

# Chapter 5

## Properties of Sums of Geological Random Variables



G. M. Kaufman

“All models are wrong. Some are useful” George E. P. Box.

**Abstract** In the absence of empirical data that allows resolution of the vexing problem of how to address probabilistic dependencies among and between elements of large sets of geologic random variables data we need methods that refocus and streamline expert geological judgment inputs along with analytical methods for modeling dependencies that go beyond pairwise correlation and its cousins. Some possibilities are reviewed.

### 5.1 Introduction

Suppose that you are given the marginal distribution of each of a set of  $n$  random variables but no other information. What can be said about the behavior of their sum? This is an old problem, extensively studied by probability theorists and statisticians (Hoeffding 1940; Fréchet 1951). There is a rich probabilistic finance and actuarial risk analysis literature devoted to calculation of bounds on sums of random variables. This question motivates our review of state of the art methods designed to reduce geologists’ cognitive load when asked to assign judgmental probabilities to uncertain geologic variables.

In a wide range of settings geologists are asked to provide personal probability judgments about a collection of uncertain quantities and, in particular, about sums of them. Probabilistic assessments of oil and gas in unexplored petroleum plays and basins are recurring examples. In the absence of hard data they deal rather well with the cognitive task of providing personal judgments about marginal distributions of geologic attributes; i.e. their assessments are, in the large, reasonably well calibrated. Geologists’ personal judgments about dependencies among uncertain geologic quantities are more problematic.

---

G. M. Kaufman (✉)

Management Emeritus, E62-437, Sloan School of Management MIT,  
50 Memorial Drive, Cambridge, MA 92142, USA  
e-mail: gkaufman@mit.edu

© The Author(s) 2018

B. S. Daya Sagar et al. (eds.), *Handbook of Mathematical Geosciences*,  
[https://doi.org/10.1007/978-3-319-78999-6\\_5](https://doi.org/10.1007/978-3-319-78999-6_5)

105

It is worthwhile to distinguish micro-assessments—assessment of dependencies among individual reservoir attributes for example—from macro-assessments—assessment of dependencies among assessment units, each of which may be a collection of anomalies, reservoirs and fields. Measurable data bearing directly on probabilistic dependencies at the micro-assessment level is often available but precise measurable data bearing on dependencies among elements in a macro-assessment is seldom available. Chen et al. (2012) point out that

Although efforts have been made to address variable dependence in both methodology and tool development, the greatest emphasis and attention have been given to resource aggregation. Until now, the impact of interdependencies among variables in volumetric resource calculations has been mostly ignored, and the implementation of variable dependency remains a challenge to petroleum resource appraisal. In practice, inadequate data commonly exist to either specify a standard multivariate distribution with an appropriate correlation structure or to quantify the resource aggregation correlation matrices. However, variable correlations are so common among geologic variables that ignoring their interdependence may lead to serious bias, affecting both the resulting resource potential estimation.

Most geologists with some training and experience in probability assessment can provide reasonable responses to questions about marginal distributions of individual attributes of a target entity. Few if any are well equipped to provide sharp coherent judgments about possible dependencies among them. Some progress has been made in understanding how to elicit sensible, coherent judgements about second order co-variability of petroleum assessment units—the recent USGS study of CO<sub>2</sub> sequestration in depleted oil and gas reservoirs is an example. However, specification of marginal distributions along with second order moments is not sufficient for identification of a joint distribution of a set of uncertain quantities. This matters when interest centers on the right tail of a sum of magnitudes of petroleum in assessment units. Excepting special cases—joint lognormality for example—the right tail of a sum of jointly dependent uncertain quantities can, both in principle and in practice differ meaningfully from the right tail of an approximation based on marginal distributions and second moment properties alone. Lillestøl and Sinding-Larsen's (2017) study of giant field probabilities based on 182 North Sea discoveries highlights the importance of accurate modeling of tail probabilities. For economists, bureaucrats and politicians right tail probabilities are often the most interesting feature of a probabilistic oil and gas assessment. What, for example, is the probability of finding at least one more giant field in a given mature petroleum province? Objectives here are first, to outline how methods currently used by geologists to impute probabilistic dependencies among uncertain geologic quantities fit (or don't fit) into a conceptual framework developed by probabilists to answer the question posed at the outset and second, to review how the probability distribution of a sum of such quantities can be bounded given knowledge of marginal distributions alone assuming they are governed by a type of functional dependency called co-monotonicity. Co-monotonicity and cupolas are conceptual twins.

Section 5.2 lays out necessary theory and definitions and calls attention to co-monotonic upper bounds on sums of random variables and lower bounds expressed in terms of conditional expectations. Section 5.3 addresses geologic case studies in two of which geologists compute a probability distribution of a sum of random geologic magnitudes in three steps: first, specify marginal distributions of each magnitude, second, elicit judgmental appraisals of pairwise correlations among magnitudes and third, combine the two using Monte Carlo simulation to arrive at a distribution of the sum. This approach might be labelled “incomplete specification” (not to be confused with the econometric definitions of just-, over- and under-specification.). Iman and Conover’s (1982) ingenious method for imputing dependencies among a set of random variables requiring only pairwise correlations among elements of that set and marginal distributions is deployed in the  $CO_2$  sequestration study cited above (Sect. 5.3.2). Chen et al. (2012) use of cupolas to capture probabilistic dependencies in geologic micro-assessments is reviewed in Sect. 5.3.3. Brief concluding remarks appear in Sect. 5.4. Blondes et al. (2013a, b) offer a sensible rationale for careful attention to dependencies:

In the Circum-Arctic aggregation of the 48 AUs, the 90-percent uncertainty interval for recoverable gas is 1,471, 2,009, or 3,515 tcf for assumptions of independence, assessor specified dependency (correlation), or total dependence respectively. Clearly, decision makers who rely on assessment results need accurate interval projections. Too broad an interval provides little information; too narrow an interval gives a false sense of precision.

Spatial modeling provides important insights into the structure of probabilistic dependencies among petroleum play attributes and deserves careful attention in parallel with methods and models discussed here. It is a topic for another day.

## 5.2 Preliminaries

Define  $F_X$  to be the distribution function of a random vector  $\mathbf{X} = (X_1, \dots, X_n)^t$  with domain  $\mathbf{R}^n$  and marginal distributions  $F_i, i = 1, \dots, n$ . Set  $F_X(\mathbf{x}) = \text{Prob}\{X_1 \leq x_1, \dots, X_n \leq x_n\}$ . Assume that each  $F_i$  is continuous and possesses a one to one inverse. Define the  $p$ th fractile of  $X_i$  as the value in the domain of  $X_i$  such that  $\text{Prob}\{X_i \leq x_p\} = p$  and its inverse as  $F_i^{-1}(p) = x_i(p)$ . In turn the  $p$ th fractile of the sum  $S_n = X_1 + \dots + X_n$  is  $s_p$  such that  $\text{Prob}\{S_n \leq s_p\} = p$  or  $F_{S_n}^{-1}(p) = s_p$ .

What conditions guarantee that fractiles are strictly additive? That is that for all  $p \in (0, 1) s_p = x_1(p) + \dots + x_n(p)$ ? Imposition of functional dependencies among  $X_1, \dots, X_n$  is one route to sufficient conditions for this to be true. To divide difficulties suppose that  $X_1, \dots, X_n$  share a common domain  $D_X$  and consider  $n$  continuous invertible functions  $h_i$ , each with domain  $D_X$ . Suppose that  $x_i = h_i(x_1)$  for all  $x_i \in D_X, i = 2, \dots, n$ . Then  $\text{Prob}\{S_n < s\} = \text{Prob}\{X_1 + h_2(X_1) + \dots + h_n(X_1) < s\}$ . The omnibus function  $g(x_1) = x_1 + h_2(x_1) + \dots + h_n(x_1), x_1 \in D_X$  is continuous and invertible so  $\text{Prob}\{g(X_1) < s\} = \text{Prob}\{X_1 < g^{-1}(s)\}$ . The  $p$ th fractile of  $S_n$  is  $s_p$  such that  $\text{Prob}\{g(X_1) < s_p\} = p$  or  $\text{Prob}\{X_1 < g^{-1}(s_p)\} = p$  leading to  $x_1(p) = g^{-1}(s_p)$ .

Equivalently  $g(x_1(p)) = s_p$ . Functional dependencies of this type are too strong to survive the rigors of modeling most real world data. In the absence of complete knowledge of a joint distribution co-monotonicity is a more flexible approach to modeling joint behavior of dependent random variables.

**Definition** The random vector  $\mathbf{X} = (X_1, \dots, X_n)^t$  is co-monotonic if and only if  $(X_1, \dots, X_n) =_d (F_1^{-1}(U), \dots, F_n^{-1}(U))$ ,  $U$  a uniform random variable with domain  $(0, 1)$ .

Here  $=_d$  means agreement in distribution. Intuitively each element of a co-monotonic random vector is a functional of a single random variable  $U$  so all elements of  $\mathbf{X}$  exhibit strong positive dependency. McNeil et al. (2005) provide a more general definition:  $\mathbf{X}$  is co-monotonic if and only if it agrees in distribution with a random vector, each of whose components is a non-decreasing function of a single random variable. If elements of  $\mathbf{X}$  are co-monotonic increasing one element of  $\mathbf{X}$  increases all others. Goovaerts et al. (2000) provide a clear readable account of properties of sums of co-monotonic random variables in an actuarial context. Deelstra et al. (2009) offer a literature review of co-monotonicity in financial economics.

Foreshadowing a possible critique by geologists that in their setting, some elements of  $\mathbf{X}$  may be independent or possibly negatively dependent (rather rare), co-monotonicity and its consequences provide upper and lower bounds on a sum of random variables with specified marginal distributions that embrace a wide range of dependence structures. When these bounds are judged to be tight enough, reasonable projections of probability distributions of aggregates can be made using marginal distributions along with specification of certain conditional expectations. (See 5.1, 5.5). They provide useful information about projections made based on information elicited from geologists about dependencies and police reasonableness of geologic probabilistic projections of uncertain geologic resources made using other methods.

## 5.2.1 Bounds

A random variable  $X$  precedes a random variable  $Y$  in convex order, denoted by  $X \geq_{cx} Y$  if and only if  $E(g(X)) \geq E(g(Y))$  for all real convex functions  $g$  for which expectations are finite. Kaas et al. (2009) use convex order to show that fractiles of co-monotonic random variables can be added in the following sense: for any random vector  $\mathbf{X} = (X_1, \dots, X_n)$  possessing marginal cumulative distribution functions  $F_1, \dots, F_n$  and  $U$  a uniform  $(0, 1)$  random variable

$$(X_1 + \dots + X_n) \leq_{cx} S_u \equiv F_1^{-1}(U) + \dots + F_n^{-1}(U). \quad (5.1)$$

If  $S_u = {}_dF_1^{-1}(U) + \dots + F_n^{-1}(U)$  it follows immediately that the  $p$ th fractile of  $S_u$  is  $F_{S_u}^{-1}(p) = F_1^{-1}(p) + \dots + F_n^{-1}(p)$ , for all  $p \in (0, 1)$ . They point out that (5.1) is a supremum in terms of convex order and is a best bound for marginal distributions in a Fréchet space. It is well known that if a random vector  $\mathbf{X}$  with marginal distributions  $F_1, \dots, F_n$  belong to a Fréchet space  $F_n$  the joint cumulative distribution function  $Prob\{X_1 \leq x_1, \dots, X_n \leq x_n\}$  of  $\mathbf{X}$  is bounded from above by  $M_n \equiv \min\{F_1(x_1), \dots, F_n(x_n)\}$ . Goovarts et al. note that  $M_n$  is reachable in  $F_n$ .

For sums of elements of  $\mathbf{X}$  introduction of a random variable  $Z$  such that distribution functions of each  $X_i$  given  $Z$  are known with certainty leads to refined upper and lower bounds. In a geologic context  $Z$  is interpretable as a latent (background) variable describing gross geologic characteristics of, for example, a petroleum assessment unit. The conditioning variable  $Z$  might be regression dependent on geologic attributes of an assessment unit and need not be scalar. These authors define  $F_{X_i|Z}^{-1}(U)$  to be a random variable  $f_i(U, Z)$  that for  $(U, Z) = (u, z)$  assumes value  $F_{X_i|z}^{-1}(u)$  and prove that for  $U$  uniform  $(0, 1)$  and  $Z$  independent of  $U$

$$(X_1 + \dots + X_n) \leq_{cx} S_u^* \equiv F_{X_1|Z}^{-1}(U) + \dots + F_{X_n|Z}^{-1}(U). \tag{5.2}$$

Jensen’s inequality leads to a lower bound

$$E(X_1|Z) + \dots + E(X_n|Z) \leq_{cx} (X_1 + \dots + X_n). \tag{5.3}$$

Kaas et al. (2009) point out that (a) the random vector  $E(X_1|Z) + \dots + E(X_n|Z)$  will not in general have marginal distributions  $F_1, \dots, F_n$  (b) If  $E(X_1|Z), \dots, E(X_n|Z)$  are either jointly non-increasing or non-decreasing functions of  $Z$  the LHS in (5.3) is a sum of co-monotonous random variables and (c)  $Var(E(X_i|Z)) < Var(X_i)$  unless  $Var(E(X_i|Z)) = 0$ . In order to create a path to direct computation of the cdf of the LHS of (5.4) suppose that (b) obtains and that each of the random variables  $E(X_1|Z), \dots, E(X_n|Z)$  are non-decreasing functions of increasing  $Z = z$ . Write the lower bound as  $E(X_1|Z) + \dots + E(X_n|Z) = E(S|Z)$  and define  $F_{E(X_i|Z)}(x) = Prob\{E(X_i|Z) \leq x\}$ . They show that, provided that the cdf of  $E(X_i|Z)$  is continuous and increasing

$$F_{E(X_1|Z)}^{-1}(F_{E(S|Z)}(x)) + \dots + F_{E(X_n|Z)}^{-1}(F_{E(S|Z)}(x)) = x, \tag{5.4}$$

a prescription for calculating a lower bound. The quality of the lower bound (5.3) depends of course on the choice of a model for  $Z$ . Kaas et al. (2002) and Goovarts et al. (2000) demonstrate that upper and lower bounds (5.1) and (5.3) provide reasonable bounds on the cumulative distribution function of certain sums of discounted cash flows as well as for the cumulative distribution function of sums of dependent lognormal random variables. Lux and Papantoleon (2017) show that upper and lower Fréchet–Hoeffding bounds such as those described above can be tightened. They demonstrate that other types of information, knowledge of

functionals of lower dimensional marginals of an  $n$ -dimensional cupola for example, also lead to improvements. The tradeoff is that the improved bounds are quasi-cupolas but not cupolas.

Comparison of predictive distributions of undiscovered mineral resources derived by conventional methods currently in use with co-monotonic bounds on them is a promising avenue of research.

### 5.3 Thumbnail Case Studies

Thumbnail sketches of three case studies serve as a template for discussion of probabilistic dependence issues discussed above: examples of the USGS approach to probabilistic dependencies among oil and gas assessment units, the USGS probabilistic assessment of CO<sub>2</sub> sequestration in mature oil and gas reservoirs in the United States and a Canadian Geological Survey study of use of cupolas to capture probabilistic dependencies among accumulations in individual oil and gas plays.

#### 5.3.1 *USGS Oil and Gas Resource Projections*

The USGS developed an assessment system in the 1980s with the acronym FASP (fast appraisal system for petroleum resources). FASP incorporated perfect positive correlation between micro-level reservoir attributes but allowed specification of any positive correlation in the course of aggregating play resources. However, the USGS 2000 World Petroleum Assessment aggregates undiscovered resource volumes from assessment unit level to regional level using perfect correlation as the argument for adding assessment unit fractiles to arrive at regional level aggregates. Recognizing that at the global level dependencies among large regional aggregates of resources are unlikely to be perfectly correlated they adopt pairwise correlation of 0.5 between pairs of eight regions (Klett et al. 2000). No sensitivity analysis of how aggregate projections vary with these particular choices is provided.

Many USGS assessment studies present tables of fractiles of individual assessment units and then add them to arrive at a fractile assessment of total resources. Addition is qualified by the statement that “Fractiles are additive under assumption of perfect positive correlation” allowing avoidance of direct assessment of dependencies among units. Table 2 in “Assessment of Undiscovered Continuous Oil and Gas Resources in the Monterey Formation, San Joaquin Basin Province, California” USGS Fact Sheet 2015-3058 September 2015 and Table 2 in USGS Fact Sheet 2014-3082 “Assessment of Potential Shale-Oil and Shale-Gas Resources in Silurian shales of Jordan” September 2014 are examples. Chen et al. (2012) cite additional examples (Klett et al. 2000, 2005; Klett 2004). It is easy to show that “perfect correlation” is not robust to variations in specification of the functional form of marginal distributions elicited from geologists. Worse, addition of fractiles

without careful attention to properties of the joint distribution of a set of uncertain quantities can lead to incoherence. On the other hand mutual independence allows specification of arbitrary marginal probability distributions without doing violence to coherence but often leads to an unacceptably narrow probability projection of sums of oil and gas magnitudes.

A salient feature of Pearson's correlation coefficient is that random variables  $X$  and  $Y$  possess correlation 1.0 or  $-1.0$  only if  $X$  and  $Y$  are linearly dependent. As Denuit and Dehaene (2003) point out, a limiting case is a bivariate normal pair of random variables for which the variance of one member of the pair is zero. If  $X$  and  $Y$  are jointly lognormal and  $\log X$  is a linear function of  $\log Y$  the Pearson correlation of  $\log X$  and  $\log Y$  is either 1.0 or  $-1.0$ . However, the Pearson correlation of  $X$  and  $Y$  is then less than 1.0. Denuit and Dehaene provide a more nuanced treatment. Suppose  $F_1$  and  $F_2$  are marginal cumulative distribution functions of  $X$  and  $Y$  respectively, each concentrated on  $(0, \infty)$  and  $U$  is a uniform random variable independent of  $X$  and  $Y$ . Using super-modularity these authors prove that if  $F_1$  and  $F_2$  lie in a Fréchet space the Pearson correlation coefficient  $r(X, Y)$  of  $X$  and  $Y$  is bounded by

$$\frac{\text{Cov}(F_1^{-1}(U), F_2^{-1}(1-U))}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} \leq r(X, Y) \leq \frac{\text{Cov}(F_1^{-1}(U), F_2^{-1}(U))}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}. \quad (5.5)$$

In this setting perfect correlation is not achievable. They also prove that it is possible for a pair of co-monotonic lognormal random variables to have pairwise correlation close to zero, contradicting the intuitive notion that small correlation implies weak dependence. Denuit and Dehaene call attention to Shih and Huang (1992) and Schechtman and Yitzhaki's (1999) observation that, for any two random variables, the achievable range of Pearson's correlation coefficient is  $(-1, 1)$  only if the functional form of the two marginal distributions differ solely in values of location and/or scale parameters. If not, the range of Pearson's  $r$  is narrower than  $(-1, 1)$  and depends on the shape of the two marginal distributions.

These authors document several important features of Kendall's  $\tau$  and Spearman's  $\rho$ . (Spearman's  $\rho$  is at the center of the Iman and Conover method deployed in the USGS (2013) study of  $CO_2$  sequestration to compute predictive probability distributions of aggregates). First, both are invariant with respect to strictly monotone transformations. Second, when one variable is a non-decreasing (non-increasing) transformation of the other they equal 1 (or  $-1$ ) at the Fréchet upper (resp. lower) bound. They note that at a value of 1.0 or  $-1.0$  Kendall's  $\tau$  and Spearman's  $\rho$  achieve Fréchet bounds. According to them Kendall's  $\tau$  and Spearman's  $\rho$  are more desirable measures of association for non-normal multivariate distributions than Pearson's  $r$  because the latter does not share Kendall and Spearman's correlation invariance properties. These invariance properties come into play in Iman and Conover's method discussed below. Denuit and Dehaene prove the non-obvious fact that if positively or negatively quadrant dependent random couples are jointly uncorrelated they are mutually independent.

All of this emphasizes that “perfect correlation” as an omnibus argument for adding fractiles has many pitfalls. Co-monotonic bounds on random sums are a conceptually satisfactory alternative that deserves much future study.

### 5.3.2 *USGS Probabilistic Assessment of CO<sub>2</sub> Storage Capacity*

A recent USGS probabilistic assessment of CO<sub>2</sub> sequestration in mature petroleum reservoirs (Blondes et al. 2013a, b) is based on both micro- and macro-assessments by geologists. Their macro-assessment aggregates storage assessment units (SAUs) at basin, regional and national levels. An objective was to provide probabilistic assessments that take into account dependencies among assessment units arising from “overlap of geologic analogs, assessment methods and assessors” using individual SAU marginal probability distributions and “...a correlation matrix obtained by expert elicitation describing interdependencies between pairs of SAUs”. The correlation matrix dimension is 192×192. Because a menagerie of marginal distributions—Beta-PERT, lognormal, truncated lognormal—were deployed at the micro-level use of standard multivariate distribution theory is not appropriate. Dependencies among storage capacity magnitudes are induced using an innovative distribution free method developed by Iman and Conover (1982) that allows marginal distribution shapes to be estimated from data sets distinct from data sets used to estimate dependency structure. Their method is designed to provide rank correlations that match assessed correlations and to translate the match into a predictive probability distributions for individual assessment units and larger aggregates. (See Blondes et al. 2013a for informative examples).

How to aggregate from basin, to region and then to a national scale is an issue. Should this be done in a single stage using the correlation matrix for all SAUs in the study or successively aggregate subsets of SAUs in multiple stages? Blondes et al. (2013b) conclude that

Although the single-stage approach requires determination of significantly more correlation coefficients, it captures geologic dependencies among similar units in different basins and it is less sensitive to fluctuations in low correlation coefficients than the multiple stage approach. Thus, subsets of one single-stage correlation matrix are used to aggregate to basin, regional, and national scales.

Successive aggregation in multiple stages drastically reduces the number of pairwise correlations that must be elicited from geologists at the expense of requiring each assessor to appraise pairwise correlations of sums of assessment unit magnitudes. Although there are no studies comparing how well geologists’ assessments calibrate when asked to appraise dependencies among sums of SAU magnitudes relative to appraisal of dependencies among individual SAUs it is reasonable to conjecture that individual SAU appraisals are much more likely to be well calibrated. Properties of single and multi-stage appraisal methods are studied in Kaufman et al. (2018).



### 5.3.3 *Copulas and Oil and Gas Resource Assessment*

Chen et al. (2012) emphasize that at an assessment micro-level, reservoir attributes such as porosity, permeability, pressure and temperature are often decisively dependent and that empirical data suggest dependencies are present among more aggregate assessment units in mature provinces—among fields in a mature play or basin for example. Their argument is that a basin’s tectonic framework exerts “strong geographic control” over many geological features and leads to geographic and spatial dependencies and that because plays in a given basin share “...petroleum system elements, such as source rocks, regional top seal, migration fairways, timing, regional tectonics for trap formation, and accumulation preservation factors” a probabilistic model of pools or fields in a play in a given basin should incorporate probabilistic dependencies among these attributes as well as between plays. They are the first to use copulas in this setting.

Sklar (1959) proved that, subject to mild restrictions a multivariate cumulative distribution can be mapped into a joint cumulative distribution of uniform random variables called a copula. As with Iman and Conover’s method, adoption of a copula model allows marginal distribution shapes to be estimated from data sets distinct from those used to estimate dependency structure.

Suppose as in Sect. 5.2 above that  $F_X$  is the distribution function of a random vector  $\mathbf{X} = (X_1, \dots, X_n)^t$  with domain  $\mathbf{R}^n$  and marginal cumulative distributions  $F_i, i = 1, \dots, n$ . Let  $\mathbf{U}_n = (U_1, \dots, U_n)$  be a vector of independent uniform  $(0, 1)$  random variables and  $\mathbf{u}_n = (u_1, \dots, u_n)$  be a realization of  $\mathbf{U}_n$ . Then with  $u_i = F_i(x_i), i = 1, \dots, n$   $Prob\{X_1 \leq x_1, \dots, X_n \leq x_n\} = Prob\{U_1 \leq u_1, \dots, U_n \leq u_n\}$ .

**Definition**  $C(u_1, \dots, u_n) = Prob\{U_1 \leq u_1, \dots, U_n \leq u_n\}$  is the copula of  $F_X$ .

Set  $dF_i = f_i, i = 1, \dots, n$  and  $dC(u_1, \dots, u_n) = c(u_1, \dots, u_n) du_1 \dots du_n$ . The joint density of  $\mathbf{X}$  can be written as  $c(u_1, \dots, u_n) \times f_1(x_1) \times \dots \times f_n(x_n)$ . The term  $c$  in the joint density captures the dependency structure of elements of  $\mathbf{X}$ . Because  $Prob\{X_1 \leq x_1, \dots, X_n \leq x_n\} = Prob\{U_1 \leq u_1, \dots, U_n \leq u_n\}$  a procedure for generating samples from  $C$  produces samples of  $\mathbf{X}$  by inversion of  $u_i = F_i(x_i), i = 1, \dots, n$ .

Computation requires choice of a copula functional form. Among a variety of choices Chen et al. chose the bivariate normal copula, a popular choice closely tied to standard multivariate normal distribution theory.

Their regional resource assessment of the Canadian Arctic’s Beaufort-McKenzie Basin is based on analysis of 48 “significant” oil and gas discoveries containing 53 distinct accumulations. Empirical data is sufficiently detailed to allow study and estimation of pairwise correlations among reservoir attributes—area, porosity, oil saturation, net pay—for plays in the three major petroleum systems. The authors treat geologic risk factors as probabilistically independent because the data is not sufficient to allow empirical estimation of them and restrict their study of dependencies to reservoir volume attributes within each play and through them to the impact of probabilistic dependencies on the distribution of total resource volumes.

Four plays, Ivik, Taglu, Kugmallit (East) and Kugmallit (West) are used to illustrate how to incorporate dependencies among individual play resources. Although no systematic method for eliciting geologists' judgments about between play dependencies are discussed the authors motivate their choice of a rather large correlations between plays (0.6) and perfect correlation (1.0) by noting that all four plays share the same source rock and petroleum system: "The resource richness of each play is basically a function of both the oil charge and the preservation of accumulations that are mostly controlled by common petroleum system elements... we infer that the resources in the four plays are highly correlated, although the pool size distributions among the four plays vary considerably." Pairwise correlations between area, net pay, porosity and oil saturation vary from a low of 0.20 to a high of 0.86. The authors call attention to the substantial difference between total ultimate oil resource medians under the assumption of independence and under the assumption of within and between play correlations: the latter is 1.6 times the former.

Principal messages are that to be realistic, probabilistic appraisal of oil and gas resources in unexplored and partially explored regions must account for multiple sources of dependencies and that copulas are useful for doing so.

## 5.4 Concluding Remarks

In the absence of empirical data that allows resolution of the vexing problem of how to address probabilistic dependencies among and between elements of large sets of geologic random variables we need methods that refocus and streamline expert geological judgment inputs as well as analytical methods for modeling dependencies that go beyond pairwise correlation and its cousins. One promising avenue is the theory of vines proposed by Bradford and (2002). Their theory broadens the range of allowable dependency structures beyond Bayesian belief networks and exploits properties of rank correlations in a fashion that leads to efficient computation.

## References

- Blondes MS, Schuenemeyer JH, Olea RA, Drew LJ (2013a) Aggregation of carbon dioxide sequestration storage assessment units. *Stoch Environ Res Risk Assess* 27:1839–1859
- Blondes MS, Scheunemeyer JH, Drew LJ, Warwick PD (2013b) Probabilistic aggregation of individual assessment units in the U.S. Geological Survey national CO<sub>2</sub> sequestration Assessment. *Energy Procedia* 37:5110–5117
- Chen Z, Osadetz KG, Dixon J, Deitrich J (2012) Using copulas to implement variable dependencies in petroleum resource assessment: example from Beaufort-Mackenzie Basin. *Can AAPG Bull* 96(30):439–457

- Deelstra G, Jan Dhaene J, Vanmaelez M (2009) An overview of comonotonicity and its applications in finance and insurance, working paper
- Denuit and Dehaene (2003) Simple characterizations of co-monotonicity and counter-monotonicity by extremal correlations, Univ of Louvain working paper
- Fréchet M (1951) Sur les tableaux de corrélation dont les marges sont donnés. *Ann Univ Lyon Sect A, Ser 3*, 14:53–77
- Goovaerts et al (2000) Research report 9914 U. of Leuven
- Hoeffding W (1940) Masstabinvariante Korrelationstheorie. In: *Schriften des mathematischen Instituts und des Instituts für angewandte Mathematik des I.hii-versitat Berlin 5*, pp 179–233
- Iman RL, Conover WJ (1982) A distribution-free approach to inducing rank correlation among input variables. *Commun Stat—Simul Comput 11(3)*:311–334
- Kaufman GM, Olea RA, Blondes MS, Faith R (2018) Probabilistic aggregation of geologic resources. Forthcoming in *Math Geosci*
- Klett TR (2004) Future petroleum supply: exploration or development? 2nd U.S. Geological Survey Conference on Reserve Growth Denver Federal Center, Lakewood, Colorado, May 6, 2004
- Klett TR, Charpentier RR, Schmoker JW (2000) Assessment operational procedures: U.S. Geological Survey Digital Data Series 60, <http://energy.cr.usgs.gov/WEcont/chaps/OP.pdf>. Accessed 1 Aug 2011
- Klett TR, Gautier DL, Ahlbrandt TS (2005) An evaluation of the U.S. Geological survey world petroleum assessment 2000. *AAPG Bulletin 89(8)*:1033–1042
- Lillestøl J, Sinding-Larsen R (2017) Creaming and the likelihood of discovering additional giant petroleum fields. *Math Geosci 49(1)*:67–83
- Lux T, Papapantoleon A (2017) Improved Fréchet-Hoeffding Bounds on dcopulas and applications in model-free finance. *Ann Appl Probab 27(6)*:3633–3671
- McNeil AJ, Frey R, Embrechts P (2005) *Quantitative risk management. concepts, techniques and tools*. Princeton series in finance. Princeton University Press, Princeton, NJ. ISBN 0-691-12255-5, MR 2175089, Zbl 1089.91037
- Schechtman E, Yitzhaki S (1999) On the proper bounds of the Gini correlation. *Econ Lett 63(2)*:133–138
- Sklar M (1959) Fonctions de répartition à n dimensions et leur marges. *Publ Inst Statist Univ Paris 8*:229–231
- Shih WJ, Huang WM (1992) Evaluating correlation with proper bounds. *Biometrics*.1207–1213

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

