Computation of Continuous Displacement Field from GPS Data - Comparative Study with Several Interpolation Methods

Belhadj ATTAOUIA, Kahlouche SALEM, Ghezali BOUALEM, and Gourine BACHIR, Algeria

Keywords: Auscultation with observations GPS, Displacement field, Interpolation, Cross-validation.

SUMMARY

it is impossible to collect observations in a comprehensive manner at any point of a site of study for practical reasons (cost, inaccessibility. Etc.). However, the continuity of the space is the basic hypothesis for subsequent analysis. The underlying problem is the interpolation. We seek through this document to defining the best method for predicting an continuous displacement field.. The methodology consists in using several all reliable interpolation methods . And through a cross-validation determine the most effective method for the data used. The application focuses on the auscultation network of tank «LNG" industrial complex "GL4Z" Arzew (Algeria). Constitute 56 points of GPS observations. The results of this comparative study interpolation of displacement, show that the best approach is the natural neighbor(RMSQs minimums)... Only the disadvantage is its irregularly representation based on delaunay triangulation. However, we retain interpolation radial basis function multilog method presents an good results with simple algoritm (comparable to kriging).

Computation of Continuous Displacement Field from GPS Data - Comparative Study with Several Interpolation Methods

Belhadj ATTAOUIA, Kahlouche SALEM, Ghezali BOUALEM, and Gourine BACHIR, Algeria

1. INTRODUCTION

The spatial interpolation is a classical problem estimating a function $\tilde{z}(S_p)$, where z = (x, y) at a site S_p plan from known values of S a number n, of surrounding points si:

$$\tilde{z}(S_p) = \sum_{i=1}^{n} w_i \; z(s_i) \;, \sum_{i=1}^{n} w_i \; = 1. \qquad (1)$$

The problem is to determine the weighting (Wi), each of the surrounding points. There are several ways to choose these weights.

- ***Inverse Distance Weighted (IDW)***

Its general idea is based on the assumption that the attribute value of an unsampled point is the weighted average of known values within the neighborhood [1].This involves the process of assigning values to unknown points by using values from a scattered set of known points. The value at the unknown point is a weighted sum of the values of N known points. Which is based on a concept of inverse distance weighting :

$$z(x, y) = \sum_{i=1}^{N} w_i(x, y) z_i \qquad (2)$$

(Where weights $w_i(x, y)$ and distance from each value to the unknown site $h_i\,(x, y)$ are given by : $w_i(x, y) = \frac{\mathrm{h}_i^{-2}(x,y)}{\sum_{j=1}^{n} \mathrm{h}_j^{-2}(x,y)}$, $h_i\,(x, y) = \sqrt{(x - x_i)^2 + (y - y_i)^2}$) .

- ***Kriging***

The unknown values are estimated from a neighborhood of points sampled. The weights are given after a study of space variability of the data to represent. The steps kriging pass by:

- The construction of semi-variogram showing the variations of the correlation between the data according to the distance (*d)* between those. The principle consists in gathering all the data per pairs and one distributes these couples in various classes according to the distance which separates them.In each class, one calculates a semi- variance [1].

$$\gamma(di) = \frac{1}{2Ni}\sum_{j=1}^{N}\sum_{k=1}^{N}\delta_{jk}^{i}\,(z_j - zk)^2 \qquad (3)$$

Where *(di)* indicates the center of class *i* , *Ni* the number of couples in this class [3] . And $\delta_{jk}^{i} = 1$ if the points *J* and *K* belong has class *i* else equal to zero in the contrary case.

- Adjustment of the analytical model to the experimental variogram [2] .Several types of models can be used such as spherical, exponential, Gaussian, etc. Within Surfer the Kriging defaults can be accepted to produce an accurate grid, or Kriging can be custom-fit to a data set, by specifying the appropriate variogram model.

- Estimate of the variance of the computed values in each point of the grid thus building the confidence intervals around the estimated values [9].

So the variogram analysis allows first to quantify the scale of data correlation, second to detect and quantify anisotropies in data variations, and lastly to quantify local effects, as well as inherent errors included within original data, and separate them from regional effects [5]. In summary, the kriging process is composed of two parts, analysis of this spatial variation and calculation of predicted values. Spatial variation is analyzed using variograms, which plot the variance of paired sample measurements as a function of distance between samples. An appropriate parametric model is then typically fitted to the empirical variogram and utilized to calculate distance weights for interpolation. Kriging selects weights so that the estimates are unbiased and the estimation variance is minimized [16].

- ***Modified Shepard's***

Uses an inverse distance weighted least squares method. Similar to the Inverse Distance to a Power interpolator, but the use of local least squares reduces the "bull's-eye" appearance of the generated contours. Used like an exact interpolator [4]. Shepard’s method was introduced in 1968 and modified to a local method by Franke and Nielson in 1980 [6].The interpolant is defined by.

$$z(x,y) = \sum_{i=1}^{N} W_k\,(x,y)Q_k(x,y) / \sum_{i=1}^{N} w_i(x,y) \qquad (4)$$

Where, the nodal function Q_k is a bivariate quadratic polynomial that interpolates the data value at node *k* and fits the data values on a set of nearby nodes in a weighted least-squares sense.

- ***Radial Basis Functions***

A radial basis function approximation takes the form

$$z(s) = \sum_{i\in I} y_i\,\varphi(\|x - i\|)\,,\ x \in \mathbb{R}^d \qquad (5)$$

Where $\varphi: (0, \infty) \to \mathbb{R}$, is a fixed univariate function and the coefficients $(y_i)i \in I$ are real numbers. We note that the norm take Euclidean form in the most common choice. Therefore the approximation s is a linear combination of translates of a fixed function $x \to \varphi(\|x\|)$ which is symmetric with respect to the given norm, in the sense that it clearly possesses the symmetries of the unit ball. We shall often say that the points $(x_j)_{j=1}^{n}$ are the centers of the radial basis function interpolant. Moreover, it is usual to refer to ' as the radial basis function, if the norm is understood [17].
All of the *Radial Basis Function* methods are exact interpolators, so they attempt to honor the data. The basis functions are analogous to variograms in Kriging. They define the optimal set of weights to apply to the data points when interpolating a grid node. They are several types of Radial Basis Function; we take example of the multilog equation [4].

$$B(h) = \log(h^2 + R^2) \qquad (6)$$

Where, h is the anisotropically rescaled, relative distance from the point to the node. R^2 an default value (length of diagonal of the data extent)^2 / 25 * number of data points).
As many methods are used in the chosen method to interpolate spatial displacement data for this article. Cross-validation is essential to validate critical parameters that could affect the interpolation accuracy of used data. This insures the overall utility of this models and enables optimal data prediction that is comparable to the observed data.

- ***Triangulation with Linear Interpolation***

This method in *Surfer* uses the optimal Delaunay triangulation. The algorithm creates triangles by drawing lines between data points. The original points are connected in such a way that no triangle edges are intersected by other triangles. The result is a patchwork of triangular faces over the extent of the grid. This method is an exact interpolator. Which supposes that the point *S* to be estimated is inside the triangle formed.
The estimate of the variable value at the point S is written [10]:

$$\tilde{z}(S) = \frac{|S_1 S S_2| Z(S_3) + |S_1 S S_3| Z(S_2) + |S_2 S S_3| Z(S_1)}{|S_1 S_2 S_3|} \qquad (7)$$

- ***Natural neighbor***

Based on the polygonation of Thiessen of the whole of irregular points (*figure 2*). To improve by Sibson [10].which proposes an estimate of *S0* using a linear combination of the values of the close sites (*Z(S0)*) balanced by surfaces *P (S0, I)*.

$$\tilde{z}(S_0) = \sum_{i=1}^{n} \frac{P(S_0, i)}{P(S_0)} Z(S_0) \qquad (8)$$

With, N number of observation. P (S0) the surface of Thiessen polygon in S0. P (S0, i) surface of the polygon intersection of the observation I and that of S0.

2. METHODOLOGY

the generation of continuous surfaces starting from irregularly distributed data is a task for many disciplines [8]. There is a variety of methods which can perform this task, but the difficulty lies in the choice of the interpolation that best reproduces the given surface.
First, it is important to note that the deterministic or probabilistic interpolation is the result of data processing. Second, there are a variety of interpolation methods. These methods use a different calculation and do not produce the same predictions (although sometimes the results between methods are not very different). The choice of the optimal method based on the one hand on statistical criteria of quality and the other on visual representation of the continuous information is needed. We seek through this document to defining the best method for predicting a continuous displacement field.We summarizes the steps uses:

- using several interpolation methods all reliable,
- And through a cross-validation determine the most effective method for the data used.
- Comparative choice of methods performance with criteria quality.

2.1 Interpolation methods used

To evaluate physical data (unknown displacement field) in a continuous domain. In this paper, a comparison of several interpolation methods was proposed.In fact, six different methods, namely : *Inverse Distance Weighted* (IDW), *kriging* (KRG), *Natural Neighbor* (NATN), *Modified Shepard's* (MS), *Radial Basis Function methods (*RBF methods with: *Multiquadratic, Inverse Multiquadratic, Multilog, Natural Cubic Spline*), and the *Triangulation with Linear Interpolation* (TL). the program used is "Scripter of Surfer software ", where all this methods can be handled simultaneously.

2.2 Leave-One-Out cross-validation

Cross-Validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the other used to validate the model. In typical cross-validation, the training and validation sets must cross-over in successive rounds such that each data point has a chance of being validated against. The basic form of cross-validation is k-fold cross-validation. Other forms of cross-validation are special cases of k-fold cross-validation or involve repeated rounds of k-fold cross-validation [11]. Leave-One-Out (LOOCV) is a *k*-fold cross-validation where the size of the validation set[12] is one. LOO involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data [1] - [7]. Leave-one-out cross-validation is a popular method for estimating prediction error for small samples [13]. So performed this quality control analysis through cross-validation, i.e., estimating a value for a measured point based on all other data except that point, and then comparing the predicted value to the measured value [16].

2.3 Comparative choice of methods

Performance measures of the method are then based on statistical summaries of errors:

- ***The mean error (ME):*** is the measurement error that indicates whether the forecasts are biased. This is a measure of reliability [7]-[15]. Representing the average difference between predictions $\tilde{z}(S_i)$ and the observations $z(S_i)$.

$$\text{ME} = \frac{1}{n}\sum_{i=1}^{n}(\tilde{z}(S_i) - z(S_i)) \qquad (9)$$

- ***The root mean squared error (RMSE):*** a robust measure of accuracy [1]-[14], is adopted to assess the interpolation models performances. RMSE is a measure frequently used on the differences between values predicted by a model or an estimator and the values actually observed.

$$\text{RMSE} = \left(\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{z}(S_i) - z(S_i)\right)^2\right)^{0.5} \qquad (10)$$

- ***The Pearson coefficient (r)*:** at the same time, Pearson coefficient (r) is also used for evaluate whether the estimated data fits observed data. Higher the value is close to ± 1, more the linear relationship is strong, and the value of r is close to 0, more the linear relationship is weak [1] - [7].

3. CASE OF APPLICATION

- **Presentation of Study Area**

The application focused on the determination of continues displacement field in an auscultation network of the reservoir GL4Z - complex LNG - at Arzew in Algeria. Two campaigns of GPS observations (2000, 2006), were performed. Auscultation network consists of 13 support points away from a hundred meters from the reservoir and 43 secondary benchmarks distributed in the field surrounding the reservoir (Fig. 1). A total of 56 points or coordinates used in the processing are expressed in the local survey reported to the World Geodetic System WGS84.This data are organized by station. This file contains: the local geodetic coordinates of a GPS station (horizontal component (E), north horizontal component (N); Vertical component or elevation (U).

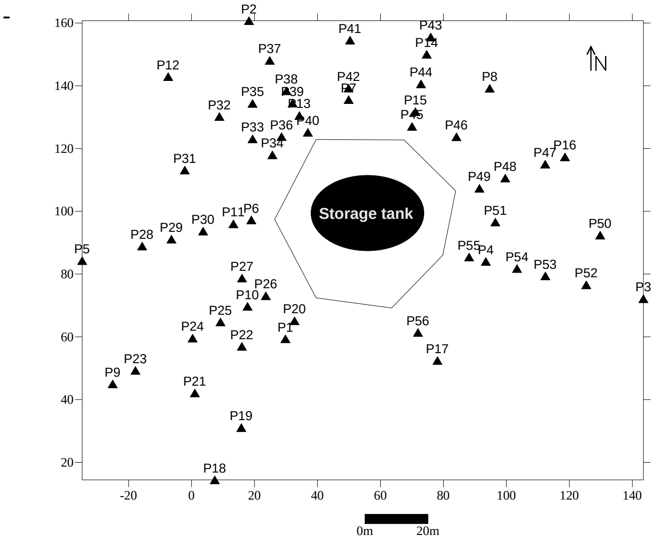


Fig. 1 Auscultation network epoch (2000-2006).

The obtained results are summaries respectively in next tables. TABLE I (shows the statistics summary of components in one epoch, e.g.2000). TABLE II (the statistics of observed data with estimated ones of LOOCV: main statistics of the Mean, Standards Deviations of displacements in the horizontal component, north horizontal and Vertical component nominated respectively DE, DN, DU).

TABLE I. STATISTICS OF DATA

	E(m)	N(m)	U(m)
observations	56	56	56
Minimum	-34,734	14,400	-0,083
Maximum	143,514	160,740	0,246
Mean	45,236	100,220	0,121
STD	44,446	35,907	0,062

TABLE II.OBSERVED AND ESTIMATED VALUES OF DISPLACEMENTS

	DE		**DN**		**DU**	
	Mean	STD	Mean	STD	Mean	STD
Observed	0,0139	0,0373	0,0253	0,0357	0,1212	0,0618
Estimated_IDW	0,0139	0,0176	0,0219	0,0150	0,1200	0,0224
Estimated_NATN	0,0136	0,0223	0,0217	0,0235	0,1183	0,0357
Estimated_TL	0,0150	0,0223	0,0206	0,0243	0,1171	0,0380
Estimated_RBF-Mutilog	0,0139	0,0293	0,0239	0,0262	0,1230	0,0431
Estimated_ KRG	0,0138	0,0294	0,0234	0,0246	0,1238	0,0434
Estimated_MS	0,0147	0,0313	0,0204	0,0236	0,1321	0,0565

Table III to V above present the errors statistics of LOOCV computed on the 56 points in three directions. With the results of different values of the Mean (the bias or mean error), root-mean-squared of error (RMSE) and Pearson's coefficient of correlation between the estimated and true values (PEARSON).

TABLE III.CROSS-VALIDATION STATISTICS (DISPLACEMENTS IN E)

ERROR (DX)	MEAN	PEARSON	RMSE
NATN	0,002	**0,250**	**0,027**
TL	0,004	0,065	0,028
RBF-MLOG	0,000	**0,015**	**0,029**
KRG	0,000	0,027	0,029
IDW	0,000	0,029	0,030
MS	0,001	0,085	0,031

TABLE IV.CROSS-VALIDATION STATISTICS (DISPLACEMENTS IN N)

ERROR (DY)	MEAN	PEARSON	RMSE
NATN	0,004	**0,295**	**0,023**
TL	0,003	0,238	0,024
KRG	-0,002	0,242	0,026
RBF-MLOG	-0,001	**0,101**	**0,027**
MS	-0,005	0,110	0,027
IDW	-0,003	0,240	0,031

TABLE V.CROSS-VALIDATION STATISTICS (DISPLACEMENTS IN U)

ERROR (DZ)	MEAN	PEARSON	RMSE
NATN	-0,009	**0,166**	**0,055**
TL	-0,010	0,139	0,056
IDW	-0,001	0,164	0,059
KRG	0,003	0,051	0,061
RBF-MLOG	0,002	**0,050**	**0,062**
MS	0,011	0,108	0,077

The classification of root mean square error results (Fig .2) displays the natural neighbor in first position. However the illustration is based on Delaunay triangulation (strictly convex geometries). for a regular meshing, the radial basis fonction-Multilog with advantage of simple algorithm . by against and te kriging gives a continues representation of displacement field without restrictions geometric.

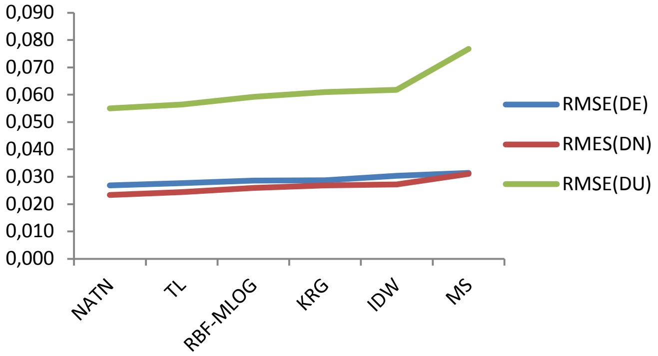
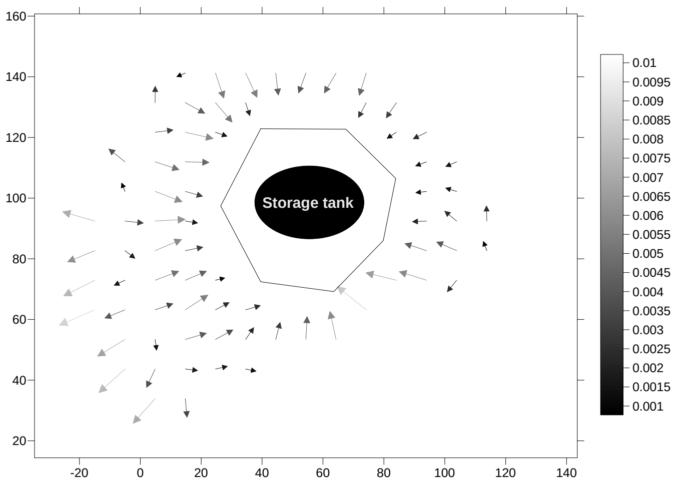


Fig .2 auscultation network.

Fig.3 shows the three bests methods for displacements fields interpolated .So Natural neighbor.,Radial basis fonction-Multilog and kriging .the two latest displayed a regular meshing. by against the first one (Natural neighbor interpolation) is irrégular.



(a)

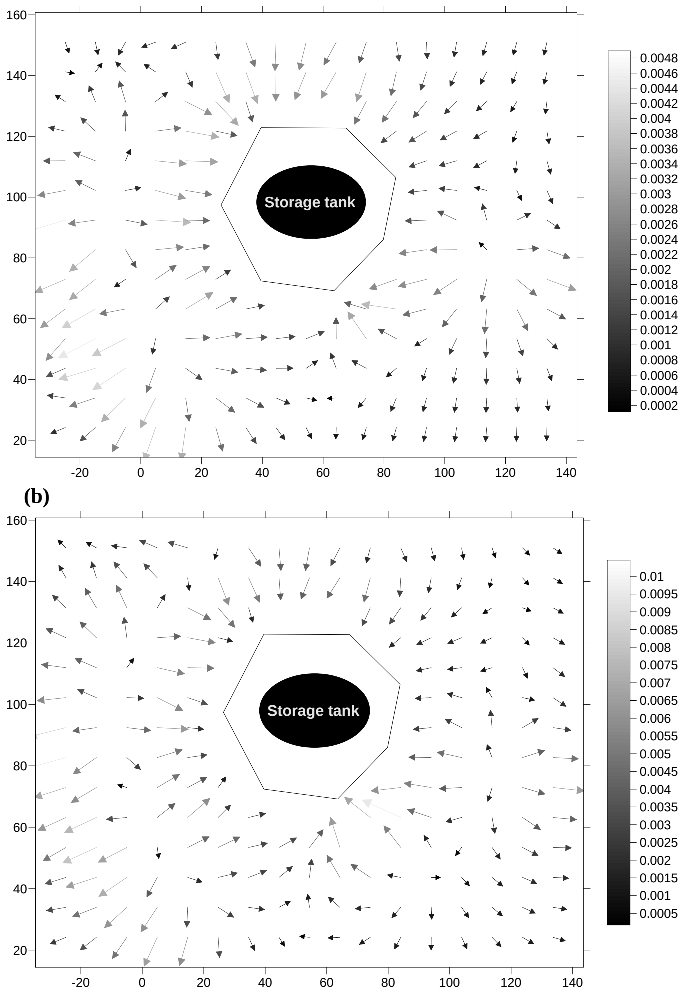


Fig. 3 Displacement field of elevation interpolated.
(a) Natural Neighbor. (b) radial basis fonction-Multilog.(c) kriging.

CONCLUSION

The spatial interpolation is used in many areas such as geodesy, meteorology, geology, etc. There is a multitude of methods with different characteristics. This operation is a priori trivial and transparent to the user a lot of. the program "Scripter of Surfer software" contain more accurate methods can be handled simultaneously. The aims of this comparative study is to Interpolate the unknown displacement field with different exact methods all reliable, namely : *Inverse Distance Weighted* (IDW), *kriging* (KRG), *Natural Neighbor* (NATN), *Modified Shepard's* (MS), *Radial Basis Function methods (*RBF methods with: *Multiquadratic, Inverse Multiquadratic, Multilog, Natural Cubic Spline*), and the *Triangulation with Linear Interpolation* (TL). However, to assess the robustness of these methods, we conduct a Leave-One-Out cross-validation, consist to cast aside one observeddata   and to compare its value with the estimate produced by applying interpolation to the remaining data. By repeating the process for each of the measures considered separately, we can calculate the root mean square error between the data and estimation. The results confirm that all methods are reliable.The results of minimal root mean square error display the natural neighbor in first position. However this illustration is based on an strictly convex geometrie .For a regular meshing, the radial basis fonction-multilog or the kriging, really give a representation of continues displacement field without restrictions geometric . In addition the radial basis fonction method offers a simple algorithm compared to kriging.

REFERENCES

[1] Feng-Wen Chen • Chen-Wuing Liu.Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. Springer-Verlag 2012.

[2] Yves Gratton.Le krigeage : la méthode optimale d'interpolation spatiale. Institut National de la Recherche Scientifique Inrs-Eau-Terre-Environnement. Juin 2002.

[3] Dr. Ir. Eric J.M. Delhez. Méthodes mathématiques d'analyse et de modélisation appliquées à l'environnement. Septembre 2008.

[4] . Surfer help. "Introduction to Gridding Methods.

[5] C. Gumiaux, D. Gapais, J.P. Brun.Geostatistics applied to best-fit interpolation of orientation data. Géosciences Rennes, UMR 6118 CNRS, Université de Rennes 1, 35042 Rennes Cedex, France. Tectonophysics 376 (2003) 241– 259.Elsevier. 2003.

[6] Robert j. Renka.Algorithm 792: Accuracy Tests of ACM Algorithms for Interpolation of Scattered. Data in the Plane. University of North Texas And Ron Brown. AISN Software. ACM Transactions on Mathematical Software, Vol. 25, No. 1, March 1999.

[7] G. Dubois. Automatic mapping algorithms for routine and emergency monitoring data. Spatial interpolation comparison 2004. Radioactivity Environmental Monitoring Emissions and Health Unit.2005.

[8] C. Caruso, F. Quarta 1998. Interpolation Methods Comparison. Computers Math. Applic. Vol. 35, No. 12, pp. 109-126, 1998. Pergamon. PII: SOS98-1221(98)00101-1**.**Elsevier Science Ltd.

[9] [8]Pierre Bosser.Interpolation spatiale. Ecole Nationale des Sciences Géographiques Département Positionnement Terrestre et Spatil. 2011-2012.

[10] Wilfried Despagne.Rapport de stage« Méthodes géostatistique pour l'interpolation et la modélisation en 2d/3d des données spatiales. » Université de Bretagne Sud.Institut Universitaire Professionnalisé .Informatique et Statistique.2006.

[11] Amaury Lendasse, Vincent Wertz, and Michel Verleysen.Model Selection with Cross-Validations and Bootstraps– Application to Time Series Prediction with RBFN Models. Université catholique de Louvain.

[12] Payam Refaeilzadeh, Lei Tang, Huan Liu. Cross-Validation Arizona State University..2008
[13] . Wenyu Jiang, Sudhir Varma and Richard Simon. Calculating Confidence Intervals for Prediction Error in Microarray Classification Using Resampling.Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National.Cancer Institute, Rockville, USA
[14] Duke. Quelle est la ligne de fond? Comment comparer les modèles[en ligne]. http://www.duke.edu/~rnau/compare.htm. (Consulted in 2013).
[15] .Eumetcal. Forecast accuracy and scores. Ordered categorical forecasts of variables. [on line].http://www.eumetcal.org/euromet/french/nwp (Accessed 2012).
[16] . Developing Spatially Interpolated Surfaces and Estimating Uncertainty. U.S. Environmental Protection Agency. November 2004 .Office of Air and Radiation. Office of Air Quality Planning and Standards. Research Triangle Park, NC 27711.
[17] . Bradley John Charles Baxter. The interpolation theory of radial basis functions. 1992.

CONTACTS

Ms Belhadj Attaouia
Centre of Space Techniques
Division of Space Geodesy
BP 13, 31200
Arzew
Algeria
Tel. +213 41472217
Fax: +213 41473665
Email:battaouia@gmail.com