Joint Porosity-permeability Stochastic Simulation and Spatial Median Regression by Nonparametric Copula Modeling

Arturo ERDELY and Martín DÍAZ-VIERA

Abstract

In petrophysics, assesment of formation permeability is a complex and challenging problem that plays a key role in reservoir forecasts and optimal reservoir management. In heterogeneous carbonate reservoirs, permeability evaluation is commonly performed using permeability-porosity relationships, which often seem to be nonlinear and complex. Copulas are marginal-free dependence functions that may capture such nonlinear relationships. In the present work we make use of a nonparametric copula approach for bivariate modeling of permeability- porosity real data, and its application for a spatial median regression.

1 Introduction

Assessment of formation permeability is a complex and challenging problem that plays a key role in oil reservoir forecasts and optimized reservoir management. Generally, permeability evaluation is performed using porosity-permeability relationships obtained by integrated analysis of various petro physical parameters from cores and well logs. In carbonate double-porosity formations with complex microstructure of pore space this problem becomes more difficult because the permeability usually does not depend on the total porosity, but on classes of porosity, such as vuggular and fracture porosity (secondary porosity). Even more, in such cases permeability is directly related to the connectivity structure of the pore system. This fact makes permeability prediction a challenging task.

Dependence relationships between pairs of petro physical random variables, such as permeability and porosity, are usually nonlinear and complex, and therefore those statistical tools that rely on assumptions of linearity and/or normality are not suitable in this case. The use of copulas for modelling petro physical dependencies is not new (DÍAZ-VIERA & CASAR-GONZÁLEZ 2005) and t-copulas have been used for this purpose. But expecting a single copula family to be able to model a any kind of bivariate dependency seems to be still too restrictive, at least for the petro physical variables under consideration in this work. Therefore, we adopted a nonparametric approach by the use of the Bernstein copula (SANCETTA & SATCHELL 2004, SANCETTA 2007).

Jekel, T., Car, A., Strobl, J. & Griesebner, G. (Eds.) (2012): GI_Forum 2012: Geovizualisation, Society and Learning. © Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach. ISBN 978-3-87907-521-8.

1.1 The problem of modelling the complex dependence pattern between porosity and permeability in carbonate formations

By far the most used permeability predictor is the porosity-permeability relationship (BALAN et al. 1995). It has long been assumed that most reservoir rocks show a reasonably linear relationship between these parameters in a semi-log scale, which allows for the estimation of permeability when a porosity profile is available. This normally requires a calibration data set that is represented by one or more key wells where comprehensive information is available in terms of core and log data. This calibration data set is used to build the predictor and to test the reliability of the results.

The regression approach, using statistical instead of deterministic formalism, tries to predict a conditional average, or expectation of permeability, corresponding to a given set of parameters. A different predictive equation must be established for each new area or new field. The main drawback of traditional regression methods is that the complex variability of data may not be effectively captured just in terms of variance or standard deviation (which may not even exist), and therefore the predicted permeability profile will be ineffective in reproducing extreme values.

What is important to note in this case is that the predicted permeability profile will be effective in estimating the average characteristics of the true profile, but will be ineffective in estimating the extreme values. These extreme values, from a fluid flow point of view, are the most important parts of the distribution, since they may represent either high permeability streaks or impermeable barriers.

Reservoir rocks show a wide spectrum of porosity-permeability relationships. In some formations, like for example homogeneous clastic rocks, these relationships show very low dispersion and can be reasonably used for prediction purposes. In other cases, as it is frequently for carbonates, this relationship is very loose and does not allow any safe regression, under traditional statistical tools.

On the other hand, model-free function estimators like artificial neural networks are very flexible tools for recognizing and reproducing the pattern of permeability distribution, but require a time consuming "learning" process which strongly depends on the amount and quality of available data.

A competitive and more systematic method for predicting permeability may be achieved by applying stochastic joint simulations, in which the correct specification of dependence pattern in the bivariate porosity-permeability distribution is crucial. This approach (DEUTSCH & COCKERHAM 1994) basically consisted on an annealing geostatistical porosity-permeability cosimulation using their empirical joint distribution. A modification of the previous methodology was proposed (DÍAZ-VIERA & CASAR-GONZÁLEZ 2005) where the basic idea was to apply a t-copula bivariate distribution instead of the empirical one, in which the permeability-porosity observed dependence pattern is specified through a rank correlation measure such as Kendall's tau or Spearman's rho.

1.2 Bivariate copula and random variables dependence

According to Sklar's Theorem (SKLAR 1959), the underlying bivariate copula associated to a bivariate random vector (*X, Y*) represents a functional link between the joint probability distribution **H** and the univariate marginal distributions *F* and *G,* respectively:

$$H(x, y) = C\,(F(x)\,,\,G(y)) \tag{1}$$

for all *x*, *y* in the extended real numbers system, where $C : [0,1]^2 \rightarrow [0,1]$ is unique whenever *X* and *Y* are continuous random variables. Therefore, all the information about the dependence between random variables is contained in their corresponding copula. Several properties may be derived for copulas (SCHWEIZER & SKLAR 1983, NELSEN 2006), and among them we have an immediate corollary from Sklar's Theorem: *X* and *Y* are independent continuous random variables if and only if their underlying copula is $\Pi(u, v) = uv$.

Let $S := \{(x_1, y_1), \ldots, (x_n, y_n)\}$ be observations of a random vector (*X*, *Y*). We may obtain empirical estimates for the marginal distributions of *X* and *Y* by means of

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{I}\{x_k \le x\}, \quad G_n(y) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{I}\{y_k \le y\}, \tag{2}$$

where I stands for an indicator function which takes a value equal to 1 whenever its argument is true, and 0 otherwise. It is well-known (BILLINGSLEY 1995) that the empirical distribution F_n is a consistent estimator of *F*, that is, $F_n(t)$ converges almost surely to $F(t)$ as $n \rightarrow \infty$, for all *t*. See Figure 1 (left, center) for the graphs of the empirical distribution functions of the data that will be used in the present work (KAZATCHENKO et al. 2006b).

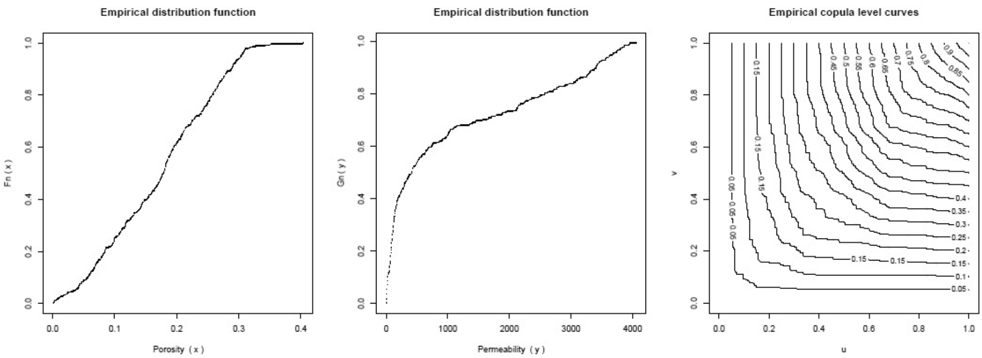

Fig. 1: Empirical marginal distributions and copula

Similarly, we have the empirical copula (DEHEUVELS 1979), a function C_n with domain $\left\{\frac{i}{n} : i = 0,1,\ldots,n\right\}^2$ defined as

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{I}\{rank(x_k) \le i,\, rank(y_k) \le j\} \tag{3}$$

and its convergence to the true copula C has also been proved (FERMANIAN et al. 2004). The empirical copula is not a copula, since it is only defined on a finite grid, not in the whole unit square $[0, 1]^2$, but by Sklar's Theorem (SKLAR 1959) C_n may be extended to a copula. See Figure 1 (right) for the graph of the level curves of the empirical copula corresponding to the data that will be used in the present work (KAZATCHENKO et al. 2006b).

2 Porosity-permeability Data Modelling

We propose a copula-based nonparametric approach to model the relationship between the permeability and porosity of the double porosity carbonate formations of a South Florida Aquifer in the western Hillsboro Basin of Palm Beach County, Florida.

The characterization of this aquifer for the borehole and field scales is given in PARRA et al. (2001) and PARRA & HACKERT (2002) and a hydrogeological situation is described by BENNETT et al. (2002). The interpretation of the borehole data and determination of the matrix and secondary porosity and secondary-pore types (shapes of spheroids approximating secondary pores) were presented by KAZATCHENKO et al. (2006b), where to determine the pore microstructure of aquifer carbonate formations the authors applied the petro physical inversion technique that consists in minimizing a cost function that includes the sum of weighted square differences between the experimentally measured and theoretically calculated logs (KAZATCHENKO et al. 2004).

In this case the following well logs were used for joint simultaneous inversion as input data: resistivity log, transit times of the P- and S-waves (acoustic log), total porosity (neutron log), and formation density (density log). To calculate the theoretical acoustic and resistivity logs the double-porosity model for describing carbonate formations was applied (KAZATCHENKO et al. 2006a).

This model treats carbonate rocks as a composite material that consists of a homogeneous isotropic matrix (solid skeleton and matrix pore system) where the secondary pores of different shapes are embedded. The secondary pores were approximated by spheroids with variable aspect ratios to represent different secondary porosity types: vugs (close-to-sphere shapes), quasi-vugs (oblate spheroids), channels (prolate spheroids), and micro fractures (flattened spheroids). For computing the effective properties the symmetrical self-consistent method of the effective medium approximation was used.

In this paper we used the results of inversion obtained by KAZATCHENKO et al. (2006b) for carbonate formations of South Florida Aquifer that includes the following petro physical characteristics: matrix porosity, secondary vuggy and crack porosities. It should be noted that the secondary-porosity system of this formation has complex microstructure and corresponds to a model with two types of pore shapes: cracks (flattered ellipsoids) with the overall porosity of 2% and vugs (close to sphere) with the porosity variations in the range of 10-30%. Such a secondary-porosity model can be interpreted as the interconnected by micro fractures and channels vuggy formation.

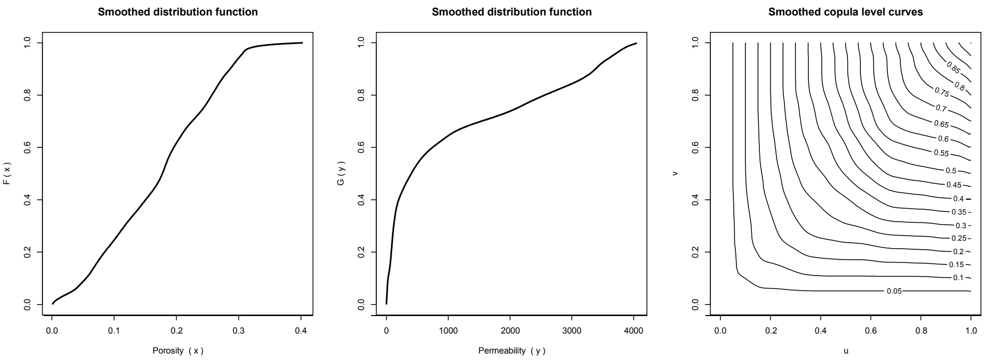

Fig. 2: Smoothed marginal distributions and copula

We model vuggy porosities as an absolutely continuous random variable X with unknown marginal distribution function F, and permeability as an absolutely continuous random variable Y with unknown marginal distribution function G. From KAZATCHENKO et al. 2006b we have bivariate observations from the random vector (X, Y). For simulation of continuous random variables, the use of the empirical distribution function estimates (2) is not appropriate since F_n is a step function, and therefore discontinuous, so a smoothing technique is needed. Since our main goal is simulation of porosity-permeability, it will be better to have a smooth estimation of the marginal quantile function $Q(u) = \mathrm{F}^{-1}(u) = \inf\{x : F(x) \geq u\}$, $0 \leq u \leq 1$, which is possible by means of Bernstein polynomials as in MUÑOZ-PÉREZ & FERNÁNDEZ-PALACÍN (1987):

$$\tilde{Q}_n(u) = \sum_{k=0}^{n} \frac{1}{2}(x_k + x_{k+1}) \binom{n}{k} u^k (1-u)^{n-k}, \tag{4}$$

and the analogous case for marginal G in terms of values y_k. For a smooth estimation of the underlying copula we make use of the Bernstein copula (SANCETTA & SATCHELL 2004, SANCETTA 2007):

$$\tilde{C}(u,v) = \sum_{i=0}^{n} \sum_{j=0}^{n} C_n\left(\frac{i}{n}, \frac{j}{n}\right) \binom{n}{i} u^i (1-u)^{n-i} \binom{n}{j} v^j (1-v)^{n-j} \tag{5}$$

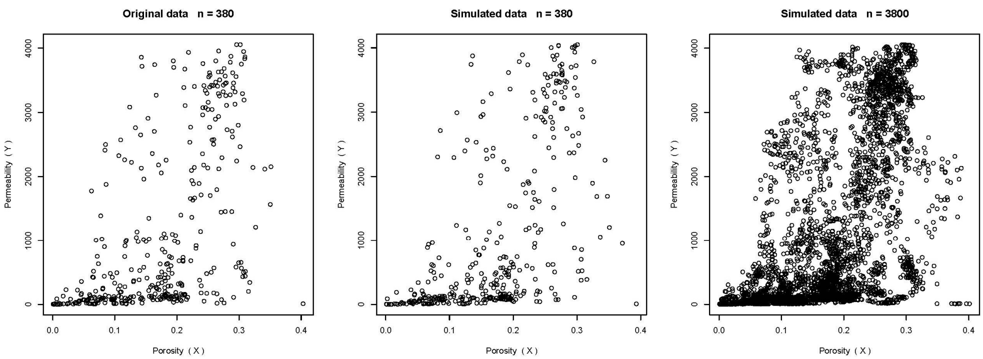

Fig. 3: Original data and simulations

for every (u, v) in the unit square $[0, 1]^2$, and where C_n is as defined in (3). By means of (4) and (5), in Figure 2 we present the smoothed counterparts of the empirical versions shown in Figure 1.

3 Simulation and nonparametric regression

With smoothed estimations of the underlying copula (dependence structure) and the marginal distribution function of the random variables involved, we have enough information to perform simulations and to make inferences.

In order to simulate replications from the random vector (X, Y) with the dependence structure inferred from the observed data $S := \{(x_1, y_1), \ldots, (x_n, y_n)\}$, accordingly to a result in NELSEN (2006) we have the following algorithm:

1. Generate two independent and continuous Uniform (0, 1) random variates u and t.
2. Set $v = c_u^{-1}(t)$ where

$$C_u(v) = \frac{\partial \tilde{C}(u,v)}{\partial u} \tag{6}$$

   and $\tilde{C}$ is obtained by (5).
3. The desired pair is $(x, y) = (\tilde{Q}_n(u), \tilde{R}_n(v))$, where $\tilde{Q}_n$ and $\tilde{R}_n$ are the smoothed estimated quantile functions of X and Y, respectively, accordingly to (4).

In Figure 3 we show: a scatterplot of porosity-permeability real data (sample size of n = 380) taken from KAZATCHENKO et al. (2006b) (left); simulation of $n = 380$ and $n = 3800$ porosity-permeability observations, accordingly with the dependence structure learned from the original data (center, right).

For a value x in the range of the random variable X and $0 < \alpha < 1$ let $y = \varphi_\alpha(x)$ denote a solution to the equation $P(Y \leq y \mid X = x) = \alpha$. Then the graph of $y = \varphi_\alpha(x)$ is the α-quantile regression curve of Y conditional on $X = x$. Recalling (6), it has been proved (NELSEN 2006) that

$$P(Y \leq y \mid X = x) = c_u(v)\big|_{u=F(x),\, v=G(x)}, \tag{7}$$

and this result leads to the following algorithm (NELSEN 2006) to obtain the α-quantile regression curve of Y conditional on $X = x$:

1. Set $c_u(v) = \alpha$.
2. Solve for the regression curve $v = g_\alpha(u)$.
3. Replace u by $\tilde{Q}_n(x)$ and v by $\tilde{R}_n(x)$.
4. Solve for the regression curve $y = \varphi_\alpha(x)$.

In Figure 4 we present: median regression curve (left), that is $\alpha = 0.50$ quantile regression curve; first and third quartile regression curves (center), that is $\alpha = 0.25$ and $\alpha = 0.75$ quantiles; first and ninth decile regression curves (right), that is $\alpha = 0.10$ and $\alpha = 0.90$ quantiles.

Finally, since the original data in KAZATCHENKO et al. (2006b) is presented in one-dimensional spatial form (depth of measurements), given the spatial data for porosity, see Figure 5 (down), we compare the original spatial data for log-permeability versus a spatial median regression based on porosity, Figure 5 (up).

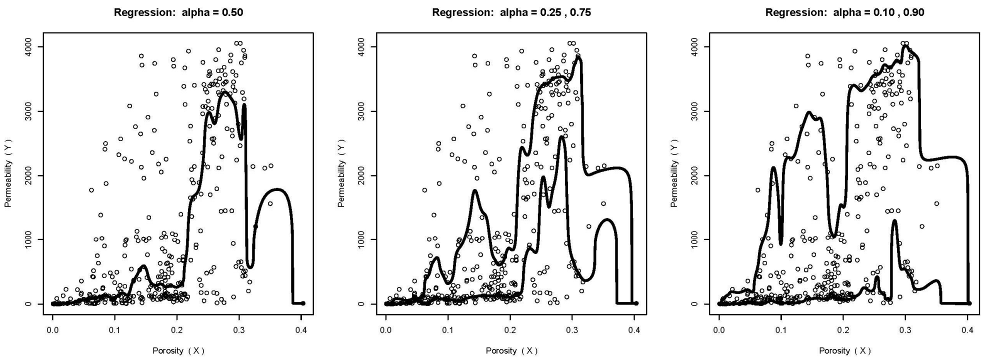

Fig. 4: Regression curves: $\alpha = 0.50$; 0.25 and 0.75; 0.10 and 0.90

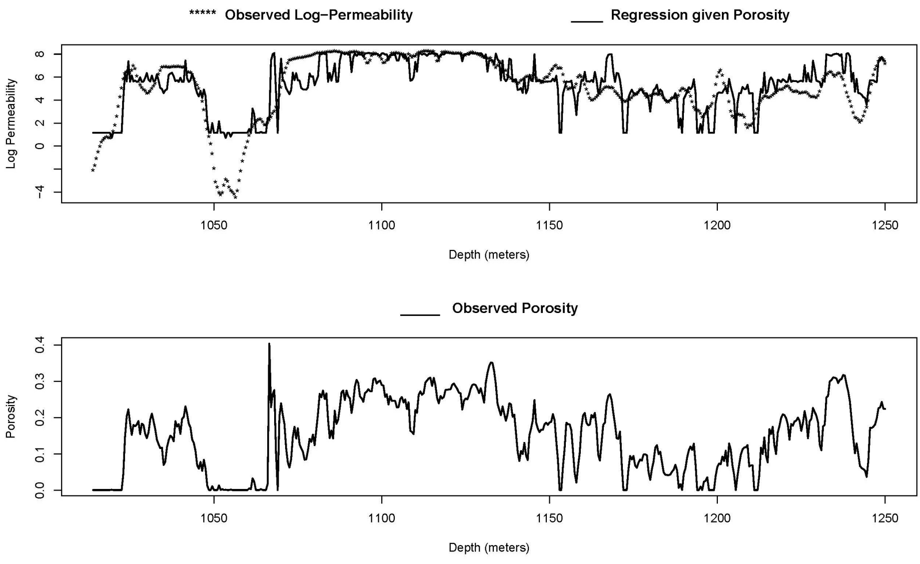

Fig. 5: Spatial regression of permeability given porosity

4 Conclusions

From a methodological point of view, this approach provides a very flexible statistical research tool to investigate the existing complex dependence relationships between pairs of petro physical properties such as porosity and permeability, without imposing strong assumptions of linearity or log-linearity, and/or normality when we are modelling them as random variables, not even the existence of first or second moments of the variables involved. The only assumption has been the random variables to be jointly absolutely continuous, and thereafter the data is allowed to speak by itself about the dependence structure.

The methodology used in this work has two main advantages: first, an easy way to simulate bivariate data with the dependence structure and marginal behaviour suggested by already observed data; second, a straightforward way to perform nonparametric quantile regression, which is useful in obtaining conditional point and interval estimates for Y given $X = x$, and without imposing or assuming functional relationships between variables.

All the information about the dependence structure is contained in the underlying copula, and its estimation is being used, instead of the extreme information reduction that is done by the use of numerical measures such as the linear correlation coefficient, which under the presence of nonlinear dependence may become useless and/or quite misleading (EMBRECHTS et al. 1999).

The nonparametric regression obtained is useful to confirm or to question prior ideas about relationships between variables, or even in proposing an appropriate model to explain such relations. The results obtained in this work were discussed with A. Mousatov (KAZATCHENKO et al. 2004, 2006a, 2006b) and the median regression (Figure 4, left) reflects what was expected: a very moderate increase in permeability while porosity increases below a level of 0.2 (percolation threshold), with an explosive increase in permeability when porosity is above such threshold, since at those levels the vug clusters finally become connected.

In relation with the geostatistical applications this method opens a promising line of research to model in a nonparametric fashion the intrinsic spatial dependence of random functions overcoming the restriction imposed by a linear coregionalization models. This methodology may be extended to more than two variables, but some intensive computing issues need to be efficiently solved.

Acknowledgement

The present work has been supported by project IN110311 from Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) at Universidad Nacional Autónoma de México.

References

BALAN, B., MOHAGHEGH, S. & AMERI, S. (1995), State-of-The-Art in Permeability Determination From Well Log Data: Part 1. A Comparative Study, Model Development, SPE 30978.

BENNETT, W. M., LINTON, P. F. & RECTENWALD, E. E. (2002), Hydrologic investigation of the Floridian aquifer system, western Hillsboro Basin, Palm Beach County, Florida, Technical Publication WS-8, South Florida Management District.

BILLINGSLEY, P. (1995), Probability and Measure, 3rd edition. Wiley, 593 p.

DEHEUVELS, P. (1979), La fonction de dépendance empirique et ses porpriétés, Acad. Roy. Belg. Bull. Cl. Sci., 65 (5), 274-292.

DEUTSCH, C. V. & COCKERHAM, P. W. (1994), Geostatistical Modeling of Permeability with Annealing Cosimulation (ACS), SPE 28413.

DÍAZ-VIERA, M. & CASAR-GONZÁLEZ, R. (2005), Stochastic simulation of complex dependency patterns of petrophysical properties using t-copulas, Proc. IAMG'05: GIS and Spatial Analysis 2, 749-755.

EMBRECHTS, P., MCNEIL, A. & STRAUMANN, D. (1999), Correlation: pitfalls and alternatives, Risk Magazine 5, 69-71.

FERMANIAN, J-D., RADULOVIĆ, D. & WEGCAMP, M. (2004), Weak convergence of empirical copula processes. Bernoulli, 10, 547-860.

KAZATCHENKO, E., MARKOV, M. & MOUSATOV, A. (2004), Joint inversion of acoustic and resistivity data for carbonate microstructure evaluation. Petrophysics, 45, 130-140.

KAZATCHENKO, E., MARKOV, M. & MOUSATOV, A. (2006a), Simulation of the acoustical velocities, electrical and thermal conductivities using unified pore structure model of double-porosity carbonate rocks. J. Appl. Geophys., 59, 16-35.

KAZATCHENKO, E., MARKOV, M., MOUSATOV, A. & PARRA, J. (2006b), Carbonate microstructure determination by inversion of acoustic and electrical data: application to a South Florida Aquifer. J. Appl. Geophys., 59, 1-15.

MUÑOZ-PÉREZ, J. & FERNÁNDEZ-PALACÍN, A. (1987), Estimating the quantile function by Bernstein polynomials. Comput. Statist. & Data Anal., 5, 391-397.

NELSEN, R. B. (2006), An Introduction to Copulas. 2nd edition. Springer, 269 p.

PARRA, J. O. & HACKERT, C. L. (2002), Permeability and porosity images based on crosswell reflection seismic measurements of a vuggy carbonate aquifer at the Hillsboro site, South Florida, 72nd Annual Meeting of the Society of Exploration Geophysicists, Paper VCD P1.2.

PARRA, J. O., HACKERT, C. L., COLLIER, H. A. & BENNETT, M. (2001), A methodology to integrate magnetic resonance and acoustic measurements for reservoir characterization, Report DOE/BC/ 15203-3. National Petroleum Technology Office, Department of Energy, Tulsa, Oklahoma.

SANCETTA, A. (2007), Nonparametric estimation of distributions with given marginals via Bernstein-Kantorovic polynomials: L1 and pointwise convergence theory. J. Multivariate Anal., 98, 1376-1390.

SANCETTA, A. & SATCHELL, S. (2004), The Bernstein copula and its applications to modeling and approximations of multivariate distributions. Econometric Theory, 20, 535-562.

SCHWEIZER, B. & SKLAR, A. (1983), Probabilistic Metric Spaces. North Holland, 313 p.

SKLAR, A. (1959), Fonctions de répartition àn dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris, 8, 229-231.