ISSN 1067-4136, Russian Journal of Ecology, 2022, Vol. 53, No. 4, pp. 295–307. © Pleiades Publishing, Ltd., 2022.
Russian Text © The Author(s), 2022, published in Ekologiya, 2022, No. 4, pp. 299–311.

Soil Respiration Variability: Contributions of Space and Time Estimated Using the Random Forest Algorithm

I. A. Smorkalov*

Institute of Plant and Animal Ecology, Ural Branch, Russian Academy of Sciences, Yekaterinburg, 620144 Russia

**e-mail: ivan.a.smorkalov@gmail.com*

Received January 25, 2022; revised February 9, 2022; accepted February 16, 2022

Abstract—Soil respiration modeling (i.e., simulation of carbon dioxide emissions from the soil surface) makes it possible to analyze and forecast changes in the carbon cycle in terrestrial ecosystems. Along with classical regression models, researchers currently use machine learning methods based on neural networks or regression tree ensembles. However, models produced using these methods are often used as 'black boxes', which hinders the analysis of process mechanisms. This paper demonstrates capabilities of the Random Forest algorithm that can be successfully used to estimate effects exercised by various factors on soil respiration based on features importance measurements. Using variance separation, predictors have been classified either as spatial (biotope type, soil type, vegetation type, and soil moisture content) or temporal (soil and air temperature, NDVI, LAI, FPAR, and SPEI). Several models were produced based on 5670 respiration measurements performed during five growing seasons (2012–2016) on 30 sampling plots in south-taiga pine forests and meadows that feature different vegetation and soil types but are confined to the same small area. Different models include different sets of predictors (all predictors, temporal predictors only, spatial predictors only, and temperature and humidity only), and their accuracy reaches $R^2 = 0.88$ ($MSE = 0.47$). It is established that soil respiration depends primarily on temporal factors whose importance ranges from 76 to 91%. In forests, the effect of spatial factors on respiration is stronger than in meadows.

Keywords: CO_2 emission, carbon cycle, machine learning, forest ecosystems, temperature, moisture content, environmental factors

DOI: 10.1134/S1067413622040051

INTRODUCTION

Mathematical modeling plays an important role in the analysis and forecast of changes in the carbon cycle. To enhance the quality of models, it is important to know the key factors that affect the intensity of the main cycle components on various temporal and spatial scales [1].

Mathematical modeling of soil respiration has been developing for more than 50 years [2]. The most frequently used predictors are soil temperature [3, 4] and/or soil moisture content [3, 5–7] since these parameters can be easily measured simultaneously with respiration measurements. The identified relationships between respiration and these predictors can be linear [8], quadratic [9], exponential [10], or power-law [11].

Authors who use only temperature and humidity as predictors present in each particular case satisfactorily describe only their own data and are applicable only on a scale of tens of meters. The nature of the dependences may differ between individual sampling points even in the framework of the same study [3]. This happens because many factors determining the spatial and temporal variability of CO_2 fluxes are not taken into account [12].

An advantage of classical multiple regression methods is the possibility to directly assess the importance of different predictors based on model coefficients and select the most significant ones (e.g., using stepwise regression [12]). However, these methods have significant limitations, including requirements to the data types used (it is difficult to operate with variables on a nominal scale) and relationships between individual factors (there must be no correlation between predictors). Therefore, an alternative approach is currently being actively developed: machine learning-based modeling.

One of the most common and popular machine learning techniques are artificial neural networks. They can use from several [16] to several dozen variables [17]. Such models describe soil respiration both at local [18] and global [19] scales. However, the authors almost never disclose the architecture of their networks; they only list variables used to identify the relationship and provide the number of nodes (blocks) in the neural network [19, 20]. As a result, no one can

use such models (e.g., verify them based on other data sets) except for their authors.

To ensure that results using a neural network are not limited to "black boxes" but give an insight into process mechanisms, a special group of artificial intelligence techniques has been developed: explainable artificial intelligence. It is even recommended to abandon the use of "black boxes" in favor of natively interpretable models (21).

One such interpretable model type is the Random Forest (RF) algorithm; it is increasingly widely used in tasks involving regression, classification, and identification of the most informative features [22]. The algorithm is fundamentally different from the approach used in neural networks. It involves using a large number (ensemble) of decision trees; each of these trees is built based on a subsample isolated from the original sample using random data selection and a portion of available predictors [22, 23]. To ensure its high accuracy, responses received from many trees are averaged. The main advantages of the method are as follows: (1) relative protection from overfitting, even if the number of features exceeds the number of observations; (2) only two parameters are sufficient for tuning (the number of trees and the maximum number of features used for separation); (3) features measured on different scales (ratio, ordinal, and nominal ones) can be used; (4) importance of the selected predictors for the model accuracy (feature importance) can be evaluated; and (5) resistance to outliers. The main disadvantages of the method are its inability to extrapolate, heavyweight models, and poor handling of linear relationships.

The RF application in soil respiration modeling has just begun [24–31]; so, its popularity is still lower compared to classical regression. Even though RF ensures a greater accuracy in comparison with classical models, not all algorithm features suitable for the analysis of effects exercised by environmental factors on soil respiration are used in modern studies. In other words, most authors do not apply interpretation tools to simulation results.

Objective—Not so much to produce yet another CO_2 emission model for a specific situation but demonstrate the capabilities of the RF algorithm suitable for the assessment of effects exercised by environmental factors on soil respiration rate. Typical south-taiga forest and meadow biotopes with contrasting vegetation were selected within a relatively small area, thus, making it possible to exclude the effect of climatic differences inevitable on a macroscale. The main idea of the study was to analyze the roles of two groups of factors: (1) factors more variable in space; and (2) factors more variable in time. Their contributions to soil respiration variability were assessed in two ways: (1) by comparing the importance of individual predictors; and (2) by comparing the quality of models with different combinations of predictors.

MATERIALS AND METHODS

Study Area

The study area is located in the south-taiga subzone, 30 km southeast of Yekaterinburg. Field studies were performed in different variants of forest and meadow biotopes (Table 1). Ten sampling sites were selected: 7 forest and 3 meadow sites (Fig. 1). The sites are located within a triangle whose side dimensions are about 1 km (vertex coordinates: 56.6072° N, 61.0480° E; 56.6072° N, 61.0682° E; and 56.5998° N, 61.0602° E). Distances between neighboring sites are 30–150 m. Three sampling plots were established on each of the sites (30 plots in total). In pine forests (except for two variants), their size was 10 × 10 m; in wood-sorrel and dead-cover pine forests, 2 × 4 m; and in meadows, 5 × 5 m.

In pine forests, sampling sites were established on similar soils (except for dead-cover and young pine forests); however, they differ significantly in the diversity and abundance of the grass–dwarf-shrub layer: from dead-cover to forb (Table 1). In addition, the fireweed pine forest shows marks of a recent ground fire; while soils under the young pine forest forming on old arable lands are mechanically disturbed and have no ground vegetation. The three meadow sites were established in different edaphic conditions: from a dry mowed meadow (M1) to a waterlogged meadowsweet meadow. All soils are slightly acidic (pH 4.7–6.0).

CO_2 Emission Measurements

The rate of CO_2 fluxes from the soil surface was measured using the standard version of the closed dynamic chamber method [32]. An Li-8100A field respirometer (LI-COR Biosciences, USA) and polypropylene rings with an inner diameter of 105 mm previously installed in the soil at a depth of 3 cm were used. On each sampling plot, measurements were performed at ten randomly selected permanent points (five permanent points in dead-cover and wood-sorrel pine forests).

Studies were performed during the growing seasons of 2012–2016 (from May to October). In total, 21 measurement rounds were performed (2012: August 23–24 and October 1–2; 2013: May 5–6, May 28–29, June 20–21, July 24–25, August 27–28, September 24–25, and October 22–23; 2014: May 10–11, May 27–28, June 26–27, July 30–31, September 3–4, and October 1–2; 2015: May 27–28 and June 29–30; and 2016: May 30–31, July 5–6, August 2–3, and October 5–6). The measurements were performed during daylight hours (from 10:00 to 16:00). Special studies have shown that soil respiration in this period does not differ significantly from the average daily values, which makes it possible to compare results at different times on different plots [33]. In total, 5670 measurements were performed.

Table 1. Descriptions of studied sites

No.	Biotope	Code	Dominants in the arborescent and grass–dwarf-shrub layers	Soil type (WRB)	Soil type (Classification and Diagnostics of Russian Soils)
			Pine forests		
1	Blueberry pine forest	Sch	*Pinus sylvestris*, single *Betula* spp. and *Sorbus aucuparia* in the undergrowth; *Vaccinium myrtillus*, gramineous species	Albic Retisols (Differentic)	Typical sod–podzolic soil
2	Eagle-fern pine forest	Spap	*P. sylvestris*, single *S. aucuparia* in the undergrowth; *Pteridium aquilinum*, *Calamagrostis arundinaceae*, *V. myrtillus*	Albic Retisols (Differentic) and Pretic Luvisols (Siltic)	Typical sod–podzolic soil and podzolized brown forest soil
3	Forb (fireweed) pine forest	SDG	*P. sylvestris*, single *Alnus* spp. in the undergrowth; *Chamaenerion angustifolium*	Albic Luvisols (Loamic)	Podzolized brown forest soil
4	Gramineous pine forest	SZL	*P. sylvestris*, single *Betula* spp. in the undergrowth; *C. arundinaceae*, *Brachypodium pinnatum*	Albic Retisols (Clayic)	Sod–podzolic podzolized soil
5	Pine forest forming on old arable lands	SSG	*P. sylvestris* undergrowth (10–15 years)	Skeletic Combisols (Densic, Turbic)	Turbated brown forest soil in combination with turbozem soil
6	Dead-cover pine forest	SDC	*P. sylvestris*	Eutric Retisols (Densic)	Typical sod–podzolic soil
7	Wood-sorrel pine forest	SA	*P. sylvestris*, single *S. aucuparia* in the undergrowth; *Oxalis acetosella, Rubus saxatilis*	Albic Luvisols (Loamic)	Podzolized brown forest soil
			Meadows		
8	Dry forb meadow	M-1	*Poa* spp., *Geum rivale, Arctium tomentosum, Trifolium* spp.	Haplic Luvisols (Densic)	Podzolized brown forest soil
9	Floodplain forb meadow	M-2	*Carex* spp., *Aegopodium podagraria, Vicia cracca,*	Eutric Fluvisols (Siltic)	Typical alluvial gray-humus soil
10	Floodplain meadowsweet meadow	M-3	*Filipendula ulmaris*	Gleyic Phaeozems (Clayic)	Gleyic alluvial gray-humus soil

Data Analysis

Statistical data processing was performed in the R v.3.6.2 software. In all cases, the sampling plot (i.e., mean of 10 (or 5) measurements) was used as a statistical unit.

Two types of predictors were used for modeling: parameters measured directly on the sampling plot and Earth remote sensing data (Table 2). Soil temperature (to an accuracy of 0.1°C) and volumetric soil moisture content (to an accuracy of 0.1 vol %) were measured by sensors connected to the respirometer control unit: an Omega 88311E thermometer (OMEGA Engineering, UK) and a ThetaProbe ML2 soil moisture sensor (Delta-T devices, UK). Air temperature was measured by a thermal sensor built into the respirometer chamber.

The NDVI, LAI, and FPAR indices were determined based on 8-day moderate-resolution composite images taken by MODIS spectroradiometers aboard the *Terra* and *Aqua* satellites and provided by the VEGA-Science service [34] (Table 2). These parameters are often used to model soil respiration, vegetation production, and vegetation gas exchange [35–39].

The Standardized Precipitation Evapotranspiration Index (SPEI) was obtained from the global database (http://spei.csic.es) [41]. It shows how dry the studied period was relative to the norm over the past several decades (in this database, since 1950). Index values in the range from 0.99 to –0.99 indicate normal humidity; from 1.00 to 1.49, moderate overwetting; over 1.50, severe overwetting; from –1 to –1.49, moderate drought; and less than –1.50, severe drought.

Fig. 1. Location of studied sites:
Pine forests: (1) blueberry (Sch), (2) eagle-fern (SP), (3) forb (SDG), (4) gramineous (SZ), (5) overgrowth (SSG), (6) dead-cover (SDC), and (7) wood-sorrel (SA); meadows: (8) dry forb (M1), (9) floodplain forb (M2), and (10) floodplain meadow-sweet (M3). Map source: www.openstreetmap.org.

To subsume a predictor under the group of spatial or temporal factors, variance decomposition into components was performed in a two-way analysis of variance (VCA package [42]). A predictor is subsumed under the group of spatial factors if the biotope type makes a greater contribution to the variance; under the group of temporal factors, if the measurement date makes a greater contribution to the variance.

To implement the RF algorithm, the entire data set was split into the training set (80%) and the test set (20%) (caret package [43]). The regression was produced in the RF package [44] with the following parameters: (1) number of trees = 500; and (2) maximum number of features used for separation = 3–5. Model training was performed in four variants: (1) complete set of predictors; (2) spatial predictors only; (3) temporal predictors only; and (4) 'standard' set used in most regression models: soil temperature, air temperature, and soil moisture content. The test set was used to evaluate the model quality based on the determination coefficient (R^2) and the mean of squared errors (*MSE*).

To estimate contributions of temporal and spatial factors, the features importance was determined based on two parameters: *%IncMSE* (increase in *MSE* in the course of random permutations of each variable) and *IncNodePurity* (average increase in 'node purity' in a tree computed based on the *MSE* value determined prior to each node splitting in each tree). In most studies, only *%IncMSE* is used to estimate importance. However, this study also uses *IncNodePurity* for models with the complete set of predictors: this parameter is additive and, therefore, makes it possible to assess relative contributions of variables.

For convenience purposes, absolute importance values were expressed in percentage points. These two parameters are not identical to each other, but do not differ much (no more than a few percentage points).

The resultant models and code are available at: github.com/IASmorkalov/Respiration_RF_2022.

Table 2. List of predictors used in modeling

Predictor	Description	Annotation
T_{soil}	Soil temperature at a depth of 5 cm	Measured in close proximity to the soil respiration measurement point
T_{air}	Air temperature	
Hum	Volumetric soil moisture content at a depth of 5 cm	
Soil type	Soil type (WRB)	See Table 1
Vegetation	Vegetation type	Forest/meadow
Biotop	Biotope type	Site (see Table 1)
NDVI	Normalized difference vegetation index (usually correlates well with above-ground phytomass)	MYD09Q1 product, resolution 250 m
LAI	Leaf area index (one-sided green leaf area per unit ground surface area)	MCD15A2 product, resolution 500 m
FPAR	Fraction of absorbed photosynthetically active radiation (sunlight fraction theoretically available for photosynthesis)	
SPEI-1	Drought index computed for one month immediately preceding the measurements	Source: SPEI Global Drought Monitor, spatial resolution 0.5°
SPEI-12	Drought index computed for 12 months preceding the measurements	

RESULTS

Throughout the entire study period, soil respiration remained within the range of 0.5–9.9 μmol CO_2/m^2 s (Fig. 2a). The maximum values were observed in the summer months; the minimum values, at the beginning and after the end of the growing season. The maximum difference between mean values computed for different biotope types is 1.9 times. Soil respiration is primarily determined by temporal variability: the measurement date explains 60.9% of the variance. The spatial variability effect is manifested to a noticeably lesser extent: 10.4% of the variance. Most predictors can be subsumed under the group of temporal factors (the measurement date contribution to the variance exceeds 75%): soil temperature, air temperature, SPEI, NDVI, LAI, and FPAR (Fig. 3). Soil moisture content was subsumed under the group of spatial factors (49.7% of the explained variance), although a significant part (36.8%) of its variance was associated with the measurement time. Such factors as vegetation type, biotope type, and soil type are fully spatial; they remained unchanged over the entire research period.

The analyzed factors differ insignificantly in different biotopes, except for soil moisture content (Fig. 2): the meadowsweet meadow is the wettest one, while the dead-cover pine forest is the driest one; the difference between them reaches 3.3 times (Fig. 2d).

In all variants, the best approximation was achieved for test samples using the full set of predictors (Figs. 4–6). Temporal factors explain 76–77% of the importance for the full data set and separately for pine forests (based on *IncNodePurity*) and 91% for meadows. The importance of the standard set of predictors (soil temperature, air temperature, and soil moisture content) is 37, 26 and 52% for the full data set, pine forests, and meadows, respectively.

In different models, the order of individual predictors sorted by their importance is different. For instance, for the full data set, soil temperature is of the greatest importance (Fig. 4). Of the spatial factors, biotope type is the most important one, while vegetation type is the least important one. In models using limited sets of factors, the determination coefficient is lower; while *MSE*, is higher. The lowest approximation quality was registered in models using only temporal predictors.

In pine forests, NDVI and soil temperature are the most important factors (Fig. 5). The importance of air temperature is significantly lower in comparison with soil temperature. Even if all the predictors are used, the model produced for pine forests is less accurate than that produced for the full data set.

The best approximation was achieved for meadows (Fig. 6). Interestingly, the results using all predictors and only temporal predictors are almost identical; while the importance of air temperature is greater in comparison with soil temperature. In models produced for meadows using only spatial factors, soil moisture content is the most important factor.

In models using the standard set of predictors, soil temperature is the most important factor in all cases, while soil moisture content is the least important factor (Figs. 4d–6d).

DISCUSSION

Absolute soil respiration values obtained at the height of the growing season for pine forests (1.3–10.7 μmol CO_2/m^2 s) are close to values normally registered in forests at temperate latitudes: up to 4.3 ± 0.75 [45], 1.9–8.8 [46], 2.8–6.7 [47], 6.3–9.5 [48], 3.9–6.9 [49], 4.4–11.4 [50, and 3.5–4.4 μmol CO_2/m^2 s [51].

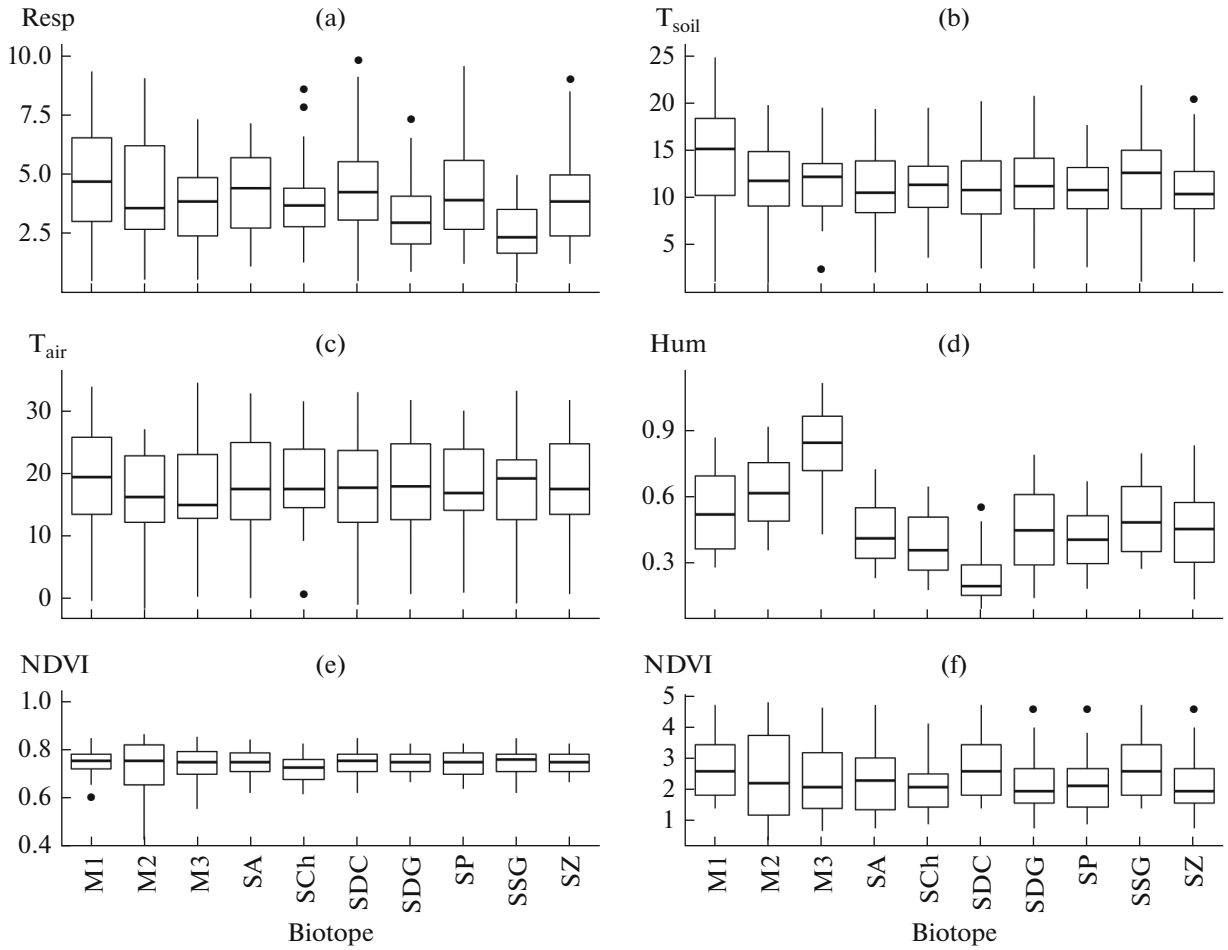

Fig. 2. Parameter values computed for the studied sites over the entire observation period: (a) soil respiration (μmol CO_2/m^2 s), (b) soil temperature (C°), (c) air temperature (C°), (d) volumetric soil moisture content (m^3/m^3), (e) NDVI, and (f) LAI. Biotope types are indicated on the horizontal axis (see Fig. 1 for codes). Horizontal lines represent median values; rectangles, 25% and 75% quartiles; bars, 1.5*interquartile ranges; and dots, outliers.

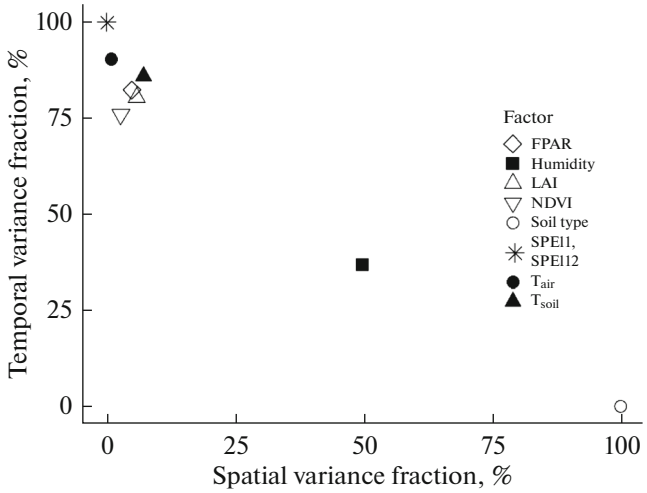

Fig. 3. Contributions of the spatial and temporal components to the variance of predictors (see Table 2 for predictor codes).

In meadows, the range of values is somewhat narrower than in pine forests (1.5–9.9 CO_2/m^2 s), but wider than ranges registered for grass ecosystems by other researchers: 2.5–8.8 [52], 3.2–5.1 [53], 3.2–5.1 [54], and 5.1–6.3 μmol CO_2/m^2 s [55]. The seasonal dynamics featuring maximum CO_2 emissions in the summer months is similar to the dynamics described for south-taiga forests [52] and north-taiga forests of Eastern Siberia [51] and Central Siberia [56].

The vegetation type affects respiration, but not always unequivocally. Analysis of the Respiration of Russian Soils database shows that in almost half of the cases, the vegetation type significantly affects the CO_2 flux magnitude [57]; for instance, soil respiration in meadow and forest coenoses is different [58]. Data collected by different authors [52, 59] indicate that for the same soil type, meadows feature a greater respiration intensity than forests; a significant effect of the soil type on respiration has been noted as well [60]. However, it was shown [61] that the growing season

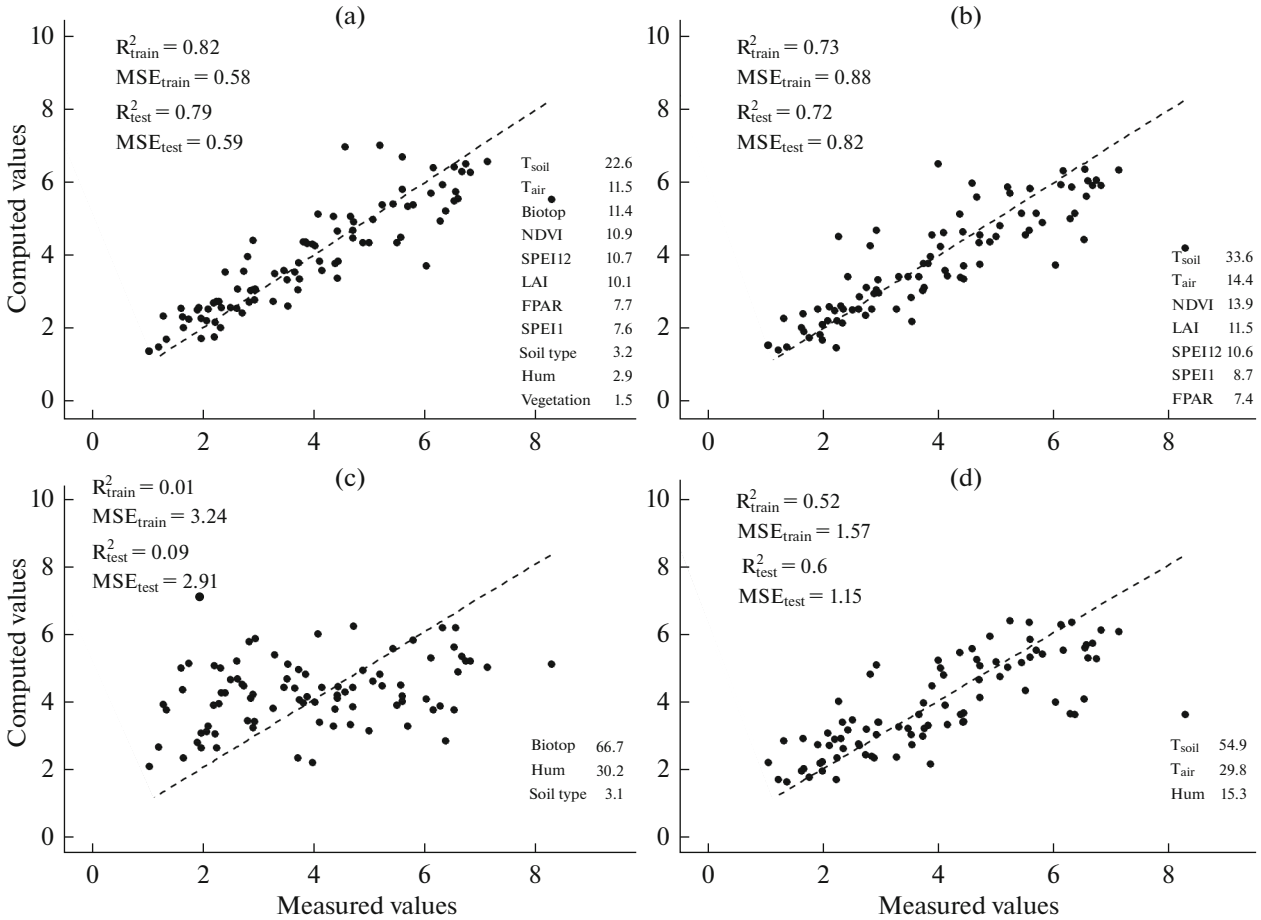

Fig. 4. RF-based approximation of soil respiration data (μmol CO_2/m^2 s) collected in all biotopes. Here and in Figs. 5–6: (a) all predictors, (b) temporal predictors, (c) spatial predictors, and (d) temperature and soil moisture content. R^2 and *MSE* values are provided for the training and test samples; importance (*%IncMSE*) values are provided for individual predictors (%). The dashed line represents the ideal forecast line.

duration also has a strong effect on the carbon dioxide flux from the soil. In this study, the effect of the biotope type reaches its maximum in models based on the full data set or produced for pine forests only. Apparently, meadow warming in early spring and cooling in autumn eliminate the difference between meadows and pine forests in average soil respiration over the entire growing season. In other words, our results do not contradict the literature data: differences between different biotopes exist, but climatic factors play the leading role.

Overall, our models demonstrate a high accuracy: R^2 reaches 0.88 (*MSE* = 0.47) for meadows and 0.79 (*MSE* = 0.59) for all biotopes. This is higher than the accuracy of RF-based models produced for mountain forests in the southern Rocky Mountains (R^2 = 0.44, *MSE* = 0.8) [25] and is comparable to models produced for sugarcane plantations (R^2 = 0.8) [62], for all forests of the northern hemisphere (R^2 up to 0.86, *MSE* = 2.16) [28], and for global respiration (R^2 = 0.89) [27].

The obtained determination coefficients are higher in comparison with most classical regression models. For instance, R^2 values obtained for soil respiration in the city of Kursk never exceed 0.6 even if a separate equation is produced for each specific location [3]. Classical regression models achieve higher accuracy levels in extreme conditions: for instance, R^2 reaches 0.9 in mountain tundra biotopes [63] and 0.8 on a transect in the mountains with a steep moisture content gradient [64]. In more complex simulation models, R^2 is 0.34 and 0.77 for dry and humid years, respectively [1]. Overall, the accuracy of models based on neural networks does not exceed the accuracy of classical regression: R^2 ranges from 0.3–0.4 on a local scale [18] to 0.6 on a global scale [19]. In other words, unlike traditional parametric regression methods and neural networks, the RF algorithm makes it possible to satisfactory interpolate data even in the absence of steep environmental gradients. Strictly speaking, R^2 is not the best efficiency criterion for comparison of different algorithms. But, unfortunately, publications

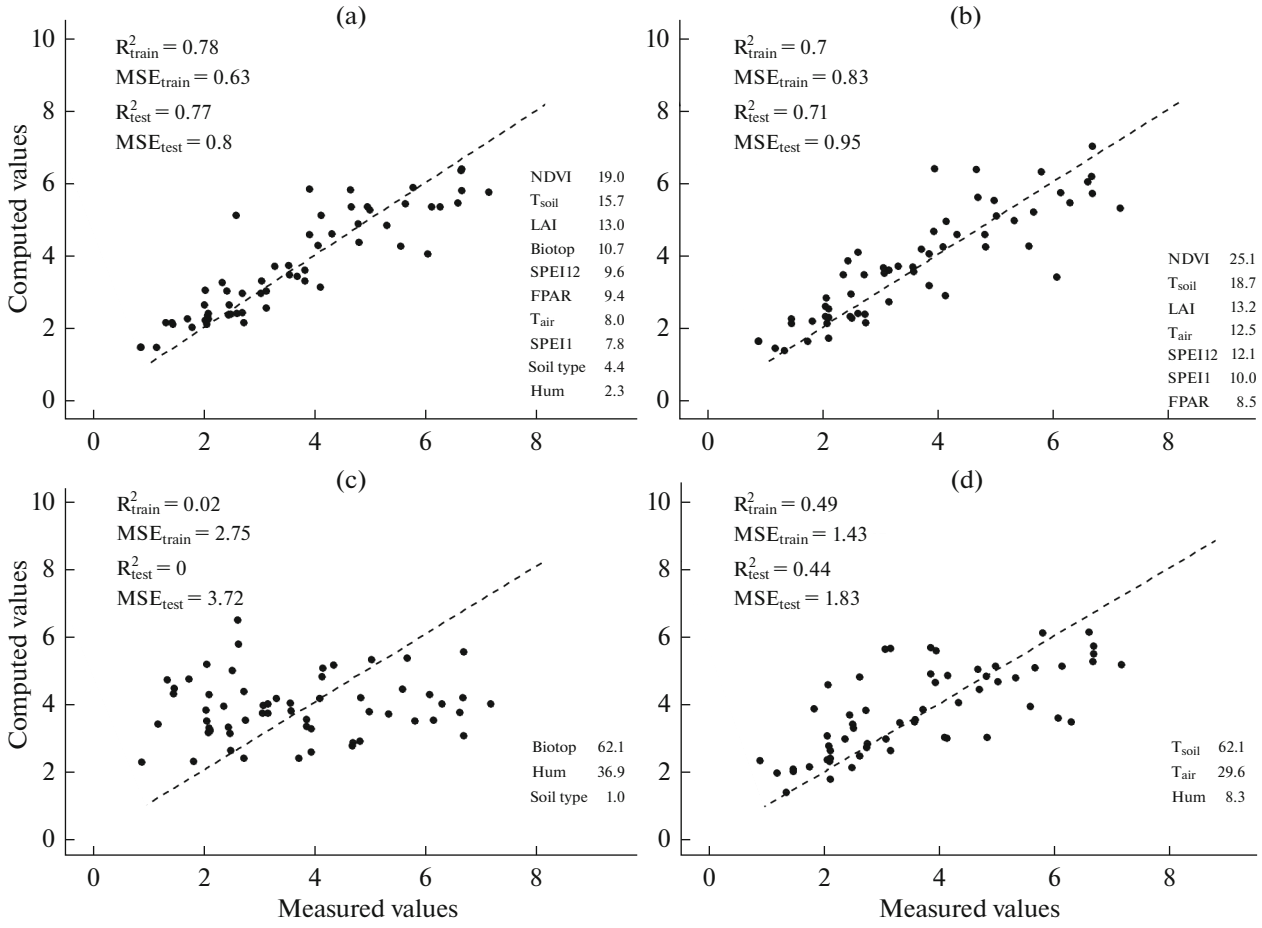

Fig. 5. RF-based approximation of the soil respiration data (μmol CO_2/m^2 s) collected in pine forests.

rarely present other techniques making it possible to assess the quality of models; therefore, this metric is used in our study.

To compare the accuracy of parametric models with the accuracy of RF, multiple regression modeling was performed for the full data set. The following model characteristics were obtained for the test sample: (1) $R^2 = 0.55$, $MSE = 1.3$ (full set of predictors); and (2) $R^2 = 0.25$, $MSE = 2.21$ (soil temperature, air temperature, and soil moisture content). In the stepwise regression implementation, only soil temperature, NDVI, and SPEI1 turned out to be significant predictors, and the model accuracy was lower in comparison with the model using all predictors ($R^2 = 0.33$, $MSE = 1.92$). Therefore, the application of the RF algorithm to the same data set made it possible to describe these data much better compared to classical regression models. In terms of efficiency, RF can be inferior to parametric regression models and even to a separate decision tree if the amount of data is small; however, if the amount of data is sufficient, then RF demonstrates better results [30].

Authors of [12] used a scheme similar to the one described in this study to divide factors into spatial and temporal ones; its criterion is as follows: mean variance computed for different dates is compared with variance between points. The results of both studies are pretty close with one exception: in the cited publication, in addition to soil and air temperature, volumetric soil moisture content is subsumed under the group of temporal factors; while in this study, it belongs to the group of spatial factors.

Results using the RF algorithm indicate that temporal variables are of greater importance, which is more typical for scales ranging from a few meters to tens of meters. Perhaps, this is because the spatial resolution of the remote sensing data is comparable to the size of the study area; as a result, the spatial spread of these parameters was not significant. Interestingly, different sites differed significantly in soil moisture content, but this did not have a strong effect on the soil respiration rate, although this parameter is traditionally considered the main determinant of emissions [32].

Spatial factors (e.g., biome type) have a significant effect on soil respiration on a regional or global scale [19, 65]. The inclusion of spatial factors into models

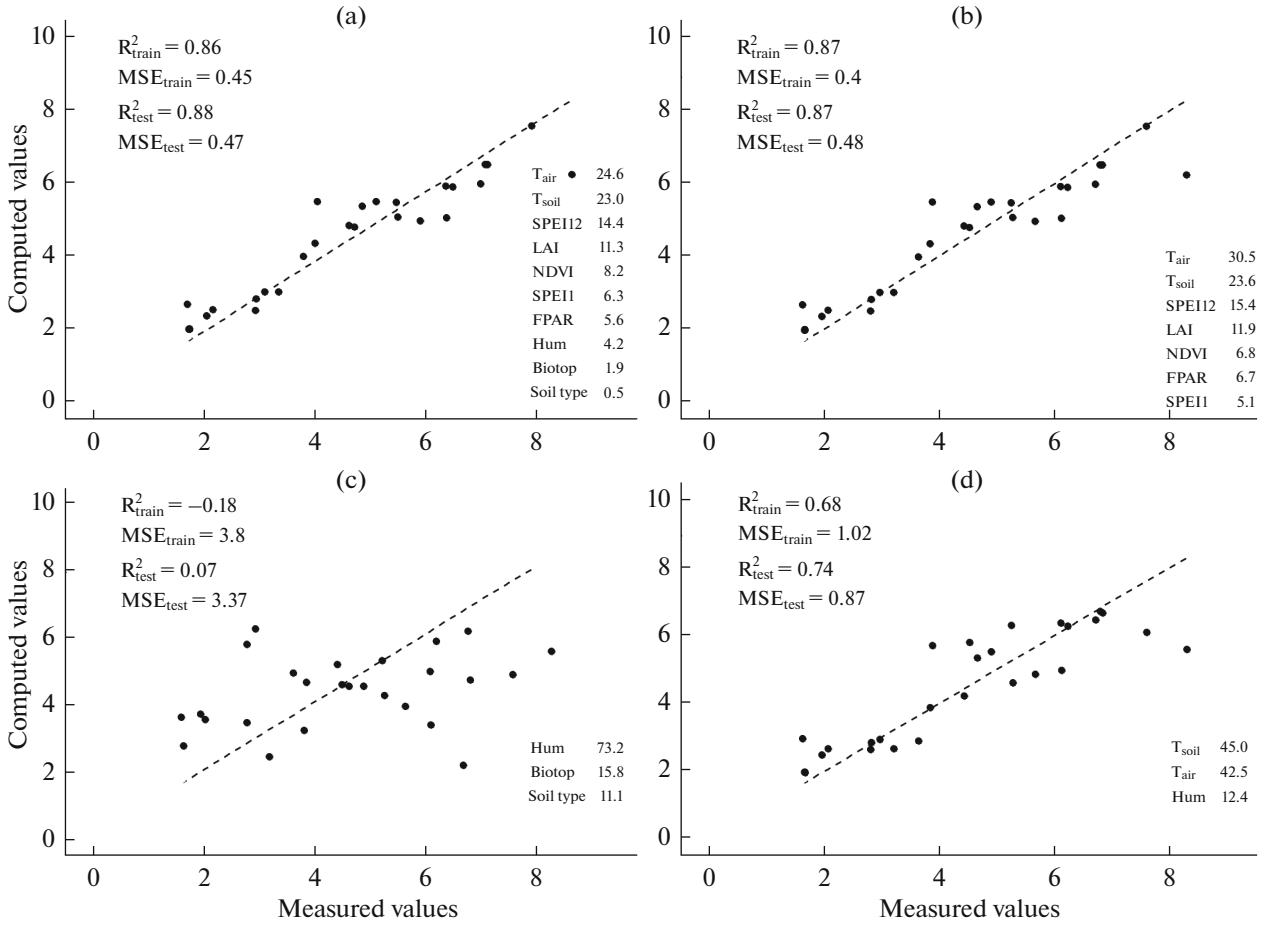

Fig. 6. RF-based approximation of the soil respiration data (μmol CO_2/m^2 s) collected in meadows.

significantly enhances their accuracy on a smaller scale as well, but only in the presence of steep environmental gradients. For instance, the inclusion of soil moisture content into models produced for a steep mountain transects on a scale of tens of meters increases R^2 from 0.5 to 0.8 [64]; while the inclusion of soil pollution data on a scale of tens of kilometers, from 0.2 to 0.74 [66].

Among the studied spatial variables, biotope type turned out to be the most important one, although this is true only for the full data set and pine forests, not for meadows. It must be emphasized that biotope type is a complex predictor that comprises all features of the studied site, both accounted and unaccounted. Therefore, models describing soil CO_2 emissions, even those produced on a scale of hundreds of meters, must take into account not only temporal factors, but also spatial ones. In meadows, the importance of all spatial factors is less than the importance of temporal ones, even if significant differences in soil moisture are taken into account. At the preliminary analysis stage, several chemical soil parameters (pH, organic carbon content, exchangeable Ca and Mg content, and mobile Fe content) were also included in the model, but this did not enhance its accuracy; therefore, the above variables were excluded from the final analysis.

Remote sensing data significantly enhance the model accuracy; a comparison with the standard set of predictors clearly demonstrates this (Figs. 4b–6b and 6d). In a similar way, they enhance models produced on a regional scale [39].

The model quality may deteriorate due to the "soil memory" effect caused by previous impacts on different temporal scales [67]. In recurrent neural networks, "short-term memory" can be implemented to reduce this effect [68]. In other modeling techniques, it is necessary to introduce variables that characterize previous states of the system. In this study, SPEI was such a variable: it characterizes wetting conditions preceding the measurements. It turned out that in terms of wetting conditions, the previous 12 months are more important than the month immediately preceding the measurement. This is somewhat consistent with the discovered effect exercised by the total precipitation amount over the spring–summer period on the total annual CO_2 emission [69].

Overall, the obtained data on the importance of individual parameters are in line with expectations: temperature plays the leading role, which has been repeatedly demonstrated earlier [32]. The inclusion of spatial predictors enhances the model quality, especially in forests; this confirms the need to take into account not only temporal, but also spatial components of soil respiration variability [12]. The high importance of NDVI indicates that models describing carbon dioxide fluxes from the soil must incorporate vegetation parameters as well. It must be noted that the above conclusions apply only to the growing season since no studies were conducted in winter.

CONCLUSIONS

Despite the significant biotopic diversity comprising two groups of biotopes (forests and meadows), the rate of CO_2 emissions from the soil primarily depends on temporal factors. For meadows, the model accuracy is identical in variants using temporal factors only and in variants using the full set of temporal and spatial factors despite significant differences in wetting conditions. By contrast, in pine forests, one of the spatial factors (biotope type) is of great importance. This indicates that soil respiration models, even those produced on a scale of hundreds of meters, must take into account not only temporal factors, but also spatial ones.

The feature importance assessment performed for individual variables using the RF algorithm makes it possible to interpret the simulation results in a meaningful way. The possibility to simultaneously use variables of different types, including highly correlated ones, makes the RF algorithm more flexible in comparison with classical regression methods. But this is not the only reason to recommend the RF algorithm for wide use in models describing carbon dioxide fluxes from soils: its undeniable capability to assess contributions of various factors to the soil respiration variability gives a better insight into underlying carbon cycle mechanisms in terrestrial ecosystems.

ACKNOWLEDGMENTS

The author is grateful to D.A. Bedin for consulting on machine learning, to E.L. Vorobeichik for useful comments and assistance in manuscript improvement, to S.Yu. Kaigorodov and T.Yu. Gabershtein for soil diagnostics and analysis, and to N.O. Sadykova for discussing the text and commenting on it.

FUNDING

The field studies were supported by the Ural Branch, Russian Academy of Sciences, project no. 12-P-4-1057; the data analysis and preparation of this article, by the Ministry of Science and Higher Education of the Russian Federation, projects no. 122021000076-9 and FEUZ-2021-0014, respectively.

REFERENCES

1. Priputina, I.V., Bykhovets, S.S., Frolov, P.V., et al., Application of mathematical models ROMUL and Romul_Hum for estimating CO_2 emission and dynamics of organic matter in albic luvisol under deciduous forest in the south of Moscow oblast, *Eurasian Soil Sci.*, 2020, no. 53, pp. 1480–1491.
2. Reiners, W.A., Carbon dioxide evolution from the floor of three minnesota forests, *Ecology*, 1968, vol. 49, no. 3, pp. 471–483. https://doi.org/10.2307/1934114
3. Nevedrov, N.P., Sarzhanov, D.A., Protsenko, E.P., and Vasenev, I.I., Seasonal dynamics of CO_2 emission from soils of Kursk, *Eurasian Soil Sci.*, 2021, vol. 54, pp. 80–88.
4. Chimner, R.A., Soil respiration rates of tropical peatlands in Micronesia and Hawaii, *Wetlands*, 2004, vol. 24, no. 1, art. ID 51. https://doi.org/10.1672/0277-5212(2004)024[0051:SRROTP]2.0.CO;2
5. Jia, B., Zhou, G., Wang, F., and Wang, Y., A comparative study on soil respiration between grazing and fenced typical Leymus chinensis steppe, Inner Mongolia, *Chin. J. Appl. Ecol.*, 2004, vol. 15, no. 9, pp. 1611–1615.
6. Orchard, V.A. and Cook, F.J., Relationship between soil respiration and soil moisture, *Soil Biol. Biochem.*, 1983, vol. 15, no. 4, pp. 447–453. https://doi.org/10.1016/0038-0717(83)90010-X
7. Rubio, V.E. and Detto, M., Spatiotemporal variability of soil respiration in a seasonal tropical forest, *Ecol. Evol.*, 2017, vol. 7, no. 17, pp. 7104–7116. https://doi.org/10.1002/ece3.3267
8. Webster, K.L., Creed, I.F., Skowronski, M.D., and Kaheil, Y.H., Comparison of the performance of statistical models that predict soil respiration from forests, *Soil Sci. Soc. Am. J.*, 2009, vol. 73, no, 4. pp. 1157–1167. https://doi.org/10.2136/sssaj2008.0310
9. Sotta, E.D., Meir, P., Malhi, Y., Donato, A., et al., Soil CO_2 efflux in a tropical forest in the central Amazon, *Global Change Biol.*, 2004, vol. 10, no. 5, pp. 601–617. https://doi.org/10.1111/j.1529-8817.2003.00761.x
10. Tufekcioglu, A., Raich, J.W., Isenhart, T.M., and Schultz, R.C., Soil respiration within riparian buffers and adjacent crop fields, *Plant Soil*, 2001, vol. 229, no. 1, pp. 117–124. https://doi.org/10.1023/A:1004818422908
11. Lloyd, J. and Taylor, J.A., On the temperature dependence of soil respiration, *Funct. Ecol.*, 1994, vol. 8, no. 3, pp. 315–323. https://doi.org/10.2307/2389824
12. Karelin, D.V., Pochikalov, A.V., Zamolodchikov, D.G., and Gitarskii, M.L., Factors of spatiotemporal variability of CO_2 fluxes from soils of southern taiga spruce forests of Valdai, *Lesovedenie*, 2014, no. 4, pp. 56–66.

13. Karelin, D.V., Zamolodchikov, D.G., and Isaev, A.S., Little-known pulse components of soil carbon dioxide emission in taiga forests, *Dokl. Ross. Akad. Nauk*, 2017, vol. 75, no. 4, pp. 473–476.

14. Zeng, X.H., Song Y.G., Zeng, C.M., et al., Partitioning soil respiration in two typical forests in semi-arid regions, North China, *Catena*, 2016, vol. 147, pp. 536–544. https://doi.org/10.1016/j.catena.2016.08.009

15. Kuzyakov, Y. and Gavrichkova, O., Time lag between photosynthesis and carbon dioxide efflux from soil: a review of mechanisms and controls, *Global Change Biol.*, 2010, vol. 16, no. 12, pp. 3386–3406. https://doi.org/10.1029/2005JG000063

16. Richardson, A.D., Braswell, B.H., Hollinger, D.Y., et al., Comparing simple respiration models for eddy flux and dynamic chamber data, *Agric. For. Meteorol.*, 2006, vol. 141, nos. 2–4, pp. 219–234. https://doi.org/10.1016/j.agrformet.2006.10.010

17. Farhate, C.V.V., Souza, Z.M., Oliveira, S.R.M., et al., Use of data mining techniques to classify soil CO_2 emission induced by crop management in sugarcane field, *PLoS ONE*, 2018, vol. 13, no. 3, art. ID e0193537. https://doi.org/10.1371/journal.pone.0193537

18. Jarvi, L., Nordbo, A., Junninen, H., et al., Seasonal and annual variation of carbon dioxide surface fluxes in Helsinki, Finland, in 2006–2010, *Atmos. Chem. Phys.*, 2012, vol. 12, no. 18, pp. 8475–8489. https://doi.org/10.5194/acp-12-8475-2012

19. Zhao, Z.Y., Peng, C.H., Yang Q., et al., Model prediction of biome-specific global soil respiration from 1960 to 2012, *Earth's Future*, 2017, vol. 5, no. 7, pp. 715–729. https://doi.org/10.1002/2016ef000480

20. Song, X.Z., Peng, C.H., Zhao, Z.Y., et al., Quantification of soil respiration in forest ecosystems across China, *Atmos. Environ.*, 2014, vol. 94, pp. 546–551. https://doi.org/10.1016/j.atmosenv.2014.05.071

21. Rudin, C., Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, *Nat. Mach. Intell.*, 2019, vol. 1, no. 5, pp. 206–215. https://doi.org/10.1038/s42256-019-0048-x

22. Chistyakov, S.P., Random forests: an overview, *Tr. Karel. Nauchn. Tsentra Ross. Akad. Nauk*, 2013, no. 1, pp. 117–136.

23. Hastie, T., Tibshirani, R., and Friedman, J., Random Forests, in *the Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New-York: Springer-Verlag, 2009, pp. 587–604.

24. Tang, X.L., Fan, S.H., Du, M.Y., et al., Spatial and temporal patterns of global soil heterotrophic respiration in terrestrial ecosystems, *Earth Syst. Sci. Data*, 2020, vol. 12, no. 2, pp. 1037–1051. https://doi.org/10.5194/essd-12-1037-2020

25. Berryman, E.M., Vanderhoof, M.K., Bradford, J.B., et al., Estimating soil respiration in a subalpine landscape using point, terrain, climate, and greenness data, *J. Geophys. Res.: Biogeosci.*, 2018, vol. 123, no. 10, pp. 3231–3249. https://doi.org/10.1029/2018jg004613

26. Jian, J.S., Steele, M.K., Thomas, R.Q., et al., Constraining estimates of global soil respiration by quantifying sources of variability, *GCB Bioenergy*, 2018, vol. 24, no. 9, pp. 4143–4159. https://doi.org/10.1111/gcb.14301

27. Lu, H.B., Li, S.H., Ma, M.N., et al., Comparing machine learning-derived global estimates of soil respiration and its components with those from terrestrial ecosystem models, *Environ. Res. Lett.*, 2021, vol. 16, no. 5, art. ID 054048. https://doi.org/10.1088/1748-9326/abf526

28. Shiri, N., Shiri, J., Kazemi, M.H., and Xu, T.R., Estimation of CO_2 flux components over northern hemisphere forest ecosystems by using random forest method through temporal and spatial data scanning procedures, *Environ. Sci. Pollut. Res.*, 2022, vol. 29, no. 11, pp. 16123–16137. https://doi.org/10.1007/s11356-021-16501-x

29. Warner, D.L., Bond-Lamberty, B., Jian, J., et al., Spatial predictions and associated uncertainty of annual soil respiration at the global scale, *Global Biogeochem. Cycles*, 2019, vol. 33, no. 12, pp. 1733–1745. https://doi.org/10.1029/2019gb006264

30. Mukhortova, L., Schepaschenko, D., Moltchanova, E., et al., Respiration of Russian soils: Climatic drivers and response to climate change, *Sci. Total Environ.*, 2021, vol. 785, art. ID 147314. https://doi.org/10.1016/j.scitotenv.2021.147314

31. Jian, J.S., Frissell, M., Hao, D.L., et al., The global contribution of roots to total soil respiration, *Global Ecol. Biogeography*, 2022, vol. 31, no. 4, pp. 685–699. https://doi.org/10.1111/geb.13454

32. Luo, Y. and Zhou, X., *Soil Respiration and the Environment*, Burlington: Acad. Press, 2006.

33. Smorkalov, I.A., The ratio of spatial and temporal variability of soil respiration, *Materialy Vserossiiskoi konferentsii molodykh uchenykh "Ekologiya: fakty, gipotezy, modeli"* (Proc. Int. All-Russian Conf. Young Scientists), Ekaterinburg, 2016, pp. 120–123.

34. Lupyan, E.A., Savin, I.Yu., Bartalev, S.A., et al., Satellite service for vegetationmonitoring Vega, *Sovrem. Probl. Distantsionnogo Zondirovaniya Zemli Kosmosa*, 2011, vol. 8, no. 1, pp. 190–198.

35. Wu, J.S. and Fu, G., Modelling aboveground biomass using MODIS FPAR/LAI data in alpine grasslands of the Northern Tibetan Plateau, *Remote Sens. Lett.*, 2018, vol. 9, no. 2, pp. 150–159. https://doi.org/10.1080/2150704x.2017.1407045

36. Liu, S., Cheng, F., Dong, S., et al., Spatiotemporal dynamics of grassland aboveground biomass on the Qinghai-Tibet Plateau based on validated MODIS NDVI, *Sci. Rep.*, 2017, vol. 7, no. 1, art. ID 4182. https://doi.org/10.1038/s41598-017-04038-4

37. Wang, X.F., Wang, H.B., Li, X., and Ran, Y.H., Photosynthesis (NPP, NEP, Respiration), *Observation and Measurement*, Li, X. and Vereecken, H., Eds., Heidelberg, 2017, pp. 1–30.

38. Reynolds, L.L., Johnson, B.R., Pfeifer-Meister, L., and Bridgham, S.D., Soil respiration response to climate change in Pacific Northwest prairies is mediated by a regional Mediterranean climate gradient, *GCB Bioenergy*, 2015, vol. 21, no. 1, pp. 487–500. https://doi.org/10.1111/gcb.12732
39. Ding, S.C., Yao, X.W., Wang, J., et al., Relationships between soil respiration and hyperspectral vegetation indexes and crop characteristics under different warming and straw application modes, Environmental Science and Pollution Research, 2021, vol. 28, no. 30, pp. 40756–40770. https://doi.org/10.1007/s11356-021-13612-3
40. Vicente-Serrano, S.M., Beguería, S., and López-Moreno, J.I., A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index, *J. Clim.*, 2010, vol. 23, no. 7, pp. 1696–1718. https://doi.org/10.1175/2009jcli2909.1
41. Beguería, S., Vicente-Serrano, S.M., Reig, F., and Latorre, B., Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring, *Int. J. Climatol.*, 2014, vol. 34, no. 10, pp. 3001–3023. https://doi.org/10.1002/joc.3887
42. VCA: Variance Component Analysis. R package version 1.4.2. https://CRAN.R-project.org/package=VCA.
43. Caret: Classification and Regression Training version 6.0-86. https://cran.r-project.org/package=caret.
44. RandomForest: Breiman and Cutler's Random Forests for classification and regression version 4.6-14. https://cran.r-project.org/package=randomForest.
45. Osipov, A.F., Effect of interannual difference in weather conditions of the growing season on the Co_2 emission from the soil surface in the middle-taiga cowberry–lichen pine forest (Komi Republic), *Eurasian Soil Sci.*, 2018, vol. 51, no. 12, pp. 1419–1426.
46. Glukhova, T.V., Vomperskii, S.E., and Kovalev, A.G., Emission of CO_2 from the surface of oligotrophic bogs with due account for their microrelief in the southern taiga of European Russia, *Pochvovedenie*, 2014, no. 1, pp. 48–57.
47. Kadulin, M.S. and Koptsik, G.N., Emission of CO2 by soils in the impact zone of the Severonikel smelter in the Kola subarctic region, *Eurasian Soil Sci.*, 2013, vol. 46, pp. 1107–1116. https://doi.org/10.7868/S0032180X13110063
48. Mashika, A.V., Carbon dioxide emission from the surface of podzolic soils, *Eurasian Soil Sci.*, 2006, vol. 39, pp. 1312–1317.
49. Smorkalov, I.A. and Vorobeichik, E.L., Soil respiration of forest ecosystems in gradients of environmental pollution by emissions from copper smelters, *Russ. J. Ecol.*, 2011, vol. 42, art. ID 464.
50. Kozlov, M.V., Zvereva, E.L., and Zverev, V.E., *Impacts of Point Polluters on Terrestrial Biota: Comparative Analysis of 18 Contaminated Areas*, Dordrecht: Springer-Verlag, 2009.
51. Takakai, F., Desyatkin, A.R., Lopez, C.M.L., et al., Influence of forest disturbance on CO_2, CH_4 and N_2O fluxes from larch forest soil in the permafrost taiga region of eastern Siberia, *Soil Sci. Plant Nutr.*, 2008, vol. 54, no. 6, pp. 938–949.
52. Kurganova, I.N., Lopes de Gerenyu, V.O., Rozanova, L.N., and Kudeyarov, V.N., Annual and seasonal CO2 fluxes from Russian southern taiga soils, *Tellus B.*, 2001, vol. 55, pp. 338–344.
53. Wang, M., Liu, X.T., Li, X.J., et al., Soil respiration dynamics and its controlling factors of typical vegetation communities on meadow steppes in the western Songnen Plain, *Chinese J. Appl. Ecol.*, 2014. vol. 25, no. 1, pp. 45–52. https://doi.org/10.1029/2009GB003538
54. Bahn, M., Rodeghiero, M., Anderson-Dunn, M., et al., Soil respiration in European grasslands in relation to climate and assimilate supply, *Ecosystems*, 2008, vol. 11, no. 8, pp. 1352–1367. https://doi.org/101029/2005JG000063
55. Perez-Quezada, J.F., Saliendra, N.Z., Akshalov, K., et al., Land use influences carbon fluxes in northern Kazakhstan, *Rangeland Ecol. Manage.*, 2010, vol. 63, no. 1, pp. 82–93.
56. Shibistova, O., Lloyd, J., Zrazhevskaya, G., et al., Annual ecosystem respiration budget for a *Pinus sylvestris* stand in central Siberia, *Tellus, Ser. B: Chem. Phys. Meteorol.*, 2002, vol. 54, no. 5, pp. 568–589.
57. Kurganova, I.N. and Kudeyarov, V.N., The assessment of carbon dioxide fluxes from Russian south taiga soils, *Eurasian Soil Sci.*, 1998, vol. 31, pp. 954–965.
58. Kudeyarov, V.N. and Kurganova, I.N. Respiration of Russian Soils: Database Analysis, Long-Term Monitoring, and General Estimates, *Eurasian Soil Sci.*, 2005, vol. 38, no. 9, pp. 983–992.
59. Raich, J.W. and Tufekcioglu, A., Vegetation and soil respiration: Correlations and controls, *Biogeochemistry*, 2000, vol. 48, no. 1, pp. 71–90.
60. Kurganova, I.N., Lopes de Gerenyu, V.O., Khoroshaev, D.A., et al., Analysis of the long-term soil respiration dynamics in the forest and meadow cenoses of the prioksko-terrasny biosphere reserve in the perspective of current climate trends, *Eurasian Soil Sci.*, 2020, vol. 53, pp. 1421–1436.
61. Kudeyarov, V.N., Khakimov, F.I., Deeva, N.F., et al., Assessment of Russian soils respiration, *Pochvovedenie*, 1995, no. 1, pp. 33–42.
62. Tavares, R.L.M., Oliveira, S.R.D., de Barros, F.M.M., et al., Prediction of soil CO_2 flux in sugarcane management systems using the Random Forest approach, *Sci. Agric.*, 2018, vol. 75, no. 4, pp. 281–287. https://doi.org/10.1590/1678-992x-2017-0095
63. Magnani, M., Baneschi, I., Giamberini, M., et al., Drivers of carbon fluxes in Alpine tundra: a comparison of three empirical model approaches, *Sci. Total Environ.*, 2020, vol. 732, art. ID 139139. https://doi.org/10.1016/j.scitotenv.2020.139139
64. Lecki, N.A. and Creed, I.F., Forest soil CO_2 efflux models improved by incorporating topographic con-

trols on carbon content and sorption capacity of soils, *Biogeochemistry*, 2016, vol. 129, no. 3, pp. 307–323. https://doi.org/10.1007/s10533-016-0233-5

65. Huang, N., Wang, L., Song, X.P., et al., Spatial and temporal variations in global soil respiration and their relationships with climate and land cover, *Sci. Adv.*, 2020, vol. 6, no. 41, art. ID eabb8508. https://doi.org/10.1126/sciadv.abb8508
66. Smorkalov, I.A. and Bedin, D.A., Accounting for industrial pollution in soil respiration models, *Materialy 6-i natsional'noi nauchnoi konferentsii s mezhdunarodnym uchastiem "Matematicheskoe modelirovanie v ekologii"* (Proc. 6th National Sci. Conf. with International Participation "Mathematical modeling in ecology"), Pushchino, 2019, pp. 191–193.
67. Parolari, A.J., Sizemore, J., and Katul, G.G., Multiscale legacy responses of soil gas concentrations to soil moisture and temperature fluctuations, *J. Geophys. Res.: Biogeosci.*, 2021, vol. 126, no. 2, art. ID e2020-JG005865. https://doi.org/10.1029/2020jg005865
68. Besnard, S., Carvalhais, N., Arain, M.A., et al., Memory effects of climate and vegetation affecting net ecosystem CO_2 fluxes in global forests, *PLoS ONE*, 2019, vol. 14, no. 2, art. ID e0211510. https://doi.org/10.1371/journal.pone.0211510
69. Kurganova, I.N., Lopes de Gerenyu, V.O., Myakshina, T.N., et al., CO_2 emission from soils of various ecosystems of the Southern Taiga Zone: Data analysis of continuous 12-year monitoring, *Dokl. Biol. Sci.*, 2011, vol. 436, no. 6, pp. 56–58.

Translated by L. Emeliyanov