Environmental Modelling and Software 193 (2025) 106611

Contents lists available at ScienceDirect

Environmental Modelling and Software

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

Introducing 'miniRECgap' R package for simple gap-filling of missing eddy covariance CO_2 flux measurements with classic nonlinear environmental response functions via GUI-supported R-scripts (case-study: In-sample gap-filling with 'miniRECgap' vs. MDS and an optimised shallow ANN in a 'challenging' peatland ecosystem)

Alina Premrov [a,b,*], Jagadeesh Yeluripati [c], Richard Slevin [a], Adam Bates [a], Magdalena Matysek [a], Stephen Barry [d,m], Kenneth A. Byrne [e,f], Rowan Fealy [g], Bernard Hyde [h], Gary Lanigan [i], Mark McCorry [d], Rachael Murphy [i,l], Florence Renou-Wilson [j], Amey Tilak [e], David Wilson [k], Matthew Saunders [a]

[a] *School of Natural Sciences, Botany Discipline, Trinity College Dublin, Ireland*
[b] *Department of Environmental Science, Faculty of Science, Atlantic Technological University, Sligo, Ireland*
[c] *Information and Computational Sciences Department, The James Hutton Institute, Aberdeen, Scotland, United Kingdom*
[d] *Bord na Móna, 2-, 4 Boora Ave., Blueball, Co. Offaly, Ireland*
[e] *Department of Biological Sciences, University of Limerick, Limerick, Ireland*
[f] *Bernal Institute, University of Limerick, Limerick, Ireland*
[g] *ICARUS & Department of Geography, Maynooth University, Ireland*
[h] *Environmental Protection Agency, Regional Inspectorate, The Glen, Monaghan Town, Ireland*
[i] *Environment, Soils and Land-Use Research Teagasc, Teagasc Johnstown Castle, Wexford, Ireland*
[j] *School of Biology & Environmental Sciences, University College Dublin, Ireland*
[k] *Earthy Matters Environmental Consultants, Donegal, Ireland*
[l] *Teagasc Climate Centre, Ireland*
[m] *School of Architecture Building, and the Environment, Technological University Dublin, Dublin, Ireland*

ARTICLE INFO

Keywords:
Eddy covariance gap-filling and flux-partitioning
CO_2 fluxes
'miniRECgap' R-package
Nonlinear environmental response functions
Artificial neural networks

ABSTRACT

Numerous tools/software exist to gap-fill missing eddy covariance (EC) data, with varying performance depending on study-site dynamics. Disturbed ecosystems like former cutaway-peatlands may be challenging for gap-filling. Researchers using gap-filling spreadsheets may benefit from transitioning to R, but may face challenges if they lack programming skills. To address these, we introduce 'miniRECgap', a user-friendly tool in R for effortless gap-filling of EC carbon dioxide flux data using well-known temperature- and light-response functions. 'miniRECgap' can model net ecosystem exchange (NEE) via GUI-supported scripts with only five code-lines and minimal inputs. A case-study on one 'classic' (forest) and one 'challenging' (rehabilitating cutaway-peatland) ecosystem indicated that standard gap-filling (MDS) performed better for the 'classic', but not for the 'challenging' ecosystem (MDS $R^2 = 0.24$; 'miniRECgap' $R^2 = 0.57$). For the rehabilitating-peatland, an optimised shallow Artificial Neural Network outperformed other two approaches ($R^2 = 0.68$). These findings demonstrate the importance of NEE gap-filling for assessing ecosystem-level carbon-dynamics, important for rehabilitating-peatlands.

* Corresponding author. School of Natural Sciences, Botany Discipline, Trinity College Dublin, Ireland; Department of Environmental Science, Faculty of Science, Atlantic Technological University, Sligo, Ireland.
E-mail addresses: apremrov@tcd.ie, alina.premrov@atu.ie (A. Premrov).

https://doi.org/10.1016/j.envsoft.2025.106611
Received 16 August 2024; Received in revised form 30 June 2025; Accepted 1 July 2025
Available online 14 July 2025
1364-8152/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Abbreviations:

A_{max}	maximum rate of CO_2 assimilation
'am'	subscript in P and O - refers to the arithmetic mean of each variable
ANN	artificial neural network
AOC	area over curve
'BFGS'	optimisation algorithm/method (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970)
BP	back propagation
C	carbon
c	circa
Co	county
CO_2	carbon dioxide
d	dimensions in the dataset (refers to the training points upper bound explained in Dubbs (2021–2022))
DI	diurnal interpolation
DNN	deep neural network
DOY (or DoY)	day of year
E	total uncertainty (Eq. (10))
EC	Eddy Covariance
Er	random error (Eq. (9))
Es	systematic error or bias (Eq. (8))
EWMA	Exponential Weighted Moving Average
GHG	greenhouse gas
GPP	gross primary productivity
GUI	graphical user interface
H	sensible heat flux
ICOS	Integrated Carbon Observation System
j	number of hidden neurons on a hidden layer in ANN
k-cv	k-cross validation (k refers to number of repetitions in k-cross validation, not to be confused with the unsupervised ML 'k-means clustering' approach)
'L-BFGS-B'	optimisation method/algorithm of Byrd et al. (1995).
LH	latent heat flux
LUT	look-up table
MAE	mean absolute error (Eq (7b))
MC	Monte-Carlo
MDS	marginal distribution sampling
MDSnight	MDS followed by nighttime net flux partitioning according to Reichstein et al. (2005).
MI	multiple imputation
ML	machine learning
MPI	Max Planck Institute
n	number of points in the dataset (refers to the training points upper bound explained in Dubbs (2021–2022))
'N3'	shallow ANN with one hidden layer and three hidden neurons (j = 3), also referred to as 'nnj = 3'
NEE	net ecosystem exchange
Np	number of predicted half-hourly fluxes for which also the observed values are available
Oi	observed flux at each half-hourly time-interval (i)
p	number of parameters in optimal training/testing ratio of $p^{0.5}$: 1 suggested by Joseph (2022).
PAR	photosynthetic active radiation
PB	process-based biogeochemical/biogeophysical modelling
Pi	predicted flux at each half-hourly time-interval (i)
PM	process modelling
PPFD	photosynthetic photon flux density
QC/QA	quality control/quality acceptance
R10	respiration rate at 10 °C
R^2	prediction coefficients of determination (Eq. (5))
'relu'	'rectified linear unit' activation function in ANN
Reco	ecosystem respiration
RF	random forest
Rg	incoming solar radiation
rH	relative humidity
RMSD (or RMSE*)	root mean squared prediction deviation (Eq. (6b))
RMSE	root mean squared prediction error (Eq. (6a))
rRMSPE	relative version of RMSE (Eq. (6c))
RROC	Regression Receiver Operating Characteristic
SPM	semi-parametric modelling
StD	standard deviation calculated from MAE (Eq. (7a))
SVM	support vector machine
T	measured temperature
T0	temperature of 230 K
Tair	air temperature
Tsoil	soil temperature
u*(or Ustar)	friction velocity
UKF	unscented Kalman filter
VPD	vapor pressure deficit
WGS84	World Geodetic System 1984 geographic coordinate system
XGBoost	parallel boosted decision trees -i.e. eXtreme gradient boosting
α	quantum yield based on incident irradiance
y	coefficient of convexity

1. Introduction

1.1. Introduction to the eddy covariance (EC) technique and flux gap-filling

Atmospheric flux measurements are central to investigating the ecosystem biosphere-atmosphere exchange of energy, water, and greenhouse gases (GHGs), such as carbon dioxide (CO_2) (Burba et al., 2007). Eddy covariance (EC) is a well-known micrometeorological technique frequently used to measure and calculate the turbulent fluxes and net flux exchange across the vegetation canopy-atmosphere boundary layer (Baldocchi, 2003; Baldocchi et al., 1988). EC technique is considered a relatively non-invasive (in-situ) technique compared to some other techniques (Baldocchi et al., 1988) and so may be particularly useful to study terrestrial ecosystems, such as peatlands. For a detailed description of the technique, see Burba and Aderson (2010), Aubinet et al. (2012), and Rebmann et al. (2018). The accuracy of the EC technique depends on various environmental conditions, with steady atmospheric conditions and homogeneous vegetation cover typically preferable (Baldocchi, 2003). Such ideal conditions are not often found in many natural ecosystems where environmental conditions can vary both temporally and spatially. This means that atmospheric storage, flux divergence and advection measurements need to be accounted for in the quantification of the GHG exchanges, such as in the case of CO_2 exchange quantifications (Baldocchi, 2003). The popularity of this method is evident from the development of EC station networks over the past three decades, such as the European Integrated Carbon Observing System (ICOS) network (Rebmann et al., 2018), which contains more than 170 sites across 16 countries (ICOS, 2023), and the worldwide FLUXNET network of networks (Papale, 2020; Pastorello et al., 2020) with over 1000 active and historic flux-site stations (FLUXNET, 2015). ICOS is a highly standardised European network that aims to adhere to sustained and consistent technical and scientific standards across network sites (ICOS, 2022).

One of the major advantages of the EC technique is its suitability to construct continuous or near-continuous data measurements of carbon

(C) exchange between the atmosphere and the biosphere over prolonged periods of time (Falge et al., 2001), which are often needed in ecosystem studies. For example, terrestrial ecosystem studies that focus on CO_2 fluxes often need to assess and compare the net ecosystem exchange (NEE) response, usually performed on annual sums of NEE obtained from EC measurements. Such comparisons of NEE response are studied in connection with the effect of various natural conditions and anthropogenic pressures on the global C balance - e.g. scientists may be interested in investigating the impact on NEE by biome type, different patterns in phenology, and various environmental and anthropogenic conditions (Falge et al., 2001). The construction of long datasets from EC networks is subject to the occurrence of data-gaps, which can occur for a variety of reasons (Baldocchi, 2003), such as instrument/sensor calibration, instrument/sensor malfunction or failure, data-rejection due to adverse weather, low turbulence, nonstationary conditions or undesirable wind conditions, or any other undesirable condition that will negatively impact the instruments and result in measurements that will fail the required data quality control (QC)/quality acceptance (QA) criteria (Baldocchi, 2003; Foken and Wichura, 1996; Menzer et al., 2015). Although, the QA/QC of raw flux data is not the main focus of this paper, it should be noted that this is usually carried out by applying the established standardised criteria and/or tools used within the scientific community. Further details on this topic can be found in relevant literature sources (AmeriFlux, 2024; Aubinet et al., 2012; FLUXNET, 2022; Sabbatini et al., 2018; Vitale et al., 2020).

Given that EC techniques, on average, provide only *c.* 65 %–68 % of annual data coverage (Falge et al., 2001; Vekuri et al., 2023), it is not surprising that various gap-filling methods have been developed over the past decades. Gap-filling is also frequently accompanied by the partitioning of the flux data into nighttime/daytime using different approaches (Lasslop et al., 2010; Richardson et al., 2006), to compute different components of the ecosystem C budget. While this study focuses on CO_2 fluxes, there are numerous gap-filling methods that can be used for different GHG EC flux measurements from various terrestrial ecosystems. These range from relatively simple to complex methods, method-combinations (Falge et al., 2001; Moffat et al., 2007; Vekuri et al., 2023), as well as various machine learning (ML) methods (Mahabbati, 2022). The application of EC flux data gap-filling in various terrestrial ecosystems is reported throughout the literature, and a number of the gap-filling methods are listed in Table 1. Many different types of computational tools/software and packages for processing EC flux data have also been developed, such as the well-known commercial software 'Tovi™'(LI-COR, 2023), and the freely available 'REddyProc' (Wutzler et al., 2018). The latter ('REddyProc') is available in the form of a R-package (Wutzler et al., 2018), and an online webtool (MPI, 2024d) hosted on the website of the Department of Biogeochemical Integration at the Max Planck Institute, Jena, Germany (MPI, 2024e). Several software/packages that can be used to perform different flux data post-processing tasks, including gap-filling and partitioning of ecosystem respiration data, are also listed on FLUXNET website (FLUXNET, 2022) and include some well-known examples applied in various studies on terrestrial ecosystems, such as:

- The 'REddyProc' package (Wutzler et al., 2018) written in R environment[1]; this package is often referred within the EC flux community as the standard approach for EC flux data processing, and has been used in various studies on terrestrial ecosystems globally, including e.g. forests and wetlands (e.g. Falge et al. (2001); Wutzler et al. (2018); Vekuri et al. (2023)).
- 'ONEFlux' (Open Network-Enabled Flux processing pipeline) package written in Python[2] that is jointly developed by some of the major flux networks and used for standard processing of their data from

[1] R: The R Foundation for Statistical Computing. URL: https://www.r-project.org/.

Table 1
Application of eddy covariance (EC) flux data gap-filling in various terrestrial ecosystems and sites: Examples of EC studies that have used various types of approaches to gap-fill missing greenhouse gas (GHG) flux data in terrestrial ecosystems.

A Grouped types of EC gap-filling methodologies	B Selected examples (studies from the literature)	C Various types of terrestrial ecosystems or sites (from selected examples/studies - literature under B)
1. Use of redundant variables, simple linear interpolation, merging values, different gap-filling methods based on the previous data or assorted conditions depending on meteorology, such as mean diurnal variations and look-up tables (LUT), etc.	Falge et al. (2001) and Jarvis et al. (1976).	Coniferous forest; deciduous forests; cropland; grassland.
2. Various semi-empirical and nonlinear regression gap-filling methods, which can vary from classic/traditional semi-empirical and parameter-optimisation nonlinear residual/least square methods based on the use of different environmental response functions, to various other multiple and non-linear regression methods allowing for different time-windows, seasonal dependencies, moving-window, dormancy during winter-periods, etc.	Falge et al. (2001); Barr et al. (2004); Hollinger et al. (2004); Desai et al. (2005), Richardson et al. (2006); Moffat et al. (2007); Jones et al. (2010); Lloyd (2010), Kiely et al. (2018); Holl et al. (2019); Murphy et al. (2022); Murphy (2022); Heiskanen (2023).	Coniferous forest; deciduous forests; old-growth forest; cropland/tillage land; grassland/grazed grassland; bogs/peatlands; subarctic ecosystems.
3. Advanced LUT approaches and interpolation methods, as well as various process/process-based models – e.g. marginal distribution sampling (MDS) for moving LUT, semi-parametric modelling (SPM) for three-dimensional LUT, diurnal interpolation (DI), multiple imputation (MI) methods, such as MI using Monte-Carlo (MC); various advanced process modelling (PM) approaches, such as using unscented Kalman filter (UKF), or process-based biogeochemical/biogeophysical models (PB), such as Agro-IBIS;	Reichstein et al. (2005) - MDS; Stauch and Jarvis (2006) - SPM; Falge et al. (2001) - DI; Hui et al. (2004) - MI with MC; Gove and Hollinger (2006) - PM (UKF); Aslan Sungur et al. (2019) - PB (Agro-IBIS).	Coniferous forest; deciduous forests; mixed forest; cropland; perennial biofuel crops; grassland; synthetic canopy data.
4. Use of natural statistical frameworks and different semi-supervised/supervised machine learning (ML) approaches, ranging from Bayesian approaches (Bayesian) and different multiple gap-filling approaches (multiple) to numerous ML approaches, such as random forests (RFs), support vector machines	Mahabbati (2022) - different MLs; Braswell et al. (2005) & Buzacott et al. (2023) - Bayesian; Lucas-Moffat et al. (2022) - multiple; Mahabbati et al. (2021), Irvin et al. (2021) & Gao et al. (2023) - RF; Gao et al. (2023) - SVM & BP; Vekuri et al. (2023) - XGBoost; Ooba et al. (2006), Evrendilek	Peatlands; agricultural/drained peatlands, mined/degraded/rewetted bog, wetlands; rice paddies; evergreen coniferous forests; deciduous broadleaf forest; mixed forest; decadal forest; grasslands; croplands; shrublands; Australian: steppe, tropical savanna, tropical &

(*continued on next page*)

Table 1 (*continued*)

A Grouped types of EC gap-filling methodologies	B Selected examples (studies from the literature)	C Various types of terrestrial ecosystems or sites (from selected examples/studies - literature under B)
(SVM) for non-linear regression, parallel boosted decision trees, i.e. extreme gradient boosting (XGBoost), Back Propagation (BP) neural networks, different types of artificial neural networks (ANNs), and many others.	(2013), Knox et al. (2015), Holl et al. (2020), Zhu et al. (2023), Melesse and Hanley (2005) - ANNs.	subtropical desert and moist broadleaf forest; Northern latitude sites: grassland, permanent wetlands, evergreen needleleaf forests; coniferous trees plantation; 'challenging' sites: managed and grazed pastures; oil palm plantations; dryland sites with experienced wildfire.

NOTE: It should be noted that Table 1 is for information purposes only. Readers who would like to further explore different types of gap-filling methods are recommended to check the relevant review-/comparison studies that cover this topic in detail, such as the works of Falge et al. (2001), Moffat et al. (2007) and Irvin et al. (2021).

various terrestrial ecosystems (ONEFlux, 2023); it is also known for its use in the creation of the FLUXNET2015 dataset (Pastorello et al., 2020).

- 'PyFluxPro' is a package that is also written in Python and integrated into GUI (PyFluxPro, 2024); its predecessor was 'OzFlux', developed by the flux community in Australia, and predominantly used for Australian terrestrial ecosystems, (Isaac et al., 2017).
- 'GaFir' (Zhao et al., 2014) is another package written in R, available at the Department of Micrometeorology, University of Bayreuth in Germany (GaFiR, 2014), which can be applied at forest-, crop-, and meadow-sites.
- 'FluxnetLSM' is a package that is also written in R environment and can be applied at FLUXNET sites; this package is suitable for land surface modelling (LSM) due to its ability to transform files into NetCDF format that can be used directly by LSM (Ukkola et al., 2017).

1.2. *Study rationale, objectives, and case-study*

1.2.1. *Rationale and objectives*

Despite the number of existing gap-filling tools/software and packages, many users/scientists still perform gap-filling of missing EC flux data outside these popular tools using spreadsheet-software, such as Microsoft Office Excel. The Excel Solver add-in program is often used to perform various optimisation tasks in different semi-empirical and nonlinear regression computations (e.g. group 2 in Table 1) applied for specific gap-filling needs that are often dictated by the specific characteristics of the ecosystem under study, and it has been reported in various publications and research dissertations (e. g. Cotten et al., 2017; Ito and Ishida, 2023; Lees et al., 2021; Lees et al., 2019; Lloyd, 2010; Myklebust et al., 2008; Šigut, 2012; Singh, 2008; Strack et al., 2018). These examples show that Excel Solver has many great applications (Fylstra et al., 1998) and its use in science and education has been well established over a number of years. However, for users/scientists, who perform such computations for flux gap-filling tasks via spreadsheet-software, it may be beneficial to transition to R or Python, particularly in cases with large datasets and where they may wish to reduce the time of their computational-runs, while some less experienced R or Python users may potentially face challenges in adapting their specific EC flux data gap-filling methods to these computing environments. Therefore, the development of a user-friendly gap-filling computational tool written in R, which would enable a more effortless application of one of the frequently used semi-empirical and nonlinear regression approaches, as well as being suitable for new users with no prior knowledge of R, is considered beneficial.

While some user-friendly gap-filling tools/software already exist, such as the 'REddyProc' webtool (MPI, 2024d), there may be also other various reasons why some users perform gap-filling by applying semi-empirical and nonlinear regression methods that may not necessarily be supported by existing software tools. This may occur in cases where the EC flux data were obtained from stations that were not equipped to operate according to standardised protocols and may not have the data for all variables that are required as inputs into some existing well-established flux gap-filling tools/software. In other cases, the reasons may be due to specific conditions, characteristics or management practices at the study sites, which may restrict the application of the standard gap-filling methods. An example is a study by Šigut (2012), which required the application of an empirically calibrated correction factor for ecosystem respiration due to the topographically complex terrain at their study site.

There can be many other reasons why some sites may experience specific conditions that could restrict the application of some standard or frequently used gap-filling methods. For example, while many popular flux gap-filling methods are well known for less-disturbed natural or near-natural ecosystems, their application can be challenging in ecosystems that have undergone significant disturbance (Zhu et al., 2023), such as converted and rehabilitated cutaway-peatlands, which are often extremely heterogenous in terms of vegetation composition, as well as the extent of bare soil/peat and open water areas (Bord na Móna, 2021). For such ecosystem sites, the key variables used in some of the standard/popular gap-filling methods may or may not sufficiently account for their specific conditions. In such cases, it can be expected that several different gap-filling approaches may need to be assessed (including more advanced methods, i.e. Group 4 in Table 1) to find the most suitable one for the given ecosystem/site. Therefore, an additional user-friendly tool/software that would allow users to evaluate and decide, in a more timely and effortless manner, whether the application of a simple semi-empirical and nonlinear regression gap-filling approach may be suitable for the given site is required.

This paper introduces a simple gap-filling R-package computational tool, 'miniRECgap'. It is based on the application of plain and robust validated empirical/semi-empirical modelling EC flux data gap-filling approaches. The 'miniRECgap' package operates via GUI-supported R-scripts and is purposely designed to be user-friendly and suitable for new users without any prior knowledge of R.

The main objective of this paper is to introduce the new 'miniRECgap' R-package to the audience that is interested in applying very simple and robust gap-filling of missing EC CO_2 flux measurements, and to provide guidance on using the selected functions. Given that this R-package was designed for users with very little or no experience in the R environment/language, the intention was to keep it as simple and as short as practically possible. This is a relatively small package, hence the word 'mini' in the package name. The flux-partitioning and gap-filling approaches applied in 'miniRECgap' are based on very simple yet robust empirical/semi-empirical modelling, with minimum required input-variables. For this, we have chosen classic light- and temperature-response functions (Gilmanov et al., 2003; Lloyd and Taylor, 1994; Rabinowitch, 1951) known to represent some of the most conventional and widely used nonlinear functions within the EC flux community, which have been applied (among others) in various terrestrial ecosystems (see Table 1 - group 2). Therefore, it should be noted that the gap-filling methodology applied in 'miniRECgap' is not meant to outperform other more sophisticated and advanced gap-filling methods or packages. As such, the main purpose in designing 'miniRECgap' was to deliver a very simple R-package that could:

- Assist scientists/users to perform a more effortless application of the selected robust empirical/semi-empirical gap-filling method used in this package (i.e. light- and temperature-response functions (Gilmanov et al., 2003; Lloyd and Taylor, 1994; Rabinowitch, 1951), to enable them to assess whether this approach may be suitable for their datasets in a potentially more effortless way than using spreadsheet-software, for example.
- Serve as a potential tool for learning purposes, such as for users who may wish to transition from performing their flux gap-filling calculations in spreadsheet-software, towards using R for their computational flux gap-filling needs.

1.2.2. Case-study

A case-study has been included in this paper to assess the EC CO_2 flux gap-filling methodology applied in 'miniRECgap' vs. selected other approaches (i.e. vs. MDS - group 3, Table 1; ANN - group 2, Table 1) using crude metrics of model performance evaluation. Given that the main focus of this paper was to introduce the 'miniRECgap' package to the audience, the assessment based on crude in-sample metrics was assumed to be sufficient for this particular study, while in-depth comprehensive model evaluation is recommended for potential future work (further explanation provided in section 2.4). The study employs a dataset from a former peat-extraction site currently undergoing rehabilitation in Ireland (Cavemount Bog), as an example of such an ecosystem, to assess and critically evaluate our simple in-sample gap-filling approach (applied with 'miniRECgap' vs. the more advanced MDS approach, and the selected ML (shallow ANN) approach). The aim was to assess how these different gap-filling approaches may impact estimation of CO_2 emissions/removals for the given site and consequently:

- Develop our understanding as to whether the given site may act as a C-sink or C-source, which is critically important for GHG inventory reporting.
- Gain insights into directing potential future studies investigating the C-dynamics and the main drivers of CO_2 fluxes in relation to peatland rehabilitation and restoration activities and management.

For comparison purposes in terms of gap-filling, this study also employs an example from a natural/less-disturbed and well-known ecosystem dataset (i.e. the so-called DE-Tha dataset from a German forest site used as a template for data-inputs in the 'REddyProc' package by Wutzler et al. (2018) – further details are provided in section 2.2). Considering that both chosen ecosystems (Cavemount Bog and DE-Tha forest) are very different in terms of vegetation, soil type, and other characteristics, it should be noted that this comparison is not intended to assess/compare the GHG emissions/removals from these two ecosystems (which is outside the scope of this study). The noted comparison between the selected two ecosystem flux datasets in this study is entirely intended for the purpose of crude in-sample assessment of the performance of gap-filling techniques ('miniRECgap' and 'MDSnight'; shallow ANN was not applied at DE-Tha). For easier reporting of flux gap-filling findings in this study, the Cavemount Bog dataset is labelled as a 'challenging' ecosystem dataset and the DE-Tha dataset is labelled as a 'classic' ecosystem dataset. As such, within the context of this study, the term 'classic' ecosystem is used for well established, less-disturbed ecosystems and for which it can be expected that the popular standard gap-filling approaches will work well. Contrary to this, the term 'challenging' is used for ecosystems that are expected to pose a challenge for some popular standard gap-filling approaches for various reasons. These ecosystems include disturbed/formerly disturbed, less established and heterogeneous ecosystems, although the challenges in gap-filling may occur for other reasons as well.

2. Methods

2.1. Flux gap-filing and flux-partitioning methodologies applied in 'miniRECgap'

2.1.1. Basic concepts

This study uses the classic/traditional robust and validated empirical/semi-empirical modelling approach for filling gaps in EC CO_2 flux data using a minimum number of input-variables (i.e. it requires time-series data for only three main input-variables; Table 2). The approach is based on the application of environmental response functions in combination with empirical/semi-empirical parameter-optimisation (i.e. methods classified under Group 2 in Table 1). To understand the gap-filling methods used in this study, the following basic terrestrial ecosystem C-cycle components and concepts need to be explained: NEE, gross primary productivity (GPP) and ecosystem respiration (Reco). NEE refers to the net CO_2 exchange between a terrestrial ecosystem and the atmosphere (Law et al., 2006), which is also obtained from EC flux measurements (Reichle, 2019) and is calculated as follows:

$$\mathrm{NEE} = \mathrm{GPP} + \mathrm{Reco} \qquad \text{Eq. 1}$$

where GPP is defined as the amount of plant-assimilated CO_2 resulting from photosynthesis (CO_2 input from photosynthesis); Reco refers to ecosystem respiration (CO_2 output from respiration) and includes both autotrophic and heterotrophic respiration (Baldocchi and Valentini, 2004; MPI, 2024c; Reichle, 2019). The units in Eq. (1) used in this study are in μ mol CO_2-C m^{-2} s^{-1}. The NEE fluxes obtained from the EC

Table 2
Main input-variables in the 'miniRECgap' and 'MDSnight' approaches.

Input-variable	Units	'miniRECgap' R-package [a, b, d] (Premrov, 2024)	'MDSnight' via 'REddyProc' webtool/R-package [a, b] (Wutzler et al., 2018)
Year	/	Yes - included in ['DateTime']	Yes ['Year']
Day of year	/	Yes ['DOY']	Yes ['DoY']
Hour	/	Yes - included in ['DateTime']	Yes ['Hour']
Net ecosystem exchange	[μmol C m^{-2} s^{-1}]	Yes ['NEE']	Yes ['NEE']
Photosynthetic photon flux density	[μmol m^{-2} s^{-1}]	Yes ['PPFD']	No [c]
Soil temperature	[^{0}C]	No [e]	Yes ['Tsoil']
Air temperature	[^{0}C]	Yes [e] ['T']	Yes ['Tair']
Latent heat flux	[W m^{-2}]	No	Yes ['LE']
Sensible heat flux	[W m^{-2}]	No	Yes ['H']
Incoming solar radiation	[W m^{-2}]	No	Yes ['Rg']
Relative humidity	[%]	No	Yes ['rH']
Vapor pressure deficit	[hPa]	No	Yes* ['VPD'] or calculated if missing
Friction velocity (or u*)	[ms^{-1}]	No	Yes** ['Ustar'] required if case of u* filtering

NOTE:
Yes* denotes that the package uses an estimated value if the variable is missing.
Yes** denotes that gap-filling without u* filtering can also be performed if the variable is missing.
[a] Variable abbreviation/name used in the package (if applicable) is provided in brackets.
[b] 'Yes/No' denotes that the variable is used/not used in the package.
[c] In this study, PPFD was used to estimate 'Rg', which is a required input in the 'REddyProc' webtool/R-package; the methodology is explained in section 2.1.2 and Supplemental Material.
[d] When 'miniRECgap' is used, a value of '-9999' is not to be used for any missing values.
[e] Air temperature was used in Eq. (3) in 'miniRECgap' (see section 2.1.3); in this study, the application of soil temperature in Eq. (3) was also tested, but gap-filling performance was inferior to air temperature.

measurements (usually at half-hourly intervals) can be either positive or negative. This study uses the standard sign-convention, where a negative NEE value refers to a flux that moves from the atmosphere into the ecosystem (i.e. ecosystem functions as a C sink) and a positive NEE value refers to a flux that moves from the ecosystem to the atmosphere (i.e. ecosystem function as a C source) (Gogo and Laurent, 2023; MPI, 2024c). Therefore, in the case of a negative NEE, Eq. (1) can be also expressed as net uptake - NEE = GPP - Reco (MPI, 2024c).

2.1.2. Flux data-partitioning into nighttime and daytime

The flux data are first partitioned into nighttime and daytime. It is well known, from terrestrial ecosystem CO_2 diurnal dynamics, that CO_2 plant-uptake (GPP) occurs during the day via photosynthesis, which is strongly affected by photosynthetic active radiation (PAR) in addition to some other factors (such as temperature, vapor pressure deficit, leaf area index) (Falge et al., 2001). The PAR that is available to plants is often measured in the form of photosynthetic photon flux density or PPFD μ mol [quanta] m^{-2} s^{-1} and can be a part of EC flux variables, depending on whether the specific site is equipped with specialised instrumentation, for example the LI-190R Quantum Sensor [LI-COR®], (LI-COR Environmental, 2025). In this study, data-partitioning was carried out by using an incoming solar radiation (Rg) threshold value of 10 W m^{-2} according to Wutzler et al. (2018), converted to PPFD with a factor of 2.02 (reported in dos Reis and Ribeiro, 2020), which resulted in a threshold PPFD value of 20.2 μ mol m^{-2} s^{-1} as the chosen threshold value between night and day, i.e. obtaining nighttime data by removing the data where:

$$\text{Daytime PPFD} > 20.2\ [\mu\ \text{mol m}^{-2}\ \text{s}^{-1}] \qquad \text{Eq. 2}$$

Nighttime fluxes are further screened to remove any negative values, as in the absence of photosynthesis all nighttime CO_2 values should be positive. The obtained nighttime flux is assumed to represent Reco, i.e. nighttime efflux of CO_2 from autotrophic and heterotrophic respiration.

2.1.3. Modelling Reco using a temperature-response function

Modelling of the nighttime CO_2 efflux (Reco) using the temperature-response function is done according to the modified Arrhenius equation adapted from Lloyd and Taylor (1994):

$$\text{Reco} = \text{R10} \cdot e^{[\text{E0}\cdot\ (1/(283.2\ -\text{To})\ -\ 1/(\text{T}-\text{To}))]} \qquad \text{Eq. 3}$$

where R10 is the respiration rate at 10 °C [μmol CO_2 m^{-2} s^{-1}], E0 is the activation energy (309 K), temperature T0 = 230 K, and T is the measured temperature converted to K (T°C +273.2).

Initial estimates of Reco (pre-modelled Reco) values are obtained by using an arbitrary R10 value in Eq. (3) (i.e. the arbitrary value of 4 was used in this study; further details are provided in supporting material SM4, Step 2 and Step 4), followed by the optimisation of R10 in the above model by minimising the residual sum of squares from initial (pre-modelled) and measured (nighttime) Reco. If the data are processed in Excel, this optimisation is usually performed using Excel Solver, whereas the optimisation in the 'miniRECgap' package is done using the function 'optim' from R-package 'stats' (R Core Team, 2024a) by applying the 'L-BFGS-B' method/algorithm of Byrd et al. (1995). It should be noted that, while the function 'optim' allows users to choose between several different optimisation methods/algorithms, the 'L-BFGS-B' algorithm was included in 'miniRECgap' for this optimisation task. 'L-BFGS-B' was chosen based on prior comparison with Excel Solver[3] on a portion of EC flux data from another peatland, which indicated the suitability of this method. It is however recommended that additional comparisons are made as part of future studies on the possible further expansion of this package. For example, future work could involve the potential extension of specific functions in this package to include different optimisation algorithms, which would allow users to select the algorithm of their choice. The optimised R10 value was then used in Eq. (3) to calculate the final modelled Reco for the full dataset (nighttime and daytime).

Eq. (3) assumes that Reco depends on a single variable, temperature (T). In reality, Reco may also depend on other factors, for example, in a study on a temperate peatland, Juszczak et al. (2013) reported a positive correlation between Reco and water table depth for some of the microsites. However, it should be noted that in this study, Eq. (3) (conventional temperature-response function) has been used deliberately to enable the 'miniRECgap' package to operate with the minimum number of input variables. As such, potential 'miniRECgap' limitations in terms of accuracy and precision, resulting solely from the temperature variable to estimate Reco, need to be recognised.

2.1.4. Modelling GPP using a light-response function

The half-hourly flux data (including modelled Reco) are partitioned into daytime by removing values below the PPFD threshold (Eq. (2); section 2.2). The daytime fluxes are further screened to remove any positive NEE values. The daytime dataset is then used to calculated GPP from measured NEE and modelled Reco according to Eq. (1) (this refers to calculated GPP). The obtained GPP is next multiplied by −1 to switch the sign convention. The modelled GPP is obtained next by using the light-response function reported in Gilmanov et al. (2003), which is adapted from the non-rectangular hyperbola model by Rabinowitch (1951):

$$\text{GPP} = ((\alpha \cdot \text{PPFD} + \text{A}_{\text{max}}) - (((\alpha \cdot \text{PPFD} + \text{A}_{\text{max}})^2)^{0.5} - ((4 \cdot \gamma) \cdot (\text{PPFD} \cdot \alpha \cdot \text{A}_{\text{max}})))) \,/\, (2 \cdot \gamma) \qquad \text{Eq. 4}$$

where α refers to the quantum yield based on incident irradiance [mol CO_2 (mol photon)$^{-1}$], γ is the coefficient of convexity, PPFD is the incident photosynthetic photon flux density, and A_{max} is the maximum rate of CO_2 assimilation [μmol CO_2 m^{-2} s^{-1}].

Initial estimates of GPP (pre-modelled GPP) values are obtained by setting the model coefficient parameters (α, A_{max} and γ) from Eq. (4) to initial values (the following initial values for model coefficients were used in this study: α = 0.08; A_{max} = 15; γ = 0.5; further details are provided in Supplemental Material SM4, Step 2 and Step 4). This was followed by optimisation of the three model coefficient parameters (α, A_{max} and γ) by minimising the residual sum of squares from the initial estimates of GPP (pre-modelled GPP). If the data are processed in Excel, this optimisation is usually performed with Excel Solver. Optimisation in the 'miniRECgap' package was again carried out with the function 'optim' from R-package 'stats'(R Core Team, 2024a) by applying the 'BFGS' algorithm/method (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970), similar to procedure outlined previously (section 2.1.3).[4] As explained in section 2.1.3, further testing of different optimisation algorithms that could be applied in 'miniRECgap', for this optimisation task, is recommended as a part of potential future studies on further expansion of this package. The optimised model coefficient parameters (α, A_{max} and γ) were then used in Eq. (4) to calculate the final estimated GPP (modelled GPP). It should be noted that these modelled GPP values are not yet processed at this stage. Processing of modelled GPP values is explained in section 2.1.5.

2.1.5. Modelling NEE, gap-filling, and potential for further package development

The obtained modelled GPP values (section 2.1.4) are used to calculate modelled NEE. Prior to this step, however, the modelled GPP values had to be first processed. This involved filtering to remove any positive modelled GPP values that occurred during nighttime conditions

[2] Python. The Python Software Foundation. URL: https://www.python.org/.

[3] The difference between two outputs from L-BFGS-B′ and Excel Solver was minimal and appeared only at/after the sixth decimal of the output value.

[4] The 'BFGS' output was again very similar to output from Excel Solver.

(using the PPFD threshold value, Eq. (2)), and then multiplying the values by −1 to convert the data back into the EC sign convention. The modelled NEE values were next calculated according to Eq. (1) using the previously modelled Reco (section 2.1.3) and processed modelled GPP. The modelled NEE values were then used to infill any missing values/gaps in the measured NEE dataset.

Here, the presented 'miniRECgap' R package utilises classic functions in Eqs. (3) and (4), which are well-known as some of the most common nonlinear functions for modelling Reco and GPP.

As monotonic functions (such as Eqs. (3) and (4)) may not always accurately capture the temperature-/light-responses of real-world conditions in different ecosystems (Chen et al., 2023; Meng et al., 2024), modified versions of such functions have been studied and reported in the literature. For example, studies across many terrestrial ecosystems that have utilised FLUXNET data have indicated a unimodal relationship between Reco and temperature (parabolic curves; Chen et al., 2023), as well as between NEE and temperature (Meng et al., 2024). Furthermore, Ye (2007) developed a modified light-response function that accounted for photo-inhibition that was, in turn, utilised by Jia et al. (2014) for daytime NEE modelling. Here, our introduced version of 'miniRECgap' is deliberately limited to the classic functions (Eqs. (3) and (4)) to remain simple and user-friendly. As such, potential future work on 'miniRECgap' could involve the incorporation of additional modelling options for Reco and GPP, to enable potential use of this package under different circumstances.

2.2. Case-study: site description and data

This study employs two examples of EC flux measurements data:

(a) A 'classic' (less-disturbed) forest ecosystem known as DE-Tha dataset (MPI, 2024d), which was used in this study mainly for comparison reasons in terms of gap-filling.
(b) A 'challenging' ecosystem dataset from Cavemount Bog (original to this study), which is a former cutaway extraction peatland site undergoing rehabilitation (Fig. 1a, b, c).

These ecosystem examples are used to test the gap-filling approaches, i.e. 'miniRECgap' approach vs. selected MDS and ML (shallow ANN) approaches. It should be noted that the labels 'classic' ecosystem and 'challenging' ecosystem were introduced for the purpose of easier reporting of findings in this study, and do not represent specifically defined ecosystem classes/categories, as they are solely based on whether the dataset from the given site was expected to potentially pose

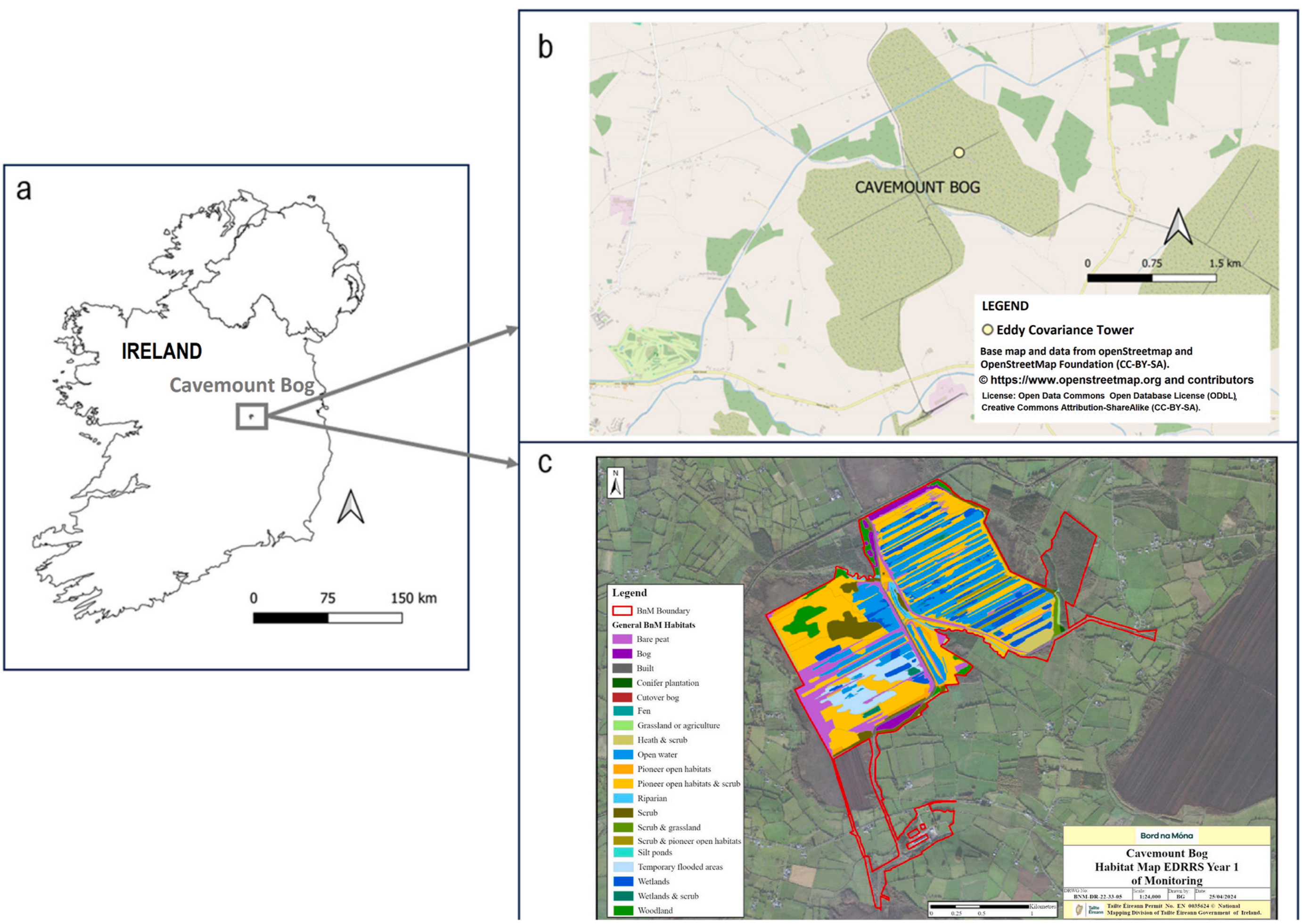

Fig. 1. Cavemount Bog: (a) Schematic presentation of Cavemount Bog location, (b) zoomed-in Cavemount Bog with indicated location of the installed Eddy Covariance tower, (c) Cavemount Habitat Map 2022. Prepared by Bord na Móna (BnM), Ireland.
[NOTE – Cavemount Bog coordinates: 53.307954° N, −7.240437° W (WGS84) (Bord na Móna, 2021). Attributions: Fig. 1a - Ireland map boundary: GADM data (version 4.1). Global Administrative Areas (GADM). © 2018–2022 GADM. URL: https://gadm.org/(GADM, 2018). Fig. 1b - Base map and data from OpenStreetMap and OpenStreetMap Foundation (CC-BY-SA). © https://www.openstreetmap.org and contributors. (OpenStreetMap & Contributors, 2024, Year accessed 2024). Fig. 1c - Cavemount Habitat Map 2022, prepared by Bord na Móna (BnM), Ireland. Red line refers to site boundary by BnM.]. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

a challenge for the selected gap-filling approaches used in this study.

The DE-Tha ecosystem dataset is well-known and has been reported in several studies (e.g. Grünwald and Bernhofer (2007), MPI (2024d) and Wutzler et al. (2018)). This dataset contains one year of half-hourly EC measurements from the Tharandt coniferous forest site in Germany, from the start of January 1998 to the end of December 1998 (MPI, 2024d). Further details on this site are outlined in Grünwald and Bernhofer (2007), MPI (2024d) and Wutzler et al. (2018). The dataset file ('Example_DETha98 EC') can be downloaded from the Max Planck Institute 'REddyProc' webtool input format website (MPI, 2024d), where this dataset is provided as an example of the data-input template format used in the 'REddyProc' webtool/R-package.

Cavemount Bog is located in Co. Offaly, Ireland [latitude 53.307954° N, longitude −7.240437° W (WGS84); elevation *ca.* 74m], with the Esker River flowing through its centre (Bord na Móna, 2021). A full description of this site is provided in Bord na Móna (2021). In brief, Cavemount Bog refers to a cutaway peatland site where peat extraction occurred from 1970 to 2015. The site is actively undergoing rehabilitation, which has involved drain blocking to increase the height of the water-table. This has resulted in the formation of different vegetated areas via natural colonisation. The site also has areas of bare (exposed) peat, and open water areas (Bord na Móna, 2021), as evident from Fig. 1c. Therefore, due to its highly heterogeneous conditions (i.e. mosaic of vegetated, bare peat and open water areas), Cavemount Bog was chosen as the "challenging" ecosystem in this study. The Cavemount dataset used in this study contains (among others) NEE data from on-site EC tower measurements at the eastern side of this site, which was installed under the former SmartBOG (2020) project (further details are outlined in Bord na Móna (2021)). This study used *c.* 8 months of EC flux measurements (from *c.* February 2022 to October 2022) from the Cavemount site, provided at half-hourly intervals. Therefore, the sums of gap-filled NEE values from both sites are therefore not directly comparable due to the shorter period of observation at Cavemount Bog compared to DE-Tha (results are reported in later section 4.1.2.2, Tables 9a and 9b). The dataset did not contain the associated meteorological data, which were obtained from Clara Bog, an adjacent peatland site located in Co. Offaly (further information on Clara Bog can be found in Ingle et al. (2023)). Prior to flux gap-filling, the raw EC flux data were pre-processed and QC/QA screened to remove poor quality data. The same methodology for pre-processing/QC/QA screening was used as described in Ingle et al. (2023).

It should be noted that, because this study mainly focuses on introducing the 'miniRECgap' package to the audience, it employs only two selected examples of EC flux measurements data from sites in Ireland and Germany. Therefore, future studies that would include additional sites from various terrestrial ecosystems and different climatic zones, is strongly recommended. Future testing could include datasets from ecosystems that could be considered as 'challenging' due to other reasons, such as dryland ecosystems and deserts, which may potentially represent another type of challenge in modelling studies due to occurrence of pulsed dynamics in precipitation events. Examples of dryland ecosystem and desert datasets can be found reported in the literature [e.g. Cong et al. (2023), Biederman et al. (2018, 2017)].

2.3. Flux gap-filling approaches and crude in-sample gap-filling performance evaluation applied in case-study

Crude evaluation of the gap-filling performance in this study was based on the NEE modelling performance, where the modelled NEE values were used to fill in the missing NEE data in the dataset.

2.3.1. Data preparation and processing

Both datasets had to be formatted/prepared to be suitable as inputs in the chosen R-package/s. The data-input template for 'miniRECgap' is provided as an example in the Supplemental Material ["SM_miniRECgap_input_template.csv"; further explanation can be found in section 3.2.1]. Gap-filling was carried out on:

- full datasets (without assigning the data into groups); and
- datasets that were assigned into separate groups/seasons.

This was undertaken to check whether the 'miniRECgap' gap-filling performance could be improved after the data was assigned into groups/seasons. Separation of the data into groups/seasons had to be done prior to using 'miniRECgap', because this package does not have a built-in function for this purpose. The dataset from the Cavemount site proved to be relatively challenging for gap-filling. Simple separation of data on a monthly/seasonal-basis did not improve the gap-filling performance for this site (results not presented). Therefore, a simple unsupervised ML clustering approach was applied to apportion the data into separate groups/clusters. This was carried out for both datasets. The k-means clustering approach was utilised, where clustering is based on the internal similarity within each cluster (Boehmke, 2018b; Ryan, 2019), which was performed using NEE and air temperature data, by applying the algorithm presented in Hartigan and Wong (1979) via 'cluster' R-package (Maechler et al., 2022). Details are provided in Supplemental Material SM1. The procedure resulted in an optimal number of three clusters.

2.3.2. Gap-filling with 'miniRECgap'

Gap-filling was first performed on the half-hourly EC flux datasets using the 'miniRECgap' package by applying the classic nonlinear environmental response functions to model the NEE values, as outlined in section 2.1. While 'miniRECgap' utilises a number of functions (Fig. 2a), the package is supported by GUI R-scripts that allow for gap-filling to be performed in five simple steps (Fig. 2b), and using only five lines of code. This makes it suitable for users with different levels of R-programming skills, including those who are new to the R environment.

The five steps are explained in section 3 and Supplemental Material SM5, which provide a very thorough explanation on how to use the 'miniRECgap' package, ranging from instructions on how to prepare the inputs, install and load the software, to specific instructions on how to use the package via the GUI supported scripts, including information on which calculations (equations from section 2.2) are performed at each step. In this study the NEE gap-filling with 'miniRECgap' was initially carried out on full datasets, as well as on the data assigned into three clusters (using k-means clustering approach), for both DE-Tha and Cavemount. Table 2 shows the input-variables for the 'miniRECgap' (as well as 'MDSnight', see section 2.3.3) used in this study.

2.3.3. Gap-filling with 'MDSnight' approach

In order to evaluate the simple technique applied in 'miniRECgap', a gap-filling procedure was also performed using the popular 'REddyProc' webtool [based on the R-package by Wutzler et al. (2018)]. 'REddyProc' was chosen because it is referred within the EC flux community as one of the standard approaches for EC flux data processing, with several studies showing better performance of this package compared to some other methods (Boudhina et al., 2018; Wang et al., 2023; Wutzler et al., 2018). The 'REddyProc' package includes a comprehensive set of tools for processing, quality control, and gap-filling of EC data, incorporating advanced statistical algorithms and modelling techniques (Wutzler et al., 2018), which are not built into 'miniRECgap'. The full datasets from both sites were used as inputs into the 'REddyProc' webtool. In accordance with this paper, hereafter, 'MDSnight' was used to refer to the gap-filling approach implemented in this study using the 'REddyProc' webtool, in order to avoid confusion with the actual 'REddyProc' package/software, which can support different settings.

Detailed explanation on gap-filling with this approach is provided in the Supplemental Material SM2. In brief, the gap-filling performed via 'REddyProc' webtool allowed for modelling and gap-filling the NEE data via the marginal distribution sampling (MDS) algorithm presented in

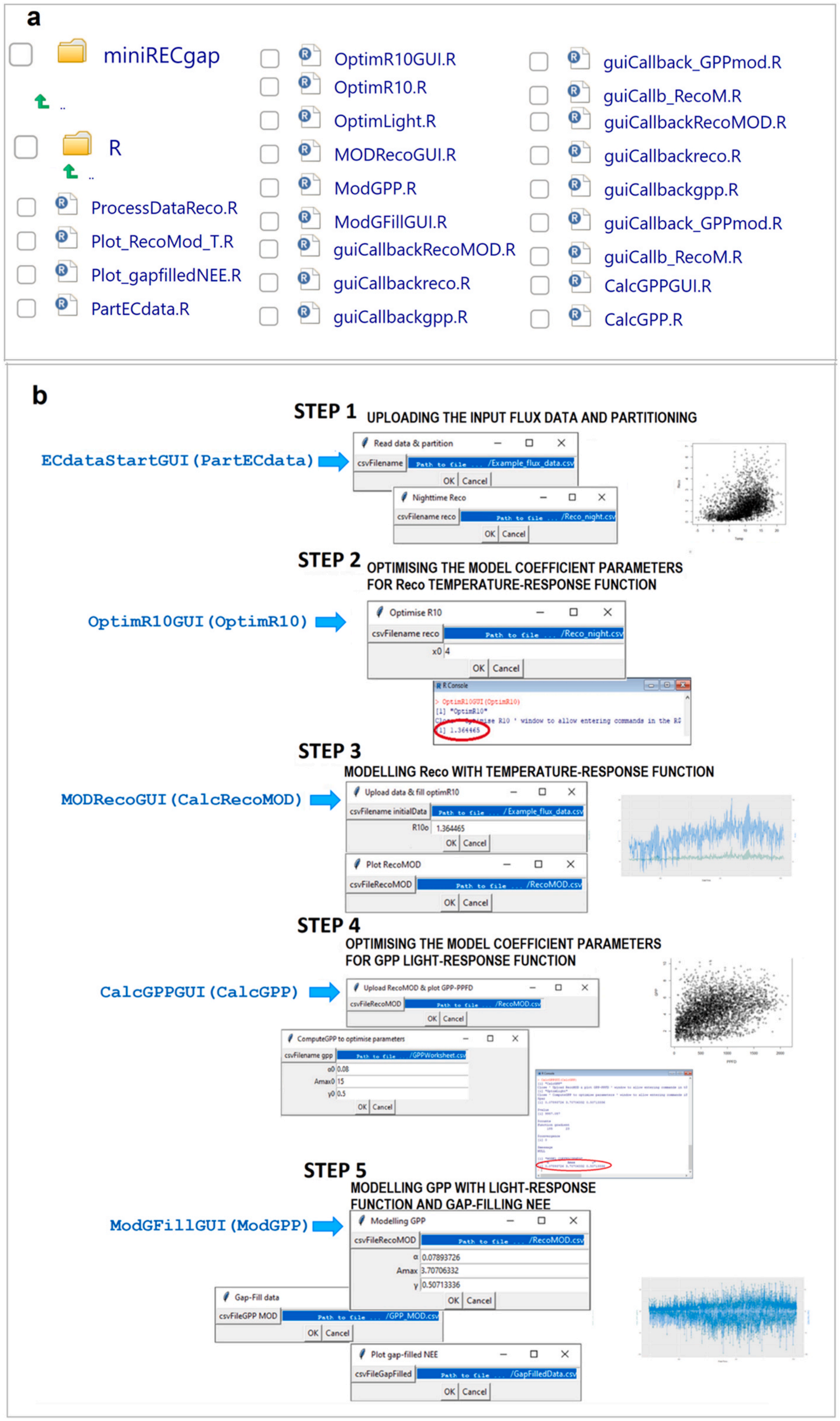

Fig. 2. Illustrative presentation of the 'miniRECgap' R-package: (a) number of functions/R-scripts inside the 'miniRECgap' package source-code, and (b) schematic presentation of five simple operating steps of the 'miniRECgap' run using only five lines of code to complete the gap-filling process. Source: Fig. 2a 'miniRECgap' (Premrov, 2024). NOTE: Schematic presentation in Fig. 2b is provided for illustrative purpose only. For additional information and improved visual clarity, the readers should refer to detailed description of the five steps in Section 3 and in Supplemental Material SM5, accompanied with larger print-screens in separate figures (Figs. SM2 to SM11).

Reichstein et al. (2005), which is the gap-filling method implemented in 'REddyProc' (Wutzler et al., 2018), followed by nighttime net flux partitioning into GPP and Reco, also according to Reichstein et al. (2005). Gap-filling was carried out both with and without ustar (u*) filtering (Wutzler et al., 2018). Used was a default u* filtering method ("UStar Threshold estimation: Moving Point Test …") available in an online tool (MPI, 2024e). The u* filtering refers to the technique that removes the EC flux data/measurements recorded during unsuitable

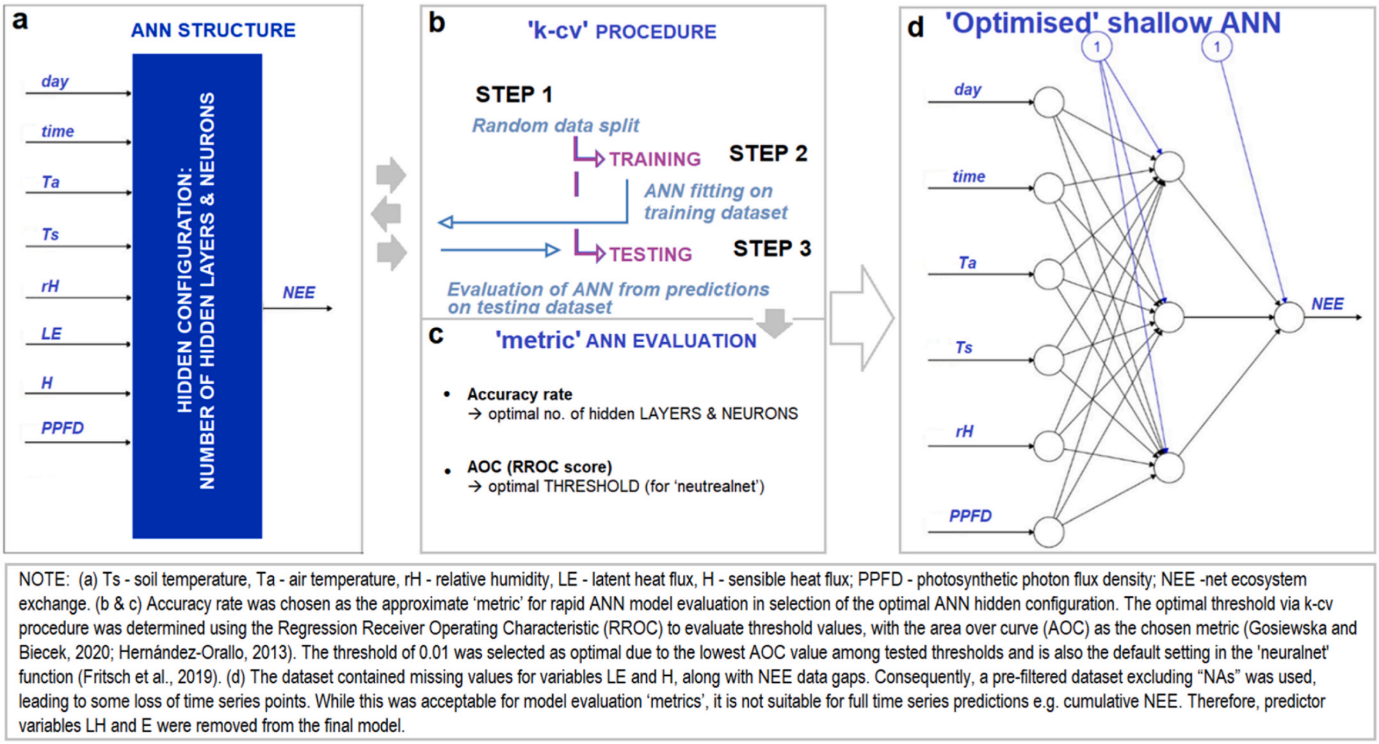

Fig. 3. Schematic presentation of hyperparameter-optimisation for a shallow ANN: (a) Initial input and output variables; (b) employed 'k-cv' procedure by repeating k-times the tree outlined steps (STEP 1, STEP 2, STEP 3); (c) 'metric' for quick evaluation of ANN as a part of optimisation procedure; (d) final 'optimised' shallow ANN.

conditions with insufficient turbulence (i.e. low friction velocity) (MPI, 2024f). In order to be able to use the 'MDSnight' approach, the Rg inputs (Table 2) were estimated from measured PPFD using a conversion factor of 2.02 (reported in dos Reis and Ribeiro, 2020), (see section 2.1.2). The extent of introduced uncertainty resulting from the estimation of Rg alone was not assessed in this study due to the lack of measured Rg data.

2.3.4. Gap-filling with shallow ANN approach applied to the 'challenging' ecosystem dataset

Given that the 'miniRECgap' package was designed to be as simple as possible, it was initially assumed that the 'MDSnight' gap-filling approach would likely outperform 'miniRECgap' for both sites used in this study. However, our findings showed that gap-filling of the Cavemount dataset was quite difficult with both approaches (see results in section 4) and therefore, an additional Artificial Neural Networks (ANN) ML-based gap-filling approach was applied solely at Cavemount. This study is limited to application of relatively shallow ANNs (i.e. ANNs with only one or two hidden layers and small number of neurons). The approach involved the following workflow, which is also summarised in Fig. 3:

1. Choice of model-inputs and ANN settings for continuous data (Fig. 3a; further details are provided in Supplemental Material SM3.1).
2. Optimisation of the shallow ANN hyperparameters (Fig. 3b and c; further details are provided in Supplemental Material SM3.2).
3. Choice of the training/testing ratio (part of Fig. 3b and c; further details are provided in Supplemental Material SM3.3).
4. Obtaining the final model for gap-filling the missing NEE values (Fig. 3d; further details are provided in Supplemental Material SM3.4).

Details on the employed shallow ANN approach are outlined in the Supplemental Material SM3. In brief, the application of shallow ANN in this study was done via the 'neuralnet' R-package using the 'neuralnet' function (Fritsch et al., 2019) and the settings are explained in SM3.1. The modelling was performed on the pre-scaled/normalised Cavemount dataset. The NEE variable was modelled from predictor variables DOY, Hour, Tair, Tsoil, rH, LE, H and PPFD (Rg was excluded, because Rg was not measured in the Cavemount dataset and had to be estimated from PPFD due to lack of measured data). These eight predictor variables were initially chosen as potentially suitable for ANN modelling, because most were used in the gap-filling methods reported in Table 2. The optimisation of ANN hyperparameters (SM3.2) as well as the choice of training/testing ratio (SM3.3) were carried out via an adapted 'fast' k-cross validation (k-cv) approach by calculating selected and simple 'metrics' for ML model evaluation (explained in detail in SM3.2 and SM3.3). It should be noted that employed optimisation procedures were limited to a relatively low number of repetitions and to the use of selected simple 'metrics' for ML model evaluation. Therefore, findings obtained from using the simplified ANN assessment procedures in this study need to be considered with a degree of caution, taking into consideration their indicative nature.

The optimisation procedures allowed for the choice of the final ANN model, which consisted of a single hidden layer and three neurons (hereafter referred to as the 'N3' model, Fig. 4), with settings for linear output and a threshold of 0.01 (explained in detail in SM3.4). The model was applied using the two training/testing random data-splits: 30:70 and 70:30. It should be noted that the Cavemount dataset also had some missing values for predictor variables LE and H (in addition to NEE data-gaps). For this reason, the pre-filtered dataset excluding "NAs" had to be used, resulting in the loss of some points in the timeseries. While this was considered acceptable for the model evaluation 'metrics' used in the optimisation procedures, it is not an appropriate approach when

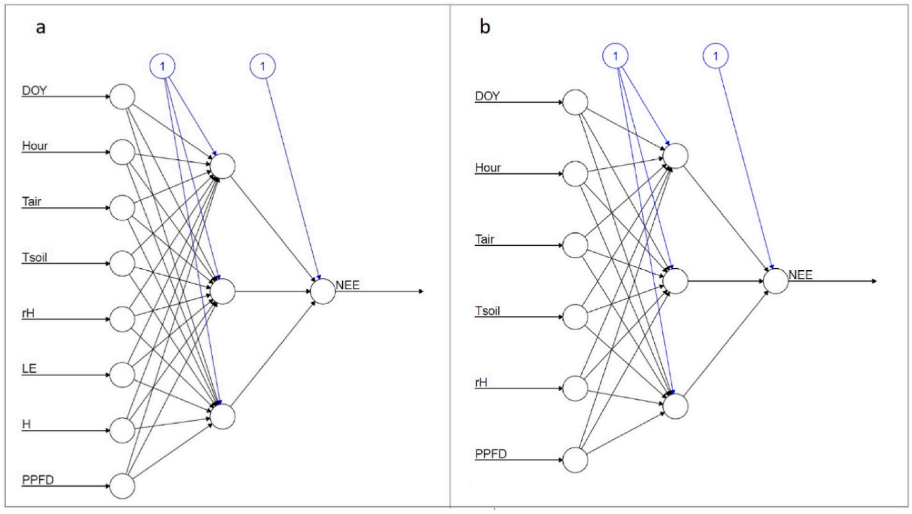

Fig. 4. Schematic Artificial Neural Network (ANN) presentation of the 'N3' model architecture: (a) model with eight predictor input-variables (DOY, Hour, Tair, Tsoil, rH, LE, H and PPFD), a single hidden layer with three neurons (j = 3), and one output (NEE), and (b) model with six predictor input-variables (DOY, Hour, Tair, Tsoil, rH and PPFD), a single hidden layer with 3 neurons (j = 3) and one output (NEE).

predictions over a full timeseries are needed. Therefore, the predictor variables LH and E were dropped from the 'N3' model, resulting in the use of only six predictor input-variables (instead of the initial eight). As such, the NEE sum and cumulative NEE could be calculated only on the gap-filled dataset using the model with six predictor input-variables. The model performance was evaluated using multiple model prediction indices (explained in section 2.4) derived from the 'N3' with eight predictors (Fig. 4a) and with six predictors (final 'N3' model, Fig. 4b) applied to the joined (training + testing) dataset (see results in section 4.1.1.2; Table 7b). In addition, the model 'N3'(Fig. 4a and b) for both eight and six predictor input-variables, was also evaluated separately for training and testing datasets, using multiple model prediction indices (see results in section 4.1.2, Table 8).

Literature sources cited in the NOTE to Fig. 3: Fritsch et al. (2019); Gosiewska and Biecek (2020); Hernández-Orallo (2013).

2.4. Crude evaluation of gap-filling performance applied in case-study

The applied crude model-performance evaluation in this study is intended to serve solely as an illustrative example, to demonstrate the application of 'miniRECgap' vs. two other selected gap-filling techniques, using a case-study approach. The evaluation of the gap-filling performance for the 'miniRECgap', 'MDSnight' and optimised shallow ANN ('N3' model) approaches used in this study was conducted using the selected multiple model prediction indices and estimates of uncertainty presented in Table 3, which mainly refer to in-sample metrics calculated from the same dataset used to either train or develop the model, unless specified otherwise. Some well-known issues and limitations of such in-sample model performance metrics are for example model overfitting, and lack of information on how the model would perform on unseen data. As such, it should be noted that this is a very crude evaluation approach, which should not be generalised, and that

Table 3
(a) Selected model prediction indices and (b) estimation of uncertainty used for crude evaluation of flux gap-filling performances.

a) Model prediction indices (equation)	**Eq. no.**	**Description**
$R^2 = [\Sigma\,(P_i - P_{am}) \cdot (O_i - O_{am})]^2 / [\Sigma\,(P_i - P_{am})^2 \cdot \Sigma\,(O_i - O_{am})^2]$	[Eq. 5]	R^2 refers to coefficient of determination, which explains the proportion of variance that is accounted for by the model (in this case gap-filling technique), P_i refers to the predicted and O_i to the observed flux at each half-hourly time-interval (i) and subscript 'am' refers to the arithmetic mean of each variable (Lucas-Moffat et al., 2022; Mahabbati et al., 2021).
$RMSE = [\Sigma\,(P_i - O_i)^2/(N_p)]^{0.5}$	[Eq. 6a]	The model prediction indices used as indicators of the magnitude of the individual errors: RMSE refers to the root mean squared error (Moffat et al., 2007); RMSE* sometimes named as the root mean squared deviation (RMSD) [in some literature it is reported as RMSE] (Mahabbati et al., 2021; Piñeiro et al., 2008); rRMSE refers to the relative RMSE (Moffat et al., 2007); in Eqs. (6a) and (6b) the Np refers to number of predicted half-hourly fluxes (Lucas-Moffat et al., 2022) for which the observed values are available.
$RMSE^* = [\Sigma\,(P_i - O_i)^2/(N_p - 1)]^{0.5}$	[Eq. 6b]	
$rRMSE = [\Sigma\,(P_i - O_i)^2/\Sigma\,(O_i)^2]^{0.5}$	[Eq. 6c]	
$StD = 2^{\circ.5}/N_p \cdot \Sigma_{Np}\,\lvert P_i - O_i \rvert$	[Eq. 7a]	StD refers to standard deviation calculated from mean absolute error (MAE) calculated as $1/N_p \cdot \Sigma_{Np}\,\lvert P_i - O_i \rvert$ (Lucas-Moffat et al., 2022; Moffat et al., 2007).
$MAE = 1/N_p \cdot \Sigma_{Np}\,\lvert P_i - O_i \rvert$	[Eq. 7b]	
b) Estimation of uncertainty (equation)	**Eq. no.**	**Description**
$Es = 1/N_p \cdot \Sigma_{Np}\,(P_i - O_i)$	[Eq. 8]	Es refers to bias or systematic error estimated from model residuals (P_i-O_i) (Lucas-Moffat et al., 2022; Moffat et al., 2007).
$Er = \{\,\Sigma_{Np}\,[\,(P_i - O_i)^2/((N_p - 1) \cdot N_p)\,]\,\}^{0.5}$	[Eq. 9]	Er refers to random error (Aurela et al., 2002; Lloyd, 2010), which can potentially include random errors due to statistical uncertainties of EC method, varying footprint, as well as NEE gap-filling (Aurela et al., 2002).
$E = (Er^2 + Es^2)^{0.5}$	[Eq. 10]	E refers to an estimate of the total uncertainty (Lloyd, 2010; Lucas-Moffat et al., 2022), obtained from previously calculated random error Er (Eq. (9)) and systematic error Es (Eq. (8)) using equation from (Lloyd, 2010)

NOTE: All of the model prediction indices and estimates of uncertainties were computed based on the observed and predicted half-hourly NEE data without further conversion of the units (i.e. this study used the same NEE units in which the observed/measured NEE were originally reported [μ mol CO_2-C m^{-2} s^{-1}]).

there are more advanced and in-depth evaluation approaches available in the literature (e.g. maximum likelihood used in Richardson et al. (2006), introducing the out-of-sample or various scenarios of artificial data-gaps of different lengths (e.g. Moffat et al. (2007)). Althoughn these approaches were not applied in this study (which is mainly focusing on introducing 'miniRECgap'), they could be considered in potential future work on various 'challenging' ecosystems.

The gap-filling model performance was first examined using simple regression analysis (i.e. comparison of predicted and observed NEE values), followed by computation of the remainder of the model prediction indices listed in Table 3. The procedure included plotting observed vs. predicted NEE values and the calculation of the coefficients of determination (R^2; Eq. (5)), the root mean squared prediction error (RMSE; Eq. (6a)) [or the root mean squared prediction deviation (RMSD*; Eq. (6b))], as well as the relative version of RMSE (rRMSPE; Eq. (6c)) and the standard deviation calculated from mean absolute error (StD, Eq. (7a) and MAE, Eq. (7b)) (Table 3a).

While there are different approaches to quantify the uncertainties in flux studies (Hollinger and Richardson, 2005), the term 'E' in this study refers to an estimate of total uncertainty obtained from systematic error or bias (Es) and random error (Er), as defined by equations Eq. (8), Eq. (9) and Eq. (10) (Table 3b) (Lloyd, 2010; Lucas-Moffat et al., 2022). As such, the variable E in this study should not be confused with the uncertainty that results entirely from the flux data, which occurs mainly due to random measurement errors. Furthermore, the random error (Er) can potentially include random errors due to the statistical uncertainties from the EC method, variable footprint, as well as NEE gap-filling, and is computed using the approach from Aurela et al. (2002), which differs somewhat from the equation used by Lucas-Moffat et al. (2022). Nevertheless, similar to Lucas-Moffat et al. (2022), the errors in this study were calculated from model residuals (Table 3b) without accounting for potential bias from measurements or for potential errors due to autocorrelation, and therefore, the estimate of total uncertainty (E) from this study should be considered as the lower range of this estimate. Further details on multiple model prediction indices and uncertainty estimation used in this study, with accompanying description and literature sources, are provided in Table 3, and accompanying notes to the table. The sum of gap-filled NEE expressed in units of [t C ha^{-1}] per total period of observation for each site, was also calculated for different gap-filling approaches.

3. Introducing the 'miniRECgap' R-package

3.1. About 'miniRECgap'

The 'miniRECgap' package is designed for the application of very basic functions for gap-filling of missing EC CO_2 flux measurements and for flux data-partitioning. It requires that the users understand the gap-filling procedure, as it is designed in five steps that need to be run in sequence, forcing the users to engage in the process. While the 'miniRECgap' package is primarily designed to be suitable for users who are less experienced in the R environment and beginners, the functions written in 'miniRECgap' may also serve as potentially useful examples for experienced R-users who may wish to write different types of functions in the R environment for some other flux data-partitioning and gap-filling methods of their own.

Before using the 'miniRECgap' package, a user will first need to prepare the input flux data in the required format, launch the R environment, set up the working directory, and install and load the 'miniRECgap' and other required R-packages. These procedures are explained in section 3.2.1 and the Supplemental Material SM4. After this, the user can proceed to section 3.2.2 and the Supplemental Material SM5, which explain how to use the new 'miniRECgap' package with the GUI supported scripts. The procedures in section 3.2.2 and Supplemental Material SM5, are explained in the form of steps, where at each step, the package generates intermediate outputs (when applicable), and not only the final output. These intermediate outputs are saved in the working directory. It is thought that this feature of the 'miniRECgap' package should be beneficial for users who may wish to explore the specific intermediate outputs and that it may also enable the users to gain a better understanding of how data processing and computing is performed, by linking to the methodologies explained in section 2.1. A more experienced user can opt to view the source code and the structure of specific functions on the GitHub portal (Premrov, 2024) where the experienced user can further explore how these functions link to the specific data processing and computing methodologies described in section 2.1. The individual 'miniRECgap' files, including the source code for individual functions, are available on the GitHub portal, where they can be viewed i.e. source: Premrov (2024) available under the 'MIT License + File License'. Details on how to reference the 'miniRECgap' package are provided in subsection 3.2.1 of this paper and Supplemental Material SM6.

3.2. Using 'miniRECgap'

3.2.1. Starting procedures

It is recommended that users who are completely new to the R environment first get familiar with the R user-interface and learn some basic concepts about R (i.e. Basic-R or R-Studio, depending on their preference/choice) from the numerous materials and manuals that are available in various sources and literature (e.g. Douglas, 2023; Douglas et al., 2023; FAO, 2022; Grolemund, 2014; Maindonald, 2008; Torfs and Brauer, 2014). Starting procedures refer to procedures on setting up the working directory, preparing the input flux data in required format, and installing and loading the 'miniRECgap' and other required R-packages. The data-format that is needed for input in the 'miniRECgap' package is provided in data-example template (see Supplemental Material "SM_miniRECgap_input_template.csv").[5] After preparation of the input data in the desired format, a user should follow the instructions on the remaining starting procedures explained in detail in the Supplemental Material SM4. In brief, a user should start by launching R, choosing the working directory, create a new script, and install the 'miniRECgap' R-package from GitHub. The code used for starting procedures involving the installation of 'miniRECgap' and other required packages, is summarised in Table 4. As outlined in Table 4d, a user will need to install and load several other R-packages that are needed to use 'miniRECgap' [i.e. packages 'stats', 'grDevices', utils' (R Core Team, 2024b), 'fgui' (Hoffmann and Laird, 2009), 'dplyr' (Wickham et al., 2023) and 'ggplot2' (Wickham, 2016)]. The code to install these packages (if not previously installed) is further explained in Supplemental Material SM4. It is the responsibility of each user to ensure that these packages are appropriately referenced and are used according to their license(s). Recommendations on retrieving the information needed for referencing the packages are provided in section 3.3 and SM6.

3.2.2. Five steps: employing 'miniRECgap' via GUI supported scripts

Use of the 'miniRECgap' R-package via GUI supported scripts can be done by running just the five lines of the code provided in Table 5 via five simple steps. For this reason, this option of employing the 'miniRECgap' package via GUI supported scripts is thought to be suitable for all users, including users who are less experienced in the R environment and beginners. The code that is provided in Table 5, needs to be run in sequence, following the procedures explained in the individual five steps (STEP 1 to STEP 5, Fig. 2b). This will enable utilisation of several

[5] The data-format that is needed for input in the 'miniRECgap' package is provided in example template Supplemental Material "SM_miniRECgap_input_template.csv", which is a fictional example.

The data for variables 'DOI', 'DateTime' 'NEE', 'PPFD' and 'T' included in the "SM_miniRECgap_input_template.csv" file are fictional (not real observations) provided entirely as an example for the given template. Accompanying information is provided in the "READ_ME_SM_miniRECgap_input_template.txt".

Table 4
Code used for starting procedures, installation of 'miniRECgap' and other required packages [details are explained in Supplemental Material SM4].

Task	Code
(a) Installation and loading of 'devtools' package	`install.packages('devtools')` `library('devtools')`
(b) Installation and loading of 'miniRECgap' package from GitHub https://github.com/APremrov/miniRECgap (Premrov, 2024)	`install_github('APremrov/miniRECgap')` `library('miniRECgap')`
(c) Coercing R object to an environment.	`as.environment('package:miniRECgap')`

(d) Installation and loading of the other required R-packages that are needed to use the 'miniRECgap'	`library('stats')`
	`library('utils')`
	`library('fgui')`
	`library('dplyr')`
	`library('ggplot2')`
	`library('grDevices')`

NOTE: (a) 'devtools' refers to package by Wickham et al. (2022).

Table 5
Code used for starting procedures, installation of 'miniRECgap' and other required packages [details are explained in Supplemental Material SM4].

Task	Code

Performing flux gap-filling procedures with 'miniRECgap' in five steps (running GUI supported scripts)	`ECdataStartGUI(PartECdata)`
	`OptimR10GUI(OptimR10)`
	`MODRecoGUI(CalcRecoMOD)`
	`CalcGPPGUI(CalcGPP)`
	`ModGFillGUI(ModGPP)`

'miniRECgap' functions (Fig. 2a), which will perform various data-processing and partitioning, and NEE gap-filling (Premrov, 2024). While the user will need to run five code lines (Table 5), it should be noted that 'miniRECgap' consists of a number of functions (listed in Fig. 2a), most of which will be called via GUI and run in the background without the need to write a separate script/code to individually run them.

Detailed instructions on how to employ the code from Table 5 to perform gap-filling using 'miniRECgap' R-package via GUI supported scripts (i.e. STEP 1 to STEP 5), are provided in the Supplemental Material SM5. The user should follow there described instructions on performing the procedures in five steps and illustrated in Fig. SM2 to Fig. SM11, which provide the print-screens of individually called GUI windows.

3.2.3. On referencing and citing relevant sources when using 'miniRECgap'

When using the 'miniRECgap' package, the user should ensure to include the references to the package/software source-code, as well as to this paper, i.e. the information on referencing the 'miniRECgap' R-package source-code is included in Table 6a. A user can also easily retrieve the information required for referencing the Arrhenius equation adapted from Lloyd and Taylor (1994) and the light-response function reported in Gilmanov et al. (2003), which is adapted from the non-rectangular hyperbola model by Rabinowitch (1951), which form part of the flux-partitioning and gap-filing methodologies used in 'miniRECgap' (explained in section 2.1). Information on both references can be obtained by running the code from Table 6b in R. It is the responsibility of each user to also include the references on all other R-packages that are required for operating 'miniRECgap'. The information needed for referencing these packages can be retrieved by using the function 'citation' explained in Table 6c (with further details explained in the note to Table 6 and Supplemental Material SM6).

Table 6
Retrieving information needed for referencing: (a) 'miniRECgap' R-package citation; (b) code, which can be used to retrieve information needed for referencing the non-linear response functions/gap-filling methods used in 'miniRECgap'; (c) other R-packages required for operating 'miniRECgap'.

Retrieving information needed for referencing	Citation (a) / Code (b), (c)
(a) 'miniRECgap' package/software source code on GitHub https://github.com/APremrov/miniRECgap (Premrov, 2024)	Premrov, Alina. 2024. 'miniRECgap': R-package for gap-filling of the missing eddy covariance CO2 flux measurements using selected classic nonlinear environmental response functions via simple user-friendly GUI supported R scripts. Copyright (c) Trinity College Dublin 2024. URL: https://github.com/APremrov/miniRECgap, DOI: https://doi.org/10.5281/zenodo.13228227.
(b) Lloyd and Taylor (1994)	`Cite_calc_miniRECgap('CalcRecoMod')`
Gilmanov et al. (2003) and Rabinowitch (1951)	`Cite_calc_miniRECgap('ModGPP')`
(c) Other packages. The `'PACKAGE NAME'` in the code needs to be replaced with the actual name of the package.	`citation('PACKAGE NAME')`

NOTE: The code in Table 6 (b), (c) can be run in R to retrieve required information after first running the code from Table 4. After running the code from Table 6 (b), (c) in R, the information on references will be loaded in R Console. The reference to the 'miniRECgap' R-package/source-code (Table 6 (a)) should also be accompanied by the references to this paper. The 'PACKAGE NAME' in the code (Table 6c) must be replaced with the actual name of the package, i.e. these packages are: 'stats', 'grDevices', 'utils' (R Core Team, 2024b), 'fgui' (Hoffmann and Laird, 2009), 'dplyr' (Wickham et al., 2023) and 'ggplot2' (Wickham, 2016); (additional explanation is provided in Supplemental Material SM6].

4. Case-study results and discussion

4.1. Gap-filling performance evaluation using data from the 'classic' and 'challenging' ecosystems

The gap-filling performances were assessed on the basis of the findings of multiple model prediction indices and estimations of uncertainty derived from the NEE modelling outputs using different modelling approaches from this study. The findings are reported based on the NEE modelling that was performed on data-inputs at half-hourly intervals (i.e. the original format of time-interval of measured data), which resulted in model outputs for the same half-hourly time-intervals. To evaluate the 'miniRECgap' gap-filling approach (i.e. temperature- and light-response functions), the results also include the findings from the standard 'MDSnight' approach (applied via the 'REddyProc' webtool by Wutzler et al. (2018)) for both ecosystem examples. Furthermore, in the case of

Cavemount Bog ('challenging' ecosystem), the NEE modelling results obtained from applying an optimised shallow ANN are also reported.

4.1.1. The 'classic' ecosystem: DE-Tha forest

4.1.1.1. Model prediction indices and estimation of uncertainty. The 'miniRECgap' package performance was first examined by plotting the observed vs. modelled/predicted NEE values for the DE-Tha 'classic' ecosystem. The R^2 value for the full dataset was 0.61 and was 0.71 for the dataset with three clusters (Fig. 5). There was a reasonably close fit of the regression line to the 1:1 line (Fig. 5b and c). The results further showed that the 'MDSnight' approach outperformed 'miniRECgap', which was evident from an even closer fit of the regression line to the 1:1 line and an R^2 value of 0.83 for the former (Fig. 5a). These results confirm our earlier

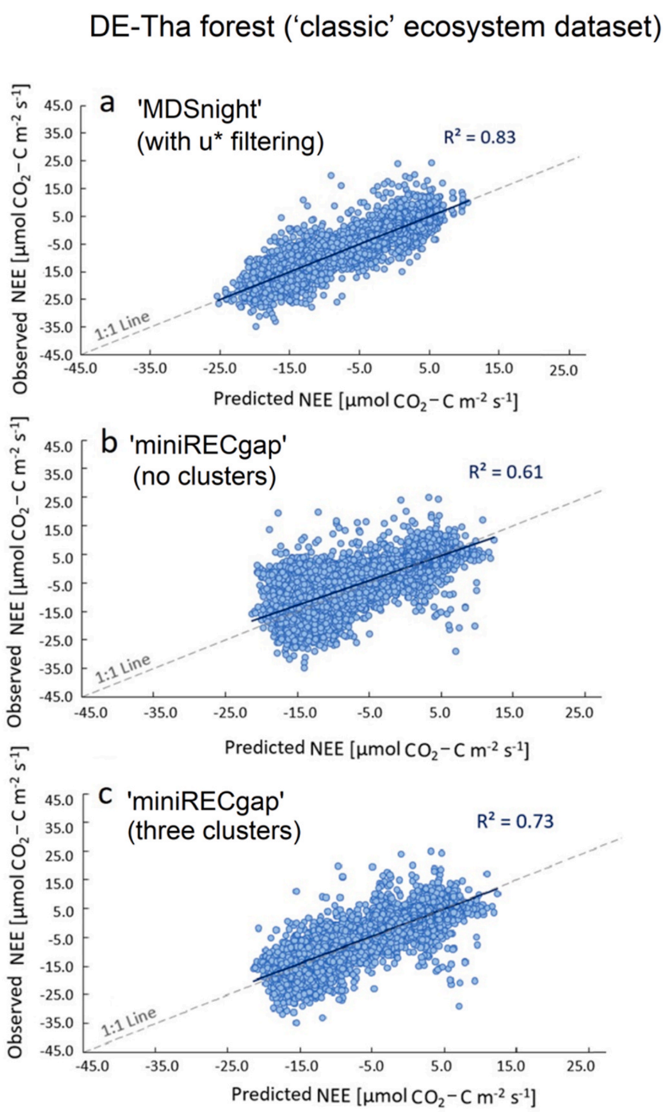

Fig. 5. Observed vs. predicted net ecosystem exchange (NEE; μ mol CO_2-C m^{-2} s^{-1}] for DE-Tha forest ('classic' ecosystem dataset) using different modelling approaches (a) 'MDSnight' - with u* filtering, (b) 'miniRECgap' - full dataset (no clusters); (c) 'miniRECgap' - dataset assigned into three clusters.
NOTE: Data-splitting into clusters was performed using k-means clustering approach. A figure for NEE modelled using 'MDSnight' without u* filtering is not presented (due to its strong similarity with Fig. 5a). Note the difference in x-axis and y-axis scale (interval-length), but not in the units. The observed values are plotted on the y axis and predicted values on the x axis (i.e. OP type of plot) in accordance with the mathematical evidence that OP regression should be used when evaluating models, following Piñeiro et al. (2008).

expectation that the 'MDSnight' approach (applied using 'REddyProc' webtool) would likely outperform 'miniRECgap' for the 'classic' ecosystem, which is in agreement with previous work that showed the better performance of 'REddyProc' compared to some other methods (Boudhina et al., 2018; Wang et al., 2023; Wutzler et al., 2018). Nevertheless, while the 'MDSnight' approach performed better than 'miniRECgap', it is also more 'data-hungry'. The advantage of the 'miniRECgap' package is that it requires (in addition to stamp/date-time variable) data for only three main input-variables (NEE, PPFD and T; Table 2), whereas the 'MDSnight' approach requires data for five to seven input-variables (NEE, Tsoil, Tair, LE, Rg, rH and optional VPD and Ustar; Table 2). Therefore, the 'miniRECgap' package may be potentially useful for datasets that lack some of the required input data needed in 'MDSnight'.

The results from multiple model prediction indices and estimation of uncertainty derived from both approaches (Table 7a) further confirmed that 'MDSnight' performed better than 'miniRECgap', with 'MDSnight' with u* filtering likely the best gap-filling approach for the DE-Tha 'classic' ecosystem ($R^2 = 0.83$; $E = 0.03$ μ mol CO_2-C m^{-2} s^{-1}, Table 7a). In the case of DE-Tha, RMSE (rRMSE) values derived from the different approaches [i.e. rRMSE = 0.39 for 'MDSnight' (with u* filtering); rRMSE = 0.50 for 'miniRECgap' (on the dataset with three clusters); Table 7a] were found to be in the range of rRMSE values reported in the literature. For example, a study by Moffat et al. (2007), which compared a variety of different EC flux gap-filling techniques on six forested European sites, reported mean rRMSE site-dependency values that ranged from *c.* < 0.35 (daytime) to > 0.7 (nighttime).

While most of the model prediction indices (RMSE, StD, MAE, Table 7a) derived from the 'MDSnight' approach without u* filtering appeared to be the same, similar or slightly lower than the ones derived with u* filtering, the observed bias or systematic error (Es) was smaller where u* filtering was used (Es = 0.009 μ mol CO_2-C m^{-2} s^{-1}) (Table 7a). The Es and relative error (Er) values indicate that the major contributor to the difference in the magnitude of uncertainty estimate (E) between the 'miniRECgap' and 'MDSnight' approaches was likely due to the observed negative Es value when the former approach was used (Table 7a). The strongest negative Es value was observed when NEE was modelled using 'miniRECgap' on the dataset with no prior separation into clusters/groups (Es = −0.46 μ mol CO_2-C m^{-2} s^{-1}; Table 7a). However, the strength of the bias almost halved after assigning the data into three clusters (Es = −0.24 μ mol CO_2-C m^{-2} s^{-1}; Table 7a), while the R^2 value increased to 0.71. These results strongly indicate that the performance of the 'miniRECgap' approach could potentially improve, provided NEE modelling was undertaken on datasets disaggregated by groups. As 'miniRECgap' is a relatively simple tool, it does not have a built-in function for this purpose, and users need to assign their own data into groups or clusters. In some cases, this may be the preferred practice because different ecosystem sites could require different approaches for such grouping or clustering of data, depending on their specific conditions or management. The advantage of 'miniRECgap' is that it operates in an effortless manner using just five lines of code, thereby allowing users to check the performance on various pre-prepared data-splits using data-splitting methods of their own choice, in order to decide on the best one for the given site.

It is known that continuous bias in gap-filling can potentially lead to over- or under-estimation of annual NEE sums (Moffat et al., 2007). 'However, bias occurrence is not unusual in gap-filling techniques and has been reported widely in the literature. For example, a negative bias has been reported for a study by Desai et al. (2005) for classic non-linear regression gap-filling technique (Moffat et al., 2007). Negative biases for gap-filling of CO_2 can be also found for a number of sites in studies by Lucas-Moffat et al. (2007, 2022) using different gap-filling techniques. The literature shows that even the widely used MDS gap-filling method is not immune to bias. The recent study by Vekuri et al. (2023) showed that the MSD gap-filling technique can cause systematic bias if applied to data from northern sites at latitudes > 60°, which can result in the systematic overestimation of CO_2 emissions. The authors further suggested that ML can be used to reduce this bias for northern sites (Vekuri

Table 7
Selected model prediction indices and estimation of uncertainty derived from using different gap-filling approaches for: (a) DE-Tha forest ('classic' example) and (b) Cavemount Bog ('challenging' example) datasets.

Gap-filling approach/package/methodology		Model prediction indices [b]					Estimation of uncertainty [b]		
		R^2 [Eq. (5)]	RMSE [Eq. 6a; 6b][b,]	rRMSE [Eq. (6c)]	StD [Eq. (7a)] [b]	MAE [Eq. (7b)] [b]	Es [Eq. (8)] [b]	Er [Eq. (9)] [b]	E [Eq. (10)] [b]
a) DE-Tha[a]									
'miniRECgap'	Full dataset	0.61	4.53	0.61	4.13	2.92	−0.46	0.04	0.47
	3 clusters	0.73	3.69	0.50	3.52	2.50	−0.24	0.04	0.25
'MDSnight' [c]	u* filtering	0.83	3.02	0.39	2.84	2.01	0.01	0.03	0.03
	No u* filtering	0.83	2.91	0.39	2.71	1.92	0.02	0.03	0.04
b) Cavemount									
'miniRECgap'	Full dataset	0.46	1.60	0.77	1.66	1.17	−0.39	0.02	0.39
	3 clusters	0.57	1.43	0.69	1.45	1.03	−0.36	0.02	0.36
'MDSnight' [c]	u* filtering	0.23	1.83	0.88	1.93	1.36	(−)0.00	0.02	0.02
	No u* filtering	0.24	1.82	0.87	1.91	1.35	(−)0.00	0.02	0.02
'N3' with **eight** predictors [d]	Trained on 30 %	0.69	1.17	0.56	1.11	0.79	−0.03	0.01	0.03
	Trained on 70 %	0.69	1.17	0.56	1.12	0.79	−0.01	0.01	0.02
'N3' with **six** predictors (final model) [d,e]	Trained on 30 %	0.68	1.18	0.57	1.13	0.80	0.02	0.01	0.03
	Trained on 70 %	0.68	1.17	0.56	1.13	0.80	0.01	0.01	0.01

NOTE:
cThe outputs from Eq. (6a) and Eq. (6b) were the same (due to high Np).
[a] DE-Tha refers to Tharandt forest site dataset, which is a well-known 'classic example' provided as a template for data input into the 'REddyProc' webtool (the dataset was downloaded from MPI (2024d)).
[b] The units (where applicable) are in μ mol CO_2-C m^{-2} s^{-1} because gap-filling was done on datasets without converting their original units (i.e. measured NEE [μ mol CO_2-C m^{-2} s^{-1}] in 1/2 h intervals). A negative sign (−) for zero Es values correspond to negative Es values at three or four decimal precision, but they were rounded to zero because they are reported at two decimals. In order to avoid losing on precision in the findings from gap-filling, the values have not been converted from original units μ mol CO_2-C to g CO_2-C m^{-2} s^{-1}.
[c] Gap-filling performed with application of 'REddyProc' by Wutzler et al. (2018), where 'MDSnight' refers to application settings/approaches in chosen in the 'REddyProc' webtool as explained in section 2.3.3 and Supplemental Material SM2.
[d] Gap-filling performed on joined dataset (training + testing) with the ANN model 'N3 'with a single hidden layer and three neurons (j = 3), via 'neuralnet' function from R-package 'neuralnet' (Fritsch et al., 2019), using settings explained in section 2.3.4 and Supplemental Material SM3.
[e] When using six predictors the joined dataset (training + testing) represents the full dataset (further explanation is provided in sections 2.3.4 and 4.1.2).

et al., 2023). Clearly, these findings indicate that users need to pay special attention when using gap-filling approaches/techniques at different sites, and should assess the systematic errors/bias to evaluate the risk of potential over-/under-estimation of CO_2 fluxes.

4.1.1.2. NEE sums and cumulative NEE. Overall, our results showed that the DE-Tha forest site acted as a C-sink (NEE sum exceeding −6 t CO_2-C ha^{-1}). These results are in agreement with similar NEE sum values reported for the same site in the literature by Grünwald and Bernhofer (2007), with some slight differences possibly occurring due to potential different settings in the gap-filling approaches. In this study, the DE-Tha annual NEE sums (Table 9) and cumulative NEE (Fig. 6) from 'miniRECgap' may need to account for potential under-estimation due to observed negative bias (Es = −0.24 μ mol CO_2-C m^{-2} s^{-1}; Table 7a).

Table 8
Selected model prediction indices and estimation of uncertainty derived from the Artificial Neural Network (ANN) model 'N3' for predicting net ecosystem exchange (NEE) for the Cavemount Bog dataset, trained on either 30 % or 70 % training data, and by applying different number of input predictor variables: (a) eight input predictor variables [DOY, Hour, Tair, Tsoil, rH, LE, H, PPFD]; (b) six input predictor variables [DOY, Hour, Tair, Tsoil, rH, PPFD].

ANN model 'N3'		Model prediction indices [a]					Estimation of uncertainty [a]		
		R^2 [Eq. (5)]	RMSE [Eq. (6a); 6b][a, b]	rRMSE [Eq. (6c)]	StD [Eq. (7a)] [a]	MAE [Eq. (7b)] [a]	Es [Eq. (8)] [a]	Er [Eq. (9)] [a]	E [Eq. (10)] [a]
a) Eight input predictor variables									
Training/testing ratio 30:70	Training	0.70	1.15	0.54	1.09	0.77	0.00	0.02	0.02
	Testing	0.68	1.17	0.56	1.13	0.80	−0.04	0.02	0.04
Training/testing ratio 70:30	Training	0.69	1.16	0.55	1.11	0.79	0.00	0.02	0.02
	Testing	0.68	1.19	0.57	1.15	0.81	−0.03	0.02	0.04
b) Six input predictor variables									
Training/testing ratio 30:70	Training	0.65	1.23	0.59	1.19	0.84	0.00	0.02	0.02
	Testing	0.69	1.16	0.56	1.11	0.79	0.03	0.02	0.03
Training/testing ratio 70:30	Training	0.67	1.18	0.57	1.13	0.80	0.00	0.02	0.02
	Testing	0.70	1.16	0.55	1.11	0.79	0.02	0.02	0.03

NOTE:
a Gap-filling performed on separate training and testing datasets for the Cavemount site ('challenging' example) with an ANN model 'N3' with a single hidden layer and three neurons (j = 3), via 'neuralnet' function from R-package 'neuralnet' (Fritsch et al., 2019), using settings explained in section 2.3.4 and Supplemental Material SM3.
d Training - refers to in-sample metrics; Testing - refers to out-of-sample metrics (see section 2.4).
[a] The units (where applicable) are in μ mol CO_2-C m^{-2} s^{-1} as gap-filling was done on datasets without conversion of their original units (i.e. measured NEE [μ mol CO_2-C m^{-2} s^{-1}] in 1/2 h intervals).
[b] The outputs from Eq. (6a) and Eq. (6b) were the same (due to high Np).

Table 9
Gap-filled net ecosystem exchange (NEE) sums using different gap-filling approaches/packages for the (a) DE-Tha ('classic' example) and (b) Cavemount ('challenging' example) sites.

a) DE-Tha[a]		Sum of gap-filled NEE [t C ha^{-1}] for 365 days	Proportion of gaps [c]
'miniRECgap' [b]	Full dataset	−6.23	36 %
	3 clusters	−6.02	36 %
'MDSnight' [d]	u* filtering	−6.12	45 %
	No u* filtering	−6.43	36 %
b) Cavemount		**Sum of gap-filled NEE [t C ha^{-1}] for 236 days**	Proportion of gaps [c]
'miniRECgap' [b]	Full dataset	−0.33	20 %
	3 clusters	−0.23	20 %
'MDSnight' [d]	u* filtering	−0.17	25 %[c]
	No u* filtering	−0.03	20 %
'N3' [e] with 6 predictors	Trained on 30 % data	0.01	20 %
	Trained on 70 % data	0.01	20 %

NOTE:
[a] DE-Tha dataset is provided as a template for data input into the 'REddyProc' webtool, and it was downloaded from MPI (2024d).
[b] Three half-hourly NEE values from the DE-Tha 'miniRECgap' gap-filled dataset were estimated with linear interpolation as gap-filling with 'miniRECgap' was not possible for the three data-points due to missing meteorological data.
[c] The number of data-gaps in the dataset increased in case if the u* filtering was applied (using 'REddyProc' webtool) resulting in an increase in the total number of NEE data-gaps. In the case of the Cavemount dataset, the u* filtering was applied on top of pre-existing QC/QA screening (further details are explained in section 2.2).
[d] Gap-filling performed with the application of 'REddyProc' by Wutzler et al. (2018), where 'MDSnight' refers to the application settings/approaches selected in the 'REddyProc' webtool as explained in section 2.3.3 and Supplemental Material SM2.
[e] Gap-filling performed using ANN via the 'neuralnet' function from R-package 'neuralnet' (Fritsch et al., 2019) using hidden configuration and other settings explained in section 2.3.4 and Supplemental Material SM3. ANN gap-filling approach with 'N3' was used for the Cavemount peatland dataset only.

However, the range of annual NEE sums from both 'miniRECgap', and 'MDSnight' approaches overlap somewhat [i.e. range for 'miniRECgap': 6.02 to −6.23 t CO_2-C ha^{-1} (depending on whether the approach used clustering or not); range for 'MDSnight': 6.12 to −6.43 t CO_2-C ha^{-1} (depending on whether the approach used u* filtering or not)]. This overlapping would indicate that the risk of potential NEE sum under-estimation when using 'miniRECgap' could lie within the range of occurring variability of modelled NEE results derived from different gap-filling approaches. This is also supported by the fact that the shape of the cumulative NEE curves derived from the two approaches closely resemble each other (Fig. 6a, b and 6c).

4.1.2. The 'challenging' ecosystem: Cavemount Bog

4.1.2.1. Model prediction indices and estimation of uncertainty. The package performance was next examined by plotting the observed vs.

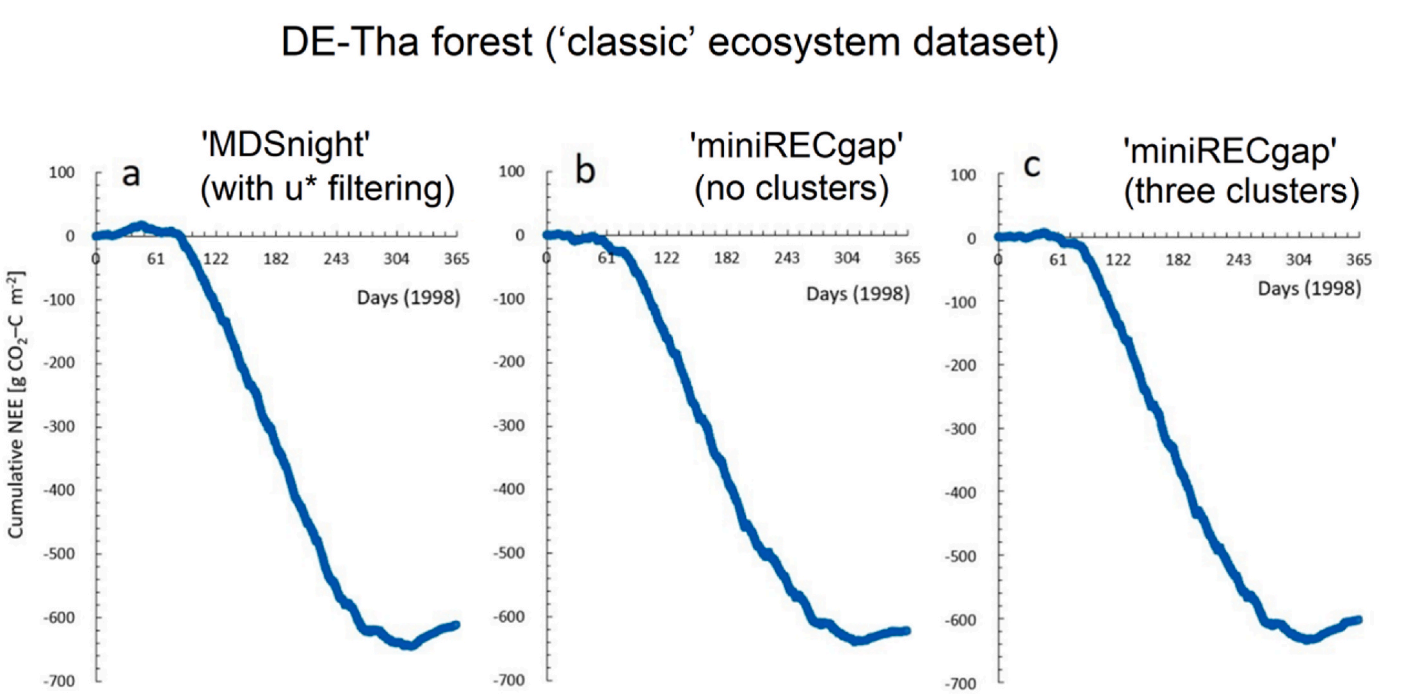

Fig. 6. Daily cumulative gap-filled net ecosystem exchange (NEE; g CO_2-C m^{-2}] for the DE-Tha forest site ('classic' ecosystem dataset) using different gap-filling approaches: (a) 'MDSnight' - with u* filtering, (b) 'miniRECgap' - full dataset (no clusters); (c) 'miniRECgap' - dataset assigned into three clusters.
NOTE: Total period of observation is 1 year (365 days). Data-splitting into clusters was performed using k-means clustering approach (k = 3 clusters). A figure for cumulative NEE obtained by using 'MDSnight' without u* filtering is not presented (strong similarity was observed between the two cumulative NEE curves obtained using the 'MDSnight' gap-filling approach with and without u* filtering).

modelled/predicted NEE values for the ‘challenging’ ecosystem example, Cavemount Bog (Fig. 6). In contrast to DE-Tha, the ‘MDSnight’ approach (applied via ‘REddyProc’ webtool) no longer showed a better performance compared to ‘miniRECgap’. In the first instance, this was evident from the results of the regression analysis, which showed low R^2 values ($R^2 \leq 0.24$) and a deviation of the regression line from the 1:1 line (Fig. 6a and b) when using the ‘MDSnight’ approach, regardless of u* filtering. With the exception of uncertainty estimation, the results from other multiple model prediction indices derived from different model approaches (Table 7b) further confirmed that the application of the ‘MDSnight’ approach may be a challenge for this rehabilitated ecosystem.

While ‘miniRECgap’ showed some advantages over ‘MDSnight’ in the ‘challenging’ ecosystem in specific context of this particular study, this does not imply its universal superiority. Rather, it indicates that the Cavemount Bog ecosystem dataset posed more of a challenge for the ‘MDSnight’ gap-filling method than if conventional temperature- and light-functions were applied in ‘miniRECgap’ in this particular example. It should be noted that the MDS method relies more on the existing data distribution. Therefore, the heterogeneous and complex nature of a ‘challenging’ ecosystem in the given example (resulting from the presence of a mosaic of vegetated areas, bare patches and open water areas) may have posed a challenge for ‘MDSnight’ if the existing dataset did not adequately represent the full range (distribution) of various conditions at the given site. This may have been the reason for somewhat better performance of ‘miniRECgap’ compared to ‘MDSnight’, considering that ‘miniRECgap’ relies more on the given functional relationships. These findings indicate that users are recommended to test and apply the ‘miniRECgap’ tool to periods of data that capture the dynamism in variable driving flux dynamics at their site, particularly those that relate to management or significant variations in climate variables. However, since this study employs only crude model performance metrics, these findings should be considered as indicative. Therefore, additional research using longer datasets (than in the given example), from extended monitoring periods, and sites from a wider range of terrestrial ecosystems, is strongly recommended. As explained earlier, there is also a need for future in-depth comprehensive model evaluation, which could include different data-gap scenarios (Moffat et al., 2007) and various terrestrial ecosystem types.

Although the multiple prediction indices indicated that ‘miniRECgap’ performed somewhat better than ‘MDSnight’, the findings also indicate that the Cavemount Bog dataset is also a challenge for the ‘miniRECgap’ approach. The observed R^2 values were lower compared to DE-Tha [i.e. $R^2 = 0.48$ for Cavemount dataset with no clusters, Fig. 7c; $R^2 = 0.57$ for Cavemount dataset with three clusters), Fig. 7d]. These results are in agreement with the findings of Zhu et al. (2023) who assessed the application of EC gap-filling approaches in a number of ‘challenging’ ecosystems. According to Zhu et al. (2023) many popular flux gap-filling methods (such as MDS employed in this study) can work well for less disturbed/natural ecosystems (such as the DE-Tha), but their application can be challenging for ecosystems that have undergone significant disturbances (e.g. Cavemount Bog). Despite this, the RMSE value [μ mol CO_2-C m^{-2} s^{-1}] derived from the ‘miniRECgap’ and ‘MDSnight’ approaches for the Cavemount site [i.e. RMSE: 1.83 (‘MDSnight’ with u* filtering); RMSE: 1.43 ‘(miniRECgap’, three clusters); Table 7b] still appears to be within the range of RMSE variability reported in the literature. For example, a study on an upland blanket bog in the north Pennines by Lloyd (2010) noted seasonally variable RMSE values in the range 1.88–2.75 μ mol CO_2-C m^{-2} s^{-1} when the PIRT[6] gap-filling modelling technique was used.

In the case of Cavemount, the ‘miniRECgap’ approach also showed the highest observed E estimate of uncertainty, which again appears to result from the negative bias Es (Table 7b). For this reason, when using ‘miniRECgap’, the interpretation of the findings of the Cavemount Bog NEE sums or cumulative NEE again need to take into consideration the potential risks of underestimation of NEE sums, resulting from the observed negative Es. Similar to earlier findings (section 4.1.1), the strongest negative bias was observed when NEE was modelled using ‘miniRECgap’ on the dataset with no prior separation into groups/clusters (Es = −0.39 μ mol CO_2-C m^{-2} s^{-1}; Table 7b). A slight reduction in the strength of the bias (Es = −0.36 μ mol CO_2-C m^{-2} s^{-1} Table 7b) can be observed when the dataset was assigned into three clusters, indicating that the performance of the ‘miniRECgap’ approach could potentially improve, provided NEE modelling was carried out on the datasets disaggregated by groups/clusters using the clustering methodology suitable for the given data (i.e. in this study the approach was the unsupervised ML k-means clustering; section 2.3.1).

It should be noted that as Cavemount Bog is a rehabilitated cutaway peatland, the conditions at this site are subject to extreme heterogeneity, represented by a mosaic of vegetation communities, as well as areas of bare soil/peat and open water (Bord na Móna, 2021). Therefore, it is not surprising that challenges in NEE gap-filling are observed for both approaches (‘miniRECgap’ and ‘MDSnight’) as these approaches may not have sufficiently accounted for the specific heterogeneous conditions found at this site. In cases when we deal with highly disturbed ecosystems that are very heterogeneous in nature, several different gap-filling approaches may need to be assessed to determine the most suitable one for the given case, including more advanced methods, such as ML (i.e. Group 4 in Table 1). The advantage of ‘miniRECgap’ is that it operates in a simple and user-friendly way, which allows the user to obtain the NEE modelled outputs more effortlessly and focus more on evaluating the model performance in order to decide whether any other gap-filling approaches may need to be tested. In this study, the application of ‘miniRECgap’ showed that for the ‘challenging’ Cavemount Bog site, other gap-filling methods should be tested. Therefore, an additional ANN ML-based gap-filling approach was applied to the Cavemount dataset (as explained in section 2.3.4).

The performance of an optimised and trained ANN model (‘N3’ consisting of a single hidden layer and three neurons) was first assessed separately on the training and testing datasets. The results showed that the best overall model performance was when six predictor input-variables [DOY, Hour, Tair, Tsoil, rH, PPFD] and the 70:30 training: testing ratio were employed ($R^2 = 0.70$; Es = 0.02 μ mol CO_2-C m^{-2} s^{-1}; the other model prediction indices and estimation of uncertainty are presented in Table 8). The performance of the optimised and trained ‘N3’ model was next assessed by applying the model to the joined dataset (training + testing). Here, the results showed that the best overall performance was when eight predictors [DOY, Hour, Tair, Tsoil, rH, LE, H, PPFD] and the 70:30 training: testing ratio were employed ($R^2 = 0.69$, Es = -0.01 μ mol CO_2-C m^{-2} s^{-1}; the other model prediction indices are presented in Table 7b). The ‘N3’ model with six predictors [DOY, Hour, Tair, Tsoil, rH, PPFD] and the 70:30 training: testing ratio resulted in similar but positive bias (Es = 0.01 μ mol CO_2-C m^{-2} s^{-1}, $R^2 = 0.68$, Table 7b) compared to the model with eight predictors, which resulted in similar but negative bias (Es = -0.01 μ mol CO_2-C m^{-2} s^{-1}; Table 7b).

As explained in section 2.3.4 and Supplemental Material SM3.4, the Cavemount dataset had some missing LE and H values, which meant that when eight predictor input-variables were used in ANN modelling [DOY, Hour, Tair, Tsoil, rH, LE, H, PPFD], the joined dataset (training + testing) also contained gaps due to the missing LE and H points in the timeseries. This was not the case when the ANN modelling was performed using only six predictor variables, as the LE and H variables were excluded from model-inputs, which meant that the joined model-input dataset (training + testing) represented the full time-series dataset for model input-variables [DOY, Hour, Tair, Tsoil, rH, PPFD]. For this reason, the ‘N3‘ model using six predictor input-variables was selected as the final ANN model, because it allowed NEE gap-filling on the full

[6] PIRT refers to the photosynthetic irradiance response and temperature sensitive respiration model explained in Lloyd (2010).

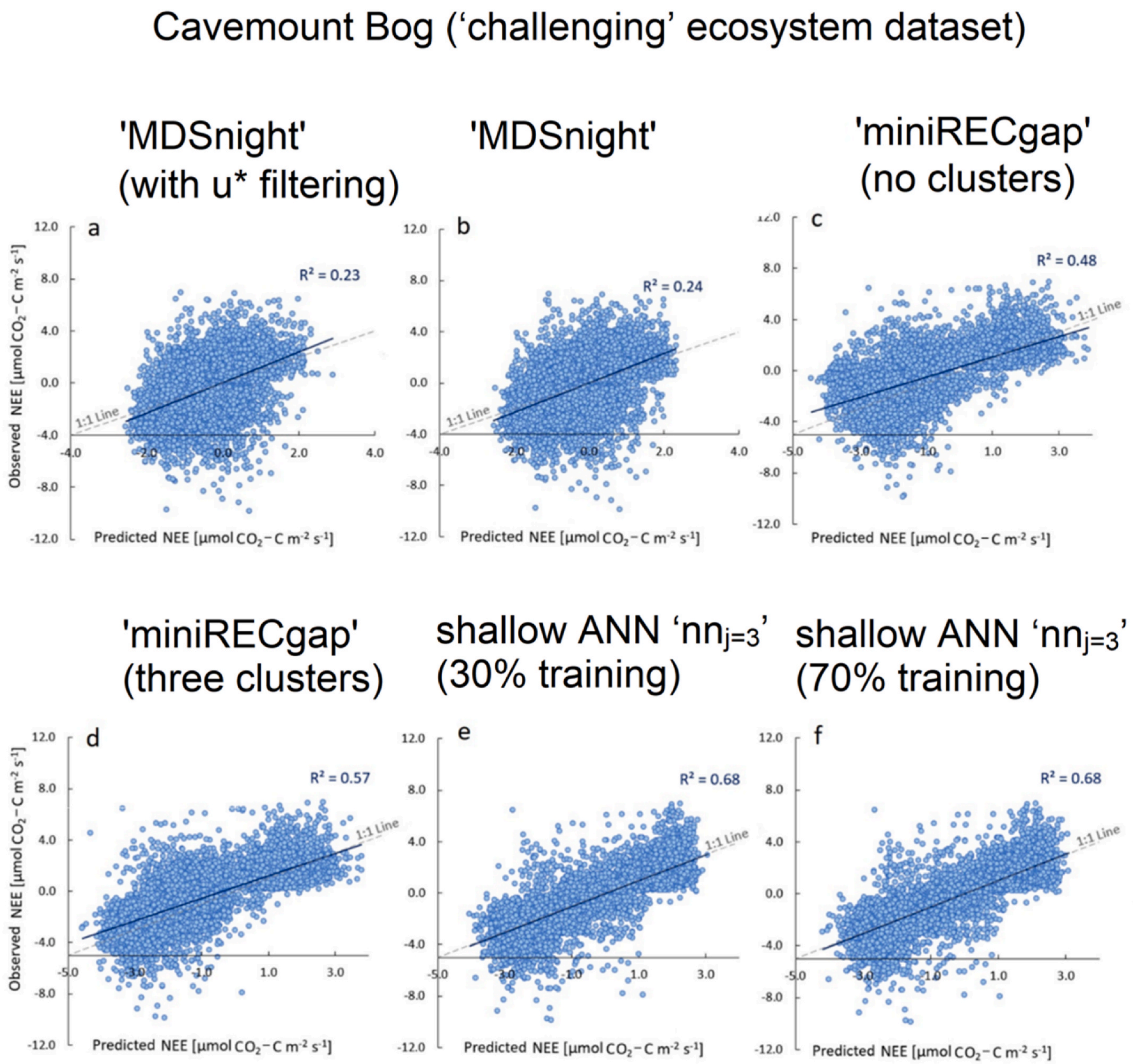

Fig. 7. Observed vs. predicted net ecosystem exchange (NEE; μ mol CO_2-C m^{-2} s^{-1}] for Cavemount Bog ('challenging' ecosystem dataset) using different approaches to model NEE: (a) 'MDSnight' with u* filtering, (b) 'MDSnight' without u* filtering, (c) 'miniRECgap' full dataset (no clusters); (d) 'miniRECgap' dataset split into three clusters; (e) ANN model 'N3' (six predictors) trained on 30 % data; (f) ANN model 'N3' (six predictors) trained on 70 % data.
NOTE: Data-splitting into clusters was performed using k-means clustering approach. Note the difference in x-axis scale and in crossing y-axis in Fig. 7a and b vs. Fig. 7c–f. The observed values are plotted on the y axis and predicted values on the x axis (i.e. OP type of plot) in accordance with the mathematical evidence that OP regression should be used when evaluating models, following Piñeiro et al. (2008).

dataset, which is needed to calculate the NEE sum and the cumulative NEE for the given period of observation (236 days at the Cavemount site).

The regression analysis showed that the final 'N3' model with six predictor input-variables (R^2: 0.68; regression line to the 1:1 line; Fig. 6e and f) outperformed the 'miniRECgap' and 'MDSnight' approaches for both the 70:30 and 30:70 training: testing ratios. Although the strength of bias for the 'MDSnight' approach (with and without u* filtering Es = 0.00 μ mol CO_2-C m^{-2} s^{-1}; Table 7b) was much lower compared to the strength of the bias observed when using the final 'N3' model (Table 7b), the remainder of the model prediction indices showed that the optimised shallow ANN gap-filling approach performed better than both 'miniRECgap' and 'MDSnight'. The final 'N3' model (using six predictors and 70 % training data) showed similar magnitude of bias (Es = 0.01 μ mol CO_2-C m^{-2} s^{-1}) compared to the model trained on only 30 % data (Table 7b).

4.1.2.2. NEE sums and cumulative NEE. For Cavemount Bog, annual NEE sums and cumulative NEE derived from different gap-filling approaches used in this study did not provide conclusive information as to whether this site acts as a C sink or C source. The range in annual NEE sums from the different approaches ('miniRECgap', 'MDSnight' and 'N3' model) did not overlap (as was the case with DE-Tha). NEE sums derived from 'miniRECgap' and 'MDSnight' were all negative [i.e. 'miniRECgap': 0.33 t CO_2-C ha^{-1} (no clusters), −0.23 t CO_2-C ha^{-1} (three clusters); 'MDSnight': 0.17 t CO_2-C ha^{-1} (no u* filtering), −0.03 t CO_2-C ha^{-1} (with u* filtering); Table 7b], which may be consistent with the observed negative Es values (Table 7b). The NEE sum derived from the 'N3' was positive [0.01 t CO_2-C ha^{-1}; Table 7b], which may be consistent with the observed positive Es value (Table 7b). Contrary to the DE-Tha site, the shapes of the cumulative NEE curves for Cavemount Bog, derived from different gap-filling approaches, showed only a slight resemblance to each other (Fig. 8a–f). These findings indicate that we may need to account for potential under-/over-estimation in NEE sums and in the cumulative NEE values derived from different gap-filling approaches, possibly influenced by the negative/positive bias (Table 7b). Despite these potential under-/over-estimations, the Cavemount NEE sums derived from all three gap-filling approaches appear to be in the low range and relatively close to zero [i.e. not exceeding ±0.50 t CO_2-C ha^{-1}]. Furthermore, during the total observation period, we can observe fluctuations between positive (C losses) and negative (C uptake) cumulative NEE at the Cavemount Bog. These results indicate that the Cavemount cutaway peatland ecosystem undergoing rehabilitation, is in a transition-period and will still need some time before it is

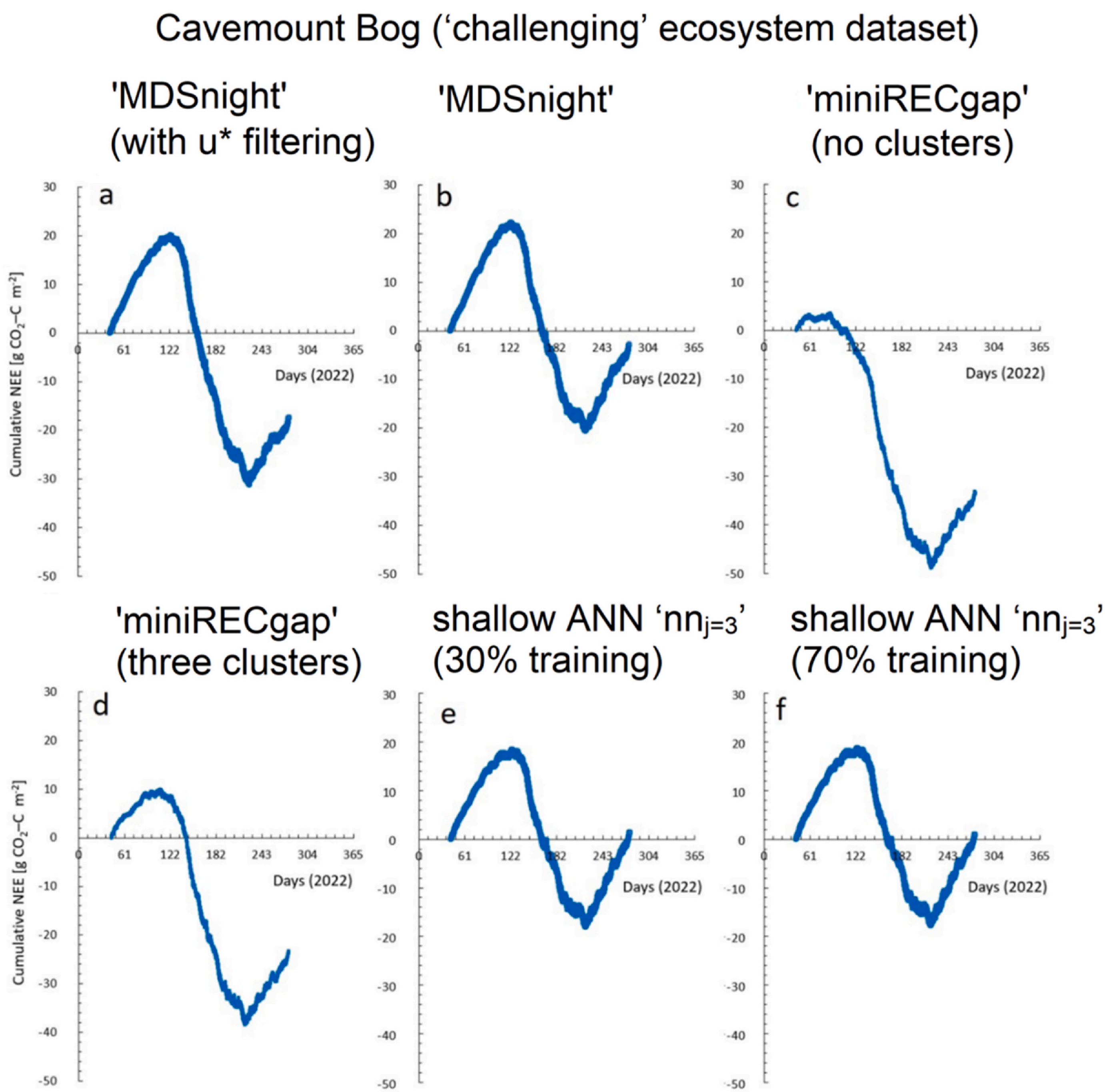

Fig. 8. Daily cumulative gap-filled net ecosystem exchange (NEE; g CO_2-C m^{-2}] for Cavemount Bog ('challenging' ecosystem dataset) using different gap-filling approaches: (a) 'MDSnight' with u* filtering, (b) 'MDSnight' without u* filtering, (c) 'miniRECgap' full dataset (no clusters); (d) miniRECgap' - dataset split into three clusters; (e) ANN model 'N3' (six predictors) trained on 30 % data; (f) ANN model 'N3'(six predictors) trained on 70 % data. NOTE: Fig. 8b and c - total period of observation is less than a year (236 days). Data-splitting into clusters was performed using k-means clustering approach.

established as an active long-term C sink.

It appears that the NEE sum and cumulative NEE results derived from the final 'N3' model most closely represent the potential transitioning conditions at Cavemount Bog, which is evident from the very low NEE sum value of 0.01 t CO_2-C ha^{-1} (closest to zero among the three approaches), as well as from the well-balanced s-shaped cumulative NEE curve, where positive and negative cumulative NEE areas appear to almost cancel each-other (Fig. 8e and f). Although direct comparison of the NEE sums with studies from the literature is not possible due to the short period of observation at Cavemount Bog (*ca.* 8 months), the observed findings are in agreement with the low NEE values observed during the initial six month period (January to June) at a drained and rewetted bare peat Bellacorrick peatland microsite in a study by Wilson et al. (2016). Based on the shapes of the cumulative NEE curves (Fig. 8e and f) during the observational period from *c.* February to October 2022, one can observe the predominantly positive cumulative NEE occurring during *c.* first half of the total observation period, to predominantly negative values occurring during the second half of this period. It would appear that Cavemount Bog switched from acting as a C source to becoming a C sink at around the end of June or early July 2022. This observed seasonal pattern in the Cavemount cumulative NEE is in agreement with the overall generally strong summer CO_2 uptake observed at some of the microsites from another Irish former peat-extraction site, Bellacorrick peatland, which was rehabilitated in 2002 (Wilson et al., 2016). These findings clearly demonstrate the importance of assessing the gap-filled NEE to gain deeper insights into C dynamics and better understanding whether the 'challenging' ecosystem has shifted from a C source to C sink. It would be interesting to further assess other variables, such as water-table fluctuations, plant biomass, length of growing season, etc. at the site during these periods, to determine the potential main drivers of C dynamics, which can be important information for management of the rehabilitation and restoration activities at this site. For peatlands undergoing rehabilitation which may be subjected to transitioning conditions, it is recommended to assess the annual cumulative NEE curves over a number of years of observation. This can provide further insights if the shapes of the annual cumulative NEE curves over the observational years are shifting towards the shape that can be observed in ecosystems that are well-established C sinks (e.g. the 'classic' ecosystem example used in this study (Fig. 6)). Furthermore, long-term observations are also crucial to assess the interannual variations in GHG fluxes, given that large variations can be an indicator that the given rehabilitated peatland ecosystem is undergoing transition (Wilson et al., 2016).

5. Conclusions

Here, we introduced the 'miniRECgap' R-package, which simplifies the workflow for flux-partitioning and gap-filling of missing eddy covariance (EC) CO_2 flux measurements by enabling researchers to apply conventional, robust empirical and semi-empirical gap-filling methods in five main steps, using GUI supported scripts in just five lines of code, and requiring the minimum number of input variables.

There is significant potential to further develop and expand this package. Future work is recommended to potentially expand some of the 'miniRECgap' functions to allow for more choice in some of their settings when the selected gap-filling methods are applied. Further studies on assessing model performance are also recommended (e.g. by applying different data-gap scenarios, the use of longer datasets and data from a wider range of terrestrial ecosystems).

The main findings from our comparative simple evaluation of the 'miniRECgap' package (against two other gap-filling methods) based on used case-study are summarised in the key points below.

- **Gap-filling performance based on case-study**

In this study, the shallow ANN ('N3' model) outperformed both 'miniRECgap' and 'MSDnight' in the 'challenging' ecosystem (Cavemount Bog). 'MDSnight' performed better in the 'classic' ecosystem (DE-Tha), while 'miniRECgap' showed somewhat better performance in the 'challenging' ecosystem (Cavemount Bog). These findings indicate that testing several gap-filling methods is recommended when dealing with 'challenging' ecosystems. The findings further indicate that 'miniRECgap' performance could potentially improve through application of appropriate data-clustering.

- **Customised flux gap-filling approach**

This study encourages the consideration and application of customised flux gap-filling approaches, where specific site conditions and data availability are assessed in order to select the most suitable method. The selection of the most appropriate gap-filling method will largely depend on:

- o <u>Data availability:</u> The choice of gap-filling approach can depend on the number of input variables required and the quantity and quality of the EC flux data available.
- o <u>Model evaluation and performance metrics:</u> In-depth comprehensive model evaluation is recommended. However, less complex model performance metrics (such as used in this study) may also provide valuable indicative insights into gap-filling method selection.
- o <u>Ecosystem type and characteristics:</u> The complexity and characteristics of the ecosystem in question can influence the choice of the most appropriate gap-filing method; therefore, testing several methods is advised when dealing with 'challenging' environments (such as in the case of disturbed or formerly disturbed ecosystems, or ecosystems that are very heterogeneous in nature, ecosystems under rehabilitation, etc. such as the Cavemount Bog example in this study).

CRediT authorship contribution statement

Alina Premrov: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Jagadeesh Yeluripati:** Writing – review & editing. **Richard Slevin:** Investigation - Field Work. **Adam Bates:** Investigation - Field Work. **Magdalena Matysek:** Investigation - Field Work. **Stephen Barry:** Writing – review & editing. **Kenneth A. Byrne:** Writing – review & editing. **Rowan Fealy:** Writing – review & editing. **Bernard Hyde:** Writing – review & editing. **Gary Lanigan:** Writing – review & editing. **Mark McCorry:** Writing – review & editing. **Rachael Murphy:** Writing – review & editing. **Florence Renou-Wilson:** Writing – review & editing. **Amey Tilak:** Writing – review & editing. **David Wilson:** Writing – review & editing. **Matthew Saunders:** Writing – review & editing, Investigation - Field Work, Data curation.

Software availability

The 'miniRECgap' package and the shallow ANN models applied in this paper are entirely written in R https://www.R-project.org/ (R Core Team, 2024b). The 'miniRECgap' (Version v0.1.0) package/software source-code is publicly available under the 'MIT License + file License' [Copyright (c) Trinity College Dublin 2024] on GitHub https://github.com/APremrov/miniRECgap (Premrov, 2024), DOI https://doi.org/10.5281/zenodo.13228227.

The 'miniRECgap' package (Version v0.1.0) is available under the 'MIT License + file License', with license provided in 'LICENSE.md' and 'LICENSE' files, which can be accessed at following links: https://github.com/APremrov/miniRECgap?tab=MIT-2-ov-file, https://github.com/APremrov/miniRECgap?tab=License-1-ov-file.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been performed under the CO2PEAT Research Project entitled "Improving methodologies for reporting and verifying terrestrial CO_2 removals and emissions from Irish peatlands" (CO2PEAT, 2024). The authors are grateful to the Irish Environmental Protection Agency for funding the CO2PEAT project 2022-CE-1100 under the EPA Research Programme 2021–2030. The EPA Research Programme is a Government of Ireland initiative funded by the Department of the Environment, Climate and Communications. Thanks also to the Smart-BoG Project (2018-CCRP-LS.2) funded by Irish Environmental Protection Agency (EPA) under the EPA Research Programme 2014–2020 and the Terrain-AI project co-funded by Science Foundation Ireland and by Microsoft (20/SPP/3705). We would also like to thank Tadhg O'Mahony from Irish Environmental Protection Agency, Strategic Environmental Assessment Unit, Office of Evidence & Assessment, Inniscarra, Co. Cork, Ireland. Thanks also to the staff members from Trinity Innovation and Enterprise, Trinity College Dublin, Ireland for their advice and support with 'miniRECgap' package/software licensing and related policies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2025.106611.

Data availability

'miniRECgap' package is publicly available under 'MIT License + file License' [© Trinity College Dublin 2024]. https://github.com/APremrov/miniRECgap. https://doi.org/10.5281/zenodo.13228227.

References

AmeriFlux, 2024. Raw data processing and QA/QC. Available at AmeriFlux website:. In: https://ameriflux.lbl.gov/resources/resource-list/tools-and-software-for-flux-scientists/raw-data-processing-and-qa-qc/ (accessed May 2024).

Aslan Sungur, G., VanLoocke, A., Moore, C., Bernacchi, C., Ferin, K., 2019. Process-based models as a gap filling method for eddy covariance measurements of net ecosystem CO 2 exchange: a case study for the perennial grass miscanthus. AGU Fall Meet. Abstr. B41K–2499.

Aubinet, M., Vesala, T., Papale, D., 2012. Eddy Covariance: A Practical Guide to Measurement and Data Analysis. Springer Sci. Bus. Media 365–376. https://www.researchgate.net/publication/254255682/ (accessed May 2024).

Aurela, M., Laurila, T., Tuovinen, J.-P., 2002. Annual CO2 balance of a subarctic fen in northern Europe: importance of the wintertime efflux. J. Geophys. Res. 107, 4607, 4610.1029/2002JD002055.

Baldocchi, D., Valentini, R., 2004. Geographic and temporal variation of carbon exchange by ecosystems and their sensitivity to environmental perturbations. In: Field, C., Raupach, M. (Eds.), Towards CO2 Stabilization: Issues, Strategies and Consequences, A SCOPE/GCP Rapid Assessment Project, pp. 295–316.

Baldocchi, D.D., 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. Glob. Change Biol. 9 (4), 479–492, 410.1046/j.1365-2486.2003.00629.x.

Baldocchi, D.D., Hincks, B.B., Meyers, T.P.J.E., 1988. Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. Ecology 69 (5), 1331–1340, 1310.2307/1941631.

Barr, A.G., Black, T.A., Hogg, E.H., Kljun, N., Morgenstern, K., Nesic, Z., 2004. 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), 237–255. https://doi.org/10.1016/j.agrformet.2004.1006.1011.

Biederman, J.A., Scott, R.L., Bell, T.W., Bowling, D.R., Dore, S., 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., Goulden, M.L., 2017. CO2 exchange and evapotranspiration across dryland ecosystems of Southwestern North America. Glob. Change Biol. 23 (10), 4204–4221.

Biederman, J.A., Scott, R.L., Arnone Iii, J.A., Jasoni, R.L., Litvak, M.E., Moreo, M.T., Papuga, S.A., Ponce-Campos, G.E., Schreiner-McGraw, A.P., Vivoni, E.R., 2018. Shrubland carbon sink depends upon winter water availability in the warm deserts of North America. Agric. For. Meteorol. 249, 407–419.

Boehmke, B., 2018b. K-means cluster analysis. UC Business Analytics R Programming Guide. University of Cincinnati. https://uc-r.github.io/kmeans_clustering/ (accessed May 2024).

Bord na Móna, 2021. Carbon measurements. Bord na Mona. Bord na Móna, Leabeg, Co. Offaly. Ireland. pp. 1-7. https://www.bnmpcas.ie/wp-content/uploads/sites/18/2021/08/Cavemount-rehab-plan-2021-V7.pdf.

Boudhina, N., Zitouna-Chebbi, R., Mekki, I., Jacob, F., Ben Mechlia, N., Masmoudi, M., Prévot, L., 2018. Evaluating four gap-filling methods for eddy covariance measurements of evapotranspiration over hilly crop fields. Geosci. Instrum. Method. Data Syst. 7 (2), 151–167, 110.5194/gi-5197-5151-2018.

Braswell, B.H., Sacks, W.J., Linder, E., Schimel, D.S., 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. Glob. Change Biol. 11 (2), 335–355, 310.1111/j.1365-2486.2005.00897.x.

Broyden, C.G., 1970. TThe convergence of a class of double-rank minimization algorithms: 2. The new algorithm. IMA J. Appl. Math. 6, 222–231, 210.1093/imamat/1096.1093.1222.

Burba, G., Aderson, D., 2010. A Brief Practical Guide to Eddy Covariance Flux Measurements: Principles and Workflow Examples for Scientific and Industrial Applications. LI-COR, Inc., Lincoln, Nebraska, USA. Copyright © 2005-2010 LI-COR, Inc., Version 1.0.1. LI-COR Biosciences. https://www.licor.com/env/pdf/eddy_covariance/Brief_Intro_Eddy_Covariance.pdf.

Burba, G., Anderson, D., Amen, J., 2007. Eddy covariance method: overview of general guidelines and conventional workflow. AGU Fall Meet. Abstr. B33D–1575.

Buzacott, A., van den Berg, M., Kruijt, B., Bataille, L., van der Velde, Y., 2023. A Bayesian approach to gapfilling fluxes from heterogenous Dutch peatlands measured by eddy covariance. EGU General Assembly 2023. https://doi.org/10.5194/egusphere-egu23-12916. Vienna, Austria, 24–28 Apr 2023, EGU23-12916.

Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. 16 (5), 1190–1208, 1110.1137/0916069.

Chen, W., Wang, S., Wang, J., Xia, J., Luo, Y., Yu, G., Niu, S., 2023. Evidence for widespread thermal optimality of ecosystem respiration. Nature Ecol. Evol. 7 (9), 1379–1387, 1310.1038/s41559-41023-02121.

CO2PEAT, 2024. Improving Methodologies for Reporting and Verifying Terrestrial CO2 Removals and Emissions from Irish Peatlands (CO2PEAT). Research project funded by Irish Environmental Protection Agency (EPA) under the EPA Research Programme 2021-2030. https://co2peat.wixsite.com/co2peat/ (accessed Aug 2024).

Cong, H., Yan-Mei, M., Tian-Sha, n.Z., Shu-Gao, Q., Peng, L., Yun, T., Xin, J., 2023. A dataset of ecosystem fluxes in a shrubland ecosystem of Mau Us Sandy Land in Yanchi, Ningxia, China (2012-2016), 47 (9), 1322–1332.

Cotten, D.L., Zhang, G., Leclerc, M.Y., Raymer, P., Steketee, C.J., 2017. Carbon dioxide fluxes from tifway bermudagrass: early results. Int. J. Biometeorol. 61 (1), 103–113, 110.1007/s00484-00016-01194-z.

Desai, A.R., Bolstad, P.V., Cook, B.D., Davis, K.J., Carey, E.V., 2005. Comparing net ecosystem exchange of carbon dioxide between an old-growth and mature forest in the upper midwest, USA. Agric. For. Meteorology 128 (1), 33–55. https://doi.org/10.1016/j.agrformet.2004.1009.1005.

dos Reis, M., Ribeiro, A., 2020. Conversion factors and general equations applied in agricultural and forest meteorology. Agrometeoros 27, 227–258, 210.31062/agrom.v31027i31062.26527.

Douglas, A., 2023. intro2R. URL: https://alexd106.github.io/intro2R/. Date Accessed: Jan. 2024.

Douglas, A., Roos, D., Francesca, M., Couto, A., Lusseau, D., 2023. An Introduction to R. URL: https://github.com/alexd106/Rbook/raw/master/docs/Rbook.pdf.

Evrendilek, F., 2013. Quantifying biosphere–atmosphere exchange of CO2 using eddy covariance, wavelet denoising, neural networks, and multiple regression models. Agric. For. Meteorology 171, 1–8. https://doi.org/10.1016/j.agrformet.2012.1011.1002.

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., 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., Wofsy, S., 2001. Gap filling strategies for defensible annual sums of net ecosystem exchange. Agric. For. Meteorology 107 (1), 43–69. https://doi.org/10.1016/S0168-1923(1000)00225-00222.

FAO, 2022. Introduction to R and RStudio. 2022 Virtual Training on Data Disaggregation and Small Area Estimation for the SDGs. Food and Agriculture Organisation of the United Nations (FAO). URL. https://www.fao.org/3/cc3216en/cc3216en.pdf.

Fletcher, R., 1970. A new approach to variable metric algorithms. Comput. J. 13, 317–322, 310.1093/comjnl/1013.1093.1317.

FLUXNET, 2015. FLUXNET network. Fig. 1 FLUXNET 2015. https://fluxnet.org/about/ (accessed Jun 2023).

FLUXNET, 2022. A rolling list of software/packages for flux-related data processing. FLUXNET.org Blog). https://fluxnet.org/2017/10/10/toolbox-a-rolling-list-of-softwarepackages-for-flux-related-data-processing/ (accessed Jun 2023).

Foken, T., Wichura, B., 1996. Tools for quality assessment of surface-based flux measurements. Agric. For. Meteorology 78 (1), 83–105, 110.1016/0168-1923(1095)02248-02241.

Fritsch, S., Guenther, F., Wright, T.M., 2019. Neuralnet: training of neural networks. R package version 1.44.2. https://CRAN.R-project.org/package=neuralnet. https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf. May 2024.

Fylstra, D., Lasdon, L., Watson, J., Waren, A., 1998. Design and use of the microsoft excel solver. Interfaces 28 (5), 29–55. https://doi.org/10.1287/inte.1228.1285.1229.

GADM, 2018. Ireland - GADM data (version 4.1). Global administrative areas (GADM). © 2018-2022 GADM. https://gadm.org/download_country.html.

GaFiR, 2014. GaFiR-Gap filling program. Department of Micrometeorology. University of Bayreuth, Germany. http://www.bayceer.uni-bayreuth.de/mm/de/software/software/software_dl.php?id_obj=124194 (accessed Sept 2023).

Gao, D., Yao, J., Yu, S., Ma, Y., Li, L., Gao, Z., 2023. Eddy covariance CO2 flux gap filling for long data gaps: a novel framework based on machine learning and time series decomposition. Remote Sens. 15 (10), 2695, 2610.3390/rs15102695.

Gilmanov, T.G., Verma, S.B., Sims, P.L., Meyers, T.P., Bradford, J.A., Burba, G.G., Suyker, A.E., 2003. Gross primary production and light response parameters of four Southern Plains ecosystems estimated using long-term CO2-flux tower measurements. Global Biogeochem.Cycles 17 (2), 1071, 1010.1029/2002GB002023.

Gogo, S., Laurent, A., 2023. Power to the peatlands: measuring methods and integrated model to predict C-emissions and sequestration in natural peatland. INTERREG CARE-PEAT. Report June, 2023. pp.1-25. https://vb.nweurope.eu/media/21532/care-peat_main_output_research-programme-and-model_wpt1.pdf/ (accessed May 2024).

Goldfarb, D., 1970. A family of variable metric methods derived by variational means. Math. Comput. 24, 23–26. https://doi.org/10.1090/S0025-5718-1970-0258249-0258246.

Gosiewska, A., Biecek, P., 2020. Auditor: an R package for model-agnostic visual validation and diagnostic. https://doi.org/10.32614/rj-2019-036.

Gove, J.H., Hollinger, D.Y., 2006. Application of a dual unscented kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange. J. Geophys. Res. 111 (D08S07). https://doi.org/10.1029/2005JD006021.

Grolemund, G., 2014. Hands-On Programming with R. URL: https://rstudio-education.github.io/hopr/ [License: CC BY-NC-ND 4.0].

Grünwald, T., Bernhofer, C., 2007. A decade of carbon, water and energy flux measurements of an old spruce forest at the anchor station tharandt. Tellus B 59, 387–396. https://doi.org/10.1111/j.1600-0889.2007.00259.x.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: a K-means clustering algorithm. J. Royal Statist. Soc. Series C (Applied Statistics) 28, 100–108. https://doi.org/10.2307/2346830.

Heiskanen, L., 2023. Environmental responses of carbon dioxide and methane fluxes of subarctic ecosystems in northern Finland. Doctoral Programme in Atmospheric Sciences, Institute for Atmospheric and Earth System Research/Physics. Faculty of Science. University of Helsinki: FINNISH METEOROLOGICAL INSTITUTE. https://doi.org/10.35614/isbn.9789523361676. CONTRIBUTIONS No. 184.

Hernández-Orallo, J., 2013. ROC curves for regression. Pattern Recogn. 46 (12), 3395–3411. https://doi.org/10.1016/j.patcog.2013.3306.3014.

Hoffmann, T.J., Laird, N.M., 2009. Fgui: a method for automatically creating graphical user interfaces for command-line R packages. J. Stat. Software 30 (2), 1–14. https://doi.org/10.18637/JSS.V18030.I18602.

Holl, D., Pancotto, V., Heger, A., Camargo, S.J., Kutzbach, L., 2019. Cushion bogs are stronger carbon dioxide net sinks than moss-dominated bogs as revealed by eddy covariance measurements on Tierra del Fuego, Argentina. Biogeosciences 16 (17), 3397–3423, 3310.5194/bg-3316-3397-2019.

Holl, D., Pfeiffer, E.M., Kutzbach, L., 2020. Comparison of eddy covariance CO2 and CH4 fluxes from mined and recently rewetted sections in a northwestern German cutover bog. Biogeosciences 17 (10), 2853–2874, 2810.5194/bg-2817-2853-2020.

Hollinger, D.Y., Aber, J., Dail, B., Davidson, E.A., Goltz, S.M., Hughes, H., Leclerc, M.Y., Lee, J.T., Richardson, A.D., Rodrigues, C., Scott, N.A., Achuatavarier, D., Walsh, J., 2004. Spatial and temporal variability in forest–atmosphere CO2 exchange. Glob. Change Biol. 10 (10), 1689–1706. https://doi.org/10.1111/j.1365-2486.2004.00847.x.

Hollinger, D.Y., Richardson, A.D., 2005. Uncertainty in eddy covariance measurements and its application to physiological models. Tree Physiol. 25 (7), 873–885. https://doi.org/10.1093/treephys/1025.1097.1873.

Hui, D., Wan, S., Su, B., Katul, G., Monson, R., Luo, Y., 2004. Gap-filling missing data in eddy covariance measurements using multiple imputation (MI) for annual estimations. Agric. For. Meteorology 121 (1), 93–111. https://doi.org/10.1016/S0168-1923(1003)00158-00158.

ICOS, 2022. ICOS Handbook 2022. Integrated Carbon Observation System (ICOS) research Infrastructure. Helsinki, Finland, 978-952-69501-5-0. https://www.icos-cp.eu/sites/default/files/2022-03/ICOS_handbook_2022_WEB.pdf/ (accessed Jun 2023).

ICOS, 2023. ICOS in a nutshell. Integrated Carbon Observation System (ICOS). https://www.icos-cp.eu/about/icos-in-nutshell/ (accessed Jun 2023).

Ingle, R., Bhatnagar, S., Ghosh, B., Gill, L., Regan, S., Connolly, J., Saunders, M., 2023. Development of hybrid models to estimate gross primary productivity at a near-natural peatland using sentinel 2 data and a light use efficiency model. Remote Sens. 15 (1), 673, 610.3390/rs15061673.

Irvin, J., Zhou, S., McNicol, G., Lu, F., Liu, V., Fluet-Chouinard, E., Ouyang, Z., Knox, S. H., Lucas-Moffat, A., Trotta, C., Papale, D., Vitale, D., Mammarella, I., Alekseychik, P., Aurela, M., Avati, A., Baldocchi, D., Bansal, S., Bohrer, G., Campbell, D.I., Chen, J., Chu, H., Dalmagro, H.J., Delwiche, K.B., Desai, A.R., Euskirchen, E., Feron, S., Goeckede, M., Heimann, M., Helbig, M., Helfter, C., Hemes, K.S., Hirano, T., Iwata, H., Jurasinski, G., Kalhori, A., Kondrich, A., Lai, D.Y. F., Lohila, A., Malhotra, A., Merbold, L., Mitra, B., Ng, A., Nilsson, M.B., Noormets, A., Peichl, M., Rey-Sanchez, A.C., Richardson, A.D., Runkle, B.R.K., Schäfer, K.V.R., Sonnentag, O., Stuart-Haëntjens, E., Sturtevant, C., Ueyama, M., Valach, A.C., Vargas, R., Vourlitis, G.L., Ward, E.J., Wong, G.X., Zona, D., Alberto, M. C.R., Billesbach, D.P., Celis, G., Dolman, H., Friborg, T., Fuchs, K., Gogo, S., Gondwe, M.J., Goodrich, J.P., Gottschalk, P., Hörtnagl, L., Jacotot, A., Koebsch, F., Kasak, K., Maier, R., Morin, T.H., Nemitz, E., Oechel, W.C., Oikawa, P.Y., Ono, K., Sachs, T., Sakabe, A., Schuur, E.A., Shortt, R., Sullivan, R.C., Szutu, D.J., Tuittila, E.-S., Varlagin, A., Verfaillie, J.G., Wille, C., Windham-Myers, L., Poulter, B., Jackson, R.B., 2021. Gap-filling eddy covariance methane fluxes: Comparison of machine learning model predictions and uncertainties at FLUXNET-CH4 wetlands. Agric. For. Meteorol., 108528, 108510.101016/j.agrformet.102021.108528.

Isaac, P., Cleverly, J., McHugh, I., van Gorsel, E., Ewenz, C., Beringer, J., 2017. OzFlux data: network integration from collection to curation. Biogeosciences 14 (12), 2903–2928, 2910.5194/bg-2914-2903-2017.

Ito, D., Ishida, S., 2023. The effect of periodical grass mowing and various meteorological factors on CO2 flux in a sod-cultured apple orchard. J. Agric. Meteorol. 79 (1), 18–27. https://doi.org/10.2480/agrmet.D-2422-00010.

Jarvis, P., James, G.B., Landsberg, J.J., 1976. Coniferous forest. In: Monteith, J.L. (Ed.), Vegetation and the Atmosphere, Vol. II. Case Studies. Academic Press, London, pp. 171–240.

Jia, X., Zha, T.S., Wu, B., Zhang, Y.Q., Gong, J.N., Qin, S.G., Chen, G.P., Qian, D., Kellomäki, S., Peltola, H., 2014. Biophysical controls on net ecosystem CO_2 exchange over a semiarid shrubland in northwest China. Biogeosciences 11 (17), 4679–4693, 4610.5194/bg-4611-4679-2014.

Jones, M., Osborne, B., Williams, M., Saunders, M., Lanigan, G., Burke, J., Davis, P., Abdalla, M., Clifton-Brown, J., Connolly, J., Kumar, S., Nagy, M., 2010. Climate change – accounting for greenhouse gas sources and sinks in major Irish land-use categories:. Towards the Establishment of a Co-ordinating. Centre for FLUX Measurements (CCFLUX. STRIVE. Report Series No.43. Environ. Protect. Agen. Program. 2007-2013. Report Series No.43, pp. 1-40. https://www.epa.ie/publications/research/climate-change/EPA_Strive43_CCFLUX_final_web.pdf/ (accessed Jun 2024).

Joseph, V.R., 2022. Optimal ratio for data splitting. Stat. Anal. Data Min.: ASA Data Sci. J. 15 (4), 531–538, 510.1002/sam.11583.

Juszczak, R., Humphreys, E., Acosta, M., Michalak-Galczewska, M., Kayzer, D., Olejnik, J., 2013. Ecosystem respiration in a heterogeneous temperate peatland and its sensitivity to peat temperature and water table depth. Plant Soil 366, 505–520. https://doi.org/10.1007/s11104-11012-11441-y.

Kiely, G., Leahy, P., Lewis, C., Sottocornola, M., Laine, A., Koehler, A.-K., 2018. GHG Fluxes from Terrestrial Ecosystems in Ireland. Environmental Protection Agency, Ireland. CCRP08Proj-1.1A). EPA Research Report. https://www.epa.ie/publications/research/climate-change/Research_Report_227.pdf/ (accessed Jun 2023).

Knox, S.H., Sturtevant, C., Matthes, J.H., Koteen, L., Verfaillie, J., Baldocchi, D., 2015. Agricultural peatland restoration: effects of land-use change on greenhouse gas (CO2 and CH4) fluxes in the Sacramento-San Joaquin Delta. Glob. Change Biol. 21 (2), 750–765. https://doi.org/10.1111/gcb.12745.

Lasslop, G., Rechstein, M., Papale, D., Richardson, A.D., Arneth, A., Barr, A., Stou, P., Wohlfahrt, G., 2010. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. Glob. Change Biol. 16 (1), 187–208. https://doi.org/10.1111/j.1365-2486.2009.02041.x.

Law, B.E., Turner, D., Campbell, J., Lefsky, M., Guzy, M., Sun, O., Tuyl, S.V., Cohen, W., 2006. Carbon fluxes across regions: observational constrains at multiple scales. In: Wu, J., Jones, K.B., Li, H., Loucks, O.L. (Eds.), Scaling and Uncertainty Analysis in Ecology. Springer Netherlands, Dordrecht, pp. 167–190, 110.1007/1001-4020-4663-1004_1009.

Lees, K.J., Khomik, M., Quaife, T., Clark, J.M., Hill, T., Klein, D., Ritson, J., Artz, R.R.E., 2021. Assessing the reliability of peatland GPP measurements by remote sensing: from plot to landscape scale. Sci. Total Environ. 766, 142613, 142610.141016/j.scitotenv.142020.142613.

Lees, K.J., Quaife, T., Artz, R.R.E., Khomik, M., Sottocornola, M., Kiely, G., Hambley, G., Hill, T., Saunders, M., Cowie, N.R., Ritson, J., Clark, J.M., 2019. A model of gross primary productivity based on satellite data suggests formerly afforested peatlands undergoing restoration regain full photosynthesis capacity after five to ten years. J. Environ. Manag. 246, 594–604. https://doi.org/10.1016/j.jenvman.2019.1003.1040.

LI-COR, 2023. Tovi. LI-COR Inc. USA. URLs: https://www.licor.com/env/support/Tovi/software.html https://www.licor.com/env/support/Tovi/manuals.html/ (accessed Jun 2024).

LI-COR Environmental, 2025. LI-190R quantum sensor LI-COR Inc., USA. URLs. www.licor.com. https://www.licor.com/products/light/quantum (accessed Jan 2025).

Lloyd, A.R., 2010. Carbon Fluxes at an Upland Blanket Bog in the North Pennines. (Phd Thesis), Department of Biological and Biomedical Sciences. Durham University: Durham University, Durham. UK. Available at: Durham E-Theses Onlin http://etheses.dur.ac.uk/192/ (accessed Jun 2023).

Lloyd, J., Taylor, J.A., 1994. On the temperature dependence of soil respiration. Funct. Ecol. 8 (3), 315–323, 310.2307/2389824.

Lucas-Moffat, A.M., Schrader, F., Herbst, M., Brümmer, C., 2022. Multiple gap-filling for eddy covariance datasets. Agric. For. Meteorology 325, 109114, 109110.101016/j.agrformet.102022.109114.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2022. Cluster: cluster analysis basics and extensions. R package version 2 (1), 4. https://CRAN.R-project.org/package=cluster.

Mahabbati, A., 2022. Investigating the Application of Machine Learning Models to Improve the Eddy Covariance Data gap-filling. The University of Western Australia. School of Agriculture and Environment. PhD Thesis. https://api.research-repository.uwa.edu.au/ws/portalfiles/portal/197310249/THESIS_DOCTOR_OF_PHILOSOPHY_MAHABBATI_Atbin_2022.pdf/ (accessed Jun 2023).

Mahabbati, A., Beringer, J., Leopold, M., McHugh, I., Cleverly, J., Isaac, P., Izady, A., 2021. A comparison of gap-filling algorithms for eddy covariance fluxes and their

drivers. Geosci. Instrum. Method. Data Syst. 10 (1), 123–140, 110.5194/gi-2020-5121.
Maindonald, J.H., 2008. Using R for Data Analysis and Graphics. Introduction, Code and Commentary. Centre for Mathematics and Its Applications. Australian National University. URL. https://cran.r-project.org/doc/contrib/usingR.pdf.
Melesse, A.M., Hanley, R.S., 2005. Artificial neural network application for multi-ecosystem carbon flux simulation. Ecol. Model. 189 (3), 305–314. https://doi.org/10.1016/j.ecolmodel.2005.1003.1014.
Meng, C., Xiao, X., Wagle, P., Zhang, C., Pan, L., Pan, B., Qin, Y., Newman, Gregory S., 2024. Exponential or unimodal relationships between nighttime ecosystem respiration and temperature at the Eddy covariance flux tower sites. Ecol. Lett. 27 (10), e14532, 14510.11111/ele.14532.
Menzer, O., Meiring, W., Kyriakidis, P.C., McFadden, J.P., 2015. Annual sums of carbon dioxide exchange over a heterogeneous urban landscape through machine learning based gap-filling. Atmos. Environ. 101, 312–327. https://doi.org/10.1016/j.atmosenv.2014.1011.1006.
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., Stauch, V.J., 2007. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. Agric. For. Meteorology 147 (3), 209–232. https://doi.org/10.1016/j.agrformet.2007.1008.1011.
MPI, 2024c. REddyProc Web online tool. FAQ. "Why does negative NEE define an net uptake of the ecosystem? (#noRg). Max Planck Institute (MPI) for Biogeochemistry. https://www.bgc-jena.mpg.de/5629512/FAQ. Jun 2023 - May 2024.
MPI, 2024d. REddyProc Web online tool. REddyProcWeb Input format. Max Planck Institute (MPI) for Biogeochemistry. https://www.bgc-jena.mpg.de/5624918/Input-Format. Jun 2023 - May 2024.
MPI, 2024e. REddyProc: eddy covariance data post-processing tool. (web-tool). Department of Biogeochemical Integration. Max Planck Institute (MPI) for Biogeochemistry. In: https://www.bgc-jena.mpg.de/REddyProc/ui/REddyProc.php. Jun 2023 - May 2024.
MPI, 2024f. REddyProcWeb: methods. Ustar filtering. Department of biogeochemical integration. Max Planck Institute (MPI) for Biogeochemistry. Germany. https://www.bgc-jena.mpg.de/5624872/Ustar-filtering. Jun 2023 - Mat 2024.
Murphy, R., 2022. Investigating the Role of Management and Measurement Technique on the Temporal and Spatial Variability of Carbon Dynamics and Nitrous Oxide Emissions from Temperate Grasslands. (Phd Thesis). , Department of Botany. School of Natural Sciences, Trinity College Dublin: Dublin. Ireland, p. 240. http://www.tara.tcd.ie/handle/2262/98488.
Murphy, R.M., Saunders, M., Richards, K.G., Krol, D.J., Gebremichael, A.W., Rambaud, J., Cowan, N., Lanigan, G.J., 2022. Nitrous oxide emission factors from an intensively grazed temperate grassland: a comparison of cumulative emissions determined by eddy covariance and static chamber methods. Agric. Ecosyst. Environ. 324, 107725, 107710.101016/j.agee.102021.107725.
Myklebust, M.C., Hipps, L.E., Ryel, R.J., 2008. Comparison of eddy covariance, chamber, and gradient methods of measuring soil CO2 efflux in an annual semi-arid grass. Bromus tectorum. Agric. and For. Meteorology 148 (11), 1894–1907. https://doi.org/10.1016/j.agrformet.2008.1806.1016.
ONEFlux, 2023. ONEFlux processing pipeline. By Ameriflux Management Project (Ameriflux), European Fluxes Database (Europe-Fluxdata), ICOS Ecosystem Thematic Centre (ICOS-ETC) (Available on GitHub. under the FLUXNET collaboration account: ONEFlux). https://github.com/fluxnet/ONEFlux/ (accessed Jun 2023).
Ooba, M., Hirano, T., Mogami, J.-I., Hirata, R., Fujinuma, Y., 2006. Comparisons of gap-filling methods for carbon flux dataset: a combination of a genetic algorithm and an artificial neural network. Ecol. Model. 198 (3), 473–486. https://doi.org/10.1016/j.ecolmodel.2006.1006.1006.
OpenStreetMap & Contributors, 2024. Base map and data from OpenStreetMap and OpenStreetMap Foundation (CC-BY-SA). https://www.openstreetmap.organdcontributors.License:OpenDataCommonsOpenDatabaseLicense(ODbL. Creative Commons Attribution-ShareAlike (CC-BY-SA). (accessed Jun 2024).
Papale, D., 2020. Ideas and perspectives: enhancing the impact of the FLUXNET network of eddy covariance sites. Biogeosciences 17 (22), 5587–5598, 5510.5194/bg-5517-5587-2020.
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., Reichstein, M., Ribeca, A., van Ingen, C., Vuichard, N., 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., 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., Cinti, B.D., Grandcourt, A.d., Ligne, A.D., De Oliveira, R.C., Delpierre, N., Desai, A.R., Di Bella, C.M., Tommasi, P.d., 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, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gitelson, A., Goded, I., Goeckede, M., Goldstein, A.H., Gough, C.M., Goulden, M.L., Graf, A., Griebel, A., Gruening, C., Grünwald, T., Hammerle, A., Han, S., Han, X., Hansen, B.U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hörtnagl, L., Hutley, L., Ibrom, A., Ikawa, H., Jackowicz-Korczynski, M., Janouš, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W.L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Y., Liddell, M., Limousin, J.-M., Lion, M., Liska, A.J., Lohila, A., López-Ballesteros, A., López-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., Lüers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H.A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J.H., Mazzenga, F., McCaughey, H., McHugh, I., McMillan, A.M.S., Merbold, L., Meyer, W., Meyers, T., Miller, S.D., Minerbi, S., Moderow, U., Monson, R.K., Montagnani, L., Moore, C.E., Moors, E., Moreaux, V., Moureaux, C., Munger, J.W., Nakai, T., Neirynck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J.E., Ourcival, J.-M., Papuga, S.A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R.P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S.M., Rambal, S., Rannik, Ü., Raz-Yaseef, N., Rebmann, C., Reed, D., Dios, V.R.d., Restrepo-Coupe, N., Reverter, B.R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S.R., Sánchez-Cañete, E.P., Sanchez-Mejia, Z.M., Schmid, H.P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R.L., Sedlák, P., Serrano-Ortíz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Šigut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R.M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., van der Molen, M., van Gorsel, E., van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J.P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Y., Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., Papale, D., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci. Data 7 (1), 225. https://doi.org/10.1038/s41597-41020-40534-41593.
Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M., 2008. How to evaluate models: observed vs. predicted or predicted vs. observed? Ecol. Model. 216 (3), 316–322. https://doi.org/10.1016/j.ecolmodel.2008.1005.1006.
Premrov, A., 2024. 'miniRECgap': R-Package for gap-filling of the Missing Eddy Covariance CO2 Flux Measurements Using Selected Classic Nonlinear Environmental Response Functions via Simple user-friendly GUI Supported R Scripts. Copyright ©, Trinity College Dublin (v0.1.0).2024. https://github.com/APremrov/miniRECgap. https://doi.org/10.5281/zenodo.13228228; https://doi.org/10.5281/zenodo.13228227.
PyFluxPro, 2024. PyFluxPro (Available on GitHub. under the OzFlux: PyFluxPro) (accessed Jun 2023 - May 2024). https://github.com/OzFlux/PyFluxPro.
R Core Team, 2024a. Optim. In: Package: Stats. R: a Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/optim.html (accessed Jun 2023 - Aug 2024).
R Core Team, 2024b. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.
Rabinowitch, E.I., 1951. Photosynthesis and Related Processes. Interscience Publishers, New York, London.
Rebmann, C., Aubinet, M., Schmid, H., Arriga, N., Aurela, M., Burba, G., Clement, R., De Ligne, A., Fratini, G., Gielen, B., Grace, J., Graf, A., Gross, P., Haapanala, S., Herbst, M., Hörtnagl, L., Ibrom, A., Joly, L., Kljun, N., Kolle, O., Kowalski, A., Lindroth, A., Loustau, D., Mammarella, I., Mauder, M., Merbold, L., Metzger, S., Mölder, M., Montagnani, L., Papale, D., Pavelka, M., Peichl, M., Roland, M., Serrano-Ortiz, P., Siebicke, L., Steinbrecher, R., Tuovinen, J.-P., Vesala, T., Wohlfahrt, G., Franz, D., 2018. ICOS eddy covariance flux-station site setup: a review. Int. Agrophys. 32 (4), 471–494. https://doi.org/10.1515/intag-2017-0044.
Reichle, D., 2019. The Global Carbon Cycle and Climate Change. Elsevier.
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., 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., Valentini, R., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Global Change Biol. 11 (9), 1424–1439. https://doi.org/10.1111/j.1365-2486.2005.001002.x.
Richardson, A.D., Braswell, B.H., Hollinger, D.Y., Burman, P., Davidson, E.A., Evans, R. S., Flanagan, L.B., Munger, J.W., Savage, K., Urbanski, S.P., Wofsy, S.C., 2006. Comparing simple respiration models for eddy flux and dynamic chamber data. Agric. For. Meteorology 141 (2), 219–234. https://doi.org/10.1016/j.agrformet.2006.1010.1010.
Ryan, P.A., 2019. K-Means clustering and related algorithms. In: Unsupervised Learning. COS 324 – Elements of Machine Learning. Princeton University. https://www.cs.princeton.edu/courses/archive/spring19/cos324/files/kmeans.pdf (accessed Jun 2023 - Jun 2024).
Sabbatini, S., Mammarella, I., Arriga, N., Fratini, G., Graf, A., Hörtnagl, L., Ibrom, A., Longdoz, B., Mauder, M., Merbold, L., Metzger, S., Montagnani, L., Pitacco, A., Rebmann, C., Sedlák, P., Šigut, L., Vitale, D., Papale, D., 2018. Eddy covariance raw data processing for CO2 and energy fluxes calculation at ICOS ecosystem stations. Int. Agrophys. 32 (4), 495–515. https://doi.org/10.1515/intag-2017-0043.
Shanno, D.F., 1970. Conditioning of quasi-newton methods for function minimization. Math. Comput. 24, 647–650. https://www.jstor.org/stable/2004840, 2004810.2002307/2004840.
Šigut, L., 2012. Improvement of CO2 eddy fluxes modelling in topographically complex terrain. SVK Conference 2012. University of Ostrava. Czech Republic. https

://konference.osu.cz/svk/sbornik2012/pdf/budoucnost/fyzika/Sigut.L.pdf/ (accessed Jun 2023).
Singh, S., 2008. Soil respiration processes in Canadian boreal forest soils following fire. Msc Thesis in Soil Science), Department of Renewable Resources. University of Alberta, Edmonton, Alberta. https://era.library.ualberta.ca/items/2e5581b4-15e6-444e-b82a-c68de21e645b.
SmartBOG, 2020. Smart Bogland Observation Group (Smartbog). Research Project Funded. Irish Environmental Protection Agency (EPA) under the EPA Research Programme 2014-2020.
Stauch, V.J., Jarvis, A.J., 2006. A semi-parametric gap-filling model for eddy covariance CO2 flux time series data. Glob. Change Biol. 12 (9), 1707–1716. https://doi.org/10.1111/j.1365-2486.2006.01227.x.
Strack, M., Softa, D., Bird, M., Xu, B., 2018. Impact of winter roads on boreal peatland carbon exchange. Glob. Change Biol. 24 (1), e201–e212. https://doi.org/10.1111/gcb.13844.
Torfs, P., Brauer, C., 2014. A (very) short introduction to R. Hydrology and Quantitative Water Management Group. Wageningen University, The Netherlands. URL. https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf.
Ukkola, A.M., Haughton, N., De Kauwe, M.G., Abramowitz, G., Pitman, A.J., 2017. FluxnetLSM R package (v1.0): a community tool for processing FLUXNET data for use in land surface modelling. Geosci. Model Dev. (GMD) 10 (9), 3379–3390, 3310.5194/gmd-3310-3379-2017.
Vekuri, H., Tuovinen, J.-P., Kulmala, L., Papale, D., Kolari, P., Aurela, M., Laurila, T., Liski, J., Lohila, A., 2023. A widely-used eddy covariance gap-filling method creates systematic bias in carbon balance estimates. Sci. Rep. 13 (1), 1720. https://doi.org/10.1038/s41598-41023-28827-41592.
Vitale, D., Fratini, G., Bilancia, M., Nicolini, G., Sabbatini, S., Papale, D., 2020. A robust data cleaning procedure for eddy covariance flux measurements. Biogeosciences 17 (6), 1367–1391. https://doi.org/10.5194/bg-1317-1367-2020.
Wang, X., Ma, Y., Li, F., Chen, Q., Sun, S., Ma, H., Zhang, R., 2023. Gap filling method and estimation of net ecosystem CO2 exchange in alpine wetland of Qinghai–Tibet Plateau. Sustainability 15 (5), 4652.
Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York. https://doi.org/10.1007/978-3-319-24277-4.
Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., 2023. Dplyr: Agrammar of data manipulation. R package version 1.1.2. Available at: https://CRAN.R-project.org/package=dplyr https://cran.r-project.org/web/packages/dplyr/dplyr.pdf.
Wickham, H., Hester, J., Chang, W., Bryan, J., 2022. Devtools: tools to make developing R packages easier. R package version 2.4.5. Available at: https://CRAN.R-project.org/package=devtools https://cran.r-project.org/web/packages/devtools/devtools.pdf.
Wilson, D., Farrell, C.A., Fallon, D., Moser, G., Müller, C., Renou-Wilson, F., 2016. Multiyear greenhouse gas balances at a rewetted temperate peatland. Glob. Change Biol. 22 (12), 4080–4095. https://doi.org/10.1111/gcb.13325.
Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O., Reichstein, M., 2018. Basic and extensible post-processing of eddy covariance flux data with REddyProc. Biogeosciences 15 (16), 5015–5030, 5010.5194/bg-5015-5015-2018.
Ye, Z.P., 2007. A new model for relationship between irradiance and the rate of photosynthesis in Oryza sativa. Photosynthetica 45 (4), 637–640. https://doi.org/10.1007/s11099-11007-10110-11095.
Zhao, P., Lü ers, J., Foken, T., 2014. Gafir: a gap-filling Package for ecosystem-atmosphere Carbon Dioxide Flux and Evapotranspiration Data: Department of Micrometeorology. University of Bayreuth. ISSN 1614-8924. https://epub.uni-bayreuth.de/id/eprint/1702/1/GaFiR_work_report_final_JL.pdf/ (accessed Jun 2023).
Zhu, S., McCalmont, J., Cardenas, L.M., Cunliffe, A.M., Olde, L., Signori-Müller, C., Litvak, M.E., Hill, T., 2023. Gap-filling carbon dioxide, water, energy, and methane fluxes in challenging ecosystems: comparing between methods, drivers, and gap-lengths. Agric. For. Meteorology 332, 109365, 109310.101016/j.agrformet.102023.109365.

Further reading including Supplemental Material References

Afendras, G., Markatou, M., 2019. Optimality of training/test size and resampling effectiveness in cross-validation. J. of Statistical Plan. and Inference 199, 286–301, 210.1016/j.jspi.2018.1007.1005.
Boehmke, B., 2018a. Artificial neural network fundamentals. UC Business Analytics R Programming Guide. University of Cincinnati. https://uc-r.github.io/ann_fundamentals/ (accessed May 2024).
Dubbs, A., 2021-2022. Test Set Sizing Via Random Matrix Theory. Preprint version 2021. https://doi.org/10.21203/rs.3.rs-3173147/v1. Preprint version of 24 Jul 2022 DOI: https://doi.org/10.48550/arXiv.2112.05977. URL: https://arxiv.org/pdf/2112.05977.pdf.
Hagan, M.T., Demuth, H.B., Beale, M.H., De Jesús, O., 2014. Neural Network Design. 2nd Edtion. eBook, second ed., p. 800 ISBN-10 0971732116. ISBN-13 978-0971732117. (print-length). https://hagan.okstate.edu/NNDesign.pdf. (Accessed 1 September 2014). https://hagan.okstate.edu/nnd.html.
Kassambara, A., Mundt, F., 2020. _factoextra: extract and visualize the results of multivariate data analyses_. R package version 1.0.7. Available at: https://CRAN.R-project.org/package=factoextra https://cran.r-project.org/web/packages/factoextra/factoextra.pdf.
Moritz, S., Bartz-Beielstein, T., 2017. imputeTS: time series missing value imputation in R. The R J 9 (1), 207–218. https://doi.org/10.32614/RJ-32017-32009.
MPI, 2024a. REddyProc Web online tool. Max Planck Institute (MPI) for Biogeochemistry. https://www.bgc-jena.mpg.de/5622399/REddyProc. Jun 2023 - May 2024.
MPI, 2024b. REddyProc Web online tool. FAQ. "I have no incoming solar radiation (Rg) - what can I do? (#noRg). Max Planck Institute (MPI) for Biogeochemistry. https://www.bgc-jena.mpg.de/5629512/FAQ. Jun 2023, May 2024.
Plant, R.E., 2020. Spatial data analysis using artificial neural networks. October 4, 2020. Additional Topic to Accompany Spatial Data Analysis in Ecology and Agriculture Using R, second ed. University of California Davis https://psfaculty.plantsciences.ucdavis.edu/plant/additionaltopics_ann.pdf/. (accessed May, 2024).
Quast, B., 2022. _sigmoid: sigmoid functions for machine learning_. R package version 1.4.0. Available at: https://CRAN.R-project.org/package=sigmoid https://cran.r-project.org/web/packages/sigmoid/sigmoid.pdf.
Riedmiller, M., 1994. Rprop - Description and implementation details. Technical Report University of Karlsruhe. Karlsruhe. Technical Report. January 1994. Available at: https://www.inf.fu-berlin.de/lehre/WS06/Musterererkennung/Paper/rprop.pdf (accessed Jun 2024 - May 2024).
Tibshirani, R., Walther, G., Hastie, T., 2002. Estimating the number of clusters in a data set via the gap statistic. J. Roy. Stat. Soc. B 63 (2), 411–423. https://doi.org/10.1111/1467-9868.00293.
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Mille, r.E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. J. Open Source Softw. 4 (43), 1686. https://doi.org/10.21105/joss.01686.
Winston, P., 2016. 12a: Neural Nets & 12b: Deep neural nets. MIT 6.034 artificial intelligence, fall 2010. Complete course. http://ocw.mit.edu/6-034F10.MITOpenCourseWare.URLs. https://www.youtube.com/watch?v=uXt8qF2Zzfo. https://www.youtube.com/watch?v=VrMHA3yX_QI. License: Creative Commons BY-NC-SA.
Wutzler, et al., 2024. R package ‘REddyProc’. Jan 25, 2024. In: https://cran.r-project.org/web/packages/REddyProc/REddyProc.pdf.