Souza, Eliana de; Pontes, Lucas Machado; Fernandes, Elpídio Inácio;
Schaefer, Carlos Ernesto Goncalves Reynaud; Santos, Eliana Elizabet dos

Spatial and Temporal Potential Groundwater Recharge: the Case of the Doce River Basin, Brazil

Revista Brasileira de Ciência do Solo

Division – Soil Use and Management | Commission – Soil and Water Management and Conservation

Spatial and Temporal Potential Groundwater Recharge: the Case of the Doce River Basin, Brazil

Eliana de Souza[(1)]* (iD), Lucas Machado Pontes[(2)], Elpídio Inácio Fernandes Filho[(3)], Carlos Ernesto Goncalves Reynaud Schaefer[(3)] and Eliana Elizabet dos Santos[(1)]

[(1)] Universidade Federal de Viçosa, Departamento de Solos, Programa de Pós-Graduação em Solos e Nutrição de Plantas, Viçosa, Minas Gerais, Brasil.
[(2)] Universidade de São Paulo, Departamento de Ciências Atmosféricas do Instituto de Astronomia, Geofísica e Ciências Atmosféricas, São Paulo, São Paulo, Brasil.
[(3)] Universidade Federal de Viçosa, Departamento de Solos, Viçosa, Minas gerais, Brasil.

ABSTRACT: Little is known about the groundwater recharge potential of weathered tropical catchments, where increasing water uptake is widespread to meet various water demands. This study aimed to estimate the volume of groundwater recharge of the Doce River Basin, Minas Gerais, Brazil. The BALSEQ model was applied to calculate the water balance over a period of two years (2007-2009). Evapotranspiration, runoff, and potential groundwater recharge (PGR) were calculated, using daily data on rainfall, potential evapotranspiration, and plant-available water. A soil survey was undertaken for all major soils occurring in the basin. Soils samples were used to determine hydraulic conductivity, bulk density, and water content at field capacity and at the permanent wilting point. Vegetation data were obtained from the literature and used to determine the following parameters: canopy interception, crop coefficient, and root depth. The estimated groundwater was spatially predicted using the Random Forests model with digital elevation, vegetation index, pedological, lithological, and climate maps. During the two years, an average of 32 % of rainfall was converted to groundwater. Annually, the percent of rainwater converted to groundwater varied between 27 and 48 % for all soil classes, highlighting the great temporal variability. The spatial prediction showed a volume of approximately 17,484 and 35,410 m^3 of rainfall being converted to groundwater for the first and second year, respectively. The BALSEQ model showed a feasibility for the water balance calculation and can be reapplied for updating the groundwater maps of the Doce River Basin. These maps could then be used to guide land use planning programs, with the aim to protect water resources.

Keywords: aquifer, infiltration, potential recharge, Random Forests Model, water balance.

INTRODUCTION

Groundwater recharge is determined by a combination of variables such as soil type and management, vegetation cover, topography, climatic, and geological conditions. In Brazil, groundwater is increasingly used to supply urban centers, industries, irrigation, agriculture, and recreation, which coexist with a lack of information regarding the total amount of water being taken from aquifers (Hirata et al., 2017).

A correct evaluation of aquifer recharge is fundamental for the calculation of water availability, especially in areas of high demand for water resources. In Brazil, groundwater has become more than a complementary water source for multiple uses. Aquifers in Brazil are used to supply water for up to 40 % of the population and play a strategic role in times of climate change with long and severe drought periods (Hirata and Conicelli, 2012; Hirata et al., 2017). Furthermore, aquifers are important groundwater reserves that should be preserved to provide water at affordable prices (Resende Filho et al., 2015).

Despite the importance of groundwater for multiples uses, knowledge of its status at a national scale is extremely limited. Regional studies on the status of the quantity and quality of groundwater are only available for specific sites in Brazil, and the data are outdated and scattered (Hirata and Conicelli, 2012). Such studies have only focused on São Paulo and the northeast of Brazil (Zoby and Matos, 2002).

Infiltration, together with rainfall, runoff, and evapotranspiration, constitutes the hydrological water cycle, which is a phenomenon of continuous circulation and distribution of water in the earth's soil, surface, atmosphere, and oceans. Infiltration represents the most important hydrological process for aquifer recharge and is determined by variables such as soil type and management, vegetation cover, topography of relief, as well as climatic and geological conditions (Sophocleous, 1992; de Vries and Simmers, 2002; Brandão et al., 2006).

Groundwater recharge is broadly defined as water that reaches an aquifer from any direction (down, up, or laterally). Both types of aquifer, (i) confined, where it is completely filled with water and separated from the surface by an overlying aquifer or an almost impermeable rock layer, and (ii) unconfined, where the water table is exposed to the earth's atmosphere through the unsaturated zone and can be recharged by the water that infiltrates and percolates the soil zone by a direct, indirect, or localized mechanism, depending on the local condition (Lerner, 1997).

Groundwater recharge can be quantified above and/or below the saturated zone. Models which operate above the saturated zone are considered predictive, giving estimates of potential recharge, while models which work below the saturated zone provide estimates of actual recharge (Rushton, 1988; Scanlon et al., 2002; Oliveira, 2004). Potential recharge, according to Rushton (1988), refers to the groundwater from the net balance of precipitation over evapotranspiration, which subsequently disappears through a local discharge system or by evaporation from the saturated zone, but may become 'permanent' recharge by lowering the water table after extraction.

Models for actual recharge estimates rely on water table levels or tracer methods. Predictive models, on the other hand, consist of soil water budget methods, which use climate and soil data to estimate the amount of water that passes the root zone. Within each zone, techniques are generally classified into physical, tracer, or numerical modeling approaches for groundwater quantification.

Approaches to quantify the direct recharge of groundwater include the following: water balance estimation, hydrogeological models, chemical tracer methods, and baseflow separation, whereas indirect recharge methods are based on water table fluctuation and the Darcy law. Direct recharge is conceptually the easiest to define and forms the basis of numerous commonly used recharge estimation techniques (de Vries and

Simmers, 2002). In addition, recharge can be calculated by direct measurements using the soil moisture budget via sensors (neutron or TDR probes) (Udawatta et al., 2011) and lysimeters. In contrast to sensors, which are relatively inexpensive, lysimeters are expensive and therefore not routinely used to calculate recharge (Scanlon et al., 2002).

For an overview of models, techniques, and challenges related to estimate groundwater recharge, the following studies provide detailed descriptions of the models and examples of applications, with case studies based on specific environmental conditions: Lerner et al. (1990), Sophocleous (1992), Lerner (1997), Simmers (1998), de Vries and Simmers (2002), Healy and Cook (2002), Scanlon et al. (2002), Oliveira (2004), among others. Despite the existence of numerous methods to calculate groundwater recharge, the hydrological conditions, quantity, and quality of data determine the selection of a desirable method (Ghandhari and Alavi Moghaddam, 2011; Kinzelbach et al., 2002). The water balance (water budget) method is the most common method applied throughout the different zones (Scanlon et al., 2002).

For calculating water recharge, we selected a numerical model which estimates the potential groundwater by means of the daily water balance calculation in the soil layer above the saturated zone. The model, BALSEQ, was developed by Lobo-Ferreira (1981) and used to estimate the recharge of unconfined aquifers in several areas in Europe (Lobo-Ferreira, 1981; Lobo-Ferreira and Delgado-Rodrigues, 1998; Paralta et al., 2003; Oliveira, 2004). The BALSEQ with modifications, developed by Oliveira (2004), was later used by Lobo-Ferreira et al. (2015) and Hugman et al. (2017).

Water balance calculations using the BALSEQ model were also applied in studies on the Bardez and Kakinada basins in India and China, respectively, by Chachadi et al. (2001, 2004). In Brazil, the model was successfully applied by Brito et al. (2009) in the state of Ceará, in the limestone (Karst) area of Minas Gerais (Camargo et al., 2011), and by Pontes et al. (2016) at the watersheds of the Serra da Mantiqueira.

Like most unsaturated-zone techniques, the numerical models calculated by BALSEQ provide point estimates of recharge. Hence, recharge is usually integrated as input data in a spatial model to allow the regional quantification of groundwater. Different modeling approaches have been used for this purpose by using spatial maps of environmental variables with strong relationships with groundwater recharge.

In recent years, several model approaches based on spatial data have been applied for mapping groundwater. Among these, we highlight the GIS-based models (Chung et al., 2010; Gaur et al., 2011; Naghibi et al., 2016; Rahmati et al., 2016), geostatistical models using kriging and regression-kriging (Noori et al., 2013; Crosbie et al., 2018), and a variety of machine learning algorithms such as Support Vector Machines, Logistic Regression, Splines, and Random Forests. The Random Forests model is often applied in groundwater pollution studies (Rodriguez-Galiano et al., 2014; Wheeler et al., 2015) and has shown good performance in mapping potential groundwater recharge (Xu and Valocchi, 2015; Naghibi et al., 2016; Rahmati et al., 2016; Naghibi et al., 2017; Golkarian et al., 2018; Sahoo et al., 2018).

Despite the success of Random Forests, due to its robustness and power of generalization demonstrated in groundwater mapping, its applications are still limited. Hence, we chose this model for the mapping of potential groundwater recharge in the unconfined aquifer of the Doce River Basin.

The Doce River is one of the largest rivers in southeast Brazil; its surface water and groundwater supply more than 3.3 million people, and the river is used for urban, mining, forestry, and industry purposes. Despite its importance, the water quality and quantity across the basin is critical due to widespread deforestation and inappropriate land use, along with soil degradation. Soil and forest degradation have led to soil erosion and, consequently, the silting of rivers and flooding, besides a direct impact on the increase in water price (Resende Filho et al., 2015).

The largest metallurgical complex of the Latin America is located in the Doce River Basin, resulting in water contamination related to mining activities, as sadly demonstrated by the environmental disaster that occurred on November 2015 (Neves et al., 2016). On this occasion, the collapse of a large dam at *Fundão*, Mariana's municipality, affected all riverine cities along the states of Minas Gerais and Espírito Santo, with a population of 2 million people left without access to clean water. Groundwater supply was particularly important for urban supplies during such an emergency.

The main objective of this study was to calculate the daily water balance for a two-year period and to develop a spatial prediction of groundwater recharge of the unconfined aquifer of the Doce River Basin in Minas Gerais State, Brazil.

MATERIALS AND METHODS

The methodological approach used, including data acquisition, processing, and modelling of groundwater estimate, and spatial prediction, is presented in figure 1. In this study, we used the BALSEQ model to obtain daily sequential estimates of water balance for the entire Doce River Basin. In addition, we used the Random Forests model for the prediction of groundwater recharge, using environmental covariates as proxies.

Study area: Doce River Basin

The Doce River Basin is located in southeastern Brazil, with 86 % of its area (71,472 km^2) in the state of Minas Gerais and 14 % in Espírito Santo. For the present study, we only considered the Minas Gerais sector. Three types of climates are found in the area, according to the Köppen classification system: Cwb in the highlands of the Espinhaço and the Mantiqueira mountain ranges, Cwa in the dissected upper highlands of the Doce River, and Aw on the middle plateau and the lower depressions (Alvares et al., 2014).

Average annual rainfall on the depression and low plateau varies from 1,000-1,200 mm, combined with high temperatures, contributing to high evaporation rates (Cupolillo, 2015). The temperatures of the coldest month in the basin are below 18 °C, but this never occurs in areas below 250-300 m. The rainy season ranges from October to March and the dry period from April to September (Cupolillo, 2015).

The basin is the nuclear domain of a hilly landscape known as "*Mares de Morros*". The landforms are dominated by convex hills on deep weathered soils with pronounced dissection. Nowadays, forest areas correspond to 26 % of the basin, while pastures occur in 67 %, and eucalypt plantations in 3 %. The remaining 8 % of the area are represented by transitional Brazilian Savanna (*Cerrados*), rupestrian vegetation (*Campos Rupestres*), and agricultural areas (Probio, 2006).

The predominant soils are *Latossolos Vermelho-Amarelos* (Rhodic Ferralsols) and *Argissolos Vermelhos* (Rhodic Acrisols), covering 66 and 23 % of the basin, respectively. *Cambissolos* (Cambisols) and *Neossolos Litólicos* (Leptosols) also occur, albeit to a lesser extent, usually on steep slopes at the plateau borders (FEAM/UFV/SEMA, 2011).

Data and methods

Pedological

The soil cover of the Doce River Basin was mapped based on 123 representative profiles of the main soil classes occurring in the area (Figure 2a). For each soil profile, data of the A and B diagnostic horizons were gathered. The soil classes mapped and the number of profiles per class were as follows: *Latossolos* (55), *Argissolos* (41), *Cambissolos* (17), and *Neossolos Litólicos* (10).
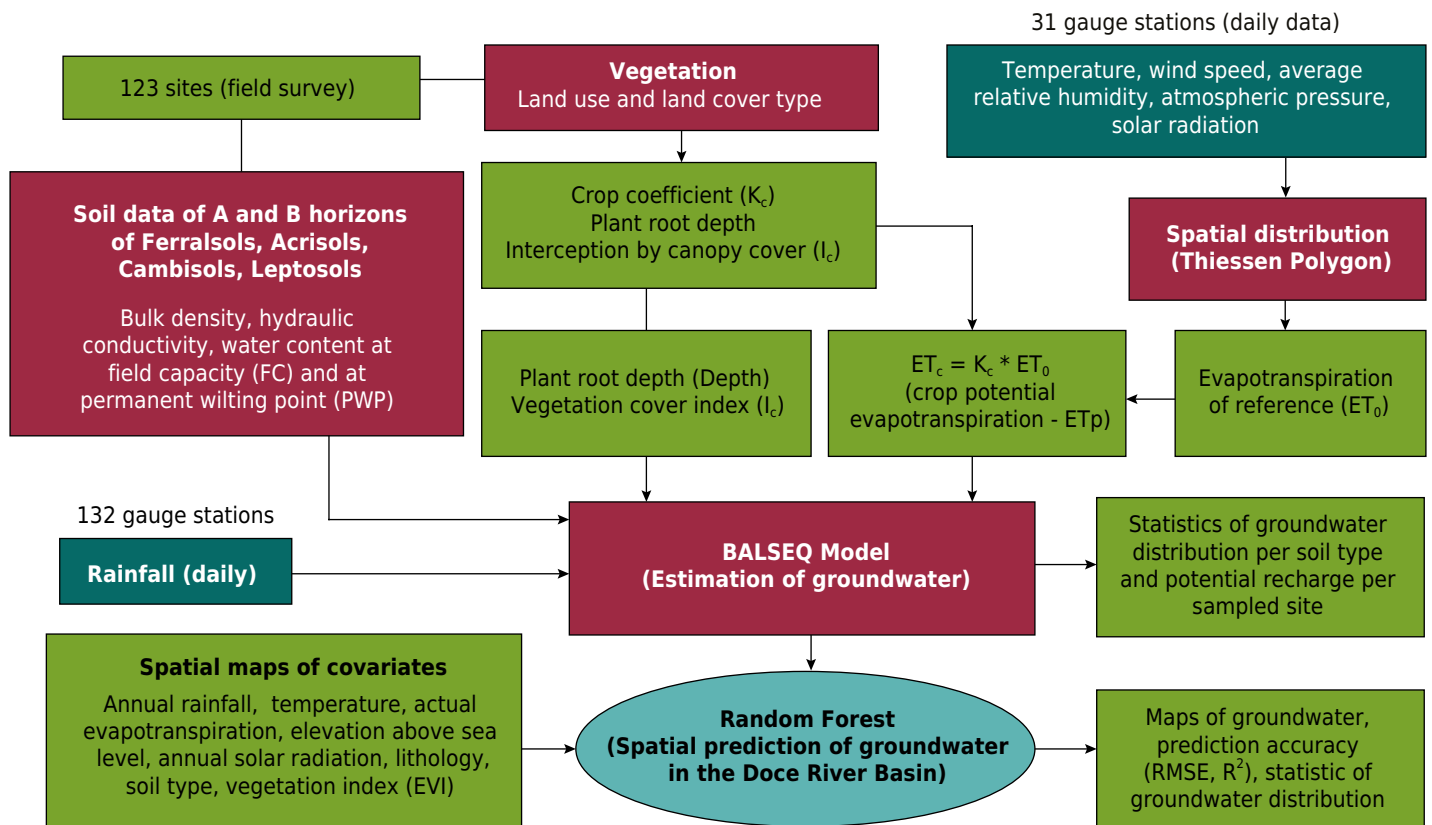
Figure 1. Flowchart of the methodological groundwater estimation and spatial prediction in the Doce River Basin, Minas Gerais, Brazil.

For each soil profile, undisturbed samples were collected for A and B horizons to determine the water content at field capacity (FC) and at permanent wilting point (PWP) by using Richard's chamber method (Richards, 1949), applying a 10-kPa and a 1,500-kPa tension, respectively. Soil bulk density was determined by the volumetric method, and the hydraulic conductivity of saturated soil (K_0) was evaluated according to Teixeira et al. (2017). Soil water recharge was calculated for both A and B horizons, accounting for a total of 231 soil samples.

Meteorological

Data of daily rainfall for the period from 1975 to 2010 was obtained from regional stations operated by the Brazilian National Waters Agency (*Agência Nacional de Águas* - ANA). We acquired data from 88 stations located inside the basin and from 44 stations in the surrounding area (Figure 2b).

Evapotranspiration of reference (ET_0) was calculated via the FAO-Penman-Monteith method (Allen et al., 1998). For the ET_0 calculation, daily climatic data of minimum and maximum temperature, wind speed, average relative humidity, atmospheric pressure, and solar radiation of 31 meteorological stations were used. The data were obtained from nine stations monitored by a private company (CENIBRA - Celulose Nipo-Brasileira S/A) and from 22 stations operated by the National Institute of Meteorology (*Instituto Nacional de Meteorologia* - INMET), with eight stations within the basin and fourteen in the surroundings (Figure 2c). For the spatial prediction of these meteorological variables, the Thiessen Polygon was used, which is a commonly used methodology for computing the mean value for catchment meteorological observations. These variables were then used to calculate the potential evapotranspiration (ETp). The distribution of the meteorological station is presented in figure 2.

The selected period for groundwater modeling was based on the available data. In the study area, the majority of rainfall stations (132 used) have a time series available of more
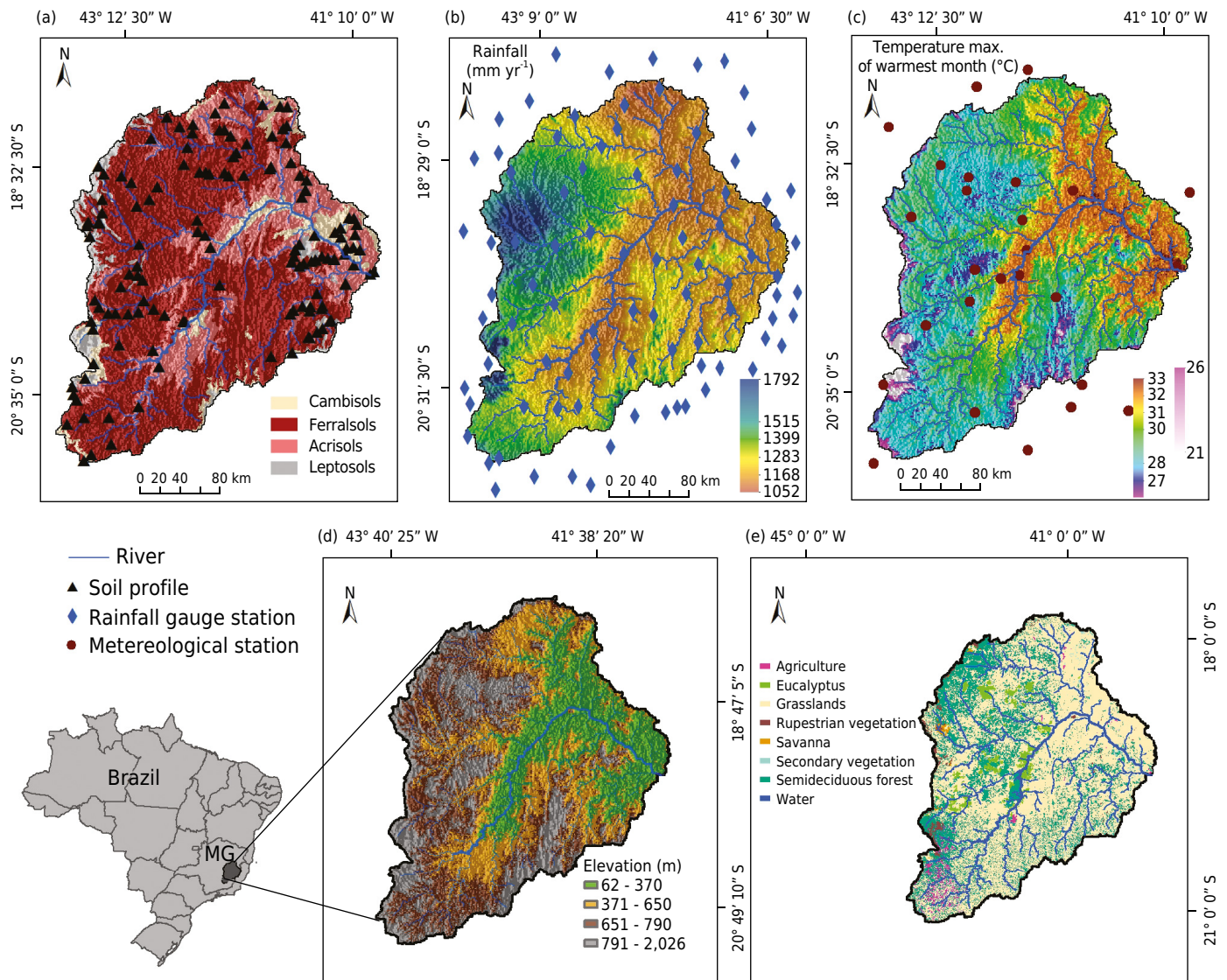
Figure 2. Location of the Doce River Basin in the state of Minas Gerais (MG), Brazil, and the thematic maps: (a) soil types and location of the soil profile used, (b) rainfall gauge stations, (c) meteorological gauge stations, (d) land coverage, and (e) elevation above sea level.

than 40 years of recorded data. However, the gauge stations for other meteorological variables used in the models applied operated in the following years: 2002/2003 (10 %), 2005 (10 %), 2006 (39 %), 2007 (33 %), and 2008 (10 %). Considering that less than 50 % of the stations have a 5-year period when the groundwater was modeled (2011), we chose to work on a 3-years period, using daily data without filling the gaps.

As we aimed to evaluate the spatial pattern of the average recharge in a large basin, we found it more consistent to select a time series with high data quality instead of the traditional approach of splitting the time series into calibration and validation periods. Furthermore, a poor spatial distribution of rainfall data is one of the main sources of uncertainty in hydrological modeling (Pluntke et al., 2014).

Vegetation

Land use and land cover were classified during the soil survey, recording the geographical position of each of the soil profiles by using a GPS positioning system. The vegetation parameters obtained were as follows: crop coefficient (K_c), plant root depth (depth), and the index of vegetation cover (Ic), which determined the percentage of rainfall intercepted by vegetation cover (Table 1).

Table 1. Vegetation types and number of profiles sampled in the basin, with respective parameters adopted for the vegetation in the water balance model

Coverage	Profile	K_c	Root depth	Ic
			m	%
Banana	1	1.1	0.40	-
Coffee	5	1	0.50[(a)]	
Corn	2	1.2	0.30	
Eucalyptus	12	1.00[(1)]	2.50[(b)]	10
Forest	14	1.00[(1)]	3.50[(b)]	15
Outcrop	3	0.75[(2)]	0.30	
Pasture	75	0.80[(3)]	0.30	
Shrub	10	1.00[(1)]	1.00[(b)]	
Sugar cane	1	1.25	0.30	

K_c and soil depth values were obtained from Allen et al. (1998), except: [(1)] adapted from conifers (Allen et al., 1998), [(2)] adapted from grazing pasture (Allen et al., 1998), [(3)] Alencar et al. (2009). [(a)] Depth for coffee (Kobayashi, 2007). [(b)] Root depth for forest, eucalyptus, and shrub (Almeida and Soares, 2003). Ic = interception by canopy cover (Almeida and Soares, 2003; Moura et al., 2009; Arcova et al., 2003). - = no value.

The calculation of effective soil depth up to which water is subject to evapotranspiration (rooting depth) was based on the literature (Allen et al., 1998; Almeida and Soares, 2003; Kobayashi, 2007), considering values for the culture mid-season stage.

The coefficient K_c was used to estimate plant water consumption throughout the development stages as it allows to calculate the crop potential evapotranspiration (ETp) by multiplying K_c by ET_0. We used a K_c value of the average stage of the culture (Allen et al., 1998), whereas the same values were used for forests as well as eucalyptus and pine forests. Similarly, the same values were used for pasture and open grassy vegetation on highlands (Table 1).

The Ic refers to the percentage of water retained in the aboveground vegetation (leaves, branches, and trunks) during a rainfall event, which evaporates without reaching the ground; values of I_C were obtained from the literature (Almeida and Soares, 2003; Arcova et al., 2003; Moura et al., 2009). An Ic of 10 and 15 % of rainfall was considered for eucalyptus and forest vegetation, respectively. These percentages were deducted from the total daily rainfall in areas with forest cover.

Model of sequential water balance - BALSEQ

Considering the variables rainfall, actual evapotranspiration, variation of water storage, surface runoff, and recharging potential, the water balance was calculated using the BALSEQ model, as presented in the workflow in figure 3. The calculation of the balance can be expressed by the equation 1:

$$P - ETR - DAi - SR - I = \varepsilon \qquad \text{Eq. 1}$$

in which P is the rainfall precipitation in mm; ETR is the actual evapotranspiration (mm); DAi is the variation (final - initial) of the soil water storage (mm); SR is the surface runoff (mm); I is the infiltration (mm); and ε is the error of the water balance.

As the aim of BALSEQ is to estimate the Potential Groundwater Recharge (PGR), PGR is calculated as the difference between the soil water content (Hi) and the water stored in the soil (TWS) (Equation 2). To calculate the sequential water balance, it is necessary to know the values of rainfall (P) and ET_0 for each time interval of the balance, as well as the total water stored in the soil (TWS):

$$TWS = FC \times Depth \times Bd \times 10^{-1}$$
<div align="right">Eq. 2</div>

in which TWS is the water in mm; FC is the soil water content at field capacity (g g^{-1}); Bd is the soil bulk density (Mg m^{-3}), Depth is the soil depth subjected to evapotranspiration in centimeters (usually the effective root depth), and the factor 10^{-1} is used to obtain the water depth in millimeters (mm).
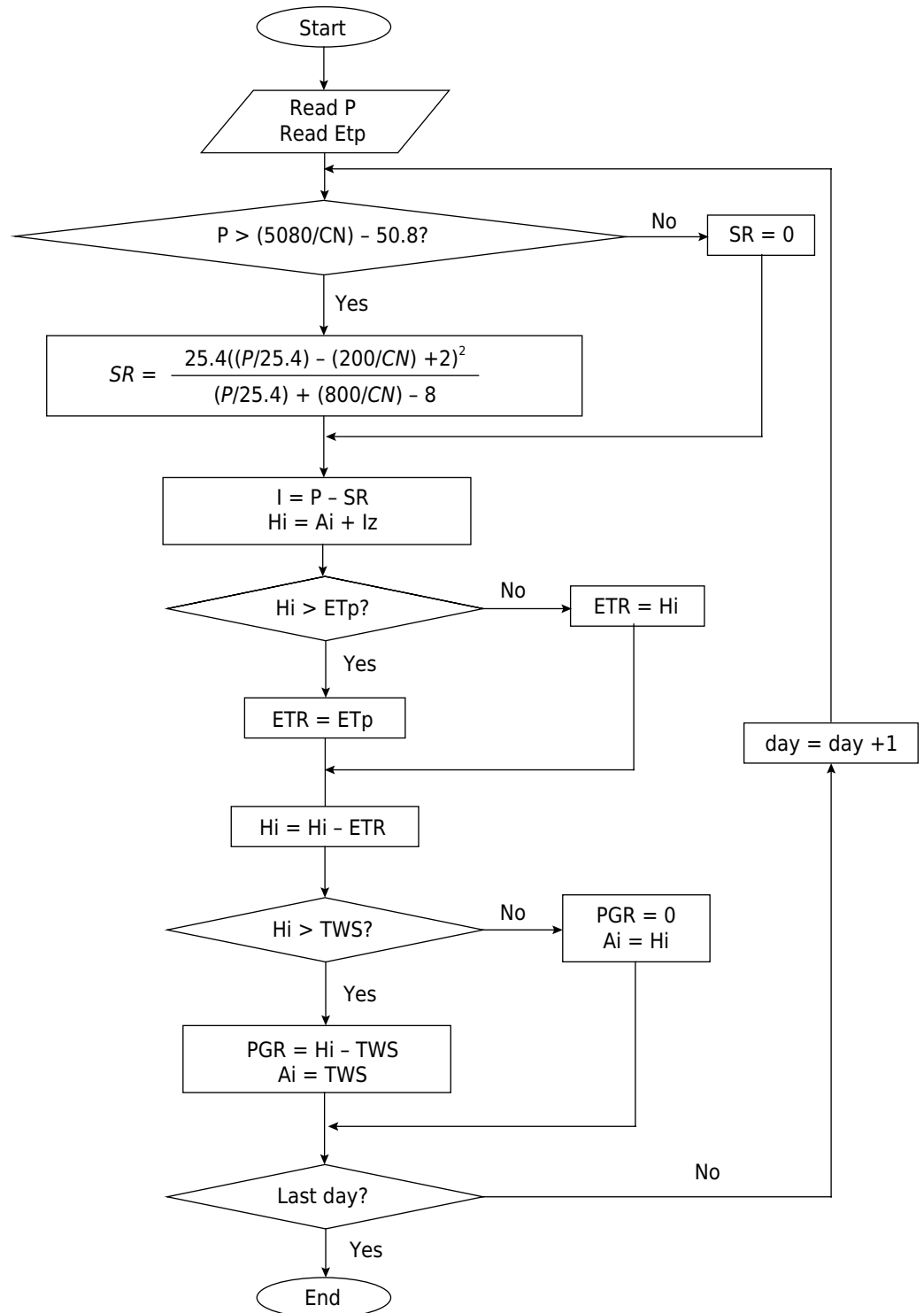
Figure 3. Schematic representation of the BALSEQ model (Modified from Lobo-Ferreira, 1981). P = rainfall (mm); ETp = potential evapotranspiration (mm); CN = coefficient of CN method; SR = surface runoff (mm); I = infiltration (mm); Hi = water content stored in the soil over the day (mm); Ai = water content stored in the soil at the end of the day (mm); ETR = actual evapotranspiration (mm); TWS = total soil water storage capacity (mm); PGR = potential groundwater recharge (mm).

The calculation of TWS in the model proposed by Lobo-Ferreira (1981) considers the soil water content at field capacity (FC) minus the water content at the permanent wilting point (PWP). In contrast, we only considered the water content at FC to obtain a more precise recharge. As the water at PWP is strongly adsorbed on soil particles, and unavailable for most plants, any increase would not effectively contribute to groundwater recharge.

Runoff was determined by the Curve Number (CN) method, which has been developed by the US Department of Soil Conservation (SCS, 1993) and uses data of K_0, soil water at FC and the soil classification according to the soil class and vegetation cover.

The calculation of the water balance on the BALSEQ model run continuously for the two-year period (2007-2009), which means the components SR, I, and ETR in the balance accounted for 100 % of the total rainfall at the end of the two-year period. However, we also calculated the annual balance, separately for each year, by subtracting the variables of each year from the balance of the total period.

The bottomland areas of *Gleissolos* (Gleysols) and *Neossolos Flúvicos* (Fluvisols) and the water bodies do not directly contribute to aquifer recharge, and thus, they were not considered in the groundwater recharge prediction, although they have been considered for runoff calculations.

The period of time used to calculate the water balance was defined based on continuous data availability, setting a maximum limit of 15 % of data missing for each station used in the model. The water balance was calculated for two hydrological years (2007-2008 and 2008-2009). The available data to calculate evapotranspiration was the main constrain for the two-year period under study.

Since it is a daily water balance model, days without data were excluded from the model, and the balance continued on the following day. Thus, the number of days within the two-year period of calculation differed between stations, which can result in non-systematic errors, especially in stations with missing data during the rainy season.

The water content stored in the soil on the first day of the balance (Ai) to the start of the groundwater recharge calculations was determined interactively, in which the soil water content determined by the model at the end of the two hydrological years was considered as the initial water content at the first day of the balance. This procedure was repeated until the end of the balance, in which the sum of SR, ETR, and I was equal to the total precipitation in this period.

The BALSEQ model was developed for seasonal regions, which require daily balancing due to the water deficit caused by the difference between rainfall and potential evapotranspiration, on a daily basis. This aspect does not limit its application in a tropical humid environment. As pointed out by Lobo-Ferreira and Delgado-Rodrigues (1998), the model is sensitive to the parameters Curve Number and Total Water Storage Capacity (TWS). Although it was designed for a different region, it has been successfully applied to tropical conditions in Brazil in a large number of hydrological studies. The parameter soil water capacity (TWS) was adopted using field data of soil and standard laboratory analyses. Plant evapotranspiration cannot be directly measured and is therefore best estimated by the information available.

The BALSEQ model does not allow any assumptions related to the local climate. The model uses local climatological data as input to calculate evapotranspiration, runoff, and infiltration. The volume of infiltrated water that is potentially converted to groundwater recharge, when it exceeds the water storage capacity at a given soil depth, depends upon the local characteristics. This depth is measured according to the effective depth of the vegetation root system.

Therefore, the model is expected to give a reliable water balance according to the assumptions adopted. The main limitation of the model is that it gives punctual estimates, and to overcome this limitation, we used spatial maps as covariates of groundwater and applied the Random Forests model for the spatial prediction of groundwater recharge.

Model assessment

A three-step assessment was performed as follows: 1) sensitivity analysis; 2) uncertainty analysis; and 3) model validation.

Sensitivity analysis

Sensitivity analysis plays an important role in the calibration of hydrological models because it enables the evaluation of which parameters should have their values changed in the calibration. In addition, it provides information on the magnitude and direction (increase or decrease) of this change. One-at-time (OAT) sensitivity analysis presents the sensitivity of a variable (in this case PGR) to changes in a given parameter when all other parameters are kept constant. We used the sensitivity index (SI) (Equation 3) proposed by McCuen and Snyder (1986):

$$SI = \frac{\dfrac{O_2 - O_1}{\bar{O}}}{\dfrac{I_2 - I_1}{\bar{I}}},$$

<div align="right">Eq. 3</div>

in which SI is the sensitivity index; O and I stand for the output and input parameter value, respectively, associated to the minimum (I_1, O_1), maximum (I_2, O_2), and average value ($\bar{I}$, $\bar{O}$) of a given parameter.

We tested the following parameters: field capacity (FC), bulk density (Bd), crop coefficient (Kc), root depth, and curve number (CN). The model sensitivity to one parameter depends on the value of other parameters, as we did not have the "correct" values of the parameters. The initial values used for the analyses were defined according to the standard deviation and the average value of each parameter tested as follows: FC = 0.25 (g g^{-1}); Bd = 1.1 (Mg m^{-3}); Kc = 1; Depth = 0.50 m; and CN = 50.

Uncertainty analysis

All models are prone to uncertainties that can be structural, related to the algorithm, to its assumptions and to the parameters used. Uncertainties are also due to input data and caused by all types of data errors, the spatial distribution of measurement data, and the temporal resolution of measured data. The rainfall data is the main source of uncertainty in hydrological models (Fischer et al., 2014).

Traditionally, uncertainty analysis in hydrological models has been made using simulations in which the model performance is assessed while changing parameter values. These techniques have high computational costs, and only the uncertainty of one parameter is assessed. In the case of BALSEQ, the advantage of the model is its simplicity, which allows a discussion of uncertainties in a comprehensive way to evaluate not only the parameters, but also the model structure and the input data used in the groundwater modeling.

Model validation

To validate the groundwater recharge from the BALSEQ model in the typical soil conditions, vegetation, and climate in the Doce River Basin, two locations at the headwaters (municipality of Araponga) were monitored by collecting soil moisture data during a period of 10 months (09/01/2009 to 07/07/2010). The locations were monitored using

a weather station (E-5000, Irriplus Equipment's) and soil moisture reflectance sensors (CS 616, Campbell Scientific) (Udawatta et al., 2011).

The land uses in these two areas are conventional coffee plantations (42° 31' 45.4" W and 20° 41' 53.9" S) and coffee plantations with trees (Agroforestry system) (42° 32' 34.5" W and 20° 39' 52.7" S); both areas are on *Latossolos*, with a total rainfall of 1,250 mm. The water balance for the 10-month period on the two locations was calculated using the BALSEQ model, applying local data (soil, vegetation, rainfall, and evapotranspiration) and the groundwater recharge compared with the soil moisture data of sensors installed at a depth of 1 m.

Spatial prediction of the groundwater using the Random Forests model

Groundwater estimated with the BALSEQ was used to predict the water recharge for locations not sampled in the basin, using the Random Forests model with maps of variables correlated with groundwater. Random Forests is a tree-based model developed by Breiman (2001), whose working principle is based on the construction of many trees, resulting in a "model of forests" without any pruning during the growth of the trees.

The setting of the model parameters [number of trees in the forest (ntree), number of predictors sampled for splitting at each node (mtry), and the number of divisions at the end of each tree (nodesize)] depends on the size of the dataset. Breiman (2002) explains how to define the parameters for obtain optimal estimates. In the case of a regression model, as in this study, basically a "nodesize" equal to 1 is usually appropriated and should not be very large to avoid a high computing demand. The "mtry" should be tested considering values in a range of the squared root of the number of predictors and a number twice larger or twice smaller, whereas the "ntree" does not affect the model accuracy; a number between 1,000 and 5,000 trees generally provides good results.

Each tree in the model is constructed from a selection of samples from the total dataset, which is done by the bootstrap method, while another subset is kept for assessing model accuracy. This selection, based on the data split, allows for a more robust error estimate, using the second set of samples called Out-Of-Bag (OOB). The OOB data set is excluded from the training dataset. The mean square error (MSE-OOB) is calculated from the aggregation of the predictions of all OOBs of the trees (Liaw and Wiener, 2018).

The MSE of the OOB is also used to evaluate the importance of covariates in the prediction. For each tree in the Random Forests model, the prediction error of the OOB portion of the data is recorded (MSE). Subsequently, another run is performed after permuting each predictor variable. The differences between the two are then averaged over all trees and normalized by the standard deviation of the differences.

The subdivision of the trees at each node is made from the best set of subdivisions, randomly selected from the predictors of the total dataset, which prevents over-fitting.

The advantages of the method, according to Breiman (2002), are as follows: (a) higher prediction performance; (b) no overfitting; (c) low correlation of individual trees, as tree diversity increases through of the use of a limited number of predictors; (d) low deviation due to the mean of a large number of trees; and (e) robust error estimates using the OOB dataset.

Covariates used as a proxy of groundwater recharge

Random Forests was run in R software (R Development Core Team, 2016), using the Random Forests Package (Liaw and Wiener, 2018). All maps used to predict water recharge of the Doce River Basin were resampled for a 1,000-m spatial resolution. The choice of the spatial resolution was based on the scale and resolution of the maps. The covariates were selected as candidates for spatial models of recharge, based on factors driving

the spatial variation of groundwater drivers identified in the literature (Lerner, 1990; Sophocleous, 1992; de Vries and Simmers, 2002; Scanlon et al., 2002), including rainfall, geology and soil, land cover, topography and landform, and groundwater condition.

To consider model parsimony, we used the RF importance rank of the OOB estimates to select the best predictor among the following terrain attributes: elevation above sea level, multiresolution index of the ridge top flatness (MrRTF), multi-resolution index of valley bottom flatness (MRVBF), slope, standardized elevation, topographic wetness index, and valley depth. The maps used as proxies for water recharge in the Random Forests model were as follows:

- Annual rainfall and temperature, spatial resolution of 1,000 m (Hijmans et al., 2005). These two variables stand for the climate factor of the model;

- Actual evapotranspiration (mm), with a spatial resolution of 920 m (Trabucco and Zomer, 2010);

- Elevation above sea level, derived from a digital elevation model (DEM) generated with SRTM images with a spatial resolution of 90 m (Jarvis et al., 2008), representing the relief;

- Annual solar radiation, generated on ArcGis using the DEM of SRTM images;

- Lithological map representing the parent material, scale 1:1,000,000 (CPRM, 2004);

- Pedological map, scale 1:500,000 (FEAM/UFV/SEMA, 2011);

- Vegetation index EVI (Enhanced Vegetation Index), used as a proxy of vegetation (MODIS, 2009).

Assessment of Random Forests performance in groundwater prediction

To assess the performance of Random Forests in terms of predicting groundwater recharge, we partitioned the dataset in model training (75 %) and model validation (25 %). The model fit was assessed by means of the accuracy metrics, Root Mean Squared Error - RMSE (Equation 4) and R-squared (R^2), calculated using the OOB estimates of model training. The R^2 of Random Forests is a measure of model-fit comparable to the R^2 values of other models. This "pseudo R^2" is labeled the "percentage of variance explained by the model". In addition to the model training, RMSE and R^2 were calculated for an external validation dataset to assess the power.

The RMSE indicates the magnitudes of the errors in the predictions by the global standardized correction included in the index calculation, as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{x}_i - x_i)^2}{n}}$$

Eq. 4

in which: x_i represents the recharge estimated by BALSEQ; $\hat{x}_i$ the recharge predicted by the Random Forest model; and n is the number of locations used for the estimates.

RESULTS AND DISCUSSION

Soil data

The statistical distribution of soil properties used as parameters to calculate the water balance for the BALSEQ model is shown in table 2. The results correspond to the expected behavior of the soil class surveyed in the basin, where *Latossolos*, due to a greater development of the granular structure, show the highest hydraulic conductivity, followed

Table 2. Statistical distribution of the soil properties used as parameters of water balance in the BALSEQ model

Statistic	FC - 10 kPa	PWP - 150 kPa	Bd	Ko	CN
	g g^{-1}		Mg m^{-3}	m s^{-1}	
Argissolos (Acrisols) (n = 80)					
Minimum	0.12	0.05	1.05	0.000000	36
Mean	0.26	0.15	1.40	0.000030	75
Maximum	0.58	0.33	1.69	0.000538	89
Standard deviation	0.08	0.05	0.17	0.000067	9
Cambissolos (Cambisols) (n = 34)					
Minimum	0.04	0.02	0.87	0.000000	49
Mean	0.24	0.12	1.32	0.000067	73
Maximum	0.37	0.21	1.67	0.000436	86
Standard deviation	0.08	0.05	0.19	0.000108	11
Latossolos (Ferralsols) (n = 107)					
Minimum	0.09	0.04	0.77	0.000000	25
Mean	0.27	0.18	1.15	0.000072	54
Maximum	0.48	0.30	1.80	0.000591	80
Standard deviation	0.07	0.05	0.20	0.000102	17
Neossolos Litólicos (Leptosols) (n = 10)					
Minimum	0.06	0.01	0.87	0.000016	25
Mean	0.18	0.07	1.41	0.000067	59
Maximum	0.41	0.22	1.87	0.000165	79
Standard deviation	0.10	0.06	0.32	0.000053	17

Field capacity (FC) and permanent wilting point (PWP) determined by Richard's *chamber* method; bulk density (Bd) (volumetric method), and hydraulic conductivity of saturated soil (Ko) according to Teixeira et al. (2017); classification of soil into hydrologic groups according to the Curve Number (CN) method (SCS, 1993).

by *Neossolos Litólicos*, and *Cambissolos*. *Argissolos*, especially in the B horizon, show a lower infiltration capacity, resulting from the textural gradient formed by clay illuviation in the subsurface. The good consistency of soil parameters with reference to soil types is the high correlation between soil properties, such as bulk density, water content, and hydraulic conductivity.

Bulk density varied between 0.87 and 1.87 Mg m^{-3}. *Neossolos Litólicos* are young, shallow soils with primary minerals and higher values of bulk density. Consistently, on average, the *Latossolos* showed the lowest values (1.15 Mg m^{-3}).

The Curve Number (CN) showed large variations for *Neossolos Litólicos* and *Latossolos* (SD = 17), with a minimum of 25 and a maximum of 79 and 80, respectively.

Model assessment

Sensitivity analysis

All parameters tested showed a negative response to the OAT sensitivity analyses (Figure 4). The parameters that presented the highest sensitivity were Kc (SI = -1.83), CN (SI = -0.89), and the variables used in the TWS (SI = -0.75).

The parameters FC, Bd, and Depth had an equivalent effect on the calculation of groundwater using the BALSEQ, as these variables are used in the calculation of water available for evapotranspiration (TWS). Hence, the sensitivity index values for these three variables are represented in the same graph (SI = -0.748).
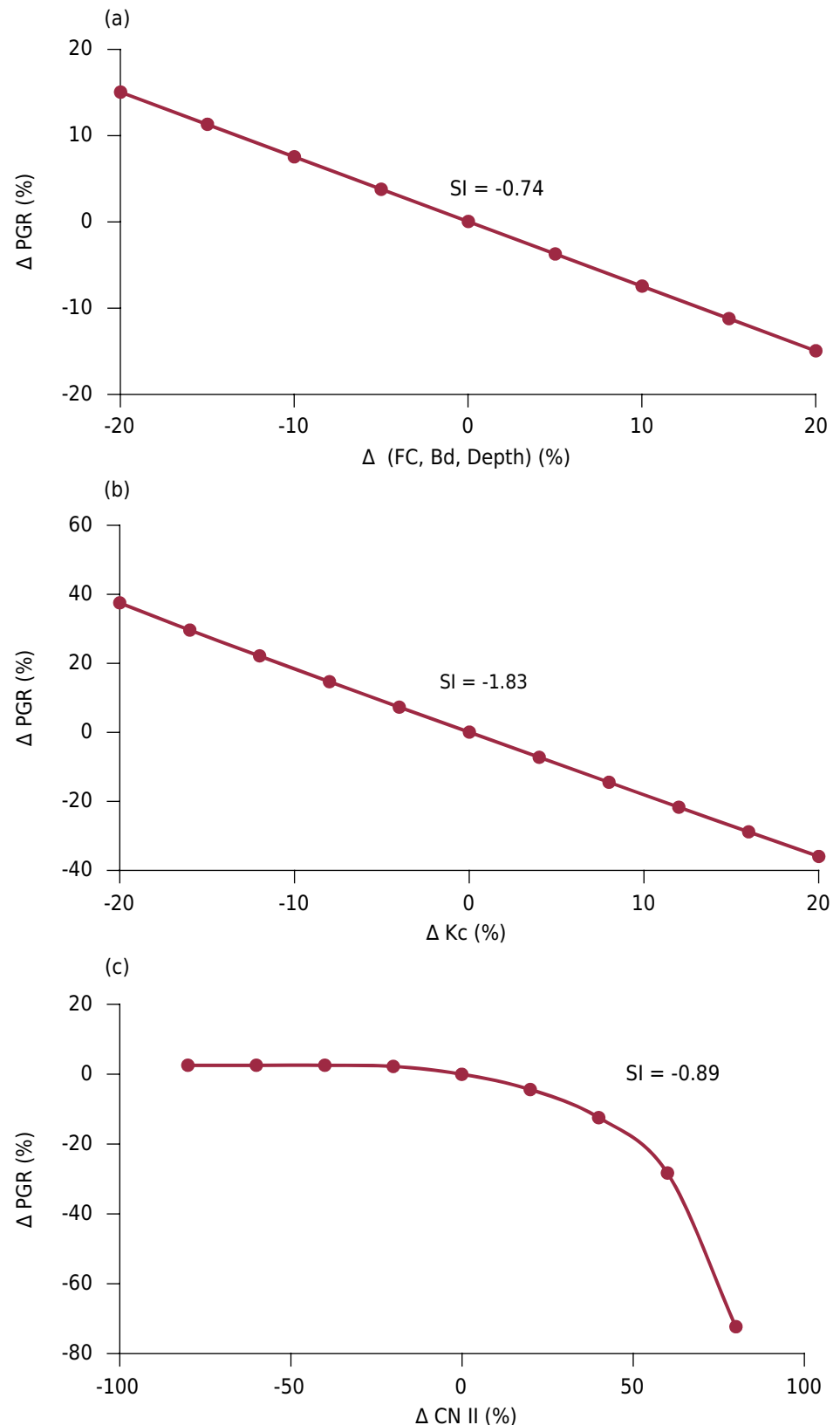
Figure 4. Variation in groundwater recharge due to the BALSEQ model parameters (a) field capacity, soil bulk density, and effective depth of the root system; (b) crop coefficient; and (c) curve number.

The TWS represents the water available for evapotranspiration. The higher the TWS, the higher the ETR and, consequently, the smaller the water available for recharge. The same can be observed for the crop coefficient (Kc), which varies according to the type of vegetation cover, directly influencing evapotranspiration. Therefore, the higher

the value of Kc, and TWS and its components, the lower the PGR, which explains the negative signal of the SI.

The variable CN also presented a negative SI value, mainly because an increase in CN causes a direct increase in surface water flow, with lower infiltration and PGR. It should be noted, however, that for a negative variation in CN, the change in PGR is small and shows a rapid decrease for variations above zero. Other parameters in the model showed a linear behavior.

The parameter Kc showed the highest sensitivity, followed by CN and the combined parameters Bd, Depth, and FC. Thus, these parameters should be used for model calibration. Similar results were obtained by Pontes et al. (2015, 2016), when testing the BALSEQ model in small catchments in the Serra da Mantiqueira.

Uncertainty analysis

Conceptual

The BALSEQ water budget equation was developed for situations in which there is no artificial recharge, no surface runoff affluent from adjacent areas, the water table is always below the soil depth subject to evapotranspiration, there are no impediment subsurface layers, and there are no preferential paths of water in the soil.

Each of these assumptions refers to possible uncertainties not computed by the model. There is no evidence to indicate the occurrence of artificial recharge or surface inflow affluent to the study area. Therefore, it can be discarded as a potential source of uncertainty.

On the other hand, the occurrence of saturated zones above the root depth, that is, places where the root system has access to the free aquifer, is an important source of uncertainty in the estimation of actual evapotranspiration, especially for forest vegetation. In this type of vegetation cover, the root system can reach saturated zones of the soil and absorb water to meet the evapotranspirative demand, which would cause an underestimation of the actual evapotranspiration. However, in this case, it would not directly affect the estimate of the potential recharge, since the saturated zone already constitutes water recharge.

Possible impermeable subsurface layers and preferential pathways are also important sources of uncertainties in PGR estimation. Given the high spatial variability of soil properties, the calculation of the effects of preferential pathways and dense layers is highly complex when included in hydrological models. Therefore, they are incorporated as uncertainties even in the most complex and robust models (Fatichi et al., 2016). One of the advantages of BALSEQ is the simplicity of its structure, which allows it to be adapted to different situations.

The preferential pathways of water flow are not traced and could be considered a limitation of the model. However, the criterion of classifying the recharge (e.g., direct, indirect, localized) is difficult to adhere to (Lerner, 1997), since the flow of infiltration through fissures, fractures, root channels, and galleries can potentially lead to direct recharge. In addition, there are other mechanisms, such as the variability of water percolation in soils with varying soil structure and pedoclimate, which can account for different recharge potentials and are not considered in the model.

Regarding the model parameters, related to soil data, it is worth highlighting the extensive work required for soil data surveys to characterize the physical soil properties (soil classification, bulk density, field capacity, and hydraulic conductivity), which contributes to minimize the uncertainties regarding soil parameters used in the model. Furthermore, FC and Bd present a small sensitivity index (-0.36) and therefore only slightly impact BALSEQ PGR estimations.

The CN is a critical variable for groundwater modeling, especially at values higher than 50, representing most values for the soils in the basin. Although CN has been developed for USA soils, it has been applied in hydrological models for tropical soils, particularly for soils in Minas Gerais (Silva et al., 2018).

The crop coefficient (Kc) is used to represent the evapotranspiration of each crop and varies according to the phenological stage of the crop. However, a fixed and constant Kc value was used for each crop due to the great complexity involved in taking this variation into account. This simplification introduces uncertainties in the estimation of evapotranspiration and, consequently, of PGR. On the other hand, the Kc values for the vegetation cover found in the basin are widely available in the literature, which reduces the degree of uncertainty regarding evapotranspiration.

Another difficulty in adopting depth values for trees is that the BALSEQ model does not work under conditions where the water table or even the capillary effect is in contact with roots, influencing evapotranspiration. Thus, it was decided to adopt an average fixed value for eucalyptus (Depth = 2.5 m) and for sub-deciduous seasonal forests (Depth = 3.5 m), which were lower than those suggested by other authors (Almeida and Soares, 2003).

Input data uncertainty

Regarding the modelling period (3 years), our choice was to exclude data of gauge stations with more than 15 % of missing data rather than filling the missing values with estimates based on the neighboring stations. Because the daily water balance uses a data series with missing data for some days (missing data were excluded, and the balance was continued on the next day), the number of days differed from season to season, which can result in non-systematic errors, especially with missing data from rainy seasons. To overcome this issue, we suggest that, as new data is collected, the groundwater model should be rerun for further estimates.

The accuracy of estimates is closely related to climatic and soil datasets used, particularly the climatic variables. Regarding the climatic variables, the number of rainfall gauge stations has a direct impact on the temporal and spatial accuracy of estimates. Rainfall is a key variable in the model since we use a water balance model in which the main input for the water budget is the amount of rain. This means that a large variation in input data will lead to increasing variability. In the Doce River Basin, the number of rainfall gauge stations was about three times that of meteorological stations.

Evapotranspiration is a variable typically estimated due to difficulties in finding measured data. The estimates use meteorological data and are hampered by missing data, which could not be estimated due to the short time series and the low number of neighbor stations. As already commented, the constraints for the time-scale used in this study were limited by the data of the meteorological stations.

Once more data is collected, a more accurate estimate will be possible, and new stations will lead to a decrease in spatial variation.

Model validation

For this comparison, it was considered that soil moisture at a depth of 1 m exceeded the water content at field capacity (FC) at that same depth, representing groundwater (Figure 5).

Qualitatively, it can be stated that the estimates of BALSEQ responded correctly to the recharge events, since the events were identified simultaneously by both methods and in both areas, agroforestry and conventional agriculture (Figure 6). The volume of recharge could not be quantified using the sensor method, since a volume of control should be
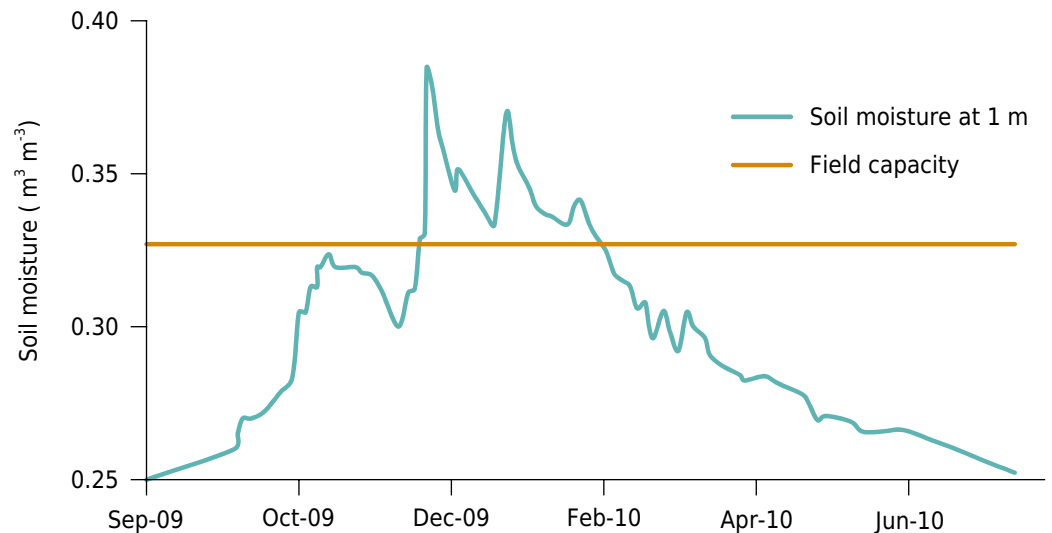
Figure 5. Soil moisture and water content at field capacity (FC) at a depth of 1 m in the agroforestry system. The area above the red line corresponds to the total volume of recharge (water per volume of soil).

adopted, which would dictate the volume of recharge infiltrated. Therefore, the analysis of validation was only of a qualitative nature.

Another observation that corroborates with the results of BALSEQ, especially for forested areas, is that in the agroforestry system, there was no recharge event between April and May, contrary to the conventional agricultural area. One possible explanation for this would be that after the rainy period (December to February), the trees in the agroforestry system increased the demand for water up to a depth of 1.00 m, leading to decreased soil moisture. In April, at the beginning of the rainy season, the subsoil under the agroforestry system should be drier, not exceeding field capacity. In addition, the rapid tree growth, after the first rains, increased the demand for water for evapotranspiration, with rapid water consumption at subsequent rainfall events.

Estimates of water balance using the BALSEQ model

The calculation of the water balance over the two-year period, using the BALSEQ model, was continuously run from the first day of the hydrological year of 2007 to the end of the hydrological year of 2009 for each point of the 123 soils. This means that the water balance considering runoff, evapotranspiration, and the potential recharge summed up 100 % of the accumulated rainfall at the end of the two-year period. The annual balance indicated the annual pattern of groundwater recharge, whereas the total period allowed to observe the long-term trend.

The statistical summary of the water balance variables (rainfall, evapotranspiration, runoff, and groundwater recharge) is presented in table 3. Overall, considering the annual and biannual time length, the highest percentage of rainfall was evapotranspirated (53-73 %), while surface runoff accounted for only 4-11 %. Groundwater recharge varied from 32-35 % of the total precipitation, while annual rainfall varied from a minimum of around 500 mm to a maximum of 1,662 and 1,998 mm.

During the two-year period, the average recharge was 757 mm, which accounts for 32 % of the rainfall. Recharge was considerably smaller in the first year compared to the second year. On average, 25 and 35 % of the rainfall were converted to recharge in the first and second year, respectively. The factors that contributed to the lower percentage of water percolating to the deeper soil layers in the first year were the lesser volume of water rainfall and the higher volume of evapotranspirated water.
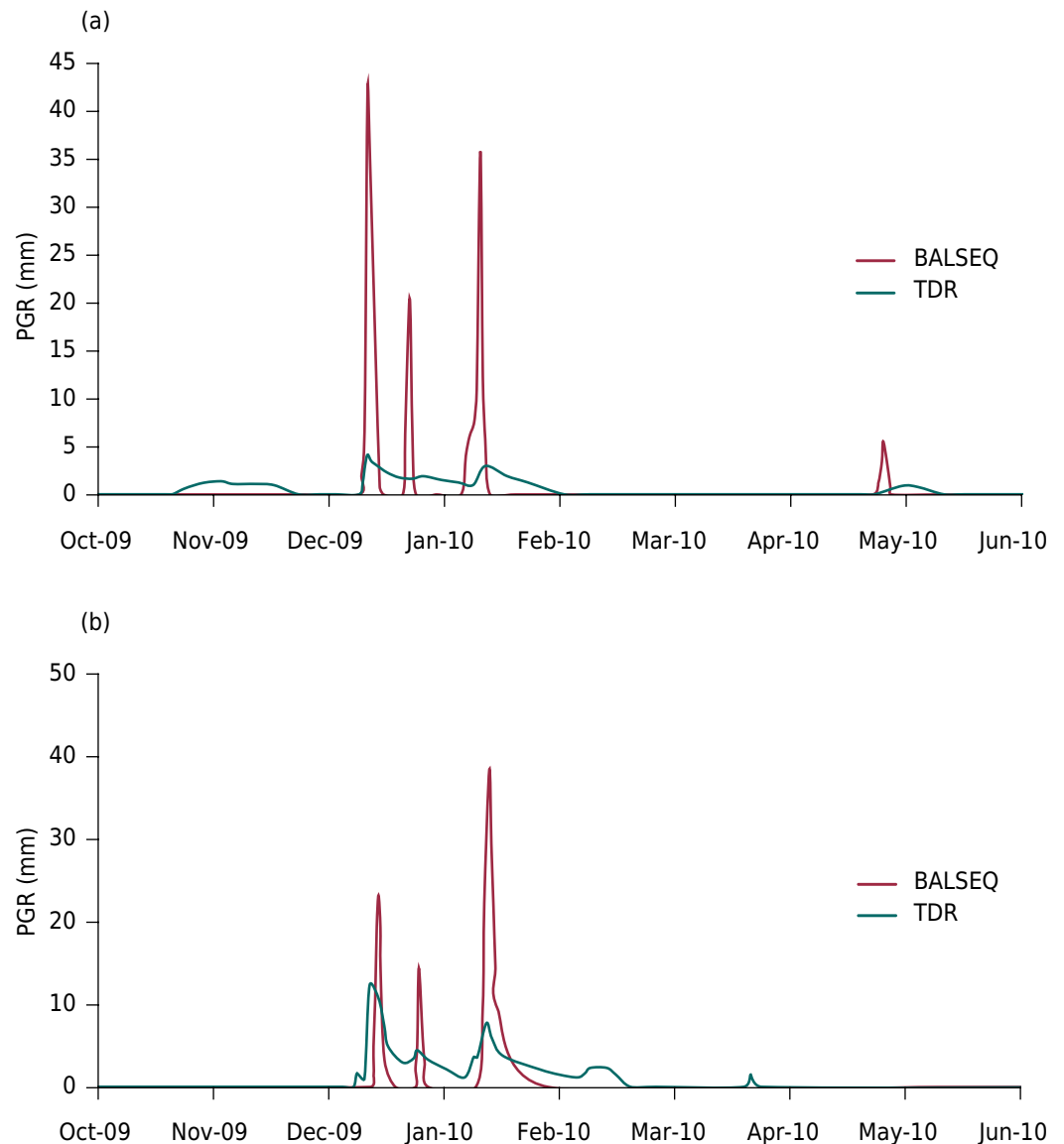
Figure 6. Comparison between recharge calculated with the BALSEQ model and recharge measured with the sensor in a conventional coffee plantation (a) and an agroforestry system (b) (PGR = potential groundwater recharge; Rp = potential recharge with BALSEQ; R sensor = recharge with the TDR sensor).

The volume of groundwater recharge observed for the Doce River Basin was similar to that observed in other studies in nearby areas. For example, Paiva (2006) estimated the recharge of the Piranga River watershed, a tributary of the Doce River, and found a potential recharge of 33 % of rainfall, using the water balance method. Also, using the water table method, a study on a watershed in the municipality of Araguari, Minas Gerais, showed a recharge of 38 % of the mean rainfall (Velasquez et al., 2007).

In a study on a karst area, using the BALSEQ model, Camargo et al. (2011), observed a higher volume of recharge (43 % of the annual rainfall) than that determined in the present study, which is most likely a result of the high rock porosity in karst areas, facilitating the percolation of water to the deeper soil layers.

Pontes et al. (2016) compared recharge estimates from the BALSEQ model with recharge calculated via the base runoff for a watershed of mountainous relief of a *Cambissolos* domain and a watershed of suave relief and intensive occupation. The percentage of rainfall quantified as recharge was 44 and 25 %, and the agreement of estimates

Table 3. Statistical summary of the water balance for the two-year period (2007-2009) and the hydrological years 2007-2008 and 2008-2009 in the Doce River Basin, obtained with the BALSEQ model for the point data estimates

Statistic	Rainfall	PGR	PGR	SR	SR	ETR	ETR
		%	mm	%	mm	%	mm
			2007-2009				
Minimum	1,510	0	0	0	0	26	681
Mean	2,270	32	757	8	188	60	1,325
Maximum	3,487	72	2,123	37	992	100	2,454
SD	487	18	518	7	170	19	392
CV (%)	21	56	68	82	91	32	30
			2007-2008				
Minimum	537	0	0	0	0	28	290
Mean	896	25	250	4	38	73	625
Maximum	1,662	68	939	25	259	133	1,352
SD	248	19	229	5	50	24	213
CV (%)	28	74	92	120	130	33	34
			2008-2009				
Minimum	458	0	0	0	0	22	316
Mean	1,374	35	508	11	149	53	700
Maximum	1,998	75	1,267	43	775	98	1,258
SD	307	18	326	8	128	18	214
CV (%)	15	25	26	19	17	19	17

PGR = potential groundwater recharge; SR = superficial runoff; ETR = evapotranspiration, values in (%) are given in percentage of rainfall; SD = standard deviation; CV = coefficient of variation.

calculated with the BALSEQ to estimates obtained with the base runoff method was considered satisfactory.

Overall, among soil classes an average of 27-48 % of the rainfall contributed to groundwater, with the lowest rate found for *Argissolos* and the largest for *Neossolos*. *Latossolos* occurs in 66 % of the basin and therefore feature areas of importance to maintain water recharge, with low levels of surface runoff and a moderate rate of recharge. On the other hand, *Argissolos* showed the lowest potential for recharge and the second largest for the generation of surface runoff; this soil type is therefore more susceptible to erosion and losses of soil and water (Figure 7).

Neossolos Litólicos, although presenting the highest potential for recharge, with up to 48 % of rainfall turning into groundwater, only covered 3 % of the basin (~5,000 km²); nevertheless, such areas are of high importance for the preservation and maintenance of groundwater levels.

Cambissolos showed the highest percentage of runoff (13 %); this was expected, as *Cambissolos* typically occur on steep reliefs, contributing to increased runoff rates (Figure 7).

Spatial prediction of groundwater recharge

The potential groundwater was spatially predicted using environmental maps considering the interactions of climate, lithology, morphology, soil type, and vegetation, based on the Random Forest model. The parameters of the Random Forest model which gave the best results were as follows: number of trees (ntree = 1,000), number of predictors sampled for splitting at each node (mtry = 2), and the number of divisions at the end of
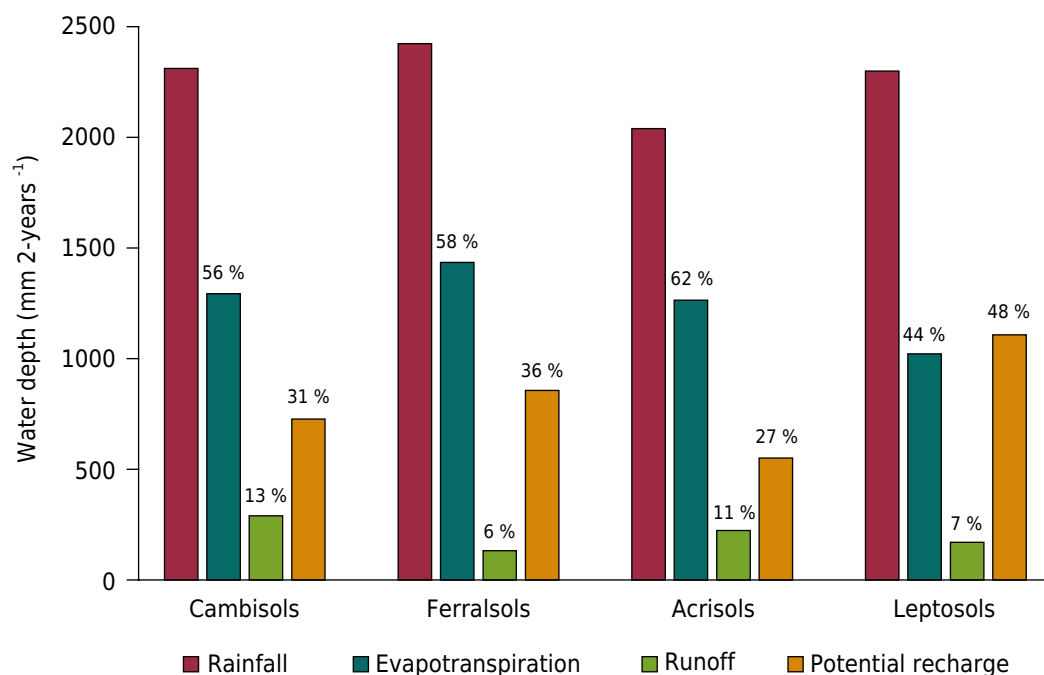
Figure 7. Mean values of the water balance components for the different soil classes for the two-year period from 2007-2009; the values in percentage refer to the (%) of rainfall in the water balance components (Potential recharge, Runoff, Evapotranspiration).

each tree (nodesize = 5). These parameters were defined after simulation using values of the intervals indicated by Breiman (2002).

The importance of each predictor variable used in the model, measured by its contribution to decreasing the MSE of prediction, is presented in figure 9.

Rainfall appears as the main driver of recharge, followed by evapotranspiration, lithology, and elevation. Rainfall contributed to 16-17 % of MSE decrease. Solar radiation, temperature, and the vegetation index EVI contributed with similar importance, while the soil map showed the lowest contribution (maximum 7 %), which can be related to the level of generalization of the map (Figure 8). The soil map was simplified based on the first component of the map unit, which corroborates with its lower importance on spatial prediction of the recharge.

Previous studies on mapping PGR conducted by Naghibi et al. (2017) and Rahmati et al. (2016) showed elevation as the main predictor, and in both studies, the Random Forest model performed better than the other models evaluated. In the study of Naghibi et al. (2017), for an area of 1,524 km^2 in Iran, the authors observed that morphometric maps, such as those showing elevation and slope, were highly important for the Random Forest model compared to lithology and land use maps. This study compared Random Forest performance with Kernel-based support vector machine methods.

Rahmati et al. (2016) mapped PGR classified in categorical classes (low, moderate, and high) for a catchment of 226 km^2. Their results showed elevation as the most important predictor, contributing to 50 % of the decrease in accuracy. Similar to the present study, the lithological map showed a relatively low contribution. Rahmati et al. (2016) compared the Random Forest model with a model of maximum entropy, which performed significantly better.

In our study, we obtained an R^2 of 0.22 and 0.27 for predicting groundwater recharge for the first and second year, respectively, and of 0.27 for the two-year period (Table 3). The accuracy of prediction based on the external validation dataset showed a higher

accuracy for the two-year period, with an R^2 of 0.34 and a lower RMSE. For the annual recharge, the accuracy showed little change (Table 4).

Considering the scale and resolution of maps used for the spatial prediction of recharge, as well as the natural limitation of soil data and climatic variables, we consider the accuracy of the prediction to be satisfactory. In addition, we highlight the fact that the regional scale of the Doce River Basin, which comprises a large diversity of environmental conditions, contributes to the high prediction variability.

A visual analysis of the spatial distribution of the groundwater recharge across the basin showed a similar annual recharge pattern over the studied period (Figure 9). The maps for the two-year period and the maps of each year separately depicted higher recharge rates on the western border of the basin (Espinhaço Range), where rainfall and elevation levels are higher. Also, the headwaters of the Doce River at the south are also areas of high potential recharge on the plateau.

Groundwater recharge rates are directly related with rainfall distribution. Areas of higher recharge (>500 mm year) are concentrated on the western side of the basin, where rainfall above 1,500 mm (Figures 2b and 2c). Temperature, on the other hand, is generally lower in areas with a maximum groundwater recharge. Maximum monthly temperature varies from 21-30 °C. The forest cover is also more preserved on the western side of the basin, as well as at the headwaters of the basin to the south (Figure 2d). The combination of higher elevation and rainfall with lower temperature and a denser vegetation is therefore favorable for soil water percolation and groundwater recharge. The identification (stratification) of zones of the basin into areas of different PGR levels can be a valuable tool for land planning.

Using the quantile interval, the groundwater maps were classified into four different groundwater potential zones, including low, medium, high, and very high groundwater
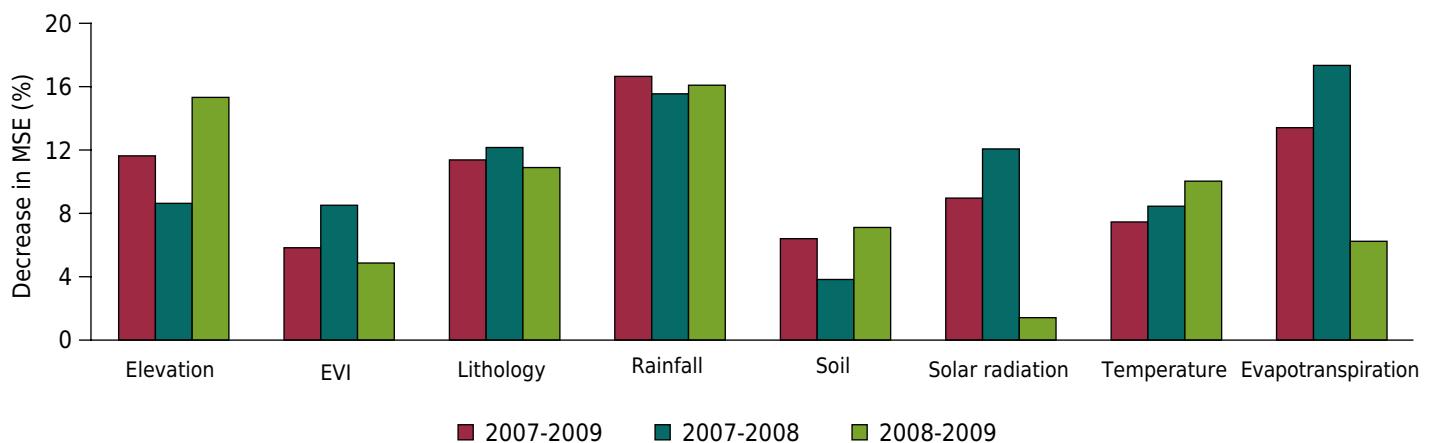

Figure 8. Percentage of covariate contributions in the spatial prediction of groundwater recharge in the Doce River Basin.

Table 4. Statistical distribution of the groundwater recharge predicted for the Doce River Basin for two hydrological years

Period	Min	Mean	Max	SD	CV	Total	R^2	RMSE	R^2	RMSE
							Training (n = 95)		Validation (n = 28)	
			m		%	m³				
2007-2009	107	880	1767	426	48	52,867	0.28	445	0.32	403
2007-2008	36	398	811	209	53	17,484	0.30	195	0.21	200
2008-2009	54	500	1001	250	50	35,410	0.22	283	0.30	273

Min = minimum; Max = maximum; SD = standard deviation; CV = coefficient of variation; R^2 = R-squared; RMSE = root mean squared error.
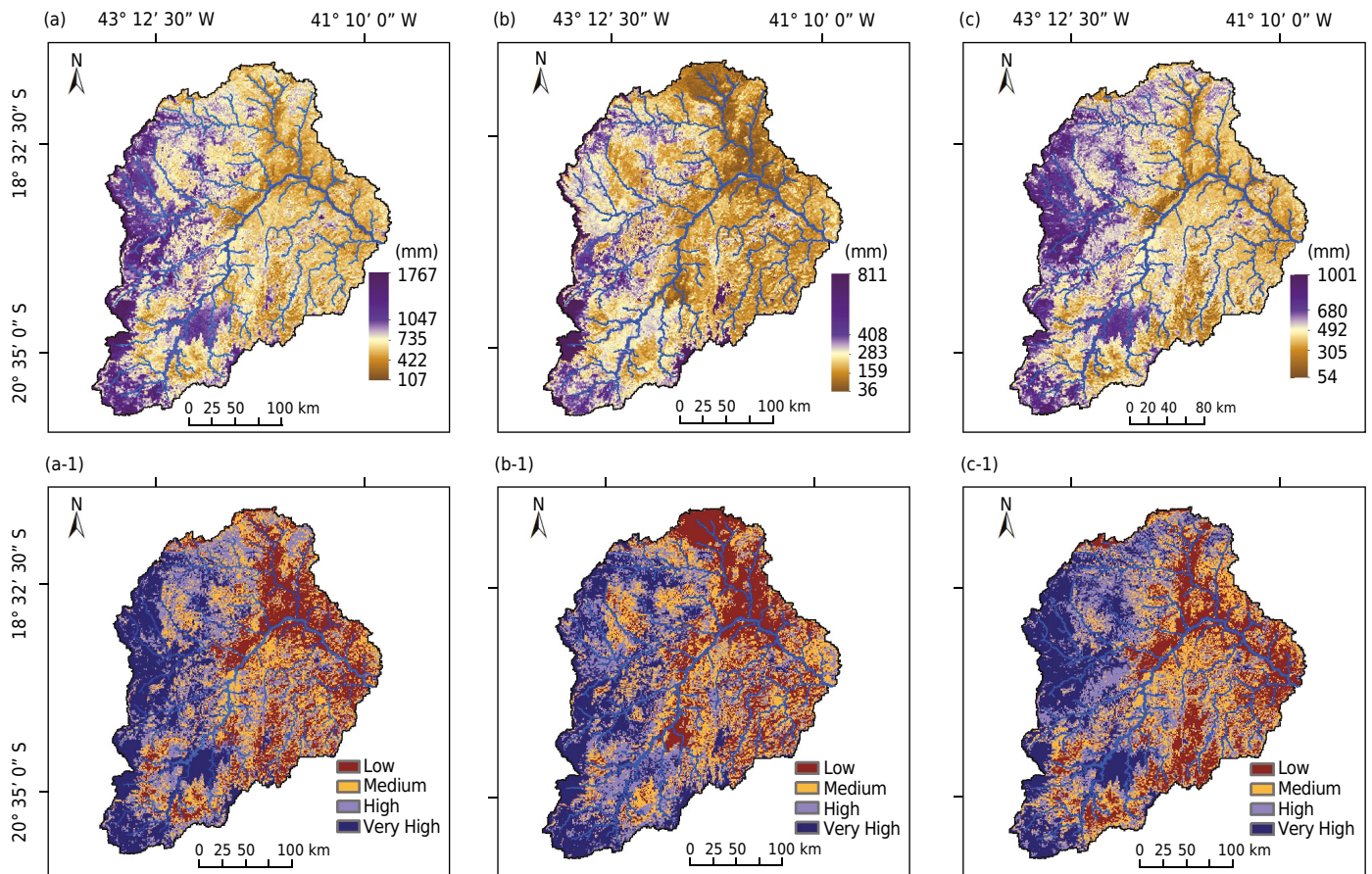
Figure 9. Potential groundwater recharge (PGR) in the Doce River Basin in the years 2007/2009 (a), 2007/2008 (b), and 2008/2009 (c) and respective maps of the PGR zones.

potential zones (Figure 9). In this regard, the use of the quantile method for stratifying the PGR into classes has effectively been applied by Nampak et al. (2014), Naghibi and Pourghasemi (2015), and Razandi et al. (2015).

CONCLUSIONS

The estimation of the water balance for the Doce River Basin, using the BALSEQ model, is feasible with a combination of soil, vegetation, precipitation, and evapotranspiration data. Hence, the amount of water that can potentially percolate and reach the water table can effectively be quantified and modelled.

An innovative aspect of this study is the spatial prediction of groundwater recharge using the statistical Random Forest model, which differs from the commonly used methods based on the weighted average method, using conventional soil maps.

The spatial prediction of groundwater recharge using the Random Forests model showed a moderate performance, which is compatible with the scale of the maps used as predictors.

Both models, the BALSEQ model to estimate the recharge and the Random Forest model for the spatial prediction of the recharge, showed feasibility of updating the groundwater map when new datasets were available.

The maps of groundwater recharge of unconfined aquifers could be used to guide regional land and water use planning to help protecting the groundwater resources in the Doce River Basin.

In future studies, data on infiltration and groundwater percolation, obtained at the field level, should be considered for the validation of estimates, ensuring a greater reliability of the groundwater recharge at the basin scale.

ACKNOWLEDGEMENTS

REFERENCE

Alencar CAB, Cunha FF, Martins CE, Cóser AC, Rocha WSD, Araújo RAS. Irrigação de pastagem: atualidade e recomendações para uso e manejo. Rev Bras Zootec. 2009;38:98-108. https://doi.org/10.1590/S1516-35982009001300012

Allen RG, Pereira LS, Raes D, Smith M. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO irrigation and drainage paper 56. Rome: Food and Agriculture Organization of the United Nations; 1998 [cited 2017 Aug 7]. Available form: www.fao.org/docrep/X0490E/x0490e0b.htm.

Almeida AC, Soares JV. Comparação entre uso de água em plantações de *Eucalyptus grandis* e floresta ombrófila densa (Mata Atlântica) na costa leste do Brasil. Rev Arvore. 2003;27:159-70. https://doi.org/10.1590/S0100-67622003000200006

Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLM, Sparovek G. Köppen's climate classification map for Brazil. Meteorol Z. 2014;22:711-28. https://doi.org/10.1127/0941-2948/2013/0507

Arcova FCS, Cicco V, Rocha PAB. Precipitação efetiva e interceptação das chuvas por floresta de Mata Atlântica em uma microbacia experimental em Cunha - São Paulo. Rev Arvore. 2003;27:257-62. https://doi.org/10.1590/S0100-67622003000200014

Brandão VS, Cecílio RA, Pruski FF, Silva DD. Infiltração da água no solo. 3. ed. atual ampl. Viçosa, MG: Editora UFV; 2006.

Breiman L. Random forest. Mach Learn. 2001;45:5-32. https://doi.org/10.1023/A:1010933404324

Breiman L. Manual on setting up, using, and understanding random forest. Version 3.1. Berkeley: Statistics Department University of California; 2002.

Brito GLM, Schuster HD, Srinivasan VS. Estimation of annual ground water recharge in the sedimentary basin of the river Peixe, Paraíba, Brazil. In: Zhang C, Tang H, editors. Advances in water resources and hydraulic engineering - Proceedings of 16th IAHR-APD Congress and 3rd Symposium of IAHR-ISHS. Berlin: Springer Heidelberg; 2009. p. 269-74.

Camargo RA, Couto EA, Alves LGS, Calijuri ML, Silva MDFM. Aplicação de modelo de balanço hídrico BALSEQ para estimativa da infiltração profunda em região cárstica. Rev Ambient Agua. 2011;6:303-17. https://doi.org/10.4136/ambi-agua.456

Chachadi AG, Chaudri BS, Naronha L, Lobo-Ferreira JP. Estimation of surface runoff and groundwater recharge in Goa mining area using daily sequential water balance model - BALSEQ. Hydrol J. 2004;27:1-15.

Chachadi AG, Raikar PS, Lobo-Ferreira JP, Oliveira MM. GIS and mathematical modelling for the assessment of groundwater vulnerability to pollution: application to an indian case study area in Goa. Lisboa: Laboratório Nacional de Engenharia Civil / Laboratório Nacional de Engenharia Civil; 2001.

Chung I-M, Kim N-W, Lee J, Sophocleous M. Assessing distributed groundwater recharge rate using integrated surface water-groundwater modelling: application to Mihocheon watershed, South Korea. Hydrogeol J. 2010;18:1253-64. https://doi.org/10.1007/s10040-010-0593-1

Crosbie RS, Peeters LJM, Herron N, McVicar TR, Herr A. Estimating groundwater recharge and its associated uncertainty: use of regression kriging and the chloride mass balance method. J Hydrol. 2018;561:1063-80. https://doi.org/10.1016/j.jhydrol.2017.08.003

Cupolillo F. Diagnóstico hidroclimatológico da bacia do rio Doce. Alemanha: Novas Edições Acadêmicas; 2015.

de Vries JJ, Simmers I. Groundwater recharge: an overview of process and challenges. Hydrogeol J. 2002;10:5-17. https://doi.org/10.1007/s10040-001-0171-7

Fatichi S, Vivoni ER, Ogden FL, Ivanov VY, Mirus B, Gochis D, Downer CW, Camporese M, Davison JH, Ebel B, Jones N, Kim J, Mascaro G, Niswonger R, Restrepo P, Rigon R, Shen C, Sulis M, Tarboton D. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. J Hydrol. 2016;537:45-60. https://doi.org/10.1016/j.jhydrol.2016.03.026

Fundação Estadual de Meio Ambiente/Universidade Federal de Viçosa/Secretaria Estadual de Meio Ambiente - FEAM/UFV/SEMA. Levantamento de solos e aptidão agrícola das terras da Bacia do Rio Doce, Estado de MG - Relatório e Mapa de Solos, escala 1:500.000. Belo Horizonte: FEAM/UFV/SEMA; 2011.

Gaur S, Chahar BR, Graillot D. Combined use of groundwater modeling and potential zone analysis for management of groundwater. Int J Appl Earth Obs. 2011;13:127-39. https://doi.org/10.1016/j.jag.2010.09.001

Ghandhari A, Alavi Moghaddam SMR. Water balance principles: a review of studies on five watersheds in Iran. J Environ Sci Technol. 2011;4:465-79. https://doi.org/10.3923/jest.2011.465.479

Golkarian A, Naghibi AS, Kalantar B, Pradhan B. Groundwater potential mapping using C5.0, random forest, and multivariate adaptive regression spline models in GIS. Environ Monit Assess. 2018;190:149. https://doi.org/10.1007/s10661-018-6507-8

Healy RW, Cook PG. Using groundwater levels to estimate recharge. Hydrogeol J. 2002;10:91-109. https://doi.org/10.1007/s10040-001-0178-0

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol. 2005;25:1965-78. https://doi.org/10.1002/joc.1276

Hirata R, Conicelli BP. Groundwater resources in Brazil: a review of possible impacts caused by climate change. An Acad Bras Cienc. 2012;84:297-312. https://doi.org/10.1590/s0001-37652012005000037

Hirata R, Zoby JLG, Oliveira FR. Ground water: strategic or emergency reserve. In: Bicudo CM, Tundisi JG, Scheuenstuhl MCB, editors. Waters of Brazil: Waters of Brazil. Cham: Springer; 2017. p. 119-36.

Hugman R, Stigter T, Costa L, Monteiro JP. Numerical modelling assessment of climate-change impacts and mitigation measures on the Querença-Silves coastal aquifer (Algarve, Portugal). Hydrogeol J. 2017;25:2105-21. https://doi.org/10.1007/s10040-017-1594-0

Jarvis A, Reuter HI, Nelson A, Guevara E. Hole-filled seamless SRTM data V4. International Centre for Tropical Agriculture (CIAT); 2008. Available from: http://srtm.csi.cgiar.org.

Kinzelbach W, Aeschbach W, Alberich C, Baba GI, Beyerle U, Brunner P, Chiang W-H, Rueedi J, Zoellmann K. A survey of methods for groundwater recharge in arid and semi-arid regions. Nairobi: United Nations Environment Programme; 2002.

Kobayashi ES. Consumo de água e produtividade de cafeeiros arábica na região de Mococa, SP [dissertação]. Campinas: Instituto Agronômico de Campinas; 2007.

Lerner DN, Issar AS, Simmers I. Groundwater recharge: a guide to understanding and estimating natural recharge. Hannover: Heise; 1990.

Lerner DN. Groundwater recharge. In: Saether OM, de Caritat P, editors. Geochemical processes, weathering and groundwater recharge in catchments. Rotterdam: AA Balkema. 1997. p. 109-50.

Liaw A, Wiener M. Classification and regression with random forests. R package versión 4.6-14; 2018. Available from: https://CRAN.R-project.org/package=randomForest.

Lobo-Ferreira JP, Delgado-Rodrigues J. BALSEQ - a model for the estimation of water balances, including aquifer recharges, requiring scarce hydrologic data. In: Simmers I, editor. Estimation of natural groundwater recharge. Dordrecht: Reidel; 1998. p. 313-20.

Lobo-Ferreira JP, Leitão TE, Oliveira MM. Portugal's river basin management plans: groundwater innovative methodologies, diagnosis, and objectives. Environ Earth Sci. 2015;73:2627-44. https://doi.org/10.1007/s12665-014-3462-8

Lobo-Ferreira JP. Mathematical model for the evaluation of the recharge of aquifers in semiarid regions with scarce (lack) hydrogeological data. In: Verruijt A, Barends FBJ. Proceedings of Euromech 143; Setp 1981; Rotterdam. Rotterdam: AA Balkema; 1981. p. 2-4.

McCuen RH, Snyder WM. Hydrologic modeling: statistical methods and applications. New Jersey: Prentice-Hall; 1986.

Moderate Resolution Imaging Spectroradiometer - MODIS. MOD13Q1, EVI - Enhanced vegetation index. South Dakota: NASA EOSDIS land processes DAAC/USGS Earth Resources Observation and Science (EROS); 2009 [cited 2018 Jul 10]. Available from: https://lpdaac.usgs.gov/data_access/glovis.

Moura AESS, Correa MM, Silva ER, Ferreira RLC, Figueiredo AC, Possas JMC. Interceptação das chuvas em um fragmento de floresta da Mata Atlântica na bacia do Prata, Recife, PE. Rev Arvore. 2009;33:461-9. https://doi.org/10.1590/S0100-67622009000300008

Naghibi SA, Ahmadi K, Daneshi A. Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. Water Resour Manag. 2017;31:2761-75. https://doi.org/10.1007/s11269-017-1660-3

Naghibi SA, Pourghasemi HR, Dixon B. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environ Monit Assess. 2016;188:44. https://doi.org/10.1007/s10661-015-5049-6

Naghibi SA, Pourghasemi HR. A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. Water Resour Manag. 2015;29:5217-36 https://doi.org/10.1007/s11269-015-1114-8

Nampak H, Pradhan B, Manap MA. Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. J Hydrol. 2014;513:283-300. https://doi.org/10.1016/j.jhydrol.2014.02.053

Neves ACO, Nunes FP, Carvalho FA, Fernandes GW. Neglect of ecosystems services by mining, and the worst environmental disaster in Brazil. Nat Conservacao. 2016;14:24-7. https://doi.org/10.1016/j.ncon.2016.03.002

Noori SMS, Ebrahimi K, Liaghat A-M, Hoorfar A-H. Comparison of different geostatistical methods to estimate groundwater level at different climatic periods. Water Environ J. 2013;27:10-9. https://doi.org/10.1111/j.1747-6593.2012.00321.x

Oliveira MJPM. Recarga de águas subterrâneas: métodos de avaliação [tese]. Lisboa: Universidade de Lisboa; 2004.

Paiva GP. Estimativa de recarga em uma unidade detrítica de uma micro-bacia do Rio Piranga (MG) [dissertação]. Viçosa, MG: Universidade Federal de Viçosa; 2006.

Paralta EA, Oliveira MM, Lubczynski MW, Ribeiro LF. Avaliação da recarga do Sistema Aquífero dos Gabros de Beja segundo critérios múltiplos - disponibilidades hídricas e implicações agro-ambientais. In: Actas do VI Simpósio de Hidráulica e Recursos Hídricos dos Países de Língua Oficial Portuguesa - SILUSBA; novembro 2003; Cabo Verde. Lisboa: Associação Portuguesa dos Recursos Hídricos; 2003. p. 501-16.

Pluntke T, Pavlik D, Bernhofer C. Reducing uncertainty in hydrological modelling in a data sparse region. Environ Earth Sci. 2014;72:4801-16. https://doi.org/10.1007/s12665-014-3252-3

Pontes LM, Coelho G, Mello CR, Silva AM, Oliveira GC, Viola MR. Avaliação de modelo de balanço hídrico com base na estimativa da recarga potencial. Rev Ambient Agua. 2016;11:915-28. https://doi.org/10.4136/ambi-agua.1856

Pontes LM, Coelho G, Mello CR, Silva AM, Oliveira GC. Análise de sensibilidade e avaliação da estrutura do modelo BALSEQ em condições distintas de clima, solo e vegetação. RBRH. 2015;20:46-54. https://doi.org/10.21168/rbrh.v20n1.p46-54

Projeto de conservação e utilização sustentável da diversidade biológica brasileira - Probio. Mapeamento da cobertura vegetal nativa dos biomas brasileiros: Bioma Mata Atlântica - Ano base 2002, Mapa digital, escala 1:250.000. Brasília: DF: Probio/MMA/UFRJ/IESB/UFF; 2006.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2016. Available from: http://www.R-project.org/.

Rahmati O, Pourghasemi HR, Melesse AM. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. Catena. 2016;137:360-72. https://doi.org/10.1016/j.catena.2015.10.010

Razandi Y, Pourghasemi HR, Neisani NS, Rahmati O. Application of analytical hierarchy process, frequency ratio, and certainty factor models for groundwater potential mapping using GIS. Earth Sci Inform. 2015;8:867-83. https://doi.org/10.1007/s12145-015-0220-8

Resende Filho MA, Correa JSO, Torres MO. Water pricing in Brazil: successes, failures, and new approaches. In: Dinar A, Pochat V, Albiac-Murillo J, editors. Water pricing experiences and innovations. Cahm: Springer; 2015. p. 41-61.

Richards LA. Methods of measuring moisture tension. Soil Sci. 1949;68:95-112. https://doi.org/10.1097/00010694-194907000-00008

Rodriguez-Galiano V, Mendes MP, Garcia-Soldado MJ, Chica-Olmo M, Ribeiro L. Predictive modeling of groundwater nitrate pollution using random forest and multisource variables related to intrinsic and specific vulnerability: a case study in an agricultural setting (Southern Spain) Sci. Total Environ. 2014;476-477:189-206. http://dx.doi.org/10.1016/j.scitotenv.2014.01.001

Rushton KR. Numerical and conceptual models for recharge estimation in arid and semi-arid zones. In: Simmers I, editor. Estimation of natural groundwater recharge. Dordrecht: Reidel; 1998. p. 223-38.

Sahoo M, Kasot A, Dhar A, Kar A. On predictability of groundwater level in shallow wells using satellite observations. Water Resour Manage. 2018;32:1225-44. https://doi.org/10.1007/s11269-017-1865-5

Scanlon BR, Healy RW, Cook PG. Choosing appropriate techniques for quantifying groundwater recharge. Hydrogeol J. 2002;10:18-39. https://doi.org/10.1007/s10040-0010176-2

Serviço Geológico do Brasil - CPRM. Carta geológica do Brasil ao milionésimo: folha SE-23, escala 1:1.000.000. Brasília: Programa Geologia do Brasil; 2004.

Silva LA, Silva AM, Coelho G, Pinto LC. Soil map units of Minas Gerais State from the perspective of Hydrologic Groups. Rev Ambient Agua. 2018;13:e2118. https://doi.org/10.4136/ambi-agua.2118

Simmers I. Estimation of natural groundwater recharge. Dordrecht: Reidel; 1998.

Soil Conservation Service - SCS. National Engineering Handbook Hydrology. Washington, DC: United States Department of Agriculture, Natural Resources Conservation Service; 1993.

Sophocleous M. Groundwater recharge estimation and regionalization: the Great Bend Prairie of central Kansas and its recharge statistics. J Hydrol. 1992;137:113-40. https://doi.org/10.1016/0022-1694(92)90051-v

Teixeira PC, Donagemma GK, Fontana A, Teixeira WG. Manual de métodos de análise de solo. 3. ed. rev. ampl. Brasília, DF: Embrapa; 2017.

Trabucco A, Zomer RJ. Global Soil Water Balance Geospatial Database. CGIAR Consortium for Spatial Information; 2010 [cited 2017 Jul 12]. Available from: http://www.cgiar-csi.org.

Udawatta RP, Anderson SH, Motavalli PP, Garrett HE. Calibration of a water content reflectometer and soil water dynamics for an agroforestry practice. Agroforest Syst. 2011;82:61-75. https://doi.org/10.1007/s10457-010-9362-3

Velasquez LNM, Carvalho Filho CA, Camargos CC, Branco OEA, Rodrigues PCH, Minardi PP, Migliorini RB, Cota SDS, Bomtempo VL, Bertol G, Diaz JI. Avaliação dos recursos hídricos do Sistema Aqüífero Guarani no município de Araguari, Minas Gerais, Brasil. In: Fondo de Universidades del Proyecto para la Protección Ambiental y Desarollo Sostenible del Sistema Acuífero Guaraní, organizador. Acuífero Guaraní Avances en el conocimiento para su gestión sustentable. Montevidéo: ALHSUD; 2007. p. 24-37.

Wheeler DC, Nolan BT, Flory AR, DellaValle CT, Ward MH. Modeling groundwater nitrate concentrations in private wells in Iowa. Sci Total Environ. 2015;536:481-8. https://doi.org/10.1016/j.scitotenv.2015.07.080

Xu T, Valocchi AJ. Data-driven methods to improve baseflow prediction of a regional groundwater model. Comput Geosci. 2015;85:124-36. https://doi.org/10.1016/j.cageo.2015.05.016

Zoby J, Matos B. Águas subterrâneas no Brasil e sua inserção na Política Nacional de Recursos Hídricos. In: Anais do XII Congresso Brasileiro de Águas Subterrâneas. São Paulo: Associação Brasileira de Águas Subterrâneas 2002.