AGU PUBLICATIONS

Radio Science

RESEARCH ARTICLE
10.1002/2014RS005470

Special Section:
Beacon Satellite Symposium 2013

Key Points:
- Kriging algorithm to interpolate data
- Scintillation map generation

Correspondence to:
P. Hamel,
pierrick.hamel@ieea.fr

Citation:
Hamel, P., D. C. Sambou, M. Darces, Y. Beniguel, and M. Hélier (2014), Kriging method to perform scintillation maps based on measurement and GISM model, *Radio Sci.*, *49*, 746–752, doi:10.1002/2014RS005470.

Received 4 APR 2014
Accepted 25 AUG 2014
Accepted article online 27 AUG 2014
Published online 12 SEP 2014

Kriging method to perform scintillation maps based on measurement and GISM model

Pierrick Hamel[1], Diagna Charles Sambou[2], Muriel Darces[2], Yannick Beniguel[1], and Marc Hélier[2]

[1]IEEA, Courbevoie, France, [2]Sorbonne Universités, UPMC Univ Paris 06, Paris, France

Abstract This paper presents an ionosphere scintillation mapping algorithm. The technique used is a Kriging technique, well known in geostatistics. It is an interpolation method applied to a set of random data, namely, in this application, to the intensity scintillation (S4) index. The paper focuses on the importance of the regression model used in the algorithm. The one-dimensional and then the two-dimensional problems have been investigated. One example of results combining measurement data recorded by two different networks in Peru is presented. The accuracy of the results has been verified using a cross-validation technique, which consists in removing part of the known data set and comparing the interpolated results to it.

1. Introduction

This paper addresses the problem of ionosphere scintillation mapping for nowcasting purposes. This problem is of strong interest for many applications such as the satellite navigation, the telecommunications, and the space weather climatology. Ionosphere scintillation characterizes the fluctuations on a signal after its propagation through the random bubbles which might develop inside the ionosphere. These bubbles may reach dimensions up to several hundreds of kilometers. They are aligned with the terrestrial magnetic field and appear after sunset, mostly at equatorial regions. Their occurrence and characteristic properties have been addressed in many papers [*Aarons*, 1982; *Basu and Basu*, 1985; *Rodrigues et al.*, 2004; *Cervera and Thomas*, 2006; *Béniguel et al.*, 2009]. The scintillation on transmitted signals may exhibit fades fluctuations up to 30 dB peak to peak in the worst cases, well above the tolerance noise level of the above mentioned systems.

The scintillation mapping aims to give a planar view at the ionosphere peak point altitude (usually set to 350 km) of the bubbles extent. At low latitudes this corresponds to a cross section of these bubbles. A climatological scintillation model such as the Global Ionospheric Scintillation Model (GISM) [*Béniguel and Hamel*, 2011] fails to accurately produce ionosphere scintillation maps as the medium is described in terms of statistical parameters which may depart from current observations. Such a model might consequently overestimate or underestimate the scintillation activity. The problem of modeling the bubbles and their time and space evolution can be addressed, simplifying the medium equations. This is a quite complicated problem as not only the instability processes at a given altitude should develop but in addition the coupling between different ionosphere altitude layers should be made possible as the bubbles might have extent of hundreds of kilometers in altitude with different instability processes and significant differences in densities of the ionosphere species depending on the altitude. A bubble model gives more insight into the electron density gradients but still remains not properly related to real-time observations. The technique presented in this paper relies on real-time measurements. It consequently properly addresses the problem of accurate medium modeling.

The algorithm developed is based on a Kriging technique. It uses real or near-real-time data recorded by a network of 50 Hz receivers. These data are supplemented on occasion by 1 Hz data provided by the International GNSS Service (IGS) network, in which GNSS is the Global Navigation Satellite System. The Kriging technique can be seen as a data assimilation technique. The accuracy of the results is related to the accuracy and quantity of the measurements. The more data that are available at a given location, the more accurate is the resulting map. The GISM model is used as a background tool to fill the gaps between the measurements data points providing the algorithm with the "regression model" which plays a major role in this technique.

Table 1. Approaches and Related Regression 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	$f_1(x) = x^3$; $f_2(x) = 4\cos(3x)$

Details on the algorithm are given in the next section. As an example, a result of a scintillation map over Peru has been plotted. The data assimilated have been recorded by the European Space Agency (ESA) Monitor network [*Prieto and Béniguel*, 2011] and the Low-Latitude Ionospheric Scintillation Network (LISN) [*Valladares et al.*, 2008]. The accuracy of the algorithm has been verified using the cross-validation technique which consists in removing some of the data points to check that the results evaluated at the same points fit within the confidence level range given by the algorithm.

2. Calculation Technique

The Kriging technique [*Lefebvre et al.*, 1996] was selected for the following reasons:

1. It is an exact interpolator. The error is null at the data points, this is a property of the algorithm;
2. It is well suited to interpolate multivariate random data;
3. It provides an estimation of the interpolation error.

The technique consequently allows plotting scintillation maps and concurrently the related error maps. It addresses both the intensity and the phase scintillation maps.

2.1. Parameterization of the Model

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

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

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

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

$$r^T(x) = \begin{pmatrix} f_1(x) & \cdots & f_M(x) \end{pmatrix} \tag{2}$$

$$p^T(x) = \begin{pmatrix} p_1 & \cdots & p_M \end{pmatrix} \tag{3}$$

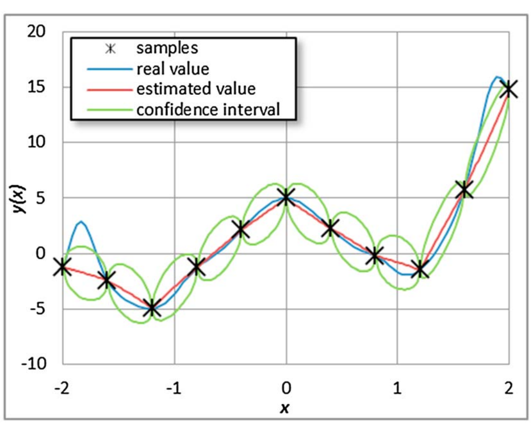

Figure 1. Approach A: no a priori on the model.

The measurements y_i^s ($i \in [1; n]$), gathered in the vector Y_s, are recorded at n samples locations $\boldsymbol{x}_i^s$, contained in the matrix 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:

$$R_s = \begin{pmatrix} f_1(\boldsymbol{x}_1^s) & \cdots & f_M(\boldsymbol{x}_1^s) \\ \vdots & \ddots & \vdots \\ f_1(\boldsymbol{x}_n^s) & \cdots & f_M(\boldsymbol{x}_n^s) \end{pmatrix} \tag{5}$$

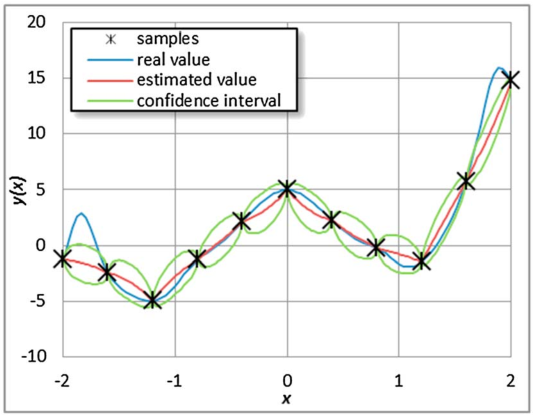

Figure 2. Approach B: interpolation of the model with a polynomial function.

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

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

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

$$\begin{aligned} C_s(i,j) &= \sigma_\omega^2 * C_{s\omega}(i,j) \\ &= \mathrm{cov}\left(\omega\left(\boldsymbol{x}_i^s\right), \omega\left(\boldsymbol{x}_j^s\right)\right) \end{aligned} \tag{7}$$

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 *Lefebvre et al.* [1996] or any other function, such as in *De Doncker et al.* [2006]. 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 *Sayin et al.* [2007]:

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

where the range a has to be optimized, using the estimated value of the weight coefficients p_{est} and σ_ω^2, given by [*Lefebvre et al.*, 1996]

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

To optimize the range a, as suggested in *Lefebvre et al.* [1996] and *Zhuo et al.* [2011], the determinant of the covariance matrix should be minimized.

2.2. 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 matrix X_e. The estimated values y_i^e constitute the vector Y_e:

$$Y_e = \begin{pmatrix} y_1^e \\ \vdots \\ y_N^e \end{pmatrix} ; X_e = \begin{pmatrix} \boldsymbol{x}_1^e \\ \vdots \\ \boldsymbol{x}_N^e \end{pmatrix} \tag{10}$$

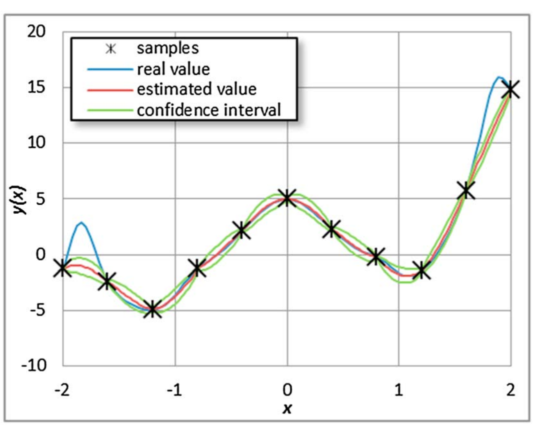

Figure 3. The model is assumed to be very well known.

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

$$R_e = \begin{pmatrix} f_1\left(\boldsymbol{x}_1^e\right) & \cdots & f_M\left(\boldsymbol{x}_1^e\right) \\ \vdots & \ddots & \vdots \\ f_1\left(\boldsymbol{x}_N^e\right) & \cdots & f_M\left(\boldsymbol{x}_N^e\right) \end{pmatrix} \tag{11}$$

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

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

Table 2. Approaches and Related Regression Models, Considering the Variables for Latitude and Longitude

Approach	Number of Models M	Model Definition
A	1	$f_1(x) = 1$
B	6	$f_i(x) = \text{lat}^i\,\text{long}^j\,; 0 \leq i+j \leq 2$
C	1	$f_1(\text{lat, long}) = \text{GISM}(\text{lat, long})$

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

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

Finally, the estimated values of the observable are given by

$$Y_e = \left[R_e A + C_e C_s^{-1} (I_d - R_s A)\right] Y_s \tag{14}$$

with A defined as follows:

$$A = \left(R_s^t C_{s\omega}^{-1} R_s\right)^{-1} R_s^t C_{s\omega}^{-1} \tag{15}$$

The equation (14) is actually the development of the equation (1), the first term corresponds to the linear combination of the models, while the second term corresponds to the random fluctuation. The interpolation error $\epsilon_e\left(\boldsymbol{x}_i^e\right)$ can also be estimated, as suggested in *Lefebvre et al.* [1996].

3. Kriging a One-Dimensional Variable

3.1. Example

The method presented in the previous section 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:

$$y(x) = x^3 + 4\cos(3x) + \exp(-x\,\sin(6x)) \tag{16}$$

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 have been considered. The first one assumes that there is no a priori known model and that the observable fluctuates around a constant. The second approach states 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 Table 1.

3.2. Comparison of the Results Obtained by Different Approaches

The results obtained are presented below in Figures 1, 2, and 3.

The results of approach A and approach B are very close. The confidence interval,

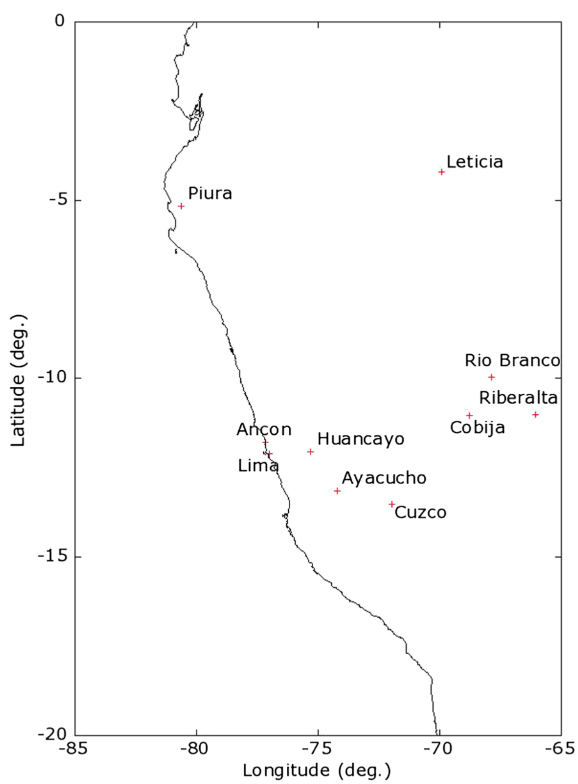

Figure 4. Stations of the LISN network and the Lima monitor station.

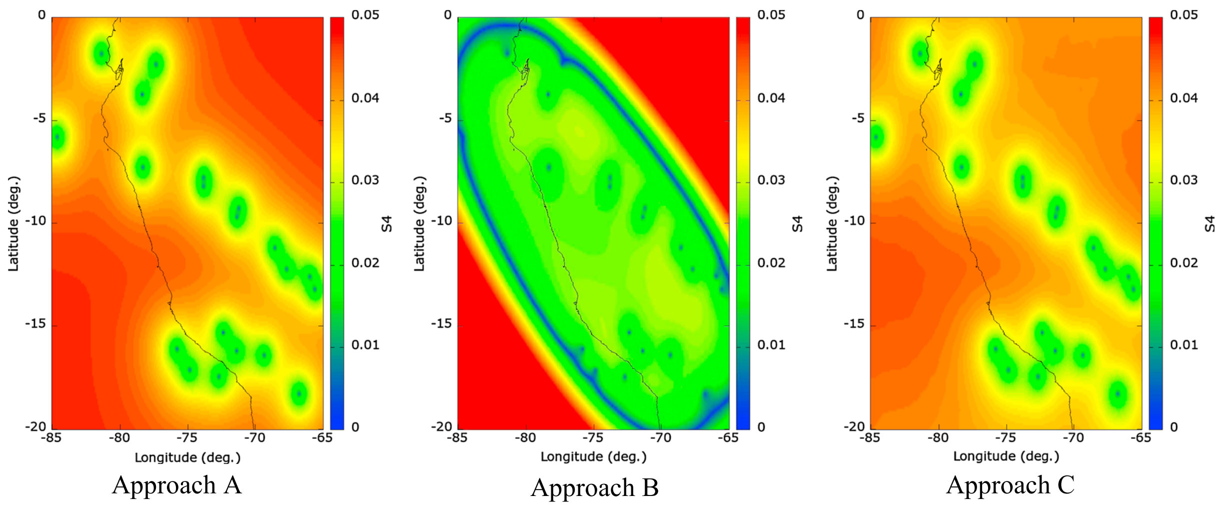

Figure 5. Estimated error.

related to the σ_ω^2 estimation, is 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 to 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 is the estimation provided by the Kriging algorithm.

4. Scintillation Map Generation

As shown in the previous section, the choice of the deterministic model used as background for the observables (currently the S4 values) is critical. The GISM scintillation model has been selected for this purpose as a regression model. It is compared to a constant model and to a polynomial model. These regression models are summarized in Table 2, considering the variable *lat* (latitude) and *long* (longitude).

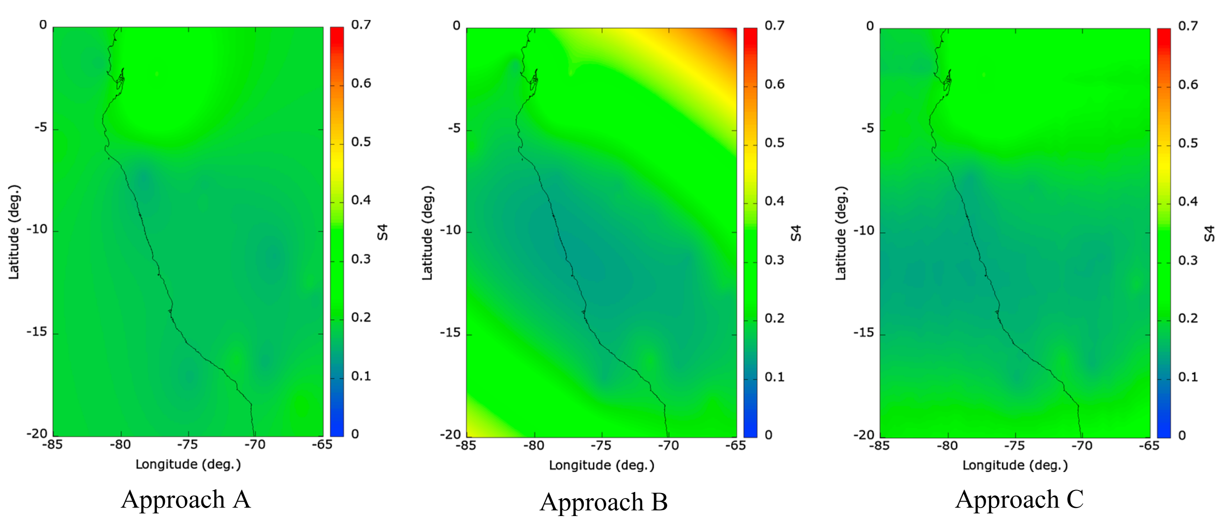

Figure 6. Estimated value.

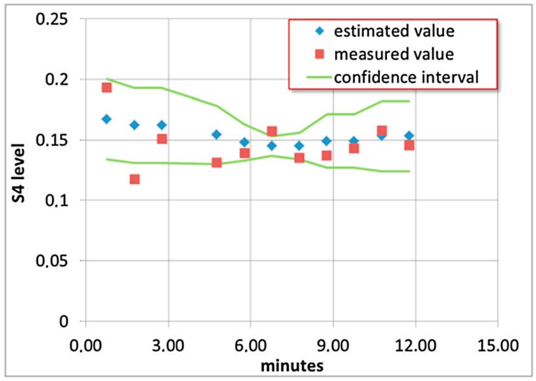

Figure 7. Estimated and real value on the PRN 29 over Piura.

The last approach is the result of a simulation performed on GISM. The S4 level has been simulated, on a grid, which dimensions are [−85°; −65°] for latitude and [−20°; 0°] for longitude. The step in latitude and longitude is 0.5°. The S4 level of the model is linearly interpolated between three values on the grid generated by GISM. The locations of the stations recording data are shown in Figure 4.

This S4 index has been measured over Peru on 1 October 2012, at 2:30 am, for 15 min. The map is plotted at the Ionosphere Pierce Points altitude (set to 350 km), assuming a planar ionosphere. To avoid multiple samples at the same location (which would result in a singular covariance matrix), the map is divided into several areas. Samples are merged in a single value in these unit cell areas. The estimated errors on the S4 index and the estimated value are depicted in Figures 5 and 6.

According to the Kriging technique, the estimated error is null at the sample points (the blue points in Figures 5), for any regression model. The first and third approaches provide equivalent results. The lowest estimated error is obtained with approach C. Approach B gives a good estimate close to the sampling data points but fails to estimate the values outside this region: using this approach, no extrapolation seems to be possible, the estimated value actually diverges over the sampling area. For the three approaches, the estimated error is lower than 0.05, which is very low as compared to the average S4 values.

5. Validation

To validate the method, 25% of the sample points were removed from the data set. One-track satellite has been removed. The Kriging technique has then been applied to estimate the values of the missing points, using GISM as a regression model. The results obtained, shown in Figures 7 and 8, give an absolute estimated error of the S4 index that is lower than 0.05 on the missing satellite track.

Seventy-six percent of the measured samples are within the confidence interval, which is a good accuracy.

6. Conclusion

A technique based on a Kriging algorithm, allowing to produce near-real-time ionosphere scintillation maps, has been presented. This technique is an interpolation technique. It uses scintillation data measurements assimilated with an algorithm developed for random variables. As an interpolation technique, there is no error at the measurement data points. The measurement data points are usually sparsely distributed, and the algorithm aims to extrapolate in space in order to build the scintillation maps. It showed the importance of taking a representative background regression model into account. The climatological GISM scintillation model was used for this purpose.

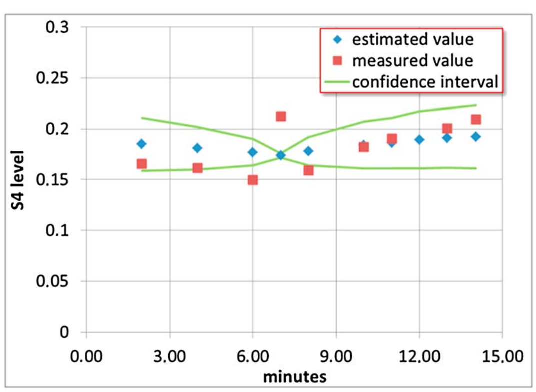

Figure 8. Estimated and real value on the PRN 29 over Cuzco.

The algorithm provides maps and concurrently error maps. The accuracy of the algorithm was checked using a cross-validation technique, removing some of the data points and running the algorithm to obtain an estimate at their location. The results obtained were shown to be within the estimated confidence level.

As mentioned in the paper, the more data that are available, the more accurate is the result. The map accuracy will increase consequently in the future as the new satellite constellations are being populated. The capability to include additional low-frequency data taken from the IGS network is also under analysis.

Acknowledgments
The authors want to express their acknowledgment to Cesar Valladares and colleagues from Boston College and to Juan Carlos Palacios from the Peru Geophysical Institute who have made available for this study the data from the LISN network. The data used to generate the maps presented in this paper were taken from the ESA Monitor website (http://monitor.estec.esa.int/Monitor/ dataset file name: lima275c.12.sct) and from the LISN website (http://lisn.igp.gob.pe/data/ dataset file names: lcuz_121001.s4; laya_121001.s4; lanc_121001.s4; lpiu_121001.s4; lhyo_121001.s4; lrib_121001.s4; lcob_121001.s4; lrbr_121001.s4; lle__121001.s4). This paper is under ESA contract 31083.

References

Aarons, J. (1982), Global morphology of ionospheric scintillations, *Proc. IEEE*, *70*(4), 360–378.

Basu, S., and S. Basu (1985), Equatorial scintillations: Advances since ISEA- 6, *J. Atmos. Sol. Terr. Phys.*, *47*, 753–768.

Béniguel, Y., and P. Hamel (2011), A global ionosphere scintillation propagation model for equatorial regions, *J. Space Weather Space Clim.*, *1*, A04, doi:10.1051/swsc/2011004.

Béniguel, Y., J.-P. Adam, N. Jakowski, T. Noack, V. Wilken, J.-J. Valette, M. Cueto, A. Bourdillon, P. Lassudrie-Duchesne, and B. Arbesser-Rastburg (2009), Analysis of scintillation recorded during the PRIS measurement campaign, *Radio Sci.*, *44*, RS0A30, doi:10.1029/2008RS004090.

Cervera, M., and R. M. Thomas (2006), Latitudinal and temporal variations of equatorial ionospheric irregularities determined from GPS scintillation observations, *Ann. Geophys.*, *24*, 3329–3341.

De Doncker, P., J.-M. Dricot, R. Meys, M. Hélier, and W. Tabbara (2006), Electromagnetic fields estimation using spatial statistics, *Electromagnetics*, *26*, 111–122.

Lefebvre, J., H. Roussel, E. Walter, D. Lecointe, and W. Tabbara (1996), Prediction from wrong models: The Kriging approach, *IEEE Antennas Propag. Mag.*, *38*(4), 35–45.

Prieto, C. R., and Y. Béniguel (2011), *The MONITOR Project: Architecture, Data and Products*, Ionospheric Effects Symposium, Alexandria, Va.

Rodrigues, F., E. R. De Paula, M. Abdu, A. Jardim, K. Iyer, P. Kintner, and D. Hysel (2004), Equatorial spread F irregularity characteristics over Sao Luis, Brazil, using VHF radar and GPS scintillation techniques, *Radio Sci.*, RS1S31, doi:10.1029/2002RS002826.

Sayin, I., F. Arikan, and O. Arikan (2007), Synthetic TEC mapping with ordinary and universal kriging, RAST 2007, *Istanbul*, 14–16.

Valladares, C. E., T. Bullett, J. L. Chau, J. V. Ecles, E. Kudeki, and R. F. Woodman (2008), *The Low-Latitude Ionospheric Sensor Network (LISN)*, CEDAR Workshop, Zermatt, Utah.

Zhuo, W., Prabhat, C. Paciorek, C. Kaufman, and W. Bethel (2011), Parallel kriging analysis for large spatial datasets. ICDMW '11 Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops Pages 38–44.