

**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

**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).

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 (Participação Virtual)
Departamento de Engenharia e Ciências Exatas (DECEX) / FCAV / UNESP - Jaboticabal

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

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

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

Prof. Dr. GENE TADEU PEREIRA (Participação Virtual)
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>i</i>
ABSTRACT	<i>iv</i>
RESUMO	<i>vi</i>
LIST OF ABBREVIATIONS AND ACRONYMS	<i>viii</i>

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

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

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 CO₂ emission from soil (FCO₂), but few studies have modeled the temporal dynamics of soil O₂ influx (FO₂). 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 FCO₂ and FO₂ 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 FCO₂ and FO₂ 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 FCO₂. The general regression neural network (GRNN) with 10 input variables explained temporal variability in FO₂ with an R^2 of 56%. For eucalyptus, in the estimation of FCO₂, the best predictive performance was obtained with MLP with validation ($R^2 = 0.59$; RMSE = 1.034 $\mu\text{mol m}^{-2}\text{s}^{-1}$). FO₂ estimation: validation ($R^2 = 0.36$; RMSE = 0.076 $\text{mg m}^{-2}\text{s}^{-1}$). 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 FCO₂; validation ($R^2 = 0.53$; RMSE = 0.990 $\mu\text{mol m}^{-2}\text{s}^{-1}$). In Native Species areas, the best results were: FCO₂ with Radial Basis Function Neural Network (RBFNN) ($R^2 = 0.54$, RMSE = 1.015 $\mu\text{mol m}^{-2}\text{s}^{-1}$) and FO₂ with RBFNN ($R^2 = 0.74$, 0.079 $\text{mg m}^{-2}\text{s}^{-1}$). Estimates of FCO₂ showed better predictive performance than FO₂. RBFNN was best estimate for FCO₂. MLPNN is the best architecture for FO₂ ($R^2 = 0.45$, RMSE = 0.94 $\text{mg m}^{-2}\text{s}^{-1}$). In relation to ANFIS, FO₂, 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 = $1,034 \mu mol m^{-2} s^{-1}$). 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 foram observados para FCO_2 ; validação ($R^2 = 0,53$; RMSE = $0,990 \mu mol m^{-2} s^{-1}$). 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 $\mu\text{mol m}^{-2}\text{s}^{-1}$) e FO_2 com RBFNN ($R^2 = 0,74$; RMSE 0,079 $\text{mg m}^{-2}\text{s}^{-1}$). 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 $\mu\text{mol m}^{-2}\text{s}^{-1}$). Em contrapartida a MLPNN foi a melhor arquitetura para o FO_2 ($R^2 = 0,45$, RMSE = 0,94 $\text{mg m}^{-2}\text{s}^{-1}$). 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
CO₂ - Carbon dioxide
Cstock - Carbon stocks
d - Index of Agreement developed by Willmott
ET_o - Evapotranspiration reference
EVI - Enhanced Vegetation Index
FCO₂ – soil CO₂ emission
FO₂ - influx of O₂ 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 (tCO₂) into the atmosphere (Alencar et al., 2020), and more than 12 thousand kilometers (km²) 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 CO₂ 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, CO₂ emission or flux (FCO₂) 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 (FO₂), 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, CO₂ and O₂ 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 O₂ 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 CO₂ emissions remains limited due to the complexity of the phenomenon. In addition, most studies focus only on the study of FCO₂, 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).

1.2 – Hypothesis

The application of machine learning techniques: ANNs, SVR, ANFIS, and RF, generates predictive models to explain the pattern of temporal variability of soil respiration in planted forest areas using environmental covariates.

1.3 – Objectives

The objective of this study was to evaluate the potential of four machine learning techniques (ANN, SVR, ANFIS, and RF) to estimate soil CO₂ emissions and oxygen influx in a planted forest area in the Cerrado using environmental covariates. Initially, we developed models of CO₂ emission and oxygen uptake using machine learning techniques for pine, eucalyptus, and native species planted in forest areas (Chapters 2, 3, and 4). Subsequently, we developed a global model considering the database of three areas for CO₂ emissions and oxygen consumption (Chapter 5).

1.4 Literature Review

1.4.1 Areas of planted forest

According to information from the Global Assessment of Forest Resources Report (Fan et al., 2020) produced by the United Nations (FAO, 2020), natural regenerating forests have suffered significant losses for 120 years, while planted forest areas have thrived. Currently, 93% of the forest area worldwide is composed of naturally regenerating forests and 7% corresponds to planted forests, totaling 123 million hectares. More than 54% of the world's forests are in just five countries: the Russian Federation (815 million hectares), Brazil (457 million hectares), Canada (347 million hectares), the United States of America (310 million hectares), China (220 million hectares), and the rest of the world (1,870 million hectares).

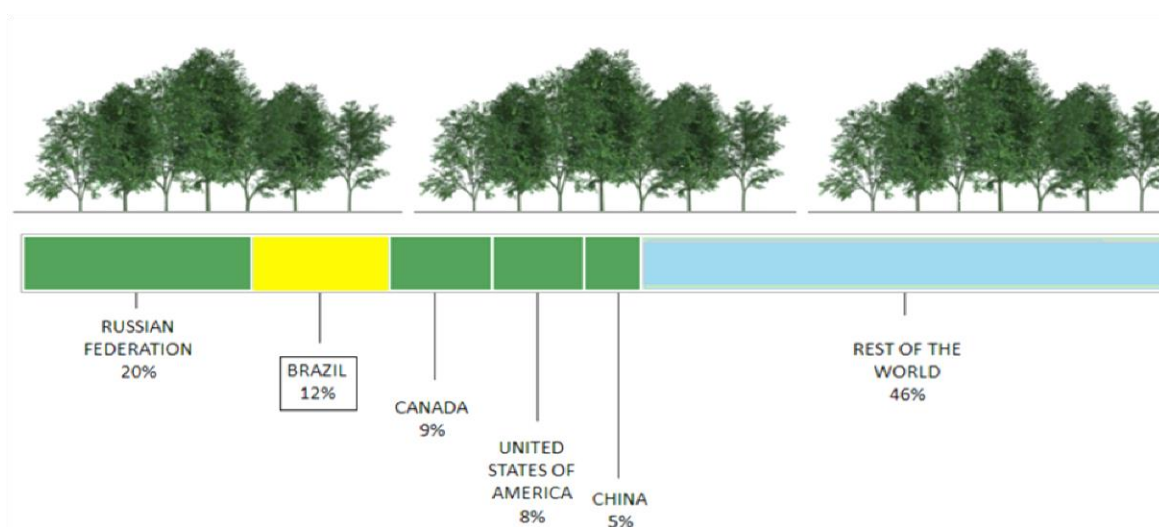


Figure 1. Forests around the world [adapted]. FAO 2020 Global Forest Resources Assessment 2020 (Rome: FAO) (available at: <https://doi.org/10.4060/ca8753en>)

In Brazil, the commercial use of forest products has been gaining prominence since the 1960s, when the federal government began granting fiscal incentives for the increase of the forest sector, aiming to meet the needs of the pulp and paper, steel, and furniture industries that required wood as a raw material (Barborsa et al., 2020).

According to recent information from the Brazilian Tree Industry (IBÁ, 2020), in 2019, 9.0 million hectares of trees were planted in the country, of which 77% were eucalyptus (6.97 million hectares), 18% were pine forest (1.64 million hectares), and the rest were species such as rubber trees, acacia, and teak. The states of Minas Gerais (MG), São Paulo (SP), Mato Grosso do Sul (MS), Paraná (PR), Rio Grande do Sul (RS), and Santa Catarina (SC) are the largest holders of planted forests in Brazil. According to the Brazilian Institute of Geography and Statistics (IBGE, 20021),

through the IBGE news agency portal, 44.3% of the eucalyptus areas are concentrated in the southeast and 84.6% of the pine forests are in the south,

1.5 Soil Respiration

Soil Respiration (R_s) is one of the most significant components of the carbon (C) cycle in forest ecosystems and represents the major flux of CO_2 from terrestrial ecosystems (Adachi et al., 2006) to the atmosphere; however, its temporal and spatial variability as well as the driving forces surrounding this process are still not well understood (Huang et al., 2020), especially for tropical forests (Gutiérrez del Arroyo and Wood, 2020). According to (Anderson, 2015) R_s has been widely used to assess the overall activity of belowground biomass since it is influenced by climate, physical and chemical properties, soil management practice and agricultural practices. That author also indicated that R_s can be used as a bioindicator to assess (C) mineralization and stabilization in relation to the type of organic matter and its rate of entry into the soil. R_s is the primary pathway by which CO_2 that has been fixed by plants returns to the atmosphere. Therefore, quantifying the temporal variability of soil respiration is important for the development of models that aid in the understanding of ecosystem carbon loss at regional and global scales to aid in understanding the mechanisms that control such losses, advancing the understanding of the terrestrial C cycle (Phillips et al., 2017; Rubio and Detto, 2017).

1.5.1 Soil properties

According to Cable et al.(2008) heterotrophic respiration (R_h) represents the loss of stored C through the decomposition of SOM, and autotrophic respiration (R_a) indicates the loss of C that was fixed during photosynthesis. Generally, R_s varies with time and space, with environmental factors such as temperature and water content as the main influencers (Reth et al., 2005; de Araújo Santos et al., 2019). Although the dynamics of these two variables are generally well understood, the complexity of R_s alludes to other factors that act directly or indirectly in the process of production and transport of CO_2 flux, including carbon stock (Panosso et al., 2011), soil chemical attributes such as CEC and iron oxides (La Scala et al., 2000),

physical attributes such as soil density and pore classes (Teixeira et al., 2012; Silva et al., 2019) and microbial activity (Xavier et al., 2019).

Studies have described the relationship between soil respiration and root biomass (Fang et al., 1998) and soil oxygen concentration (Ferraz de Almeida et al., 2018) reflecting CO₂ fluxes. As Lagomarsino et al. (2006) observed, changes in the quantity and quality of labile C substrates play an important role in the seasonal variation of FCO₂. Estruch et al. (2020), highlighted the influence of plant species on FCO₂ through specific effects on soil moisture, temperature, and microbial communities, as well as plant-soil system interactions with burlap production and root exudates, which provide improvements in autotrophic and heterotrophic soil respiration. Pell et al. (2005) described R_s as a fundamental process for carbon flux to the atmosphere, with soil SWC, oxygen concentration, and carbon bioavailability as the main regulating dynamics.

1.5.2 Soil aeration

Gas emissions in soil are driven by the production and consumption processes, along with the pore structure of the soil and the physical processes of gas transport (Laemmle et al., 2019). According to Fan et al. (2020) in soil, the movement of gas and water is influenced not only by the number of pores, but also by their distribution, size, and other geometric characteristics and indirectly by the clay and organic matter content (Eden et al., 2011). Soil aeration is an important aspect of the physical quality of the soil. This is since the vast majority of biological activity in soil requires a continuous supply of oxygen for organisms (Czyż, 2004). Watanabe et al. (2008) observed that physical quality for plant growth is dependent on soil moisture dynamics.

Adachi et al. (2006) observed that soil respiration is negatively correlated with soil water content in tropical forest soils because of the higher water content, as liquid-filled pores hinder gas exchange with the atmosphere. Diffusion and mass flow are primarily responsible for the movement of CO₂ in the soil and its release into the atmosphere, and diffusion is required for oxygen influx (Neira et al., 2015).

Maier et al. (2019) affirmed

“(...) Gas transport in the soil and across the soil-atmosphere boundary is assumed to be entirely driven by diffusive transport; therefore, air mass flux due to pressure gradients or other transport processes (boiling or plant-mediated gas transport) is considered negligible” ...

Under a steady state condition, gas diffusion follows Fick's First Law, Equation 1 (Eq.1), as described by (Román et al., 2021)

$$q = -D_p \left(\frac{\partial C}{\partial x} \right) \quad (\text{Eq 1}),$$

where q is the gas flux ($\text{cm}^3 \text{ air cm}^{-2} \text{ soils}^{-1}$), D_p ($\text{cm}^3 \text{ air cm}^{-2} \text{ soils}^{-1}$) is the effective diffusion coefficient of the gas in the soil, and $\partial C / \partial x$ is the gas concentration gradient ($\text{cm}^3 \text{ air cm}^{-1} \text{ soil}$). (Román et al., 2021).

The gas diffusivity (D_p / D_0) corresponds to the ratio of D_p to the diffusion coefficient of gas in free air at atmospheric pressure (D_0 , $\text{cm}^2 \text{ s}^{-1}$), and p is a function of the air-filled porosity related to transport efficiency, continuity-tortuosity, and the constraints on the soil pore matrix. Recently, some important studies have shown the need to understand oxygen influx to explain the dynamics of organic matter in soil (Ferraz de Almeida et al., 2018; Vicentini et al., 2019).

1.5.3 Soil respiration associated photosynthesis

Han et al. (2014) noted that on a diurnal scale, soil respiration can be modulated by net photosynthesis (difference between gross photosynthesis and the sum of mitochondrial and photorespiration) and different vegetation types contribute to variations. Gutiérrez del Arroyo and Wood (2020), reported that other biophysical factors, such as photosynthetically active radiation (PAR) and relative humidity (RH), play an important role in R_s . Liu et al. (2020) found that R_s plays a specific role in adjusting crop photosynthesis, but that it also increases CO_2 emissions by contributing greenhouse gases crop photosynthesis, but that it also increases CO_2

emissions by contributing greenhouse gases. The production of FCO_2 is a result of soil biochemical activities and the transport of the gas to the atmosphere (Silva et al., 2019), both of which are influenced by the aerobic and anaerobic conditions of the area (Fan et al., 2020). Siqueira et al. (1994) reported that aerobic microorganisms use oxygen as the final electron receptor in soil biochemical processes. Both soil metabolic processes such as aerobic respiration by microorganisms, respiration, and root growth require well-structured soils.

1.5.4 Soil respiration and the climate

Soil respiration is also influenced by photosynthesis, and precipitation regimes control this effect (Thomey et al., 2011). For Reichstein et al. (2003) and Xiao et al (2020), R_s is sensitive to climate warming and precipitation change in various ecosystems, and the importance of precipitation to R_s occurs through two pathways, the first refers to the direct influence through water limitation of soil microbes and the second relates to the indirect influence via root productivity.

1.6 Machine Learning (ML)

The American Arthur Samuel was a pioneer in computer games and artificial intelligence. In 1959, he defined machine learning (ML) as a field of study giving computers the ability to learn without being explicitly programmed (Awad; Khanna, 2015). Machine learning consists of a subset of artificial intelligence that relates directly to algorithms and methods endowed with the ability to learn from a database (Figure 2). Although this mode of cognition differs from the human cognitive process, its fundamentals are identical. By experiencing errors, it is possible to improve a given response, that is, learning is obtained only through experience (Azevedo, 2018).

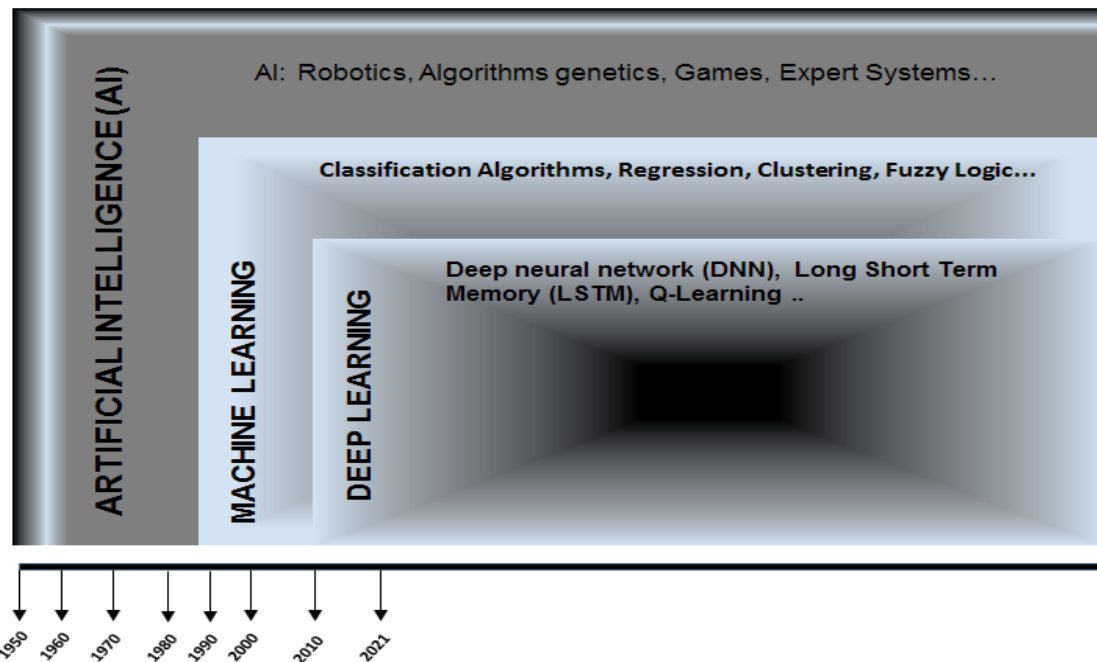


Figure 2. Relationship between artificial intelligence, machine learning, deep learning and data science.

With machine learning, algorithms are developed that can learn to solve two types of problems: classification when discrete results are preferred and regression when continuous results are preferred. Zhang et al. (2020) indicated that “ML aims to establish a regressor or classifier by learning from the training set and then evaluating the performance of the regressor or classifier through a test set.” For learning to be effective, the algorithms are subjected to training processes in which a series of information is analyzed, allowing the development of learning models to solve new situations and/or instances (Carvalho, 2012).

According to Santos (2016), the aforementioned procedures are conducted in two distinct ways, i.e., unsupervised, when information is combined based on similarity, or supervised, in which labeled examples are employed (e.g., example inputs and their desired outputs). Zhang et al. (2020) exemplified: “*supervised learning as a teacher exposing a problem (finding the mapping relationship between inputs and outputs) and its solutions (labeled output data) to a student and asking that student to figure out how to solve other similar problems*”. According to the author, unsupervised learning, which consists of unlabeled datasets, is similar to giving a student a set of patterns and asking the student to determine the underlying reasons generating the patterns. Methodologies that consider machine learning

provide excellent results for many different fields of science and are of significant value to research being conducted around the world (Rocha et al., 2018)

According to Padarian et al. (2020), the modeling and validation process can be used to formulate a model for explaining a soil distribution, thereby improving our understanding of the causes of soil variation and informing the environmental variables that control the soil distribution.

The application of ML techniques such as artificial neural networks (ANNs), support vector regression (SVR), an adaptive neuro fuzzy inference system (ANFIS), and random forest, have been applied to soil science during the last decade. In studies conducted in the areas of agriculture and native forests. Rivera and Bonilla (2020), estimated the aggregate soil stability in the Mediterranean using two machine learning techniques: an ANN model and a generalized linear model. The authors observed a superior performance when applying an ANN and concluded that the model was reliable for predicting the aggregate stability.

Ghorbani et al. (2015) estimated the cation exchange capacity (CEC) in Iran in areas of native forests, reforested regions, pastures, and other land areas using an ANNs multilayer perceptron (MLPNN), radial basis function (RBFNN), and ANFIS, and compared the results to those of traditional multiple regression. The results showed that the accuracies of the ANN and ANFIS models were similar with respect to their statistical parameters. The authors pointed out that machine learning techniques are powerful tools that can be used to predict CEC, particularly in arid grassland ecosystems.

According to Chen et al. (2017), a review of the literature showed that different machine learning models have particular strengths and weaknesses, and their behavior generally depends on the characteristics of the different study areas. For example, in a supervised study comparing linear and Gaussian support vector regression models with ordinary least squares (OLS) and a linear ridge regression (LRR) to estimate the different physical soil properties from a dataset of chemical properties. Kovačević et al. (2010) found that the performances of the models are similar, except for a regression model applied to the pH, where both SVR kernel functions significantly outperformed those of linear regression. According to these authors, the low correlation between the chemicals and physics in the dataset compromised the model performance.

1.6.1 Neural Networks

Artificial neural networks are inspired by the biological nervous system and are commonly used in regression and classification problems. Minasnye and McBratney (2002) describe an ANN as an attempt to build a mathematical model that aims to function analogously to the human brain. ANNs are increasingly used as valuable approaches to the modeling, prediction, and optimization of complex processes (Sardoo et al. 2020). As vigorous data modeling tools, they gain knowledge through experience. According to Yilmaz et al. (2011), a trained ANN can be considered an “expert” in the category of the information it was given for analysis. In addition Emadi et al. (2020) reported that networks are able to detect patterns and elicit results; therefore, they can be used to predict data with correlations such as the soil properties.

According to Debanshi and Pal (2020),

“Each unit of a neural network is entrusted as a layer and remains connected through unidirectional flows of information, from the input layer to the layer located as the hidden layer and finally to the output layer. Thus, the first input layer generates a signal value responding to an element that is followed by the yielding of an output value from each neuron, which can act as an input value to the subsequent layer”.

The following diagram illustrates a neural network with three layers. The input and hidden layers contain three nodes, and the output layer contains a single node (Milon, Imdadul Haque, 2019). Here, w represents the weights assigned to the synapses between the input layer and the hidden layer and $O1$ is the output of the network (Figure 3).

According to (Haykin, 2001) the architecture is closely linked with the learning algorithm used to train the network. For Braga (2000), defining the ideal number of neurons in a hidden layer is extremely important owing to its computational complexity and generalization capacity. Neural networks consist of a large class of architectures. The most widely used architectures in the literature for classification or regression problems are a multilayer perceptron neural network (MLPNN), radial

basis function (RBFNN), and generalized regression neural network. (Kashi et al., 2014).

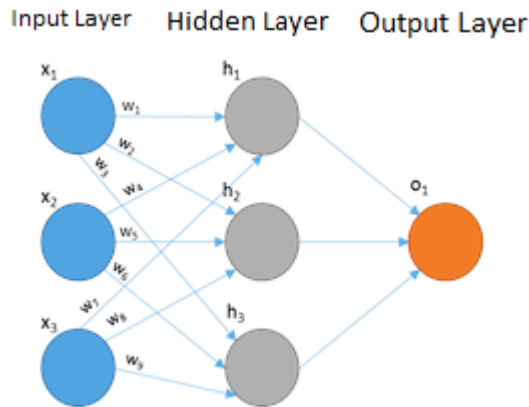


Figure 3. Artificial neural network architecture

1.6.1.1 Multilayer Perceptron Neural Network (MLPNN)

The most commonly used neural model for prediction is an MLPNN, which in engineering problems consists of three layers, i.e., an initial input layer, a final output layer, and a hidden intermediate layer, and in the case of nonlinear mapping is called a universal approximator. Such networks are feedforward-type networks, that is, no output from a neuron in a layer k will be an input signal to a neuron in a layer less than or equal to k . In addition, they are completely connected, and each neuron therefore provides its output to each unit in the next layer (Han et al., 2005). This type of network uses a supervised learning process, most commonly by applying the back-propagation algorithm.

1.6.1.2 Radial Basis Function Neural Network (RBFNN)

RBFNNs are alternatives to MLPNNs, consisting of three layers: an input layer, a hidden layer (kernel), and an output layer. The nodes of each layer are fully connected to the previous layer. The radially symmetric basis function is used as an activation function for the hidden nodes. The transformation from input to hidden nodes is non-linear, and the training of this part of the network is usually conducted in

an unsupervised manner (Bagheri et al., 2015). The training of the network parameters (weight) between the hidden and output layers occurs in a supervised fashion based on target outputs (Kashaninejad et al., 2009).

1.6.1.3 Generalized Regression Neural Network (GRNN)

The technique is based on a probabilistic model, trains quickly, and allows the modeling of nonlinear functions. The topology of the GRNN developed by Specht consists of four layers: an input layer, a pattern layer, a summation layer, and an output layer (F.Specht, 2007). The input layer collects information and transmits it to the pattern layer. The pattern layer is used to apply clustering during the training process. Typically, the number of neurons in the pattern layer is equal to the number of training datasets (Ghritlahre and Prasad, 2018). This class of network paradigm has the distinctive features of learning quickly, working with simple and straightforward training algorithms, and being discriminative against infrequent outliers and erroneous observations (Leung et al., 2000). ANNs have been successfully employed to predict soil properties (Minasny and Mcbratney, 2002.)

1.7 Regression with Support Vectors

Support vector machines (SVM) can be divided into two main techniques, namely, classification (SVC) and regression (SVR), which are based on the theory developed by (Vapnik et al., 1995). SVMs are basically a linear machine, aiming at the theory of machine learning techniques that seek to maximize generalization ability. An SVM was initially used to solve problems with binary linear classification. However, the concept has been expanded, allowing it to be adopted to regression, multiclass classification, information clustering (unsupervised learning), and nonlinear classification problems (Filgueiras, 2014).

An SVM installs a learning algorithm relevant to understanding more refined patterns in a complex information set. Through the algorithm, discriminative classification is applied to the development of various activities, such as predicting the classification of information that has not been seen before. The construction of an SVM starts from the training complexity, which is cubic or quadratic as a function of

the number of samples used in training, which is one of the main disadvantages of this type of algorithm (Lafeta et al., 2020).

SVR aims to find a function that approximates a given set of information. With this, it seeks to find a function $f(x)$ responsible for approximating a given element with less than a previously established prediction error. In addition, it simultaneously reduces the margin reciprocal and allows the insertion of variables with slack (Koide, 2016), as shown in Figure 4. SVR seeks to minimize an upper bound on the generalization error rather than minimize the training error, which is the principle followed by ANNs (Wu et al., 2010). The main idea is to build a hyperplane as a decision surface to maximize the margin of separation between the positive and negative points is maximum (Lin, 2018) SVR is a promising new approach to data regression.

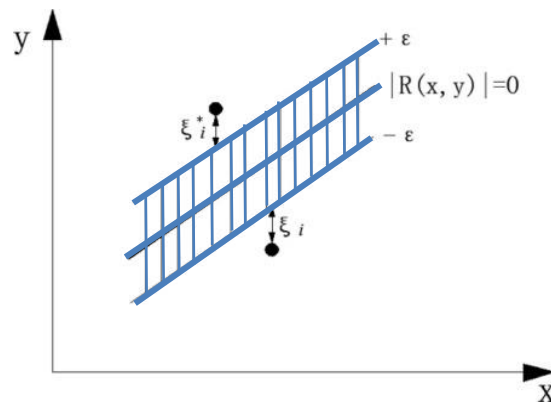


Figure 4. Graphical representation of a linear regression function (Chen et al., 2017),

SVR performance is highly dependent on kernel functions. Among the most used kernel functions in the SVR algorithm stand out the kernel: linear, polynomial, sigmoidal and Gaussian.

Table 1. Support vector machine and kernel function

Kernel	Function $K(x_i, x_j)$	Parameters
Linear kernel	(x_i, x_j)	-
Polynomial kernel	$(\delta(x_i, x_j) + k)^d$	$\delta, k \text{ e } d$
Gaussian radial basis	$\exp(-\gamma \ x_i - x_j\ ^2)$	γ
Sigmoid kernel	$\tanh(\delta(x_i x_j) + k)$	$\delta \text{ e } k$

1.8. Adaptive Neuro-Fuzzy Inference System (ANFIS)

In 1965, Zadeh of the University of California proposed the fuzzy logic theorem to describe complicated systems and help solve various problems. ANFIS is a hybrid model that combines the adaptive ability of ANNs (multi-layer feed-forward networks) with a fuzzy interface system (FIS) (Jang, 1993), and constructs a hybrid intelligent system that can be more efficient than either an ANN or fuzzy logic (Wali et al., 2012).

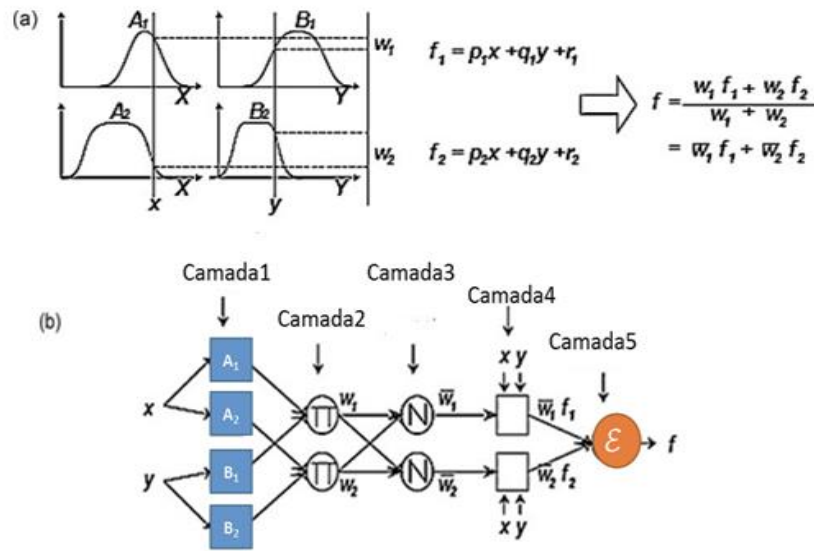


Figure 5. (a) First-order Sugeno fuzzy reasoning. (b) Equivalent ANFIS architecture. The output of (a) is inserted into layer 4 of (b). Adapted from Singh Source: Adapted (Jang, 1993)

According to Mohandes et al. (2011), the neuro-diffuse model exploits the capabilities of both neural networks and fuzzy logic systems. Singh et al. (2012) indicated that ANFIS builds an FIS whose membership function parameters are tuned using a back-propagation algorithm alone or in combination with a type of least squares method. (Yegireddi and Uday Bhaskar, 2009). The ANFIS architecture is composed of five layers: layer 1, input nodes; layer 2, rule nodes; layer 3, average nodes; layer 4, consequent nodes; and layer 5, output nodes (Jang, 1993). ANFIS has been widely used in the field of environmental science to solve nonlinear and difficult problems in many parts of the world. Fuzzy systems contain three basic sections, i.e., fuzzification, fuzzy databases, and defuzzification, whereas a fuzzy rule

base and inference system are two sections of the fuzzy database parts (Safari et al., 2020).

1.9 Random Forest (Regressor)

A random forest (RF) is one of the best-known approaches for categorical (classification) and continuous (regression) prediction problems. (Taghizadeh-Mehrjardi et al., 2020) The random forest (RF) induction algorithm was created by Cutler and Breiman (2001), and combines random selection with the idea of bagging the explanatory variables in a tree induction process. As one of the main benefits of random forests, they can accurately explain the composite connections between independent and dependent variables (Emadi et al., 2020).

Breiman (2001), described the basic premise of this algorithm as the building of a certain number of decision trees; thus, the decision trees are then combined according to certain criteria to generate random forests, and the final value of the predictions will be the average of the predictions of each individual tree. In many situations, random forests show a high level of fitness when compared with other types of algorithms, contributing to its efficient use in large databases, favoring the processing of a number of variables.

An RF is easily tunable with two main tuning hyperparameters: the number of trees (ntree) and the number of different predictors tested at each node (mtry) (Taghizadeh-Mehrjardi et al., 2020). However, it is necessary to determine the exact number of decision trees that exist in the random forest because, through the joint use of the variety provided by the decision tree, it converges from an error value to one that has not undergone an overfitting or overfitting depending on the information provided. Thus, one can see that there is an initial dataset, later divided into other random subsets and with repetition, favoring the realization of induction in each of the equivalent decision trees. In this way, the final classification occurs through voting, through which each of the decision trees provides a classification. The class that obtains more votes is called the final element of the random forest (Sullivan, 2018).

1.10 References

- Adachi M, Bekku YS, Rashidah W, Okuda T, Koizumi H (2006) Differences in soil respiration between different tropical ecosystems. **Applied Soil Ecology** 34:258-65. <https://doi.org/10.1016/j.apsoil.2006.01.006>
- Adachi M, Ito A, Yonemura S, Takeuchi W (2017) Estimation of global soil respiration by accounting for land-use changes derived from remote sensing data. **Journal of Environmental Management**. 200:97-104.
<https://doi.org/10.1016/j.jenvman.2017.05.076>
- Alencar A Zimbres B, Shimbo J, Brandão Jr. A, Arruda V, Lenti F, Souza Jr. C, Azevedo T (2020). **Nota Metodológica SEEG 8 Setor Mudança de Uso da Terra e Florestas**.76p.
- Anderson JPE (2015): Soil Respiration. 9:831-71.
<https://doi.org/10.2134/agronmonogr9.2.2ed.c41>
- de Araújo Santos GA, Moitinho MR, de Oliveira Silva B, Xavier CV, Teixeira DDB, Corá JE, Júnior NLS (2019) Effects of long-term no-tillage systems with different succession cropping strategies on the variation of soil CO₂ emission. **Science of the Total Environment** 686:413-24. <https://doi.org/10.1016/j.scitotenv.2019.05.398>
- Awad, M, Khanna, R. (2015). **Máquinas de aprendizagem eficientes: teorias, conceitos e aplicações para engenheiros e designers de sistemas** Natureza Springer. 268p.
- Azevedo T, Rosa MR, Shimbo JZ, Oliveira MG de. Relatório Anual Do Desmatamento No Brasil 2020. 2021:93
- Azevêdo, L. L. A. **Métodos estatísticos em aprendizado de máquinas para problemas de classificação**. 2018. 131 f. Dissertação (Mestrado em Estatística) – Universidade de Brasília, Brasília, 2018.
- Barbosa HP de O. Detecção De Phishing No Twitter Baseada Em Algoritmos De Aprendizagem Online. 2018.
- Braga, A. de P (2000). **Redes neurais artificiais: teoria e aplicações**. Livros Técnicos e Científicos.
- Breiman L. Random forests. *Mach Learn* (2001) 45:5-32.
<https://doi.org/10.1023/A:1010933404324>
- Cable JM, Ogle K, Williams DG, Weltzin JF, Huxman TE (2008). Soil texture drives responses of soil respiration to precipitation pulses in the sonoran desert: Implications for climate change. **Ecosystems**, 11:961-79.
<https://doi.org/10.1007/s10021-008-9172-x>
- Carvalho, A. L. C (2012) **Aplicação de técnicas de aprendizagem de máquina na geração de índices para sistemas de busca**. 101 f. Tese (Doutorado em

Informática) – Universidade Federal do Amazonas, Manaus.

<https://tede.ufam.edu.br/handle/tede/4517>

Chambers JQ, Artaxo P (2017) Biosphere-atmosphere interactions: Deforestation size influences rainfall. *Nature Climate Change* 7:175-6.

<https://doi.org/10.1038/nclimate3238>

Chen S, Huang Y, Zou J, Shen Q, Hu Z, Qin Y, Chen H, Pan G (2010) Modeling interannual variability of global soil respiration from climate and soil properties. *Agricultural and Forest Meteorology*. 150:590-605.

<https://doi.org/10.1016/j.agrformet.2010.02.004>

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. **Appl Energy**; 195:659-70.

<https://doi.org/10.1016/j.apenergy.2017.03.034>

Czyż EA (2004) Effects of traffic on soil aeration, bulk density and growth of spring barley. **Soil and Tillage Research** 79:153-66.

<https://doi.org/10.1016/j.still.2004.07.004>

Davidson EA, Belk E, Boone RD (1998) Soil water content and temperature as independent or confounded factors controlling soil respiration in a temperate mixed hardwood forest. **Global Change Biology**, 4:217-27. <https://doi.org/10.1046/j.1365-2486.1998.00128.x>

Debanshi S, Pal S (2020) Modelling water richness and habitat suitability of the wetlands and measuring their spatial linkages in mature Ganges delta of India. **Journal of Environmental Management** ;271:110956.

<https://doi.org/10.1016/j.jenvman.2020.110956>

DeLucia EH, Hamilton JG, Naidu SL, Thomas RB, Andrews JA, Finzi A, Lavine M, Matamala R, Mohan JE, Hendrey GR, Schlesinger WH (1999). Net primary production of a forest ecosystem with experimental CO₂ enrichment. **Science**; 284:1177- 9. <https://doi.org/10.1126/science.284.5417.1177>

Del Rosario Uribe M, Dukes JS. Land cover change alters seasonal photosynthetic activity and transpiration of Amazon forest and Cerrado (2021). **Environmental Research Letters**, 16. <https://doi.org/10.1088/1748-9326/abf60d>

Eden M, Schjønnig P, Moldrup P, de Jonge LW (2011) Compaction and rotovation effects on soil pore characteristics of a loamy sand soil with contrasting organic matter content. **Soil Use and Management** ;27:340-9. <https://doi.org/10.1111/j.1475-2743.2011.00344.x>

Emadi M, Taghizadeh-Mehrjardi R, Cherati A, Danesh M, Mosavi A, Scholten T (2020) Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. **Remote Sensing** 12. <https://doi.org/10.3390/rs12142234>

Estruch C, Macek P, Armas C, Pistón N, Pugnaire FI. (2020) Species identity improves soil respiration predictions in a semiarid scrubland. **Geoderma** ;363:114153. <https://doi.org/10.1016/j.geoderma.2019.114153>

F. Specht D. General Regression Neural Network. IEEE Int Conf Neural Networks - Conf Proc (2007) ;2:1959-64. <https://doi.org/10.1109/IJCNN.2007.4371258>

Fan LC, Han WY (2018) Soil respiration in Chinese tea gardens: autotrophic and heterotrophic respiration. **European Journal of Soil Science** ;69:675-84. <https://doi.org/10.1111/ejss.12670>

Fan R, Zhang B, Li J, Zhang Z, Liang (2020) A. Straw-derived biochar mitigates CO₂ emission through changes in soil pore structure in a wheat-rice rotation system. **Chemosphere**.;243:125329. <https://doi.org/10.1016/j.chemosphere.2019.125329>

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>

Fernandes MMH, Coelho AP, Fernandes C, Silva MF da, Dela Marta CC (2019) Estimation of soil organic matter content by modeling with artificial neural networks. **Geoderma**, 350:46-51. <https://doi.org/10.1016/j.geoderma.2019.04.044>

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 and Tillage Research* ;56:373. <https://doi.org/10.1071/sr16344>

Filgueiras, P. R (2014) **Regressão por vetores de suporte aplicado na determinação de propriedades físico-químicas de petróleo e biocombustíveis**. 115 f. Tese (Doutorado em Ciências) – Universidade Estadual de Campinas, Campinas.

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 H, Kashi H, Hafezi Moghadas N, Emamgholizadeh S (2015). Estimation of Soil Cation Exchange Capacity using Multiple Regression, Artificial Neural Networks, and Adaptive Neuro-fuzzy Inference System Models in Golestan Province, Iran. **Communications in Soil Science and Plant Analysis**, 46:763-80. <https://doi.org/10.1080/00103624.2015.1006367>

Gutiérrez del Arroyo O, Wood TE. Significant Diel (2020) Variation of Soil Respiration Suggests Aboveground and Belowground Controls in a Tropical Moist Forest in

Puerto Rico. **Journal of Geophysical Research-Biogeosciences**; 125:1-14.
<https://doi.org/10.1029/2019JG005353>

Han G, Xing Q, Luo Y, Rafique R, Yu J, Mickle N (2014) Vegetation types alter soil respiration and its temperature sensitivity at the field scale in an estuary wetland. **PLoS One**;9 <https://doi.org/10.1371/journal.pone.0091182>

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

Huang N, Niu Z, Zhan Y, Xu S, Tappert MC, Wu C, Huang W, Gao S, Hou X, Cai D (2012) Relationships between soil respiration and photosynthesis-related spectral vegetation indices in two cropland ecosystems. **Agricultural and Forest Meteorology**, 160:80-9. <https://doi.org/10.1016/j.agrformet.2012.03.005>

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>

IBÁ. Relatório anual do IBÁ (2020) .Indústria Brasileira De Árvores. Relatório Anual Iba 2020. São Paulo: Café Art, 2020. Available:
<https://iba.org/datafiles/publicacoes/relatorios/relatorio-iba-2020.pdf>.

Kashaninejad M, Dehghani AA, Kashiri M (2009) Modeling of wheat soaking using two artificial neural networks (MLP and RBF). **Journal of Food Engineering**, 91:602-7. <https://doi.org/10.1016/j.jfoodeng.2008.10.012>

Koide, R. M (2016) **Aplicação de regressão de vetores de suporte na otimização em flambagem e pós-flambagem de estruturas compósitas laminadas**. 137 f. Tese (Doutorado em Engenharia Mecânica e de Materiais) – Universidade Tecnológica Federal do Paraná, Curitiba,

Kuzyakov Y(2006) Sources of CO₂ efflux from soil and review of partitioning methods. **Soil Biology and Biochemistry** ;38:425-48.
<https://doi.org/10.1016/j.soilbio.2005.08.020>

Laemmel T, Mohr M, Longdoz B, Schack-Kirchner H, Lang F, Schindler D, Maier M (2019) From above the forest into the soil – How wind affects soil gas transport through air pressure fluctuations. **Agricultural and Forest Meteorology**; 265:424-34. <https://doi.org/10.1016/j.agrformet.2018.11.007>

Lafeta, B. O. et al (2020) Modelagem morfométrica de Licania tomentosa (Benth.) por regressão logística e máquinas vetor de suporte. **Scientia Plena**, 16: 6.

Lagomarsino A, Moscatelli MC, De Angelis P, Grego S (2006) Labile substrates quality as the main driving force of microbial mineralization activity in a poplar plantation soil under elevated CO₂ and nitrogen fertilization. **Science of the Total Environment** ;372:256-65. <https://doi.org/10.1016/j.scitotenv.2006.08.031>

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

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](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](https://doi.org/10.1016/S0929-1393(03)00065-9)

Li Y-L, Otieno d, Owen K, Zhang Y, Tenhunen J Rao X Q Lin Y-B (2008) Temporal Variability in Soil CO₂ Emission in an Orchard Forest Ecosystem. **Pedosphere** 18:273-83. [https://doi.org/10.1016/s1002-0160\(08\)60017-x](https://doi.org/10.1016/s1002-0160(08)60017-x)

Li Y, Wang Yunqi, Wang Yujie, Wang B (2019) Effects of simulated acid rain on soil respiration and its component in a mixed coniferous-broadleaved forest of the three gorges reservoir area in Southwest China. **Forest Ecosystems** ;6. <https://doi.org/10.1186/s40663-019-0192-0>

Liu Z, Liu W tao, Liu H tao, Gao T ping, Zhao H xiang, Li G, Han H fang, Li Z jia, Lal R, Ning T yuan (2020) Capture of soil respiration for higher photosynthesis with lower CO₂ emission. **Journal of Cleaner Production**, 246:119029. <https://doi.org/10.1016/j.jclepro.2019.119029>

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

Maier M, Mayer S, Laemmel T. Rain and wind affect chamber measurements. *Agric For Meteorol*. 2019;279. <https://doi.org/10.1016/j.agrformet.2019.107754>

Moitinho MR, Padovan MP, Panosso AR, Teixeira DDB, Ferraudo AS, La Scala N (2015) On the spatial and temporal dependence of CO₂ emission on soil properties in sugarcane (*Saccharum* spp.) production. **Soil and Tillage Research**. 148:127-32. <https://doi.org/10.1016/j.still.2014.12.012>

Moitinho MR, Teixeira DDB, Bicalho E da S, Panosso AR, Ferraudo AS, Pereira GT, Tsai SM, Borges BMF, La Scala N (2021) Soil CO₂ emission and soil attributes associated with the microbiota of a sugarcane area in southern Brazil. **Scientific Reports**. 11:1-15. <https://doi.org/10.1038/s41598-021-87479-2>

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>

Padarian, J., Minasny, B., and McBratney, A. B (2020). Machine learning and soil sciences: a review aided by machine learning tools, **SOIL**, 6, 35–52, <https://doi.org/10.5194/soil-6-35-2020>, 2020.

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>

Pell, Mikael; Stenström, J. O. H. N.; Granhall, U. 7.2 (2005). **Soil Respiration**. Bloem, J; DW Hopkins and A Benedetti, 117-126p.

Phu Nguyen, Lee Y-C (2017). An Application of Support Vector Machines to Customer Loyalty Classification of Korean Retailing Company Using R Language 1). *J Inf Syst* ;26:17-37. <https://doi.org/10.5859/KAIS.2017.26.4.17>

Pouladi N, Møller AB, Tabatabai S, Greve MH (2019). Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. **Geoderma**. 342:85-92. <https://doi.org/10.1016/j.geoderma.2019.02.019>

Raich JW, Potter CS,(2002). Bhagawati D. Interannual variability in global soil respiration, 1980-94. **Global Change Biology**. 8:800-12. <https://doi.org/10.1046/j.1365-2486.2002.00511.x>

Reichstein M, Rey A, Freibauer A, Tenhunen J, Valentini R, Banza J, Casals P, Cheng Y, Grünzweig JM, Irvine J, Joffre R, Law BE, Loustau D, Miglietta F, Oechel W, Ourcival JM, Pereira JS, Peressotti A, Ponti F, Qi Y, Rambal S, Rayment M, Romanya J, Rossi F, Tedeschi V, Tirone G, Xu M, Yakir D (2003). Modeling temporal and large-scale spatial variability of soil respiration from soil water availability, temperature and vegetation productivity indices. **Global Biogeochem Cycles**.17. <https://doi.org/10.1029/2003gb002035>

Reth S, Reichstein M, Falge E (2005). The effect of soil water content, soil temperature, soil pH-value and the root mass on soil CO₂ efflux - A modified model. **Plant Soil** 268:21-33. <https://doi.org/10.1007/s11104-005-0175-5>

Rivera JI, Bonilla CA (2020) Predicting soil aggregate stability using readily available soil properties and machine learning techniques. **Catena**; 187:104408. <https://doi.org/10.1016/j.catena.2019.104408>

Rocha SJSS da, Torres CMME, Jacovine LAG, Leite HG, Gelcer EM, Neves KM, Schettini BLS, Villanova PH, Silva LF da, Reis LP, Zanuncio (2018) JC. Artificial neural networks: Modeling tree survival and mortality in the Atlantic Forest biome in Brazil. **Science of the Total Environment** ;645:655-61. <https://doi.org/10.1016/j.scitotenv.2018.07.123>

Román JN, Ortiz M, Rolston D, Morales-Salinas L, Seguel OS, Riveros-Burgos C, Acevedo E (2021). Validation of Gas Diffusivity Models with Chilean Soil Samples. **Journal of Soil Science and Plant Nutrition**, 21:404-14. <https://doi.org/10.1007/s42729-020-00369-w>

Safari MJS, Mohammadi B, Kargar K. Invasive weed optimization-based adaptive neuro-fuzzy inference system hybrid model for sediment transport with a bed deposit. **J Clean Prod**. 2020;276:124267.

<https://doi.org/10.1016/j.jclepro.2020.124267>

Santos, K. N (2016) **A utilização de técnicas de aprendizado de máquinas para predição de crises epiléticas**. 2016. 74 f. Dissertação (Mestrado em Sistemas e Computação) – Universidade Federal do Rio Grande do Norte, Natal.

Savage K, Davidson EA, Tang J (2013) Diel patterns of autotrophic and heterotrophic respiration among phenological stages. **Global Change Biology** ;19:1151-9. <https://doi.org/10.1111/gcb.12108>

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>

Singh R, Kainthola A, Singh TN (2012) Estimation of elastic constant of rocks using an ANFIS approach. **Applied Soft Computing – Journals**, 12:40-5. <https://doi.org/10.1016/j.asoc.2011.09.010>

Siqueira, J. O., Moreira, F. D. S., Grisi, B. M., Hungria, M., and Araújo, R. S. (1994). **Microrganismos e processos biológicos do solo**. Brasília: Embrapa, 1994, 42.

Sullivan, W (2018). **Decision Tree and Random Forest: Machine Learning and Algorithms**: The Future Is Here! Califórnia: Createspace Independent Publishing Platform.

Taghizadeh-Mehrjardi R, Mahdianpari M, Mohammadimanesh F, Behrens T, Toomanian N, Scholten T, Schmidt K (2020) Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in central Iran. **Geoderma**. 376:114552. <https://doi.org/10.1016/j.geoderma.2020.114552>

Tang J, Baldocchi DD, Xu L (2005). Tree photosynthesis modulates soil respiration on a diurnal time scale. **Global Change Biology**. 11:1298-304. <https://doi.org/10.1111/j.1365-2486.2005.00978.x>

Teixeira DDB, Bicalho E da S, Panosso AR, Perillo LI, Iamaguti JL, Pereira GT, La Scala Jr N (2012). Uncertainties in the prediction of spatial variability of soil CO₂ emissions and related properties. **Revista Brasileira de Ciência do Solo**. ;36:1466-75. <https://doi.org/10.1590/s0100-06832012000500010>

Thomey ML, Collins SL, Vargas R, Johnson JE, Brown RF, Natvig DO, Friggens MT (2011). Effect of precipitation variability on net primary production and soil respiration in a Chihuahuan Desert grassland. **Global Change Biology** ;17:1505-15. <https://doi.org/10.1111/j.1365-2486.2010.02363.x>

Tuomi M, Vanhala P, Karhu K, Fritze H, Liski J (2008). Heterotrophic soil respiration-Comparison of different models describing its temperature dependence. **Ecological Modelling** ;211:182-90. <https://doi.org/10.1016/j.ecolmodel.2007.09.003>

Vapnik, Vladimir; Guyon, Isabel; Hastie (1995) Trevor. Support vector machines. **Mach. Learn**, v. 20, n. 3, p. 273-297.

Vargas R, Detto M, Baldocchi DD, Allen MF (2010). Multiscale analysis of temporal variability of soil CO₂ production as influenced by weather and vegetation. **Global Change Biology**, 16:1589-605. <https://doi.org/10.1111/j.1365-2486.2009.02111.x>

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 Soil.**;444:193-211. . <https://doi.org/10.1007/s11104-019-04262-z>

Wali WA, Al-Shamma'A Al, Hassan KH, Cullen JD (2012). Online genetic-ANFIS temperature control for advanced microwave biodiesel reactor. **Journal of Process Control** ;22:1256-72. <https://doi.org/10.1016/j.jprocont.2012.05.013>

Wang W, Peng S, Fang J (2010). Root respiration and its relation to nutrient contents in soil and root and EVI among 8 ecosystems, northern China. **Plant Soil**. 333:391-401. <https://doi.org/10.1007/s11104-010-0354-x>

Winkler K, Fuchs R, Rounsevell M, Herold M (2021). Global land use changes are four times greater than previously estimated. **Nature Communications** ;12:1-10. <https://doi.org/10.1038/s41467-021-22702-2>

Wright SJ, Muller-landau HC. (2006). The Future of Tropical Forest Species 1 Amazon Basin and Guiana Shield , or the Congo Basin and western Given the importance of habitat loss to estimates of future. **Biotropica** ;38:287-301

Xavier CV, Moitinho MR, De Bortoli Teixeira D, André de Araújo Santos G, de Andrade Barbosa M, Bastos Pereira Milori DM, Rigobelo E, Corá JE, La Scala JúniorN.(2019). Crop rotation and succession in a no-tillage system: Implications for CO₂ emission and soil attributes. **Journal of Environmental Management** ;245:8-15. <https://doi.org/10.1016/j.jenvman.2019.05.053>

Xiao Y, Liu S, Zhang M, Chen B, Xu Z, Pan Y, Shi X, Wu Z, Luo T (2020). Biotic and abiotic properties most closely associated with subtropical forest soil respiration differ in wet and dry seasons: A 10-year in situ study. **Agricultural and Forest Meteorology**. 292-293:108134. <https://doi.org/10.1016/j.agrformet.2020.108134>

Yegireddi S, Uday Bhaskar G (2009) Identification of coal seam strata from geophysical logs of borehole using Adaptive Neuro-Fuzzy Inference System. **Journal of Applied Geophysics**. ;67:9-13. <https://doi.org/10.1016/j.jappgeo.2008.08.009>

Zhang X-D (2020). **Machine Learning. A Matrix Algebra Approach to Artificial Intelligence**, vol. 45. Singapore: Springer Singapore. 223-440p. https://doi.org/10.1007/978-981-15-2770-8_6

Zhu Y, Dyck M, CAI H jie, Song L bing, Chen H (2019). The effects of aerated irrigation on soil respiration, oxygen, and porosity. **Journal of Integrative Agriculture** 18:2854-68. [https://doi.org/10.1016/S2095-3119\(19\)62618-3](https://doi.org/10.1016/S2095-3119(19)62618-3)

CHAPTER 2 – ESTIMATION OF SOIL RESPIRATION IN A PINE FOREST PLANTED IN THE CERRADO USING ARTIFICIAL NEURAL NETWORKS AND SUPPORT VECTOR REGRESSION

Abstract - The present study evaluated the predictive performance of artificial neural networks (ANNs) and support vector regression (SVR) in modeling the temporal variability of CO₂ flux (FCO₂) and O₂ influx (FO₂) in a pine forest planted in the Brazilian Cerrado. The dataset was composed of soil physicochemical, climatic, and enhanced vegetation index (EVI) variables, with respectively 500 and 176 observations of FCO₂ and FO₂. The ANN and SVR input variables were optimized with principal component analysis (PCA), and the model adjustments were tested with analysis of variance (F test; $p < 0.01$). Overall, for both estimates, ANNs showed better predictive performance than SVR, with fewer errors associated with the model. The multilayer perceptron with 12 input variables explained temporal variability in FCO₂ with R^2 of 42% and MSE of 1.392. The general regression neural network with 10 input variables explained temporal variability in FO₂ with an R^2 of 56% and MSE of 0.045. SVR with radial basis function kernel was the best estimate of FCO₂ and FO₂ ($R^2 = 51\%$, MSE = 0.8664 and $R^2 = 42\%$, MSE = 0.008, respectively). The model Polynomial kernel SVR and sigmoid kernel SVR does not fit the given data set. Polynomial kernel SVR models and sigmoid kernel SVR was not significant. Environmental variables; soil silt content, bulk density, and temperature; and EVI better explained soil respiration in this ecosystem.

Keywords: Soil gases flux; soil CO₂ emission, O₂ uptake, tropical soil, reforestation, soil-atmosphere relationship, agrometeorological variables.

2.1. Introduction

At the global scale, changes in the land and forest use (particularly deforestation) are the major sources of greenhouse gas (GHG) emission. According to the data of the Greenhouse Gas Emission and Removal Estimating System (SEEG), deforestation accounted for 44% of the total emissions in 2020.

The consequences of these anthropic actions include the loss of soil organic matter (SOM), nitrogen (N), and carbon stock (Cstock), thus compromising the chemical, physical, and biological quality of soil (Cerri et al., 2003). These implications further increase soil CO₂ emission (FCO₂), significantly affecting the concentration of this gas in the atmosphere. Meanwhile, reforestation is the key source of carbon absorption (Moreira et al., 2016), mitigating the net GHG levels. In this context, we use *Pinus caribaea* var. *hondurensis*—an exotic forest species widely distributed in the tropical regions—for reforestation in Brazil, given its socioeconomic importance (V. E. Silva et al., 2019). In recent years, this pine species has served as a viable option for forestation in the Cerrado regions, mainly because it can well adapt to soils with low natural fertility.

Soil respiration is the sum of heterotrophic respiration related to the microbial decomposition of litter, SOM, and fauna as well as autotrophic respiration related to root activity, resulting in oxygen influx (FO₂) and CO₂ flux (Six et al., 2006); (Luo and Zhuo 2006). In soil, oxygen is fundamental for aerobic microbial metabolism, and its concentration is mainly controlled through consumption by microorganisms and roots and diffusion movement (Schurgers et al., 2006). . FO₂ is the best indicator of all metabolic activities of soil microorganisms.

Soil respiration varies in response to system characteristics, including land use (Ebrahimi et al., 2019). Several studies have described the association of soil respiration with soil physicochemical characteristics (Panosso et al., 2012), climate (La Scala et al., 2003), vegetation type (Bahn et al., 2008), and canopy photosynthesis (Kuzyakov and Gavrichkova, 2010), and the interactions among these multiple factors can hinder the cause–effect relationships (Raich and Tufekcioglu, 2000).

In an attempt to describe and estimate soil CO₂ emissions, many models with empirical and semi-empirical approaches have been built for different ecosystems using predictive variables of climate, soil, and vegetation characteristics (Hanson et

al., 1993; La Scala et al., 2003; Chen et al., 2014), and biophysical parameters of surface albedo and enhanced vegetation index (EVI) (Yan et al., 2020), however, there remain uncertainties regarding the interrelationships between soil respiration and forest ecosystem parameters (Yan et al., 2006), and developing predictive models for such ecosystems represents one of the biggest challenges (Liu et al., 2018).

Recently, robust mathematical models developed using artificial intelligence (AI), such as artificial neural networks (ANNs) and vector support regression (SVR), have achieved satisfactory performance in predicting soil respiration in different agricultural ecosystems (Song et al., 2014; Farhate et al., 2018; Huang et al., 2020)

To this end, we proposed that the application of machine learning approaches, such as ANN and SVR, can help build predictive models to explain the patterns of temporal variability in soil respiration based on environmental covariates in pine forests. Therefore, the objectives of the present study were (1) to develop a generalized regression neural network (GRNN), radial basis function (RBFNN), and multilayer perceptron (MLPNN) to predict FCO_2 and FO_2 ; (2) to build an SVR model (i.e., sigmoid function and polynomial RBF) to predict FCO_2 and FO_2 ; and (3) to select the most appropriate model based on the best performance criteria. Simultaneously, we sought to elucidate the environmental variables that best explained the temporal variability in soil respiration.

2. Materials and methods

2.1 Data

The data were collected during a field study in a pine (*Pinus caribaea* var. *hondurensis*) forest in the Cerrado regions of the state of Mato Grosso do Sul in Brazil between November 2015 and May 2016. During the study, the mean FCO_2 was $4.059 \text{ } (\mu\text{mol m}^{-2}\cdot\text{s}^{-1})$ mean FO_2 was $0.371 \text{ } (\text{mg m}^{-2}\cdot\text{s}^{-1})$, soil temperature was 25.99°C , and soil water content (SWC) was 11.56 (%) (Vicentini et al., 2019).

2.2 Study location

The study site is located on the experimental farm of the Faculty of Engineering (UNESP), Selvíria, Mato Grosso do Sul, Brazil (20°20'53.41"S, 51°23'55.50"W; 354 masl), on the banks of the Paraná River. (Vicentini et al. 2019) The original vegetation in the area was described as typical Cerrado, with predominance of shrub forms and low soil fertility. Between 1988 and 1989, the native vegetation cover was replaced by homogeneous plantation of *Pinus caribaea* var. *hondurensis*. The initial population of pines included 3,000 plants in a 3×3 arrangement. Twelve years after planting, 60% thinning was performed. In 2014, the area harbored 1,484 plants (Santos, 2014).

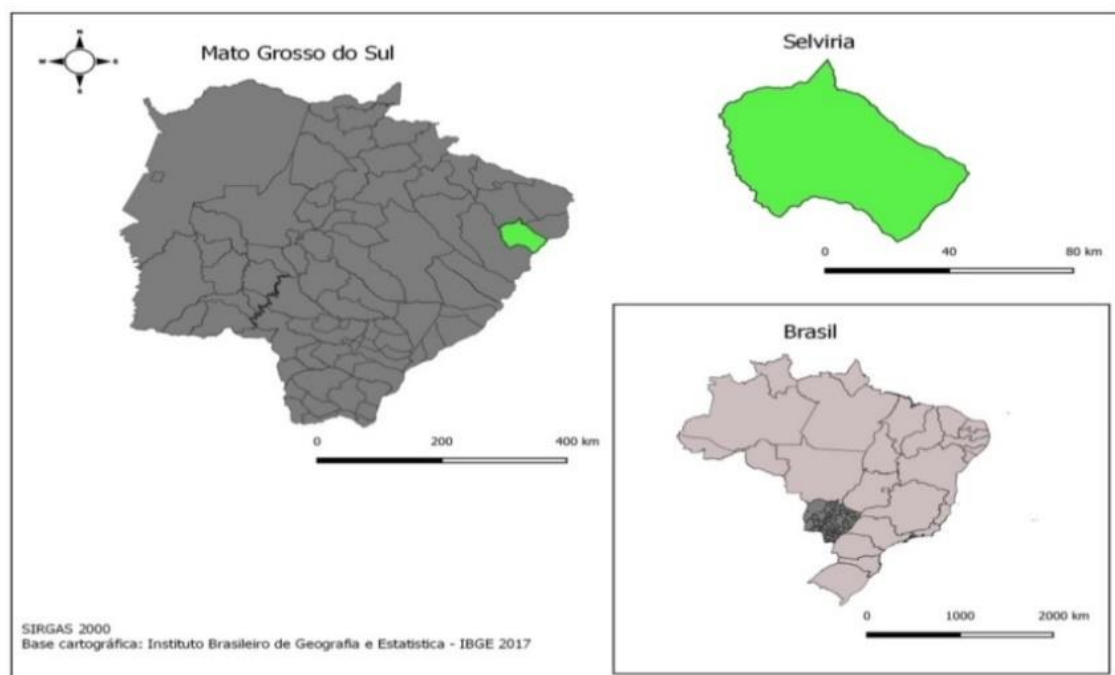


Figure 1. Location of the experimental area in Selvíria, Mato Grosso do Sul (MS), Brazil. SIRGAS: Sistema de Referência Geocêntrico para Las Américas; IBGE: Brazilian Institute of Geography and Statistics. Satellite images were obtained from Google Earth.

The climate in the region is classified as the C1dAa type according to the Thornthwaite System (De Souza Rolim et al., 2007), characterized by a dry sub-humid region, without water surplus and with mega-thermal with evapotranspiration in summer, annual evapotranspiration of <48%, mean annual temperature of 23.5°C, and mean annual rainfall of 1,300 mm. The soil is classified as Latossolo Vermelho distrófico (LVd) according to the Brazilian Classification System (Embrapa Solos, 2018) and as Oxisol (Haplic Acrustox) according to the U.S. Soil Taxonomy (Soil Survey Staff 2014). The topography of the region is characterized as moderately flat.

Soil physicochemical characteristics during the evaluation of FCO_2 and FO_2 were as follows: sand content = 648.27 g kg^{-1} , silt content = 25.76 g kg^{-1} , and clay content = 325.95 g kg^{-1} in the 0–0.10 m layer, as determined using the pipette method (Embrapa, 2017); macroporosity = $0.042 \text{ m}^3 \text{ m}^{-3}$; microporosity = $0.368 \text{ m}^3 \text{ m}^{-3}$; bulk density = 1.57 g dm^{-3} ; and P (resin) content = 6.32 mg dm^{-3} ; Al content = $13.92 \text{ mmolc dm}^{-3}$; SOM content = $20.04 \text{ g} \cdot \text{dm}^{-3}$; pH in $1 \text{ mol L}^{-1} \text{ CaCl}_2$ = 4; cation exchange capacity (CEC) = 83.91; and sum of bases = $8.56 \text{ mmolc dm}^{-3}$, as determined according to the methodology proposed by van Raij et al. (2001). Information described by Vicentini et al. (2019).

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

From November 2015 to May 2016, soil respiration was measured using the Li-8100 Soil CO_2 Flux System (LI-COR 8100, Li-Cor, Inc., Lincoln, Nebraska, USA). The system monitors changes in CO_2 concentration within a chamber using optical absorption spectroscopy in the infrared spectrum (IRGA), as described by (Panosso et al., 2011; de Araújo Santos et al., 2019).

All readings are obtained in the morning. To measure soil respiration, we selected 25 sampling points (repetitions), demarcated by polyvinyl chloride (PVC) collars (0.10 m in diameter and 0.85 m in height) and installed on the ground at a depth of 5 cm in each plot. At all points, soil temperature at a depth of 0.10 m was monitored using a digital thermometer simultaneously with soil respiration. SWC was measured using the TDR-Campbell® probe (Hydrosense TM, Campbell Scientific, Australia), inserted at a depth of 0.12 m, close to the soil respiration collars.

To determine FO_2 , 10 points were selected, and readings were obtained using a portable UV flux sensor (0%–25%) (CO_2 Meter Inc. Ormond Beach, FL, USA). This system uses the principle of fluorescence with ultraviolet (UV) light to determine O_2 concentration in the environment. The system is coupled to a microcomputer using GasLab to verify the O_2 capture rate by soil (Jassal et al., 2012; Giacomo et al., 2014). The PVC collars served as a chamber to calculate FO_2 , and the reading time of O_2 concentration was 5 min (120 s). Soil FO_2 was calculated using taking atmospheric pressure, ambient temperature, and chamber volume into account.

Additional information is available elsewhere (Ferraz de Almeida et al., 2018; Vicentini et al., 2019).

Agrometeorological variables were obtained from the historical records of UNESP-CANAL CLIMA for the interval between November 2015 and May 2016 (<http://clima.feis.unesp.br/index.php>). MODIS-EVI data were obtained using SATVeg - Temporary Vegetation Analysis System, developed by Embrapa Informática Agropecuária. MODIS images used for SATVeg development were obtained from the LP DAAC Data Pool (<http://lpdaac.usgs.gov>). These images are composed of the maximum value of the vegetation indices every 16 days. The spatial resolution of the MODIS sensor for the product used was 250 m. To minimize the effects of the noise of the time series, the data were obtained by pre-filtering and filtering available in the tool. The Savitzky–Golay filter was selected.

2.4 Datasets

In the present study, the database of dependent variables for modeling was composed of respectively (n= 500) and (n =176) observations of FCO₂ and FO₂.

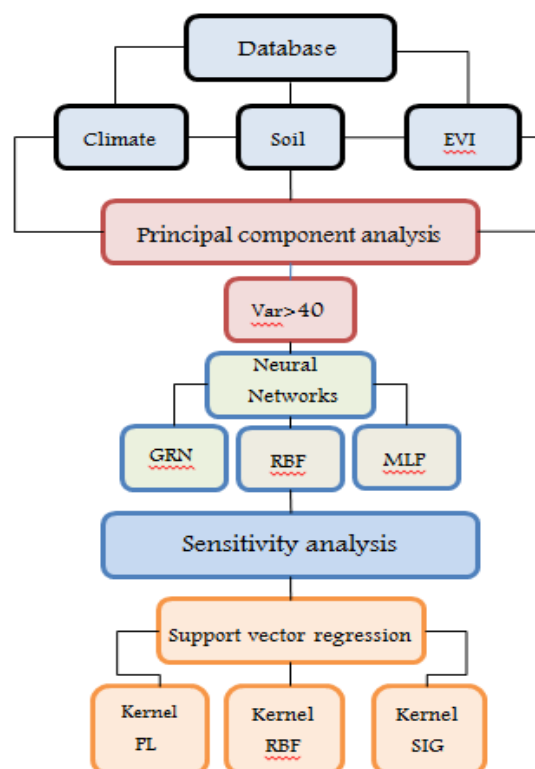


Figure 2. Flow chart of methodology used in this study.

Agrometeorological variables, soil physicochemical attributes, and EVI were selected as the quantitative environmental variables for the prediction of soil FCO_2 and FO_2 . Multivariate outliers were identified based on Mahalanobis distance. Data mining was performed using principal component analysis (PCA). The proposed AI approaches to predict FCO_2 and FO_2 included ANNs, GRNN, RBF networks, MPLs, and SVR with sigmoid RBF and polynomial kernels.

2.5. Data preprocessing and analysis

2.5.1 Principal Component Analysis

Principal component analysis was performed with the central objective of reducing the dimensionality of the database and summarizing the variability in some principle components (PCs) characterizing soil respiration and representing variables that best fit the tested models. Variables with loads greater than 0.40 were considered relevant for each PC, because they were the highest correlations. We selected the best variables related to PC1 and PC2. A total of 15 variables were used as the inputs for the ANN models to estimate the FCO_2 and FO_2 [soil temperature (T_s), SOM, P content (P), sum of bases (Bases), pH; sand content (Sand), silt content (Silt), clay content (Clay), CEC, microporosity (Micro), bulk density (Bd), EVI, evapotranspiration reference (ET_0), air temperature/ ambient temperature (T_{air}), photosynthetically active radiation (PAR), and global solar radiation (GSR),].

Although SWC was weakly correlated with PC1 and PC2, we retained it in the dataset and added it as an input variable to the ANN model, considering that SWC regulates the intensity and frequency of soil respiration. More information on the exploratory analysis and pre-selection of variables is available elsewhere (Li, 2019) . Thus, a total of 16 input variables were used for ANN model training. The three models (GRNN, RBFNN, and MLPNN) were compared, and the statistical performance indicators for each model are shown in Table 1. All ANN models were statistically significant ($p < 0.01$). PCA was performed using R (R Core Team, 2019) with the FactoMineR_2.3.tar.gz and factoextra_1.0.7.tar.gz packages.

2.5. 2 Machine learning algorithms

In the construction of all ANN and SVR models, 70% of the data were used for training and 30% for validation. Artificial neural networking is a computational technique that simulates human learning processes, recognizes patterns through examples, and is efficient and accurate for determining nonlinear relationships among a series of data inputs and one or more outputs (Ebrahimi et al., 2019).

In the present, we evaluated three different architectures: GRNN, RBFNN, and MPLNN. The GRNN has a fixed structure, with one input layer, two intermediate layers, and one output layer, and the number of neurons in each layer depends on the number of training patterns for the network inputs and outputs. (Arévalo et al. 2016). The RBF network is a multilayer feedforward network with neurons following the RBF at all nodes of a hidden layer as the activation function. Generally, it uses the Gaussian function. The MLP contains an input layer, a hidden layer, and an output layer. Each neuron in the network uses a nonlinear activation function (sigmoid function, defined by the logistic function) (Haykin, 2001) .

In this study, all ANN architectures were developed using *the Intelligent Problem Solver (IPS)* tool in Statistica 7.0® (StatSoft Dell Software Inc., Tulsa, OK, USA). The models were constructed to estimate FCO₂ and FO₂. PCA was used to identify the input variables for training the different ANN architectures. For each architectures model tested, 100 ANN were trained. The software in the IPS mode returned 10 ANNs with the lowest root mean square error (RMSE). The best neural network architecture for each estimate was selected to assess its predictive performance. The criterion for interrupting the calibration algorithm was the total number of cycles equal to 100 (Fernandes et al. 2019).

Support vector machines (SVM), as proposed by (Vapnik et al., 1995), can be used to solve problems of binary classification and regression as well as nonlinear problems. SVM follows the same principles for both classification and regression. The main idea is to build ideal hyperplanes that maximize the margin between the classes of data and minimize the unexpected errors (Zhang et al. 2020). In SVR, the hyperparameters determine the support vectors, while the kernel functions map the data into a new hyperspace, where complex nonlinear patterns can be basically represented (Xuemei, et al. 2013). The theory behind SVR in machine learning is comprehensively described by (Smola and Schölkopf, 2004)

The performance of SVR closely depends on the kernel functions. The following are the most commonly used kernel functions:

$$K(X_i, X_j) \left\{ \begin{array}{ll} (-\gamma X_i \cdot X_j + k)^d & \text{Polynomial Eq. (3)} \\ \exp(-\gamma |X_i \cdot X_j|^2) & \text{Radial Basis Function (RBF) Eq. (4)} \\ \tanh(-\gamma X_i \cdot X_j + k) & \text{Sigmoid Eq. (5)} \end{array} \right\}$$

where γ (gamma), d (degree), and k (coefficient) represent the adjustable parameters for SVR settings (Souza et al., 2019)

In the present study, SVR was used to estimate FCO_2 and FO_2 . The training of all SVR models was supervised. The algorithm was applied using Statistica 7.0[®] (StatSoft Dell Software Inc.) (Demidova et al., 2016), to explore the full potential of the software package. The cross-validation program optimization option was used to determine the values of the method parameters (capacity and epsilon value) and kernel type (Ulenberg et al. 2015). After 100 runs of optimization with the training set, a set of parameters that produced the smallest RMSE was selected. The type of SVR was as follows:

Tipo I (ϵ -SVM for regression or ϵ -SVR)

$$\text{Minimize } \frac{1}{2} W^T W + C \sum_{i=1}^n (\xi_i + \xi_i^n) \quad \text{Eq. (6)}$$

$$\text{Subject to } \begin{cases} y_i - W^T \phi(X_i) - b \leq \epsilon + \xi_i \\ W^T \phi(X_i) + b - y_i \leq \epsilon + \xi_i^n \\ \xi_i, \xi_i^n \geq 0, i = 1, \dots, n \end{cases} \quad \text{Eq. (7, 8, 9)}$$

where C is the regulation term, w is the vector of the parameters, b is a constant, and ξ is the variable slack of the error outside the accepted range ϵ , calculated by the support vectors. Index i labels n training cases. The term $\phi(X_i)$ represents the transformation of the input data (independent) by the kernel function $K(X_i, X_j)$ in the characteristic space:

$K(X_i, X_j) = \phi(X_i) \cdot \phi(X_j)$. Eq. (10) adapted from (Souza et al., 2019) In addition to the type of kernel function, the performance of SVR models depends on parameters, such as the insensitive loss function (ϵ) and capacity and function (C), which control the trade-off between maximizing the margin and minimizing the training error. If C is

very large, the algorithm will overload the training data; if C is very small, training will be insufficient (Papadopoulou et al., 2013).

2.6 Evaluation criteria for model performance

The predictive capacity of the models was assessed based on the performance of the following statistical indices: Pearson correlation coefficient (r) (Eq. 11) mean absolute percentage error (MAPE) (Eq. 12), mean square error (MSE) (Eq. 13), agreement index (d) (Eq. 14) (Willmott, 1981), and confidence coefficient (c) (Eq. 15) (Camargo and Sentelhas, 1997a). Analysis of variance was performed using the F test at the probability level of 0.01.

$$r = \frac{(N \sum_{i=1}^N (X_{obs1} Y_{obs1}) - \sum_{i=1}^N X_{obs1} \sum_{i=1}^N Y_{obs1})}{\sqrt{((N \sum_{i=1}^N X_{obs1})^2 - (\sum_{i=1}^N X_{obs1})^2)((N \sum_{i=1}^N Y_{obs1})^2 - (\sum_{i=1}^N Y_{obs1})^2)}} \quad \text{Eq. (11)}$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left(\frac{|(A_i - P_i)|}{A_i} 100 \right) \quad \text{Eq. (12)}$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2 \quad \text{Eq. (13)}$$

$$d = 1 - \frac{\sum_{i=1}^N (Y_{obs1} - Y_{est1})^2}{\sum_{i=1}^N (|Y_{est1} - \bar{Y}| + |Y_{obs1} - \bar{Y}|)^2} \quad \text{Eq. (14)}$$

$$c = r * d \quad \text{Eq. (15)}$$

where n is number of data points; X_{obs1} is the observed value of X; Y_{obs1} is the observed value of Y; Y_{est1} is the estimated value of Y; and $\bar{Y}$ is the observed mean of Y.

2.7. Results and Discussion

2.7.1 PCA

Principal component analysis identified three significant PCs that explained 78.93% of the total normalized variance of the 20 selected variables (Table 1). PC1 explained 49.54% of the total variance and was significantly correlated with the following soil physicochemical properties: CEC ($r = 0.99$), Bd ($r = -0.98$), Base ($r =$

0.97), SOM ($r = 0.97$), Sand ($r = -0.98$), Clay ($r = 0.97$), Silt ($r = -0.95$), Micro ($r = 0.92$), P ($r = 0.87$), E_{t_0} ($r = -0.51$); T_{atm} ($r = -0.040$).

Table 1. Correlations of soil attributes, climatic variables, and EVI with FCO_2 and FO_2 along the first three principal components (PC1, PC2, and PC3).

Principal componente	PC1	PC2	PC3
Explained variance (%)	49.54	18.47	10.92
Correlation			
FCO_2	0.333	0.315	-0.301
FO_2	0.197	0.475	0.529
EVI	-0.034	0.706	0.549
E_{t_0}	-0.513	0.792	-0.084
T_{air}	-0.401	0.621	-0.04
PAR	-0.361	0.775	-0.207
P_{atm}	0.463	-0.299	-0.391
Bd	-0.984	-0.152	0.03
GSR	-0.466	0.767	-0.126
Micro	0.923	0.253	0.084
Clay	0.971	0.072	0.057
Sand	-0.982	0.101	0.041
CEC	0.985	0.158	-0.007
Bases	0.978	0.18	0.006
Silt	-0.956	-0.046	0.072
pH	0.278	-0.32	0.738
P	0.869	0.286	0.081
SOM	0.971	0.072	-0.057
SWC	-0.156	-0.072	0.779
T_s	-0.253	0.647	-0.256

FCO_2 = soil CO_2 emission; FO_2 = O_2 influx; T_s ($^{\circ}C$) = soil temperature; SWC = soil water content; T_{air} = ambient temperature; evapotranspiration reference (E_{t_0}), EVI = enhanced vegetation index; P_{atm} (kPa) = atmospheric pressure; PAR ($\mu mol \cdot m^{-2}$); photosynthetically active radiation; GSR ($MJ \cdot m^{-2} \cdot day^{-1}$) = global solar radiation; SOM ($g \cdot dm^{-3}$) = soil organic matter; Micro ($m^3 \cdot m^{-3}$) = microporosity; Bd = bulk density; ($g \cdot dm^{-3}$); pH in $CaCl_2$ = hydrogen potential; Bases ($mmolc \cdot dm^{-3}$) = sum of bases (K^+ , Ca^{2+} , and Mg^{2+}); CEC ($mmolc \cdot dm^{-3}$) = cation exchange capacity; P ($mg \cdot dm^{-3}$) = soil phosphorus content; Clay = clay content ($g \cdot kg^{-1}$); Silt = silt content ($g \cdot kg^{-1}$); Sand = sand content ($g \cdot kg^{-1}$).

PC2 explained 18.47% of the total variance and was significantly correlated with soil temperature, climatic variables, and EVI. In some studies, EVI has been used as a substitute for gross primary production (GPP) (Huete et al. 2006). The variables that best correlated with PC2 were E_{t_0} ($r = 0.79$), PAR ($r = 0.77$), GRS ($r = 0.76$), EVI ($r = 0.70$), T_s ($r = 0.65$), T_{atm} ($r = 0.62$) and FO_2 ($r = 0.48$). PC3 explained 10.92 % of the total variance. It was positively correlated with SWC ($r = 0.77$), pH ($r = 0.73$), EVI ($r = 0.54$), and FO_2 ($r = 0.30$) but negatively correlated with FCO_2 ($r = -$

0.30. The FCO_2 contributed positively to PC1 ($r = 0.33$) and PC2 ($r = 0.31$) but negatively to PC3.

In recent years, computational techniques have become important in predictive modeling. However, it is challenging to select an appropriate method and find the most accurate predictive model for a given set of data (Li, 2019), on environmental variables because of its complexity. In the present study, we used PCA as an exploratory analysis to detect the correlations between the response and predictive variables. Because of multiple interdependent correlations among the variables related to soil respiration, particularly in the studied forest ecosystem, it was difficult to precisely identify the predictive variables that best explained the temporal dynamics of FCO_2 and FO_2 in this biome.

Because of the attributes retained on PC1 and their respective relationships, we expected a greater contribution of this component to FCO_2 and FO_2 , suggesting processes related to CO_2 production, resulting from the mineralization of OM by aerobic decomposer microorganisms, as corroborated by soil fertility (Six et al., 2006) and gas exchange as a function of soil texture and soil. Despite the weak correlations of FCO_2 and FO_2 with PC1, SOM was associated with FCO_2 and FO_2 , with vectors located on the same side of the biplot (Figure 3). Meanwhile, FCO_2 and FO_2 were indirectly associated with soil density and silt and sand fractions, with vectors located in on the opposite sides of the biplot.

The studied soil has low fertility and acidity (see Materials and Methods); therefore, we believe that the weak correlations of FCO_2 and FO_2 with PC1 can be explained as a result of a lower biological activity in the area, resulting in lower O_2 consumption and CO_2 production; in addition, the quality of pine residues deposited in the soil, expressed as the C/N ratio, makes it difficult to decompose by the soil microbiota (Vicentini et al., 2019). Regarding the physical properties of soil in the studied area, we observed low macroporosity ($0.042 \text{ m}^3 \text{ m}^{-3}$) and bulk density above the ideal of apparent density (1.57 dm^{-3}) for clayey soil (Reichardt and Timm 2004), suggesting soil compaction. Soil compaction compromises total porosity and, consequently, limits O_2 and CO_2 diffusion through macropores, thus affecting the water retention capacity of soil (Brady and Weill 2013). Compact soils tend to have a higher proportion of medium and small pores, which retain less water than large pores, and this water is not available to plants. Accordingly, the high microporosity of soil in the studied area ($0.368 \text{ m}^3 \text{ m}^{-3}$) explains the its inverse correlation with soil

water content (Figure 3) and weak correlations with other components with a greater explanatory power along PC1 and PC2.

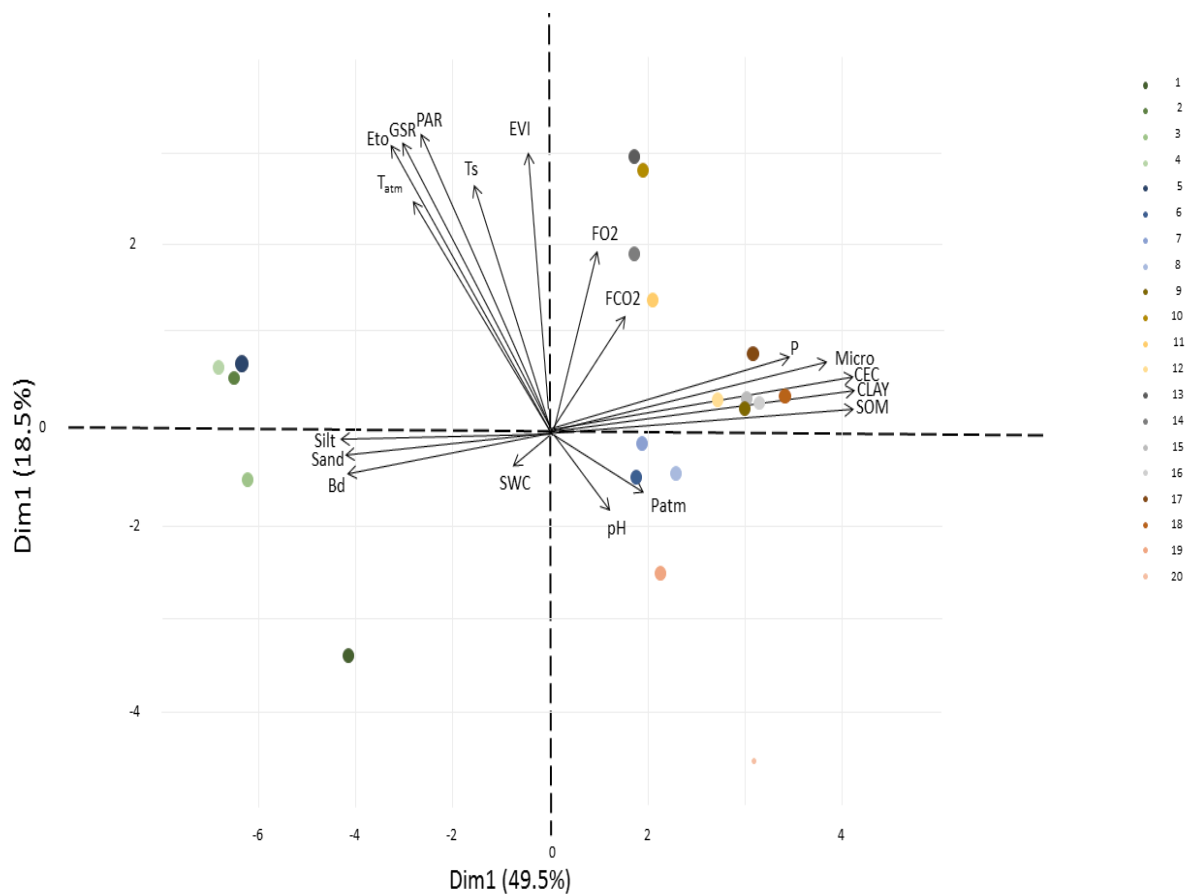


Figure 3. Biplot of first and second principal components (PC1 and PC2)

PC2 and its respective correlations indicate processes governed by the radiation and energy balance, which regulate the temporal variability of soil temperature, water evaporation (Brady and Weil 2013), evapotranspiration, and photosynthesis (Tubelis), thus controlling plant physiological activities and soil biological activity and ultimately CO₂ production and emission (Ouyang and Zheng, 2000) and O₂ consumption. The positive correlation between soil respiration and canopy photosynthesis has been well documented (Högberg et al., 2001; Ryan and Law, 2005; Wang et al., 2010) observed that in forests aged 15, 25, and 35 years, respiration by fine roots constituted the majority of the soil respiration in the area, presenting strong correlations with soil temperature and EVI.

2.8 Artificial neural networks (ANNs)

In the training phase, all ANNs showed good correlation coefficients (r) and agreement indices (d) (Willmott, 1981). Based on the performance criteria, GRNN and MLPNN were classified as “very good” (Camargo and Sentelhas, 1997a) and RFB as “bad” in the training phase. (Table 2).

Considering the original data set, the correlation coefficients for different architectures varied (GRNN: $r = 0.89$; MLP: $r = 0.81$; and RBF: $r = 0.67$). The MSE ranged from (0.61 to $1.192 \mu\text{mol m}^{-2}\cdot\text{s}^{-1}$). (Table 1). All models showed low percentages that varied between MAPE = 16% and 26%.

In the validation, the precision and error of the ANNs were similar. MSE values ranged from 1.287 to $1.558 \mu\text{mol m}^{-2} \text{ s}^{-1}$. MSE is a measure of the quality of an estimator, which penalizes the values that are the farthest from the real ones, indicating the model with the best accuracy. MSE was in the order of $\text{RBF} < \text{MLP} < \text{GRNN}$. Regarding precision, all ANNs presented similar correlation coefficients (GRNN: $r = 0.63$; MLP: $r = 0.64$; RBF: $r = 0.59$). According to the performance criteria (Camargo and Sentelhas, 1997a) GRNN and RFB were classified as “bad” and MLPNN as “suffering.” The MAPE values varied but were low, ranging from 25% to 28%, indicating that there were no major errors associated with the developed models. The explanatory power for the temporal variability in FCO_2 of MLP was slightly superior to that of the other models (adjusted $R^2 = 0.42\%$). Based on the performance of statistical indicators, MLPNN was considered be the best model (Figure 4), yielding a good agreement between the estimated and observed values. Furthermore, sensitivity analysis was used to determine the effects of the selected input variables on the dependent variable (Chen et al. 2018). The most effective inputs were T_s , Bd, Silt, EVI, Bases, SOM, T_{air} , PAR, SWC, P, Sand, and CEC.

For FO_2 all networks performed well in the training phase, with good correlation coefficients and agreement indices (d) (Willmott, 1981). The correlation coefficients for different architectures varied (GRNN: $r = 0.95$; RFB: $r = 0.81$; and MLP: $r = 0.79$). According to the performance criteria (Camargo and Sentelhas, 1997a), GRNN was classified as “very good” and RBFNN and MLPNN were classified as “good.” The MSE values varied but were low, ranging from 0.010 to $0.033 (\text{mg m}^{-2}\text{s}^{-1})$. However, the MAPE values were high, ranging from 39% to 54 %. (Table 2).

In the validation phase, the correlation coefficients for different architectures varied (GRNN: $r = 0.74$; RBF: $r = 0.58$; and MLP: $r = 0.67$). The agreement index (d)

(Willmott, 1981) ranged from 0.70 to 0.81. Based on these values, GRNN was classified as “average,” RBFNN as “bad,” and MLPNN as “suffering,” In the validation phase, the MSE values ranged from 0.045 to 0.067 $\text{mg m}^{-2}\cdot\text{s}^{-1}$), Based on the coefficient of determination, GRNN was selected as the best model, which explained 56% of the temporal variability in soil FO_2 in the studied region, followed by MLPNN ($R^2 = 46$) and RBFNN ($R^2 = 33$). However, for all models, the percentage error was very high (>50%), with values ranging from 65% to 70%.

Overall, GRNN was the best model for FO_2 prediction in the studied forest. In sensitivity analysis, the variables that most contributed to this model were Silt, GSR, Ts, EVI, T_{atm} , Bd, Sand, PAR, Bases, and Micro.

Among the ANN models for estimating FCO_2 , MLPNN trained with the backpropagation (BP) algorithm provided the best estimate of the temporal variability of FCO_2 , followed by GRNN. Both ANN models used 12 input variables during the training phase. The main advantage of MLP is the feasibility of generalization using the delta rule, which adjusts synaptic weights in a way that minimizes the quadratic errors of the network, thus calculating the difference between the observed and expected values of FCO_2 . These architectures are the most comprehensive and commonly used neural networks owing to their satisfactory performance.

In diverse ecosystems, MLPNN has shown great predictive potential to understand processes involved in the temporal dynamics of soil respiration (Song et al., 2014). In tropical soil, (Freitas et al., 2018) estimated the temporal variability in FCO_2 based on soil attributes. In the present study, the performance of MLPNN was superior to that of GRNN, with greater precision (adjusted $R^2 = 0.91$ and MAPE = 18.29%).

In the present study, the performance of RBFNN was slightly poorer than that of MLP, and it used 13 input variables during the training phase. Although RBFNN and MLPNN are three-layer feed-forward ANNs, the former network differs from the latter in that it includes only one set of trainable weights from the hidden layer to the output layer. The hidden and output layers of MLPNN are classified as non-linear. Even in its most basic form, the three layers of RBFNN have different roles, and only the output layer is linear (Haykin 2008). Likewise, RBFNN underperformed in predicting FO_2 .

Regarding FO_2 , GRNN based on a probability density function showed the best predictive performance, with preferential accuracy. The performance of a neural

network depends on the complexity of its architecture. GRNN includes more hidden layers, resulting in a greater complexity of the model and thus providing a greater capacity for non-linear estimation (Lotfinejad et al., 2018), as evidenced by the predictive performance of this architecture for FO_2 in the present study.

Sensitivity analysis of ANN models showed that the input variables Silt, Bd, Ts, and EVI stood out in both selected models. Soil texture determined by the fraction of sand, silt, and clay has a great effect on soil gas diffusion, water processes (infiltration and retention), and chemical properties (fertility). Soil compaction limits gas exchange between the soil and atmosphere (Brady and Weill 2013). In the studied region, land use change due to deforestation and intense traffic of agricultural machinery at the beginning of seedling plantation may have reduced in the porous space in soil and increased its density. The association of soil compaction and silt content with the temporal variability of FCO_2 is known (Herbst et al., 2012; Moitinho et al., 2021) .

Soil temperature also showed satisfactory performance as an input variable in the ANN models. This variable is affected by meteorological conditions, such as GSR and ambient temperature, as well as the physical parameters of soil and physical barrier formed by litter (Reichardt; Timm, 2012) In the present study, only GRNN considered SWC for estimating FO_2 . Water is one of the main barriers to O_2 diffusion (Neira et al., 2015). In a previous study, (Wang et al., 2010) used EVI as a substitute for gross primary productivity and observed that the availability of nutrients, specifically nitrogen, in forest soil enhanced root activity and, consequently, photosynthesis.

Overall, the performance of ANNs in estimating FCO_2 was better than that in estimating FO_2 , with smaller errors. However, this result may be attributed to the larger size and greater representativeness of the training dataset ($n = 500$) for FCO_2 than those for FO_2 ($n = 176$). The small number of observations probably limited the ability to generalize and recognize patterns with ANNs for FO_2 . Thus, complex models tend to overfit when there are little data (Haykin,2001). The following table and graphs shows the performance of the models.

Table 2. Performance of statistical indexes during the training and validation phases of artificial neural networks

FCO₂ (μmol·m⁻²·s⁻¹)									
Architecture	Phase	r	MAPE	MSE	d	c	Classification	F	p-value
GRNN	Calibration	0.894	0.167	0.610	0.901	0.813	Very good	979	p < 0.0001
12-250-2	Validation	0.633	0.282	1.558	0.646	0.409	Bad	75.62	p < 0.0001
RBF	Calibration	0.678	0.26	1.192	0.783	0.531	Bad	199.9	p < 0.0001
13-45-1	Validation	0.591	0.253	1.287	0.714	0.422	Bad	59.66	p < 0.0001
MLP	Calibration	0.818	0.186	0.747	0.893	0.73	Very good	492.9	p < 0.0001
12-13-8	Validation	0.648	0.271	1.392	0.784	0.507	Sufferable	87.39	p < 0.0001
FO₂ (mg m⁻²·s⁻¹)									
GRNN	Calibration	0.952	0.393	0.01	0.953	0.907	Excellent	840.9	p < 0.0001
10-89-2	Validation	0.748	0.654	0.045	0.81	0.606	Medium	49.51	p < 0.0001
RBF	Calibration	0.81	0.498	0.021	0.892	0.722	Good	131.6	p < 0.0001
14-20-1	Validation	0.58	0.786	0.067	0.706	0.41	Bad	19.29	p < 0.0001
MLP	Calibration	0.791	0.549	0.033	0.868	0.687	Good	139	p < 0.0001
14-10-1	Validation	0.67	0.701	0.063	0.779	0.522	Sufferable	31.83	p < 0.0001

The sequence of numbers indicates the number of input variables, the number of neurons in the first layer, and the number of neurons in the second layer, respectively; r: Pearson's correlation coefficient; MAPE: mean absolute percentage error; MSE: mean squared error; d: Willmott's agreement index; c: confidence coefficient; p-value: probability of the occurrence of H0.

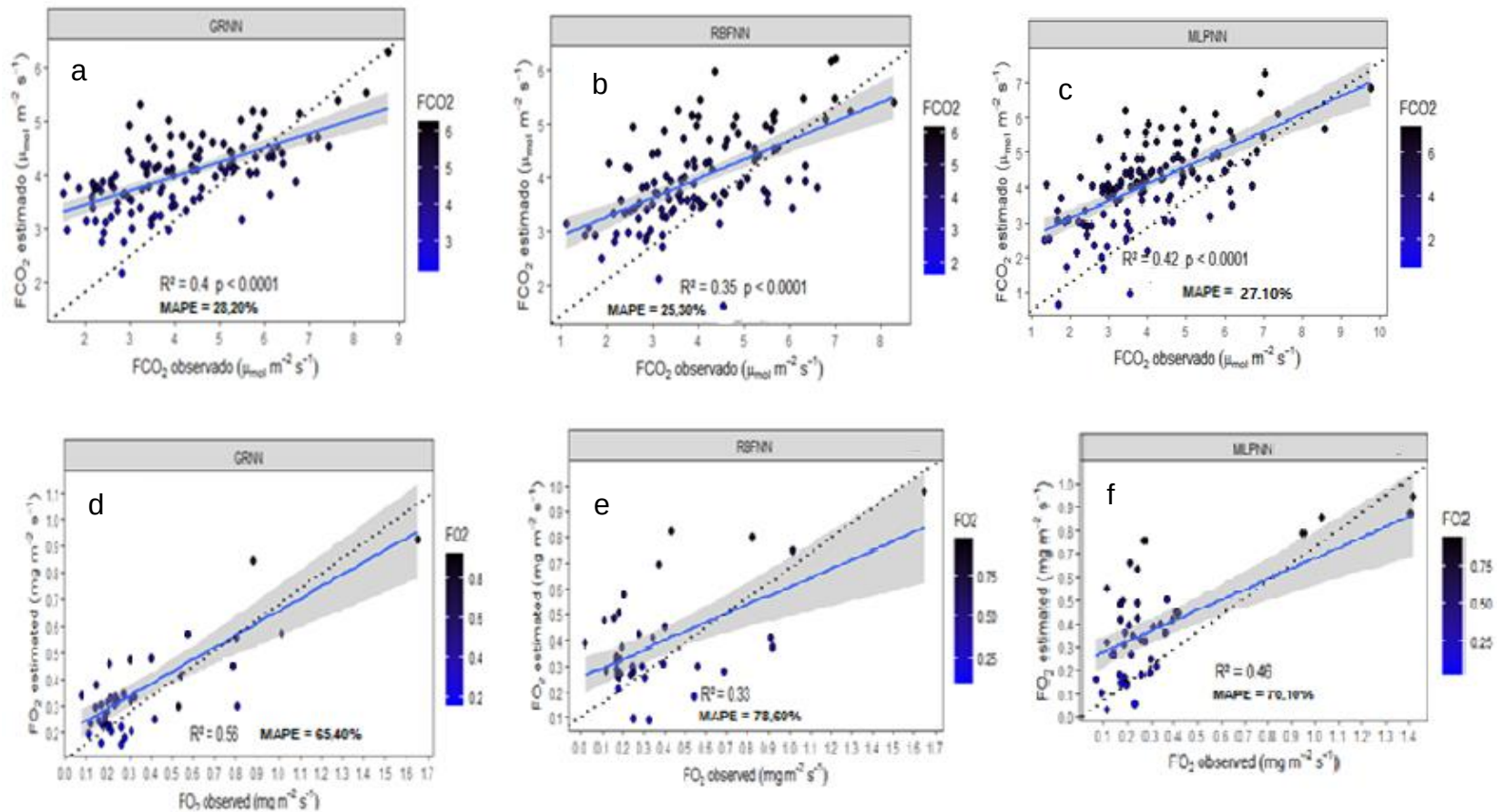


Figure 4. Performance graphs for the validation of soil CO₂ emission (FCO₂) and influx O₂ (FO₂). Coefficient of determination (R^2) and MAPE: mean absolute percentage error. a) GRNN soil CO₂ flux; b) RBFNN soil CO₂ flux; c) MLPNN soil CO₂ flux; d) GRNN O₂ influx; e) RBF O₂ influx; and f) MLP O₂ influx.

2.9 Support vector regression (SVR)

To optimize the models, the input variables of the SVR models were mined from the sensitivity analysis of the best ANN model for FCO_2 and FO_2 . For FCO_2 estimation, the dataset was composed of 12 independent variables: T_s , B_d , Silt, EVI, Bases, SOM, T_{atm} , PAR, SWC, P, Sand, and CEC. For FO_2 estimation, the dataset was composed of 10 variables: Silt, GSR, T_s , EVI, T_{air} , B_d , Sand, PAR, Bases, and Micro. The performance of the predictive models for FCO_2 and FO_2 , along with the configuration of their kernel functions, namely SVR- RBF- second-degree polynomial (SVR-PL), and sigmoid (SVR-SIG,) were as follows: $\{\text{FCO}_2[\text{SVR-RBF: } (C, = 10.00 \ \varepsilon = 0.100 \ \gamma = 0.15); \text{SVR-PL } (C, = 10.00 \ \varepsilon = 0.100 \ \gamma = 0.25) \text{ and SVR-SIG: } (C, = 10.00 \ \varepsilon = 0.100 \ \gamma = 0.50)]\}; \{\text{FO}_2[\text{SVR-RBF: } (C, = 10.00 \ \varepsilon = 0.100 \ \gamma = 0.50); \text{SVR-PL: } (C, = 20.00 \ \varepsilon = 0.100 \ \gamma = 0.25) \text{ and SVR-SIG } (C, = 20.00 \ \varepsilon = 0.100 \ \gamma = 0.16)]\}.$

Of the three models developed SVR-BRF presented the most acceptable performance (Table 3 and Figure 5). In both training and validation phases, SVR - RBF-achieved good correlation coefficients ($r = 0.84$ and 0.71 , respectively. Based on the estimates, the model showed good precision (MAPE = 21%) and (MSE = $0.869 \ \mu\text{mol}\cdot\text{m}^{-2} \ \text{s}^{-1}$). Many studies on the use of SVM in soil science suggested the improved performance by RBF kernel-based SVR (Shi et al., 2013; Jafarzadeh et al., 2016). Appropriate kernel selection allows clear data separation in the resource space, even those it is not separable in the original input space; this allows for obtaining nonlinear algorithms from algorithms previously restricted to the linear treatment (Tabari et al., 2013). SVR-PL was the second-best model due to the associated errors. Good precision in the training and validation phases ($r = 0.66$ and 0.61 respectively). The MSE values in the training and validation phases were 1.161 and 1.103 , respectively.

According to the performance criteria (Camargo and Sentelhas, 1997a), SVR- RBF was classified as “Suffering”, SVR-PL was classified as “bad” and SVR- SIG was classified as “very bad.”

For FO_2 , despite the broader applicability of SVR owing to SVM construction with polynomial, sigmoid, and RBF kernels, SVR- RBF- showed the best accuracy with minimal associated errors. in the training and validation phases ($r = 0.68$ and 0.64 , respectively). The MSE values in the training and validation phases were 0.073 and 0.087 , respectively. Li et al. (2017) evaluated and compare the performance of

multiple linear regression (MLR), ANN and SVR for the prediction dissolved oxygen concentration based on multiple water quality parameters. The results suggested that the SVM model was superior to the MLR and ANN in the prediction of dissolved oxygen concentration. The RBF kernel was chosen as the most appropriate kernel function in the oxygen concentration model. In all models the MSE values were >1 . Ideally, the MSE values should be close to zero.

SVR- PL- was the second-best model, with good precision in the training and validation phases ($r = 0.53$ and 0.55 respectively). The MSE values in the training and validation phases were 0.0901 and 0.151 , respectively, and MAPE 125% . The model fails to fit in the dataset provided. For SVR- SIG, the differences in predicted and observed values were not significant based on analysis of variance (F-test).. Probably due to the large number of explanatory variables, and the number of the training data set being too small, overfitting occurred. According to the results discussed here, overall, the ANN model provided better fit to the studied attributes and better generalization the SVR model. The models achieved fewer errors in FCO_2 estimation. The FCO_2 and FO_2 tests were run with different sample sizes. In the literature, most cases with the studied ANN and SVR algorithms are analyzed from existing big data.

Table 3. Metrics generated for the global support vector regression model phase developed for the prediction soil CO₂ emission (FCO₂) and O₂ influx (FO₂)

FCO ₂ (μmol m ⁻² s ⁻¹)									
Kernel		MAPE	MSE	r	d	c	Classification	F	p-value
SVR-RBF	Calibration	0.191	0.707	0.844	0.883	0.745	Good	893.1	p < 0.0001
	Validation	0.210	0.869	0.717	0.812	0.582	Suffering	125.7	p < 0.0001
SVR-PL	Calibration	0.268	1.161	0.665	0.774	0.515	Suffering	280.2	p < 0.0001
	Validation	0.245	1.103	0.612	0.746	0.456	Bad	70.05	p < 0.0001
SVR-SiG	Calibration	0.332	2.083	0.324	0.268	0.087	Very bad	42.32	Ns
	Validation	0.271	1.317	0.226	0.247	0.056	Very bad	6.01	Ns
FO ₂ (mg m ⁻² ·s ⁻¹)									
SVR -RBF	Calibration	0.702	0.073	0.683	0.731	0.499	Bad	111.2	p < 0.01
	Validation	0.841	0.087	0.649	0.697	0.452	Bad	27.61	p < 0.01
SVR -PL	Calibration	0.901	0.090	0.533	0.728	0.388	Very bad	50.4	p < 0.01
	Validation	1.256	0.151	0.554	0.715	0.396	Very bad	18.58	p < 0.01
SVR- SiG	Calibration	0.684	0.059	0.632	0.571	0.361	Very bad	81.92	Ns
	Validation	1.306	0.092	0.396	0.398	0.157	Very bad	7.796	Ns

FCO₂: Soil CO₂ emission; FO₂: O₂ influx; SVR- RBF radial basis function; SVR-PL Kernel polynomial; SVR- SiG, Kernel sigmoid ; MAPE: mean absolute percentage error; MSE: mean square error; r: Pearson correlation coefficient; d: agreement index; c: confidence coefficient. Ns: not significant.

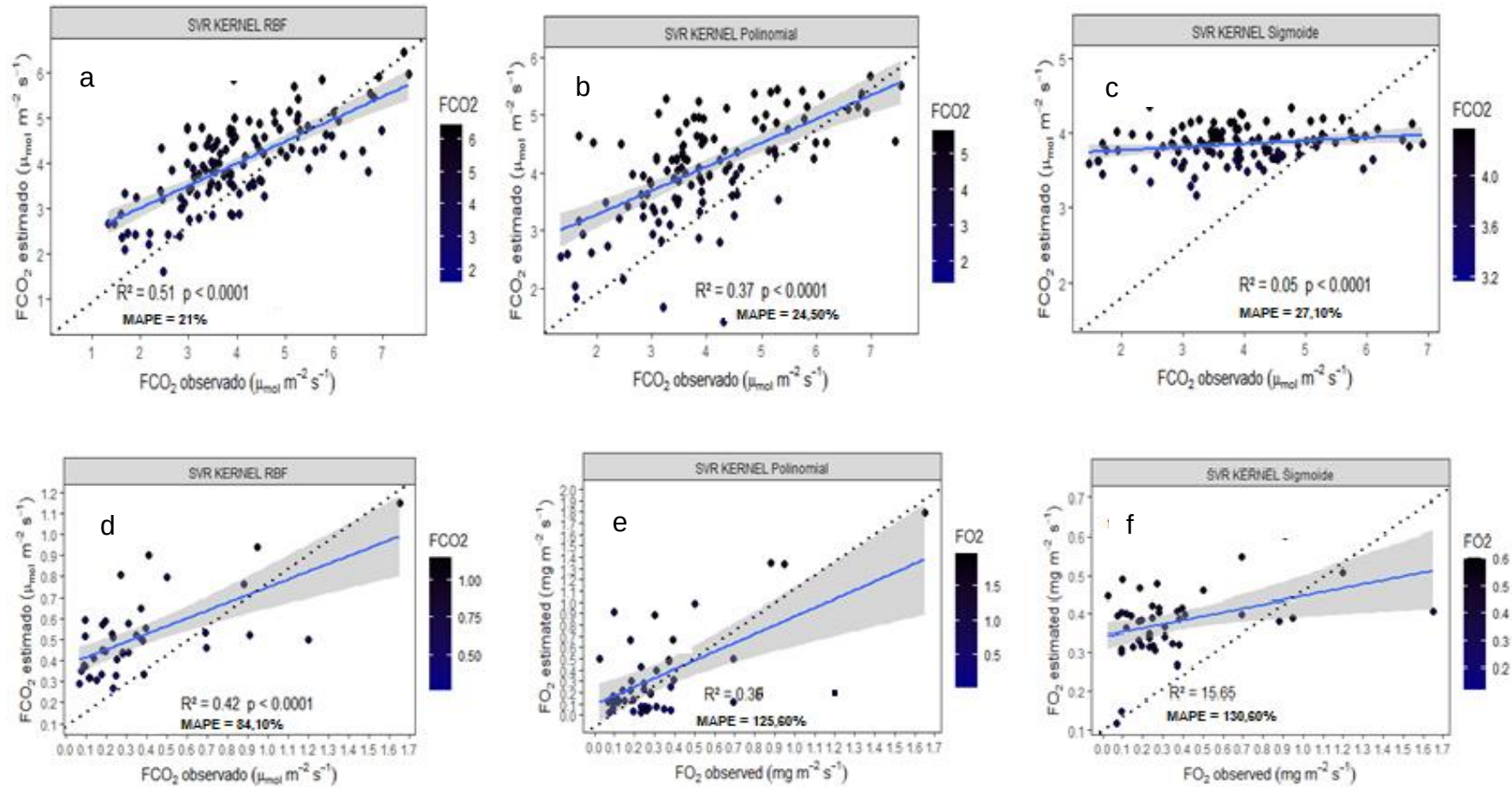


Figure 5. Performance graphs of the validation of soil CO₂ emission (FCO₂) and O₂ influx (FO₂) with radial basis function, polynomial, and sigmoid kernels. Coefficient of determination (R²) and MAPE: mean absolute percentage error. FCO₂: a) SVR-kernel radial basis function; b) SVR- Kernel polynomial; c) SVR Kernel sigmoid. FO₂: d) SVR-kernel radial basis function; e) SVR-Kernel polynomial; f) SVR Kernel sigmoid.

2.10. Concluding remarks

In the present study, using three different neural network architectures and three SVR models with different kernel functions, we sought to determine the best model for estimating the temporal variability in soil FCO₂ and FO₂ in a pine forest conserved for over 30 years.

The predictor variables in all models included soil physicochemical attributes, climate data, and EVI. PCA was performed to assess the correlations of all tested variables with FCO₂ and FO₂ as well as to reduce the size of the database. The effects of land use change and duration of conversion on pine forest soil respiration were evidenced based on soil density, soil texture, and SWC, which have a great influence of the dynamics of the gases, such as CO₂, and O₂, in soil. Our results suggest that root respiration contributes more to the temporal variability of FCO₂.

ANNs showed better performance than SVR models in explaining the patterns of temporal variability in FCO₂, specifically, MLPNN with 12 input variables (Ts, Bd, Silt, EVI, bases, SOM, T_{air}, PAR, SWC, P, Sand, and CEC) best explained the temporal variability of soil respiration ($R^2 = 42\%$). Among the ANN models, GRNN with 10 input variables (Silt, RSG, T_s, EVI, T_{air}, Bd, Sand, PAR, Bases, and Micro) showed the best performance in estimating FO₂, ($R^2 = 56\%$); however, this estimate is larger.

Environmental variables, silt content, bulk density, soil temperature, and EVI best explained soil respiration in this ecosystem. SWC did not contribute to the regressions. SVR- RBF- was the best estimate of FCO₂ and FO₂ ($R^2 = 51\%$ and 42% , respectively). The predictive models for FCO₂ were more efficient than those for FO₂. Further research on soil respiration is essential. Due to the complex interactions in the Cerrado biome, other data-mining tools must be tested or combined approached be implemented.

2.11 References

Arévalo A., Niño J., Hernández G., Sandoval J. (2016) High-Frequency Trading Strategy Based on Deep Neural Networks. In: Huang DS., Han K., Hussain A. (eds) Intelligent Computing Methodologies. ICIC 2016. Lecture Notes in Computer Science 9773. Springer, Cham. https://doi.org/10.1007/978-3-319-42297-8_40

Brady NC, Weil RR (2013). **Elementos da natureza e propriedades dos solos**. 3^a ed. Porto Alegre: Bookman. 704p.

Bahn M, Rodeghiero M, Anderson-Dunn M, Dore S, Gimeno C, Drösler M, Williams M, Ammann C, Berninger F, Flechard C, Jones S, Balzarolo M, Kumar S, Newesely C, Priwitzer T, Raschi A, Siegwolf R, Susiluoto S, Tenhunen J, Wohlfahrt G, Cernusca A (2008). Soil respiration in European grasslands in relation to climate and assimilate supply. **Ecosystems** ;11:1352-67. <https://doi.org/10.1007/s10021-008-9198-0>

Camargo AP de, Sentelhas PC. Performance evaluation of different potential evapotranspiration estimating methods in the State of São Paulo, Brazil. *Rev Bras Agrometeorol*. 1997;5:89-97

Cerri CEP, Coleman K, Jenkinson DS, Bernoux M, Victoria R, Cerri CC (2003). Modeling Soil Carbon from Forest and Pasture Ecosystems of Amazon, Brazil. **Soil Science Society of America Journal** ;67:1879-87. <https://doi.org/10.2136/sssaj2003.1879>

Chen S, Zou J, Hu Z, Chen H, Lu Y (2014). Global annual soil respiration in relation to climate, soil properties and vegetation characteristics: Summary of available data. **Agricultural and Forest Meteorology**, 198:335-46. <https://doi.org/10.1016/j.agrformet.2014.08.020>

Demidova L, Ivkina M, Zhdankina E, Krylova O, Sofyin E, Reshetova V, Stepanov N, Tyart N (2016). **Software Package STATISTICA and Educational Process**. SHS Web Conf ;29:02011. <https://doi.org/10.1051/shsconf/20162902011>

Dogan E, Sengorur B, Koklu R (2009). Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *J Environ Manage* ;90:1229-35. <https://doi.org/10.1016/j.jenvman.2008.06.004>

Ebrahimi M, Sarikhani MR, Safari Sinigani AA, Ahmadi A, Keesstra S (2019). Estimating the soil respiration under different land uses using artificial neural network and linear regression models. **Catena**; 174:371-82. <https://doi.org/10.1016/j.catena.2018.11.035>

Embrapa Solos (2018) Sistema brasileiro de classificação de solos. <https://livimagens.sct.embrapa.br/amostras/00085890.pdf>

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>

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>

Hanson PJ, Wullschleger SD, Bohlman SA (1993). Todd DE. Seasonal and topographic patterns of forest floor CO₂ efflux from an upland oak forest. **Tree Physiology** ;13:1-15. <https://doi.org/10.1093/treephys/13.1.1>

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

Haykin, S. (2008) **Neural Networks and Learning Machines**. 3rd Edition, Prentice Hall, Upper Saddle River.

Herbst M, Bornemann L, Graf A, Welp G, Vereecken H, Amelung W (2012). A geostatistical approach to the field-scale pattern of heterotrophic soil CO₂ emission using covariates. **Biogeochemistry** ;111:377-92. <https://doi.org/10.1007/s10533-011-9661-4>

Högberg P, Nordgren A, Buchmann N, Taylor AFS, Ekblad A, Högberg MN, Nyberg G, Ottosson-Löfvenius M, Read DJ (2001). Large-scale forest girdling shows that current photosynthesis drives soil respiration. **Nature**. 411:789-92.
<https://doi.org/10.1038/35081058>

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>

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**. 13:87-96. <https://doi.org/10.1007/s13762-015-0856-4>

Jassal RS, Black TA, Nesic Z, Gaumont-Guay (2012) D. Using automated non-

steady-state chamber systems for making continuous long-term measurements of soil CO₂ efflux in forest ecosystems. **Agricultural and Forest Meteorology**, 161:57-65. <https://doi.org/10.1016/j.agrformet.2012.03.009>

Kuzyakov Y, Gavrichkova O. REVIEW: Time lag between photosynthesis and carbon dioxide efflux from soil: A review of mechanisms and controls. **Global Change Biology**. 16:3386-406. <https://doi.org/10.1111/j.1365-2486.2010.02179.x>

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](https://doi.org/10.1016/S0929-1393(03)00065-9)

Li J (2019) A critical review of spatial predictive modeling process in environmental sciences with reproducible examples in R. **Applied Sciences** 9. <https://doi.org/10.3390/app9102048>

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>

Lotfinejad MM, Hafezi R, Khanali M, Hosseini SS, Mehrpooya M, Shamshirband S. (2018). A comparative assessment of predicting daily solar radiation using Bat Neural Network (BNN), Generalized Regression Neural Network (GRNN), and Neuro-Fuzzy (NF) system: A case study. **Energies**. ; 11:1-15. <https://doi.org/10.3390/en11051188>

Luo, Y., Zhou, X (2006). Soil respiration and 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, La Scala N (2021) Soil CO₂ emission and soil attributes associated with the microbiota of a sugarcane area in southern Brazil. **Scientific Reports**. 11:1-15. <https://doi.org/10.1038/s41598-021-87479-2>

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>

Ouyang Y, Zheng C.(2000) Surficial processes and CO₂ flux in soil ecosystem. **J Hydrol**. 234:54-70. [https://doi.org/10.1016/S0022-1694\(00\)00240-7](https://doi.org/10.1016/S0022-1694(00)00240-7)

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 CO₂ 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>

Papadopoulou OS, Panagou EZ, Mohareb FR, Nychas GJE (2013) Sensory and microbiological quality assessment of beef fillets using a portable electronic nose in tandem with support vector machine analysis. *Food Research International*; 50:241-9. <https://doi.org/10.1016/j.foodres.2012.10.020>

Raich JW, Tufekcioglu A (2000) Vegetation and soil respiration: Correlations and controls. **Biogeochemistry** ;48:71-90. <https://doi.org/10.1023/A:1006112000616>

Raij B, Andrade JC, Cantarella H, Quaggio JÁ (2001). **Análises químicas para avaliação da fertilidade de solos tropicais**. 1a Edição Campinas, Instituto Agrônômico de Campinas. 285 p.

Ranković V, Radulović J, Radojević I, Ostojić A, Čomić L.(2010). Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. *Ecological Modelling*. 221:1239-44. <https://doi.org/10.1016/j.ecolmodel.2009.12.023>

Reichardt, K; Tilm, L. C (2004) **Solo, planta e atmosfera: conceitos, processos e aplicações**. Barueri: Manole.

Reichardt, K.; Tilm, L. C (2012) **Solo, planta e atmosfera: conceitos, processos e aplicações**. 2 ed. Barueri: Manole 500p.

Ryan MG, Law BE (2005). Interpreting, measuring, and modeling soil respiration. **Biogeochemistry**, 73:3-27. <https://doi.org/10.1007/s10533-004-5167-7>

Santos, W. S (2014). **Variação genética e desempenho de progênis de *Pinus caribaea* var. *hondurensis* para produção de madeira e resina** 99 f. Dissertação (Mestrado em Sistemas de Produção) - Universidade Estadual Paulista, Ilha Solteira.

Sistema de Estimativas de Emissões e Remoções de gases de efeito estufa (SEEG) (2020). **Análise das emissões brasileiras de gases de efeito estufa e suas implicações para as metas de clima de Brasil 1970 – 2019**.
http://energiaeambiente.org.br/wp-content/uploads/2020/11/SEEG8_DOC_ANALITICO_SINTESE_1990-2019.pdf

Schurgers G, Dörsch P, Bakken L, Leffelaar P, Haugen LE (2006). Modelling soil anaerobiosis from water retention characteristics and soil respiration. *Soil Biology and Biochemistry*. 38:2637-44. <https://doi.org/10.1016/j.soilbio.2006.04.016>

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>

Silva VE, da Silva PRT, Montanari R, Lisboa SD da S, Batello ERB, Aguilar JV, Lisboa LAM, Albertini MM (2019). Productivity of *Pinus caribaea* VAR. *Hondurensis* and its relations with chemical attributes of soils in the Brazilian Cerrado region. *Ciência Florestal*. 29:292-307. <https://doi.org/10.5902/1980509829367>

Six J, Frey SD, Thiet RK, Batten KM (2006) Bacterial and Fungal Contributions to Carbon Sequestration in Agroecosystems. *Soil Science Society of America Journal* ;70:555-69 <https://doi.org/10.2136/sssaj2004.0347>

Smola AJ, Schölkopf B (2004). A tutorial on support vector regression. *Statistics and Computing*. 14:199-222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>

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 – Journals**, 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 systems and their applicability in the determination of agroclimatic zoning for the state of São Paulo, Brazil. *Bragantia* ;66:711-20. <https://doi.org/10.1590/s0006-87052007000400022>

Tabari H, Martinez C, Ezani A, Hosseinzadeh Talaee P (2013). Applicability of support vector machines and adaptive neurofuzzy inference system for modeling potato crop evapotranspiration. **Irrigation Science**, 31:575-88. <https://doi.org/10.1007/s00271-012-0332-6>

Vapnik, Vladimir; Guyon, Isabel; Hastie (1995) Trevor. Support vector machines. **Mach. Learn**, v. 20, n. 3, p. 273-297

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 Soil**; 444:193-211. . <https://doi.org/10.1007/s11104-019-04262-z>

Wang W, Peng S, Fang J (2010). Root respiration and its relation to nutrient contents in soil and root and EVI among 8 ecosystems, northern China. **Plant Soil**. ;333:391-401. <https://doi.org/10.1007/s11104-010-0354-x>

Willmott CJ (1981). On the validation of models. **Physical Geography**, 2:184-94. <https://doi.org/10.1080/02723646.1981.10642213>

Williams, G.J. (2017). **The Essentials of Data Science: Knowledge Discovery Using R** (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315151458>

Xuemei, L.; Jianshe, L (2013) Measurement of soil properties using visible and short wave- near infrared spectroscopy and multivariate calibration. *Meas. J. Int. Meas. Confed*. 46, 3808–381. <https://doi.org/10.1016/j.measurement.2013.07.007>

Yan J, Wang Y, Zhou G, Zhang D (2006). Estimates of soil respiration and net primary production of three forests at different succession stages in South China. **Global Change Biology**, 12:810-21. <https://doi.org/10.1111/j.1365-2486.2006.01141.x>

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>

Zhang X-D (2020). Machine Learning. A Matrix Algebra Approach to Artificial Intelligence, vol. 45. Singapore: **Springer Singapore** 223-440p.
https://doi.org/10.1007/978-981-15-2770-8_6

CHAPTER 3 - ESTIMATION OF SOIL CO₂ EMISSIONS AND OXYGEN INFLUX IN A EUCALYPTUS FOREST USING ARTIFICIAL NEURAL NETWORKS AND SUPPORT VECTOR REGRESSION MACHINE LEARNING

Abstract - The dynamics of C at the soil-atmosphere interface relate the soil C content and stock with the CO₂ emission from the soil surface to the atmosphere (soil respiration). In recent years, artificial intelligence has helped interpret complex problems, and machine learning techniques have been effective in modeling soil and environmental sciences. Therefore, we hypothesize that machine learning techniques can predict soil respiration in areas reforested with eucalyptus over a conversion period of 30 years. This study aimed to estimate soil CO₂ emissions (FCO₂) and oxygen uptake (FO₂) using artificial neural networks (ANNs) and support vector regression (SVR). We compared the performance of three different network architectures: a general regression neural network (GRNN), multilayer perceptron (MLPNN), and radial basis function (RBFNN) neural network. We tested three different kernels in SVR (radial basis function, sigmoid, and polynomial). The database consisted of 23 original variables for climate, soil, photosynthetically active radiation (PAR) sensor, and enhanced vegetation index (EVI). The numbers of FCO₂ and FO₂ recordings were $n = 500$ and $n = 175$, respectively. Principal component analysis (PCA) was used to reduce the dimensionality of the data. The PCs were the inputs to all the models. For the estimation of FCO₂, the best predictive performance was obtained with MLPNN during the calibration ($R^2 = 0.75$; RMSE = $0.932 \mu\text{mol m}^{-2} \text{s}^{-1}$) and validation ($R^2 = 0.59$; RMSE = $1.034 \mu\text{mol m}^{-2} \text{s}^{-1}$) phases. MLPNN also showed the most accuracy ($R^2 = 0.73$; and 0.36 , calibration and validation). FO₂ estimation: calibration ($R^2 = 0.73$; RMSE = $0.073 \text{ mg m}^{-2} \text{s}^{-1}$) and validation ($R^2 = 0.36$; RMSE = $0.076 \text{ mg m}^{-2} \text{s}^{-1}$). In SVR, the RBF kernel was superior to the sigmoid and polynomial kernels, with the following values for FCO₂; calibration ($R^2 = 0.58$; RMSE = $0.820 \mu\text{mol m}^{-2} \text{s}^{-1}$) and validation ($R^2 = 0.53$; RMSE = $0.990 \mu\text{mol m}^{-2} \text{s}^{-1}$). For FO₂ the validation regressions were not significant. The estimates of FCO₂ were superior to those of FO₂. The principal component analysis with FCO₂ and FO₂ indicated that the organic matter in the eucalyptus forest had stabilized. The dynamics of FCO₂ and FO₂ were related to root activity.

Keywords: Soil Respiration; artificial intelligence; Land-use change, tropical ecosystems; soil-atmosphere relationship; agrometeorological variables.

3.1 Introduction

Tropical forests are an important component of the global carbon cycle, acting as sources or sinks for carbon dioxide (CO₂) exchange between the forest and the atmosphere (Cerri et al., 2007; Soong et al., 2020). The conversion of native forests to monocultures directly affects the microclimate, alters energy flow, evapotranspiration (Biudes et al., 2015), and nutrient cycling (Davidson *et al.* 2004) and promotes soil organic matter depletion (Don *et al.* 2010). This leads to a reduction in the soil organic C stocks (Bonini et al., 2018), and concomitant changes in soil nitrogen (N) content (Conant *et al.* 2005). Carbon and N are relevant elements in soil organic matter studies as nutrients with the most enunciated dynamics in the system (d' Andréa *et al.* 2004).

There are many uncertainties regarding the effects of deforestation and land use change (LUC) on biogeochemical cycles and their consequences on the regional carbon balance (Bonini et al., 2018). These include stand age and the time needed after afforestation to reach a new equilibrium in soil respiration (Wu et al., 2020) necessary to ameliorate climate change (Alves et al., 2021). Climate change has caused an increase in soil temperature, which can directly affect the soil microbial population, intensifying the decomposition of organic matter, consequently emitting CO₂ into the atmosphere, and exacerbating global warming (Heimann and Reichstein, 2008; Milfont et al., 2021). In this process (heterotrophic respiration), aerobic microorganisms consume oxygen (FO₂) as part of their metabolic activities, and as a result, CO₂ is released into the atmosphere. When respiration is associated with the roots, it is known as autotrophic respiration, which is related to photosynthetic activity (Tang et al., 2005; Moyano et al., 2008). In forest ecosystems, the contribution of living roots to soil CO₂ emissions is higher than the decomposition of litter (Rodeghiero et al., 2013).

Many studies have correlated R_s with agrometeorological (La Scala et al., 2003; Kuzyakov and Gavrichkova, 2010) and microclimate variables (Rodeghiero et al., 2013), soil temperature and moisture (Lal, 2009; Panosso et al., 2009), vegetation characteristics (Bahn et al., 2008), chemical attributes (Lal, 2009), soil physical properties (Silva et al., 2020), enhanced vegetation index (EVI) (Yan et al., 2020), substrate availability, and CN ratio (Xavier et al., 2019; Chen et al., 2021). However, few studies have correlated RS with oxygen availability (Ferraz de Almeida

et al., 2018; Vicentini et al., 2019). Oxygen has an important influence on the speed of aerobic decomposition in its role as an electron acceptor in the degradation of organic matter (Greenwood, 1961). Decomposition rates tend to increase with increasing temperature, moisture, and oxygen (Sierra et al. 2017), The result is flux C is released from the soil system to the atmosphere (Bond-Lamberty and Thomson, 2010), which is then known as soil CO₂ emission (FCO₂). In soil water content controls oxygen diffusion (Neira et al., 2015).

Brazil is threatened by deforestation. The Cerrado biome, Brazil, is classified as a global biodiversity hotspot and has been considered as the agricultural frontier of Brazil since the early 1970s (Batlle-Bayer et al., 2010). Between 2016 and 2019, 46% of its native area was devastated and transformed for use in agriculture and livestock farming (Moreira et al., 2021). In this naturally low C soil, organic carbon plays an important role in practically all edaphic processes (Zinn et al., 2002). Some studies have described reforestation with fast-growing exotic tree species (such as *Eucalyptus* spp.) as a good alternative for increasing soil organic carbon content and as a viable option for C sequestration in relatively stable forms (Zinn et al., 2002; Gatto et al., 2010; Souza et al., 2019). However, little is known about the dynamics of soil respiration after a long period of reforestation.

In the last 50 years, the study of modeling has improved. Solving nonlinear problems of high complexity has led to many advances in all areas of science, including the adoption of machine learning techniques (Jha et al., 2019). Machine learning is the "field of study that gives computers the ability to learn without being explicitly programmed (Samuel, 1959). Among the best-known techniques used to solve supervised learning problems are regression artificial neural networks (ANN), and support vector regression (SVR). Neural networks have demonstrated good performance in estimating and classifying FCO₂ in different ecosystems (Farhate et al., 2018; Freitas et al., 2018; Ebrahimi et al., 2019); good results have also been observed using a support vector regression machine (Huang et al., 2020).

Regarding FO₂, algorithms have been applied to estimate dissolved oxygen in water (Dogan et al. 2009; Ranković et al. 2010; Yuan et al., 2019). However, an investigation and comparison of these techniques for FO₂ in tropical soils with *Eucalyptus* forests has not been performed. We hypothesize that such learning techniques are efficient in estimating FCO₂ and FO₂ in an area of *Eucalyptus*-planted forest. The objective of this work was to evaluate the predictive performance of three

different ANN architectures (GRNN: regression neural network; MLPNN: multilayer perceptron and RBFNN radial basis function) and three different kernels in SVR (radial basis function, sigmoid, and polynomial) in modeling the temporal variability of FCO_2 and FO_2 in mature *Eucalyptus*-planted forest in Cerrado from environmental covariates and soil attributes

3.2 Materials and methods

3.2.1 Study area

The database is the result of field studies conducted in an area (*Eucalyptus camaldulensis*) (Euc). The research was conducted over the 193 days, spanning from November 2015 to May 2016. The experimental area is located in the municipality of Selvíria, Mato Grosso do Sul, Central Brazil, (20° 20' 53.41" South, 53.41" South and 51° 23' 55.50" West, altitude 354 m). The climate of the region was classified as C1dAa' by the Thornthwaite system (Rolim et al. 2007), with mean annual temperature of 23.5 °C and mean annual rainfall of 1,300 mm, and relative humidity between 70 and 80%.

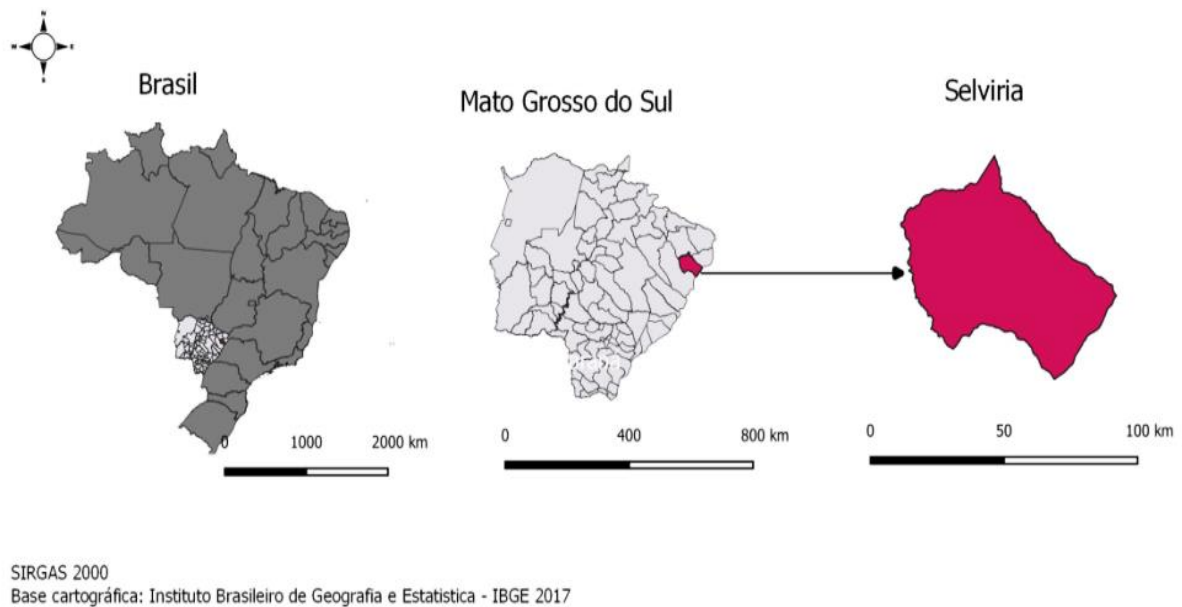


Figure 1. Location of the experimental area in Selvíria, Mato Grosso do Sul (MS), Brazil. SIRGAS: Sistema de Referencia Geocéntrico para Las Américas; IBGE: Brazilian Institute of Geography and Statistics

The local soil was classified as dystrophic Red Latosol (LVd), according to the Brazilian Soil Classification System (Embrapa, 2018). The relief is characterized as

moderately flat and undulating, with sand content 610 g kg⁻¹, silt 56 gkg and clay 332 g kg⁻¹. The studied area had soil covered by native Cerrado vegetation until the 1970s; it was deforested. In 1978 for annual agricultural production (soybeans, corn, cotton, rice, and green manures) until 1986. In 1986, the areas were converted to eucalyptus forest use.

3.3. Soil data

Soil physical analyses were performed in the 0–0.10 m layer after the FCO₂ and FO₂ readings. The following results were observed: macroporosity (macro) = 0.08 ± 0.006 (m³ m⁻³), microporosity (micro) = 0.31 ± 0.006 (m³ m⁻³), and soil density (BD) = 1.47 ± 0.02 (g dm⁻³) (according to the methodology proposed by Embrapa (1997)). For the chemical analyses the following results were observed: phosphorus content (resin) (P) = 6.52 ± 0.50 (mg dm⁻³), potential acidity H + Al = 55.68 ± 02.612 (mmolc dm⁻³), soil organic matter (SOM) = 29.76 ± 1.12 (g dm⁻³), pH (0.1 mol L⁻¹ CaCl₂) = 4.31 ± 0.07 , cation exchange capacity (CEC) = 86.99 ± 3.10 and base sum (SB) = 25.98 ± 3.75 (mmolc dm⁻³), according to the methodology proposed by Van Raij et al. (2001). Carbon stocks (Cstock) = 56.26 t ha⁻¹ and N organic total (0.95 t ha⁻¹). The carbon stock was calculated on the basis of the equivalent soil mass (Carvalho et al. 2009). Litter samples were collected in May of 2016. The organic carbon and nitrogen contents were determined using the methods described by Tedesco et al. (1995) and Bataglia et al. (1983), respectively.

3.3.1 Determination of soil CO₂ emission, O₂ uptake, soil temperature, moisture and Environmental data

Evaluations of CO₂ emission (FCO₂), influx oxygen (FO₂), soil temperature (T_s) and soil water content were performed throughout the 193 days of the study, in each of the 25 experimental plots, thus totaling 20 evaluations. The readings always took place in the morning (7 am to 12 noon). FCO₂ was measured using a portable LI-8100 system (LI-COR Inc., Lincoln, NE). Soil temperature was determined for all points studied with the aid of a digital thermometer at a depth of 0.10 m and soil water content was obtained with a portable TDR (Time Domain Reflectometry) sensor - Campbell® (Hydrosense TM, Campbell Scientific, Australia), which

evaluates the available soil moisture (% in volume). To determine O_2 uptake (FO_2) 10 sample plots were selected and with the help of a portable UV Flux (0-25%) sensor (CO₂Meter, Inc., Ormond Beach, FL, USA) the O_2 concentration was captured.

Further information was described by (Ferraz de Almeida et al., 2018; Vicentini et al., 2019). In this period the mean soil respiration values for FCO₂ was 4.059 ($\mu\text{mol m}^{-2} \text{s}^{-1}$) and for FO_2 0.369($\text{mg m}^{-2}\text{s}^{-1}$) and soil temperature averaged (26.5°C) and soil water content was 10.6(%).

The meteorological data was obtained through the historical records of UNESP - CANAL CLIMA (<http://clima.feis.unesp.br>) for the interval from (November 2015 to October 2017). The following variables were selected: precipitation (PP-mm); wind velocity w (m/s^{-1}); photosynthetically active radiation (PAR - $\mu\text{mol m}^{-2}\text{s}^{-1}$); global radiation (RSG - $\text{MJ m}^2 \text{day}^{-1}$); atmospheric pressure (P_{atm} - Kpa) and air temperature (T_{air} °C). Regarding the study period, we used a time series of EVI data obtained from the time series of the MODIS sensor on board the Terra satellite, product MOD13Q1 with the help of SATVeg. Temporal Vegetation Analysis System (Embrapa Informática Agropecuária, 2015), which can be accessed online: (<https://www.satveg.cnptia.embrapa.br/satveg/login.html>). The images are acquired from the Land Processes Distributed Active Center (LP-DAAC), a center linked to NASA's Earth Observing System and the EVI are available in maximum 16-day compositions, with a spatial resolution of approximately 250m (Embrapa, 2014).

3.4 Data processing

For this study, the dataset was composed of 23 variables relevant to help explain soil respiration: soil-related (T_s ; SWC; P; pH; H+Al; SB; CEC; Macro; Micro; Bd; C stock; N stock; CN ; sand;silt and clay), agrometeorological (T_{air} ; P_{atm} ; RSG; PAR; rain; and W) and biophysical parameters (EVI vegetation index). FCO₂ totaled $n= 500$ observations and FO_2 $n = 175$ observations. Multivariate outliers were assessed using the Mahalanobis distance at 1% significance. As a data mining technique, principal component analysis (PCA) was used to reduce the size of the dataset. Subsequently, the components were used to train the three different neural network architectures, namely, generalized regression neural network (GRNN), radial basis function (RBF), and multilayer perceptron (MLPNN) networks, and three support vector kernel regression of RBF, sigmoid (SIG), and polynomial (PL). PCA

was performed using R software (version 4.0.2; R Core Team 2019, R Foundation for Statistical Computing, Vienna, Austria), using the R packages FactoMinerR (Le, Josse, and Husson, 2008), and factoextra (Kassambara and Mundt, 2017).

Data mining was performed using principal component analysis. The proposed AI approaches to predict FCO₂ and FO₂ included ANNs, GRNN, RBFNN, MLPNN, and SVR with sigmoid, radial basis function and polynomial kernels.

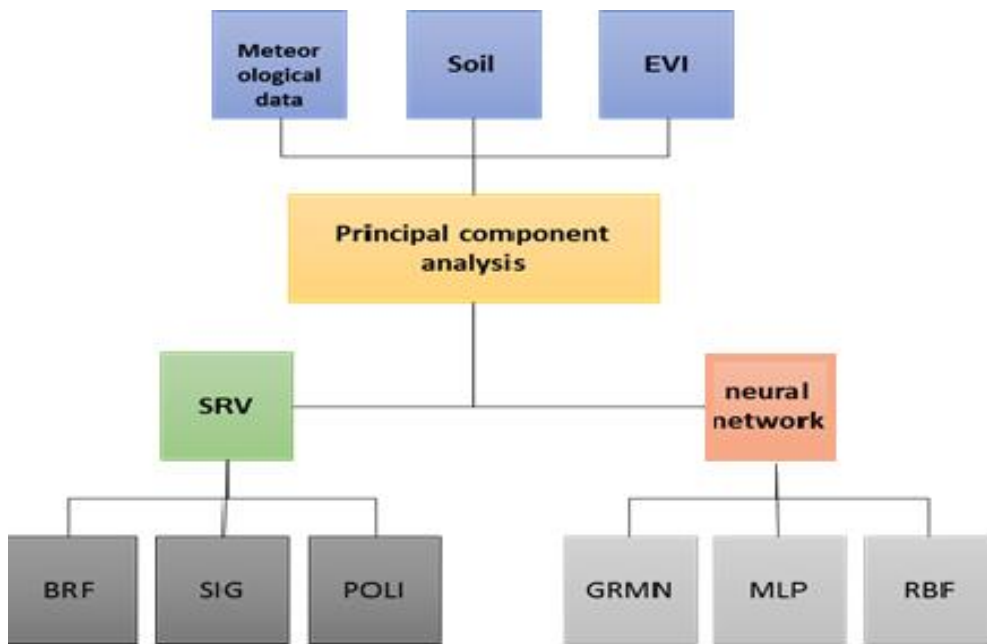


Figure 2: Flow chart of methodology used in this study.

3.4.1 Principal component analysis

The PCA replaces a set of correlated variables with new variables called principal components (PCs), which are uncorrelated and obtained as linear combinations of the original variables. Thus, the first PC should have the highest possible variance, and this component will explain part of the inertia of the data (Abdi and Williams, 2010). One of the main objectives of this algorithm is to reduce the size of the dataset (Jolliffe, 2002), discard variables, and exclude outliers.

The database consisting of the 23 variables was determined as the main components: PC1, PC2, PC3, and PC18. Neural networks and support vector regression (SVR) were used as inputs to the models. Table 1 presents the results of

the variances (eigenvalues) and explained and accumulated variances of the PCs obtained

Table 1. Variances and explained and accumulated variance of the PCs

Principal Component	Eigenvalue	Percentage of variance	Cumulative percentage of variance
PC1	7.62	33.15	33.15
PC2	3.49	15.16	48.31
PC3	2.12	9.23	57.55
PC4	1.77	7.71	65.25
PC5	1.66	7.24	72.49
PC6	1.46	6.37	78.86
PC7	1.02	4.41	83.27
PC8	0.89	3.86	87.13
PC9	0.76	3.31	90.44
PC10	0.67	2.91	93.35
PC11	0.5	2.17	95.52
PC12	0.35	1.52	97.04
PC13	0.23	1.01	98.05
PC14	0.17	0.73	98.78
PC15	0.13	0.58	99.36
PC16	0.08	0.35	99.71
PC17	0.06	0.25	99.96
PC18	0.01	0.04	100

The main PCs, those with variance greater than 0.70 (Jolliffe, 1972), correspond to the first nine PCs, which express 90.44% of the preserved variance of the original data. Therefore, the original 23 variables could be analyzed from the nine CPs. However, in the tests performed, we realized that keeping the 18 PCs for this dataset provided a better performance for ANNs and SVRs, especially for the GRNN model, which is one of the main disadvantages of this architecture.

3.4.2 Artificial neural networks

An artificial neural network is a set of mathematical processes and computational algorithms that simulate the human nervous system. Owing to its ability to recognize complex nonlinear patterns, it has strong error tolerance. The following is a brief description of the architecture used in this study.

The architecture of an MLPNN is variable, but in general, it consists of several layers of neurons. Intermediate layers are known as hidden layers between the input and output layers, which are trained by the backpropagation (BP) algorithm. The basic idea is to propagate the error signal calculated in the training step back to all neurons and the activation function of the sigmoidal type (Haykin, 2009).

RBFNN has three layers: an input layer, a hidden layer, and an output layer with linear processing units. The number of neurons in the hidden layer can be adjusted according to actual needs. Learning can be supervised, unsupervised, or blended, but when training is supervised, the weights between the hidden layer and output layer are calculated to minimize the error between the desired and obtained output. The most common activation function of the RBF is Gaussian. A GRNN was proposed as a variant of RBFNN by Specht in 1990 (Specht, 1991). It is an ANN with a fixed structure and consists of an input layer, a hidden layer, a summation layer, and an output layer. Thus, the number of neurons in each layer depends on the number of inputs and outputs of the ANN. The GRNN estimates continuous variables by performing a non-parametric Parzen regression (Jobran Al-Mahasneh et al., 2018) and subsequently computes the output of the network according to the maximum likelihood principle (Xiao et al., 2020). All training of ANNs is supervised.

In the present study, we trained 100 ANNs with MLP architecture and 100 RBF and 100 GRNN in FCO₂ and FO₂ predictions. In all ANN models, 70% of the data were used for calibration and 30% for validation. The training and validation data were randomly selected using the program. All ANN architectures were developed using the *Intelligent Problem Solver (IPS)* software Statistica® 7.0 (Statsoft Inc., Tulsa, OK, USA). We selected the model with the best performance for each architecture.

3.4.3 Support vector regression

Different support machines can be built depending on the learning algorithm, such as polynomial, radial basis, and sigmoid, defined in the choice of some important parameters of SVR, such as kernels. The kernel function helps to find the best hyperplane in space, and when applied in SVR, the best fit occurs when the hyperplane has the maximum number of points.

The generalization ability of the SVR model depends on the correct selection of parameters C and ε and the type of kernel function and its parameters. In this work,

the SVM package used Statistica® 7.0 (Statsoft Inc., Tulsa, OK, USA), which automatically adjusts the parameters C (cost), γ (gamma), and ε (epsilon) through cross-validation. We developed Regression SVM Type 1, offered by the SVM package.

Tipo I (ϵ -SVM for regression or ϵ -SVR)

$$\text{Minimize } \frac{1}{2} W^T W + C \sum_{i=1}^n (\xi_i + \xi_i^n) \quad \text{Eq. (4)}$$

$$\text{Subject to } \begin{cases} y_i - W^T \phi(X_i) - b \leq \varepsilon + \xi_i \\ W^T \phi(X_i) + b - y_i \leq \varepsilon + \xi_i^n \\ \xi_i \cdot \xi_i^n \geq 0, i = 1, \dots, n \end{cases} \quad \text{Eq. (5, 6, 7)}$$

where C is the regulation term, w is the vector of the parameters, b is a constant, and ξ is the variable slack of the error outside the accepted range ε , calculated by the support vectors. Index i labels n training cases. The term $\phi(X_i)$ represents the transformation of the input data (independent) by the kernel function $K(X_i, X_j)$ in the characteristic space:

$$K(X_i, X_j) = \phi(X_i) \cdot \phi(X_j). \quad \text{Eq. (8) adapted from Aquino de Souza et al. (2019).}$$

3.4.4 Evaluation criteria for model performance

The predictive capacity of the models was assessed based on the performance of the following statistical indices: Pearson correlation coefficient (r) (Eq. 9) mean absolute percentage error (MAPE) (Eq. 10), agreement index (d) (Eq. 11) (Willmott, 1981), and confidence coefficient (c) (Eq. 12) (Camargo and Sentelhas, 1997b). Analysis of variance was performed using the F test at the probability level of 0.01.

$$r = \frac{(N \sum_{i=1}^N (X_{obs1} Y_{obs1}) - \sum_{i=1}^N X_{obs1} \sum_{i=1}^N Y_{obs1}) \left((N \sum_{i=1}^N Y_{obs1})^2 - (\sum_{i=1}^N Y_{obs1})^2 \right)}{\sqrt{\left((N \sum_{i=1}^N X_{obs1})^2 - (\sum_{i=1}^N X_{obs1})^2 \right) \left((N \sum_{i=1}^N Y_{obs1})^2 - (\sum_{i=1}^N Y_{obs1})^2 \right)}} \quad \text{Eq. (9)}$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left(\frac{|(A_i - P_i)|}{A_i} 100 \right) \quad \text{Eq. (10)}$$

$$d = 1 - \frac{\sum_{i=1}^N (Y_{obs1} - Y_{est1})^2}{\sum_{i=1}^N (|Y_{est1} - \bar{Y}| + |Y_{obs1} - \bar{Y}|)^2} \quad \text{Eq. (11)}$$

$$c = r * d \quad \text{Eq. (12)}$$

where n is number of data points; X_{obs_1} is the observed value of X ; Y_{obs_1} is the observed value of Y ; Y_{est_1} is the estimated value of Y ; and $\bar{Y}$ is the observed mean of Y .

3.5 Results and Discussion

3.5. 1 Model performance for FCO₂

Table 2 presents the error rates for the ANN and SVR models in the calibration and validation phases. All models were significant ($p < 0.01$). In this study, as explained earlier, we used principal component analysis to decrease the size of the dataset. In the set of 23 variables, nine principal components demonstrated higher explanatory power ($r = >70$) (Table 1). However, by performing preliminary tests, we realized that for ANNs, it was more predictive to retain the set of 18 CPs, thereby contributing to the best model performance. The variables correlated with these components and were part of the input set of the ANN were T_s ; SWC; P; pH; H + Al; SB; CEC; macro; micro; Bd, Cstock; StockN; CN; sand; silt; clay; RSG; and PAR.

In the calibration phase, the ANNs showed low RMSE indices (0.780 to 0.952 $\mu\text{mol m}^{-2} \text{s}^{-1}$) and MAPE ranged from 16.83% to 20.23% (Table 2). As for the accuracy, all models showed good agreement with agreement indices d (Willmott, 1981) having values above 0.83. The adjustment index d ranges from 0 to 1, where 0 and 1 represent total disagreement or perfect agreement between the measurements and estimates, respectively. MLPNN had the best index with $d = 0.92$. According to the classification index (Camargo and Sentelhas, 1997), the good fit between the observed and estimated values in GRNN, MLPNN, and RBFNN were classified as average, very good, and good, respectively. Furthermore, it can be observed that MLPNN presented the highest determination coefficient (0.65) followed by RBFNN (0.63) and GRNN (0.60).

For validation (Table 2), the RMSE values ranged from 0.820 to 1.034 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Generally, RMSE is proportional to the square of the errors and is sensitive to large deviations between the observed and predicted values. The MAPE values varied between 21.39% and 23.17%. Regarding d -index and coefficient of determination, MLP showed the best indices with $d = 0.867$ and $R^2 = 0.59$. Moreover,

MLPNN obtained a "good" rating, whereas GRMN and RBF were considered to be "average". The R^2 values for GRNN and RBFNN were 0.58 and 0.57, respectively.

Regarding the architecture of GRNN, the number of neurons in the input layer was equal to the number of predictors, i.e., 16. According to (Ghritlahre and Prasad, 2018), for GRNNs, the number of input variables depends on the number of datasets observed in the data collection. Typically, the number of neurons in the pattern layer is equal to the number of training datasets. However, GRNN and RBFNN showed a higher number of neurons in the intermediate layer compared to that of MLPNN. This indicates that both the RNA models were complex. Li et al. (2020) described that it is a characteristic of RBF to use a large number of highly interconnected neurons in the intermediate layer, especially when there is low correlation in the dataset, because the network requires a larger number of neurons to recognize linear patterns. Thus, 12 parameters in the input layer and 10 neurons in the intermediate layer was a good configuration to estimate FCO_2 .

Regarding SVR models, generally, SVR- RBF showed superior performance in the calibration and validation phases compared to SVR- SIG and SVR - PL-. The Gaussian kernel is generally more accurate in terms of the data used (VanDeventer et al., 2019). The SVR models showed lower RMSE values for the calibration and validation phases compared to the ANN models. The calibration data ranged from 0.690 to 0.820 $\mu\text{mol m}^{-2} \text{s}^{-1}$. The validation data ranged from 0.77 to 0.990 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Although SVR-RBF showed the highest RMSE value, it obtained the lowest value for MAPE calibration and validation (21% and 22%) and the highest value for d -index calibration and validation (0.82 and 0.817), respectively.

Overall, the MAPE values in SVR ranged from 21.60% to 23% for calibration and 22.23% to 24.60% for validation. The index d ranged from 0.735 to 0.820 for calibration and 0.714 to 0.817 for validation. These results corroborated the performance rating for SVR –RBF of average and sufferable for calibration and validation, respectively. SVR-SIG was rated sufferable in both phases, whereas SIG PL was rated sufferable and bad.

In all ANN training phases, the correlations (r) were higher than 0.75, whereas in SVR they ranged from 0.68 to 0.76. The adjusted R^2 penalizes the inclusion of poorly explained regressions and demonstrates that all ANNs obtained higher precision in both training phases. In validation, the models explained 57% to 59% of

the temporal variability of FCO_2 , while the SVR models explained 46% to 53% of the variability of FCO_2 . For estimating FCO_2 , MLPNN was better than SVR- RBF.

3.5.2 Model performance for FO_2

For this estimation FO_2 we used 18 PSCs. Table 3 presents the error rates for ANNs and SVRs. For ANNs, the models were significant ($p < 0.01$). First, the values of the RBF kernel function parameters were determined. The RBF parameters, C , ϵ , and γ , were fit.

The performance of the predictive models for FCO_2 and FO_2 , along with the configuration of their kernel functions, namely SVR RBF, second-degree polynomial (SVR - PL), and sigmoid (SVR - SIG), were as follows: $\{\text{FCO}_2[\text{SVR- RBF} (C = 4000, \epsilon = 0.200, \gamma = 0.056), \text{SVR - PL:} (C = 4000, \epsilon = 0.200, \gamma = 0.056), \text{ and SVR - SIG}(C = 10.00, \epsilon = 0.100, \gamma = 0.50)]\}$; $\{\text{FO}_2[\text{SVR - RBF:} (C = 10.00, \epsilon = 0.500, \gamma = 0.056); \text{SVR - PL}(C = 10.00, \epsilon = 0.500, \gamma = 0.056, \text{ degree} = 3), \text{ and SVR - SIG}(C = 10.00, \epsilon = 10.00, \gamma = 0.56)]\}$.

For SVR, the models were significant only during the calibration phase. In this study, all the models presented high MAPE values. For an ANN with a calibration greater than 46% and a calibration of 100%. For the SVR models in calibration and validation, higher than 100%. Such a result would make our estimates unfeasible. However, it is important to note that the dataset of the dependent variable presents very small values close to zero. In the original dataset, the FO_2 values ranged from 0.0148 to 1.796 $\text{mg m}^{-2}\text{s}^{-1}$ with a mean value of 0.369 $\text{mg m}^{-2}\text{s}^{-1}$.

According to (Kim and Kim, 2016), one of the disadvantages of MAPE is that it produces infinite or undefined values when the actual values are zero or close to zero. Thus, if the actual values are very small (usually less than one), MAPE will produce extremely large percentage errors (outliers), whereas zero actual values will result in infinite MAPEs. Therefore, for this training FO_2 we considered only the RMSE and d -index metrics as measures of estimation accuracy (Willmott, 1981).

The agreement index is a complement to non-square measures, such as RMSE, as it is based on absolute errors. Unlike RMSE, d is less sensitive to outliers (Achieng, 2019). In the calibration phase, ANNs achieved low RMSE values (0.037 to 0.074 $\mu\text{mol m}^{-2} \text{s}^{-1}$) and d indices (0.835 to 0.911). For validation, RMSE values ranged from 0.076 to 0.089 $\text{mg m}^{-2}\text{s}^{-1}$, whereas d indices ranged from 0.546 to

0.730. The good fit of the models according to the classification index (Camargo and Sentelhas, 1997b) classified GRNN, MLPNN, and RBFNN as very good, very good, and average, respectively. For validation, all performances were “bad”. MLPNN had the best accuracy with the validation value of coefficient of determination being 0.36 compared to 0.34 of GRNN and RBFNN (Table 2; Figure 4). For SVR, as the regression in validation was not significant, it is not discussed here.

Table 2. Performance of statistical indices during the training and validation phases of artificial neural networks and support vector regression

Soil CO ₂ emission (μmol m ⁻² s ⁻¹) (FCO ₂)											
	Architecture		<i>r</i>	<i>R</i> ²	RMSE	MAPE	d	Classification (C)		<i>F</i>	<i>p</i> -valor
RNA 1	GRNN	Calibration	0.77	0.6	0.78	20.233	0.836	0.646	Median	349	<i>p</i> < 0.0001
	15-251-2-1	Validation	0.76	0.58	0.82	21.393	0.829	0.63	Median	144.7	<i>p</i> < 0.0001
RNA 2	MLPNN	Calibration	0.86	0.75	0.932	16.839	0.923	0.797	Very Good	708.2	<i>p</i> < 0.0001
	12--12--10-1	Validation	0.77	0.59	1.034	22.745	0.867	0.668	Good	172.7	<i>p</i> < 0.0001
RNA 3	RBFNN	Calibration	0.8	0.63	0.952	18.729	0.883	0.703	Good	400.7	<i>p</i> < 0.0001
	11-44-1-1	Validation	0.75	0.57	1.008	23.179	0.848	0.638	Median	146.3	<i>p</i> < 0.0001
SVR											
Soil CO ₂ emission (μmol m ⁻² s ⁻¹) (FCO ₂)											
			<i>r</i>	<i>R</i> ²	RMSE	MAPE	d	Classification (C)		<i>F</i>	<i>p</i> -valor
SVR 1	SVR-RBF	Calibration	0.76	0.58	0.82	0.216	0.82	0.623	Median	493.8	<i>p</i> < 0.0001
		Validation	0.73	0.53	0.99	0.223	0.817	0.599	Suffering	141.9	<i>p</i> < 0.0001
SVR 2	SVR-SIG	Calibration	0.69	0.47	0.8	0.23	0.763	0.525	Suffering	311.3	<i>p</i> < 0.0001
		Validation	0.68	0.46	0.92	0.243	0.761	0.521	Suffering	107.7	<i>p</i> < 0.0001
SVR 3	SVR-PL	Calibration	0.69	0.48	0.69	0.23	0.735	0.51	Suffering	321.1	<i>p</i> < 0.0001
		Validation	0.68	0.46	0.77	0.246	0.714	0.488	Suffering	107.2	<i>p</i> < 0.0001

The sequence of numbers indicate the number of input variables, number of neurons in the first layer, and number of neurons in the second layer; *r*: Pearson's correlation coefficient; MAPE: mean absolute percentage error; *d*: Willmott's agreement index; *C*: confidence coefficient; *p*-value: probability of the occurrence of H0. SVR-RBF = support vector machine kernel radial basis function; SVR-SIG = sigmoid; SVR-PL.

Table 3. Performance of statistical indices during the training and validation phases of artificial neural networks and support vector regression for influx O_2

			Influx O_2 (mg m ⁻² s ⁻¹) (FO ₂)								
	Architecture		<i>r</i>	<i>R</i> ²	RMSE	MAPE	<i>D</i>	Classification (C)		F	<i>p</i> -valor
ANN 1	GRNN	Calibration	0.89	0.79	0.074	46.32	0.871	0.775	Very Good	331	<i>p</i> < 0.0001
	18-89-2-1:1	Validation	0.57	0.34	0.084	0.741	0.546	0.316	Bad	19.66	<i>p</i> < 0.0001
ANN 2	MLPNN	Calibration	0.85	0.73	0.730	0.73	0.911	0.781	Very Good	232.4	<i>p</i> < 0.0001
	13--11--11	Validation	0.59	0.36	0.076	57.28	0.675	0.405	Bad	19.61	<i>p</i> < 0.0001
ANN 3	RBFNN	Calibration	0.74	0.55	0.037	71.06	0.835	0.618	Mediano	96.82	<i>p</i> < 0.0001
	14--22-1	Validation	0.58	0.34	0.089	100.5	0.730	0.424	Bad	16.87	<i>p</i> < 0.01
			Influx O_2 (mg m ⁻² s ⁻¹) (FO ₂)								
SVR 5	SVR - RBF	Calibration	0.35	0.11	0.121	1.702	0.515	0.180	Bad	17.04	<i>p</i> < 0.0001
		Validation	0.13	0.01	0.080	1.124	0.331	0.045	Bad	0.95	0.3352
SVR 6	SVR - SIG	Calibration	0.31	0.09	0.079	1.718	0.464	0.464	Bad	13.34	<i>p</i> < 0.0001
		Validation	0.14	0.00	0.055	1.146	0.317	0.045	Bad	1.01	0.319
SVR 7	SVR- PL	Calibration	0.33	0.105	0.047	1.703	0.441	0.148	Bad	15.42	<i>p</i> < 0.0001
		Validation	0.0	0.004	0.039	1.151	0.292	0.018	Bad	0.2052	0.6526

The sequence of numbers indicate the number of input variables, number of neurons in the first layer, and number of neurons in the second layer, respectively; *r*: Pearson's correlation coefficient; MAPE: mean absolute percentage error; *d*: Willmott's agreement index; C: confidence coefficient; *p*-value: probability of the occurrence of H₀. SVR - RBF = support vector machine kernel radial basis function; SVR-SIG = sigmoid; SVR - PL = polynomial

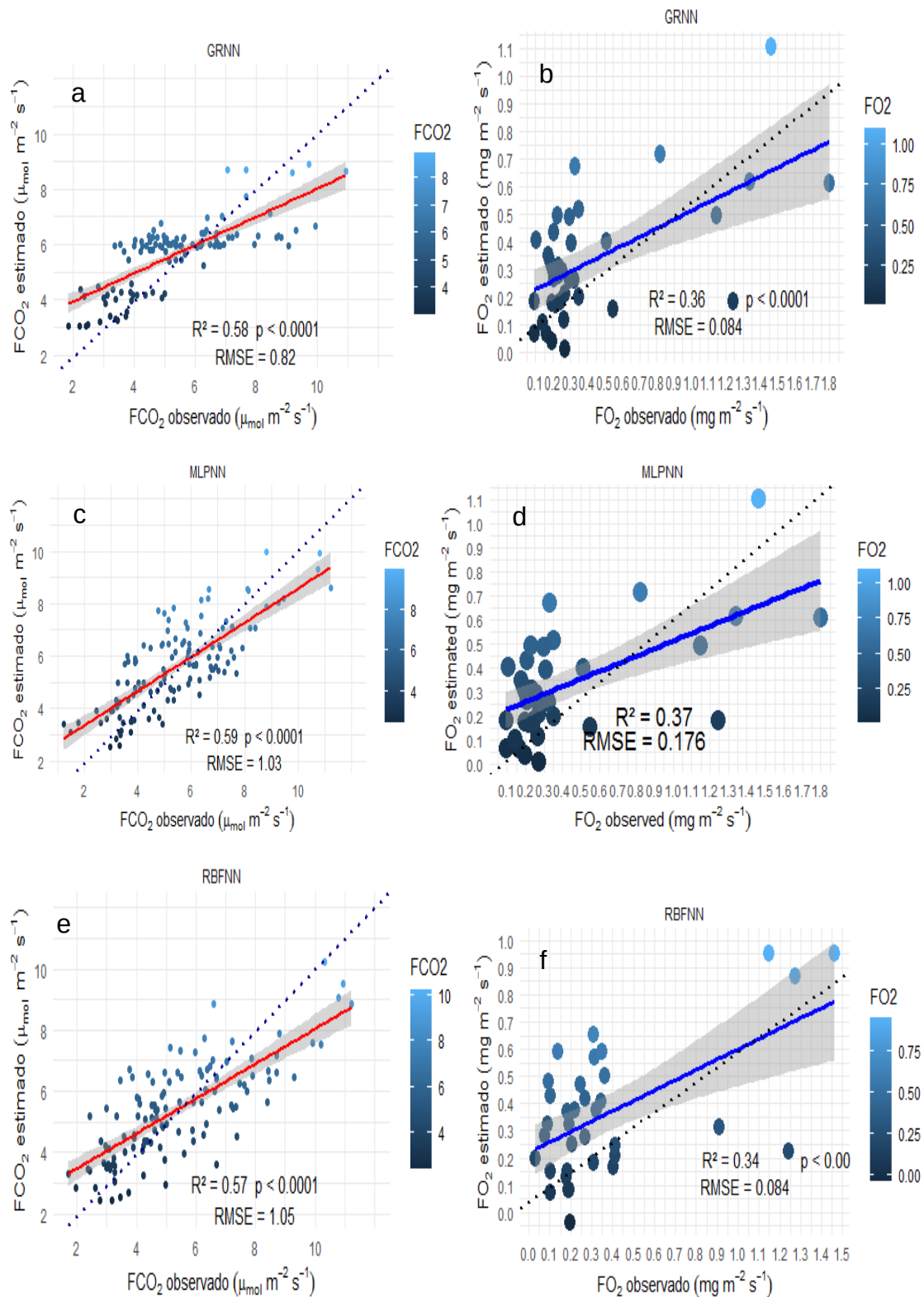


Figure 4. Performance of soil CO₂ emission ($\mu\text{mol m}^{-2} \text{s}^{-1}$) (FCO₂) and Influx O₂ ($\text{mg m}^{-2} \text{s}^{-1}$) (FO₂) during the validation phases of artificial neural networks (GRNN; MLP and RFB). Coefficient of determination (R^2) and RMSE = mean absolute percentage error. a) GRNN – FCO₂; b) GRNN – FO₂; c) MLPNN - FCO₂; d) MLPNN - FO₂; e) RBFNN - FCO₂; f) RBFNN - FO₂.

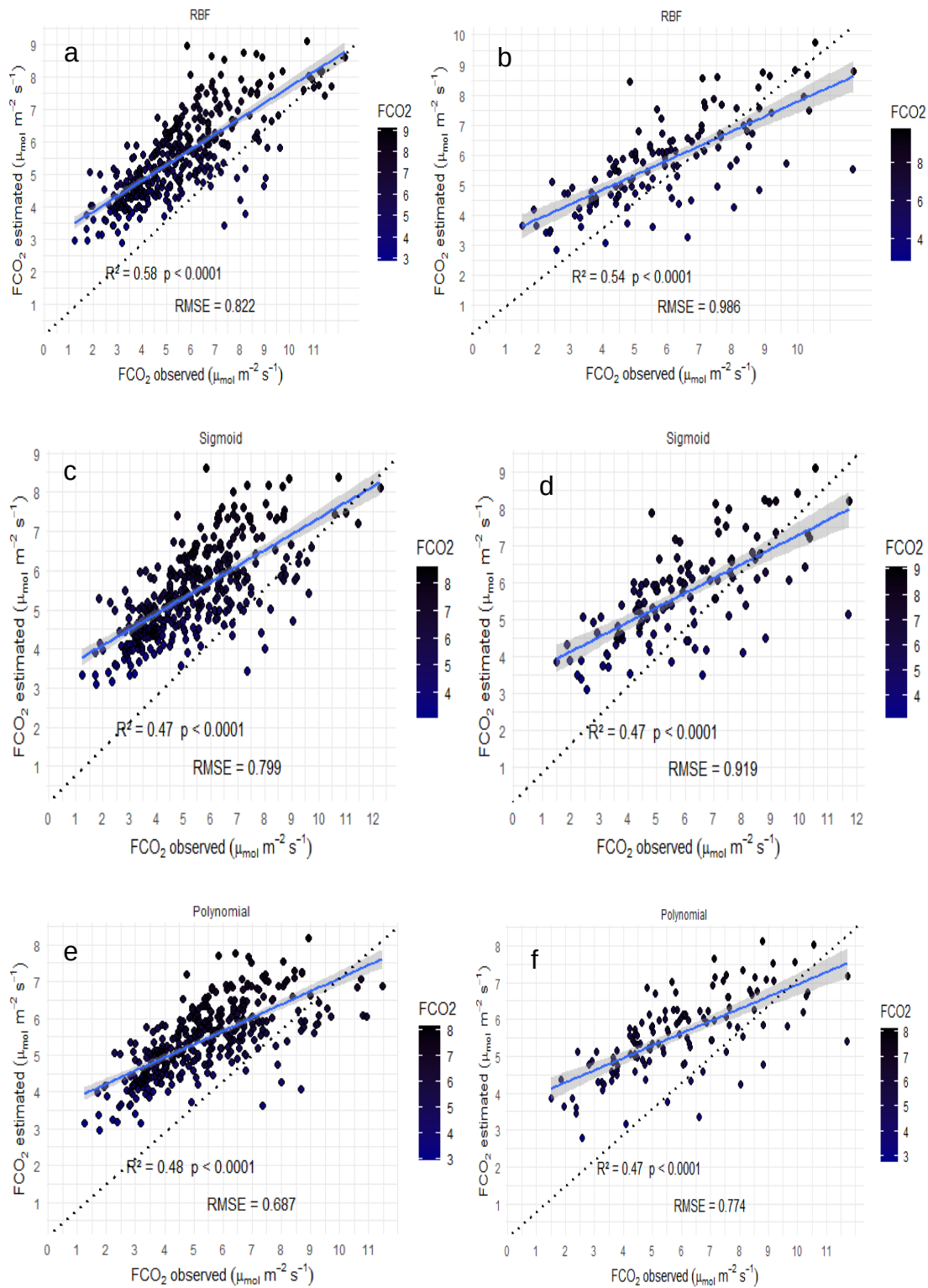


Figure 5. Performance graphs of the calibration and validation of soil CO₂ emission (FCO₂) with radial basis function (RBF), sigmoid and polynomial kernels. Coefficient of determination (R^2) and RMSE = mean absolute percentage error RBF: a) calibration and b) validation. Sigmoid: c) calibration and d) validation. Polynomial: e) calibration and f) validation.

3.5.3 Importance of Predictor Variables

In our study, we used a set of 23 original variables (climate, soil, PAR and EVI) to model the prediction of FCO_2 and FO_2 . Principal component analysis assisted in reducing the dimensionality of the dataset, resulting in 18 PCs. To understand how each variable impacted each component, and analyze the inter-relationships between the large number of variables with FCO_2 and FO_2 dynamics in the eucalyptus forest soil, we generated a biplot (Figure 3) from the original 23 variables, and FCO_2 and FO_2 , resulting in a dataset with 25 variables.

The first two components represented 30.5% (PC1) and 14.4% (PC2) of the total variance of the original data (Figure 3), thus explaining 44.9% of the original variance. Among the 25 variables, the ones that contributed the most to PC1 in order of relevance were: clay (0.90), pH (0.88), SB (0.80), macro (0.75), silt (0.66), P (0.65), Cstock (0.59), CEC (0.58), and Stock N (0.54) and inversely correlated (-0.94), (-0.89), (-0.83), (-0.59), and (-0.38). For PC2, in order of relevance, the attributes that contributed the most were: T_{atm} (0.89), GSR (0.78), PAR (0.75), T_s (0.72), EVI (0.65), FCO_2 , and FO_2 (0.23) and inversely correlated (-0.61) and (-0.35) (Table 4). The attributes FCO_2 , FO_2 , T_s , PAR, EVI, RSG, and T_{atm} were all projected onto PC2 in the same direction.

In general, it is evident that the processes related to PC1 indicate the stability of C in the studied ecosystem. The lack of correlation between FCO_2 and FO_2 could be a result of land use change and the long period of reforestation which contributed to the maintenance of the C stock. Probably, after deforestation of the native vegetation for agricultural purposes in the study area, an intense decay of organic matter occurred in the soil. This would have caused a rapid decline in C stocks (Lal et al. 1997), especially in the first 10 years of conversion, owing to organic matter with high lability (Melillo et al. 2002) promoting changes in N stocks. During this period, it is possible that soil respiration was controlled by biological activity.

Table 4. Correlation between FCO₂, FO₂, environmental covariates, and the first two principal components (PC1 and PC2) in the eucalyptus forest

Principal componente	PC1	PC2
Explained variance (%)	30.5	14.4
Correlation		
FCO ₂	-0.07	0.32
FO ₂	-0.05	0.23
T _s	0.03	0.72
SWC	-0.08	0.02
P	0.65	-0.03
pH	0.88	-0.03
H.Al	-0.59	0.00
SB	0.86	-0.05
CEC	0.58	-0.08
Macro	0.75	0.06
Micro	-0.89	-0.06
Bd	-0.83	-0.03
C stock	0.59	-0.03
N stock	0.54	0.08
CN	-0.38	-0.13
Sand	-0.94	-0.03
Silt	0.66	0.14
Clay	0.90	0.01
T _{air}	-0.05	0.89
P _{atm}	0.06	-0.61
RSG	-0.05	0.78
PAR	-0.04	0.75
Rain	0.01	-0.05
W	0.01	-0.35
EVI	-0.01	0.65

FCO₂= soil CO₂ emission; FO₂= oxygen uptake; T_s (°C) = soil temperature; SWC = soil water content; P (mg dm⁻³); pH in CaCl₂ = hydrogen potential; H+Al; SB - sum of bases (K⁺, Ca²⁺, and Mg²⁺); CEC (mmolc·dm⁻³) = cation exchange capacity; Macro (m³·m⁻³) = macroporosity; Micro (m³·m⁻³) = microporosity; Bd = bulk density; (g dm⁻³); C stock = ; N stock = ; CN ratio; Sand = sand content (g kg⁻¹); Silt = silt content (g kg⁻¹); Clay = clay content (g kg⁻¹); T_{air} = air temperature; P_{atm} (kPa) = atmospheric pressure; GSR (MJ·m⁻²·day⁻¹) = global solar radiation; PAR (μmol·m⁻²) = photosynthetically active radiation; Rain; W = wind velocity w (m/s⁻¹); EVI.

However, over the years, subsequent to the humidification processes of microorganisms, aggregation, and sedimentation, most of the organic matter was transformed into humidified material of low lability and resistance to decomposition, corroborating the C stock in the soil (Lal 2007). Thus, soil respiration in this ecosystem was not because of heterotrophic respiration and may be inert from a biological point of view (Brady and Weil 2013). Therefore, the high C stock and low FCO₂ (Silva et al., 2020) explains the lack of correlation between FCO₂ and FO₂ with PC1.

As mentioned earlier, the database was from an area that had undergone more than 30 years of reforestation; thus, in mid-aged or mature forests, soil C stocks increase until equilibrium in their dynamics is established (Pussinen et al., 2002). In this ecosystem, the eucalyptus burlap is responsible for nutrient cycling, maintenance of natural soil stability, and is the main source of energy for microorganisms (Brady and Weil 2012). The C/N ratio influences the speed of burlap decomposition, favoring the ability of decomposer microorganisms to produce assimilable forms of nitrogen (Xavier et al., 2019), which explains the inverse correlation in PC1. In soil, the C stock is determined by the input of organic matter and land use (Dawson; Smith, 2007), decomposition, mineralization, and leaching (Zhou et al., 2013). However, this dynamic is dependent on several factors, such as substrate quality, vegetation type, soil fertility, and climate and soil physical structure (Smith, 2008).

As shown in Figure 3, there were direct associations between C and N stocks and soil fertility-related attributes, and all vectors projected to the same side. Soil texture (silt and clay) is related to the stabilization of C (Aduan, 2003; Nair et al. 2011) and N (Van Noordwijk et al., 1997), and to improvements in soil fertility attributes (Barlow et al., 2018) which reflect the improvement of soil physical quality and climate (Dawson and Smith 2007; Smith, 2008).

In relation to PC2, the identified associations supported that the dynamics of FCO_2 and FO_2 are related to root respiration. Previous studies indicate that solar radiation indirectly influences soil respiration through direct effects on photosynthesis (Moyano et al., 2008), and by influencing the amount of root-exuded C (Nakayama and Tatenno, 2018)(Liu et al. 2006; and the vegetation index (EVI) (Wang et al., 2010). The low contribution of water content to soil respiration can be explained by covariation between water content and soil temperature (Adachi 2013).

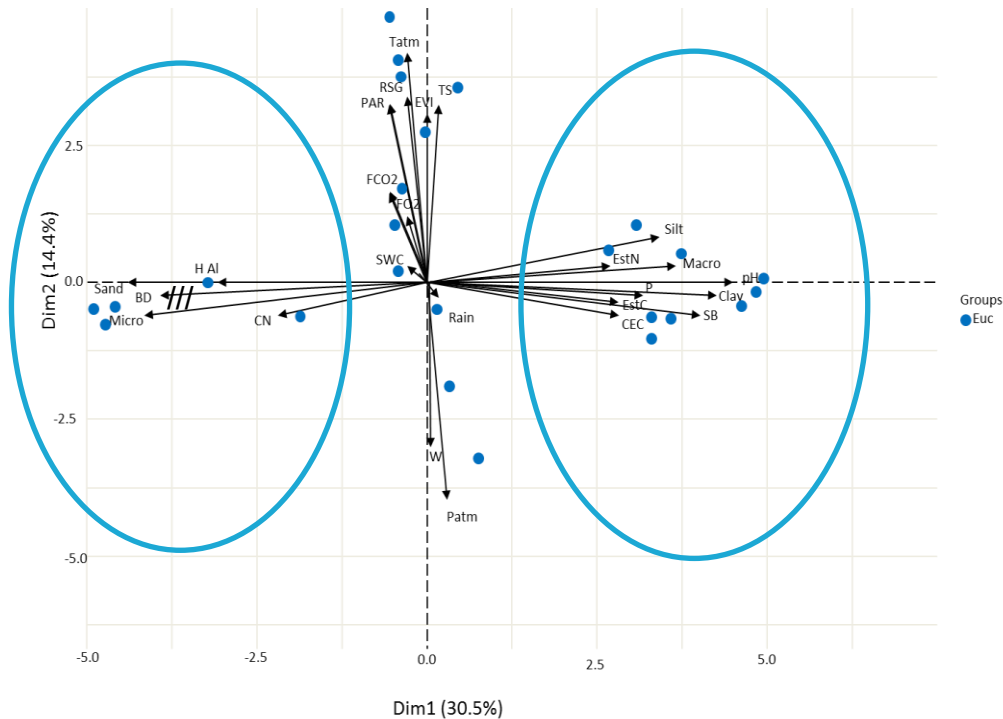


Figure 3. Biplot of first and second principal components (PC1 and PC2)

3.6 Conclusion

Regarding the machine learning models, our study showed that ANNs outperformed SRV models in both FCO_2 and FO_2 estimations. An MLP ANN showed the best predictive performance for FCO_2 ($R^2 = 0.75$ and 0.59) when compared to the other ANNs. For FO_2 , GRNN and MLPNN performed similarly. However, the GRNN architecture was more complex. Thus, we consider MLPNN the best model for estimating FCO_2 ($R^2 = 0.75$ and 0.36). In all estimates in the SRV, the best kernel was RBF. The SVR validation regressions were not significant for the FO_2 .

Although the principal component analysis summarized the dimensions of the data, all models were calibrated with a higher number of components than indicated by the analysis. This highlights the complexity of the data used in our study. Principal component analysis indicated that soil C was stabilized in a eucalyptus forest more than 30 years after land use conversion. FCO_2 and FO_2 were associated with root respiration, as evidenced by the variables related to the PAR and EVI indices. Further studies with new data mining techniques.

3.7 References

Abdi H, Williams LJ (2010) Principal component analysis. *Wiley Interdiscip Rev Comput Stat* ;2:433-59. <https://doi.org/10.1002/wics.101>

Achieng KO (2019) Modelling of soil moisture retention curve using machine learning techniques: Artificial and deep neural networks vs support vector regression models. **Computers & Geosciences – Journals**. 133:104320. <https://doi.org/10.1016/j.cageo.2019.104320>

Adachi M, Bekku YS, Rashidah W, Okuda T, Koizumi H (2006) Differences in soil respiration between different tropical ecosystems. **Applied Soil Ecology** 34:258-65. <https://doi.org/10.1016/j.apsoil.2006.01.006>

Alves JDN, Ribeiro A, Rody YP, Loos RA, Hall KB (2021) Carbon uptake and water vapor exchange in a pasture site in the Brazilian Cerrado. **Journal of Hydrology**. 594. <https://doi.org/10.1016/j.jhydrol.2020.125943>

Bahn M, Rodeghiero M, Anderson-Dunn M, Dore S, Gimeno C, Drösler M, Williams M, Ammann C, Berninger F, Flechard C, Jones S, Balzarolo M, Kumar S, Newesely C, Priwitzer T, Raschi A, Siegwolf R, Susiluoto S, Tenhunen J, Wohlfahrt G, Cernusca (2008). A. Soil respiration in European grasslands in relation to climate and assimilate supply. **Ecosystems** ;11:1352-67. <https://doi.org/10.1007/s10021-008-9198-0>

Barlow J, França F, Gardner TA, Hicks CC, Lennox GD, Berenguer E, Castello L, Economo EP, Ferreira J, Guénard B, Gontijo Leal C, Isaac V, Lees AC, Parr CL, Wilson SK, Young PJ, Graham NAJ (2018). The future of hyperdiverse tropical ecosystems. **Nature**. 559:517-26. <https://doi.org/10.1038/s41586-018-0301-1>

Batlle-Bayer L, Batjes NH, Bindraban PS (2010) Changes in organic carbon stocks upon land use conversion in the Brazilian Cerrado: A review. **Agriculture, Ecosystems & Environment** ;137:47-58. <https://doi.org/10.1016/j.agee.2010.02.003>

Biudes MS, Vourlitis GL, Machado NG, de Arruda PHZ, Neves GAR, de Almeida Lobo F, Neale CMU, de Souza Nogueira J(2015). Patterns of energy exchange for tropical ecosystems across a climate gradient in Mato Grosso, Brazil. **Agricultural and Forest Meteorology** , 202:112-24. <https://doi.org/10.1016/j.agrformet.2014.12.008>

Bonini I, Hur Marimon-Junior B, Matricardi E, Phillips O, Petter F, Oliveira B, Marimon BS (2018). Collapse of ecosystem carbon stocks due to forest conversion to soybean plantations at the Amazon-Cerrado transition. **Forest Ecology and Management** ;414:64-73. <https://doi.org/10.1016/j.foreco.2018.01.038>

Bond-Lamberty B (2018). New Techniques and Data for Understanding the Global Soil Respiration Flux. **Earth's Future** ;6:1176-80. <https://doi.org/10.1029/2018EF000866>

Brady NC, Weil RR (2013). **Elementos da natureza e propriedades dos solos**. 3ª ed. Porto Alegre: Bookman, 2013. 704p

Camargo AP de, Sentelhas PC (1997) Performance evaluation of different potential evapotranspiration estimating methods in the State of São Paulo, Brazil. **Revista Brasileira de Agrometeorologia**. 5:89-97

Cerri CEP, Easter M, Paustian K, Killian K, Coleman K, Bernoux M, Falloon P, Powlson DS, Batjes NH, Milne E, Cerri CC (2007). Predicted soil organic carbon stocks and changes in the Brazilian Amazon between 2000 and 2030. **Agriculture, Ecosystems & Environment** ;122:58-72.
<https://doi.org/10.1016/j.agee.2007.01.008>

Conant RT, Paustian K, Del Grosso SJ, Parton WJ (2005) Nitrogen pools and fluxes in grassland soils sequestering carbon. **Nutrient Cycling in Agroecosystems** ;71:239-48. <https://doi.org/10.1007/s10705-004-5085-z>

Chen LF, He Z Bin, Wu XR, Du J, Zhu X, Lin PF, Tian QY, Kong JQ (2021) Linkages between soil respiration and microbial communities following afforestation of alpine grasslands in the northeastern Tibetan Plateau. **Applied Soil Ecology** ;161:103882.
<https://doi.org/10.1016/j.apsoil.2021.103882>

D'Andréa AF, Naves Silva ML, Curi N, Guimarães Guilherme LR (2004) Carbon and nitrogen storage, and inorganic nitrogen forms in a soil under different management systems. **Pesquisa Agropecuária Brasileira** ;39:179-86.
<https://doi.org/10.1590/s0100-204x2004000200012>

Dawson JJC, Smith P (2007) Carbon losses from soil and its consequences for land-use management. **Science of the Total Environment** ;382:165-90.
<https://doi.org/10.1016/j.scitotenv.2007.03.023>

Dogan E, Sengorur B, Koklu R (2009). Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. **Journal of Environmental Management** ;90:1229-35.
<https://doi.org/10.1016/j.jenvman.2008.06.004>

Ebrahimi M, Sarikhani MR, Safari Sinigani AA, Ahmadi A, Keesstra S. Estimating the soil respiration under different land uses using artificial neural network and linear regression models. **Catena**. 2019;174:371-82.
<https://doi.org/10.1016/j.catena.2018.11.035>

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

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>

Gatto A, Barros NF de, Novais RF, Silva IR da, Leite HG, Leite FP, Villani EM de A (2010). Estoques de carbono no solo e na biomassa em plantações de eucalipto.

Revista Brasileira Ciência Do Solo ;34:1069-79. <https://doi.org/10.1590/s0100-06832010000400007>

Greenwood, D. J (1961) The effect of oxygen concentration on the decomposition of organic materials in soil. **Plant and Soil**, 360-376 p.

Ghritlahre HK, Prasad RK (2018) Investigation of thermal performance of unidirectional flow porous bed solar air heater using MLP, GRNN, and RBF models of ANN technique. **Thermal Science and Engineering Progress (TSEP)**;6:226-35.

<https://doi.org/10.1016/j.tsep.2018.04.006>

Heimann M, Reichstein M. Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature*. 2008;451:289-92. <https://doi.org/10.1038/nature06591>

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>

Jha K, Doshi A, Patel P, Shah M (2019). A comprehensive review on automation in agriculture using artificial intelligence. **Artificial Intelligence in Agriculture**; 2:1-12. <https://doi.org/10.1016/j.aiia.2019.05.004>

Jobran Al-Mahasneh A, Anavatti S, Garratt and Mahardhika Pratama M (2018) Applications of General Regression Neural Networks in Dynamic Systems.

Digital Systems. <https://doi.org/10.5772/intechopen.80258>

Jolliffe I.T (1972) **Discarding Variables in a Principal Component Analysis. I: Artificial Data**. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 2, 2,160- 173Pp.

Jolliffe, I. T (2002). **Principal component analysis for special types of data. Principal component analysis**, 338-372 p.

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

Kim S, Kim H (2016) A new metric of absolute percentage error for intermittent demand forecasts. **International Journal of Forecasting**, 32:669-79. <https://doi.org/10.1016/j.ijforecast.2015.12.003>

Kuzyakov Y, Gavrichkova (2010). O. REVIEW: Time lag between photosynthesis and carbon dioxide efflux from soil: A review of mechanisms and controls. **Global Change Biology** ;16:3386-406. <https://doi.org/10.1111/j.1365-2486.2010.02179.x>

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

Lal, R (2007) Carbon management in agricultural soils. **Mitigation and adaptation strategies for global change**,12, 2, 303-322. <https://doi.org/10.1007/s11027-006-9036-7>

Lê, Sébastien; Josse, Julie; Husson, François (2008) FactoMineR: an R package for multivariate analysis. **Journal of statistical software**, 25, 1, 1-18. <https://doi.org/10.18637/jss.v025.i01>

Milfont TL, Abrahamse W, MacDonald EA (2021) Scepticism of anthropogenic climate change: Additional evidence for the role of system-justifying ideologies. **Personality and Individual Differences**; 168:110237. <https://doi.org/10.1016/j.paid.2020.110237>

Moreira CP, Bertini SCB, Ferreira AS, Azevedo LCB (2021). Biochemical activity and microbial biomass in wetlands (Vereda) and well-drained soils under native vegetation types in Brazilian Cerrado. **Applied Soil Ecology** ;160. <https://doi.org/10.1016/j.apsoil.2020.103840>

Moyano FE, Kutsch WL, Rebmann C (2008). Soil respiration fluxes in relation to photosynthetic activity in broad-leaf and needle-leaf forest stands. **Agricultural and Forest Meteorology**, 148:135-43. <https://doi.org/10.1016/j.agrformet.2007.09.006>

Nakayama M, Tateno R (2018) Solar radiation strongly influences the quantity of forest tree root exudates. **Trees**; 32:871-9. <https://doi.org/10.1007/s00468-018-1685-0>

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>

Van Noordwijk M, Cerri C, Woomer PL, Nugroho K, Bernoux M (1997) Soil carbon dynamics in the humid tropical forest zone. **Geoderma** ;79:187-225. [https://doi.org/10.1016/S0016-7061\(97\)00042-6](https://doi.org/10.1016/S0016-7061(97)00042-6)

Panosso AR, Marques J, Pereira GT, La Scala N (2009) Spatial and temporal variability of soil CO₂ emission in a sugarcane area under green and slash-and-burn

managements **Soil and Tillage Research**, 105:275-82.
<https://doi.org/10.1016/j.still.2009.09.008>

Pussinen A, Karjalainen T, Mäkipää R, Valsta L, Kellomäki S (2002) Forest carbon sequestration and harvests in Scots pine stand under different climate and nitrogen deposition scenarios. **Forest Ecology and Management**; 158:103-15.
[https://doi.org/10.1016/S0378-1127\(00\)00675-7](https://doi.org/10.1016/S0378-1127(00)00675-7)

Ranković V, Radulović J, Radojević I, Ostojić A, Čomić L.(2010). Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. **Ecological Modelling**. 221:1239-44. <https://doi.org/10.1016/j.ecolmodel.2009.12.023>

Rodeghiero M, Churkina G, Martinez C, Scholten T, Gianelle D, Cescatti (2013) A. Components of forest soil CO₂ efflux estimated from $\Delta^{14}\text{C}$ values of soil organic matter. **Plant Soil** ;364:55-68. <https://doi.org/10.1007/s11104-012-1309-1>

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](https://doi.org/10.1016/S0929-1393(03)00065-9)

Samuel, Arthur (1959) Machine Learning: **Field of study that gives computers the ability to learn without being explicitly programmed.**

Silva PA da, De Lima BH, La Scala N, Peruzzi NJ, Chavarette FR, Panosso AR (2020). Spatial variation of soil carbon stability in sugarcane crops, central-south of Brazil. **Soil and Tillage Research** ;202:104667.
<https://doi.org/10.1016/j.still.2020.104667>

Smith P, Fang C, Dawson JJC, Moncrieff JB (2008) Impact of Global Warming on Soil Organic Carbon. **Advances in Agronomy**;97:1-43.
[https://doi.org/10.1016/S0065-2113\(07\)00001-6](https://doi.org/10.1016/S0065-2113(07)00001-6)

Soong JL, Janssens IA, Grau O, Margalef O, Stahl C, Van Langenhove L, Urbina I, Chave J, Dourdain A, Ferry B, Freycon V, Herault B, Sardans J, Peñuelas J, Verbruggen E.(2020). Soil properties explain tree growth and mortality, but not biomass, across phosphorus-depleted tropical forests. **Scientific Reports** ;10:1-13.
<https://doi.org/10.1038/s41598-020-58913-8>

Souza GSA de, Soares VP, Leite HG, Gleriani JM, do Amaral CH, Ferraz AS, Silveira MV de F, Santos JFC dos, Velloso SGS, Domingues GF, Silva S (2019) Multi-sensor prediction of Eucalyptus stand volume: A support vector approach. ISPRS J Photogramm. **Remote Sensing** ;156:135-46.
<https://doi.org/10.1016/j.isprsjprs.2019.08.002>

Tang J, Baldocchi DD, Xu L (2005). Tree photosynthesis modulates soil respiration on a diurnal time scale. **Global Change Biology** ;11:1298-304.
<https://doi.org/10.1111/j.1365-2486.2005.00978.x>

VanDeventer W, Jamei E, Thirunavukkarasu GS, Seyedmahmoudian M, Soon TK, Horan B, Mekhilef S, Stojcevski A (2019).Short-term PV power forecasting using

hybrid GASVM technique. **Renewable Energy** ;140:367-79.
<https://doi.org/10.1016/j.renene.2019.02.087>

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 Soil**;444:193-211. <https://doi.org/10.1007/s11104-019-04262-z>

Wang W, Peng S, Fang J(2010). Root respiration and its relation to nutrient contents in soil and root and EVI among 8 ecosystems, northern China. **Plant Soil.** ;333:391-401. <https://doi.org/10.1007/s11104-010-0354-x>

Willmott CJ (1981). On the validation of models. **Physical Geography**, 2:184-94.
<https://doi.org/10.1080/02723646.1981.10642213>

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>

Xavier CV, Moitinho MR, De Bortoli Teixeira D, André de Araújo Santos G, de Andrade Barbosa M, Bastos Pereira Milori DM, Rigobelo E, Corá JE, La Scala JúniorN.(2019). Crop rotation and succession in a no-tillage system: Implications for CO2 emission and soil attributes. **Journal of Environmental Management** ;245:8-15. <https://doi.org/10.1016/j.jenvman.2019.05.053>

Xiao Y, Liu S, Zhang M, Chen B, Xu Z, Pan Y, Shi X, Wu Z, Luo T (2020). Biotic and abiotic properties most closely associated with subtropical forest soil respiration differ in wet and dry seasons: A 10-year in situ study. **Agricultural and Forest Meteorology.** ;292-293:108134. <https://doi.org/10.1016/j.agrformet.2020.108134>

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.** ;178:40-8.
<https://doi.org/10.1016/j.catena.2019.03.001>

Zhou D, Zhao SQ, Liu S, Oeding J (2013) A meta-analysis on the impacts of partial cutting on forest structure and carbon storage. **Biogeosciences** ;10:3691-703.
<https://doi.org/10.5194/bg-10-3691-2013>

Zinn YL, Resck DVS, Da Silva JE (2002). Soil organic carbon as affected by afforestation with Eucalyptus and Pinus in the Cerrado region of Brazil. **Forest Ecology and Management** ;166:285-94. [https://doi.org/10.1016/S0378-1127\(01\)00682-X](https://doi.org/10.1016/S0378-1127(01)00682-X)

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

Abstract - Soil respiration is one of the most important components of carbon (C) cycle in forest ecosystems. Estimating and understanding all the factors involved in this dynamics is fundamental to predicting the intensity of climate change. The objective of our study was to estimate the temporal variability of CO₂ emission from soil (FCO₂) and oxygen uptake (FO₂) from environmental covariates and estimate the enhanced vegetation index (EVI) and attributes of soil in reforested area with native species, in the Brazilian biome Cerrado, using regression from machine learning techniques such as Artificial Neural Networks (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS). Data set comprised observations of FCO₂ (n= 500) and FO₂ (n= 175). Two ANN architectures: multilayer perceptron (MLPNN) and radial basis function (RBFNN) were trained and compared. Sensitivity analysis was applied to each ANN. Predictive performance was evaluated based on mean absolute error (MEA), root mean square error (RMSE), mean absolute percentage error (MAPE), agreement index (d), confidence coefficient (c), and coefficient of determination (R²). The best estimation results for validation were: FCO₂ with MLPNN (R² = 0.42, RMSE = 1.069 μmol m⁻² s⁻¹) and RBF (R² = 0.54, RMSE = 1.015 μmol m⁻² s⁻¹) and FO₂ with MLPNN (R² = 0.45, RMSE = 0.093 mg m⁻²s⁻¹) and RBFNN (R² = 0.74, 0.079 mg m⁻² s⁻¹). The largest percentage errors were found in FO₂ estimation (MAPE = 0.90 calibration and 0.86 validation). Soil temperature and macroporosity are important predictors in the estimation of FCO₂ and FO₂. The best combination of variables for training the ANFIS was selected based on trial and error. The results were FCO₂ (R² = 16) and FO₂ (R² = 29). Nitrogen stock is an important variable for predicting FCO₂ and atmospheric pressure for FO₂. For all models, FCO₂ outperformed FO₂. A primary factor analysis was performed and FCO₂ and FO₂ correlated best with weather attributes and soil attributes, respectively.

Keywords: Soil CO₂ emissiom; oxygen influx; reforestation, tropical ecosystems; artificial intelligence.

4.1 Introduction

The tropics contain the highest biodiversity on Earth, yet they are the hardest hit by anthropogenic actions (Barlow et al., 2018). Brazil has contributed to the loss of tropical biodiversity and currently leads the world ranking of countries that have deforested the majority of primary forests located within the country. Specifically, in recent decades, the removal of native vegetation in the Cerrado biome has increased considerably, reaching higher rates than those related to deforestation in the Amazon region (Trigueiro et al., 2020).

The Cerrado occupies approximately 24% of the Brazilian territory and although it is a key component of the regional carbon (C) cycle, is one of the most threatened biomes globally, owing to the conversion of forests into agricultural areas (Strassburg et al., 2017). Notably, land use change (LUC) in this biome has also caused changes in the microbial community, which plays a fundamental role in the biogeochemical cycles of C, nitrogen (N) and phosphorus (P), as well as in maintaining soil quality (de Souza and Procópio, 2021). The degradation of vegetation cover and the replacement of native vegetation directly influences C and N stocks, primarily by reducing the input of organic material into the soil, thereby decreasing soil quality (Nourbakhsh, 2007; Yuan et al., 2019).

Soil respiration (R_s) is the sum of autotrophic (root) and heterotrophic (microbes, soil fauna) respiration (Hanson et al., 2000). Basal (R_B) and substrate-induced respiration (S_{RB}) is a major bioindicator of soil quality (Sparling 1997). As carbon dioxide is produced and emitted during soil respiration (FCO_2), it is a major contributor to atmospheric CO_2 . Furthermore, in this dynamic, oxygen is consumed (FO_2) as a result of the metabolic processes of microbial activity and root.

Notably, soil respiration varies with time and space and is influenced by soil water content and temperature (Adachi et al. 2006; Kim et al., 2012). In recent years, a number of studies have correlated R_s with meteorological data (Ouyang and Zheng, 2000; La Scala et al. 2003); oxygen availability (Ferraz de Almeida et al., 2018), soil physical attributes such as class and pore size (Silva et al., 2019), the quantity, quality, and distribution of plant residues (de Araújo Santos et al., 2019), as well as with vegetation indices such as the enhanced vegetation index (EVI) (Huang et al. 2012). Specifically, (Raich and Tufekcioglu, 2000) observed that vegetation type, forest microclimate, and net primary productivity are positively correlated with

R_s . This is because vegetation and canopy age affect several soil properties such as the C/N ratio and pH (Lucas-Borja et al., 2019). Furthermore, the type of canopy cover (monoculture or mixed) has an effect on the physical, chemical, and biological characteristics of soil, as well as on the contribution of organic material. In addition, studies have shown that composite forests with different species increase microbial respiration (Kooch and Noghre, 2020; Kooch and Ghaderi, 2021) and influence soil CO_2 flux (Laganière et al., 2012). In this context, it is important to estimate and understand the dynamics of FCO_2 and FO_2 with environmental variables and soil attributes. This especially, in areas of reforestation with mixed vegetation, including areas of the Cerrado.

Recently, modeling via machine learning (ML) has increased rapidly, while the field of soil science has helped improve our understanding of the causes of soil variation temporal, the environmental variables that control soil distribution (Padarian et al., 2020), and the estimation of soil respiration (Bond-Lamberty, 2018; Freitas et al., 2018; Ebrahimi et al., 2019). Machine learning algorithms include artificial neural network (ANN) and adaptive network-based fuzzy inference system (ANFIS) models. Machine learning algorithms include artificial neural network (ANN) and adaptive network-based fuzzy inference system (ANFIS) models. Although ANFIS models have been successfully used to explain of the variability of R_s in different ecosystems (Zhao et al., 2017), few studies have used ANFIS in the prediction of R_s in planted forest areas (Canteral, 2020).

Concerning oxygen modeling, although the majority of estimates have focused on predicting dissolved oxygen concentration in water with ANN, the ANFIS system has shown good results (Heddam et al., 2014; Zounemat-Kermani et al., 2014; Lamzouri et al., 2017). However, some estimates have recorded large mean absolute percentage errors (MAPE) associated with these models (Liu et al., 2019). Notably, studies on soil O_2 influx in soil using machine learning are scarce.

Given the above, we hypothesize that of this study is that environmental variables such as EVI and soil attributes can be used as predictor variables to model the temporal variability of CO_2 emissions and O_2 influx in reforested areas using machine-learning techniques. The objective of the study is to evaluate and compare two ANN architectures, namely multilayer perceptrons (MLPNN) and radial basis function (RBFNN) networks and an ANFIS model, in predicting CO_2 emissions and O_2 influx in Cerrado areas reforested with native species.

4.2 Materials and methods

In the present study, the database was created based on an evaluation of the temporal variability of soil respiration (FCO_2 and FO_2) in an area inhabited by native species. The mean FCO_2 value was determined as $4.059 \text{ } (\mu\text{mol m}^{-2} \text{ s}^{-1})$, mean FO_2 was $0.371 \text{ } (\text{mg m}^{-2} \text{ s}^{-1})$, soil temperature was $25.99 \text{ } ^\circ\text{C}$, and soil water content (SWC) was 11.56% (Vicentini et al., 2019).

4. 2.2 Study area

The study area was a riparian forest in the Cerrado, composed of the following 21 native species: *Albizia lebeck*, *Chorisia speciosa*, *Cydonia oblonga*, *Enterolobium contortisiliquum*, *Eugenia jambolana*, *Ficus clusiifolia*, *Holocalyx glaziovii*, *Hovenia dulcis*, *Jacaranda semisenata*, *Koelreuteria paniculata*, *Moquilea tomentosa*, *Morus nigra*, *Myroxylon balsamum*, *Peltophorum dubium*, *Psidium guajava*, *Pterocarpus quercinus*, *Peltophorum vogelianum*, *Spondias venulosa*, *Tabebuia chrysotricha*, *Tabebuia impetiginosa*, and *Tabebuia odontodiscus*. The species were randomly distributed in a spacing of $3 \times 2 \text{ m}$. The forest is located in the Fazenda Experimental da Faculdade de Engenharia (UNESP), in the city of Selvíria, Mato Grosso do Sul, Central Brazil, on the banks of the Paraná River (Figure 1), with central coordinates of $20^\circ 20' 53.41'' \text{ S}$ and $51^\circ 23' 55.50'' \text{ W}$, and an altitude of 354 m above sea level.

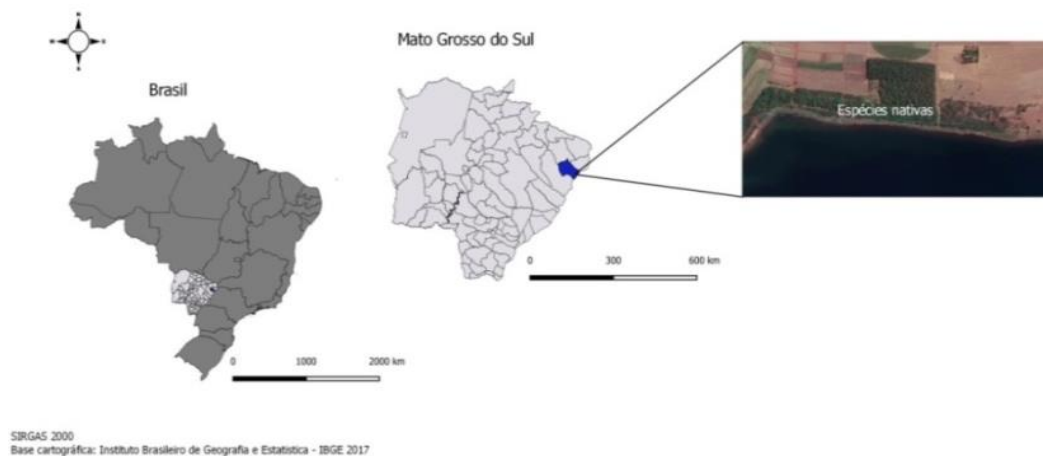


Figure 1. Location of the experimental area. Selvíria, MS - Mato Grosso do Sul, km - kilometers, N - north, S - south, E - east, W - west. SIRGAS - Geocentric Reference System for the Americas and IBGE - Brazilian Institute of Geography and Statistics. The satellite images were obtained by Google Earth.

The climate of the region was classified as C1dAa' by the Thornthwaite system (De Souza Rolim et al., 2007), with a mean annual temperature of 23.5 °C, mean annual rainfall of 1,300 mm, and relative humidity between 70% and 80%. The local soil was classified as dystrophic Red Latosol (LVd) according to the Brazilian Soil Classification System (Embrapa Solos, 2018). The relief was moderately flat and undulating with 542.38 g kg⁻¹ sand, 65.21 g kg⁻¹ silt, and 392.39 g kg⁻¹ clay in the 0–0.10 m layer. For approximately ten years, the area was managed with rice cultivation.

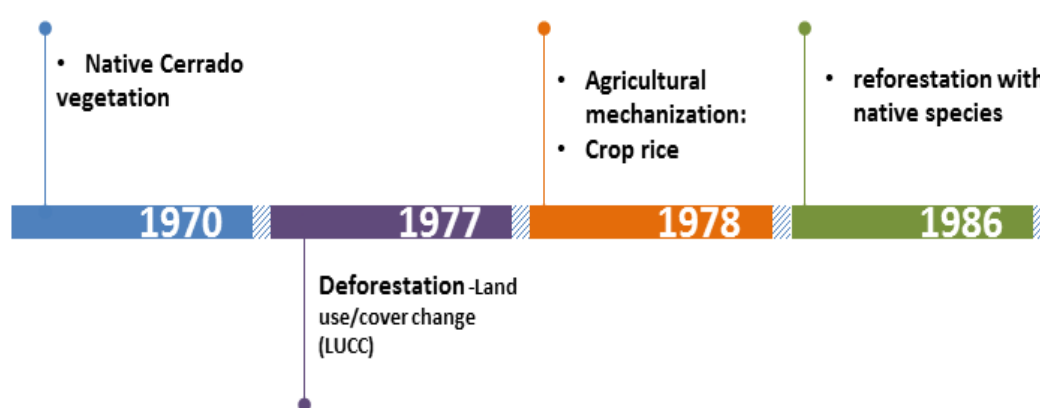


Figure 2: History of land use and change in the experimental area

4.2.3 Determination of soil CO₂ emission, O₂ influx, soil temperature, and water content

Soil CO₂ emissions were recorded between 07:00 - 12:00 using a LI-8100 portable system (LI-COR, NE, USA). Twenty-five sampling points were demarcated with polyvinyl chloride (PVC) collars (0.10 m diameter and 0.85 m height) and installed in the soil at a depth of 5 cm in each plot. The reading time was 120 s. Simultaneously with the FCO₂ readings, soil temperature (T_s) and soil water content (SWC) were determined. Soil temperature was recorded for all studied points in the 0–0.12 m layer using a digital thermometer, while water content was evaluated using a portable TDR-Campbell® system (Hydrosense™, Campbell Scientific, Australia), consisting of a probe with two 20 cm rods, which were inserted into the soil perpendicular to the surface. Further information was detailed by (Panosso et al. 2012).

The O₂ influx was determined with the aid of a portable UV Flux (0%–25%) sensor (CO₂Meter, Inc., Ormond Beach, FL, USA) at ten selected points in the area. To determine the O₂ uptake in the soil, 10 points in each area were selected using the portable UV flux (0%–25%) sensor (CO₂Meter, Inc). The O₂ influx into the soil was calculated by considering the atmospheric pressure, air temperature, and chamber volume. Further information was detailed by (Ferraz de Almeida et al., 2018; Vicentini et al., 2019).

4. 2.4 Soil data and Environmental variables

Soil physical analyses were performed in the 0–0.10 m layer after the FCO₂ and FO₂ readings. The following results were observed: macroporosity = 0.049 (m³ m⁻³), microporosity = 0.394 (m³ m⁻³), and soil density = 1.34 (g dm⁻³) (according to the methodology proposed by Embrapa (1997). For the chemical analyses the following results were observed: phosphorus content (resin) = 6.16 (mg dm⁻³), aluminum = 5.88 (mmolc dm⁻³), soil organic matter = 25.2 (g dm⁻³), pH (01 mol L⁻¹ CaCl₂) = 4.44, cation exchange capacity (CEC) = 82.98 and base sum = 33.06 (mmolc dm⁻³), according to the methodology proposed by (Raij et al., 2001). Litter samples were collected in May of 2016. The organic carbon and nitrogen contents were determined using the methods described by Tedesco et al. (1995) and Bataglia et al. (1983), respectively.

The meteorological database was obtained from the historical records of UNESP-CANAL CLIMA for the interval from November 2015 to October 2017. (<http://clima.feis.unesp.br/index.php>) including PP (mm) = rainfall, w (m/s) = wind speed, ET_o PN-M (mm/day) = reference evapotranspiration by Penman-Monteith, P_{atm} (kPa) = atmospheric pressure, T_{air} = air temperature, PAR (μmol m⁻²) = photosynthetically active radiation, and RSG (MJm⁻² day⁻¹) = global solar radiation. Biophysical data EVI = spectral index; MODIS-EVI data were obtained using SATVeg – Temporary Vegetation Analysis System, developed by Embrapa Informática Agropecuária. The MODIS images used for the development of SATVeg were obtained from the LP DAAC Data Pool (<http://lpdaac.usgs.gov>). These images were composed of the maximum values of the vegetation indices every 16 days. The spatial resolution of the MODIS sensor for the product used was 250 m. To minimize

the effects of time series noise, the data were obtained with pre-filtering and filtering available in the tool. A Savitzky-Golay filter was selected.

4.2.5 Data mining

The dependent variables for modeling constituted $n = 500$ and $n = 175$ observations for FCO_2 and FO_2 , respectively. agrometeorological (climatic) variables, EVI vegetation index, and soil physicochemical attributes were selected as quantitative environmental variables for predicting soil FCO_2 and FO_2 . Multivariate outliers were identified based on the Mahalanobis distance.

4.2.5.1 Artificial neural network (ANN)

An MLP network comprises an input layer, which has no adjustable weights, one or more hidden layers, and output layer. Each layer has multiple processing units, and each unit is fully interconnected with weighted connections to the units in the subsequent layer. The input signal propagates the network layer-by-layer (Haykin, 2001). Its training is supervised by an algorithm based on an error backpropagation learning rule (Rumelhart et al. 1986).

The model for each node in the network includes an activation function. Among those used by the neurons of MLP networks, logarithmic and tangent sigmoid functions are used in neurons in the hidden and output layers (Yilmaz et al. 2009). In this study, a logistic sigmoid activation function was used (Eq. 1), where $f(v)$ is the sigmoid activation function, a is the estimation of the parameter that determines the sigmoid slope, and v is the activation potential of the function.

$$f(v) = \frac{1}{1+e^{-av}}, \quad (\text{Eq.1}).$$

An RBF neural network is based on supervised learning. Its construction, in its most basic form, involves three layers. The first is an input layer with source nodes that connect the network to its environment; however, it does not perform any function on the input data. The second is a single hidden layer, in which a nonlinear

transformation from the input space to the hidden space occurs. The neurons in the hidden layer perform an activation function by means of a nonlinear kernel (e.g., Gaussian transfer function), whose outputs are inversely proportional to the distance from the center of the neurons (Dębska et al. 2011). Third is the output layer, which is the weighted sum of the output samples of the hidden layer (Haykin, 2001; Hamed and Dehdashti Jahromi, 2021).

An MLP neural model is generated using the back propagation (BP) method and is optimized using the conjugate gradient algorithm. For the RBF, the Gaussian density function is used in the hidden layer as an activation function. For each estimate (i.e., FCO_2 and FO_2) of different architectures, we observe that 100 epochs are sufficient for error convergence.

Step 1: We trained 100 RBF and 100 MLPs. From the original data, the training set consisted of 24 variables: T_s ; SWC; P; pH; H+Al; SB; CEC; Macro; Micro; Bd; Cstock, StockN; CN; Sand; Silt; Clay; T_{atm} ; P_{atm} ; GRS; PAR; ET_0 ; Rain; W; and EVI. Correspondingly, we used ANN1 and ANN2 for FCO_2 and ANN5 and ANN6 for FO_2 (Table 1). The architectures were developed using the Intelligent Problem Solver (IPS) tool of Statistica v.7.0. to return the 10 ANNs having the lowest root mean square error (RMSE), randomly presented. The best model from each architecture was selected for each estimate (i.e., FCO_2 and FO_2). The criterion for stopping the training algorithm was one when the RMSE was less than 1 % (Haykin, 2001).

Step 2: The IPS provides a sensitivity analysis as part of the ANN simulation results. Thus, in an attempt to improve ANN performance, we used the results of the sensitivity analyses performed by ANN1 and ANN2, ANN5, ANN6, which determine the most important predictors in modeling FCO_2 and FO_2 . The software follows the sensitivity analysis (SA) methodology proposed by (Hunter et al. 2000). According to the procedure described, the SA used in this study iteratively rejects an input variable and evaluates the network error. The measure of a variable's sensitivity is the ratio of the error without the variable to the original error, which provides the sensitivity score (Hunter et al., 2000; Anagu et al. 2009). The variables were ranked in order of importance based on the ratio. From the sensitivity results, a new subset of data was formed, and new MLPNN and RBFNN training was performed to estimate FCO_2 and FO_2 .

From the sensitivity results (ANN1; ANN2 for FCO_2 and ANN3; ANN4 for FO_2), a new subset of data was formed, and a new training of MLPNN and RBFNN was

performed for estimation of FCO_2 and FO_2 . The subset inputs of ANN3 were T_s ; CN; Clay; pH; P; ET_0 ; SB; Macro; EVI; Sand; Micro; P_{atm} ; and Silt. For ANN4, they were T_s ; Cstock CN; SWC; EVI; PAR; Macro; P; CEC; Clay; pH; Micro; Silt; Sand; SB; H+Al; and w. For FO_2 , the ANN7 MLPNN used Micro; EVI; Silt; Cstock; Macro; SWC; pH; H+Al; CN; SB; CTC; Sand; T_s , P; PAR; and W. The ANN8 used SWC; CEC; Macro; Sand; Silt; Clay; T_{air} ; P_{atm} ; W and CN

4.2.5.2 Adaptive Neuro-diffuse FIS (ANFIS)

Jang (1993) developed a hybrid multilayer feed-forward neural network and an FIS (ANFIS), which combines the advantages of both sources, making it the most powerful artificial intelligence technique (Kharb et al., 2014). The major advantage of the ANFIS model is its computational efficiency and adaptability (Ghadernejad et al. 2018). Hence, it can approximate any function. In this study, the Takagi and Sugeno fuzzy system is employed to formalize a systematic approach to generate fuzzy rules from a set of input and output data (Buragohain and Mahanta, 2008). The basic structure of an FIS is divided into fuzzification, a rule base, inferencing, and fuzzification.

A disadvantage of ANFIS is the restriction of the number of parameters defined in the model. If the number of inputs exceeds five layers, the training time takes too long to assign the appropriate values to the parameters of the membership function (Dehghani et al., 2019)(Kaab et al., 2019;. The most common membership functions are triangular, trapezoidal, and Gaussian (Klir; Yuan, 1995). In this study, the trapezoidal type is used (Riza, 2015). Although ANFIS has a high mapping ability of input to output similar to a black-box model, it suffers from a long training time that inhibits appropriate value assignments to the membership function parameters (the basis of fuzzy set theory).

Thus, several combinations of inputs were used for FCO_2 and FO_2 . For FCO_2 , these included pH; H+Al; ET_0 ; SRG and PAR. For FO_2 , these included T_s ; SWC; P_{atm} ; N stock, and SB. In this study, the ANFIS was performed using R software (version 4.0.2; R Core Team 2019, R Foundation for Statistical Computing, Vienna, Austria), using the R packages frbs: Fuzzy Rule-Based Systems for Classification and Regression Tasks. (Riza, Bergmeir, Herrera, and Benitez 2015).

4.2.5.3 Criteria for model assessment

The evaluation of the efficiencies of the developed models, including accuracy and agreement, was assessed using statistical criteria of the Pearson correlation coefficient (r) (Eq. 2), mean absolute error (MEA) (Eq. 3), mean absolute percentage error (MAPE) (Eq. 4), agreement index (d) (Eq. 5) (Williams, 1981), and confidence coefficient (c) (Eq. 6) (Camargo and Sentelhas, 1997b). Analysis of variance was performed using the F-test at a probability level of 0.01.

$$r = \frac{(N \sum_{i=1}^N (X_{obs1} Y_{obs1}) - \sum_{i=1}^N X_{obs1} \sum_{i=1}^N Y_{obs1} ((N \sum_{i=1}^N Y_{obs1})^2 - (\sum_{i=1}^N Y_{obs1})^2))}{\sqrt{((N \sum_{i=1}^N X_{obs1})^2 - (\sum_{i=1}^N X_{obs1})^2)}}, \text{ Eq. (2)}$$

$$MEA = \frac{\sum_{i=1}^N (Y_{obs1} - Y_{est1})}{N}, \text{ Eq. (3)}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left(\frac{|(A_i - P_i)|}{A_i} 100 \right), \text{ Eq. (4)}$$

$$d = 1 - \frac{\sum_{i=1}^N (Y_{obs1} - Y_{est1})^2}{\sum_{i=1}^N (|Y_{est1} - \bar{Y}| + |Y_{obs1} - \bar{Y}|)^2}, \text{ Eq. (5)}$$

$$c = r * d, \text{ Eq. (6)}$$

where n is number of data points, X_{obs1} is the observed value of X , Y_{obs1} is the observed value of Y , Y_{est1} is the estimated value of Y , and $\bar{Y}$ is the observed mean of Y .

4.2.5.4 Summary of the Methodology Used in This Study

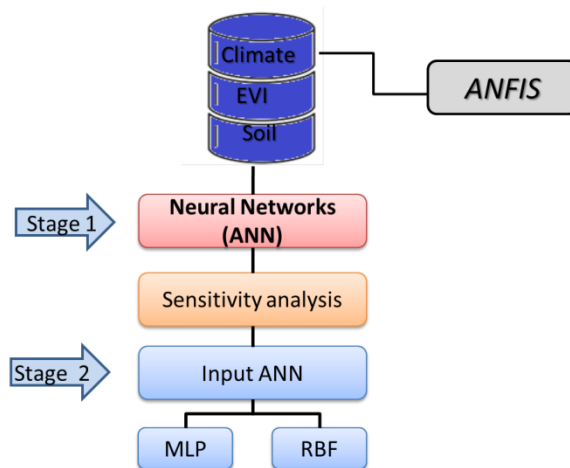


Figure 3. The flowchart of methodology used in this stud

Data collection (Section 4.2.4 and 4.2.5), We next used several combinations of inputs were used for FCO₂ and FO₂ (Section 4.2.5.2). Stage 1 and stage 2 (Section 4.2.5.1).

4.2.6 Results and Discussion

All the models generated by ANN and ANFIS were significant ($p < 0.01$) (Table 1). The architecture of the neural network is a key factor that influences the learning time and performance of the network. The two models (i.e., MLPNN and RBFNN) were compared. Note that in both stages of ANN processing (before and after SA), the best-performing MLPNN for FCO₂ was obtained with a small reduction in the number of neurons in its structure (ANN3), as compared with ANN1. This characteristic is common to MLP networks.

In contrast, RBFNN generally requires a larger number of neurons and has a tendency to use more parameters from the learning dataset (Boniecki et al., 2020). Therefore, in this study, the RBFNN in both phases presented a more complex architecture, requiring more neurons in the intermediate layer (ANN2) even after SA (ANN4) using 17 variables to train the entire dataset. The RMSE and MAE evaluation criteria indicate how close the data points are to the line of best fit. After SA, there was a decrease in RMSE and MAE, and the estimates were better correlated (Table 1), ranging from (r) 0.78 to 0.79 at calibration and 0.59 to 0.74 at validation for FCO₂ and from 76 to 86 at calibration and 0.40 to 0.67 at validation for FO₂.

According to (Hamedi and Dehdashti Jahromi, 2021), the choice of the best network parameters is related to the smallest errors of calibration and validation and the maximum of the coefficient of determination (R^2), which describes how much of the observed scatter is predicted so that an appropriate network size can be determined. Figure 4 shows the results of the best estimates. In this study, the best performance was observed for FCO₂. The R^2 value ranged from 0.46 to 0.63, and from 0.32 to 0.54, respectively calibration and validation.

During calibration, the values of FO₂ ranged from 0.15 to 0.59; during validation, they ranged from 0.095 to 0.74. Both phases had high associated perceptual errors. Regarding FCO₂, according to Table 1, the RMSE, MEA, and values were within the acceptable range; during calibration, they were below one. The ANN4 showed better performance indices in terms of the R^2 and RMSE and was

the most accurate model, with MAPE values indicating errors in 16% of the data during calibration and 19 % during validation. The concordance index, which relates the distance of the observed values from the estimated values of the standard model, showed that during calibration the values were 0.86 0.88 ANN3 and ANN4. These results corroborate the classification performance as “good” according to the classification index (Camargo and Sentelhas, 1997a) for both ANNs. Regarding the validation phase, the MLPNN presented a bad calibration ($d = 0.75$), whereas RBF ($d = 0.84$) was good.

For FO_2 , all models showed inferior performance compared with those of FCO_2 . Despite the low RMSE values for calibration (0.122 to 0.144), and validation ranged from 0.079 to 0.133. There are high percentage errors associated with the area at FO_2 , ranging from 51 to 86% in calibration and 61 to 90% in validation, indicating a lower accuracy. We believe that the size of the dataset for neural-network training ($n = 176$) affected the prediction ability of the ANNs. The FCO_2 dataset is compact ($n = 500$). In an ANN, generalization, or the ability to respond appropriately to previously unseen patterns, is influenced primarily by the size of the training set and its representativeness with respect to estimation, problem complexity, and ANN architecture (Haykin, 2001). Additionally, the observed FO_2 values in the field are very low. In the original dataset, over the 193 study days, the values ranged from 0.030 to 2.02 ($\text{mg m}^{-2} \text{s}^{-1}$, with a mean value of $0.371 \text{ mg m}^{-2} \text{s}^{-1}$), and in the soil, the temporal variability of oxygen mainly related to the frequency of precipitation events (Liptzin et al. 2011). The diffusion of O_2 in soil is conditioned by its physical properties (Neira et al., 2015). MAPE is often used in model performance evaluation because it is easy to understand (Myttenaere et al. 2016).

However, the very small values close to zero observed in the dataset present a disadvantage, because in the calculation, MAPE divides the absolute error by the actual data. Thus, values close to zero can significantly increase the MAPE. We believe that this may have caused the values observed when training the ANNs of FO_2 . Like the FCO_2 ANN, after reducing the dataset with SA, there was an improvement in the correlation of the estimates. For FO_2 , the estimation using MLPNN performed slightly better than the RBFNN.

When comparing the performance of the ANNs to ANFIS (Table 1, Figure 4, and 5), the RMSE, MEA, and MAPE values indicate an inferior performance of the ANFIS. In the ANFIS, the agreement index value in the FCO_2 estimate ranged from

0.63 to 0.66 for the calibration and validation periods, indicating a low relationship between observed and modeled values. The same was found for FO_2 , with values ranging from 0.46 to 0.68. In both estimates, the phases (i.e., calibration and validation) were classified as "bad" according to the classification index (Camargo and Sentelhas, 1997a). The ANFIS obtained RMSE and MEA errors higher than those of MLPNN and RBFNN. For the FCO_2 ANN, the RMSE values ranged from 1 to $1.18 \mu\text{mol m}^{-2} \text{s}^{-1}$ and MEA from 1.22 to $127 \mu\text{molm}^{-2}\text{s}^{-1}$. These results indicate low accuracy with coefficient of determination values equal to 0.19 and 0.16 for calibration and validation phases, respectively. For FO_2 , the performance was also poor in both validation and calibration. Despite the low RMSEs (0.11 to $1.12 \text{ mg m}^{-2} \text{s}^{-1}$) and MEAs (0.13 to $0.19 \text{ mg m}^{-2}\text{s}^{-1}$), validation and calibration respectively, those observed for the FO_2 ANN had high MAPE values, reflecting a low accuracy of 55 % for calibration and 88% for validation. We expected a better performance for ANFIS compared with the two ANNs, owing to the higher degree of robustness and fault tolerance provided by the model. Owing to the complexity of FCO_2 and FO_2 modeling, the limitation regarding the number of system inputs compromised the estimation performance.

Table 1. Classification of performance of statistical performance indicator (Camargo; Sentelhas, 1997).

Soil CO ₂ emission (μmol m ⁻² s ⁻¹)												
			r	R ² adj	RMSE	MEA	MAPE	d	Performance (c)		F	p valor
ANN1	MLPNN	Calibration	0.76	0.579	0.906	0.935	0.187	0.846	0.643	Median		< 0.0001
		13-8-1:1 Validation	0.64	0.415	1.069	1.067	0.217	0.787	0.507	Bad	84.31	< 0.0001
ANN2	RBF	Calibration	0.68	0.462	0.867	0.938	0.183	0.796	0.541	Sufferable		< 0.0001
		17-20-1:1 Validation	0.57	0.321	0.996	1.179	0.244	0.717	0.406	Bad	52.5	< 0.0001
Sensitivity analysis												
ANN3	MLPNN	Calibration	0.781	0.590	0.886	0.845	0.168	0.861	0.663	Good	349.9	< 0.0001
		12-7-1:1 Validation	0.592	0.350	1.148	1.115	0.239	0.758	0.449	Bad	64.63	< 0.0001
ANN4	RBF	Calibration	0.796	0.634	0.875	0.818	0.167	0.88	0.701	Good	418.8	< 0.0001
		17-46-1:1 Validation	0.742	0.540	1.015	0.966	0.198	0.848	0.629	Good	138.2	< 0.0001
O ₂ influx (mg m ⁻² s ⁻¹)												
ANN5	MLPNN	Calibration	0.387	0.150	0.156	0.144	0.717	0.612	0.237	Insufficient	14.12	< 0.0001
		14-10-1:1 Validation	0.323	0.104	0.196	0.271	1.079	0.491	0.158	Insufficient	4.647	< 0.01
ANN6	RBF	Calibration	0.452	0.204	0.093	0.109	0.607	0.646	0.292	Insufficient	21.3	< 0.0001
		13-16-1:1 Validation	0.309	0.095	0.114	0.281	0.69	0.346	0.107	Insufficient		< 0.01
Sensitivity analysis												
ANN7	MLPNN	Calibration	0.769	0.591	0.142	0.142	0.516	0.855	0.658	Median	123.1	< 0.0001
		13-11-1:1 Validation	0.677	0.458	0.133	0.139	0.616	0.793	0.537	Sufferable	69.43	< 0.0001
ANN8	RBF	Calibration	0.407	0.166	0.122	0.229	0.905	0.512	0.209	Insufficient	15.71	< 0.0001
		10-3-1:1 Validation	0.861	0.741	0.079	0.147	0.861	0.517	0.178	Insufficient	5.231	< 0.01
Soil CO ₂ emission (μmol m ⁻² s ⁻¹)												
ANN9	ANFIS	Calibration	0.43632	0.19	1.18734	1.2744	0.25857	0.66615	0.29066	Insufficient	776.80	< 0.01
		Validation	0.41002	0.16	1.02002	1.22425	0.25714	0.63716	0.26125	Insufficient	385.80	< 0.01
O ₂ influx (mg m ⁻² s ⁻¹)												
ANN10	ANFIS	Calibration	0.38544	0.14	0.12553	0.19891	0.55487	0.462	0.17807	Insufficient	6.414	< 0.01
		Validation	0.54245	0.29	0.11001	0.13441	0.81404	0.68359	0.37081	Insufficient	7.581	< 0.01

r: pearson correlation coefficient ; R² adj, RMSE-Observation Standard Deviation Ratio; MEA; mean absolute error; MAPE; mean absolute percentage error; d, agreement index

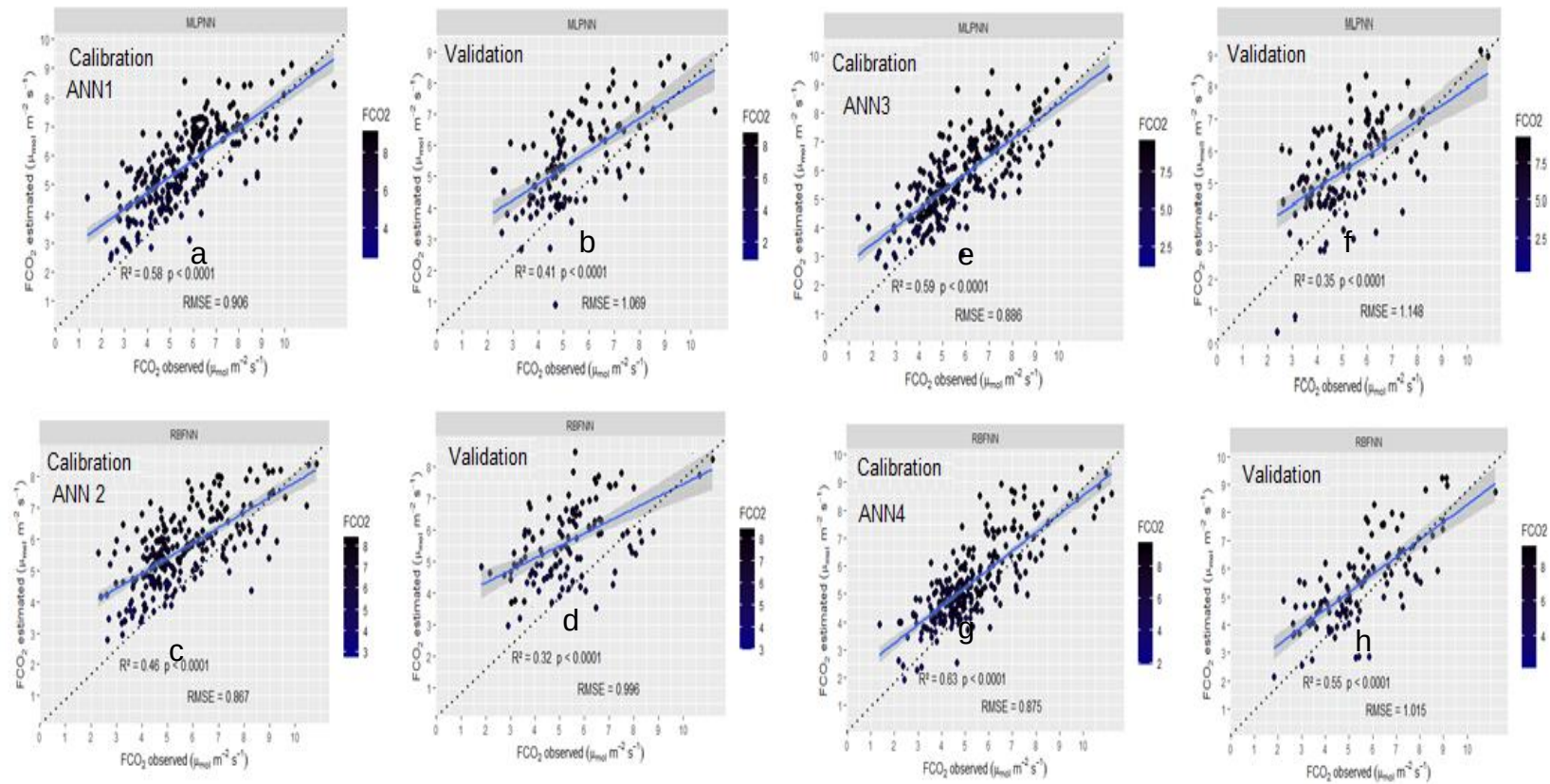


Figure 4. Performance of soil CO₂ emission ($\mu\text{mol m}^{-2} \text{s}^{-1}$) during the calibration and validation phases of artificial neural networks. a) ANN1-MLP calibration, b) ANN1- MLP validation, c) ANN2 – RBF calibration, d) ANN2 – RBF validation, e) ANN3 – MLP calibration, f) ANN3 – MLP validation, g) ANN4 – RBF calibration, h) ANN4 – RBF validation.

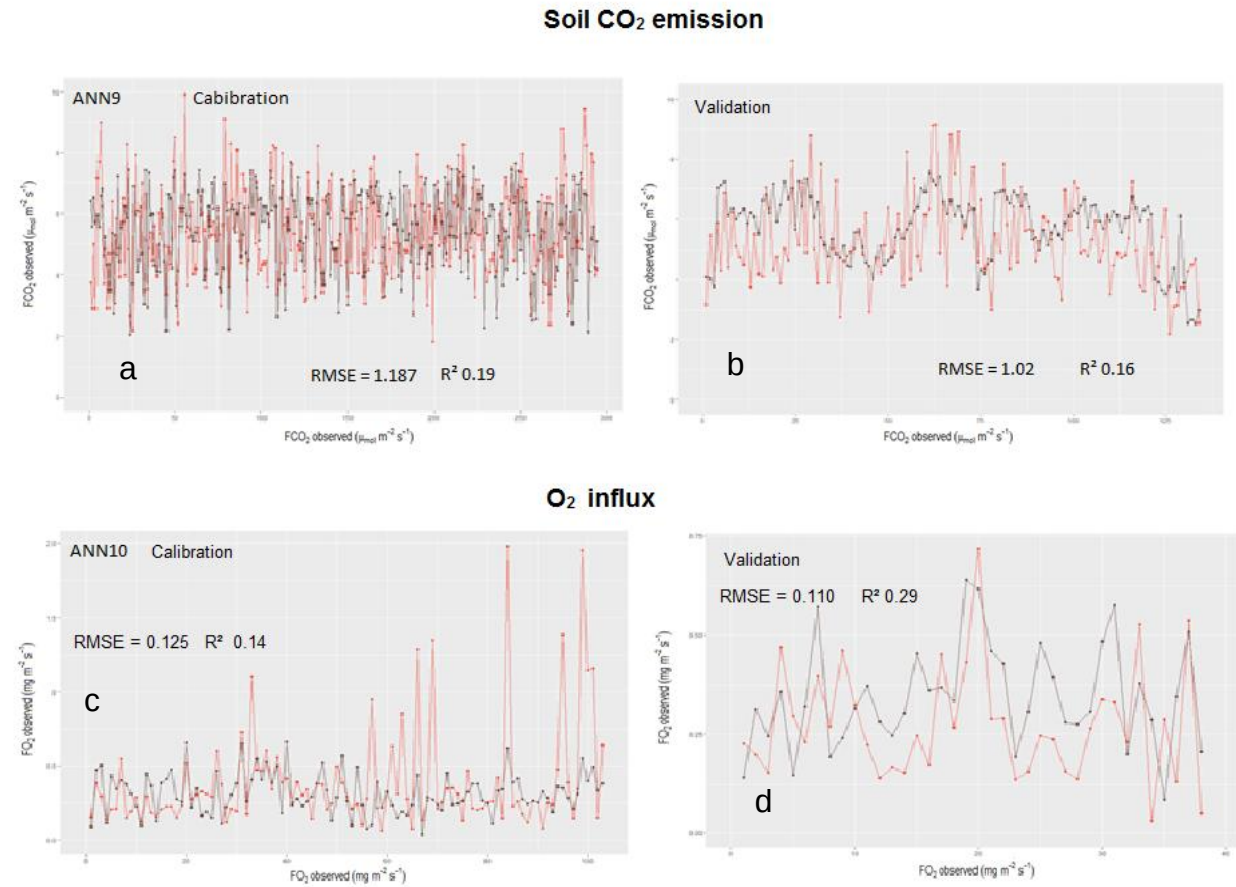


Figure 5. Model soil CO₂ emission ($\mu\text{mol m}^{-2} \text{s}^{-1}$) and O₂ influx ($\text{mg O}_2 \text{m}^{-2} \text{s}^{-1}$) during the calibration and validation phases of the adaptive neuro-diffuse inference system. Black line: observed data; red line: estimated. a) ANN9- FCO₂ calibration, b) ANN9- FCO₂ validation. c) ANN10- FO₂ calibration, b) ANN10- FO₂ validation.

Overall, it is interesting that in most of the models studied, soil temperature and macroporosity formed part of the dataset in the calibration of the models, except for ANFIS, which only considered soil temperature in the estimation of FO_2 and macroporosity, and did not contribute to any of the estimates. Macroporosity is related to the diffusion of CO_2 and O_2 gases into the soil, and the soil temperature is a determining factor in the decay of organic matter (Yuste et al., 2007), with high temperatures accelerating microbial and root activity (Silva-Olaya et al., 2013), except in tropical soils (Adachi et al., 2006).

Although several studies have reported temperature and moisture as the controlling factors on the temporal dynamics of soil respiration (Davidson et al., 1998; Lal, 2009), in this study, soil water content did not contribute to the dataset during model calibration. This observation may be related to the covariation that exists between soil temperature, evapotranspiration, and precipitation, inhibiting the determination of the individual effects of each variable (Wu et al., 2014)(Davidson et al., 1995).

It is worth noting that although ANFIS did not incorporate soil temperature in its calibration dataset to FCO_2 ; solar radiation played an important role in this estimation. Solar radiation and air temperature are among the main factors influencing the variation in soil temperature (Reichardt and Timm, 2012), in addition to soil water content and texture, and the surface area covered by leaf litter and plant canopies (Paul et al., 2004). Moreover, ANFIS was the only model in which evapotranspiration played an important role in the calibration dataset for the estimation of FCO_2 .

It is worth noting that, in addition to the physiological process of plant transpiration, evapotranspiration is the result of physical processes such as energy exchange between the earth's surface and the atmosphere, and is related to solar radiation, soil water content, precipitation, and wind (Liang et al., 2010). We also observed that only ANFIS was sensitive to the N stock in the model calibration dataset for the estimation of FO_2 . In the soil, heterotrophic, anaerobic, and aerobic microorganisms that consume oxygen are responsible for the mineralization of organic N, which constitutes the process by which N becomes available to plants. Thus, although the C stock did not contribute to the calibration of the dataset, it is known that soil organic matter is closely related to the amount of soil N, and temporal changes in the soil C and N stocks of reforested areas tend to show similar temporal

patterns (Li et al., 2021). In addition, during the calibration of the ANFIS model for FO_2 estimation, atmospheric pressure was observed to be a highly sensitive model parameter. The entry of oxygen into the soil occurs by diffusion and, as described by Fick's law, is pressure dependent (Neira et al., 2015).

Studies conducted on various ecosystems and soils demonstrated good correlations for FCO_2 estimation using physicochemical properties; for example, for your study (Freitas et al., 2018) observed good MLP results with a backpropagation algorithm for a sugarcane area situated in Brazil. The R^2 of the model was 0.91, with a MAPE error of 18%. For ANFIS, (Canteral, 2020) determined an R^2 value ranging from 0.53–0.71 and a MAPE error of 28.84–50.46% for tropical forest soils. However, soil is a complex and spatially heterogeneous mixture of minerals, carbon stock, roots, and microorganisms, and climatic factors such as precipitation and radiation contribute to the formation of different microclimates that drive temporal variability in soil respiration (Rubio and Detto, 2017). Therefore, comparing our results with other estimates does not necessarily provide insight into the dynamics of respiration in our ecosystem. Regarding oxygen consumption, thus far, we have not observed estimates with machine learning for tropical soils.

4.2.6.1 Principal Component Analysis

We performed a PCA to aid in the temporal dynamic compression of FCO_2 and FO_2 and identify the main factors involved in this process. The first three dimensions of the analysis expressed 53.4% of the total dataset inertia, indicating that 53.4% of the individual (or variable) cloud total variability was explained by the plane. Generally, the best results are obtained when the original variables are highly or inversely correlated (Manly, 2008). Moreover, the first component was the most important of the 10 components (Figure 6).

However, since FCO_2 and FO_2 contributed more to the first two components, we have only discussed the linear combinations of the variables in these two components (Table 2 and Figure 7). According to Table 2, the variables that contributed the most to PC1 in order of relevance are represented by SB (0.94), pH (0.93), N stock (0.87), macro (0.59), C stock (0.55), clay (0.47), P (0.40), CEC (0.34), FO_2 (0.26), inversely correlated H+Al (-0.87), Bd (-0.74), silt (-0.69), micro (-0.61), sand (-0.24), and CN (-0.23). This selection of variables indicated that this process

might be related to heterotrophic soil respiration, derived from the decomposition of soil organic matter, influenced by the substrate quality, and controlled by the soil physical characteristics.

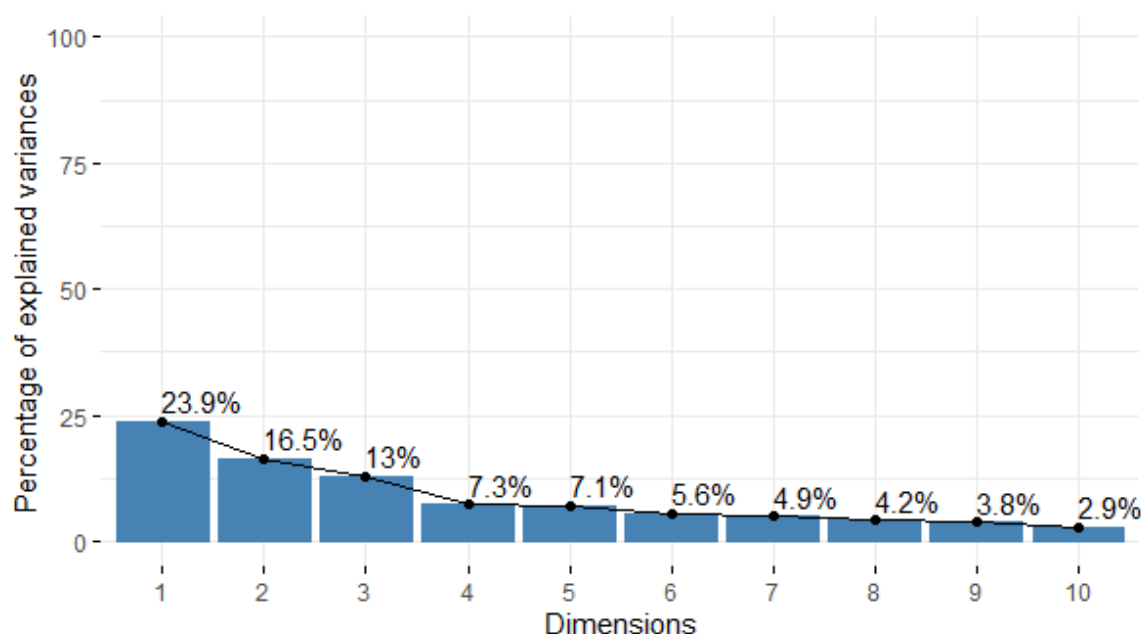


Figure 6. Decomposition of total inertia.

For PC2, the following variables were retained in order of relevance: E_t (0.89), T_{air} (0.87), GSR (0.83), PAR (0.75), T_s (0.70), EVI (0.67), FCO_2 (0.51), and inversely correlated P_{atm} (-0.53), W (-0.20), and FO_2 (0.14). Despite the low explanation value of the component, we believe that this process may be related to root activity (Tabl2).

In forest ecosystems, leaf litter is the primary source of organic matter. The organic compounds in these plant residues are composed of carbohydrates with different degrees of complexity, as well as lignin and polyphenol molecules, which are resistant to decomposition. The quality of these molecules is expressed by the CN ratio, and both the quality and quantity of the residues influence the dynamics of dissolved organic carbon and nitrogen in the soil (Zhou et al., 2015). When biochemical processes such as the decomposition of litter occur in well-aerated soil, microorganisms consume oxygen and CO_2 is released. Moreover, nutrients such as phosphorus, nitrogen, and sulfur are immobilized or released, which is fundamental for the recycling of soil nutrients (Moreira and Siqueira, 2006). This was revealed by the direct relationship between SB, CEC, P, and pH with FCO_2 (Figure7). Among the abiotic factors affecting soil microorganisms, pH is an important factor in the

formation, size, and distribution of soil microbial diversity (Liu et al., 2014; Cao et al., 2017), and also influences nutrient availability. In our study, the measured pH value was 4.4, which is within the expected range for this soil (Moreira and Siqueira 2006).

Table 2. Correlation between FCO₂ and FO₂ with environmental variables

Principal component	PC1	PC2	PC3
Explained variance (%)	23.9	16.5	13
Correlation			
FCO ₂	0.10	0.51	-0.14
FO ₂	0.26	0.14	-0.02
Ts	-0.05	0.70	-0.01
SWC	0.02	0.09	-0.11
P	0.40	-0.04	0.58
pH	0.93	0.01	0.01
H+.Al	-0.87	-0.04	0.30
SB	0.94	0.02	0.08
CEC	0.34	-0.06	0.62
Macro	0.59	0.03	-0.37
Micro	-0.61	0.02	-0.02
Bd	-0.74	-0.08	-0.16
CStock	0.55	-0.03	0.82
NStock	0.86	0.01	-0.10
CN	-0.24	-0.05	0.84
Sand	-0.23	-0.01	0.71
Silt	-0.69	-0.07	-0.16
Clay	0.47	0.04	-0.55
T _{air}	-0.05	0.87	0.06
Pa _{tm}	0.04	-0.53	-0.06
GSR	-0.05	0.83	0.06
PAR	-0.04	0.75	0.06
ET _o	-0.05	0.89	0.06
Rain	0.01	-0.04	-0.02
W	0.01	-0.20	0.10
EVI	-0.01	0.67	0.01

FCO₂ = soil CO₂ emission; FO₂ = O₂ influx; T_s(°C) = soil temperature; SWC = soil water content; ; P (mg dm⁻³) = soil P content; pH in CaCl₂ = hydrogen potential; H+Al = potential acidity; sum of bases (K⁺, Ca²⁺, and Mg²⁺); CEC (mmolc·dm⁻³) = cation exchange capacity; macro(m³·m⁻³) = macroporosity; ;Micro (m³·m⁻³) = microporosity; Bd = bulk density; (g dm⁻³); CStock =(mg ha⁻¹); carbon stocks(mg ha⁻¹); NStock (mg ha⁻¹) = nitrogen stocks; CNr atio; Sand = sand content (g kg⁻¹); Silt = silt content (g kg⁻¹); Clay = clay content (g kg⁻¹); T_{atm} = temperature atmosphere; Patm (kPa) = atmospheric pressure; GSR (MJ·m⁻²·day⁻¹) = global solar radiation; PAR (μmol·m⁻²); photosynthetically active radiation; (ET_o),EVI = enhanced vegetation index; Rain (mm); w = wind speed- w (m s⁻¹); ; EVI = enhanced vegetation index.

In addition, during decomposition, rearrangements and formation of material that are unaffected by microbial action may occur because of the dynamics of soil

aggregates. This process alter the storage of C and N in the soil and in the humic fractions of organic matter (Golchin et al., 1994); Reis Silva et al., 2016), reflecting the temporal dynamics of FO_2 and FCO_2 .

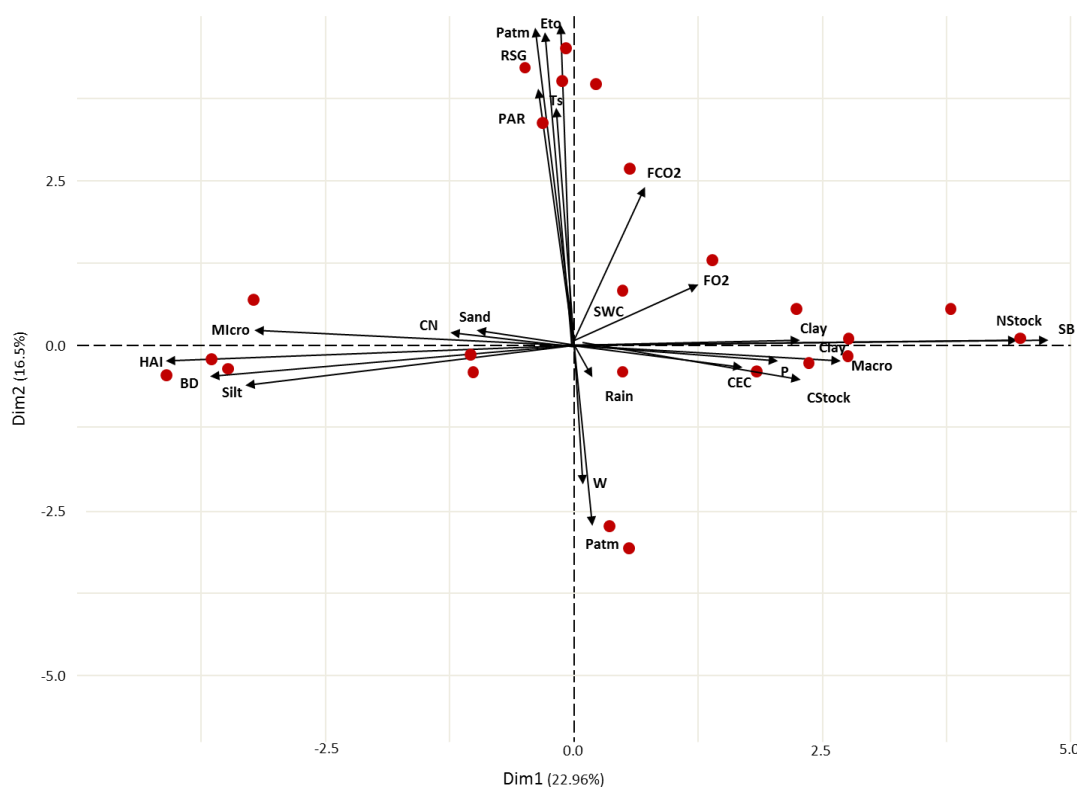


Figure 7 . Biplot graph of environmental variables and soil characteristics.

In the ecosystem analyzed in this study, the temporal dynamics of FO_2 and FCO_2 are more complex when compared with monoculture reforestation (Li et al., 2012), mainly due to phenological differences. The mixed forest examined in this study, constituting various colonizing plants and 21 species of trees with deciduous or subcaducifolic species consisting of pioneer, secondary, and late species, is characterized by substrates with different degrees of lability and seasonal variations in plant residues (Davidson et al., 2006). The CN ratio of the study area revealed an indirect association with C and N stocks, as well as FCO_2 and FO_2 , indicating that the vegetation type significantly influences the temporal dynamics of soil respiration.

Physical factors influence both the quality of the residue and the transport of gases in the soil (Brady and Weil, 2013). As mentioned previously, based on our observations we believe that soil temperature was not a limiting factor in this study. Moreover, studies on forest ecosystems suggest that autotrophic respiration is more

sensitive to temperature than heterotrophic respiration (Mäkiranta et al., 2008; Chen et al., 2017). Therefore, we did not consider the PC2 results anomalous since the existing linear correlations in this PC indicate that FCO₂ is most likely a by-product of root respiration. Moreover, considerable evidence exists to support the correlation between the physiological processes of roots and photosynthetic activity (Tang et al., 2005; Kuzyakov, 2010; Wang et al., 2010).

In addition, these studies also suggest significant correlations between the hyperspectral vegetation indices, air temperature, and soil respiration, for which 82% of the temporal variability of respiration can be attributed in broadleaved evergreen and mature broadleaved deciduous forests (Wu et al., 2014). The low correlation between the water content and both components can be explained by either covariation with ETo or soil temperature (Davidson et al., 1998; Schwendenmann et al., 2003). We believe that the high negative contribution of microporosity to PC1 is related to soil compaction. Note that both microporosity and soil density have direct and opposite associations with macroporosity. The physical characterization of the area described in the methodology indicates a density value of 1.34 g dm⁻³, which is higher than the value expected for clayey soils (Arshad et al. 1996). It is known that soil density influences the total porosity and distribution of pore sizes, influencing water retention in the soil and hindering the process of gas transport in the soil by consuming oxygen and CO₂ output from the soil into the atmosphere (Tormena et al., 1998; Teixeira et al. 2012; Bicalho et al., 2014).

The biplot figure indicate substantiating information. Note that the vectors and dispersion points of SWC, FO₂, FCO₂, clay C, and N stock participate in Group I, located on the right side of the biplot. In tropical soils, latosols tend to be rich in clays, which are directly linked to water availability; thus, water availability controls the oxygen dynamics of these soils and may lead to periods when the O₂ demand exceeds supply, influencing the dynamics of microbial community activity (Liptzin et al., 2011). A possible group (Group II), formed to the left of the bilpot, would be predominantly related to the dynamics of soil gas transport, except for the CN ratio.

4.2.7 Conclusion

This study aimed to estimate the temporal variability of FCO₂ and FO₂ of soil attributes and environmental covariates. With a dataset consisting of 24 variables, we

trained and compared two neural network architectures (MLP and RFB) for each estimate (FCO_2 and FO_2). The best estimates were observed in FCO_2 from neural networks. A sensitivity analysis was used to select the most predictive variables. After sensitivity analysis, the predictive ability of FCO_2 for MLPNN varied from $R^2 = 0.35$ to $R^2 = 0.42$ validation, while that for RBFNN varied from $R^2 = 0.32$ to $R^2 = 0.54$ validation. The best explanation of RBFNN occurred when the number of neurons in the intermediate layer was increased. However, large errors were associated with the estimation. FO_2 for MLPNN varied from $R^2 = 0.10$ to $R^2 = 0.45$ validation, while that for RBFNN varied from $R^2 = 0.09$ to $R^2 = 0.74$ validation. The size of the FO_2 database was not sufficient for the network to generalize the models.

For AFINS, we obtained an estimation of $R^2 = 0.16$ and $R^2 = 0.29$ for FCO_2 and FO_2 , respectively. However, for FO_2 , there were large errors associated with the models (MAPE = 21 to 85%). The input variable set for each model had a significant variation when the variables were combined. Soil temperature and macroporosity are good predictors of FCO_2 . The ANFIS model was very sensitive to N stock and atmospheric pressure in calibrating and estimating FCO_2 and FO_2 , respectively. Principal component analysis showed associations between heterotrophic and autotrophic respiration. FCO_2 was best correlated with variables that are linked to root respiration, while FO_2 correlated with variables that are linked to heterotrophic respiration. The quality of litters, which was expressed with CN ratio, is one of the factors that control this dynamics. Given the complexity of the data obtained in this study, neural networks outperformed ANFIs in both estimates (FCO_2 and FO_2). New studies and data mining analytics need to be tested. ANNs have the potential to generate predictive FCO_2 and FO_2 models.

4,8 References

Adachi M, Bekku YS, Rashidah W, Okuda T, Koizumi H (2006) Differences in soil respiration between different tropical ecosystems. **Applied Soil Ecology** 34:258-65. <https://doi.org/10.1016/j.apsoil.2006.01.006>

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)

de Araújo Santos GA, Moitinho MR, de Oliveira Silva B, Xavier CV, Teixeira DDB, Corá JE, Júnior NLS (2019); Effects of long-term no-tillage systems with different succession cropping strategies on the variation of soil CO₂ emission. **Science of the Total Environment** ;686:413-24. <https://doi.org/10.1016/j.scitotenv.2019.05.398>

Barlow J, França F, Gardner TA, Hicks CC, Lennox GD, Berenguer E, Castello L, Economo EP, Ferreira J, Guénard B, Gontijo Leal C, Isaac V, Lees AC, Parr CL, Wilson SK, Young PJ, Graham NAJ (2018). The future of hyperdiverse tropical ecosystems. **Nature**. 559:517-26. <https://doi.org/10.1038/s41586-018-0301-1>

Bicalho ES, Panosso AR, Teixeira DDB, Miranda JGV, Pereira GT, La Scala N (2014). Spatial variability structure of soil CO₂ emission and soil attributes in a sugarcane area. **Agriculture, Ecosystems & Environment** ;189:206-15. <https://doi.org/10.1016/j.agee.2014.03.043>

Bond-Lamberty B (2018). New Techniques and Data for Understanding the Global Soil Respiration Flux. **Earth's Future** ;6:1176-80. <https://doi.org/10.1029/2018EF000866>

Boniecki P, Zaborowicz M, Pilarska A, Piekarska-Boniecka H (2020) Identification process of selected graphic features apple tree pests by neural models type MLP, RBF and DNN. **Agriculture** ;10:1-9. <https://doi.org/10.3390/agriculture10060218>

Brady NC, Weil RR (2013). **Elementos da natureza e propriedades dos solos**. 3^a ed. Porto Alegre: Bookman. 704p

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>

Camargo AP de, Sentelhas PC (1997a). Performance evaluation of different potential evapotranspiration estimating methods in the State of São Paulo, Brazil. **Revista Brasileira de Meteorologia**, 5:89-97

Camargo AP de, Sentelhas PC(1997b) Performance evaluation of different potential evapotranspiration estimating methods in the State of São Paulo, Brazil. **evista Brasileira de Meteorologia** ;5:89-97

Canteral, K.F.F. **Aprendizado de máquina na modelagem temporal da emissão de CO₂ do solo em áreas agrícolas no Cerrado brasileiro** 2020. 41f. Dissertação (Mestrado). Universidade Estadual Paulista (Unesp). Faculdade de Ciências Agrárias e Veterinárias. Jaboticabal, 2020.

Cao C, Zhang Y, Qian W, Liang C, Wang C, Tao S (2017). Land-use changes influence soil bacterial communities in a meadow grassland in Northeast China. *Solid Earth* ;8:1119-29. <https://doi.org/10.5194/se-8-1119-2017>

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, Belk E, Boone RD (1998); Soil water content and temperature as independent or confounded factors controlling soil respiration in a temperate mixed hardwood forest. **Global Change Biology** ;4:217-27. <https://doi.org/10.1046/j.1365-2486.1998.00128.x>

Davidson EA, Janssens IA, Lou Y (2006). On the variability of respiration in terrestrial ecosystems: Moving beyond Q₁₀. **Global Change Biology** ;12:154-64. <https://doi.org/10.1111/j.1365-2486.2005.01065.x>

Dehghani M, Riahi-Madvar H, Hooshyaripor F, Mosavi A, Shamshirband S, Zavadskas EK, Chau K wing (2019). Prediction of hydropower generation using Grey wolf optimization adaptive neuro-fuzzy inference system. **Energies** ;12:1-20. <https://doi.org/10.3390/en12020289>

Ebrahimi M, Sarikhani MR, Safari Sinegani AA, Ahmadi A, Keesstra S (2019). Estimating the soil respiration under different land uses using artificial neural network and linear regression models. **Catena** ;174:371-82. <https://doi.org/10.1016/j.catena.2018.11.035>

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

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>

Golchin A, Oades JM, Skjemstad JO, Clarke P (1994). Soil structure and carbon cycling. **Australian Journal of Soil Research**; 32:1043-63.

<https://doi.org/10.1071/SR9941043>

Hamed S, Dehdashti Jahromi H (2021). Performance analysis of all-optical logical gate using artificial neural network. **Expert Systems With Applications** ;178:115029. <https://doi.org/10.1016/j.eswa.2021.115029>

Hanson PJ, Edwards NT, Garten CT, Andrews JA (2000). Separating root and soil microbial contributions to soil respiration: A review of methods and observations. **Biogeochemistry** .;48:115-46. <https://doi.org/10.1023/A:1006244819642>

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

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

Heddami, S (2014) Modeling hourly dissolved oxygen concentration (DO) using two different adaptive neuro-fuzzy inference systems (ANFIS): a comparative study. *Environ Monit Assess* 186, 597–619. <https://doi.org/10.1007/s10661-013-3402-1>

Kharb RK, Shimi SL, Chatterji S, Ansari MF(2014). Modeling of solar PV module and maximum power point tracking using ANFIS. **Renewable and Sustainable Energy Reviews** ;33:602-12. <https://doi.org/10.1016/j.rser.2014.02.014>

Kim DG, Vargas R, Bond-Lamberty B, Turetsky MR. Effects of soil rewetting and thawing on soil gas fluxes.(2012).A review of current literature and suggestions for future research. **Biogeosciences** ;9:2459-83. <https://doi.org/10.5194/bg-9-2459-2012>

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>

Kooch Y, Noghre N (2020). The effect of shrubland and grassland vegetation types on soil fauna and flora activities in a mountainous semi-arid landscape of Iran. **Science of the Total Environment** ;703. <https://doi.org/10.1016/j.scitotenv.2019.135497>

Kuzyakov Y(2010). Priming effects:Interactions between living and dead organic matter. **Soil Biology and Biochemistry** .;42:1363-71. <https://doi.org/10.1016/j.soilbio.2010.04.003>

Laganière J, Paré D, Bergeron Y, Chen HYH (2012).The effect of boreal forest composition on soil respiration is mediated through variations in soil temperature and C quality. **Soil Biology and Biochemistry** ;53:18-27. <https://doi.org/10.1016/j.soilbio.2012.04.024>

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

Lamzouri K, Latrach L, Mahi M, Ouattar S, Bartali HE, Masunaga T, Mandi L. (2017). Controlling biochemical oxygen demand in the multi-soil-layering using neural

network tool. **Journal of Materials and Environmental Science**; 8:2013
8.http://www.jmaterenvironsci.com/Document/vol8/vol8_N6/214-JMES-3059-Lamzouri.pdf

Li D, Niu S, Luo Y(2012).Global patterns of the dynamics of soil carbon and nitrogen stocks following afforestation: A meta-analysis. **New Phytologist** ;195:172-81.
<https://doi.org/10.1111/j.1469-8137.2012.04150.x>

Li M, Liu T, Duan L, Ma L, Wang Y, Zhou Y, Li Y, Zhao X, Wang X, Wang G, Lei H (2021). Hydrologic gradient changes of soil respiration in typical steppes of Eurasia. **Science of the Total Environment** ;794:148684.
<https://doi.org/10.1016/j.scitotenv.2021.148684>

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>

Liu Y, Zhang Q, Song L, Chen Y(2019). Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction. **Computers and Electronics in Agriculture**; 165:104964.
<https://doi.org/10.1016/j.compag.2019.104964>

Lucas-Borja ME, Hedo de Santiago J, Yang Y, Shen Y, Candel-Pérez D (2019). Nutrient, metal contents and microbiological properties of litter and soil along a tree age gradient in Mediterranean forest ecosystems. **Science of the Total Environment** ;650:749-58. <https://doi.org/10.1016/j.scitotenv.2018.09.079>

Myers, N., Mittermeier, R., Mittermeier, C. et al (2000). Biodiversity hotspots for conservation priorities. **Nature** 403, 853–858 <https://doi.org/10.1038/35002501>

Moreira, F. M. de S., Siqueira, J. O (2006). **Microbiologia e bioquímica do solo**. Lavras: Editora da UFLA, 2.ed., 729p.

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>

Nourbakhsh F (2007). Decoupling of soil biological properties by deforestation. **Agriculture, Ecosystems & Environment** ;121:435-8.
<https://doi.org/10.1016/j.agee.2006.11.010>

Ouyang Y, Zheng C (2000). Surficial processes and CO₂ flux in soil ecosystem. **Journal of Hydrology** ;234:54-70. [https://doi.org/10.1016/S0022-1694\(00\)00240-7](https://doi.org/10.1016/S0022-1694(00)00240-7)

Padarian J, Minasny B, McBratney AB (2020). Machine learning and soil sciences: A review aided by machine learning tools. **Soil**. ;6:35-52. <https://doi.org/10.5194/soil-6-35-2020>

Raich JW, Tufekcioglu A (2000). Vegetation and soil respiration: Correlations and controls. **Biogeochemistry**. 48:71-90. <https://doi.org/10.1023/A:1006112000616>

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

Reichardt, K; Tilm, L. C (2004) **Solo, planta e atmosfera: conceitos, processos e aplicações**. Barueri: Manole.

Riza LS, Bergmeir C, Herrera F, Benítez JM (2015). “frbs: Fuzzy Rule-Based Systems for Classification and Regression in R.” **Journal of Statistical Software**, 65:6 1–30. <http://www.jstatsoft.org/v65/i06/>.

Rubio VE, Detto M (2017) Spatiotemporal variability of soil respiration in a seasonal tropical forest. **Ecology and Evolution** ;7:7104-16. <https://doi.org/10.1002/ece3.3267>

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>

Silva-Olaya AM, Cerri CEP, La Scala Jr N, Dias CTS, Cerri CC (2013). Carbon dioxide emissions under different soil tillage systems in mechanically harvested sugarcane. **Environmental Research Letters**. 8:1-8. <https://doi:10.1088/1748-9326/8/1/015014>

de Souza LC, Procópio L (2021). The profile of the soil microbiota in the Cerrado is influenced by land use. **Applied Microbiology and Biotechnology**; 105:4791-803. <https://doi.org/10.1007/s00253-021-11377-w>

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

Strassburg BBN, Brooks T, Feltran-Barbieri R, Iribarrem A, Crouzeilles R, Loyola R, Latawiec AE, Oliveira Filho FJB, De Scaramuzza CAM, Scarano FR, Soares-Filho B, Balmford A (2017). Moment of truth for the Cerrado hotspot. **Nature Ecology & Evolution**, 1:1-3. <https://doi.org/10.1038/s41559-017-0099>

Tang J, Baldocchi DD, Xu L (2005). Tree photosynthesis modulates soil respiration on a diurnal time scale. **Global Change Biology**, 11:1298-304. <https://doi.org/10.1111/j.1365-2486.2005.00978.x>

Teixeira DDB, Bicalho E da S, Panosso AR, Perillo LI, Iamaguti JL, Pereira GT, La Scala Jr N (2012). Uncertainties in the prediction of spatial variability of soil CO₂ emissions and related properties. **Revista Brasileira Ciência Do Solo**. 36:1466-75. <https://doi.org/10.1590/s0100-06832012000500010>

Trigueiro WR, Nabout JC, Tessarolo G (2020). Uncovering the spatial variability of recent deforestation drivers in the Brazilian Cerrado. *Journal of Environmental Management*; 275. <https://doi.org/10.1016/j.jenvman.2020.111243>

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 W, Peng S, Fang J (2010). Root respiration and its relation to nutrient contents in soil and root and EVI among 8 ecosystems, northern China. **Plant Soil**. 333:391-401. <https://doi.org/10.1007/s11104-010-0354-x>

Willmott CJ (1981). On the validation of models. **Physical Geography**, 2:184-94. <https://doi.org/10.1080/02723646.1981.10642213>

Wu C, Gaumont-Guay D, Andrew Black T, Jassal RS, Xu S, Chen JM, Gonsamo A (2014) Soil respiration mapped by exclusively use of MODIS data for forest landscapes of Saskatchewan, Canada. **ISPRS J Photogramm Remote Sens** ;94:80-90. <https://doi.org/10.1016/j.isprsjprs.2014.04.018>

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** ;178:40-8. <https://doi.org/10.1016/j.catena.2019.03.001>

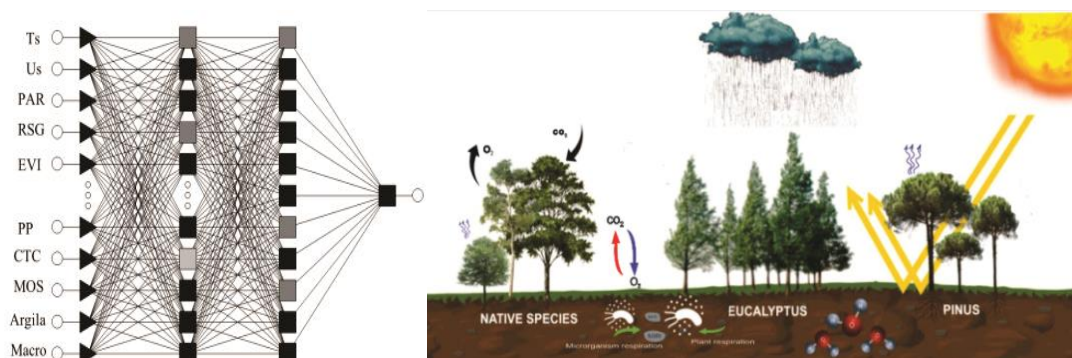
Zhao Z, Peng C, Yang Q, Meng FR, Song X, Chen S, Epule TE, Li P, Zhu Q. (2017). Model prediction of biome-specific global soil respiration from 1960 to 2012. *Earth's Futur*. 2017;5:715-29. <https://doi.org/10.1002/2016EF000480>

Zhou WJ, Sha LQ, Schaefer DA, Zhang YP, Song QH, Tan ZH, Deng Y, Deng XB, Guan HL.(2015). Direct effects of litter decomposition on soil dissolved organic carbon and nitrogen in a tropical rainforest. **Soil Biology and Biochemistry** ;81:255-8. <https://doi.org/10.1016/j.soilbio.2014.11.019>

Zounemat-Kermani M, Scholz M (2014). Modeling of Dissolved Oxygen Applying Stepwise Regression and a Template-Based Fuzzy Logic System. *Journal of Environmental Engineering*, 140:69-76 [https://doi.org/10.1061/\(asce\)ee.1943-7870.0000780](https://doi.org/10.1061/(asce)ee.1943-7870.0000780)

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

Graphical Abstract



Abstract - Changes in land and forest use influence the dynamics of soil organic matter, altering soil respiration. Using four models of machine learning (neural network, vector machine support, neuro-fuzzy inference system, and random forest), we proposed to model the temporal variability of soil CO_2 emission (FCO_2) and oxygen consumption (FO_2) in forests planted in the Cerrado, using agrometeorological variables, soil, and enhanced vegetation index (EVI) variables. 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. The predictive performance was evaluated based on the mean absolute error (MEA), root mean square error (RMSE), mean absolute percentage error (MAPE), agreement index (d), confidence coefficient (c) and determination coefficient (R^2). Estimates of FCO_2 showed better predictive performance than FO_2 . RBF neural network was best estimate for FCO_2 . RBFNN is the best architecture ($R^2 = 0.51$, $\text{RMSE} = 0.97 \mu\text{mol m}^{-2} \text{s}^{-1}$). MLPNN is the best architecture for FO_2 ($R^2 = 0.45$, $\text{RMSE} = 0.94 \text{ mg m}^{-2} \text{s}^{-1}$). ANFIS did not show good generalizability and presented the worst performance, showing the highest errors and lowest accuracy ($R^2 = 0.12$, $R^2 = 0.28$, respectively). Analyzing the performance of the two estimates, SVR-RBF for FCO_2 performed better than SVR-RBF for FO_2 . The SVR-RBF for FCO_2 in the calibration and validation phases presented similar values of the error indices and coefficient of determination ($R^2 = 0.53$ and $R^2 = 0.50$), calibration and validation respectively. The RF for FCO_2 in the calibration and validation phases presented values of the error indices: MEA (0.92 and $0.95 \mu\text{mol m}^{-2} \text{s}^{-1}$), RMSE (1.04 to $1.27 \mu\text{mol m}^{-2} \text{s}^{-1}$). The accuracy was $R^2 0.80$ and $R^2 0.60$ respectively. For FO_2 , the performance was lower than that of FCO_2 . The RMSE values for calibration and validation were 0.101 ($\text{mg m}^{-2} \text{s}^{-1}$) and 0.110 ($\text{mg m}^{-2} \text{s}^{-1}$), with accuracy shown by $R^2 0.57$ and 0.30 . The type of forest influenced temporal variability in soil respiration. We found that 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 .

Keywords: Soil gases flux; carbon loss; soil-atmosphere relationship; land-use change; artificial intelligence.

5.1. Introduction

Land use and land cover change (LULCC), are considered the second largest anthropogenic source of greenhouse gas (GHG) emissions. In recent years, deforestation in the tropics and subtropics has intensified, especially in the Cerrado biome in Brazil. This is motivated by the expansion of agricultural frontiers, especially monocultures. The Cerrado is considered a global hotspot for biodiversity conservation (Myers et al., 2000) with an area of approximately 2 million km², corresponding to approximately 24% of the Brazilian territory. The degradation of the biome has led to the loss of carbon stocks (Cstock) in the aboveground and soil biomass (Don et al., 2011), corroborating the emission of carbon dioxide to the atmosphere and loss of biodiversity (de Espindola et al., 2021).

On the other hand, management practices such as reforestation can lead to an increase in soil organic carbon (SOC) stocks and potentially mitigate high GHG emission rates, especially in tropical regions (Pulrolnik et al., 2009). Forests play a key role in the cycle of C and nitrogen (N). In the soil, the C balance is dependent on the relationship between additions of photosynthesized C and losses to the atmosphere (Gourlez de la Motte et al., 2016). Landscape changes, vegetation type, canopy compositions, and stand age promote changes in substrate quantity and quality, influencing the C and N fractions of organic matter (Kooch and Ghaderi, 2021) in the microbial community and influencing soil respiration (R_s) dynamics (Alekseev et al., 2018; Wu et al., 2020).

Soil organic matter is the main source of energy for microorganisms and is an indicator of soil quality and microbial activity (Doran and Parkin, 1994). In soil, the decomposition of this organic matter, when promoted by aerobic microorganisms in their metabolic process, consumes oxygen (FO_2) and soil CO_2 emission (FCO_2) (Luo and Zhuo, 2006). R_s is the second largest component of carbon flux after photosynthesis in the Earth's biosphere (Zhao et al., 2021); is a crucial ecosystem process that releases C from the soil in the form of CO_2 (Liang et al., 2010); and can be divided into autotrophic (root and associated mycorrhizae) and heterotrophic soil respiration (R_a and R_h , respectively), associated with microbial decomposition of organic matter (Raich and Schlesinger, 1992.). It is correlated with climatic factors such as solar radiation and air temperature and is highly sensitive to variables such as precipitation, soil temperature, soil water content evapotranspiration (La Scala et

al., 2003; Liang et al., 2010) LI et al., 2021), soil-related factors such as soil fertility (Ward et al., 2017), physical attributes (Wang et al., 2003; Panosso et al., 2011), oxygen (O₂) availability (Ferraz de Almeida et al., 2018), and substantially with variations in space and time (Han et al., 2018; Huang et al., 2020) Tang et al. (2015) stated that the pattern of temporal variability of soil respiration in forested areas is modulated by photosynthetic activity, and spectral vegetation indices (VIs) have been used in these studies (Muraoka et al., 2013; Yan et al., 2020).

In recent years, artificial intelligence techniques combining different algorithms have been successfully applied to model soil respiration (Farhate et al., 2018; Freitas et al., 2018; Zhao et al., 2020), reducing the uncertainty of these predictions (Liu et al., 2018). According to Shadrin et al. (2020), supervised machine learning, such as support vector regression (SVR) and artificial neural networks (ANN), is considered a promising and efficient tool for interpreting high variability and nonlinear data, especially environmental data. The random forest (RF) algorithm is widely considered one of the best methods for modeling complex interactions, offering resistance to overfitting and relative robustness with respect to noise features (Chagas et al. 2016).

One popular artificial intelligence is neuro-fuzzy technique which is a hybrid combination of artificial neural networks (ANN) and fuzzy inference system (FIS) Adaptive neuro-fuzzy inference system (ANFIS) (Buragohain and Mahanta, 2008). ANFIS is a potent approach to modeling the input and output relationship in non-linear and complex systems (Afroozeh et al., 2018), although, depends heavily on the choice of process variables involved as well as the available data set and the domain used for training purposes (Buragohain and Mahanta, 2008).

However, studies have focused only on CO₂ emissions from soil, and few studies have modeled oxygen consumption, which is essential for soil respiration dynamics. Thus, the study of models can assist in estimating FCO₂ and FO₂ in Brazil's tropical soils, where few studies of FCO₂ prediction using machine learning algorithms are available. In this context, we hypothesized that environmental variables, vegetation index (EVI), and soil attributes are effective in predicting the temporal variability of soil CO₂ emission and O₂ influx using four machine learning techniques. The objective of this work was to evaluate the predictive performance of artificial neural networks (ANNs), using three different architectures (general regression neural network (GRNN), multi-layer perceptron (MLPNN), and radial basis

function (RBFNN), support vector regression (SVR), radial basis function kernel (RBF), adaptive neuro-diffuse inference system (ANFIS), and random forest (RF) to model the temporal variability of FCO_2 and FO_2 in planted forest areas in the Brazilian Cerrado.

5.2 Materials and methods

5.2.1 Data used

The temporal variability datasets of soil respiration and oxygen consumption used in this study were from two field studies conducted in adjacent areas of forests planted with *Eucalyptus camaldulensis* (EUC), *Pinus caribaea* var. *Hondurensis* (PI), 21 native species (NS), and Brazilian Cerrado (Table 1 of Supplementary Material). The first study was conducted between November 2015 and May 2016, totaling 193 studies (Vicentini et al., 2019). The second study was conducted from March 2017 to October 2018, with a total of 273 evaluation days (Ramos, 2018).

5.2.2 Site description

The areas are localized in the municipality of Selvíria, Mato Grosso do Sul, Central Brazil, (20° 20' 53.41" South, 53.41" South and 51° 23' 55.50" West, altitude 354 m) (Figure 1). The climate of the region was classified as C1dAa' by the Thornthwaite system (De Souza Rolim et al., 2007), with a mean annual temperature of 23.5 °C, mean annual rainfall of 1,300 mm, and relative humidity between 70% and 80%. The local soil was classified as dystrophic Red Latosol (LVd) based on the Brazilian Soil Classification System (Embrapa, 2018). The relief was characterized as moderately flat and undulating.

In the early 1970s, the policies and projects of land use and occupation by the rulers of the country accelerated the deforestation of native vegetation throughout the state of Mato Grosso do Sul. Reforestation occurred only in 1986 in the three areas investigated (Figure 2).

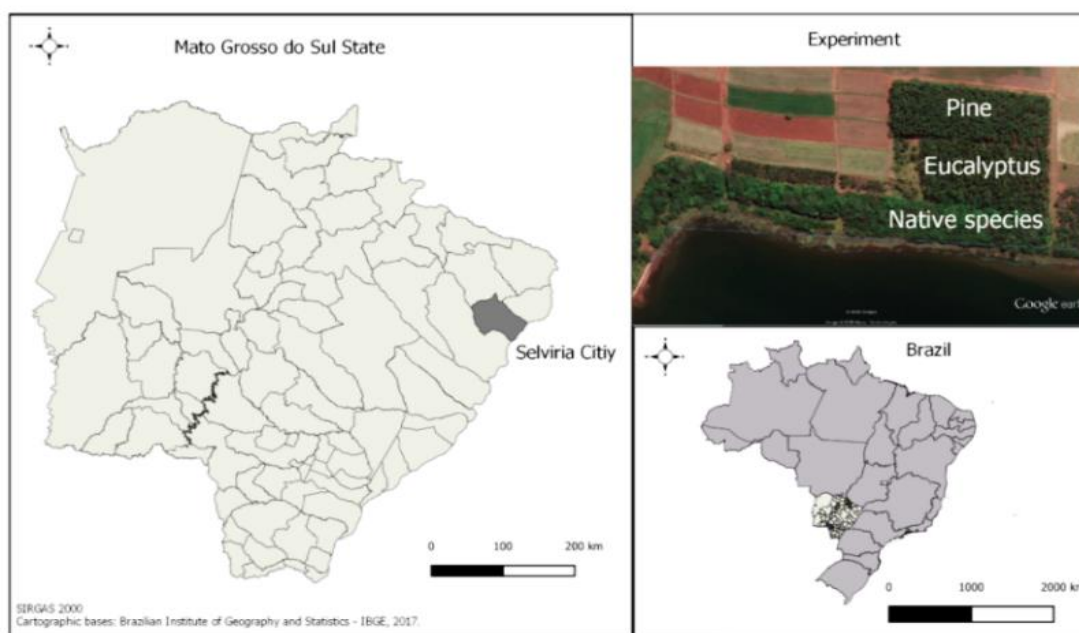


Figure 1. Location of the experimental area in Selvíria, Mato Grosso do Sul (MS), Brazil. SIRGAS - Geocentric Reference System for the Americas and IBGE - Brazilian Institute of Geography and Statistics. The satellite images were obtained by Google Earth.

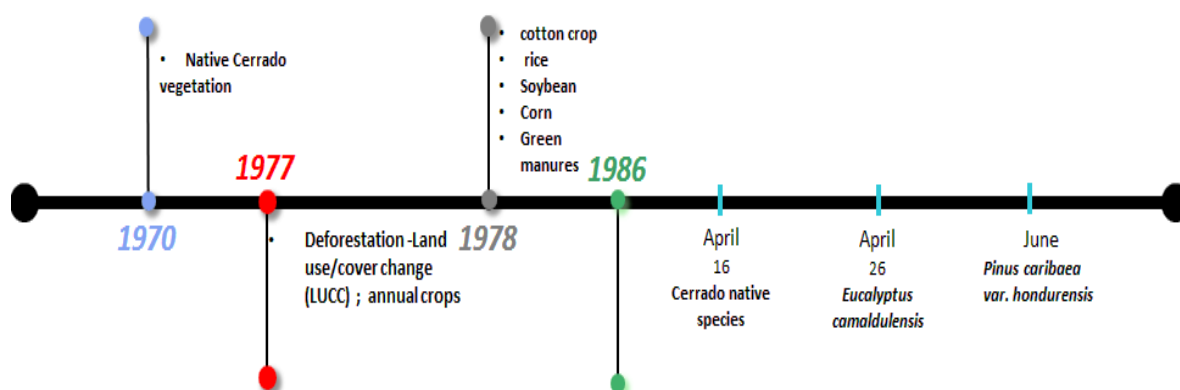


Figure 2. A timeline with the history of the area

5.2.3 Environmental data and Soil sampling and analysis of physico-chemical properties of soils

The meteorological data (precipitation (PP- mm); wind speed- w (m s^{-1}); evapotranspiration by Penman_Monteith (ET_0 - mm day^{-1}); photosynthetically active radiation (PAR - $\mu\text{mol m}^{-2}\text{s}^{-1}$); global radiation (RSG - $\text{MJ m}^2 \text{day}^{-1}$); atmospheric pressure (P_{atm} - Kpa) and air temperature (T_{atm} - $^{\circ}\text{C}$), were obtained at <http://clima.feis.unesp.br/index.php>. MODIS-EVI data were obtained using SATVeg - Temporary Vegetation Analysis System, developed by Embrapa Informática

Agropecuária. MODIS images used for SATVeg development were obtained from the LP DAAC Data Pool (<http://lpdaac.usgs.gov>). These images are composed of the maximum value of the vegetation indices every 16 days. The spatial resolution of the MODIS sensor for the product used was 250 m. To minimize the effects of the noise of the time series, the data were obtained by pre-filtering and filtering available in the tool. The Savitzky–Golay filter was selected.

Soil samples were obtained at a depth from 0.0 to 0.10 m. Subsequently, based on the methodology proposed by (RAIJ et al., 2001), the following were determined: concentrations of soil organic matter (MOS), phosphorus (P), calcium (Ca), magnesium (Mg), potassium (K), and aluminum (Al); potential acidity (H + Al) at pH 7.0 and pH (CaCl₂); cation exchange capacity (CEC); and base sum (SB). Physical analyses of soil density (Bd), macroporosity (macro), microporosity (micro), and granulometric analysis (sand, silt, and clay) were performed based on the methodology proposed by Embrapa (1997).

5.2.4 Measurements of soil respiration, soil temperature, soil water content and Oxygen Uptake

Soil respiration was measured using an LI-8100 portable system (LI-COR, Nebraska, USA), description in (Panosso et al., 2011; de Araújo Santos 2019). In each area, 25 sampling points (repetitions) were installed with polyvinyl chloride (PVC) collars (diameter, 0.10 m; height, 0.085 m; depth, 3 cm). In all areas, evaluations were performed on the same day and in the morning (07:00 AM to 12:00 PM). In addition to soil respiration, we measured the soil temperature and soil water content. The soil temperature was measured using a digital thermometer at a depth of 0.12 m. The soil water content was determined based on the estimated dielectric constant of water, using a portable TDR-Campbell® system (Hydrosense TM, Campbell Scientific, Australia), which comprised a probe with two 0.12 m rods inserted into the soil near the PVC collars (Ramos et al., 2018; Vicentini et al., 2019).

To determine the O₂ uptake in the soil, 10 points in each area were selected using a portable UV flux (0%–25%) sensor (CO₂Meter, Inc., Ormond Beach, FL, USA). The soil O₂ influx was calculated using Equation (1), where the atmospheric pressure, air temperature, and chamber volume were considered. Further information is available in the paper by (Ferraz de Almeida et al., 2018)

$$FO_2(t) = \frac{dO_2}{dt} \cdot A \quad \text{Eq. (1)}$$

where $FO_2(t)$ is the amount of O_2 measured in time, and dO_2 is the change in concentration in relation to the unit of time (dt) on the surface area of the collar (A) (Jassal et al., 2012; Giacomo et al., 2014).

The initial concentrations of the readings were obtained in parts per million (ppm). The volume of the PVC camera was equal to 0.00066 m^3 , with an area equal to 0.008 m^2 . The volume measured using the sensor (ppm) was converted to mol of O_2 by using the ideal gas equation as follows

$$P(\Delta V) = (\Delta n)RT \quad \text{Eq. (2)}$$

The soil FO_2 rate (dO_2/dt) was calculated via linear interpolation of the concentration values as a function of time, taking into account the atmospheric pressure, temperature and volume of the gas trapped in the chamber, using FO_2 (Ferraz de Almeida et al., 2018)

$$FO_2 (g \cdot m^{-2} s^{-1}) = \frac{dO_2 \cdot 10^{-6} PM}{dt RT} H \quad \text{Eq. (3)}$$

where dO_2/dt is the amount of O_2 (ppm) measured at time t (s); P is the atmospheric pressure (Pa); M is the O_2 molar mass ($g \cdot m^{-3}$); R is the universal gas constant ($8.31 \text{ J mol}^{-1} \text{ K}^{-1}$); T is the absolute temperature; (K) and $H=V/A$:the ratio of the volume (V) = 0.00066 m^3 to cross-sectional area (A) = 0.008 m^2 of the camera above the ground (soil surface). ((Ferraz de Almeida et al., 2018; Vicentini et al., 2019)

5.2.5 Statistical Analysis

In this study, the database of dependent variables for modeling FCO_2 comprised ($n = 2125$) observations and FO_2 ($n = 78$). To obtain better estimations of FO_2 , we worked with the average values of oxygen consumption. Therefore, our data

set only had 78 observations, referring to the evaluations between 11/07/2015 and 15/07/2016. The average FCO₂ value in the database ranged from 2.98 (μmol m⁻²s⁻¹) in the pine forest to 5.61 (μmolm⁻²s⁻¹) eucalyptus forest. For FO₂ value ranged from 0.15 (mg m⁻²s⁻¹) for in pine forest at 0.37 (mg m⁻²s⁻¹) for the native species (Table 3).

Table 3. Mean and coefficient of variation of soil CO₂ emission, oxygen uptake, soil temperature, and soil water content in planted forest area. Source: Ramos (2018) and (Vicentini et al., 2019)(adapted).

	FCO ₂		CV		FO ₂		CV		TS		CV		CAS		CV		Referência	
Eucalyptus	3.96	a	48	0.19	ab	58	24	b	8.5	5.55	B	60	(Ramos, 2018)*					
Pinus	2.98	b	51	0.15	b	40	24	b	9.9	5.94	B	61						
Eucalyptus	5.61	a	38	0.37	a	85	27	a	3.8	10.67	A	41	(Vicentini et al. 2019)**					
Pinus	4.06	b	38	0.37	a	87	26	a	5	11.57	A	33						
Native species	5.53	a	33	0.37	a	93	26	a	5.1	15.62	B	33						

FCO₂: soil CO₂ emission (μmol m⁻² s⁻¹);FO₂: = Oxygen Uptake(mg m⁻²s⁻¹);Ts=soil temperature(°C); soil water content(%)CV: coefficient of variation (%). FCO₂=N: 20 e FO₂= N:10;** FCO₂=N: 25 e FO₂= N:10 Means followed by the same letter (lower case) in the columns do not differ (Tukey test; p < 0.05).

Table 4. Average of chemical and physical soil attributes in 0–0.10 m layer in Eucalyptus, Pinus and Native Forest areas.

	Eucalyptus		Pinus		Native species		CV(%)
Atributos							
P (mg dm ⁻³)	6.52 ± 0.50	a	6.32 ± 0.51	a	6.16 ± 0.46	a	24.53
SOM (g dm ⁻³)	29.76 ± 1.12	a	20.04 ± 0.56	c	25.2 ± 0.59	b	14.74
pH (CaCl ₂)	4.31 ± 0.07	a	4.00 ± 0.02	b	4.44± 0.05	a	6.7
SB (mmolc dm ⁻³)	25.98 ± 3.75	b	8.56 ± 0.52	c	33.06 ± 2.89	a	18.8
CTC (mmolc dm ⁻³)	86.99 ± 3.10	a	83.91 ± 3.63	a	82.98 ± 1.76	a	14.74
Al+3 (mmolc dm ⁻³)	5.88 ± 0.93	a	13.92 ± 0.94	b	5.88 ± 0.93	a	37.92
Macro (m3 m ⁻³)	0.08 ± 0.006	a	0.042 ± 0.003	b	0.049 ± 0.007	b	52.27
Micro (m3 m ⁻³)	0.31 ± 0.006	c	0.368 ± 0.004	b	0.394 ± 0.006	a	7.39
Bd (g dm ⁻³)	1.47 ± 0.02	b	1.57 ± 0.01	a	1.34 ± 0.01	c	6.54
Sand g kg ⁻¹)	610.94 ± 9.88	b	648.27 ± 5.24	a	542.38 ± 5.19	c	5.27
Silt (g kg ⁻¹)	56.471 ± 1.70	a	25.76 ± 3.66	b	65.21 ± 2.77	a	32.13
Clay (g kg ⁻¹)	332.58 ± 9.38	b	325.95 ± 6.52	b	392.39 ± 6.00	a	9.67

N = 25; Means followed by the same letter (lowercase) in the columns do not differ by Tukey's test (p < 0.05), P= phosphorus; SOM= soil organic matter; pH = hydrogen potential; Al = exchangeable aluminum concentration; Macro= macroporosity; Micro = microporosity; Bd = bulk soil density; sand content; silt content; clay content and CV= coefficient of variation. Source: Ramos (2018); Vicentini et al., 2019 (adapted).

5.2.6 Principal Component Analysis (PCA)

PCA was used in the preprocessing stage to extract the most important patterns of temporal variability in soil respiration such that the dataset dimension can be reduced. Variables with a higher power of contribution in the components were selected as inputs: neural networks (ANN); support vector regression (SVR); ANFIS; and random forest (RF). In this study, we used 22 variables, including FCO_2 and FO_2 , as inputs in PCA. PCA was performed using R software (version 4.0.2; R Core Team 2019, R Foundation for Statistical Computing, Vienna, Austria), using the R packages FactoMinerR (Le, Josse, and Husson, 2008), and factoextra (Kassambara and Mundt, 2017).

5.2.7 Artificial Neural Network (ANN)

An ANN can assimilate the main patterns of information within a multidimensional information domain (Kalogirou et al.1999.) and perform predictions during its learning process. The mathematical neuron (Figure 4) receives one or more information (input signals-independent variables) from other sources and then performs nonlinear operations, which results in only one output signal (dependent variable), and an algorithm maps the input space to the output space (Kalogirou et al.1999).

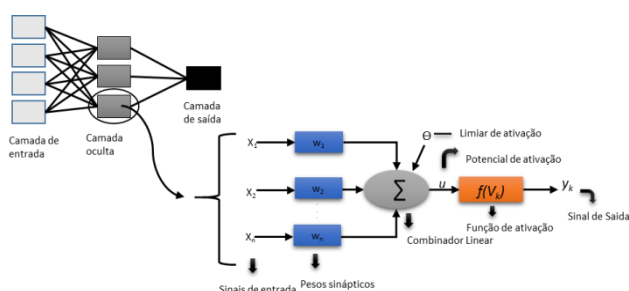


Figure 4. Model of neural network and McCulloch–Pitts-type artificial neuron [adapted from (Haykin, 2001) and from (Vendruscolo et al., 2017)].

This model comprises three basic elements:

- a set of n input connections ($x_1, x_2, \dots, x_n$) characterized by weights ($w_1, w_2, \dots, w_n$);
- an adder (Σ) to accumulate input signals;

- an activation function (ϕ) that limits the allowable amplitude range of the output signal (y) to a fixed value.

The architecture of the ANN reflects the manner in which the neurons are structured, both in terms of the number of layers and number of neurons in each layer. The RBF neural network has a single hidden layer, in which the neurons therein constitute only the radial basis function (RBF). Its learning can be performed in two distinct and independent stages: unsupervised and supervised clustering (Schwenker et al., 2001). They are considered hybrids. The most typically used radial basis is the Gaussian function (Haykin, 2009). Eq. (4)

$$\phi(V) = e^{\left(-\frac{V^2}{2\sigma^2}\right)}$$

Onde $V = \|x - t\|$

x: input vector

Eq. (4)

t: center of radial function

σ : width of radial function

Unlike RBF neural networks, multi-layer perceptron (MLP) comprise an input layer, one or more hidden layers, and an output layer. They are trained using the backpropagation algorithm, which is one of the most widely used algorithms for training nets. This algorithm iteratively adjusts the weights to reduce the error between the actual and desired network outputs (Kalogirou et al.1999).

The activation function used in this study was the sigmoid (logistic) continuous function, also known as the “S-shape,” as it is the most typically used function in this area. Eq. (5).

This function assumes a continuous range of values from zero to one. According to Haykin et al. (2009), the activation function, which is denoted by $\phi(\cdot)$, defines the output of a neuron based on the input activity level.

$$\phi(v) = \frac{1}{1 + \exp(-pv)} \quad \text{Eq. (5)}$$

where $\phi(v)$ is the sigmoid activation function, and p is the slope of the function. By varying p , sigmoid functions for different slopes can be obtained.

One hundred MLPNNs and 100 BRNNs were trained for each prediction of FCO_2 and FO_2 . In all ANN models, 70% and 30% of the data were used for calibration and validation, respectively. The training and validation data were randomly selected using the program. All ANN architectures were developed using the Intelligent Problem Solver (IPS) tool of Statistica®, software, version 7.0. (StatSoft Dell Software Inc., Tulsa, Oklahoma). The IPS module replaces the traditional heuristic process of an ANN design with interleaved search algorithms, and it determines the selection of inputs (if necessary), number of hidden units, and other key factors in the network design (Landeras et al., 2008).

The IPS evaluates the most predictive variables and selects the optimal number of neurons per layer to avoid underfitting. When the number of neurons in the hidden layer is small, may not be sufficient, whereas an overfitted network due to an excessive number of neurons may cause the training data to be memorized. In this study, the IPS mode was configured to return the 10 ANNs with the smallest root mean square error (RMSE), presented randomly by the program; subsequently, the best one from each architecture was selected. The criterion to stop the calibration algorithm was a total number of cycles of 100, or when the mean square error was less than 1% (Haykin, 2001). During the training of the ANNs, the IPS performed a sensitivity analysis, which aimed to classify the variables that affected network training the most.

5.2.8. Regression by Support Vector Machine (SVM)

The SVR is a supervised machine-learning algorithm for solving classification and regression problems (Drucker et al., 1997). It is popularly known as a nonlinear method that refines its structure based on the input data. As described by (Ghorbani et al., 2017), for certain training data represented by $(x_1, y_1); \dots; (x_N, y_N)$, where x is an input vector and y is a corresponding output vector, the regression function of the SVM can be expressed as

$$f(x) = \omega \cdot \phi(x) + b, \quad \text{Eq. (6)}$$

where ω is a weight vector, b a bias, and ϕ a nonlinear transfer function. The purpose of ϕ is to nonlinearly map the input vectors to a high-dimensional feature space by

(Ghorbani et al., 2017). Readers are referred to (Drucker et al., 1997) for more details.

The accuracy of the SVR model with RBF depends on the selection of parameter C (cost) and positive regularization, which correspond to the weight parameters in the optimization problem for minimizing the empirical error. From the training penalty variable ε (epsilon) and the kernel parameter γ (gamma), For the two models (FCO₂ and FO₂), the algorithm was executed using Statistica®, software, version 7.0. (StatSoft Dell Software Inc., Tulsa, Oklahoma)

Tipo I (ϵ -SVM para regressão ou ϵ -SVR)

$$\frac{1}{2}W^TW + C \sum_{i=1}^n (\xi_i \xi_i^*) \quad \text{Eq. (7)}$$

$$\begin{cases} y_i - W^T \phi(x_i) - b \leq \varepsilon + \xi_i \\ W^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i \xi_i^* \geq 0, i = 1, \dots, n \end{cases} \quad \text{Eq. (8)}$$

where C is the regulation term, w is the vector of parameters, b is a constant, and ξ_i is the error clearance variable outside the acceptance margin.

The kernel selected was the RBF. The SVM package performs an automatic adjustment by cross-validating the parameters (C; and ε). The cross-validation algorithm had a capacity constant C value, which was estimated to be 1. The maximum number of iterations selected was 100, with an error of 0.001. Hence, the different configurations were evaluated for the prediction of FCO₂ and FO₂, and the best performances were selected based on the following statistical indicators: correlation coefficient (r), MSE, and RMSE. An analysis of variance of SRVs was performed using the F-test (0.01 probability level), and the statistical measures were evaluated by R software (version 4.0.2; R Core Team 2019, R Foundation for Statistical Computing, Vienna, Austria), using the R packages Metrics (Ben Hamner [aut, cph], Michael Frasco [aut, cre], Erin LeDell [ctb]).

5.2.9 Adaptive Neuro-Fuzzy Inference System (ANFIS)

Fuzzy theory is an algorithm inspired by human-life decision-making and has gained popularity among researchers in different fields (Chen et al., 2017). The ANFIS is a hybrid model based on the combination of an ANN (typically an RBF neural network) and a fuzzy logic system; one of its main advantages is its ability to capture the benefits of both models in a single framework (Hipni et al., 2013). The ANFIS applies neural learning rules to identify and adjust the parameters and structure of a *fuzzy* inference system (FIS) (Jang, 1993; Laib et al. 2017), therefore, it can yield models with better results than an FIS alone in solving nonlinear problems.

The most important feature that distinguishes an ANFIS from an ANN is that the former uses a set of “If... then...” fuzzy rules. To exemplify the operation of an ANFIS, a first-order Sugeno fuzzy model is presented herein. As described by (Teimouri and Baseri, 2014), we assume an FIS with (m inputs) and an output (y) (as shown in Figure 5); this inference system is composed of five layers, and each layer comprises several nodes.

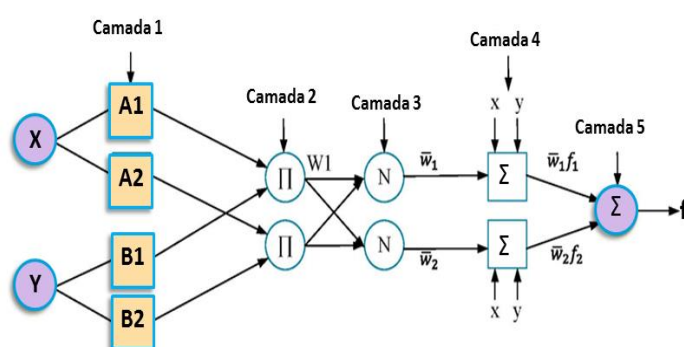


Figure 5. ANFIS structure with two inputs and one output. [adapted (Khosravi et al., 2018).

- The first layer is the fuzzification layer, in which the signal obtained from each node is transferred to the other layer. The nodes in this first layer are adaptive, which implies that the outputs of the nodes depend on the parameters belonging to these nodes and are known as premise parameters. (Shu and Ouada, 2008):
- The second layer obtains the rule with the degrees of pertinence, and each node output demonstrates the fuzzy rule firing strength.
- The third layer calculates the normalized activation degree; each node in this layer is a fixed node labeled N.

- The fourth layer performs defuzzification; the output value for each rule is calculated from the value of the previous layer.
- The fifth layer calculates the overall output by summing all input signals.

In the ANFIS, the input membership function is “gaussmf,” and the output membership function is “linear.” R software (version 4.0.2; R Core Team 2019, R Foundation for Statistical Computing, Vienna, Austria), using the R packages `frbs`: Fuzzy Rule-Based Systems for Classification and Regression Tasks. Riza LS, Bergmeir C, Herrera F, Benítez JM (2015).

5.2.10 Random Forest Regression

Random Forest models, which are tree-structured models, are used for both classification and regression (Breiman, 2001). It combines predictor (forest) trees that are generated from a random, independently sampled vector with the same distribution for all trees in a forest (Horst-Heinen et al., 2021). To construct a regression tree, the explanatory variables are partitioned based on their similarities and dissimilarities (Wang and Witte, 1997), and the final result is based on the average of the results of all trees (Chagas et al., 2016). This is a nonparametric statistical learning technique. Sena et al., (2021), described the Classification and regression by randomForest as reported (Liaw and Wiener (2002): During learning, three hyperparameters must be optimized: the number of trees (`ntree`), minimum amount of data in each terminal node (node size), and number of variables tested in each tree (`mtry`). The analysis was performed in Python 3.6, Jupyter Notebook and the library scikit-learn's `sklearn` ensemble for the development of a random forest regressor model (version 0.21) (<http://scikit-learn.org/stable/modules/ensemble.html>).

5.2.11 Statistical Measures of Performance Evaluation of Different Models

For each machine-learning method used in this study, the performance was evaluated using the following statistical indicators: r (9), coefficient of determination (R^2) (10), RMSE (11), mean absolute error (MAE) (12), and mean absolute percentage error (MAPE) (13).

$$r = \frac{\sum \sigma_{ht_i} \times \sigma_{hp_i} - \bar{\sigma}_{ht} \times \bar{\sigma}_{hp}}{(n-1) S_{\sigma_{ht}} S_{\sigma_{hp}}} \quad \text{Eq.(9)}$$

$$R^2 = 1 - \frac{\sum_i (\sigma_{hp_i} - \sigma_{ht_i})^2}{\sum_i (\sigma_{hp_i} - \bar{\sigma}_{hp_i})^2} \quad \text{Eq.(10)}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\sigma_{ht_i} - \sigma_{hp_i})^2} \quad \text{Eq.(11)}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\sigma_{ht_i} - \sigma_{hp_i}| \quad \text{Eq.(12)}$$

$$MAPE = \left[\frac{1}{n} \sum_{i=1}^n \left| \frac{\sigma_{ht_i} - \sigma_{hp_i}}{\sigma_{ht_i}} \right| \right] \times 100 \quad \text{Eq.(13)}$$

Note that σ_{ht} , σ_{hp} are, respectively observed and estimated value, n and estimated value, n number of observations,, $\bar{\sigma}_{ht}$, $\bar{\sigma}_{hp}$ predicted mean values.

5.2.12 Summary of the Methodology Used in This Study

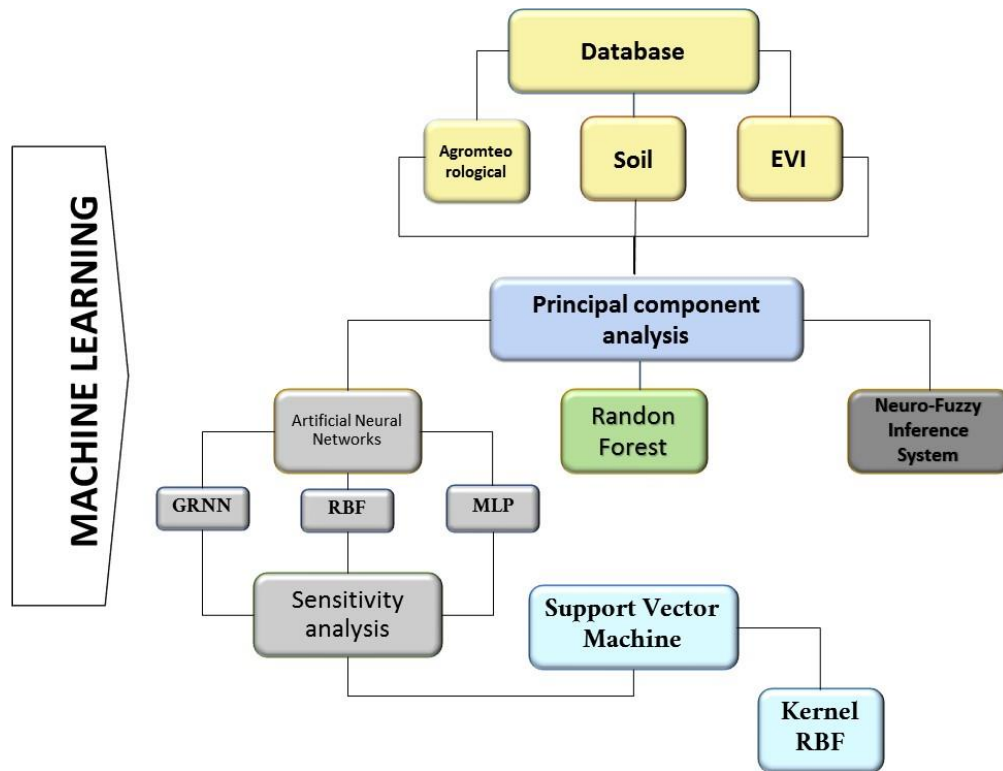


Figure 6. Diagram illustrating techniques investigated for predicting soil CO₂ emission (FCO₂) and oxygen uptake (FO₂).

5.3 Results

5.3.1 Principal Component Analysis

The major principal components explained 31.3% (PC_1) and 20.14% (PC_2) of the total variance of the original data (Figure 7), thus together explaining 51.44% of the original variance (Table 5). PC_1 had the highest percentage of explained variance. The variables that contributed the most to this component in order of relevance were SOM (0.97), SB (0.92), m (-0.87), Macro (0.84), Micro (-0.74), Silt (0.74), P (0.70), and FCO_2 (0.50). The relationships between these attributes and PC_1 may be characterized as reflecting the processes associated with the production and transport of CO_2 . Regarding PC_2 , the following variables were retained, in order of relevance: RSG (0.88%), ET_o (0.88%), PAR (0.82%), T_{atm} (0.76%), EVI (0.65%), T_s (0.63%), and FO_2 (0.44%). PC_2 is related to climatic variables, biophysical ET_o , soil temperature, and FO_2 .

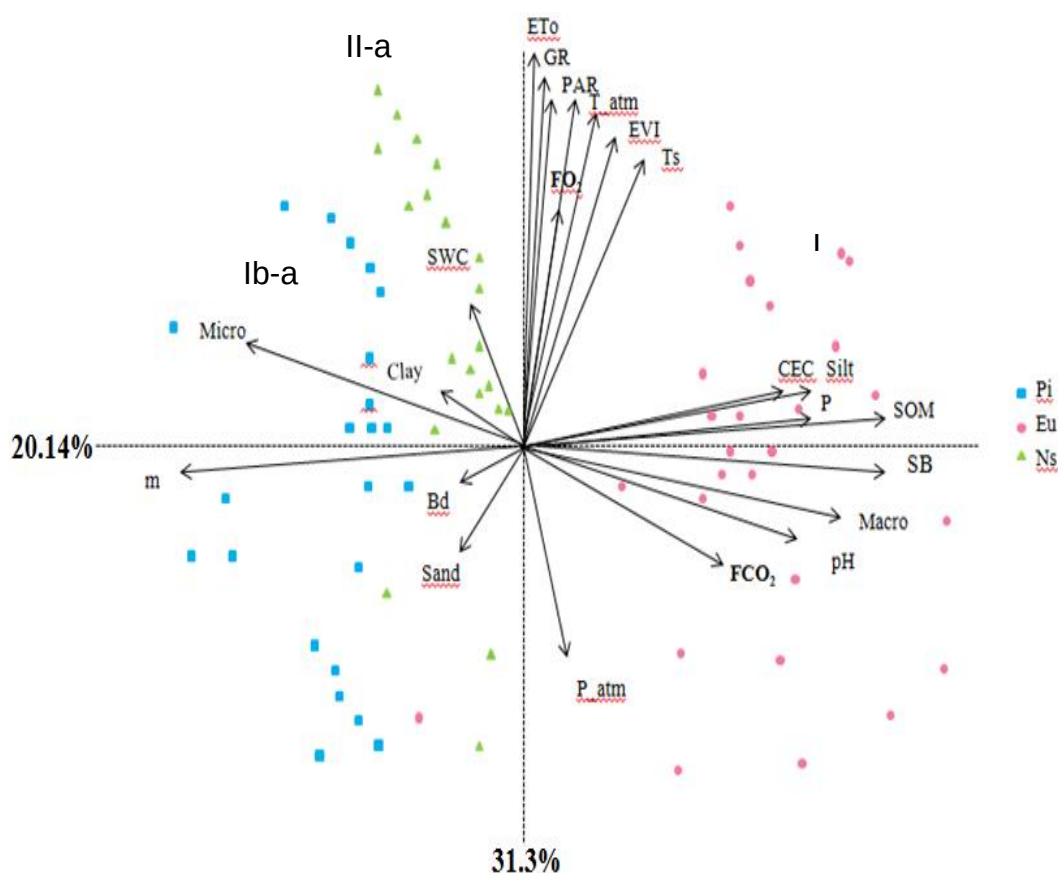


Figure 7: Biplot of the principal components

Table 5. Correlations between soil attributes, climate variables, and spectral indices EVI, FCO₂, and FO₂ and the first two principal components (PC₁ and PC₂) in forest areas.

Principal Components	CP ₁	CP ₂
Variance explained (%)	31.3*	20.14*
Correlation		
FCO ₂	0.50	-0.27
FO ₂	0.06	0.44
T _s	0.17	0.63
SWC	-0.10	0.40
T _{air}	0.06	0.76
P _{atm}	0.03	-0.58
GSR	0.10	0.88
PAR	0.01	0.82
ET _o	0.07	0.88
P	0.70	0.03
SOM	0.97	0.01
pH	0.77	-0.22
SB	0.92	-0.07
CEC	0.65	0.14
m	-0.87	-0.03
Macro	0.84	-0.17
Micro	-0.74	0.19
BD	-0.06	-0.09
Sand	-0.12	-0.17
Silt	0.74	0.12
Clay	-0.17	0.15
EVI	0.09	0.65

Analysis ACP 1 and ACP 2 with the meteorological variables: T_{atm} (°C) = atmospheric temperature; P_{atm} (Kpa)= atmospheric pressure; GR(MJ m⁻² day⁻¹) = global radiation; PAR(μmoles/m²) = photosynthetically active; ET_o PN-M (mm/day) = evapotranspiration Penman Monteith. Physical and chemical soil attributes: T_s (°C) = soil temperature; SWC (%) = soil water content. Covariates in (mmolc dm⁻³), CEC= cation exchange capacity; SB = base sum. P (mg dm⁻³) = exchangeable phosphorus; SOM (g dm⁻³) = organic matter; pH (CaCl₂) = hydrogen potential; m (%) = aluminum saturation. Covariates in (m³m⁻³), macro= macroporosity; micro= microporosity; Bd (g cm⁻³) = soil bulk density. Covariates in (g Kg⁻¹), Sand = sand content; Sil = silt content; Clay = clay content. EVI = spectral index (enhanced vegetation index); FCO₂ (μmol m⁻²s⁻¹), CO₂ emission from soil, FO₂ (mg m⁻²s⁻¹), O₂ uptake in pine (PI), eucalyptus (EU) and native species (NS) areas.

The biplot graph (CP₁ and CP₂; Figure 6) indicates the existence of two groups; thus, we characterized the variables that most discriminated them. Group I, on the right side of the biplot with greater dispersion, was represented by the eucalyptus forest. With the exception of aluminum saturation and microporosity, which characterize the pine area, the attributes retained in CP₁ characterize the eucalyptus area. Group II, located on the left side of the biplot, consisted of two subgroups of native species: (II-a), represented by the clay fraction and content of water, and pine forest (IIb).

5.3.2 Neural Networks

Based on the results of the PCA, the following variables were used as inputs for the models in the estimation of FCO₂ and FO₂: soil temperature; atmospheric pressure; solar radiation; photosynthetically active radiation; evapotranspiration; EVI; soil organic matter; macroporosity; microporosity; pH; aluminum; saturation bases; phosphorus; cation exchange capacity; silt; SWC. Although the contribution of soil water content was low under PCA, we included this variable in the training data set of all ANNs because soil water content is an important factor that controls the diffusion of gases in the soil and is essential in soil and root biological processes. In addition, we added the forest management type (trat) as a predictor variable in the estimation of FCO₂ and FO₂ in the neural networks. Thus, the data set constitutes 17 input variables for the models. Several RBF and MLP networks were trained to predict FCO₂ and FO₂.

Table 6 presents the best results for each estimate. Different combinations of input variables were tested using ANN. The best calibration results were observed when 16 variables were maintained as inputs of the models, thus evidencing the importance of these variables in the learning process of ANN in the prediction of the temporal variability of FCO₂ and FO₂, and the large number probably reflects the heterogeneity of the database. All ANN models studied were significant ($p < 0.01$; Table 6). The best performance was observed for ANN 4 with FO₂ (Figure 7).

In the calibration phase and validation the ANN-FCO₂ presented MEA indices with values ranging from 0.93 to 1.056 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and 1.021 to 1.086 $\mu\text{mol m}^{-2} \text{s}^{-1}$, respectively. The MEA value best serves for comparing different models because it reflects the ability of the network to predict the correct output. Regarding the

accuracy of the models, all showed good correlation with values in calibration ($r = 69.0$ to 78.0) and validation ($r = 0.67$ to 0.78), respectively. Figure 7 presents a performance graph for the validation phase. In addition, RBFNN was more accurate in estimating the temporal variability of FCO_2 ($R^2 = 0.60$; $R^2 = 0.51$), calibration, and validation with the lowest absolute and percentage errors (MAPE = 21.91 and 26.29%), calibration and validation respectively. The values for MLPNN were $R^2 = 0.48$, $R^2 = 0.44$, with (MAPE = 28.65 and 30.75%), calibration and validation respectively, indicating that there were more errors associated with this model. For the two ANNs, in validation, the differences between the predicted and observed data were low, as indicated by the variation in RMSE from 0.940 to $0.970 \mu\text{mol m}^{-2} \text{s}^{-1}$ (Figure 8).

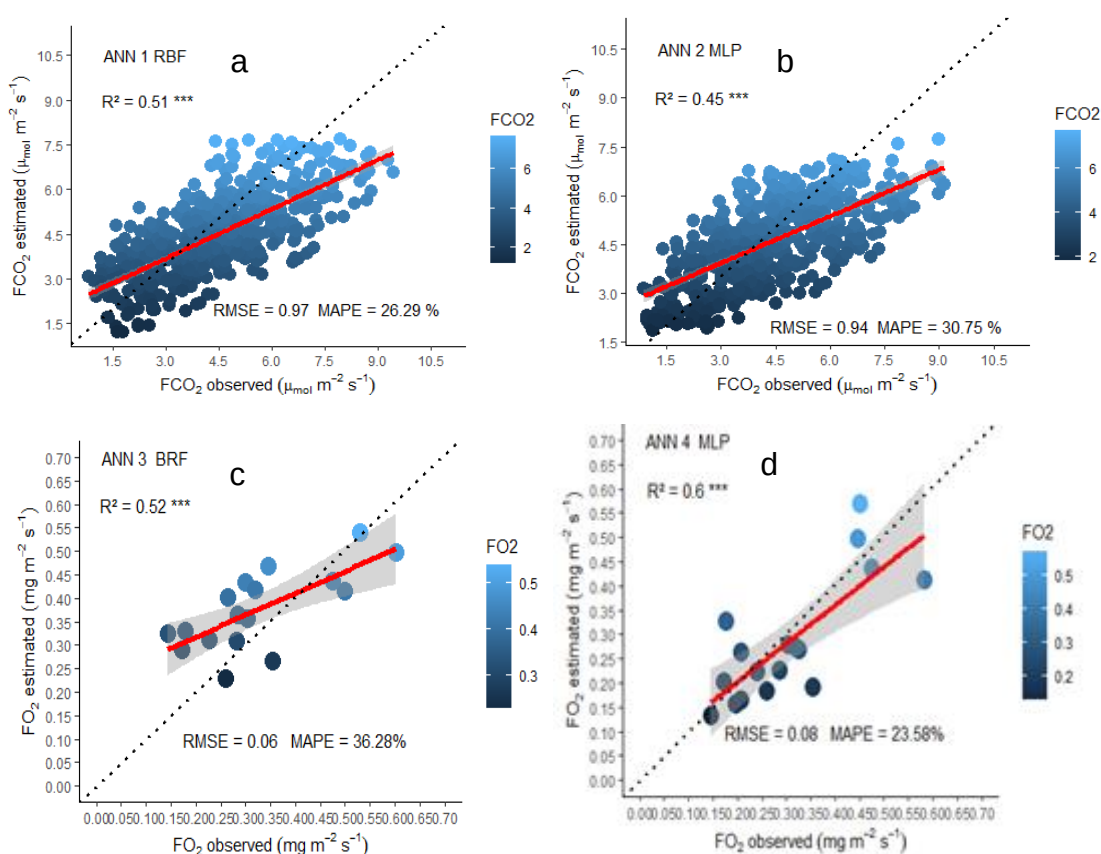


Figure 8: Validation performance plots of soil CO₂ emission (FCO₂) and Oxygen Uptake (FO₂). Coefficient of determination (R^2); mean absolute percentage error (MAPE); square root of the mean error (RMSE). a) ANN1: FCO₂ –RBF; b) ANN2: FCO₂ – MLP; c) ANN3: FO₂ - RBF; d) ANN4: FO₂ - MLP.

The major difference between RBFNN and MLPNN lies in the transfer function inside the hidden layer (sigmoid tangent in MLP, Gaussian in RBF). For this dataset,

the RBFNN structure showed advantages in the learning process resulting in a better generalization of the model. The sensitivity analysis investigated the most effective variables at the entrance of the ANNs. For the RBFNN, the analysis indicated: T_{rat} , T_s , RSG, Macro, EVI, SOM, PAR, pH, Micro, SB, m, P, Silt, P_{atm} .

As we can observe from the 18 input variables, 14 were necessary to predict the temporal variability of FCO_2 , emphasizing the complexity of modeling this variable. Temporal variability is the result of production by biochemical processes and transport to the atmosphere. The physical characteristics of soils thus influence this dynamic.

For FO_2 in the calibration, the ANN presented MEA indices with values ranging from 0.045 to 0.069 $mg\ m^{-2}\ s^{-1}$ and validation, 0.068 to 0.091 $mg\ m^{-2}\ s^{-1}$, respectively. indicating the "better" (i.e., vs. RBF) performance of the MLP (Table 6). The ANNs of FO_2 presented better correlation values than did those of FCO_2 with values of r ranging from 0.74 to 0.92 for calibration and 0.72 to 0.77 for validation. In the validation, the ANN presented similar low RMSE values (0.06 to 0.08 $mg\ m^{-2}\ s^{-1}$; Figure 8). MLPNN gave the best estimates for FO_2 , unlike the FCO_2 model. The adjusted R^{2adj} values indicated an precision of calibration and validation ($R^2 = 0.83$, $R^{2adj} = 0.57$, respectively) and low associated errors MAPE (23.58%) in validation, indicating that this network can adequately calculate the output for an input that is not present in the learning process

5.3.4 Support Vector Regression

The set of variables for the calibration of the FCO_2 model comprised the variables found to be most significant under the sensitivity analysis of ANN BRF training: T_{rat} , T_s , RSG, Macro, EVI, MOS, PAR, pH, Micro, SB, m, P, Silt, and P_{atm} , for a total of 14 variables.

However, to calibrate the FO_2 and improve the performance of the model, we need to add P_{atm} to the training data set by the trial error method. The performance criteria of the SVR model for the training and test datasets are listed in Table 4. The SVR- FCO_2 and SVR- FO_2 models were significant ($p < 0.01$). To improve the choice of appropriate values of hyperparameters, k -fold cross-validation was performed by subdividing the dataset into k random training subsets. The search grid shows the optimal values of the parameters for SVR.

In this study, the radial basis function (RBF) is used as the basic kernel function of SVR. Thus, for all SVR models with the RBF kernel function in this study, the parameters (C) had values of 1 and $\epsilon = 0.1$. For FCO₂, the optimal value of parameter was $\gamma = 0.25$, and for FO₂ $\gamma = 0.46$.

The SVR-RBF for FCO₂ in the calibration and validation phases presented similar values of the error indices: MEA (1.069 and 1.058 $\mu\text{mol m}^{-2} \text{s}^{-1}$), RMSE (1.072 to 1.109 $\mu\text{mol m}^{-2} \text{s}^{-1}$), and MAPE (13.40% to 25.45%; Table 7). Regarding precision, it showed correlation coefficients ($r = 0.72$ and $r = 0.71$) and coefficient of determination ($R^2_{\text{adj}} = 0.53$ and $R^2_{\text{adj}} = 0.50$), calibration and validation respectively.

SVR-RBF for FO₂ performed better in calibration. MEA values ranged from 0.053 to 0.078 $\text{mg m}^{-2} \text{s}^{-1}$ for calibration and validation, and the RMSE presented similar errors in both phases (0.05 and 0.06 $\text{mg m}^{-2} \text{s}^{-1}$). The MAPE values ranged from 21.23 to 23.57, indicating the good accuracy of the method. . The adjusted R^2_{adj} values indicated an precision of calibration and validation (0.76 and 0.48) respectively. Analyzing the performance of the two estimates, SVR-RBF for FCO₂ performed better than SVR-RBF for FO₂

Table 6. Performance indices for prediction of soil CO₂ emission and oxygen uptake

			Soil CO ₂ Emission (μmol m ⁻² s ⁻¹)					
Architecture			<i>r</i>	MEA	MAPE	<i>R</i> ² .adj	<i>F</i>	<i>p</i> -value
ANN1	RBFNN 14-210-1	Calibration	0.78	0.933	21.91	0.606	977	< 0.001
		Validation	0.72	1.021	26.29	0.512	481	< 0.001
ANN 2	MLPNN 11-10-4	Calibration	0.69	1.056	28.65	0.483	899.5	< 0.001
		Validation	0.67	1.086	30.75	0.449	371.3	< 0.001
Oxygen Uptake (mg m ⁻² s ⁻¹)								
ANN 3	RBFNN 14-90-1	Calibration	0.74	0.069	22.44	0.531	32.04	< 0.001
		Validation	0.72	0.091	36.28	0.483	15.93	< 0.001
ANN 4	MLPNN 14-13-7	Calibration	0.92	0.045	21.11	0.836	199.6	< 0.001
		Validation	0.77	0.068	23.58	0.572	22.37	< 0.001

r = Pearson correlation; RMSE = root mean square error; MAPE = mean absolute percentage error; MEA = mean absolute error, R^2 .adjusted

Table 7. SRV metrics in predicting soil CO₂ emission and oxygen uptake

		Soil CO ₂ Emission (μmol m ⁻² s ⁻¹)						
		<i>R</i>	RMSE	MAPE	MEA	<i>R</i> ² .adj	<i>F</i>	<i>p</i> -value
SVR_RBF	Calibration	0.730	1.072	13.40	1.069	0.53	1561	< 0.001
	Validation	0.710	1.109	25.45	1.058	0.50	473.5	< 0.001
		Oxygen Uptake (mg m ⁻² s ⁻¹)						
SVR_RBF	Calibration	0.807	0.05	21.23	0.053	0.76	134.8	< 0.001
	Validation	0.690	0.06	23.57	0.078	0.48	16.52	< 0.001

r = Pearson correlation; RMSE = root mean square error; MAPE = mean absolute percentage error; MEA = mean absolute error, R^2 .adjusted

5.3.5 Neuro-Diffuse Adaptive Inference System

From the set of 16 variables selected using PCA, we selected the best ones for training the models using a trial-error approach. However for the FO_2 estimate, we noticed that when we added wind velocity (w) to the training data set, the variable helped improved performance the FO_2 estimation performance. Although wind velocity was not part of the PCA, the variable was part of the meteorological data proposed for modeling FCO_2 and FO_2 . The FCO_2 dataset comprised T_s , SWC, EVI, and GSR, and the FO_2 dataset comprised ET_0 , w , GSR, T_s , SWC, and pH. The quality of adjustment of FCO_2 was very similar for the calibration and validation phases, as observed by the MAPE value of 30% (Figure 9 and Table 8).

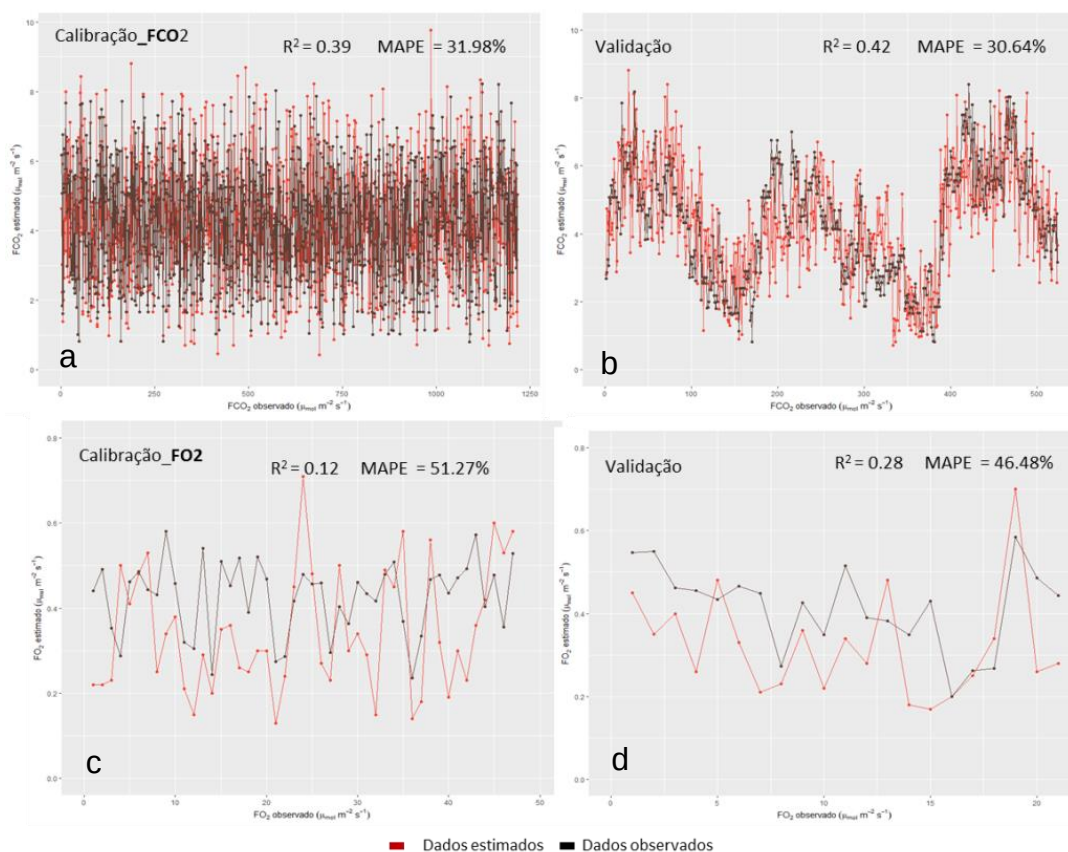


Figure 9. Performance of ANFIS metrics in predicting soil CO_2 emission and oxygen uptake. a) FCO_2 calibration, b) FCO_2 validation, c) FO_2 calibration, d) FO_2 validation.

The calibration yielded RMSE and MEA errors of 1.237 and 1.149 ($\mu\text{mol m}^{-2}\text{s}^{-1}$), respectively. The precision presented a coefficient of determination of 0.39%. The performance of the model in validation was better, with lower errors

associated with RMSE and MEA of 1.238 and 1.128 ($\mu\text{mol m}^{-2}\text{s}^{-1}$) and explaining 42% of the temporal variability in FCO_2 .

For FO_2 , the performance was lower than that of FCO_2 . In both phases, the MAPE values were 51.27 and 40.48%, respectively, indicating large errors associated with the estimate. The RMSE values for calibration and validation were 0.079 ($\text{mg m}^{-2}\text{s}^{-1}$) and 0.085 ($\text{mg m}^{-2}\text{s}^{-1}$), respectively, and those of MEA ranged from 0.139 ($\text{mg m}^{-2}\text{s}^{-1}$) to 0.125 ($\text{mg m}^{-2}\text{s}^{-1}$), with precision shown by $r = 0.53$ e $R^2 = 28\%$.

5.3.6 Random Forest

The RF- FCO_2 and RF- FO_2 models were significant ($p < 0.01$).—From the set of 16 variables selected using PCA, we chose the best set for training the models using the trial-error method. However, for the FCO_2 and FO_2 estimate, we noticed that when we added soil density to the training data set, the variable helped the estimation performance. —However, by trial and error method, the best results were obtained for FCO_2 for: Micro, Macro, BD, SB, CEC, P, T_s , P_{atm} , PAR, ETo, SOM, pH, M, and EVI; and for FO_2 : Micro, Bd, CEC, P, T_s , T_{atm} , RSG, P_{atm} , PAR, ETo, SOM, pH, m, and EVI. For both estimates, we added variables to the set of predictors that made important contributions to the model. The best results for FCO_2 were obtained when the number of tree estimators in the forest was equal to 300 and the maximum tree depth (DM) was 10. For FO_2 , the optimal number of tree estimators in the forest was 500 and the maximum tree depth (DM) was 9. The number of variables used in each tree (mtry) was one-third of the total number of samples, as suggested by Breiman et al. (1996).

The RF for FCO_2 in the calibration and validation phases presented values of the error indices: MEA (0.92 and 0.95 $\mu\text{mol m}^{-2}\text{s}^{-1}$), RMSE (1.04 to 1.27 $\mu\text{mol m}^{-2}\text{s}^{-1}$), and MAPE (30.61 to 33.42). The accuracy was R^2 0.80 and R^2 0.60 respectively. For FO_2 , the performance was lower than that of FCO_2 . The RMSE values for calibration and validation were 0.101 ($\text{mg m}^{-2}\text{s}^{-1}$) and 0.110 ($\text{mg m}^{-2}\text{s}^{-1}$), respectively, and those of MEA ranged from 0.097 ($\text{mg m}^{-2}\text{s}^{-1}$) to 0.095 ($\text{mg m}^{-2}\text{s}^{-1}$), with precision shown by R^2 0.57 and 0.30. The MAPE value in the calibration was 40.27 and validation 38.02, and $R^2 = 0.57$ calibration and $r = 0.54$ e $R^2 = 0.30$ validation respectively.

Table 8 . Performance of ANFIS metrics in predicting soil CO₂ emission and oxygen uptake

Soil CO ₂ Emission ($\mu\text{mol m}^{-2} \text{s}^{-1}$)							
Phase	<i>r</i>	RMSE	MAPE	MEA	<i>R</i> ² ad	F	p-value
Calibration	0.651	1.237	31.98	1.149	0.39	776.80	< 0.001
Validation	0.642	1.238	30.64	1.128	0.42	385.80	< 0.001
Oxygen Uptake ($\text{mg m}^{-2} \text{s}^{-1}$)							
Calibration	0.353	0.079	51.27	0.139	0.125	6.414	< 0.001
Validation	0.534	0.085	46.48	0.125	0.285	7.581	< 0.001

r = Pearson correlation; RMSE = root mean square error; MAPE = mean absolute percentage error; MEA = mean absolute error, *R*².adjusted

Table 9. Performance of random forest metrics in predicting soil CO₂ emission and oxygen uptake

Random Forest							
Soil CO ₂ Emission ($\mu\text{mol m}^{-2} \text{s}^{-1}$)							
	Phase	<i>r</i>	RMSE	MAPE	MAE	<i>R</i> ² .adj	p-value
RF1	Calibration	0.88	1.04	40.27	0.92	0.75	< 0.001
	Validation	0.80	1.27	38.02	0.95	0.61	< 0.001
Oxygen Uptake ($\text{mg m}^{-2} \text{s}^{-1}$)							
RF2	Calibration	0.77	0.10	40.27	0.1	0.57	< 0.001
	Validation	0.54	0.11	38.02	0.1	0.30	< 0.001

r = Pearson correlation; RMSE = root mean square error; MAPE = mean absolute percentage error; MEA = mean absolute error, *R*².adjusted

5 3.7 Comparison of Models

The selection of model input variables plays an important role in the development of predictive models. From the data set formed based on the results of the PCA, the models (ANN, SVR, and RF) used 14 predictor variables for the calibration of FCO₂ and 15 variables for FO₂ on average, indicating the complexity of the models. The ANFIS model yielded the best results, with four input variables for FCO₂ and six for FO₂.

According to the results of Table 10 (Supplementary Material), in general, in calibration and validation (RBFNN and RF) the estimation of FCO₂ showed the lowest RMSE and MEA errors and highest accuracy. It should be noted that RF better explains the temporal variability of FCO₂ ($R^2 = 60\%$) and has a low MEA value. However, the mean absolute error, whose values indicate high reliability between the observed and predicted values, only considers positive values, in that in RMSE errors (observed predictions) are squared before the mean is calculated. Thus, we penalized the presence of possible outliers in the data set (Chai and Draxler, 2014) and chose the lowest RMSE values for the model tiebreaker criteria. Thus, RBF is the best model for estimating the FCO₂.

For FO₂, the MLPNN and SVR-RBF models were the most satisfactory in the calibration and validation phases. The MLPNN stood out because of the greater accuracy of the estimation of the temporal variability of FO₂ ($R^2 = 0.60$) and the lower RMSE and MAE errors associated with this estimate.

ANFIS did not show good generalizability and presented the worst performance, showing the highest errors and lowest accuracy ($R^2 = 0.12$, $R^2 = 0.28$, respectively) in calibration and validation (MAPE = 51.27; 46.48%); however, high percentage errors were associated with the model. Based on the comparison of the MEA, RMSE, and Adjusted R^2 indices (Table 10), this study found RBFNN to be the best predictive model for the estimation of FCO₂ and the MLPNN model the best for estimating FO₂.

Table 10. Statistical summary of machine learning techniques

Soil CO₂ emission ($\mu\text{mol m}^{-2} \text{s}^{-1}$)				
		MEA	RMSE	R^2
RBFNN 14-210-1	Calibration	0.933	0.901	0.60
	Validation	1.021	0.971	0.50
SVR1	Calibration	1.069	1.072	0.50
	Validation	1.058	1.109	0.50
ANFIS	Calibration	1.149	1.237	0.39
	Validation	1.128	1.238	0.42
RF	Calibration	0.92	1.040	0.80
	Validation	0.95	1.271	0.60
Oxygen Uptake ($\text{mg m}^{-2} \text{s}^{-1}$)				
MLPNN14-13-7	Calibration	0.045	0.059	0.75
	Validation	0.068	0.080	0.60
SVR2	Calibration	0.053	0.050	0.80
	Validation	0.078	0.06	0.50
ANFIS	Calibration	0.139	0.079	0.12
	Validation	0.125	0.085	0.28
RF2	Calibration	0.097	0.101	0.57
	Validation	0.095	0.110	0.30

r = Pearson correlation, RMSE = root mean square error, MEA = mean absolute error, R^2 .adjusted

5. 3.8. Discussion

5.3.8.1 Principal Component Analysis

The change in land use, especially due to the substitution of native vegetation by monocultures results in soil degradation, impacting its chemical, physical, and biological quality (Adachi et al. 2006; Young et al. 2010), and the microbial diversity and composition (Ding et al., 2021). This compromises the dynamics of soil organic carbon, especially its flow and stock, which is directly related to the type of management adopted and the local environmental conditions (Islam and Weil, 2000; Pulrolnik et al., 2009);

In our study, the results of PCA indicated that eucalyptus forests may be associated with various production factors and CO₂ emission, represented in the biplot by Group I. Soil organic matter is the main source of energy for microorganisms (Lal, 2009), and the variable showed a direct association with FCO₂ and the highest contribution (0.97) towards CP1, followed by the sum of bases (0.92).

The chemical structure of eucalyptus litter may be attributed to its greater lability due to its lower C/N ratio relative to the litter of the pine forest (Vicentini et al., 2019), and this may facilitate the availability of organic C for microorganisms and promote nutrient cycling, which is a fundamental factor for maintaining soil fertility. The benefits of the presence of organic matter include changes in soil organic carbon, soil water retention capacity, aggregate formation, pH, cation exchange capacity, base saturation (e.g. (Gruba and Mulder, 2015), reflect the improvement in soil physical quality as fertilization (Xuebin et al. 2019)

Therefore, despite the apparent density value of the soil in the eucalyptus forest, which is above ideal (1.47 g dm⁻³) for clay soils (Arshad et al. 1996), the highest percentage of mean macropore values (0.08 m³ m⁻³) combined with lower microporosity contents (0.31 m³ m⁻³) was favorable for gas exchange between the soil and atmosphere. This may contribute to the influx of oxygen for consumption by microorganisms and result in higher CO₂ emissions (Table 3). These factors may influence soil microbial activity and the production and emission of CO₂, as was observed in the study by Ramos et al. (2018) (3.96 µmol m⁻² s⁻¹) and (Vicentini et al., 2019) (5.61 µmol m⁻² s⁻¹) (Table 3).

It was interesting to observe the contribution of the silt fraction towards PC1 (0.74). This may be related to soil microbial interactions, texture, and soil mineralogy.

Other researchers have identified a relationship between FCO_2 , soil mineralogy (La Scala et al., 2000), and silt (Moitinho et al., 2021).

In general, the formation of the subgroups IIa (native species) and IIb (pines) indicated that the forest type influenced the physical properties of the soil. The high apparent density value of pine soil indicates the occurrence of soil compaction (1.57 g dm^{-3}) which resulted in a decrease in macropore volume ($0.042 \text{ m}^3 \text{ m}^{-3}$) and was partially compensated by an increase in microporosity ($0.368 \text{ m}^3 \text{ m}^{-3}$) (Table 3). This impaired the infiltration of water into the soil, compromising gas exchange with the atmosphere and consequently oxygen availability. According to (Neira et al., 2015) gases, including oxygen, move in the soil according to diffusion laws ; larger pores generally allow more rapid movement of gases than smaller pores.

Soil with pine trees has low organic matter content. This results due to the specific characteristics of the litter, which consists of aciculae with high structural complexity (lignin and polyphenols), resulting in greater resistance to microbial degradation (Yuan et al., 2019). These factors may contribute to the lower organic C content in soil and cause soil acidification, as indicated by the low pH value (Table 4). This may affect the dynamics of the heterotrophic microbial population and impact the production of CO_2 . (Oertel et al., 2016), resulting in lower CO_2 emissions over time as observed in the data (Table 3).

In this study, climatic parameters were not directly associated with the formation of groups but had a high contribution towards CP2. In forest ecosystems, canopy structure influences the energy balance, relative humidity, wind, and barometric pressure (Sotta et al. 2004), which results in the formation of different microclimates, and in combination with the quantity and quality of litter control the speed of decomposition of plant residues (Yuan et al., 2019). Despite the low contribution of FCO_2 towards CP1, the factor was inversely associated with FO_2 , soil temperature, soil water content, and other attributes retained in CP2. Several studies in different ecosystems have identified soil temperature and content as regulators of the temporal variability of FCO_2 (Luo and Zhou, 2006; Panosso et al., 2012); Silva et al., 2019).

We hypothesized that because soil temperature was relatively constant as shown by our results, and as this is expected for tropical forests (Davidson et al., 2000), it may not be a limiting factor for biological activity.

Regarding water content, despite the differences between the three managements Eucalyptus, Pinus, native species in 2019, the higher soil water content observed in the forest with native species (Table 1, **Supplementary**) may probably result due to the higher clay content of this soil (Table 4). In 2018, this difference was not observed between eucalyptus and pine forests. The study was performed in the adjacent areas, and the mean rainfall was within the expected range for the region during the study period (Ramos, 2018; (Vicentini et al., 2019)

Evapotranspiration and soil water content are important components of the hydrological cycle. The evapotranspiration estimates depend on the expressed meteorological variables alone. the maximum evapotranspiration of the plants and objective to evaluate different evapotranspiration demands. Balogh et al (2015) identified strong negative correlations between evapotranspiration at CO₂ emission rates and Grossiord et al. (2012) have shown that this may relate to reduced root respiration. Given the results of our study, we hypothesize that the contribution of global radiation, photosynthetically active radiation, EVI, and oxygen consumption are indicators for the respiratory activity of roots. Evapotranspiration may serve as a factor regulating the intensity of this respiratory activity.

5.3.8.2 Development and comparison of models

Modeling the temporal variability of soil respiration is a major challenge because of the complex interactions between the plants, atmosphere, and soil system, Using four models of machine learning (neural network, vector machine support, neuro-fuzzy inference system, and random forest), we proposed to model the temporal variability of soil CO₂ emission (FCO₂) and oxygen consumption (FO₂) in forests planted in the Cerrado

In contrast with a previous study describing the dynamics of the temporal variability of soil respiration as being governed by temperature and water content in the soil (Panosso et al. 2009), the results obtained using our dataset indicated that the chemical and physical attributes of the soil and the EVI influence these dynamics remarkably.

As previously discussed in the manuscript, organic matter is an indicator of C dynamics. In general, the organic matter content in the studied ecosystems is low; consequently, soils have low fertility with high acidity and compaction. The

degradation of litter is one of the main mechanisms responsible for nutrient cycling in forest systems and the main source of organic matter. Depending on the quantity and quality of the residues, expressed by the concentration of carbon in relation to nitrogen (C/N ratio), the lignin and polyphenol contents hinder the process of mineralization of organic matter and nutrient availability. We also believe that part of the organic matter present in the soil may be in more stable forms, resistant to microbial action. Further, the forest stand age influences both carbon stocks and CO₂ fluxes (Bradford et al. 2008). The long reforestation period of over 30 years might have contributed to the storage of carbon; for this reason, the heterotrophic biological activity, which occurs in these soils, would be only plant residues with labile fractions. Therefore, we believe that the temporal variability in soil respiration is predominantly autotrophic.

The ANFIS model required great computational effort, and it was necessary to work with several reduced variables. Thus, when selecting variables for the training dataset in all combinations tested, organic matter was not a good predictor variable for estimating FCO₂ and FO₂. However, EVI presented a satisfactory performance in the estimation of FCO₂. Regarding the ANNs, the best RBFNN architecture selected by the IPS to estimate FCO₂ consisted of a combination of 14 vectors in the input layer and a hidden layer with 210 neurons (Table 6). The greater number of neurons in each layer contributes to the ability of the network to solve more complex problems. This also occurs when the number of hidden layers increases. Thus, ANN improves efficiency over the nonlinear dataset of the input layer (Scapino et al., 2019). By analyzing sensitivity, the results of the model revealed what type of management was the dominant factor.

In contrast to our results, (Freitas et al., 2018) estimated the spatial variability of FCO₂ in a semi-perennial crop using MLPNN (7-18-1) with backpropagation algorithm; this method was highly effective, explaining R² 80% of the spatial variability. According to (Song et al., 2014), MLPNN with a three-layer architecture trained with the backpropagation algorithm was also effective in modeling the temporal variability of FCO₂ in Chinese forests. The model explained 74% of the annual variation in soil respiration.

When estimating FO₂, MLPNN provided the best results for BRFNN. By analyzing sensitivity, we determined that soil temperature stimulates root activities, resulting in oxygen consumption. Macroporosity is essential for gas exchange

between the soil and the atmosphere (Tavares et al. 2015). The movement of gases in the soil is a complex process. Soil texture, structure, density, and water content are the main factors that influence soil aeration and consequently, gas exchange between the soil and atmosphere. The intake of oxygen into the ground is the result of a pressure gradient of O_2 between the atmospheric air and ground air. According to Fick's law, a dynamic balance is established by spreading O_2 in which the consumption rate of O_2 in the soil is equal to the flow of O_2 on its surface (Jong van Lier, 2001)..

Interestingly, in the training of regression by support vectors when estimating FO_2 , the addition of pressure as an input variable contributes to performance. Many studies on the use of SVM in soil science have demonstrated improved performance using RBF kernel-based SVR (Shi et al., 2013; Jafarzadeh et al., 2016) SVR is the best model for predicting carbon stocks in forest reserve soils, when compared to random forest and ANN (Were et al. 2015).

Random forest achieved a good fit for the FCO_2 model. According to Grimm et al. (2008), random forest has several advantages compared to other modeling approaches, such as the fact that it is not sensitive to excessive adjustments or noise resources. Decision tree methods consider the relationships and interactions between variables. The tree structure was obtained based on these relationships and interactions. The technique is efficient in modeling the treatment of large numbers of variables with small numbers of observations (Grömping, 2009); however, we did not observe this advantage in the estimation of FO_2 .

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. FCO₂ modeling provides more predictive models than FO₂.

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. **Advances 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) **Defining 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, Shamsirband 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](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) Adaptive-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 CO₂ 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](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](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 CO₂ 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 access 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** ;94:546-551.<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 systems and their applicability in the determination of agroclimatic zoning 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. **Applied 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/s1104-019-04262-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](https://doi.org/10.1016/S0038-0717(02)00274-2)

Wang, Y.; Witten, I. H (1996). Induction of model trees for predicting continuous classes. Working Paper 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>

5.11 Attachment: Supplementary Material

Table 1. Composição da área de espécies nativas do Cerrado

Nome Popular	Nome Científico
Albizia	<i>Albizia lebeck</i> L. Benth
Alecrim-do-campo	<i>Holocalyx balansae</i> Micheli
Amora ou amora-preta	<i>Morus nigra</i> L
Angico-roxo	<i>Anadenanthera peregrina</i> var. <i>falcata</i>
Aroeira	<i>Myracrodruon urundeuva</i> Fr. All
Bálsamo	<i>Myroxylon peruiferum</i> L.f
Café-de-bugre	<i>Cordia ecalyculata</i> Vell
Cajá-mirim	<i>Spondias lutea</i> L.
Canafístula	<i>Peltophorum dubium</i> (Sprengel) Taubert
Canudo-de-pito	<i>Mabea fistulifera</i> Mart
Capitão-do-campo	<i>Terminalia argentea</i> Mart. et. Succ
Figueira	<i>Ficus guaranitica</i> Schodat
Goiaba	<i>Psidium guajava</i> L
Guaruaia	<i>Parapiptadenia rigida</i> (Benth.) Brenan
Ingá	<i>Inga</i> sp
Ipê-amarelo	<i>Tabebuia</i> sp
Ipê-branco	<i>Tabebuia roseo-alba</i> (Ridl.) Sand
Ipê-rosa	<i>Tabebuia heptaphylla</i> Loren. ex. Griseb
Ipê-roxo	<i>Tabebuia impetiginosa</i> (Mart.) Standl.
Jaboticaba	<i>Plinia trunciflora</i> Berg
Jacarandá-caroba	<i>Jacaranda cuspidifolia</i> Mart
Jacarandá-do-campo	<i>Machaerium acutifolium</i> Vog
Jambolão	<i>Syzygium cumini</i> (L.) Skeels
Jenipapo	<i>Genipa americana</i> L
Leucena	<i>Leucana leucocephala</i> (Lm.) R. de Wit
Magnólia-amarela	<i>Michelia champaca</i> L.
Marmelo ou marmelada	<i>Alibertia edulis</i> Rich. (Rubiaceae)
Nespereira	<i>Eriobotrya japonica</i> Lindl
Oiti	<i>Licania tomentosa</i> (Benth.) Fritsch
Olho-de-dragão	<i>Adenanthera pavonina</i> L
Paineira	<i>Chorisia speciosa</i> St. Hil
Quereutéria	<i>Koeleruteria paniculata</i> Laxm
Santa Bárbara	<i>Melia azedarach</i> L
Tamboril	<i>Enterelobium contortisiliquum</i> (Vell.) Morong
Uva japonesa	<i>Hovenia dulcis</i> Thunb

CHAPTER 6 - FINAL REMARKS

Soil respiration is a complex phenomenon comprising of nonlinear and interdependent relationships. Using machine learning techniques, such as neural network, support vector regression, adaptive neuro-fuzzy inference system, and random forest, we estimated the FCO_2 and FO_2 of three ecosystems of planted forests: pine, eucalyptus, and native species. As predictor variables, we used meteorological data, EVI, and chemical and physical soil attributes.

First, we evaluated each forest management individually then created a database for the generation of a global model. We chose principal component analysis as the data mining technique. Although this reduced the size of the data set in general, it did not rule out the selection of variables by trial and error. Therefore, we use a large number of predictor variables in all estimates. This showed the complexity of the data.

Only the ANFIS model presented a reduced data set, however presented higher computational expense. Regression methods by machine learning are efficient models to estimate the FCO_2 ; however, the selection of the training data set is paramount. Thus, for future studies adopting new data mining techniques and association of different analyses can increase the performance of the models. Regarding FO_2 , the small number of observations of dependent variable might not have been enough for the models to learn from examples of the desired inputs and outputs or to extract relevant information from the data set. Overall, neural networks and random forest provided the most predictive models for estimating and to explain the temporal variability patterns of FCO_2 and FO_2 in planted forest areas using environmental covariates (soil and meteorological). Development of new studies, especially in reforested areas, would be beneficial, as there are many uncertainties about the dynamics of soil respiration in these ecosystems. In addition, it is important to note that studies on the temporal dynamics of oxygen influx and estimation are lacking; thus, all studies are important in order to help understand the temporal dynamics of soil respiration.