RESSALVA

Atendendo solicitação do(a)
autor(a), o texto completo desta
Tese será disponibilizado somente a
partir de 09/08/2023.

SÃO PAULO STATE UNIVERSITY – UNESP
JABOTICABAL CAMPUS

MACHINE LEARNING MODELING IN TEMPORAL VARIABILITY OF SOIL RESPIRATION IN PLANTED FOREST AREAS

Maria Elisa Vicentini
Agronomic Engineering

2021

SÃO PAULO STATE UNIVERSITY – UNESP

JABOTICABAL CAMPUS

MACHINE LEARNING MODELING IN TEMPORAL VARIABILITY OF SOIL RESPIRATION IN PLANTED FOREST AREAS

Maria Elisa Vicentini

Advisor: Prof. Dr. Alan Rodrigo Panosso

Co-Advisor: Prof. Dr. Glauco de Souza Rolim

The thesis presented to the College of Agricultural and Veterinarian Sciences – UNESP, Jaboticabal Campus, as partial fulfillment of the Doctor degree in Agronomy (Soil Science).

2021

V633m

Vicentini, Maria Elisa

Machine learning modeling in temporal variability of soil respiration in planted forest areas / Maria Elisa Vicentini. -- Jaboticabal, 2021

170 p. : il., tabs., mapas

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal

Orientador: Alan Rodrigo Panosso

Coorientador: Glauco de Souza Rolim

1. Carbon dynamics. 2. Soil atmosphere. 3. Land use. 4. Greenhouse gases. 5. Mathematical models. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: MACHINE LEARNING MODELING IN TEMPORAL VARIABILITY OF SOIL RESPIRATION IN PLANTED FOREST AREAS

AUTORA: MARIA ELISA VICENTINI
ORIENTADOR: ALAN RODRIGO PANOSSO
COORIENTADOR: GLAUCO DE SOUZA ROLIM

Aprovada como parte das exigências para obtenção do Título de Doutora em AGRONOMIA (CIÊNCIA DO SOLO), pela Comissão Examinadora:

Prof. Dr. ALAN RODRIGO PANOSSO (Participaçao Virtual)
Departamento de Engenharia e Ciências Exatas (DECEx) / FCAV / UNESP - Jaboticabal

Pesquisador Dr. RICARDO DE OLIVEIRA BORDONAL (Participaçao Virtual) p/
CTBE/CNPEM / Campinas/SP

Prof. Dr. NEWTON LA SCALA JUNIOR (Participaçao Virtual) p/
Departamento de Engenharia e Ciências Exatas (DECEx) / FCAV / UNESP - Jaboticabal

Prof. Dr. DANIEL DE BORTOLI TEIXEIRA (Participaçao Virtual) p/
UNIMAR / Marília/SP

Prof. Dr. GENER TADEU PEREIRA (Participaçao Virtual) p/
Departamento de Engenharia e Ciências Exatas / FCAV / UNESP - Jaboticabal

Jaboticabal, 09 de agosto de 2021

AUTHOR'S CURRICULUM DATA

MARIA ELISA VICENTINI - Daughter of José Vicentini and Rejane Maria Vicentini, Maria Elisa Vincentini was born on January 13, 1983, in the city of Jales, São Paulo State, Brazil. She studied Agronomic Engineering at the State University of Londrina (UEL), Paraná State. At the same institution, she obtained the titles of Specialist in Environmental Analysis and Education in Earth Sciences in 2012 and Master in Bioenergy in 2014. In February 2017, she obtained a Master of Science degree in Agronomy from the School of Natural Sciences and Engineering of Ilha Solteira (FEIS)/São Paulo State University (UNESP). In March 2017, she enrolled in a PhD course in Agronomy (Soil Science) from the Faculty of Agrarian and Veterinary Sciences (FCAV-UNESP), Campus of Jaboticabal, SP. In August 2021, she submitted the doctoral thesis to an examination panel, and received his Ph.D. degree in Agronomy (Soil Science) at UNESP/FCAV.

“For God has not given us a spirit of fear, but of power and of love and of a sound mind”

(Bible, 2 Timothy 1:7)
New King James Version®.

I dedicate this work to my beloved father:

José Vicentini *(in memoriam).*

ACKNOWLEDGMENTS

The Lord Jesus Christ, author and finisher of my faith! *Thank you Father!*

My family for never letting me give up and always supporting me. Special thanks to my loving mother, Rejane.

Professor Dr. Alan Rodrigo Panosso, for his guidance, patience, friendship, incentives, and assistance during this project.You were always available to help me. In this journey that is now concluding, I leave my admiration for your professional competence.

Professor Dr. Glauco de Souza Rolim, who was always available to help, for his co-orientation, attention, and incentives, and mainly for his theoretical teaching.

Professors of the Qualification Committee: Dr. Newton La Scala Júnior and Dr. Nelson José Peruzzi, for their suggestions and attention.

The defense board members, Professors Dr. Daniel de Bortoli Teixeira, Dr. Gener Tadeu Pereira, Dr. Newton La Scala Júnior, and Ricardo de Oliveira Bordonal for their suggestions and criticisms that contributed to the conclusion of this study.

I also thank the professors at the Department of Engineering and Exact Sciences, who always elucidated my doubts in a kind manner.

The employees of the Engineering and Exact Sciences Department: Maria José Trizólio (Zezé), Adriana Takakura, and Shirley de Sousa, Thank you very much!

The colleagues and friends in the Department of Engineering and Exact Sciences , especially those with whom I had a close relationship: Bruna, Deise Nogueira, Gustavo André, Kamila Meneses, Kleve Canteral, Láis Teixeira, Ludhanna Veras, Mary Jane Nunes, Wanderson Lucena, and Paulo Alexandre Silva. Thank you for the harmonious coexistence, exchange of experiences, and knowledge shared.

Dear friends Angelina Pedro Chitlhango and Mayara Germana Gomes. Thanks for your friendship, for the encouragement, support in this journey!

The College of Agricultural and Veterinary Sciences, Paulista State University "Júlio de Mesquita Filho," Jaboticabal campus, especially the Graduate Program in Agronomy (Soil Science), for the opportunity offered.

The Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (process no. 2016/03861-5), for funding the Project.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

And to all who contributed directly and indirectly to the conduct of my study. Thank you very much!

SUMMARY

SUMMARY ***i***

ABSTRACT ***iv***

RESUMO ***vi***

LIST OF ABBREVIATIONS AND ACRONYMS ***viii***

CHAPTER 1 – General considerations **1**

1 Introduction 1

1.2 – Hypothesis 3

1.3 – Objectives 3

1.4 Literature Review 3

1.4.1 Areas of planted forest 3

1.5 Soil Respiration 5

1.5.1 Soil properties 5

1.5.2 Soil aeration 6

1.5.3 Soil respiration associated photosynthesis 7

1.5.4 Soil respiration and the climate 8

1.6 Machine Learning (ML) 8

1.6.1 Neural Networks 11

1.6.1.1 Multilayer Perceptron Neural Network (MLPNN) 12

1.6.1.2 Radial Basis Function Neural Network (RBFNN) 12

1.6.1.3 Generalized Regression Neural Network (GRNN) 13

1.7 Regression with Support Vectors 13

1.8. Adaptive Neuro-Fuzzy Inference System (ANFIS) 15

1.9 Random Forest (Regressor) 16

1.10 References 17

CHAPTER 2 – Estimation of soil respiration in a pine forest planted in the Cerrado using artificial neural networks and support vector regression **25**

2.1. Introduction 26

2. Materials and methods 27

2.1 Data 27

2.2 Study location 27

2.3 Determination of soil CO_2 emission, oxygen influx, temperature, soil water content, climate and remote sensing 29

2.4 Datasets 30

2.5. Data preprocessing and analysis 31

2.5.1 Principal Component Analysis 31

2.5. 2 Machine learning algorithms 31

2.6 Evaluation criteria for model performance ____ 34
2.7. Results and Discussion ____ 34
2.7.1 PCA ____ 34
2.8 Artificial neural networks (ANNs) ____ 37
2.9 Support vector regression (SVR) ____ 43
2.10. Concluding remarks ____ 47
2.11 References ____ 48

CHAPTER 3 - Estimation of soil CO2 emissions and oxygen influx in a eucalyptus forest using artificial neural networks and support vector regression machine learning ____ 54

3.1 Introduction ____ 55
3.2 Materials and methods ____ 57
3.2.1 Study area ____ 57
3.3. Soil data ____ 58
3.3.1 Determination of soil CO_2 emission, O_2 uptake, soil temperature, moisture and Environmental data ____ 58
3.4 Data processing ____ 59
3.4.1 Principal component analysis ____ 60
3..4.2 Artificial neural networks ____ 61
3..4.3 Support vector regression ____ 62
3.5 Results and Discussion ____ 64
3.5. 1 Model performance for FCO_2 ____ 64
3.5.2 Model performance for FO_2 ____ 66
3.5.3 Importance of Predictor Variables ____ 72
3.6 Conclusion ____ 75
3.7 References ____ 76

CHAPTER 4 – Regression using artificial neural networks and Adaptive Neuro-Fuzzy Inference Systems in predicting soil respiration in areas of native species in the Cerrado ____ 82

4.1 Introduction ____ 83
4.2 Materials and methods ____ 85
4. 2.2 Study area ____ 85
4.2.3 Determination of soil CO_2 emission, O_2 influx, soil temperature, and water content ____ 86
4. 2.4 Soil data and Environmental variables ____ 87
4.2.5 Data mining ____ 88
4.2.5.1 Artificial neural network (ANN) ____ 88
4.2.5.2 Adaptive Neuro-diffuse FIS (ANFIS) ____ 90
4.2.5.3 Criteria for model assessment ____ 91
4.2.5.4 Summary of the Methodology Used in This Study ____ 91

4.2.6 Results and Discussion 92
4.2.6.1 Principal Component Analysis 112
4.2.7 Conclusion 116
4,8 References 118

CHAPTER 5 - MACHINE LEARNING MODELING IN THE TEMPORAL VARIABILITY OF SOIL RESPIRATION IN PLANTED FOREST AREAS WITH EUCALYPTUS, PINE, AND NATIVE SPECIES IN THE BRAZILIAN CERRADO 124

5.1. Introduction 126
5.2 Materials and methods 128
5.2.1 Data used 128
5. 2.2 Site description 128
5.2.3 Environmental data and Soil sampling and analysis of physico–chemical properties of soils 129
5.2.4 Measurements of soil respiration, soil temperature, soil water content and Oxygen Uptake 130
5.2.5 Statistical Analysis 131
5.2.6 Principal Component Analysis (PCA) 133
5.2.7 Artificial Neural Network (ANN) 133
5..2.8. Regression by Support Vector Machine (SVM) 135
5.2.9 Adaptive Neuro-Fuzzy Inference System (ANFIS) 136
5.2.10 Random Forest Regression 138
5..2.11 Statistical Measures of Performance Evaluation of Different Models 138
5.2.12 Summary of the Methodology Used in This Study 139
5.3.1 Principal Component Analysis 140
5.3.2 Neural Networks 142
5.3.4 Support Vector Regression 144
5.3.5 Neuro-Diffuse Adaptive Inference System 147
5.3.6 Random Forest 148
5 3.7 Comparison of Models 150
5. 3.8. Discussion 152
5.3.8.1 Principal Component Analysis 152
5.3.8.2 Development and comparison of models 154
5.3.9 Conclusion 156
5.3.10 References 157

CHAPTER 6 - FINAL REMARKS 165

MACHINE LEARNING MODELING IN TEMPORAL VARIABILITY OF SOIL RESPIRATION IN PLANTED FOREST AREAS

ABSTRACT - Understanding the temporal dynamics of land respiration in tropical ecosystems is challenging, especially when it is associated with Land Use, Land-Use Change and Forestry (LULUCF). Many studies have modeled the dynamics of CO_2 emission from soil (FCO_2), but few studies have modeled the temporal dynamics of soil O_2 influx (FO_2). Therefore, the objective of this study was to evaluate the predictive performance of artificial neural networks (ANN), support vector regression (SVR), adaptive neuro-fuzzy inference system (ANFIS), and random forest (RF) machine learning (ML) techniques in modeling the temporal variability of FCO_2 and FO_2 in forests planted in three ecosystems of planted forests: Pinus (*Pinus spp*), Eucalyptus (*Eucalyptus spp*), and native species, converted more than 30 years ago in the Cerrado biome, Brazil. We used a database composed of agro-meteorological data, improved vegetation index (EVI), and soil chemical and physical attributes as predictor variables and principal component analysis as the main data mining technique. For each monoculture and native species forest the numbers of FCO_2 and FO_2 recordings were (n = 500) and (n = 175), respectively. For pine forest, ANNs showed better predictive performance than SVR. The multilayer perceptron (MLPNN) with 12 input variables explained R^2 = 42% of the temporal variability in FCO_2. The general regression neural network (GRNN) with 10 input variables explained temporal variability in FO_2 with an R^2 of 56%. For eucalyptus, in the estimation of FCO_2, the best predictive performance was obtained with MLP with validation (R^2 = 0.59; RMSE = 1.034 µmol m-^{2}s-^{1}). FO_2 estimation: validation (R^2 = 0.36; RMSE = 0.076 mg m-^{2}s-^{1}). SVR with radial basis function kernel (SVR-RBF) was superior to the sigmoid (SVR-SIG), and polynomial kernels (SVR-PL), with the following values for FCO_2; validation (R^2 = 0.53; RMSE = 0.990 µmol m-^{2}s-^{1}).In Native Species areas, the best results were: FCO_2 with Radial Basis Function Neural Network (RBFNN) (R^2 = 0.54, RMSE = 1.015 µmol m-^{2}s-^{1}) and FO_2 with RBFNN (R^2 = 0.74, 0.079 mg m-^{2}s-^{1}). Estimates of FCO_2 showed better predictive performance than FO_2. RBFNN was best estimate for FCO_2. MLPNN is the best architecture for FO_2 (R^2 = 0.45, RMSE = 0.94 mg m-^{2}s-^{1}). In relation to ANFIS, FO_2, did not show good generalizability and presented the worst performance, showing the highest mean absolute percentage error and lowest accuracy (R^2 = 0.12, MAPE 51.27% and R^2 = 0.28, MAPE 47.48%)

calibration and validation respectively). Analyzing the performance of the two estimates, SVR-RBF for FCO_2 performed better than SVR-RBF for FO_2. The RF for FCO_2 in the calibration and validation phases presented values of the (R^2 = 0.80 and R^2 = 0.60 respectively). The type of forest influenced temporal variability in soil respiration. We found that soil temperature (T_s) EVI, global solar radiation (GSR), macroporosity (macro), and organic matter (SOM) were the variables that most influenced the two estimates.

Keywords: Carbon dynamics, Soil-atmosphere, Land use, Greenhouse gases, Mathematical models.

APRENDIZADO DE MÁQUINA NA MODELAGEM DA VARIABILIDADE TEMPORAL DA RESPIRAÇÃO DO SOLO EM ÁREA DE FLORESTA PLANTADA

RESUMO - Compreender a dinâmica temporal da respiração do solo (R_S) nos ecossistemas tropicais é desafiador, principalmente quando está associada à Mudança do uso da terra e Florestas (MUTF). Os diferentes tipos de manejo podem ter impactos na mudança do fluxo CO_2 (ou emissão de CO_2 do solo - FCO_2), e do influxo de oxigênio no solo (FO_2). Muitos estudos modelaram a dinâmica temporal da FCO_2, mas poucos estudos modelaram a dinâmica temporal do FO_2. Portanto, o objetivo deste estudo foi avaliar o desempenho preditivo de quatro técnicas de aprendizado de máquina: redes neurais artificiais (RNA), Regressão por vetores suporte (RVS), Sistema de Inferência Adaptativo Neuro-Difuso (ANFIS) e Random Forest (RF) na modelagem da variabilidade temporal da FCO_2 e FO_2 em áreas de florestas plantadas convertidas há mais de 30 anos no bioma Cerrado, Brasil, com *Eucalyptus* (*Eucalyptus* spp.), *Pinus* (*Pinus* spp) e espécies nativa, O banco de dados foi composto pelas seguintes variáveis preditoras: dados agro-meteorológicos, índice de vegetação melhorado (EVI), e atributos químicos e físicos do solo. A análise de componentes principais foi utilizada como técnica de mineração dos dados. Para cada área de monocultivo e reflorestamento com espécies nativas o número de observações das variáveis respostas para FCO_2 e FO_2 foram (n= 500) e (n = 175) respectivamente. Na floresta de pinus, as RNas, demonstraram melhor desempenho preditivo do que as RVS. Uma rede neural Multilayer Perceptron (MLPNN), constituída com 12 variáveis de entrada explicou 42% da variabilidade temporal da FCO_2. A rede neural de regressão geral (GRNN) com 10 variáveis de entrada explicou (R^2 = 56%) da variabilidade temporal FO_2. Para floresta de eucalipto, na estimativa da FCO_2, o melhor desempenho foi obtido com MLPNN na fase de validação (R^2 = 0,59;) e Raiz quadrada do erro-médio (RMSE = 1,034 µmol m^{-2} s^{-1}). Para FO_2 os valores foram: validação (R^2 = 0,36; RMSE = 0,076 mg m^{-2} s^{-1}). Em relação à SVR o desempenho dos modelos com o kernel de função de base radial (SVR-RBF) foi superior ao sigmóide (SVR-SIG), e kernel polinomial (SVR-PL). Os seguintes valores forma observados para FCO_2; validação (R^2 = 0,53; RMSE = 0,990 µmol m^{-2} s^{-1}). A dinâmica da FCO_2 e FO_2 nessa área foi associada à respiração das raízes. Na área com espécies nativas, o desempenho mais preditivo para FCO_2 foi com Rede Neural de Função de Base Radial (RBFNN) (R^2 = 0,54, RMSE = 1,015 µmol m-2s-1) e FO_2 com RBFNN (R^2 = 0,74; RMSE 0,079 mg $m^{-2}s^{-1}$). Na segunda fase do estudo, desenvolvemos um modelo global considerando a base de dados das três áreas para estimativa da FCO_2 e FO_2. As estimativas para FCO_2 foram mais perditivas que $FO_{2.}$ .A RBFNN foi a melhor modelo para na FCO_2 (R^2 = 0,51; RMSE = 0,97 µmol m-^{2}s^{-1}). Em contrapartida a MLPNN foi a melhor arquitetura para o FO_2 (R^2 = 0,45, RMSE = 0,94 mg m-^{2}s-^{1}). A ANFIS não resultou em um modelo com boa capacidade de generalização para FO_2 e apresentou o pior desempenho na calibração (R^2 = 0,12, MAPE 51, 27%) e validação (R^2 = 0,28 e MAPE 47, 48%) com elevados erros percentuais associados. De uma forma geral o modelo SVR-RBF teve melhor desempenho na $FCO_{2,}$ quando comparado com FO_2. O modelo RF foi satisfatório para FCO_2 nas fases de calibração e validação R^2 0,80 e R^2 0,60 respectivamente. O tipo de floresta influenciou a variabilidade temporal na respiração do solo. Verificámos que a temperatura do solo (T_s) EVI, radiação solar (RSG), macroporosidade (macro), e matéria orgânica (MOS) foram as variáveis que mais influenciaram as duas estimativas.

Palavras-chave: Dinâmica do carbono, Atmosfera do solo, Uso do solo, Gases do efeito estufa, Modelos matemáticos.

LIST OF ABBREVIATIONS AND ACRONYMS

AI - Artificial intelligence

ANN - Artificial neural network

ANFIS - Adaptive neuro-fuzzy inference system

Bases - sum of bases

Bd - Soil bulk density

c - Confidence coefficient

CEC - Cation exchange capacity

CO_2 - Carbon dioxide

Cstock - Carbon stocks

d - Index of Agreement developed by Willmott

ET_o - Evapotranspiration reference

EVI - Enhanced Vegetation Index

FCO_2 – soil CO_2 emission

FO_2 - influx of O_2 in the soil

GSR - Global solar radiation

GRNN - Generalized Regression Neural Network

Macro - Macroporosity

MAPE- Mean absolute percentage error

MEA - Mean absolute error

Micro - Microporosity

ML - machine learning

MLPNN - multilayer perceptron

MSE - Mean square error

NStock - Nitrogen stock

P - Phosphorus

PAR - Photosynthetically active radiation

P_{atm} - Atmospheric pressure

PCA - Principal component analysis

pH - hydrogen potential

SVR - PL Polynomial kernel

r - Pearson correlation coefficient

SVR – RBF Radial basis function kernel

RH- relative humidity

RF - Random forest

R_s - Soil respiration

SA- Sensitivity analysis

SOC- Soil organic C

SOM- Soil organic matter

SVR - SIG Sigmoid kernel

SVMs - Support vector machines

SVC - Support vector classification

SVR - Support vector regression

SWC - Soil water content

T_s - Soil temperature

T_{air} - air temperature

w - wind velocity

CHAPTER 1 – GENERAL CONSIDERATIONS

1 Introduction

Over the past 60 years (1960–2019), land use change (LUC) has affected nearly one-third (32%) of the global surface area. During this period, 0.8 million square kilometers of forests is estimated to have been cleared worldwide (Winkler et al., 2021); therefore, tropical ecosystems are under constant human pressure (Wright and Muller-landau, 2006; Barbosa et al. 2021). In 2019 in Brazil, land use, land-use change and forestry (LULUCF) sector was responsible for emitting 425.92 million tons of carbon dioxide (tCO_2) into the atmosphere (Alencar et al., 2020), and more than 12 thousand kilometers (km^2) of native vegetation near indigenous lands and areas of legal reserve or permanent protection were devastated, mainly by illegal means. The occurrences were concentrated in the Amazon and Cerrado regions (Azevedo et al., 2021). Actions of this type are influenced by economic factors and the neglect of environmental policies by managers. The main consequences of the removal and replacement of native vegetation are related to the biogeochemical cycle and irrigation, particularly the seasonality of precipitation (Chambers and Artaxo, 2017).

Forests contribute approximately half of the global net primary production (NPP), have significant abilities to sequester and store carbon (C) in plant biomass and soil, and can offset anthropogenic CO_2 emissions (DeLucia et al., 1999). The benefits of reforestation go beyond biodiversity conservation and affect the region's microclimate and precipitation events through its relationship with the atmosphere, including seasonal patterns of photosynthesis and transpiration (Del Rosario Uribe and Dukes, 2021), which are fundamental to minimizing the effects of climate change.

The benefits of reforestation go beyond biodiversity conservation and influence the microclimate of the region, consequently the occurrence of precipitation events and their relationship with the atmosphere, including seasonal patterns of photosynthesis and transpiration (Del Rosario Uribe and Dukes, 2021), which are key to minimizing the effects of climate change.

Brazil contains some of the largest planted forests in the world, according to the Annual Report of the Brazilian Tree Industry (IBÁ, 2020). In 2019, the total area of planted trees totaled 9.0 million hectares, 90% of which were composed of *Eucalyptus* (*Eucalyptus* spp.) and *Pinus* (*Pinus* spp.), species used for the production of charcoal, lumber, paper, energy, and others, stimulating the bioeconomy of the states located in the Cerrado region, such as Mato Grosso do Sul (MS).

In the soil, CO_2 emission or fllux (FCO_2) is the result of two factors: production and gas transport. Production is derived from the respiration of roots (autotrophic) and the biochemical processes that occur in soil due to the mineralization of organic matter, which is the main source of energy in the respiration of microorganisms (heterotrophic) (Lal, 2009; Moitinho et al., 2015), whereby aerobic forms consume oxygen (FO_2), which is fundamental for the metabolism of microorganisms.

Thus, soil respiration (R_s) is one of the major carbon fluxes between the land surface and the atmosphere (Adachi et al., 2017). Regarding transport, CO_2 and O_2 move mainly by diffusion, as described by Fick's Law. It is the main mechanism of soil aeration; therefore, it is conditioned by the physical properties of the soil (Neira et al., 2015). Consumption by plant roots and microbial populations causes the partial pressure of oxygen to decrease, resulting in an O_2 pressure gradient between the atmosphere and the soil (Zhu et al., 2019).

In the field it is difficult for us to separate autotrophic respiration from heterotrophic respiration (Fan and Han, 2018), both of which participate in multiple interactions with climate variables such as solar radiation (La Scala et al., 2003), soil chemical and physical attributes, (de Araújo Santos et al., 2019; Moitinho et al., 2021), photosynthesis (Savage et al., 2013), and oxygen availability (Ferraz de Almeida et al., 2018).

Studies indicate that soil temperature and moisture are often associated with temporal variability in R_s (Davidson et al., 1998; Li et al., 2008). Temperature can directly affect soil microbial functions through increased rates of microbial activity and soil organic matter (SOM) decomposition (Luo, 2006), while soil water content may be related to gas diffusion and microbial mobility (Silva et al., 2019). Tang et al. (2005) stated that the pattern of temporal variability of R_s in forest areas, on a diurnal scale, is modulated by photosynthetic activity. Biophysical parameters, such as the Improved Vegetation Index (EVI), have been shown to correlate well with R_s (Huang et al., 2012).

For better monitoring of soil carbon dynamics, many studies have been conducted with different models and predictor variables to estimate and understand the temporal dynamics of soil respiration in different ecosystems (Tuomi et al., 2008; Chen et al., 2010). According to Vargas et al. (2010), our ability to predict CO_2 emissions remains limited due to the complexity of the phenomenon. In addition, most studies focus only on the study of FCO_2, and few have modeled oxygen consumption, which is essential for these dynamics to occur (Ferraz de Almeida et al., 2018; Vicentini et al., 2019).

Given the above, artificial intelligence (AI) with the adoption of machine learning techniques such as artificial neural networks (ANNs), support vector regression (SVR), adaptive neuro-fuzzy inference system (ANFIS) and random forest (RF), have been successfully applied to model various processes in the environmental sciences, due to their ability to solve complex nonlinear problems (Fernandes et al., 2019; Taghizadeh-Mehrjardi et al., 2020) particularly in relation to soil respiration (Farhate et al., 2018; Freitas et al., 2018; Liu et al., 2020).

5.3.9 Conclusion

The neural and random forest networks provided predictive models to explain the patterns of temporal variability of FCO_2 from climate, soil, and EVI variables. However, these networks were associated with minor errors. RBF is the best architecture. In relation to FO_2, networks and regression of support vectors performed well. For this estimate, MLP returned the best result. However, new data mining methodologies need to be employed to improve dataset selection for training. Further studies need to be carried out as machine learning is rarely used in the prediction of soil respiration, especially when predicting oxygen consumption.

The type of forest influenced temporal variability in soil respiration in ANN. We found that the spectral index (EVI), soil temperature, solar radiation, macroporosity, and organic matter were the variables that most influenced the two estimates. FCO_2 modeling provides more predictive models than FO_2.

5.3.10 References

Afroozeh M, Reza Sohrabi M, Davallo M, Mirnezami SY, Motiee F, Khosravi M (2018). Application of Artificial Neural Network, Fuzzy Inference System and Adaptive Neuro-Fuzzy Inference System to Predict the Removal of Pb(II) Ions from the Aqueous Solution by Using Magnetic Graphene/Nylon 6. **Chemical Science journal** ;09. https://doi.org/10.4172/2150-3494.1000185

Alekseev A, Alekseeva T, Kalinin P, Hajnos M (2018). Soils response to the land use and soil climatic gradients at ecosystem scale: Mineralogical and geochemical data. **Soil and Tillage Research** ;180:38-47. https://doi.org/10.1016/j.still.2018.02.008

Arshad MA, Lowery B, Grossman B (1996) Physical tests for monitoring soil quality. In: Doran W, Jones J (eds) methods for assessing soil quality, Madison, **Soil Science Society of America Journal**, 49:123–141 (special publication).

Breiman L. Random forests. Mach Learn (2001) 45:5-32. https://doi.org/10.1023/A:1010933404324

Balogh J, Fóti S, Pintér K, Burri S, Eugster W, Papp M, Nagy Z (2015) Soil CO2 efflux and production rates as influenced by evapotranspiration in a dry grassland. **Plant and Soil.** 388:157-73. https://doi.org/10.1007/s11104-014-2314-3

Buragohain M, Mahanta C (2008) A novel approach for ANFIS modelling based on full factorial design. **Applied Soft Computing - Journals,** 8:609-25. https://doi.org/10.1016/j.asoc.2007.03.010

Chagas C da S, de Carvalho Junior W, Bhering SB, Calderano Filho B (2016) Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. **Catena** ;139:232-40. https://doi.org/10.1016/j.catena.2016.01.001

Chai T, Draxler RR (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. **Geoscientific Model Development** ;7:1247-50. https://doi.org/10.5194/gmd-7-1247-2014

Chen Y, Xu P, Chu Y, Li W, Wu Y, Ni L, Bao Y, Wang K (2017) Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. **Applied Energy** ;195:659-70. https://doi.org/10.1016/j.apenergy.2017.03.034

Davidson EA, Verchot L V., Henrique Cattânio J, Ackerman IL, Carvalho JEM.

(2000). Effects of soil water content on soil respiration in forests and cattle pastures of eastern Amazonia. **Biogeochemistry**; 48:53-69. https://doi.org/10.1023/A:1006204113917

Ding X, Liu G, Fu S, Chen HYH (2021). Tree species composition and nutrient availability affect soil microbial diversity and composition across forest types in subtropical China. **Catena**. 2021;201:105224. https://doi.org/10.1016/j.catena.2021.105224

Drucker H, Surges CJC, Kaufman L, Smola A, Vapnik V (1997) Support vector regression machines. A**dvances in neural information processing systems;** 1:155-61.

Don A, Schumacher J, Freibauer A (2011). Impact of tropical land-use change on soil organic carbon stocks - a meta-analysis. **Global Change Biology** ;17:1658-70. https://doi.org/10.1111/j.1365-2486.2010.02336.x

Doran JW, Parkin TB (1994) Defining and assessing soil quality. In: Doran JW, Coleman DC, Bezdicek DF, Stewart BA (eds) D**efining soil quality for a sustainable environment**. Madison, SSSA 3–21p.

Embrapa Solos (2018). Sistema brasileiro de classificação de solos.http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1094003

de Espindola GM, de Silva Figueredo E, Picanço Júnior P, dos Reis Filho AA (2021) Cropland expansion as a driver of land-use change: the case of Cerrado-Caatinga transition zone in Brazil. **Environment, Development and Sustainability.** https://doi.org/10.1007/s10668-021-01387-z

Farhate CVV, De Souza ZM, De Medeiros Oliveira SR, Tavares RLM, Carvalho JLN (2018) Use of data mining techniques to classify soil CO2 emission induced by crop management in sugarcane field. **PLoS One**, 13:1-18. https://doi.org/10.1371/journal.pone.0193537

Ferraz de Almeida R, de Bortoli Teixeira D, Montanari R, Bolonhezi AC, Teixeira EB, Moitinho MR, Panosso AR, Spokas KA, La Scala Júnior N (2018) Ratio of CO2 and O2 as index for categorising soil biological activity in sugarcane areas under contrasting straw management regimes. **Soil Research** ;56:373. https://doi.org/10.1071/sr16344

Freitas LPS, Lopes MLM, Carvalho LB, Panosso AR, La Scala Júnior N, Freitas RLB, Minussi CR, Lotufo ADP (2018). Forecasting the spatiotemporal variability of soil CO2 emissions in sugarcane areas in southeastern Brazil using artificial neural networks. **Environmental Monitoring and Assessment** ;190. https://doi.org/10.1007/s10661-018-7118-0

Ghorbani MA, Shamshirband S, Zare Haghi D, Azani A, Bonakdari H, Ebtehaj I. (2017). Application of firefly algorithm-based support vector machines for prediction of field capacity and permanent wilting point. **Soil and Tillage Research,** 172:32-8. https://doi.org/10.1016/j.still.2017.04.009

Giacomo G, Angelo F, Fabio B, Stefano B, Riccardo M (2014). Measurements of soil carbon dioxide emissions from two maize agroecosystems at harvest under different tillage conditions. **The Scientific World Journal** https://doi.org/10.1155/2014/141345

Gourlez de la Motte L, Jérôme E, Mamadou O, Beckers Y, Bodson B, Heinesch B, Aubinet M (2016). Carbon balance of an intensively grazed permanent grassland in southern Belgium. **Agricultural and Forest Meteorology**, 228-229:370-83. https://doi.org/10.1016/j.agrformet.2016.06.009

Gruba P, Mulder J (2015). Tree species affect cation exchange capacity (CEC) and cation binding properties of organic matter in acid forest soils. **Science of the Total Environment** ;511:655-62. https://doi.org/10.1016/j.scitotenv.2015.01.013

Grossiord C, Mareschal L, Epron D (2021). Transpiration alters the contribution of autotrophic and heterotrophic components of soil CO2 efflux. **New Phytologist** ;194:647-53. https://doi.org/10.1111/j.1469-8137.2012.04102.x

Han G, Sun B, Chu X, Xing Q, Song W, Xia J (2018). Precipitation events reduce soil respiration in a coastal wetland based on four-year continuous field measurements. **Agricultural and Forest Meteorology** ; 256-257:292-303. https://doi.org/10.1016/j.agrformet.2018.03.018

Haykin SS (2001). Redes Neurais: Princípios e prática. 2. ed. Porto Alegre, Rs.

Horst-Heinen TZ, Dalmolin RSD, ten Caten A, Moura-Bueno JM, Grunwald S, Pedron F de A, Rodrigues MF, Rosin NA, da Silva-Sangoi DV (2021) Soil depth prediction by digital soil mapping and its impact in pine forestry productivity in South Brazil. **Forest Ecology and Management** ;488. https://doi.org/10.1016/j.foreco.2021.118983

Hipni A, El-shafie A, Najah A, Karim OA, Hussain A, Mukhlisin M. Daily(2013). Forecasting of Dam Water Levels: Comparing a Support Vector Machine (SVM) Model With Adaptive Neuro Fuzzy Inference System (ANFIS). **Water Resources Management** ;27:3803-23. https://doi.org/10.1007/s11269-013-0382-4

Huang N, Wang Li, Song XP, Andrew Black T, Jassal RS, Myneni RB, Wu C, Wang Lei, Song W, Ji D, Yu S, Niu Z (2020). Spatial and temporal variations in global soil respiration and their relationships with climate and land cover. **Science Advances** ;6:1-12. https://doi.org/10.1126/sciadv.abb8508

Islam KR, Weil RR (2000) Land use effects on soil quality in a tropical forest ecosystem of Bangladesh. **Agriculture, Ecosystems & Environment** 79:9-16. https://doi.org/10.1016/S0167-8809(99)00145-0

Jafarzadeh AA, Pal M, Servati M, FazeliFard MH, Ghorbani MA (2016). Comparative analysis of support vector machine and artificial neural network models for soil cation exchange capacity prediction. **International Journal of Environmental Science and Technology;** 3:87-96. https://doi.org/10.1007/s13762-015-0856-4

Jang JR. ANFIS (1993) Adap tive-Ne twork-Based Fuzzy Inference System;23.

Jassal RS, Black TA, Nesic Z, Gaumont-Guay D (2012) Using automated non-steady-state chamber systems for making continuous long-term measurements of soil CO 2 efflux in forest ecosystems. **Agricultural and Forest Meteorology**; 161:57-65. https://doi.org/10.1016/j.agrformet.2012.03.009

Kalogirou S, Neocleous C, Pashiardis S, Schizas C, Box PO(1999). **Wind Speed Prediction Using Artificial Neural Networks** * Ministry of Agriculture , Natural Resources and Environment. Complement.

Kassambara, A (2017). Mundt, F. Factoextra: Extract and visualize the results of multivariate data analyses. R package version 1

Khosravi A, Machado L, Nunes RO (2018). Estimation of density and compressibility factor of natural gas using artificial intelligence approach. **Journal of Petroleum Science and Engineering;** 168:201-16.https://doi.org/10.1016/j.petrol.2018.05.023

Kooch Y, Ghaderi E (2021).Soil function can sensitively respond to different canopy composition of Crataegus and Berberis. **Applied Soil Ecology** ;167:104112. https://doi.org/10.1016/j.apsoil.2021.104112

Laib A, Melit M, Nekhoul B, Boudjefdjouf H, Bouchekara H (2017). Localisation of faults in wiring networks using time domain reflectometry and adaptive neuro-fuzzy inference system. **Electronics Letters** ;53:600-2. https://doi.org/10.1049/el.2017.0566

Lal R (2009). Challenges and opportunities in soil organic matter research. **European Journal of Soil Science** ;60:158-69. https://doi.org/10.1111/j.1365-2389.2008.01114.x

Landeras G, Ortiz-Barredo A, López JJ.(2008). Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). **Agricultural Water Management**;95:553-65. https://doi.org/10.1016/j.agwat.2007.12.011

La Scala N, Marques J, Pereira GT, Corá JE.(2000). Carbon dioxide emission related to chemical properties of a tropical bare soil. **Soil Biology and Biochemistry** ;32:1469-73. https://doi.org/10.1016/S0038-0717(00)00053-5

La Scala N, Panosso AR, Pereira GT (2003).Modelling short-term temporal changes of bare soil CO2 emissions in a tropical agrosystem by using meteorological data. **Applied Soil Ecology** ;24:113-6. https://doi.org/10.1016/S0929-1393(03)00065-9

Lê, Sébastien; Josse, Julie; Husson, François (2008). FactoMineR: an R package for multivariate analysis. **Journal of statistical software**, 25,1, 1-18p.

Liang L, Peng S, Sun J, Chen L, Cao Y(2010). Estimation of annual potential evapotranspiration at regional scale based on the effect of moisture on soil respiration. **Ecological Modelling** ;221:2668-74.

https://doi.org/10.1016/j.ecolmodel.2010.08.010

Liaw A, Wiener M(2002). Classification and Regression by randomForest. R News ;2:18-22.

Liu YR, Delgado-Baquerizo M, Wang JT, Hu HW, Yang Z, He JZ.(2018). New insights into the role of microbial community composition in driving soil respiration rates. **Soil Biology and Biochemistry** ;118:35-41. https://doi.org/10.1016/j.soilbio.2017.12.003

Luo, Y., Zhou, X. (2006). Soil respiration a 486 nd the environment. San Diego: Academic Press. 316p.

Moitinho MR, Teixeira DDB, Bicalho E da S, Panosso AR, Ferraudo AS, Pereira GT, Tsai SM, Borges BMF(2021). La Scala N. Soil CO2 emission and soil attributes associated with the microbiota of a sugarcane area in southern Brazil. Scientific Reports. https://doi.org/10.1038/s41598-021-87479-2

Muraoka H, Noda HM, Nagai S, Motohka T, Saitoh TM, Nasahara KN, Saigusa N (2013). Spectral vegetation indices as the indicator of canopy photosynthetic productivity in a deciduous broadleaf forest. **Journal of Plant Ecology.** 6:393-407. https://doi.org/10.1093/jpe/rts037

Neira J, Ortiz M, Morales L, Acevedo E (2015) Oxygen diffusion in soils: Understanding the factors and processes needed for modeling. **Chilean Journal of Agricultural Research**. 75:35-44. https://doi.org/10.4067/S0718-58392015000300005

Oertel C, Matschullat J, Zurba K, Zimmermann F, Erasmi, S (2016) Greenhouse gas emissions from soils—A review. Chemie der Erde—**Geochemistry.** https://doi.org/10.1016/j.chemer.2016.04.002

Panosso AR, Marques J, Milori DMBP, Ferraudo AS, Barbieri DM, Pereira GT, La Scala N (2011). Soil CO2 emission and its relation to soil properties in sugarcane areas under Slash-and-burn and Green harvest. **Soil and Tillage Research**;111:190-6. https://doi.org/10.1016/j.still.2010.10.002

Panosso AR, Perillo LI, Ferraudo AS, Pereira GT, Vivas Miranda JG, La Scala N. (2012). Fractal dimension and anisotropy of soil CO 2 emission in a mechanically harvested sugarcane production area. **Soil and Tillage Research**;124:8-16. https://doi.org/10.1016/j.still.2012.04.005

Panosso AR, Ribeiro CER, Zanini JR, Pavani LC, Pereira GT, La Scala N (2009) Variabilidade espacial da emissão de CO2, da temperatura e umidade de um latossolo desprovido de vegetação sob diferentes lâminas de molhamento. **Semina Ciencias Agrarias.**https://doi.org/10.5433/1679-0359.2009v30n4Sup1p1017

Pulrolnik K, de Barros NF, Silva IR, Novais RF, Brandani CB. (2009).Carbon and nitrogen pools in soil organic matter under eucalypt, pasture and savanna vegetation in Brazil. **Sociedade Brasileira de Ciência do Solo** ;33:1125-36. https://doi.org/10.1590/s0100-06832009000500006

Raich JW, Schlesinger WH (1992) The global carbon-dioxide flux in soil respiration and its relationship to vegetation and climate. **Tellus B: Chemical and Physical Meteorology.** https://doi. org/10.1034/j.1600-0889.1992.t01-1-00001.x

Raij B, Andrade JC, Cantarella H, Quaggio JA.(2001).Análises químicas para avaliação da fertilidade de solos tropicais. 1a Edição Campinas, Instituto Agronômico de Campinas 284p.

Ramos, J.C.A. **Emissão de CO2, quantidade e qualidade do carbono do solo em sistemas agrícolas na região do cerrado do Mato Grosso do Sul.** 2018. 83 f.Dissertação de Mestrado (Agronomia). Universidade Estadual Paulista, Faculdade de Engenharia, Ilha Solteira, 2018.

Sena NC, Veloso GV, Lopes AO, Francelino MR, Fernandes-Filho EI, Senra EO, Silva Filho LA da, Condé VF, Silva DL de A, Araújo RW de (2021)Soil sampling strategy in areas of difficult acess using the cLHS method. **Geoderma Regional** ;24. https://doi.org/10.1016/j.geodrs.2020.e00354

Sotta ED, Meir P, Malhi Y, Nobre AD, Hodnett M, Grace J (2004) Soil CO2 efflux in a tropical forest in the central Amazon. **Global Change Biology** https://doi.org/10.1111/j.1529-8817.2003.00761.x

Shi T, Cui L, Wang J, Fei T, Chen Y, Wu G (2013). Comparison of multivariate methods for estimating soil total nitrogen with visible/near-infrared spectroscopy. **Plant Soil.** 366:363-75. https://doi.org/10.1007/s11104-012-1436-8

Shu C, Ouarda TBMJ(2008).Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. **Journal of Hydrology**; 349:31-43. https://doi.org/10.1016/j.jhydrol.2007.10.050

Silva BO, Moitinho MR, Santos GAA, Teixeira DDB, Fernandes C, La Scala N (2019) Soil CO2 emission and short-term soil pore class distribution after tillage operations**.** **Soil and Tillage Research** 186: 224–232. https://doi.org/10.1016/j.still.2018.10.019

Song X, Peng C, Zhao Z, Zhang Z, Guo B, Wang W, Jiang H, Zhu Q (2014).Quantification of soil respiration in forest ecosystems across China. **Atmospheric Environment** ;94:546-51.https://doi.org/10.1016/j.atmosenv.2014.05.071

De Souza Rolim G, De Camargo MBP, Grosselilania D, De Moraes JFL (2007). Climatic classification of köppen and thornthwaite sistems and their applicability in the determination of agroclimatic zonning for the state of São Paulo, Brazil. **Bragantia;** 66:711-20. https://doi.org/10.1590/s0006-87052007000400022

Tavares RLM, de Souza ZM, La Scala N, Castioni GAF, de Souza GS, Torres JLR (2016). Spatial and temporal variability of soil CO2 flux in sugarcane green harvest systems. **Revista Brasileira de Ciência do Solo** ;40:1-14. https://doi.org/10.1590/18069657rbcs20150252

Teimouri R, Baseri H (2014).Optimization of magnetic field assisted EDM using the continuous ACO algorithm. A**pplied Soft Computing - Journals** ;14:381-9. https://doi.org/10.1016/j.asoc.2013.10.006

Vendruscolo DGS, Chaves AGS, Medeiros RA, Silva RS da, Souza HS, Drescher R, Leite HG (2017). Estimativa da altura de árvores de Tectona grandis L.f. utilizando regressão e redes neurais artificiais. **Nativa** . 5:52-8. https://doi.org/10.5935/2318-7670.v05n01a09

Vicentini ME, Pinotti CR, Hirai WY, de Moraes MLT, Montanari R, Filho MCMT, Milori DMBP, Júnior NLS, Panosso AR (2019) . CO2 emission and its relation to soil temperature, moisture, and O2 absorption in the reforested areas of Cerrado biome, Central Brazil. **Plant Soi;** 444:193-211. . https://doi.org/10.1007/s11104-019-04262-z

Wang WJ, Dalal RC, Moody PW, Smith CJ (2003). Relationships of soil respiration to microbial biomass, substrate availability and clay content. Soil Biology and Biochemistry, 35:273-84. https://doi.org/10.1016/S0038-0717(02)00274-2

Wang, Y.; Witten, I. H (1996). Induction of model trees for predicting continuous classes. WorkingPaper 96/23. Available:https://researchcommons.waikato.ac.nz/bitstream/handle/10289/1183/uow-cs-wp-1996-23.pdf?sequence=1

Ward D, Kirkman K, Hagenah N, Tsvuura Z. (2017).Soil respiration declines with increasing nitrogen fertilization and is not related to productivity in long-term grassland experiments. **Biology and Biochemistry** ;115:415-22. https://doi.org/10.1016/j.soilbio.2017.08.035

Wu X, Xu H, Tuo D, Wang C, Fu B, Lv Y, Liu G (2020). Land use change and stand age regulate soil respiration by influencing soil substrate supply and microbial community. **Geoderma** ;359:113991. https://doi.org/10.1016/j.geoderma.2019.113991

Yan J, Zhang X, Liu J, Li H, Ding G (2020). MODIS-Derived estimation of soil respiration within five cold temperate coniferous forest sites in the eastern Loess Plateau, China. **Forests** ;11. https://doi.org/10.3390/f11020131

Yuan ZQ, Jiang XJ, Liu GJ, Jin HJ, Chen J, Wu QB (2019). Responses of soil organic carbon and nutrient stocks to human-induced grassland degradation in a Tibetan alpine meadow. **Catena.** 2019;178:40-8. https://doi.org/10.1016/j.catena.2019.03.001

Zhao J, Lange H, Meissner H.(2020). Gap-filling continuously-measured soil respiration data: A highlight of time-series-based methods. **Agricultural and Forest Meteorology** ;285-286:107912. https://doi.org/10.1016/j.agrformet.2020.107912

Zhao X, Liang N, Zeng J, Mohti A (2020). A simple model for partitioning forest soil respiration based on root allometry. **Biology and Biochemistry** ;152:108067. https://doi.org/10.1016/j.soilbio.2020.108067