Incorporation of ancillary information derived from satellite images applied on environmental variables evaluation.

Rodrigo Lilla Manzione[1]
Martin Knotters[2]
Gerard B. M. Heuvelink[2]

[1]Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brazil
manzione@dpi.inpe.br

[2]Alterra – Soil Science Center
Droevendaalsesteeg 3, 6708PB – Wageningen, The Netherlands
{Martin.Knotters, Gerard.Heuvelink}@wur.nl

Abstract. Geostatistical models for spatial prediction are based on observations points. When ancillary information related to the target variable is available, it can be used to improve the estimations. Some sources of information, like satellite images or digital elevation models, provide full information for the whole study area, improving even more the point estimation. It is useful when the target variable is difficult or expensive to sample. Also when is knew the observations contain errors and ancillary information can contribute to estimation procedure. The results are maps more realistic, incorporating physical knowledge about the processes under study. The aim of this work is to show how incorporate ancillary information derived from remote sensing products in spatial prediction model. Here is presented a study case using trends in water table depths as target variable and land use classification derived from Landsat image as ancillary information. The results were evaluated by cross validation and the use of ancillary information contributed to improve the spatial prediction.

Key-words – geostatistics, ancillary information, universal kriging, groundwater.
Palavras-chave – geoestatística, informação auxiliar, krigagem universal, água subterrânea.

1. Introduction

Spatial prediction is the procedure of estimating the values of a target quantity at unvisited locations. When applied to a whole study area, it is also referred to as spatial interpolation or mapping (Hengel, 2004). Several studies have been made combining remote sensing, GIS, statistical analysis, DEM and ancillary information to map vegetation (Dymond et al., 1992; Hoersch et al., 2002; Lees & Ritman, 1991; Michaelsen et al., 1994; Moore et al., 1991). Land cover classification is one of the principal motivations and successes of satellite remote sensing. This classification is obtained by supervised classification from some ground-control points. The interest for digital soil mappers is to detect areas of bare soil, or of particular crops representing where humans have picked out soil with particular qualities.

In the 1990s, with emerging GIS and remote sensing technologies, analysts became interested to use exhaustively mapped secondary variables to map directly environmental variables. The first applications were based on the use of simple linear regression models between terrain attribute maps and soil parameters (Moore et al., 1993; Gessler et al., 1995). In the next phase, the predictors were extended to spatial prediction by multiple regression with auxiliary variables (Odeh et al., 1994, 1995) or a set of environmental variables and remote sensing images (McKenzie & Ryan, 1999). In the last decade, many 'hybrid' interpolation techniques, which combine kriging and use of auxiliary information, has been developed and tested. A spatial prediction technique, which jointly employs correlation with auxiliary maps and spatial correlation is universal kriging (UK), originally described by Matheron (1969). Other variants of UK are Kriging with External Drift (KED) and Regression Kriging (RK). In fact, UK, KED and RK are equivalent methods and should,

under the same assumptions, yield the same predictions. Many authors (Deutsch & Journel, 1992; Wackernagel, 1995), however, agree the term UK should be reserved for the case where the drift (or trend) is modeled as a function of the coordinates only. In several studies (Odeh et al., 1994, 1995; Goovaerts, 1999; Bishop & McBratney, 2001, Hengel et al., 2004), combination of kriging and correlation with auxiliary data outperformed ordinary kriging, co-kriging and plain regression.

The aim of this study is to show how incorporate ancillary information derived from remote sensing products in spatial prediction model. Here are presented a study case using linear trends in water table depths as target variable and land use classification derived from satellite image as ancillary information.

2. Materials e Methods

2.1. Study area

The Jardim River watershed is a representative Cerrado area in the eastern part of the Brazilian D.C., latitudes $15^o40^{'}$S and $16^o02^{'}$S and longitudes $47^o20^{'}$W and $47^o40^{'}$W. The dry and the wet season are well-defined, with the rain concentrated between October and April. During the past years, almost all natural vegetation present in this area was replaced by agricultural crops, and the use of irrigation systems has substantially increased in this region during the past years. The main cultivations present in this area are soybeans, cotton and corn crops, as well as pasture and horticultural crops.

To monitor the water table depths, 40 wells were drilled in the area. The locations were selected purposively, to cover the range of soil types in the area (Lousada, 2005). The water table was observed semimonthly from October 2003 until August 2006, resulting in series of 70 more or less regularly spaced semi-monthly observations. Series of precipitation and potential evapotranspiration were available from a climate station close to the basin, from 1974 until 1996 with a monthly frequency, and from 1996 until August 2006 with a daily frequency. Figure 1 shows a map of the study area and the well locations.

Ancillary information related to the sources of systematic changes was derived from Landsat images. An image from July 23, 2005, orbit/point 221/71, was classified for the actual land use in the region. The image classification results in a land use surface, divided in three classes: Agricultural Crops, Pasture and Cerrado. This classification was created based on expert knowledge and manual delineation of the classes. The class Agricultural Crops includes all kinds of agricultural products cultivated in the area: small areas cultivated with horticultural products, such as carrots, lettuce, tomatoes, and big areas cultivated with products such as corn, soybeans, cotton, coffee, sugarcane. All these crops demand more water than the original vegetation. The agricultural activities are intensive, resulting in three production cycles during one year when irrigation is applied. Also, the land use in the class Agricultural Crops is very dynamic because of agronomical recommendations, rotation schemes or simply prices. The class Pasture is considered to be less water demanding than Agricultural Crops, but more demanding than the natural Cerrado vegetation. These areas are not as dynamic in land use changes as the Agricultural Crops. **Figure 1** gives the classified land use map.
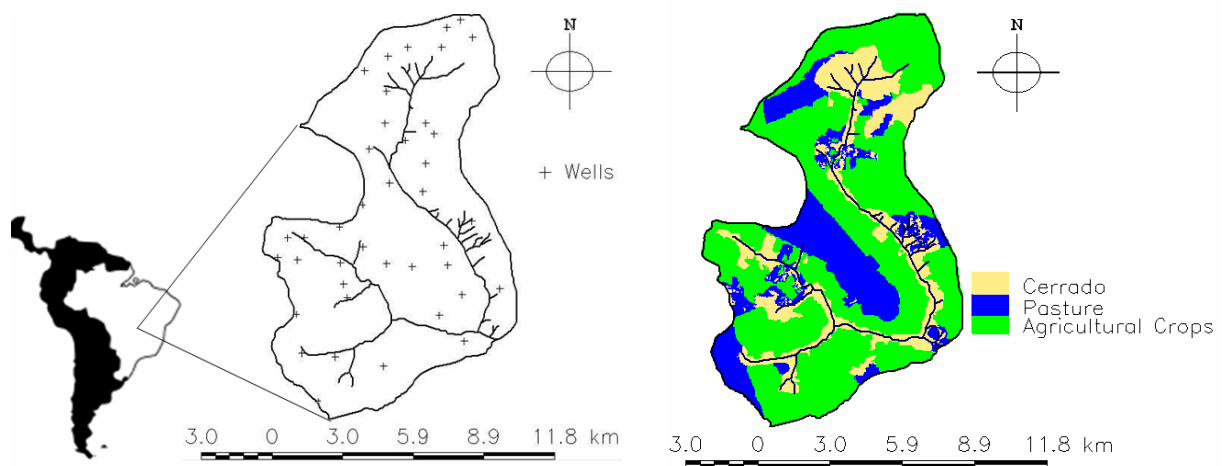
Figure 1: Jardim River watershed and location of the observation wells (+) (right) and the Image classification for actual land use (left).

The land use map shown in **Figure 1** has sharp boundaries. For hydrological studies, this does not make sense because water levels do not have abrupt variations related to land use. Therefore, the land use map was smoothed by computing the average presence of land use within a window with 500m radius. The choice of radius was based on expert knowledge and chosen after several tests.

2.2. The PIRFICT-model

The PIRFICT-model, introduced by Von Asmuth et al. (2002), was applied in this study because the model can describe a wide range of response times with differences in sampling frequency between input series and output series. Being the most important driving forces of water table fluctuation, precipitation and evapotranspiration are incorporated as exogenous variables into the model. Besides precipitation and evapotranspiration, a linear trend component is incorporated to model systematic changes in the water table system. It is an alternative to discrete-time TFN models. In the PIRFICT-model a block pulse of the input is transformed into an output series by a continuous-time transfer function. The coefficients of this function do not depend on the observation frequency. The following single input continuous TFN model can be used to model the relationship between water table dynamics and precipitation surplus/deficit. Manzione et al. (2006) give more details about developing the PIRFICT-model and its application in the Cerrado area.

2.3. Regionalizing the linear trend parameter of the time series model

PIRFICT-models were calibrated to the 40 series of water table depths, using the program Menyanthes. Next, the trend parameters reflecting systematic changes of water table depths were mapped. We followed the hypothesis the actual land use can lead to systematic changes in water table depths. The trend parameter of the PIRFICT-model was interpolated spatially using universal kriging. This works as follows. Let the 'observed' trend parameters be denoted as $z(x_1)$, $z(x_2)$, ..., $z(x_n)$, where x_i is a (two-dimensional) well location and n is the number of observations (i.e., n=40). At a new unvisited location x_0 in the area, $z(x_0)$ is predicted by summing the predicted drift and the interpolated residual (Odeh et al., 1994; Hengl et al., 2004):

$$\hat{z}(x_0) = \hat{m}(x_0) + \hat{e}(x_0) \tag{1}$$

where the drift m is fitted by linear regression analysis, and the residuals e are interpolated using kriging:

$$\hat{z}(x_0) = \sum_{k=0}^{p} \hat{\beta}_k \cdot q_k(x_0) + \sum_{i=1}^{n} w_i(x_0) \cdot e(x_i); \tag{2}$$
$$q_0(x_0) = 1$$

Here, the β_k are estimated drift model coefficients, $q_k(x_0)$ is the kth external explanatory variable (predictor) at location x_0, p is the number of predictors, $w_i(x_0)$ are the kriging weights and $e(x_i)$ are the zero-mean regression residuals.

The general universal kriging technique was used to interpolate the linear trend parameter (*LTP*) of the PIRFICT model. The Land Use (*LU*) map was used as predictor. The model was formulated as follows:

$$LTP(x_0) = \beta_0 + \beta_1 \cdot LU1(x_0) + \beta_2 \cdot LU2(x_0) + \beta_3 \cdot LU3(x_0) + e(x_0) \tag{3}$$

where *LU*1 is land use class 1 (Agricultural Crops), *LU*2 is land use class 2 (Pasture), *LU*3 is land use class 3 (Cerrado Vegetation) and e is a zero-mean spatially correlated residual. The semivariogram characterizes the spatial correlation structure. The geostatistical package GSTAT (Pebesma, 2004) was used in these analyzes.

3. Results e discussion

Including the land use variables into the geostatistical model caused a decrease in the semivariance. Here is compared the semivariogram for *LTP* with and without inclusion of *LU* as ancillary information in the spatial model (**Figure 2**).
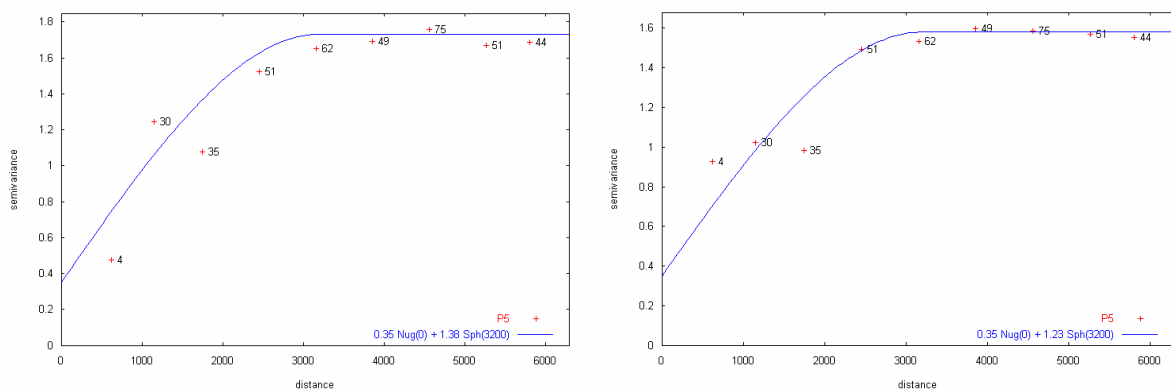

Figure 2: Semivariograms fitted for the linear trend parameter without including a trend that depends on land use (right) and with including a trend (left).

The spatial dependence at small distances is poorly estimated because of the small number of observation wells that are fairly uniformly spread across the area. The nugget parameter of the semivariogram reflects the precision of the *LTP* and the short-distance spatial variation in *LTP*. The variance at small distances increase including *LU* variables in the model because we are including more sources of uncertainty in the estimation. These sources of

uncertainty come from errors at observations and misclassification of the satellite image. The criteria adopted to incorporate *LU* information was first the relationship between actual land use and the *LTP*. The regression model presented a $R^2=0,37$. This poor relationship indicated that other variables are responsible for eventual changes in water table depths, not included in this analysis. So, we investigated if the variance of the model is reduced including *LU* variables. The sill of the semivariograms revealed a small decrease in the variance including *LU* variables, what convinced us to use it in estimation of systematic changes in water table depths.

These semivariograms presented on the **Figure 2** were used to perform ordinary kriging and universal kriging. Ordinary kriging consider only the *LTP* calibrated from PIRFICT-model (**Figure 3**) and universal kriging include *LU* variables as ancillary information (**Figure 4**). Positive values in the interpolated map of systematic changes in water table depth indicate a rise of the water table during the last three years, and negative values lowering. The map of the kriging standard deviations reflects the accuracy of the predicted systematic changes of water table depth. The large standard deviations reflect the large uncertainty in the *LTP* parameters which were estimated from relatively short time series. The large uncertainty implies that observed lowerings and risings of the water table depth may not be statistically significant.
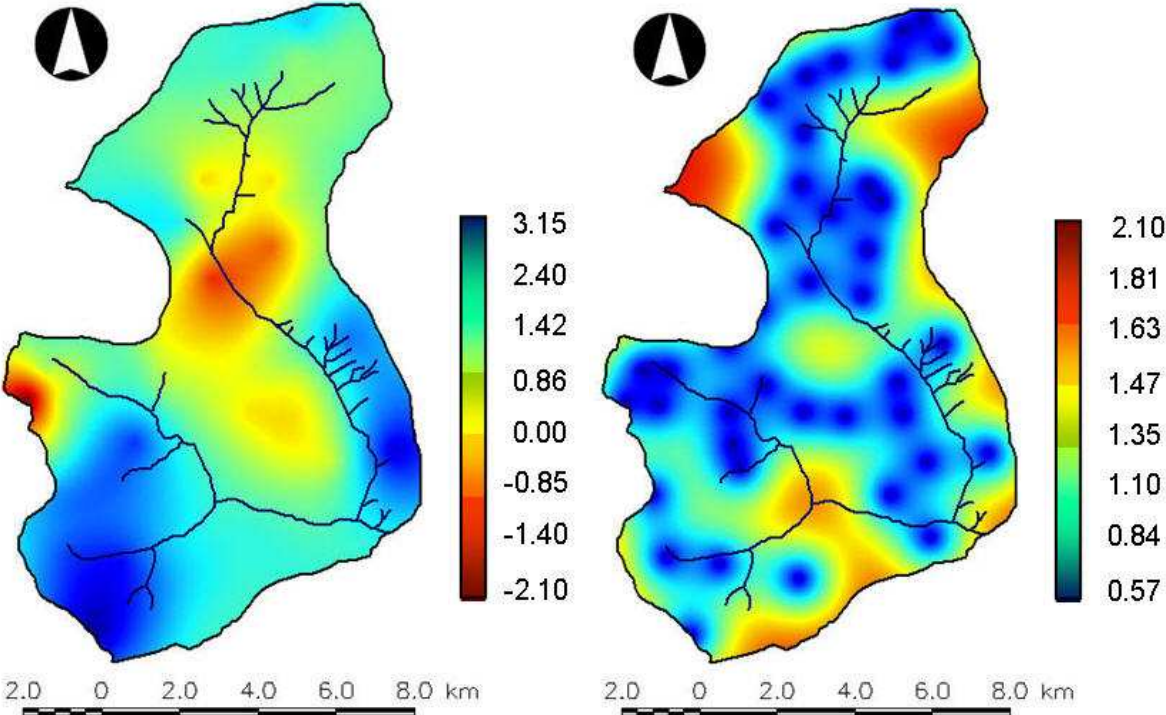
Figure 3: Systematic changes of the water table depths (m) during the period from October 2003 to August 2006 (Left), and the corresponding ordinary kriging standard deviations (Right).
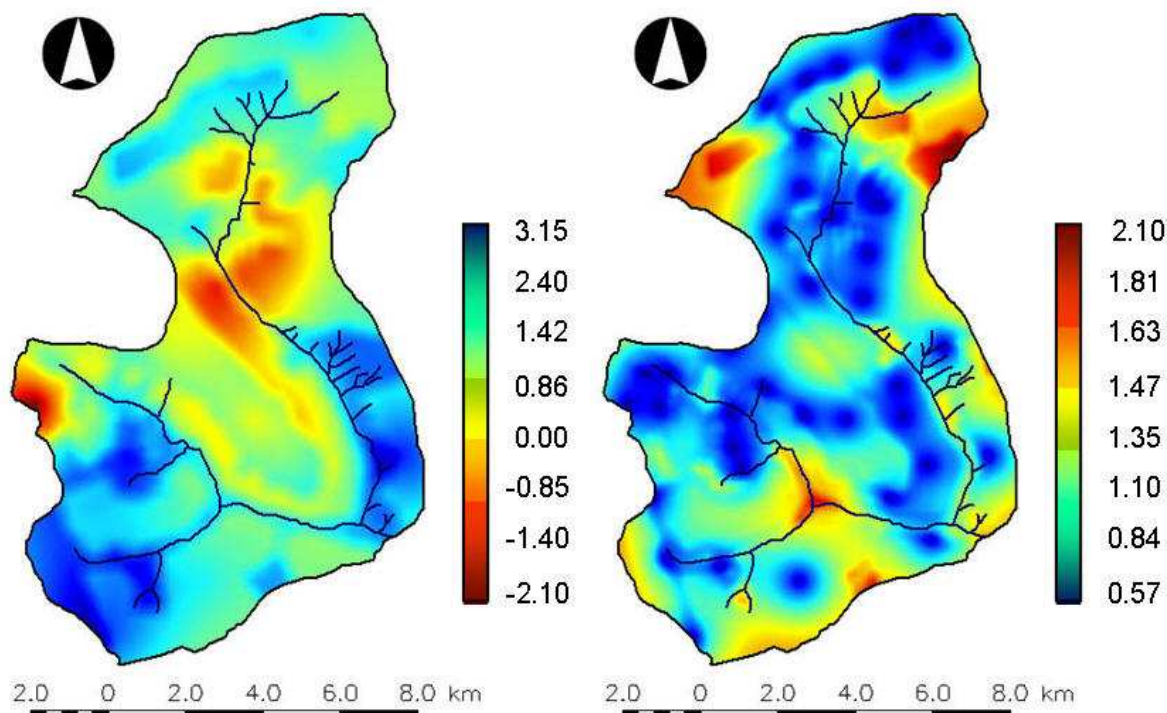
Figure 4: Systematic changes of the water table depths (m) during the period from October 2003 to August 2006 (Left), and the corresponding universal kriging standard deviations (Right).

The results of the ordinary and universal kriging were evaluated by cross-validation. **Table 1** gives the results for ordinary kriging and **Table 2** gives the results for universal kriging.

Table 1: Cross-validation for the spatial interpolation of *LTP* using ordinary kriging.

	Observed	Predicted	Pred. – Obs.	Pred. SD	Z-score
Min	-3.3	-0.7568	-2.86	0.9801	-2.214
1st Q	-0.08762	0.1829	-0.8409	1.064	-0.7972
Median	0.6235	0.744	-0.05554	1.108	-0.05244
3rd Q	1.408	1.054	0.5291	1.179	0.4711
Max	3.63	2.09	4.061	1.332	3.729
Mean	0.6877	0.6404	-0.04728	1.13	-0.01864
SD	1.376	0.662	1.283	0.09335	1.13

Pred.=Predicted; Obs.=Observed; Min=Minimum; 1st Q=First quantil; 3rd Q=Third quantil; Max=Maximum; SD=Standard deviation; *Z*-score=(Pred-Obs) / Kriging variance.

The cross-validation results indicate large interpolation errors, because the standard deviation of the prediction errors is only a bit smaller than that of the observations themselves. These errors can be explained from the uncertainty about the *LTP* parameters at the 40 well locations, the poor relationship between land use and *LTP* and the poor spatial correlation structure in the stochastic residual of the kriging models. Comparing both interpolation methods, the mean standard deviation of universal kriging is a smaller than ordinary kriging. It confirms the choice for the model including LU variables to explain the

spatial distribution of systematic changes in water table depths in Jardim River watershed. The Z-score mean and standard deviation of the Z-score indicate a good performance of the kriging systems, with values close to zero and one, respectively.

Table 2: Cross-validation for the spatial interpolation of *LTP* using universal kriging.

	Observed	Predicted	Pred. – Obs.	Pred. SD	Z-score
Min	-3.3	-0.8321	-2.988	0.9632	-2.4
1st Q	-0.08762	0.2526	-0.8921	1.059	-0.7134
Median	0.6235	0.7263	-0.08821	1.111	-0.0744
3rd Q	1.408	1.02	0.837	1.174	0.7784
Max	3.63	2.661	3.899	1.321	3.641
Mean	0.6877	0.661	-0.02672	1.12	-0.00976
SD	1.376	0.734	1.301	0.09012	1.161

Pred.=Predicted; Obs.=Observed; Min=Minimum; 1st Q=First quantil; 3rd Q=Third quantil; Max=Maximum; SD=Standard deviation; Z-score=(Pred-Obs) / Kriging variance.

The map on **Figure 4** shows a large area near the river where systematic lowering occurs. These areas are covered with traditional agricultural crops, using irrigation systems that catch water directly from the river (surface water). Also, this region has a barrier to stop the river flow and to create a water reservoir for the irrigation systems. For some areas systematic rising of the water table depths was estimated. These risings can be explained as follows. The years 2001, 2002 and 2003 were very dry with 24.4, 41.02 and 33.2% less rainfall than the annual average over the last 31 years, respectively. During 2004 and 2005, rainfall was 8.54 and 4.6% larger than the annual average of the last 31 years, respectively. Therefore, the groundwater system could recharge during the latter period in some areas, resulting in rising water tables.

In the northern part of the basin some areas are found where the Cerrado vegetation still remains, and where systematic risings are indicated. These locations have shallow soils, with slightly fluctuating water tables close to the ground surface. The contribution of this subsystem to the groundwater system of the Jardim river watershed is restricted (Lousada, 2005). The same is true for the area with risings in the eastern part of the basin, which belongs to the same geological system. The degradation of the Cerrado vegetation in these areas could also be a reason of systematic rising of the water table depths, because the degradated vegetation does not use all the water volume that could be explored by the original biomass.

4. Conclusions

The inclusion of satellite images on trends in water table depths estimation reduced the variance of the spatial prediction model. The procedure presents maps more realistic because some physical knowledge about the hydrological process could be incorporated in the analyses.

5. Acknowledgements

program Menyanthes on time series analysis, and Jos Von Asmuth (Kiwa Water Research) for his advices.

6. References

Bishop, T.; McBratney, A. A comparison of prediction methods for the creation of field-extent soil property maps. **Geoderma**, v. 103, p. 149–160, 2001.

Deutsch, C.; Journel, A. G. **Geostatistical Software Library and User's Guide.** New York: Oxford University Press, 1992. 369 p.

Dymond, J. R.; Stephens, P. R.; Newsome, P. F.; Wilde, R. H. Percent vegetation cover of a degrading rangeland from SPOT. **International Journal of Remote Sensing**, v. 13, p. 1999– 2007, 1992.

Gessler, P.; Moore, I.; McKenzie, N.; Ryan, P. Soil-landscape modelling and spatial prediction of soil attributes. **International Journal of Geographical Information Systems**, v. 9, n. 4, p. 421– 432, 1995.

Goovaerts, P. Using elevation to aid the geostatistical mapping of rainfall erosivity. **Catena**, V. 34, p. 227– 242, 1999.

Hoersch, B.; Braun, G.; Schmidt, U. Relation between landform and vegetation in alpine regions of Wallis, Switzerland. A multiscale remote sensing and GIS approach. **Computers, Environment and Urban Systems**, v. 26, p. 113– 139, 2002.

Lees, B. G.; Ritman, K. Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. **Environmental Management**, v. 15, p. 823–831, 1991.

Lousada, E. O. **Estudos hidrológicos e isotópicos no Distrito Federal: Modelos conceituais de fluxo.** 2005. 127 p. PhD Thesis - University of Brasília, Brasília. 2005.

Manzione, R. L.; Knotters, M.; Heuvelink, G. M.B. Mapping trends in water table depths in a brazilian cerrado area. In: Caetano, M. & Painho, M.(Eds.), **Procedings of Accuracy 2006.** Lisboa: Instituto Geográfico Português, 2006, p. 449-458.

Matheron, G. **Le krigeage universel. Cachiers du Centre de Morphologie Mathematique**, v. 1. Fontainebleau: Ecole des Mines de Paris, 1969. 83 p.

McKenzie, N.; Ryan, P. Spatial prediction of soil properties using environmental correlation. **Geoderma**, v. 89, p. 67–94, 1999.

Michaelsen, J.; Schimel, D. S.; Friedl, M. A.; Davis, F. W.; Dubayah, R. C. Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys. **Journal of Vegetation Science**, v. 5, p. 673– 686, 1994.

Moore, D. M.; Lees, B. G.; Davey, S. M. A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. **Environmental Management**, v. 15, p. 59– 71, 1991.

Moore, I.; Gessler, P.; Nielsen, G.; Peterson, G. Soil attribute prediction using terrain analysis. **Soil Science Society of America Journal**, v. 57, p. 443– 452, 1993.

Odeh, I.; McBratney, A.; Chittleborough, D. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. **Geoderma,** v. 63, p. 197– 214, 1994.

Odeh, I.; McBratney, A.; Chittleborough, D. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. **Geoderma**, v. 67, p. 215– 226, 1995.

Pebesma, E. J. Multivariate geostatistics in S: the gstat package. **Computer & Geosciences**, v. 30, p. 683-691, 2004.

Von Asmuth, J. R.; Bierkens, M. F. P.; Maas, C. Transfer function noise modeling in continuous time using predefined impulse response functions. **Water Resources Research**, v. 38 (12), p. 23.1-23.12, 2002.

Wackernagel, H. **Multivariate geostatistics: an introduction with applications**. Berlin: Springer, 1995. 256 p.