Applied Geography 94 (2018) 272–281

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Environmental factors and groundwater behavior in an agricultural experimental basin of the Brazilian central plateau

Leandro de Almeida Salles[a,c,*], Jorge Enoch Furquim Werneck Lima[b], Henrique Llacer Roig[c], Juaci Vitória Malaquias[b]

[a] Instituto Brasília Ambiental – IBRAM, SEPN 511 – Bloco C – Edifício Bittar, CEP: 70.750-543, Brasília – DF, Brazil
[b] Empresa Brasileira de Pesquisa Agropecuária, Centro de Pesquisa Agropecuária dos Cerrados – EMBRAPA/CPAC, BR 020, Km 18, Caixa 08223, Brazil
[c] Instituto de Geociências, Universidade de Brasilia, Campus Universitário Darcy Ribeiro, Asa Norte, CEP: 70910-900, Brasilia, DF, Brazil

ARTICLE INFO

Keywords:
Geostatistics
Statistical modeling
Groundwater recharge/water budget
Geographic information system

ABSTRACT

Knowledge of groundwater storage in tropical soils of the Cerrado biome is relevant because of its importance for stream flow regime in Brazil's central plateau, where rivers are perennial during the drought period of the year. The purpose of this study was to: (i) identify environmental variables capable of explaining the amplitude of groundwater variation in points of an agricultural basin in the Cerrado; and (ii) develop a statistical model to simulate the amplitude of water table variation from the chosen variables. The study was developed with monthly data obtained from 35 piezometric wells located at the Upper Jardim experimental river basin (105 km^2), collected between 17/02/2004 and 01/10/2012. Two multivariate analysis methodologies were used: (i) Ordinary Least Squares regression (OLS); and (ii) Geographical Weighted Regression (GWR). Groundwater amplitude in the studied period was correlated with independent variables related to soil's physical properties and related to the basin's geomorphology, where six explanatory variables were selected to compose the statistical model. The results indicate that both OLS and GWR methodologies were capable of establishing correlations between the response and the explanatory variables. However, GWR managed to capture local correlations between the explanatory and the dependent variable that were not obtained by the OLS. Despite the complexity of groundwater behavior in tropical soils, this study demonstrates the possibility to estimate with data collected near the surface, how much water is stored in porous aquifers and groundwater behavior in soils of the Brazilian central plateau, being of great use to unmonitored river basins.

1. Introduction

Knowledge of flow and water storage in tropical soils of the Cerrado biome is relevant because of the importance of these hydrological processes for the stream flow regime in this biome. The fact that many rivers in the region are perennial, even with seasonal fluctuations of rainfall throughout the year, where 80% of the precipitation occurs from October to April (Brasil, 1984), corroborates this assertion. According to Lima (2010), based on data collected from small experimental watersheds in the Cerrado biome, base flow represent's 90% of streamflow, with the other 10% resulting from runoff generated by some rain events.

Monitoring soil's hydrological processes is not a trivial activity, demanding infrastructure and equipment which, added to the large spatial variability of environment and phenomena, end up discouraging such efforts. In this context, experimental basins are an important option to collect data and produce knowledge on these hydrological processes and their relationship with the environment. Hydrological models are presented as important assessment tools to help comprehend these complex systems and to simulate these processes, from collected data and information that are easier to obtain.

Several studies have looked for relationships between environmental features that are easy to obtain and groundwater hydrological processes occurring in tropical soils (Campos & Freitas-Silva, 1998; Lousada & Campos, 2005; Almeida, Resende, Rodrigues, & Campos, 2006; Fiori, Campos, & Almeida, 2010; Lima, 2010; Lima, Silva, Strauch, & Lorz, 2013; Gonçalves, Lohe, & Campos, 2015). Different hydrological models have been used in the Cerrado biome to simulate soil water behavior such as Feflow (Lima, 2010), Modflow (Santos, 2012) and SWAT (Minoti, Silva, Lombardi Neto, Koide, & Crestana, 2011; Salles, 2012; Strauch et al., 2012, 2013).

Methodologies of multivariate analysis are also applied in water

* Corresponding author. Instituto Brasília Ambiental – IBRAM, SEPN 511 – Bloco C – Edifício Bittar, CEP: 70.750-543, Brasília – DF, Brazil.
E-mail addresses: leandro.almeida.salles@gmail.com (L.d.A. Salles), jorge.werneck-lima@embrapa.br (J.E.F.W. Lima), roig@unb.br (H.L. Roig), juaci.malaquias@embrapa.br (J.V. Malaquias).

https://doi.org/10.1016/j.apgeog.2018.02.007
Received 8 September 2017; Received in revised form 7 February 2018; Accepted 11 February 2018

resources management. These methods can be grouped into three classes: Classification, Summarization and spatial relationships analysis (Bierman, Lewis, Ostendorf, & Tanner, 2011). Some of these techniques do not take into account the location of sampling points or the existence of spatial relationships between the measured data, for example, discriminant analysis and principal components analysis, while others seek to understand if spatial relation is significant, such as Geographic Weighting Regression, GWR (Bierman et al., 2011).

GWR has also been used in different types of studies, such as: epidemiology (Gilbert & Chakraborty, 2011; Leyk, Norlund, & Nuckols, 2012); remote sensing (Erdogan, 2010; Bayramov, Buchroithner, & McGurty, 2012; Gao, Huang, Li, & Li, 2012); economy (Himmelberger, Pearsall, & Rakshit, 2009; Breetzke, 2012 .; Deller, 2011; Grubesic, Mack, & Kaylen, 2012); forestry (Kupfer & Farris, 2007); transport (Blainey, 2010; Ibeas, Cordera, Dell'Olio, & Moura, 2011); and water quality (Mills et al., 2009; Wu & Driscoll, 2009; Harris & Brundson, 2010).

Given the importance of studying flow and water storage in tropical soils of the Cerrado biome and the difficulty to collect data capable of representing groundwater's spatial and temporal variation, in order to identify how much water is stored in porous aquifers, this study looked for relationships between the Upper Jardim experimental river basin's geomorphology and soil's physical properties, with the amplitude of groundwater's variation in the basin, using spatial statistics and geoprocessing methodologies.

2. Materials and methods

2.1. Study area

The Upper Jardim experimental river basin is located in the eastern region of the Federal District, between 15.71° and 15.86° S latitudes and 47.55° and 47.64° W longitudes (Fig. 1), in the center of Brazil. The basin has approximately 105 km^2 drainage area, and can be divided into two sub-basins: the Estanislau's and the Jardim's river basin, with 50 and 55 km^2, respectively (Lima, 2010). The Jardim's river is a tributary of the Preto's river, which is part of the Paracatu's watershed, an important contributor to the left side of the São Francisco river basin.

The climate in the Upper Jardim river basin has seasonality pattern in rainfall distribution, with two seasons well defined: dry winter, between April and October; and rainy summer in the remaining months of the year, with average annual rainfall ranging from 1,000 mm to 1,700 mm, and more than 80% of the precipitation occurring during this period. According to the classification proposed by Koppen, the prevailing climate in the region is Tropical Savanna (Brasil, 1984).

The Upper Jardim river basin is located in Brasilia's rural area. Therefore, agriculture is the principal land use (76.4%), followed by native vegetation (22.4%), which is divided in cerrado savannah (14.9%), and Gallery or Riparian forest (7.5%). The other 1.2% area is occupied by buildings, reservoirs and bare soil (Lima et al., 2007). Crops (soy, beans, cotton, corn and sorghum), citrus, coffee, manioc, vegetables, poultry, pigs and cattle are the main agricultural activities developed in the basin (Lima, 2010). Comparing satellite images from the past 20 years, it is possible to notice that land use has not changed significantly.

According to the semi-detailed survey at 1:50,000 scale produced by Reatto et al. (2000) the soils in the Upper Jardim river basin (Fig. 1) are classified as: Latosol (76.38%); Cambisol (16.68%); Plinthosol (2.54%); Gleysol (2.41%); and Quartzarenic Neosol (2.09%). These soil classes correspond, respectively, to Ferralsols, Cambisols, Plinthosols, Gleysols and Arenosols using the World Reference Base (WRB) for Soil Resources (Food and Agriculture Organisation, 2015). In addition to these types of soils, there are also outcropping rocks (0.24%).

2.2. Ordinary Least Squares regression – OLS

The general model for Ordinary Least Squares, OLS, can be expressed by the following equation:

$$Y_i = \beta_0 + \sum_k \beta_k X_{ik} + \varepsilon_i \tag{1}$$

where Y_i is the value of the dependent variable Y at point *i* (response variable), β_0 is the point that the regression line intercepts the y-axis, β_k is the slope coefficient of the k^{th} independent variable, X_{ik} is the value of k^{th} independent variable at point *i* (explanatory variable), and ε_i is the independent random error at point *i*, with normal distribution, N (0, σ^2).

Stepwise regression technique, with backward direction and based on the Akaike information criterion (AIC), was conducted to support the model's selection. The stepwise regression analysis is available in the R-Commander package (Fox, 2005). A final model was selected after evaluating the significance level of each estimated coefficient, and removing those with values above the established criteria (p-value ≤ 0.05).

An important part of the OLS methodology is the residual analysis. This assessment is relevant to evaluate if the linear model's assumptions are been achieved (Souza, 1998). The Residual analysis was conducted with the following procedures which are suggested on the literature (Souza, 1998; Erdogan, 2010; Cardozo, García-Palomares, & Gutiérrez, 2012): F-test to assess the significance of the model; Koenker's studentized Breusch-Pagan statistic, to evaluate the model stationarity; Bonferonni test for outliers analysis; Shapiro-Wilks and Jarque-Bera test to assess the residuals normality; Global Moran's I statistic for autocorrelation analysis. The Variance Inflator Factor, VIF, was also evaluated for each explanatory variable in the selected model.

2.3. Geographically Weighted regression – GWR

Geographically weighted regression, GWR, was conducted to include spatial variability of the variables in the selected statistic model used to correlate environmental factors with the amplitude of water table variation in the Upper Jardim river basin. The general model for GWR can be expressed by the following equation:

$$Y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) X_{ik} + \varepsilon_i \tag{2}$$

where (u_i, v_i) is the coordinate of the i^{th} point in space and β_k (u_i, v_i) is a realization of the continuous function β_k (u, v) at the point *i*, and ε_i is the independent random error at the point *i* with normal distribution, N (0, σ^2) (Fotheringham, Brunsdon, & Charlton, 2002). It is possible to notice that Equation (2) is a special case of Equation (1), where the parameters do not vary in space (Fotheringham et al., 2002).

The regression's parameters, for each observed point *i*, can be calculated with the following matrix equation:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \tag{3}$$

where $\hat{\beta}$ is the estimate of β, and W (u_i, v_i) is a *n* by *n* matrix that the diagonal elements, w_{in}, are spatial weights for each of the *n* observed data for regression point *i*, and the remaining elements of the matrix are zero. Equation (3) resembles OLS estimator with the difference that the GWR has weights which vary according to point *i* location, while OLS have constant weights (Fotheringham et al., 2002).

This study used the bi-square spatial weighting function with an adaptive spatial kernel. The weights are calculated as follow:

$$w_{ij} = \left\{1 - \left(\frac{d_{ij}}{b}\right)^2\right\}^2 \qquad \text{if } d_{ij} < b \tag{4}$$

$$w_{ij} = 0 \qquad \text{if } d_{ij} \geq b \tag{5}$$

where *i* is the point where the regression parameter is been estimated, *j*