DOI: http://dx.doi.org/10.1590/1678-992X-2017-0095

Scientia Agricola

ISSN 1678-992X

Prediction of soil CO_2 flux in sugarcane management systems using the Random Forest approach

Rose Luiza Moraes Tavares[1]*, Stanley Robson de Medeiros Oliveira[2], Flávio Margarito Martins de Barros[3], Camila Viana Vieira Farhate[3], Zigomar Menezes de Souza[3], Newton La Scala Junior[4]

[1]University of Rio Verde/ESUCARV, C.P. 104 – 75901-970 – Rio Verde, GO – Brazil.
[2]Embrapa Agricultural Informatics – Artificial Intelligence Lab., Av. André Tosello, 209 – 13083-886 – Campinas, SP – Brazil.
[3]University of Campinas/FEAGRI, Av. Cândido Rondon, 501 – 13083-875 – Campinas, SP – Brazil.
[4]São Paulo State University/FCAV – Dept. of Exact Sciences, Via de acesso Prof. Paulo Donato Castellane, s/n – 14884-900 – Jaboticabal, SP – Brazil.
*Corresponding author <roseluiza@unirv.edu.br>

Edited by: Eduardo Alvarez Santos

Received December 05, 2016
Accepted May 17, 2017

ABSTRACT: The Random Forest algorithm is a data mining technique used for classifying attributes in order of importance to explain the variation in an attribute-target, as soil CO_2 flux. This study aimed to identify prediction of soil CO_2 flux variables in management systems of sugarcane through the machine-learning algorithm called Random Forest. Two different management areas of sugarcane in the state of São Paulo, Brazil, were selected: burned and green. In each area, we assembled a sampling grid with 81 georeferenced points to assess soil CO_2 flux through automated portable soil gas chamber with measuring spectroscopy in the infrared during the dry season of 2011 and the rainy season of 2012. In addition, we sampled the soil to evaluate physical, chemical, and microbiological attributes. For data interpretation, we used the Random Forest algorithm, based on the combination of predicted decision trees (machine learning algorithms) in which every tree depends on the values of a random vector sampled independently with the same distribution to all the trees of the forest. The results indicated that clay content in the soil was the most important attribute to explain the CO_2 flux in the areas studied during the evaluated period. The use of the Random Forest algorithm originated a model with a good fit (R^2 = 0.80) for predicted and observed values.

Keywords: *Saccharum officinarum*, soil respiration, green sugarcane, clay

Introduction

The management conversion process from burned sugarcane to green sugarcane has been studied on several aspects, such as the increase of greenhouse gases emissions (Panosso et al., 2009), greenhouse gas balance (Figueiredo and La Scala Jr., 2011), productivity (Castro et al., 2014), soil quality (Sant'Anna et al., 2009). However, little focus has been given to CO_2 emissions related to soil quality, more specifically to carbon stock and microbial activity. The analysis of CO_2 emission by classical statistics is not always enough to understand how this process occurs in the soil as well as interactions with other soil attributes. In this sense, the Random Forest algorithm is a data mining technique capable of classifying attributes observed in order of importance to explain the variation in an attribute-target, soil CO_2 flux in this study.

This algorithm was proposed by Breiman (2001) and its main advantages are: (a) non-parametric nature; (b) one of the most accurate learning algorithms available; (c) handling thousands of input variables without variable deletion; (d) estimating important variables in classification; (e) effective method to estimate missing data with accuracy when a large proportion of data is missing; (f) robustness in the presence of noise and unimportant variables; (g) high flexibility to perform several types of data analyses, namely regression, classification, and unsupervised learning (Rodriguez-Galiano et al., 2012). Therefore, the Random Forest analysis can be used in a wide range of fields, including the study on soil attributes that shows a variability or anisotropic trend. On the other hand, the Random Forest algorithm overfits for some datasets with noisy classification/regression tasks.

The use of the Random Forest classifier in CO_2 flux prediction was effective to predict new cases. Grimma et al. (2008) worked with digital soil mapping to predict spatial distribution of soil organic carbon and concluded that the Random Forest-based digital approach to map soil organic carbon (SOC) provided SOC estimations of high spatial resolution with estimated error and predictor of importance. Rodriguez-Galiano et al. (2012) evaluated the performance of the Random Forest classifier for land cover classification of a heterogeneous area and observed a good performance of this algorithm in the context of classifications. It was higher than the standard classification as a simple decision tree, for allowing a greater differentiation between the different categories of the study area. Moreover, the authors reported that the variables identified by the classifier as the most important corresponded to expectations. This study sought to identify prediction properties of soil CO_2 flux in sugarcane management systems using the Random Forest classification model.

Materials and Methods

Study site

The study was conducted in sugarcane areas located in northern São Paulo State, Brazil, near the coordinates latitude: 21°19′8″ South and longitude 48°7′24″ West, and approximately 500 m above sea level. The climate in the region is classified as $B_2rB'4a'$, which means humid with small water, according to the Thornthwaite system (Rolim et al., 2007) and topography of areas ranges from flat to undulating.

Description and history of experimental sites

The areas evaluated were managed under burned sugarcane (BS) and green sugarcane (GS) harvesting systems. GS has a history of 10 years after converting from the burning to the green system of sugarcane cropping. The soil of the areas was classified as Haplustox, clay texture.

The BS area was managed under the burning system since the 1980s and during the evaluation period of the experiment, between 2011 and 2012, the sugarcane variety planted was at the 5th ratoon with an average productivity of 67 t ha^{-1}. On the other hand, the GS area began without the burning system and with mechanized harvesting in early 2001. In 2011, it was at the 5th sugarcane ratoon with an average productivity of 75 t ha^{-1}. In the sugarcane plantation reform, 2 t ha^{-1} of dolomitic limestone and 480 kg ha^{-1} of NPK were applied. Over the years (with the exception of the 2011-2012 period), 100 m^{-3} ha^{-1} of vinasse and 300 kg ha^{-1} of area were applied on average in the areas.

In each area, a grid containing 81 sampling points was installed in an area of 1 ha, whose points were georeferenced with the support of a total station and DGPS (L1/L2 Hiper Lite Plus).

Soil properties observed

The soil penetration resistance test (PR) and soil sampling for analyses were performed at the sampling grid points. For the PR test (Stolf, 1991), an impact penetrometer (model IAA/Planalsucar) was used, with a 30° cone angle.

The response variable of this study is soil CO_2 flux (FCO_2), which was evaluated in the field in 81 points of each sugarcane management area, simultaneously, for 10 d during the dry period (23 Aug 2011 – 21 Sept 2011) and rainy period (26 June 2012, 04 July 2012) in the morning (7-11 a.m.). The sampling in the grid was previously planned to start and finish the CO_2 evaluated by the same orientation in the grid of both areas aiming to standardize the sampling time and save chamber battery power.

The CO_2 evaluation was performed using soil chamber. The system detects changes in CO_2 concentration within the chamber by spectroscopy in the infrared region (IRGA - Infrared Gas Analyzer). The equipment is a closed system with internal volume of 991 cm^3, with soil contact area of 71.6 cm^2 and placed on PVC pipes previously inserted into the soil at 3 cm deep and in the green cane area, the PVC was installed after removing cane wastes. More details on the sampling protocol for the chamber fluxes can be found in Tavares et al. (2016b).

Soil Temperature (St) and moisture (Sm) were evaluated simultaneously with the measurement of CO_2 concentration by a temperature sensor attached to the soil chamber system. For water content in the soil, we used the portable Hydrosense system on the soil layer 0-20 cm.

The variable microbial biomass carbon (MBC) was evaluated in the same period and consisted of collecting deformed soil samples at 0-20 cm layer, which were stored at 4 °C after collection to maintain the moisture until the analysis of MBC proposed by Jenkinson and Powlson (1976). Other deformed samples were collected in a single evaluation period, and used for the analysis of organic carbon - C (Nelson and Sommers, 1982), pH, P, K, S (Raij et al., 2001) and mean diameter weight – MDW (Kemper and Chepil, 1965). Undeformed samples were also collected with the aid of volumetric ring volume of 100 cm^3 for soil density analysis (Sd), macroporosity (Macro), and microporosity (Micro), according to the manual of physical analysis (Embrapa, 2011).

Statistical modeling

Part of the data of this study was used in the geostatistical analyses aiming to study CO_2 spatial and temporal variability (Tavares et al., 2016a). The principal component analysis (PCA) was used to identify the factors that explain the variance observed (Tavares et al., 2015; 2016b). Other analytical tools were used to find approaches that were not previously detected, which can elucidate the process under study.

For data interpretation, descriptive statistics (mean and standard deviation) was used considering that for the CO_2 flux, mean daily measurements were obtained during the dry period of 2011 and the wet period of 2012. The mean differences of attributes between the two management systems were tested by the t test ($p < 0.05$), using the SAS program, v.9.2.

Some soil properties were selected to explain soil CO_2 flux from a correlation matrix between all data in which we excluded variables with null or constant variance, or variables strongly correlated to each other and that did not contribute with information to the statistical model (Grimma et al., 2008; Rodriguez-Galiano et al., 2012). The algorithm used for elaborating the predictive model of soil CO_2 flux was the Random Forest, which is a classification and regression technique developed by Breiman (2001). It consists of a set of combined decision trees to solve classification problems. Each decision tree is built using a random initial data sampling and, at each division of these data, a random subset of m attributes is used to choose the most informative attributes. Random Forest generates a list of the most important attributes in forest development, which are determined by the accumulated importance of the attribute at node divisions of each forest tree (Hastie et al., 2009). The main steps of the Random Forest algorithm are shown in Figure 1.

A decision tree is a graphic model represented by nodes and branches in which intermediate or decision nodes represent the attribute tests (independent variables), while the branches represent the results of these tests. The node located at the treetop represents its beginning and is called the root node. The external node, on the other hand, which does not have a descendant node, located at the lower end, is called the leaf or terminal node and represents the prediction value for the target attribute or class (Han et al., 2011).

To elaborate the classification model using the Random Forest algorithm, we selected 17 soil variables or properties (65 % of the total observed variables) from a correlation matrix in which we excluded those with null or constant variance, or those that show strong correlation among each other, represented by dark circles in Figure 2. These properties were removed because they do not provide relevant information to generate the statistical model. Thus, the properties selected were: Clay, MDW, S, PR, P, MBC, K, Cu, Sd, Macro, C, pH, $CaCl_2$, Micro, St, management system, Sm, and precipitation.

For model validation, we split the data in training and test sets, where ¾ of data was used for training and ¼ for testing. In the training part, we used cross-validation with 10 folds, and this method was used to adjust hiperparameters of the model. After fitting the model, the R^2 and RMSE measures were evaluated in the test set, which are standard metrics used in model evaluation of machine learning algorithms with numeric outcomes (Hastie et al., 2009).

Results

Temporal variability was not significant for the GS area, according to Tavares et al. (2016a); therefore, the soil CO_2 flux mean was combined both periods evaluated (winter/2011 and summer/2012). Thus, the soil CO_2 flux was higher ($p < 0.05$) in the GS management system (2.68 µmol m^{-2} s^{-1}) compared to the BS management system (1.53 µmol m^{-2} s^{-1}) (Table 1).

More than 70 % of selected properties presented order of importance of 80-100 % in the Random Forest model range to explain the soil CO_2 flux of sugarcane management systems. Soil temperature and the effect of

Given a data set $X = x_1, x_2, \ldots, x_j$ and $Y = y_1, y_2, \ldots, y_k$.

For b = 1, 2, 3,…B, repeat:

a. Create a bootstrap sample (X_b, Y_b) with n examples of (X, Y).
b. Adjust an f^b decision tree for the training set (X_b, Y_b), using m attributes for the choice of each node.

End of repetition.

Generate the final model: $\hat{f}(x) = \sum_{b=1}^{B} f^b(x)$, which calculates the votes obtained by the f^b models, resulting in a final classification according to the majority vote.

Figure 1 – Sketch of the Random Forest algorithm. Source: Breinman (2001).

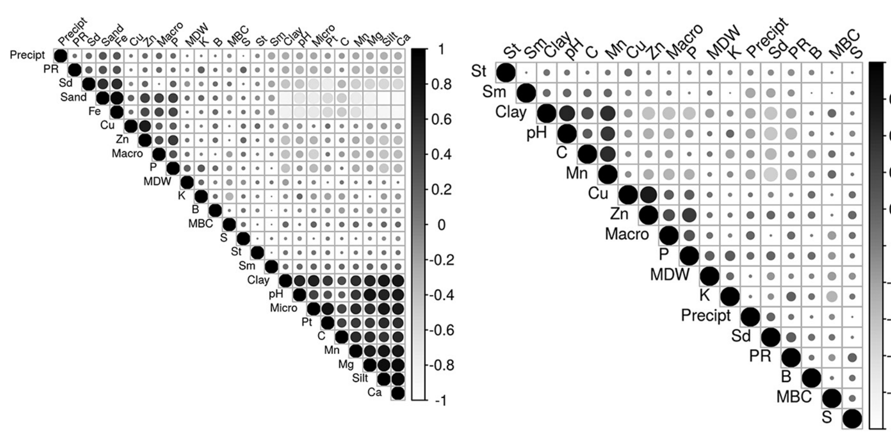

Figure 2 – Selection of soil properties used in the Random Forest classification model; Sm = soil moisture; C = carbon; Mn = manganese content; St = soil temperature; MBC = microbial biomass carbon; S = sulphur content; Precipt = precipitation; Sd = soil bulk density; CO_2 = soil CO_2 emissions; MDW = mean diameter weight; Cu = copper content; Zn = Zinc content; Macro = macroporosity; P = phosphorous content; B = boron content; PR = soil penetration resistance; K = potassium content.

the management system presented order of importance of 50-60 % in the Random Forest model range, and precipitation and Sm showed no or little influence on soil CO_2 flux (Figure 3).

Eight properties showed statistically significant differences ($p < 0.05$) between sugarcane management areas, considering that four (PR, P, Sd, and Macro) showed higher values in the GS area and four (clay, C, pH, and Micro) showed higher values in the BS area (Table 1).

The clay content at had greater relevance due to its classification as the most important attribute in the ranking of the model to explain the variations in CO_2 flux in the studied areas (Figure 3). Hence, clay content in the soil was higher ($p < 0.05$) in the BS area (552.14 g kg^{-1}) than in the GS area (531.57 g kg^{-1}).

Figure 4 shows the variation in CO_2 flux due to clay content in the management areas. In the GS area, the clay content is lower and CO_2 flux is higher, while the opposite is observed for the BS area.

Model validation using ¼ of observed data presented good coefficient of determination, $R^2 = 0.80$, between observed and predicted data (Figure 5).

Discussion

Among the variables selected for the model generated by the Random Forest algorithm, the clay content showed the greatest importance as the main attribute to explain the CO_2 flux for the sugarcane management areas evaluated in this study. In the PCA, the clay content

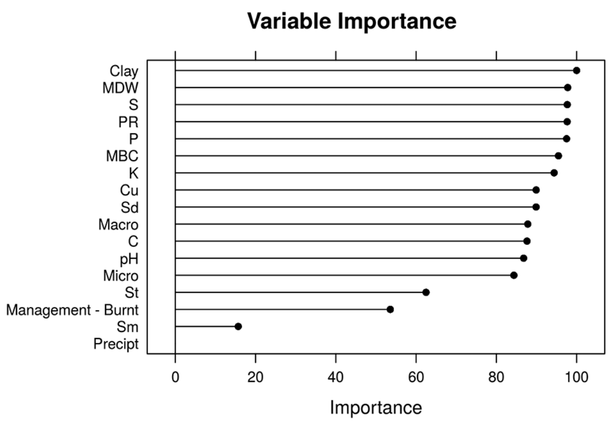

Figure 3 – Order of importance of properties selected for use in the model generated by the Random Forest algorithm to explain variations in the soil CO2 flux of sugarcane management systems. Sm = soil moisture; C = carbon; St = soil temperature; MBC = microbial biomass carbon; S = sulphur content; Precipt = precipitation; Sd = soil bulk density; MDW = mean diameter weight; Cu = copper content; Macro = macroporosity; P = phosphorous content; PR = soil penetration resistance; K = potassium content.

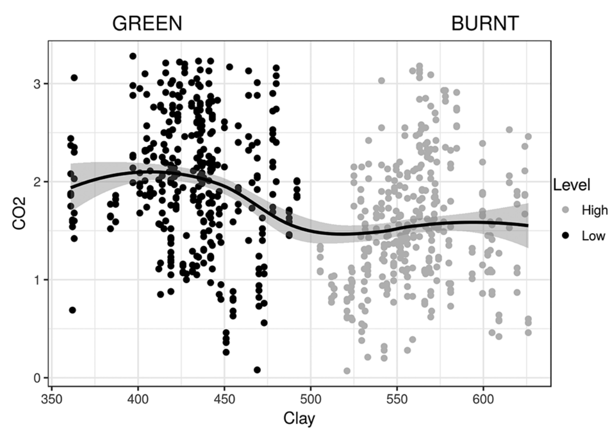

Figure 4 – Effect of different levels of clay (g kg^{-1}) content on soil CO_2 flux (μmol CO2 m^{-2} s^{-1}) in sugarcane management systems.

Table 1 – Descriptive statistics of the response variable (CO_2 flux) and other soil properties in sugarcane management systems.

Soil properties	Unit	Burned sugarcane		Green sugarcane	
		Mean	SD	Mean	SD
CO_2 flux	μmol CO_2 m^{-2} s^{-1}	1.53 b	0.69	2.68 a	1.30
Clay content	g kg^{-1}	552.14 a	39.20	531.57 b	31.50
Mean diameter weight	Mm	1.64 a	0.46	1.75 a	0.50
S content	mg dm^{-3}	7.34 a	4.75	7.26 a	5.80
Soil penetration resistance	Mpa	2.81 b	0.96	3.46 a	1.31
P content	mg dm^{-3}	15.89 b	5.54	32.34 a	20.29
Microbial biomass carbon	ugC g^{-1}d^{-1}	189.94 a	51.38	199.78 a	114.50
K content	cmolc dm^{-3}	5.24 a	2.40	4.86 a	2.99
Soil bulk density	g cm^{-3}	1.19 b	0.12	1.38 a	0.14
Macroporosity	cm^{-3} cm^{-3}	0.20 b	0.03	0.23 a	0.05
Carbon content	g kg^{-1}	29.5 a	3.2	25.30 b	3.60
pH	-	5.22 a	0.13	4.82 b	0.26
Microporosity	cm^{-3} cm^{-3}	0.37 a	0.02	0.26 b	0.06
Soil temperature	°C	24.73 a	3.55	21.85 b	6.14
Soil moisture	%	11.96 a	1.73	21.05 a	3.86

SD = standard deviation.

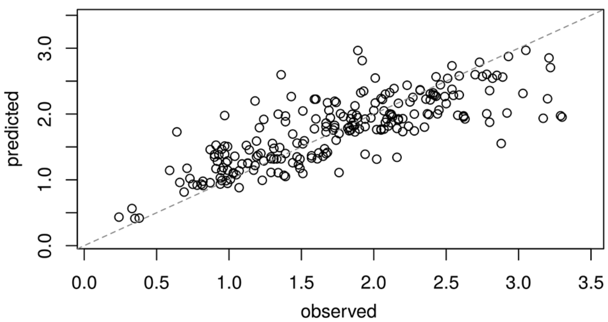

Figure 5 – The x and y axes represent, respectively, the observed and predicted soil CO_2 flux values. The results confirm how the Random Forest classification model fits well for prediction of soil CO_2 flux in sugarcane management systems, through a regression whose coefficient of determination (R^2) was equal to = 0.80.

explained 4-15 % of the variance in soil CO_2 flux (Tavares et al., 2015). Thus, regardless of the type statistical analysis, the clay content is an important soil attribute to explain CO_2 emissions.

The clay content in the soil was higher in the BS area when compared with the GS area. This difference is an intrinsic characteristic of the soil type in each area, resulting from the formation process with no or little influence of the management type. In addition, in the GS area, where the clay content is lower, the CO_2 flux was higher, while in the BS area, the clay content is higher and the CO_2 flux was lower.

One possible explanation of the clay content effect on the CO_2 flux is porosity, that is, soils with higher clay content tend to have more micropores and fewer macropores, thus, macroporosity is the main attribute for spatial distribution of soil gases. According to the Fick's law, macroporosity provides a less tortuous path for the CO_2 molecule in the soil (Alvenäs and Jansson, 1997; Brito et al., 2009). In turn, soil microporosity provides less linearity of porous space, with most devious paths, hindering transportation of CO_2 from the soil into the atmosphere. Therefore, possibly, higher macroporosity in the GS provided greater CO_2 emissions when compared with the BS area.

Another hypothesis refers to one of the clay properties in soil, which is to promote the aggregation of particles. It is an important process of soil carbon accumulation and stabilization, because clustering provides a physical protection to soil organic matter (SOM) against microbial attack, as widely reported in the literature (Carbonell-Bojollo et al., 2012; Edwards and Bremner, 1967; Elliott, 1986; Six et al., 2000; Tisdall and Oades, 1982). This physical protection promoted by the clay content in the soil prevents OM loss in the form of CO_2. According to Cerri et al. (2007), the increase in the SOM amount reduces gas emissions into the atmosphere, primarily CO_2, CH_4, and N_2O.

In the BS area, the carbon content was higher than in the GS area possibly because of cane ash deposition, which can increase carbon contents in the soil. On the other hand, in the BS area, where clay content is higher, there was a higher soil carbon content compared with the GS area, where the clay content is lower. Studies of Cerri et al. (2011) and Silver et al. (2000) showed the effects of soil texture on carbon stocks, and observed a direct correlation between clay content attributes and carbon stocks, which highlights the importance of clay content for the potential of carbon sequestration in the soil and consequent reduction in CO_2 emissions. As reported by Carbonell-Bojollo et al. (2012), CO_2 emissions in three sites was explained by soil textural differences. The area with higher clay content (62 %) showed lower CO_2 emissions, while areas with lower clay content (44 and 9 %) showed higher CO_2 emissions.

The second variable with greater degree of importance was MDW, which evaluates the structural quality of the soil. In both management areas, the MDW value was statistically equal, but with great influence on CO_2 emissions. Other attributes directly related to soil porosity, such as Sd and PR, showed greater importance, meaning a direct effect of CO_2 emissions on the soil. Some studies also showed direct or indirect influence of density and porosity on soil CO_2 emissions (Bicalho et al., 2014; Epron et al., 2004; Xu and Qi, 2001).

Other attributes also showed great importance, such as sulphur and phosphorus contents that are directly related to microbial performance, and consequently, to CO_2 production (Nordgren, 1992; Tate, 1995). Schwendenmann et al. (2003) investigated the spatial and temporal variation in CO_2 emissions in forests and reported the influence of phosphorus content in soil CO_2 emissions. Corroborating this, the MBC, which represents the number of microorganisms per soil unit mass, also had degree of importance in the prediction of CO_2 emissions. Ball and Virginia (2015) observed soil CO_2 emissions moderately increased with microbial biomass, demonstrating a sometimes small, but significant role of biological fluxes. A study of Xu and Qi (2001) proved the direct relationship between CO_2 emissions and MBC while monitoring both attributes.

Several studies indicate temperature and soil moisture as the main factors that influence CO_2 emissions (Carbonell-Bojollo et al., 2012; Epron et al., 2004; Epron et al., 2006; Kosugi et al., 2007; La Scala Jr. et al., 2010; Xu and Qi, 2001) by providing a favorable microclimate for the development of microorganisms. However, in this study, the effect of these attributes on CO_2 flux was not very important because rainfall events did not occur during the evaluation periods of CO_2 emissions in the field, which kept CO_2 emissions stable over time.

The effect of the management system on CO_2 flux showed an importance degree of 50 % on soil CO_2 flux. The GS area presented CO_2 emissions higher than the BS area did, possibly due to the presence of sugarcane trash on the soil in the GS system. This is because as sugarcane trash is decomposed, part of the plant residue carbon is released as CO_2 into the atmosphere and part is incorporated into the soil, increasing carbon stocks (Cerri et al., 2011).

The same effect has been proven in various studies. For instance, the use of sugarcane trash on the soil surface increased CO_2 emissions in pine (Fang et al., 1998) and eucalyptus (Epron et al., 2004) crops with the highest CO_2 emission in regions of higher concentration of sugarcane trash. Medeiros et al. (2011) detected higher soil CO_2 flux in soil with straw (no tillage) when compared with conventional management, because sugarcane trash keeps better conditions of moisture and temperature, which favor microbial development. Lenka and Lal (2013) also detected higher CO_2 emissions in soils with greater amounts of wheat straw (16 t ha^{-1}) when compared with areas that had added 8 t ha^{-1} and 0 t ha^{-1}.

Conclusion

The clay content in the soil was the most important variable affecting CO_2 production in the studied areas. The use of the Random Forest algorithm enabled to create a model that showed a good fit in relation to the predicted and observed values, showing high potential to predict new cases.

References

Alvenäs, G.; Jansson, P.E. 1997. Model for evaporation, moisture and temperature of bare soil: calibration and sensitivity analysis. Agricultural and Forest Meteorology 88: 47-56.

Ball, B.A.; Virginia, R.A. 2015. Controls on diel soil CO_2 flux across moisture gradients in a polar desert. Antarctic Science 27: 527-534.

Bicalho, E.S.; Panosso, A.R.; Teixeira, D.D.B.; Miranda, J.G.V.; Pereira, G.T.; La Scala Jr., N. 2014. Spatial variability structure of soil CO_2 emission and soil attributes in sugarcane area. Agriculture, Ecosystems and Environment 189: 206-215.

Breiman, L. 2001. Random forests. Journal of Machine Learning 45: 5-32.

Brito, L.F.; Marque Júnior, J.; Pereira, G.T.; Souza, Z.M. 2009. Soil CO_2 emission of sugarcane fields as affected by topography. Scientia Agricola 66: 77-83.

Carbonell-Bojollo, R.; Torres, M.A.R.R.; Rodriguez-Lizana, A.; Ordónez-Fernández, R. 2012. Influence of soil and climate conditions on CO_2 emissions from agricultural soils. Water, Air & Soil Pollution 223: 3425-3435.

Castro, S.G.Q.; Franco, H.C.J.; Mutton, M.A. 2014. Harvest managements and cultural practices in sugarcane. Revista Brasileira de Ciência do Solo 38: 299-306.

Cerri, C.E.P.; Sparoveki, G.; Bernoux, M.; Easterling, W.E.; Melillo, J.M.; Cerri, C.C. 2007. Tropical agriculture and global warming: impacts and mitigation options. Scientia Agricola 64: 83-99.

Cerri, C.C.; Galdos, M.V.; Maia, S.M.F.; Bernoux, M.; Feigl, B.J.; Powlson, D.; Cerri, C.E.P. 2011. Effect of sugarcane harvesting systems on soil carbon stocks in Brazil: an examination of existing data. European Journal of Soil Science 62: 23-28.

Edwards, A.P.; Bremner, J.M. 1967. Microaggregates in soils. European Journal of Soil Science 18: 64-73.

Elliott, E.T. 1986. Aggregate structure and carbon, nitrogen, and phosphorus in native and cultivated soils. Soil Science Society of America Journal 50: 627-633.

Empresa Brasileira de Pesquisa Agropecuária [Embrapa]. 2011. Manual methods of soil analysis = Manual de análise de solos. 2ed. Embrapa, Rio de Janeiro, RJ, Brazil (in Portuguese).

Epron, D.; Bosc, A.; Bonal, D.; Freycon, V. 2006. Spatial variation of soil respiration across a topographic gradient in a tropical rain forest in French Guiana. Journal of Tropical Ecology 22: 565-574.

Epron, D.; Nouvellon, Y.; Roupsard, O.; Mouvondy, W.; Mabiala, A.; Saint-Andre, L.; Joffre, R.; Jourdan, J.; Bonnefond, J.M.; Berbigier, P.; Hamel, O. 2004. Spatial and temporal variations of soil respiration in a eucalyptus plantation in Congo. Forest, Ecology and Management 202: 149-160.

Fang, C.; Moncrieff, J.B.; Gholz, H.L.; Clark, K.L. 1998. Soil CO_2 efflux and its spatial variation in a Florida slash pine plantation. Plant and Soil 205: 135-146.

Figueiredo, E.B.; La Scala Jr., N. 2011. Greenhouse gas balance due to the conversion of sugarcane areas from burned to green harvest in Brazil. Agriculture, Ecosystems and Environment 141: 77-85.

Grimm, R.; Behrens, T.; Märker, M.; Elsenbeer, H. 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island: digital soil mapping using random forests analysis. Geoderma 146: 102-103.

Han, J.; Kamber, M.; Pei, J. 2011. Data Mining: Concepts and Techniques. 3ed. Morgan Kaufmann, San Francisco, CA, USA.

Hastie, T.; Tibshirani, R.; Friedman, J. 2009. 2ed. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, London, UK.

Jenkinson, D.S.; Powlson, D.S. 1976. The effects of biocidal treatments on metabolism in soil. V. Method for measuring soil biomass. Soil Biology and Biochemistry 8: 209-213.

Kemper, W.D.; Chepil, W.S. 1965. Size distribution of aggregates. p. 499-510. In: Black, C.A., ed. Methods of soil analysis: physical and mineralogical properties, including statistics of measurement and sampling. American Society of Agronomy, Madison, WI, USA.

Kosugi, Y.; Mitani, T.; Itoh, M.; Noguchi, S.; Tani, M.; Matsuo, N.; Takanashi, S.; Ohkubo, S.; Nik, A.R. 2007. Spatial and temporal variation in soil respiration in a Southeast Asian tropical rainforest. Agricultural and Forest Meteorology 147: 35-47.

La Scala Jr., N.; Mendonça, E.S.; Souza, J.J.; Panosso, A.R.; Simas, F.N.B.; Schaefer, C.E.G.R. 2010. Spatial and temporal variability in soil CO_2-C emissions and relation to soil temperature at King George Island, Maritime Antarctica. National Institute of Polar Research 4: 479-487.

Lenka, N.K.; Lal, R. 2013. Soil aggregation and greenhouse gas flux after 15 years of wheat straw and fertilizer management in a no-till system. Soil and Tillage Research 126: 78-89.

Medeiros, J.C.; Silva, A.P.; Cerri, C.E.P.; Fracetto, F.J.C. 2011. Linking physical quality and CO_2 emission under long-term no-till and conventional-till in a subtropical soil in Brazil. Plant and Soil 338: 5-15.

Nelson, D.W.; Sommers, L.E. 1982. Total carbon, organic carbon and organic matter. In: Page, R.H.; Kenny, D.R., eds. Methods of soil analysis. II. Chemical and microbiological properties. 2ed. Soil Science Society of America, Madison, WI, USA.

Nordgren, A. 1992. A method for determining microbially available N and P in an organic soil. Biology and Fertility of Soils 13: 195-199.

Panosso, A.R.; Marques Júnior, J.; Pereira, G.T.; La Scala Jr., N. 2009. Spatial and temporal variability of soil CO_2 emission in a sugarcane area under green and slash-and-burn managements. Soil and Tillage Research 105: 275-282.

Raij, B.V.; Andrade, J.C.; Cantarella, H.; Quaggio, J.A. 2001. Chemical analysis to evaluate the fertility of tropical soils = Análise química para avaliação da fertilidade de solos tropicais. Instituto Agronômico, Campinas, SP, Brazil (in Portuguese).

Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 67: 93-104.

Rolim, G.S.; Camargo, M.B.P.; Lania, D.G.; Moraes, J.F.L. 2007. Climate classification of Koppen and Thornthwaite systems and their applicability in the determination of agroclimatic zoning for the state of São Paulo, Brazil. Bragantia 66: 711-729.

Sant'anna, S.A.C.; Fernandes, M.F.; Ivo, W.M.P.M.; Costa, J.L.S. 2009. Evaluation of soil quality indicators in sugarcane management in sandy loam soil. Pedosphere 19: 312-322.

Schwendenmann, L.; Veldkamp, E.; Brenes, T.; O'brien, J.J.; MacKensen, J. 2003. Spatial and temporal variation in soil CO2 efflux in an old-growth neotropical rain forest, La Selva, Costa Rica. Biogeochemistry 64: 111-128.

Silver, W.L.; Neff, J.; McGroddy, M.; Veldkamp, E.; Keller, M.; Cosme, R. 2000. Effects of soil texture on belowground carbon and nutrient storage in a lowland Amazonian forest ecosystem. Ecosystems 3: 193-209.

Six, J.; Paustian, K.; Elliott, E.T.; Combrink, C. 2000. Soil structure and soil organic matter. I. Distribution of aggregate size classes and aggregate associated carbon. Soil Science Society of America Journal 64: 681-689.

Stolf, R. 1991. Theory and test of formulas for transforming impact penetrometer data in soil resistance. Revista Brasileira de Ciência do Solo 15: 229-235 (in Portuguese, with abstract in English).

Tate, R.L. 1995. The sulfur and related biogeochemical cycles. p. 359-373. In: Tate, R.L., ed. Soil microbiology. John Wiley, New York, NY, USA.

Tavares, R.L.M.; Souza, Z.M.; La Scala Jr., N.; Castioni, G.A.F.; Souza, G.S.; Torres, J.L. 2016a. Spatial and temporal variability of soil CO2 flux in sugarcane green harvest systems. Revista Brasileira de Ciência do Solo 40: 195-203.

Tavares, R.L.M.; Siqueira, D.S.; Panosso, A.R.; Castioni, G.A.F.; Souza, Z.M.; La Scala Jr., N. 2016b. Soil management of sugarcane fields affecting CO2 fluxes. Scientia Agricola 73: 543-551.

Tavares, R.L.M.; Souza, Z.M.; Siqueira, D.S.; La Scala Jr., N.; Panosso, A.R.; Campos, M.C.C. 2015. Soil CO2 emission in sugarcane management systems. Acta Agriculturae Scandinavica, Section B – Soil & Plant Science 65: 755-762.

Tisdall, J.M.; Oades, J.M. 1982. Organic matter and water-stable aggregates in soils. European Journal of Soil Science 33: 141-163.

Xu, M.; Qi, Y. 2001. Soil surface CO2 efflux and its spatial and temporal variations in a young ponderosa pine plantation in northern California. Global Change Biology 7: 667-677.