Numerical analysis of the accuracy of bivariate quantile distributions utilizing copulas compared to the GUM supplement 2 for oil pressure balance uncertainties

Vishal Ramnath[*]

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

Abstract. In the field of pressure metrology the effective area is $A_e = A_0 \left(1 + \lambda P\right)$ where A_0 is the zero-pressure area and λ is the distortion coefficient and the conventional practise is to construct univariate probability density functions (PDFs) for A_0 and λ. As a result analytical generalized non-Gaussian bivariate joint PDFs has not featured prominently in pressure metrology. Recently extended lambda distribution based quantile functions have been successfully utilized for summarizing univariate arbitrary PDF distributions of gas pressure balances. Motivated by this development we investigate the feasibility and utility of extending and applying quantile functions to systems which naturally exhibit bivariate PDFs. Our approach is to utilize the GUM Supplement 1 methodology to solve and generate Monte Carlo based multivariate uncertainty data for an oil based pressure balance laboratory standard that is used to generate known high pressures, and which are in turn cross-floated against another pressure balance transfer standard in order to deduce the transfer standard's respective area. We then numerically analyse the uncertainty data by formulating and constructing an approximate bivariate quantile distribution that directly couples A_0 and λ in order to compare and contrast its accuracy to an exact GUM Supplement 2 based uncertainty quantification analysis.

Keywords: pressure balance / uncertainty / copula / bivariate quantile function / GUM Supplement 2

1 Introduction

In the field of pressure metrology when piston-cylinder operated pressure balances are used to measure applied pressures defined as $P = F/A_e$ where F is a generalized force and A_e is the effective area, the most widespread model to characterize a pressure balance is that of a two parameter model such that $A_e = A_0 \left(1 + \lambda P\right)$ where A_0 is the zero-pressure area and λ is the distortion coefficient. As a result knowledge of both the expected values as well as the associated covariance of a pressure balance's effective area model parameters A_0 and λ are intrinsically necessary in order to fully quantify the uncertainty in A_e for subsequent application in pressure generation and measurement tasks. The conventional practise by many metrologists when performing an uncertainty quantification (UQ) analysis of a piston-cylinder operated pressure balance is to characterize the model parameters A_0 and λ based on an analysis of the combination of the available experimental data such as cross-floating measurements and theoretical data such as finite element simulations. This

analysis is usually performed with the application of the classical sensitivity coefficient based formulation of the GUM [1], or alternatively with the GUM Supplement 1 (GS1) Monte Carlo based UQ technique [2] in cases where the measurand model is either too complex or too nonlinear.

When these UQ techniques are applied the resulting multivariate UQ data for the pressure balance model, either implicitly in terms of a multivariate Gaussian distribution if generated with the GUM or potentially in terms of a multivariate non-Gaussian distribution if generated with the GS1, is then typically approximated with univariate probability density function (PDF) distributions for A_0 and λ due to the complexity of the models, whilst the covariance information through the available experimental and theoretical available data in terms of the coupling between A_0 and λ is then usually either neglected, estimated or occasionally indirectly approximated.

Univariate PDFs of measurands have traditionally been analytically specified either as Gaussian or Student t-distributions in the case of a GUM based analysis, or alternately through discrete representations of either the underlying distribution function or equivalently the PDF

* ramnav@unisa.ac.za

when the uncertainty analysis was performed using the GS1 approach. In GS1 based uncertainty analysis approaches generated Monte Carlo data has in a majority of cases been utilized for subsequent analysis by retrieving previously generated and stored numerical data from electronic files. Recent work by Harris et al. [3] considered the problem of how to analytically represent GS1 numerical data by investigating and studying different families of quantile functions and approaches in order to estimate the respective quantile function parameters. Their investigations demonstrated the utility of extended lambda distribution (ELD) based quantile function representations to accurately summarize arbitrary PDFs generated using Monte Carlo simulation data. This methodology has recently been applied by Ramnath [4] in the field of gas pressure metrology through the use of an ELD based quantile function for the characterization of a pressure balance's effective area in the low pressure range $5\,\mathrm{kPa} \leq p \leq 7\,\mathrm{MPa}$ where the distortion coefficient λ is considered negligible and the pressure balance characteristics is fully specified in terms of just a univariate PDF of the zero-pressure area A_0. In this particular investigation a non-continuum gas flow was present and as a result the GUM was not appropriate due to the presence of a highly non-linear measurand model. The measurand model of the gas pressure balance's effective area which was numerically solved with a corresponding uncertainty PDF determined through a GS1 Monte Carlo methodology demonstrated the feasibility of utilizing quantile functions for characterizing univariate PDFs which may exhibit deviations from more traditional Gaussian shaped deviations such as asymmetry and/or skewness in more complex fluid based primary scientific measurement systems.

Unfortunately for pressure balances operated at higher pressures where the effect of the distortion coefficient is significant due to fluid structural interaction effects the uncertainties and correlation effects between A_0 and λ for hydraulic oil operated gauge pressure balances in the medium pressure range $7\,\mathrm{MPa} \leq P \leq 30\,\mathrm{MPa}$ must be considered. These effects are considered particularly critical and essential in the high pressure range $30\,\mathrm{MPa} \leq P \leq 500\,\mathrm{MPa}$ when calculating the uncertainty of a pressure balance's effective area. As a result based on this observation from an UQ perspective it is considered desirable and beneficial if a pressure balance model can be characterized directly in terms of a joint bivariate PDF in order to preserve intrinsic statistical uncertainty analysis information that may utilized be in order to accurately determine and quantify the magnitudes of the correlation effects that are present in physical pressure balances operated at high hydraulic pressures where there is strong coupling between A_0 and λ.

Based on these motivating factors our approach in this paper is to firstly utilize an implicit multivariate GUM methodology which is used to solve for and characterize generated applied pressure data values and uncertainties for an oil pressure balance laboratory standard by sampling from appropriate PDF and joint PDF distributions where available. We then utilize the known generated pressure values and uncertainties as inputs into a full GS1 Monte Carlo simulation which is used to numerically generate multivariate data for the cross-floating of a transfer standard. Once the multivariate data for the transfer standard has been generated we then post-process it using the GS2 [5] methodology for further analysis in order to determine the actual ellipsoidal and smallest coverage regions for the coupling of the transfer standard's zero-pressure area A_0 and distortion coefficient λ in a joint PDF. The exact coverage regions obtained with the GS2 methodology are then compared with results obtained with approximate bivariate quantile distributions that we formulate in this paper in order to investigate and determine the potential accuracy of bivariate quantile distributions for modelling pressure balance characteristics in terms of joint PDFs.

2 Mathematical models

The mathematical model for a pressure balance using a working fluid in a liquid state may be formulated as a pressure measurand equation through the analysis of a free-body diagram such that the applied pressure $P = (p - p_a)$ is of the form

$$p - p_a = \frac{1}{S}\left(\left\{\sum_{i=1}^{n} m_i\left(1 - \frac{\rho_a}{\rho_i}\right) - V_s(\rho_f - \rho_a)\right.\right.$$
$$\left.\left. + H(\rho_f - \rho_a)S\right\}g + \sigma C\right) \tag{1}$$

where p is the working fluid pressure, and p_a is the surrounding ambient pressure as per the original formulation by Dadson et al. [6]. In this model m_i and ρ_i are the actual mass and density values for the system of weights used to generate the respective pressures if there are n total mass pieces, ρ_a is the ambient density, V_s is the submerged volume of the piston into the working fluid, ρ_f is the density of the working fluid at an actual pressure of p, g is the local gravitational acceleration constant, σ is the surface tension of the working fluid where C is the circumference of the wetted perimeter of the piston in contact with the working fluid, H is a hydrostatic height term to account for any potential differences in elevation between the reference datums of generated pressues and measured pressures, and S is the effective area of the pressure balance. In the absence of additional information a useful approximation that may be used to estimate the circumference is $C = \sqrt{4\pi S}$.

Due to the fact that physical measurements may occur in which the temperature of the pressure balance may vary through a combination of ambient conditions and operating conditions the effective area is usually adjusted to correspond to a known reference temperature such that $S = A_0\left[1 + \lambda P\right]\phi\left(t, t_{ref}\right)$ and $\phi\left(t, t_{ref}\right) = 1 + \alpha\left(t - t_{ref}\right)$. As per the discussion by Dadson et al. α is the sum of the thermal expansion coefficients of the piston and cylinder so that $\alpha = (\alpha_p + \alpha_c)$, and the reference temperature is fixed to a constant value which is usually just set to $t_{ref} = 20\,^\circ C$ in many commonwealth countries, and to $23\,^\circ C$ in other countries. Following the general approach by many pressure metrologists the ambient air density may be specified using the CIPM-2007 formula as $\rho_a = \frac{pM_a}{ZRT}$

$\left[1 - x_v\left(1 - \frac{M_v}{M_a}\right)\right]$ as discussed by Picard et al. [7] where $M_a = 28.96546 \times 10^{-3}\,\mathrm{kg\,mol^{-1}}$ is the molar mass of dry air, $M_v = 18.01528 \times 10^{-3}\,\mathrm{kg\,mol^{-1}}$ is the molar mass of water, Z is the air compressibility, $R = 8.314472\,J\mathrm{mol^{-1}}$ $\mathrm{K^{-1}}$ is the CODATA-2006 recommended value for the universal gas constant, and x_v is the mole fraction of water vapour for the corresponding pressure, temperature and relative humidity conditions for the surrounding ambient air. The previous formulae may now be combined and rearranged as a nonlinear function $f^{\star}(p, q_1^{\star}, \ldots, q_m^{\star})$ where p is the unknown working fluid pressure p which is to be determined and $q_j^{\star}$, $j = 1, \ldots, m$ are laboratory standard parameters such that

$$f^{\star}(p, q_1^{\star}, \ldots, q_m^{\star}) = \frac{1}{S^{\star}}\left[\sum_{i=1}^{n} m_i^{\star} g\left(1 - \frac{\rho_a}{\rho_i^{\star}}\right)\right.$$
$$\left. + g[H^{\star}S^{\star} - V_s^{\star}][\rho_f^{\star} - \rho_a] + \sigma^{\star}C^{\star}\right] - (p - p_a) \quad (2)$$

The generated pressure p of the working fluid from a laboratory standard may then be obtained by solving $f^{\star}(p, q_1^{\star}, \ldots, q_m^{\star}) = 0$ with any suitable solver by noting that ρ_f, C and S formally depend on the pressure p whilst all the other terms are independent of the working fluid's pressure. The above calculations are performed if we assume that the air density $\rho_a = f_a(p_a, t_a, h_a)$ and fluid density $\rho_f = f_f(p, t)$ are calculated by known functions in terms of the independent variables, and where $f^{\star}(p, q_1^{\star}, \ldots, q_m^{\star})$ formally has units of applied pressure which allows for the numerical solution to be obtained for any specified applied pressure accuracy level.

A simple approach to solve for the unknown generated pressure p is to assume that $p_{\min} \leq p \leq p_{\max}$ and search within this pressure range where $p_{\min}$ and $p_{\max}$ are rough estimates of the nominal pressure. As an example for an applied pressure of 250 MPa with a nominal atmospheric pressure of $p_a = 101.325\,\mathrm{kPa}$ first calculate the approximate pressure as $p_0 = P + p_a$ and then for example set $p_{\min} = 0.95p_0$ and $p_{\max} = 1.05p_0$ as specifications to search within this interval or alternately to solve the full non-linear solver of $f^{\star}(p, q_1^{\star}, \ldots, q_m^{\star}) = 0$ using p_0 as a starting solution. The next issue which naturally arises is how to solve this particular non-linear equation noting that for the generating pressure equation both the density ρ_f and the wetted circumference $C = \sqrt{4\pi S}$ depend on the pressure p. Ideally we would use for example Newton's method $x_{n+1} = x_n - [f(x_n)]/[f'(x_n)]$, $n = 0, 1, 2, \ldots$ for successive approximations to solve $f(x) = 0$ where $f'(x_n)$ denotes the derivative of the function where the benefits of using an iterative solver is that the generating pressure may be calculated to any desired numerical accuracy level. In our particular case it is however cumbersome to analytically calculate the derivative due to the choice of the fluid's equation of state. Fortunately there are derivative free iterative formulae as reported by Dehghan and Hajarian [8] to solve a non-linear equation $f(x) = 0$ and we may to use a forward difference formula originally developed by Steffensen such that $x_{n+1} = x_n - [f(x_n)]^2/[f(x_n + f(x_n))] - f(x_n)]$. As a result through the use of the Steffensen formula we may control the numerical

precision of the solved for applied pressures P_k, $k = 1$, $\ldots$, n_p where n_p are the number of generated pressures by specifying the number of iterations.

In pressure metrology practise there is generally only have a limited indirect level of covariance information available and it is therefore common to disregard covariance terms in the $f^{\star}(p, q_1^{\star}, \ldots, q_m^{\star}) = 0$ model by setting $\mathrm{cov}(q_i, q_j) = 0$, $i \neq j$ in the generated pressure calculation. The only potential exception where $\mathrm{cov}(q_i, q_j) \napprox 0$ $i \neq j$ is in terms of the correlation between A_0 and λ i.e. for $\mathrm{cov}(A_0, \lambda)$, and in the correlations between the weights and densities used i.e. $\mathrm{cov}(m_i, m_j)$, $\mathrm{cov}(\rho_i, \rho_j)$ and $\mathrm{cov}(m_i, \rho_j)$ respectively for $i \neq j$ unless otherwise specified.

The issue of how to approximate the correlation for weights was previously investigated by Bich [9] who utilized the multivariate GUM approach of Cox and Harris [10] in a slightly modified equation for a mass measurement as a general implicit equation. Later investigations by Palencar et al. [11] followed a similar multivariate GUM framework but refined the analysis slightly by using a least squares formulation to deduce the optimal values of the parameters in a calibration of mass pieces. Their investigation concluded that covariance terms can actually contribute non-negligible terms in the uncertainty analysis of mass measurements and that ideally covariance information should be included in mass measurement calibration certificates.

Whilst the approach of Palencar et al. [11] does allow for the covariance information to be directly estimated under certain restricting assumptions for mass pieces traceable through a sequence of measurements to a particular country's national kilogram prototype standard or equivalent national metrology institure primary mass standard, in this paper we will disregard the correlation effects between mass and density terms used by the set of weights in the laboratory standard pressure generation model for simplicity. This modelling simplification is due to the fact that in most practical cases calibrated mass certificates do not provide any covariance information for the weights used by a pressure balance. In addition for simplicity we will also disregard the covariance between the laboratory standard zero-pressure area and distortion coefficient $\mathrm{cov}(A_0, \lambda)$ in the pressure balance that is used to generate the known applied pressure due to that fact that many existing pressure calibration certificates do not generally provide any specific explicit covariance information.

Our approach of disregarding the covariance term $\mathrm{cov}(A_0, \lambda)$ for the generated pressure in the absence of specific known covariance information is consistent with experimental recommendations for national metrology institutes for pressure balances operated in free deformation mode as documented in for example EURAMET CG3 [12], and results from theoretical finite element based studies for pressure balances operated in both free deformation and controlled clearance modes such as the EUROMET Project 463 by Sabuga et al. [13] and in controlled clearance modes only by Dogra et al. [14] respectively. As a result of our choice to disregard covariance effects in the generated pressure calculation when implementing a GS1 Monte

Carlo approach as discussed by Cox and Siebert [15] for the pressure generation calculations, it will then usually suffice to sample with appropriate numerical techniques for the inputs q_i from univariate PDFs $g_i(\xi_i)$ associated with the corresponding input parameters q_i where ξ_i is a random variable of q_i unless information of a joint PDF is specifically available. Unless otherwise specified the PDFs for the input parameters will usually be Gaussian distributions however following the more recent work by Harris et al. it is now in principle possible to model the inputs as arbitrary PDFs using a univariate ELD quantile distribution as originally investigated by Willink [16].

We comment that even though we disregard covariance terms in the inputs of the generated pressure balance model in the absence of available information that when these generated pressures are used as inputs in the cross-floated pressure balance model to deduce the corresponding effective areas that due to the complexity of the mathematical model the covariance information in the parameters used to model the effective areas will still be present in an implicit form of the associated Monte Carlo data of the cross-floated effective areas.

Once the generated pressures from the laboratory standard (LS) pressure balance model are obtained and the GS1 information appropriately processed, the generated pressure PDFs $g_i^\star(\xi_i^\star)$ where $\xi_i^\star$ is a random variable for generated pressure p_i may then used as inputs for when the LS is cross-floated against another pressure balance such as a unit-under-test or transfer standard (TS) in order to determine the other pressure balance's effective area. When two pressure balances are cross-floated against each other the point of equilibrium occurs when the applied pressure $P = (p - p_a)$ in both pressure balances are equal in which case the measurand equation for a cross-floated TS is

$$f(S, q_1, \ldots, q_m) = \frac{1}{(p - p_a)} \left[\sum_{i=1}^{n} m_i g \left(1 - \frac{\rho_a}{\rho_i} \right) \right.$$
$$\left. + g[HS - V_s][\rho_f - \rho_a] + \sigma C \right] - S \qquad (3)$$

In the above equation S is the unknown TS effective area which must be determined, and q_j, $j = 1, \ldots, m$ are corresponding parameters associated with the cross-floated TS pressure balance measurand equation. The benefit of this particular equation formulation is that $f(S, q_1, \ldots, q_m)$ formally has units of cross-floated area which then allows for the numerical solution accuracy to be controlled by specifying the number of iterations. Whilst a simple estimate to solve for the generated pressure is to assume a range of possible generated pressure values using for example $0.95 p_0 \leq p \leq 1.05 p_0$ where p_0 is the approximate nominal pressure $p_0 = P + p_a$ as previously discussed such a straightforward estimate of the range of possible cross-floated areas of the TS is not possible since we have no prior information in order to determine a possible range of values $S_{\min} \leq S_0 \leq S_{\max}$ to search within.

A simple initial approximation to overcome this difficulty is to ignore the effect of the oil surface tension and TS submerged volume on the TS cross-floated area so that a rough cross-floated area estimate is $S_0 \approx [\sum_{i=1}^{n} m_i g \left(1 - \frac{\rho_a}{\rho_i} \right)]/[(p - p_a) - Hg(\rho_f - \rho_a)]$ since the hydro-

static pressure head term would not be negligible and to then use a search range of say $0.95 S_0 \leq S \leq 1.05 S_0$ in order to solve the full non-linear equation $f(S, q_1, \ldots, q_m) = 0$ or alternately by using for example the Steffensen formula as previously discussed.

In this paper the $\star$ super-scripts signifies terms specific to the LS pressure balance which is used to generate the pressure whilst the unstarred terms represent quantities that are specific to the TS pressure balance which is cross-floated in order to determine corresponding effective areas. The model $f(S, q_1, \ldots, q_m) = 0$ using the previous generated pressure PDF $g_{P_k}(\xi_{P_k})$ inputs may then be solved in another GS1 simulation to obtain the PDF distribution $g_{S_k}(\eta_{S_k})$ for the cross-floated effective areas S_k for each pressure P_k where η_{S_k} is a random variable of the cross-floated effective area S_k.

Since the cross-floated effective areas S_k are univariate data it follows that an ELD-QF distribution may also be used to conveniently summarize S_k in terms of parameters $\{a_k, b_k, c_k, d_k\}$ for $k \in [1, \ldots, n_P]$ for each of the n_P cross-float effective area measurements corresponding to the n_P generated pressures where as previously discussed the cross-floated effective area is modelled in terms of a two parameter model such that $S = A_0 (1 + \lambda P)$ where P is the applied pressure.

For our particular pressure balance problem a GS1 based uncertainty analysis is utilized for conceptual simplicity for the LS generated pressures in order to generate data for further analysis when the LS is cross-floated against the TS pressure balance as this avoids some of the subjectivity in certain aspects of the correlation effects modelling in the GUM based matrix analysis.

The GS1 methodology is focused on multivariate input models with an input $X = [X_1, \ldots, X_N]^T$ and a single output $Y = [Y_1]$ defined in terms of an implicit equation $h(y, x) = 0$ such that the expected value $\tilde{y}$, variance $u^2(\tilde{y})$ and distribution function $G_Y(\eta)$ are approximated using M Monte Carlo simulations such that $h(y_r, x_r) = 0$ where $r = 1, 2, \ldots, M$, $\tilde{y} = \frac{1}{M} \sum_{r=1}^{M} y_r$, $u^2(\tilde{y}) = \frac{1}{M-1} \sum_{r=1}^{M} (y_r - \tilde{y})^2$, $\tilde{G}_Y(\eta) = \frac{r - \frac{1}{2}}{M} + \frac{\eta - y_r}{M(y_{r+1} - y_r)}$ where $y_r \leq \eta \leq y_{r+1}$ for $r = 1, 2, \ldots, M - 1$. When the GS1 methodology is implemented for the LS generated pressure P_k, $k = 1, \ldots, n_P$ there will be a set of Monte Carlo univariate simulation data $P_k = \{P_k^{(1)}, \ldots, P_k^{(M)}\}$ for each generated pressure which may then be used as an input for the cross-float of the TS.

A GS2 methodology on the other hand when contrasted to the GS1 method also considers the case for an input $\boldsymbol{X}$ but where there are now multiple outputs such that $Y = [Y_1, \ldots, Y_m]^T$ defined in terms of an implicit vector equation $h(y, x) = 0$. This methodology is also implemented by generating Monte Carlo data by first sampling from the input vector PDF for $x_r \sim g_X(\xi)$ or relevant associated joint PDFs if this information is available such that $h(y_r, x_r) = 0$ for $r = 1, 2, \ldots, M$, $\tilde{y} = \frac{1}{M} [y_1 + \cdots + y_M]$, $U_{\tilde{y}} = \frac{1}{M-1} [(y_1 - \tilde{y}) \cdot (y_1 - \tilde{y})^T + \ldots + (y_M - \tilde{y}) \cdot (y_M - y)^T]$ and $G = [y_1, \ldots, y_M]$ where M is the number of Monte Carlo simulation events in order to produce a dataset $\boldsymbol{y}_r$. In our particular case when the GS2 methodology is implemented for the TS cross-floated area S_k, $k = 1, \ldots, n_P$ there will be a set of Monte Carlo bivariate simulation data $S_k = \{[(A_0)_k^{(1)}, \lambda_k^{(1)}]^T, \ldots, [(A_0)_k^{(M)}, \lambda_k^{(M)}]^T\}$.

When the GS2 uncertainty analysis method is implemented and the Monte Carlo simulation data post-processed an implicit multivariate model may be used to construct the average vector value $\tilde{y}$ and corresponding covariance matrix $U_{\tilde{y}}$. As a result when using a GS2 approach the method as per the office $y \sim \mathcal{N}(\mu, V)$ implicit assumption that the output follows a multivariate Gaussian distribution such that $y \sim \mathcal{N}(\mu, V)$ where the expected value is $\mu \approx \tilde{y}$ and the covariance matrix is $\sum \approx U_{\tilde{y}}$ unless otherwise specified. This feature is common to both the GS1 and GS2 due to the extreme mathematical complexity and numerical challenges in directly solving the Markov integral which provides for the exact PDF or joint PDF without any assumptions for both univariate as well as multivariate models. For the case of the GS2 the joint PDF for a multivariate Gaussian may then be written in matrix notation such that

$$g_X(\xi) = \frac{\exp\left(-\frac{1}{2}(\xi - \mu)^T V^{-1}(\xi - \mu)\right)}{\sqrt{(2\pi)^N |V|}} \qquad (4)$$

where $\mathbf{V}^{-1}$ is the inverse of the matrix $\mathbf{V}$, and $|V| = det(V)$ denotes the determinant of the matrix $\mathbf{V}$. Using this notation in the case of a bivariate distribution with an input $X = [X_1, X_2]^T$ and a corresponding random variable $\xi = [\xi_1, \xi_2]^T$ the expected vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$ may then be specified as

$$\mu = \begin{bmatrix} \mu_{X_1} \\ \mu_{X_2} \end{bmatrix} \qquad (5a)$$

$$V = \begin{bmatrix} \sigma_{X_1}^2 & \rho\sigma_{X_1}\sigma_{X_2} \\ \rho\sigma_{X_1}\sigma_{X_2} & \sigma_{X_2}^2 \end{bmatrix} \qquad (5b)$$

where $\sigma_{X_1}^2$ and $\sigma_{X_2}^2$ are the variances for X_1 and X_2 respectively, and ρ is an indication of the covariance between the inputs X_1 and X_2. The use of a Pearson r instead of ρ i.e. Spearman's rho as an indication of the correlation is usually the more common approach in mechanical metrology problems however other approaches such as Kendall's tau are also possible. The main advantage of these approaches are that the quantification of the extent of the correlation between the random variables x_1 and x_2 do not depend of the choice of parametrization for the model. Some additional benefits of using either a Kendall's tau or Spearman's rho instead of the Pearson's correlation coefficient are that these options are usually a more accurate indicator of correlation when the random variables X_1 and X_2 do not follow a Gaussian joint PDF since a Pearson's correlation can sometimes give misleading results if the random variables do not follow multivariate normal distributions as discussed in more detail by Goda [17]. If $(x_1, y_1), ..., (x_n, y_n)$ are a set of n observations then Kendall's tau τ_K may be calculated as

$$\tau_K = \frac{C - D}{n(n-1)/2} \qquad (6)$$

where C is the number of concordant pairs and D is the number of discordant pairs for the set of observations where a particular pair (x_i, y_i) and (x_j, y_j) for $i \neq j$ is

concordant if $x_i > x_j$ and $y_i > y_j$ or alternately if $x_i < x_j$ and $y_i < y_j$, whilst the same pair is discordant if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. In the special case if $x_i = x_j$ and $y_i = y_j$ then the pair is neither concordant nor discordant and does not contribute in the calculation of τ_K. The use of Kendall's tau τ_K provides a useful simplification later in the paper for constructing particular formulations of copulas.

The benefit of using a GS2 approach to directly calculate the cross-floated effective areas is that any covariance information is immediately present and available from the covariance matrix $U_{\tilde{y}}$, and as a result this approach avoids any additional modelling assumptions for covariance terms. We will utilize the above $S_k = \{[(A_0)_k^{(1)}, \lambda_k^{(1)}]^T, \ldots, [(A_0)_k^{(M)}, \lambda_k^{(M)}]^T\}$ multivariate dataset in order to investigate the accuracy of a bivariate quantile distribution in approximating a joint PDF for A_0 and λ.

Once the univariate generated pressure Monte Carlo data $\wp_k$ has been generated it is then necessary to summarize them. In the case of a univariate PDF $g(\eta)$ with a random variable η and output Y one particular definition of defining a quantile function $Q(\rho)$ following the approach of Harris et al. is by setting the distribution function as $G(\eta) = \rho$ where $0 \leq \rho \leq 1$ such that the quantile function $Q(\rho)$ is defined such that $\eta = Q(\rho) \equiv G^{-1}(\rho)$ from which it may be demonstrated that $g(\eta) = 1/[Q'(\rho)]$. The practical consequences of this definition is that if ρ is sampled from the rectangular distribution $R(0, 1)$ then $Q(\rho)$ corresponds to a sampling from $g(\eta)$, $[Q(p_1), Q(p_2)]$ is a probabilistically symmetric coverage interval for p if $p_1 = \frac{1}{2}(1 - p)$ and $p_2 = \frac{1}{2}(1 + p)$, and that the expectation and variance for Y may be calculated as $E(Y) = \int_0^1 Q(\rho)\,d\rho$ and $V(Y) = \int_0^1 [Q(\rho) - E(Y)]^2 d\rho$ respectively. For the univariate case where a quantile function distribution following the approach of Willink [16] utilizing an ELDs is implemented it has been demonstrated by Willink that the corresponding QF and ELD-QF for the PDF may be simply represented in terms of real parameters a, b, c and d. Technical details to determine the values of a, b, c and d for an ELD were discussed in full and complete detail by Willink who compared the method of using four moments, four quantiles, and a modified method using the mean, variance and two quantiles. For pressure metrology purposes the method of using four quantiles is considered relatively simple and straightforward for pressure balances with univariate effective area models such as gas operated pressure balances at low pressure as discussed by Ramnath [4]. The approach of using quantile functions such as ELDs to summarize Monte Carlo uncertainty analysis data was further developed by Harris et al. [3] where they extended the earlier work of Willink for univariate distributions of a quantity Y to summarizing an arbitrary parent probability distributions denoted as P in terms of a generalized quantile function. This approach involved the utilization of the distribution function $G_P(\zeta)$ and PDF $G_P(\zeta)$ where $\zeta \in Z \sim P$ is a random variable for the parent probability distribution and $\eta \in Y$ is a random variable for the output Y. According to this approach where $\eta = F(\zeta)$ the random variable is calculated as $\eta = F(G_P^{-1}(\rho))$ where $\rho \in U \sim$

$R(0, 1)$ is a random variable following a rectangular distribution and the PDF of Y is calculated as $g(\eta) = [g_P(\zeta)]/[F'(\zeta)]$. The key simplification of this approach is that the generalized quantile function defined in terms of $F(\zeta)$ simply requires knowledge of $F(\zeta)$ and a mechanism of sampling random variables from the parent distribution $\zeta \in Z \sim P$. In this paper we will opt for this general approach of Harris et al. since the ELD approach of Willink using four parameters to model the PDFs of the generated pressures and cross-floated areas may not necessarily be adequate. For simplicity we will use a random variable $\rho \in R(0, 1)$ following a rectangular distribution as this is convenient with readily available random number generators. Although it is possible to use an increasing number of parameters over and above the four parameters in for example an ELD based QF through the calculation of higher order moments by working out for example the coefficients of B-spline approximations for $F(\zeta)$ with open source software routines [18] our approach will simply directly utilize the actual GS1 distribution functions for each of the generated pressures and cross-floated areas when constructing the joint PDF.

Earlier approaches such as that by Ramnath [19] utilized statistical linear regression analysis techniques developed by Krystek and Anton [20] for straight line data with correlation in plots of applied pressure $[P \pm u(P)]$ data versus cross-floated effective area $[S \pm u(S)]$ data with associated error bars. In this approach the generated pressure and cross-floated areas uncertainties were calculated with the classical sensitivity coefficient based GUM and the straight line fit for the curve $S = A_0(1 + \lambda P)$ allowed for A_0 and $A_0\lambda$ to be estimated, and which were then in turn used to estimate the distortion λ and indirectly estimate the correlation between A_0 and λ. In this paper our approach is different since our objective is to instead construct an approximation to the actual joint PDF $g_X(\xi)$ with $X = [A_0, \lambda]^{\mathrm{T}}$ so that we may directly sample random values ξ_{A_0} and ξ_λ from the underlying joint bivariate PDF similar to how univariate random variables ξ_P may be sampled from the generated pressure PDF.

The GS2 documentation specifically focuses on multivariate Gaussian distributions $x \sim \mathcal{N}(\mu, \mathrm{V})$ for a model with an output x, an expected value μ and an uncertainty matrix $\mathbf{V}$. Explicit guidelines for sampling for such a joint PDF are provided such that a sampled draw may be estimated as $\xi = x + \mathrm{R}^{\mathrm{T}}z$ where $x = [\mu_1, ..., \mu_N]^{\mathrm{T}}$, and $z_i \sim \mathcal{N}(0, 1)$ if $X = [X_1, ..., X_N]^{\mathrm{T}}$ and $\mathbf{R}$ is an upper triangular matrix formed from a Cholesky decomposition such that $\mathrm{U}_x = \mathrm{R}^{\mathrm{T}}\mathrm{R}$. This focus is achieved through the assumptions that $\mu \approx \tilde{y}$ and $\mathrm{V} \approx \mathrm{U}_x$, however in general it may not necessarily be the case that a multivariate Gaussian is the best choice of PDF.

Although there may be some merits in particular cases of modelling the multivariate Monte Carlo data in terms of non-Gaussian distributions the benefits of doing so may be outweighed by the necessity of selecting, deciding on and choosing alternative probability distribution information formulations as discussed by Tang et al. [21] who studied these effects in certain applications using the equivalent probability distribution information contained in appropriate choices of copulas. The theoretical approach of

constructing bivariate distributions as discussed by Tang et al. for constructing the joint cumulative distribution function (CDF) and PDF of variables X and Y is possible if the marginal distributions of X say $F(x)$ and that of Y say $G(y)$ where x and y are corresponding random variables are both known. If a particular choice of copula function is specified then the application of Sklar's theorem as discussed by Tang et al. allows for the joint PDF $f(x, y)$ to be constructed as

$$f(x, y) = F(x)G(y)c(F(x), G(y); \theta) \tag{7a}$$

$$c(F(x), G(y); \theta) = \frac{\partial^2 C}{\partial u \partial v} \tag{7b}$$

where $u = F(x)$, $v = G(y)$ and the copula is defined as $C: I \times I \to I$ where $c(F(x), G(y))$ is known as the copula density and $I = [0, 1]$. The copula C occasionally written as $C_\theta(u, v)$ where θ is a fixed parameter is formally defined as a mapping from the unit square I^2 to the unit interval I. In this formulation θ is a parameter associated with the choice of the copula function for the underlying data specific to the particular model data as discussed by Genest and Favre [22]. One particular example of a copula are those belonging to the Farlie–Gumbel–Morgenstern family such that $C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v)$ for $u, v \in [0, 1]$ where the copula is constrained by the Frechet–Hoeffding bounds such that $W(u, v) \leq C(u, v) \leq M(u, v)$ where $W(u, v) = \max\{0, u + v - 1\}$ and $M(u, v) = \min\{u, v\}$. These bounds are used if X and Y are not independent, whilst a choice of $C = \Pi = uv$ may be used if for example X and Y are fully independent where the degree of independence between the variables X and Y may be estimated in terms of either the Spearman's rho or Kendall's tau values as previously discussed. In the particular case of bivariate distributions the formal definition of a copula is that of a mapping such that

$$C: I^2 \to I \tag{8a}$$

$$C(0, x) = C(x, 0) = 0 \tag{8b}$$

$$C(1, x) = C(x, 1) = x \quad \forall x \in I \tag{8c}$$

$$\begin{aligned} 0 &\leq [C(b, d) + C(a, c) - C(a, d) \\ &- C(b, c)] \quad \forall a, b, c, d \in I, a \\ &\leq b \, \text{and} \, c \leq d \end{aligned} \tag{8d}$$

Different types of copulas are considered later in the paper when we construct the exact joint PDF from a GS2 Monte Carlo simulation and utilize various choices of copulas in approximating the actual bivariate joint PDF in order to investigate the suitability of various choices of bivariate quantile constructions of which copulas are one particular choice amongst several for modelling and summarizing a pressure balance's joint PDF.

Regardless of the particular choice of copula $C_\theta(u, v)$ it is seen that the construction of the joint PDF essentially reduces to that of the choice of a mapping function $\Phi I^2 \to I$.

This choice of mapping function is necessary in order to relate how variables $0 \le p, r \le 1$ may be used as inputs for the mapping function to generate the corresponding random variables consistent with the joint PDF $f(x, y)$. As a result although the absence of guidance from the GS2 for sampling from non-Gaussian multivariate PDFs is potentially problematic this aspect is not an issue since it turns out that sampling from a bivariate quantile function based distribution may still be achieved using sampling from associated univariate PDFs and two dimensional coordinate transformation mappings.

In this paper our main motivation is to summarize the GS2 based joint PDF using a bivariate quantile smoothing spline that was originally developed by He et al. [23] where a response surface Z depends on two variables X and Y where it is assumed that the observations z_{ij} is known at each (x_i, y_j) for $i = 1, \dots, m$ and $j = 1, \dots, n$ such that $x_1 < \dots < x_m$ and $y_1 < \dots < y_n$ for convenience. In the original paper by He et al. it was demonstrated that the optimal solution for fitting the surface and they considered the special case where the covariates X_1 and X_2 were in the domain $[0, 1] \times [0, 1]$ which corresponds to our particular problem for random variables $0 \le p, r \le 1$.

One approach to construct the joint PDF is based on the direct use of the Markov formula as discussed by Cox and Siebert [15] where the PDF denoted as $g_Y(\eta)$ for a model $Y = f(X)$ is formally defined in terms of the joint PDF denoted as $g_X(\xi)$ as $g_Y(\eta) = \int_{-\infty}^{\infty} g_X(\xi)\delta(\eta - f(\xi)) \, \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_N$. For our model the cross-floated area of the pressure balance is simply $S = A_0(1 + \lambda P)$ so this may be formally expressed as $S \sim g_S(\eta)$ so that

$$g_S(\eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_{A_0,\lambda}(\xi_{A_0}, \xi_\lambda) g_P(\xi_P) \times \delta(\eta - f(\xi_{A_0}, \xi_\lambda, \xi_P)) \, \mathrm{d}\xi_{A_0} \, \mathrm{d}\xi_\lambda \, \mathrm{d}\xi_P \quad (9)$$

In the above formulation $g_{A_0,\lambda}(\xi_{A_0}, \xi_\lambda)$ is the unknown joint PDF which is not necessarily a multivariate Gaussian where the applied pressure is independent since in practical terms it can be varied by simply setting the working fluid pressure. As a result the PDF for the pressure can be uncoupled from the joint PDF for the area. It follows that the above continuous multivariate integral may then be approximated as a system of simultaneous equations such that

$$g_{S_k}(\eta_{S_k}^{(\ell)}) \approx \sum_{i=1}^{N_{A_0}} \sum_{j=1}^{N_\lambda} \sum_{r=1}^{N_P} \left[g_{A_0,\lambda}\left(\xi_{A_0}^{(i)}, \xi_\lambda^{(j)}\right) g_{P_k}\left(\xi_{P_k}^{(r)}\right) \right.$$
$$\left. \delta\left(\eta_\ell - \xi_{A_0}^{(i)}\left\{1 + \xi_\lambda^{(j)}\xi_P^{(r)}\right\}\right) \right], \quad k = 1, \dots, n_P \text{ and } \ell \in [1, \dots, N_\eta] \quad (10)$$

Whilst the Markov formula may in principle be used to calculate the exact joint PDF it is however not generally feasible in practical terms due to finite arithmetic precision resolution errors. These issues are due to the large difference in magnitudes from the cross-floated area PDF $g_S(\eta)$, the joint PDF $g_{A_0,\lambda}(\xi_{A_0}, \xi_\lambda)$ and pressure PDF $g_P(\xi_P)$ which when utilized to build up an equivalent

system to solve the Markov integral equation will result in an ill-conditioned linear system of the form $\mathrm{A}x = B$ where the coefficient matrix $\mathbf{A}$ is built up in terms of the PDF $g_P(\xi_P)$, the known vector $\boldsymbol{B}$ is built up in terms of the PDF $g_S(\eta)$ and the unknown x is the values of the joint PDF $g_{A_0,\lambda}(\xi_{A_0}, \xi_\lambda)$ at fixed coordinates for a chosen grid of ξ_{A_0} and ξ_λ points. Extensions beyond standard IEEE 32-bit and 64-bit accuracies to variable precision arithmetic (VPA) accuracies using Fortran/C++ based implementations to mitigate against finite arithmetic resolution errors are discussed in more detail by Bailey and Borwein [24].

Our approach in this paper is to instead use the GS2 approach for multivariate measurand models of the form $h(y, x) = 0$ for an input $\boldsymbol{x}$ and output $\boldsymbol{y}$. If $x \sim g_X(\xi)$ and $y \sim g_Y(\eta)$ where $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are random variables then model must also satisfy $h(\eta, \xi) = 0$. This means that if the model undergoes a Monte Carlo simulation then $\boldsymbol{\eta}$ can also be post-processed in order to determine its probability distribution as per the GS2 documentation if $\boldsymbol{\eta}$ can be recovered from the equation $h(\eta, \xi) = 0$.

The utilization of the GS2 to determine A_0 and λ where in our case $\eta = [A_0, \lambda]^{\mathrm{T}}$ is slightly more complicated since the pressure balance model for the TS $f(S, q_1, \dots, q_m) = 0$ is not in the standard form where there is an explicit system of equations for the parameters A_0 and λ. If we however consider the entire set of cross-float measurements as one system we may then construct a new vector equation $h(y, x) = 0$ by employing a unweighted linear least squares which is discussed in more technical detail by White and Saunders [25].

For our purposes we will utilize the well known results for a general linear unweighted least squares problem where the function is constructed as $y(x) = \sum_{i=1}^{N} a_i X_i(x)$ in terms of an independent variable x, and parameters $a_1, \dots, a_N$ where $X_i(x)$ are specified basis functions as discussed by Press et al. [26]. Following this approach a merit function is constructed as $\chi^2 = \sum_{j=1}^{M} \left[y_j - \sum_{i=1}^{N} a_i X_i(x) \right]^2$ and minimized by calculating the parameter values to satisfy $\partial \chi^2 / \partial a_i = 0$, $i = 1, \dots, N$. Implementing this approach to our particular problem where n_P is the number of generated pressures and associated cross-floats with $S = A_0(1 + \lambda P)$ then yields

$$\chi^2 = \sum_{j=1}^{n_P} [S_j - \{A_0(1 + \lambda P_j)\}]^2 \quad (11)$$

so that the simultaneous system of equations $\partial \chi^2 / \partial A_0 = 0$ and $\partial \chi^2 / \partial \lambda = 0$ may then be used to implement the GS2 methodology where

$$h(y, x) = \begin{bmatrix} \dfrac{\partial \chi^2}{\partial A_0} \\ \dfrac{\partial \chi^2}{\partial \lambda} \end{bmatrix} \quad (12a)$$

$$\frac{\partial \chi^2}{\partial A_0} = \sum_{j=1}^{n_P} \left\{ -2A_0(1 + \lambda P_j)[S_j - A_0(1 + \lambda P_j)] \right\} \quad (12b)$$

$$\frac{\partial \chi^2}{\partial \lambda} = \sum_{j=1}^{n_P} \left\{ -2A_0 P_j[S_j - A_0(1 + \lambda P_j)] \right\} \quad (12c)$$

where in our case $y = [A_0, \lambda]^T$ and $x = [q_1, \ldots, q_m]^T$ as per our earlier mathematical modelling approach. In practical terms the approach used would be to sample random variables $q_1^\star, \ldots, q_m^\star$ in order to solve the LS equation $f^\star(p, q_1, \ldots, q_m) = 0$ for the generated pressure, and to then use this as input to solve the TS equation $f(S, q_1, \ldots, q_m) = 0$ for the cross-floated area S. As a result there will be a set of cross-floated areas $\mathcal{S} = [S_1, \ldots, S_M]^T$ for the M Monte Carlo simulation events where each simulation event will have a known applied pressure. Since for each simulation event the solved for pressures and cross-float areas are statistically valid it follows that the model would also need to be satisfied for these values.

As a result the GS2 Monte Carlo simulation of unweighted least squares of statistically sampled values in accordance with the underlying probablity distributions then becomes equivalent to a conventional multivariate regression practice as discussed by Press et al. [26] since the sampled values are formally a statistically valid possibility based on the underlying probability distribution. This system may then be used to determine a Hyper-ellipsoidal coverage region which in our particular problem will correspond to a two dimensional region for the bivariate joint PDF.

In order to have a better conceptual mathematical understanding of the statistical definition of a quantile we first make the geometric observation that in the case of a univariate PDF $g(\eta)$ that the corresponding univariate quantile $Q_{1D}(\rho)$ function is technically a one dimensional mapping

$$Q_{1D} : \rho \rightarrow \eta \, , \text{s.t.} \, \rho \in R(0, 1), \quad \eta \in g(\xi)$$

that transforms a random variable ρ sampled from a rectangular distribution into a corresponding random variable ξ such that the sampled value is $\eta = g(\xi)$. We may then using this conceptual tool in this paper then use the approach of Gilchrist [27] to generalize the definition of a bivariate quantile $Q_{2D}([\rho_1, \rho_2]^T)$ as a corresponding two dimensional mapping

$$f(x, y) = \frac{1}{\|\frac{\partial x}{\partial p}\frac{\partial y}{\partial r} - \frac{\partial x}{\partial r}\frac{\partial y}{\partial p}\|}$$

In particular if when implementing the previously discussed geometrical mapping mathematical definition for bivariate quantiles by taking $h(p, r) = 1$ where p and r are random variables such that $0 \leq p \leq 1$ and $0 \leq r \leq 1$ a natural consequence which results is that the bivariate quantile may be considered to be the mapping from the unit rectangle to the corresponding surface of the joint PDF $f(x, y)$ formally defined following the approach by Gilchrist as

$$f(x, y) = \frac{1}{\|\frac{\partial x}{\partial p}\frac{\partial y}{\partial r} - \frac{\partial x}{\partial r}\frac{\partial y}{\partial p}\|} \tag{13a}$$

$$= [h(p, r)] \cdot \|J(p, r | x, y)\| \tag{13b}$$

In the above formula $J(p, r | x, y) = 1/[J(x, y | p, r)]$ is the corresponding Jacobian of the transformation from the unit rectangle $h(p, r) = 1$ to the joint PDF surface $f(x, y)$.

The random variables x and y as outlined above may be assumed to be equivalent to $x \equiv \xi_{A_0}$ and $y \equiv \xi_\lambda$ for brevity later in the paper where they correspond to random variables associated with the zero-pressure area A_0 and distortion coefficient λ respectively for our particular pressure balance cross-floated effective area model. If an appropriate two dimensional mapping is correctly chosen then the variables p and r may be independently sampled from a rectanguar distribution $R(0, 1)$ and then the choice of mapping will allow for random variables to be sampled from the corresponding bivariate joint PDF. This purely mathematical transformation result is consistent with the earlier statistical based observation of the need for a choice of copula that may be used with the marginal distributions in order to construct the joint PDF.

As a result of this observation we note that a bivariate quantile is not necessarily unique as there are in principle an arbitary number of mappings between the two surfaces however if we fix the type of mapping which is equivalent to a particular choice of copula then we can construct a unique associated bivariate quantile distribution. Our approach in this paper is to use the multivariate GS1 methodology to generate the actual bivariate joint distribution of the pressure balance TS $g_X(\xi)$ where $X = [A_0, \lambda]^T$ and $\xi = [\xi_{A_0}, \xi_\lambda]^T$ are associated random variables sampled from a PDF corresponding to **X**. The actual joint PDF $f(\xi_{A_0}, \xi_\lambda)$ surface constructed out of the GS1 Monte Carlo simulation data is then numerically approximated with a bivariate PDF $\varphi(\xi_{A_0}, \xi_\lambda)$ using a selection of two dimensional transformation mappings. The accuracy of the bivariate quantile distribution may then be investigated by determining how the corresponding effective area uncertainty calculated in terms of $\varphi(\xi_{A_0}, \xi_\lambda)$ compares to the actual effective area uncertainties obtained through the actual GS1 Monte Carlo simulation cross-floating data.

The previous definition of bivariate quantiles by Gilchrist was developed in more formal mathematical statistical details in an earlier work by Chaudhuri [28] and may be formally extended to systems other than bivariate distributions with potential application to metrology problems involving higher dimensional non-Gaussian multivariate distributions. One particular model discussed by Gilchrist is the generalized circular model of form

$$Q_x(p, r) = \lambda_x + \eta_x \sqrt{Q(p; \beta)}\cos(2\pi r) \tag{14a}$$

$$Q_y(p, r) = \lambda_y + \eta_y \sqrt{Q(p; \beta)}\sin(2\pi r) \tag{14b}$$

where $0 \leq p, \ r \leq 1$ as previously discussed and β is an appropriate constant parameter that must be determined. The above system of equations is a particular form for the general bivariate quantile transformation where each random variable has a corresponding equation. Another strategy that may be used to construct bivariate quantiles with more generality than the generalized circular model is that of the marginal/conditional form where

$$x = Q_x(p) \tag{15a}$$

$$y = Q_y(p, r) \tag{15b}$$

In this formulation the first function $x = Q_x(p)$ is a univariate quantile function, whilst in the second function $y = Q_y(p, r)$ if x is fixed then p is in turn fixed so that the second function is actually also a univariate quantile function for y for a given choice of x. The practical consequence of choosing to model a bivariate quantile in terms of a marginal/conditional formulation is that the previously developed ELD formulation can in principle also be utilized to construct the associated level curves, however a potential drawback of a conventional marginal distribution approximation of a bivariate joint PDF is that in general only a fixed number of contour curves can be analytically constructed and hence interpolation is required in order to calculate random variables for arbitrary choices of specified values of p and r. Whilst marginal distributions for the zero-pressure area A_0 and distortion coefficient λ can be constructed relatively easily using for example extended lambda based univariate quantile distributions a newer approach of fitting the actual bivariate joint PDFs with copulas is now available. In the copula approach one uses known specified marginal distributions which are relatively easy to calculate, and then optimizes parameters $\theta = [\theta_1, \ldots, \theta_n]^T$ associated with the particular choice of copula family $C_\theta(u, v)$ in order more closely match the copula constructed joint PDF with the actual known joint PDF. As a result fitting copulas to joint PDFs can potentially offer considerable algebraic simplifications where for bivariate distributions the main choices that may be investigated are usually elliptical copulas and generalized Archimedean copula families.

Whilst these particular copula based approaches have their own respective merits we comment that a generalized multivariate quantile function approach originally developed by Chaudhuri [28] may also be implemented for bivariate distributions. The starting point is to generate multivariate data points $X_1, X_2, \ldots, X_n \in \mathbb{R}^d$, $2 \leq d \in \mathbb{Z}$ in $\mathbb{R}^d$ which are assumed to be known through for example Monte Carlo simulations. For our purposes this Monte Carlo data will be generated with the GS2 approach as discussed earlier with $h(y, x) = 0$ which will formally provide a sequence of data points $y_1, \ldots, y_M \in \mathbb{R}^2$ where M is the number of Monte Carlo simulation events and $y_q = \left[\xi_{A_0}^{(q)}, \xi_\lambda^{(q)} \right]^T$, $q = 1, \ldots, M$ and where we make the observation that we do not assume that $\boldsymbol{\mu} \approx \boldsymbol{y}$ and $\sum \approx U_{\boldsymbol{y}}$ but simply assume that $y_1, \ldots, y_M$ follows some underlying probability distribution which may not necessarily be Gaussian.

By defining the set as $B^{(d)} = \{u : u \in \mathbb{R}^d, \|u\| < 1\}$ i.e. $B^{(d)}$ is an open unit ball, and by defining a functional as $\Phi(u, t) = \|t\| + \langle u, t \rangle$ where $u \in B^{(d)}$, $t \in \mathbb{R}^d$ where $\langle \cdot, \cdot \rangle$ represents the Euclidean inner product, a multivariate quantile function may then be formally mathematically defined as

$$\hat{Q}_n(u) = \min_{Q \in \mathbb{R}^d} \sum_{i=1}^n \Phi(u, X_i - Q) \qquad (16)$$

The above is a multivariate generalization for $2 \leq d \in \mathbb{Z}$ of the univariate case corresponding to $d = 1$ where the univariate quantile maps values of α for $0 \leq \alpha \leq 1$ with an associated parameter $u = (2\alpha - 1)$ to the one dimensional interval $(-1, 1)$.

According to the above multivariate definition extreme quantiles would correspond to $\|\boldsymbol{u}\| \approx 1$ and central quantiles would correspond to $\|\boldsymbol{u}\| \approx 0$ respectively. An iteration algorithm in order to construct the quantile function was originally provided by Chaudhuri where for $X_1, \ldots, X_n$ the algorithm Step #1 is to compute for all of the corresponding data points X_i for $1 \leq i \leq n$ and check if the degeneracy condition

$$\begin{aligned} | \sum_{j=1, j \neq i}^n \{\|X_j - X_i\|^{-1}(X_j - X_i)\} + (n-1)u| \\ \leq (1 + \|u\|) \end{aligned} \qquad (17)$$

is valid for some $1 \leq i \leq n$. If the degeneracy condition is satisfied for some i then just set $\hat{Q}_n(u) = X_i$. Alternately if the degeneracy condition cannot be satisfied for any $1 \leq i \leq n$ then move to Step 2 by constructing $\hat{Q}_n(u)$ by solving the equation

$$\sum_{i=1}^n \{|X_i - \hat{Q}_n(u)|^{-1}[X_i - \hat{Q}_n]\} + nu = 0 \qquad (18)$$

The above equation must be solved by using for example the method of successive approximations of which one possible approach is to use a starting solution composed of a vector of means such that

$$\hat{Q}_n^{(1)} \approx [\mu_1, \ldots, \mu_n]^T \qquad (19)$$

where each mean is simply the mean of the respective set of components. Once this starting value has been constructed successive approximations for $m = 2, 3, \ldots$ may then be generated using iterations so that

$$\hat{Q}_n^{(m+1)}(u) = \hat{Q}_n^{(m)}(u) + \Phi^{-1}\Delta \qquad (20)$$

where

$$\Phi = \sum_{i=1}^n \Big(|X_i - \hat{Q}_n^{(m)}|^{-1}[I_d - |X_i - \hat{Q}_n^{(m)}|^{-2} \\ \{X_i - \hat{Q}_n^{(m)}\}\{X_i - \hat{Q}_n^{(m)}\}^T]\Big) \qquad (21)$$

$$\Delta = \sum_{i=1}^n |X_i - \hat{Q}_n^{(m)}|^{-1}\Big\{X_i - \hat{Q}_n^{(m)}\Big\} + n\boldsymbol{u} \qquad (22)$$

and I_d is the $d \times d$ identity matrix. Further aspects of how this definition of quantile function may be used to quantify and describe skewness and kurtosis in higher dimensional multivariate distributions is discussed in more detail in the original paper by Chaudhuri [28].

This geometric generalization of a quantile function in higher dimensional spaces may thus have some further technical potential for analysing metrology systems than cannot be modelled in terms of either univariate and bivariate probability distributions as a future topic of research study by metrologists particularly when non-Gaussian multivariate distributions may occur. In the context of pressure metrology such potential future research problems might include higher dimensional

multivariate joint PDF's such as for pressure balances in free deformation mode or controlled clearance modes. Examples of higher dimensional multivariate probability distributions for pressure balances in free deformation mode operation include situations when non-linear elasticity theory is utilized and two distortion coefficient parameters are then necessary so that the effective area is modelled as $A = A_0 \left(1 + \lambda_1 p + \lambda_2 p^2\right)$ and a joint PDF $g_{A_0, \lambda_1, \lambda_2}(\xi_{A_0}, \xi_{\lambda_1}, \xi_{\lambda_2})$ must be constructed. On the other hand for pressure balances operated in controlled clearance mode such as when the Heydemann–Welch method is implemented for a primary standard scale realization for hydraulic pressures and the pressure balance effective area is modelled as $A = A_0 \left(1 + \alpha \left(t - t_{ref}\right)\right) \left(1 + \lambda P\right)\left[1 + d_j \left(p_{j0} - p_j\right)\right]$ as discussed by Kajikawa et al. [29,30] additional parameters such as the jacket pressure coefficient d_j and zero-clearance jacket pressure p_{j0} are then present in which case a higher dimensional joint probability distribution $g_{A_0, \lambda, d_j, p_{j0}}(\xi_{A_0}, \xi_\lambda, \xi_{d_j} \xi_{p_{j0}})$ must be approximated to fully characterize a pressure balance.

3 Numerical simulations

In this paper we perform an analysis by utilizing the experimental data set previously reported by Ramnath [19] as indicated in Appendix A which provides full technical details and supporting data where the working fluid is assumed to be Di(2)-ethyl-hexylsebacate defined in terms of an oil equation of state reported by Kocas et al [31]. Numerical experiments were performed on a Toshiba laptop with an Intel Pentium B950 CPU operating at 2.10 GHz with 2 GB of RAM running on a Microsoft Windows 7 64-bit operating system using Gnu Octave 4.2.0. Simulations were undertaken with $M = 10\,000$ Monte Carlo simulation events for each generated pressure and cross-floated area data point. The full GS1 simulation for ten generated pressures and cross-floats from 50 MPa to 500 MPa therefore consisted of $10\,000 \times 10 = 100\,000$ total simulation events and took approximately 115 minutes to solve. When the GS2 was implemented by post-processing the GS1 data by using the cross-float data to fit a curve through the data points $[P_1, \ldots, P_{10}]^{\mathrm{T}}$ and $[S_1, \ldots, S_{10}]^{\mathrm{T}}$ in order to construct the χ^2-optimization for the best fit curve $S = A_0 \left(1 + \lambda P\right)$ the equivalent number of $M = 10\,000$ simulation events took less than a minute to solve when using the built-in polyfit routine of Gnu Octave. As a result Monte Carlo simulations for $10 \times 10^3 \leq M \leq 25 \times 10^3$ simulation events may be is considered a feasible option with single-core laptops/ workstations as these calculations would typically be able to performed in a single working day. On the other hand if higher counts of simulation events in the range $25 \times 10^3 \leq M \leq 100 \times 10^3$ are required we comment that this would in most practical circumstances require access to either multi-core computers or alternatively high performance computing parallel computing computers or clusters to avoid simulation times of a few weeks.

Our approach in this paper as outlined earlier in the paper first solves the underlying equations for the generated pressure $f^\star(p, q_1^\star, \ldots, q_m^\star) = 0$ and cross-floated area $f(S, q_1, \ldots, q_m) = 0$ using sampled random values $\boldsymbol{q}^\star$

and $\boldsymbol{q}$ which are drawn from the respective input PDFs which produce Monte Carlo GS1 data for the generated pressure $\mathrm{MCP} = [P_1, \ldots, P_{10}]$ and cross-floated area $\mathrm{MCS} = [S_1, \ldots, S_{10}]$ which are both $M \times 10$ matrices since each generated pressure is used to cross-float the TS. The Monte Carlo data of these matrices are built up in terms of column vectors of the form $P_k = [\xi_{P_k}^{(1)}, \ldots, \xi_{P_k}^{(M)}]^{\mathrm{T}}$ and $S_k = [\eta_{S_k}^{(1)}, \ldots, \eta_{S_k}^{(M)}]^{\mathrm{T}}$ where each column contains the Monte Carlo data for that particular generated pressure or cross-floated area so that the matrix data is then stored in data files for convenience. The distribution functions $G_{P_k}(\xi_{P_k})$ and $G_{S_k}(\eta_{S_k})$ may if necessary be calculated and summarized in terms of ELDs built up in terms of fours parameters a, b, c, and d such that $\eta = Q(\rho)$ where

$$Q(\rho) = \begin{cases} d + \dfrac{c}{b}[a\rho^b - (1-\rho)^b + 1 - a], & b \neq 0 \\ d + c[a\ln(\rho) - \ln(1 - \rho)], & b = 0 \end{cases} \quad (23a)$$

$$g(\eta) = \left\{ c\left[a\rho^{b-1} + (1-\rho)^{b-1} \right] \right\}^{-1} \quad (23b)$$

and $0 \leq \rho \leq 1$ is a random variable to generate the PDFs as originally discussed by Willink [16]. We comment that the generated pressures and cross-floated areas may if necessary also be calculated directly as $\mu_{P_k} = \int_0^1 Q_{P_k}(\rho)\,\mathrm{d}\rho$ and $\sigma_{P_k}^2 = \int_0^1 (Q_{P_k} - \mu_{P_k})^2\,\mathrm{d}\rho$ where Q_{P_k} is the quantile function for pressure P_k as summarized in Table 1 with similar expressions for the cross-floated areas as summarized in Table 2.

Once the expected values μ_{P_k} and μ_{S_k} are calculated this set of GS1 data which is consistent with the associated underlying PDFs may then be used to estimate the approximate nominal zero-pressure area μ_{A_0} and distortion coefficient μ_λ from the pressure versus area graph as illustrated in Figure 1a which is based on a GS1 simulation with $M = 10\,000$ total simulation events as previously discussed. When this original GS1 data is then further processed in the GS2 simulation using the original GS1 data in the χ^2 optimization to extract the values for A_0 and λ it then results in the scatter plot as illustrated in Figure 1b which may then be further post-processed in order to construct the bivariate joint PDF.

An approximate visualization of the bivariate joint PDF using a kernal density estimate (KDE) approach along with visualizations of the associated marginal distributions for A_0 and λ are illustrated in Figure 1c using a Gaussian kernal.

Our approach to construct the bivariate PDF uses the built-in histogram2d function from the Python numpy and scipy scientific computing libraries for convenience since we wish to fit the actual joint PDF $f(x, y)$ where x and y are random variables. This choice of using histogram2d to explicitly build up a discrete approximation of the actual PDF with explicit bin counts that are appropriately weighted for a two dimensional normalization of the joint PDF is used over a KDE approach since the implementation of a KDE implicitly requires a choice of kernal function as discussed by Diwekar and David [32] where the PDF for univariate data $X_1, \ldots, X_n$ is constructed such that $f(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$ where h is the bandwidth and K

Table 1. Summary of ELD parameters for generated applied pressures P_k/[Pa] for $k \in [1, ..., 10]$. The index k indicates the pressures P_k from $P_1 = 50$ MPa through to $P_{10} = 500$ MPa so that the extended lambda distribution parameter a_k, b_k, c_k and d_k can be used to summarize the probability density function distribution for P_k.

k	a_k	b_k	c_k	d_k	μ_k	σ_k
1	9.3820626800E−001	1.3981985600E−001	1.7989736500E+003	4.9846059000E+007	4.9884703500E+007	2.5334462000E+003
2	9.5043178500E−001	1.7397850600E−001	3.7529649900E+003	9.9764041800E+007	9.9763883300E+007	5.0638896000E+003
3	1.0415714700E+000	1.5838605300E−001	5.2460166500E+003	1.4963997400E+008	1.4963978500E+008	7.5752196400E+003
4	9.8571523900E−001	1.4237782100E−001	7.1297741600E+003	1.9951174900E+008	1.9951183900E+008	1.0247772600E+004
5	1.0117402400E+000	1.4394410000E−001	8.7463788100E+003	2.4938016600E+008	2.4938007600E+008	1.2707088000E+004
6	1.0190847600E+000	1.4582407200E−001	1.0800616100E+004	2.9924473600E+008	2.9924455600E+008	1.5705943200E+004
7	1.0062382000E+000	1.2392021700E−001	1.2333996600E+004	3.4910510600E+008	3.4910503700E+008	1.8404232700E+004
8	9.8810397400E−001	1.6710459100E−001	1.5405216600E+004	3.9896251000E+008	3.9896266700E+008	2.1394147200E+004
9	1.0091905000E+000	1.5889771100E−001	1.7023106000E+004	4.4881671600E+008	4.4881658100E+008	2.4173042100E+004
10	9.6490400800E−001	1.5693017600E−001	1.9758337100E+004	4.9866538100E+008	4.9866598000E+008	2.7516728000E+004

Table 2. Summary of ELD parameters for generated areas S_k/[m²] for $k \in [1, ..., 10]$. The index k indicates the area S_k corresponding to the cross-floated pressures from $P_1 = 50$ MPa through to $P_{10} = 500$ MPa so that the extended lambda distribution parameter a_k, b_k, c_k and d_k can be used to summarize the probability density function distribution for S_k.

k	a_k	b_k	c_k	d_k	μ_k	σ_k
1	1.0644929800E+000	1.4094731700E−001	6.6609457400E−011	1.9615108500E−006	1.9615070800E−006	9.9751602100E−011
2	1.0357306900E+000	1.6799847700E−001	7.0124204300E−011	1.9616024900E−006	1.9616003500E−006	9.9593893700E−011
3	9.5004383900E−001	1.5462372500E−001	7.1808028900E−011	1.9616757400E−006	1.9616788500E−006	9.9579873300E−011
4	9.8612791000E−001	1.4724149200E−001	7.0151858100E−011	1.9617509800E−006	1.9617510700E−006	1.0077042300E−010
5	9.8377620500E−001	1.4480840700E−001	6.9866561300E−011	1.9618245500E−006	1.9618255400E−006	9.9968476500E−011
6	9.8618925800E−001	1.4526018300E−001	7.1943126600E−011	1.9618980700E−006	1.9618989400E−006	1.0299717600E−010
7	9.6417022200E−001	1.2782696900E−001	7.0093421000E−011	1.9619690400E−006	1.9619692700E−006	1.0347690300E−010
8	1.0292755700E+000	1.6451349400E−001	7.3959082800E−011	1.9620409700E−006	1.9620391100E−006	1.0522668900E−010
9	9.6877215700E−001	1.6122393100E−001	7.6162710100E−011	1.9621033000E−006	1.9621053500E−006	1.0562591100E−010
10	1.0341895200E+000	1.4988393500E−001	7.4577940000E−011	1.9621833400E−006	1.9621811200E−006	1.0862103400E−010

(a) Cross-float data of expected values   **(b)** Scatter plot of GS2 data   **(c)** Bivariate KDE & marginal distribution
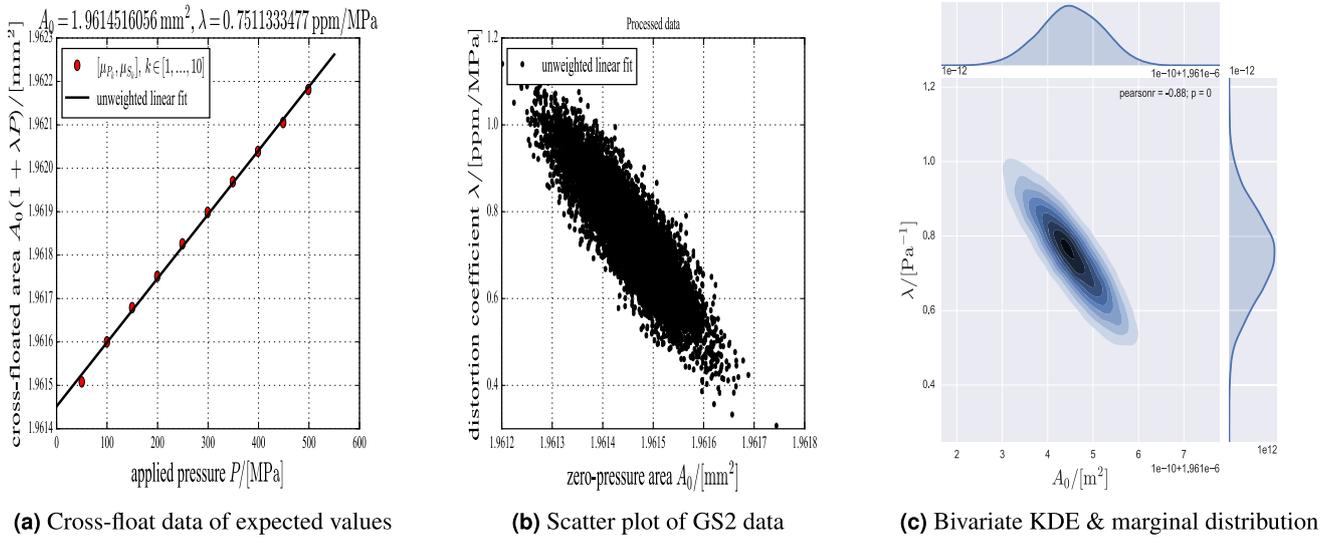
Fig. 1. Illustration of how the Monte Carlo data from a GUM Supplement uncertainty analysis is utilized in order to determine the joint probability density function for the transfer standard pressure balance. In (a) the univariate GUM Supplement 1 (GS1) Monte Carlo data of the cross-floated pressures P_k and cross-floated areas S_k is first pre-processed by averaging and plotted in order to get a rough qualitative estimate of the model parameters. In (b) the actual GS1 data points are then processed with the GUM Supplement 2 (GS2) using a suitable χ^2 optimization to extract the actual model parameters and the discrete model parameter values plotted. In (c) the discrete model parameters are then post-processed using a kernal density estimation scheme in order visualize the mathematically continuously variable joint bivariate probability density function behaviour of the underlying model.

is the kernel function. The use of a kernal function also occurs in the case of multivariate data $X_i \in \mathbb{R}^d$ with data points $X_1, \ldots, X_n$ where the multivariate PDF is constructed such that $f(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$ and as a result a similar issue will arise if we opt to use a KDE formulation to construct the bivariate PDF since the choice of kernal used for the KDE construction will introduce a level of subjectivity in the bivariate joint PDF. Typically when implementing a KDE to construct a univariate PDF a Gaussian kernal function is usually used however other particular choices of kernals such as an Epanechnikov and biweight kernal can also be used. As a result different PDF estimates may occur when the underlying Monte Carlo data is post-processed using a KDE approach.

Since we do not have any a priori knowledge of the most appropriate underlying kernal our approach is to simply construct the actual discrete approximation of the joint PDF using the Python histogram2d function which correctly normalizes the two dimensional PDF such that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1$ where in our particular problem the random variables are chosen such that $x \equiv \xi_{A_0}/[m^2]$ and $y \equiv \xi_\lambda/[Pa^{-1}]$. This is performed by using the well known Freedman–Diaconis statistical rule to estimate the bandwidth as

$$h_x = 2\Big(IQR(x)\Big) N^{-1/3} \qquad (24)$$

where $IQR(x) = Q_3(x) - Q_1(x)$ is the interquartile range for the the random variable x where $Q_3(x)$ and $Q_1(x)$ are the third and first quartiles. The number of bins h_x associated with x using this choice of bandwidth is then calculated using the ceiling function as

$$N_x = \lceil \frac{\max(x) - \min(x)}{h_x} \rceil \qquad (25)$$

with similar associated expressions for the other random variable y. The final actual bivariate joint PDF from the GS2 simulation is illustrated in in natural physical units. Referring to the two dimensional discrete histogram we observe that the limits for a joint PDF $g_{A_0, \lambda}(\xi_{A_0}, \xi_\lambda)$ are approximately $\min(\xi_{A_0}) = 1.9612 \, \text{mm}^2$, $\max(\xi_{A_0}) = 1.9617 \, \text{mm}^2$, $\min(\xi_\lambda) = 0.30 \, \text{ppm/MPa}$ and $\max(\xi_\lambda) = 1.10$ ppm/M. In this joint PDF the random variable for area ξ_{A_0} is in units of m^2 so that when the pressure is in units of Pa the random variable for the distortion coefficient ξ_λ is in units of Pa^{-1} for dimensional consistency in order to satisfy the normalization condition $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_{A_0, \lambda}(\xi_{A_0}, \xi_\lambda) \, d\xi_{A_0} d\xi_\lambda = 1$.

For brevity let the random variable ξ_{A_0} be represented by x and the random variable ξ_λ be represented by y so that the joint PDF $f(\xi_{A_0}, \xi_\lambda)$ is just $f(x, y)$ for convenience. If we construct variables such that

$$p = \frac{\xi_{A_0} - \min(\xi_{A_0})}{\max(\xi_{A_0}) - \min(\xi_{A_0})} \qquad (26a)$$

$$r = \frac{\xi_\lambda - \min(\xi_\lambda)}{\max(\xi_\lambda) - \min(\xi_\lambda)} \qquad (26b)$$

$$\varphi(p, r) = \frac{\xi_{A_0}, (\xi_{A_0}, \xi)}{\max(\xi_{A_0}, (\xi_{A_0}, \xi))} \qquad (26c)$$

the original random variables are mapped to equivalent scaled variables such that $0 \le p \le 1$ and $0 \le r \le 1$. As a result the original joint PDF $g_{A_0}, (\xi_{A_0}, \xi)$ may now be defined in terms of transformed random variables p and r.

(a) Scatter plot indicating bin counts   **(b)** Actual bivariate joint PDF   **(c)** Normalized bivariate joint PDF
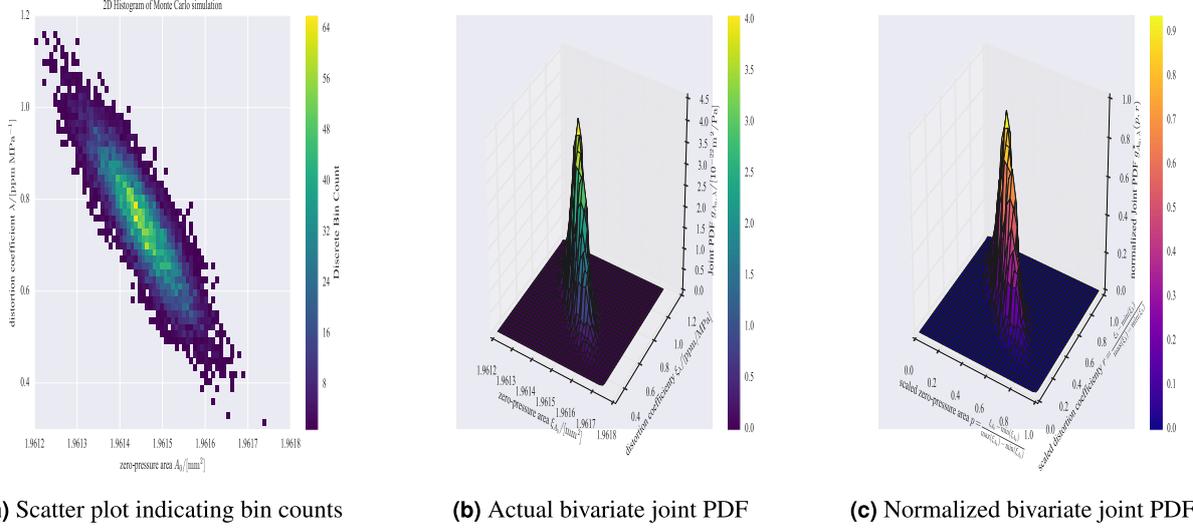
Fig. 2. Illustration of Monte Carlo GUM Supplement 2 processed data showing a visualization of the transfer standard pressure balance bivariate joint PDF where the random variables are plotted in physical units such that $10^6 A_0/[\text{m}^2] = A_0/[\text{mm}^2]$ and $10^{12}\lambda/[\text{Pa}^{-1}] = \lambda/[\text{ppm}/\text{MPa}]$ for easy visualization so that the joint PDF has the correct magnitude for a two dimensional normalization such that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{A_0,\lambda}(\xi_{A_0}, \xi_\lambda)\, \mathrm{d}\xi_{A_0}\mathrm{d}\xi_\lambda = 1$ when constructed from the two dimensional histogram data using $N_{A_0} = N_x = 63$ and $N_\lambda = N_y = 58$ bins respectively with $M = 10\,000$ Monte Carlo simulation events. In (a) the actual discrete data of the bivariate joint probability density function is first plotted for a qualitative visualization. In (b) the discrete data is post-processed to produce the mathematically continuous joint probability density function in a three dimensional space where each space coordinate corresponds to a random variable. In (c) the joint probability density function is normalized in order to more clearly view the behaviour of the underlying model.

The normalization of variables is utilized due to scaling effects since it will be generally be more convenient to fit the normalized joint PDF such that $\varphi(p, r) = [f(x, y)]/[\max(f(x, y)]$. As a result by fitting the normalized parameters $0 \leq \varphi(p, r) \leq 1$ with $0 \leq p \leq 1$, $0 \leq r \leq 1$ and $f_{\max} = \max\{f(x, y)\}$ the scaled values can then be recovered such that

$$x = p(x_{\max} - x_{\min}) + x_{\min} \qquad (27\text{a})$$

$$y = r(y_{\max} - y_{\min}) + y_{\min} \qquad (27\text{b})$$

$$f(x, y) = f_{\max}\varphi(p, r) \qquad (27\text{c})$$

This fit may be implemented using the marginal distributions form of the generalized bivariate quantile distributions for the zero-pressure area designated as $g_{A_0}^{\star}(\xi_{A_0})$ in Figure 3a and for the distortion coefficient designated as $g_\lambda^{\star}(\xi_\lambda)$ in Figure 3b respectively. We comment that due to the shape of the PDFs for these marginal distributions as observed from the histogram for A_0 in Figure 3c and the histogram for λ in Figure 3d respectively which are constructed with a Freedman–Diaconis choice of bin size that it would generally be convenient and reasonably accurate to approximate these PDFs in terms of either univariate ELDs as summarized in Table 3 or alternatively in terms of univariate splines where the degree of interpolation can be adjusted to refine the desired accuracy level.

The approach of sampling points x and y from a marginal distribution approximation of a joint PDF $f(x, y)$ is to first generate a point p using a rectangular random number generator distribution which can then be used to calculate x from the corresponding marginal distribution $Q_x(p)$. Next by using this known value of x we solve for y from $Q_y(p, r)$ using the previously determined value of p however a practical implementation scheme is needed to construct the marginal distribution for $Q_y(p, r)$. In this paper for simplicity we opt to perform this implementation by constructing a set of functions for a set of contour curves for specified known values of x so that any arbitrary values can be estimated using for example bilinear or bicubic interpolation for simplicity.
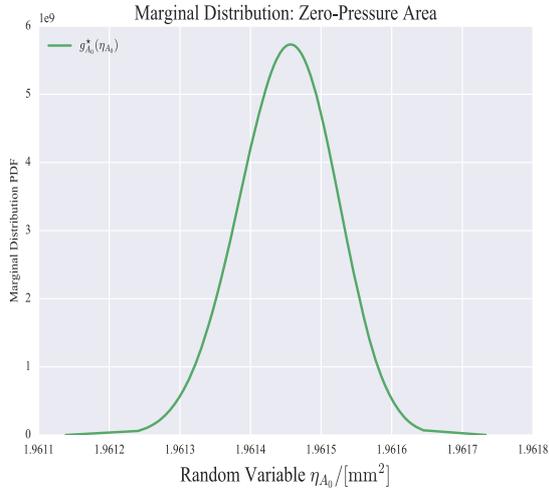
A conventional classical marginal distribution approximation of the joint PDF may therefore be specified by the following three steps where a convenient functional form would in most practical cases be that of a spline fit such as for example a B-spline. The general algorithm would then implement the following general steps such that:

Step 1. Fit the marginal distribution $Q_p(p)$ for the scaled variate p with any convenient practical functional form such that $Q_p(p) = M(p, \boldsymbol{\alpha})$ where $M(p, \boldsymbol{\alpha})$ is a marginal distribution function and $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_m]$ is a corresponding fitted parameter
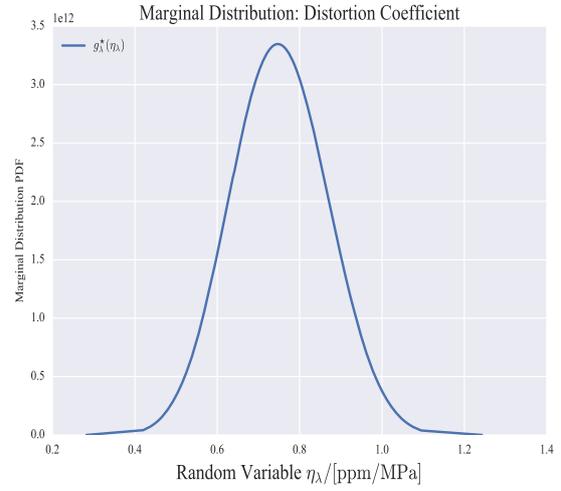
Step 2. Construct a set $X = \{p_1, \ldots, p_m\}$ for some finite integer $m \in \mathbb{Z}$ where $0 = p_1 < p_2 < \ldots < p_{m-1} < p_m = 1$

Step 3. For each $p_i \in X$ extract the corresponding contour curve $C_i = \varphi(p_i, r)$ from the scaled joint PDF $\varphi(p, r)$ and fit any convenient practical functional form for each curve C_i such that $C_i = N_i(r, \beta_i)$ where β_i is a corresponding fitted parameter
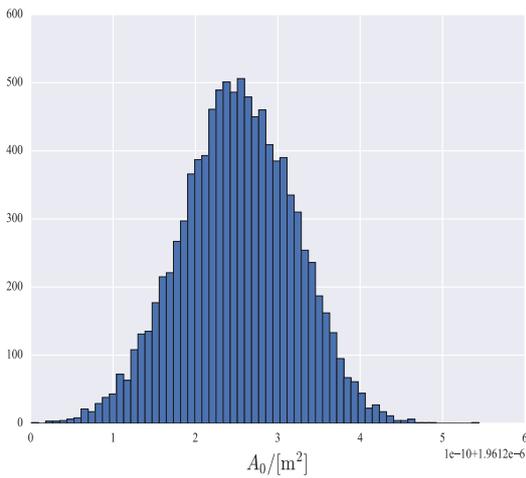
As a result the quantification and specification of the parameter $\boldsymbol{\alpha}$ for the marginal distribution $Q_x(p; \boldsymbol{\alpha})$ and the set of parameters $\beta_1 \ldots \beta_m$ for the contour curves for the
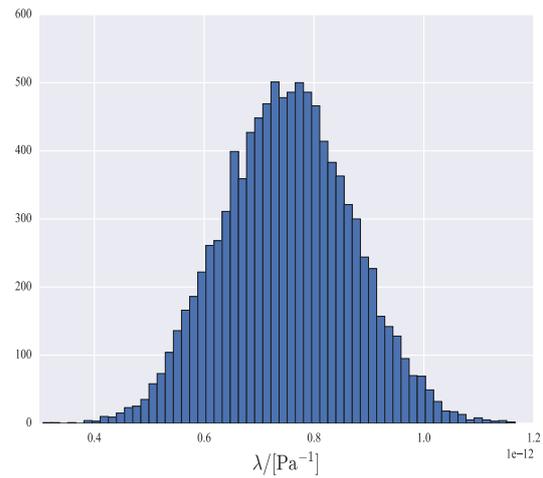
(a) ELD fit for marginal distribution $g^\star_{A_0}(\xi_{A_0})$

(b) ELD fit for marginal distribution $g^\star_\lambda(\xi_\lambda)$

(c) Histogram for $A_0/[\mathrm{m}^2]$ using $N_{A_0} = 63$ bins

(d) Histogram for $\lambda/[\mathrm{Pa}^{-1}]$ using $N_\lambda = 58$ bins

Fig. 3. Monte Carlo GUM Supplement 2 marginal distributions. In (a) the probability density function distribution for the continuous random variable ξ_{A_0} of the zero-pressure area is visualized. In (b) the probability density function distribution for the continuous random variable ξ_λ of the distortion coefficient is visualized. In (c) the underlying discrete data from the GUM Supplement 2 analysis that is used to build the PDF for the zero-pressure area is shown. In (d) the underlying discrete data from the GUM Supplement 2 analysis that is used to build the PDF for the distortion coefficient is shown

Table 3. Marginal distribution ELD parameters for zero-pressure area A_0 and distortion coefficient λ variates. The underling probability density function distribution for the zero-pressure area random variable ξ_{A_0} and the distortion coefficient random variable ξ_λ is modelled and approximated with an extended lambda distribution based on the discrete data obtained from the GUM Supplement 2 uncertainty analysis.

Variate	a	b	c	d
$A_0/[\mathrm{m}^2]$	1.16239968	$1.64611221 \times 10^{-1}$	$4.52707068 \times 10^{-11}$	$1.96145805 \times 10^{-6}$
$\lambda\,/[\mathrm{Pa}^{-1}]$	$9.33904128 \times 10^{-1}$	$1.75753525 \times 10^{-1}$	$8.72674858 \times 10^{-14}$	$7.46168899 \times 10^{-13}$

(a) Actual bivariate surface   **(b)** Marginal contour curve approx.   **(c)** Scattered data surface approx.
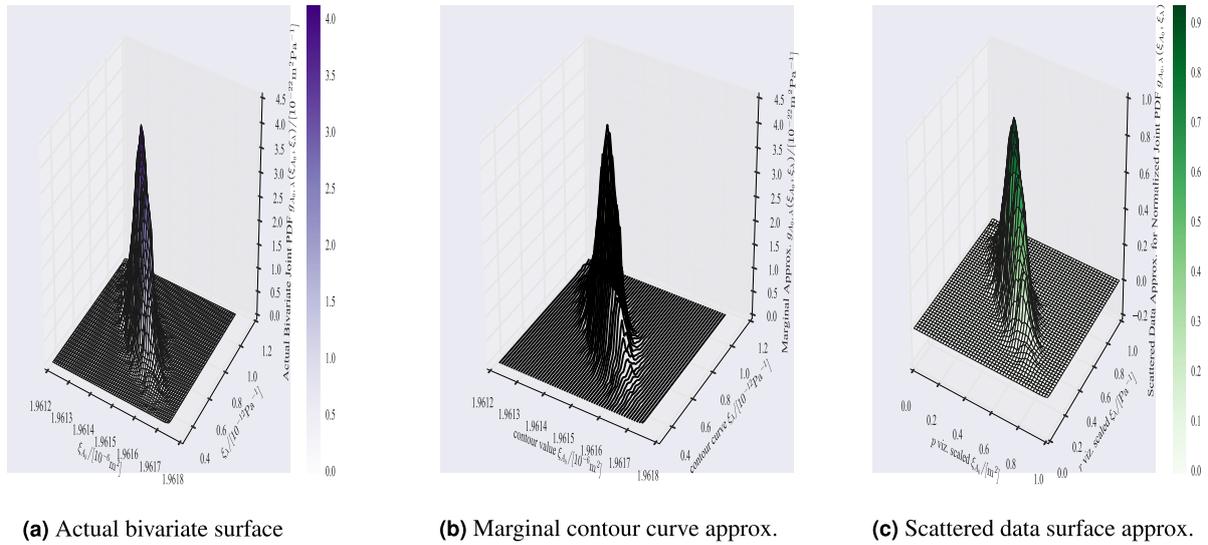
Fig. 4. Illustration of comparison between the actual bivariate joint PDF surface with that of a conventional classical marginal distribution approach approximation showing the discrete set of contour curves visualized by three dimensional lines that are used to construct joint PDF surface, and that of a scattered data approximation of the corresponding normalized bivariate surface. In (a) the actual continuous probability density function distribution is shown which must be modelled and approximated through a visualization in a three dimensional space where the x coordinate corresponds to the random variable ξ_{A_0}, the y coordinate corresponds to the random variable ξ_{λ} and the z coordinate corresponds to the joint PDF $g_{A_0\lambda}(\xi_{A_0}, \xi_{\lambda})$. In (b) the modelling and approximation is shown where the three dimensional surface is approximated by a series of three dimensional lines within the three dimensional space. In (c) the modelling and approximation is visualized for the equivalent surface that is built up from discrete three dimensional scattered points.

marginal distribution $Q_y(p_i, r; \beta_i)$ would actually be sufficient to fully characterize the TS pressure balance's behaviour for test calibrations and metrology inter-comparisons. The challenge with this traditional marginal distribution approximation of a bivariate distribution approach is that whilst we could easily explicitly directly generate values for the x random variable corresponding to A_0 we would have to implicitly indirectly generate values for the y random variable corresponding to λ by choosing a fixed number of contour curves and then employing an interpolation scheme.

In order to clarify how to sample random variables ξ_{A_0} and ξ_λ from the marginal distribution approximation of the joint PDF $g_{A_0,\lambda}(\xi_{A_0}, \xi_\lambda)$ suppose for illustration purposes that we wish to generate a random variable of the zero-pressure area ξ_{A_0} as x_i for $p_i = 0.7$ and a random variable for the distortion coefficient λ as y_i for $r_i = 0.82$. We can immediately calculate x as $x_i = Q_x(0.7)$ but it will be more difficult to calculate y_i since in general we would have a set of contour curves for fixed values of p, say $p \in [0.0, 0.25, 0.5, 0.75, 1.0]$. Selecting the nearest p_ℓ that is smaller than p_i i.e. $p_\ell = 0.5$ and the nearest p_u that is larger than p_i i.e. $p_u = 0.75$ allows us to calculate the corresponding y for p_ℓ as $y_\ell = Q_y(p_\ell, r_i)$, and that for p_u as $y_u = Q_y(p_u, r_i)$. Then these data points can be used to interpolate for the required y_i corresponding to $r_i = 0.82$ using linear interpolation so that $y_i = y_\ell + [(y_u - y_\ell)(p_i - p_\ell)]/[p_u - p_\ell]$.

This particular illustrative example for how to perform a two dimensional interpolation to calculate samples of random variables from the underlying joint PDF is mathematically equivalent to a bilinear interpolation, although in principle any particular interpolation

scheme such as for example bicubic interpolation may also be used to obtain smoother interpolated fit values. The marginal distribution approach is illustrated by comparing the actual bivariate surface in Figure 4a with the surface approximation as shown in Figure 4b which demonstrates how a set of contour curves can be used to approximate the two dimensional joint PDF surface. In this figure there are a total of $N_x = 60$ curves for evenly spaced values of the random variable $x = \xi_{A_0}$ from 1.9612 mm^2 to 1.9617 mm^2 and the $y = \xi_\lambda$ varies from 0.3145 ppm/MPa to 1.1581 ppm/MPa in evenly spaced values using a total of $N_y = 75$ points. Whilst the marginal distribution approach for constructing a bivariate PDF is theoretically possible it does not unfortunately offer a convenient and practical means of modelling and summarizing a bivariate PDF for reporting purposes such as in a pressure balance calibration certificate or inter-comparison report.

An alternative to the marginal distribution approach for arbitrary specifications of p_i and r_i without any need for additional interpolations in order to immediately directly sample random variables from the joint PDF is to formulate the surface fitting of the joint PDF as a scattered data interpolation problem. This problem which is common in the field of statistics is formally specified as that given a set of discrete sampled data points (x_j, y_j) for $j = 1, 2, \ldots, N$ with $x_j \in \mathbb{R}^s$ and $y_j \in \mathbb{R}$ where $0 < s \in \mathbb{Z}$ is the dimension of the data, then find a continuous function P_f such that $P_f(x_j) = y_j \quad \forall j \in [1, 2, \ldots, N]$. A particular well-posed solution to the scattered data interpolation problem is to use the distance matrix formulation such that

$$P_f(x) = \sum_{k=1}^{N} c_k \|x - x_k\|_2, \quad x \in [0,1]^2 \qquad (28)$$

where $\| \cdot \|_2$ is the ℓ^2-norm defined as $\|x\| = \sqrt{x_1^2 + x_2^2}$ for $x = [x_1, x_2]^T$ in the two dimensional case when $s = 2$. The coefficients c_k for $k = 1, 2, \ldots, N$ where N is the total number of known discrete points $\{x_j, y_j\}$ is obtained by solving the linear system specified as

$$\mathrm{D}c = b \qquad (29a)$$

$$D_{ij} = \|x_i - x_j\|_2 \qquad (29b)$$

$$b = [y_1, \ldots, y_N]^T \qquad (29c)$$

$$c = [c_1, \ldots, c_N]^T \qquad (29d)$$

We comment that the scattered data interpolation solution using the distance matrix formulation can be refined for an arbitrarily large number of supplied known data points. A useful benefit of this approach is that it is not strictly necessary to first normalize and scale the random variables to fit $x = [p_j, r_j]^T$ and $y_j = \varphi(p_j, r_j)$ for $j = 1, \ldots, N$ since the scattered data interpolation scheme can be applied directly with $x_j = [(\xi_{A_0})_j, (\xi_\lambda)_j]^T$ and $y_j = f((\xi_{A_0})_j, (\xi_\lambda)_j)$. Nevertheless although a normalization is not theoretically necessary in our simulations we opt to apply the fit to the normalized variables $0 \leq p \leq 1$, $0 \leq r \leq 1$ and $0 \leq \varphi(p, r) \leq 1$ to avoid ill-conditioning issues when solving the matrix equation $\mathrm{D}c = b$. Results for a scattered data interpolation scheme using the same underlying data is illustrated by comparing the actual bivariate surface as previously shown in Figure 4a with the scattered data surface approximation as shown in Figure 4c.

Whilst this construction of the conventional scattered data two dimensional joint PDF is better than that of a marginal distribution approach which is restricted by the fixed number of discrete contour curves chosen, the main issue is that the number of constants used to construct the surface is still very large since in this case there are $N = 3534$ constants $c_1, \ldots, c_N$ for the underlying discrete data points $\{x_j = [p, r]^T, y_j = \varphi(p, r)\}, \quad j = 1, \ldots, N$. As a result the main issue with a scattered data interpolation scheme apart from the large number of constants to approximate the surface in our particular physical pressure balance metrological system, is that although it is guaranteed to adequately model the underlying data that the associated system of equations is usually ill-conditioned.

Alternative meshless techniques that avoid ill-conditioning encompass the radial basis function (RBF) approach as discussed by Larsson and Fornberg [33] which may be considered a generalization of the scattered data interpolation scheme $P_f(x) = \sum_{k=1}^{N} c_k B_k(x)$. In this approach the constants are calculated by solving the matrix equation $\mathrm{A}c = y$ where $A_{ij} = \phi(\varepsilon, r)$ where $\phi = r^2 \log(r)$ in the case of a thin plate spline or where $\phi = \exp(-(\varepsilon r)^2)$ in the case of a Gaussian RBF where $r = \|x_j - x_i\|$ and $c = [c_1, \ldots, c_N]^T$ as previously discussed, however the same issue of a large number of constants is still present.

This problem again occurs when low order spline approximations of bivariate functions are considered where the bivariate function is traditionally approximated as a tensor product spline surface of the form

$$f(x,y) \approx \sum_{k=1}^{K} \sigma_k u_k(x) v_k(y) \qquad (30)$$

where K is an integer corresponding to the order of the approximation, σ_k are constants, and $u_k(x)$ and $v_k(y)$ are univariate functions each with their own individual corresponding parameters particular to the characteristics of the respective basis functions as recently developed by Georgieva and Hofreither [34]. The main issue with such a traditional low rank approximation in our particular problem is that although it can very accurately model and generate a reconstructed surface for suitable choices of σ_k, $u_k(x)$, $v_k(y)$ and K, is that in general a very large number of coefficients are generally still necessary in order to adequately reconstruct the basis functions $u_k(x)$ and $v_k(y)$. Whilst a large number of coefficients is not an issue in a computer implementation the reporting of a large number ranging from a few hundred to a few thousand coefficient values is nevertheless however not considered practical for a calibration paper certificate or written inter-comparison report.

Due to these practical implementation issues we therefore consider a copula approach $C_\theta(u, v)$ as previously discussed in order to construct the bivariate distribution for our particular physical metrological measurand system since for many practical copula families only a small number of parameters are required in the term $\theta = [\theta_1, \ldots, \theta_n]^T$. As a result due to generally small number of parameters to adequately characterize bivariate PDF distributions this renders the reporting of the physically characterized pressure balance joint PDF in calibration certificates and inter-comparison reports a practical and feasible alternative for oil based pressure balances.

In this paper for brevity we restrict our numerical investigation to the normal, Student-t, Gumbel, Frank and Clayton bivariate copulas $C_\theta(u, v)$ as summarized in Table 4 where the respective copula parameter θ is obtained from the Kendall tau parameter τ_K as previously discussed. When constructing the copula we use x as the random variable for the zero-pressure area A_0 and y as the random variable for the distortion coefficient λ in order to construct the joint PDF $f(x, y)$ such that

$$f(x,y) = uv \frac{\partial^2 C(u,v)}{\partial u \partial v} \qquad (31)$$

$$u = \int_{-\infty}^{x} g_x(\xi_x) \, d\xi_x \qquad (32)$$

$$v = \int_{-\infty}^{y} g_y(\xi_y) \, d\xi_y \qquad (33)$$

In the above formulae $g_x(\xi_x)$ is the marginal PDF for x as previously approximated with an ELD as illustrated in Figure 3a, whilst $g_y(\xi_y)$ is the marginal PDF for y as illustrated in Figure 3b both of which are fully parame-

Table 4. Specification of common parameter based bivariate copula families (Adapted from Goda [17]). In our particular problem the variable u corresponds to the random variable ξ_{A_0} and the variable v corresponds to the random variable ξ_λ so that the bivariate copula $C(u, v)$ can be used to construct the joint probability density function as $g_{A_0,\lambda}(\xi_{A_0}, \xi_\lambda) = uv\frac{\partial^2 C}{\partial u \partial v}$ for convenience.

Copula name	Copula function $C(u, v)$	Copula parameter relationship with Kendall's tau
Normal	$\Phi_\rho\Big(\Phi^{-1}(u), \Phi^{-1}(v)\Big)$	$\rho = \sin\left(\dfrac{\pi\tau_K}{2}\right)$
Student-t	$t_{\rho,\nu}\Big(t_\nu^{-1}(u), t_\nu^{-1}(v)\Big)$	$\rho = \sin\left(\dfrac{\pi\tau_K}{2}\right)$
Gumbel	$\exp\Big(-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}\Big)$	$\theta = \frac{1}{(1-\tau_K)}$ and $\theta \geq 1$
Frank	$-\dfrac{1}{\theta}\ln\left(1 + \dfrac{\{e^{-\theta u}-1\}\{e^{-\theta v}-1\}}{e^{-\theta}-1}\right)$	$\tau_K = 1 - \frac{4}{\theta} + \frac{4}{\theta^2}\int_0^\theta \frac{t}{e^t-1}dt$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$\theta = \frac{2\tau_K}{(1-\tau_K)}$ and $\theta \in [0,\, \infty]$

trized in terms of respective univariate quantile parameters a, b, c, d. By exploiting the properties of ELDs as previously discussed by Willink [16] the distribution function formally defined as $F(x) = \int_{-\infty}^x f(u)\,\mathrm{d}u$ for the random variable x with a similar expression for the random variable y may in the special case of an ELD approximation be simplified as
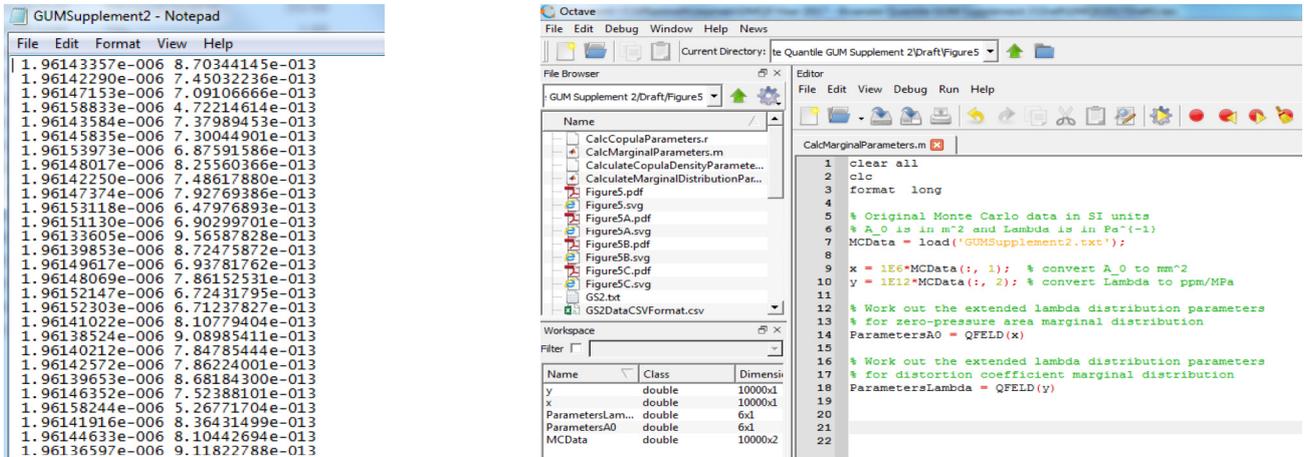
$$Q(\rho) = \begin{cases} d + \frac{c}{b}[a\rho^b - (1-\rho)^b + 1 - a] & if\, b \neq 0 \\ d + c\{a\ln\rho - \ln(1-\rho)\} & if\, b = 0 \end{cases} \quad (34)$$

As a result if the marginal distribution for the zero-presssure area A_0 random variable x has ELD parameters a_x, b_x, c_x, d_x, and the marginal distribution for the distortion coefficient λ random variable y has ELD parameters a_y, b_y, c_y, d_y it follows that we can immediately utilize the previous analytical expressions in the copula construction for the bivariate joint PDF in terms of the ELD parameters. In order to construct a copula function $C_\theta(u, v)$ for the formulae listed in for example Table 4 we must in the case of the Gaussian copula use $\Phi_\rho(\bullet)$ which denotes the bivariate standard normal distribution with correlation ρ whilst $\Phi^{-1}(\bullet)$ denotes the inverse standard normal distribution. On the other hand for the bivariate Student-t copula $t_{\rho,\nu}$ denotes the bivariate t-distribution with parameters ρ and ν where the degrees of freedom ν must typically be calculated using the Akaike Information Criterion (AIC) where the smallest AIC value (which depends on ν) is considered to be the best-fit copula choice, and $t_\nu^{-1}(\bullet)$ is the corresponding inverse t-distribution value for the associated degrees of freedom ν parameter for the copula. Although the conventional t-distribution copula usually has one degree of freedom parameter ν more modern alternatives can incorporate multiple degrees of freedom parameters $\nu_1, \ldots, \nu_n$ for bivariate models as discussed in Luo and Shevchenko [35].

As a result whilst it is certainly possible to estimate bivariate copula parameters $\theta = [\theta_1, \ldots, \theta_n]^\mathrm{T}$ directly with 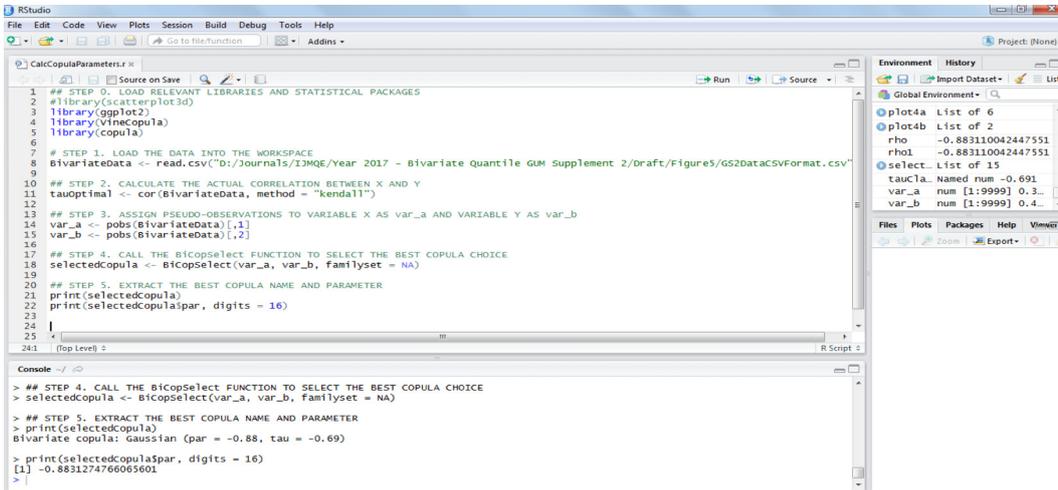custom written computer code through various parameter optimization approaches for a range of copula families of which Table 4 is only a small selection of the very extensive range of possible copula families this is no longer strictly essential. This is due to the fact that Yan [36] developed a R based package copula which is now readily available to researchers worldwide, and which has subsequently been expanded by Kojadinovic and Yan [37] for multivariate distribution modelling using copulas. As a result the use of R based open source statistical software from the Comprehensive R Archive Network is now commonly available to researchers worldwide and is accepted as a standard statistical tool within the statistics community. In this paper we will therefore utilize the copula [38], copBasic [39], and VineCopula [40]. R packages to simplify the analysis in order to construct the best choice of a copula function for our underlying pressure balance Monte Carlo bivariate data.

The use of the above-mentioned R packages allows us to simply load the bivariate data from our post-processed GS2 Monte Carlo simulations and to then implement the following computer code using the VineCopula library which automatically selects the most appropriate bivariate copula function drawing on a selection of copula families more extensive that the most common families previously outlined in Table 4. When performing the copula selection to avoid numerical ill-conditioning due to scaling effects since $\mathcal{O}(\xi_{A_0}) = 10^{-6}$ and $\mathcal{O}(\xi_\lambda) = 10^{-12}$ in SI units it may depending on the available computer systems be advantageous when fitting the copula function to utilize the variate data ξ_{A_0} for A_0 and ξ_λ for λ such that they are first converted so that x is in mm and y is in ppm/MPa. We comment that the choice of units is not problematic since our main objective is to model the bivariate joint PDF and the form of the fit should simply allow us to adequately sample values for the A_0 and λ variates x and y. Further technical implementation details for the VineCopula library are available in the official documentation however we briefly comment that the user would in most practical situations select from ml for maximum likelihood, mpl for maximum pseudo-likelihood, itau for inversion of Kendall's

(a) Original GS2 Monte Carlo source data   (b) Univariate ELD marginal distribution fit in GNU Octave/Matlab

(c) Bivariate copula distribution fit in RStudio

Fig. 5. Illustration of how a GUM Supplement 2 (GS2) Monte Carlo uncertainty analysis would be implemented in practice for constructing the marginal distribution and copula distribution parameters. In (a) an extract of the actual $M = 10000$ data points is shown from the GS2 which were determined with a suitable χ^2 optimization. In (b) the computer code implementation of how to process the GS2 data to fit the marginal distributions in terms of an extended lambda distribution is shown. In (c) the computer code implementation of how to determine the copula parameters using open source easily accessible standard statistical software is shown.

tau, or irho for inversion of Spearman's rho as choices for the parameter optimization choice for the software code to determine the optimal copula parameter fit.

The process to fit the bivariate data starts with the GS2 Monte Carlo data GUMSupplement2.txt that is obtained with any suitable computer code simulation environment that is saved in a neutral ASCII txt file format. Once the data is saved it can then be processed to firstly extract the univariate ELD quantile parameters in for example GNU Octave for the marginal distributions, and then secondly processed in RStudio using the VineCopula library to determine the optimal bivariate copula and associated copula parameters as shown in Figure 5.

Results for the univariate marginal distribution parameter fits with GNU Octave v4.2.0 by running computer program CalcMarginalParameters.m with the m-code shown in Figure 6 are shown below in Listing 1.

Results for the bivariate copula parameter fits with RStudio by running computer program CalcCopulaParameters.r are shown below in Listing 2.

The above computer code console outputs are actually sufficient to fully characterize the ELD marginal distribution and copula bivariate distribution parametrizations. In order to construct the joint PDF the second order mixed partial derivative of the copula $C(u, v; \theta)$ must be calculated in order to calculate the copula density $c(u, v; \theta)$. Further technical implementation details to directly numerically construct the joint PDF from a general copula that is not amenable to closed form analytical expressions of the copula density are outlined in Appendix B.

For the particular case of a Gaussian copula Meyer [41] provides explicit analytical closed form expressions such that for the univariate case the cumulative distribution function $\Phi^{-1}(\bullet)$ is defined as

GNU Octave/Matlab m-code

```
1   function qfvalues = QFELD(ydata)
2   y = sort(ydata); M = length(y); r = 1:1:M; p = (1/M)*(r − 0.5);
3   eta = y; rho = p; Q = eta;
4   p1 = 0.025; p2 = 0.25; p3 = 0.75; p4 = 0.975;
5   theta1 = interp1(rho, Q, p1); theta2 = interp1(rho, Q, p2);
6   theta3 = interp1(rho, Q, p3); theta4 = interp1(rho, Q, p4);
7   xi1 = (theta4 − theta1)/(theta3 − theta2);
8   xi11 = (theta4 − theta2)/(theta3 − theta1);
9   bmin = −10.5; bmax = 1.50; N = 1000;
10  delta = (bmax − bmin)/(N − 1); bvector = bmin:delta:bmax;
11
12  for i = 1:N
13      bval = bvector(i);
14      fvector(i) = (p1^bval − (1 − p1)^bval)/(p2^bval − (1 − p2)^bval) − xi1;
15  end
16
17  b = interp1(fvector, bvector, 0);
18  a = ( xi11*((1 − p1)^b − (1 − p3)^b) + (1 − p4)^b − (1 − p2)^b )/( xi11*(p1^b − p3^b) + p4^b − p2^b );
19  c = ( b*(theta2 − theta1) )/( (a*(p2^b) − (1 − p2)^b) − (a*(p1^b) − (1 − p1)^b) );
20  d = theta1 − (c/b)*(a*(p1^b) − (1 − p1)^b + 1 − a);
21  mu = d + (c*(1 − a))/(b + 1);
22
23  if (b ~= 0)
24      sigma_squared = ( (c/b)^2 )*( (a^2 + 1)/(2*b + 1) − ( 2*a*gamma(b + 1)^2 )/( gamma(2*b + 2) ) − ( (a − 1)/(b +
            1) )^2 );
25  end
26
27  if (b == 0)
28      sigma_squared = (c^2)*(a^2 + ((pi^2)*a)/3 − 2*a + 1);
29  end
30
31  sigma = sqrt(sigma_squared);
32  qfvalues = [a; b; c; d; mu; sigma];
```

Fig. 6. Illustration of m-code for estimating an ELD parametrization for univariate Monte Carlo data. The computer program works by loading the Monte Carlo univariate data as an input $ydata = [y_1, \ldots, y_M]^T$ where $M = 10\,000$ is the number of Monte Carlo simulation events and it then uses this data to determine the extended lambda distribution parameters a, b, c and d so that the probability density function can be analytically specified.

GNU Octave console output

```
1   >>
2   ParametersA0 =
3
4     1.16239967645405e+000
5     1.64611220787361e−001
6     4.52707068305772e−005
7     1.96145804545895e+000
8     1.96145173266717e+000
9     6.86495890320276e−005
10
11  ParametersLambda =
12
13    0.9339041282953079
14    0.1757535247523311
15    0.0872674857691966
16    0.7461688986804182
17    0.7510747062514455
18    0.1164645833607343
19
20  >>
```

Listing 1. Computer output used to calculate the univariate ELD parameters for A0 and .

RStudio console output

```
1   >
2   > print(selectedCopula)
3   Bivariate copula: Gaussian (par =
        −0.88, tau = −0.69)
4   > print(selectedCopula$par, digits
        =16)
5   [1] −0.8831274766065601
6   >
```

Listing 2. Computer output used to calculate the copula parameter with open source standard statistical software.

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{35}$$

$$\Phi(h) = \int_{-\infty}^{h} \varphi(x)\, \mathrm{d}x \tag{36}$$

The univariate standard normal CDF $\Phi(h)$ is then used to construct the bivariate standard normal CDF $\Phi_2(a, b)$ such that

$$C(u, v; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}}$$
$$\times \int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\phi^{-1}(v)}\exp\left(\frac{-(s^2 - 2\rho st + t^2)}{2(1-\rho^2)}\right)ds\,dt \tag{37}$$

$$c(u, v; \rho) = \frac{1}{\sqrt{1-\rho^2}}$$
$$\times \exp\left(\frac{2\rho\Phi^{-1}(u)\Phi^{-1}(v) - \rho^2(\Phi^{-1}(u)^2 + \Phi^{-1}(v)^2)}{2(1-\rho^2)}\right) \tag{38}$$

When the above expressions are combined we then obtain the final expression for a Gaussian copula formulation of the bivariate joint PDF as

$$f(x, y) = F(x)G(y)c(u, v; \rho) \tag{39}$$

In our particular case since the optimal copula is a Gaussian we simply specify the copula as

$$C(u, v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \rho) \tag{40a}$$

$$\rho = -0.88312747 \tag{40b}$$

The above copula function informations is then completed by also specifying the associated quantile function parameters for $F(x) = u$ and $G(y) = v$ which using our ELD approximation is of the form

$$u = \begin{cases} d_x + \frac{c_x}{b_x}[a_x\rho^{b_x} - (1-\rho)^{b_x} + 1 - a_x], & b_x \neq 0 \\ d_x + c_x\{a_x\ln\rho - \ln(1-\rho)\}, & b_x = 0 \end{cases} \tag{41a}$$

$$v = \begin{cases} d_y + \frac{c_y}{b_y}[a_y\rho^{b_y} - (1-\rho)^{b_y} + 1 - a_y], & b_y \neq 0 \\ d_y + c_y\{a_y\ln\rho - \ln(1-\rho)\}, & b_y = 0 \end{cases} \tag{41b}$$

where

$$a_x = 1.1623996764 \tag{42a}$$

$$b_x = 1.6461122078 \times 10^{-1} \tag{42b}$$

$$c_x = 4.5270706830 \times 10^{-5} \tag{42c}$$

$$d_x = 1.9614580454 \tag{42d}$$

while the quantile function parameters for G (y) are

$$a_y = 0.9339041282 \tag{43a}$$

$$b_y = 0.1757535247 \tag{43b}$$

$$c_y = 0.0872674857 \tag{43c}$$

$$d_y = 0.7461688986 \tag{43d}$$

The above set of four equations for the copula function $C(u, v; \theta)$, marginal distributions $F(x)$ and $G(y)$, and parameters associated with the respective marginal distributions $\{a_x, b_x, c_x, d_x\}$ and $\{a_y, b_y, c_y, d_y\}$ are then actually sufficient to model the bivariate joint PDF for the TS pressure balance, by noting the ELD parameter limits are by definition

$$d - \frac{ac}{b} \leq \eta \leq d + \frac{c}{b} \quad if\,b > 0 \tag{44a}$$

$$-\infty < \eta > \infty \quad if\,b \leq 0\,\text{and}\,a \neq 0 \tag{44b}$$

$$0 \leq \eta < \infty \quad if\,b \leq 0\,\text{and}\,a = 0 \tag{44c}$$

Once the joint PDF has been summarized using $f(x, y) = F(x)\,G(y)\,c(u, v)$ with $u = F(x)$ and $v = G(y)$ as indicated above, it can then be used in any further uncertainty analysis numerical simulations for the TS pressure balance. As an example we would in a practical physical pressure balance calibration like to sample random variables ξ_{A_0} and ξ_λ from $g_{A_0,\lambda}(\xi_{A_0}, \xi_\lambda)$ in order to calculate the uncertainty in generated pressure. The general numerical procedure we propose in this paper adapted from the univariate sampling procedure for arbitrary distributions is as follows for bivariate distributions:

Step 1. Sample a random variable $\rho_x \sim R$ from the rectangular distribution so that $0 \leq \rho_x \leq 1$

Step 2. Solve the equation $F(x^\star) = \rho_x$ in order to deduce a value for $x^\star$ and then simply set $\xi_{A_0} = x^\star$

Step 3. Construct the contour curve $\mathcal{C}(y) = f_X(x^\star, y)$ by fixing $x = x^\star$ in the joint PDF $f(x, y)$ and normalize as appropriate

Step 4. Build the CDF such that $H(y) = \int_{-\infty}^{y}\mathcal{C}(\eta)\,d\eta$ and normalize as appropriate

Step 5. Sample a random variable $\rho_y \sim R$ from the rectangular distribution so that $0 \leq \rho_y \leq 1$

Step 6. Solve the equation $H(y^\star) = \rho_y$ in order to deduce a value for $y^\star$ and then simply set $\xi_\lambda = y^\star$

If the sampled variates are $x^\star$ and $y^\star$, and ρ_x and ρ_y are sampled rectangular random numbers then the above procedure can be summarized as solving the following system of equations such that

$$x^\star \leftarrow F(x^\star) = \rho_x \tag{45a}$$

Table 5. Extended lambda distribution (ELD) parameters for transfer standard pressure balance generated applied pressures P_k/[Pa] using bivariate copula joint PDF. The index k corresponds to the pressures P_k for 50 MPa through to 500 MPa and the parameters a_k, b_k, c_k and d_k can then be used to construct analytical expressions for the probability density function distribution for each of the P_k pressures.

k	a_k	b_k	c_k	d_k	μ_k	σ_k
1	1.00826638e+000	2.32831909e−001	1.05511163e+003	4.98845255e+007	4.98845185e+007	1.35218460e+003
2	8.25935529e−001	1.27475710e−001	1.83660496e+003	9.97643451e+007	9.97646287e+007	2.48272214e+003
3	8.57266846e−001	2.50998814e−001	2.38151166e+003	1.49640123e+008	1.49640395e+008	2.75632616e+003
4	8.84814329e−001	1.07192634e−001	2.31987524e+003	1.99511888e+008	1.99512129e+008	3.33569791e+003
5	1.05502324e+000	2.07251770e−001	3.05943658e+003	2.49379382e+008	2.49379243e+008	4.15319862e+003
6	1.60989614e+000	2.21883068e−002	2.52668161e+003	2.99244536e+008	2.99243028e+008	5.79009224e+003
7	1.23624636e+000	4.89697314e−002	4.07133564e+003	3.49102761e+008	3.49101844e+008	7.61453960e+003
8	1.64812777e+000	5.67217654e−002	5.47850213e+003	3.98960634e+008	3.98957274e+008	1.20351081e+004
9	1.01291867e+000	2.60882836e−001	1.16876849e+004	4.48812421e+008	4.48812301e+008	1.44686784e+004
10	1.82861497e+000	2.23641442e−001	1.09786651e+004	4.98667444e+008	4.98660009e+008	2.01532077e+004

$$y^\star \leftarrow \frac{\int_{-\infty}^{y^\star} F(x^\star)G(y)c(\rho_x, G(y); \rho)\, \mathrm{d}y}{\int_{-\infty}^{\max(y)} F(x^\star)G(y)c(\rho_x, G(y); \rho)\, \mathrm{d}y} = \rho_y \qquad (45b)$$

We comment that the above bivariate sampling scheme is more general than a traditional bivariate Gaussian sampling scheme since it is not restricted to a conventional Gaussian bivariate PDF, and can sample from arbitrary bivariate joint PDFs since the approach will work for arbitrary marginal distributions and arbitrary bivariate copulas. As a result the benefit of using a copula to model a bivariate distribution over a traditional GS2 based Gaussian joint PDF is that one is no longer specifically restricted to bivariate Gaussian distributions as is presently the case according to the official GS2 guidelines. With the application of copulas as a particular modelling approach for bivariate quantile distributions one may therefore now in principle utilize any other alternative physical or mathematically plausible choice of probability distributions modelled in terms of copulas to summarize bivariate Monte Carlo uncertainty analysis simulation data as outlined in Appendix B.

We can now test the extent to which our bivariate joint PDF for the TS pressure balance can estimate the generated pressures. This is performed by implementing a GS1 simulation of a generating pressure model for the TS pressure balance as

$$f(p) = \frac{1}{S}\left[\sum_{j=1}^{n} m_j g\left(1 - \frac{\rho_a}{\rho_j}\right)\right.$$
$$\left. + g(HS - V_s)(\rho_f - \rho_a) + \sigma C \right] - (p - p_a) \quad (46a)$$

$$S = A_0(1 + \lambda P)\phi(t, t_{ref}) \qquad (46b)$$

$$P = (p - p_a) \qquad (46c)$$

where $[\xi_{A_0,\xi_\lambda}]^{\mathrm{T}} \sim g_{A_0,\lambda}(\xi_{A_0}, \xi_\lambda)$ is sampled using our bivariate joint PDF modelling approach using the temperature compensation function $\phi(t, t_{ref}) = 1 + \alpha(t - t_{ref})$.

When this Monte Carlo simulation for the TS pressure balance is performed using the data in Appendix A and the Monte Carlo generated pressure data is obtained, then this generated pressure data for each of the cross-floated pressure points 50 MPa, ..., 500 MPa may be compared to the actual generated pressures from the LS pressure balance. Numerical results were performed using a direct simulation single language implementation in GNU Octave with the above generalized sampling scheme for $x^\star$ and $y^\star$ as shown in Table 5. These results were obtained on a computer workstation using an Intel Xeon E5-1650 v3 CPU running at 3.50 GHz with 32 GB of RAM for $M = 500$ Monte Carlo simulation events by directly sampling from the bivariate joint PDF by solving the respective equations for $x^\star$ and $y^\star$.

Alternatively in the special case where the analytical expression for the copula density $c(u, v)$ is explicitly known beforehand then a mixed language approach using R and Octave can offer considerable computational time savings. This approach involves first using R to sample from the Gaussian copula, or any other suitable copula family available from the RStudio available copula library that adequately approximates the bivariate data Kendall tau value, with the specified parameter value θ and then saving the marginal distribution values u and v to a text file as indicated below in Listing 3.

The R computer code shown in Listing 3 writes the u and v values that are consistent with the specified copula density function $c(u, v)$ to a text file CopulaDataUV.txt which is an array of dimension 500×2 where the first column is the u data and the second column is the v data, whilst the corresponding copula density $c(u, v)$ is saved to the text file CopulaDataC.txt. Afterwards the corresponding variate data x and y can then be simply recovered by using one dimensional interpolations from the previously generated ELD based univariate marginal distributions $u = F(x)$ and $v = G(y)$. This mixed language process for $n = 10\,000$ samples using RStudio and GNU Octave

R console input

```
1   setwd("D:/Journals/IJMQE/")
2   getwd()
3
4   theta <- -0.8831274766065601
5
6   # Create Gaussian copula
7   mycopulafunction <- normalCopula(
        theta, dim = 2)
8
9   # {u, v} points (marginal
        distributions)
10  mycopulapoints <- rCopula(500,
        mycopula)
11
12  # c(u, v) points (copula density)
13  mycopuladensity <- dCopula(
        mycopulapoints,
        mycopulafunction)
14
15  write.table(mycopulapoints, "
        CopulaDataUV.txt", sep = "\t",
        row.names = FALSE, col.names =
        FALSE)
16
17  write.table(mycopuladensity, "
        CopulaDataC.txt", sep = "\t",
        row.names = FALSE, col.names =
        FALSE)
```

Listing 3. Computer output used to calculate sampled values from the copula with open source statistical software.

takes approximately 85.36 s to generate, post-process and then recover the variate ξ_{A_0} and ξ_λ sampled data on the previously mentioned Toshiba laptop. As a result the numerical simulations to test the bivariate statistical sampling approach demonstrates that a copula based bivariate quantile joint distribution sampling scheme is computationally feasible for both single computer and well as mixed computer language implementations when used as inputs in subsequent Monte Carlo uncertainty simulations.

In order to verify and validate (V&V) our proposed bivariate joint PDF modelling approach in terms of the standard quality engineering V&V methodology we must compare and contrast these numerical results which utilize the bivariate statistical sampling of the pressure balance',s zero-pressure area and distortion coefficient values with the exact Monte Carlo numerical results previously summarized in Table 1. We can implement this comparison by determining the normalized errors E_n defined as

$$E_n = \frac{x_{lab} - x_{ref}}{\sqrt{U_{lab}^2 + U_{ref}^2}} \qquad (47)$$

for all our generated/cross-floated data points between the LS which generates the known applied pressures and the TS which is cross-floated against the LS. According to this approach x_{ref} would correspond to the LS generated applied pressure $P_k^{(LS)}$, $k = 1, \ldots, 10$ and U_{lab} would correspond to the expanded uncertainty of the generated pressure $U_k^{(LS)}$, whilst x_{lab} would correspond to the TS equivalent generated pressure $P_k^{(TS)}$ using the copula based cross-

floated area parameters where $U_k^{(TS)}$ are the corresponding TS expanded uncertainties for the respective applied pressure.

In order to calculate the normalized errors knowledge of the respective expanded uncertainties $U_k^{(LS)}$ and $U_k^{(TS)}$ are necessary noting that we have used ELDs to summarize the respective Monte Carlo data for both the LS and TS. We recall that the expected value μ and standard uncertainty σ for an ELD are specified as

$$\mu = d + \frac{c(1-a)}{b+1} \qquad (48a)$$

$$\sigma^2 = \begin{cases} \left(\frac{c}{b}\right)^2 \left[\frac{a^2+1}{2b+1} - \frac{2a\Gamma(b+1)^2}{\Gamma(2b+2)} - \left(\frac{a-1}{b+1}\right)^2\right] & if\, b \neq 0 \\ c^2[a^2 + \frac{\pi^2 a}{3} - 2a + 1] & if\, b = 1 \end{cases}$$

$$(48b)$$

As a result knowledge of the ELD parameters immediately gives us the values of the expected values μ_k for the LS and TS but we do not have direct knowledge of the respective expanded uncertainties which must be calculated from the ELD based distribution function. In general for a specified confidence level p, say $p = 0.95$ corresponding to a 95% confidence level, for a possibly non-symmetric and/or skew PDF $g(\eta)$ a value of α is first calculated by minimizing $[G^{-1}(p+\alpha) - G^{-1}(\alpha)]$, however if symmetry with an absence of skewness is assumed as a simplification then $\alpha = \frac{(1-p)}{2}$.

Under these circumstances for a measurand y with a distribution function $G(\eta)$ the confidence interval for the specified confidence level is just $Y_{\min} = G^{-1}(\alpha)$ and $Y_{\max} = G^{-1}(p+\alpha)$. Consequently we may then just approximate the expanded uncertainty as

$$U \approx \frac{1}{2}(Y_{\max} - Y_{\min}) \qquad (49)$$

where

$$Y_{\min} = Q(\alpha) \qquad (50a)$$

$$Y_{\max} = Q(p+\alpha) \qquad (50b)$$

$$Q(\rho) = \begin{cases} d + \frac{c}{b}[a\rho^b - (1-\rho)^b + 1 - a] & if\, b \neq 0 \\ a + c[a\ln(\rho) - \ln(1-\rho)] & if\, b = 0 \end{cases} \qquad (50c)$$

When the above V&V formulation is implemented it results in the data summarized in Figure 7 using $M = 500$ Monte Carlo simulation events by sampling from the bivariate joint PDF. Referring to the normalized errors E_n it is observed that $-1 \leq E_n \leq 1$ for all the applied pressures 50 MPa, $\ldots$, 500 MPa.

As a result of these numerical simulations we therefore conclude that the proposed method of using bivariate quantiles with ELDs for the marginals with parameter based optimized copulas is mathematically and statistically consistent for pressure balance calibrations and intercomparisons at a primary scientific metrology standards level.
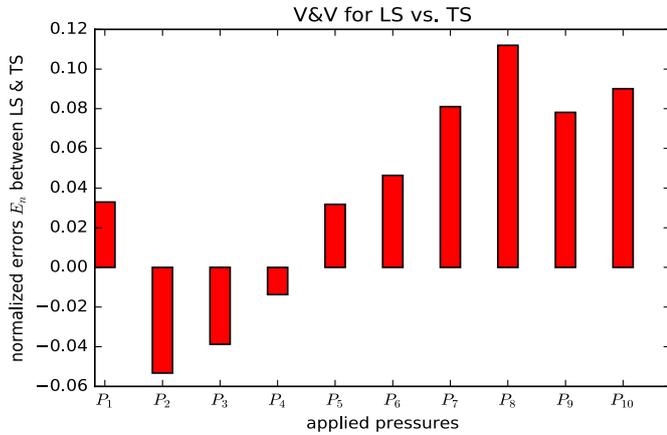
Fig. 7. Verification and validation of proposed method using normalized errors of applied pressures demonstrating that $\|E_n\| \leq 1 \, \forall \, n \in [1, \ldots, 10]$ laboratory standard pressure balance generated pressures and transfer standard pressure balance cross-floated area measurements. The pressures P_1 through to P_{10} correspond to pressures from 50 MPa through to 500 MPa at 50 MPa steps and the E_n values for each of these pressures indicates the corresponding normalized error. When calculating the E_n value for each pressure the known reference pressure P_{ref} is determined from the laboratory standard pressure balance whilst the calibrated pressure P_{cal} is determined from the transfer standard pressure balance using the bivariate copula that models and summarizes the information of the calibrated pressure balance's zero-pressure area A_0 and distortion coefficient λ. Due to the fact that all of the E_n values are smaller than unity this then means that the results and uncertainties that were calculated for the calibrated pressure balance when tested against the reference pressure balance are statistically consistent and hence provides proof that the proposed method of using copulas to model and summarize bivariate probability density functions has been verified and validated.

4 Discussion

In this paper we have investigated the feasibility and utility of extending and applying quantile functions to systems which naturally exhibit bivariate PDF's and considered the particular case of oil pressure balances where the area takes the form $A = A_0 \left(1 + \lambda P\right)$ and knowledge of the bivariate joint PDF $g_{A_0,\lambda}(\xi_{A_0}, \xi_\lambda)$ in terms of the zero-pressure area ξ_{A_0} and distortion coefficient ξ_λ random variables is necessary for elevated pressures. Our research approach involved the implementation of a GS2 multivariate uncertainty analysis of a TS pressure balance where we mathematically formulated an approach to post-process the original Monte Carlo data in order to recover the underlying ξ_{A_0} and ξ_λ bivariate random variable statistical data for further analysis.

Numerical simulations were then considered and performed for a variety of mathematical modelling approaches in order to study how to adequately model, summarize and reconstruct the underlying bivariate statistical uncertainty analysis data. Based on these numerical simulations we sought to investigate the extent to which a combination of univariate ELD quantile functions for the marginal distributions $u = F(x)$ and $v =$

$G(y)$ of $x = \xi_{A_0}$ and $y = \xi_\lambda$ respectively when coupled with a suitable bivariate copula family $C_\theta(u, v)$ selection with an optimized copula parameter θ would be sufficient for pressure calibrations and inter-comparisons.

Results of generated pressures for a variety of conditions were then considered and analysed in order to perform bench-mark studies using conventional metrology statistical tests. When these results were analysed to assess the level of necessary verification and validation measures for the method proposed in this paper it was concluded that a bivariate quantile distribution of a pressure balance PDF of the form $g_{A_0,\lambda}(\xi_{A_0}, \xi_\lambda) = uvc(u, v)$ is indeed sufficient to accurately model and summarize a pressure balance's metrological characteristics at a primary standards scientific metrology level. As a result we conclude that the measurement modelling technique proposed compares very favourably to an exact GS2 UQ analysis and may thus offer benefits to pressure metrologists at national metrology institutes involved in high accuracy client calibration work and in participation in inter-comparisons at a primary scientific standards metrology level.

A potential future topic of metrology research in the area of uncertainty analysis for pressure balances is therefore the possible application of vine copulas for the modelling of multivariate uncertainty analysis data for summarizing GS2 Monte Carlo for higher dimensions. Alternatives to vine copulas are empirical copulas which are copulas note defined in terms of a parameter $\theta = [\theta_1, \ldots, \theta_n]^{\mathrm{T}}$ but which are defined directly in terms of the underlying data. In the case of two dimensional variates $[x_i, y_j]^{\mathrm{T}}$ the bivariate empirical copula when there are n variates is defined as

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) \overset{\text{def}}{=} \frac{\text{pairs } (x, y) \text{ s.t. } x \leq x_i \text{ and } y \leq y_j}{n} \quad (51)$$

The above working definition is usually expressed mathematically as

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{1}{n} \sum_{i=1}^{n} 1\left(\frac{R_i}{n} \leq u_i, \frac{S_i}{n} \leq v_i\right) \quad (52a)$$

$$R_i = \text{rank of point } x_i \quad (52b)$$

$$S_i = \text{rank of point } y_i \quad (52c)$$

$$1\left(\frac{R_i}{n} \leq u_i, \frac{S_i}{n} \leq v_i\right) = \begin{cases} 1 \text{ if condition true} \\ 0 \quad \text{otherwise} \end{cases} \quad (52d)$$

Different types of implementations for an empirical copula are possible as discussed in the copBasic package [39] and include the Hazen, Weibull and Berstein empirical copulas. More recent recent research in the area of empirical copulas by Segers et al. [42] using Berstein polynomials concluded that beta empirical copulas can fully meet the formal mathematical specifications for a copula and that these offer very good performance in terms of estimating bias and variance.

As a result these types of copulas are potentially attractive choices for metrologists for future UQ research in terms of modelling and summarizing higher dimensional

PDFs where multivariate Gaussian PDFs as per the GS2 are problematic due to higher dimensional probability distribution asymmetries and skewness characteristics.

5 Implications and influences

The main implication of this paper is that we have demonstrated that quantile functions may be used to accurately and completely model the bivariate joint probability density distribution function for a pressure balance's effective area in terms of its zero-pressure area A_0 and distortion coefficient λ. As a result the specification of nine numerical parameters, four for each of the marginal distributions plus one for the bivariate copula parameter, now enables pressure metrologists to simply summarize pressure balance bivariate PDF information in calibration certificates and inter-comparison reports, and to have an increased level of confidence in the behaviour and uncertainty of their oil pressure balance laboratory primary standards at elevated pressures which was previously limited due to the complexity of incorporating the uncertainty in the distortion coefficient.

Based on the results reported in this paper the wider influences which are now possible is an increased awareness of the utility of multivariate higher dimensional uncertainty analysis with the GS2, and how copulas may now be used to simply and conveniently summarize higher dimensional uncertainty analysis results in high accuracy calibrations and scientific metrology inter-comparisons in other metrology areas and fields of work.

Appendices
Appendix A: Experimental pressure balance data

We use the representative experimental data set previously reported by Ramnath [19] as indicated in Table A.1 for the laboratory standard (LS) and TS respective mass and density values, Table A.2 for the temperature and ambient conditions, Table A.3 for the utilized mass pieces of the LS, and Table A.4 for the utilized mass pieces of the TS during the cross-floating experiments where points 1 through to 10 correspond to nominal applied pressures such that $P_i/$[MPa] $\approx 50i$ for $i \in [1,10]$.

In this data set the working fluid is assumed to be Di(2)-ethyl-hexylsebacate i.e. DHS oil for which the fluid density is modelled as

$$\rho_f = [912.8 + 0.752(p \times 10^{-6}) - (1.65 \\ \times\ 10^{-3})(p \times 10^{-6})^2 + (1.5 \times 10^{-6})(p \times 10^{-6})^3][1 \\ - (7.8\ \times\ 10^{-4})(t - 20)]$$

as discussed by Kocas et al. [31].The parameters and inputs for the pressure balance model that generates the known pressures are $A_0 = 1.96151 \times 10^{-6}\,\mathrm{m}^2$, $u(A_0) = \pm 9.89581 \times 10^{-11}\,\mathrm{m}^2$, $\lambda = 7.25 \times 10^{-13}\,\mathrm{Pa}^{-1}$, $u(\lambda) = \pm 4.5 \times 10^{-14}\,\mathrm{Pa}^{-1}$, $\alpha = 9.10 \times 10^{-6}\,\mathrm{K}^{-1}$, $u(\alpha) = \pm 0.45 \times 10^{-6}\,\mathrm{K}^{-1}$, $V_s = 0\,\mathrm{m}^3$, $u(V_s) = \pm 10^{-10}\,\mathrm{m}^3$, $\sigma = 31.2 \times 10^{-3}\,\mathrm{N\ m}$, $u(\sigma) = \pm 0.001 \times 10^{-3}\,\mathrm{N\ m}^{-1}$, $H = 173.1 \times 10^{-3}\,\mathrm{m}$, $u(H) = \pm 0.2 \times 10^{-3}\,\mathrm{m}$, $g = 9.7860994\,\mathrm{m\ s}^{-2}$, and $u(g) = 10^{-7}\,\mathrm{m\ s}^{-2}$ respectively where the gravitational acceleration is common to both the pressure balance that generates the pressure as well as the pressure balance which undergoes the cross-floating. For the particular data we assume that the ambient conditions are measured such that $u(t_a) = \pm 0.5\,^\circ C$, $u(p_a) = \pm 15\,\mathrm{Pa}$ and $u(h_a) = \pm 0.05$ respectively, and that the temperature of the respective pressure balances can be measured such that $t = \pm 0.015\,^\circ\mathrm{C}$.

The associated parameters and inputs for the pressure balance model that undergoes the cross-floating are $V_s = 0\,\mathrm{m}^3$, $u(V_s) = \pm 10^{-10}\,\mathrm{m}^3$, $H = 0\,\mathrm{m}$, $u(H) = \pm 0.2 \times 10^{-3}\,\mathrm{m}$ and $\alpha = 1.45 \times 10^{-5}\,\mathrm{K}^{-1}$ respectively where the surface tension for the oil is the same as for the previously specified pressure balance since the working fluid is common to both the LS and TS pressure balances.

Appendix B: Parameter fits of specific copulas

To fit specific types of copulas consider for example explicitly using a Clayton distribution. The corresponding R computer code to fit a Clayton copula and extract the associated parameter values would be:

```
copulaModelClayton <- claytonCopula(dim = 2)
mClayton <- pobs(BivariateData)
fitClayton <- fitCopula(copulaModelClayton,
             mClayton, method = 'itau')
copulaClaytonParameter <- coef(fitClayton)
# Check Kendall's tau value for
# the Clayton copula with theta
tauClayton <- tau(claytonCopula(param =
             copulaClaytonParameter))
```

The parameter value θ associated with a user's particular choice of copula family may then be obtained by simply printing out the value as

```
print(copulaClaytonParameter, digits=16)
```

When the respective fits are completed for a particular choice of copula family we could then compare the Kendall's tau values for a particular fit with that of the optimal copula fit in order to determine how closely the chosen fit matches the actual data correlation by using the following code:

Table A.1. Pressure balance laboratory standard and transfer standard mass and density data.

Mass #	$\dfrac{m^{(LS)}}{[\text{kg}]}$	$\dfrac{u(m^{(LS)})}{[\text{mg}]}$	$\dfrac{\rho^{(LS)}}{[\text{kg m}^{-3}]}$	$\dfrac{u(\rho^{(LS)})}{[\text{kg m}^{-3}]}$	$\dfrac{m^{(TS)}}{[\text{kg}]}$	$\dfrac{u(m^{(TS)})}{[\text{mg}]}$	$\dfrac{\rho^{(TS)}}{[\text{kg m}^{-3}]}$	$\dfrac{u(\rho^{(TS)})}{[\text{kg m}^{-3}]}$
1	0.002735	0.000005	8000	20	0.200007	0.000003	7920	20
2	1.497245	0.000005	8000	20	0.800047	0.000001	7914	23.742
3	0	0.000002	8000	25	5.000008	0.000002	7920	25
4	5.00008	0.000002	8000	25	4.999995	0.000002	7920	25
5	5.00004	0.000002	8000	25	4.999985	0.000002	7920	25
6	5.00005	0.000002	8000	25	4.999983	0.000002	7920	25
7	5.00004	0.000002	8000	25	4.999991	0.000002	7920	25
8	5.00005	0.000002	8000	25	4.999979	0.000002	7920	25
9	5.00004	0.000002	8000	25	4.999981	0.000002	7920	25
10	5.00006	0.000002	8000	25	4.999979	0.000002	7920	25
11	5.00004	0.000002	8000	25	4.99997	0.000002	7920	25
12	5	0.000002	8000	25	4.999979	0.000002	7920	25
13	5	0.000002	8000	25	4.999976	0.000002	7920	25
14	5	0.000002	8000	25	4.999974	0.000002	7920	25
15	5	0.000002	8000	25	4.999979	0.000002	7920	25
16	5	0.000002	8000	25	4.999988	0.000002	7920	25
17	5	0.000002	8000	25	4.99999	0.000002	7920	25
18	5	0.000002	8000	25	5.000003	0.000002	7920	25
19	5	0.000002	8000	25	4.999982	0.000002	7920	25
20	5	0.000002	8000	25	4.999996	0.000002	7920	25
21	5	0.000002	8000	25	4.000001	0.000002	7920	25
22	5	0.000002	8000	25	2.00001	0.000002	7920	25
23	3.00005	0.000002	8000	25	2	0.000002	7920	25
24	2.000002	0.000002	8000	25	1.000001	0.000002	7920	25
25	1.000003	0.000002	8000	25	0.5	0.000002	7920	25
26	0.500001	0.000002	8000	25	0.199998	0.000002	7920	25
27	0.3	0.000002	8000	25	0.199999	0.000002	7920	25
28	0.200002	0.000002	8000	25	0.099999	0.000002	7920	25

Table A.2. Pressure balance laboratory standard and transfer standard temperatures and ambient data.

Point #	$\dfrac{t^{(LS)}}{[^\circ\text{C}]}$	$\dfrac{p_a}{[\text{kPa}]}$	$\dfrac{h_a}{[\%\text{RH}/100]}$	$\dfrac{t_a}{[^\circ\text{C}]}$	$\dfrac{t^{(TS)}}{[^\circ\text{C}]}$
p_1	19.9	86.707	0.63	19	19.33
p_2	20	86.69	0.63	19	19.57
p_3	20	86.67	0.64	19	19.57
p_4	20.1	86.637	0.66	19	19.82
p_5	20.1	86.622	0.66	19	19.98
p_6	20.15	86.606	0.66	19	20.01
p_7	20.3	86.825	0.67	19	20.33
p_8	20.4	86.489	0.68	19	20.28
p_9	20.6	86.442	0.67	19	20.26
p_{10}	20.6	86.435	0.67	19	20.28

Table A.3. Boolean values to determine pressure balance laboratory standard and transfer standard operating conditions (weights for pressure generation).

Mass #	Point #									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	0	0	0	0	0	0	0	0	0	0
4	1	1	1	1	1	1	1	1	1	1
5	0	1	1	1	1	1	1	1	1	1
6	0	1	1	1	1	1	1	1	1	1
7	0	0	1	1	1	1	1	1	1	1
8	0	0	1	1	1	1	1	1	1	1
9	0	0	0	1	1	1	1	1	1	1
10	0	0	0	1	1	1	1	1	1	1
11	0	0	0	0	1	1	1	1	1	1
12	0	0	0	0	1	1	1	1	1	1
13	0	0	0	0	0	1	1	1	1	1
14	0	0	0	0	0	1	1	1	1	1
15	0	0	0	0	0	0	1	1	1	1
16	0	0	0	0	0	0	1	1	1	1
17	0	0	0	0	0	0	0	1	1	1
18	0	0	0	0	0	0	0	1	1	1
19	0	0	0	0	0	0	0	0	1	1
20	0	0	0	0	0	0	0	0	1	1
21	0	0	0	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	0	1
23	1	1	1	1	1	1	1	1	1	1
24	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0
26	1	1	1	1	1	1	1	1	1	1
27	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0

Table A.4. Boolean values to determine pressure balance laboratory standard and transfer standard operating conditions (weights for cross floating).

Mass #	Point #									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	0	1	1	1	1	1	1	1	1	1
5	0	1	1	1	1	1	1	1	1	1
6	0	0	1	1	1	1	1	1	1	1
7	0	0	1	1	1	1	1	1	1	1
8	0	0	0	1	1	1	1	1	1	1
9	0	0	0	1	1	1	1	1	1	1
10	0	0	0	0	1	1	1	1	1	1
11	0	0	0	0	1	1	1	1	1	1

Table A.4. (continued).

Mass #	Point #									
12	0	0	0	0	0	1	1	1	1	1
13	0	0	0	0	0	1	1	1	1	1
14	0	0	0	0	0	0	1	1	1	1
15	0	0	0	0	0	0	1	1	1	1
16	0	0	0	0	0	0	0	1	1	1
17	0	0	0	0	0	0	0	1	1	1
18	0	0	0	0	0	0	0	0	1	1
19	0	0	0	0	0	0	0	0	1	1
20	0	0	0	0	0	0	0	0	0	1
21	1	1	1	1	1	1	1	1	1	1
22	0	0	0	0	0	0	0	0	0	1
23	0	0	0	0	0	0	0	0	0	1
24	0	0	0	0	0	0	0	0	0	1
25	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0

```
> print(tauOptimal, digits=16)
                  x                 y
x  1.000000000000000 -0.690606986191983
y -0.690606986191983  1.000000000000000
> print(tauClayton, digits=16)
             param
-0.6906069861919829
> print(copulaClaytonParameter,digits=16)
             param
-0.8169929402072855
```

In our particular case using Kendall's tau as an indication of the variate correlation from the above VineCopula library computer results we observe that the Clayton copula approximation diverges from the optimal Gaussian copula and therefore it would not be beneficial to use a Clayton's copula to model the pressure balance joint distribution. If simulations are performed on a 64-bit computer using double floats then the practical numerical precision is usually 16 digits unless variable precision arithmetic (VPA) libraries are used as previously discussed.

In the event that the copula function determined either from an optimal fit using the VineCopula optimal fit routine or a custom copula fit based on physical or mathematical requirements is too complicated to algebraically differentiate then a numerical differentiation of a function $u(x, y)$ may be performed so that

$$\frac{\partial^2 u(x_i, y_j)}{\partial x \partial y} \approx \frac{1}{4(\Delta x)(\Delta y)} \times [u_{i+1,j+1} - u_{i+1,j-1} - u_{i-1,j+1} + u_{i-1,j-1}] + \mathcal{O}\left((\Delta x)^2, (\Delta y)^2\right)$$

The copula density may then be explicitly calculated directly from the analytical copula function $C(u, v; \theta)$ by using the above numerical approximation so that

$$c(u, v; \theta) = \frac{\partial^2 C(u, v; \theta)}{\partial u \partial v}$$

As a result as long as an analytical algebraic expression for the copula function $C(u, v; \theta)$ is specified along with the respective value for the parameter θ in for example a calibration certificate or regional/international intercomparison report, then the joint PDF distribution information for the measurement instrument is completely specified through the reported copula function and appropriate parametrizations for the associated marginal distributions.

Extensions to non-parametric copulas in terms of empirical copulas are possible using the R package copBasic [39]. More recent research in the area of empirical copulas tends to favour an empirical beta copula as discussed by Segers et al [42] due to certain statistical technical reasons, and the use of empirical beta copulas may therefore in principle be considered for utilization by metrologists for modelling higher dimensional joint probability density distributions where multivariate Gaussian probability density functions are problematic.

References

1. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OILM, Guide to the expression of uncertainty in measurement. Technical report, Bureau International de Poids et Mesures (BIPM), 2008, www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

2. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OILM, Evaluation of Measurement data − Supplement 1 to the "Guide to the expression of uncertainty in measurement" − propagation of distributions using a Monte Carlo method. Technical report, Bureau International de Poids et Mesures (BIPM), 2008, http://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf

3. P.M. Harris, C.E. Matthews, M.G. Cox, A.B. Forbes, Summarizing the output of a Monte Carlo Method for uncertainty evaluation, Metrologia **51**, 243–252 (2014)

4. V. Ramnath, Application of quantile functions for the analysis and comparison of gas pressure balance uncertainties, Int. J. Metrol. Qual. Eng. **8**, 1–18 (2017)

5. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OILM, Evaluation of measurement data − supplement 2 to the "Guide to the expression of uncertainty in measurement" − extension to any number of output quantities. Technical report, Bureau International de Poids et Mesures (BIPM), 2011, http://www.bipm.org/utils/common/documents/jcgm/JCGM_102_2011_E.pdf

6. R.S. Dadson, S.L. Lewis, G.N. Peggs, *The Pressure Balance: Theory and Practice* (HMSO, London, 1982)

7. A. Picard, R.S. Davis, M. Glaser, K. Fujii, Revised formula for the density of moist air (CIPM-2007), Metrologia **45**, 149–159 (2008)

8. M. Dehghan, M. Hajarian, Some derivative free quadratic and cubic convergence iterative formulas for solving nonlinear equations, Comput. Appl. Math. **29**, 19–30 (2010)

9. W. Bich, Interdependence between measurement uncertainty and metrological traceability, Accredit. Qual. Assur. **14**, 581–586 (2009)

10. M.G. Cox, P.M. Harris, Software support for metrology best practice guide no. 6–uncertainty evaluation. Technical report, National Physical Laboratory (United Kingdom), 2006. NPL Report DEM-ES-011 (ISSN 1744-0475), http://publications.npl.co.uk/npl_web/pdf/dem_es11.pdf

11. R. Palencar, G. Wimmer, M. Halaj, Determination of the uncertainties and covariances in the calibration of the set of weights, Meas. Sci. Review **2**, 9–20 (2002)

12. EURAMET, Calibration of Pressure Balances. Technical report, European Association of National Metrology Istitutes, 2011. Guide 3 Version 1.0, ISBN 978-3-942992-02-2

13. W. Sabuga, G. Molinar, G. Buonanno, T. Esward, J.C. Legras, Finite element method used for calculation of the distortion coefficient and associated uncertainty of a PTB 1 GPa pressure balance − EUROMET project 463, Metrologia **43**, 311–625 (2006)

14. S. Dogra, S. Yadav, A.K. Bandyopadhyay, Computer simulation of a 1.0 GPa piston-cylinder assembly using finite element analysis (FEA), Measurement **43**, 1345–1354 (2010)

15. M.G. Cox, B.R.L. Siebert, The use of a Monte Carlo method for evaluating uncertainty and expanded uncertainty, Metrologia **43**, S178–S188 (2006)

16. R. Willink, Representing Monte Carlo output distributions for transferability in uncertainty analysis: modelling with quantile functions, Metrologia **46**, 154–166 (2009)

17. K. Goda, Statistical modeling of joint probability distribution using copula: application to peak and permanent displacement seismic demands, Struct Saf. **32**, 112–123 (2010)

18. Scipy v0.19.0 reference guide − scipy.interpolate.bspline, https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.BSpline.html#scipy.interpolate.BSpline (accessed: 2017 /02/04)

19. V. Ramnath, Determination of pressure balance distortion coefficient and zero-pressure effective area uncertainties, Intern. J. Metrology Quality Eng. **2**, 101–119 (2011)

20. M. Krystek, M. Anton, A least-squares algorithm for fitting data points with mutually correlated coordinates to a straight line, Meas. Sci. Technol. **22**, 035101 (2011)

21. X.-S. Tang, D.-Q. Li, C.-B. Zhou, K.-K. Phoon, L.-M. Zhang, Impact of copulas for modelling bivariate distributions on system reliability, Struct. Saf. **44**, 80–90 (2013)

22. C. Genest, A.-C. Favre, Everything you always wanted to know about copula modeling but were afraid to ask, J. Hydrol. Eng. **12**, 347–368 (2007)

23. X. He, P. Ng, S. Portnoy, Bivariate quantile smoothing splines, J. R. Stat. Soc. Ser. B: Stat. Methodol. **60**, 537–550 (1998)

24. D.H. Bailey, J.M. Borwein, Hight-precision arithmetic in mathematical physics, Mathematics **3**, 337–367 (2015)

25. D.R. White, P. Saunders, The propogation of uncertainty with calibration equations, Meas. Sci. Technol. **18**, 2157–2169 (2007)

26. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, in *Numerical Recipes: The Art of Scientific Computing*, 3rd edn. (Cambridge University Press, 2007)

27. W.G. Gilchrist, in *Statistical Modelling with Quantile Functions*, edited by Chapman and Hall, 1st edn. (2000) ISBN 1-58488- 174–7

28. P. Chaudhuri, On a geometric notion of quantiles for multivariate data, J. Am. Stat. Assoc. **91,** 862–872 (1996), http://www.jstor.org/stable/2291681

29. H. Kajikawa, T. Kobata, K. Ide, A. Ooiwa. Precise determination of the jacket pressure coefficient of controlled clearance pressure balances, in 3rd IMEKO TC16 International Conference on Pressure Measurement (Curran Associates, Inc., 2007), pp. 125–134, ISBN 978-1-61567-635-4

30. H. Kajikawa, K. Ide, T. Kobata, Methods of precisely estimating the jacket pressure coefficient of controlled clearance piston cylinders at pressures up to 1 GPa, Metrologia **48**, 352–358 (2011)

31. I. Kocas, W. Sabuga, M. Bergoglio, A. Eltaweel, C. Korasie, P. Farar, J. Setina, B. Waller, Y. Durgut, Final report on key comparison euramet.m.p-k13 in the range 50 mpa to 500 mpa of hydraulic gauge pressure, Metrologia **52**, 07008 (2015)

32. U. Diwekar, A. David, in *BONUS Algorithm for Large Scale Stochastic Nonlinear Programming Problems*, edited by U. Diwekar, A. David (Springer, 2015), pp. 27–34, ISBN 978-1-4939-2281-9, http://link.springer.com/content/pdf/10.1007%2F978-1-4939-2282-6_3.pdf

33. E. Larsson, B. Fornberg, A numerical study of some radial basis function based solution methods for elliptic PDEs, Comput. Math. Appl. **46**, 891–902 (2003)

34. I. Georgieva, C. Hofreither, An algorithm for low-rank approximation of bivariate functions using splines, J. Comput. Appl. Math. **310**, 80–91 (2017), DOI:10.1016/j.cam.2016.03.023

35. X. Luo, P.V. Shevchenko, The t copula with multiple parameters of degrees of freedom: bivariate charactersistics and application to risk management, Quant. Finance **10**, 1039–1054 (2010)

36. J. Yan, Enjoy the joy of copulas: with a package copula, J. Stat. Softw. **21**, 1–21 (2007)

37. I. Kojadinovic, J. Yan, Modelling multivariate distributions with continous margins using the copula **R** package, J. Stat. Softw. **34**, 1–20 (2010)

38. M. Hofert, I. Kojadinovic, M. Maechler, J. Yan, The Comprehensive R Archive Network − Multivariate Dependence with Copulas Version 0. 999-15, 2017, https://cran.r-project.org (accessed: 05/29/ 2017)

39. W. Asquith, The Comprehensive R Archive network − General Bivariate Copula Theory and Many Utility Functions Version 2.0.4, 2017, https://cran.r-project.org (accessed: 05/29/2017)

40. U. Schepsmeier, E. C. Brechmann, B. Graeler, T. Nagler, T. Erhardt, C. Almeida, A. Min, C. Czado, M. Hofmann, M. Killiches, H. Joe, T. Vatter, The Comprehensive R Archive Network − Statistical Inference of Vine Copulas Version 2.1.2, 2017, https://cran.r-project.org (accessed: 05/29/2017)

41. C. Meyer, The bivariate normal copula, Commun. Stat.: Theory Methods **42**, 2402–2422 (2013)

42. J. Segers, M. Sibuya, H. Tsukahara, The empirical beta copula, J. Mult ivar. Anal. **155**, 35–51 (2017)