

Gap-filling continuously-measured soil respiration data: A highlight of time-series-based methods

Junbin Zhao*, Holger Lange, Helge Meissner

Department of Biogeochemistry and Soil Quality, Division of Environment and Natural Resources, Norwegian Institute of Bioeconomy Research, Ås, Norway

ARTICLE INFO

Keywords:

Carbon flux

Non-linear least squares

Artificial neural networks

Singular spectrum analysis

Expectation-maximization

Imputation

ABSTRACT

Soil respiration is an important ecosystem process that releases carbon dioxide into the atmosphere. While soil respiration can be measured continuously at high temporal resolutions, gaps in the dataset are inevitable, leading to uncertainties in carbon budget estimations. Therefore, robust methods used to fill the gaps are needed. The process-based non-linear least squares (NLS) regression is the most widely used gap-filling method, which utilizes the established relationship between the soil respiration and temperature. In addition to NLS, we also implemented three other methods based on: 1) artificial neural networks (ANN), driven by temperature and moisture measurements, 2) singular spectrum analysis (SSA), relying only on the time series itself, and 3) the expectation-maximization (EM) approach, referencing to parallel flux measurements in the spatial vicinity. Six soil respiration datasets (2017–2019) from two boreal forests were used for benchmarking. Artificial gaps were randomly introduced into the datasets and then filled using the four methods. The time-series-based methods, SSA and EM, showed higher accuracies than NLS and ANN in small gaps (<1 day). In larger gaps (15 days), the performance was similar among NLS, SSA and EM; however, ANN showed large errors in gaps that coincided with precipitation events. Compared to the observations, gap-filled data by SSA showed similar degree of variances and those filled by EM were associated with similar first-order autocorrelation coefficients. In contrast, data filled by both NLS and ANN exhibited lower variance and higher autocorrelation than the observations. For estimations of the annual soil respiration budget, NLS, SSA and EM resulted in errors between –3.7% and 5.8% given the budgets ranged from 463 to 1152 g C m⁻² year⁻¹, while ANN exhibited larger errors from –11.3 to 16.0%. Our study highlights the two time-series-based methods which showed great potential in gap-filling carbon flux data, especially when environmental variables are unavailable.

1. Introduction

The quantification of carbon (C) exchanges between natural ecosystems and the atmosphere, in the form of carbon dioxide (CO₂), has attracted great attention over the last few decades in the context of climate change (IPCC, 2013). With the development of automatic and continuous measurements at high temporal resolutions (e.g., eddy covariance), it is now possible to estimate the C budget of the ecosystems to a certain degree of accuracy and evaluate their contributions to climate change (e.g., Lu et al., 2017; Reichstein et al., 2007).

Among the key components of the C cycle in an ecosystem, soil respiration is the process that releases the CO₂ produced by roots, soil organisms and chemical oxidation into the atmosphere (Lloyd and Taylor, 1994). Soil respiration is usually measured using either static or dynamic chambers (Bekku et al., 1997; Jensen et al., 1996; Rochette et al., 1992). Automated soil respiration measurements have

been performed in many studies at a temporal interval of 10–60 min (e.g., Suh et al., 2006; Wu et al., 2016). Among the approaches, the forced diffusion (FD) technique was developed based on the theory of dynamic chambers but has the advantage of low power consumption (see details in Risk et al., 2011). At the same time, FD chambers require no mechanical movement between or during measurements which allows them to operate even under snow cover (Lavoie et al., 2015). The FD chambers have been demonstrated to produce comparable results to conventional chambers under the same conditions (Risk et al., 2011). The FD technique, by producing data continuously throughout the year, provides a firm ground for estimating budgets of soil respiration in an ecosystem.

As with all automatic field measurements, missing data, due to e.g., power failure, instrumentation malfunction, disturbance during maintenance, data corruption, etc. are almost inevitable. Depending on the size and number of gaps, large uncertainties may be introduced to the

* Corresponding author at: Høgskoleveien 8, N-1433 Ås, Norway.

E-mail address: junbin.zhao@nibio.no (J. Zhao).

Table 1

Summary of the 6 studied soil respiration datasets.

Site	Plot #	Period	# of data points	
			Available	Missing
Hurdal	H1	2017/7/27–2019/4/29	89,290	2972
	H2	2018/5/9–2019/3/20	39,974	5355
	H3	2017/9/18–2019/4/29	70,646	14,036
	H4	2017/9/8–2019/4/29	81,208	4905
Løten	L1	2018/6/1–2019/4/18	46,291	8
	L2	2018/3/14–2019/4/29	51,062	8148

Date format: year/month/day. Data temporal resolution: 10-min.

Table 2

Summary of the analysis of variance (ANOVA) for the mixed effects models with RMSE as dependent variable and sampling window size, method, season, gap size and the two-way interactions between window size and the other three factors as independent variables.

Independent variable	SS	F-value	P-value
Window size	3.63	8.42	<0.001
Method	39.76	184.32	<0.001
Gap size	79.20	367.16	<0.001
Season	333.39	4636.62	<0.001
Window size*Method	2.44	1.89	0.013
Window size*Gap size	1.17	0.91	0.572
Window size*Season	0.70	1.62	0.136

The plot number is included in the models as random effects. SS: type III sums of squares with Satterthwaite approximation for the degrees of freedom. RMSE: root mean squared error.

estimated C budget (Falge et al., 2001). Therefore, robust and suitable gap-filling methods are critical to ensure the accuracy of C budget estimations.

The gap-filling of carbon flux data usually takes auxiliary variables measured within or near the flux footprint to generate models for predicting missing values in the gaps. For soil respiration, its exponential relationship with soil temperature has long been acknowledged (Lloyd and Taylor, 1994). This relationship is thus most widely used as the basis of the non-linear least squares (NLS) method for soil respiration gap-filling (Gomez-Casanovas et al., 2013). Apart from soil temperature, soil moisture is another driver of soil respiration (Cook and Orchard, 2008). However, the relationship between soil moisture and soil respiration is more complicated because other factors are heavily entangled in it (e.g., time of soil rewetting) (Cook and Orchard, 2008). Many types of models have been used to describe the relationship, such as linear functions (Cook and Orchard, 2008), quadratic polynomials (Hersh et al., 2017), or the moisture effect was even included in the temperature-based model as a limitation term (Reichstein et al., 2003). Thus, it is challenging to choose the proper mathematical form to incorporate soil moisture in the models for soil respiration gap-filling. As a data-driven alternative, artificial neural networks (ANNs) use machine learning algorithms to simulate the output variable (soil respiration) based on the inputs (e.g., soil temperature and moisture) without explicitly specifying the mathematical relationships (Zhang et al., 1998). ANNs have been successfully implemented in gap-filling the ecosystem CO₂ exchange data from eddy covariance observations with multiple environmental variables (including soil moisture) (Moffat et al., 2007); however, their performance in soil respiration data gap-filling has yet to be evaluated.

Despite low temperatures in the winter, the soil respiration usually remains active under the snow cover (Brooks et al., 2011). Due to the long snow covering period at high latitudes, winter soil respiration can represent a substantial proportion of the annual ecosystem CO₂ budget (Alm et al., 1999; Aurela et al., 2002; Groffman et al., 2006; Zhao et al., 2016). However, since soil temperature and moisture exhibit very limited variations under the snow cover, their control over soil

respiration is usually not significant in the winter (Alm et al., 1999; Liptzin et al., 2009; Merbold et al., 2012). As a result, gap-filling methods driven by environmental variables might perform poorly when being applied to winter time fluxes. This has not been investigated before.

Alternatively, methods using time series analysis can predict the missing data based on the measured soil respiration time series itself, independent of auxiliary variables. For example, singular spectrum analysis (SSA) (Ghil et al., 2002) implements nonparametric spectral algorithms to analyze time series for forecasting and has been applied in diverse fields such as hydrometeorology, remote sensing and space sciences for data analysis and gap-filling purposes (Golyandina and Osipov, 2007; Kondrashov et al., 2010; Mahecha et al., 2011; Schoellhamer, 2001; Shun and Duffy, 1999; Wang and Liang, 2008). Compared to other time series analysis methods, SSA is fully data-adaptive and has no prior knowledge requirements on the periodicities of the data, making it rather flexible (Golyandina et al., 2018). While the performances of all gap-filling methods are to some degree influenced by the gap size (Kunwor et al., 2017; Moffat et al., 2007), the accuracy of the SSA method is expected to be especially sensitive to gap size increase, due to the self-referencing nature of this method. In contrast, methods that uses the expectation-maximization (EM) algorithm can take other soil respiration time series measured in parallel as reference and impute the missing values based on the relationships between the multiple time series (Junger and Ponce de Leon, 2015). This method can be effective when multiple chambers are deployed at the same site as replicates. Overall, these time-series-analysis-based methods have never been applied to C flux gap-filling but may have great potential for future applications, especially in cases when the environmental variables are not measured or missing.

All the gap-filling methods require a sampling window around the gap. Using the available data within the window, the corresponding models are trained to predict the data in the gap. The sizes of this sampling window are usually arbitrarily chosen, ranging from one half-month to a full year (e.g., Kunwor et al., 2017; Moffat et al., 2007; Zhao and Huang, 2015). Theoretically, a large sampling window could contain more information on the dynamics and patterns of the time series investigated but may produce an overgeneralized model that hardly represents the local gaps. At the same time, a small sampling window that captures the local conditions close to the gap may nonetheless cause large uncertainties in the model itself due to less training data. Thus, an optimal sampling window should have the smallest size possible that minimizes the prediction errors for the missing data. Whether the optimal window size varies among different methods, seasons and gap sizes needs to be further studied.

In this study, we implemented the four gap-filling methods described above, namely NLS, ANN, SSA, and EM, on artificially created data gaps in six soil respiration datasets from two boreal forest sites in Norway. The aims of the study are 1) to determine the best sampling window size for each gap-filling method, 2) to evaluate the performance of the gap-filling methods in gaps of different sizes (1-, 6-h, 1- and 15-day) and seasons (winter vs. non-winter), and 3) to test the suitabilities of the four methods in estimating annual soil respiration budgets under mixed-gap scenarios. Through these analyses, we accessed the strengths and weaknesses of these methods in soil respiration data gap-filling and enrich the toolbox for future C flux data gap-fittings.

2. Material and methods

2.1. Datasets

The soil respiration datasets used in this study were collected at two forest sites in Hurdal (60°22'21" N, 11°04'41" E, 284 m a.s.l.) and Løten (60°51'33" N, 11°26'08" E, 262 m a.s.l.), southern Norway. Both sites are dominated by Norway spruce (*Picea abies* (L) Karst.). At Hurdal, the

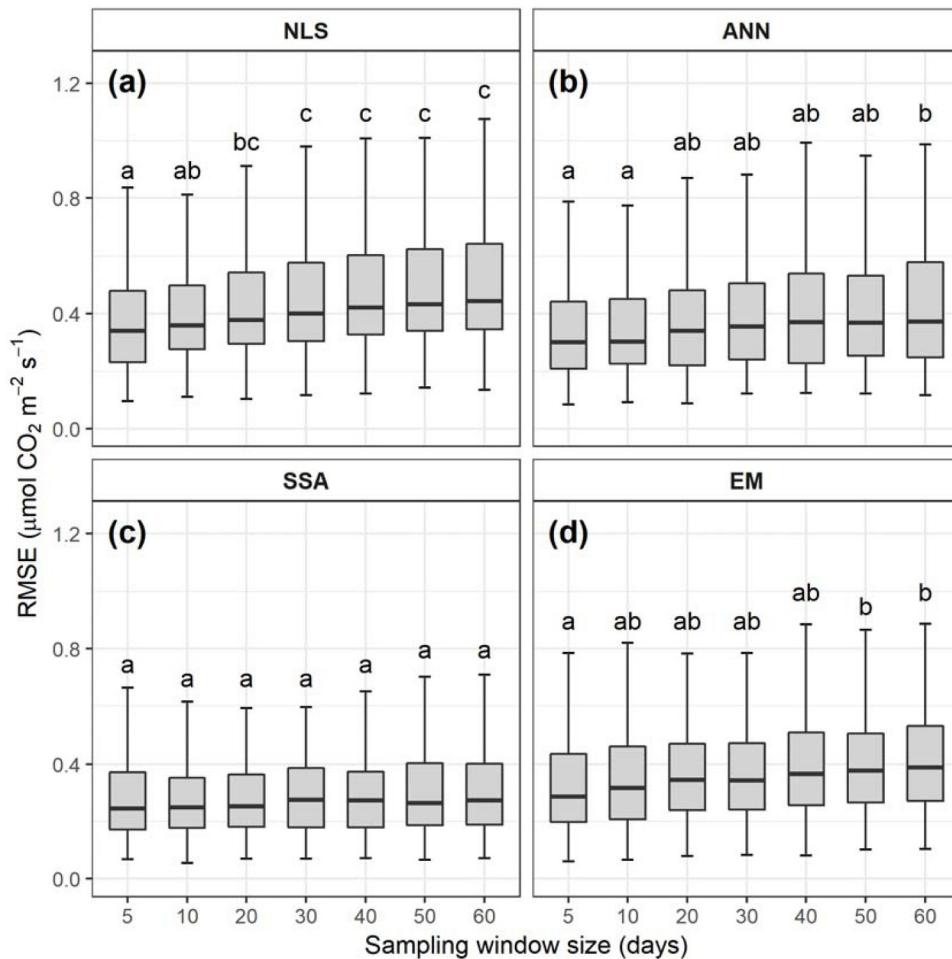


Fig. 1. The RMSE for the gap-filled values predicted by NLS (a), ANN (b), SSA (c) and EM (d) as a function of the sampling window size. The bottom and top of a box indicate the lower and upper quartiles, respectively, and the horizontal line within the box is the median. The lower and upper whiskers are the 10th and 90th percentiles, respectively. Different lowercase letters indicate significant differences in the RMSE among different sampling window sizes ($P < 0.05$).

terrain is slightly sloping, and the soils are podzolic in the upper part and rather hydrogenic (partially waterlogged) in the lower part. The site at Løten is rather flat and contains a bog surrounded by mixed forests. The soils are loamy with a high water-holding capacity. Both sites are typically covered with snow for 4–5 months per year, with a long-term (1957–2018) annual mean air temperature of 3.9 °C for Hurdal and 3.1 °C for Løten, respectively according to the seNorge2018 database (Lussana et al., 2019). The soil respiration measurements were carried out automatically every 10-min using forced diffusion (FD) chambers (eosFDCO₂, Eosense Inc., Dartmouth, NS, Canada). Specifically, each chamber is equipped with membranes allowing gases to freely diffuse into it. With different compartments, it separately measures the CO₂ concentrations that represent the ambient air (C_{ambient}) and the air that mixes the ambient air with gases from the soil (C_{mix}). The soil respiration rate (R_s) is then calculated as (Risk et al., 2011):

$$R_s = G \cdot (C_{\text{mix}} - C_{\text{ambient}}) \quad (1)$$

where G is an empirical positive constant (a function of the gas diffusion rate through the membrane and measured soil surface area; unit: m s^{-1}) provided by the manufacturer. More details of the method are outlined in Risk et al. (2011). The C_{mix} is larger than the C_{ambient} under typical circumstances, implying a loss of CO₂ from the soil (i.e., a positive flux). Since the FD chamber measures the fluxes with the same CO₂ source from the soil (i.e., soil respiration) as static chambers and produces comparable soil respiration data (Risk et al., 2011), we suggest that the methods and results in this study can also be applied to the

soil respiration measured by the widely used static chambers.

Along with the soil respiration, corresponding soil temperature and moisture (TMS-4, TOMST inc., Prague, Czech Republic) at 10, 20 and 50 cm depths were measured within <1 m from each chamber at the same temporal resolution (i.e., 10-min). In addition, the ambient temperatures were measured at ~5 cm above the soil surface, which represented either the temperature of the snowpack during winter time or air temperature when snow was absent. Data from a total of 6 plots (4 from Hurdal and 2 from Løten) were used for analysis. The datasets overall covered the period between July 2017 and April 2019 but varied among the plots (see details in Table 1). The soil respiration data were quality-checked before being analyzed and were rejected (marked as missing values): 1) during instrument malfunction or power failure, and 2) with implausible values (soil respiration rate < -2 or >10 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$). In total, 11% of the data were discarded (see Table 1 for the summary for each plot).

2.2. Artificial gaps

In order to compare the gap-filled data to the original observations, we created artificial gaps (flagged) in the soil respiration time series in all the 6 datasets. Specifically, we used each dataset to create four separate benchmark datasets: one with 80 one-hour gaps (6 observations per gap), a second one with 40 six-hour gaps (36 observations per gap), a third one with 30 one-day gaps (144 observations per gap), and the last one with 10 fifteen-day gaps (2160 observations per gap). The gaps

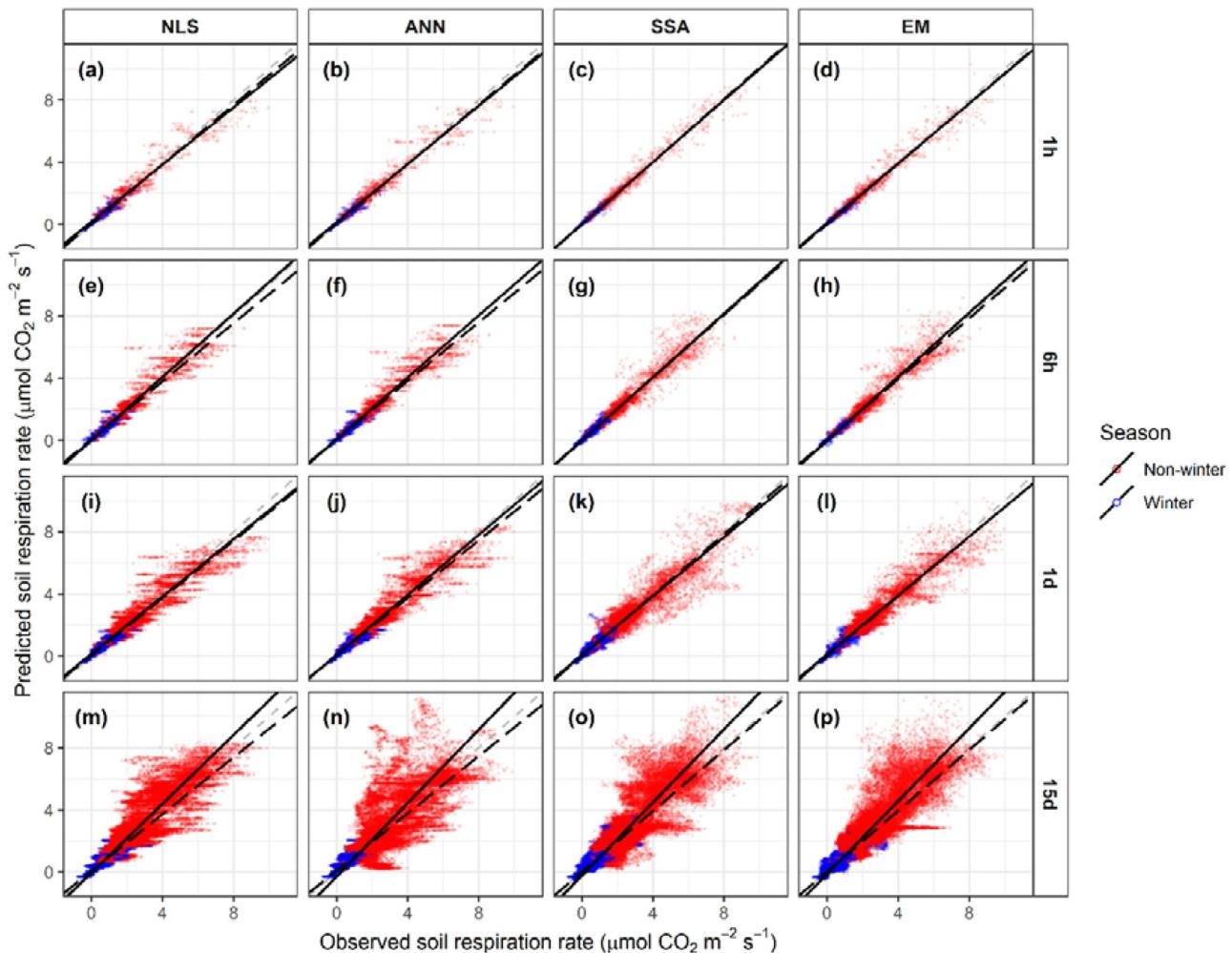


Fig. 2. Predicted values for the gaps of 1-h. (1 h, a–d), 6-h. (6 h, e–h), 1-day (1d, i–l) and 15-day (15d, m–p) by NLS (a, e, i, m), ANN (b, f, j, n), SSA (c, g, k, o) and EM (d, h, l, p) as a function of the corresponding observations. Non-winter and winter seasons are plotted in red and blue, respectively. Total least squares regression was used to accommodate the uncertainty of both dependent and independent variables. The black solid and dashed lines indicate the regressions for non-winter and winter seasons, respectively. The gray dashed lines indicate the 1:1 ratio. The coefficients of the regressions are presented in Table 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3

Coefficients of the linear regressions between the observations and the gap-filled data that were predicted by the four methods for different gap sizes and seasons.

Gap size	Season	NLS		ANN		SSA		EM	
		Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
1-h	Non-winter	0.170 (0.017)	0.919 (0.009)	0.139 (0.016)	0.936 (0.009)	0.014 (0.013)	0.998 (0.008)	0.074 (0.016)	0.960 (0.009)
	Winter	0.043 (0.007)	0.956 (0.010)	0.038 (0.007)	0.958 (0.010)	-0.003 (0.004)	1.007 (0.005)	0.014 (0.005)	0.971 (0.007)
6-h	Non-winter	0.012 (0.010)	1.016 (0.006)	0.059 (0.010)	0.996 (0.006)	-0.023 (0.010)	1.021 (0.006)	-0.033 (0.011)	1.028 (0.009)
	Winter	0.045 (0.004)	0.935 (0.006)	0.046 (0.004)	0.942 (0.006)	-0.009 (0.003)	1.010 (0.004)	0.012 (0.009)	0.978 (0.006)
1-day	Non-winter	0.149 (0.007)	0.923 (0.004)	0.075 (0.006)	0.966 (0.004)	0.062 (0.008)	0.949 (0.005)	0.097 (0.003)	0.948 (0.004)
	Winter	0.051 (0.002)	0.920 (0.005)	0.045 (0.002)	0.925 (0.005)	0.006 (0.002)	0.977 (0.004)	0.025 (0.007)	0.955 (0.004)
15-day	Non-winter	-0.055 (0.004)	1.108 (0.003)	-0.297 (0.009)	1.182 (0.005)	-0.236 (0.006)	1.177 (0.004)	-0.130 (0.006)	1.170 (0.004)
	Winter	0.066 (0.002)	0.914 (0.003)	0.097 (0.002)	0.920 (0.003)	0.010 (0.002)	0.979 (0.003)	0.029 (0.002)	0.977 (0.003)

Standard errors of the coefficients are shown in the parentheses.

were added at random locations of the time series while avoiding overlapping with each other as well as with the original gaps in the dataset. These benchmark datasets were used to test the performance of the gap-filling methods in gaps of different sizes.

In addition, to evaluate the overall influences of different gap-filling methods on the estimations of annual soil respiration budget, we selected the datasets that were longer than 1 year (i.e., H1, H3 and H4 in Hurdal and Løten, Table 1) and extracted the data of a one-year period each from April 30, 2018 to April 29, 2019. Based on each of the

extracted datasets, we created a random number of gaps with random lengths (uniformly distributed) up to 15 days (mixed gaps) that resulted in benchmark datasets with 5 levels of the gap proportion (10, 20, 30, 40 and 50%). For each of the 5 levels, 10 replicate datasets were created with different random gap arrangements (i.e., 200 benchmark datasets in total for the 4 plots, see Fig. A1 in the Appendices for an example of gap summary in the dataset). These mixed gap benchmark datasets were thereafter used for the estimation of the annual soil respiration sum after the artificial gaps have been filled.

Table 4.

Summary of the analysis of variance (ANOVA) for the mixed effects models with RMSE, relative error, AR (1) and variance as dependent variables and method, season, gap size and their two-way interactions as independent variables.

Dependent variable	Independent variable	SS	F-value	P-value
RMSE	Method	0.27	1.72	0.161
	Season	41.81	791.66	<0.001
	Gap size	15.43	97.411	<0.001
	Method*Season	0.30	1.90	0.128
Relative error	Method*Gap size	1.57	3.29	<0.001
	Method	0.40	1.95	0.119
	Season	14.98	217.81	<0.001
	Gap size	6.35	30.75	<0.001
Variance	Method*Season	0.21	1.00	0.394
	Method*Gap size	1.14	1.83	0.057
	Method	3.37	25.99	<0.001
	Season	3.89	119.77	<0.001
AR (1)	Gap size	11.12	114.28	<0.001
	Method*Season	1.61	12.37	<0.001
	Method*Gap size	4.28	10.98	<0.001
	Method	25.62	47.32	<0.001
	Season	39.06	288.60	<0.001
	Gap size	271.34	668.25	<0.001
	Method*Season	18.56	34.28	<0.001
	Method*Gap size	18.33	11.28	<0.001

The plot number is included in the models as random effects. SS: type III sums of squares with Satterthwaite approximation for degrees of freedom. RMSE: root mean squared error. AR (1): first order autocorrelation coefficient. The sampling window length of 5 days was used.

2.3. Sampling window length

Each of the gap-filling methods has its own algorithm to build models trained by available data around each gap, and the models were then used to predict the values in the gap. Since the data sampled around the gap for training the models could have a significant influence on the accuracy of the gap-filled values, choosing a sampling window length that includes proper training data is important. In this study, we compared the performance of the gap-filling methods using sampling windows of 5, 10, 20, 30, 40, 50 and 60 days (i.e., 720, 1440, 2880, 4320, 5760, 7200, 8640 observations, respectively), arranged symmetrically around the gap. A sampling window of 10 days, for example, included data of 5 days before and after the gap, regardless of the gap length. The shortest window length that achieved best accuracy was chosen for each method. This sampling window is more flexible than the traditional fixed window (Moffat et al., 2007), which also includes the gap itself and may suffer from insufficient training data if the gap size approaches the sampling window size.

2.4. Gap-filling methods

In this study, we used four methods to fill the artificially created gaps.

Firstly, the artificial gaps were filled with the predictions from the nonlinear least squares (NLS) regression based on the relationship between soil temperature and soil respiration. Since NLS is the most commonly used method for gap-filling (Gomez-Casanovas et al., 2013; Kunwor et al., 2017; Moffat et al., 2007), we took it as a reference to compare other methods with. The Lloyd-Taylor function was used for making the models (Lloyd and Taylor, 1994):

$$R_s = R_{10} \cdot e^{E_0 \left(\frac{1}{T_{ref} - T_0} - \frac{1}{T_s - T_0} \right)} \quad (2)$$

where R_{10} and E_0 are parameters to be estimated, T_{ref} and T_0 take the values of 283.15 and 227.13 K, respectively, following Lloyd and Taylor (1994), and T_s is the soil temperature measured at 10 cm depth (K). The NLS regression was performed by applying the Levenberg-

Marquardt algorithm using the “minpack.lm” package (Elzhov et al., 2016) in the program R 3.5.2 (R Development Core Team, 2018). In cases when the variance of soil temperature in the winter was too low for the NLS function to fit with, the mean value of the soil respiration rate within the sampling window was taken to fill the gap (71% of the gaps in the winter).

Secondly, to include environmental variables other than the soil temperature, artificial neural network (ANN) models similar to Moffat et al. (2010) were implemented to fill the gaps. The inputs of the ANN models were soil temperature, soil moisture and ambient temperature (i.e., either snow temperature in the winter or air temperature in non-winter periods), and the output was soil respiration rate. The input variables were normalized to be centered around 0 with a standard deviation of 1 before used to train the models. We included one hidden layer with two neurons in the models and a back-propagation algorithm was performed using the R package “neuralnet” (Fritsch et al., 2019).

Thirdly, being independent of the variations of other environmental variables, the singular spectrum analysis (SSA) method analyzes the structure of the soil respiration time series and predicts the missing values in the gap. In this study, we used the iterative approach suggested by Kondrashov and Ghil (2006) to determine the major periodic component of the time series in the sampling windows and projected this periodic component into the gaps. In addition to the major periodic (i.e., low frequency) component, we also extracted the high frequency component (i.e., the eigentriples with periodicity band < 10, unit: 10-min) of the time series in the sampling window and added its average diurnal variations to the projected low frequency values in the gap at the corresponding time of the day. This approach takes both low and high frequency components derived from the time series into account for the gap-filled data. The SSA was conducted using the R packages “spectral.methods” (Buttlar, 2015) and “Rssa” (Golyandina et al., 2015).

Lastly, we implemented the method developed by Junger and Ponce de Leon (2015) based on the expectation-maximization (EM) algorithm (Dempster et al., 1977). This method takes multiple other parallelly measured soil respiration time series as references for the target gap-filling series and makes predictions accounting for both the correlation among the series and temporal structure of individual series. In this study, all the parallel soil respiration time series (no artificial gaps) that reached a certain degree of correlation with the target series (i.e., $r > 0.75$; see Table A1 for the correlations among the plots) from the same site (Hurdal or Løten) were used as references. We used the “spline” method to fit the time series at each iteration with a degree of freedom of 10 (chosen by cross validation). Due to the unequal data temporal range among the plots, the gaps located out of the temporal range of the references were not filled and excluded in further analyses (< 5% in total). Similarly, EM was not applied to gap-fill the mixed gap benchmark datasets from the plot L2 in Løten for determining the annual budget, due to the lack of a full year data at plot L1 as the reference. The EM gap-filling was performed using the R package “mtsdi” (Junger and Ponce de Leon, 2018).

2.5. Statistical analysis

The performance of the gap-filling methods was investigated separately in the winter and other periods of the year. Since the snow was not monitored at the sites, we defined the winter as the period with soil temperature < 1.5 °C and the remaining period as non-winter. This criterion well separated each year into two seasons and characterized the winter by low soil temperature with a narrow variation range (mostly 0–1 °C).

To select the optimal sampling window length for each gap-filling method, we calculated the root-mean-squared error (RMSE) between gap-filled values and the original observations for each gap. RMSE has been widely used as an indication of the model accuracy and it was

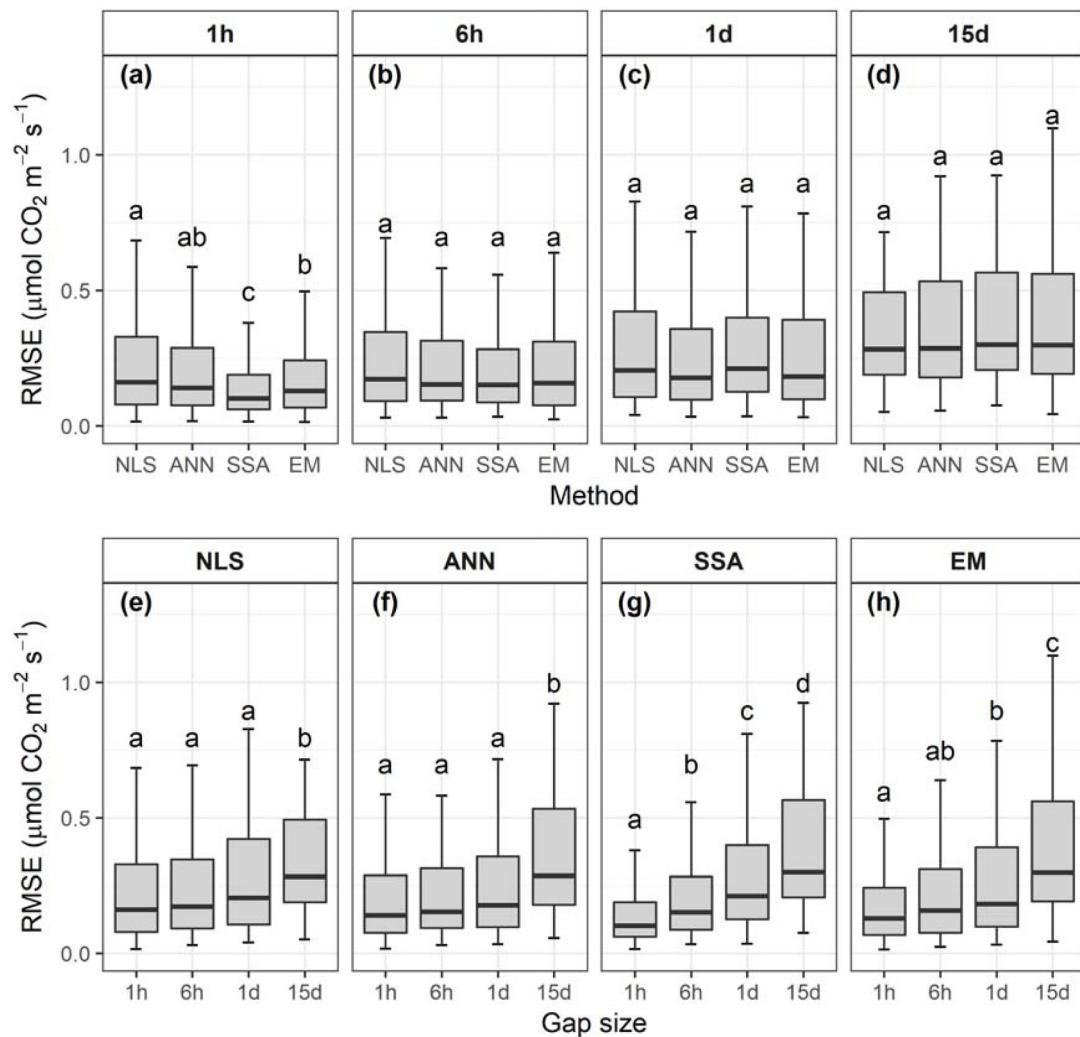


Fig. 3. The RMSE for the gap-filled values of different gap sizes (1 h: 1-h, 6h: 6-h, 1d: 1-day, 15d: 15-days) predicted by different methods. The boxplots are plotted as in Fig. 1. Different lowercase letters indicate significant differences in RMSE among the different methods in (a-d) and among different gap sizes in (e-h) ($P < 0.05$).

calculated as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (3)$$

where n is the number of data points in the gap, and P_i and O_i are the predicted (gap-filled) and observed soil respiration rates ($\mu\text{mol CO}_2 \text{m}^{-2} \text{s}^{-1}$), respectively. A linear mixed-effects model was used to investigate the effects of sampling window size, gap-filling method, season, gap size and the interactions between window size and each of the other three factors on RMSE. Plot ID was included as random effect in the model. Normality and homoscedasticity of the model were evaluated visually by plotting residuals. The post-hoc Tukey HSD test was used to further investigate the differences among the groups of the significant effects. The smallest window size associated with the lowest RMSE was chosen for each method for further analyses.

With the gaps filled using the optimal sampling windows, we further calculated relative error, variance and first order autocorrelation coefficient (AR (1)), in addition to RMSE (Eq. (3)), for each gap to evaluate the performance of the gap-filling methods. They are given as follows:

$$\text{Relative error} = \frac{\sum_{i=1}^n |P_i - O_i|}{\sum_{i=1}^n |O_i|} \quad (4)$$

$$\text{Variance} = \frac{1}{n} \sum_{i=1}^n (PO_i - \overline{PO})^2 \quad (5)$$

$$AR(1) = r(PO_t, PO_{t+1}) \quad (6)$$

where PO_i is either the predicted (gap-filled) or observed soil respiration rate ($\mu\text{mol CO}_2 \text{m}^{-2} \text{s}^{-1}$) and \overline{PO} is the mean of the rates in each gap. AR (1) is the Pearson Correlation Coefficient (r) between the time series in the gap (OP_t) and the series itself with a lag of one-time step (i.e., 10-min) (PO_{t+1}). Variance and AR (1) were calculated separately for the gap-filled and observed values. The gaps that were filled with the mean value of the sampling window in the winter were excluded in the AR (1) calculations for NLS. Linear mixed-effects models were established to determine the effects of the method, season, gap size and their two-way interactions on RMSE, relative error, variance and AR (1) with the random effect of plot ID.

The relationships between the observations and the gap-filled values were fitted by linear regressions using the total-least-squares approach to account for variances of both the dependent and independent variables. We assumed that the residue variance of the gap-filled values was the same as the observations (i.e., the variance ratio was set to 1). The regression slopes were used to indicate systematic bias of each method in different gap sizes and seasons. The standard errors of the regression coefficients were estimated by bootstraps (1000 random replicates). In addition, we also calculated the Nash-Sutcliffe Efficiency (NSE)

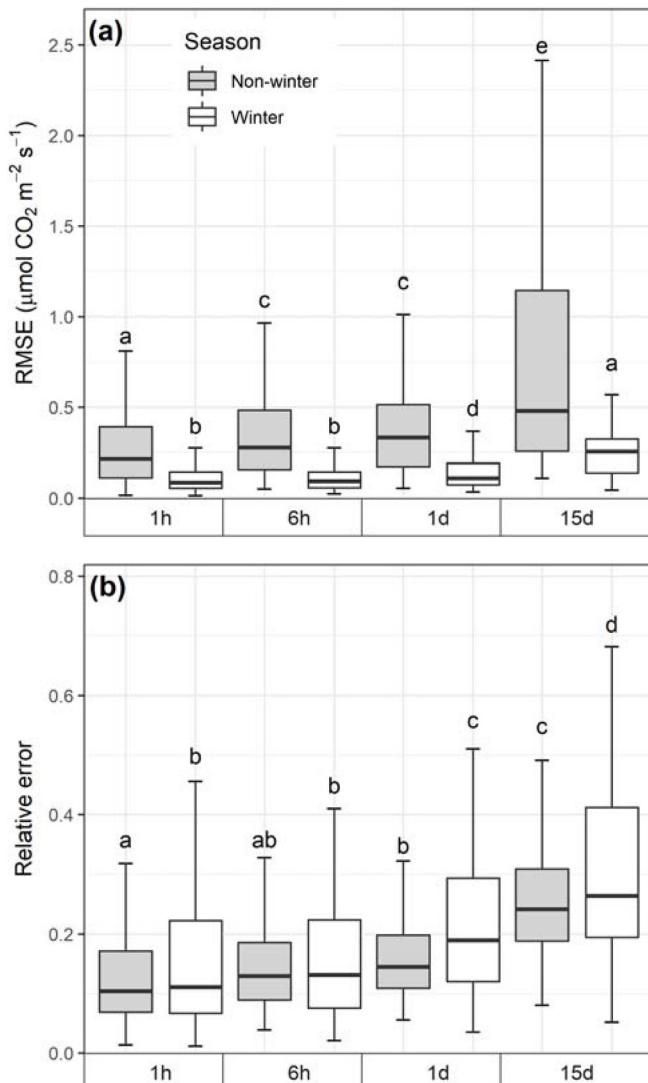


Fig. 4. The RMSE (a) and relative error (b) for gaps of different sizes (1 h: 1-h, 6 h: 6-h, 1d: 1-day, 15d: 15-days) in different seasons. The boxplots are plotted as in Fig. 1. Different lowercase letters indicate significant differences in RMSE or relative error among the different gap sizes and seasons ($P < 0.05$).

(Nash and Sutcliffe, 1970) and Kling-Gupta Efficiency (KGE) (Gupta et al., 2009; Kling et al., 2012) in the gaps to evaluate the overall performance of the four methods in different gap sizes and seasons, as follows:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (7)$$

$$\text{KGE} = 1 - \sqrt{(r - 1)^2 + (cv - 1)^2 + (\beta - 1)^2} \quad (8)$$

where r is the Pearson correlation coefficient, cv is the ratio of the coefficient of variation between the predicted and the observed values, and β is the ratio of the means between the predicted and the observed values. The calculations were conducted using the R package “hydroGOF” (Zambrano-Bigiarini, 2017).

To investigate the influences of gap-filling methods on the annual flux budget, the four methods were also applied to the one-year benchmark datasets with mixed gaps. We then compared the annual sum of soil respiration including gap-filled values to those summed from original observations using the Wilcoxon tests. The original gaps in the datasets were not taken into account for the budget estimations.

All the data processing and analyses were carried out using the

program R 3.5.2 (R Development Core Team, 2018). In addition, the packages “lme4” (Bates et al., 2015) and “lmerTest” (Kuznetsova et al., 2016) were used to estimate linear mixed-effects models, the “emmeans” package was used for the post-hoc Tukey HSD test (Lenth, 2019), the “MethComp” package was used to perform regressions with the total-least-squares approach (Carstensen, 2015), and “ggplot2” package was used for plotting graphs (Wickham, 2016). The R code to perform the gap-filling with the four methods in this study is incorporated into the R package “FluxGapsR” freely available at <https://github.com/junbinzhao/FluxGapsR/>.

3. Results

3.1. Effect of sampling window size

According to the mixed-effects model (Table 2), the sampling window size showed a significant effect on the RMSE of the gap-filling results ($P < 0.01$). This window size effect differed among the methods ($P = 0.01$), but not among the gap sizes ($P = 0.57$) or the seasons ($P = 0.14$). Of the four methods, NLS, ANN and EM showed slight increases in RMSE along the window size gradient from 5 to 60 days with significantly higher values of RMSE being present in the window of ≥ 30 , 50 and 60 days, respectively, compared to the 5-day window ($P < 0.05$) (Fig. 1a, b and d). In contrast, SSA showed no significant changes in RMSE with different sampling window sizes ($P > 0.05$) (Fig. 1c). Therefore, the 5-day window, which is the smallest window size associated with the lowest RMSE for all the four gap-filling methods, was chosen in this study for further analysis.

3.2. Performance of the gap-filling methods

The scatterplots of the gap-filled (predicted) soil respiration values versus the original observations generally followed the 1:1 line in the gaps ≤ 1 day (Fig. 2). While the corresponding regression slopes for SSA and EM were between 0.95 and 1.03, the slopes for NLS and ANN occasionally dropped down to 0.92 (Table 3), indicating a slight underestimation. For the 15-day gaps, all the methods generally overestimated the values in the non-winter gaps (i.e., regression slope > 1.1 , Fig. 2m-p and Table 3). For the winter gaps, SSA and EM exhibited slopes of 0.98 but NLS and ANN slightly underestimated the flux values in the gaps (i.e., slopes were 0.91 and 0.92, respectively).

According to the mixed-effects model (Table 4), the RMSE differed among different seasons ($P < 0.01$) and gap sizes ($P < 0.01$) but not among different methods in general ($P = 0.16$). However, differences of the RMSE were present among the methods when considering the interaction with the gap size ($P < 0.01$) but not when considering the season ($P = 0.13$). The difference was present in the 1-h gap where the results from SSA showed significantly lower RMSE than those from the other 3 methods ($P < 0.01$) (Fig. 3a). While the RMSE of EM was also significantly lower than that of NLS ($P < 0.05$), ANN resulted in similar RMSE as NLS ($P > 0.05$). For the gap sizes of 6-h, 1- and 15-day, the RMSE from ANN, SSA and EM were not significantly different from NLS ($p > 0.05$) (Fig. 3b-d).

All the four methods generally showed significantly greater RMSE as gap size increased from 1 h to 15 days (Fig. 3e-h). For NLS and ANN, the RMSE remained similar in gaps ≤ 1 day ($P > 0.05$) while the RMSE became significantly higher in the 15-day gaps ($P < 0.01$) (Fig. 3e and f). The RMSE of SSA showed much higher sensitivity to gap size changes with a significant RMSE increase with every step of gap enlarging from 1-h to 15-day ($P < 0.01$) (Fig. 3g). The RMSE of EM showed a similar pattern with significantly lower values exhibited in 1-h gaps than in 15-day gaps ($P < 0.01$) while 15-day gaps had the greatest RMSE ($P < 0.01$) (Fig. 3h).

The relative error was significantly different among seasons ($P < 0.01$) and gap sizes ($P < 0.01$), but not among the methods ($P = 0.12$) (Table 4). The interactions of method with season

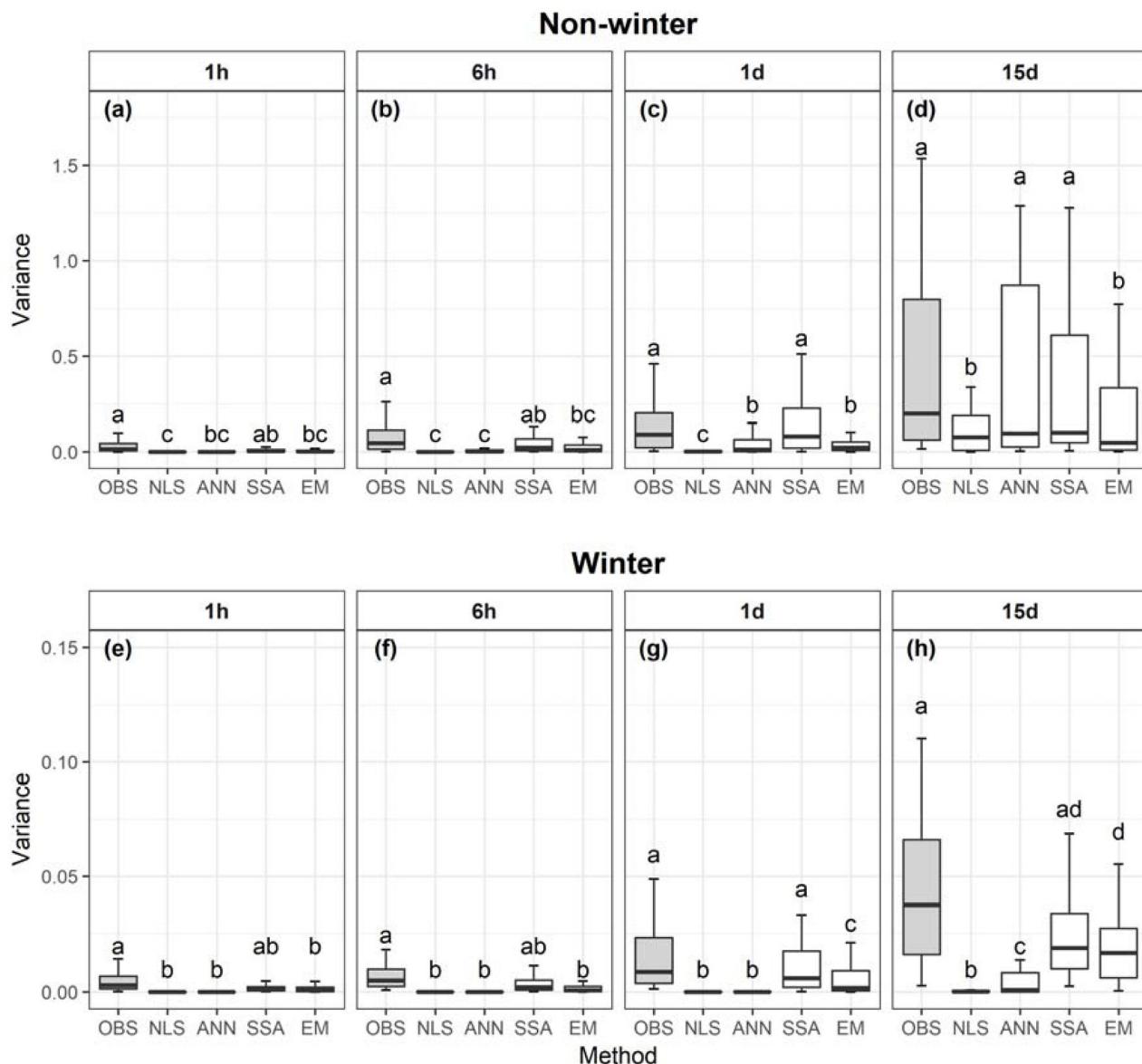


Fig. 5. The variance of the gap-filled values predicted by different methods in gaps of different sizes (1h: 1-hour, 6h: 6-h, 1d: 1-day, 15d: 15-days) as compared to variance of the observations. The variances are shown separately for the winter (a-d) and the non-winter (e-h) with different scales. The boxplots are plotted as in Fig. 1. Different lowercase letters indicate significant differences in variances among different methods ($P < 0.05$).

($P = 0.39$) and gap size ($P = 0.06$) were not significant either. Furthermore, even though the RMSE was generally lower in the winter gaps compared to those in the non-winter ($P < 0.01$) (Fig. 4a), the relative errors were significantly higher in the winter than in the non-winter for the 1-h, 1- and 15-day gaps ($P < 0.05$) (Fig. 4b). The greatest relative error was present in the winter 15-day gaps.

The mixed-effects model for the variance showed significant effects of method, season, gap size as well as the two-way interactions of method with season and gap size ($P < 0.01$) (Table 4). Among the four methods, only SSA resulted in similar variance in the gap-filled values compared to the original observations across the seasons and the gap sizes ($P > 0.05$) (Fig. 5). The variances of the values filled by other 3 methods were overall much lower than those of the observations ($P < 0.05$), particularly for the winter gaps (Fig. 5e–h). Exceptions are the non-winter 15-day gaps filled by ANN, which also showed similar level of variance as the observations ($P > 0.05$) (Fig. 5d).

The AR (1) were affected by method, season, gap size as well as the two-way interactions of method with season and gap size ($P < 0.01$) (Table 4). Of the four methods, EM outputted data series with similar AR

(1) to the original observations ($P > 0.05$), except for the 6-h and 1-day gaps in the winter (Fig. 6). NLS also resulted in similar AR (1) structure as the observations in the winter gaps of ≤ 1 day ($P > 0.05$) when excluding those filled using the mean values in the sampling windows (Fig. 6e–g). AR (1) of the data series generated by SSA were similar to the observations only in 1-h gaps (Fig. 6a and e) and the winter 6-h gaps (Fig. 6f). ANN produced data series with significantly higher AR (1) than the observations in all cases ($P < 0.01$) (Fig. 6).

The NSE were > 0.85 in the gaps of ≤ 1 day for all the methods, among which the NSE values of SSA and EM were > 0.95 in the 1-h gaps (Table 5). The NSE values dropped below 0.8 in the 15-day gaps for all the methods with the values being < 0.7 for ANN in the winter and EM in the non-winter periods and being even < 0.6 for ANN in the non-winter periods. The more comprehensive index KGE, which considers accuracy, bias and variance of the gap-filled data, showed a similar pattern as the NSE (Table 5). Specifically, KGE values were > 0.9 for gaps ≤ 1 day for all the methods and the values were even ≥ 0.95 for SSA and EM in the 1- and 6-h gaps. In the 15-day gaps, lower KGE values were present, which ranged between 0.8 and 0.9.

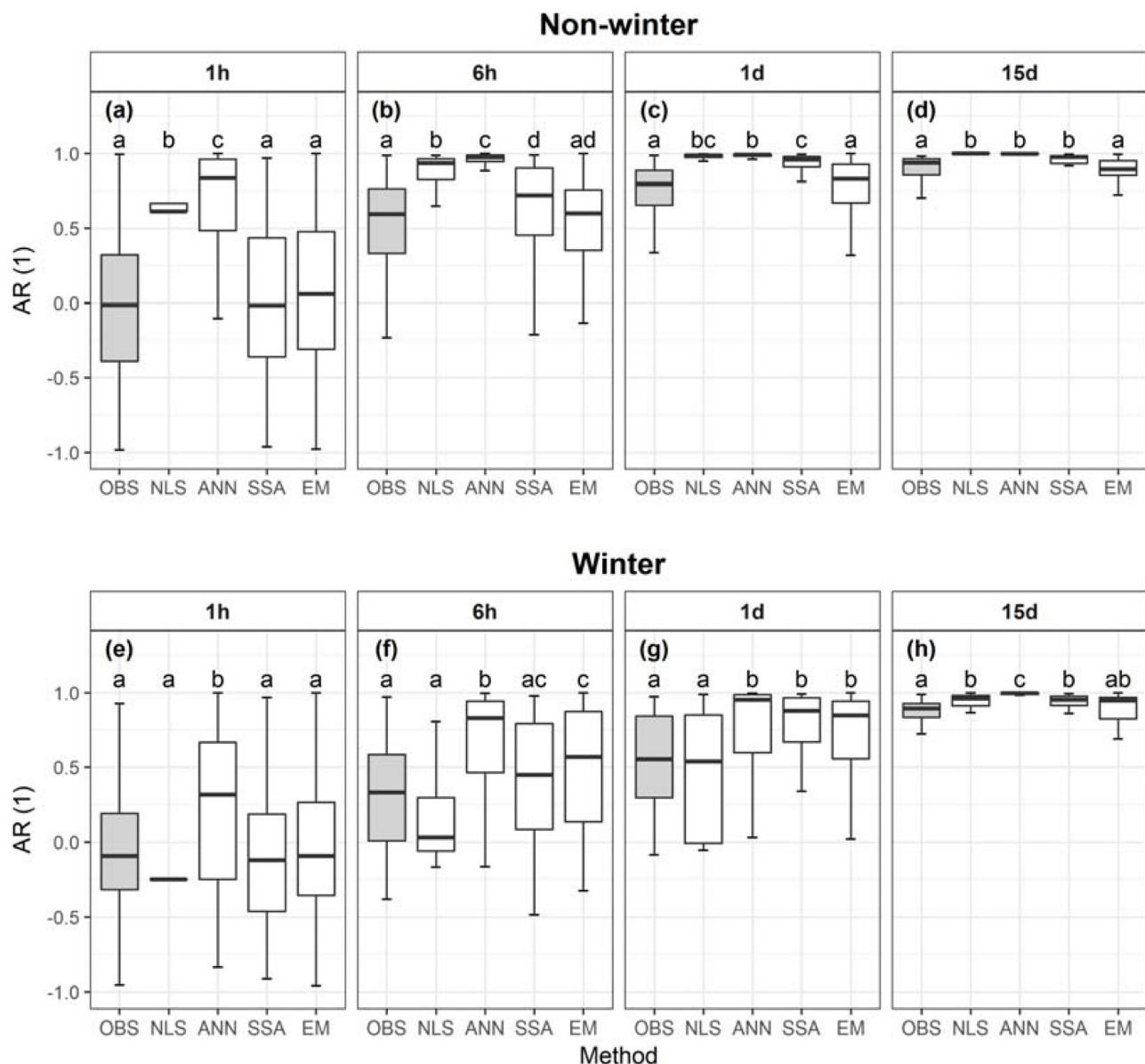


Fig. 6. The AR (1) coefficient of the gap-filled values predicted by different methods in gaps of different sizes (1 h: 1-h, 6 h: 6-h, 1d: 1-day, 15d: 15-days) as compared to variance of the observations. The AR (1) coefficients are shown separately for the winter (a–d) and the non-winter (e–h) with different scales. The boxplots are plotted as in Fig. 1. Different lowercase letters indicate significant differences in AR (1) coefficient among different methods ($P < 0.05$). Note that the gaps failed to be filled with NLS (but filled with the average of the sampling window) in the winter are not included in the AR (1) calculations (e–h).

Table 5

Nash-Sutcliffe efficiency (NSE) and Kling-Gupta Efficiency (KGE) for the gap-filled values predicted using different methods.

Gap size	Season	NSE				KGE			
		NLS	ANN	SSA	EM	NLS	ANN	SSA	EM
1-h	Winter	0.89	0.90	0.97	0.96	0.92	0.93	0.98	0.97
	Non-winter	0.93	0.94	0.97	0.96	0.91	0.93	0.99	0.95
6-h	Winter	0.89	0.90	0.94	0.96	0.92	0.92	0.97	0.97
	Non-winter	0.91	0.92	0.94	0.93	0.95	0.95	0.97	0.96
1-day	Winter	0.89	0.89	0.88	0.91	0.91	0.92	0.94	0.94
	Non-winter	0.90	0.92	0.89	0.91	0.92	0.95	0.93	0.94
15-day	Winter	0.76	0.69	0.76	0.79	0.85	0.83	0.88	0.89
	Non-winter	0.78	0.58	0.72	0.69	0.88	0.85	0.80	0.84

3.3. Annual budget estimation

The annual soil respiration budget was estimated to be 1152, 463, 627 and 865 g C m⁻² year⁻¹, respectively, at the plot H1, H3, H4 in Hurdal and L2 in Løten during April 30, 2018–April 29, 2019 based on the observations (Fig. 7). By filling the mixed artificial gaps in the datasets, the corresponding annual soil respiration budget estimations ranged 1108–1136, 449–475, 598–646 and 767–955 g C m⁻² year⁻¹, respectively, for the four plots. Overall, the errors of budget estimations tended to be larger as the proportion of gaps increased. For different gap-filling methods, the overall budget errors were –3.7 to 3.6%, –11.3 to 16.0%, –3.7 to 5.8% and –3.2 to 5.3%, respectively, for NLS, ANN, SSA and EM. Significant biases only occurred in the H3 and L2 datasets with 50% gaps (Fig. 7b and d). Specifically, NLS and ANN overestimated the budget in the L2 dataset ($P < 0.05$), while NLS and SSA underestimated the budget in the H3 dataset ($P < 0.05$).

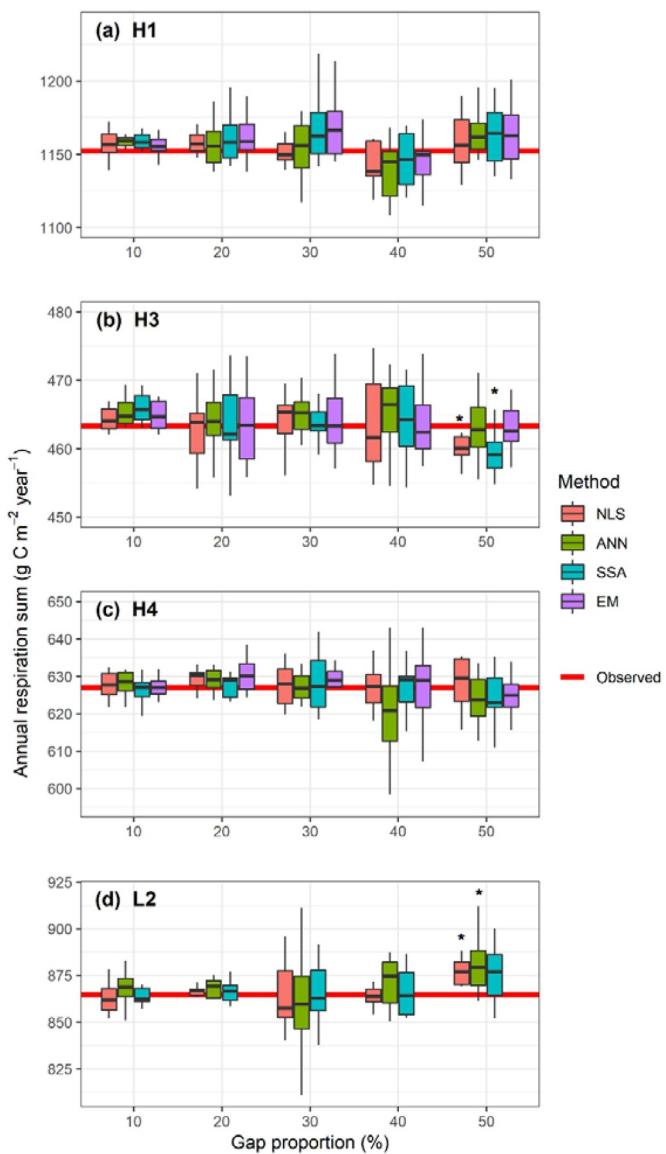


Fig. 7. The annual soil respiration sum estimated based on the observations (red lines) and included the gap-filled values predicted by different methods (boxplots) during April 30, 2018–April 29, 2019 for the mixed gap benchmark datasets from plot H1 (a), H3 (b), H4 (c) in Hurdal and L2 (d) in Løten. The boxplots are plotted as in Fig. 1. The asterisks (*) indicate significant differences of the estimated annual sum from the observed sum ($P < 0.05$). The datasets from L2 were not gap-filled by EM due to the lack of reference for the full year. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Our study evaluated the performances of four gap-filling methods on soil respiration datasets from two boreal forest sites. For carbon flux studies, gap-filled data are mostly used for estimating the sum of the flux budget, usually on the scale of one year. According to our estimations, NLS, SSA and EM resulted in annual budget errors that ranged from -3.7 to 5.8% , which are fully acceptable given that this result represents a wide range of soil respiration budget (i.e., 463 – 1152 g C m^{-2} year $^{-1}$) in high latitude ecosystems (Bahn et al., 2010; Raich and Schlesinger, 1992). Among the methods used, SSA and EM are based on the time series analysis and had never been used for gap-filling in carbon flux data. Particularly, they are implemented independent of other environmental measurements (e.g., soil temperature). With this

advantage and their accurate annual budget estimations, we suggest SSA and EM to be considered as alternatives to the more conventional methods (e.g., NLS) for gap-filling in soil respiration data and possibly in other carbon flux data (e.g., net ecosystem CO₂ exchange measured by eddy covariance, methane fluxes, etc.), especially when the controlling environmental variables are unavailable, missing or unknown.

4.1. The accuracy of the gap-filling methods

In our study, we selected the sampling window size of 5 days, because this window size already achieved the most accurate results for all the methods. Previous studies commonly use sampling windows ranging from half a month up to a full year depending on the method used (e.g., Kunwor et al., 2017; Moffat et al., 2007; Zhao and Huang, 2015), which might not be necessary according to our result. Our result implies that a smaller sampling window that captures the local data structures or relationships around the gap can be more representative for the gap than one that includes data of longer periods. Using a smaller sampling window also significantly reduces the computing time for the methods involving iterations (e.g., ANN). However, we do not recommend windows that are smaller than 5 days, because they may simply underrepresent the diurnal variation of the variables in the gap if e.g., only 1 day before and 1 day after the gaps are considered. In addition, it is worth to note that our datasets have a temporal resolution of 10 min, which is three times as many as the commonly used 30-min interval flux datasets in a 5-day window. Thus, we further aggregated the 10-min-interval data by averaging data within each half hour and generated datasets with 30-min intervals. The gap-filling results on the aggregated datasets also indicated that the 5-day sampling window had the lowest RMSE among the window sizes for all the methods (see Fig. A2 in the Appendices). Therefore, we suggest a 5-day sampling window could be suitable for most gap-filling procedures on carbon flux data.

Among the four methods, NLS is the most widely accepted method for carbon flux gap-filling since it is usually supported by well-acknowledged relationships between the flux and environmental factors, such as the one between soil respiration and temperature used in this study (Gomez-Casanovas et al., 2013; Lloyd and Taylor, 1994). This method provides the basis to evaluate the other methods. Compared to NLS, ANN models were trained against more environmental variables (i.e., soil moisture and ambient temperature); however, it showed no improvement over NLS in terms of the RMSE (Fig. 3a–d). A previous study that included soil moisture in soil respiration gap-fillings also reached the same conclusion (Gomez-Casanovas et al., 2013). This is mainly because the controls of soil moisture over soil respiration are complicated and depend on many other factors (e.g., time, microbial communities, etc.) (Cook and Orchard, 2008). Particularly in this study, we found large prediction errors by ANN associated with the pulse in soil moisture induced by precipitation events that occurred in the gaps or the sampling windows (see Fig. 8 for examples). Depending on when those pulses occurred as well as their magnitude, gap-filled values predicted by ANN could greatly over- or under-estimate the fluxes in the gaps due to a mismatch of the soil moisture magnitudes in the gaps to those in the sampling windows. These large errors were also reflected in the lower NSE values in the 15-day gaps (Table 5) and the large uncertainties in soil respiration budget estimations by ANN (Fig. 7). Yet, models that better represent soil moisture in soil respiration gap-filling are still needed.

The SSA yielded the smallest errors among the methods in the small gaps (i.e., 1-hour) and outcompeted NLS (Fig. 3a), suggesting that SSA is a highly robust method for filling small gaps. In larger gaps, SSA could also capture the diurnal pattern and well construct the data series in the gap close to the observations if the variation pattern persisted throughout the gap (see Fig. 9a for an example). However, drifts in the diurnal patterns and magnitudes of soil respiration, which are driven by environmental factors, often occur in larger gaps (e.g., Fig. 9b). Without

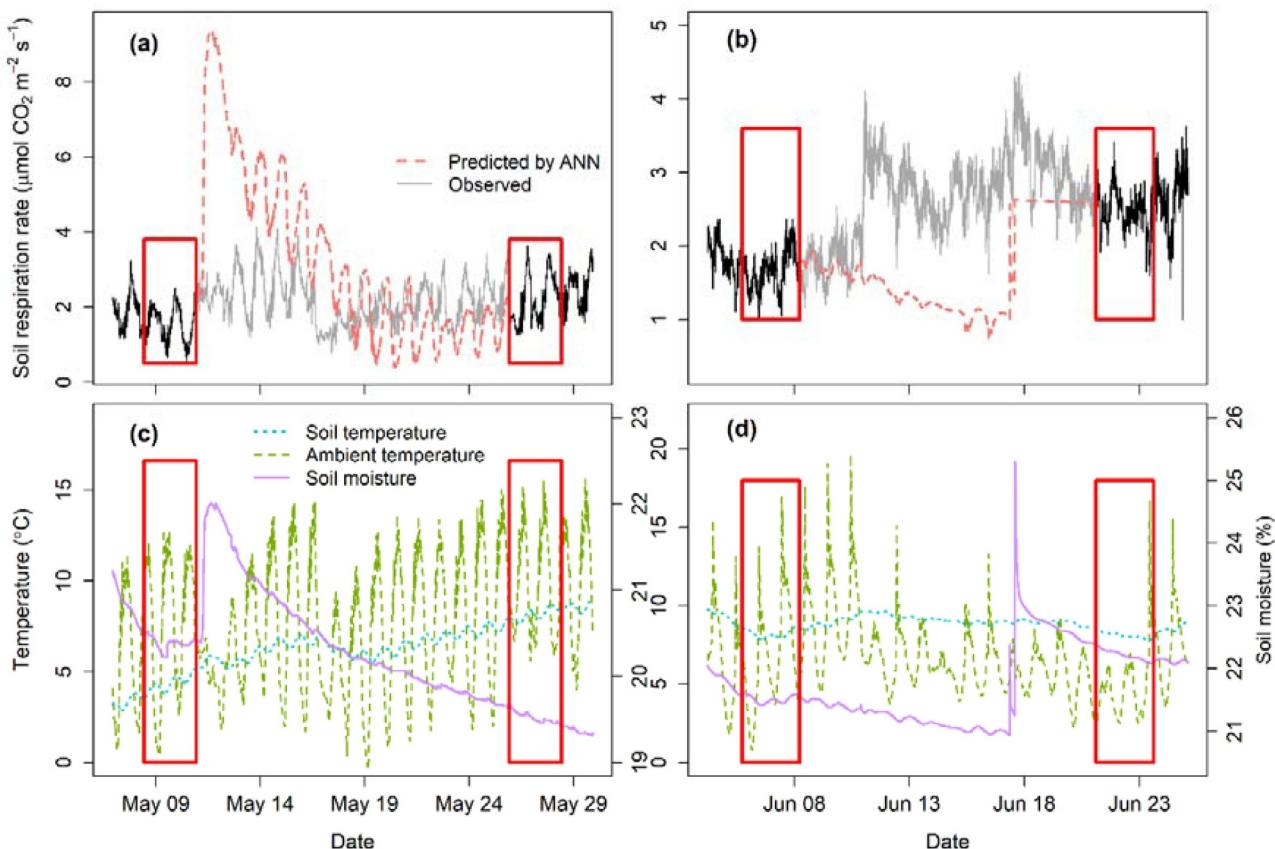


Fig. 8. Examples of the gap-filled soil respiration rates predicted by ANN (pink dashed lines) as compared to the original observations (gray solid lines) (a–b) in response to the pulse of the (purple solid lines, c–d) induced by the precipitation events. Soil temperature and ambient temperature are shown in blue dotted and green dashed lines, respectively (c–d). The black lines are observed values out of the gap. The red boxes indicate the corresponding sampling window for each gap. The data were derived from the mixed gap benchmark datasets at the plot H4 in Hurdal in 2018. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

referencing to environmental variables, these drifts in the gap could be hardly predicted by SSA (but could be predicted by NLS and ANN), and, thus, introduced large errors in the gap-filled values. In addition to the gap size, the completeness of the data in the sampling windows also plays a big role in the SSA predictions. Without sufficient data to construct the diurnal pattern, a “blind guess” by SSA can deviate well away from the observations (e.g., Fig. 9c). A high gap proportion (e.g., 50%) significantly increases the chances of incomplete data in the sampling window. This can increase the errors of the budget estimation with SSA, as shown, for example, in the H3 mixed-gap dataset (Fig. 7b). Therefore, caution needs to be taken when applying SSA for large gaps or highly fragmented datasets.

EM is a method whose performance relies on the relationship between the target time series and the references (Junger and Ponce de Leon, 2015). For our datasets, the accuracy of EM was higher than that of NLS in the 1-hour gaps but not in larger gaps (Fig. 3a–d). This may be partly due to the heterogeneity of our sites where time series considerably deviated from each other ($r < 0.85$, Table A1). A better accuracy from EM is expected in cases when more closely related time series are available to be used as references. Nevertheless, EM provides a practical approach to gap-fill missing data when parallel measurements are available, which is common for measurements such as soil respiration when multiple replicates are needed.

Due to the limited variations under the snow cover, soil temperature and soil respiration usually decouple from each other during the winter (Alm et al., 1999; Liptzin et al., 2009; Merbold et al., 2012). As expected, the NLS could not be applied to up to 71% of the small winter gaps in this study because of the low variance of the soil temperature. Compared to NLS, the time-series-analysis-based methods (i.e., SSA and

EM), which were implemented independent of the soil temperature, achieved a better accuracy in gaps ≤ 6 h and less biased predictions in the 15-day gaps in the winter. Even though the RMSE in the winter was generally much lower than in the non-winter season, larger relative errors were present in the winter gaps. This suggests that the errors introduced by filling winter gaps can account for a large fraction of the annual budget in higher latitude ecosystems when the winter contribution is a significant component (Alm et al., 1999; Aurela et al., 2002; Groffman et al., 2006; Zhao et al., 2016). According to our results, gap-filling methods that are based on time series analysis (i.e., SSA and EM), rather than the more conventional NLS approach, are recommended in ecosystems with substantial winter contributions.

4.2. Data structure restoration

The time series analyses are useful tools in ecology studies; however, they are usually very sensitive to gaps in the data (Barnea et al., 2008). In addition to achieving low errors, gap-filled data that restore the original structure of the time series can potentially be further used in time series analysis for, e.g., forecasting (Kohn and Ansley, 1986). Variance and AR (1) are two important structural properties of time series data (Box et al., 2015). In our study, the conventional method NLS as well as ANN largely underestimated the variances, especially during the winter time (Fig. 5). This is because the structures of their predicted data series closely followed the corresponding environmental variables, which usually have much lower variances than the flux data (Kunwor et al., 2017). The data series produced by EM were also low in variances and this is mainly due to the relatively low degrees of freedom (i.e., 10) we chose for spline constructions to achieve better

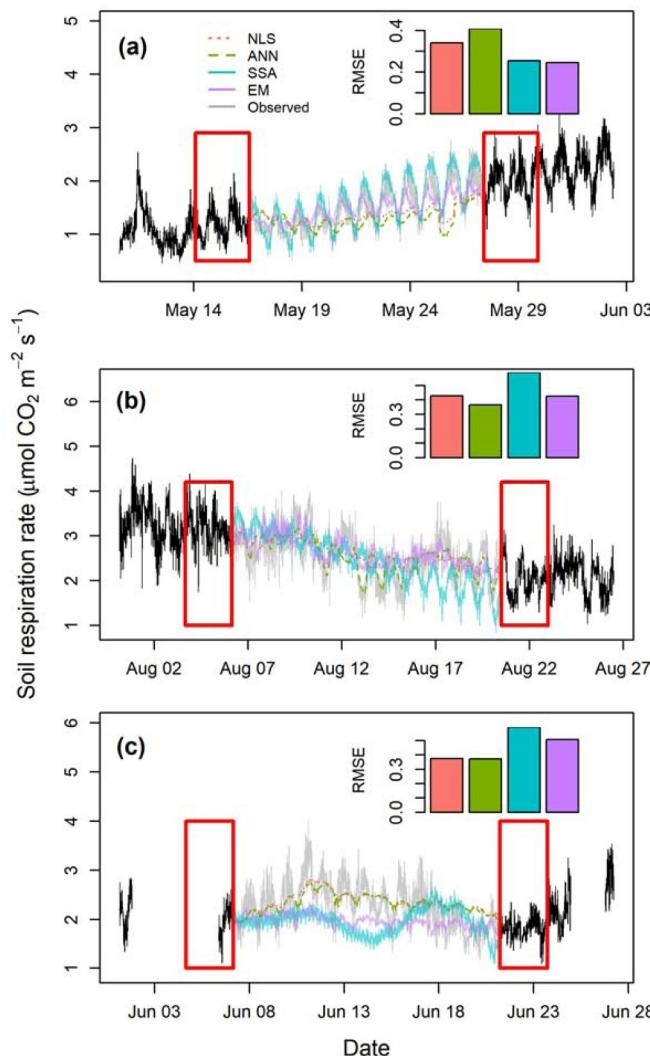


Fig. 9. Examples of the gap-filled soil respiration rates as compared to the original observations (gray lines) for gaps > 10 days derived from the mixed gap benchmark datasets at the plot H3 in Hurdal in 2018. The examples include scenarios where soil respiration diurnal pattern in the gap followed (a) and did not follow (b) the patterns in the sampling windows and where missing data were present in the sampling windows (c). Values predicted by NLS, ANN, SSA and EM are plotted in pink dotted, green dashed, blue solid and purple solid lines, respectively. The black lines are observed values out of the gap. The red boxes indicate the corresponding sampling window for each gap. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

accuracy. In contrast, only SSA, which considered the high frequency signal of the data in the sampling window, restored the variance to the same level as the observations.

For AR (1), we found the gap size showed a significant effect, which was mainly due to the systematic errors introduced to the auto-correlation coefficient estimations when sample size was small, known as finite size effects. The finite size effect is most evident in the observations of 1-hour gaps ($n = 5$) which had large variances that covered the full range from -1 to 1 , and, thus, hardly represented the AR

(1) structure of the original time series. Apart from the 1-h gaps, only EM attained the same level as the observations in most cases (Fig. 6) mainly because of the similar AR (1) structures between the target series and the references. In contrast, the other 3 methods all overestimated the AR (1). It is worth to note that even though the predictions from NLS showed lower AR (1) in the winter time to roughly the same degree as the observations, this low AR (1) was largely caused by the low winter soil temperature variances and, also, that 71% of the gaps with extremely low temperature variances were excluded in AR (1) calculations for NLS. Given that the AR (1) became much larger than the observations in the non-winter season, we suggest the capability of NLS to reproduce the AR (1) is limited, which agrees with Kunwor et al. (2017).

Overall, SSA and EM can restore the original structure of the data in certain respects. Nevertheless, to restore the full structural properties of the data is challenging and new methods are still needed.

5. Conclusions

All the four studied methods achieved good accuracy in estimating the annual soil respiration budget. However, they all have different strengths and weaknesses which need to be considered when used for gap-filling in different situations. NLS is a widely used method; however, due to its poor performance in the winter (i.e., failing to fit the equation and lack of variance), ecosystems with substantial winter contributions should consider other alternatives for gap-filling, at least during the winter. In contrast, SSA has great advantages (i.e., good accuracy, variance and less bias) for winter gaps, especially the small ones. Also, because SSA is independent from environmental variables, it can be implemented even when the driving variables are unavailable, missing or even unknown. Compared to SSA, implementation of EM is more limited by the availability of the corresponding reference data. With a highly related reference dataset, EM is also a reliable option to be used without involving environmental variables. ANN is a flexible approach to include other potential driving factors in the gap-filling models without specifying the exact mathematical formulation. However, its performance may suffer under certain conditions (e.g., a precipitation event). At the same time, the long computing time of ANN compared to its (usually disproportional) improvement in the accuracy over NLS must also be considered.

Funding

This work has been supported by the Norwegian Research Council (grant no. NFR-255061). All measured data used in this paper are produced within this project.

Acknowledgements

The authors are extremely grateful to Manon Sueur (ENSAIA, 2 avenue de la Forêt de Haye, 54505 Vandoeuvre-Les-Nancy, France) who took the initiative to gap-fill the soil respiration data using SSA during an internship at the Norwegian Institute of Bioeconomy Research in 2018.

Appendices

Figs. A1-A2 and Table A1.

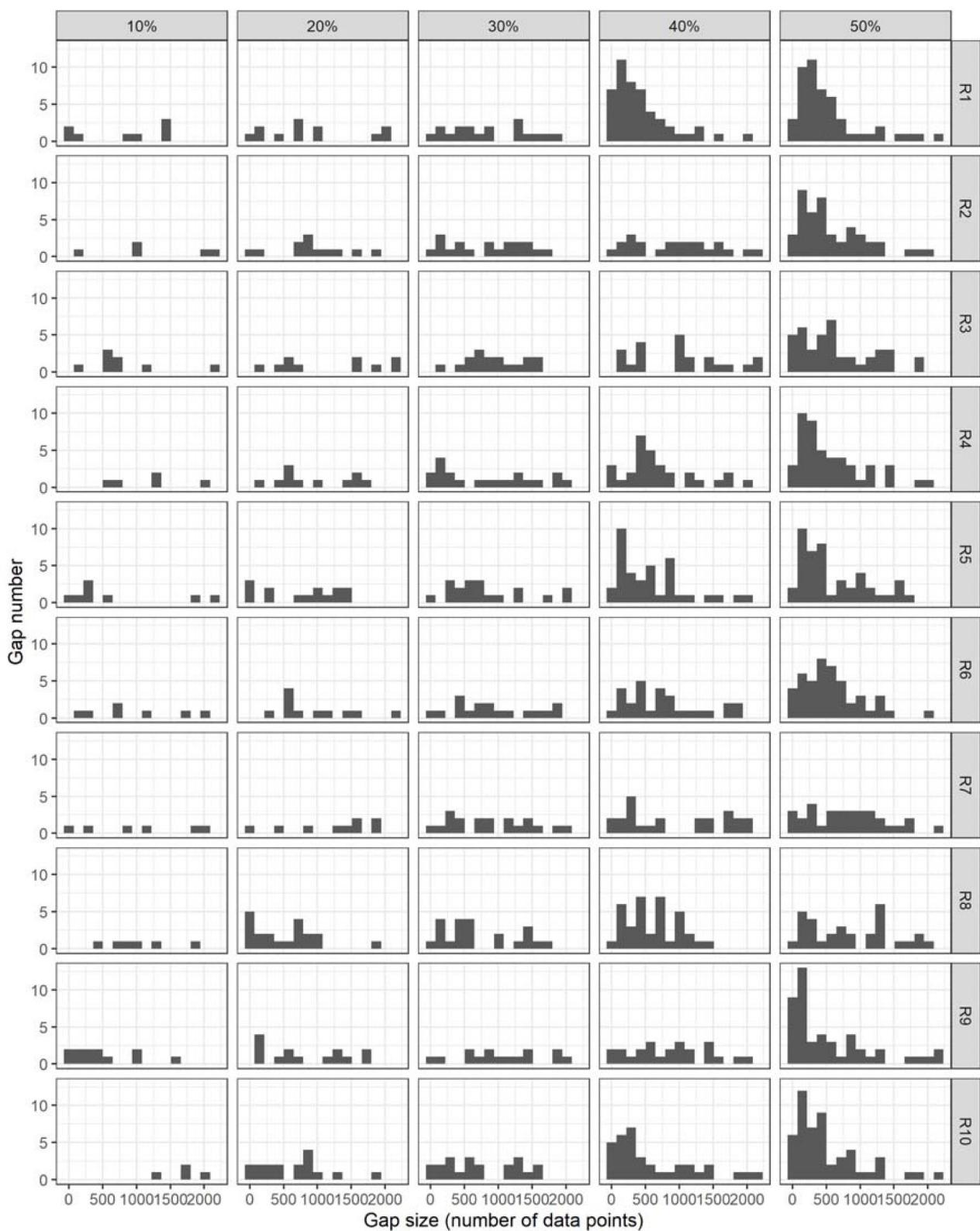


Fig. A1. Summary of the gap number and their sizes in the benchmark mixed gap datasets at the plot H1 in Hurdal. Columns indicate gap proportion from 10% to 50% and rows indicate 10 replicate datasets (R1–R10) for each gap proportion level. See Method and Materials [Section 2.2](#) for details.

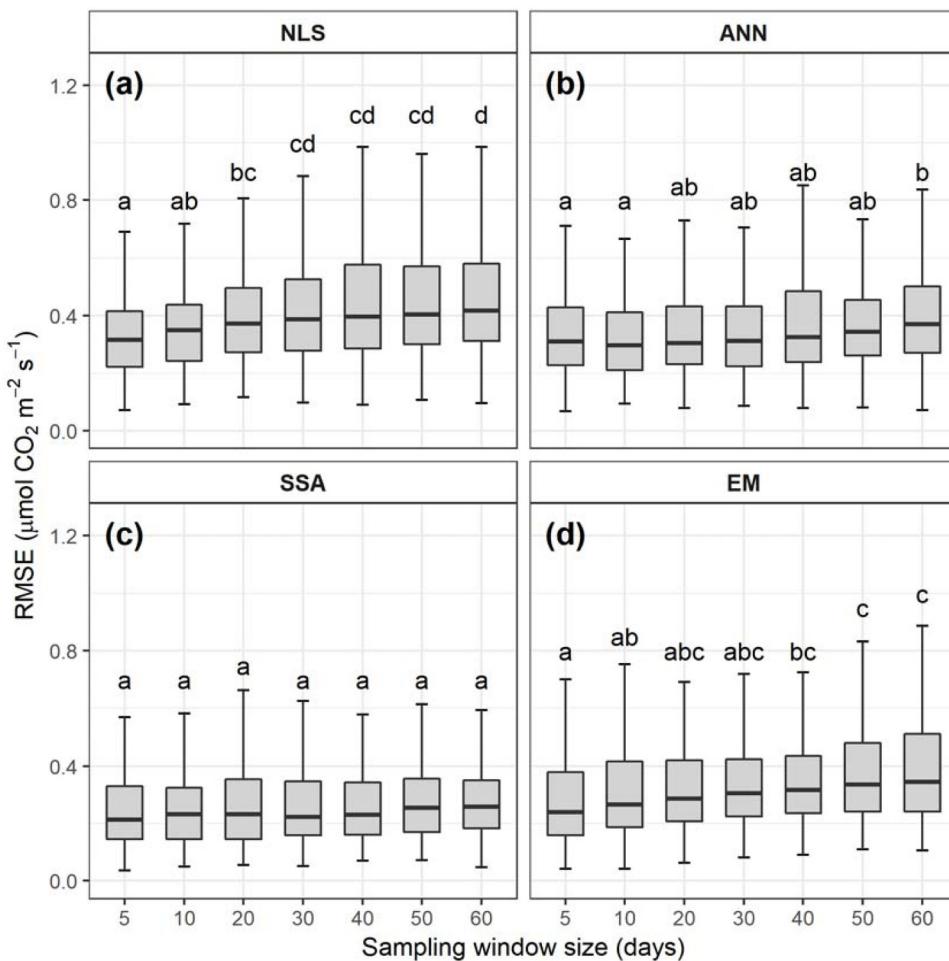


Fig. A2. The RMSE for the gap-filled values as a function of the sampling window size using different methods for the datasets that were aggregated to achieve a temporal resolution of 30-min. The bottom and top of a box indicate the lower and upper quartiles, respectively, and the horizontal line within the box is the median. The lower and upper whiskers are the 10th and 90th percentiles, respectively. Different lowercase letters indicate significant differences in the RMSE among different sampling window sizes ($P < 0.05$).

Table A1
Correlation matrix of the soil respiration time series.

Site	Plot ID	Hurdal				Løten	
		H1	H2	H3	H4	L1	L2
Hurdal	H1	–	0.57	0.81	0.78	–	–
	H2	0.57	–	0.84	0.65	–	–
	H3	0.81	0.84	–	0.83	–	–
	H4	0.78	0.65	0.83	–	–	–
Løten	L1	–	–	–	–	–	0.83
	L2	–	–	–	–	0.83	–

Pearson correlation coefficients (r) are shown among plots from the same site. Only those relationships with $r > 0.75$ (in bold) were taken for gap-filling in the expectation-maximization (EM) method (see details in the Method and Materials Section 2.4).

References

- Alm, J., Saarnio, S., Nykanen, H., Silvola, J., Martikainen, P.J., 1999. Winter CO₂, CH₄ and N₂O fluxes on some natural and drained boreal peatlands. *Biogeochemistry* 44 (2), 163–186.
- 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.-Atmos* 107 (D21) ACH 17-1-ACH 17-12.
- Bahn, M., et al., 2010. Soil respiration at mean annual temperature predicts annual total across vegetation types and biomes. *Biogeosciences* 7 (7), 2147–2157.
- Barnea, O., Solow, A.R., Stone, L., 2008. On fitting a model to a population time series with missing values. *Isr. J. Ecol. Evolut.* 52 (1), 1–10.
- Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.
- Bekku, Y., Koizumi, H., Oikawa, T., Iwaki, H., 1997. Examination of four methods for measuring soil respiration. *Appl. Soil. Ecol.* 5 (3), 247–254.
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M., 2015. Time series analysis: forecasting and control.
- Brooks, P.D., et al., 2011. Carbon and nitrogen cycling in snow-covered environments. *Geogr. Compass* 5 (9), 682–699.
- Buttlar, J.v., 2015. Spectral.methods: singular spectrum analysis (SSA) tools for time series analysis. R package version 0.7.2.133. <https://CRAN.R-project.org>.
- Carstensen, B., 2015. MethComp: analysis of method comparison studies. R package version 2.0. <http://BendixCarstensen.com/MethComp/>.
- Cook, F.J., Orchard, V.A., 2008. Relationships between soil respiration and soil moisture. *Soil Biol. Biochem.* 40 (5), 1013–1018.
- Dempster, A.P., Laird, N.M. and Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm journal of the Royal Statistical Society: series B (Methodological).
- Elzhov, T.V., Mullen, K.M. and Bolker, A.N.S.B., 2016. Minpack.lm: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds support for bounds. R package version 1.2-1. <https://CRAN.R-project.org>.
- Falge, E., et al., 2001. Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agric. For. Meteorol.* 107 (1), 43–69.
- Fritsch, S., Guenther, F. and Wright, M.N., 2019. Neuralnet: training of neural networks. R package version 1.44.2. <https://CRAN.R-project.org>.
- Ghil, M., et al., 2002. Advanced spectral methods for climatic time series. *Rev. Geophys.* 40 (1) 3–1–3–41.
- Golyandina, N., Korobeynikov, A., Shlemov, A. and Usevich, K., 2015. Multivariate and 2D extensions of singular spectrum analysis with the Rssa package. *J. Stat. Softw.*, 67(2): 1–78.
- Golyandina, N., Korobeynikov, A. and Zhigljavsky, A., 2018. Singular spectrum analysis with R.
- Golyandina, N., Osipov, E., 2007. The “Caterpillar”-SSA method for analysis of time series

- with missing values. *J. Stat. Plan. Inference* 137 (8), 2642–2653.
- Gomez-Casanovas, N., Anderson-Teixeira, K., Zeri, M., Bernacchi, C.J., Delucia, E.H., 2013. Gap filling strategies and error in estimating annual soil respiration. *Glob. Change Biol.* 19, 1941–1952.
- Groffman, P.M., Hardy, J.P., Driscoll, C.T., Fahey, T.J., 2006. Snow depth, soil freezing, and fluxes of carbon dioxide, nitrous oxide and methane in a northern hardwood forest. *Glob. Change Biol.* 12 (9), 1748–1760.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *J. Hydrol.* 377 (1–2), 80–91.
- Hursh, A., et al., 2017. . the sensitivity of soil respiration to soil temperature, moisture, and carbon supply at the global scale. *Glob. Change Biol.* 23, 2090–2103.
- IPCC, 2013. The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Jensen, L.S., et al., 1996. Soil surface CO₂ flux as an index of soil respiration in situ: a comparison of two chamber methods. *Soil Biol. Biochem.* 28 (10–11), 1297–1306.
- Junger, W. and Ponce de Leon, A., 2018. mtsdi: multivariate time series data imputation.: R package version 0.3.5. <https://CRAN.R-project.org>.
- Junger, W.L., Ponce de Leon, A., 2015. Imputation of missing data in time series for air pollutants. *Atmos. Environ.* 102, 96–104.
- Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.* 424–425, 264–277.
- Kohn, R., Ansley, C.F., 1986. Estimation, prediction, and interpolation for ARIMA models with missing data. *J. Am. Stat. Assoc.* 81 (395), 751–761.
- Kondrashov, D., Ghil, M., 2006. Spatio-temporal filling of missing points in geophysical data sets. *Nonlinear Process. Geophys.* 13 (2), 151–159.
- Kondrashov, D., Shprits, Y., Ghil, M., 2010. Gap filling of solar wind data by singular spectrum analysis. *Geophys. Res. Lett.* 37 (15).
- Kunwor, S., Starr, G., Loescher, H.W., Staudhammer, C.L., 2017. Preserving the variance in imputed eddy-covariance measurements: alternative methods for defensible gap filling. *Agric. For. Meteorol.* 232, 635–649.
- Kuznetsova, A., Brockhoff, P. and Christensen, R., 2016. LmerTest: tests in linear mixed effects models. R package version 2.0–30. <https://CRAN.R-project.org/package=LmerTest>.
- Lavoie, M., Phillips, C.L., Risk, D., 2015. A practical approach for uncertainty quantification of high-frequency soil respiration using forced diffusion chambers. *J. Geophys. Res.* 120, 128–146.
- Lenth, R., 2019. Emmeans: estimated marginal means, aka least-squares means. R package version 1.3.4. <https://CRAN.R-project.org>.
- Liptzin, D., et al., 2009. Process-level controls on CO₂ fluxes from a seasonally snow-covered subalpine meadow soil, Niwot Ridge, Colorado. *Biogeochemistry* 95 (1), 151–166.
- Lloyd, J., Taylor, J.A., 1994. On the temperature dependence of soil respiration. *Funct. Ecol.* 8, 315–323.
- Lu, W.Z., et al., 2017. Contrasting ecosystem CO₂ fluxes of inland and coastal wetlands: a meta-analysis of eddy covariance data. *Global. Change Biol.* 23 (3), 1180–1198.
- Lussana, C., Tveito, O.E., Dobler, A., Tunheim, K., 2019. SeNorge_2018, daily precipitation, and temperature datasets over Norway. *Earth Syst. Sci. Data* 11 (4), 1531–1551.
- Mahecha, M.D., Lange, H., Lischeid, G., 2011. Long-term structures in southern german runoff data. *Extremis: Disruptive Events and Trends in Climate and Hydrology*. Springer, pp. 250–265.
- Merbeld, L., Rogiers, N., Eugster, W., 2012. Winter CO₂ fluxes in a sub-alpine grassland in relation to snow cover, radiation and temperature. *Biogeochemistry* 111 (1–3), 287–302.
- Moffat, A.M., Beckstein, C., Churkina, G., Mund, M., Heimann, M., 2010. Characterization of ecosystem responses to climatic controls using artificial neural networks. *Glob. Change Biol.* 16 (10), 2737–2749.
- Moffat, A.M., et al., 2007. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agric. For. Meteorol.* 147, 209–232.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - a discussion of principles. *J. Hydrol.* 10 (3), 282–290.
- R. Development Core Team, 2018. R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria URL. <http://www.R-project.org/>.
- Raich, J.W., Schlesinger, W.H., 1992. The global carbon-dioxide flux in soil respiration and its relationship to vegetation and climate. *Tellus B* 44 (2), 81–99.
- Reichstein, M., et al., 2007. Determinants of terrestrial ecosystem carbon balance inferred from European eddy covariance flux sites. *Geophys. Res. Lett.* 34 (1).
- Reichstein, M., et al., 2003. Modeling temporal and large-scale spatial variability of soil respiration from soil water availability, temperature and vegetation productivity indices. *Glob. Biogeochem. Cy.* 17 (4).
- Risk, D., Nickerson, N., Creelman, C., McArthur, G., Owens, J., 2011. Forced diffusion soil flux: a new technique for continuous monitoring of soil gas efflux. *Agric. For. Meteorol.* 151, 1622–1631.
- Rochette, P., Gregorich, E.G., Desjardins, R.L., 1992. Comparison of static and dynamic closed chambers for measurement of soil respiration under field conditions. *Can. J. Soil Sci.* 72 (4), 605–609.
- Schoellhamer, H., 2001. Singular spectrum analysis for time series with missing data. *Geophys. Res. Lett.* 28, 3187–3190.
- Shun, T., Duffy, C.J., 1999. Low-frequency oscillations in precipitation, temperature, and runoff on a west facing mountain front: a hydrogeologic interpretation. *Water Resour. Res.* 35 (1), 191–201.
- Suh, S.U., et al., 2006. A chamber system with automatic opening and closing for continuously measuring soil respiration based on an open-flow dynamic method. *Ecol. Res.* 21 (3), 405–414.
- Wang, D., Liang, S., 2008. Singular Spectrum Analysis for Filling Gaps and Reducing Uncertainties of Modis Land Products. International Geoscience and Remote Sensing Symposium (IGARSS).
- Wickham, H., 2016. ggplot2: Elegant Graphics For Data Analysis. Springer-Verlag, New York.
- Wu, C., et al., 2016. Heterotrophic respiration does not acclimate to continuous warming in a subtropical forest. *Sci. Rep.-Uk.* 6, 21561.
- Zambrano-Bigiarini, M., 2017. hydroGOF: goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.3–10. URL <http://hzambran.gith>.
- Zhang, G., Eddy Patuwo, B.Y., Hu, M., 1998. Forecasting with artificial neural networks: the state of the art. *Int. J. Forecast.* 14 (1), 35–62.
- Zhao, J., Peichl, M., Oquist, M., Nilsson, M.B., 2016. Gross primary production controls the subsequent winter CO₂ exchange in a boreal peatland. *Global. Change Biol.* 22 (12), 4028–4037.
- Zhao, X., Huang, Y., 2015. A Comparison of Three Gap Filling Techniques for Eddy Covariance Net Carbon Fluxes in Short Vegetation Ecosystems. *Advances in Meteorology* 2015.