Revista Brasileira de Cartografia
ISSN 1808-0936 | DOI. 10.14393/revbrascartogr
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto

Spatial Modeling of Groundwater Potential in the North of Minas Gerais, Brazil: An Integrated Approach Using Machine Learning and Environmental Data

Modelagem Espacial do Potencial Hídrico Subterrâneo no Norte de Minas Gerais, Brasil: Uma Abordagem Integrada Usando Machine Learning e Dados Ambientais

Cristiano Marcelo Pereira Souza[1], Lucas Augusto Pereira da Silva [2], Marcos Esdras Leite[3], Gustavo Vieira Veloso[4] e Elpídio Inácio Fernandes Filho[5]

1 Universidade Estadual de Montes Claros, Montes Claros, Brasil. souzacristianomp@gmail.com
ORCID: http://orcid.org/0000-0001-7692-1613
2 Universidade Federal de Uberlândia, Uberlândia, Brasil. lucaskaio1605@gmail.com
ORCID: http://orcid.org/0000-0001-5504-9029
3 Universidade Estadual de Montes Claros, Montes Claros, Brasil. marcos.leite@unimontes.br
ORCID: http://orcid.org/0000-0002-9020-6445
4 Universidade Federal de Viçosa, Viçosa, Brasil. gustavo.v.veloso@gmail.com
ORCID: http://orcid.org/0000-0002-9451-2714
5 Universidade Federal de Viçosa, Viçosa, Brasil. elpidio@ufv.br
ORCID: http:// orcid.org/0000-0002-9484-1411

Recebido: 02.2022 | Aceito: 09.2023

Abstract: In arid and semi-arid regions, like the North of Minas Gerais (NMG) in Brazil, groundwater serves as a crucial resource. Due to the anticipated surge in demand for these resources, devising effective strategies for managing and analyzing water resources is vital. This study aims to model the spatial distribution of potential groundwater areas in the NMG by evaluating six Machine Learning Algorithms based on water flow data from 4,028 tubular wells (Groundwater Information System - SIAGAS). The modeling was supported by environmental covariates connected with water dynamics (climate, geology, relief, soil, and vegetation). The covariate selection technique (RFE-Recursive Feature Elimination) selected the most important ones. The Random Forest (RF) model was the most efficient in the prediction (R^2 0.16 and an RMSE of 17.50 m^3/h). The model captured the influence of critical environmental covariates. The central and western regions of the NMG exhibited the highest groundwater potential, with flow values from tubular wells in these areas 620% higher than the eastern regions. This disparity can be attributed to the significant presence of psamitic, and carbonate sedimentary rocks characterized by high porosity and fissures, extensive plateaus (recharge zones), and higher rainfall levels observed in the central and western regions. The mapping results can serve as a valuable tool for public management, especially to define areas suitable for groundwater use in the NMG. We encourage future studies for advances and improvements in groundwater modeling processes in the region.
Keywords: Water resource. Hydrogeology. Minas Gerais semiarid. Spatial modeling. Random forest.

Resumo: Em regiões áridas e semiáridas, como o Norte de Minas Gerais (NMG) no Brasil, as águas subterrâneas servem como um recurso crucial. Devido ao aumento previsto na procura destes recursos, é vital conceber estratégias eficazes para a gestão e análise dos recursos hídricos. O objetivo desse estudo foi realizar predição espacial de áreas potenciais de água subterrânea no NMG empregando seis Algoritmos de Aprendizado de Máquina baseados em dados de 4.028 poços tubulares. A modelagem foi auxiliada por covariáveis ambientais que possuem conexão com dinâmica da água (clima, geologia, relevo, solo e vegetação). A técnica de seleção de covariáveis (RFE- Recursive Feature Elimination) selecionou as mais importantes. O modelo Random Forest (RF) foi o mais eficiente na predição (R^2 0,16 e um RMSE de 17,50 m^3/h), o modelo capturou a influência de importantes covariáveis ambientais. Espacialmente, as regiões central e oeste do NMG possuem maior potencial de água subterrânea, com valores de fluxo de poços tubulares nessas áreas 620% maiores em comparação com as regiões situadas a leste. As variações possuem conexão ao predomínio de rochas sedimentares psamíticas e carbonáticas com alta porosidade e fissuras, extensos planaltos facilitando a recarga e maior pluviosidade nas regiões centro e oeste. Os resultados do mapeamento realizado podem servir como uma ferramenta para a gestão pública, definindo áreas aptas para uso das águas subterrâneas no NMG. Incentivamos estudos futuros para possíveis avanços e melhorias nos processos de modelagem de águas subterrâneas na região.
Palavras-chave: Recurso hídrico. Hidrogeologia. Semiárido mineiro. Modelagem espacial. Random forest.

1 INTRODUCTION

The sustainability of water resources will become an increasingly crucial point for development in various regions worldwide, including those with vast water reserves. Brazil has the largest hydrological reserve in the world (SHIKLOMANOV, 2000), but it suffers from an uneven distribution of water resources across its territory. Approximately 70% of water reserves are concentrated in the Amazon basin, while the Southeast region, more populated (48.58 habitants/km²), has only 6%. Historically, due to high population growth rates, the Southeast region has experienced an expansion of deforestation in the Atlantic Forest and Cerrado biomes (KLINK; MACHADO, 2005; CERQUEIRA et al., 2019; PAIVA et al., 2020). The maintenance of this situation implies a drastic reduction in water availability from 2050 onwards, with levels below the recommended (UNEPE, 2002; GHISI, 2006). The ongoing challenge is developing water resource analysis and management strategies to enhance water use efficiency. The challenges are even more remarkable in regions with low rainfall, where the water crisis is more severe (MARANGON et al., 2020; YAO et al., 2020; IPCC, 2022).

In Southeast Brazil, there are occasional records of droughts, as seen in 2015, which was declared the "worst drought recorded in southeastern Brazil" (MEDEIROS et al., 2020). The current situation remains challenging, with 130 municipalities in Minas Gerais facing a severe water crisis in 2023 (INMET, 2023). Besides, some parts of Southeast Brazil are semi-arid, such as Northern Minas Gerais (NMG) (NOBRE et al., 2016; EMPINOTTI et al., 2019). Water scarcity drives groundwater demand in semi-arid zones (TAYLOR et al., 2013). This practice is widespread in NMG, where in some areas, groundwater serves as the primary water source (CÂNDIDO et al., 2019). However, overexploitation of groundwater can result in depletion, and in regions with low rainfall, this problem is more dramatic (GONÇALVES et al., 2020; TENENWURCEL et al., 2020; NANEKELY et al., 2023). The definition of potential areas for groundwater exploration emerges as a fundamental fact (DAS et al., 2017).

Identifying potential areas for groundwater resources at the regional level requires an integrated analysis of various conditioning factors (DAS et al., 2017; TENENWURCEL et al., 2020). The dynamics of infiltration and evapotranspiration are influenced by the interaction of climatic, geological, pedological, topographic, and phytophysiognomic factors (SOPHOCLEOUS, 2002). Complex models are necessary to adequately capture this heterogeneity of factors (ZHOU; LI, 2011; ASHER et al., 2015). These models establish potential groundwater contexts by analyzing environmental aspects, such as climatic geology, soil, and vegetation, and using flow simulations, hydraulic conductivity, and Geographic Information Systems as a tool. However, some models require extensive sample collection, data processing, time-consuming processes, and expensive equipment for spatial prediction.

In spatial prediction, Machine Learning Algorithms (ML) have gained prominence in studies of various fields of science. Predictive geospatial studies with ML are promising due to the increase in software/hardware capacity and, at the same time, greater availability of data serving as explanatory covariates of a given phenomenon (BERGEN et al., 2019; PADARIAN et al., 2020). Therefore, ML can generate computationally robust simulations supported by geo-datasets of various environmental factors, including studies in the spatial modeling of variables linked to groundwater dynamics (HUSSEIN et al., 2020; SAHOUR et al., 2020; MOSAVI et al., 2021; SINGHA et al., 2021).

The NMG possesses accessible data concerning the flow of tubular wells, which are part of the SIAGAS database (Groundwater Information System) (SIAGAS, 2023). However, these data have not yet been utilized for spatial prediction through Machine Learning (ML) and environmental covariates for prediction purposes. This methodology has the potential to identify areas with significant groundwater extraction potential. Noteworthy studies employing exploratory data analysis have demonstrated this interpretive trend (CÂNDIDO et al., 2019). Thus, the main objectives of this study were (i) to identify areas with high groundwater potential based on variations in the flow values of tubular wells using predictive ML models; (ii) to define the most effective ML model and the most important covariates for predicting the outflow values of tubular wells; and (iii) investigate more explanatory environmental aspects for the distribution of values in the predicted map.

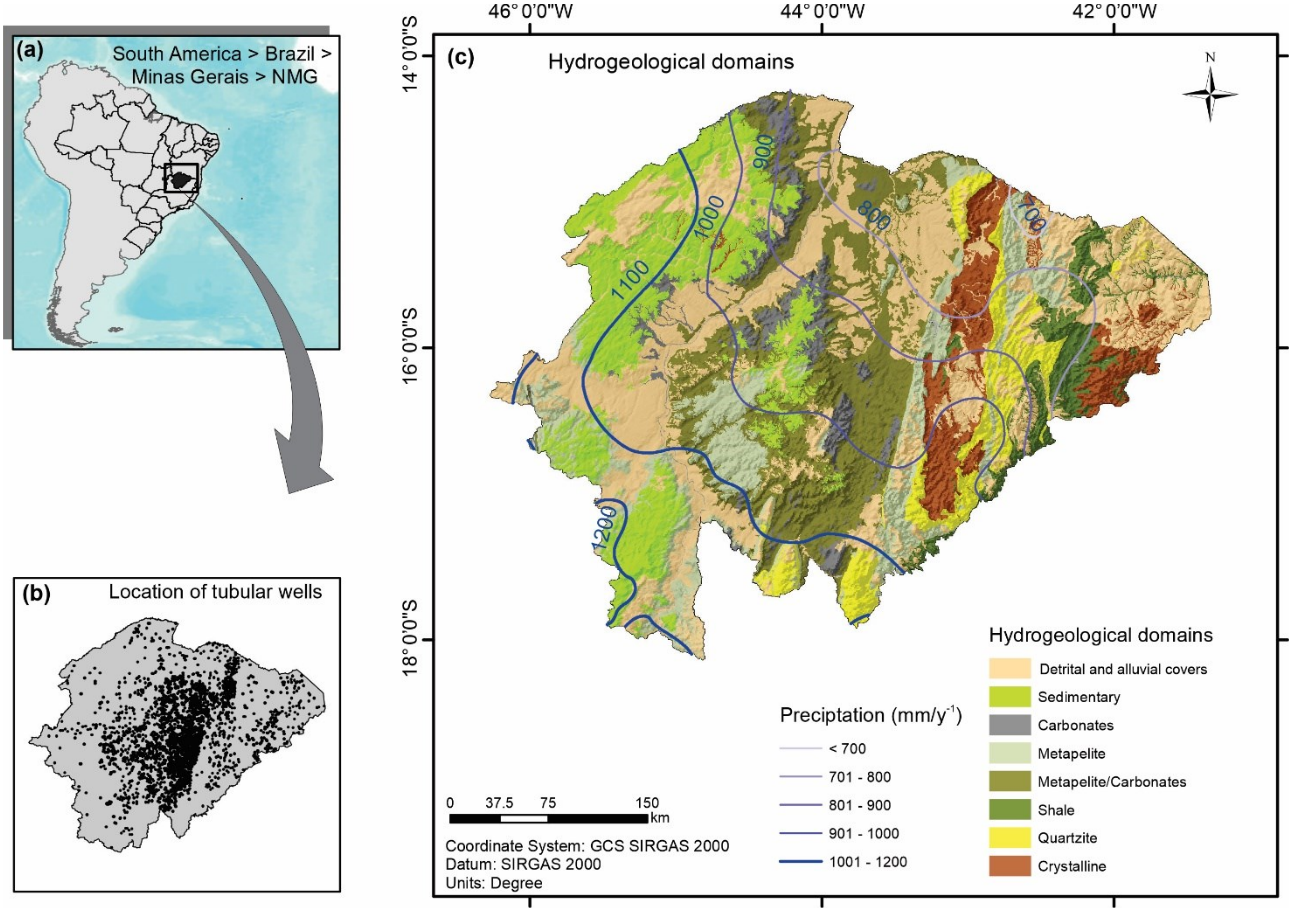
2 MATERIAL AND METHODS

2.1 Overview of the study area

The North of Minas is a region located in the Southeast of Brazil, between the coordinates of -41º33' to -46º28' W and -14º24' to -18º33' S. It encompasses 89 municipalities and has a population of 1,722,156 inhabitants according to the 2010 survey (IBGE, 2010). Montes Claros is the largest city in the region the sixth largest city in the state, with a population density of 13.4 inhabitants/km² (IBGE, 2010). Two municipalities were not considered in the study area, due to the lack of some data (Várzea da Palma and part of Pirapora).

The region includes three major hydrographic basins: São Francisco, Pardo, and Jequitinhonha. The prevailing climate in the area is classified as Aw, according to the Köppen system, and is characterized by an irregular distribution of precipitation. The region experiences eight dry and four rainy months, with rainfall ranging from 600 to 1200 mm year^{-1}. Rainfall decreases from southwest to northeast (Figure 1).

Figure 1 – (a) Location of the NMG in South America, Brazil, and the state of Minas Gerais. (b) Spatial distribution of tubular well points in the NMG. (c) Hydrogeological domains and precipitation ranges.



Elaboration: The authors, adapted PINTO et al. (2003).

The region is characterized by a predominance of sedimentary or metasedimentary rocks, occasionally interspersed with crystalline rocks. The relationships between geological domains and water dynamics are summarized in the following characteristics: the Detrital and Alluvial Coverage domain is formed by genetically heterogeneous systems; however, they have similar hydrogeological infiltration characteristics. The sedimentary domain has granular psammite rocks, and groundwater circulation occurs in spaces conditioned by unconsolidated sediments. The Carbonates domain comprises limestone and dolomite rocks, which have a hydrological potential directly related to dissolution processes (PESSOA et al., 2020).

The Metapelite domain comprises the fractured metapelites associated with the metasilitic facies, presenting secondary permeability. The Metapelite/Carbonate domain offers faciological variations, with intercalations and interdigitations between carbonate and pelitic lithotypes (ATMAN et al., 2011). In addition, they are predominantly fractured and have discontinuities of karstification zones.

Shale domains are fractured and discontinuous systems, with the primary source of recharge from the pluvial input and watercourses that intercept the quartz levels. The Quartzite domain is formed by lithofacies of the São Francisco Supergroup, occurring at altitudes ranging from 1200 to 1400 m (Espinhaço Mountain Range), with a structure marked by folds, faults, and fractures. Finally, the Crystalline domain comprises weathered mantle compartments above fractured crystalline rocks.

2.2 Database, models and procedures

The methodology employed in this study involved using Machine Learning Algorithms (ML) to spatially predict water flow from tubular wells and identify areas with high groundwater potential based on variations in these values. We evaluated six different ML with varying statistical analysis capabilities for groundwater potential spatial modeling. Linear Regression Model (LM), Inverse Distance Weighting (IDW), Multivariable Adaptive Regression Splines (EARTH - MARS), Support Vector Machine (SVM), Cubist (CB), and Random Forest (RF) models.

The input variable used in modeling groundwater potential was flow data from tubular wells (measured in m^3/h) collected from 4028 wells in the NMG region (SIAGAS, 2023). The tubular wells have flow values ranging from 0 to 100 m^3/h and 92% of the samples are below 40 m^3/h. This data consists of the dependent variable, which will be predicted using ML models and with dataset support of environmental covariates (independent variables). The entire methodological structure was implemented in the R software environment (RCORE, 2023).

Among the models considered, the LM model is based on a straightforward statistical principle, a desirable quality in machine learning studies to meet the parsimony principle (PADARIAN et al., 2020). The IDW model is a widely used method of spatial interpolation modeling, which estimates unsampled cell values by a linear weighted combination of the dataset and sample points (NISTOR et al., 2020). The EARTH-MARS algorithm is an advanced nonlinear regression technique that involves two steps in the analysis, the forward, which comprises the creation of an over-fitted model with various knots with candidate features, and the backward, applying a pruning technique to remove redundant knots (MILBORROW et al., 2014). The SVM aims to find a hyperplane in n-dimensional space, that is, the number of features that comprise the patterns from the database (CORTES; VAPNIK, 1995). CB is based on the logic of decision tree rules, where iterative regression trees are created to adjust previous errors of weak trees (QUILAN, 2010). Finally, RF is based on creating uncorrelated decision trees from different datasets, and the final prediction is given by averaging all trees (BREIMAN, 2001).

We fit machine learning models from an environmental covariates (independent variables) dataset to understand the spatial pattern of flow values from tubular wells. These covariates were used for model training, associating the covariate data and the flow data from tubular wells (i.e., dependent variable). We created a dataset of 177 covariates encompassing climatic, geological, topographic, pedological, and vegetation contexts. The database and modeling results available at <https://zenodo.org/records/10008537>. These covariates are routinely used in modeling studies, including spatial modeling of groundwater (RAHMATI et al., 2016; GOMES et al., 2019; SENA et al., 2020; MOSAVI et al., 2021; SOUZA et al., 2022).

The covariates were selected based on their explanatory power in recharge dynamics and water dynamics, which means that they contain information about runoff, infiltration, evapotranspiration, or processes generating recharge in surface or shallow soil or geology. The climate covariates comprised 103 covariates from the WorldClim project, including monthly temperature, precipitation, bioclimatic data, and mean and annual accumulated rates of these covariates (FICK; HIJMANS, 2017).

We used the Digital Elevation Model – SRTM (Shuttle Radar Topography Mission) to automatically extract 36 topographic covariates using SAGA-GIS software (OLAYA; CONRAD, 2009; SENA et al., 2020). Because the attributes of the terrain greatly influence runoff and infiltration rates, aspects directly linked to the maintenance of groundwater (RAHMATI et al., 2016; MOSAVI et al., 2021).

Other covariates were also prioritized, totaling thirty-eight raster data: Gamma-spectrometry data (K, Th, U) were entered, representing levels of K, Th, and U elements and their combination, totaling eight covariates. Gamma-spectrometry data are essential information and are connected to some soil and rock properties (PINTO et al., 2003). Furthermore, categorical maps of geology, soil, climate, vegetation and geomorphology also participated as covariates (PINTO et al., 2003; CPRM, 2007; UFV et al., 2010) – (12 covariates). Studies show a relatively more significant contribution to the prediction of groundwater potential when inserted into categorical datasets (RAHMATI et al., 2016; MOSAVI et al., 2021).

Covariates linked to land cover were inserted, in this case, vegetation indices for winter and summer periods (average) – (Landsat Soil Adjusted Vegetation Index-SAVI and Normalized Difference Vegetation Index-NDVI) (USGS, 2023) – (four covariates). This temporal frame better captures the behavior of vegetation in different periods, given that vegetation changes behavior in different seasons of the year. Data of covariates related to groundwater, erosivity and basin indices for the NMG were also inserted, totaling fourteen covariates (CÂNDIDO et al., 2019; SOUZA et al., 2022). Finally, X and Y coordinate grids were inserted to capture some influence of the spatial dependence of the data.

In the end, the values of the covariates for each tubular well sample were extracted to form the data matrix (dependent variable and independent variables). In this matrix, we apply a criterion to select the most relevant covariates for predicting the flow values of tubular wells. This step is essential, as the presence of redundant data can create problems in the modeling processes, such as increasing training time and overfitting problems (COELHO et al., 2019; GOMES et al., 2019; SOUZA et al., 2022). Therefore, from the findcorrelation function, we removed from the database pairs of covariates with high correlation (Pearson > 95%) (KUHN et al., 2020).

The newly generated data matrix associates the variable (tubular well points) and low correlation covariate values (Pearson < 95%). The next steps included training processes with 75% samples and testing (25%). The training phase uses known data to teach the model to make predictions. In this phase, we apply the Recursive Feature Elimination – RFE to train the models and simultaneously select the most relevant set of covariates for improved predictions. This is a crucial step, as it reduces the size of the database by considering the adjustment of the hyperparameters of the evaluated models. Initially, the models were trained with all covariates. Then, following a regressive elimination logic, less critical features were removed successively until we obtained the smallest number of covariates without negatively affecting the model metrics (GOMES et al., 2019; SOUZA et al., 2022).

In the testing phase of the model, we utilized a Holdout-test, where 25% of randomly selected samples were reserved exclusively for testing, and they were not included in the training phase (GOMES et al., 2019). The evaluation of the model's performance in Machine Learning relied on the holdout-test, a commonly used technique to assess how well the model generalizes on unseen data, i.e., data not used during training. The statistical indexes R^2 and RMSE – root mean squared error expressed the accuracy. To prevent biased predictions based on a single sample group, we conducted the prediction process 100 times, utilizing random samples for training and testing at each iteration (KUHN; JOHNSON, 2013; GOMES et al., 2019).

We compared the training and testing (holdout-test) phase metrics to assess the overfitting effect. Furthermore, we used the Kruskal-Wallis test with Dunn’s multiple comparison tests to determine if there were significant differences in the performance of these algorithms (95% confidence interval) (KRUSKAL; WALLIS, 1952). Finally, we carefully evaluated the generated maps to verify the result's consistency with the regional landscape's logic. We interpreted data based on climatic, geological, pedogeomorphological, and land cover aspects.

3 RESULTS

3.1 Model performance and covariate selection

We evaluated the performance of models with cross-validation and holdout-test. The RF showed the best performance in predicting groundwater ($R^2 = 0.16$ and RMSE = > 17 m^3/h), followed by CB, EARTH-

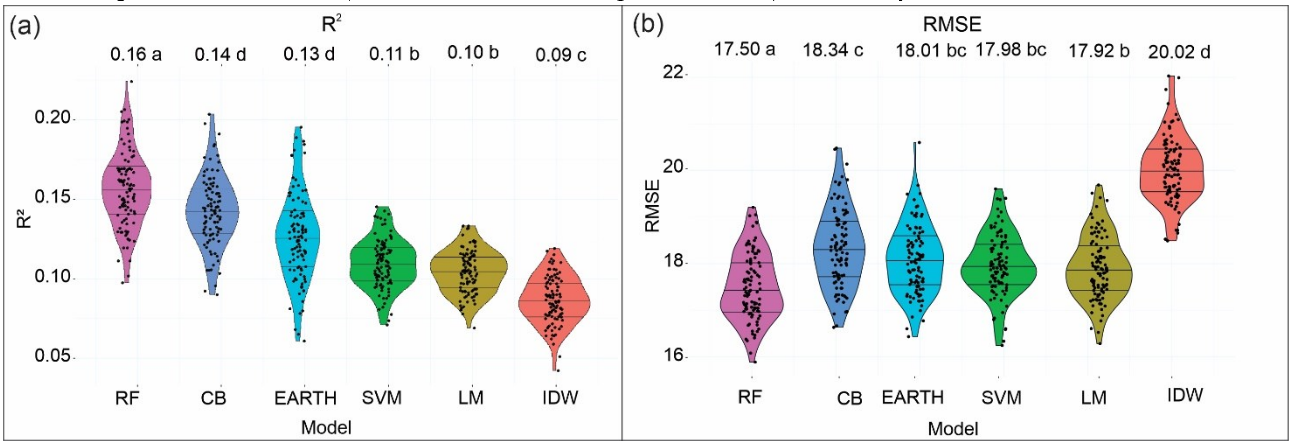
MARS, SVM, LM, and IDW (Table 1 and Figure 2). The high RMSE shows the complexity of groundwater mapping. Additionally, the low variation of the metrics (RMSE and R^2) in the training and testing phases indicates modeling with low overfitting.

Table 1 – Performance of models to predict the spatial distribution of discharge values from tubular wells. Metrics refer to the mean after 100 predictions. Models: Linear Regression Model (LM), Inverse Distance Weighting (IDW), Multivariable Adaptive Regression Splines (EARTH), Support Vector Machine (SVM), Cubist (CB), and Random Forest (RF). Metrics: R-squared metric (R^2), RMSE root mean squared error. SD – standard deviation, CV –coefficient of variation.

			Training – (5-fold cross-validation)			
Model	**(R^2) R-squared**	**R-squared SD**	**R-squared CV**	**RMSE**	**RMSE SD**	**RMSE CV**
RF	0.17	0.02	10.99	17.30	0.38	2.19
LM	0.11	0.01	8.53	17.76	0.32	1.78
SVM	0.12	0.01	9.42	18.16	0.40	2.23
Cubist	0.15	0.02	10.62	17.79	0.40	2.26
EARTH	0.14	0.02	14.29	18.20	0.40	2.20
IDW	0.13	0.03	21.85	20.02	0.70	3.49
			Testing – (holdout-test)			
Model	**(R^2) R-squared**	**R-squared SD**	**R-squared CV**	**RMSE**	**RMSE SD**	**RMSE CV**
RF	0.16	0.02	15.12	17.50	0.71	4.06
LM	0.10	0.01	12.73	17.92	0.68	3.81
SVM	0.11	0.02	14.15	18.34	0.84	4.55
Cubist	0.14	0.02	15.50	17.98	0.66	3.68
EARTH	0.14	0.02	14.29	17.20	0.40	2.33
IDW	0.13	0.03	21.85	20.02	0.70	3.49

Elaboration: The authors (2022).

Figure 2 – Violin plot Chart for R^2 and RMSE values of models after 100 predictions. Equal letters represent non-significant differences (confidence level of 95% p-value < 0.05) assessed by Kruskal-Wallis statistical.

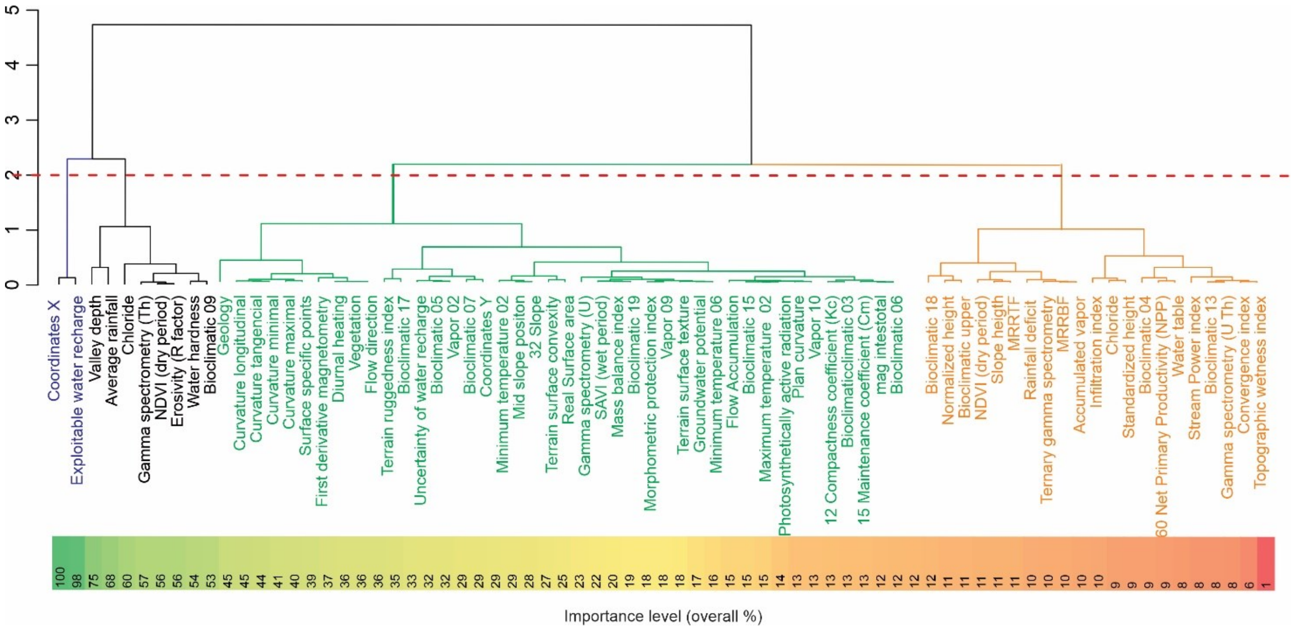


Elaboration: The authors (2022).

Initially, our study used a dataset with 177 input covariates. However, studies indicate that a large dataset can decrease the performance of the models and lead to overfitting (KUHN; JOHNSON, 2013). Thus, only ~41% of the covariates remained after applying the correlation cutoff and selecting covariates by RFE-RF (Figure 3). In addition, the RFE-RF hierarchically ranked the covariates based on their importance (% overall) (Figure 3).

The covariates selected by RFE can be separated into four large clusters using the K-means method (Figure 3). The first cluster consists of only longitude and exploitable recharge data, which exhibit prominent levels of importance at 100% and 98%, respectively. The second cluster comprises eight covariates with importance levels ranging from 75% to 53%, including indicators of climate, topography, land cover, geochemistry, and water properties. The third cluster encompasses a more extensive and diverse set of covariates, with importance levels ranging from 45% to 12%. Like the third, the fourth cluster contains various environmental aspects but with lower importance levels (< 9.8%).

Figure 3 – Hierarchy of covariates based on relative importance (%overall), grouping based on importance values using the K-means method.



Elaboration: The authors (2022).

3.2 Spatial distribution: groundwater potential

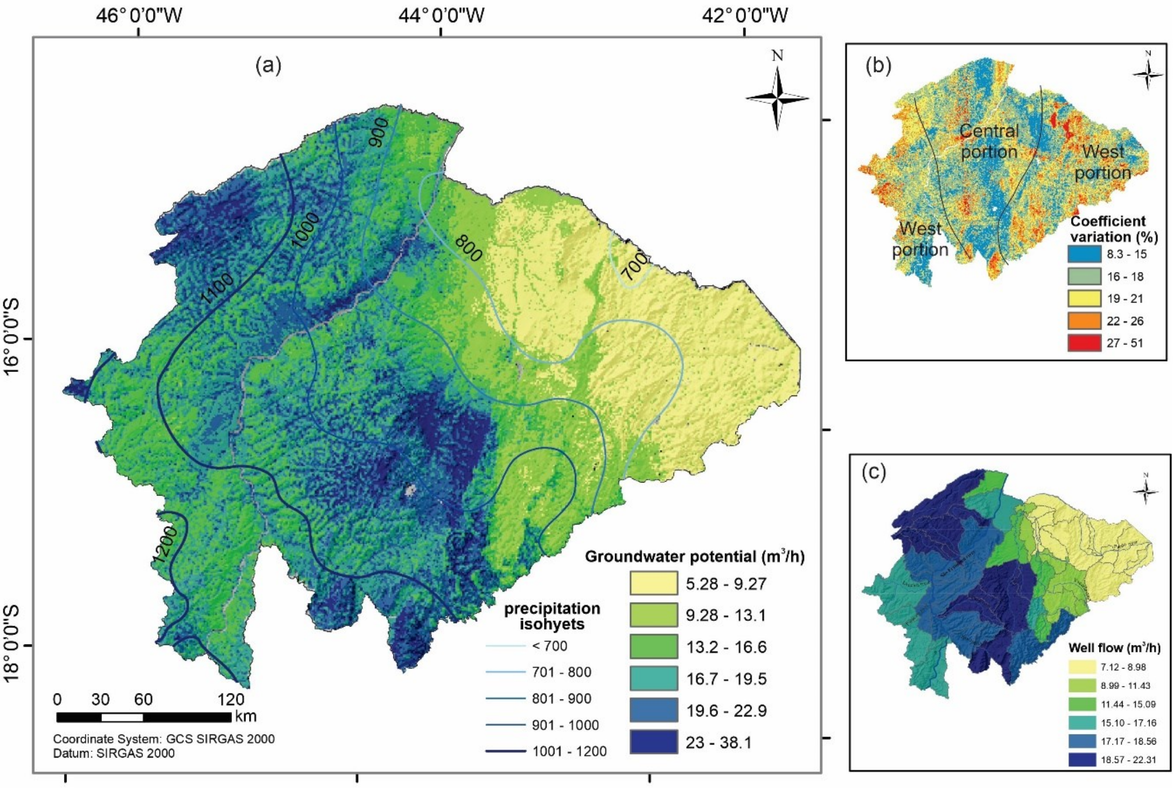
The RF model prediction was based on data from 4,028 tubular wells, with flows ranging from 0 to 100 m^3/h. However, in the prediction with the RF, the flow range represented on the map was 5 to 38 m^3/h. The Random Forest model is renowned for its ability to manage outliers, even though it does not explicitly eliminate them. This is because the way decision trees are constructed in the random forest makes the model less sensitive to outliers (BREIMAN, 2001). As a result, when combining multiple decision trees in the random forest, predictions are determined by averaging (in the case of regression), which leads to the exclusion of high values in the map representation. Consequently, 8.7% of the data (350 points) with flow values above 38.1 m^3/h were not represented on the map (Figure 4). The final map obtained by RF is based on the average of 100 spatial predictions (Figure 4a). The coefficient of variation of the prediction was higher in areas with slopes and or with low sample density (Figure 4b).

The central-western region of the study area has the most significant potential for groundwater, with flow rates from tubular wells of 13 to 38 m^3/h (Figure 4a). The average flow values of the tubular wells represented by sub-basin confirm this argument (Figure 4c). The hydrogeological domains of this region are sedimentary and metasedimentary rocks with low metamorphism and high fissure density (Figure 5). The relief is formed by plateaus and depressions, which have low slopes, except for the steep areas of the highlands. Soils in this relief have rapid permeability, including Oxisols (Latossolos) in most parts and Quartzipsamments (Neossolos Quartzarênicos).

The eastern region of the study area exhibits lower flow rates, with an average ranging from 5 to 13 m^3/h. The hydrogeological units are metamorphic, composed of schists and quartzites (Figure 5). The relief is characterized by mountain relief, with slopes ranging from 20 to 75%. The pedology presents Oxisols on the plateaus, while on the steep slopes, there are occurrences of Inceptisols, Entisols (Lithics), and rocky outcrops. This litho-pedological complexity associated with low sample coverage (Figure 1c) supports a high coefficient of variation for this zone (Figure 4b).

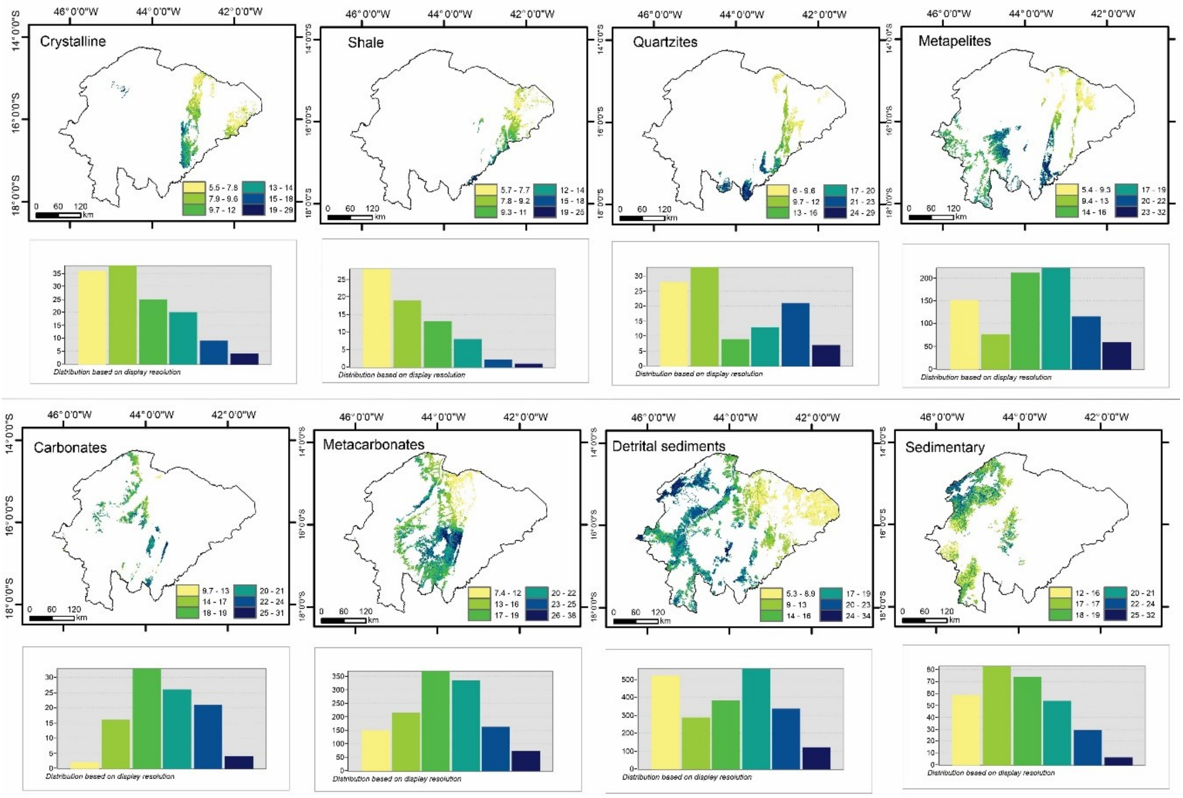
 DOI: http://dx.doi.org/10.14393/rbcv75n0a-65381

Figure 4 – (a) Groundwater potential map (b) coefficient of variation of the 100 predictions. (c) average flow values of tubular wells in sub-basins.



Elaboration: The authors (2022).

Figure 5 – Distribution of groundwater potential values in the hydrogeological domains with the representation of the histogram of values.



Elaboration: The authors (2022).

4 DISCUSSION

4.1 Model performance

We evaluated six machine learning (ML) models to predict groundwater potential in NMG from tube well flow. More complex models have greater efficiency in this modeling (RF and CB). The IDW interpolation model generated lower accuracy metrics. Generally, poor performance may be influenced by the spatial distribution of samples, point density, and the presence of outliers (STEIN, 2012). In the case of NMG tubular wells, their highest concentration coincides with the highest flow values (central and western parts) (Figure 1b). Thus, this characteristic of the data must also have affected the quality of predictions with the LM and EARTH-MARS models. Previous studies show that complex data regarding environmental characteristics require equally complex predictive models in modeling (KUHN; JOHNSON, 2013; GOMES et al., 2019; PADARIAN et al., 2020).

The best metrics in the predictions were observed in more complex models, including RF and CB. However, the RF model presented significantly higher R^2 values than the CB (K-W: p-value <0.05) (Figure 2). These results confirm a standard behavior of the RF in previous studies, that is, better performance compared to less robust models (BREIMAN, 2001; LIAW; WIENER, 2002; NAGHIBI et al., 2017; GOMES et al., 2019; SOUZA et al., 2020). Despite the better accuracy, R^2 values were low and RMSE high, confirming the spatial complexity of the predicted variable.

The acceptability of the RMSE depends on the specific context of the problem. There is an intrinsic challenge to measuring the variable in the case of flow from tubular wells. Data are momentary and reflect flow stability. Thus, this implies that even close wells can present different discharge values. Several factors, such as rainy seasons or periods of less precipitation, can influence this variation in flow. Furthermore, the study area presents an environmental complexity that makes it even more challenging to obtain an accurate prediction (MACHADO; SILVA, 2010; CÂNDIDO et al., 2019; COSTA, 2021). However, the spatial distribution of groundwater potential corresponded to the environmental context of the study area despite the limitation of metrics.

RF modeling uses a set of non-overestimated correlation environmental covariates, and this is an advantage in ML for making the model more straightforward and more reproducible (COELHO et al., 2019; SOUZA et al., 2022). In addition, the RFE function allowed for selecting a subset of relevant covariates for groundwater potential prediction (Figure 3). However, the RFE selected many covariates, confirming the complexity of the predicted variable. The chosen covariates are associated with hydrological cycle factors and are potentially promising for future groundwater modeling studies (RAHMATI et al., 2016).

The longitude covariate emerged as the most significant, capturing an effect of the spatial distribution of the data. There is a spatial pattern with wells with higher flows in the west and southwest areas and high rainfall (MELLO et al., 2007; GUIMARÃES et al., 2010). Areas with higher precipitation are expected to have higher water table recharge and flow from wells. However, it is essential to consider that other environmental factors can also play a significant role in this process.

The second covariate is a data that has a connection with water recharge and exploitation potential (Exploitable Water Recharge); this explains the high level of importance, as the covariate estimates the precipitated water available for groundwater recharge and, which can be considered an exploitable resource (CÂNDIDO et al., 2019).

Topographic covariates derived from the SRTM-DEM were frequently used in modeling (Figure 3). These covariates can potentially characterize spatial patterns of water distribution on the ground, influencing runoff and water infiltration variations and impacting the flow values in tubular wells (MOSAVI et al., 2021). In general, steep slopes lead to poor infiltration that hinders groundwater recharge, while low-slope areas, especially in highlands, have a higher rate of groundwater recharge (MAKONYO; MSABI, 2021).

Climatic factors were crucial in the modeling and included climatic data from WorldClim (FICK; HIJMANS, 2017). These covariates show the spatial behavior of climatic variables, and in the case of the study area, there are well-marked spatial patterns of these variables (GUIMARÃES et al., 2010). Regions positioned further west have higher precipitation levels, implying a more significant source of aquifer recharge. Therefore, the covariates that capture this climate behavior gained greater importance in the prediction. These factors explain the greater importance of covariates annual erosivity, bioclimatic, average precipitation, as they indicate climate trends (Figure 3) (MELLO et al., 2007; GOMES et al., 2019; SOUZA et al., 2022).

The gamma-spectrometry and geological map covariates were selected with an importance level between 45% and 57%. These covariates explain the variation and litho-pedological characteristics as a function of rock class (i.e., geology map) and geochemical characteristics (gamma-spectrometry). The different litho-pedological features have a strong influence on infiltration rates, which affects the flow level of tubular wells (RAHMATI et al., 2016; CÂNDIDO et al., 2019; MAKONYO; MSABI, 2021).

The covariates linked to the land cover aspects of the landscape were also relevant in the prediction indicated by the selection of the NDVI and SAVI indices. The indices are sensitive to detect vegetation patterns. The physiognomies of the NMG are predominantly deciduous and semi-deciduous, expressing a contrast in the landscape with vegetation that remains evergreen throughout the year (RIBEIRO; WALTER, 1998; SILVA, L. A. et al., 2020). This gradient of phytophysiognomies is a response to climatic seasonality and the geological/pedological context, and all these factors (climate geology, pedology, and land cover) are fundamental to influence the dynamics of infiltration and surface runoff (RAHMATI et al., 2016).

4.2 Groundwater potential mapping by RF model

The spatial prediction performed by RF showed consistency with the region's environmental conditions. The eastern areas, where shale, crystalline, and quartzite hydrogeologies predominate, have the lowest flow values in tubular wells (Figure 5). These hydrogeologies are represented by fissured/fractured aquifer systems where water circulation occurs in rock fissures, fractures, and faults. The primary characteristic of these systems is the absence or minimal presence of empty spaces in the rocks (BANKS et al., 1996). Studies suggest that these characteristics result in uncertainties in water productivity in wells, mainly when technical criteria for this aquifer are not implemented (INOCÊNCIO et al., 2021).

Geomorphological conditions of the areas to the eastern also affects water production from the tubular wells in that region. High slope reliefs in the eastern region contribute to surface runoff; this behavior is exemplified by a complex of mountains (Espinhaço Mountain Range) (COSTA, 2021). Water recharge in the eastern part is also reduced due to the predominance of small plateaus (COSTA, 2021), implying a smaller area for water infiltration. Despite this, these plateaus play a crucial role in the recharge of adjacent regions with lower elevations.

The low flow of tubular wells in the eastern region is also influenced by climatic factors, such as less precipitation (Figure 4). The characteristics of a semi-arid climate, characterized by scarce and irregularly distributed rainfall alongside prolonged periods of severe drought, further compound the challenges, leading to a diminished water potential compared to other regions (CÂNDIDO et al., 2019; SILVA, J. L. et al., 2020). Other environmental factors, such as the presence of shallow soils, also contribute to reducing infiltration rates (BOSQUILIA et al., 2019). In areas with deeper soils, evergreen forest vegetation predominates and has a prominent level of evapotranspiration, but these areas are sustainable.

Several environmental factors in the eastern areas can potentially explain the reduced water production levels observed in tubular wells within this region. Furthermore, land use practices, such as plateaus for eucalyptus plantations, can reduce water production. Studies in the NMG show that this activity can have high real evapotranspiration rates in summer, like wetlands such as Palm Swamp Veredas (SILVA, L. A. et al., 2020). Considering the climatic context of low rainfall in the eastern portion is also essential. Studies around the world suggest that eucalyptus plantations in watersheds under low precipitation regimes can negatively affect the water balance due to their water consumption (LITTLE et al., 2009; GONÇALVES et al., 2017; OUYANG et al., 2021).

The central and western areas of NMG have a predominance of tubular wells with higher flow rates, suggesting more significant groundwater potential. This region's climatic, geological, and geomorphological factors are crucial for this behavior. The hydrogeology of the central region is dominated by the karst-fissure system (Carbonates, Metacarbonates, Metapelites), which has a high potential for underground water resources depending on the structural arrangement of the rock (PESSOA et al., 2020). The rocks in this domain are overly complex due to the development of preferential infiltration paths through the dissolution of carbonates (sinks, dissolution conduits). This increases groundwater productivity in wells intersecting fracture and karstification zones (ATMAN et al., 2011; PESSOA et al., 2020).

Higher levels of rainfall in the central and western areas are a vital factor for aquifer recharging and maintaining water flow levels in tubular wells in the region. The area receiving this volume of rain consists of plateaus that function as a recharge zone and supply water to lower topography areas through infiltration and percolation. These plateaus have facilitated infiltration due to the predominance of sedimentary rocks (TENENWURCEL et al., 2020). For example, the Urucuia Aquifer system in the NW can reach up to 80 m in thickness with a high potential for water recharge (GASPAR et al., 2012).

In areas of plains and depressions, such as the São Francisco River depression, high groundwater potential values follow the direction of the main river channel (São Francisco River). Sedimentary geology and the influence of recharge zones (plateaus) contribute to expanding groundwater potential (GASPAR et al., 2012). Additionally, because the river has a significant flow, it controls water table levels during periods of low rainfall. Due to this potential, cities and agricultural activities extract water from drilled wells and rivers. However, overexploitation can influence flow reduction in the São Francisco River basin (LUCAS et al., 2021).

Although the modeling results suggest that the central and western regions exhibit the highest groundwater potential based on tubular well flow values, several important considerations must be considered. Firstly, it is essential to acknowledge that the modeling is based on well flow data measured during drilling, which may not accurately represent the current groundwater conditions. Furthermore, these regions' predominant land use and cover should be considered, as certain land uses exert remarkable pressure on groundwater resources. Additionally, identifying areas with high groundwater potential can be influenced by the number of sampling points available for analysis. For instance, Montes Claros, the most populous NMG with increased water demand (DURÃES et al., 2022), can significantly impact the results.

5 CONCLUSIONS

Random Forest model outperformed the SVM, EARTH, LM, and IDW models in predicting groundwater potential based on well flow values. The analysis utilized data from 4028 groundwater wells with flow values and seventy-three key covariates encompassing climatic, geological, topographic, and vegetation aspects to define the areas of groundwater potential. These covariates provided crucial insights into infiltration rates, surface runoff, and aquifer recharge processes.

Central and western regions of the NMG exhibited the highest groundwater potential, with flow values from tubular wells in these areas being 620% higher compared to the eastern regions. That means the central and western regions as key areas with significant potential. This disparity can be attributed to the significant presence of psammitic and carbonate sedimentary rocks characterized by high porosity and fissures, extensive plateaus, and higher rainfall levels observed in the central and western regions of the NMG.

The generated map is a valuable tool for public management and can be utilized as a variable in future models aiming to assess groundwater resource potential. However, it is essential to acknowledge certain limitations of the modeling approach, including the relatively low R^2 values obtained and the reliance on a single variable (well flow) to indicate groundwater potential. In the future, these models' accuracy can be increased by inserting additional covariates more linked to groundwater dynamics and exploring new modeling techniques.

This study provides valuable insights into the groundwater potential of the NMG, highlighting the central and western regions with groundwater potential. Further research and refinement of modeling techniques will contribute to a more comprehensive understanding of groundwater resources in the NMG.

The database and modeling results available at <https://zenodo.org/records/10008537>

Acknowledgment

The first author thanks the National Council for Scientific and Technological Development (CNPq) for the postdoctoral fellowship. The 3rd author thanks CNPq for providing a productivity grant. Thanks, the Research Support Foundation of the State of Minas Gerais (FAPEMIG – Process Number: BPD-00233-22). We would like to thank the anonymous reviewers of the Revista Brasileira de Cartografia.

Authors contribution

All authors contributed to the preparation of the paper. C. M. P. S.: conceptualization, research, methodology, writing; L. A. P. S.: writing, validation, formal analysis; G. V. V.: methodology, software, visualization; M. E. L. and E. I. F. F.: supervision, review.

Conflict of Interest

The authors declare no conflicts of interest.

References

ASHER, M. J.; CROKE, B. F. W.; JAKEMAN, A. J.; PEETERS, L. J. M. A review of surrogate models and their application to groundwater modeling. **Water Resources Research**, v. 51, n. 8, 5957-5973, 2015. DOI. 10.1002/2015WR016967

ATMAN, D.; VELÁSQUEZ, L. N. M.; FANTINEL, L. M. Controle estrutural na circulação e composição das águas no sistema aquífero cárstico-fissural do grupo bambuí, norte de minas gerais. **Águas Subterrâneas**, v. 25, n. 1, 2011. DOI. 10.14295/ras.v25i1.21023

BANKS, D.; ODLING, N. E.; SKARPHAGEN, H.; ROHR-TORP, E. Permeability and stress in crystalline rocks. **Terra Nova**, v. 8, n. 3, 223-235, 1996. DOI. 10.1111/j.1365-3121.1996.tb00751.x

BERGEN, K. J.; JOHNSON, P. A.; DE HOOP, M. V.; BEROZA, G. C. Machine learning for data-driven discovery in solid earth geoscience. **Science**, v. 363, n. 6433, eaau0323, 2019. DOI. 10.1126/science.aau0323

BOSQUILIA, R. W. D.; NEALE, C. M. U.; DUARTE, S. N.; LONGHI, S. J.; FERRAZ, S. F. D. B.; MULLER-KARGER, F. E. Evaluation of evapotranspiration variations according to soil type using multivariate statistical analysis. **Geoderma**, v. 355, 113906, 2019. DOI. 10.1016/j.geoderma.2019.113906

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, 5-32, 2001. DOI. 10.1023/a:1010933404324

CÂNDIDO, M. O.; BEATO, D. A. C.; FIUME, B.; SCUDINO, P. C. B.; CARNEIRO, F. A.; NASCIMENTO, F. M. F.; COUTINHO, M. M.; ALMEIDA, C. S. C.; SOCORRO, A. S.; SANTANA, M. S. **Projeto águas do norte de minas, panm: Estudo da disponibilidade hídrica subterrânea do norte de minas gerais: Relatório de integração**. CPRM. 2019.

CERQUEIRA, T. C.; MENDONÇA, R. L.; GOMES, R. L.; DE JESUS, R. M.; DA SILVA, D. M. L. Effects of urbanization on water quality in a watershed in northeastern brazil. **Environmental Monitoring and Assessment**, v. 192, n. 1, 65, 2019. DOI. 10.1007/s10661-019-8020-0

COELHO, M. T. P.; DINIZ-FILHO, J. A.; RANGEL, T. F. A parsimonious view of the parsimony principle in ecology and evolution. v. 42, n. 5, 968-976, 2019. DOI. 10.1111/ecog.04228

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, v. 20, n. 3, 273-297, 1995. DOI. 10.1007/BF00994018

COSTA, L. R. F. Considerações sobre as macrounidades geomorfológicas do estado de minas gerais–brasil. **William Morris Davis–Revista de Geomorfologia**, v. 2, n. 1, 11-18, 2021. DOI. 10.48025/ISSN2675-6900.v2n1.2021.109

CPRM. **-serviço geológico do brasil- mapa de domínios/subdomínios hidrogeológicos do brasil**. 2007. Available at: https://rigeo.cprm.gov.br/handle/doc/10323?mode=full. Acess date: 16/10/2023.

DAS, S.; GUPTA, A.; GHOSH, S. Exploring groundwater potential zones using mif technique in semi-arid region: A case study of hingoli district, maharashtra. **Spatial Information Research**, v. 25, n. 6, 749-756, 2017. DOI. 10.1007/s41324-017-0144-0

DURÃES, R. C. F.; GOMES, A. J. D. L.; GOMES, J. L. D. S. Lowering of the water table in the urban spaces of the city of montes claros-mg. **Research, Society and Development**, v. 11, n. 7, e20211730018, 2022. DOI. 10.33448/rsd-v11i7.30018

EMPINOTTI, V. L.; BUDDS, J.; AVERSA, M. Governance and water security: The role of the water institutional framework in the 2013–15 water crisis in são paulo, brazil. **Geoforum**, v. 98, 46-54, 2019. DOI. 10.1016/j.geoforum.2018.09.022

FICK, S. E.; HIJMANS, R. J. Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, v. 37, n. 12, 4302-4315, 2017. DOI. 10.1002/joc.5086

GASPAR, M. T. P.; CAMPOS, J. E. G.; MORAES, R. Determinação das espessuras do sistema aquífero urucuia a partir de estudo geofísico. **Revista Brasileira de Geociências**, v. 42, n. Supl 1, 154-166, 2012. DOI. 10.5327/Z0375-75362012000500013

GHISI, E. Potential for potable water savings by using rainwater in the residential sector of brazil. **Building and Environment**, v. 41, n. 11, 1544-1550, 2006. DOI. 10.1016/j.buildenv.2005.03.018

GOMES, L. C.; FARIA, R. M.; SOUZA, E.; VELOSO, G. V.; SCHAEFER, C. E. G. R.; FERNANDES FILHO, E. I. Modelling and mapping soil organic carbon stocks in brazil. **Geoderma**, v. 340, 337-350, 2019. DOI. 10.1016/j.geoderma.2019.01.007

GONÇALVES, J. L. M.; ALVARES, C. A.; ROCHA, J. H. T.; BRANDANI, C. B.; HAKAMADA, R. Eucalypt plantation management in regions with water stress. **Southern Forests: a Journal of Forest Science**, v. 79, n. 3, 169-183, 2017. DOI. 10.2989/20702620.2016.1255415

GONÇALVES, R. D.; STOLLBERG, R.; WEISS, H.; CHANG, H. K. Using grace to quantify the depletion of terrestrial water storage in northeastern brazil: The urucuia aquifer system. **Science of The Total Environment**, v. 705, 135845, 2020. DOI. 10.1016/j.scitotenv.2019.135845

GUIMARÃES, D.; REIS, R.; LANDAU, E. **Índices pluviométricos em minas gerais - boletim de pesquisa e desenvolvimento**. Sete Lagoas: Embrapa, available at <https://www.infoteca.cnptia.embrapa.br/handle/doc/879085 > acess date: 16/10/2023 2010. 88 p.

HUSSEIN, E. A.; THRON, C.; GHAZIASGAR, M.; BAGULA, A.; VACCARI, M. Groundwater prediction using machine-learning tools. **Algorithms**, v. 13, n. 11, 300, 2020. DOI. 10.3390/a13110300

IBGE. **Censo demográfico**. Rio de Janeiro, 2010. Available at: https://censo2010.ibge.gov.br/. Acess date: 16/10/2023.

INMET. **Instituto nacional de meteorologia - seca deixa 130 cidades em situação de emergência em minas gerais**. 2023. Available at: https://www.defesacivil.mg.gov.br/index.php?option=com_content&view=article&id=14. Acess date: 21 de maio de 2023.

INOCÊNCIO, T. D. M.; RIBEIRO NETO, A.; OERTEL, M.; MEZA, F. J.; SCOTT, C. A. Linking drought propagation with episodes of climate-induced water insecurity in pernambuco state - northeast brazil. **Journal of Arid Environments**, v. 193, 104593, 2021. DOI. 10.1016/j.jaridenv.2021.104593

IPCC. Impacts of 1.5 c global warming on natural and human systems. v. 2022. https://helda.helsinki.fi/handle/10138/311749

KLINK, C. A.; MACHADO, R. B. Conservation of the brazilian cerrado. **Conservation Biology**, v. 19, n. 3, 707-713, 2005. DOI. 10.1111/j.1523-1739.2005.00702.x

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. **Journal of the American Statistical Association**, v. 47, n. 260, 583-621, 1952. DOI. 10.1080/01621459.1952.10483441

KUHN, M.; JOHNSON, K. **Applied predictive modeling**. 26 ed. New York: Springer, 2013. DOI. 10.1007/978-1-4614-6849-3.

KUHN, M.; WING, J.; WESTON, S.; WILLIAMS, A.; KEEFER, C.; ENGELHARDT, A.; COOPER, T.; MAYER, Z.; KENKEL, B.; TEAM, R. C.; BENESTY, M.; LESCARBEAU, R.; ZIEM, A.; SCRUCCA, L.; TANG, Y.; CANDAN, C.; HUNT, T. Package ‘caret’. **The R Journal**, v. 2020.

LIAW, A.; WIENER, M. J. R. N. Classification and regression by random forest. **R news**, v. 2, n. 3, 18-22, 2002.

LITTLE, C.; LARA, A.; MCPHEE, J.; URRUTIA, R. Revealing the impact of forest exotic plantations on water yield in large scale watersheds in south-central chile. **Journal of Hydrology**, v. 374, n. 1, 162-170, 2009. DOI. 10.1016/j.jhydrol.2009.06.011

LUCAS, M. C.; KUBLIK, N.; RODRIGUES, D. B. B.; MEIRA NETO, A. A.; ALMAGRO, A.; MELO, D. D. C. D.; ZIPPER, S. C.; OLIVEIRA, P. T. S. Significant baseflow reduction in the sao francisco river basin. **Water**, v. 13, n. 1, 2, 2021. DOI. 10.3390/w13010002

MACHADO, M. F.; SILVA, S. F. D. **Geodiversidade do estado de minas gerais**. 2010. ISSN 8574990914. Available at: https://rigeo.cprm.gov.br/handle/doc/14704. Acess date: 16/10/2023.

MAKONYO, M.; MSABI, M. M. Identification of groundwater potential recharge zones using gis-based multi-criteria decision analysis: A case study of semi-arid midlands manyara fractured aquifer, north-eastern tanzania. **Remote Sensing Applications: Society and Environment**, v. 23, 100544, 2021. DOI. 10.1016/j.rsase.2021.100544

MARANGON, B. B.; SILVA, T. A.; CALIJURI, M. L.; ALVES, S. D. C.; DOS SANTOS, V. J.; OLIVEIRA, A. P. D. S. Reuse of treated municipal wastewater in productive activities in brazil’s semi-arid regions. **Journal of Water Process Engineering**, v. 37, 101483, 2020. DOI. 10.1016/j.jwpe.2020.101483

MEDEIROS, F. J.; OLIVEIRA, C. P.; TORRES, R. R. Climatic aspects and vertical structure circulation associated with the severe drought in northeast brazil (2012–2016). **Climate Dynamics**, v. 55, n. 9, 2327-2341, 2020. DOI. 10.1007/s00382-020-05385-1

MELLO, C. R.; SÁ, M. A. C.; CURI, N.; MELLO, J. M.; VIOLA, M. R.; SILVA, A. M. Erosividade mensal e anual da chuva no estado de minas gerais. **Pesquisa Agropecuária Brasileira**, v. 42, n. 4, 537-545, 2007. DOI. 10.1590/S0100-204X2007000400012

MILBORROW, S.; HASTIE, T.; TIBSHIRANI, R. Earth: Multivariate adaptive regression spline models. **R package version**, v. 3, 2-7, 2014.

MOSAVI, A.; SAJEDI HOSSEINI, F.; CHOUBIN, B.; GOODARZI, M.; DINEVA, A. A.; RAFIEI SARDOOI, E. Ensemble boosting and bagging based machine learning models for groundwater potential prediction. **Water Resources Management**, v. 35, n. 1, 23-37, 2021. DOI. 10.1007/s11269-020-02704-3

NAGHIBI, S. A.; AHMADI, K.; DANESHI, A. Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. **Water Resources Management**, v. 31, n. 9, 2761-2775, 2017. DOI. 10.1007/s11269-017-1660-3

NANEKELY, M.; AL-FARAJ, F.; SCHOLZ, M. Analysis of a joint impact of climate change and anthropogenic interventions on groundwater depletion in arid and semi-arid areas. *In*: SHERIF, M.;SINGH, V. P., *et al* (Ed.). **Water resources management and sustainability: Solutions for arid regions**. Cham: Springer Nature Switzerland, 2023. p. 167-203. DOI. 110.1016/j.jhydrol.2015.1009.1074.

NISTOR, M. M.; RAHARDJO, H.; SATYANAGA, A.; HAO, K. Z.; XIAOSHENG, Q.; SHAM, A. W. L. Investigation of groundwater table distribution using borehole piezometer data interpolation: Case study of singapore. **Engineering Geology**, v. 271, 105590, 2020. DOI. 10.1016/j.enggeo.2020.105590

NOBRE, C. A.; MARENGO, J. A.; SELUCHI, M. E.; CUARTAS, L. A.; ALVES, L. M. Some characteristics and impacts of the drought and water crisis in southeastern brazil during 2014 and 2015. **Journal of Water Resource Protection**, v. 8, n. 2, 252-262, 2016. DOI. 10.4236/jwarp.2016.82022

OLAYA, V.; CONRAD, O. Chapter 12 geomorphometry in saga. *In*: HENGL, T. e REUTER, H. I. (Ed.). **Developments in soil science**: Elsevier, 2009. v. 33, p. 293-308.

OUYANG, L.; WU, J.; ZHAO, P.; LI, Y.; ZHU, L.; NI, G.; RAO, X. Consumption of precipitation by evapotranspiration indicates potential drought for broadleaved and coniferous plantations in hilly lands of south china. **Agricultural Water Management**, v. 252, 106927, 2021. DOI. 10.1016/j.agwat.2021.106927

PADARIAN, J.; MINASNY, B.; MCBRATNEY, A. B. Machine learning and soil sciences: A review aided by machine learning tools. **SOIL**, v. 6, n. 1, 35-52, 2020. DOI. 10.5194/soil-6-35-2020

PAIVA, A. C. D. E.; NASCIMENTO, N.; RODRIGUEZ, D. A.; TOMASELLA, J.; CARRIELLO, F.; REZENDE, F. S. Urban expansion and its impact on water security: The case of the paraíba do sul river basin, são paulo, brazil. **Science of The Total Environment**, v. 720, 137509, 2020. DOI. 10.1016/j.scitotenv.2020.137509

PESSOA, P.; ATMAN, D.; KIMURA, G. Environmental problems in the lagoa santa karst. *In*: S. AULER, A. e PESSOA, P. (Ed.). **Lagoa santa karst: Brazil's iconic karst region**. Cham: Springer International Publishing, 2020. p. 283-303. DOI. 210.1007/1978-1003-1030-35940-35949_35913.

PINTO, C.; SILVA, L.; LEITE, C. Mapa geológico do estado de minas gerais, 1: 1.000. 000. CPRM-CODEMIG, Belo Horizonte. Belo Horizonte: CODEMIG. 2003.

RAHMATI, O.; POURGHASEMI, H. R.; MELESSE, A. M. Application of gis-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at mehran region, iran. **CATENA**, v. 137, 360-372, 2016. DOI. 10.1016/j.catena.2015.10.010

RCORE, T. **R: A language and environment for statistical computing**. Vienna, Austria (2016), 2023. Available at: http://www.R-project.org/. Acess date: Access Date: 15 jan 2023.

RIBEIRO, J. F.; WALTER, B. M. T. Fitofisionomias do bioma cerrado. *In*: SANO, S. M. e ALMEIDA, S. P. (Ed.). **Cerrado: Ambiente e flora**. Planaltina: EMBRAPA-CPAC, 1998.

SAHOUR, H.; GHOLAMI, V.; VAZIFEDAN, M. A comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer. **Journal of Hydrology**, v. 591, 125321, 2020. DOI. 10.1016/j.jhydrol.2020.125321

SENA, N. C.; VELOSO, G. V.; FILHO, E. I. F.; FRANCELINO, M. R.; SCHAEFER, C. E. G. R. Analysis of terrain attributes in different spatial resolutions for digital soil mapping application in southeastern brazil. **Geoderma Regional**, v. e00268, 2020. DOI. 10.1016/j.geodrs.2020.e00268

SHIKLOMANOV, I. A. Appraisal and assessment of world water resources. **Water International**, v. 25, n. 1, 11-32, 2000. DOI. 10.1080/02508060008686794

SIAGAS. **Sistema de informações de águas subterrâneas**. Rio de Janeiro. Serviço Geológico do Brasil - CPRM. Available in <http://siagasweb.cprm.gov.br/layout/> Acess Date 25 July 2023. , 2023. Available at: http://siagasweb.cprm.gov.br/layout/. Acess date: 25/07/2023.

SILVA, J. L.; RIBEIRO, E. M.; LIMA, V. M. P.; HELLER, L. As secas no jequitinhonha: Demandas, técnicas e custos do abastecimento no semiárido de minas gerais. **Revista Brasileira de Estudos Urbanos e Regionais**, v. 22, 2020. DOI. 10.22296/2317-1529.rbeur.202013

SILVA, L. A.; SOUZA, C. M. P.; LEITE, M. E.; FILGUEIRAS, R. Estimating water loss in an environmental protection area -minas gerais, southeast brazil. **Caderno de Geografia**, v. 30, n. 62, 768-768, 2020. DOI. 10.5752/P.2318-2962.2020v30n62p768

SINGHA, S.; PASUPULETI, S.; SINGHA, S. S.; SINGH, R.; KUMAR, S. Prediction of groundwater quality using efficient machine learning technique. **Chemosphere**, v. 276, 130265, 2021. DOI. 10.1016/j.chemosphere.2021.130265

SOPHOCLEOUS, M. Interactions between groundwater and surface water: The state of the science. **Hydrogeology Journal**, v. 10, n. 1, 52-67, 2002. DOI. 10.1007/s10040-001-0170-8

SOUZA, C. M. P.; FIGUEIREDO, N. A.; COSTA, L. M.; VELOSO, G. V.; ALMEIDA, M. I. S.; FERREIRA, E. J. Machine learning algorithm in the prediction of geomorphic indices for appraisal the influence of landscape structure on fluvial systems, southeastern - brazil. **Revista Brasileira de Geomorfologia**, v. 21, n. 2, 363-378, 2020. DOI. 10.20502/rbg.v21i2.1671

SOUZA, C. M. P.; VELOSO, G. V.; MELLO, C. R.; RIBEIRO, R. P.; SILVA, L. A. P.; LEITE, M. E.; FERNANDES FILHO, E. I. Spatiotemporal prediction of rainfall erosivity by machine learning, minas gerais state, brazil. **Geocarto International**, v. 37, n. 2, 1-19, 2022. DOI. 10.1080/10106049.2022.2060318

STEIN, M. L. **Interpolation of spatial data: Some theory for kriging**. New York, NY: Springer, 1999// 2012. DOI. 10.1007/978-1-4612-1494-6_1.

TAYLOR, R. G.; SCANLON, B.; DÖLL, P.; RODELL, M.; VAN BEEK, R.; WADA, Y.; LONGUEVERGNE, L.; LEBLANC, M.; FAMIGLIETTI, J. S.; EDMUNDS, M.; KONIKOW, L.; GREEN, T. R.; CHEN, J.; TANIGUCHI, M.; BIERKENS, M. F. P.; MACDONALD, A.; FAN, Y.; MAXWELL, R. M.; YECHIELI, Y.; GURDAK, J. J.; ALLEN, D. M.; SHAMSUDDUHA, M.; HISCOCK, K.; YEH, P. J. F.; HOLMAN, I.; TREIDEL, H. Ground water and climate change. **Nature Climate Change**, v. 3, n. 4, 322-329, 2013. DOI. 10.1038/nclimate1744

TENENWURCEL, M. A.; MOURA, M. S.; COSTA, A. M.; MOTA, P. K.; VIANA, J. H. M.; FERNANDES, L. F. S.; PACHECO, F. A. L. An improved model for the evaluation of groundwater recharge based on the concept of conservative use potential: A study in the river pandeiros watershed, minas gerais, brazil. **Water**, v. 12, n. 4, 1001, 2020. DOI. 10.3390/w12041001

UFV; CETEC; UFLA; FEAM. Mapa de solos do estado de minas gerais - escala 1: 500.000. v. 49, 2010.

UNEPE. Global environment outlook 3: Past, present and future perspectives. **Environmental Management and Health**, v. 13, n. 5, 560-561, 2002. DOI. 10.1108/emh.2002.13.5.560.1

USGS. **Earthexplorer** Disponível em:< http://earthexplorer.usgs.gov>, 2023. Available at: Disponível em:< http://earthexplorer.usgs.gov>. Acess date: Access Date: 27 jan 2023.

YAO, Y.; SUN, J.; TIAN, Y.; ZHENG, C.; LIU, J. Alleviating water scarcity and poverty in drylands through telecouplings: Vegetable trade and tourism in northwest china. **Science of The Total Environment**, v. 741, 140387, 2020. DOI. 10.1016/j.scitotenv.2020.140387

ZHOU, Y.; LI, W. A review of regional groundwater flow modeling. **Geoscience Frontiers**, v. 2, n. 2, 205-214, 2011. DOI. 10.1016/j.gsf.2011.03.003

 DOI: http://dx.doi.org/10.14393/rbcv75n0a-65381

Main author biography

Bachelor's degree in Geography from the State University of Santa Cruz (2011); Master's in Soil Science and Plant Nutrition from the Federal University of Viçosa (2013); Ph.D. in Soil Science and Plant Nutrition from UFV (2018); Currently holds a position of professor in the Department of Geosciences at Unimontes. Also participates in the Geography Postgraduate Program (Master's) at Unimontes. Conducting postdoctoral research at the same institution as part of a project funded by FAPEMIG. Primarily engaged in the following areas: modeling with machine learning algorithms, soil-landscape relationships, GIS, geomorphology and geology, geo-environmental mapping, zoning, and watershed management.