithms for estimating 16 meteorological drivers as well as nine
677    algorithms for the three key ecosystem turbulent fluxes (sensible heat flux (Fh), latent heat flux (Fe),
678    and net carbon flux (Fc)) were investigated, and their performance evaluated based on the datasets of
679    five towers in Australia. Overall, three ML algorithms, XGB, RF and ANNs, performed nearly equally
680    well and significantly better than their linear rivals (the CLR, PD, and ELN) in estimating the flux
681    values. However, the linear algorithms performed almost equally well as the ML algorithms in
682    assessing the meteorological drivers. Amongst these nine algorithms, the RF and XGB showed the
683    highest level of robustness and reliability in estimating the Fc, Fe, and Fh. The PD was expected to
684    perform better than the linear methods, and it was hoped to compete with the ML algorithms in
685    estimating the fluxes, but it failed to do so. The SVR was the only ML algorithm that did not perform
686    at the same level as the rest ML algorithms that we suspect were due to over-fitting issues, while the
687    MDS performed somewhere in between. Considering the outcomes of previous research undertaken
688    in the OzFlux Network (e.g. (Cleverly et al., 2013; Isaac et al., 2017)), none of the ML algorithms used

in this research was proven to provide substantially better flux estimations compared with the standard method (ANNs). Nonetheless, amongst the algorithms tested in this research, the RF showed potential capabilities as an alternative due to its more consistent performance regarding the long gaps. Finally, we recommend suggestions below to improve the results for similar prospective researchers, as well as the QC and gap-filling procedure for flux networks:

1) Since the RF was more consistent than its competitors, including the ANNs, we suggest it is a good idea to use RF alongside the commonly used algorithms in challenging scenarios, such as long gaps, to figure out whether this superiority can be generalised.

2) It appears that even after three levels of quality control process done by the flux processing software (e.g. PyFluxPro), the data is still quite noisy. These noisy data are an essential source of both uncertainty and inaccuracy of the outcome, regardless of the algorithm used to gap-fill the data. As a result, another level of quality control methods, such as Wavelets or Matrix Factorisation, in addition to the current classical ones used by the PyFluxPro and other similar platforms, can probably improve the data quality and thereby improve the final imputation results.

3) For future researchers, using recurrent neural networks (RNNs) instead of feedforward neural networks (FFNN) could improve the estimations. That is likely because RNNs help the model to consider temporal dynamic behaviour of time series, as unlike FFNN, wherein the activations flow only from the input layer to the output layer, RNNs also have neuron connections pointing backwards (Géron, 2019). There is a demand for an algorithm capable of considering time has been mentioned in previous research as one of the reasons why testing the new algorithms is needed (Richardson and Hollinger, 2007).

4) Developing ensemble models using algorithms with different weaknesses and strengths may also enhance the results where a single algorithm shows performance deficiency.

## 6.    Data availability

The data were used in this research are available through the following sources: The L3 and L4 data are accessible from the OzFlux data portal (http://data.ozflux.org.au/portal). Current ACCESS-R and data are available from the BoM OPeNDAP server (https://www.opendap.org/). Likewise, the data coming from the BoM AWS are accessible from (http://www.bom.gov.au/climate/data). Lastly, the BIOS2 data are accessible from the ECMWF datasets portal (https://www.ecmwf.int/en/forecasts/datasets). All data used in this research are available in this repository address: (https://research-repository.uwa.edu.au/en/datasets/a-comparison-of-gap-filling-algorithms-for-eddy-covariance-fluxes); DOI: 10.26182/5f292ee80a0c0.

*Author contributions.* The ideas for this study originated in discussions with A. Mahabbati, J. Beringer, and M. Leopold. A. Mahabbati carried out the analysis, supported by I. McHugh and P. Isaac. The paper was prepared with contributions from all authors.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* The authors would like to acknowledge the Terrestrial Ecosystems Research Network (TERN) (www.tern.gov.au) and the OzFlux Network as a part of TERN for supporting the grants and providing the required data, respectively. A. Mahabbati also personally thanks Prajwal Kalfe, Caroline Johnson and Cacilia Ewenz for their support regarding Python programming, English academic writing and PyFluxPro technical issues.

# References

Allison, P. D.: Multiple Imputation for Missing Data: A Cautionary Tale, Sociol. Methods Res., 28(3), 301–309, doi:10.1177/0049124100028003003, 2000.

Altman, D. G. and Bland, J. M.: Missing data, Br. Med. J., 334(7590), 424, doi:10.1136/bmj.38977.682025.2C, 2007.

Aubinet, M., Grelle, A., Ibrom, A., Rannik, Ü., Moncrieff, J., Foken, T., Kowalski, A. S., Martin, P. H., Berbigier, P., Bernhofer, C., Clement, R., Elbers, J., Granier, A., Grünwald, T., Morgenstern, K., Pilegaard, K., Rebmann, C., Snijders, W., Valentini, R. and Vesala, T.: Estimates of the Annual Net Carbon and Water Exchange of Forests: The EUROFLUX Methodology, Adv. Ecol. Res., 30, 113–175, doi:10.1016/S0065-2504(08)60018-5, 1999.

Aubinet, M., Vesala, T. and Papale, D.: Eddy Covariance: A Practical Guide to Measurement and Data Analysis., 2012.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, U. K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K. and Wofsy, S.: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities, Bull. Am. Meteorol. Soc., 82(11), 2415–2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2, 2001.

Baltagi, B.: Econometric analysis of panel data, [online] Available from: http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=book2.xis&method=post&formato=2&cantidad=1&expresion=mfn=001143 (Accessed 13 March 2018), 1995.

Barr, A. G., Black, T. A., Hogg, E. H., Kljun, N., Morgenstern, K. and Nesic, Z.: Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production, Agric. For. Meteorol., 126(3–4), 237–255, doi:10.1016/J.AGRFORMET.2004.06.011, 2004.

Barr, A. G., Richardson, A. D., Hollinger, D. Y., Papale, D., Arain, M. A., Black, T. A., Bohrer, G., Dragoni, D., Fischer, M. L., Gu, L., Law, B. E., Margolis, H. A., Mccaughey, J. H., Munger, J. W., Oechel, W. and Schaeffer, K.: Use of change-point detection for friction-velocity threshold evaluation in eddy-covariance studies, Agric. For. Meteorol., 171–172, 31–45, doi:10.1016/j.agrformet.2012.11.023, 2013.

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D. and Andreassian, V.: Characterising performance of environmental models, Environ. Model. Softw., 40, 1–20, doi:10.1016/j.envsoft.2012.09.011, 2013.

Beringer, J., Hutley, L. B., McHugh, I., Arndt, S. K., Campbell, D., Cleugh, H. A., Cleverly, J., De Dios, V. R., Eamus, D., Evans, B., Ewenz, C., Grace, P., Griebel, A., Haverd, V., Hinko-Najera, N., Huete, A., Isaac, P., Kanniah, K., Leuning, R., Liddell, M. J., MacFarlane, C., Meyer, W., Moore, C., Pendall, E., Phillips, A., Phillips, R. L., Prober, S. M., Restrepo-Coupe, N., Rutledge, S., Schroder, I., Silberstein, R., Southall, P., Sun Yee, M., Tapper, N. J., Van Gorsel, E., Vote, C., Walker, J. and Wardlaw, T.: An introduction to the Australian and New Zealand flux tower network - OzFlux, Biogeosciences, 13(21), 5895–5916, doi:10.5194/bg-13-5895-2016, 2016a.

Beringer, J., McHugh, I. and KLJUN, N.: Dynamic INtegrated Gap filling and partitioning for Ozflux (DINGO), Biogeosciences Discuss., OzFlux spe(In prep), 1457–1460, doi:doi:10.5194/bg-2016-188, 2016b.

Beringer, J., McHugh, I., Hutley, L. B., Isaac, P. and Kljun, N.: Technical note: Dynamic INtegrated Gap-filling and partitioning for OzFlux (DINGO), Biogeosciences, 14(6), 1457–1460, doi:10.5194/bg-14-1457-2017, 2017.

773 Burba, G. and Anderson, D.: A brief practical guide to eddy covariance flux measurements: principles and workflow examples for
774     scientific and industrial applications. [online] Available from:
775     https://books.google.com/books?hl=en&lr=&id=mCsI1_8GdrIC&oi=fnd&pg=PA6&dq=A+Brief+Practical+Guide+to+Eddy+Covaria
776     nce+Flux+Measurements&ots=TKTg25Yq5X&sig=eBYc819N7Jh3gNhJInfEL1e40eM (Accessed 11 February 2020), 2010.

777 Chen, T. and Guestrin, C.: XGBoost: A scalable tree boosting system, Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 13-
778     17-Augu, 785–794, doi:10.1145/2939672.2939785, 2016.

779 Cleverly, J., Boulain, N., Villalobos-Vega, R., Grant, N., Faux, R., Wood, C., Cook, P. G., Yu, Q., Leigh, A. and Eamus, D.: Dynamics
780     of component carbon fluxes in a semi-arid *Acacia* woodland, central Australia, J. Geophys. Res. Biogeosciences, 118(3), 1168–1185,
781     doi:10.1002/jgrg.20101, 2013.

782 Devore, J. L.: Probability and Statistics for Engineering and the Sciences., Biometrics, 47(4), 1638, doi:10.2307/2532427, 1991.

783 Dragoni, D., Schmid, H. P., Grimmond, C. S. B. and Loescher, H. W.: Uncertainty of annual net ecosystem productivity estimated
784     using eddy covariance flux measurements, J. Geophys. Res., 112(D17), D17102, doi:10.1029/2006JD008149, 2007.

785 Dreyfus, S. E.: Artificial neural networks, back propagation, and the kelley-bryson gradient procedure, J. Guid. Control. Dyn., 13(5),
786     926–928, doi:10.2514/3.25422, 1990.

787 Drucker, H., Surges, C. J. C., Kaufman, L., Smola, A. and Vapnik, V.: Support vector regression machines, in Advances in Neural
788     Information Processing Systems, vol. 1, pp. 155–161., 1997.

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

794 Farley, B. G. and Clark, W. A.: Simulation of self-organizing systems by digital computer, IRE Prof. Gr. Inf. Theory, 4(4), 76–84,
795     doi:10.1109/TIT.1954.1057468, 1954.

796 Freedman, D. A.: Statistical Models: Theory and Practice. Cambridge University Press - 2nd edition. [online] Available from:
797     https://www.cambridge.org/au/academic/subjects/statistics-probability/statistical-theory-and-methods/statistical-models-theory-
798     and-practice-2nd-edition?format=PB (Accessed 21 March 2020), 2009.

799 Friedman, J. H.: Stochastic gradient boosting, Comput. Stat. Data Anal., 38(4), 367–378, doi:10.1016/S0167-9473(01)00065-2, 2002.

800 Gani, A., Mohammadi, K., Shamshirband, S., Altameem, T. A., Petković, D. and Ch, S.: A combined method to estimate wind speed
801     distribution based on integrating the support vector machine with firefly algorithm, Environ. Prog. Sustain. Energy, 35(3), 867–
802     875, doi:10.1002/ep.12262, 2016.

803 Géron, A.: Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent
804     systems. [online] Available from:
805     https://books.google.com.au/books?hl=en&lr=&id=HHetDwAAQBAJ&oi=fnd&pg=PP1&dq=hands-
806     on+machine+learning+with+&ots=0KvfZqlgOo&sig=5tH2IHRsUaTMTy6CfQ6lw3UDKa4 (Accessed 7 February 2020), 2019.

807 Hagen, S. C., Braswell, B. H., Linder, E., Frolking, S., Richardson, A. D. and Hollinger, D. Y.: Statistical uncertainty of eddy flux -
808     Based estimates of gross ecosystem carbon exchange at Howland Forest, Maine, J. Geophys. Res. Atmos., 111(8), 1–12,
809     doi:10.1029/2005JD006154, 2006.

810 Harrell, F. E.: Regression Modeling Strategies: With Applications to Linear Models, Logistic, in books.google.nl. [online] Available
811     from:
812     https://books.google.com.au/books?hl=en&lr=&id=94RgCgAAQBAJ&oi=fnd&pg=PR7&dq=regression+modeling+strategies+frank
813     +harrell&ots=ZAt4RsaS1r&sig=mikE1s9G4IXzqZKEie-iVA9GTV0&redir_esc=y#v=onepage&q=regression modeling strategies
814     frank harrell&f=false (Accessed 11 February 2020), 2014.

815 Harvey, A. C. and Peters, S.: Estimation procedures for structural time series models, J. Forecast., 9(2), 89–108,
816     doi:10.1002/for.3980090203, 1990.

817 Haverd, V., Briggs, P., Trudinger, C., Nieradzik, L. and Canadell, P.: BIOS2 – Frontier Modelling of the Australian Carbon and
818     Water Cycles, 2015.

819 Ho, T. K.: Random decision forests, Proc. Int. Conf. Doc. Anal. Recognition, ICDAR, 1, 278–282, doi:10.1109/ICDAR.1995.598994,

820     1995.

821     Ho, T. K.: 00709601.Pdf, , 20(8), 832–844, 1998.

822     Hollinger, D. Y., Goltz, S. M., Davidson, E. A., Lee, J. T., Tu, K. and Valentine, H. T.: Seasonal patterns and environmental control of
823         carbon dioxide and water vapour exchange in an ecotonal boreal forest, Glob. Chang. Biol., 5(8), 891–902, doi:10.1046/j.1365-
824         2486.1999.00281.x, 1999.

825     Hsiao, C., Hashem Pesaran, M. and Kamil Tahmiscioglu, A.: Maximum likelihood estimation of fixed effects dynamic panel data
826         models covering short time periods, J. Econom., 109(1), 107–150, doi:10.1016/S0304-4076(01)00143-9, 2002.

827     Hui, D., Wan, S., Su, B., Katul, G., Monson, R. and Luo, Y.: Gap-filling missing data in eddy covariance measurements using
828         multiple imputation (MI) for annual estimations, Agric. For. Meteorol., 121(1–2), 93–111, doi:10.1016/S0168-1923(03)00158-8, 2004.

829     Hutley, L. B., Leuning, R., Beringer, J. and Cleugh, H. a: The utility of the eddy covariance technique as a tool in carbon accounting:
830         tropical savanna as a case study, Aust. J. Bot., 53, 663–675, 2005.

831     Isaac, P., Cleverly, J., McHugh, I., Van Gorsel, E., Ewenz, C. and Beringer, J.: OzFlux data: Network integration from collection to
832         curation, Biogeosciences, 14(12), 2903–2928, doi:10.5194/bg-14-2903-2017, 2017.

833     Izady, A., Davary, K., Alizadeh, A., Moghaddam Nia, A., Ziaei, A. N. and Hasheminia, S. M.: Application of NN-ARX Model to
834         Predict Groundwater Levels in the Neishaboor Plain, Iran, Water Resour. Manag., 27(14), 4773–4794, doi:10.1007/s11269-013-0432-
835         y, 2013.

836     Izady, A., Abdalla, O. and Mahabbati, A.: Dynamic panel-data-based groundwater level prediction and decomposition in an arid
837         hardrock–alluvium aquifer, Environ. Earth Sci., 75(18), 1–13, doi:10.1007/s12665-016-6059-6, 2016.

838     Jerome H. Friedman: Greedy Function Approximation: A Gradient Boosting Machine on JSTOR, Ann. Stat., 29, 1189–1232 [online]
839         Available from: https://www.jstor.org/stable/2699986?seq=1#metadata_info_tab_contents (Accessed 27 August 2019), 2001.

840     Kang, H.: The prevention and handling of the missing data, Korean J. Anesthesiol., 64(5), 402–406, doi:10.4097/kjae.2013.64.5.402,
841         2013.

842     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
843         eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal
844         component analysis, Glob. Chang. Biol., 26(3), 1499–1518, doi:10.1111/gcb.14845, 2020.

845     Kock, N. and Gaskins, L.: Simpson's paradox, moderation and the emergence of quadratic relationships in path models: an
846         information systems illustration, Int. J. Appl. Nonlinear Sci., 2(3), 200, doi:10.1504/ijans.2016.077025, 2016.

847     Kunwor, S., Starr, G., Loescher, H. W. and Staudhammer, C. L.: Preserving the variance in imputed eddy-covariance measurements:
848         Alternative methods for defensible gap filling, Agric. For. Meteorol., 232, 635–649, doi:10.1016/j.agrformet.2016.10.018, 2017.

849     Law, B. E., Falge, E., Gu, L., Baldocchi, D. D., Bakwin, P., Berbigier, P., Davis, K., Dolman, A. J., Falk, M., Fuentes, J. D., Goldstein,
850         A., Granier, A., Grelle, A., Hollinger, D., Janssens, I. A., Jarvis, P., Jensen, N. O., Katul, G., Mahli, Y., Matteucci, G., Meyers, T.,
851         Monson, R., Munger, W., Oechel, W., Olson, R., Pilegaard, K., Paw U H, K. T., Thorgeirsson, H., Valentini, R., Verma, S., Vesala,
852         T., Wilson, K. and Wofsy, S.: Jourassess2, Agric. For. Meteorol., 113(113), 97–120, 2002.

853     Lee, X., Fuentes, J. D., Staebler, R. M. and Neumann, H. H.: Long-term observation of the atmospheric exchange of CO2 with a
854         temperate deciduous forest in southern Ontario, Canada, J. Geophys. Res. Atmos., 104(D13), 15975–15984,
855         doi:10.1029/1999JD900227, 1999.

856     Little, R. J. A.: Statistical analysis with missing data, 2nd ed., edited by D. B. Rubin, Wiley, Hoboken, N.J., 2002.

857     Mahabbati, A., Izady, A., Mousavi Baygi, M., Davary, K. and Hasheminia, S. M.: Daily soil temperature modeling using 'panel-data'
858         concept, J. Appl. Stat., 44(8), 1385–1401, doi:10.1080/02664763.2016.1214240, 2017.

859     Menzer, O., Moffat, A. M., Meiring, W., Lasslop, G., Schukat-Talamazzini, E. G. and Reichstein, M.: Random errors in carbon and
860         water vapor fluxes assessed with Gaussian Processes, Agric. For. Meteorol., 178–179, 161–172, doi:10.1016/j.agrformet.2013.04.024,
861         2013.

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

865   doi:10.1016/j.agrformet.2007.08.011, 2007.

866   Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., Verbeke, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A. and
867      Verbeke, G.: Handbook of Missing Data Methodology, Chapman and Hall/CRC., 2014.

868   Ogle, K., Barber, J. J., Barron-Gafford, G. A., Bentley, L. P., Young, J. M., Huxman, T. E., Loik, M. E. and Tissue, D. T.: Quantifying
869      ecological memory in plant and ecosystem processes, Ecol. Lett., 18(3), 221–235, doi:10.1111/ele.12399, 2015.

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

872   Pilegaard, K., Hummelshøj, P., Jensen, N. O. and Chen, Z.: Two years of continuous $CO_2$ eddy-flux measurements over a Danish
873      beech forest, Agric. For. Meteorol., 107(1), 29–41, doi:10.1016/S0168-1923(00)00227-6, 2001.

874   Reichle, R. H., Koster, R. D., Dong, J. and Berg, A. A.: Global soil moisture from satellite observations, land surface models, and
875      ground data: Implications for data assimilation, J. Hydrometeorol., 5(3), 430–442, doi:10.1175/1525-
876      7541(2004)005<0430:GSMFSO>2.0.CO;2, 2004.

877   Richardson, A. D. and Hollinger, D. Y.: A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps
878      in the $CO_2$ flux record, Agric. For. Meteorol., 147(3–4), 199–208, doi:10.1016/j.agrformet.2007.06.004, 2007.

879   Richardson, A. D., Braswell, B. H., Hollinger, D. Y., Burman, P., Davidson, E. A., Evans, R. S., Flanagan, L. B., Munger, J. W., Savage,
880      K., Urbanski, S. P. and Wofsy, S. C.: Comparing simple respiration models for eddy flux and dynamic chamber data, Agric. For.
881      Meteorol., 141(2–4), 219–234, doi:10.1016/J.AGRFORMET.2006.10.010, 2006.

882   Richardson, A. D., Aubinet, M., Barr, A. G., Hollinger, D. Y., Ibrom, A., Lasslop, G. and Reichstein, M.: Uncertainty Quantification,
883      in Eddy Covariance, pp. 173–209., 2012.

884   Sahoo, A. K., Dirmeyer, P. A., Houser, P. R. and Kafatos, M.: A study of land surface processes using land surface models over the
885      Little River Experimental Watershed, Georgia, J. Geophys. Res. Atmos., 113(20), doi:10.1029/2007JD009671, 2008.

886   Scanlon, T. M. and Kustas, W. P.: Partitioning carbon dioxide and water vapor fluxes using correlation analysis, Agric. For.
887      Meteorol., 150(1), 89–99, doi:10.1016/j.agrformet.2009.09.005, 2010.

888   Scanlon, T. M. and Sahu, P.: On the correlation structure of water vapor and carbon dioxide in the atmospheric surface layer: A
889      basis for flux partitioning, Water Resour. Res., 44(10), doi:10.1029/2008WR006932, 2008.

890   Staebler, M.: Long-term observation of the atmospheric exchange of $CO_2$ with a temperate deciduous forest in southern Ontario ,
891      Canada ecosystem ( net ecosystem production turbulence is turbulent, Data Process., 104, 975–984, 1999.

892   Tannenbaum, C. E.: The empirical nature and statistical treatment of missing data., Diss. Abstr. Int. Sect. A Humanit. Soc. Sci., 70(10-
893      A), 3825 [online] Available from: http://gateway.proquest.com/openurl?url_ver=Z39.88-
894      2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&res_dat=xri:pqdiss&rft_dat=xri:pqdiss:3381876%5Cnhttp://ovidsp.ovid.com/o
895      vidweb.cgi?T=JS&PAGE=reference&D=psyc7&NEWS=N&AN=2010-99071-044, 2010.

896   Taylor, S. J. and Letham, B.: Business Time Series Forecasting at Scale, , doi:10.7287/peerj.preprints.3190v2, 2017.

897   Taylor, S. J. and Letham, B.: Forecasting at Scale, Am. Stat., 72(1), 37–45, doi:10.1080/00031305.2017.1380080, 2018.

898   Tenhunen, J. D., Valentini, R., Köstner, B., Zimmermann, R. and Granier, A.: Variation in forest gas exchange at landscape to
899      continental scales, Ann. des Sci. For., 55(1–2), 1–11, doi:10.1051/forest:19980101, 1998.

900   Wooldridge, J. M.: Econometric Analysis of Cross Section and Panel Data., 2008.

901   Ye, J., Chow, J.-H., Chen, J. and Zheng, Z.: Stochastic gradient boosted distributed decision trees, in Proceeding of the 18th ACM
902      conference on Information and knowledge management - CIKM '09, p. 2061, ACM Press, New York, New York, USA., 2009.

903   Zhao, X. and Huang, Y.: A comparison of three gap filling techniques for eddy covariance net carbon fluxes in short vegetation
904      ecosystems, Adv. Meteorol., 2015, 1–12, doi:10.1155/2015/260580, 2015.

905   Zou, H. and Hastie, T.: Regularization and variable selection via the elastic net. [online] Available from:
906      https://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=22250F01CC77D55C54B6BAFF4512C9E3?doi=10.1.1.124.4696&rep=rep
907      1&type=pdf (Accessed 28 August 2019), 2005.

908