Received: 19 February 2018 | Revised: 22 January 2019 | Accepted: 5 February 2019

DOI: 10.1111/ejss.12805

SPECIAL ISSUE ARTICLE

European Journal of Soil Science WILEY

Evaluating CENTURY and Yasso soil carbon models for CO_2 emissions and organic carbon stocks of boreal forest soil with Bayesian multi-model inference

Boris Ťupek[1,2] | **Samuli Launiainen[1]** | **Mikko Peltoniemi[1]** | **Risto Sievänen[1]** | **Jari Perttunen[1]** | **Liisa Kulmala[2]** | **Timo Penttilä[1]** | **Antti-Jussi Lindroos[1]** | **Shoji Hashimoto[3]** | **Aleksi Lehtonen[1]**

[1]Bioeconomy and environment, Natural Resources Institute Finland, Helsinki, Finland

[2]Department of Forest Sciences, University of Helsinki, Helsinki, Finland

[3]Department of Forest Soils, Forestry and Forest Products Research Institute, Tsukuba, Japan

Correspondence

Boris Ťupek, Bioeconomy and environment, Natural Resources Institute Finland, Latokartanonkaari 9, FI - 00790 Helsinki, Finland.
Email: boris.tupek@luke.fi
Aleksi Lehtonen.
Email: aleksi.lehtonen@luke.fi

Funding information

Academy of Finland, Grant/Award Number: Climoss /296116; FutMon programme, Grant/Award Number: LIFE07 ENV/DE/000218; Climforisk, Grant/Award Number: LIFE09 ENV/FI/000571; Finnish Ministry of the Environment and the Ministry of Agriculture and Forestry, Grant/Award Number: Developing methods for GHG-inventory

We can curb climate change by improved management decisions for the most important terrestrial carbon pool, soil organic carbon stock (SOC). However, we need to be confident we can obtain the correct representation of the simultanous effect of the input of plant litter, soil temperature and water (which could be altered by climate or management) on the decomposition of soil organic matter. In this research, we used regression and Bayesian statistics for testing process-based models (Yasso07, Yasso15 and CENTURY) with soil heterotrophic respiration (Rh) and SOC, measured at four sites in Finland during 2015 and 2016. We extracted climate modifiers for calibration with Rh. The Rh values of Yasso07, Yasso15 and CENTURY models estimated with default parameterization correlated with measured monthly heterotrophic respiration. Despite a significant correlation, models on average underestimated measured soil respiration by 43%. After the Bayesian calibration, the fitted climate modifier of the Yasso07 model outperformed the Yasso15 and CENTURY models. The Yasso07 model had smaller residual mean square errors and temperature and water functions with fewer, thus more efficient, parameters than the other models. After calibration, there was a small overestimate of Rh by the models that used monotonic moisture functions and a small generic underestimate in autumn. The mismatch between measured and modelled Rh indicates that the Yasso and CENTURY models should be improved by adjusting climate modifiers of decomposition or by accounting for missing controls in, for example, microbial growth.

Highlights

- We tested soil carbon models against monthly soil Rh fluxes and amounts of SOC stock.
- The models accurately reproduced most of the seasonal Rh trends and amounts of SOC.
- Under autumn temperature and moisture, Rh was mismatched before and even after the parameterization.
- The seasonality of the temperature and water functions should be adjusted in models.

 © 2019 British Society of Soil Science

1 | INTRODUCTION

Forest soil has been a carbon sink over millennia because of its slightly larger ecosystem CO_2 sequestration than decomposition. However, the positive offset of the carbon balance might be unlikely in future climates (Crowther et al., 2016), especially in the northern latitudes where the soil carbon stocks are largest and climate change is most rapid (Delworth et al., 2016). Accurate predictions are needed to adopt the most appropriate strategies for preserving soil carbon stocks (Smith, 2005). However, an accurate and precise estimate of SOC and its change is still a major challenge.

A warmer climate could promote soil carbon sources instead of sinks (Crowther et al., 2016; Kirschbaum, 2000; Lal, 2009). Prolonged droughts could alter photosynthetic uptake or modify the soil respiration response to temperature (Davidson & Janssens, 2006; Gaumont-Guay et al., 2006). Neither soil carbon data nor soil carbon models show consensus on the response of decomposition to temperature and moisture extremes (Sierra et al., 2015; Van Gestel et al., 2018).

The Yasso07 (Tuomi et al., 2009) and CENTURY (Parton et al., 1994) are two state of the art soil carbon models widely applied for simulating changes in SOC and soil CO_2 emissions. For example, these models have a similar form but differ in various conceptual ways (e.g. pools, processes and interactions) representing organic matter decomposition and its dependence on environmental conditions (temperature and moisture and other variables) (Figure 1). However, prediction of the magnitude and spatial

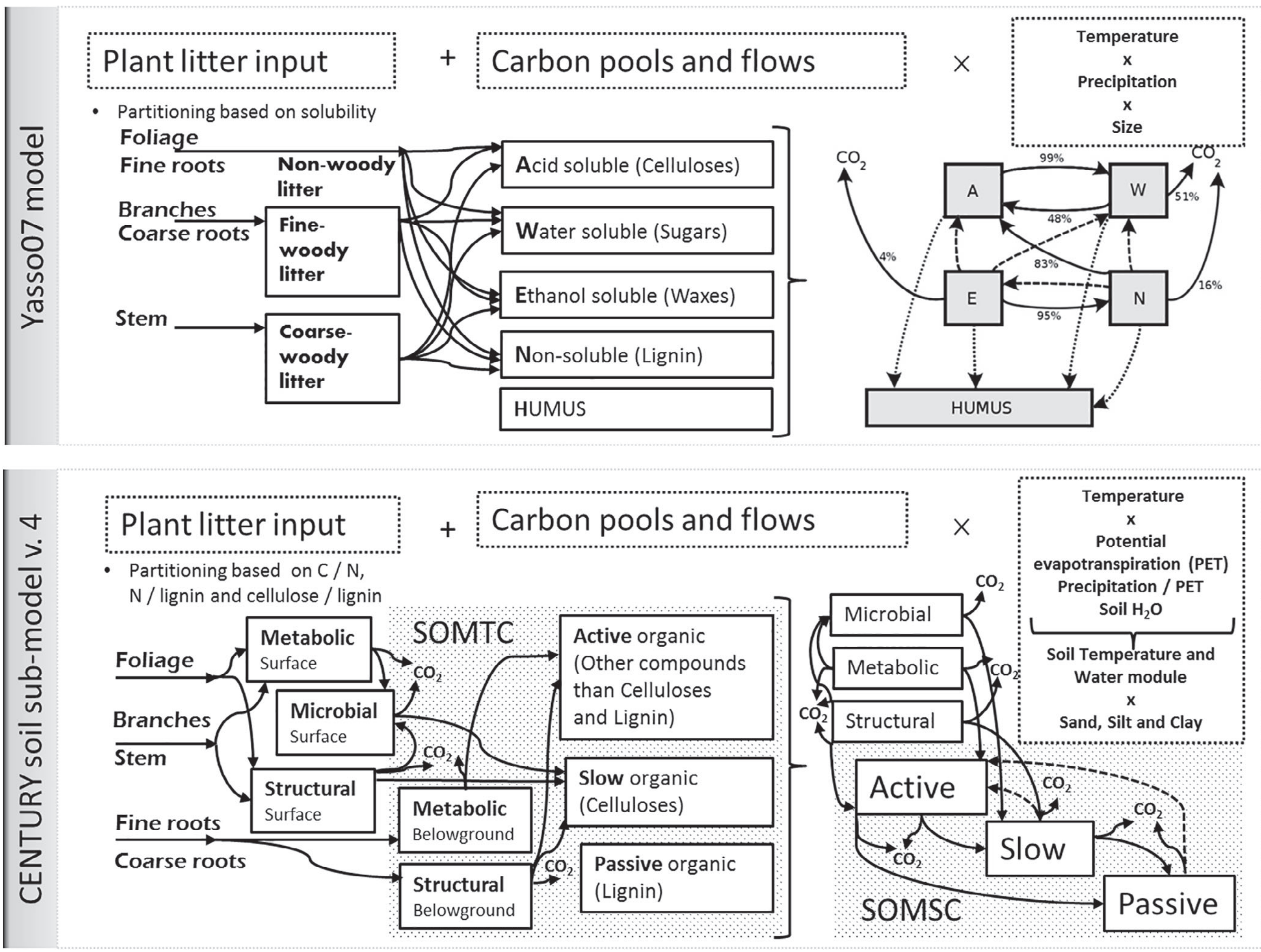

FIGURE 1 Conceptual representations of soil organic matter decomposition of the Yasso07 and Century models described in a general way as carbon (plant litter) entering the n number of time-dependent carbon pools and cross-pool flows controlled by a state of environmental conditions. The models differ in terms of their structure (pools and flows) and environmental dependence. The active pool of CENTURY in Yasso07 is represented by three pools (A, W and E) and rates of decomposition in Yasso07 are controlled by temperature and precipitation, but not explicitly by soil properties such as soil moisture and texture as in CENTURY. More detailed mathematical representations of the models are given in the Supporting Information File S1

variation of SOC is far from perfect (Hashimoto et al., 2017; Lehtonen et al., 2016; Todd-Brown et al., 2013; Ťupek et al., 2016). Model uncertainties hinder conclusions on both changes of SOC and CO_2 emissions (Lehtonen & Heikkinen, 2015). The imbalance between observed and modelled soil carbon stocks can be caused by incorrectly represented or missing biotic and abiotic factors of long-term soil carbon accumulation (Schmidt et al., 2011; Todd-Brown et al., 2013).

The Yasso07 model (Tuomi et al., 2009) was developed mainly to quantify changes in carbon stock of mineral soils. The model predicts changes in carbon stock and heterotrophic soil respiration from the balance between decomposing soil organic matter and litter input. Yasso07 was calibrated with almost 10,000 items of litter bag data from Europe, North and South America, and relatively few soil C stocks from Finnish forests (Tuomi et al., 2009). The model has been widely used for reporting SOC change in Finland to the United Nation Framework Convention on Climate Change (UNFCCC) and is also used together with Earth system models (Thum et al., 2011). Compared to Yasso07, Yasso15 (Järvenpää et al., 2015) has more detailed dependence of decomposition on temperature and has been calibrated against a larger dataset. CENTURY (Parton et al., 1994) is one of the most widely used soil carbon models of the Earth system models and is also used by Canada, Japan and the USA for soil carbon reporting to the UNFCCC. CENTURY was initially developed for grassland systems by Parton et al. (1994) and modified later to be applied also to boreal forests (e.g. Nalder & Wein, 2006). Unlike the Yasso models that do not need soil data, the CENTURY soil submodel v.4 requires soil input data (sand, silt and clay content, and bulk density) and by default operates at weekly rather than annual time-steps.

The Yasso07 and CENTURY models are used for national greenhouse gas reporting; however, neither has been tested with soil respiration and SOC data simultaneously. Furthermore, the models are mostly used with default calibrations. We aimed to test the performance of the soil carbon models Yasso07, Yasso15 and CENTURY for soil organic carbon stocks and heterotrophic respiration with and without calibration at four sites in Finland. We aimed to test the models with default parameters and to evaluate whether the expected mismatch between data and models is caused by parametrization or by the mathematical formulation of temperature and moisture functions. To test our hypothesis, we ran the models with the same litter fall data and separated the effects of functional forms from model parametrization of dependence on temperature and moisture with Bayesian inference.

2 | MATERIAL AND METHODS

2.1 | Study sites

Two Scots pine (*Pinus sylvestris* L.)-dominated and two Norway spruce (*Picea abies* L.)-dominated forest sites (Table 1) in the southern boreal zone of Finland (Figure S1a in Supporting Information File S1) were selected for this study. The sites are part of intensive monitoring of forest ecosystems (Level II) of the International Co-operative Programme on Assessment and Monitoring of Air Pollution Effects on Forests (ICP) (www.metla.fi/metinfo/forest-condition/programmes/intensive-monitoring.htm; Merilä et al., 2014). In October 2014, we trenched three square plots (1 m × 1 m) at each site to be able to subtract the autotrophic respiration of the tree roots from total forest floor CO_2 efflux to obtain soil heterotrophic respiration (Rh). The plots were divided further into four sub-squares (Figure S1b). The ingrowth of tree roots was prevented. On eight plots around the trench, we measured reference soil respiration, which included autotrophic and heterotrophic respiration (Figure S1b). At each forest site, we established three groups of trenched and control plots, yielding in total 12 trenched and 24 control plots. Respiration from the trenched plot (Rh) was used for comparison with the soil carbon models and total respiration from the control plots was used as a reference only.

2.2 | Soil respiration measurements and ancillary data

Forest floor respiration was measured once a week during the growing season (April–October) in 2015 and 2016, both on trenched and control plots (Figure S1). We used a portable infrared CO_2 analyser (EGM4, SRC-1 PP systems Inc., Amesbury, MA) connected to a closed-path ventilated non-transparent chamber (volume = 14.1 L; diameter = 30 cm). The measurements were made between 08.00 and 17.00, and the order in which the plots were measured at each station was random. The CO_2 concentration was measured every 4.8 s during 120 s of chamber closure and CO_2 fluxes (Figure S2) were calculated from the raw data (Jurasinski et al., 2014). During the flux measurements, we also measured the soil temperature (T) and moisture (SWC) with a portable thermometer and portable ThetaProbe (Delta-T Devices Ltd) at 5-, 10-, 15- and 20-cm depths. Soil temperature and moisture were also measured continuously by permanently installed sensors (iButton® temperature loggers from Maxim Integrated (San Jose, CA, USA); soil moisture sensors from Delta-T devices and Soil Scout Oy (Helsinki, Finland) (Figure S2).

TABLE 1 The characteristics of four ICP-level II sites used in this study (data from Merilä et al. (2014); Ťupek et al. (2015); Finnish Meteorological Institute)

Site name	Tammela pine	Tammela spruce	Punkaharju pine	Punkaharju spruce
Latitude /°	60.62	60.65	61.77	61.81
Longitude /°	23.84	23.81	29.33	29.32
Soil type [a]	Albic Arenosol	Haplic Arenosol	Rustic Podzol	Haplic Regosol
Sand content /%	98	59	97	68
Silt content /%	2	40	2	31
Clay content /%	0	1	1	1
Bulk density /g cm^{-3}	1.4	1.0	1.5	1.2
Humus C/N	32	30	35	31
Soil C/N	26	20	37	19
Total SOC up to 0.5 m /t C ha^{-1}	83.2	84	45	88.7
Stems /ha^{-1}	619	663	741	370
Stem volume /m^3 ha^{-1}	306	360	362	435
Basal area /m^2 ha^{-1}	29	33	32	34
Height /m	22	22	24	28
Diameter at 1.3 m /cm	25	26	24	35
Age /year	70	70	90	80
Annual temperature /°C	4.38	4.32	3.62	3.62
Annual precipitation /mm	627	625	593	594

[a]According to IUSS Working group WRB (2006) as cited in Merilä et al. (2014). SOC: soil organic carbon.

2.3 | Nonlinear least squared regression analysis

We used nonlinear least squared regression analysis (NLS) models for (a) evaluating responses of the instantaneous measurements of soil respiration (R, g CO_2 m^{-2} hr^{-1}) to environmental factors (T_5 and SWC_{10}), (b) the flux gap filling and (c) upscaling R to the monthly level (g CO_2 m^{-2} $month^{-1}$). The immediate R values were fitted to the corresponding T_5 and SWC_{10} (Figure S2 in File S1) separately for each site and treatment. We used a Q_{10}-based temperature response curve (Equation 1) modified by a response to soil water content (Davidson et al., 2012) (Equation 2):

$$R_{ij} = R_{\mathrm{ref}} Q_{10}^{\left(\frac{T_5-10}{10}\right)}, \tag{1}$$

$$R_{ij} = R_{\mathrm{ref}} d^{\left(\mathrm{SWC_{opt}} - \mathrm{SWC_{10}}\right)^2} Q_{10}^{\left(\frac{T_5-10}{10}\right)}, \tag{2}$$

where R is soil respiration, T_5 is soil temperature at 5-cm depth (°C) and SWC_{10} is volumetric soil moisture at 10-cm depth (%, m^3 m^{-3}). The R_{ref}, Q_{10}, SWC_{opt} and d are calibrated parameters for the i^{th} forest site and j^{th} treatment. The R_{ref} is the reference R (g CO_2 m^{-2} hr^{-1}) at 10°C, Q_{10} the relative increase in R per 10°C change in temperature and SWC_{opt} the optimum soil water content for respiration. The goodness of fit statistics and the parameter values are in Supporting Information, Table S1. To obtain monthly R (g CO_2 m^{-2} $month^{-1}$) we first estimated the continuous hourly R from continuous site-specific T_5 and SWC_{10} with Equation (2) (Figure S2). The monthly standard error of forest floor CO_2 fluxes was estimated as the standard deviation of model residuals divided by the square root of the number of CO_2 measurements and multiplied by the number of hours in a month.

2.4 | Soil carbon stock and CO_2 efflux modelling

We used Yasso07, Yasso15 and CENTURY to estimate initial SOC (January 1, 2014), monthly and annual SOC change and heterotrophic soil CO_2 respiration in 2014–2016. The initial SOC values were set to match the estimated equilibrium state between the litter input and decomposition for each site. The Yasso07 and Yasso15 models had a 3000-year spin up period, whereas for CENTURY it took 5000 years to reach equilibrium.

2.5 | Short model descriptions

Yasso07 is a reasonably simple soil carbon model (Tuomi et al., 2009) where soil C is divided into five pools based on plant litter chemistry (Figure 1). The rates of decomposition and C flows are affected by temperature and precipitation. The central assumptions of Yasso07 have been challenged in the Yasso15 model (Järvenpää et al., 2015), which (a) assumes different temperature and precipitation sensitivity between pools and (b) is calibrated against global SOC measurements. To allow inter-comparison between Yasso07, Yasso15 and CENTURY, we used the CENTURY soil sub-model only (Parton et al., 1994). In CENTURY, the soil carbon flows between structural, metabolic, active, slow and passive C pools with different turnover rates (Figure 1). Temperature and moisture modify the rates of decomposition. The rates of decomposition of the slow and passive pools also rely on lignin to N and C to N ratios. In the active pool, the rate of decomposition is modified by soil texture.

The models differ in their representation of soil temperature and moisture responses. CENTURY runs on air temperature and soil moisture (Kelly et al., 2000) or precipitation (Adair et al., 2008). The Yasso models run just on air temperature and precipitation (Järvenpää et al., 2015; Tuomi et al., 2009). The CENTURY model, unlike the Yasso models, has a sub-routine that computes soil temperature and water balance. The environmental modifier of the CENTURY model was altered in different applications. In CENTURY (Kelly et al., 2000) soil water function has an optimum, but temperature increases exponentially. In CENTURY (Adair et al., 2008) the temperature function has an optimum, whereas water function only saturates. Furthermore, in Adair et al. (2008) precipitation is relative to evapotranspiration but the ratio is limited to one, with the result that more precipitation than evapotranspiration does not reduce decomposition. Both Yasso models use the same equations for temperature and precipitation functions. Temperature dependence is exponential in both models, but in Yasso07 (Tuomi et al., 2009) it reaches an optimum and declines, unlike in Yasso15 (Järvenpää et al., 2015). Precipitation functions in the Yasso models are similar to CENTURY in Adair et al. (2008), in that they reach saturation although they do not account for potential evapotranspiration. More detail on the mathematical representation of the models is given in the Supporting Information, File S1.

2.6 | Model inputs

The Yasso07, Yasso15 and CENTURY models require air temperature, precipitation and litter as either monthly or annual input data. We used the same input for litterfall for all three soil carbon models (Figure S3). The daily weather data originated from the Finnish Meteorological Institute (www.fmi.fi). The litter input originated from the litterfall measurements for needles and branches (Ťupek et al., 2015; Liisa Ukonmaanaho, unpublished data), whereas stem, root and stump litter were modelled with data from Merilä et al. (2014) following Lehtonen et al. (2016). The spruce and pine needles were distributed in time (Figure S3). The annual litterfall of other components was equally distributed throughout the year (Figure S3). After trenching, we regarded fine and coarse tree roots as litterfall (Figure S3). The site-specific soil data required by CENTURY were available from Merilä et al. (2014).

2.7 | Model simulations

The Yasso07 and Yasso15 models are designed for simulations in annual time-steps. It is also possible to apply the model with a monthly time-step because of monthly timespans of litter-bag mass-loss measurements and calibration with global data, which account for considerable variation in climate. We ran the Yasso07 model using global parameters from Tuomi et al. (2009) and the Yasso15 model with parameters from Järvenpää et al. (2015) in annual and monthly time-steps. Running Yasso07 and Yasso15 in monthly time-steps (1/12 of yearly) required a transformation of monthly input data to representative ‘annual’ numbers. Monthly litter input and precipitation were multiplied by 12. The mean monthly air temperature was used directly without annual approximation. According to our tests of the feasibility of running the Yasso models in monthly time-steps, the predicted SOC and annual CO_2 respiration were not sensitive to the model time-step used.

We ran CENTURY using general parameters from the parameter file ‘tree.100,’ parameters of the site ‘AND H_J_ANDREWS’ for conifers and site ‘CWT Coweeta’ for deciduous trees (the file was available online at http://www.nrel.colostate.edu/projects/century/century-description.php from the model source code). The model accounted for topsoil N and plant litter C:N ratio, despite N being held constant during the simulations. The sensitivity of SOC stock to topsoil N and plant C:N ratio was weak compared to the sensitivity to litter input (Ťupek et al., 2016). We ran CENTURY simulations using two alternative temperature and moisture response functions for the rate of decomposition: Kelly et al. (2000) and Adair et al. (2008) (Table S2), later referred to as CENTURY.K and CENTURY.A, respectively. CENTURY estimated SOC and soil CO_2 emissions for the top 20 cm; thus to account for the deep soil carbon we increased the estimates by 40% following Jobbágy & Jackson (2000).

2.8 | Comparison of model outputs and measurements

To support the visual comparison of seasonal trends, we evaluated the performance of the models in predicting the annual and monthly soil heterotrophic respiration by linear regression statistics (slope, root mean square error and coefficient of determination) and Pearson correlation coefficient. The distributions of CO_2 values were near normal because of the seasonal character of the data. We assumed that monthly CO_2 values from separate sites were independent. In the comparison of annual SOC, we assumed uncertainty around the measured mean $\pm$ 12.8% (Häkkinen et al., 2011).

2.9 | Bayesian inference

To clarify whether the mismatch between the models' outputs and measurements originated from the formulation of temperature and moisture dependencies or their default parametrization, we constructed the empirical formulation of model matching (Equation 2). Each empirical model formulation consisted of the original temperature $f(T)$ and moisture $f(W)$ dependencies multiplied by the 'lumped' parameter of reference respiration R_{ref} (Equation 3):

$$\text{Empirical Model}_i = R_{\text{ref}} f(T) f(W), \tag{3}$$

where i represents the Yasso07, Yasso15, CENTURY.A and CENTURY.K models. The original temperature and moisture functions and their default parameters are in Tables S2 and S3. The prior values of R_{ref} were medians of monthly respiration (Table S3) estimated by each model in default settings for the model structure describing rates of decomposition for each pool, but with no climatic effect on rates of decomposition (Equation (S1) in File S1). In other words, for the R_{ref} simulations, the $\mathbf{A}(t)$ matrix describing carbon transfers and feedbacks between pools was set to default values, but the climatic effect on rates of decomposition $\xi(t)$ was equal to one. Prior and posterior change in the R_{ref} parameter accounted for changes in parametrization of the model structure separately from environmental functions.

For parametrization of the empirical models, we used measured soil respiration data, the general purpose Markov chain Monte Carlo (MCMC) sampler and Bayesian multimodel inference (Hartig et al., 2018). The median posterior parameters were used for simulations of the calibrated empirical models. The calibrated parameters of empirical models were intended only to estimate the best fit between Rh data and could not be applied to running full versions of the models. In empirical models the lumped parameter represented the base rate of carbon decomposition, which corresponds to respiration unaffected by environmental conditions in the original models. However, the calibrated lumped parameter does not apply to model runs with the original model's structure. We have not opted for calibrated equilibrium estimates of SOC, which would require full model calibrations. We compared the annual and seasonal trends of respiration simulated by calibrated empirical models with the same statistics for the models with default parameters. In addition, we also compared the models based on the deviance information criterion (DIC), which accounts for degrees of freedom by trying to estimate the effective number of parameters from MCMC outputs and is similar to Akaike's information criterion (AIC) (Spiegelhalter et al., 2002). We used R software for all data analyses (R core team, 2017).

3 | RESULTS

The predicted heterotrophic respiration (Rh) by Yasso07, Yasso15 and CENTURY identified the seasonal course of the observed Rh fluxes and environmental conditions (Figure 2, and Figures S2 and S3). As expected, the calibrated empirical models improved the absolute Rh values compared to the models with default parametrization. However, both default and calibrated empirical models showed a mismatch for Rh in both the summer and autumn.

3.1 | Models with default parametrization

The annual and monthly Rh values were, for models with default parameters, typically underestimated at all sites (Figures 2 and 3). The mean predicted annual Rh was on average 1.0 kg CO_2 m^{-2} $year^{-1}$; 44% only of the mean measured annual Rh (2.3 kg CO_2 m^{-2} $year^{-1}$) (Figure 3). The modelled monthly Rh accorded with the smallest but underestimated mean and the largest values.

The monthly predictions were correlated with the measured Rh (mean $r = 0.79$, $p < 0.001$). However, during the summer months the models failed to correlate significantly with the soil respiration measurements (Figure S2, Table S4). On average, the models underestimated observed summer Rh by 38% (Figure 3). Underestimation by the models with default settings clearly increased with temperature (Figure 4). The Yasso models showed a better fit to measurements and smaller residual error than the CENTURY models (Table S4). CENTURY simulations that used air temperature and precipitation as controlling factors (CENTURY.A) outperformed those that used air temperature and soil moisture (CENTURY.K) (Figures 2–4, and Table S4).

The equilibrium state forest SOCs estimated in the range from 7.0 to 11.6 kg C m^{-2} compared well to the measurements of Merilä et al. (2014), except for those at the Punkaharju pine site (Figure 2). At that site, all the models

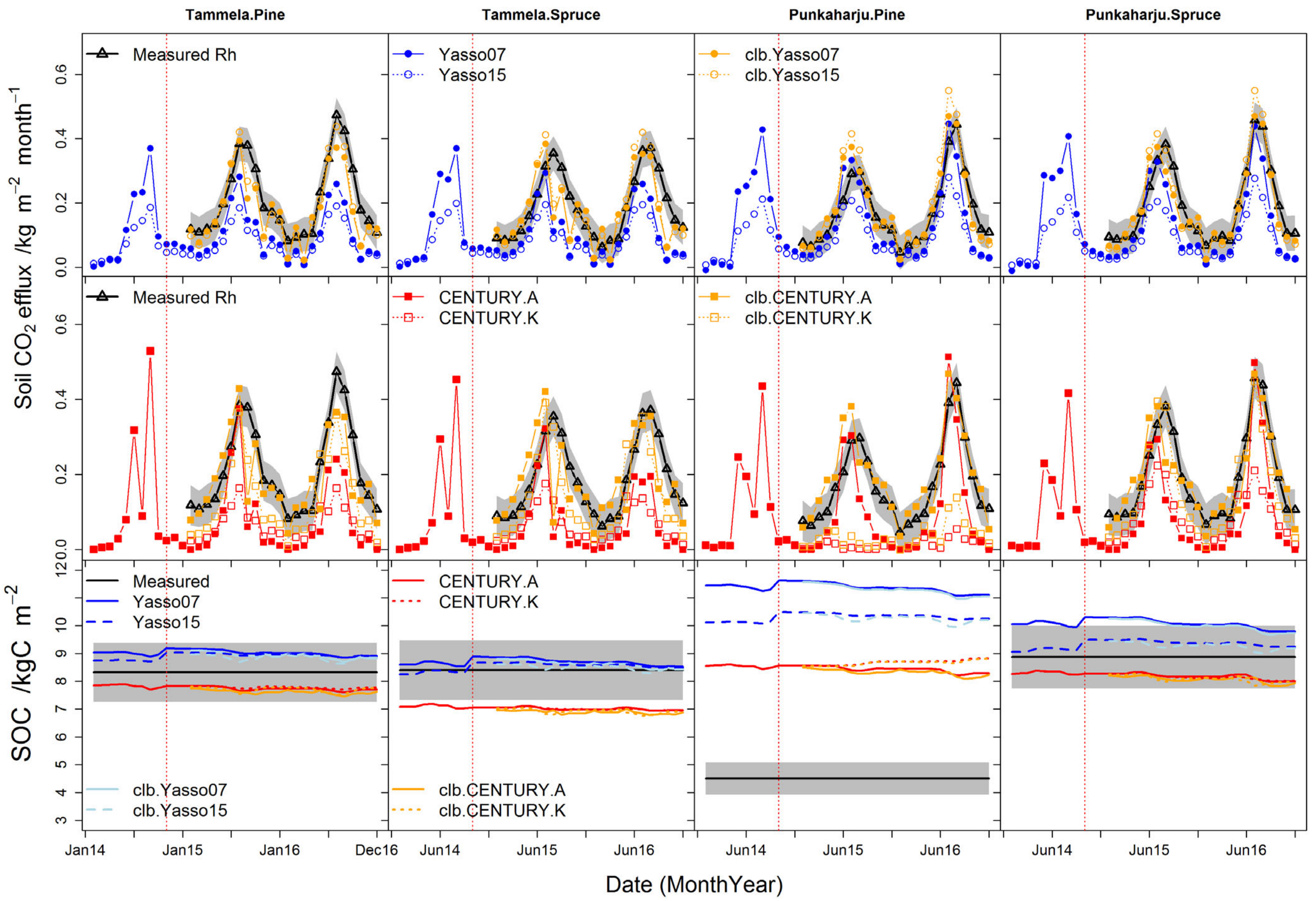

FIGURE 2 Simulated and measured monthly heterotrophic respiration (Rh, kg m^{-2} $month^{-1}$) for trenched plots from 2014 to 2016. Orange lines show Rh of calibrated empirical models. The lower panel shows simulated monthly soil organic carbon stocks, the effect of calibrated Rh on soil organic carbon (SOC) and the measured amount of SOC by Merilä et al. (2014). The grey shaded areas represent the uncertainty bounds of Rh and SOC stock measurements. The red dotted vertical line (October 2014) indicates the trenching date

estimated larger SOC stock (from 8.5 to 11.6 kg C m^{-2}) than that observed 4.5 kg C m^{-2}. The SOC stock of the Yasso models was within the error bounds of observations at three sites and that from CENTURY was in accord with the observations at two sites. The Yasso models showed more abrupt changes in SOC than CENTURY after trenching and the subsequent increase of the litter input from tree roots (Figure 2 and Figure S3). The small increase in CENTURY SOC at pine and not spruce sites before the trenching (Figure 2) was related to the different phenology of the foliar litterfall (pine maximum in the autumn and spruce maximum in the spring) (Figure S3).

3.2 | Calibrated empirical models

Mean posterior base respirations were twice as large for Yasso and four times larger for the CENTURY model (Table S3). The annual and monthly Rh of calibrated empirical models agreed well with measurements (Figures 2 and 3). However, autumn Rh was still underestimated by 26% on average (Figure 3, Table S4). Calibrated CENTURY.K Rh was especially underestimated at the Punkaharju pine site (Figure 3), the site with smaller amounts of soil water than the average for the others (Figure S2). The Rh residuals of calibrated empirical models did not show a clear relation with temperature (Figure 4). In relation to SWC, the calibrated empirical models slightly overestimated Rh values outside the moisture optimum (Figure 4). The Rh correlation statistics of calibrated empirical models favoured Yasso over CENTURY (Table S4). Model comparison by DIC also favoured the Yasso07 and Yasso15 models (−299 and −297, respectively) over CENTURY.A (−248) and CENTURY.K (338).

The empirical models comprising temperature and moisture functions and reference respiration showed almost identical Rh estimates to the soil carbon models with default parameters (Figure S4). The Rh estimates of empirical models in the climate space had a similar distribution to Rh for the NLS model based on observations (Figure S4). The models differed in their estimated Rh values and in their forms (e.g. whether they had or had not accounted for the reduction with moisture saturation). As expected, the

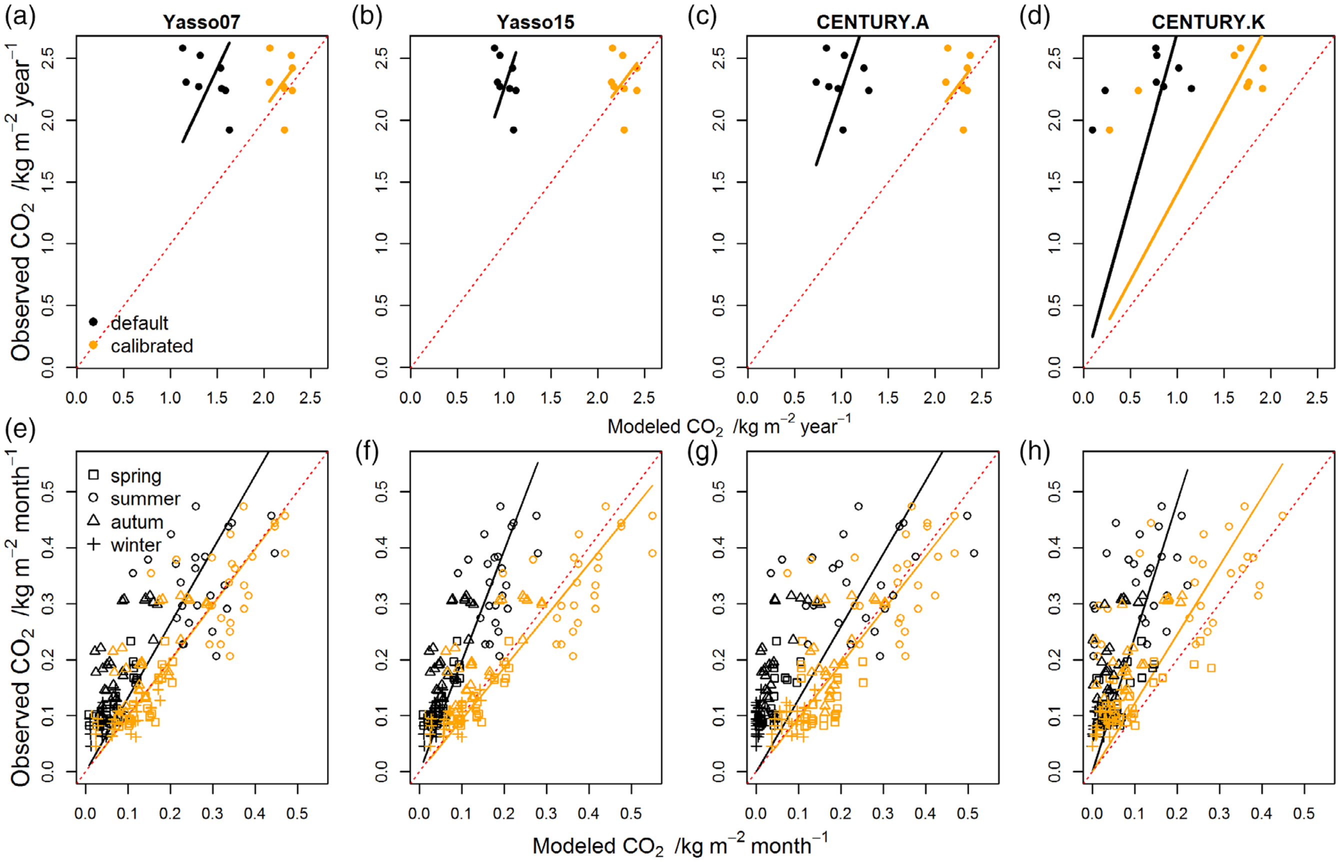

FIGURE 3 One-to-one plots between measured and modelled heterotrophic soil respiration (Rh, (a–d) kg CO_2 m^{-2} $year^{-1}$ and (e–h) kg m^{-2} $month^{-1}$). Orange points and trend lines correspond to calibrated empirical models. The annual and seasonal correlation statistics are in Table S2 in Supporting Information

recalibrated empirical models matched the distributions of the measured Rh data and of the NLS models (Figure S5). However, depending on a specific model's use of temperature and water functions, the Rh predictions showed little agreement outside the climate space of measured data (for air temperature over 20°C and for SWC over 45%) (Figure S5).

4 | DISCUSSION

We need to test process-based models with observations to increase our confidence in projected soil CO_2 emissions (Powlson, 1996). In this study, we evaluated measured SOC and heterotrophic respiration against estimates by the Yasso07, Yasso15 and CENTURY soil carbon models at monthly and annual intervals. The weak correlations between the measured and modelled CO_2 fluxes of Yasso07, Yasso15 and CENTURY soil carbon models (Figures 2 and 3) with their default parameters indicated a reduced ability to map the development of Rh according to the seasonal trends (weather and vegetation).

The models with default settings correlated with monthly Rh observations for the part of the year with lower temperatures, but there was no significant correlation for the summer months when soil moisture is likely to play its most important role. At the annual level, the models underestimated observed heterotrophic respiration by 43% on average. The diffence in Rh could be partly a result of parametric and structural uncertainty or errors in measured data (e.g. contributions of autotrophic respiration). In our study, forest floor vegetation was undisturbed; however, it contributed only slightly to the forest soil CO_2 observations (Kolari et al., 2009) and trenching excluded the main proportion of autotrophic respiration from the tree roots.

Bayesian calibration reduced the parameter uncertainty of all the models and greatly improved the fit for annual and monthly intervals (slopes close to 1). Temperature and water functions for Yasso and CENTURY models, as well as type and quality of input data, proved to be essential for the best fit. Regardless of whether temperature functions had or had not included the optimum and further decline of respiration, the calibrated empirical models slightly underestimated the observed data, mainly in autumn by 26% on average for all the models. This mismatch was probably related to water

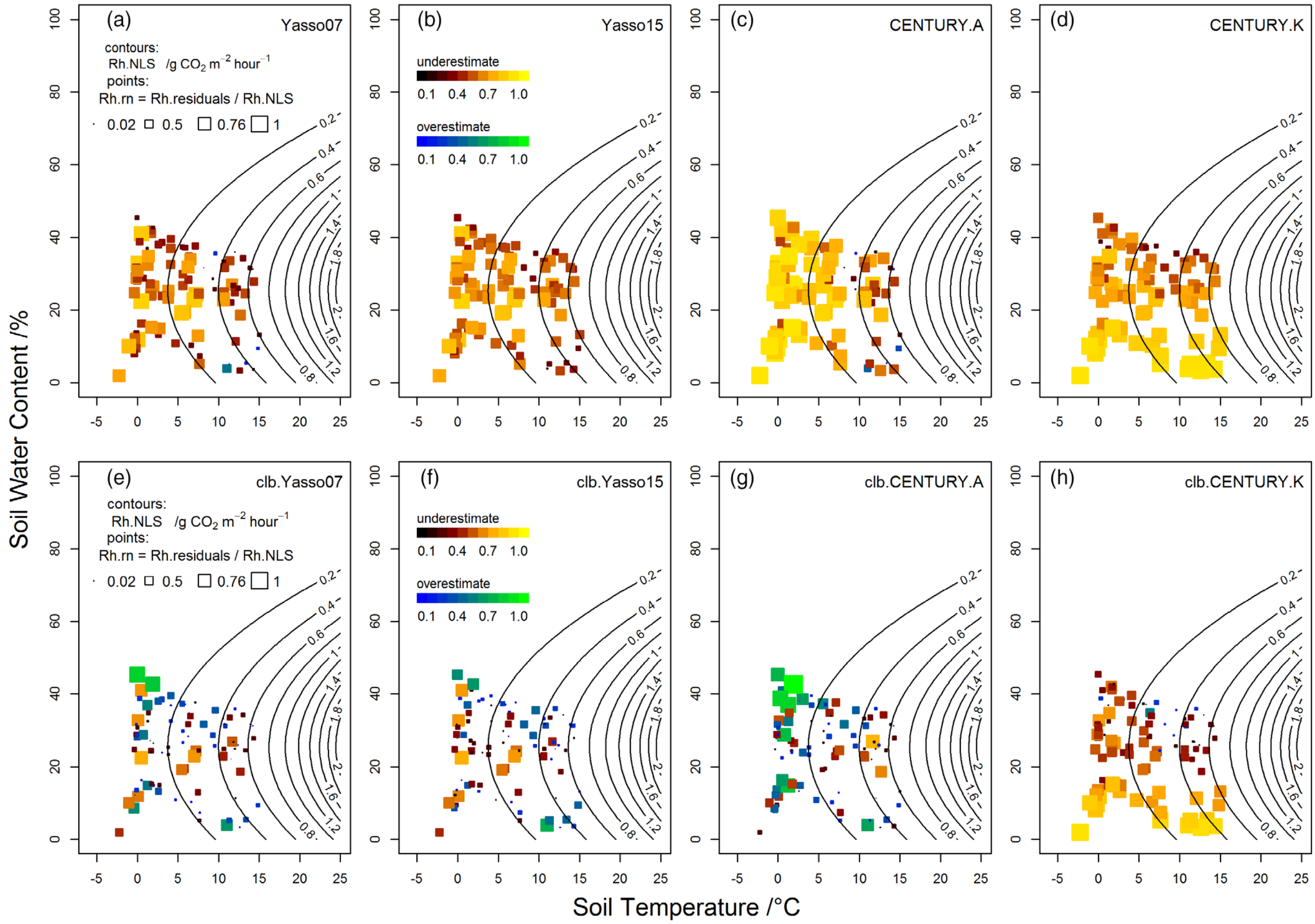

FIGURE 4 Normalized residuals (Rh.rn) between measured and modelled heterotrophic soil respiration (Rh, g CO_2 m^{-2} hr^{-1}) plotted in a climate space for soil carbon models. Contour lines show interpolated Rh.NLS (nonlinear least squares regression) values based on Equation (2) derived from Rh measurements. Note that Rh.residuals were normalized (Rh.rn) with Rh.NLS values. The Rh in panels (a–d) were modelled with default parameters and in panels (e–h) with calibrated empirical models

functions that did not acount for the reduction in Rh with large moisture content (unlike the moisture function fitted to measurements). In calibrated models the large change of prior and posterior parameters for base respiration suggested a strong influence of the model structure on the fit between measured and modelled soil respiration. However, the model structure represented the absolute difference in respiration rather than the difference in seasonal trend, which was reflected by the environmental functions.

Bayesian multi-model comparison by DIC identified Yasso models to be more plausible than CENTURY, probably because of fewer but more efficient parameters and smaller residual errors. Model ranking might have been different if SWC for CENTURY.K was generated with a water balance module. The CENTURY.K model using measured SWC data could have been biased at the Tammela pine forest because of exceptionally small SWC measurements compared to such data at the other sites.

Although the CENTURY.K model (Kelly et al., 2000) has double the number of parameters, it has temperature and water functions that are most similar to those used for interpolation of measurements (Davidson et al., 2012). When comparing CENTURY.K and CENTURY.A climate modifiers (DEFAC, a product of temperature and moisture modifiers) of Kelly et al. (2000) and Adair et al. (2008), CENTURY.K was more prone to reducing respiration under dry conditions. As a result, the climate modifier based on measured soil data overly limited potential decompositon and modelled Rh.

Differences in residuals and correlations between monthly predictions and observations, notably for spring and autumn respiration for the models, could be associated with differences in functional model formulation and or missing processes (Todd-Brown et al., 2013). Recent studies suggest that microbes represent a missing pathway in modelling soil carbon sequestration (Averill et al., 2014; Luo et al., 2016; Wieder et al., 2013). Increased root carbon allocation associated with increased carbon exudates and root turnover favours microbial and fungal development (Kaiser et al., 2010). In late summer microbial activity could increase with reallocation of carbon storage to roots after the allocation of new photosynthetic carbon to foliage and stem growth ceases (Kuptz et al., 2011). We assumed that adding representation of seasonality, for example modifying the temperature response of decomposition by accounting for the time lag of temperature-related Rh diffusion from the deep soil, could improve estimates of late summer respiration. On the other hand, we suggest that the autumn mismatch between the calibrated empirical models and observations could also indicate changes in microbial growth efficiency (MGE) because of newly shed foliar litterfall. The MGE dependence on decomposition and SOC accumulation is missing in first-order substrate-decomposing models such as CENTURY or Yasso, but could be decisive for soil carbon loss in a warming climate (Wieder et al., 2013).

Differences between the estimated soil carbon stocks for the equilibrium state forest and the SOC measurements of Merilä et al. (2014) might originate not only from uncertainty in the models but also from the uncertainty in measurements. The Punkaharju pine forests are less productive than spruce forests and small SOC values might still have reflected extensive slash and burn cultivation in the 19th century. The similarity between modelled and measured SOC on spruce sites, and the more considerable difference in pine sites, might also result from differences in plant litter production, which is a predominant factor for the models. The essential role of plant nutrient status in SOC accumulation (Fernández-Martínez et al., 2014), but its underrepresentation in soil C models (Ťupek et al., 2016), could partly explain the difference in measured SOC. The pine forest sites differed in the C/N ratio of the mineral soil, and the soil in the Tammela Pine forest was more moist and fertile than that in the Punkaharju pine forest.

Although CENTURY accounted for site-specific differences in both litterfall and soil characteristics, CENTURY SOC showed little variation between the sites, which was comparable to the Yasso models that do not use specific soil information. These spatially unchanging amounts of SOC were consistent with testing of the CENTURY model with data from a Swedish forest soil inventory, where its SOC differed only for soils with large clay content (Ťupek et al., 2016).

Monthly SOC followed the seasonal patterns of litter input, temperature and precipitation in all models; however, the SOC values from individual models differed. On an annual timescale, Yasso07 stored slightly more carbon in the soil thanYasso15. Such a difference between the pools and fluxes could have resulted from more CO_2 emissions from the pool with a slower rate of turnover (Kuzyakov, 2011). The difference in SOC and heterotrophic respiration between the two CENTURY versions was caused by the temperature response formulation because the model structure remained the same otherwise. The exponential temperature function used by Kelly et al. (2000) resulted in smaller summer CO_2 emissions and larger SOC than that of the Gaussian function of Adair et al. (2008). Although CENTURY has been found to be sensitive to litter input from the fine roots (McCormack et al., 2015), its SOC did not increase abruptly after trenching. The difference in CENTURY SOC development after trenching was a result of more gradual litter transfer between the carbon pools than for the Yasso07 and Yasso15 models.

5 | CONCLUSIONS

Our research has shown that soil carbon models developed for changes in SOC estimates with their default parametrization could not reliably predict the seasonal and long-term pattern of heterotrophic respiration. Despite the correlation between the observed soil heterotrophic respiration (Rh) and the monthly Yasso and CENTURY Rh estimates, the predicted Rh accounted for only half of the measured annual respiration. A better fit between measured and modelled soil respiration was obtained by Bayesian parametrization of the empirical models (model's empirical climate modifiers of the reference respiration). Based on a smaller underestimate and smaller deviance information criterion, the Yasso-based climate modifier was more plausible than CENTURY at the forest sites considered in this study.

We found that similar differences between the models that run with default parameters persisted after calibration of the functions of the environmental rate modifiers. The Yasso models with simpler functions for environmental modifiers fitted respiration data better than the CENTURY model with more parameters in the modifiers. The change of prior and posterior parameters for base respiration also suggested that the model structure had a strong influence on the fit between measured and modelled soil respiration. For more detailed comparison of the model structure rather than the environmental modifiers, however, more base respiration data from low temperature conditions and calibration with full versions of the models would be necessary.

We demonstrated that soil CO_2 emissions estimated based on changes in SOC from the Yasso and CENTURY models in default settings might be underestimated in greenhouse gas reporting. In addition, we clarified how estimates of soil respiration differ between these models depending on the type and parameterization of the temperature and moisture functions used. The data mismatch after calibration indicated that further improvement in the representation of environmental functions and accounting for missing processes (e. g. deep soil respiration, microbial controls) in the models is needed for accurate predictions of CO2 emissions under changing temperature and precipitation regimes of future climates.

ACKNOWLEDGEMENTS

This work was part of the project "Developing methods for GHG-inventory" funded by the Finnish Ministry of the Environment and the Ministry of Agriculture and Forestry. We acknowledge the following projects for support: Climforisk (LIFE09 ENV/FI/000571), FutMon programme (LIFE07 ENV/DE/000218), Finnish UN-ECE ICP Forests Level II monitoring program, particularly Liisa Ukonmaanaho for providing branch litter data, and Climoss (Academy of Finland no 296116). We thank our field team for collecting valuable data, Donald Smart for English language editing, and the reviewers and editors for valuable comments. We have no conflict of interest to declare.

ORCID

Boris Ťupek https://orcid.org/0000-0003-3466-0237
Aleksi Lehtonen https://orcid.org/0000-0003-1388-0388

REFERENCES

Adair, E. C., Parton, W. J., Del Grosso, S. J., Silver, W. L., Harmon, M. E., Hall, S. A., … Hart, S. C. (2008). Simple three-pool model accurately describes patterns of long-term litter decomposition in diverse climates. *Global Change Biology*, *14*, 2636–2660.

Averill, C., Turner, B. L., & Finzi, A. C. (2014). Mycorrhiza-mediated competition between plants and decomposers drives soil carbon storage. *Nature*, *505*, 543–545.

Crowther, T. W., Todd-Brown, K., Rowe, C., Wieder, W. R., Carey, J. C., Machmuller, M. B., … Allison, S. D. (2016). Quantifying global soil carbon losses in response to warming. *Nature*, *540*, 104–108.

Davidson, E. A., & Janssens, I. A. (2006). Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature*, *440*, 165–173.

Davidson, E. A., Samanta, S., Caramori, S. S., & Savage, K. (2012). The Dual Arrhenius and Michaelis–Menten kinetics model for decomposition of soil organic matter at hourly to seasonal time scales. *Global Change Biology*, *18*, 371–384.

Delworth, T. L., Zeng, F., Vecchi, G. A., Yang, X., Zhang, L., & Zhang, R. (2016). The North Atlantic Oscillation as a driver of rapid climate change in the Northern Hemisphere. *Nature Geoscience*, *9*, 509–512.

Fernández-Martínez, M., Vicca, S., Janssens, I. A., & Campioli, M. (2014). Nutrient availability as the key regulator of global forest carbon balance. *Nature Climate Change*, *4*, 471–476.

Gaumont-Guay, D., Black, T. A., Griffis, T. J., Barr, A. G., Jassal, R. S., & Nesic, Z. (2006). Interpreting the dependence of soil respiration on soil temperature and water content in a boreal aspen stand. *Agricultural and Forest Meteorology*, *140*, 220–235.

Häkkinen, M., Heikkinen, J., & Mäkipää, R. (2011). Soil carbon stock increases in the organic layer of boreal middle-aged stands. *Biogeosciences*, *8*, 1279–1289.

Hartig, F., Minunno, F., & Paul, S. (2018). *BayesianTools: General-Purpose MCMC and SMC Samplers and Tools for Bayesian Statistics.* R package version 0.1.6. Retrieved from https://cran.r-project.org/web/packages/BayesianTools/BayesianTools.pdf

Hashimoto, S., Nanko, K., Ťupek, B., & Lehtonen, A. (2017). Data-mining analysis of the global distribution of soil carbon in observational databases and Earth system models. *Geoscientific Model Development*, *10*, 1321–1337.

Järvenpää, M., Repo, A., Akujärvi, A., Kaasalainen, M., & Liski, J. (2015). *Bayesian calibration of Yasso15 soil carbon model using global-scale litter decomposition and carbon stock measurements.* Oral presentation at the 5th Nordic–Baltic Biometric Conference, Reykjavik, June 8–10 2015. Retrieved from http://math.tut.fi/inversegroup/material/jarvenpaa2015bayesian.pdf

Jobbágy, E. G., & Jackson, R. B. (2000). The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications*, *10*, 423–436.

Jurasinski, G., Koebsch, F., Guenther, A., & Beetz, S. (2014). *R Package "flux". Flux Rate Calculation from Dynamic Closed Chamber Measurement.* Retrieved from https://cran.r-project.org/web/packages/flux/flux.pdf

Kaiser, C., Koranda, M., Kitzler, B., Fuchslueger, L., Schnecker, J., Schweiger, P., … Richter, A. (2010). Belowground carbon allocation by trees drives seasonal patterns of extracellular enzyme activities by altering microbial community composition in a beech forest soil. *New Phytologist*, *187*, 843–858.

Kelly, R. H., Parton, W. J., Hartman, M. D., Stretch, L. K., Ojima, D. S., & Schimel, D. S. (2000). Intra-annual and interannual variability of ecosystem processes in shortgrass steppe. *Journal of Geophysical Research*, *105*, 20093–20100.

Kirschbaum, M. U. (2000). Will changes in soil organic carbon act as a positive or negative feedback on global warming? *Biogeochemistry*, *48*, 21–51.

Kolari, P., Kulmala, L., Pumpanen, J., Launiainen, S., Ilvesniemi, H., Han, P., & Nikinmaa, E. (2009). CO_2 exchange and component CO_2 fluxes of a boreal Scots pine forest. *Boreal Environment Research*, *14*, 761–783.

Kuptz, D., Fleischmann, F., Matyssek, R., & Grams, T. E. (2011). Seasonal patterns of carbon allocation to respiratory pools in 60-yr-old deciduous (*Fagus sylvatica*) and evergreen (*Picea abies*) trees assessed via whole-tree stable carbon isotope labeling. *New Phytologist*, *191*, 160–172.

Kuzyakov, Y. (2011). How to link soil C pools with CO_2 fluxes? *Biogeosciences*, *8*, 1523–1537.

Lal, R. (2009). Challenges and opportunities in soil organic matter research. *European Journal of Soil Science*, *60*, 158–169.

Lehtonen, A., & Heikkinen, J. (2015). Uncertainty of upland soil carbon sink estimate for Finland. *Canadian Journal of Forest Research*, *46*, 310–322.

Lehtonen, A., Linkosalo, T., Peltoniemi, M., Sievänen, R., Mäkipää, R., Tamminen, P., … Heikkinen, J. (2016). Forest soil carbon stock estimates in a nationwide inventory: Evaluating performance of the ROMULv and Yasso07 models in Finland. *Geoscientific Model Development*, *9*, 4169–4183.

Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., … Finzi, A. (2016). Toward more realistic projections of soil carbon dynamics by Earth system models. *Global Biogeochemical Cycles*, *30*, 40–56.

McCormack, M. L., Crisfield, E., Raczka, B., Schnekenburger, F., Eissenstat, D. M., & Smithwick, E. A. (2015). Sensitivity of four ecological models to adjustments in fine root turnover rate. *Ecological Modelling*, *297*, 107–117.

Merilä, P., Mustajärvi, K., Helmisaari, H., Hilli, S., Lindroos, A.-J., Nieminen, T. M., … Ukonmaanaho, L. (2014). Above-and belowground N stocks in coniferous boreal forests in Finland: Implications for sustainability of more intensive biomass utilization. *Forest Ecology and Management*, *311*, 17–28.

Nalder, I. A., & Wein, R. W. (2006). A model for the investigation of long-term carbon dynamics in boreal forests of western Canada: I. Model development and validation. *Ecological Modelling*, *192*, 37–66.

Parton, W. J., Schimel, D. S., Ojima, D. S., & Cole, C. V. (1994). A general model for soil organic matter dynamics: Sensitivity to litter chemistry, texture and management. In R. B. Bryant & R. W. Arnold (Eds.), *Quantitative modeling of soil forming processes* (pp. 147–167. SSSA Special Publication 39). Madison, WI: ASA, CSSA, and SSA.

Powlson, D. S. (1996). Why evaluate soil organic matter models? In D. S. Powlson, P. Smith, & J. U. Smith (Eds.), *Evaluation of soil organic matter models* (pp. 3–11). Berlin, Heidelberg, Germany: Springer–Verlag.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Schmidt, M. W. I., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., … Trumbore, S. E. (2011). Persistence of soil organic matter as an ecosystem property. *Nature*, *478*, 49–56.

Sierra, C. A., Trumbore, S. E., Davidson, E. A., Vicca, S., & Janssens, I. (2015). Sensitivity of decomposition rates of soil organic matter with respect to simultaneous changes in temperature and moisture. *Journal of Advances in Modeling Earth Systems*, *7*, 335–356.

Smith, P. (2005). An overview of the permanence of soil organic carbon stocks: Influence of direct human-induced, indirect and natural effects. *European Journal of Soil Science*, *56*, 673–680.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, *64*, 583–639.

Thum, T., Räisänen, P., Sevanto, S., Tuomi, M., Reick, C., Vesala, T., … Altimir, N. (2011). Soil carbon model alternatives for ECHAM5/JSBACH climate model: Evaluation and impacts on global carbon cycle estimates. *Journal of Geophysical Research: Biogeosciences*, *116*, G02028. https://doi.org/10.1029/2010JG001612

Todd-Brown, K., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. A. G., & Allison, S. D. (2013). Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences*, *10*, 1717–1736.

Tuomi, M., Thum, T., Järvinen, H., Fronzek, S., Berg, B., Harmon, M., … Liski, J. (2009). Leaf litter decomposition—Estimates of global variability based on Yasso07 model. *Ecological Modelling*, *220*, 3362–3371.

Ťupek, B., Mäkipää, R., Heikkinen, J., Peltoniemi, M., Ukonmaanaho, L., Hokkanen, T., … Lehtonen, A. (2015). Foliar turnover rates in Finland–comparing estimates from needle-cohort and litterfall-biomass methods. *Boreal Environment Research*, *20*, 283–304.

Ťupek, B., Ortiz, C. A., Hashimoto, S., Stendahl, J., Dahlgren, J., Karltun, E., & Lehtonen, A. (2016). Underestimation of boreal soil carbon stocks by mathematical soil carbon models linked to soil nutrient status. *Biogeosciences*, *13*, 4439–4459.

Van Gestel, N., Shi, Z., Van Groenigen, K. J., Osenberg, C. W., Andresen, L. C., Dukes, J. S., … Reich, P. B. (2018). Predicting soil carbon loss with warming. *Nature*, *554*, E4–E5. https://doi.org/10.1038/nature25745

Wieder, W. R., Bonan, G. B., & Allison, S. D. (2013). Global soil carbon projections are improved by modelling microbial processes. *Nature Climate Change*, *3*, 909–912.

SUPPORTING INFORMATION

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

How to cite this article: Ťupek B, Launiainen S, Peltoniemi M, et al. Evaluating CENTURY and Yasso soil carbon models for CO_2 emissions and organic carbon stocks of boreal forest soil with Bayesian multi-model inference. *Eur J Soil Sci*. 2019; 70:847–858. https://doi.org/10.1111/ejss.12805