Eurasian J Soil Sci 2020, 9 (1) 18 - 23

Eurasian Journal of Soil Science

Journal homepage : http://ejss.fesss.org

Soil hydraulic properties: A simple and practical approach to estimate the number of samples

Luiz Felipe Salemi [a,b,*], Rafael Pires Fernandes [b], Robson Willians da Costa Silva [b] Lara Gabrielle Garcia [c], Jorge Marcos de Moraes [b,d], Juliano Daniel Groppo [b,e] Luiz Antonio Martinelli [b]

[a] Núcleo de Estudos e Pesquisas Ambientais e Limnológicos, Faculdade de Planaltina, Universidade de Brasília. Área Universitária 01, Bairro Vila Nossa Senhora de Fátima, CEP 73345-010, Brasília (Planaltina), Distrito Federal, Brasil

[b] Laboratório de Ecologia Isotópica, Centro de Energia Nuclear na Agricultura, Universidade de São Paulo. Avenida Centenário, 303, Bairro São Dimas, CEP 13400-970, Piracicaba, São Paulo, Brasil

[c] Laboratório de Hidrologia Florestal, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo. Avenida Pádua Dias, 11, São Dimas, CEP 13418-900, Piracicaba, São Paulo, Brasil

[d] Escola de Engenharia de Piracicaba. Avenida Monsenhor Martinho Salgot, 560 A, Bairro Areião, Piracicaba, São Paulo, Brasil

[e] Universidade Vale do Rio Doce, Rua Israel Pinheiro, 2000, Bairro Universitário, CEP 35020-220, Governador Valadares Minas Gerais, Brasil

Abstract

There have been a number of studies dealing with soil hydraulic properties. Yet, there is a poor discussion on the number of samples necessary to represent such variables that usually vary orders of magnitude in space. In the present paper, we examine the adequate number of samples for two soil saturated hydraulic conductivity (Ksat) data sets: (1) normal distribution (a 40 year-old pasture) and (2) non-normal distribution (primary forest). To assess the adequate number of samples in each case, we used for normal distribution, an statistical criterion of standard deviation lower than 5% compared to a high sampling effort (n = 25) as an indicative of a proper representation of Ksat variability. In the case of non-normal distribution, we used the same criterion but using median absolute deviation (a non-parametric statistics). Both data sets were available in Salemi et al. (2013) and were Ksat measured at 0.15 m soil depth for medium-textured inceptisols in São Paulo State, Brazil. For each data set, we simulated 10 'new' samplings in which we calculated mean and standard deviation from sample 1 to 25 (for normal data) and median and median absolute deviation (for non-normal data). We found that, on average, at least 17 to 22 samples had to be collected to meet the adopted criterion for normal data whereas 20 to 25 had to be collected for non-normal data. Such numbers of samples exceed those used in a number of papers. Additional examples of this method with a light modification are given to establish number of samples in new study areas as well as to estimate number of samples when comparing two (or more) land-uses. Simple and practical procedures like those presented here could estimate the number of samples that adequately represents soil hydraulic properties variability.

Keywords: Inceptisols, sampling, variation, water movement.

Article Info

Received : 08.05.2019
Accepted : 20.09.2019

Introduction

Soil hydraulic properties such as infiltration capacity and saturated hydraulic conductivity are important to understand water movement within the soil (Reichardt and Timm, 2012). Such variables have application in

* Corresponding author.
Núcleo de Estudos e Pesquisas Ambientais e Limnológicos, Faculdade de Planaltina, Universidade de Brasília. Área Universitária 01, Bairro Vila Nossa Senhora de Fátima, CEP 73345-010, Brasília (Planaltina), Distrito Federal, Brasil
Tel.: +55 61 8363 7041
E-mail address: lfsalemi@gmail.com
e-ISSN: 2147-4249
DOI: 10.18393/ejss. 629344

different areas of soil, agricultural, forestry and environmental sciences such as soil and water conservation (Bertoni and Lombardi Neto, 1990), irrigation (Barreto, 1986), hillslope hydrology (Kirkby, 1978) which, in turn, can help to predict peak flows (Hewlett, 1982).

Infiltration capacity and soil saturated hydraulic conductivity generally present substantial spatial variability (Mesquita and Moraes, 2004; Bonell et al., 2010). Values of such variables can vary up to three orders of magnitude for a given soil type (Salemi et al., 2013; Ghimire et al., 2013; Ghimire et al., 2014). Such variation along with laborious field and laboratory methods make its proper representation a hard task. For example, one single measurement of infiltration took 6 hours using the double ring method (Bono et al., 2012). For this reason, papers that characterized soil hydraulic properties used a quite variable number of samples ranging from 1 to 87 samples (e.g. Elsenbeer et al., 1999; Souza and Alves, 2003; Moraes et al., 2006; Bonell et al., 2010; Salemi et al., 2013; Vilarinho et al., 2013; Ghimire et al., 2013; 2014). Yet, there is a poor discussion on the number of samples necessary to represent such variables that usually vary orders of magnitude in space. In other words, how many sampling units are needed for an accurate assessment?

In this context, the objective of the present paper was to present a simple method used to estimate the number of samples (also known as sample size) to properly characterize the soil saturated hydraulic conductivity.

Material and Methods

Study area

The soil saturated hydraulic conductivity (Ksat) measurements of pasture and forest have been obtained from two sites. First, pasture (hereafter in the present paper 'normal data') was a 40-yr old pasture (*Brachiaria decumbens* Stapf.) located in Natividade da Serra – SP (Brazil) (Latitude: 23º 24' 53,9" S Longitude: 45 º 15' 04,0" W). This pasture is sited on steep slopes dominated by medium textured soils classified as cambissolos háplicos distróficos (Brazilian soil classification system by EMBRAPA) or Inceptisols (US Soil Taxonomy). The same soil type was present in Forest (hereafter in the present paper 'non-normal data'). The forest site is located inside the State Park of Serra do Mar in São Luiz do Paraitinga – SP (Brazil) (Latitude: 23º 17' - 23º 24´ S Longitude: 45º03' - 45º11'W). Mean slope for pasture and forest were 37 ± 17% and 28 ± 14% (mean ± standard deviation) respectively. For more details on soil see Salemi et al. (2013).

Methods

Ksat measurements used in the present paper was measured at 0.15 m soil depth at both sites using a constant head permeameter known as Amoozemeter (Amoozegar, 1992). Twenty five samples have been collected following linear transects from the water divide to approximately 30 meters relative to a first order stream (Salemi et al., 2013). The distance between measurements was ~ 2 meters. We decided to use the observations available in Salemi et al. (2013) for two main reasons: (1) direct access to original observations and (2) a high number of samples (n = 25).

Aiming to estimate an adequate sampling effort, we assumed that the 25 samples used in Salemi et al. (2013) are the best representation of the two sites. For this reason, the statistics calculated with such measurements should be taken as a reference in determining the adequate number of samples. For this, we simulated 10 'new' sampling events for each site. In each simulated sampling, we assumed the 25 samples have been collected randomly as this is a practical and advantageous sampling design compared to others (Hassler et al., 2014). To simulate random sampling, we used a simple on line randomizer (www.randomizer.org). In doing the randomization, mean and standard deviation have been calculated from the first (x_1) until the last sample (x_{25}) for normal data whereas median and median absolute deviation have been calculated for non-normal data. Thus, 25 measures of central tendency and 24 of variability have been calculated for each of the 10 simulations in each site/data set. Median absolute deviation (MAD) was calculated as follows:

$$MAD = \text{median}\ (|\ xi - \text{median}\ |) \qquad (1)$$

Where: *xi* is the observation

Once performed the 10 simulations in each site (data set), a representative number of samples was considered achieved when variability statistics values presented consistently (i.e. for three times consecutively), a maximum of ± 5% of the variation of the statistics obtained from the maximum number of samples (n = 25). Put another way, when standard deviation or median absolute deviation of a given number of samples met such criterion for three times in row, we admitted a representative number of samples had been achieved. This method is an adaptation of the one ecologists use when determining the need for additional samples to define species richness in a given ecosystem (see, for example, Gotelli and Chao, 2013).

Analysis

Aiming to demonstrate graphically the effect of increasing number of samples on data variability, we arbitrarily selected 3 of the 10 simulations (randomization 2, 4 and 6) for both data sets ('normal' and 'non-normal').

The basis for the comparison of variability statistics is that when one increases the number of samples (also known as 'sample size'), it is possible to obtain statistics closer to population parameters. In the present paper, we assumed that the best representation of the soil had been achieved with a Ksat sample of 25 originally available in Salemi et al. (2013).

Results

The statistics for the high number of samples (n = 25) were 29.30 ± 17.09 mm hr^{-1} (mean ± standard deviation) and 60.97 ± 44.24 mm hr^{-1} (median ± median absolute deviation) for normal and non-normal data respectively. In the case of normal data, 17 ± 5 samples had to be collected to reach the criterion (i.e. 5% of the standard deviation of the maximum sampling effort). Number of samples varied from 9 to 23 (Table 1).

Table 1. Number of samples, mean, standard deviation and their absolute difference when compared to the maximum sample size for normal data. All statistics are presented in mm hr^{-1}.

Simulation	Number of Samples	Mean	Standard Deviation	Deviation when n = 25	
				Mean	Standard Deviation
1	19	27.57	17.01	1.73	0.07
2	12	30.52	17.65	-1.22	-0.57
3	19	30.39	17.71	-1.09	-0.63
4	13	30.25	17.76	-0.95	-0.68
5	10	29.56	17.54	-0.26	-0.46
6	9	27.05	16.51	2.25	0.57
7	12	36.46	17.61	-7.16	-0.53
8	19	29.41	17.75	-0.11	-0.67
9	18	27.01	16.63	2.29	0.45
10	23	29.64	17.41	-0.34	-0.33

*Negative values indicate mean and/or standard deviation were higher than those obtained compared to the maximum sample size (n = 25).

As for non-normal data, 20 ± 5 samples had to be collected. Number of samples varied from 7 to 25 (Table 2). In normal and non-normal data sets, there was a relatively low absolute deviation of the variability statistics compared to the one of the maximum number of samples (Table 1 and 2).

Table 2. Number of samples, median, median absolute deviation and their absolute difference when compared to the maximum number of samples for non-normal data. All statistics are presented in mm hr^{-1}.

Simulation	Number of Samples	Median	Median Absolute Deviation	Median	Deviation when n = 25
					Median Absolute Deviation
1	24	60.97	42.44	0	1.8
2	24	60.97	42.44	0	1.8
3	15	50.81	43.16	10.16	1.08
4	21	50.81	40.65	10.16	3.59
5	20	60.97	42.44	0	1.8
6	25	60.97	42.44	0	1.8
7	24	60.97	42.44	0	1.8
8	24	60.97	44.86	0	-0.62
9	7	60.97	16.63	0	-1.25
10	21	60.97	17.41	0	0

*Negative values indicate median and/or median absolute deviation were higher than those obtained compared to the maximum sample size (n = 25).

For normal data, the simulations 2, 4 and 6 reached the criterion at 12, 8 and 9 samples respectively (Figure 1). As for non-normal data, simulations 3, 4 and 6 reached at 15, 21 and 25 samples respectively (Figure 2).

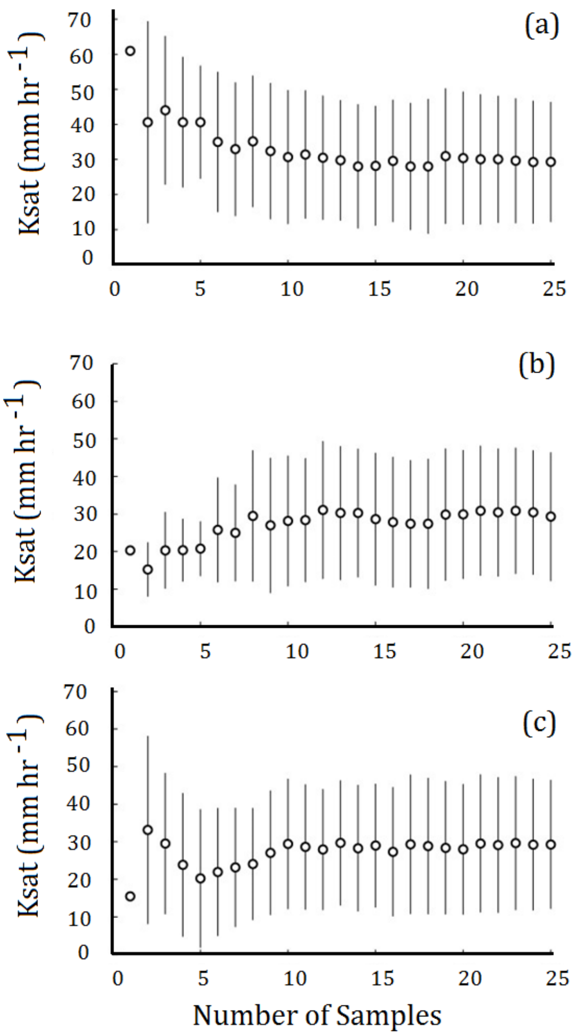

Figure 1. *Normal* data statistics. Circles indicate *mean* whereas bars represent *standard deviation* of the soil saturated hydraulic conductivity (Ksat) for randomizations 2 (a), 4 (b) e 6 (c) as a function of number of samples (x axis)

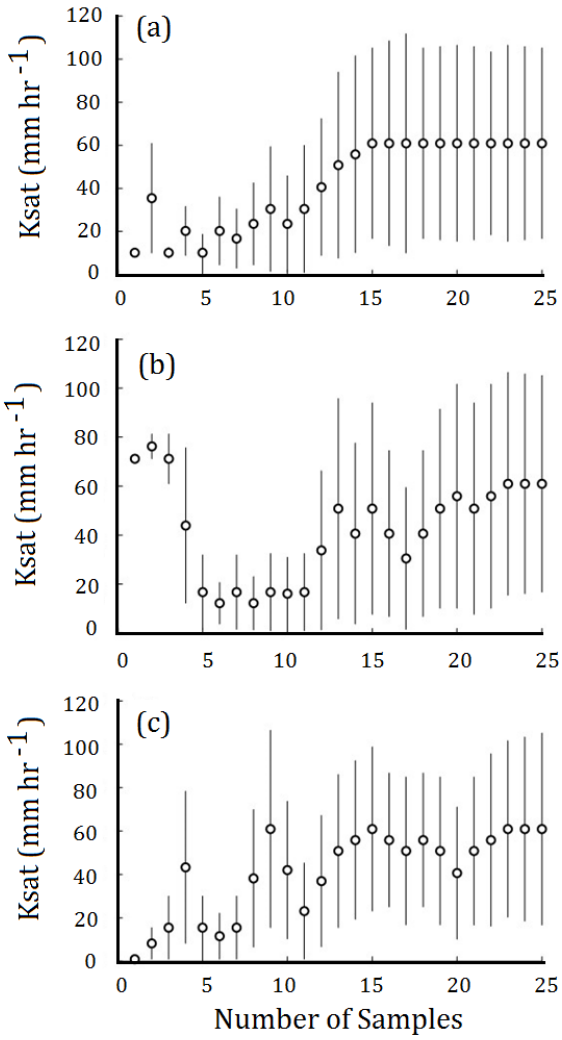

Figure 2. *Non-normal* data statistics. Circles indicate *median* whereas bars represent *median absolute deviation* of the soil saturated hydraulic conductivity (Ksat) for randomizations 3 (a), 4 (b) e 6 (c) as a function of number of samples (x axis).

Discussion

The criterion to reach a representative sampling were, on average, 17 and 20 samples for normal and non-normal data, respectively. In other words, after collecting these numbers of samples, there was no substantial modification in the variability statistics when adding an additional sample. Such number of samples are higher than the ones used in a number of papers (e.g. Silva and Kato, 1998; Beutler et al., 2001, Araújo et al., 2007; Borges et al., 2009; Bono et al., 2012;). In case our Ksat variation presented for inceptisols also represent other soil types (oxisols, ultisols and others), there have been probably a misrepresentation of such variable in the referred studies. For this reason, such papers probably over or underestimate the actual Ksat variability. As a matter of the fact, there were lower (Figure 1b) and higher (Figure 1a,c) standard deviation when using number of samples equal or lower than 5 which was substantially lower compared to the one indicated by the present approach (n=17). Number of samples lower than 5 have been used in many papers.

The same rationale just described applies to the case for non-normal data. That is to say, low number of samples (i.e. n = 5 or lower) generated variability statistics that do not approached the 'actual' population variability (Figure 2).

Regarding normal data, number of samples equivalent to 17 or even 22 (that is, the mean, 17, plus 5, the higher end of standard deviation) and in the case of non-normal data 20 or even 25 provided reasonable substance to accept that number of samples (n = 21) used by Elsenbeer et al. (1999) and Moraes et al. (2006) as well as the one used by Bonell et al. (2010), which varied from 22 to 87, could probably be considered a well-represented sample, though no explanation on the basis of the number of samples is provided in all these papers and other including Salemi et al. (2013).

We acknowledge that in doing the procedure shown here with maximum number of samples higher than 25, it is possible to obtain different results. This means that, as one increases the number of samples, there is an approximation between the statistics (estimates based on samples) and parameters (actual numbers regarding the population). Thus, increasing the number of samples leads to a more representative sample. That is specially the case when one gets samples varying from more than 2 orders of magnitude which are not uncommon in studies dealing with soil hydraulic properties under certain land-uses (e.g. Moraes et al., 2006; Scheffler et al., 2011; Hassler et al., 2012; Ghimire et al., 2013; Salemi et al., 2013; Ghimire et al., 2014). For instance, inceptisols and luvisols (Ghimire et al., 2013, Salemi et al., 2013; Ghimire et al., 2014), as well as plinthic soils (Moraes et al., 2006) presented Ksat varying 3 orders of magnitude.

The procedure presented here could also be utilized for verifying the number of samples when sampling soil hydraulic properties in a new study area (that is, there is no previous information on soil hydraulic properties of the area). In this case, a slightly different manner is proposed, that is, when collecting samples, if the addition of new samples do not produce substantial variation on the variability statistics for, at least, 3 subsequent additions of new samples, an adequate number of samples can be considered met. Furthermore, the procedure could also be applied in case a comparison between land-uses is the goal as it was the case of many studies (see, for example, Zimmermann et al., 2006; Scheffler et al., 2011; Salemi et al., 2013). The procedure used here could point to a single number of samples that should be collected in the two (or more) land-uses being compared. To exemplify such case, we performed calculations with pasture data ('normal data') using the median and median absolute deviation (the statistics used for non-normal data, that is, forest). In doing so, pasture median absolute deviation stabilized only at 25 samples whereas forest stabilized at 20 samples. Thus, in this case, 25 samples was the standard number of samples that have to be collected in order to have a balanced comparison between land-uses.

Though the criterion (5% around variability statistics) used here is, to some extent, arbitrary, it seems reasonable to assume that no substantial modification of variability indicated a proper representation of soil hydraulic property variation. In addition, such accuracy is well within the limits of accuracy that are required to estimate hydraulic properties in the field (Reynolds and Elrick, 1990).

Lastly, the idea behind such method is not entirely new. Ecologists use similar criterion when deciding the need for additional samples to determine species richness in a given ecosystem (Gotelli and Chao, 2013). In this case, when species samplings do not add new species in the area, the number of species of the ecosystem is considered met.

Conclusion

A simple and practical procedure to examine number of samples has been provided using variation stability as the main criterion. When sample variability stabilized, it indicated that sample variation probably approached population variability generating an adequate representation. Though the results presented here might be site-specific, the demonstrated procedure can be utilised for estimating number of samples in other areas in future studies.

Acknowledgements

The authors would like to thank Sao Paulo State Research Foundation (FAPESP) for the financial support of the present study (grant #06/54292-9).

References

Amoozegar A., 1992. Compact constant head permeameter: a convenient device for measuring hydraulic conductivity. In: Advances in measurement of soil physical properties: bringing theory into practice. Topp, C.G., Reynolds, W.D., Green, R.E. (Eds.). Soil Science Society of America, Madison, WI, USA. pp. 31-42.

Araújo, R., Goedert, W.J., Lacerda, M.P.C., 2007. Soil quality under different uses and native cerrado. *Revista Brasileira de Ciência do Solo* 31: 1099-108.

Barreto, G.B., 1986. Irrigação: princípios, métodos e prática. Campinas: Instituto Campineiro de Ensino Agrícola. 236p.

Bertoni, J.; Lombardi Neto, F., 1990. Conservação do Solo. Ícon, São Paulo. 394p.

Beutler, A.N., Silva, M.L.N., Curi, N., Ferreira, M.M., Cruz, J.C., Pereira Filho, I.A., 2001. Resistence to penetration and permeability of a typic dystrophic red latosol under management systems in the cerrado region. *Revista Brasileira de Ciência do Solo* 25(1): 167-177.

Bonell, M., Purandara, B.K., Venkatesh, B., Krishnaswamy, J., Acharya, H.A.K., Singh, U.V., Jayakumar, R., Chappell, N, 2010. The impact of forest use and reforestation on soil hydraulic conductivity in the Western Ghats of India: Implications for surface and sub-surface hydrology. *Journal of Hydrology* 391(1-2): 47-62.

Bono, J.A.M., Macedo, M.C.M., Tormena, C.A., Nanni, M.R., Gomes, E.P., Müller, M.M.L., 2012. Water infiltration into an oxisol in the South-west Cerrado region under different use and management systems. *Revista Brasileira de Ciência do Solo* 36: 1845-1853.

Borges, T.A., Oliveira, F.A., Silva, E.M., Goedert, W.J., 2009. Evaluation of soil-water parameters of a Red Latosol under pasture and 'cerrado'. *Revista Brasileira de Engenharia Agrícola e Ambiental* 13(1):18-25.

Elsenbeer, H., Newton, B.E., Dunne, T., de Moraes, J.M., 1999. Soil hydraulic conductivities of latosols under pasture, forest and teak in Rondonia, Brazil. *Hydrological Processes* 13(9): 1417-1422.

Ghimire, C.P., Bonell, M., Bruijnzeel, L.A., Coles, N., Lubczynski, M.W., 2013. Reforesting severely degraded grassland in the Lesser Himalaya of Nepal: effects on soil hydraulic conductivity and overland flow production. *Journal Geophysical Research Earth Surface* 118(4): 2528-2545.

Ghimire, C.P., Bruijnzeel, L.A., Bonell, M., Coles, N., Lubczynski, M.W., Gilmour, D.A. 2014. The effect of sustained forest use on hillslope soil hydraulic conductivity in the Middle Mountains of Central Nepal. *Ecohydrology* 7(2):478-495.

Gotelli, N.J., Chao, A., 2013. Measuring an estimating species richness, species diversity, and biotic similarity from sampling data. In: Encyclopedia of Biodiversity. Levin, S.A., (Ed.), Second edition. Vol.5, Waltham, MA, Academic Press, pp.195-211.

Hassler, S.K., Lark, R.M., Zimmermann, B., Elsenbeer, H., 2014. Which sampling design to monitor saturated hydraulic conductivity? *European Journal of Soil Science* 65(6): 792-802.

Hassler, S.K., Zimmermann, B., van Breugel, M., Hall, J.S., Elsenbeer, H. 2011. Recovery of saturated hydraulic conductivity under secondary succession on former pasture in the humid tropics. *Forest Ecology and Management* 261(10): 1634-1642.

Hewlett, J.D., 1982. Principles of Forest Hydrology. The University of Georgia Press, Athens, Georgia, USA. 183p.

Kirkby, M.J., 1978. Hillslope Hydrology. Chichester: John Wiley & Sons. 375p.

Mesquita, M.G.B.F., Moraes, S.O., 2004. The dependence of the saturated hydraulic conductivity on physical soil properties. *Ciência Rural* 34(3): 963-969.

Moraes, J.M., Schuler, A.E., Dunne, T., Figueiredo, R.O., Victoria, R.L., 2006. Water storage and runoff processes in plinthic soils under forest and pasture in Eastern Amazonia. *Hydrological Processes* 20(12): 2509-2525.

Reichardt, K., Timm, L.C., 2012. Solo, Planta e Atmosfera: Conceitos e Aplicações. Manole, São Paulo. 594p.

Reynolds, W.D., Elrick, D.E., 1990. Ponded infiltration from a single ring: I. Analysis of steady flow. *Soil Science Society American Journal* 54(5): 1233-1241.

Salemi, L.F., Groppo, J.D., Trevisan, R., Moraes, J.M., Ferraz, S.F.B., Villani, J.P., Duarte-Neto, P.J., Martinelli, L.A. 2013. Land-use change in the Atlantic rainforest region: Consequences for the hydrology of small catchments. *Journal of Hydrology* 499: 100-109.

Scheffler, R, Neill, C, Krusche, AV, Elsenbeer, HA. 2011. Soil hydraulic response to land-use change associated with the recent soybean expansion at the Amazon agricultural frontier. *Agriculture Ecosystem and Environment* 144(1): 281-289.

Silva, CL, Kato, E. 1998. Evaluation of models for the prevision of water infiltration in the soil under savanna. *Pesquisa Agropecuária Brasileira* 33(7): 1149-1158.

Souza, Z.M., Alves, M.C., 2003. Water movement and resistence to penetration in a distrophic Red Latosol of cerrado under different uses and management. *Revista Brasileira de Engenharia Agrícola e Ambiental* 7(1): 18-23.

Vilarinho, R.K., Koetz, M., Schlichting, A.F., Silva, M.C., Bonfum-Silva E.M., 2013. Determination of water steady infiltration rate in native soil from Cerrado. *Revista Brasileira de Agricultura Irrigada* 7(1):17-26.

Zimmermann, B., Elsenbeer, H., Moraes, J.M., 2006. The influence of land-use changes on soil hydraulic properties: implications for runoff generation. *Forest Ecology and Management* 222(1-3): 29-38.