Ground Network Optimization Using Satellite Information: Application to the French Heliographic Network

S GALLE AND J D CREUTIN

Abstract—This paper deals with the setting up of new sunshine recording stations using satellite information Most statistical criteria used to select new measurement locations rely on knowledge of the correlation structure of the measured phenomenon As this structure is generally not known extensively, it must be modeled under homogeneity assumptions An alternative solution is proposed to avoid such modeling by using instead the correlation structure deduced from re-

Manuscript received September 21, 1987

S Galle was with the Institut de Mécanique de Grenoble, Domaine Universitaire, B P No 53X, 38041 Grenoble Cédex, France She is now with CEMAGREF, Division Hydrologie, BP 121-92164, Antony Cédex, France

J D Creutin is with the Institut de Mécanique de Grenoble, Domaine Universitaire, B P No 53X, 38041 Grenoble Cédex, France

IEEE Log Number 8927106

0196-2892/89/0700-0459$01 00 © 1989 IEEE

mote-sensed measurements of the studied phenomenon. A case study is presented.

***Keywords*—Network optimization, clustering, solar radiation.**

I. Introduction

A classic problem in environmental sciences is the rational locating of ground measurement devices for optimum measurement network design. This problem presents itself in two different contexts: i) The selection of a few devices within a redundant network, or ii) the addition of new devices to reinforce an existing network or to create a new one. Statistical literature offers a wide variety of methods capable of dealing with the first context: Data analyses and clustering techniques provide criteria for a rational selection based on the correlations between the existing measurement stations. The selected stations are then sufficient to preserve most of the information originally collected by the redundant network. The second design context is drastically different because no redundant information is available.

Addressing this problem, numerous studies have defined general rules concerning the optimal spacing of devices for a given phenomenon. The best way to obtain such rules is to consider the measured phenomenon as a random process. For any given network configuration the error variance can be computed for point or space-averaged estimates in the studied area [3]–[5], [8]. New device locations are selected to minimize such error variances.

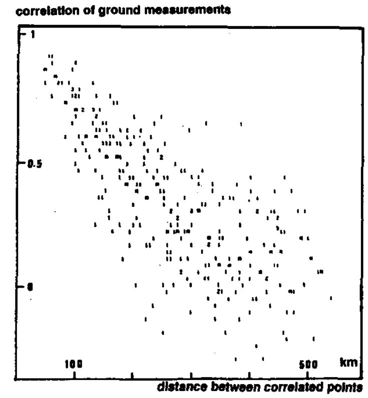
The basic concept of random process theory is the statistical homogeneity of the studied phenomenon, meaning that the statistical properties are invariant in space. Unfortunately, most natural phenomena do not satisfy this assumption; especially the correlation between two point time series is not specifically connected with the distance between the two points (cf. the dispersion of most experimental autocorrelation functions).

This paper proposes an alternative solution to process modeling when remote-sensing information is available. The high spatial density of this information allows the remote system to play the role of a very dense measurement network. Assuming that the ground and remote measurements of the studied phenomenon have the same statistical properties, particularly concerning the correlation structure, then the problem of reinforcing the ground network becomes a problem of selecting stations among the redundant "remote network" (i.e., the remote-sensing grid). The main advantage of this approach is that it avoids homogeneity assumptions. Its main weakness is that it requires i) a reasonably long series of remote measurements to provide robust estimates of the correlations, and ii) some clear evidence of the good agreement between the statistical properties of the two kinds of measurements.

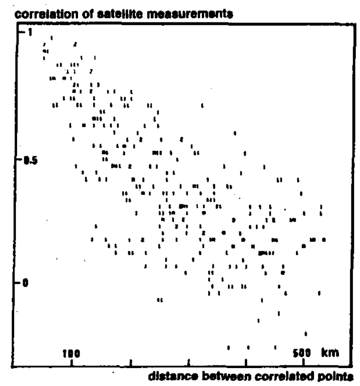
A real world example is presented. It concerns daily sunshine ratio measurement over the Southwestern part of France (from 43° to 46° latitude and from 1°30 W to 1°30 E longitude) where a network of 23 heliographs already exists. The daily sunshine ratios are the relative numbers of hours with sunshine. The satellite estimates come from Meteosat imagery through a threshold method computing percentages of cloud coverage over a 20 × 20 km square grid, named the Cactus method of the French Met. Office [7]. The period studied ran from the beginning of March to the end of May, 1984.

II. Evidence of the Correlation Structure Consistency Between Ground and Remote Measurements

The similarity of the correlation structures of the two measurement sets can initially be checked by comparing their autocorrelation functions. Such functions are obtained experimentally by relating the classical coefficient of correlation, computed for each pair of measurement stations, with the geographical distance between the stations. Due to sampling fluctuations and heterogeneity of the statistical properties, the original experimental forms of the autocorrelation functions are, in this case, widely scattered point swarms (see Fig. 1).



(a)



(b)

Fig. 1. Experimental autocorrelation swarms for (a) ground, and (b) remote-sensing measurements.

Computation of the mean correlation over classes of distance provides smoother forms of these experimental functions, from which we can draw conclusions concerning the similarity between the mean correlation structures of the two measurement sets. This conclusion can be fulfilled by examining the cross-correlation function, obtained in the same way, except that the coefficients of correlation are computed between ground measurements at the first station and remote measurements at the second station of each pair. Fig. 2 shows the good agreement of the three functions except for very short distances, indicating that the two kinds of measurements are only separated by a random noise of low variance. Thus the mean autocorrelation function of either ground or remote measurements can be used for network design under a homogeneity assumption.

However, our purpose is to avoid the homogeneity assumption by taking advantage of the detailed remote-sensing knowledge of the correlation structure.

Thus the similarity check must be pursued one step further by comparing the correlation fluctuations due to heterogeneity. In other words, we have to compare experimental swarms rather than their means values. The simplest way to do this is to plot the ground correlation versus the remote correlation for the pairs of points common to both networks. Fig. 3 shows the obtained scattergram. Note that the dispersion of this swarm is remarkably narrow, especially with respect to the autocorrelation function swarm. This can be considered as clear evidence that the correlation structure deduced from remote-sensing information is sufficiently consistent with the ground correlation structure to be substituted for it in network design.

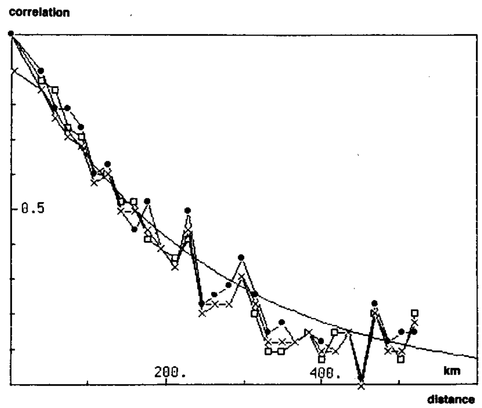


Fig. 2. Mean autocorrelation functions for (□) ground and (●) remote-sensing measurements and their corresponding cross-correlation function (×) (20 classes of distance are used; an exponential model is proposed).

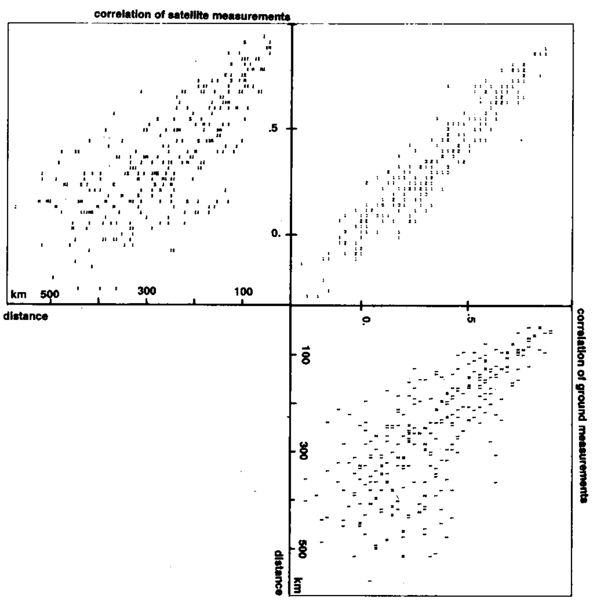


Fig. 3. Scattergram of ground versus remote-sensing coefficients of correlation for the 253 pairs of measurement points common to the two networks.

III. Criteria Used for Selecting New Measurement Sites

In so far as we can consider that ground and remote measurements have the same statistical properties, the problem is now reduced to selecting among the points of the satellite grid (set G) the most appropriate point p to reinforce the ground network N. As noted in the Introduction, this selection is based on the satellite information due to its higher measurement density. Naturally, some geographical approximations are needed since the ground network stations are not generally located at grid points.

Various methods for selecting variables have been presented in the literature. Some are based on the notion of distance and are aimed at constituting clusters of variables (e.g., in [2]). Selection then involves choosing one variable in each cluster. Other methods use the information concept and provide a direct selection of variables satisfying criteria of explained variance (see [1] or [6]). Since the first method can provide clusters with a poor geographical consistency, we prefer the second, which has the additional advantage of being more straightforward.

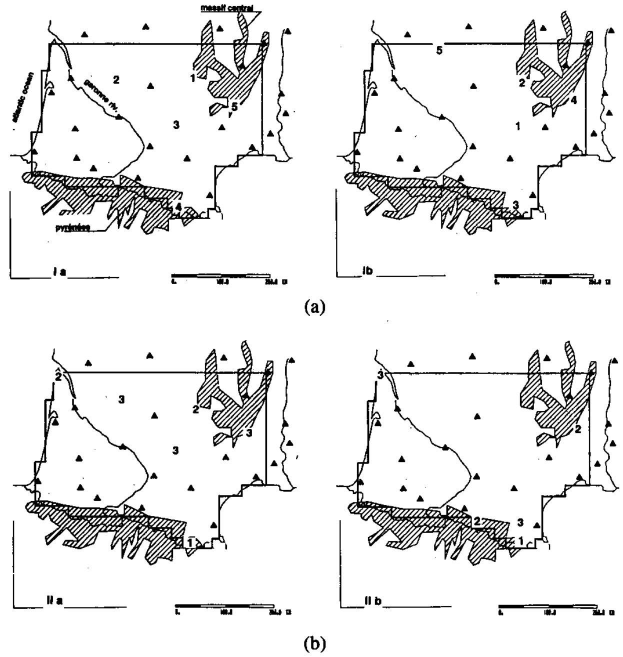


Fig. 4. Location of the new devices, numbered in accordance with their rank of selection, selected by the first criterion (I) or the second (II) using a modeled autocorrelation function (a) or the correlations computed from the remotely sensed measurements (b). The original network location is given by the black triangles.

Within these methods two selection procedures can be distinguished: i) A descending procedure which discards, at each step, the least interesting variable, and ii) a ascending procedure which selects, at each step, the most interesting variable. The necessity of taking into account the existence of a given network implies the use of an ascending method, for which the first steps are forced to select the existing network points.

Only two selection criteria are applied in the present study.

The first one selects at the first step a point p belonging to the set G–N (set of the grid points after removing the network points). This point is selected so that the mean partial correlation between p and the points j from G–N–p is as strong as possible. In other words, p is such that the mean residual variance $S^2_{j/N+p}$ when j describes G–N–p is minimum (this variance is the mean square value of the residuals obtained when the variable j is explained by the variable $N + p$ through a multiple regression). The second step selects the point m so that

$$\text{Min}_{m\ G\text{-}N\text{-}p}\left(\text{Mean}_{j\ G\text{-}N\text{-}p\text{-}m} S^2_{j/N+p+m}\right)$$

and so on.

The second criterion proceeds in a different manner. The first point selected p shows the weakest partial correlation with the network N. Thus, p from G–S is such that $S^2_{p/N}$ is maximum. In the same way, m satisfies at the second step:

$$\text{Max}_{m\ G\text{-}N\text{-}p} S^2_{m/N+p}.$$

This criterion corresponds in fact to the minimization of the estimation variance commonly used in objective analysis (see [5] or [3]).

IV. Results

The above criteria have been applied to the studied area in order to reinforce the 23-heliograph network. The correlation used by both criteria are successively deduced i) directly from remote-measurement correlations computed over the satellite grid in order to show the results of the proposed method, and ii) from the mean autocorrelation function in order to show the influence of the classical homogeneity assumption.

The obtained results are presented in Fig. 4, where the new measurement sites proposed are numbered in accordance with their rank of selection.

These results suggest two general comments:

- Depending on the chosen criterion, the selected sites are preferably situated either in the lowest density areas inside the domain (such locations are chosen by the first criterion to obtain a new network minimizing the average mean square error) or on the boundaries of the domain (where the larger variances of error are computed by the second criterion).
- Depending on the correlation pattern used, the selected sites are distributed either in a purely geometrical way (when homogeneity is assumed) or in a more physical manner (when the remote-sensing correlation structure is used). Thus in the latter case the mountainous regions of the Massif Central (in the northeast corner of the area) and of the Pyrénées (along the southern boundary of the area) are preferred to the lower region of the Garonne River basin (western half on the area) for the selection of new sites.

V. Conclusion

This study must be considered as a simple illustration of a possible application of a remote-sensing device in a classic climatological problem. The obtained results are consistent with *a priori* ideas on the heterogeneity of the phenomenon which would be expected to present a higher variability in hilly terrains. Nevertheless, some doubt remains concerning the effect of sampling fluctuation on this kind of approach. As these fluctuations play the same role as phenomenon heterogeneity in the correlation structure, it is a heavy task to separate their respective effects. Attempts are under way to solve this problem, but the best answer is clearly to use a longer time series of measurements.

References

[1] E. M. Beale, M. G. Kendall, and D. W. Mann, "The discarding of variables in multivariate analysis," *Biometrika*, vol. 54, pp. 357-366, 1967.

[2] J. D. Creutin and C. Obled, "Elimination de variables et optimisation de réseaux de mesure," in *Data Analysis and Informatics*, E. Diday *et al.*, Eds., pp. 759-775, 1980.

[3] J. P. Delhomme and P. Delfiner, "Application du krigeage à l'optimisation d'une campagne pluviométrique en zone aride," Colloque UNESCO-OMM-AIHS sur l'élaboration des projets d'utilisation des ressources en eau sans données suffisantes Tome I, Madrid, 1973, pp. 191-210.

[4] O. A. Drozdov, "A method for setting up a network of meteorological stations in a level region," *Trudy GGO*, vol. 12, no. 3, 1936.

[5] L. S. Gandin, *Objective Analysis of Meteorological Fields.* Jerusalem: Israel Program for Scientific Translations, 1965, 242pp.

[6] I. T. Jollife, "Discarding variables in a principal components analysis," *J. Appl. Stat.*, vol. 21, no. 2, pp. 160-173, 1972.

[7] C. Pastre, "Developpement d'une méthode de détermination du rayonnement solaire global à partir des données Météosat," *La Météorologie*, vol. 24, pp. 6-15, 1981.

[8] I. I. Zawadzky, "Errors and fluctuations of rain-gauge estimates of areal rainfall," *J. Hydrology*, vol. 18, pp. 243-255, 1973.

Erratum

The following errors appeared in the paper "Invertible Canopy Reflectance Modeling of Vegetation Structure in Semiarid Woodland" by Janet Franklin and Alan H. Strahler in the November 1988 issue of the IEEE Transactions on Geoscience and Remote Sensing, vol. 26, no. 6, pp. 809–825:

The equation on p. 810 should appear as follows:

$$\Gamma = \pi + \frac{\pi}{\cos\theta'} - A_0$$

where

$$A_0 = \left(\beta - \frac{1}{2}\sin 2\beta\right)\left(1 + \frac{1}{\cos\theta'}\right)$$

if $(b + h)\tan\theta > r(1 + 1/\cos\theta')$, otherwise $A_0 = 0$. In this expression

$$\beta = \cos^{-1}\left[\left(1 + \frac{h}{b}\right)\left(\frac{1 - \cos\theta'}{\sin\theta'}\right)\right]$$

and

$$\theta' = \tan^{-1}\left(\frac{b}{r}\tan\theta\right).$$

The correct formulae for equations (4), (5), and (6) on pp. 811 and 812 are

$$V(m) \approx (N + C_{r^2}^2 N + C_{r^2}^2)(\boldsymbol{R}^2)^2$$
$$= (M + C_{r^2}^2 M + C_{r^2}^2 \boldsymbol{R}^2)\boldsymbol{R}^2 \quad (4)$$

$$\boldsymbol{R}^2 = \frac{\left[(1 + C_{r^2}^2)^2 M^2 + 4V(m)C_{r^2}^2\right]^{1/2} - (1 + C_{r^2}^2)M}{2C_{r^2}^2}. \quad (5)$$

$$\boldsymbol{R}^2 \approx \frac{V(m)}{(1 + C_{r^2}^2)M}. \quad (6)$$

The last sentence on p. 814 (which continues on p. 818) should read: "To determine whether or not the *mean and* variance should be based on within-site measurements, ..."

The last sentence of that paragraph on p. 818 should read: "Accordingly, the *means and* standard deviations shown . . . ."

Fig. 5 (a)-(f) (pp. 815-817) should appear at a reduced scale, on one page.

The second complete paragraph in the second column of p. 822 should begin: "For several of the sites (2, 4, 5, 7), ..." In the last sentence of that same paragraph the equation should read:

$$X_0' = 0.98X_0.$$

X. Li's name was spelled incorrectly in the Acknowledgment section.

There were other minor errors that do not obscure the meaning of the paper.