

RESEARCH ARTICLE

OPEN ACCESS

# Analysis of approximations of GUM supplement 2 based non-Gaussian PDFs of measurement models with Rosenblatt Gaussian transformation mappings

Vishal Ramnath\*

Department of Mechanical and Industrial Engineering, University of South Africa, Private Bag X6, Florida 1710, South Africa

Received: 12 November 2019 / Accepted: 10 December 2019

**Abstract.** In scientific metrology practise the application of Monte Carlo simulations with the aid of the GUM Supplement 2 (GS2) technique for performing multivariate uncertainty analyses is now more prevalent, however a key remaining challenge for metrologists in many laboratories is the implicit assumption of Gaussian characteristics for summarizing and analysing measurement model results. Whilst non-Gaussian probability density functions (PDFs) may result from Monte Carlo simulations when the GS2 is applied for more complex non-linear measurement models, in practice results are typically only reported in terms of multivariate expected and covariance values. Due to this limitation the measurement model PDF summary is implicitly restricted to a multivariate Gaussian PDF in the absence of additional higher order statistics (HOS) information. In this paper an earlier classical theoretical result by Rosenblatt that allows for an arbitrary multivariate joint distribution function to be transformed into an equivalent system of Gaussian distributions with mapped variables is revisited. Numerical simulations are performed in order to analyse and compare the accuracy of the equivalent Gaussian system of mapped random variables for approximating a measurement model's PDF with that of an exact non-Gaussian PDF that is obtained with a GS2 Monte Carlo statistical simulation. Results obtained from the investigation indicate that a Rosenblatt transformation offers a convenient mechanism to utilize just the joint PDF obtained from the GS2 data in order to both sample points from a non-Gaussian distribution, and also in addition which allows for a simple two-dimensional approach to estimate coupled uncertainties of random variables residing in higher dimensions using conditional densities without the need for determining parametric based copulas.

**Keywords:** metrology uncertainty / GUM Supplement 2 / Monte Carlo / non-Gaussian PDF / Rosenblatt transformation / pressure balance

## 1 Introduction

### 1.1 Research motivation

The original Guide to the Uncertainty of Measurement (GUM) [1] utilizing an algebraic sensitivity coefficient and mixed frequency and Bayesian statistics approach has found widespread utilization in metrology practise for uncertainty analysis, but is however only strictly valid for linearised measurement models. An uncertainty analysis of non-linear models without any limiting approximations may instead be achieved with the application of the Monte Carlo simulation (MCS) technique using the GUM Supplement 1 (GS1) [2] and GUM Supplement 2 (GS2) [3] approaches for univariate and multivariate models which supersede the original GUM technique.

When the GS1 and GS2 are utilized a fully Bayesian statistics framework is utilized for modelling consistency, whilst newer mathematical refinements to modernise and update the original GUM are being further developed as discussed by Bich et al. [4].

As a result both the GS1 and GS2 techniques are considered suitable for application to generalized and non-linear measurement models which are problematic for the original GUM in its earlier form, where a key limitation in the application of both the GS1 and the GS2 is in the reporting of the summaries of the MCS simulation results for a measurement model's uncertainty if this deviates from a univariate or multivariate Gaussian distribution.

In the case of the GS1 the use of extended lambda distributions (ELDs) as developed by Harris et al. [5] has achieved good success, and has subsequently been applied to non-Gaussian univariate models such as that for a gas pressure balance as discussed by Ramnath [6]. The case

\* Corresponding author: [ramnav@unisa.ac.za](mailto:ramnav@unisa.ac.za)

for non-Gaussian multivariate models is more complex but is in principle directly amenable with the use of multivariate copulas as originally developed by Possolo [7] for metrology uncertainty analysis problems. This approach has subsequently been implemented using parametrized vine copulas for an oil pressure balance problem as discussed by Ramnath [8] in order to demonstrate the utility of this approach for summarizing a non-Gaussian joint PDF in mechanical metrology applications.

Consequently, although the use of copulas thus presents an attractive theoretical tool for directly summarizing arbitrary joint PDFs, either with parametrized vine copulas as outlined by Nagler and Czado [9] or with non-parametrized empirical beta copulas as outlined by Segers et al. [10], the construction and utilization of copulas is still a mathematically complex and challenging application in the absence of readily available software routines and technical guidelines for many practising metrologists.

As a result the formulation of a higher dimensional joint PDF into an equivalent readily reported system of mapped variables which follow a Gaussian distribution with expected values and variances using a Rosenblatt based transformation, thus presents a potentially simpler and more convenient approach for many practising metrologists for analysing GS2 data which may potentially exhibit non-Gaussian characteristics if a functional form of the joint PDF is available.

## 1.2 Research contributions

This paper focuses on using numerical simulations in order to investigate and compare the accuracy of a mathematically equivalent sequential system of mapped Gaussian PDFs with that of an exact non-Gaussian trivariate joint PDF, that is obtained with representative synthesized GS2 MCS statistical data of a pressure balance.

Techniques for conveniently constructing the equivalent system which is composed of univariate marginals and conditional multivariate distributions that are generated with a Rosenblatt transformation of mapped random variables from the GS2 MCS data are demonstrated using proposed developed numerical schemes. Mechanisms for sampling from non-Gaussian distributions are developed and validated by working out the Kullback-Leibler divergence between the actual non-Gaussian joint PDF and the equivalent system of Gaussian PDFs.

Finally, alternative techniques for estimating the uncertainties of the model results directly in terms of the conditional densities from the mapped measurement model variables without the utilization of copulas are demonstrated for practical metrology problems that may be encountered in high accuracy calibrations and laboratory inter-comparisons.

## 2 Literature review

### 2.1 Non-Gaussian PDFs in metrology studies

The conventional practise in reliability and quality engineering studies as an overlapping field of metrology

uncertainty analysis is to approximate a multivariate PDF in terms of statistical moments.

One approach by Acar et al. [11] uses the additive decomposition approach by Rahman and Xu [12] to approximate a univariate function  $Y = f(\mathbf{X})$  for an input  $\mathbf{X} = [X_1, \dots, X_N]^T$  as a summation of one-dimensional functions such that the approximation  $\hat{Y}(\mathbf{X})$  is  $\hat{Y}(\mathbf{X}) = \sum_{j=1}^N Y(\mu_1, \dots, \mu_{(j-1)}, X_j, \mu_{(j+1)}, \dots, \mu_N) - (N-1)y(\mu_1, \dots, \mu_N)$  where  $\mu_j, j = 1, \dots, N$  are the mean values of the random variables  $X_j$ . For this approach each of the independent random variables follows an underlying probability distribution  $g_j(X_j)$  and the problem is equivalent to the conventional metrology problem of determining the uncertainty of a measurand model specified as  $Y = f(\mathbf{X})$  using the GS1, where the principle difference is that Higher Order Statistics (HOS) information may be incorporated.

The application of HOS moments may also be extended to approximating the joint PDF for multivariate functions  $\mathbf{Y}(\mathbf{X})$  i.e. for a measurement model of the form  $\mathbf{Y} = \mathbf{f}(\mathbf{X})$  where  $\mathbf{Y} = [Y_1, \dots, Y_m]^T$  is a vector of outputs as discussed by Bretthorst [13] using the maximum entropy method of moments (MEM). The application of the MEM has however until very recently been constrained by the following five key issues in metrology studies, namely:

- The data samples themselves are not used but rather the statistical moments which is not consistent with a Bayesian approach
- The number of statistical moments and associated Lagrange multipliers are unknown
- There is no consistent optimization search algorithm
- There is no consistent approach to assign uncertainties to the Lagrange multipliers in accordance with a Bayesian framework
- There is no consistent approach to assign uncertainties to the PDF in accordance with a Bayesian framework

The above combination of issues have subsequently recently been resolved for metrology uncertainty analysis of measurement systems using a combined maximum entropy and Bayesian statistics approach developed by Armstrong et al. [14]. This new MaxEnt/Bayesian approach can therefore now in principle be applied to the MCS method either for problems in which there is a heavy computational cost, or alternately for highly non-linear problems for which non-Gaussian PDFs may be present. Investigations performed by Armstrong et al. [14] indicate that the MaxEnt/Bayesian method is able to correctly recover the measurement model's PDF for Gaussian, non-Gaussian and lognormal distributions for sample sizes larger than  $M > 60$  simulation events when using sixth order HOS moments.

This more recent work may thus be contrasted with many earlier metrology studies using the GS2 method that were specifically restricted to second order statistical moments i.e. a first order expected value and a second order covariance. In the conventional GS2 approach although HOS information is implicitly encoded and

readily available within the MCS data typically denoted as  $\Omega$ , a multivariate Gaussian PDF for a measurement model is nevertheless always implicit based on maximum statistical entropy reasons if only second order statistical moments are available. This joint PDF summary approach in the GS2 is therefore common both to models with a low computational intensity such as algebraic based models that may generate a high number of Monte Carlo simulation events, as well as to measurement systems with an intrinsic high computational intensity such as finite element method (FEM) and finite volume method (FVM) based models that may only generate a small number of Monte Carlo simulation events due to limited computational resources, since only second order statistical moments are reported for summarizing the joint PDF.

Computationally intensive models such as those governed by FEM and/or FVM based partial differential equation solutions in mechanical and electrical metrology systems generally make an implicit working approximation to justify a Gaussian distribution when only the expected value  $\tilde{\mathbf{y}}$  vector as a first order statistical moment and the covariance  $\mathbf{U}_{\tilde{\mathbf{y}}}$  matrix as a second order statistical moment for a measurement model  $\mathbf{y} = \mathbf{f}(\mathbf{x})$  are available, due to the finite number of simulation events that are available due to limited computational resources. An additional implicit assumption of a Gaussian PDF is also utilized for many low computational intensity models that could generate a larger number of Monte Carlo simulation events since both GUM supplements are presently explicitly restricted to only utilizing first and second order statistical moments. Due to this limitation the GS2 whilst theoretically allowing for the possibility of non-Gaussian PDFs within a Bayesian statistics modelling framework as encoded within  $\Omega$ , and which may in principle be used to generate HOS information, is therefore unable to offer any specific technical guidance on how to utilize HOS information for modelling, summarizing and analysing non-Gaussian characteristics in uncertainty analysis work.

If a metrologist thus allows for the possibility of a measurement model that may exhibit a non-Gaussian PDF that cannot adequately be approximated with a Gaussian PDF, it then logically becomes necessary to either model and summarize non-Gaussian PDFs either directly with multivariate copulas, or indirectly with models that utilize HOS information such as third, fourth and higher order statistical moments, or alternately to completely avoid approximations altogether but to simply utilize the actual joint PDF obtained from the MCS GS2 data  $\Omega$ .

In this latter approach, an equivalent system that can model and summarize the non-Gaussian characteristics using mapped random variables of transformed Gaussian PDFs as statistical ‘building blocks’ from the actual joint PDF, thus presents an appealing theoretical alternative for summarizing and analysing a non-Gaussian joint PDF.

## 2.2 Review of Rosenblatt transformation method

The development of the transformation that maps random variables from a general and possibly non-Gaussian distribution with coupling effects between the random

variables, to an equivalent system of new mapped random variables that follow Gaussian distributions was originally discovered by Rosenblatt [15].

In the classical paper by Rosenblatt [15]  $\mathbf{X} = [X_1, \dots, X_n]^T \in \mathbb{R}^n$  was a random vector with a cumulative distribution function  $F(X_1, \dots, X_n)$  such that a new random vector  $\mathbf{Z} = [Z_1, \dots, Z_n]^T$  was generated using a transformation  $\mathbf{T}$  such that  $\mathbf{Z} = \mathbf{T}(\mathbf{X})$ . This transformation was specified such that the corresponding random variables  $z_1, \dots, z_n$  were

$$\begin{aligned} z_1 &= P(X_1 \leq x_1) \\ z_2 &= P(X_2 \leq x_2 | X_1 = x_1) \\ &\vdots \\ z_n &= P(X_n \leq x_n | X_1 = x_1, \dots, X_{n-1} = x_{(n-1)}) \quad (1) \end{aligned}$$

Subsequent studies by Lebrun and Dotfoy [16] using results from more recent copula developments established the equivalence between the original Rosenblatt transformation and later refinements such as the generalized Nataf transformation. As a result in this paper we will restrict the investigation to the classical Rosenblatt transformation which maps from the original distribution to a system of Gaussian distributions.

The Rosenblatt transformation using a more modern approach by Lebrun and Dotfoy [16] is defined as a composition of two transformations such that if  $\mathbf{X} \in \mathbb{R}^n$  is a continuous random variable defined by univariate marginal cumulative distribution functions and its corresponding copula  $C$ , then the Rosenblatt transformation denoted as  $\mathbf{T}^R$  is defined by a composition of two sequential transformations such that  $\mathbf{U} = \mathbf{T}^R(\mathbf{X}) = \mathbf{T}_2^R \circ \mathbf{T}_1^R(\mathbf{X})$ .

The first transformation  $\mathbf{T}_1^R : (\mathbf{X} \subseteq \mathbb{R}^n) \rightarrow (\mathbf{Y} \subseteq \mathbb{R}^n)$  maps the original variate points  $\mathbf{x} \in \mathbf{X}$  into a corresponding space of points  $\mathbf{y} \in \mathbf{Y}$  where the transformation  $\mathbf{T}_1^R(\mathbf{X}) = \mathbf{Y}$  is defined as  $\mathbf{Y} = [F_{\mathbf{X}}(x_1), \dots, F_{\mathbf{X}}(x_n | x_1, \dots, x_{(n-1)})]^T$ . In this approach  $F_{\mathbf{X}}(x_j | x_1, \dots, x_{j-1})$  is the conditional cumulative distribution function (CDF) for a random variable  $x_j$  for  $j = 1, \dots, n$  and it follows from standard statistical theory that any conditional CDF automatically follows a rectangular PDF i.e.  $F_{\mathbf{X}}(x_j | x_1, \dots, x_{j-1}) \sim U[0, 1]$  for  $j = 1, \dots, n$ .

The second transformation is  $\mathbf{T}_2^R : (\mathbf{Y} \subseteq \mathbb{R}^n) \rightarrow (\mathbf{U} \subseteq \mathbb{R}^n)$  where  $\mathbf{U} = \mathbf{T}_2^R(\mathbf{Y})$  and is defined as  $\mathbf{U} = [\Phi^{-1}(Y_1), \dots, \Phi^{-1}(Y_n)]^T$  where  $\Phi^{-1}$  denotes the inverse normal transformation, formally defined as  $\Phi^{-1}(x) = \sqrt{2} \operatorname{erf}^{-1}(2x - 1)$  where  $\operatorname{erf}(x)$  denotes the error function as discussed by van Albada and Robinson [17]. In this second transformation it logically follows that the final transformed variable follows a Gaussian PDF such that  $U_j = \Phi^{-1}(Y_j) \sim N(\mu = 0, \sigma^2 = 1)$  for  $j = 1, \dots, n$ , due to the explicit choice of a mapping function.

The essential utility of the Rosenblatt transformation is therefore that it allows for a random variable  $v$  that has some arbitrary distribution  $P_v(v)$  to be mapped to the cumulative distribution which is always uniformly i.e. rectangularly distributed on

the interval  $R[0,1]$ , and then the uniform distribution is in turn mapped to a standard normal distribution such that  $\mathsf{T}_1^R : (v \sim P_v(v)) \rightarrow (F \sim R[0,1])$  and then  $\mathsf{T}_2^R : (F \sim R[0,1]) \rightarrow (u \sim N(\mu = 0, \sigma^2 = 1))$ .

Since it is clear that the distributions always refer to  $\mathbf{X}$  the subscript in the conditional distribution  $F_{\mathbf{X}}(x_j|x_1, \dots, x_{(j-1)})$  may be dropped for notational brevity in the rest of the paper so that the transformed variables may then be combined and the Rosenblatt transformation  $\mathsf{T}_2^R \circ \mathsf{T}_1^R(\mathbf{X})$  in the present context may then simply be written as

$$\left. \begin{aligned} u_1 &= \Phi^{-1}(F(x_1)) \\ u_2 &= \Phi^{-1}(F(x_2|x_1)) \\ &\vdots \\ u_n &= \Phi^{-1}(F(x_n|x_1, x_2, \dots, x_{(n-1)})) \end{aligned} \right\} \quad (2)$$

If the normal distribution  $\Phi$  is applied to all of the mapped Gaussian variables i.e.  $\Phi(u_j) = \Phi(\Phi^{-1}F(x_j|x_1, \dots, x_{j-1}))$  for  $j = 1, \dots, n$  an equivalent system of mapped variables corresponding to  $\mathsf{T}_1^R(\mathbf{X})$  that follow a rectangular distribution may be written as

$$\left. \begin{aligned} r_1 &= F(x_1) \\ r_2 &= F(x_2|x_1) \\ &\vdots \\ r_n &= F(x_n|x_1, x_2, \dots, x_{(n-1)}) \end{aligned} \right\} \quad (3)$$

Noting that  $u_j = \Phi^{-1}(F(x_j|x_1, x_2, \dots, x_{(j-1)}))$  with equation (2) and  $r_j = F(x_j|x_1, x_2, \dots, x_{(j-1)})$  with equation (3) are mathematically interchangeable since a known mapping and inverse mapping function exist between these random variables, it may be concluded that either system may be used to specify a Rosenblatt transformation. This is due to the fact that the normal distribution  $\Phi$  and inverse normal distribution  $\Phi^{-1}$  may be used to conveniently map between the random variables  $r_j$  and  $u_j$  such that  $u_j = \Phi^{-1}(r_j)$  and  $r_j = \Phi(u_j)$  for  $j = 1, \dots, n$ .

The above system of conditional CDFs that underpin the Rosenblatt transformation are connected to the equivalent decomposition of an arbitrary joint PDF in terms of conditional PDFs such that

$$f(x_1, \dots, x_n) = f(x_1) \times f(x_2|x_1) \times f(x_3|x_1, x_2) \times \dots \times f(x_n|x_1, \dots, x_{n-1}). \quad (4)$$

In the above equation the order of the decomposition is arbitrary and the equation may equivalently be written in the case of a trivariate joint PDF  $f(x_1, x_2, x_3)$  as  $f(x_1, x_2, x_3) = f(x_3)f(x_2|x_3)f(x_1|x_3, x_2)$  with similar generalizations to higher dimensional joint PDFs.

Referring to the above system it may be seen that if a random sample for  $u_j$ ,  $j = 1, \dots, n$  is generated from a standard normal distribution  $N(\mu = 0, \sigma^2 = 1)$  then the corresponding random variable for  $v$  may be obtained from solving the equation  $\Phi(u) = F(v)$ . As a result a joint PDF  $f(x_1, \dots, x_n)$  with an arbitrary and potentially non-Gaussian distribution may with the aid of the Rosenblatt transformation be modelled as an equivalent system of

Gaussian distributions using the underlying conditional distributions.

Extensions for a generalized Rosenblatt transformation that allow for mappings of arbitrary distributions where  $u_j$  does not necessarily follow an underlying Gaussian distribution with  $N(\mu = 0, \sigma^2 = 1)$  are also theoretically possible as discussed by Chang [18], but are not considered further in this paper since this is conceptually straightforward if the first step  $r_j = F(x_j|x_1, \dots, x_{j-1}) \sim R[0,1]$  has been obtained and another mapping function  $\Psi : r \rightarrow u$  is specified.

From a literature search it may be observed that the inverse cumulative distribution function for a normal distribution  $\Phi^{-1}$  is readily available in both commercial as well as open source software packages. As an example in the commercial package Matlab [19]  $\Phi^{-1}$  is computed as  $\mathbf{u} = \text{icdf}(\text{'Normal'}, \mathbf{p}, \mathbf{A})$  where  $\mathbf{p}$  is a probability and  $\mathbf{A} = [\mu, \sigma]$  is an input of the distributions parameters which in the case of a normal distribution is simply the mean  $\mu$  and standard deviation  $\sigma$ , whilst in the open source package GNU Octave [20]  $\Phi^{-1}$  is computed as  $\mathbf{u} = \text{norminv}(\mathbf{p}, \mu, \sigma)$ .

Since  $\Phi^{-1}$  is relatively straightforward to calculate the practical application of utilizing a Rosenblatt transformation as specified by equation (3) for a metrology uncertainty analysis of non-Gaussian PDFs obtained with a GS2 simulation, then reduces to the following two conceptual and functional steps, namely:

- Selecting a convenient technique/software for calculating the numerical values of the conditional distributions  $F(x_j|x_1, \dots, x_n)$  from the MCS data
- Selecting a convenient technique/software for algebraically summarizing the conditional distributions  $F(x_j|x_1, \dots, x_n)$  from the previous numerical data

## 3 Mathematical models

### 3.1 Generating MCS GS2 data

In a GS2 approach multivariate Monte Carlo data are generated for a model that may have an explicit form  $\mathbf{Y} = \mathbf{f}(\mathbf{X})$  where  $\mathbf{X} = [X_1, \dots, X_N]^T$  is a model input and  $\mathbf{Y} = [Y_1, \dots, Y_m]^T$  a model output, or an equivalent implicit form  $\mathbf{h}(\mathbf{Y}, \mathbf{X}) = \mathbf{0}$  for generality as discussed by Harris et al. [5] if the model cannot be simplified to an explicit form. The input  $\mathbf{X}$  has a corresponding joint PDF  $g_{\mathbf{X}}(\boldsymbol{\xi})$  where  $\boldsymbol{\xi} = [\xi_1, \dots, \xi_N]^T$  is a random variable, whilst the output  $\mathbf{Y}$  has a corresponding joint PDF  $g_{\mathbf{Y}}(\boldsymbol{\eta})$  where  $\boldsymbol{\eta} = [\eta_1, \dots, \eta_m]^T$  is a random variable.

Regardless of the form of the model if a GS2 technique is implemented for  $M$  Monte Carlo simulation events by solving  $\mathbf{y}_r = \mathbf{f}(\mathbf{x}_r)$  or  $\mathbf{h}(\mathbf{y}_r, \mathbf{x}_r) = \mathbf{0}$  for  $r = 1, \dots, M$  for sampled random inputs  $\mathbf{x}_1, \dots, \mathbf{x}_M$  then corresponding random outputs  $\mathbf{y}_1, \dots, \mathbf{y}_M$  will be generated from the solution of the model. The outputs  $\mathbf{y}_1, \dots, \mathbf{y}_M$  may then be conveniently grouped as a  $m \times M$  dimensional matrix  $\boldsymbol{\Omega}$  to represent the MCS model output data. Changing symbols and writing the matrix as  $\boldsymbol{\Omega} = [\mathbf{x}_1, \dots, \mathbf{x}_M]^T$  for  $M$  Monte Carlo simulation events where each random



variable has a dimension of  $m \times 1$  without loss of generality for notational convenience we consider the MCS data to take the general form

$$\Omega = [\mathbf{x}_1, \dots, \mathbf{x}_M]. \quad (5)$$

In a conventional GS2 approach the matrix  $\Omega$  contains all the encoded information of the simulation and the expected value  $\tilde{\mathbf{x}}$  and covariance matrix  $\mathbf{U}_{\tilde{\mathbf{x}}}$  are calculated. If the only known information is  $\tilde{\mathbf{x}}$  and  $\mathbf{U}_{\tilde{\mathbf{x}}}$  then according to maximum statistical entropy theory the PDF will be a multivariate Gaussian distribution in the absence of HOS information such as statistical moments, although the HOS information is theoretically already encoded in  $\Omega$  for finite order moments.

The information encoded in the MCS data represented by  $\Omega$ , may therefore after suitable statistical post-processing be represented as a joint PDF  $f(x_1, \dots, x_n)$  if  $\Omega$  is processed either using scaled histograms as discussed by Oliphant [21] or alternately with kernel density estimate (KDE) approaches as discussed by Scott [22].

A mathematically equivalent representation of the joint PDF is the CDF  $F(x_1, \dots, x_n)$  which is related to the joint PDF by the equation  $f(x_1, \dots, x_n) = \frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n}$ . The CDF may in turn also be defined in terms of the joint PDF by the equation  $F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \dots dt_n$ .

As a result either the joint PDF  $f(x_1, \dots, x_n)$  or equivalently the CDF  $F(x_1, \dots, x_n)$  are both automatically known if the MCS GS2 data in  $\Omega$  is appropriately processed using standard statistical techniques, however the key remaining challenge is in summarizing and analysing a non-Gaussian joint PDF for metrology problems which is the focus of this paper.

### 3.2 Calculating marginal distributions

In order to implement the Rosenblatt transformation in equation (3) it is first necessary to calculate the marginal distributions  $F(x_1), \dots, F(x_n)$  and then the conditional distributions  $F(x_j | x_1, \dots, x_{j-1})$  for  $j = 2, \dots, n$  which in turn rely on the earlier marginal distribution calculations.

Noting that the terms  $F(x_j)$ ,  $j = 1, \dots, n$  are univariate distributions it is observed that these functions may simply be calculated using earlier work by Willink [23] with the aid of extended lambda distributions (ELDs) if the MCS data  $\Omega$  is available. The use of such ELDs which are based on fourth order statistical moments was later adapted and extended by Harris et al. [5] using parametrized families of quantile functions and monotonic splines for greater accuracy if higher statistical moments are necessary. Other possible lambda distributions apart from ELDs include a combined generalized lambda distribution (GLD) and a generalized beta distribution (GBD) to account for problematic skewness issues in a normal GLD approach, where the combination of these approaches is known as an EGLD which was originally developed by Karian et al. [24]. An EGLD

may determined using four parameters  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  that are specified in terms of optimizations using statistical moments if the univariate  $x$ -data points are supplied as an input. Since EGLDs may take slightly different forms many researchers opt to utilize the Freimer-Mudholkar-Kollia-Lin generalized lambda distribution (FMKL-GLD) which is the most common EGLD function which takes the general form  $F^{-1}(u; \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \lambda_2^{-1}[\lambda_3^{-1}(u^{\lambda_3} - 1) - \lambda_4^{-1}((1-u)^{\lambda_4} - 1)]$ .

In the FMKL-GLD univariate model scheme all of the parameters are free where the only restriction is that  $\lambda_2 > 0$ . An additional constraint that may be posed to have finite  $k^{\text{th}}$  statistical moments is to impose the additional restriction that  $\min(\lambda_3, \lambda_4) > -1/k$ . From a review of the literature a maximum likelihood method as discussed by Corlu and Meterelloyoz [25] is generally preferred and may conveniently be obtained using the readily available open source software package fitgbd that has recently been developed by Wang [26].

A possible disadvantage of a more complicated curve fit such as a FMKL-GLD is that the optimization to determine the optimal values for  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  may be more difficult to obtain than using simpler approaches such as a least squares regression, and hence require more advanced approaches such as maximum likelihood estimation (MLE) schemes that are not necessarily available to practicing metrologists.

For this reason in many earlier metrology studies univariate curve fits for the marginal distributions  $F(x_1), \dots, F(x_n)$  typically utilized the more standard ELD form of Willink [23] which takes the simpler form for the quantile function as

$$Q(\rho) = \begin{cases} d + (c/b)[a\rho^b - (1-\rho)^b + 1 - a], & b \neq 0 \\ d + c[a \log(\rho) - \log(1-\rho)], & b = 0 \end{cases}. \quad (6)$$

The above ELD equation is based on fourth order statistical moments and may be used to fit the data for the marginal distribution  $F(x_1)$  by sampling a random point  $\rho$  where  $0 \leq \rho \leq 1$  corresponding to  $\rho = \Phi(u_1)$  if  $u_1$  is a sampled point from  $U \sim N(0, 1)$  and then working out the corresponding random variable  $x_1$  from the equation  $x_1 = Q(\rho)$ . A similar approach may also be used for all of the remaining marginal distributions  $F(x_2), \dots, F(x_n)$  so that all of the marginal distributions  $F(x_j)$  are modelled using parameters  $a_j, b_j, c_j, d_j$  for  $j = 1, \dots, n$ . In all of the above lambda distribution schemes it is generally assumed that a unimodal behaviour is present i.e. there is a 'single peak' of the random variable with an element of skewness and kurtosis. In the event that a more complicated behaviour is present with multiple peaks then univariate splines as developed by Harris [5] may instead be utilized.

If the underlying MCS data  $\Omega$  is not available, or if a parameter based functional fit using fourth order statistical moments is inadequate, and only the joint PDF  $f(x_1, \dots, x_n)$  is known then the marginal distributions, either with single or multiple peaks, may instead be calculated directly from the joint PDF with the HOS

information using the fundamental equation such that

$$f(x_j) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) \times du_1 \cdots du_{j-1} du_{j+1} \cdots du_n. \quad (7)$$

### 3.3 Calculating conditional distributions

Since  $f(x_1, \dots, x_n)$  or  $F(x_1, \dots, x_n)$  are assumed to be known for a MCS data set  $\Omega$  which may be Gaussian or non-Gaussian, the only remaining requirement to convert the problem into the standard form specified by equation (3) is to calculate the conditional distributions  $F(x_j|x_1, \dots, x_{j-1})$  for  $j = 2, \dots, n$  since the marginal distribution corresponding to the case for  $j = 1$  has already been specified in the previous section. Two approaches are possible, namely an indirect approach that uses copulas, and a direct approach which uses numerical and symbolic integrations of the joint PDF.

The indirect approach which utilizes copulas to calculate the conditional distribution makes use of a result by Nagler [27] and just the MCS data  $\Omega$ . In this approach the conditional univariate/multivariate distribution is specified using the copula  $C$  from  $F(x_1, \dots, x_n) = C(F(x_1), \dots, F(x_n))$  such that

$$F(x|\mathbf{v}) = \frac{\partial C(F(x|\mathbf{v}_{-j}), F(\mathbf{v}_j|\mathbf{v}_{-j}))}{\partial F(\mathbf{v}_j|\mathbf{v}_{-j})}. \quad (8)$$

In the above formula  $\mathbf{v} = [v_1, \dots, v_k]^T$ ,  $v_j$  is any convenient arbitrary element of  $\mathbf{v}$  e.g.  $v_j = v_1$  or  $v_j = v_k$ ,  $\mathbf{v}_{-j} = \mathbf{v} \setminus \{v_j\}$  i.e. the vector  $\mathbf{v}$  with the element  $v_j$  removed, and the copula is constructed with  $F(x|\mathbf{v}_{-j}) = z_1$  as a first argument and  $F(\mathbf{v}_j|\mathbf{v}_{-j}) = z_2$  as a second argument i.e. the copula  $C(z_1, z_2)$ .

For conditional univariate/univariate distributions the above formula simplifies to

$$F(x, v) = \frac{\partial C(F_x(x), F_v(v))}{\partial v}. \quad (9)$$

When these formulae are applied to equation (3) it follows that the conditional distributions may be calculated as

$$F(x_2|x_1) = \frac{\partial C(F(x_2), F(x_1))}{\partial F(x_1)} \quad (10)$$

$$\begin{aligned} & F(x_j|x_1, \dots, x_{j-1}) \\ &= \frac{\partial C(F(x_j|x_1, \dots, x_{j-2}), F(x_{j-1}|x_1, \dots, x_{j-2}))}{\partial F(x_{j-1}|x_1, \dots, x_{j-2})}, \\ & j = 3, \dots, n \end{aligned} \quad (11)$$

Noting that the conditional in equation (11) for an index  $j$  is defined in terms of the previous conditional for  $j - 1$  and that equation (10) has an explicit expression for  $j = 2$  it may be observed that equation (10) and equation (11) then allow for all the conditionals for  $j = 2, \dots, n$  to be calculated. As a result the calculation of

conditional distributions reduces to fitting bivariate copulas for  $C(z_1, z_2)$  for specified marginals  $z_1$  and  $z_2$ , and conditional variate data.

Different approaches to fitting bivariate copulas are possible, all of which essentially reduce to parameter optimizations for choices of bivariate copula families, however the simplest approach is to use the open source R software package VineCopula by Nagler et al. [28] which performs the optimization with the following R code:

```
C12 <- BiCopSelect(u1, u2, familyset = NA)
# c(1, 2, 3, 4, 5, 6, 7)
```

At the present time of writing, although some other routines in Matlab and Python for fitting copulas are available, the R-based VineCopula package is considered the most advanced and practical software package for fitting and estimating copulas for multivariate data.

In the above code fragment  $u1$  and  $u2$  are the marginals for random variables  $x_1$  and  $x_2$  calculated as  $u_1 = F(x_1)$  and  $u_2 = F(x_2)$  and  $C12$  is an object that contains the properties of the bivariate copula which may be used to specify its mathematical equation by looking up the properties of the object and in the software package's reference documentation.

Although forty different choices of parameter based bivariate copula may be fitted with the VineCopula package not all of these copula families are fully supported. As a result from a practical implementation perspective it is sometimes necessary to restrict the best copula fit to the first seven standard copulas which are fully supported viz. the independence Archimedean, Gaussian, Student- $t$ , Clayton, Gumbel, Frank, Joe, and BB1 copulas by specifying the selection as `familyset = c(1,2,3,4,5,6,7)`. The effect of this restriction is that the most accurate choice of copula may not be fitted due to the combination of the mathematical complexity of the copula's functional form and the associated numerical challenges with performing the parameter optimization fit.

In the event that a supported standard bivariate copula is fitted and its parameters determined, then the partial derivatives  $\frac{\partial C(u_1, u_2)}{\partial u_1}$  and  $\frac{\partial C(u_1, u_2)}{\partial u_2}$  may be calculated for specified values of  $u_1$  and  $u_2$  using the following code:

```
dC12du1 <- BiCopHfunc1(u1val, u2val, C12)
dC12du2 <- BiCopHfunc2(u1val, u2val, C12)
```

Whilst the above indirect approach is conceptually simple it is critically reliant on the use of the VineCopula software library, and may therefore not be accessible to metrologists who are not familiar with and proficient in programming with the R-language for statistical computing. As a result to cater for a variety of metrologists who may utilize other software languages and packages it is thus beneficial to utilize a direct approach for generality.

The direct approach utilizes numerical/symbolic integrations of the joint PDF  $f(x_1, \dots, x_n)$  which is assumed to be known in a form that is amenable to calculations, where  $f(x_1, \dots, x_n)$  may be obtained from results using scaled histograms or KDE schemes from the MCS data  $\Omega$  as previously discussed.

In order to calculate the conditional distributions  $F(x_j|x_1, \dots, x_{j-1})$  it is first necessary to calculate the conditional density  $f(x_j|x_1, \dots, x_{j-1})$ . This may be achieved by first assuming that  $f(x_1, \dots, x_n)$  is known so that

$$\begin{aligned} f(x_1, \dots, x_j) \\ = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_j, u_{j+1}, \dots, u_n) \\ \times du_{j+1} \times \cdots \times du_n, \quad 1 \leq j \leq n. \end{aligned} \quad (12)$$

The above result may then be used to calculate the conditional density such that

$$f(x_j|x_1, \dots, x_{j-1}) = \frac{f(x_1, \dots, x_j)}{f(x_1) \times \cdots \times f(x_{j-1})}. \quad (13)$$

The conditional distribution is then finally calculated from the conditional density such that

$$F(x_j|x_1, \dots, x_{j-1}) = \int_{-\infty}^{x_j} f(u_j|x_1, \dots, x_{j-1}) du_j. \quad (14)$$

In the above formula for  $F(x_j|x_1, \dots, x_{j-1})$  no simplifying assumptions are utilized, and the conditional distributions implicitly utilizes all of the HOS information inclusive of potential non-Gaussian characteristics if the joint PDF  $f(x_1, \dots, x_n)$  has been correctly determined from the GS2 data  $\Omega$  obtained from a MCS.

As a result, the conditional distribution may be conveniently calculated either symbolically or numerically using any convenient software package such as Matlab, Octave, Python or R if the joint PDF  $f(x_1, \dots, x_n)$  is known without the necessity of utilizing specialist statistical packages.

### 3.4 Comparing GS2 and Rosenblatt PDFs

In order to compare two different PDFs,  $f_{\mathbf{X}}(\mathbf{x})$  corresponding to the GS2 based exact non-Gaussian joint PDF and  $f_{\mathbf{Y}}(\mathbf{x})$  corresponding to the Rosenblatt equivalent joint PDF, the Kullback-Leibler divergence defined as

$$D_{KL}(\mathbf{X}, \mathbf{Y}) = \int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x}) \log \left\{ \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{x})} \right\} d\mathbf{x} \quad (15)$$

is used as a measure to estimate the ‘distance’ or deviation between the two PDFs.

If the joint PDFs are equal every such that  $f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{Y}}(\mathbf{x})$  for all points  $\mathbf{x} \in \Omega$  it then follows that  $\log(f_{\mathbf{X}}(\mathbf{x})/f_{\mathbf{Y}}(\mathbf{x})) = \log(1) = 0$ . As a result when this function is integrated for all points  $\mathbf{x} \in \Omega$  it follows that  $D_{KL}(\mathbf{X}, \mathbf{Y}) = 0$ . If on the other hand the functional evaluations are not equal every where then  $D_{KL}(\mathbf{X}, \mathbf{Y}) \neq 0$ . As a result the closer  $D_{KL}(\mathbf{X}, \mathbf{Y})$  is to zero then the better the ‘fit’ is between the two different joint PDFs.

## 4 Numerical simulations

### 4.1 Generating a MCS GS2 non-Gaussian dataset

Earlier work by Ramnath [8] studied a bivariate pressure balance model of the form  $A = A_0(1 + \lambda P)$  where  $A$  is the effective area of the pressure balance,  $\lambda$  is a distortion coefficient and  $P$  is an independently varied applied pressure in the range  $50 \leq P/[\text{MPa}] \leq 500$ . In this model the parameters  $A_0$  and  $\lambda$  are coupled to each other and a bivariate PDF  $g_{A_0, \lambda}(\eta_{A_0}, \eta_{\lambda})$  is used to model the characteristics of the pressure balance. This physical pressure balance model may be extended to a three parameter model of the form

$$A = A_0(1 + \lambda_1 P + \lambda_2 P^2) \quad (16)$$

as discussed by Yadav et al. [29] where the parameters  $A_0$ ,  $\lambda_1$  and  $\lambda_2$  are coupled in a trivariate joint PDF  $g_{A_0, \lambda_1, \lambda_2}(\eta_{A_0}, \eta_{\lambda_1}, \eta_{\lambda_2})$ , however relatively few experimental data sets are available of pressure balances which exhibit strongly non-Gaussian characteristics. In the earlier work by Ramnath [8] the experimental data-set exhibited weakly non-Gaussian characteristics i.e. the non-Gaussian joint PDF could be roughly approximated with a Gaussian joint PDF for a reduced level of accuracy, and as a result there exists a general research gap in the area of metrology uncertainty analysis for summarizing and analysing non-Gaussian joint PDFs that cannot be adequately approximated with Gaussian joint PDFs.

In order to avoid the unnecessary complexity of solving for a full measurand model which may potentially produce a multivariate weak non-Gaussian distribution in this paper we therefore opt to instead directly model a joint PDF using an Azzalini multivariate skew-normal (AMSN) distribution. The AMSN which was developed by Azzalini and Capitanio [30] is denoted as  $\mathbf{Y} \sim SN_k(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$  and is utilized in this paper in order to specifically ensure a strongly non-Gaussian distribution that cannot be adequately summarized and approximated with a GS2 multivariate Gaussian joint PDF. Under these circumstances the actual non-Gaussian joint PDF for a random variable  $\mathbf{y} \in \mathbb{R}^k$  is defined as

$$g_{\mathbf{Y}}(\mathbf{y}) = 2\phi_k(\mathbf{y} - \boldsymbol{\xi}; \boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}^T \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})). \quad (17)$$

In the above equation  $\boldsymbol{\xi}$ ,  $\boldsymbol{\Omega}$  and  $\boldsymbol{\alpha}$  are parameters for the AMSN where  $\boldsymbol{\xi}$  is a center for the distribution,  $\boldsymbol{\Omega}$  is a covariance matrix, and  $\boldsymbol{\alpha}$  is a shape parameter.

Open source software routines have subsequently been developed by Azzalini [31] to easily and conveniently sample from the AMSN distribution if the equivalent quantities for the mean  $\boldsymbol{\mu}$ , covariance  $\boldsymbol{\Omega}$  and shape parameters  $\boldsymbol{\alpha}$  are provided. The use of these open source routines thus allows for a convenient approach to generate synthetic non-Gaussian Monte Carlo data in order to build up  $\Omega$  for subsequent analysis in order to construct either  $f(x_1, \dots, x_n)$  or  $F(x_1, \dots, x_n)$ .

Simulations in this paper are performed for a skew-normal  $SN_3$  distribution in order to artificially emulate a pressure balance model that has a non-Gaussian joint

**Table 1.** Parameters for a generating a non-Gaussian pressure balance joint PDF model using a multivariate skew-normal distribution .

$\mu_1$	$\mu_2$	$\mu_3$	$r_{12}$	$r_{13}$	$r_{23}$	$\sigma_1$	$\sigma_2$	$\sigma_3$	$\delta_1$	$\delta_2$	$\delta_3$
1.961926	2.611996	6.409901	0.26	0.12	0.06	$0.05\mu_1$	$0.035\mu_2$	$0.04\mu_3$	0.4	1.0	0.6

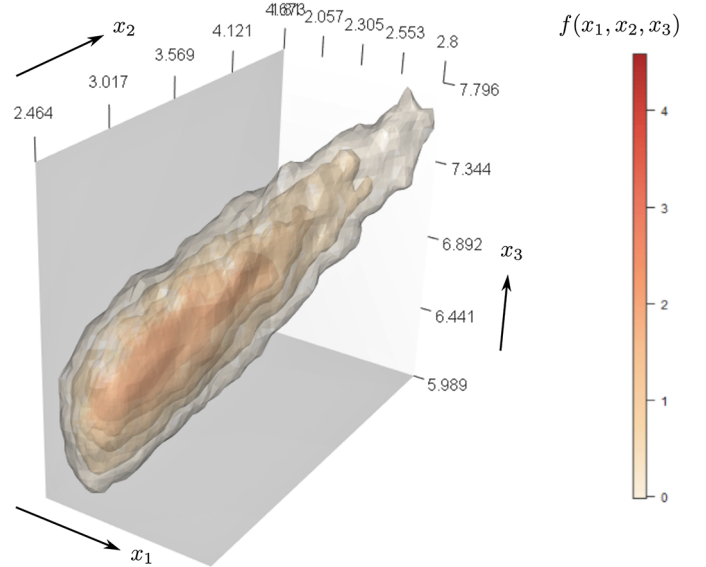
PDF equation where the three-dimensional measurement model has the form  $A = A_0(1 + \lambda_1 P + \lambda_2 P^2)$ , where the random variables are  $x_1 = \eta_{A_0}$ ,  $x_2 = \eta_{\lambda_1}$  and  $x_3 = \eta_{\lambda_2}$  so that the joint PDF  $g_{A_0, \lambda_1, \lambda_2}(\eta_{A_0}, \eta_{\lambda_1}, \eta_{\lambda_2})$  is represented as  $f(x_1, x_2, x_3)$  for notational brevity and simplicity. The parameters for the corresponding  $SN_3$  function are reported in Table 1 where the covariance matrix is built up in terms of the correlation coefficients such that  $\Sigma_{ij} = r_{ij}\sigma_i\sigma_j$ .

Using the data in Table 1 then allows for the construction of the mean  $\mu = [\mu_1, \mu_2, \mu_3]^T$ , covariance  $\Sigma$ , and shape parameter  $[\delta_1, \delta_2, \delta_3]^T$  so that the Monte Carlo data  $\text{data}\Omega = \Omega$  and corresponding joint PDF  $\text{data}f = f(\mathbf{x})$  data may be generated by running the following computer code in a RStudio version 1.2.5001 developer environment using a R version 3.6.1 code base running in Microsoft Windows 10 as shown below:

```
library(EMMIXskew)
n <- 15000
dataOmega <- rdmsn(n, p, mean, cov, del)
dataf <- ddmsn(y, n, p, mean, cov, del)
write.table(dataOmega,
  file = paste(getwd(), '/Omega.txt', sep = ""),
  append = FALSE, sep = " ", dec = ".",
  row.names = FALSE, col.names = FALSE)
write.table(dataf,
  file = paste(getwd(), '/Density.txt', sep = ""),
  append = FALSE, sep = " ", dec = ".",
  row.names = FALSE, col.names = FALSE)
```

In the above code fragment EMMIXskew is the library developed by Azzalini [31] that must first be loaded in order to call the functions `rdmsn` that samples random points from the distribution and `ddmsn` that conveniently calculates the probability density function values for the corresponding sampled points.

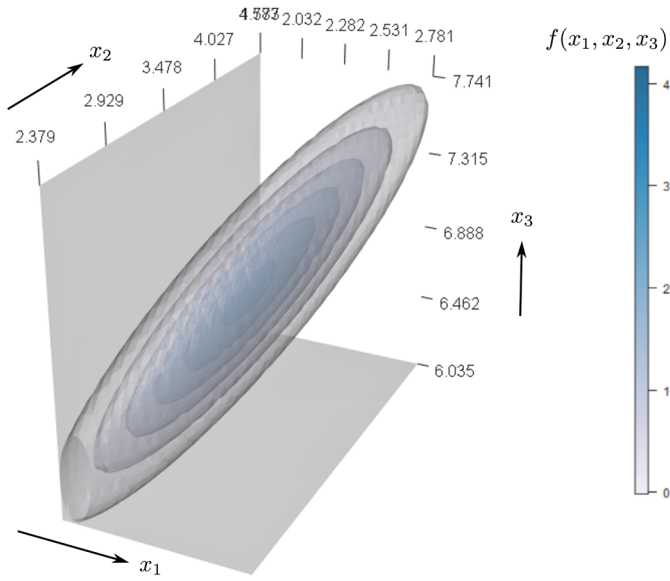
Considering the synthetic MCS non-Gaussian data set, the variable  $x_1$  emulates a scaled random variable of zero-pressure area  $\eta_{A_0}$ , the variable  $x_2$  emulates a scaled random variable of the first distortion coefficient  $\eta_{\lambda_1}$ , and the variable  $x_3$  emulates a scaled random variable of the second distortion coefficient  $\eta_{\lambda_2}$ . Typical physical units for these physical parameters are  $A_0 = \mathcal{O}(10^{-6}\text{m}^2)$ ,  $\lambda_1 = \mathcal{O}(10^{-12}\text{Pa}^{-1})$  and  $\lambda_2 = \mathcal{O}(10^{-24}\text{Pa}^{-2})$  if the applied pressure is in the range  $50 \leq P/[\text{MPa}] \leq 500$  as reported by Ramnath [32] in later studies. As a result the physical meaning of the synthetic data that is generated with the AMSN is that the Monte Carlo data corresponds to random variables which have scaled units of  $x_1/[\text{mm}^2]$ ,  $x_2/[\text{ppm}/\text{MPa}]$  and  $x_3/[(\text{ppm}/\text{MPa})^2]$  respectively for a pressure balance measurement model of the form  $A = x_1(1 + x_2P + x_3P^2)$ .

**Fig. 1.** Volumetric isosurfaces visualization of the actual GUM supplement 2 non-Gaussian joint PDF obtained from synthetic sampled Monte Carlo simulation data-points generated from a three-variable Azzalini multivariate skew-normal joint PDF.

For the example shown  $n = 15000$  random multivariate points are sampled from the AMSN distribution and the corresponding  $\Omega$  and  $f(\mathbf{x})$  data are saved to text files `omega1.txt` and `density1.txt`. The actual non-Gaussian joint PDF is visualized in Figure 1 using an isosurface volumetric perspective with 8 isodensity levels from the minimum to maximum joint PDF values. In this figure, the skew and asymmetric behaviour due to the non-Gaussian distribution behaviour may be qualitatively but subjectively observed.

Referring to the non-Gaussian joint PDF plot it is observed that the points are clustered near the ‘bottom left’ as seen by the darker colors and are then “stretched” towards the ‘upper right’. This non-Gaussian behaviour may be visually contrasted with the corresponding Gaussian approximation as shown in Figure 2 also for 8 isodensity levels from the minimum to maximum joint PDF values, from which it is observed that the points are clustered in the ‘middle region’, and are symmetrically “stretched” in an equal manner towards the ‘bottom left’ and ‘upper right’ respectively as a symmetric ellipsoid. An additional observation is that the Gaussian approximation “cuts-off” part of the joint PDF for the same range of  $x_1$ ,  $x_2$  and  $x_3$  points as the regions are ‘equally stretched’ out in both directions, whereas the original non-Gaussian exhibits an ‘unequal stretching’. Consequently when comparing the two visualizations of the actual non-Gaussian



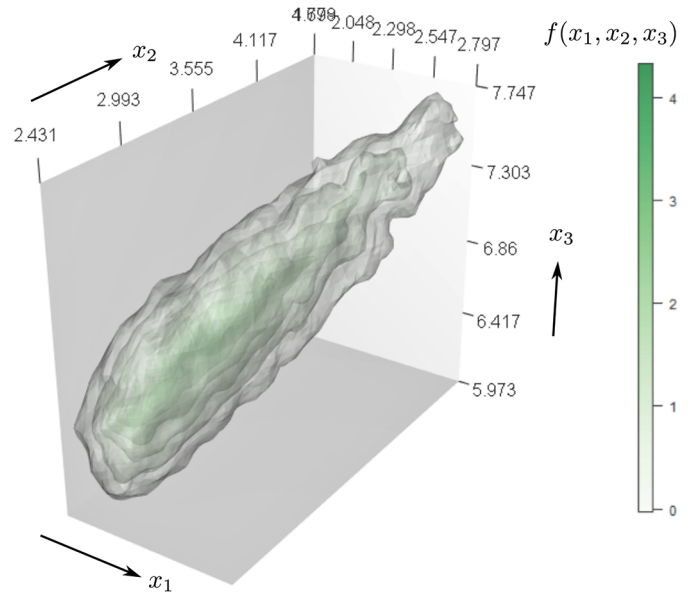


**Fig. 2.** Volumetric isosurfaces density visualization of the approximate GUM supplement 2 Gaussian joint PDF obtained from synthetic Monte Carlo data sampled from a Gaussian distribution using second order statistical moments based expected value and covariance information.

and approximate Gaussian joint PDFs, it may be observed and concluded that the Gaussian joint PDF with low order second order statistical moments i.e. an expected value  $\tilde{\mathbf{y}}$  vector and covariance  $\mathbf{U}_{\tilde{\mathbf{y}}}$  matrix cannot adequately approximate a non-Gaussian joint PDF.

The equivalent Rosenblatt transformation system constructed with  $N = 15000$  sampled points from the marginal and conditional density equations outlined later in the paper is shown in Figure 3. In this plot it is clearly observed that whilst there are minor fluctuations in the isodensity values in different regions of points  $\mathbf{x} \in \Omega$  due to the stochastic nature in which both the GS2 and Rosenblatt systems were generated that the overall characteristics of the GS2 system in Figure 1 is broadly equivalent to that of the Rosenblatt system in Figure 3 as both systems exhibit strong non-Gaussian characteristics. Both the exact and equivalent non-Gaussian characteristics of the GS2 and Rosenblatt systems when contrasted with the GS2 Gaussian approximation constructed with the expected value and covariance in Figure 2 are found to exhibit differences. As a result it may be qualitatively observed that the traditional GS2 approach for summarizing a multivariate distribution with second order statistical moments using a multivariate Gaussian, is problematic and inaccurate if there is a strong non-Gaussian behaviour in the actual data.

These numerical results are suggestive that it may be advantageous for a metrologist to first test for multivariate normality in order to decide whether a multivariate Gaussian distribution is adequate instead of automatically assuming this condition by only specifying the first order statistical moment in  $\tilde{\mathbf{y}}$  and the second order statistical moment in  $\mathbf{U}_{\tilde{\mathbf{y}}}$ . In the event that such tests determine that a Gaussian distribution is problematic



**Fig. 3.** Volumetric isosurfaces density visualization of the equivalent Rosenblatt transformation system.

then the alternative would be to construct the Rosenblatt transformation system so this could instead be utilized to appropriately model and analyse a joint PDF where second order statistical moments are inaccurate.

Noting that in a practical metrology uncertainty analysis such as a laboratory primary standard calibration or a laboratory inter-comparison, where additional quantitative statistical tests are necessary to make an objective assessment as to whether the distribution is sufficiently strongly non-Gaussian, a choice of quantitative test to objectively check for multivariate normality then becomes necessary in order to decide whether the non-Gaussian joint PDF may be approximated with a Gaussian joint PDF.

Different quantitative types of statistical tests to check for normality of univariate data are available and include the Shapiro-Wilk, D'Agostino  $K$ -squared, and Anderson-Darling tests amongst others as discussed by Ghasemi and Zahediasl [33], with related extensions for multivariate data. In this paper we opt to test for multivariate normality using the approach originally proposed by Mardia [34] for simplicity. The Mardia normality approach tests for multivariate skewness and kurtosis, and is chosen for its simplicity and ready availability in the open source R-package MVN package by Korkmaz et al. [35].

Quantitative results when this normality test is applied to the GS2 data-set from Figure 1 are summarized in Figure 4 and similarly for the Rosenblatt transformation system in Figure 5. In this statistical summary it may be observed that both distributions do not satisfy the Mardia skewness and kurtosis normality tests i.e. neither the actual GS2 nor the equivalent Rosenblatt data can be approximated with a three-dimensional Gaussian distribution. This information objectively and quantitatively demonstrates that the GS2 data fails both the normality criteria checks as the  $p$ -values are significantly below

```

library(MVN)
omega1 <- read.table(paste(getwd(),'/omega1.txt',sep = ""), header = FALSE, sep = "", dec = ".")
x1 <- omega1[, 1]
x2 <- omega1[, 2]
x3 <- omega1[, 3]
nsample <- 5000
spoints <- sample(1:length(x1),nsample, replace = T)
omegasample <- matrix(nrow = nsample, ncol = 3)
for (i in seq(1, nsample)){
  omegasample[i, 1] <- x1[spoints[i]]
  omegasample[i, 2] <- x2[spoints[i]]
  omegasample[i, 3] <- x3[spoints[i]]
}

> summary1
$multivariateNormality
      Test      Statistic      p value Result
1 Mardia Skewness 706.823091075035 2.15355991406288e-145 NO
2 Mardia Kurtosis 4.43037939619507 9.40674444249723e-06 NO
3          MVN          <NA>          <NA>      NO

$univariateNormality
      Test Variable Statistic  p value Normality
1 Shapiro-Wilk Column1      0.9684 <0.001      NO
2 Shapiro-Wilk Column2      0.9380 <0.001      NO
3 Shapiro-Wilk Column3      0.9850 <0.001      NO

$Descriptives
      n      Mean Std.Dev  Median    Min    Max   25th   75th   Skew
1 5000 2.272033 0.2566293 2.234310 1.679273 3.425019 2.083853 2.425495 0.7142794
2 5000 3.389026 0.5976767 3.258939 2.404850 6.353510 2.920255 3.744544 0.9121056
3 5000 6.874244 0.4324934 6.833387 5.518941 8.716747 6.570891 7.130277 0.5080655

      Kurtosis
1 0.4367556
2 0.5797602
3 0.3287921

```

**Fig. 4.** Statistical summary of multivariate non-Gaussian characteristics of the GUM supplement 2 Monte Carlo data using the Mardia normality test.

the cut-off thresholds that requires  $p > 0.05$  in order to satisfy the normality criteria. Referring to these results an additional benefit that may be noted when using the standard Mardia test with the MVN package, is that this statistical test also in addition reports on the normality of the marginal distributions of  $f(x_1)$ ,  $f(x_2)$  and  $f(x_3)$  in the case of a joint PDF  $f(x_1, x_2, x_3)$  using the Shapiro-Wilk test. A consequence of this is that caution should also be exercised in attempting to model the marginal distributions with extended lambda distributions (ELDs) as ELDS are generally utilized to model weakly non-Gaussian distributions that do not deviate too far from univariate Gaussian distributions. This result is suggestive that the distributions may need to be modelled either with parametrized high order statistical moments or alternately with non-parametrized kernel density estimates if there is a strong non-Gaussian nature in the distribution.

Referring to these additional results it is observed that all of the marginal distributions also do not pass the normality tests since all of  $p$ -values are again significantly

smaller than 0.001, where a value of  $p > 0.05$  is similarly required in order for the marginal distribution to be considered sufficiently Gaussian in nature. As a result, it may be concluded that the GS2 non-Gaussian joint PDF data under consideration cannot be approximated with a multivariate Gaussian distribution, and hence a measurement uncertainty analysis must either directly utilize the non-Gaussian joint PDF or alternately indirectly utilize the same data in an equivalent and more convenient form for further analysis.

Due to the complexity of the joint PDF under consideration for generality a functional form for  $f(x_1, x_2, x_3)$  must be specified in order to determine the marginal distribution  $F(x_1)$ , and the associated conditional distributions  $F(x_2|x_1)$  and  $F(x_3|x_1, x_2)$  for mathematical completeness. Since the data  $x_1, x_2, x_3$ , and  $f(x_1, x_2, x_3)$  are known for discrete points from the MCS data  $\Omega$  and exhibit strong non-Gaussian characteristics, this may be conveniently achieved using a KDE approach which avoids low order statistical moment inaccuracies with any appropriate software choice such as Matlab, Octave, Python or

```

library(MVN)
omega1 <- read.table(paste(getwd(),'/omega1.txt',sep = ""), header = FALSE, sep = "", dec = ".")
x1 <- omega1[, 1]
x2 <- omega1[, 2]
x3 <- omega1[, 3]
nsample <- 5000
spoints <- sample(1:length(x1),nsample, replace = T)
omegasample <- matrix(nrow = nsample, ncol = 3)
for (i in seq(1, nsample)){
  omegasample[i, 1] <- x1[spoints[i]]
  omegasample[i, 2] <- x2[spoints[i]]
  omegasample[i, 3] <- x3[spoints[i]]
}

> summary1
$multivariateNormality
      Test      Statistic      p value Result
1 Mardia Skewness 648.553976639205 6.87258646542482e-133 NO
2 Mardia Kurtosis 3.26293623835234 0.00110264324993303 NO
3          MVN          <NA>          <NA>      NO

$univariateNormality
      Test Variable Statistic  p value Normality
1 Shapiro-Wilk Column1      0.9641 <0.001      NO
2 Shapiro-Wilk Column2      0.9445 <0.001      NO
3 Shapiro-Wilk Column3      0.9864 <0.001      NO

$Descriptives
      n      Mean Std.Dev  Median    Min    Max   25th   75th   Skew
1 5000 2.276262 0.2652242 2.235216 1.683189 3.312966 2.079300 2.433417 0.7508712
2 5000 3.406918 0.6105936 3.290420 2.384752 6.040160 2.926692 3.784753 0.8709299
3 5000 6.882829 0.4417191 6.842274 5.725811 8.744649 6.573122 7.153655 0.4815667
      Kurtosis
1 0.4388796
2 0.4892281
3 0.2190818

```

**Fig. 5.** Statistical summary of multivariate non-Gaussian characteristics of the Rosenblatt transformation system Monte Carlo data using the Mardia normality test.

R for which KDE software routines and best practises are readily available.

## 4.2 Calculating a Rosenblatt transformation system

The GS2 non-Gaussian data  $\Omega$  as visualized in terms of its joint PDF  $f(x_1, x_2, x_3)$  in Figure 1 has a corresponding GS2 Gaussian joint PDF approximation as shown in Figure 2. This approximation which only utilizes the expected value  $\tilde{\mathbf{y}}$  and covariance  $\mathbf{U}_{\tilde{\mathbf{y}}}$  obtained from  $\Omega$  may be utilized to observe the differences in density variations over the domain. When these quantitative differences are obtained with the aid of the Marsia multivariate normality tests from the previous section, it may be concluded that the variations are too severe for a Gaussian approximation to be utilized as the GS2 data exhibits strong non-Gaussian characteristics. As a result only the actual non-Gaussian joint PDF data will be considered, contrasted and compared with the equivalent Rosenblatt transformation based joint PDF data.

In order to proceed with the analysis the information in  $\Omega$  must first be post-processed in order to calculate the marginal distributions for  $F(x_j)$  and the conditional distributions for  $F(x_j|x_1, \dots, x_{j-1})$  for  $j = 1, 2, 3$  in order to work out the corresponding values of the Rosenblatt transformations. Expanding out it follows that

$$r_1^* = F(x_1^*) \quad (18)$$

$$r_2^* = F(x_2^*|x_1^*) \quad (19)$$

$$r_3^* = F(x_3^*|x_1^*, x_2^*). \quad (20)$$

The application of the Rosenblatt transformation method then reduces to finding appropriate values of  $x_1^*, x_2^*, x_3^*$  that solve the above system of conditional distribution equations for specified random variables  $r_1^*, r_2^*, r_3^* \sim R[0, 1]$  using equation (14), for which explicit forms of these equations are derived in the Appendix.

This can be sequentially achieved by first solving equation (18) which produces the  $x_1^*$  value, then using the known values of  $x_1^*$  in equation (19) which produces the  $x_2^*$  value, and then finally using the known values of

$x_1^*$  and  $x_2^*$  in equation (20) to produce the  $x_3^*$  value so that  $[x_1^*, x_2^*, x_3^*]^T$  corresponds to a sampled point from the  $f(x_1, x_2, x_3)$  distribution. This process may be extended in a straightforward manner for higher-dimensional systems of random variables  $\mathbf{x} \in \mathbb{R}^d$  in measurement models for GS2 Monte Carlo simulation studies where a joint PDF  $f(x_1, \dots, x_d)$  for  $3 \leq d \in \mathbb{N}$  is known.

### 4.3 Calculation of the Kullback-Leibler divergence

Once all of the solved values of  $x_1^*, x_2^*, x_3^*$  have been obtained for a sufficiently large number of points then a new joint PDF  $f_{\mathbf{Y}}(\mathbf{x})$  may be constructed as visualized in Figure 3 in order to compare it with the original joint PDF  $f_{\mathbf{X}}(\mathbf{x})$  as shown in Figure 1. This information may be processed by calculating the Kullback-Leibler divergence  $D_{KL}(\mathbf{X}, \mathbf{Y})$  specified in equation (15) in order to obtain a quantitative indication of the accuracy of the equivalent Rosenblatt transformation system of the original non-Gaussian joint PDF. If the scalar value of  $D_{KL}(\mathbf{X}, \mathbf{Y})$  which is obtained by integrating over the hyper-volume that envelopes all of the generated variables that were used in the construction is sufficiently close to zero, then the equivalence of the original non-Gaussian  $f_{\mathbf{X}}(\mathbf{x})$  and the equivalent Rosenblatt system of mapped variables will have been established.

In order to evaluate the Kullback-Leibler divergence it is necessary to perform a multivariate integral of the form  $\int_{\xi \in \Omega} \phi(\xi) d\xi$  over the space of the random variables. Due to the complexity of the underlying integrand function  $\phi$  this may be more conveniently and simply achieved through the use of a Monte Carlo integration scheme such that

$$\int_V \phi(\xi) d\xi = V \times \frac{\sum_{i=1}^N f(\mathbf{x}_i)}{N}. \quad (21)$$

In the above equation  $N$  is a number of points within the hyper-volume  $V$  that is defined by the collection of all the points  $\mathbf{x} \in \Omega$  such that  $V$  is calculated as

$$V = \prod_{i=1}^d [\max(x_i) - \min(x_i)] \quad (22)$$

where  $x_i$  is the random variable for a dimension obtained from  $\mathbf{x} = [x_1, \dots, x_d]^T$  and  $d$  is the number of dimensions. When this is implemented a final value of  $D_{KL}(\mathbf{X}, \mathbf{Y}) = 0.045199$  results where  $\mathbf{X}$  denotes the actual GS2 distribution and  $\mathbf{Y}$  denotes the equivalent Rosenblatt transformation system distribution.

Since  $D_{KL} \approx 0$  even for a relatively small number of  $n = N = 15000$  Monte Carlo sampled points from the underlying distributions it may reasonably be concluded that the mapped Gaussian variables obtained from the Rosenblatt transformation is indeed equivalent to the original non-Gaussian GS2 distribution. This observation thus provides quantitative proof of the validity of the numerical scheme used to construct the conditional densities using equation (12), equation (13) and equation (14), and also

in addition provides a convenient mechanism to sample from non-Gaussian multivariate distributions.

### 4.4 Analysis of Rosenblatt transformations

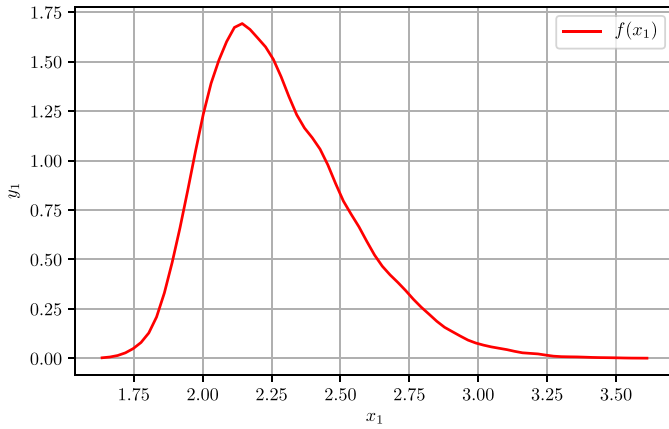
Although a mechanism of sampling from a non-Gaussian joint PDF  $f(\mathbf{x})$  is a convenient result by itself, a further additional benefit of the Rosenblatt transformation is that it allows for the study of uncertainties of random variables based on the coupling effect of other random variables without the need for a complicated determination of the underlying copula. This effect may be investigated by noting that the original joint PDF may be decomposed as  $f(x_1, x_2, x_3) = f(x_1)f(x_2|x_1)f(x_3|x_1, x_2)$  and as a result when the conditional densities  $f(x_2|x_1)$  and  $f(x_3|x_1, x_2)$  are calculated they may then be used to obtain estimates for the corresponding variations in  $x_2$  and  $x_3$  respectively.

Although the variation in  $x_1$  is immediately determined by the marginal distribution  $f(x_1)$ , it is not immediately obvious which values of  $x_1$  to utilize in  $f(x_2|x_1)$  for the variation of  $x_2$ , and similarly which combination of values of  $x_1$  and  $x_2$  to utilize in  $f(x_3|x_1, x_2)$  for the variation of  $x_3$ . Motivated by the earlier discussion of fitting extended lambda distributions for univariate distributions it is noted that quartiles and quantiles can offer a convenient initial choice of limits in the random variables in order to study their coupling effect on the uncertainties of the other random variables.

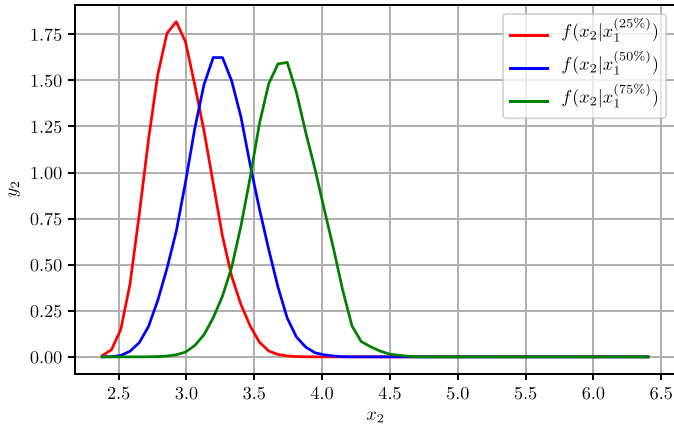
By standard statistical convention, there are three quartiles in any univariate distribution, namely  $Q_1$  which splits the lowest 25% of the data from the highest 75%,  $Q_2$  which simply gives the median value that splits the data in half by separating the lower 50% and upper 50% of the data, and finally  $Q_3$  which splits the highest 25% of the data from the lowest 75% of the data. Quantiles function in a similar manner and split the data into five equal portions, and as a result this suggests that in general percentiles may offer a convenient mechanism to test the “spread” in uncertainty for the previously calculated conditional density distributions  $f(x_2|x_1)$  and  $f(x_3|x_1, x_2)$ . Due to the variety of choices of possible percentiles to consider we restrict the selection to the 25%, 50% and 75% percentiles for brevity, noting that any choice of percentile may be utilized in a conditional density function. Referring to the earlier Mardia normality tests in Figure 5 it follows that  $x_1^{(25\%)} = 2.079300$ ,  $x_1^{(50\%)} = 2.235216$ ,  $x_1^{(75\%)} = 2.433417$ ,  $x_2^{(25\%)} = 2.926692$ ,  $x_2^{(50\%)} = 3.290420$ , and  $x_2^{(75\%)} = 3.784753$  respectively.

When the above choices of limits are incorporated into the Rosenblatt transformed system a set of a marginal density curve for  $f(x_1)$  in Figure 6, and conditional density curves for  $f(x_2|x_1)$  in Figure 7 and  $f(x_3|x_1, x_2)$  in Figure 8 respectively, may be generated in order to determine the “spread” in uncertainties for  $u(x_1)$ ,  $u(x_2)$  and  $u(x_3)$  under different operating conditions. This combination of marginal and conditional densities in terms of simple  $xy$ -graphs may then be used to estimate the “spread” of likely values of the random variables for different regions within  $\Omega$  without the attendant difficulties of

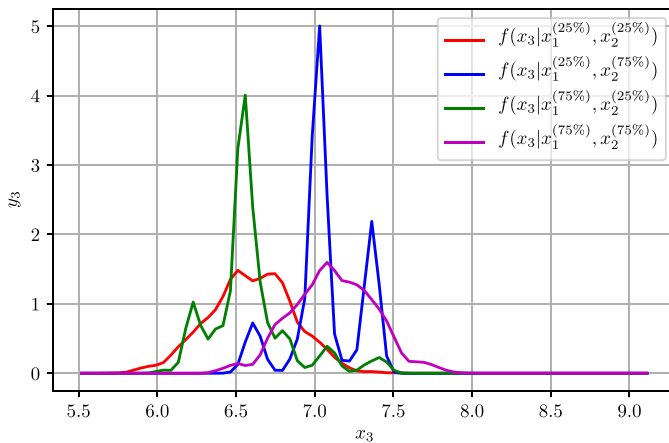




**Fig. 6.** Visualization of the Rosenblatt marginal density  $f(x_1)$ .



**Fig. 7.** Visualization of the Rosenblatt conditional density  $f(x_2|x_1)$ .



**Fig. 8.** Visualization of the Rosenblatt conditional density  $f(x_3|x_1, x_2)$

constructing the copulas and attempting to visualize the coupling effects in higher dimensional spaces.

One particular example of how the Rosenblatt system presents advantages is for the representative pressure

balance problem data-set that was studied when the zero-pressure area  $\eta_{A_0} = x_1$  is known with a greater confidence so that this influences the likely uncertainty of the distortion coefficient  $\eta_\lambda = x_2$ . In this particular case for a trivariate joint PDF where  $\eta_{A_0} = x_1$ ,  $\eta_{\lambda_1} = x_2$  and  $\eta_{\lambda_2} = x_3$ , the conditional densities such as  $f(x_2|x_1^{(50\%)})$  and  $f(x_3|x_1^{(50\%)}, x_2^{(50\%)})$  may be used to study the likely effect on the uncertainties of  $u(x_2) = u(\lambda_1)$  and  $u(x_3) = u(\lambda_2)$ . Similarly, the effect of prior known information from for example earlier instrument calibrations and/or inter-comparison reports may be used as inputs into conditional densities in order to analyse variations in uncertainties.

This functionality thus presents potential new advantages to metrologists for performing advanced uncertainty analyses of laboratory measurement standards and inter-comparisons which was not previously possible for GUM Supplement 2 based Monte Carlo simulations which exhibited non-Gaussian characteristics, and for which it was previously difficult to investigate conditional probabilities for metrology based uncertainty calculations. A potential additional benefit of the new conditional density approach that is implicit within the Rosenblatt transformation system is that it becomes easier to determine possible “double peaks” that may affect uncertainties by inspecting the  $xy$ -graphs, where this information was previously “hidden” in the higher dimensional space.

## 5 Discussion

This paper reported on research work that performed a mathematical analysis with corresponding numerical simulations in order to investigate the practicality of utilizing a mathematically equivalent sequential system of mapped Gaussian PDFs for modelling, summarizing and analysing a non-Gaussian joint PDF in metrology uncertainty analysis studies.

Techniques for conveniently constructing the Rosenblatt transformation system based univariate marginal density and multivariate conditional densities for non-Gaussian distributions with a direct integration approach were therefore considered and investigated in this paper. Results with this approach where the joint PDF was constructed with a kernel density estimate (KDE) modelling scheme of the original GUM supplement 2 data were found to offer a convenient, practical and accurate approach. When this approach was extended it was further determined that it also offered a convenient mechanism for sampling from non-Gaussian distributions.

An additional benefit of a KDE approach over lambda distributions for fitting functions was observed when the variables exhibited multiple peaks and strongly deviated from Gaussian distributions, for which conventional low order statistical moment based lambda distributions have difficulties addressing but which are straightforward for KDE schemes.

Based on the approach developed an alternative technique for estimating the uncertainties of the model results directly in terms of the conditional densities without

the utilization of constructed copulas was considered using representative quartiles and expected values of the random variables. This approach presents a new novel method for metrologists for analysing non-Gaussian PDFs in practical measurement uncertainty problems that may be encountered in high accuracy calibrations and laboratory inter-comparisons.

## 6 Conclusions

Based on the investigation reported in this paper the following conclusions were determined:

- The Rosenblatt transformation offers a convenient mechanism for decomposing a multivariate distribution into an equivalent system of marginal/conditional distributions that can be used to conveniently sample points from higher-dimensional non-Gaussian joint PDFs
- A direct numerical integration of the joint PDF that is approximated with a kernel density estimate (KDE) is recommended for calculating both the marginal and conditional densities
- The marginal and conditional densities may conveniently be used to estimate uncertainties in higher-dimensional multivariate non-Gaussian distributions using sequences of two-dimensional  $xy$ -graphs without the need for constructing parameter and/or empirical based copulas.

## 7 Influences and implications

Based on the research reported in this paper the following influences and implications were determined:

- Metrologists are no longer constrained in summarizing and approximating GUM Supplement 2 measurement uncertainties as multivariate Gaussian distributions if non-Gaussian characteristics are present
- More accurate and non-linear/multi-modal measurement uncertainty effects for metrology applications may now be directly considered using kernel density estimate based schemes without prior linear/uni-modal behavior approximations that was previously necessary with low order statistical moment models such as lambda distributions.

This work was performed with funds provided by the Department of Higher Education, Science and Technology (DHST) on behalf of the South African government for research by public universities.

## Appendix A: Derivation of conditional joint PDFs

Assume that the joint PDF  $f(x_1, x_2, x_3)$  is known and can be computed using either interpolations from scaled histograms or a kernel density estimate function formulation. It immediately follows that the marginal

distributions are

$$f(x_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, u_2, u_3) du_2 du_3 \quad (\text{A.1})$$

$$f(x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u_1, x_2, u_3) du_1 du_3 \quad (\text{A.2})$$

$$f(x_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u_1, u_2, x_3) du_1 du_2. \quad (\text{A.3})$$

The bivariate joint PDF  $f(x_1, x_2)$  may be obtained by integrating out the effect of  $x_3$  such that

$$f(x_1, x_2) = \int_{-\infty}^{\infty} f(x_1, x_2, u_3) du_3. \quad (\text{A.4})$$

The conditional joint PDF  $f(x_2|x_1)$  is then

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)}. \quad (\text{A.5})$$

The conditional joint PDF  $f(x_3|x_1, x_2)$  is then

$$f(x_3|x_1, x_2) = \frac{f(x_1, x_2, x_3)}{f(x_1)f(x_2)}. \quad (\text{A.6})$$

In the above formulae when these formulae are numerically approximated all of the densities should be normalized such that  $\int_{-\infty}^{\infty} f(x_1) dx_1 = 1$ ,  $\int_{-\infty}^{\infty} f(x_2|x_1) dx_2 = 1$  and  $\int_{-\infty}^{\infty} f(x_3|x_1, x_2) dx_3 = 1$  in order to ensure mathematical consistency so that  $0 \leq F(x_1), F(x_2|x_1), F(x_3|x_1, x_2) \leq 1 \forall [x_1, x_2, x_3]^T \in \Omega$ .

As a result the conditional joint distributions are then calculated in terms of the previous conditional densities as

$$F(x_1) = \int_{-\infty}^{x_1} f(u_1) du_1 \quad (\text{A.7})$$

$$F(x_2|x_1) = \int_{-\infty}^{x_2} f(u_2|x_1) du_2 \quad (\text{A.8})$$

$$F(x_3|x_1, x_2) = \int_{-\infty}^{x_3} f(u_3|x_1, x_2) du_3. \quad (\text{A.9})$$

The above mathematical equations are sufficiently general to be utilized in any appropriate numerical integration technique using either commercial or open source software packages.

One particular choice using open source software routines is to implement the above formulae using the Python scipy based KDE function `gaussian_kde` which may be used to calculate the kernel density estimates for one, two and three dimensions depending on the context. If the GS2 data has previously been saved in a text file `Omega.txt` from another simulation, then the marginal density  $f(x_1)$  and the joint PDF  $f(x_1, x_2, x_3)$  may then be conveniently constructed using the following Python computer code:

```

data = np.loadtxt('Omega.txt')
Omega = data
X1 = Omega[:, 0]
X2 = Omega[:, 1]
X3 = Omega[:, 2]
d1 = stats.gaussian_kde(X1)
def f1(x1):
    q = d1(x1)
    return q[0]
d123 = stats.gaussian_kde(Omega.T)
def f123(x1, x2, x3):
    q = d123([x1, x2, x3])
    return q[0]

```

The above formulation when produced using the `stats.gaussian_kde` function from the Python statistics package is useful as it simplifies the programming effort required and allows for a convenient function call to calculate the density  $f(x_1, x_2, x_3)$ . Whilst a simplification in the programming code implementation for producing a Rosenblatt transformation is certainly beneficial, a downside is that this functional approach will tend to slow down a computer code implementation due to the overhead when calling the functions which is disadvantageous when sampling large numbers of random variables.

An alternative approach is to use the existing statistical functionality in R in order to calculate discrete values for an underlying mesh as shown below.

```

library(misc3d)
library(R.matlab)
n1 <- nclass.FD(omega[, 1])
n2 <- nclass.FD(omega[, 2])
n3 <- nclass.FD(omega[, 3])
dens <- kde3d(omega[, 1], omega[, 2],
              omega[, 3], n = c(n1, n2, n3),
              lims = c(min(omega[, 1]),
                        max(omega[, 1]),
                        min(omega[, 2]), max(omega[, 2]),
                        min(omega[, 3]), max(omega[, 3])))

v1 <- dens$x
v2 <- dens$y
v3 <- dens$z
g123 <- dens$d
f1 <- density(omega[, 1])
f2 <- density(omega[, 2])
f3 <- density(omega[, 3])
g1 <- approx(f1$x, f1$y,
             v1, method = 'linear', 0, 0, rule = 2)$y
g2 <- approx(f2$x, f2$y,
             v2, method = 'linear', 0, 0, rule = 2)$y
g3 <- approx(f3$x, f3$y,
             v3, method = 'linear', 0, 0, rule = 2)$y
writeMat(paste(getwd(),
               '/datav1.mat', sep = ""),
         v1 = v1)

```

In the code fragment the Monte Carlo data after being generated and loaded into the work space is processed using the `kde3d` function in R in order to generate the

$f(x)$  discrete points and saved to a `mat` file format. The `mat` file may then be saved to file and loaded in GNU Octave or Matlab in order to implement a pure numerical implementation which depending on how the computer code is generated may be significantly faster.

Once expressions have been obtained to formulate the functional behaviour of the densities  $f(x_1)$ ,  $f(x_1, x_2)$  and  $f(x_1, x_2, x_3)$ , then the final Rosenblatt transformation may be implemented using standard numerical analysis techniques.

## References

1. BIPM, IEC, IFCC, ILAC, ISO, IUPAP, OIML, Tech. Rep., JCGM/WG1 GUM (2008). Revised 1st edition – <https://www.bipm.org/en/publications/guides/>
2. BIPM, IEC, IFCC, ILAC, ISO, IUPAP, OIML, Tech. Rep., JCGM/WG1 GUM Supplement 1 (2008). 1st edition – <https://www.bipm.org/en/publications/guides/>
3. BIPM, IEC, IFCC, ILAC, ISO, IUPAP, OIML, Tech. Rep., JCGM/WG1 GUM Supplement 2 (2011). 1st edition – <https://www.bipm.org/en/publications/guides/>
4. W. Bich, M.G. Cox, C. Michotte, *Metrologia* **53** (5), S149–S159 (2016) DOI: [10.1088/0026-1394/53/5/S149](https://doi.org/10.1088/0026-1394/53/5/S149)
5. P.H. Harris, C.E. Matthews, M.G. Cox, *Metrologia* **51** (3), 243–252 (2014) DOI: [10.1088/0026-1394/51/3/243](https://doi.org/10.1088/0026-1394/51/3/243)
6. V. Ramnath, *Int. J. Metrol. Qual. Eng.* **8** (4), 1–18 (2017) DOI: [10.1051/ijmqe/2016020](https://doi.org/10.1051/ijmqe/2016020)
7. A. Possolo, *Metrologia* **47** (3), 262–271 (2010) DOI: [10.1088/0026-1394/47/3/017](https://doi.org/10.1088/0026-1394/47/3/017)
8. V. Ramnath, *Int. J. Metrol. Qual. Eng.* **8** (29), 1–29 (2017) DOI: [10.1051/ijmqe/2017018](https://doi.org/10.1051/ijmqe/2017018)
9. T. Nagler, C. Czado, J. Multivar. Anal. **151**, 69–89 (2016) DOI: [10.1016/j.jmva.2016.07.003](https://doi.org/10.1016/j.jmva.2016.07.003)
10. J. Segers, M. Sibuya, H. Tsukahara, J. Multivar. Anal. **155**, 35–51 (2017) DOI: [10.1016/j.jmva.2016.11.010](https://doi.org/10.1016/j.jmva.2016.11.010)
11. E. Acar, M. Rais-Rohani, C.D. Eamon, *Int. J. Reliab. Saf.* **4** (2/3), 166–187 (2010) DOI: [10.1504/IJRS.2010.032444](https://doi.org/10.1504/IJRS.2010.032444)
12. S. Rahman, H. Xu, *Prob. Eng. Mech.* **19** (4), 393–408 (2004) DOI: [10.1016/j.pro bengmech.2004.04.003](https://doi.org/10.1016/j.pro bengmech.2004.04.003)
13. G.L. Bretthorst, *AIP Conf. Proc.* **1553** (3), 3–15 (2013) DOI: [10.1063/1.4819977](https://doi.org/10.1063/1.4819977)
14. N. Armstrong, G.J. Sutton, D.B. Hibbert, *Metrologia* **56** (1), 015019 (15pp) DOI: [10.1088/1681-7575/aa7d11](https://doi.org/10.1088/1681-7575/aa7d11)
15. M. Rosenblatt, *Ann. Math. Stat.* **23** (3), 470–472 (1952) DOI: [10.1214/aoms/1177729394](https://doi.org/10.1214/aoms/1177729394)
16. R. Lebrun, A. Dutfoy, *Prob. Eng. Mech.* **24** (4), 577–584 (2009) DOI: [10.1016/j.pro bengmech.2009.04.006](https://doi.org/10.1016/j.pro bengmech.2009.04.006)
17. S.J. van Albada, P.A. Robinson, *J. Neurosci. Methods* **161** (2), 205–211 (2007) DOI: [10.1016/j.jneumeth.2006.11.004](https://doi.org/10.1016/j.jneumeth.2006.11.004)
18. K.H. Chang, *Reliability analysis*, in *e-Design Computer-Aided Engineering Design*, edited by K.H. Chang (Academic Press, 2015), chap. 10, pp. 523–595
19. Mathworks, Tech. Rep., Mathworks (2019), <https://www.mathworks.com/help/stats/prob.normaldistribution.icdf.html>
20. J.W. Eaton, Tech. Rep., Sourceforge (2019), <https://octave.sourceforge.io/nan/function/norminv.html>
21. T.E. Oliphant, Tech. Rep., SciPy.org (2019), <https://docs.scipy.org/doc/numpy/user/index.html>
22. D.W. Scott, S.R. Sain, in *Handbook of statistics – Data mining and data visualization*, edited by C.R. Rao, E.J.

- Wegman, J.L. Solka (Elsevier, Oxford, 2005), Chap. 9, pp. 229–261
23. R. Willink, *Metrologia* **46** (3), 154–166 (2009) DOI: [10.1088/0026-1394/46/3/002](https://doi.org/10.1088/0026-1394/46/3/002)
  24. Z.A. Karian, E.J. Dudewicz, P. McDonald, *Commun. Stat. Simul. Comput.* **25** (3), 611–642 (1996) DOI: [10.1080/03610919608813333](https://doi.org/10.1080/03610919608813333)
  25. C.G. Corlu, M. Meterelliyozy, *Commun. Stat. Simul. Comput.* **45** (7), 2276–2296 (2015) DOI: [10.1080/03610918.2014.901355](https://doi.org/10.1080/03610918.2014.901355)
  26. B. Wang, Tech. Rep., CRAN (2019), <https://cran.r-project.org/web/packages/gb/index.html>
  27. T. Nagler, C. Schellhase, C. Czado, *Depend. Models* **5** (1), 99–120 (2017) DOI: [10.1515/demo-2017-0007](https://doi.org/10.1515/demo-2017-0007)
  28. T. Nagler, U. Schepsmeier, J. Stoeber, E.C. Brechmann, B. Graeler, T. Erhardt, C. Almeida, A. Min, C. Czado, M. Hofmann, M. Lilliches, H. Joe, T. Vatter, Tech. Rep., CRAN (2019) <https://cran.r-project.org/web/packages/VineCopula/index.html>
  29. S. Yadav, V.K. Gupta, A.K. Bandyopadhyay, *MAPAN* **26** (2), 133–151 (2011) DOI: [10.1007/s12647-011-0014-5](https://doi.org/10.1007/s12647-011-0014-5)
  30. A. Azzalini, A. Capitanio, *J. Roy. Stat. Soc. Ser. B* **61** (3), 579–602 (1999) DOI: [10.1111/1467-9868.00194](https://doi.org/10.1111/1467-9868.00194)
  31. A. Azzalini, Tech. Rep., CRAN (2019) <https://cran.r-project.org/web/packages/sn/index.html>
  32. V. Ramnath, *MAPAN* **34** (3), 387–402 (2019) DOI: [10.1007/s12647-019-00324-w](https://doi.org/10.1007/s12647-019-00324-w)
  33. A. Ghasemi, S. Zahediasl, *Int. J. Endocrinol. Metab.* **10** (2), 486–489 (2012) DOI: [10.5815/ijem.3505](https://doi.org/10.5815/ijem.3505)
  34. K.V. Mardia, *Biometrika* **57** (3), 519–530 (1970) DOI: [10.1093/biomet/57.3.519](https://doi.org/10.1093/biomet/57.3.519)
  35. S. Korkmaz, D. Goksuluk, G. Zararsiz, Tech. Rep., CRAN (2019) <https://cran.r-project.org/web/packages/MVN/index.html>

**Cite this article as:** Vishal Ramnath, Analysis of approximations of GUM supplement 2 based non-Gaussian PDFs of measurement models with Rosenblatt Gaussian transformation mappings, *Int. J. Metrol. Qual. Eng.* **11**, 2 (2020)