Springer**Plus**

Open Access

CrossMark

Geostatistical characterization of the soil of Aguascalientes, México, by using spatial estimation techniques

Ricardo Magdaleno-Márquez[1], María de la Luz Pérez-Rea[1] and Víctor M. Castaño[2]*

*Correspondence: meneses@unam.mx
[2] Centro de Física Aplicada y Tecnología Avanzada, Universidad Nacional Autónoma de México, Boulevard Juriquilla, 3001 Santiago De Querétaro, QRO, Mexico
Full list of author information is available at the end of the article

Abstract

Four spatial estimation techniques available in commercial computational packages are evaluated and compared, namely: regularized splines interpolation, tension splines interpolation, inverse distance weighted interpolation, and ordinary Kriging estimation, in order to establish the best representation for the shallow stratigraphic configuration in the city of Aguascalientes, in Central Mexico. Data from 478 sample points along with the software ArcGIS (Environmental Systems Research Institute, Inc. (ESRI), ArcGIS, ver. 9.3, Redlands, California 2008) to calculate the spatial estimates. Each technique was evaluated based on the root mean square error, calculated from a validation between the generated estimates and measured data from 64 sample points which were not used in the spatial estimation process. The present study shows that, for the estimation of the hard-soil layer, ordinary Kriging offered the best performance among the evaluated techniques.

Keywords: Spatial interpolation, Splines, Standard Kriging, Geostatistics, Geographical infromation systems, Aguascalientes

Background

The city of Aguascalientes, in a state of the same name, in central Mexico, is nestled in a valley alluvial (Aranda-Gómez and Carrillo 1983; Aranda-Gómez and Aranda-Gómez 1985), so subsoil consists of a mixture of coarse and fine soils, the latter being posed a challenge to builders (Arroyo-Contreras et al. 2004). In the Valley of Aguascalientes the problem is the presence of surface deposits collapsible settlements and causing damage to structures.

Geographic information systems are effective tools for geotechnical engineering as all information used has spatial attributes (Player 2006), thus facilitating the implementation of spatial interpolation techniques and geostatistics for estimating soil properties as the collapsible layer thickness surface. There are a wide range of spatial estimation techniques, but not easy to determine which one will yield better results as the phenomenon to be studied.
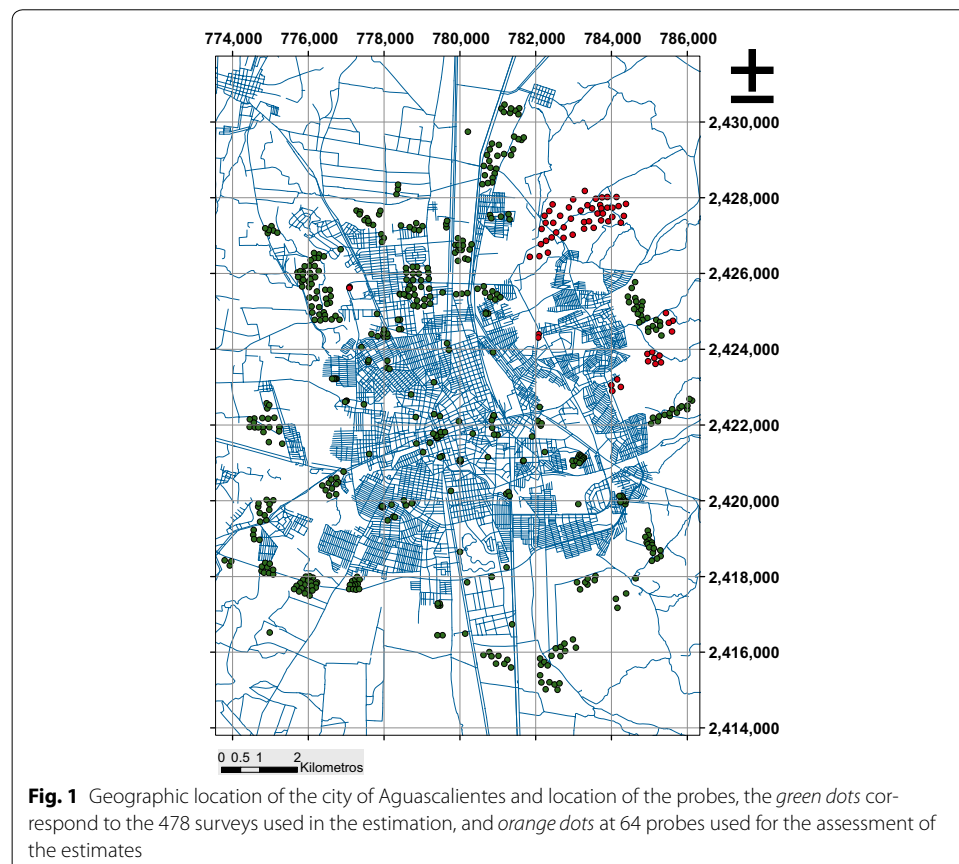
In the scientific literature has tried to resolve this issue, using two basic methods. The first of these studies consists of reconciliation, or to compare the results of the estimation results with different methods of production (Healey 1993; Sides 1992). The other

Magdaleno-Márquez *et al. SpringerPlus (2016) 5:918*

Page 2 of 13

method is to simulate a sector, testing different estimation techniques and compare the results with the simulated reality (Bell and Whateley 1994).

The objective of this study is to evaluate and compare four spatial estimation techniques: the regularized spline interpolation, by splines in tension, by the inverse distance weighted and ordinary Kriging estimation to determine which best suits the surface stratigraphy settings in the city of Aguascalientes.

Geographical location

The city of Aguascalientes is located between the extreme projected coordinate X1 = X2 = 2416000 and 2430000 north latitude, and between Y1 = Y2 = 774.000 and 786.000 west longitude coordinate system Universal Transverse Mercator, North American Datum of 1927 (UTM-NAD27, Zone 13 North) (Fig. 1), with an average elevation of 1870 m above mean sea level, nestled in the Valley of Aguascalientes which is a morphological subunit crosses the east-central part of the state, with an orientation substantially NS, and that an elongated depression is between 10 and 20 km. wide, extending beyond state boundaries, from south of Zacatecas to the eastern end of Jalisco (Arroyo-Contreras et al. 2004). Depression is an alluvial plain formed by more or less regular reaching an altitude between 1850 and 1900 m levels, where the river runs central Aguascalientes (also known as San Pedro River), which flows from north to south and is the most important state flow.

Fig. 1 Geographic location of the city of Aguascalientes and location of the probes, the *green dots* correspond to the 478 surveys used in the estimation, and *orange dots* at 64 probes used for the assessment of the estimates

Magdaleno-Márquez *et al. SpringerPlus* (2016) 5:918

Page 3 of 13

The Valle de Aguascalientes corresponds to a tectonic graben structure, known in geological literature as the "Graben of Aguascalientes", bounded on the east and west by two large faults oriented north–south, this is not no rock formation emerges as filling materials are unconsolidated alluvial deposited on a bedrock these fillers thickness reaches more than 600 m. in the deeper parts and not less than 200 m. in the more shallow (Aranda-Gómez and Aranda-Gómez 1985; Arroyo-Contreras et al. 2004).

From 1979 cracks were detected in the urban area of the city of Aguascalientes and other towns in the surrounding area as Jesus Mary and Puertecito Chicalote of the Virgin, all located in the Quaternary alluvial fill (Aranda-Gómez and Aranda-Gómez 1985). The direction of these cracks coincides with the general direction of the valley with maximum vertical displacement of 1 m or more. Invariably, with few exceptions, the downthrown the crack is located to the west of the city, that is, toward the San Pedro River. On the eastern edge of the city is located a normal fault with approximately NS direction, which is considered that originated the Graben of Aguascalientes, from this failure, the surface of the city has an east–west slope greater than 1 %.

Methods

The computer software ArcGIS (ESRI 2008) was used to implement the spatial estimation techniques by regularized splines, Splines in tension, weighted inverse distance and ordinary kriging estimation, and determine which one best represents the stratigraphy surface conditions of the city of Aguascalientes.

The data used in estimating the records consist of 478 polls open pit type (PCA) performed within the urban area of the city (Fig. 1), which were obtained from the files of the Department of Splits Development Secretariat Aguascalientes urban township, and company Integral Civil Works Supervisor, S. A. de CV, from the same sources were obtained records of 64 polls PCA type (Fig. 1) with additional assessment was made of each technique.

The open pits are scattered all over, with no obvious pattern of the spacing between them, as they were made for particular studies and do not share the same information among each other. The depth of the open holes varies between 3.00 and 5.00 m. Generally speaking, two layers can be identified: a collapsible surface layer composed mainly of clay or silt of low compressibility (CL or ML) and a second resistant layer, with high stiffness, composed by cemented silty sands or clayey materials (SM or SC). The depth of the hard layer is the basic parameter for the analysis. The mechanical properties of the soil obtained from a stratigraphic profile were used to establish the depth of this layer.

The data collected were synthesized in a database created with the coordinates of all the polls and the depth of the hard layer (surface layer thickness) in meters represented negative (−m) to characterize the relief of the hard layer (composed silty sand or gravel-sand). For the location of the probes was used UTM projected coordinate system, with the North American Datum of 1927 (Fig. 1).

The hard layer depth was estimated using four techniques, obtaining for each estimate a raster file which was processed and compared with data from the probes 64 prepared for evaluation, using the calculation of the root mean square error (RMSE) to evaluate and compare each technique. RMSE was chosen as a benchmark as it retains the same units as the input data.

Magdaleno-Márquez *et al. SpringerPlus* (2016) 5:918

Page 4 of 13

Spatial estimation techniques

The hard layer depth was estimated using the four techniques mentioned above. Since the base of each appears widely described in the literature (Isaaks and Srivastava 1989; Armstrong 1998; Webster and Oliver 2001), only given general elements each.

Regular splines and splines in tension

These are techniques which estimate deterministic interpolation values using a function that minimizes the overall curvature of the surface, resulting in a smooth surface that passes exactly through all known points of entry.

The basic shape of the spline interpolation imposes two conditions on the interpolation function:

1. The surface must pass exactly through the input data.
2. The surface must have a minimum curvature: the sum of the squares of the second derivatives calculated at each point of the surface should be as short as possible.

Splines There are two techniques: the regularized and the voltage in the regularized creates a smooth, gradual change with values that may be generated outside the range of the data sample, and the tensile stiffness is controlled surface according to the nature of the phenomenon to model, it creates a smooth surface with values less linked to the range of the sample data.

Splines regularized technique modifies the criterion of minimizing the curvature, incorporating the third derivative and assigning a weight, referred to as τ in the literature. This method is especially useful if required the calculation of the second derivatives of the surface of interpolation.

Splines technique used in tension arising in the first criterion of minimizing the curvature, assigning a weighting referred to as ϕ in the literature. The generated surface is continuous and smooth, while the first derivative is continuous, but is not smooth.

Both techniques use the following expression Splines to generate surface interpolation:

$$S(x,y) = T(x,y) + \sum_{j=1}^{N} \lambda_j R(r_j) \tag{1}$$

where N, is the number of input points, λ_j are the weight coefficients and r_j is the distance between the point to be estimated *(x, y)* and the *j input point.*

$T(x, y)$ y $R(r_j)$ are defined according to the specific Spline used, as follows:for the regular splines:

$$T(x,y) = a_1 + a_2 x + a_3 y \tag{2}$$

$$R(r) = \frac{1}{2\pi} \left\{ \frac{r^2}{4} \left[\ln\left(\frac{r}{2\tau}\right) + c - 1 \right] + \tau^2 \left[K_0\left(\frac{r}{\tau}\right) + c + \ln\left(\frac{r}{2\pi}\right) \right] \right\} \tag{3}$$

for the splines in tension:

$$T(x,y) = a_1 \tag{4}$$

Magdaleno-Márquez *et al. SpringerPlus (2016) 5:918*

Page 5 of 13

$$R(r) = -\frac{1}{2\pi\varphi^2}\left[\ln\left(\frac{r\varphi}{2}\right) + c + K_0(r\varphi)\right] \tag{5}$$

where τ y ϕ are the weighing parameters, r is the distance between the point to be estimated and the input point, K_0 is the modified Bessel function, $c = 0.577215$, and a_i are coefficients of the associated system of linear equations.

Weighted inverse of distance

This method implicitly assumes that parts of the data that are nearby are more similar than those further away. To estimate a value at a given point using a linear combination of neighboring data, so it is assumed that each input has a local influence that decreases with distance.

The general term used for this method is:

$$z^*(x_0) = \sum_{i=1}^{N} \lambda_i z(x_i) \tag{6}$$

where x_0 is the point where the estimation is to be made, $z(x_i)$, $i = 1, 2, ..., N$, are the actual measurements of the point x_i, and λ_i are the weight coefficients associated to each measurement; the estimation itself is symbolized by $z^*(x_0)$.

The weighting coefficients are calculated based on the inverse of the distance between the target point and the locations $x_0\ x_i$ input measurements:

$$\lambda_i = 1/|x_i - x_0|^\beta \quad \text{con}\beta > 0 \tag{7}$$

The preferred choice for β is 2, thus, the estimation corresponds to the square of the inverse of the distance.

The only requirement of this method is that the sum of weights equals λ_i unit. Thus, if $x_0\ x_i$ match then the weight for that point becomes unitary and $x\ z^*$ takes the value $0\ z\ (x_i)$, so that interpolation is accurate.

An attractive feature of this technique is that the weights decrease rapidly as the distance increases, so that the local estimation is significantly further since the weights never become null no discontinuities. However, the selection of the weighting function is arbitrary, and there is a measure of the estimation error. Furthermore, the created surface always has a zero gradient in the entry points, so that the maxima and minima occur only at those sites.

Standard Kriging

Kriging methods rely on mathematical and statistical models, and it is precisely the addition of statistical models which separates deterministic Kriging methods, the Kriging uses or functions of semi-variance or variance as a function of distance, which is characteristic geostatistics distinctive and allows models based on this semi-variance.

The technique of ordinary Kriging is to use estimators that are linear combinations of the data: for point estimates of the variable V (X) at a point x0 we use the estimator $V^*\ \boldsymbol{x}$ **0**, as a linear combination of known data $V\ (x_i)$:

Magdaleno-Márquez *et al. SpringerPlus* (2016) 5:918

Page 6 of 13

$$V^*(x_0) = \sum_{i=1}^{n} \lambda_i V(x_i) \tag{8}$$

To ensure that the estimator is unbiased, the sum of the weights must be equal to unity:

$$\sum_{i=1}^{n} \lambda_i = 1$$

And the expected error should be zero:

$$E[V(X) - V^*(X)] = 0$$

Thus, the variance of the estimate is:

$$Var[V(x_0)] = E\left\{(V(X) - V^*(X))^2\right\} = 2\sum_{i=1}^{N} \lambda_i \gamma(x_i, x_0) - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \gamma(x_i, x_j) \tag{9}$$

where $\gamma(xi, xj)$ is the semi-variance of V (X) between the data points $x\,i$ and $x\,j$, and $\gamma(x\,i, x0)$ is the semi-variance between the i-th data point and the target point $x\,0$, with the following expression for the semi-variance:

$$\gamma(h) = \frac{1}{2}E\left\{[V(x_i) - V(x_i + h)]^2\right\} \tag{10}$$

where h is a scalar equal to the distance between the points and $x\,j\,x\,i$.

For each estimated value exists kriging variance associated, which we denote as $\sigma^2(0)$, defined by Eq. (9). The next step is to determine the coefficients which minimize $\lambda\,i$ variance value, this is done using the method of Lagrange multipliers.

First we define an auxiliary function $f(\lambda\,i, \psi)$ containing that we want to minimize the variance plus a term that contains a multiple of Lagrange, ψ, for ordinary Kriging this function is:

$$f(\lambda_i, \psi) = Var[V(x_0) - V^*(x_0)] - 2\psi\left\{\sum_{i=1}^{n} \lambda_i - 1\right\} \tag{11}$$

After defining the partial derivatives of the function with respect to the coefficients λ and ψ regarding:

$$\frac{\partial f(\lambda_i, \psi)}{\partial \lambda_i} = 0,$$

$$\frac{\partial f(\lambda_i, \psi)}{\partial \psi} = 0,$$

for i = 1, 2,..., N. This leads to a system of N + 1 equations with the same number of unknowns:

$$\sum_{i=1}^{N} \lambda_i \gamma(x_i, x_j) + \psi(x_0) = \gamma(x_j, x_0) \tag{12}$$

Magdaleno-Márquez *et al. SpringerPlus* (2016) 5:918

Page 7 of 13

The solution of this system of equations provides the coefficients $\lambda\,i$, which are used in Eq. (8) for the estimates.

Evaluation and comparison

To compare the results obtained by each technique was used root mean square error (RMSE):

$$\text{RECM} = \sqrt{\frac{1}{N} * \sum_{i=1}^{n} \left(v_i^* - v_i\right)^2} \tag{13}$$

where N is the number of validation points, and $v\,i$ and $v\,i^*$ are estimated and measured values at each location i.
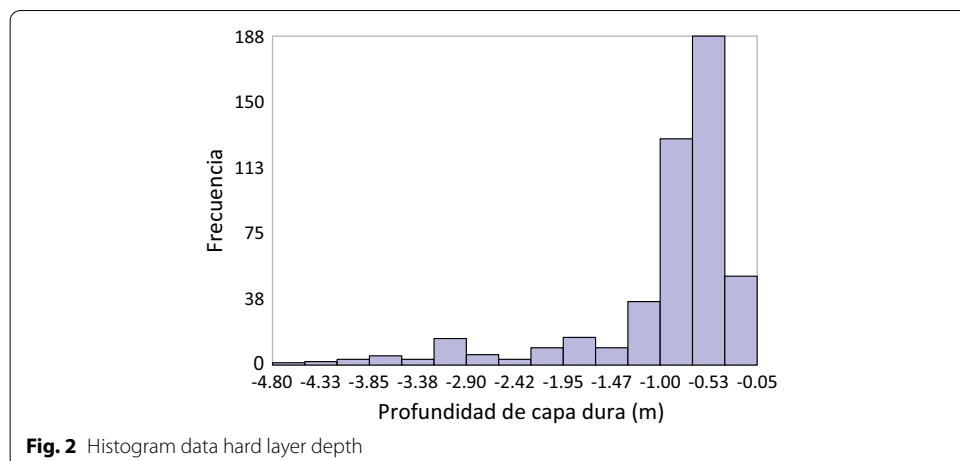
Results and discussion

Table 1 shows the basic statistical data of the layer depth.
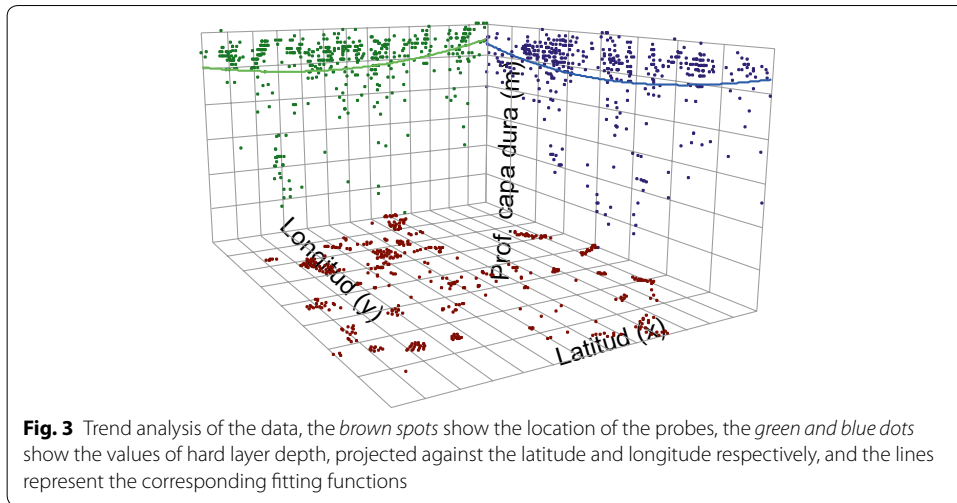
The data distribution is asymmetrical negative as the average is less than the median, having a bias towards the left, it can be seen that the shape of the histogram (Fig. 2).

We also carried out an analysis of spatial trend, showing a clear trend in the data, which was modeled by a quadratic polynomial fit (Fig. 3). The trend analysis and the analysis of the distribution were useful for the implementation of the estimation by

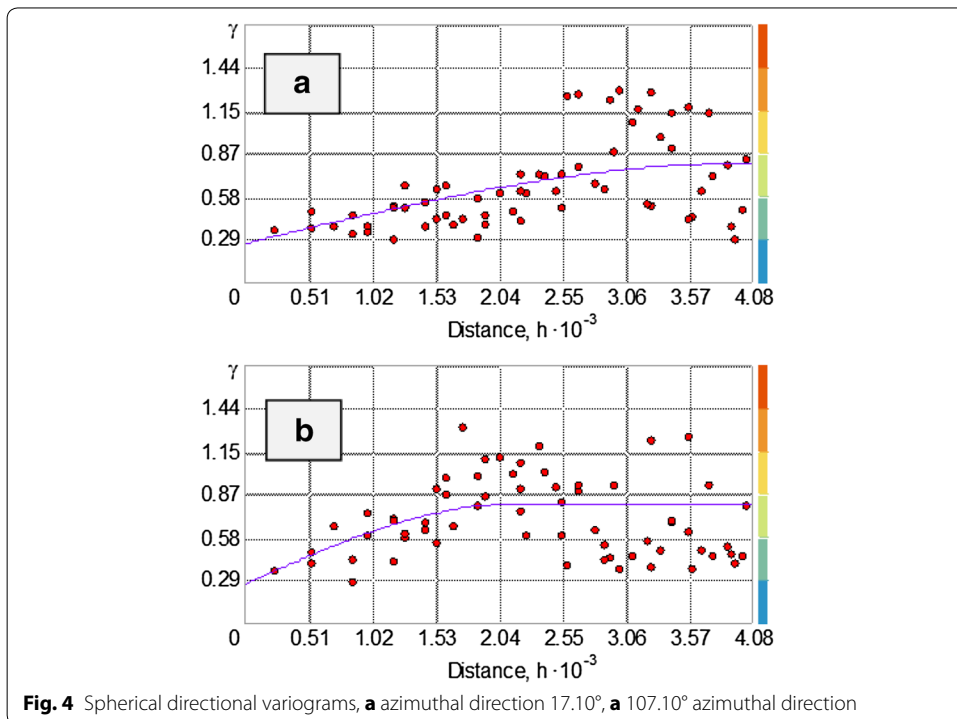### Table 1 Statistical description of depth of hard layer (m)

Media	−0.9323
Median	−0.675
Mode	−0.50
Standard deviation	0.8111
Variance	0.6579
Range	4.75
Minumum	−4.80
Maximum	−0.05
Count	478

Fig. 2 Histogram data hard layer depth

Magdaleno-Márquez *et al. SpringerPlus* (2016) 5:918

Page 8 of 13

Fig. 3 Trend analysis of the data, the *brown spots* show the location of the probes, the *green and blue dots* show the values of hard layer depth, projected against the latitude and longitude respectively, and the lines represent the corresponding fitting functions
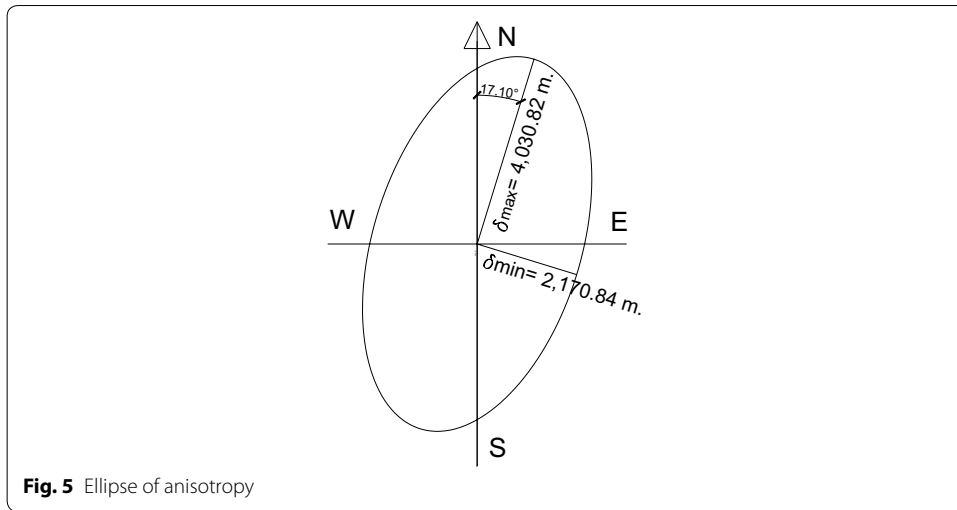
ordinary kriging, since this technique is based on statistical parameters and spatial variability (Isaaks and Srivastava 1989; Webster and Oliver 2001).

The variograms adjusted (Fig. 4) permitted the identification anisotropy in the spatial variability since not all the directions have the same distance phenomenon correlation, the correlation distance is minimum is in the direction of azimuthal = 107.10° and is 2170.84 m. and the higher correlation distance of 4030.82 m. in the azimuth = 17.10° (Figs. 4, 5) could also capture the variability in distance below the sample, through the "nugget effect" (or "nugget" in English) found 0256.

Fig. 4 Spherical directional variograms, **a** azimuthal direction 17.10°, **a** 107.10° azimuthal direction

Magdaleno-Márquez *et al. SpringerPlus* (2016) 5:918
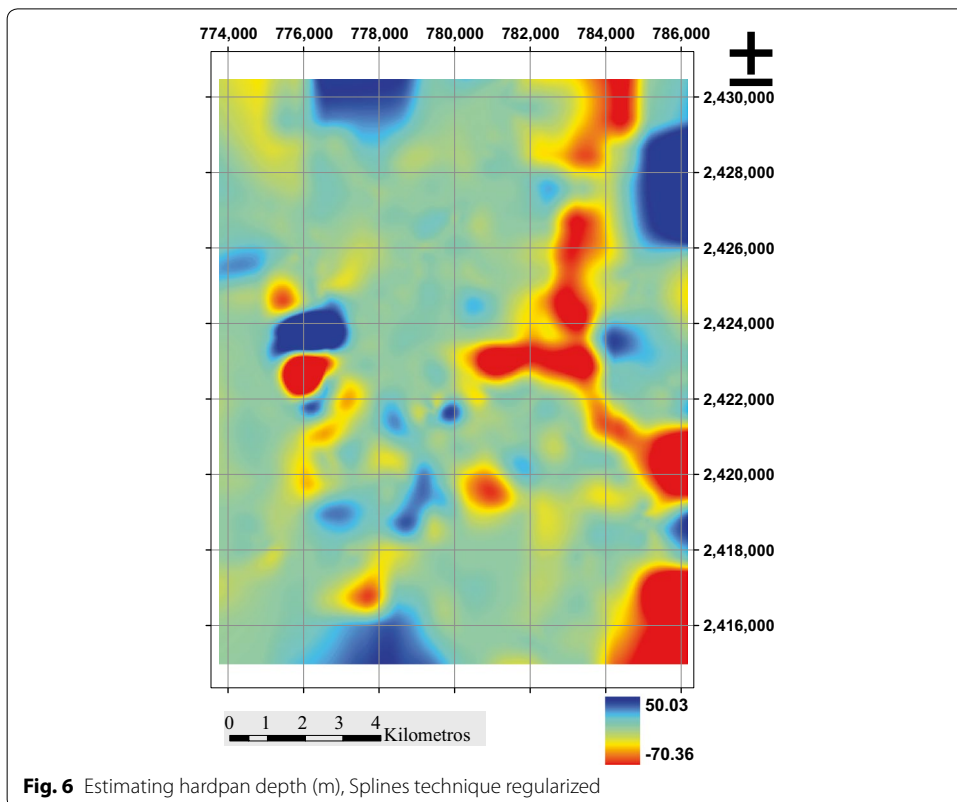
Page 9 of 13

Fig. 5 Ellipse of anisotropy

Regularized splines

The estimate generated by this technique presents very erratic values and even positive (Fig. 6), the phenomenon being modeled is hard layer depth in meters negative (-m). As RMSE values, this technique had the highest value of four, with RMSE = 4.937 (Table 2), so that this technique is more yielding inaccurate results.
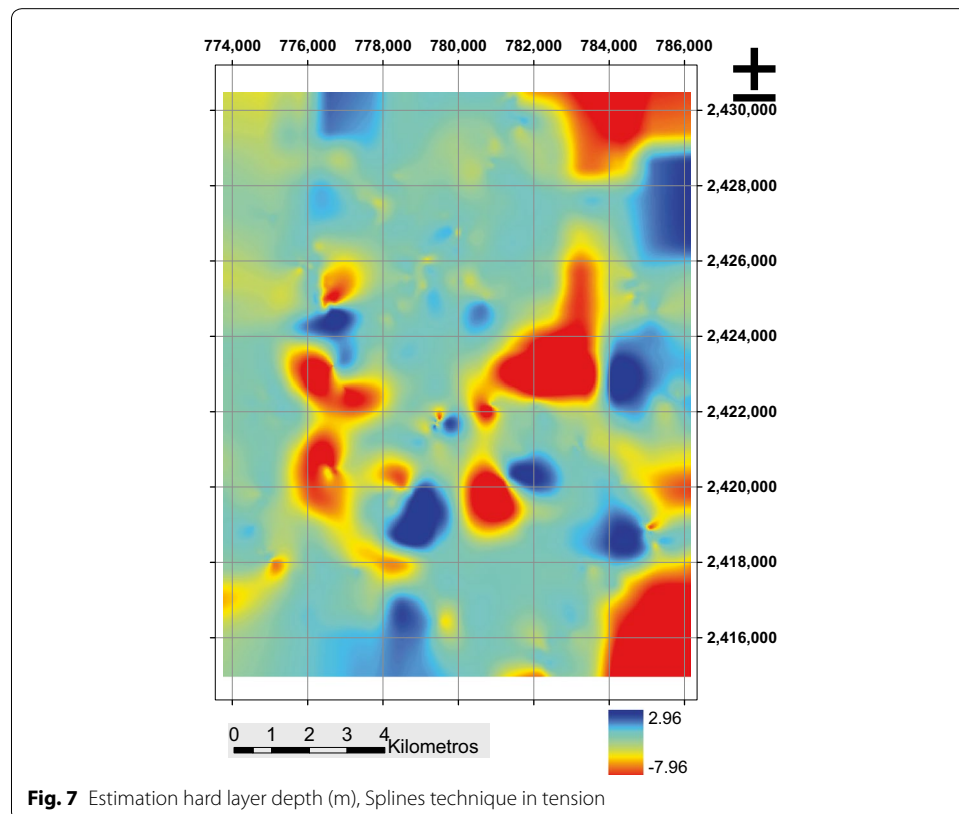

Fig. 6 Estimating hardpan depth (m), Splines technique regularized

Magdaleno-Márquez *et al. SpringerPlus* (2016) 5:918

Page 10 of 13

Table 2 Square root coefficients (RECM)

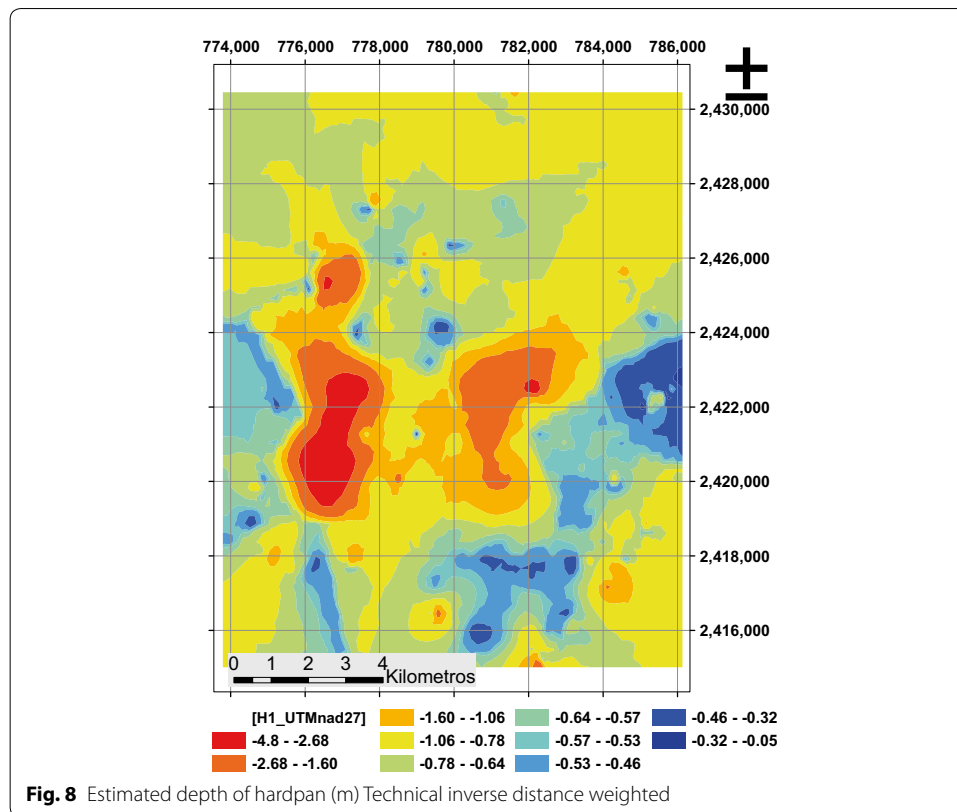| Technique	RECM
Kriging ordinary	0.814
Weiughted inverse of distance	0.843
Spline regular	4.937
Spline in tension	1.237

Splines in tension

The estimate generated by this technique has less erratic values, but still has some positive values (Fig. 7). For RMSE, this technique had the second highest value of 1.237 (Table 2) so the relief representation of the hard layer was acceptable either.

Weighted inverse technique

The estimate generated by the technique of Inverse weighted distance presents quite acceptable values at first, since they fail to generate positive (Fig. 8), but from the RMSE a value of 0.843 was obtained (Table 2), which is the third highest value, so, despite being more accurate than the two other techniques, Splines does not offer an accurate representation of the hard layer.

Fig. 7 Estimation hard layer depth (m), Splines technique in tension

Magdaleno-Márquez *et al. SpringerPlus* (2016) 5:918

Page 11 of 13

Fig. 8 Estimated depth of hardpan (m) Technical inverse distance weighted

Standard Kriging

The estimate generated from ordinary Kriging technique has very acceptable values, as with the inverse distance weighted positive values are not generated, and as to the RMSE have the lowest value of all 0.814 (Table 2), so is the characterization technique that generated the relief of the hard layer more accurately (Fig. 9).
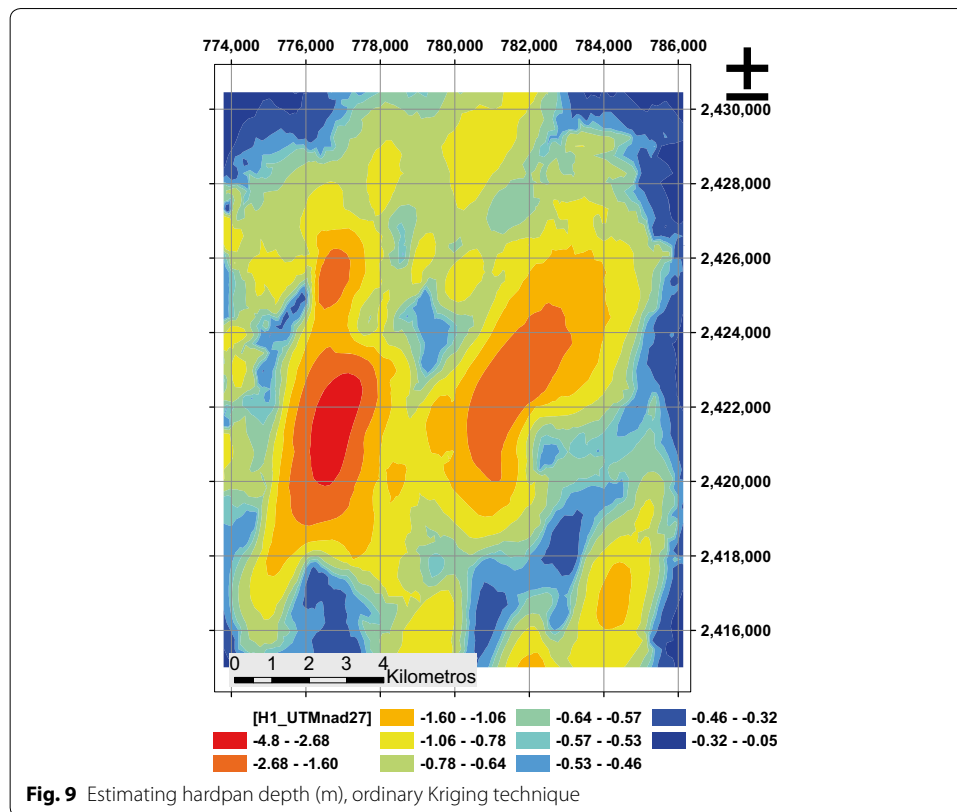
Conclusions

The methodology used to evaluate the accuracy of different estimation techniques and spatial interpolation in order to decide which one provides the most accurate results.

The study shows that within deterministic methods to the case study, the technique of inverse distance weighted yielded the most accurate results, with a value of root mean square error (RMSE) of 0.843, while the technical regularized spline was the most inaccurate of all, with a RMSE of 4.937.

Noteworthy is the large difference between the RMSE values reported by regularized spline technique (RMSE = 4.937) and spline in tension (RMSE = 1.237).

The ordinary kriging estimator is the space that offered the most accurate results among the four techniques evaluated, with the lowest value of the root mean square error RMSE = 0.814.

These methods were chosen, among several techniques available in the literature, namely inverse weighted distance, regularized spline, tension spline and ordinary kriging, to estimate the depth of the hard layer of the Valley of Aguascalientes due to its popularity and robustness as interpolation methods for spatial prediction of variables

Magdaleno-Márquez *et al. SpringerPlus* (2016) 5:918

Page 12 of 13

Fig. 9 Estimating hardpan depth (m), ordinary Kriging technique

in similar cases reported. Indeed, the technique of weighted inverse distance provides average values depending on the distance, perhaps this technique works better when the unknown points are closer to the sampling, so it is necessary to have a high density of known points. Regularized spline and tension spline tension interpolation functions, even when passing through the points of the model, are not able to predict the values of the checkpoints because of the roughness of the functions. These techniques provide good results when the variation of the surface is slight. Tension spline technique provides a closer approximation as modifies the minimization criterion incorporating the terms of the first derivative considering more abrupt variations without sacrificing the smoothness of the function. This technique is used when you want to preserve shape properties.

Ordinary Kriging technique, unlike the previous ones, uses random field theory to establish the spatial correlation of the variables achieving better prediction with a stochastic model. The advantage of the Kriging is that ensures that the average errors is zero. In ordinary Kriging, local media should not necessarily resemble the average of the entire population, so just take the closest points to make the interpolation, so the variables are correlated locally and not with the entire population. For this reason, ordinary Kriging is widely used for the prediction of variables whose value depends less on all samples collected, but the nearest properties such as geographic regions samples.

The hard layer depth in the valley of Aguascalientes has a clear spatial structure, capable of being modeled by adjusting variograms for estimation with ordinary Kriging technique.

Magdaleno-Márquez *et al. SpringerPlus* (2016) 5:918

Page 13 of 13

Generally, it is recommended for the prediction of properties of clearly defined geographical areas, where properties to predict not depend greatly from the distant properties, but local, such as processes of formation of soil, deposition sequences and climatic focalized actions, so ordinary Kriging be preferred over other techniques.

Authors' contributions

RM-M collected data and made the in-field measurements, MLP-R collected in-field data and registered and interpreted the georeferenced data, VMC designed the experiment, contributed to the interpretation of all data and the mathematical modelling. All authors read and approved the final manuscript.

Author details

[1] División de Estudios de Posgrado, Facultad de Ingeniería, Universidad Autónoma de Querétaro, Cerro de las Campanas s/n, Col. Niños Héroes, 76010 Querétaro, QRO, Mexico. [2] Centro de Física Aplicada y Tecnología Avanzada, Universidad Nacional Autónoma de México, Boulevard Juriquilla, 3001 Santiago De Querétaro, QRO, Mexico.

References

Aranda-Gómez JM, Aranda-Gómez JJ (1985) Análisis del Agrietamiento en la Ciudad de Aguascalientes. Aguascalientes, Ags., Universidad Autónoma de Aguascalientes, Centro Tecnológico, Departamento de Geotecnia, reporte anual

Aranda-Gómez JM, Carrillo JM (1983) El Subsuelo de la ciudad de Aguascalientes", Aguascalientes, Ags., Universidad Autónoma de Aguascalientes, Centro Tecnológico, Departamento de Geotecnia

Armstrong M (1998) Basic linear geostatistics. Springer, Berlin

Arroyo-Contreras GM, Zermeño-de León ME, Castañeda-Alvarado JA (2004) El Agrietamiento en Aguascalientes: Aguascalientes, Universidad Autónoma de Aguascalientes, Instituto del Agua del Estado de Aguascalientes, Colegio de Ingenieros Civiles de Aguascalientes, Universidad Nacional Autónoma de México, Universidad Autónoma de Querétaro

Bell TM, Whateley KG (1994) Evaluation of grade estimation techniques. In Whateley KG (ed) Case histories and methods in mineral resource evaluation II: Londres, Geological Society Special Publication no. 79, pp 67–86

Environmental Systems Research Institute, Inc. (ESRI) (2008) ArcGIS, ver. 9.3, Redlands, California

Healey CM (1993) Performance of reserve estimation techniques in the presence of extremely high grade samples, jasper gold mine, saskatchewan: exploration mining. Geology 2(1):41–47

Isaaks EH, Srivastava RM (1989) An introduction to applied geostatistics. Oxford University Press, New York

Player RSV (2006) Geographic Information System (GIS) Use in Geotechnical Engineering. In: American Society of Civil Engineers, GEOCongress 2006: Geotechnical Engineering in the Information Technology Age. Atlanta, GA: EUA, American Society of Civil Engineers

Secretaría de Desarrollo Urbano del Municipio de Aguascalientes (SEDUM), Departamento de Fraccionamientos, 1990–2007, Estudios Geotécnicos varios, Archivo

Sides EJ (1992) Reconciliation studies. In Annels AE (ed) Case histories and methods in mineral resource evaluation. Geological Society Special Publication, Londres, vol 63, pp 197–218

Supervisor Integral de Obra Civil, S. A. de C.V. (SIDOC) (2008) Estudios Geotécnicos varios, Archivo

Webster R, Oliver M (2001) Geostatistics for environmental scientists. Wiley, New York