Kriging Method Combined with GISM to Generate Ionospheric Scintillation Map

Pierrick Hamel[1], Diagna Charles Sambou[2], Muriel Darces[2], Yannick Beniguel[1], Marc Hélier[2]
[1] IEEA, Courbevoie, France
[2] Sorbonne Universités, UPMC Univ Paris 06, UR2, L2E, F-75005 Paris, France

***Abstract*—A kriging method has been used to interpolate the value of the scintillation indices over a region, using a limited number of samples. This method has been combined with the Global Ionospheric Scintillation Model (GISM) to estimate the scintillation level over a region, using data collected by GNSS receiver.**

***Index Terms*—kriging method, ionospheric scintillations, propagation, Global Navigation Satellite System (GNSS), space weather.**

I. INTRODUCTION

The variation of electron density in ionosphere may affect the stability of RF signal propagating through it. This phenomenon, called scintillation, is especially strong after sunset, during the equinoxes, at low latitude. It can be of interest for VHF or L band frequency (used for GNSS). Computer codes, such as GISM [1] or the Wideband ionospheric scintillation model (WBMOD) [2], have been developed to simulate such a phenomenon, based on measurement campaigns and on its physical modeling.

Like for meteorological prediction, these codes may provide trends over days and seasons but they cannot exactly predict the level of the scintillation, at one location and one given time. For other applications, the kriging method has been developed in [3] and applied in [4], [5], [6] and [7]; it aims to estimate the value of a regionalized random variable around measurement points.

The kriging algorithm is first described. Then an example is presented for a pseudo-random observable depending on a one-dimensional variable. Finally, the kriging method combined with GISM is applied to generate one scintillation map. The quality of the kriging algorithm is evaluated using the cross validation technique. It consists in removing some of the data to check that the estimated results fit in the confidence level range.

II. PRESENTATION OF THE KRIGING ALGORITHM

This algorithm is well described in [4], all the notations are introduced in this section. The parameterization of the model is first described and then the interpolation of the data is detailed. The kriging technique was selected for the following reasons:

- It is an exact interpolator;
- It is well suited to interpolate multivariate random data;
- It provides an estimation of the interpolation error.

A. Parametrization of the model

The purpose of kriging is to interpolate a random observable $y(\boldsymbol{x})$ depending on a variable $\boldsymbol{x}$. This variable can be a scalar for a one-dimensional variation or a vector, for a multi-dimension variation.

The observable fluctuates around a linear combination of M models $f_i(\boldsymbol{x})$, assumed to be well known. The fluctuation $\omega(\boldsymbol{x})$ is random and is also dependent on the variable $\boldsymbol{x}$. This can be summarized by the following equation:

$$y(\boldsymbol{x}) = \boldsymbol{r^T}(\boldsymbol{x}) * \boldsymbol{p} + \omega(\boldsymbol{x}) \tag{1}$$

with the so called regression vector $\boldsymbol{r^T}(\boldsymbol{x})$ and the weight coefficients $\boldsymbol{p}$ of the models (which have to be estimated from the measurement):

$$\boldsymbol{r^T}(\boldsymbol{x}) = (f_1(\boldsymbol{x}) \quad \cdots \quad f_M(\boldsymbol{x}))\,; \tag{2}$$

$$\boldsymbol{p^T}(\boldsymbol{x}) = (p_1 \quad \cdots \quad p_M) \tag{3}$$

The measurements y_i^s, gathered in the vector Y_s, are recorded at n samples $\boldsymbol{x_i^s}$, contained in the vector X_s:

$$Y_s = \begin{pmatrix} y_1^s \\ \vdots \\ y_n^s \end{pmatrix}\,;\; X_s = \begin{pmatrix} \boldsymbol{x_1^s} \\ \vdots \\ \boldsymbol{x_n^s} \end{pmatrix} \tag{4}$$

The regression matrix, associated to the samples, is calculated. Its coefficients correspond to the value of the models at the sample locations:

$$R_s = \begin{pmatrix} f_1(\boldsymbol{x_1^s}) & \cdots & f_M(\boldsymbol{x_1^s}) \\ \vdots & \ddots & \vdots \\ f_1(\boldsymbol{x_n^s}) & \cdots & f_M(\boldsymbol{x_n^s}) \end{pmatrix} \tag{5}$$

Then the sample locations $\boldsymbol{x_i^s}$ are replaced by their normalized counterparts $\hat{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{s}}$. A matrix, collecting the distance of the normalized sample value is defined:

$$H_s = \begin{pmatrix} \|\hat{x}_1^s - \hat{x}_1^s\| & \cdots & \|\hat{x}_1^s - \hat{x}_n^s\| \\ \vdots & \ddots & \vdots \\ \|\hat{x}_1^s - \hat{x}_n^s\| & \cdots & \|\hat{x}_n^s - \hat{x}_n^s\| \end{pmatrix} \tag{6}$$

The random fluctuation $\omega(\boldsymbol{x})$ is a Gaussian process with zero mean. Its covariance matrix C_s is associated to the samples. The coefficients of indices (i,j) are expressed as:

$$\begin{aligned} C_s(i,j) &= \sigma_\omega^2 * C_{s\omega}(i,j) \\ &= cov\big(\omega(\boldsymbol{x_i^s}), \omega(\boldsymbol{x_j^s})\big) \end{aligned} \quad (7)$$

This matrix has to be estimated. If there are sufficient reliable data, it is possible to calculate what is called a variogram and make the covariance matrix fit with a polynomial function as explained in [4] or any other function, such as in [5]. A variogram is a function quantifying the degree of spatial dependence of a random quantity. In this paper, the variogram has not been estimated and the covariance matrix is defined as in [6]:

$$C_{s\omega}(i,j) = \exp\left(-\frac{H_s(i,j)}{a}\right) \quad (8)$$

where the range a has to be optimized, using the estimated value of the weight coefficients p_{est} and σ_ω^2, given by [4] :

$$\begin{aligned} \boldsymbol{p_{est}} &= (R_s^t C_{s\omega}^{-1} R_s)^{-1} * R_s^t C_{s\omega}^{-1} Y_s \\ \sigma_\omega^2 &= \frac{1}{n}\left[\left(Y_s - R_s \boldsymbol{p_{est}}\right) * C_{s\omega}^{-1} * \left(Y_s - R_s \boldsymbol{p_{est}}\right)\right] \end{aligned} \quad (9)$$

To optimize the range a, as suggested in [4] and [7], the determinant of the covariance matrix should be minimized.

B. Interpolation of the data

Once the range a is optimized, it is possible to interpolate the observable $y(\boldsymbol{x})$ between the samples at N locations $\boldsymbol{x_i^e}$ recorded in the vector X_e. The estimated values y_i^e constitute the vector Y_e:

$$Y_e = \begin{pmatrix} y_1^e \\ \vdots \\ y_N^e \end{pmatrix} ; \; X_e = \begin{pmatrix} \boldsymbol{x_1^e} \\ \vdots \\ \boldsymbol{x_N^e} \end{pmatrix} \quad (10)$$

The regression matrix, related to the chosen locations, is calculated:

$$R_e = \begin{pmatrix} f_1(\boldsymbol{x_1^e}) & \cdots & f_M(\boldsymbol{x_1^e}) \\ \vdots & \ddots & \vdots \\ f_1(\boldsymbol{x_N^e}) & \cdots & f_M(\boldsymbol{x_N^e}) \end{pmatrix} \quad (11)$$

The normalized distance between the samples and the points to interpolate are set in the following distance matrix:

$$H_e = \begin{pmatrix} \|\hat{x}_1^e - \hat{x}_1^s\| & \cdots & \|\hat{x}_1^e - \hat{x}_n^s\| \\ \vdots & \ddots & \vdots \\ \|\hat{x}_N^e - \hat{x}_1^s\| & \cdots & \|\hat{x}_N^e - \hat{x}_n^s\| \end{pmatrix} \quad (12)$$

The estimated covariance matrix is computed similarly to the covariance matrix of the samples ((7)-(8)) using the estimated range a and σ_ω^2 optimized in (9). Its coefficients of indices (i,j) are:

$$C_e(i,j) = \sigma_\omega^2 * \exp\left(-\frac{H_e(i,j)}{a}\right) \quad (13)$$

Finally, the estimated values of the observable are given by:

$$Y_e = [A + C_e * C_s^{-1} * (I_d - A)]Y_s \quad (14)$$

with A defined as follows:

$$A = R_e * (R_s^t C_{s\omega}^{-1} R_s)^{-1} R_s^t C_{s\omega}^{-1} \quad (15)$$

The interpolation error $\epsilon_e(\boldsymbol{x_i^e})$ can also be estimated, as suggested in [4].

III. Application to a one-dimensional variable

A. Example

The method presented in section II is first applied to interpolate a pseudo-random observable, depending on a single variable (in this section, the dimension of $\boldsymbol{x}$ is 1). In the chosen example, this observable mimics one realization of a random variable, and is defined analytically by the formula:

$$y(x) = x^3 + 4\cos(3x) + \exp(-x * \sin(6x)) \quad (16)$$

This function is estimated in the interval [-2;2]. Eleven equally distributed samples are taken. A random distribution of the samples is also possible. In this case, the accuracy of the estimation might be worst but the conclusions would be the same.

Three approaches are presented. The first one considers that there is no a priori known model and that the observable fluctuates around a constant. The second approach assumes that the observable oscillates around a curve which can be approximated by a polynomial function. In the last approach, the model around which the observable fluctuates is very well known and the linear combination of the model will be very close to the observable.

These approaches and the considered models are summarized in the following table:

TABLE I. Approaches and associated models

Approach	Number of models M	Model definition
A	1	$f_1(x) = 1$
B	4	$f_i(x) = x^i \,; 0 \le i \le 3$
C	2	$f_1(x) = x^3 \; ; \; f_2(x) = 4cos(3x)$

B. *Comparison of the results obtained by different approaches*

The results of the kriging method are presented in the following figures:

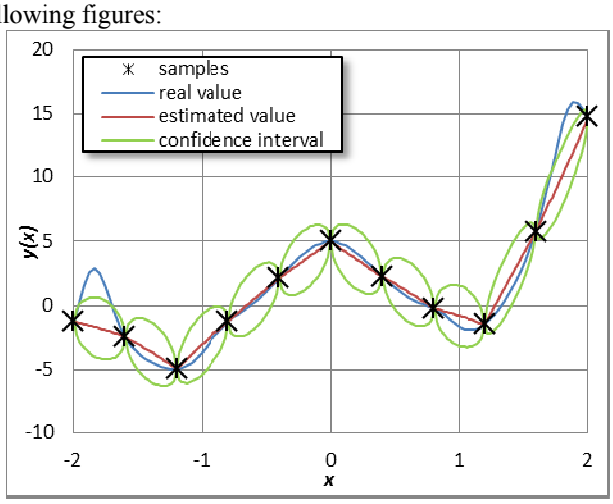

Fig. 1. Approach A : no a priori on the model

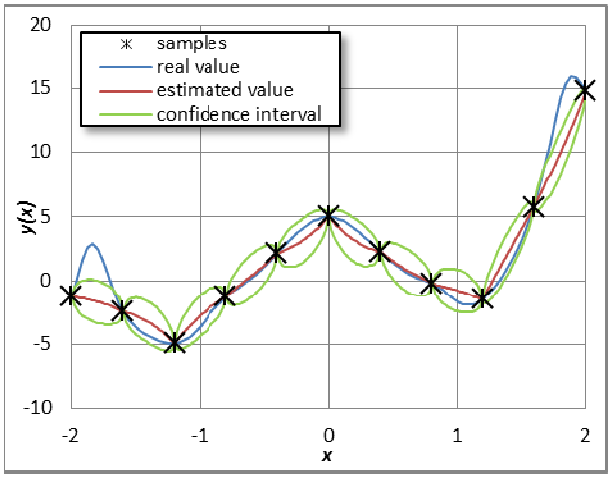

Fig. 2. Approach B : interpolation of the model with a polynomial function

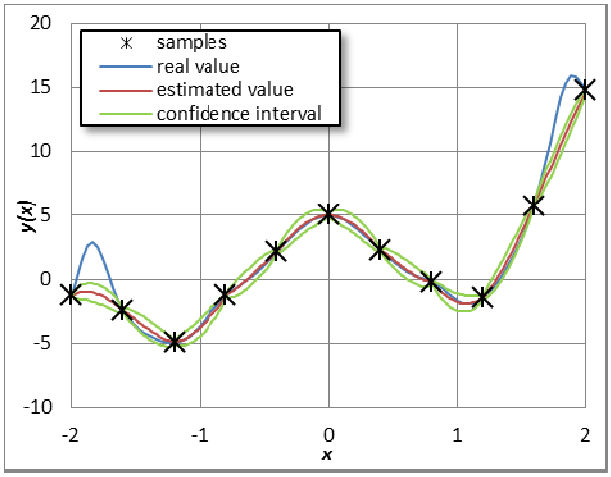

Fig. 3. Approach C : the model is assumed to be very well known

The results of approach A and approach B are very close. The confidence interval, related to the σ_ω^2 estimation, is a bit smaller for approach B than for approach A. This means that the estimated error is lower for approach B than for approach A. Consequently, for this example, a polynomial model would fit better to the observable, than a simple constant model.

The approach C provides better results than approach B, with a smaller estimated error and an estimated observable value which fits better to the observable. A polynomial function of degree 10, is required to obtain the same accuracy as the one of approach C. That can be explained by the fact that the cosine function, in this small interval, is well approximated by its Taylor series.

As a conclusion, the better the models fit to the observable, the best will be the estimation done by the kriging method.

IV. SCINTILLATION MAP GENERATION

Data have been recorded in Lima (Peru), during year 2012, using a Novatel® 50 Hz receiver. On the first of October 2012, strong scintillations were detected. The scintillation index defined as the RMS value of the normalized intensity and referenced as S4 was recorded during one hour. The location of the scintillation is assumed to be the pierce point (IPP) of the line of sight from the GNSS satellite to the ground station, with a planar ionosphere (at altitude of about 350 km). The sample points are depicted on the map shown in Fig. 4:

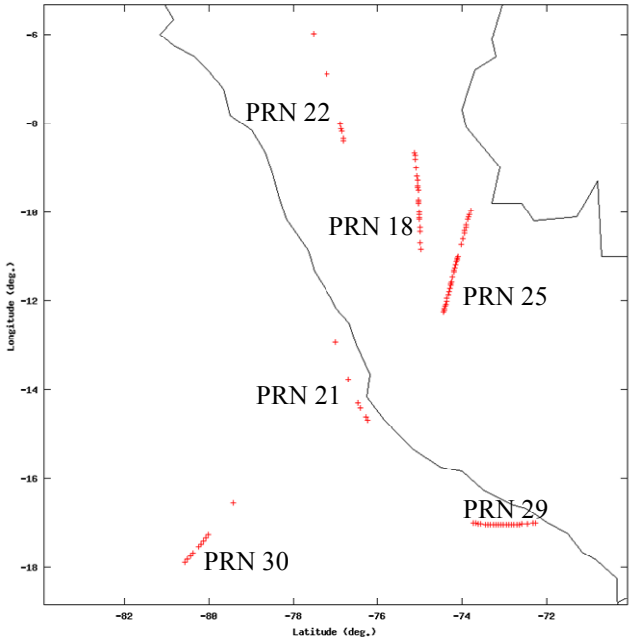

Fig. 4. Pierce points of the scintillation events.

The distribution of the samples is not regular and follows the ground track of 6 different PRNs (GPS satellites). To apply a kriging method, it has been shown in the previous section that the choice of the deterministic model around which the observable value fluctuates (in this case, the S4 level) is critical. For this reason, the chosen model is GISM's one presented in [1].

The regression matrices are the following:

$$R_e = \begin{pmatrix} f_1(\boldsymbol{x_1^e}) \\ \vdots \\ f_1(\boldsymbol{x_N^e}) \end{pmatrix} \qquad R_s = \begin{pmatrix} f_1(\boldsymbol{x_1^s}) \\ \vdots \\ f_1(\boldsymbol{x_n^s}) \end{pmatrix} \tag{17}$$

where f_1 is the estimation given by GISM at the corresponding locations $\boldsymbol{x_i^e}$ and $\boldsymbol{x_i^s}$ (each one depending on latitude and longitude).

The S4 level has been simulated, on a grid, which dimensions are [-85°;-69°] for latitude and [-20°;-4°] for longitude. The step in latitude and longitude is 1°. The S4 level of the model is linearly interpolated between 3 values on the grid generated by GISM. The final estimated values of the S4 level are presented in Fig. 5, the estimated relative accuracy is defined as:

$$\frac{|y_i^e(\boldsymbol{x_i^e}) - \epsilon_e(\boldsymbol{x_i^e})|}{y_i^e(\boldsymbol{x_i^e})} \tag{18}$$

is shown in Fig. 6. The same analysis could have also been performed on the phase scintillation index (Sigma Phi).

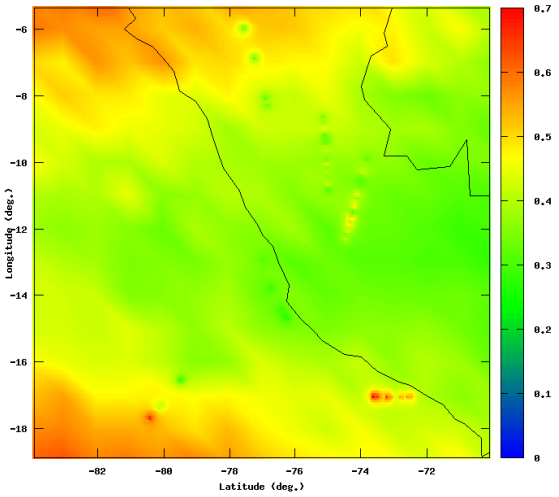

Fig. 5. S4 level estimated by a kriging method, using GISM as a reference model

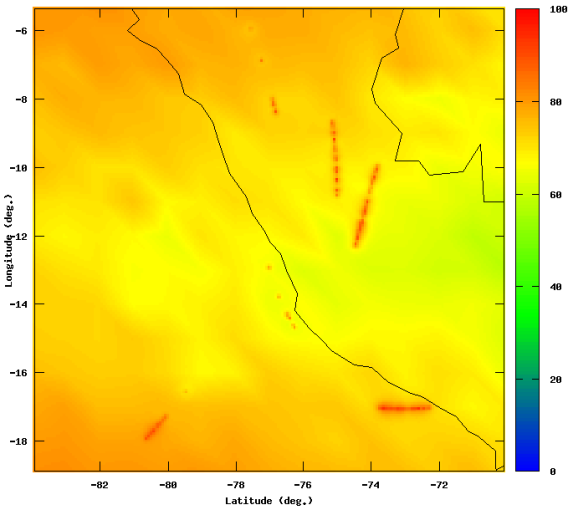

Fig. 6. Estimated relative accuracy (in percentage) of the result.

The estimated accuracy is quite good (about 75%) even in the area where no data is recorded. The absolute estimated error is almost constant on the map and is about 0.13. As a result, the relative accuracy is better when the estimated S4 level is high. Hopefully, this corresponds to the locations where the highest accuracy is required.

There are two ways to improve the accuracy of the interpolated data:

- Firstly, it would be of interest to include additional stations, in order to increase the number of samples.
- Secondly, the determination of the variogram, using the recorded data, might be also useful without increasing the number of samples.

V. VALIDATION

To validate the method, 25% of the sample points are removed from the algorithm. They correspond to the PRN 18 IPPs. The kriging method has then been applied to estimate their values. The absolute estimated error and the estimated accuracy on the S4 coefficient remain almost the same (respectively 0.12 and 75% on the track of the satellite PRN 18). The comparison between the estimated and measured values is presented in Fig. 7:

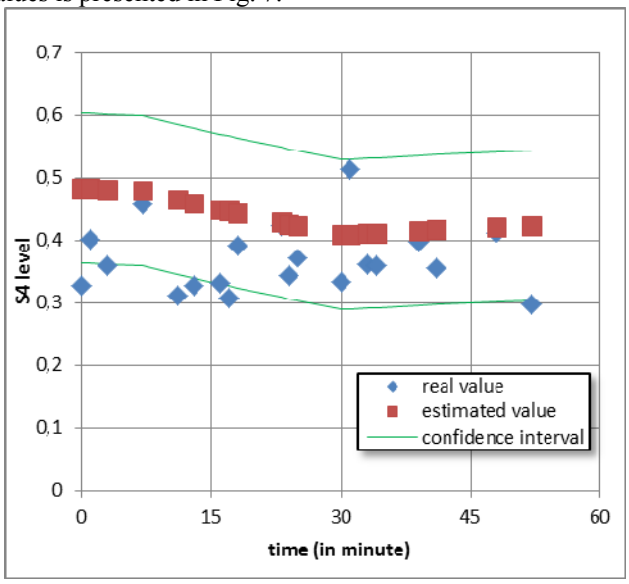

Fig. 7. Estimated and real S4 level on the PRN 18 track

75% of the measured samples are within the confidence interval, which is coherent with what was expected.

VI. CONCLUSION

The kriging method was presented. An example has been detailed for a one-dimensional problem and the choice of the model appears as a critical point. The method has then been applied to create a dense map of S4 using a limited set of theoretical data computed with GISM. It provides interpolated S4 values with a quite good accuracy. The calculations

performed were done with a poor data set. This will be improved in the future, including additional data, in particular from the IGS network. This would bring another way of improvement as an accurate determination of the variogram based on the data would be made possible as we will have enough points to do it.

The calculation technique provides maps and concurrently error maps. As a validation, the algorithm has been checked by removing some data points and estimating their values from the remaining data. The estimated values at the missing points were found to be within the confidence level range

ACKNOWLEDGMENT

The data used to generate the map, have been collected during the ESA Monitor project [8].

REFERENCES

[1] Béniguel Y., P. Hamel, “A Global Ionosphere Scintillation Propagation Model for Equatorial Regions”, Journal of Space Weather Space Climate, 1, 2011.

[2] Secan, J. A., R. M. Bussey, E. J. Fremouw, and Sa. Basu, “An improved model of equatorial scintillation”, Radio Sci., 30, 607-617, 1995.

[3] Matheron, G. (1963). Principles of geostatistics. Economic Geology, 58,1246–1266.

[4] Lefebvre, J. Roussel, H. ; Walter, E. ; Lecointe, D. ; Tabbara, W., "Prediction from wrong models: the Kriging approach" Antennas and Propagation Magazine, IEEE, vol. 38, no 4, p.35-45, 1996.

[5] P. De Doncker, J.-M. Dricot, R. Meys, M. Hélier, and W. Tabbara, "Electromagnetic Fields Estimation Using Spatial Statistics," Electromagnetics, vol. 26, pp. 111-122, 2006.

[6] Sayin, I. ; Arikan, F. ; Arikan, O. "Synthetic TEC Mapping with Ordinary and Universal Kriging", RAST 2007, Istanbul, 14-16 June, 2007.

[7] Wei Zhuo, Prabhat ; Paciorek, C. ; Kaufman, C. ; Bethel, W. "Parallel Kriging Analysis for Large Spatial Datasets" ICDMW '11 Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops Pages 38-44.

[8] R. Prieto Cerdeira, Y. Béniguel, "The MONITOR project: architecture, data and products", Ionospheric Effects Symposium, Alexandria VA, May 2011.