Scanning Microscopy

Volume 1992
Number 6 *Signal and Image Processing in Microscopy and Microanalysis*

Article 13

1992

Nonlinear Statistical Filtering and Applications to Segregation in Steels from Microprobe Images

C. Daly
Ecole Nationale Superieure des Mines de Paris

D. Jeulin
Ecole Nationale Superieure des Mines de Paris

D. Benoît
IRSID

Follow this and additional works at: https://digitalcommons.usu.edu/microscopy

Part of the Biology Commons

Recommended Citation

Daly, C.; Jeulin, D.; and Benoît, D. (1992) "Nonlinear Statistical Filtering and Applications to Segregation in Steels from Microprobe Images," *Scanning Microscopy*: Vol. 1992 : No. 6 , Article 13.
Available at: https://digitalcommons.usu.edu/microscopy/vol1992/iss6/13

This Article is brought to you for free and open access by the Western Dairy Center at DigitalCommons@USU. It has been accepted for inclusion in Scanning Microscopy by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.

UtahStateUniversity
MERRILL-CAZIER LIBRARY

Scanning Microscopy Supplement 6, 1992 (pages 137-145) 0892-953X/92$5.00+.00
Scanning Microscopy International, Chicago (AMF O'Hare), IL 60666 USA

NONLINEAR STATISTICAL FILTERING AND APPLICATIONS TO SEGREGATION IN STEELS FROM MICROPROBE IMAGES

C. Daly[*,1], D. Jeulin[1], D. Benoît[2]

[1]Centre de Géostatistique, Ecole Nationale Superieure des Mines de Paris, 35 rue St-Honoré, 77305 Fontainebleau, France

[2]IRSID, 34 rue de la Croix de Fer, 78105 Saint-Germain-en-Laye, France

Abstract

Microprobe images of solidification studies are well known to be subject to a Poisson noise. That is, the radiation count at a pixel x for a certain element may be considered to be an observation of a Poisson random variable whose parameter is equal to the true chemical concentration of the element at x. By modeling the image as a random function, we are able to use geostatistical techniques to perform various filtering operations. This filtering of the image itself may be done using linear kriging. For explicitely nonlinear problems such as the estimation of the underlying histogram of the noisy image, or the estimation of the probability that locally the concentration passes a certain value (this probability is needed for segregation studies), it is usually not possible to use linear techniques as they give biased results. For this reason, we applied the nonlinear technique of Disjunctive Kriging to these nonlinear problems. Linear kriging needs only second order statistical models (covariance functions or variograms) while disjunctive kriging needs bivariate distribution models. This approach is illustrated by examples of filtering of various X−ray mappings in steel samples.

Key Words : Statistical filters, probability map, chemical segregation, isofactorial models, disjunctive kriging.

*Address for correspondence:
C. Daly
British Petroleum, Sunbury Research Centre,
Chertsey Road, Sunbury-on-Thames,
Middlesex TW16 7LN, U.K.
Telephone No.: 44 (932) 763474
Fax No.: 44 (932) 764459

Introduction

A common problem in image analysis is that of image filtering. For example in Figure 1, we see an image of chemical concentration obtained in a steel sample using a microprobe. Despite a high acquisition time (often about 7 hours), the image remains quite noisy for elements of fairly low concentration. The simplest type of image filtering we might want to perform is to try to remove the noise from the image leaving an estimate of the underlying signal. For images of a diffuse type, such as that of the chemical concentration, a number of different types of filters give fairly good results. For example, amongst the statistical filters we have the linear kriging filter, the disjunctive kriging (see the following sections), or the filters associated with Markov random fields [1]. The kriging procedure was developed in the field of geostatistics [7] in the fifties to estimate unknown or noisy data, accounting for information about the spatial structure of the underlying phenomena. The early applications of geostatistics concerned the estimation of local concentrations in orebody deposits from probes at a given number of sites. The terminology "kriging" was used in honour of Dr. D.G. Krige who initiated this practice for gold deposits. Amongst the non statistical filters we have the median filter and the morphological filters [10].

However if instead of estimating simply the underlying maps, we are interested in estimating a more complex function such as a local probability map (probability that within a small neighbourhood v, a point chosen at random passes a certain value), then the type of method chosen seems to be of more importance. Later, we will develop the non linear disjunctive kriging filter and will apply it to estimating the local probability map. Using our estimate of the local probability, we will then be able to construct local segregation maps for steel samples which allow us to quantitatively compare the segregation properties of different types of steels. In the case of a chemical element with very low concentration but with very high influence on the in−use properties, it is important to know the true highest values occurring in the specimen. It requires performing an unbiased estimation of the underlying histogram (local or global) from noisy images.

Linear Kriging

Let us suppose that the true underlying chemical concentration at point x is Y(x) and the observed data is given as Z(x). The linear kriging filter of Y(x) from the known data {Z(x)} in a neighbourhood of x is an estimator

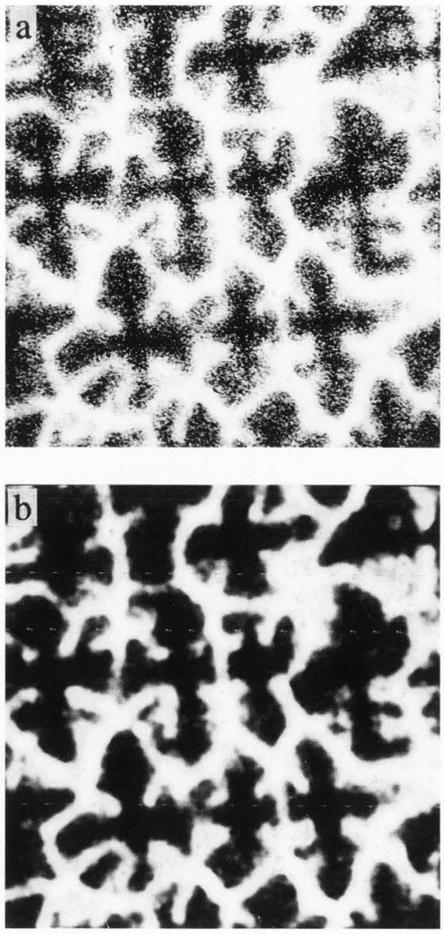


Figure 1. Improvement of a noisy microsegregation image obtained on a microprobe (**1a**) by a linear kriging filter (**1b**). $SNR_0 = 2.4$; $SNR = 83.8$; image 512 x 512, 3 μm/pixel.

$Y^*(x)$ of the unknown $Y(x)$:

$$Y^*(x) = \sum_i \lambda_i \, Z(x_i) \qquad (1)$$

where the $\{\lambda_i\}$ are chosen so that the variance of error $Var(Y(x) - Y^*(x))$ is a minimum subject to the constraint that the estimate be non biased, $E[Y(x) - Y^*(x)] = 0$, where E is the expectation (or mean) of the random variable $(Y(x) - Y^*(x))$.

If the covariance Cov of Z(x) is known,

$$\begin{aligned} Cov(Z(x), Z(y)) &= E[Z(x)\ Z(y)] - m^2 \\ m &= E[Z(x)] \end{aligned} \qquad (2)$$

and if Cov(Z(x), Y(y))is known also, then it can be shown that the $\{\lambda_i\}$ are given by :

$$\begin{aligned} \sum_i \lambda_i \, Cov(Z(x_i), Z(x_j)) + \mu &= Cov(Z(x_j), Y(x)) \\ \sum_i \lambda_i &= 1 \end{aligned} \qquad (3)$$

μ being a Lagrange parameter resulting from the imposed non bias condition.

For more information on linear kriging see [7] and for its application to filtering, see [4, 9] in the fields of microprobe and EELS (Electron Energy Loss Spectroscopy) electron microscopy imaging. The optimal (and adpative, since it depends on the covariance of the image) linear filter is a generalization of the Wiener filter, well known in the literature of signal processing.

The covariance function Cov(Z(x),Z(y)) may be calculated from the experimental data by making the hypothesis of stationarity, that is, by assuming that the covariance depends only on the distance (x − y) between the points, and then by calculating an experimental version of the covariance function on the observed data. The covariance function Cov(Z(x),Y(y)) may be found exactly if we have two independent acquisitions of noisy data $Z^1(x)$ and $Z^2(x)$ in which case it can be shown that :

$$Cov(Z^1(x), Z^2(x)) = Cov(Z(x), Y(y)). \qquad (4)$$

Failing this, it is sometimes possible to know Cov(Z(x),Y(y)) using only one acquisition. For example in the case of Poisson noise (eg. for microprobe images), we assume that each pixel x is obtained as a realization of a Poisson random variable with parameter Y(x). The covariance function for Z is equal to the cross covariance between Y and Z except at the origin where Var Z = Cov(Y,Z) + m. For more details see [2].

An illustrative example of application is given in Figure 1 for the case of a microprobe image in a steel specimen. It appears that the filtered image shows very fine details concerning secondary arms with dendrites which are not visible on the initial noisy image.

Recent extensions of the linear kriging filter include its implementation for the deconvolution of noisy images [6] which might be very useful to improve the spatial resolution of high magnification X−ray images, among others, obtained by combining a smoothing process (X−ray data arise from a non punctual source) and a noise due to the detection of photons.

The variance of estimation of the estimator Y^*, σ_K^2, is given by

$$\sigma_K^2 = D^2[Y - Y^*] = E[Y^2(x)] - \sum \lambda_i \, Cov[Z(x_i)Y(x)] \qquad (5)$$

From the variance of estimation, it is easy to calculate a signal to noise ratio (SNR) :

$$SNR = \frac{Var[Y(x)]}{\sigma_K^2} \qquad (6)$$

$$SNR_0 = \frac{Var[Y(x)]}{Var\ Z(x) - Var\ Y(x)}, \qquad (7)$$

SNR_0 being the initial SNR, before filtering. The improvement resulting from the filter is measured by $\frac{SNR}{SNR_0} = \frac{Var\ Z(x) - Var\ Y(x)}{\sigma_K^2}$.

Disjunctive Kriging (D.K.)

The principal advantage of the linear kriging considered in the previous section is that it only depends on all the covariance functions Cov(Z(x),Z(y)) and Cov(Z(x),Y(y)). As such, since these functions are relatively easy to estimate from the data, the linear kriging is a robust estimate which is not usually prone to serious error. Its major drawback is that it is linear and as such is not well adapted to the estimation of nonlinear functions such as the local law. A non linear estimate which has the merit of not making very strong demands on the data (and so hopefully remaining fairly robust) is the disjunctive kriging (D.K.) filter. The D.K. filter depends only on the bivariate distributions. Nonetheless this hypothesis is stronger than that used by linear kriging which only depends on the covariance functions. It is by making these extra hypotheses that we allow ourselves the possibility of estimating non–linear functions of the true image such as the local law, but we must be careful not to make these hypotheses too strong or we might not be able to test them even partially with the data and so risk problems of robustness.

In this case, we estimate an arbitrary function of Y(x), say $\phi(Y(x))$ by :

$$\phi\,(Y(x))^* = \sum_i f_i(Z(x_i)) \tag{8}$$

so that our estimate is now a sum of functions of the data points rather than simply a linear sum as was the case for linear kriging. It can be shown [8] that the equations to be solved to determine the functions f_i are :

$$E[\phi(Y(x))\,|\,Z(x_j)] = \sum_{i=1}^{n} E[f_i(Z(x_i))\,|\,Z(x_j)] \quad j = 1,\dots,n \tag{9}$$

where $E[\phi(Y)\,|\,Z]$ is the conditional expectation of $\phi(Y)$ given Z. Unfortunately, these are a rather complicated system of integral equations whose numerical solution is quite demanding, and so we consider an approximate solution using an isofactorial model [8]. For full details see [2]. The main lines of this approach, as explained in the next section for the Gaussian model, consist in transforming the initial values into data admitting a given distribution (eg. a Gaussian distribution, but other models were introduced in [8]). Important simplifications of Equation (9) occur when the chosen distribution is the marginal distribution of a bivariate distribution admitting an isofactorial development, over orthogonal functions (usually polynomials). In that case, the nonlinear system (9) is replaced by a sequence of linear systems (21).

The minimal variance criterion used to design the linear and disjunctive kriging filters can be interpreted as the orthogonal projection of the variable to be estimated on a Euclidean space E, using the covariance functions as a scalar product :
– for the linear kriging, E is generated by the linear combinations of the data $Z(x_i)$,
– for the disjunctive kriging, E is generated by the linear combinations of the measurable functions of the data $f_i(Z(x_i))$.

The most general projection estimator, the conditional expectation, may be found by projecting the variable to be estimated on the space of all the measurable multivariate functions of the data $f(Z(x_1),\dots,Z(x_n))$. This operator requires the knowledge of multivariate distributions of order n, which are not accessible from the data, and therefore must be introduced from some *a priori* knowledge. As a consequence, it may be subject to strong discrepancies with the data, and hence might lack robustness in applications. For this reason, we prefer to concentrate on the two previous estimators, that are implemented with less requirements.

We can note in passing that other filtering or estimation techniques, such as the so–called Maximum Entropy algorithms, while not being based on a projection operation, still use in fact an *a priori* knowledge of multivariate distributions of very high order.

The Gaussian Isofactorial Model for Disjunctive Kriging

The only example of an isofactorial model that we wish to consider in this paper is the Gaussian model, whereby we assume that the noisy image (after a suitable trans–formation) has bivariate Gaussian laws.

The principal disadvantage in using an isofactorial model is that it is usually impossible to model both the spatial law for the images and the type of noise process at the same time. Thus, if we are using a Gaussian disjunctive kriging model, it will not be rigorously possible to have Poisson noise.

However for most image filtering purposes in steel samples, the Gaussian isofactorial model approximation to the Poisson noise characteristics appears to be quite good, which enables us to use the model heuristically. In the general case, that is for an arbitrary bivariate law with an arbitrary noise law, it will be necessary to solve the direct disjunctive kriging equations (1). In fact, these are easily calculated from the knowledge of the bivariate distributions of the underlying signal Y(x) and of the noise generation (which in this case is Poisson). Therefore a general approach for the D.K. is feasible if numerically somewhat demanding (an example was developed in [2]).

We will now consider the Gaussian isofactorial laws. Let us first suppose that Y(x) (the underlying chemical concentration) has a bivariate Gaussian law and a correlation function $\varrho(h)$. Then the bivariate law for a separation of h, $(Y(x) = y_1,\ Y(x+h) = y_2)$ is given by :

$$g_h^Y(y_1,y_2) = \frac{1}{\sqrt{1-\varrho^2}}\, e^{\frac{-1}{2(1-\varrho^2)}(\varrho^2 y_1^2 + \varrho^2 y_2^2 - 2\varrho y_1 y_2)}\, \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y_1^2}{2}}\, \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y_2^2}{2}} \tag{10}$$

If we call $g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$, then it can be shown that :

$$g_h^Y(y_1,y_2) = \sum_{n=0}^{\infty} \varrho^n(h)\, \eta_n(y_1)\, \eta_n(y_2)\, g(y_1)\, g(y_2) \tag{11}$$

where $\eta_n(y)$ is the normalised Hermite polynomial of order n. These polynomials are easily generated using the formulae :

$$\eta_0(y) = 1\ ;\ \eta_1(y) = -y$$

$$\eta_{n+1}(y) = -\frac{1}{\sqrt{n+1}}\, y\, \eta_n(y) - \sqrt{\frac{n}{n+1}}\, \eta_{n-1}(y) \tag{12}$$

and are also defined by the Rodrigues formula

$e^{-\frac{y^2}{2}} \eta_n(y) = \frac{1}{\sqrt{n!}} \frac{d^n}{dy^n} e^{-\frac{y^2}{2}}$. An important result for the bigaussian couple (Y(x),Y(x+h)) of correlation ϱ is that

$$E[\eta_n(Y(x+h)) \mid Y(x)] = \varrho^n \eta_n(Y(x)) \quad (13)$$

Next we consider the bivariate law between the noisy Z(x+h) and the underlying Y(x). We suppose that the randomisation due to the noise is of the form :

$$Z(x) = r\,Y(x) + \sqrt{1-r^2}\,\varepsilon(x) \quad (14)$$

where $\varepsilon(x)$ is a white noise random function independent of Y(x) and $0 \le r \le 1$. We assume that Y(x), Z(x) and $\varepsilon(x)$ are all gaussian with mean 0 and standard deviation 1 (N(0,1)) so that (Y(x), Z(x+h)) is bigaussian with correlation $r\varrho(h)$; so its bivariate distribution can be immediately written in the isofactorial form :

$$g_h^{Y,Z}(y,z) = \sum_{n=0}^{\infty} (r\varrho(h))^n \eta_n(y)\,\eta_n(z)\,g(y)\,g(z) \quad (15)$$

Moreover the couple (Z(x), Z(x+h)) is bigaussian with the correlation :

$$\varrho_Z(h) = \begin{cases} r^2\varrho(h) & \text{if } h > 0 \\ 1 & \text{if } h = 0 \end{cases} \quad (16)$$

so that its bivariate law is written as :

$$g_h^Z(y,z) = \sum \varrho_Z(h)^n \eta_n(y)\,\eta_n(z)\,g(y)\,g(z) \quad (17)$$

Now suppose we wish to estimate an arbitrary function $\phi(Y(x))$ from the $Z_\alpha(=Z(x_\alpha))$ data. Firstly, we express $\phi(Y(x))$ in terms of the Hermite polynomials. That is, we find the coefficients ϕ_n so that :

$$\phi(Y(x)) = \sum_0^{\infty} \phi_n\,\eta_n(Y(x)) \quad (18)$$

and then it can be shown that our D.K. estimate of $\phi(Y(x))$, namely $\phi^*(Y(x))$ is given as :

$$\phi^*(Y(x)) = \sum_0^{\infty} \phi_n\,\eta_n^*(Y(x)) \quad (19)$$

where

$$\eta_n^*(Y(x)) = \sum \lambda_\alpha^n\,\eta_n(Z_\alpha) \quad (20)$$

and where the $\{\lambda_\alpha^n\}$ are given as solutions to the equations

$$\sum_{\alpha=1}^{N} \lambda_\alpha^n\,\varrho_Z^n(x_\alpha - x_\beta) = r^n\varrho^n(x - x_\beta) \qquad \beta = 1,\ldots,N \quad (21)$$

The system (9) was simplified into the system (21) in the case of an isofactorial model (namely the Gaussian bivariate model in the present case).

The N points $\{x_\alpha\}^N$ are points in a neighbourhood of the point x where we want to estimate $\phi(Y(x))$. Theoretically we would have to solve these equations an infinite number of times to obtain our estimate $\phi^*(Y(x)) = \Sigma\ \phi_n\,\eta_n^*(Y(x))$. However in practice, the convergence is usually quick and we need no more than 9–10 terms of the development at the worst. Each term is solution of a particular kriging system (2), similar to the linear kriging system presented at the beginning of this paper.

The variance of estimation is given by :

$$\mathrm{Var}\left[\phi(Y(x)) - \phi^*(Y(x))\right] = \sum_{n=0}^{\infty} \phi_n\left[\phi_n - \sum_{\alpha=1}^{N} \lambda_\alpha^n\,r^n\varrho^n(x - x_\alpha)\right]. \quad (22)$$

We have noted that we can estimate an arbitrary function $\phi(Y(x))$. In practice two such functions are especially of interest.

The first such function is the anamorphosis function. This is of interest if we simply want to filter the noise from the noisy image (although we repeat here that for most steel sample images a linear filter is usually sufficient). In this case, it is most unlikely that the true underlying image Y(x) has a Gaussian law. So that we must find a transformation or anamorphosis function ϕ such that $Y(x) = \phi(Y(x))$ with Y(x) a N(0,1) variable. In addition, we make the assumption that every pair $\phi(Y(x))$, $\phi(Y(x+h))$ is bigaussian. No additional assumption concerning higher order distributions is required. Of course fitting the anamorphosis function requires a knowledge of the true underlying law of Y(x). This is not readily available from the data and must be estimated. The technique of regularisation, which in fact turns out to be equivalent to a form of linear kriging, may be used to estimate the underlying law [2, 3]. It cannot be represented here.

Once the anamorphosis function has been calculated, we may estimate $\phi(Y(x))$, and so obtain a nonlinear D.K. filter, as illustrated in the following section.

Estimation of a Local Probability Distribution by Disjunctive Kriging and Examples of Applications

Another nonlinear function which we may wish to estimate is the local histogram. This is the function we will use to construct a measure of segregation. We are interested in calculating the local histogram for support size v : we would like to estimate the probability that a point chosen at random in the neighbourhood of size v about a point x will pass a given cutoff. Formally, we wish to estimate :

$$P\left[Y(x) < y \,\middle|\, x \in v\right] = \frac{1}{v}\int_v \mathbf{1}_{\{Y(x)<y\}}\,dx \quad (23)$$

where $\mathbf{1}_{\{Y(x)<y\}} = 1$ if $Y(x)<y$, else 0.

The indicator term in the integral admits the development :

$$\mathbf{1}^*_{\{Y(x)<y\}} = \sum \phi_n\,\eta_n(Y(x)) \quad (24)$$

with the ϕ_n easily calculated from the Rodrigues formula for the Hermite polynomials.

Once this local probability function has been calculated, we can use it to define a local segregation ratio. A possible choice is to find the upper and lower 5% quantiles for the local law $z_n(x)$ and $z_e(x)$ respectively and then to define

$$TS(x) = \frac{z_n(x) - z_e(x)}{m} \quad (25)$$

where m is the global mean of the image. This choice of a definition has the advantage that it is scalar and so may be compared from one image to another. The estimate of TS(x) given by D.K. is unbiased and the influence of the

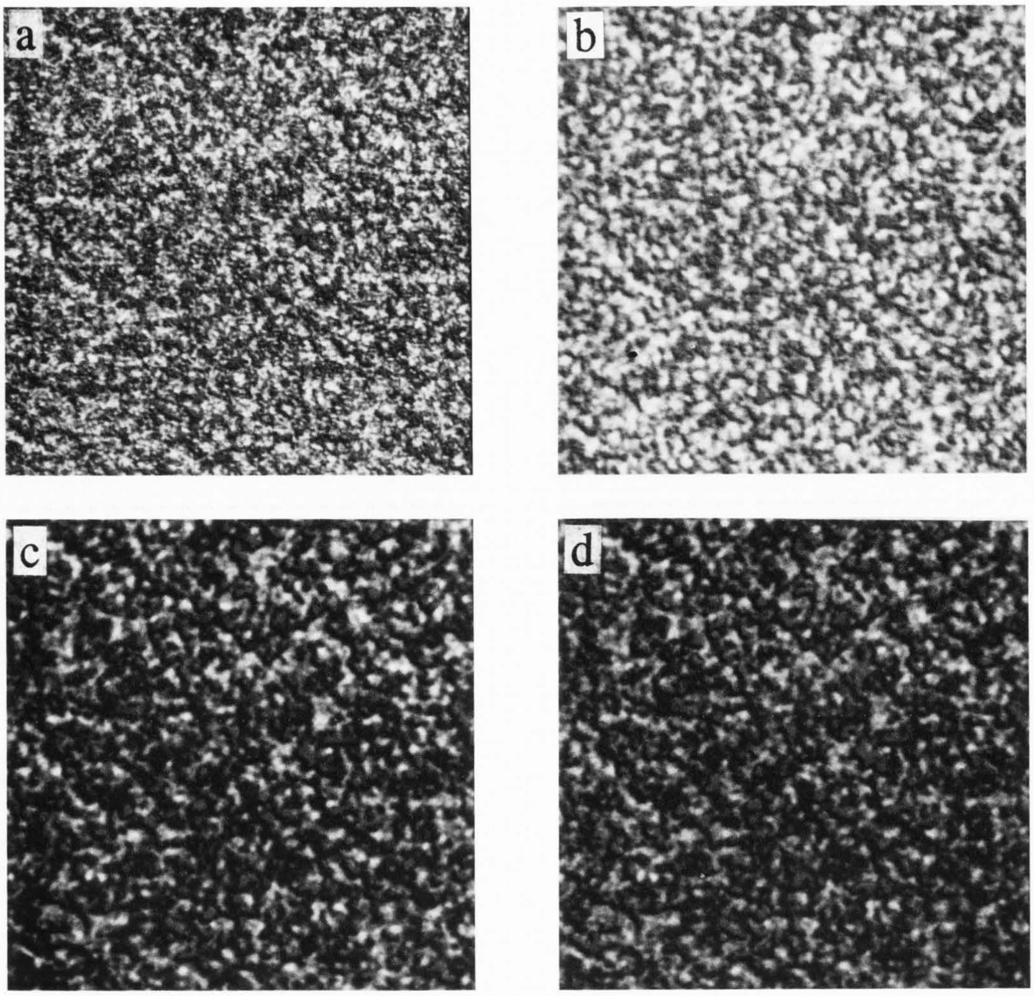


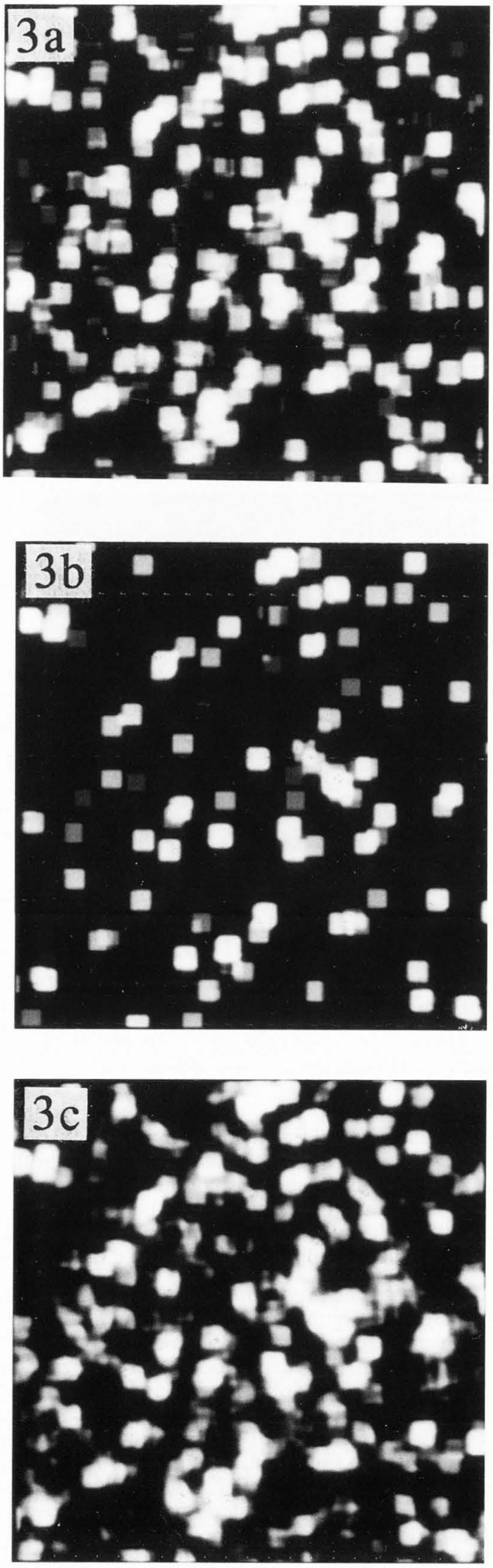
Figure 2. Comparison between linear kriging filters and disjunctive kriging filter on a simulation. **(a)** Simulated image; **(b)** Poisson noise from image in Fig. 2a (SNR = 0.66); **(c)** linear kriging filter of image in Fig. 2b (SNR = 2.820; **(d)** D.K. filter of image in Fig. 2b (SNR = 2.94).

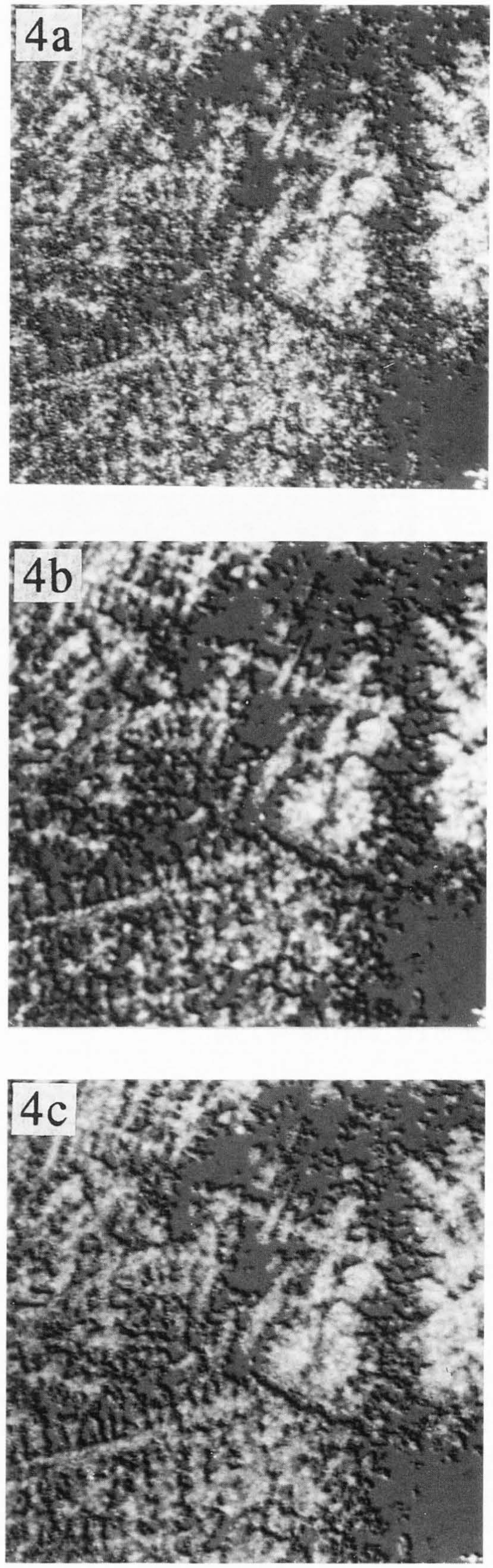
noise in the estimate is greatly reduced with respect to the simplest estimates which are obtained by post−processing of previously filtered images [5].

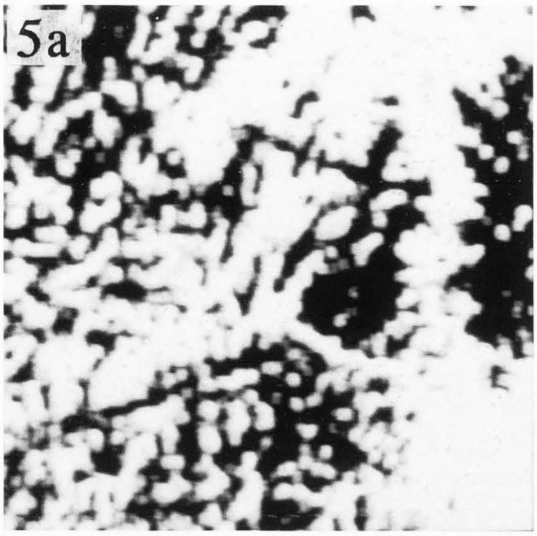
To illustrate these notions, we use both simulated images (where we can compare the estimation to the "true" histograms), and microprobe images.

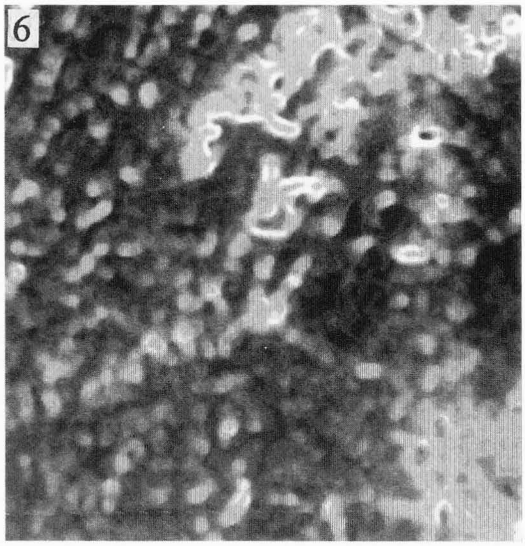
In Figure 2a, we show a simulated image, which we use as a reference image. It was spoiled by Poisson noise (Figure 2b), then filtered by linear kriging (7×7) in Figure 2c and disjunctive kriging in Figure 2d. For both filters, the SNR improvement is larger than 4, which gives a satisfactory result for the restoration of the image. In this case, there is no significant advantage using the D.K. In addition, we estimated the local probability distribution in a 11×11 neighbourhood. Figure 3 gives the probability of passing the value 250 : calculated as a frequency from the original image (Figure 3a), from the linear kriged image (Figure 3b), or estimated by disjunctive kriging (Figure 3c). We observe in this case, as in many other examples [5], that we give a biased estimation of the local probability using the linear filtered image, due to a smoothing effect of the filter. This example, and others, seem to validate the implementation of the D.K. to give a correct estimation of the local probability distribution function (p.d.f.).

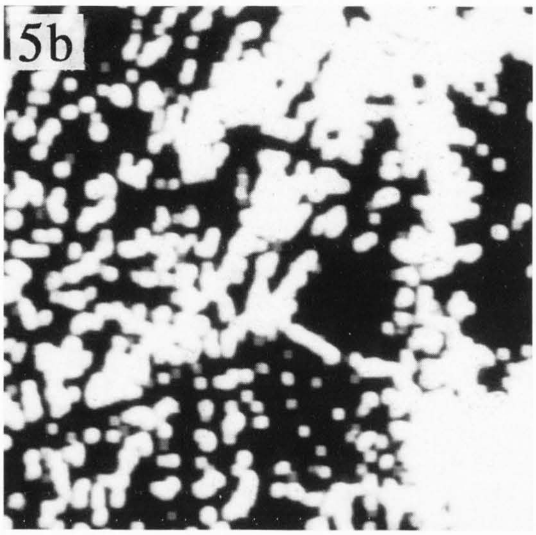
The same procedure was applied to noisy microprobe images. In Figure 4a, we see an example of an initial image which corresponds to the chemical concentration for Mn in the centre of the steel plate. The linear (Figure 4b) and D.K. (Figure 4c) filters provide similar results (the SNR ratio is improved by a factor close to 2.5). The local p.d.f. is estimated for the level 1070 in a 5×5 neighbourhood, and presented in Figure 5a, b, c, for the noisy image, the linear filtered image, and by D.K. The previously observed trend in the case of a simulated image seems to occur for this experimental situation : the D.K. p.d.f. image lies between the p.d.f. obtained on the noisy and the filtered images, as was expected.


3a
3b
3c


4a
4b
4c







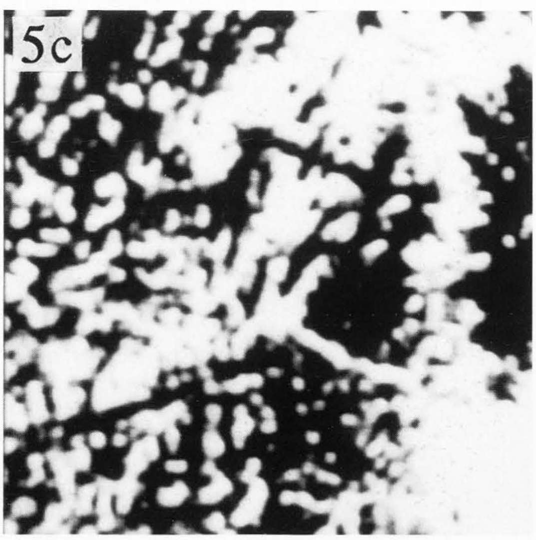


Figure 3. Local histogram estimation for the image of Figure 2: (**a**) local probability (P{Y(x) > 250}) calculated in a 11 x 11 neighbourhood from the original image (Fig. 2a); (**b**) local probability estimated from the linear filtered image (Fig. 2c); (**c**) local probability estimated by D.K.

Figure 4. Comparison between linear kriging filter and D.K. filter on a microprobe image: (**a**) noisy image (element Mn; 256 x 256 pixels, 1 pixel = 25 μm); SNR = 5.52; (**b**) filtered image (linear kriging) SNR = 11.82; (**c**) filtered image (D.K.).

Figure 5. Local histogram estimation for the images of Figure 4 (P{Y(x) > 1070}) in a 5 x 5 neighbourhood of x: (**a**) on the noisy image of Fig. 4a; (**b**) from the linear filtered image (Fig. 4b); (**c**) by D.K. estimation.

Figure 6. Estimation of a local segregation ratio map by D.K. {using the parameter TS(x)} from a noisy image with same characteristics as Figures 4 and 5.

In Figure 6, we see a map of the local segregation obtained by D.K., using the parameter TS(x) defined previously.

Conclusion

From both linear, and nonlinear (D.K.) kriging techniques, Geostatistics provides very efficient filters to improve noisy images, such as seen in chemical mappings obtained from the electron microprobe or electron microscopy.

In addition, the D.K. estimator provides unbiased estimates of nonlinear functions of the data, such as local histograms or local segregation ratio, which have very important repercussions in characterizing physical properties of materials from noisy microscopical data.

References

[1] Besag J. (1986). On the statistical analysis of dirty pictures. J. Statist. Soc. **B48**, 259--302.

[2] Daly C. (1991). Applications de la géostatistique à quelques problèmes de filtrage (Applications of geostatistics to some filtering problems). These de Docteur en Géostatistique, Paris School of Mines.

[3] Daly C. (1991). A geostatistical approach to inverse problem theory. Acta Stereologica **11** (Supplement 2), 285--290.

[4] Daly C, Lajaunie C, Jeulin D. (1989). Application of multivariate kriging to the processing of noisy images. In: Proc. 3rd Int. Geostatistics Congress, 1988. Geostatistics, Vol. 1. Armstrong M (ed.). Kluwer Academic Publ., Dordrecht, Netherlands. pp. 749-760.

[5] Daly C, Jeulin D, Benoit D, Auclair G. (1990). Application of multivariate geostatistics to macroprobe mappings in steels. ISIJ International **30**(7), 529-534.

[6] Jeulin D, Renard D. (1991). Practical limitations of the deconvolution of images by kriging. In Compte-Rendu des Journées de Géostatistique, 6-7 Juin 1991, Fontainebleau, Cahiers de Géostatistique, Fasc. 1, ENSMP, Paris School of Mines, Paris. pp. 173-203.

[7] Matheron G. (1970). The theory of regionalized variables and its applications. Cahiers du Centre de Morphologie Mathematique, Fasc. 5, Paris School of Mines, Paris. 212 p.

[8] Matheron G. (1976). A simple substitute for conditional expectation: the disjunctive kriging. In: Proc. NATO ASI, Advanced Geostatistics in the Mining Industry. Guarascio M, *et al.* (Eds.). D. Reidel Publ. Co., Dordrecht, Netherlands. pp. 221-236.

[9] Pinnamaneni BP, Jeulin D, Daly C, Mory C, Tence M, Colliex C. (1991). Multispectral linear filtering of high resolution EELS images by geostatistics. Microsc. Microanal. Microstruct. **2**, 107-127.

[10] Serra J. (1988). Image Analysis and Mathematical Morphology, Vol. 2. Academic Press, London. 411 p.

Discussion with Reviewers

N. Bonnet: *What is the size of the neighbourhood used for computing Fig. 1b ? Do the results largely depend upon this size ?*

Authors : We used a 15 × 15 window. The size of the window is selected as a function of the image, according to the following criteria : in principle, the final SNR increases with the number of data points, provided that their correlation to the unknown point is significant. So, the optimal size of the window is of the order of the range (or correlation length) of the covariance. In practice, experiments often show that when the size of the neighbourhood is increased beyond the range of the covariance, the weights λ_i, found as a solution of the kriging system, find a stable configuration, with negligible values for points located beyond the range. This behaviour illustrates the fact that, unlike conventional filters such as moving averages or median filters, the kriging filters are adaptive, and that they are quite robust with respect to the size of the neighbourhood.

N. Bonnet : *Could you expand your comment about maximum entropy algorithms, at the end of section 3 ? More generally, could you compare the underlying of your geostatistical methods to other attempts to restore noisy images ?*

P.J. Statham : *Much of the success of the D.K. approach would appear to depend on successful statistical modelling of the underlying signal. Can the approach still be used if there is no prior knowledge of expected signal variations ?*

Authors : A systematic and fair comparison of the geostatistical methods with the other filtering algorithms would require a large amount of work, using the same data set (eg. provided by simulations of the degradations of reference images). We can only mention here the main assumptions and prerequisites used in some cases. Maximum entropy algorithms cover in fact various approaches, at different levels : on the lowest level, each pixel of the image is modified in order for the restored image to recover a given histogram, which is assumed to be the underlying histogram. At this level, no account is made for the spatial variation of the data, since the same algorithm could be applied (and would provide the same results) after any permutation of the pixels. In addition, it was shown in [2, 3] how different the histogram of a noisy image is from the underlying histogram. The latter should be restored in a first step, if not provided from *a priori* knowledge.

On the highest level, maximum entropy is based on *a priori* assumptions on the multivariate distribution function involving all the pixels of the image. This assumption is so strong that it can produce impressive results, but it lacks robustness and may provide artefacts (emerging from this assumption) from a pure white noise. In other cases, maximum entropy provides a regularization function for regularization algorithms involved in the solution of ill-posed problems like deconvolution.

The geostatistical tools used in this paper are based on the second order statistics : the covariance of a stationary signal (or more generally the variogram of increments of order k for non stationary intrinsic random functions) are sufficient to implement the linear kriging.

The histogram of the image and the covariance of the data after making an anamorphis (to recover the histogram of an isofactorial model) are needed to use the D.K. algorithms. In the presented approach, we only assume that the underlying signal is a realization of a stationary random function, and we introduce assumptions about the formation of the noise from the signal (addition of a noise, or randomization of a Poisson process). These assumptions are based on the physics of the instrument (presently X-ray detection in a microprobe), and can be checked from experiments. We then make an appropriate choice of a model for the statistical tool (covariance, or choice of an isofactorial model), which is fitted to the experimental data (covariance, histogram) calculated from the image (or averaged over a population of similar images). Therefore most of the information used in the filtering technique comes from the images, which makes it adaptive, and not sensitive to some *a priori* and ill-defined knowledge.

N. Bonnet : *In your comments on Figures 2 and 3, you attribute the bias in the estimation of the local histogram by linear filtering to its smoothing effect. Do you mean that this smoothing effect is negligible when using D.K. ? More generally, could you give a qualitative explanation of the advantages of these nonlinear filters compared to linear filters and the way they work ?*

P.J. Statham : *How robust is the D.K. approach ? For example, if an inappropriate model were chosen for the underlying signal, would D.K. produce any greater bias than Wiener filtering or a simple median filter ?*

Authors : The source of the bias in the estimate of the local probability law using the truncation of the linear kriging (extension of the Wiener filter, or of any filter including the Median filter and the D.K. filter) is precisely due to the truncation : an unbiased estimate of a signal is not an unbiased estimate of binary images obtained by a threshold ! Indeed, the D.K. was introduced [8] with the aim to remove this bias and to directly estimate any nonlinear function ϕ. This sensitivity of the linear kriging to bias in the local histogram is illustrated in Figure 3.

As far as robustness is concerned, the less assumptions are made on the image, the more robust results are expected. As is recalled above, only second order statistics are used in our approach. The D.K. requires more assumptions than the linear kriging, and should be more sensitive to the choice of a model. However, it was seen in many experiments made from different simulations [2] that the D.K. filter provided a stable estimate of the local histogram, which advocates in favour of a robust behaviour.