Characterising 2D Beam Distributions:

Generalised Normal

Distributions and Copulas

**Duncan Scott**

**Astec**

**30-9-2016**

**Abstract**

Beam-size characterisation is an important technique for comparing simulated, experimental and theoretical distributions of particle beams. Mostly, the characterisation is done in terms of rms parameters and Normal distributions [[[1]](#endnote-1), [[2]](#endnote-2)]. This may not be the best way to characterise non-Normal distributions, such as those observed on VELA. Here we will begin to investigate other characterisations including Generalised Normal Distributions (GND) that include extra fitting parameters to describe the distribution skewness and kurtosis. These extra parameters allow for better fits to measured distributions as evidenced by lower fit residuals and quantile-quantile plots. Analytic extensions of 1D GNDs to 2D, although possible, does not seem trivial therefore new techniques were investigated. One technique that looks promising is the application of ‘*copulas*’. Copulas are a measure of the dependency structures between the marginal (projected) distributions. An example analytic copula is used to fit a beam distribution with GND marginals gaining ~10% improvement in terms of the fit residuals over previous methods. This technique will be considered in future characterisations of measured and simulated Vela / Clara beams.

# Introduction

Previously, work has been done on finding robust methods to take a raw camera image, find the beam distribution and characterise it with a bivariate normal distribution [1,2]. Similar techniques are used for discrete particle distributions from simulations, they will be considered in a future note.

## Review of Bivariate Normal Distribution (BVN)

The BVN distribution is completely characterised by a sample mean vectorand a covariance matrix comprised of the variances, and the covariance, and , we assume giving:

The standard deviations, , are the square root of the variances:

Another important term is the population correlation coefficient, For the random variables and , is defined using the covariance and the standard deviations:

These characterisations of a BVN are linear transforms of an identity matrix with variances = 1 and covariance = 0. Contours of the probability distribution function are elliptical with the ellipse major and minor axes (often denoted as axes 1 and 2) rotated an angle relative to the axes (when there is non-zero covariance). The angle can be found by applying to the components of the Eigenvectors of (taking care to use for the correct quadrant.) Figure 1 shows an example BVN distribution and it’s projections onto the axes. The projected distributions are often known as the *marginal probability distributions,* they are probability distributions of one variable, their combination is known as the *joint probability distribution* and is a function of two variables.

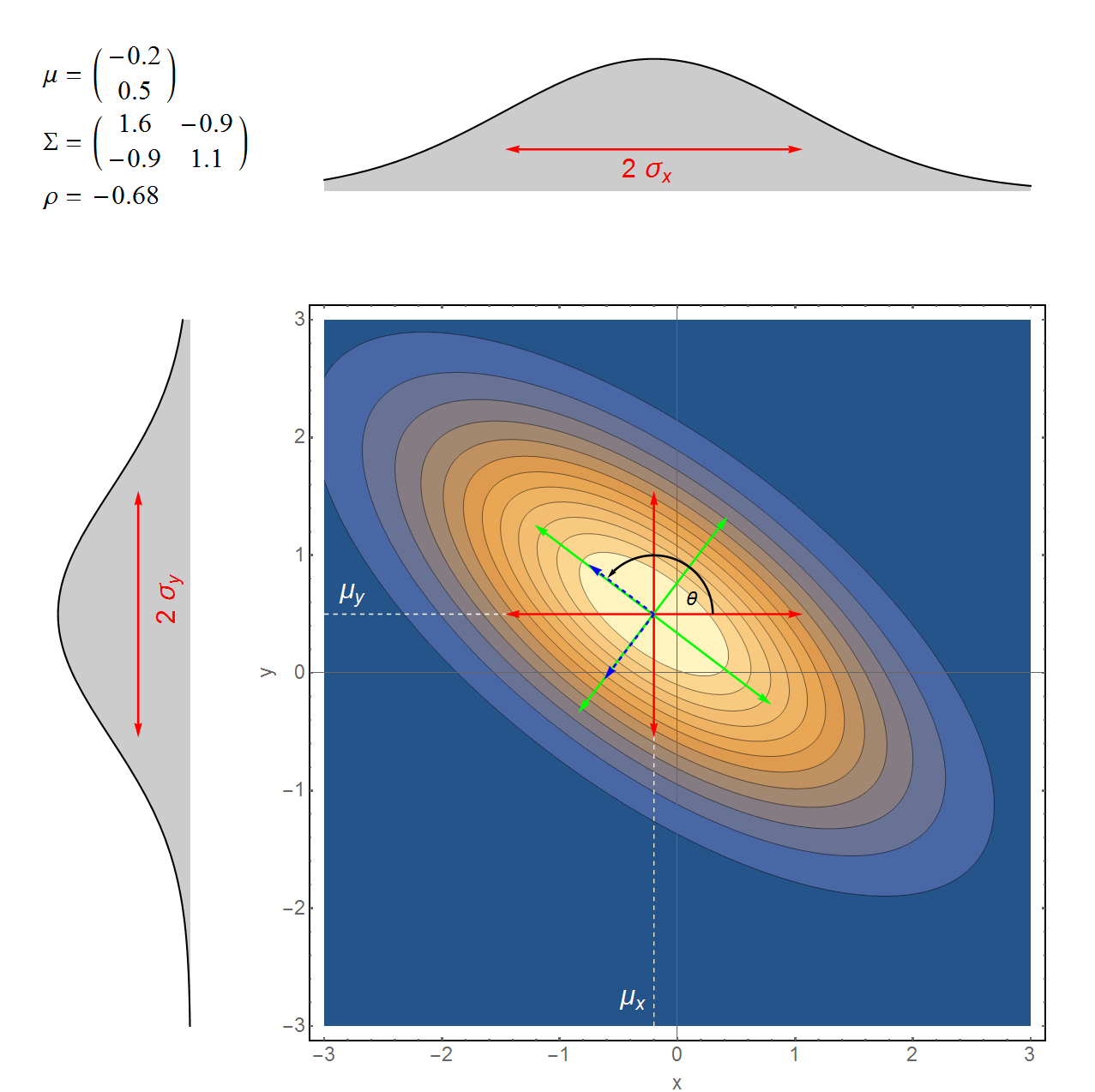


Figure : Example BVN, showing **,** the marginal 2 widths, (red), ellipse major and minor axes widths (green), eigenvectors of (blue), and their angle with the axes, .

### A VELA Example

Figure 2 gives an example of a real VELA image and the BVN fit using the existing method [red]. The marginal distributions are similar to a Normal distribution, but perhaps we can find a better function to fit the data. We will consider two generalisations of the Normal distribution that introduce shape parameters that add a *skewness* or change the *kurtosis* of the Normal distribution.

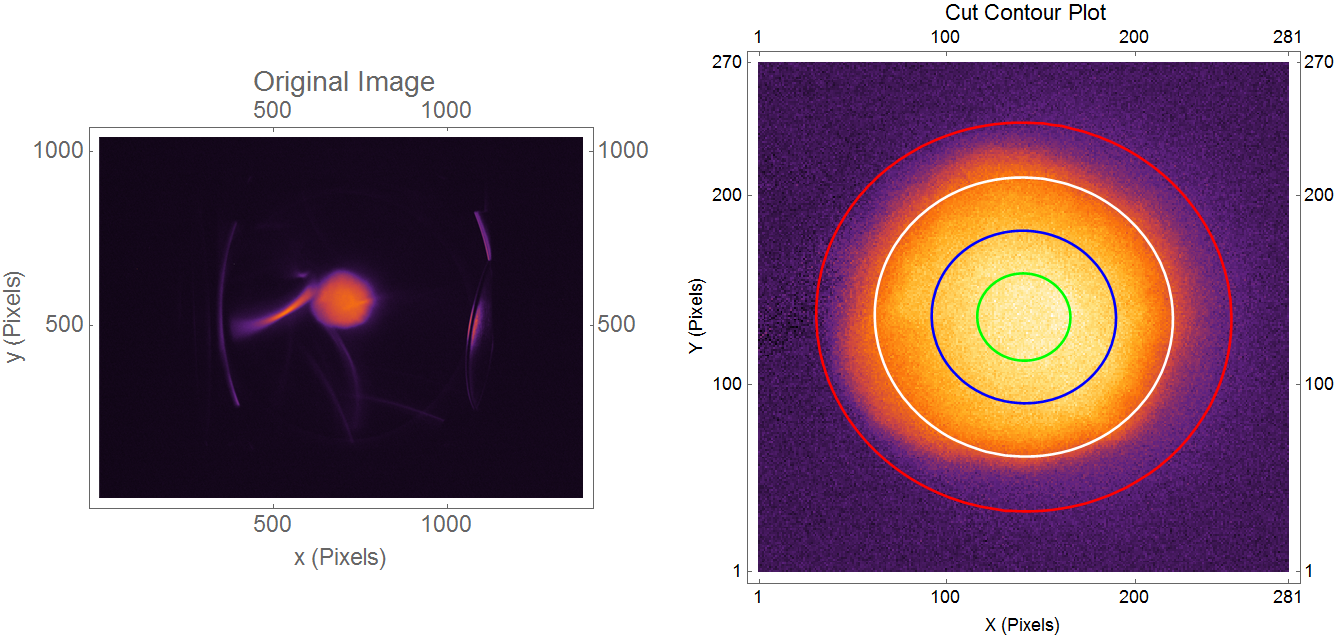


Figure : VELA camera image (left) cropped and fitted image (right).

# Generalised Normal Distributions

Skewness is a measure of the asymmetry in the distribution, it is related to the 3rd order moments. A Normal distribution has skewness = 0. Kurtosis is a measure of the ‘taildness’ or how peaked the distribution is, it is related to the 4th order moments. The Kurtosis of a Normal distribution is 3. We will consider two distributions known as the Skewed Normal Distribution () and Exponential Power Distribution (). Their probability density functions (PDFs) are given below and some examples plotted in Figure 3:

where is the shape a parameter and is the complimentary error function.

where is the shape parameter and is the gamma function.

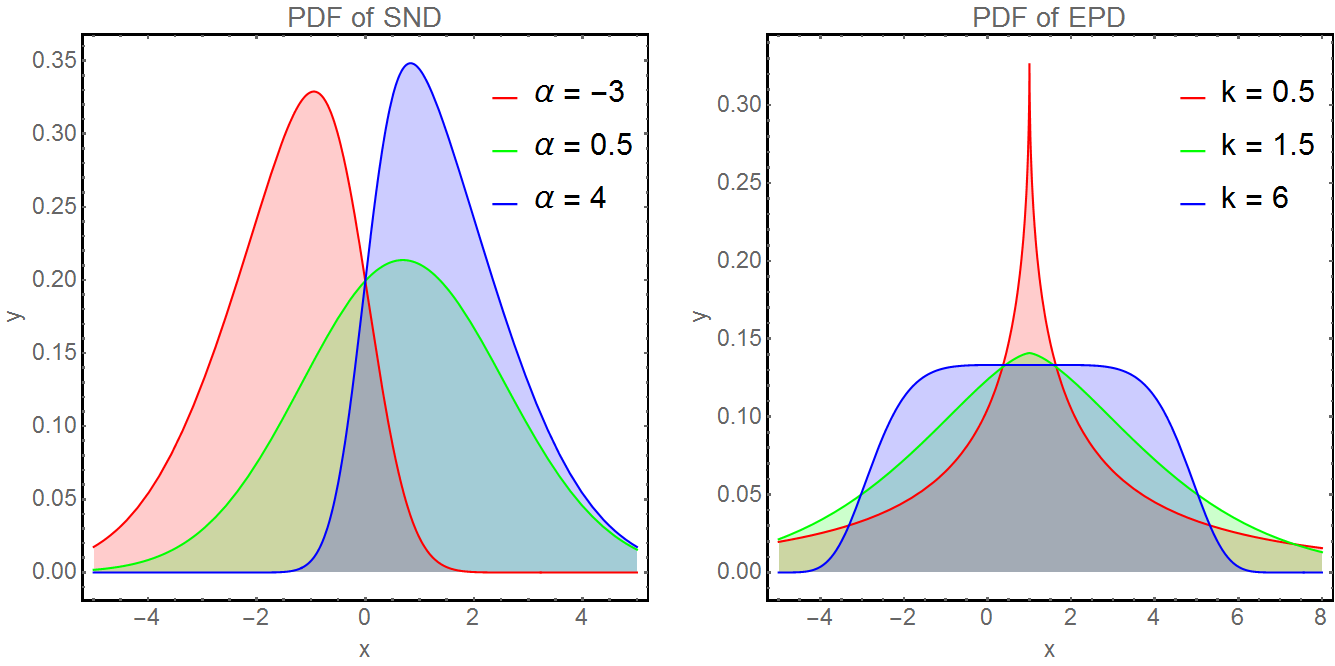


Figure : PDF plotted for different SND and .

## Fitting to Real Data

We can use these distributions for fitting and compare with a Normal Distribution, shown in Figure 4 and Table 1 for the marginals of Figure 1.

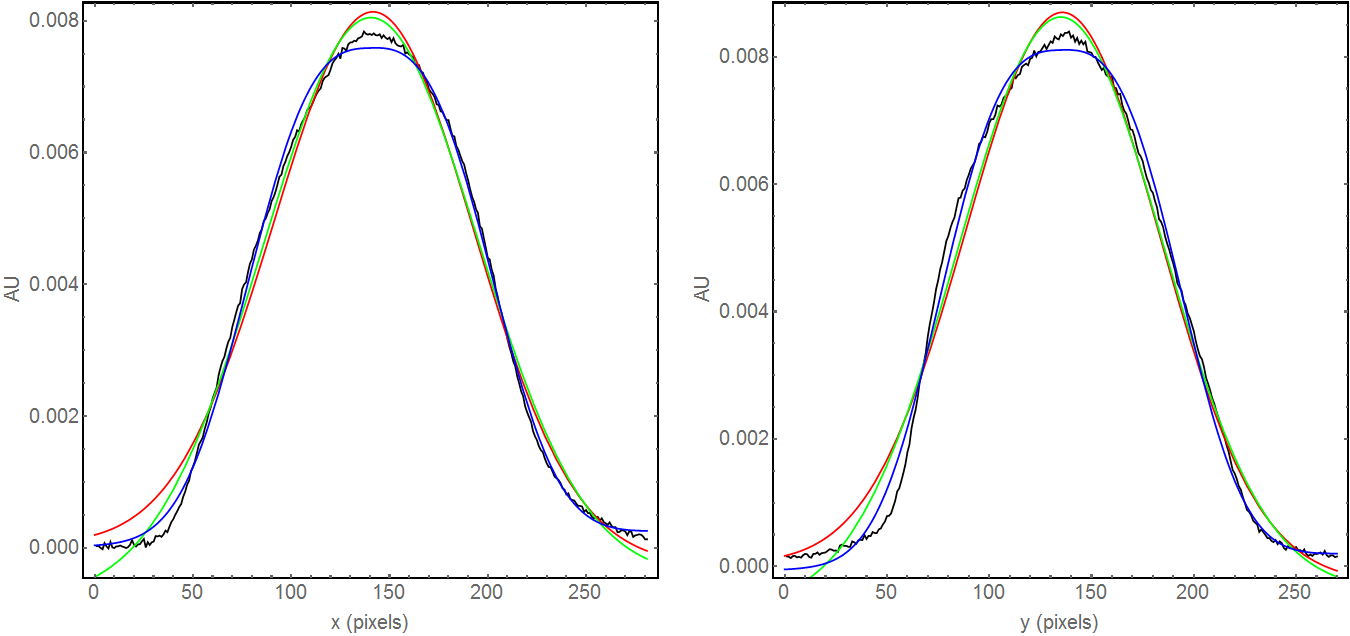


Figure : x and y marginal (black lines) and fits for Normal (red) SND (green) and EPD (blue) distributions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter** |  |  |  |  | **Residual** |
| **Unit** | **pixels** | **pixels** |  |  | **% of Normal Value** |
| x Normal | 142 | 50.8 | - | - | 100 |
| X SND | 146 | 55.0 | -0.14 | - | 80 |
| X | 140 | 50.2 | - | 2.88 | 43 |
| y Normal | 136 | 47.4 | - |  | 100 |
| Y SND | 136 | 50.2 | -0.03 | - | 89 |
| Y | 135 | 47.3 | - | 2.93 | 51 |

Table : fitting parameters for distributions in Figure 4.

## Quantile-Quantile (QQ) Plots

Another visual way to compare fitted distributions with the data is with a QQ*-*plot. A QQ plot compares two data sets and is a plot of the quantiles of the first data set against the quantiles of the second:

“*A QQ plot is used to compare the shapes of distributions, providing a graphical view of how properties such as location, scale, and skewness are similar or different in the two distributions. A point  on the plot corresponds to one of the quantiles of the second distribution ( – coordinate) plotted against the same quantile of the first distribution ( - coordinate). Thus the line is a parametric curve with the parameter which is the (number of the) interval for the quantile.*”[[[3]](#endnote-3)]

A quantile is the fraction of points below the given value. The 0.3 quantile is the value at which 30% percent of the data fall below and 70% fall above that value. QQ plots can be used to check:

* Do data sets come from populations with a common distribution?
* Do data sets have common location and scale?
* Do data sets have similar distributional shapes?
* Do data sets have similar tail behaviour?

The QQ plots for the projected data and 3 types of fits are shown in Figure 5. If the models were a perfect match to the data all the points would lie along the reference line of the plot. By inspection we see non-Normal tails that are better modelled with an .

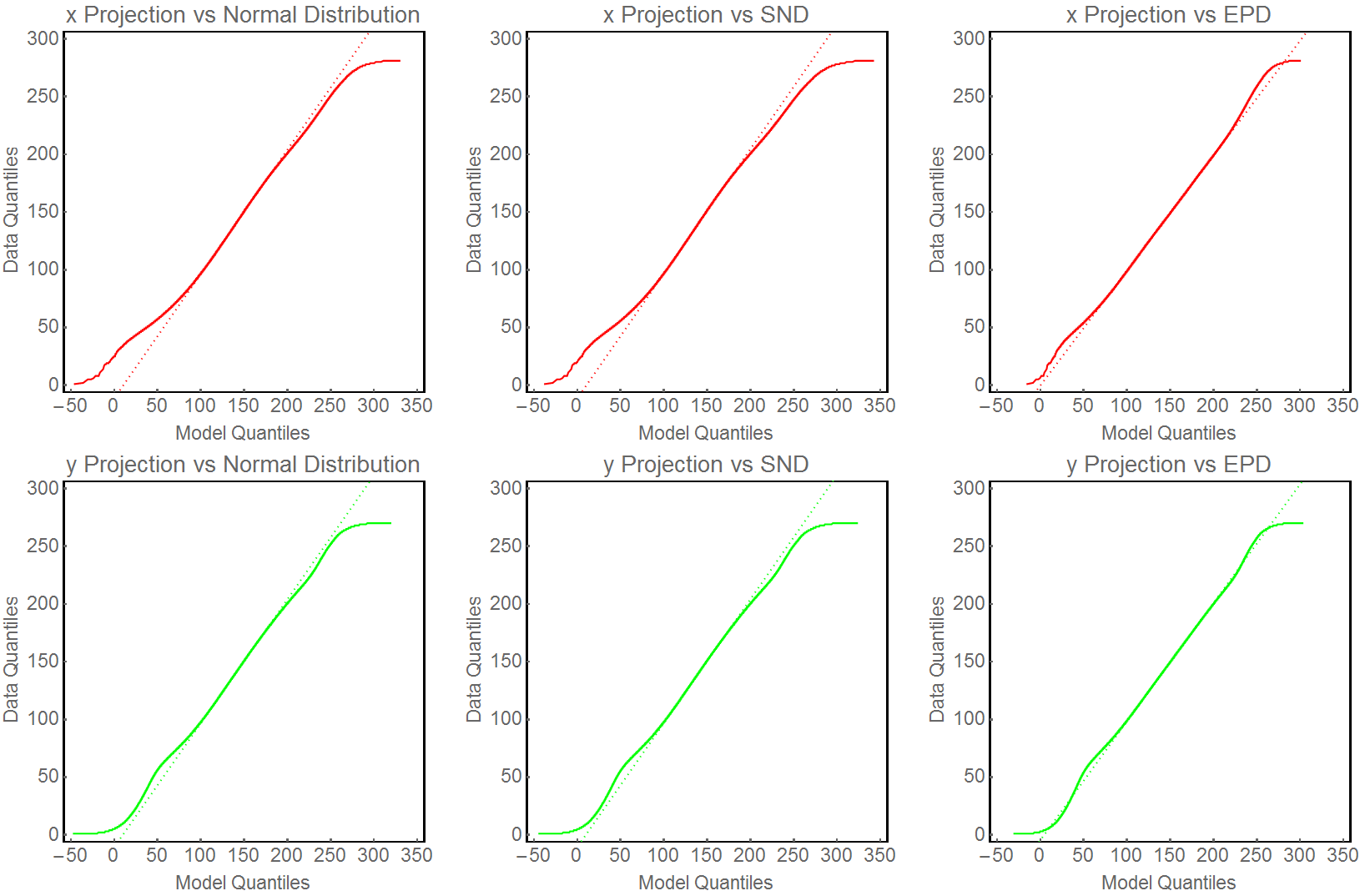


Figure : QQ-plots for projected data and different fitted models.

# Dependency Structures and Copulas

We have seen that there are analytic distributions that better match the projections by including an extra term to fit, the *shape-parameter*. However, analysing the marginal distributions takes no account of their dependency with one another (their dependency structure). The *Joint Probability Distribution* includes the dependency between two marginal distributions. Previous work included the dependency of Normal marginals through the Bivariate Normal Distribution, outlined above. Devising analytic expressions for General Normal Distributions in two dimensions, i.e. generalised normal marginal with a given dependency structure, does not seem to be a trivial task, although it probably can be done at some level. However, our measured distributions probably have no simple analytic form. Using the concept of *copulas* will allow us to extractfrom the data the dependency structure without reference to the marginals.

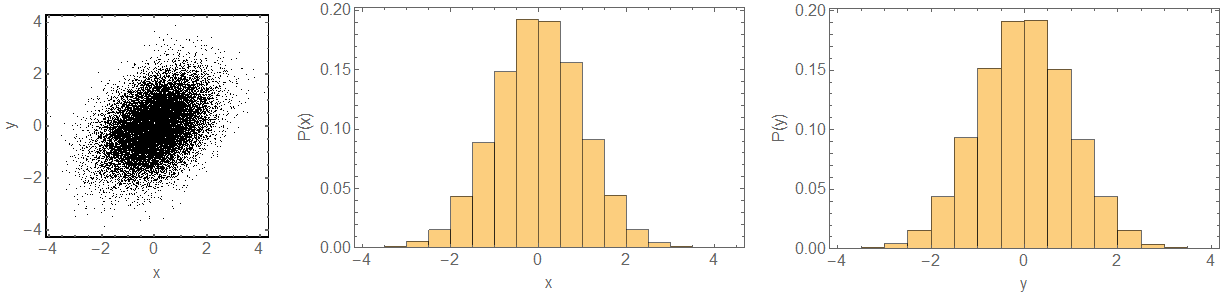
*“A copula is a multivariate probability distribution for which the marginal probability distribution of each variable is uniform. Copulas are used to describe the dependence between random variables.”*[[[4]](#endnote-4)]

The copula describes the dependency structure between the marginals. The marginal distributions of the copula are uniform in the [0,1] interval. For a 2D distribution the copula is a surface on the unit square. A short example will illustrate the method.

## A Toy Example with a Bivariate Normal Distribution

Assuming:

let be a random variable of samples froma BVN distribution. Figure 6 shows and histograms of and .

 Figure : random sampling of a BVN, , and histograms of and .

The Spearman, and Pearson, correlation matrices of are [[[5]](#endnote-5)]:

Next we use the fact that:

“*If is a random variable with distribution then is uniformly distributed in the interval [0, 1].*“

We make a transformation such that the marginals of lie on the [0,1] interval. This is achieved by using the cumulative distribution function of the marginal, which we already know [[[6]](#endnote-6)]. The result is shown in Figure 7, notice how the marginal distributions are uniform over the [0,1] interval and the correlation of has been preserved. Basically, we are left with what is often referred to as the *dependence structure* (LHS of Figure 7).

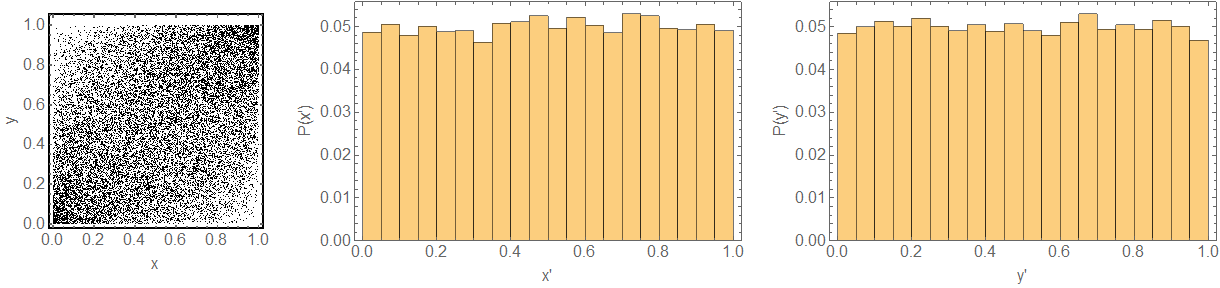


Figure : , and histograms of the transformed and

We are now free to transform the marginals to another distribution using , *and the correlation will be preserved*. For example, using an with and and with gives Figure 8.

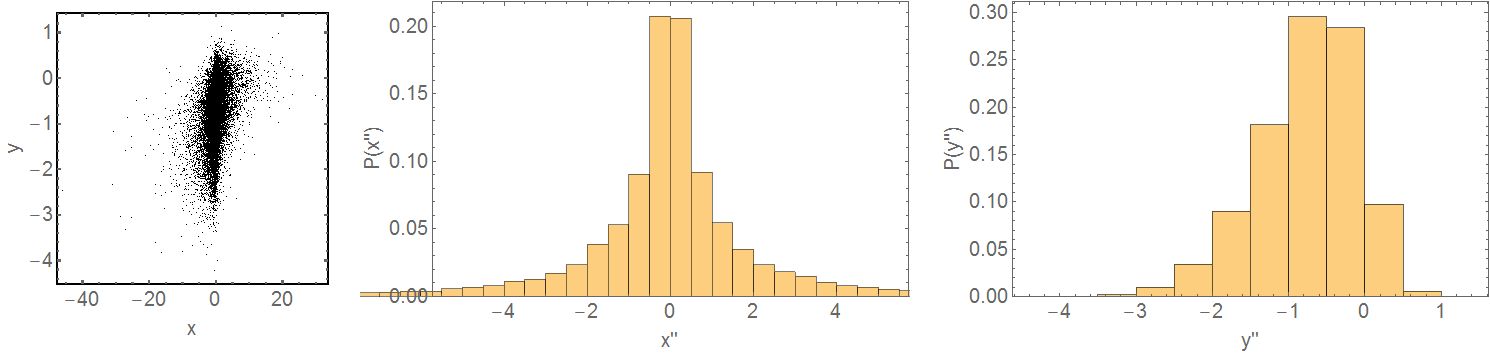


Figure histograms of the transformed and

By starting from a BVN sample we have built a sample with a chosen and fixed dependence structure and, basically, ***arbitrary marginals***.

## A Real-Life Example

Using the data from Figure 2 and Table 1 we can try and construct a distribution that *better fits* the measurement. For demonstration purposes we will consider an analytic bi-normal copula with the same correlation factor as the measured beam distribution and the marginal fits. A comparison between the distribution derived from the copula, the BVN fit and the measurement is shown in Figure 9. The distribution derived from the copula has a 10% lower fit residual. It seems with further work and more understanding more improvements may be gained.

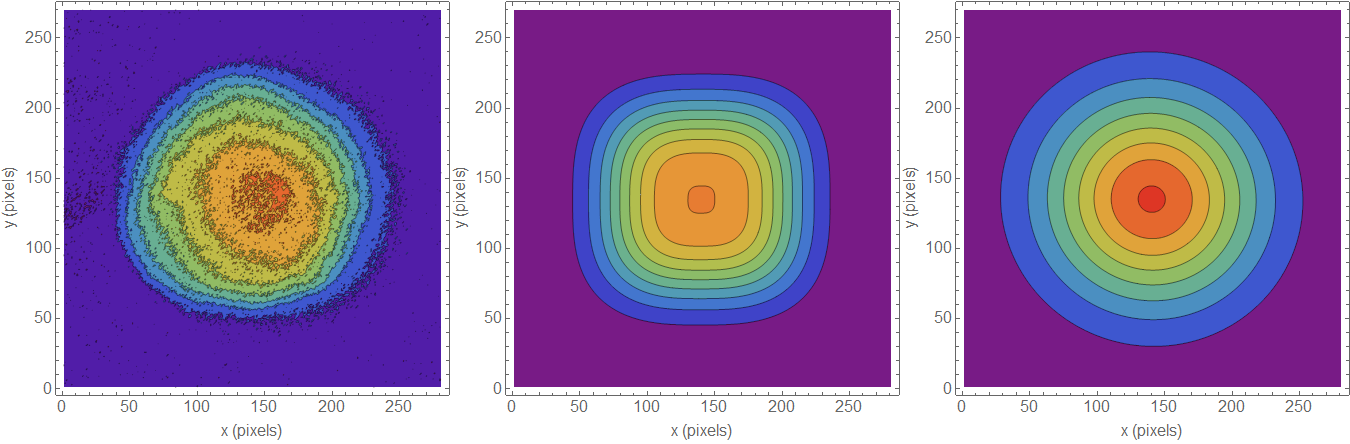
****

Figure 9: Contour plots of measured data (left) copula derived distribution (middle) and original BVN distribution (right).

# Summary and Further Work

Much of the work presented here is exploratory: we are trying to find more accurate characterisations for measured and simulated beam distributions. This is worthwhile due to the non-Normal distributions we typically see on machines such as VELA. More accurate characterisations have been achieved by increasing the number of fitting parameters, which works well for 1-D marginal distributions. However extending these to more complicated 2D analytic distributions is not easy. One way this can be achieved is by transforming a copula to the desired marginals and has been demonstrated for an example beam distribution. Further work should investigate different copulas to generate analytic expressions for a general 2D distribution.

# References

1. [] D. J. Scott *Calculating Beam Sizes From Screen Images on VELA* [↑](#endnote-ref-1)
2. [] D. J. Scott & M. Toplis Method of Estimating Beam Size from VELA Screen Images [↑](#endnote-ref-2)
3. [] <https://en.wikipedia.org/wiki/Q%E2%80%93Q_plot> [↑](#endnote-ref-3)
4. [] <https://en.wikipedia.org/wiki/Copula_(probability_theory)> [↑](#endnote-ref-4)
5. [] The Spearman correlation is less sensitive than the Pearson correlation to strong outliers that are in the tails of samples. This is because limits the outlier to the value of its rank. In this example, we are going to use an elliptical copula, the difference is not important, when we use non-elliptical copulas we should use . [↑](#endnote-ref-5)
6. [] For measured data with an arbitrary marginal the cumulative distribution function can easily be constructed. [↑](#endnote-ref-6)