Well-yield as training points to model groundwater favorability in a crystalline region of Brazil's semi-arid region

Oderson Antônio de Souza Filho [1]
Adalene Moreira Silva [2]
Mônica Mazzini Perrotta [3]
Anne Elizabeth McCafferty [4]

[1] Serviço Geológico do Brasil – CPRM/REFO
Av. Santos Dumont 1418, 60135-101 - Fortaleza - CE, Brasil
Doutorando do Instituto de Geociências - UNICAMP
oderson@fo.cprm.gov.br

[2] Serviço Geológico do Brasil – CPRM/DISERE
Rua Costa 55, 01304-010, São Paulo - SP, Brasil
perrotta@sp.cprm.gov.br

[3] Laboratório de Geofísica Aplicada, Instituto de Geociências – UnB
Campus Darcy Ribeiro, 70910-900, Brasília – DF, Brasil
adalene@unb.br

[4] United States Geological Survey/USGS
Box 25046, Denver Federal Center, Mail Stop 964, 80225, Denver - CO, US
anne@usgs.gov

Abstract. This work presents a statistical approach to determine how productivities of wells are related to variables of airborne geophysical data, spectral characteristics extracted from ETM$^+$/Landsat-7 images, and water well yield in an area of Brazil's semi-arid region, comprising crystalline rocks. Because groundwater occurrence in the study-area is controlled by fracturing and weathering, the geophysical and spectral signatures of groundwater zones may be different from the host rock. Weights-of-Evidence (WofE) (Bonham-Carter *et al.*, 1989) technique has the ability to quantify the spatial association of an evidence themes with known occurrences of the subject that is studied, therefore, it helped determine a relation for locating regions of groundwater favorability. Favorable groundwater areas were mapped but the few training points diminished the accuracy to identify potential areas located around high-yield wells. Yet, it was determined the relative influence of each theme in model building. The applied methodology is susceptible to be replicated everywhere helping groundwater drilling programs as long as enough training points are available.

Key words: spatial modeling, airborne geophysics, hydrogeology, Ceará, modelagem espacial, geofísica aérea, hidrogeologia, Ceará.

1. Introduction

Airborne geophysical surveys provide data suitable for groundwater exploration, if one or more physical properties can be measured appropriately. Geophysical data interpretation typically involves many processing methodologies, assessment of other geologic and hydrologic information, and integration of all information into meaningful products.

The Juá study-area is located in the Northeast Brazil, at the State of Ceará, at Irauçuba municipality. The geology at Juá comprises migmatites and granites bodies to the north, separated from a sequence with gneisses, amphibolites, granite sheetings and marbles to the south. Quaternary soil cover is limited in thickness to about 2 m near the main creek and about 0.5 m of combined alluvium-colluvium cover across the region (Figure 1).

1.1 Premises

Magnetic and electromagnetic data do not measure the presence or absence of water directly. However, the data have the ability to enhance and map geologic features that can be used to identify probable water-bearing structures and lithologies in crystalline rocks. There are three main points that also illustrate the usefulness of airborne geophysics to map favorable hydrogeological conditions in the crystalline domain:

a)   Magnetic anomaly gradients refer to boundaries between geomagnetic units and define shear zones. Non magnetic anomaly gradients may be interpreted as brittle-related tectonic discontinuities;

b)   Electrically conductive anomaly gradients may be water-filled discontinuities, moisture, salt-rich soil, clay-mineral or graphite content in rock. Non conductive anomaly gradients may represent quartz-filled discontinuities (a hydrological barrier) or non water-filled structures;

c)   Structural lineaments interpreted from remote sensing techniques may be correlated to any of the above geophysical features and they are important for groundwater storage and flow.
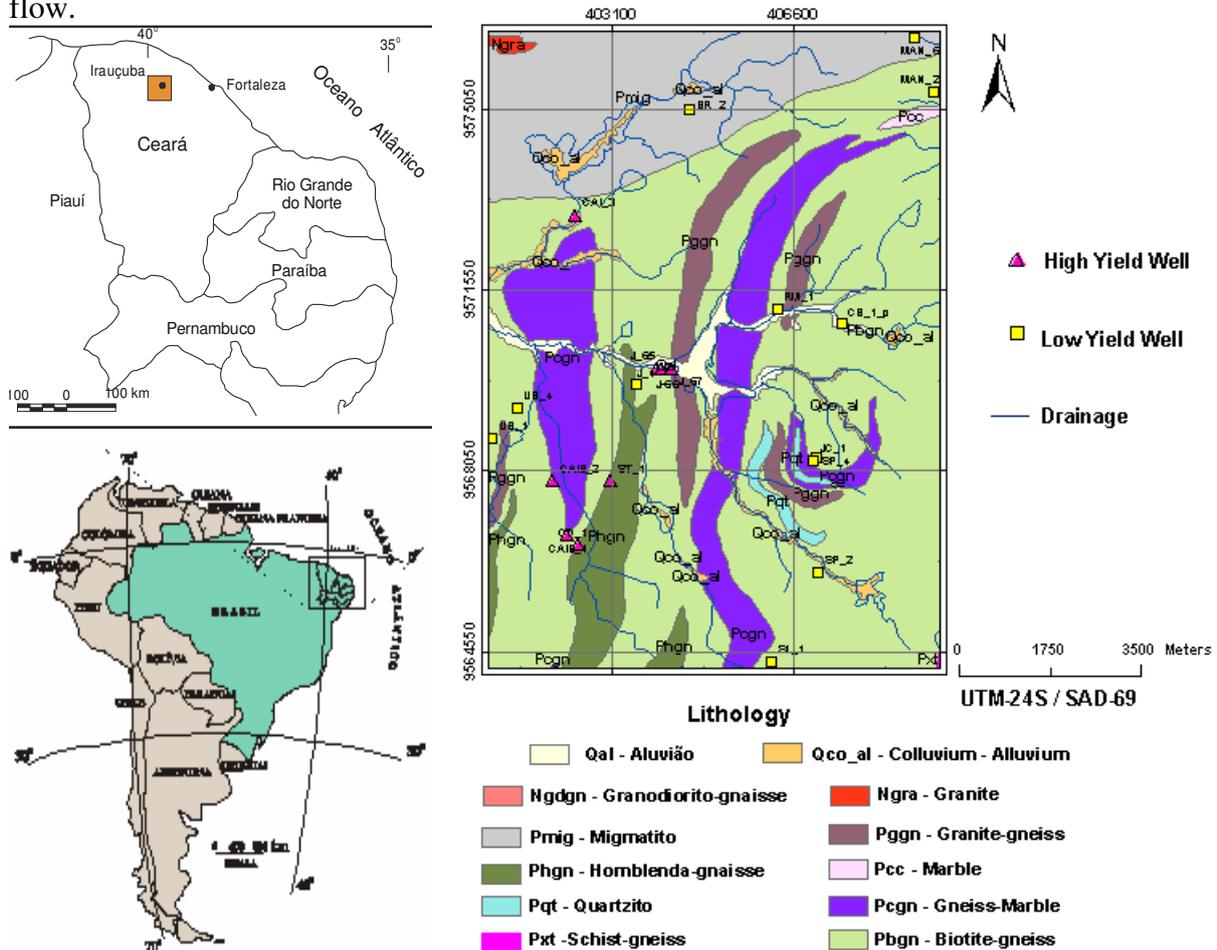
Figure 1. Location of Juá area and, detailed geologic map. (Modified from Souza Filho, 1998).

2. Method

This work was focused in two main procedures: i) data set preparation, comprising geophysical processing techniques, remoter sensing processing, classification of water-wells; ii) the spatial analysis between evidences and well production.

2.1 Data Set Preparation

Two airborne geophysical surveys with different resolution were processed. The lower resolution (Lasa, 1978) provided the radiometric data represented as a 250 m cell-size grid and later re-interpolated to 25 m cell size. The higher resolution helicopter magnetic and electromagnetic-HEM survey (Lasa, 2001) provided grids with 25 m cell size. The magnetic data were acquired with sensibility of 0.001 nT. Using Phillips' (1997) algorithms, the magnetic data were reduced to the magnetic north pole and high-pass filtered to emphasize shallow magnetic sources estimated to be located within 60 m of the topographic surface.

The electromagnetic data were acquired with coaxial coils at 4,500 Hz and, 900 Hz nominal frequencies and, coplanar array at 33,000 Hz, 4,500 Hz and, 900 Hz nominal frequencies. This HEM data was given as apparent electrical conductivity (mS/m) by the contractor (Lasa, 2001) and, lately they were corrected for spurious values and interpolated to a grid with 25m cell-size.

The filtered magnetic and electromagnetic data were transformed to physical property maps of magnetization and electrical conductivity (Cordell and McCafferty, 1989). Positive and negative gradient trends were interpreted from geophysical data and extended 50 m wide as proximity zones and finally, classified in azimuth range. Details from filtering and enhancing techniques applied to the airborne geophysical data of Juá area is found in Souza Filho (2008).

Flight-line data from the regional radiometric uranium (U), thorium (Th), and potassium (K) data were corrected for spurious values, filtered, and gridded.

The ETM+/Landsat-7 images were analyzed by the technique of principal components (Crosta and Moore, 1989) to provide the relative concentration of weathering minerals, basically Fe-oxide (goethite, limonite) and hydroxyl minerals (clay-minerals). Each related image was divided into 3 classes of high, medium, and low content, and low-pass filtered to eliminate outliers and homogenize pixel values. Interpreted lineaments from aerial-photos and ETM+ were classed in to azimuth range and extended by proximity zones with 100 m width then, converted to raster images. All raster data layers have 25-m cell size.

To assess groundwater favorability, well yield-classes were used as the related parameter and their locations were the training sites (Table 1 and Figure 2). It is worth to say that yield (Q, m^3/h) is not the most appropriate hydrogeological parameter of well productivity because it does not account for the drawdown within the well but it was the most representative parameter in the Juá well data.

2.2 Spatial Analysis: the Weights of Evidences technique

The weights of evidence (WofE) method (Bonham-Carter *et al.*, 1989) is a data-driven modeling approach that uses the spatial distribution of known occurrences as training sites to create multi-map signatures from weighted multi-classes or categorical input evidence layers (Bonham-Carter *et al.*, 1989). In this study, the training sites consist of water well locations classified by yield, and the evidence layers are the geophysical, the ETM[+]/Landsat-7 images and the structural geologic data.

The model is based on the idea of prior and posterior probability. The former is the probability that a yield class to occur in the study area, or before considering any existing predictor variables. The latter is an improved response of the prior probability, taking into account distribution patterns from other evidences such as geophysical, spectral, and structural products. The posterior-probability results are then employed to map groundwater favorability.

The prior odd of a yield class is its natural chance to occur within the study area, and it is expressed by the area occupied by the yield class divided by the total area of study. The posterior odd is the probability of a yield class to occur given the presence or absence of a

pattern. The positive and negative weights (w^+ and w^-) are calculated using the Spatial Data Modeler (SDM, Sawatzky *et al.,* 2008). The SDM utilize spatial cross-tabulation between locations of well yield classes and generalized evidential themes (such as a variable being present or absent). A positive weight indicates areas where training points are likely to occur, while negative weight refers to areas where training points and not likely to occur.

The variance of posterior probability can be estimated based on variance of the weights. The difference between w^+ and w^- defines the contrast (C), an overall measure of how well the generalized evidential themes predict training points (Bonham-Carter *et al.*, 1989). Contrast is as a measure of the spatial association between each class of the evidences and the response variable as yield classes.

The uncertainty due to the absence of predictor maps is measured by a confidence level, a ratio of the contrast-C to its standard deviation. It is also called the Studentized-T (Sawatzky *et al.,* 2008), and it is used to test if the hypothesis that the contrast is not zero. The user selects the confidence level appropriate to the problem being considered. For this work, it was considered the confidence level of 1.5 (equivalent to 92% of confidence) to constrain the use of evidence themes that are the least likely to be associated with high-yield wells, without being so restrictive to model building at the same time. To model low-yield favorability areas, a confidence level of 2 was chosen to represent the 98% confidence limit. Following this process, the evidential themes were generalized into multi-classes to assess areas of the evidence that share a greater association with locations of training points.

2.3 Training Sites

Within the Juá area, there are 20 water wells where yield was the most frequent information about water productivity (Table 1a,b; Figure 1). Yield in crystalline rocks is particularly variable and wells from Juá area follow this pattern. Therefore, the wells were grouped into two classes in order to account for the best characterization of well locations: 12 wells as low-yield class (0 to 1 m^3/h) which have and windmill or compressor as pump system; and 5 wells as high-yield class (1 to 6 m^3/h) which are installed with submersible, injection pumps or compressors. Another 3 high-yield wells were used as test points to validate the model.

High-yield class wells have a greater range of values and therefore, they have more dispersed yield values, (see standard deviation and mean parameters in Table 1a,b). Also they are positioned on the western part of the study-area, whereas, low-yield wells, with a lower range and less variability, occur all over the area.

2.4 Evidential Themes

Table 2 comprises all data sets used to modeling high-yield favorability. The tables contain also the original values, the reclassified ranges and the applied generalization for each evidence layer.

In this work, an algorithm by Sawatzky *et al.* (2008) was used to build two models for high-yield favorability and two models for low-yield favorability in ArcMap v. 9.2/ESRI®, with the following goals: i) to test the importance of HEM data, the only geophysical theme directly related to ground-water occurrence due to the positive effect of water on bedrock conductivity and; ii) to cross-validate favorable areas by comparing high- and low- yield favorability models.

Table 1. Juá well data: Yield (m^3/h); EC = electrical conductivity (mS/m). Also identified if used for modeling as training or test point. (Adapted from Veríssimo and Feitosa, 2002 and own data).

a) High-yield wells class. Statistics: mean yield: 3.38 m^3/h; median: 3.25; standard deviation: 2.06

High-Yield Well	Drilled Rock	EC (mS/m)	Yield (m^3/m)	Depth (m)
CAIB_1 (training)	Biotite-gneiss with quartz lens	415.0	1.10	60
CAIB_2 (training)	Biotite-gneiss with quartz lens	558.2	4.50	60
ST_1 (training)	Biotite-gneiss	360.0	1.25	78
CD_1 (test)	Biotite-gneiss, calc-silicate	1050.0	2.00	72
CAI_3 (training)	Biotite-gneiss, granite sheet	203.0	1.20	60
J_65 (test)	Biotite-gneiss, granite-gneiss	274.4	6.0	48
J_66 (training)	Biotite-gneiss, granite-gneiss	603.0	6.0	54
J_67 (test)	Biotite-gneiss, granite-gneiss	117.0	5.0	42

Table 1b) Low-yield wells class. Statistics: mean: 0.36 m^3/h; median: 0.30; stand. dev.: 0.24

Low-Yield Well	Drilled Rock	EC (mS/m)	Yield (m^3/m)	Depth (m)
SL_1	Biotite-gneiss marble lens	200.0	0.3	60
SP_2	Biotite-gneiss, amphibolite	234.0	0.3	7
SP_4	Biotite-gneiss with quartz lens	-----	0.3	-----
JC_1	Biotite-gneiss	150.0	0.7	75
BR_2	Biotite-gneiss, cataclasite	-----	0.7	72
RM_1	Garnet-biotite-gneiss	-----	0.5	-----
CB_1_P	Biotite-gneiss, granite-gneiss	1104.1	0.7	60
UB_1	Biotite-gneiss, granite-gneiss	-----	0.3	-----
UB_4	Biotite-gneiss, marble	-----	0.03	-----
J_1	Biotite-gneiss	-----	0.03	-----
MAN_2	Biotite-gneiss, marble	416.5	0.3	80
MAN_6	Biotite-gneiss	750.0	0.1	60

The evidential themes that did not achieve confidence level > 1.95 (97%) were: magnetic gradient trends from 60 m depth; HEM 4500 Hz; the difference between HEM 4500 – 900 Hz data; non-conductive electromagnetic anomaly trends; radiometric data (K, Th, U); density of structural lineaments; abundances of Fe (OH)-mineral and clay-mineral.

3. Results and Discussions

The calculated post-probability weights were classified according arbitrary divisions as presented in Table 3 and that representation as a favorability map (Figure 2). The Model indicates that the most potential region is concentrated on the WNW-ESE structure that captures the main drainage stream in the central area. Several smaller regions of medium favorability are concentrated to the southwest quadrant of the study-area.

To test the success of the model in mapping groundwater potential, high- and low-yield wells, not used in the formulation of the model were compared with the favorability classes. One high-yield test well was located in a medium favorability area and two others were located in areas of low favorability. Nevertheless, none of the low-yield test wells overlap favorable areas. The model does not satisfactory map groundwater potential because only one high-yield well coincides with highly favorable area. The fact that all of the low-yield wells

overlap not-favorable areas only guarantee that non-favorable areas are most likely to be associated with this class of wells.

Table 2. Evidential themes used for modeling groundwater favorability and their association to high-yield wells.

Evidence theme	Reclassified theme	Generalization threshold and confidence level	Class with positive association	Class with negative association
Terracing of magnetization	30 natural breaks	Cumulative ascending > 1.95	Values < 0.0025	----
Fe(OH) oxide + clay-minerals abundance	3 classes 0< Low < mean; Med. < mean+1SD; High > mean+1 SD	Cumulative ascending > 2	Lower class	Medium and higher classes
Proximity zone of 100 m from structural lineaments	18 classes of azimuth	Categorical > 2	20-40 Az	0-20; 60-80; 100-140 Az
HEM 33000 Hz	10 natural breaks	Cumulative descending > 2	Values > 153 mS/m	Values < 61 mS/m
HEM 900 Hz	10 natural breaks	Cumulative descending > 2	Values > 20	----
Electromagnetic trends	36 classes	Categorical > 2	0-10 Az	Area without gradients
Slope aspect	10 equal classes	Categorical >2	70 - 100	----

Table 3. Posterior-probability values for the groundwater model, classified according to arbitrary divisions. The presence of high- and low- yield wells was quantified for each class.

Probability classes	Pprb Weight Intervals	Cumul. Area (%)	Training wells High-yield n = 5	Test wells High-yield n=3	Test wells Low-yield n = 12
Improbable	0 - 0.06	87.59	1 (20 %)	-	10 (83 %)
Low	0.061 - 0.46	98.50	3 (60 %)	2 (67 %)	1 (17 %)
Medium	0.47 – 0.81	99.35	-	1 (33 %)	-
High	0.82 - 0.95	99.66	-	-	-
Very High	0.96 – 1.00	100	1 (20 %)	-	-

Pprb = posterior probability; Cumul. area = cumulative area

On the other hand, conditional dependences were not verified among the evidential themes used for modeling. The tests followed the CI-ratio calculation from Bonham-Carter (1994) and the Agterberg and Cheng (2002) CI-test (Table 4). These tests are necessary because WofE is an additive Bayesian method, so conditional independence of the evidences regarding training points is mandatory, otherwise the model is overestimated. Expected points are derived from the model and indicate the minimum number of training points to provide that model. Therefore, if T-n = 0 the hypothesis of independence fails.

Comparing these results with a similar modeling approach made by Souza Filho (2008) to the same area, where all the eight high-yield training wells were used instead of five training wells, it appears that the five training wells is not enough to accurately map groundwater potential. That is comprehensible due to heterogeneities inherent to fractured aquifers in crystalline rocks.
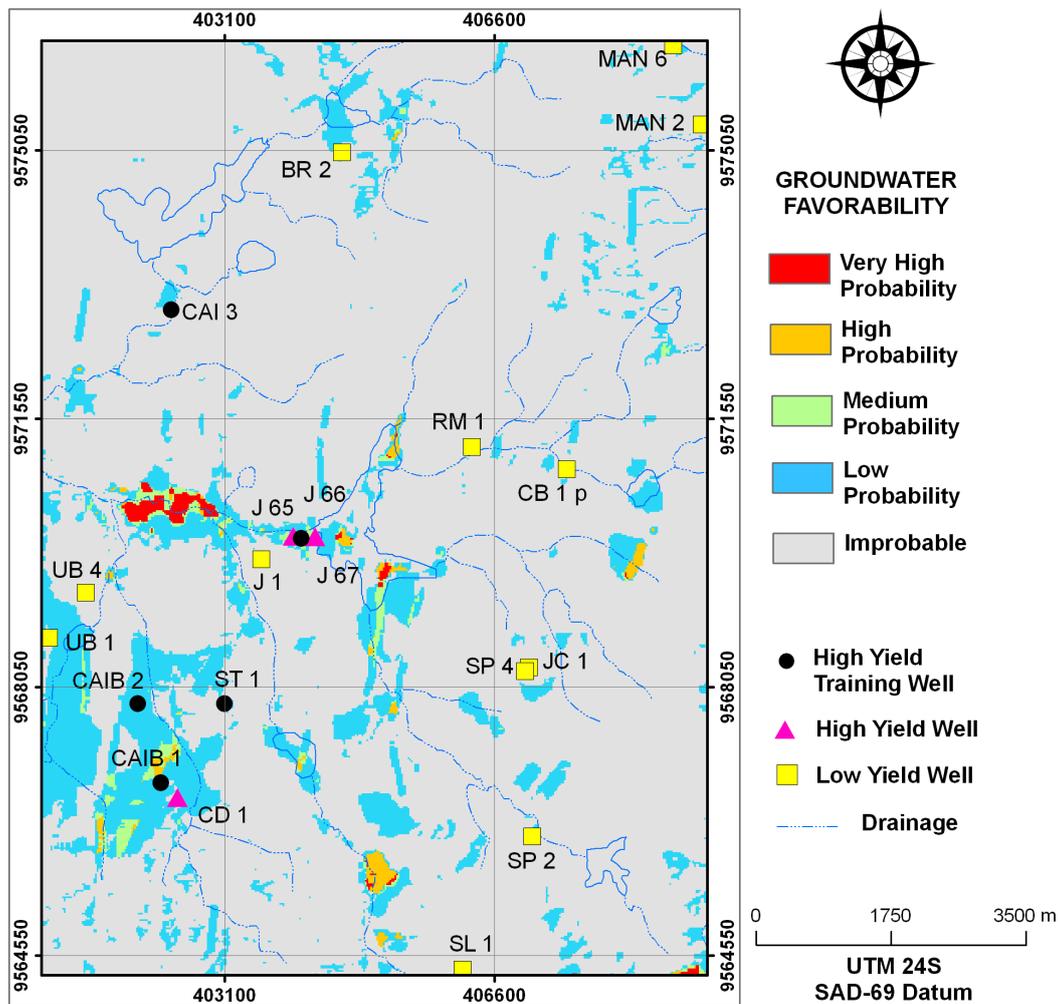
Figure 2. Groundwater favorability model for Juá area calculated by the weights-of-evidences technique. Representation as posterior probabilities ranges: 0 – 0.06, gray; 0.061 – 0.46, blue; 0.47 – 0.81, green; 0.082 – 0.95, orange; 0.96 – 1, red. Best favorability is located to the west and southwest part of the study area.

Table 4. Tests of conditional independence (CI) among two or more evidences calculated for model A1. Training points n = 5; Expected Points, T = 3.4; Difference T-n = -1.6; Standard Deviation of T, Tsd = 1.459.

CI tests	Formula	Result	Observation
CI-ratio	n/T	1.45	Dependence if CI-r < 0.85
CI-test	(T-n)/Tsd	-1.07 (14.3 %)	Dependence if CI-t > 50%

The few training points used to build the model must directly influenced in the spatial associations with evidential themes so that the fact that many evidential themes did not achieved the confidence level of 1.95 does not mean that they can not characterize groundwater potential. Souza Filho (2008) has found more evidence themes associated with groundwater because he used more training points.

4. Conclusions

Groundwater favorable regions were mapped along the main river, which corresponds to a brittle structure with geophysical signature is characterized by non-magnetic but highly

conductive gradient trends at the surface. Radiometric data and terrain elevation data were not able to characterize the differences between high- and low-yield wells, so they did not participate in model building. Low-yield wells were not located in favorable regions, however, the few training points did not accurately map known potential areas, which restrains the model applicability. The use of three more training points considerably raises the model accuracy.

5. References

Agterberg, F. P.; Cheng, Q. Conditional Independence Test for Weights-of-Evidence Modeling. **Natural Resources Research**, v. 11, n. 4, p. 249-255, 2002.

Bonham-Carter, G. F.; Agterberg, F. P.; Wright D. F. Weights of Evidence modeling: a new approach to mapping mineral potential. In Agterberg, F. P., and Bonham-Carter, G. F., eds., **Statistical Applications in the Earth Sciences**: Geol. Survey Canada, Paper 89-9, p. 171–183, 1989.

Bonham-Carter, G. F. Tools for Map Analysis: Multiple Maps. In: Bonham-Carter, G. F., **Geographic Information Systems for Geoscientists**, ch. 9. Ed. Pergamon, Oxford, 398 p, 1994

Cordell, L.; McCafferty, A. E. A terracing operator for physical property mapping with potential field data. **Geophysics**, v. 54, n. 17, p.621-634, 1989.

Crósta, A. P.; Moore, J. McM. Enhancement of Landsat Thematic Mapper Imagery for Residual Soil Mapping in SW Minas Gerais State, Brazil: A prospecting case history in Greenstone Belt Terrain. In: **Thematic Conference on Remote Sensing for Exploration Geology**, Calgary, p.1173-1187, 1989.

LASA Engenharia e Prospecções S.A. **Projeto Aerogeofísico Itatira**, Empresas Nucleares Brasileiras S.A.-NUCLEBRÁS, 3 volumes (Subáreas A, B e C), texto e anexos, Rio de Janeiro, 1978.

LASA Engenharia e Prospecções S/A. **Projeto aerogeofísico água subterrânea no Nordeste do Brasil, blocos Juá (CE), Samambaia (PE) e Serrinha (RN). 2001 - Relatório final do levantamento e processamento dos dados magnetométricos e eletromagnetométricos e seleção das anomalias eletromagnéticas.** Cooperação Canadá-Brasil**,** 3 volumes, 3 CD-Rom. 2001.

Phillips, J.D. Potential-Field geophysical software for the PC, version 2.2: U.S. Geological Survey Open-File Report 97-725, 34 p, 1997.

Sawatzky, D. L.; Raines, G. L.; Bonham-Carter, G. F.; Looney, C. G. **Spatial Data Modeler (SDM): ArcMAP geoprocessing tools for spatial data modelling using weights of evidence, logistic regression, fuzzy logic and neural network**, 2008. Available in <http://www.ige.unicamp.br/sdm/default_e.htm>.

Souza Filho, O. A. de. **Geologia e Mapa de Previsão de Ocorrência de Água Subterrânea. Folha SA.24-Y-D-V Irauçuba, Ceará**. Dissertação de Mestrado. Universidade Federal de Ouro Preto, Brasil. Ouro Preto 99 p. il. e mapas, 1998.

Souza Filho, O. A. de. Dados aerogeofísicos e geológicos aplicados à seleção de áreas favoráveis par água subterrânea no domínio cristalino do Ceará, Brasil. Tese (Doutorado em Ciências) – Universidade Estadual de Campinas, Campinas. 2008.

Veríssimo, L. S.; Feitosa, F. A. C. As Águas Subterrâneas no Nordeste do Brasil. Região de Irauçuba - Estado do Ceará, Brasil. In: Congresso da Associação Internacional de Hidrogeologia, 32. e Congresso da Associação Latino-Americana de Hidrologia Subterrânea, 6., 2002, Mar Del Plata, Argentina. **Anais**...Mar Del Plata, 2002., p. 889-896. CD-Rom.