

A dependent circular-linear model for multivariate biomechanical data: Ilizarov ring fixator study

Statistical Methods in Medical Research

2024, Vol. 33(9) 1660–1672

© The Author(s) 2024



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09622802241268654

journals.sagepub.com/home/smm

Priyanka Nagar¹ , Andriette Bekker^{2,3}, Mohammad Arashi⁴,
Cor-Jacques Kat⁵ and Annette-Christi Barnard⁶

Abstract

Biomechanical and orthopaedic studies frequently encounter complex datasets that encompass both circular and linear variables. In most cases (i) the circular and linear variables are considered in isolation with dependency between variables neglected and (ii) the cyclicity of the circular variables is disregarded resulting in erroneous decision making. Given the inherent characteristics of circular variables, it is imperative to adopt methods that integrate directional statistics to achieve precise modelling. This paper is motivated by the modelling of biomechanical data, that is, the fracture displacements, that is used as a measure in external fixator comparisons. We focus on a dataset, based on an Ilizarov ring fixator, comprising of six variables. A modelling framework applicable to the six-dimensional joint distribution of circular-linear data based on vine copulas is proposed. The pair-copula decomposition concept of vine copulas represents the dependence structure as a combination of circular-linear, circular-circular and linear-linear pairs modelled by their respective copulas. This framework allows us to assess the dependencies in the joint distribution as well as account for the cyclicity of the circular variables. Thus, a new approach for accurate modelling of mechanical behaviour for Ilizarov ring fixators and other data of this nature is imparted.

Keywords

Circular-linear data, directional statistics, fracture displacement, multivariate models, vine copulas, well-being

1 Introduction

External fixators are medical instruments used to immobilise fractures or heavy damage to the bone structure. Healing of a fracture is influenced by the amount of strain at the fracture site.¹ The strain is related to the amount of motion of the fracture under loading. The motion of the fracture relies on the combination of rings, bars, pins and wires. Different configurations of external fixators can be compared by measuring the displacement of the fracture. Fracture healing is inevitably influenced by the complex interplay of biology and biomechanics, that is, inter-fragmentary motion and inter-fragmentary biomechanics. The stiffness of a construct of an external fixator is determined by the configuration of the hardware. The construct's configuration may lead to different inter-fragmentary motion and it is, therefore, important to obtain the most appropriate

¹Faculty of Economic and Management Sciences, Department of Statistics and Actuarial Science, Stellenbosch University, Stellenbosch, South Africa

²Department of Statistics, University of Pretoria, Pretoria, South Africa

³Centre for Environmental Studies, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa

⁴Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

⁵Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa

⁶Walk-A-Mile Centre for Advanced Orthopaedics, Centurion, South Africa

Corresponding author:

Priyanka Nagar, Faculty of Economic and Management Sciences, Department of Statistics and Actuarial Science, Stellenbosch University, Bosman Street, Stellenbosch 7602, South Africa.

Email: pnagar@sun.ac.za

construct for optimal healing. Iobst et al.² provided an overview of the various circular external fixators based on current choices available within the field, accompanied by a comprehensive comparative assessment of seven prevalent hexapod circular external fixator systems. Pertinent attributes pertaining to the hardware, software components, and educational potentials of each system are meticulously delineated. Noteworthy is the systematic refinement of information procured from system manufacturers, subject to rigorous and unbiased editorial scrutiny to align with the structural requirements of this review, devoid of any extraneous subjective commentary or recommendations. The authors of the review contend that the development of varied systems has had a positive impact on the advancement of this technical domain.

In the literature various studies can be found which compare configurations of constructs to determine strain at a fracture site and thus the most appropriate configuration for optimal healing.³⁻⁷ To the best of the authors' knowledge, all studies in this area follow a similar statistical procedure: the assumption of independence among the variables and all variables are considered linear variables. There are two main shortfalls of this approach: (i) disregard of the dependence structure between the variables which may impair the accuracy of the model and (ii) the cyclicity of the rotational variables is neglected and mistreated as a linear variable which may produce misleading results. In numerous studies, the analysis of variance (ANOVA) test is frequently employed to facilitate comparisons among various fixator frames, as supported by research conducted by authors such as Corona et al.⁶ and Watts et al.,⁷ among others. These investigations are grounded in the analysis of clinical data stemming from patient cohorts. In contrast, this study diverges from this approach by utilising a virtual model to generate simulated data, which emulates the behaviour of the fixator frame under scrutiny.

In the biomechanical domain, studies including the use of directional statistical techniques have been limited to the univariate and bivariate setting.⁸⁻¹¹ The primary objective of the study by Pataky et al.¹⁰ was to conduct a comparative analysis between directional analysis and uni- as well as multivariate Cardan analysis with regard to representative joint kinematic data collected during gait. In the work authored by Telschow et al.,¹¹ they addressed the issue of gait reproducibility by examining rotations of the tibia and femur at the knee joint. This investigation takes into account both the spatial influence of marker placement and the temporal variability introduced by different walking speeds in the experimental setup. The study conducted by Rivest et al.⁹ centres their attention on estimating the orientations of the two rotation axes at the ankle joint. While Rivest et al.⁸ developed a score statistic to evaluate the adequacy of the fixed-axis model. Furthermore, they proposed techniques to address the auto-correlation of errors among adjacent data points. These studies mentioned above represent the most immediate applications of directional statistics in the biomechanical field, indicating the existence of a significant research gap that needs to be addressed. The biomechanical domain is rich with angular data, however, the use of directional statistics techniques is overlooked.

The remainder of the article is organised as follows: Section 2 provides some background on the statistical techniques required to build our proposed model defined in Section 4. The motivation for the proposed model emanates from a specific data application which is discussed in Section 3. Section 5 contains the results and discussion of the data application. Section 6 concludes with some final remarks.

2 Background

There are a large number of applications which require the analysis of data not realised in the Euclidean space, but rather on some manifold (circles, spheres, hyperspheres, cylinders, torus etc.). This type of data arises in many fields such as life sciences, image analysis, astronomy, meteorology and earth sciences. Circular data is employed in numerous applications, as shown by Ley and Verdebout and¹² SenGupta and Arnold.¹³ Directional statistics constitutes a specialised branch within statistical analysis that focuses on angular observations. An inherent challenge in handling directional data lies in the non-linearity of the sample space, typically represented as a circle or circular manifold. This unique feature has garnered increased attention in the past two decades, primarily driven by the increase of data and the corresponding demand for tailored statistical methodologies, as highlighted by Ley and Verdebout.¹⁴ Existing directional models have been summarised in the works of Ley and Verdebout,¹⁴ Mardia and Jupp,¹⁵ and Pewsey and García-Portugués.¹⁶ A fundamental query arises when considering the necessity of directional statistics: 'Is there a need for specialised techniques due to the curvature of the sample space? Why are standard linear statistical methods insufficient?' Mardia and Jupp¹⁵ and Mardia¹⁷ provided a simple example by considering the metric of the arithmetic mean to illustrate why it is necessary to account for the curvature of the sample space and how assuming linear techniques leads to inaccurate results. However, in many studies, the cyclicity of a circular variable is often neglected and mistreated as a linear variable.¹⁸⁻²¹ The biomechanical domain is one example where angular data is prevalent, however, the use of directional statistics techniques is overlooked.

Modelling of circular-linear (C-L) data is limited to the bivariate C-L joint distribution. Various bivariate C-L distributions have been proposed and studied, specifically focusing on meteorology and climatology studies. A well-known method to obtain a C-L distribution is via the conditional modelling approach. Mardia and Sutton²² combined the von Mises and

normal distribution to obtain a joint model with conditional independence. Thereafter, many other models have been proposed, however, this approach restricts the choice of the marginal distribution and may neglect the dependence structure between the variables. This concern becomes even more complicated when extending a C-L distribution to the multivariate setting. Even with the assumption of independence among the circular and linear variables, extending a C-L distribution past the bivariate setting is difficult due to the normalising constant being intractable in most cases as well as the restricted circumstances for the normalising constant to be approximated. Thus, resulting in additional complexity with regards to efficient estimation methods. An attempt to extend the model proposed by Mardia and Sutton²² to the multivariate setting is given by Luengo-Sanchez et al.²³ In these studies, dependencies between linear variables were considered however, every circular variable was considered independent of all the other variables.

One of the most commonly used C-L distributions is the angular-linear model proposed by Johnson and Wehrly,²⁴ known as the JW model. This model was constructed using copulas. Copulas are used to link the marginal distributions of variables to their multivariate distributions.²⁵ Using copulas to build multivariate distributions is a flexible and convenient approach. In the linear domain, a substantial amount of literature can be found on multivariate copulas.^{25,26} However, these copulas cannot directly be applied to our problem as we need to consider the directional variables and account for their cyclicity. In the directional domain, there are limited studies on multivariate copulas. Bivariate copulas have been studied for circular-circular (C-C) variables^{27,28} and C-L variables, recently a trivariate circular copula was proposed.²⁹ The JW model being the most extensively considered and used. The JW model allows the distribution of the C-L variable pair to be written as a function of the marginals and the JW copula function. Since copula functions in the directional domain are limited to the bivariate case an alternative model needs to be considered to obtain our joint six-dimensional (6D) model.

A possible solution for the shortfall of multivariate C-L copulas is vine copulas. The origin of vine copulas stems from the hierarchical copula-based structure, specifically the pair copula construction proposed by Joe³⁰ and later investigated by Bedford and Cooke.^{31,32} There has been a growing interest in vine copulas due to their flexibility and intuitive decomposition. Vine copulas allow the construction of multivariate joint probability distributions by decomposing the multivariate function into simple building blocks comprising of bivariate copulas (pair copulas) based on the conditional probabilities.³³ Hence, the need for a multivariate C-L copula function is avoided. A vine is usually divided into D-vine (drawable vine) and C-vine (canonical vine) structures where the latter is more suitable for situations where a variable is key in controlling the dependency. A d -dimensional vine contains $(d - 1)$ trees, where the trees are represented with nodes and edges. The nodes are used to determine the edge label. The edge label corresponds to the subscript of each pair copula density and each edge corresponds to one pair copula. Thus the final d -dimensional copula can be obtained from the product of the pair copula densities given by all the edges. Section 4.3 provides an illustrative representation of the vine copula structure considered in this study. From this illustrative structure, the corresponding density function is provided.

By considering vine copulas with our linear-linear (L-L) variable pairs, C-C variable pairs and C-L variable pairs we can construct a multivariate joint probability model while also taking into account the cyclicity of the directional variables. For the trivariate case of a circular-linear-linear joint model this approach has proven to be useful in meteorology and oceanography.^{18,34}

With the flexibility offered by vine copulas comes an increase in complexity in larger dimensions. This results in the computational effort required to estimate all the parameters growing exponentially with dimension. Given that our dataset is 6D we consider a truncated vine copula. Truncated vine copulas have been proposed by Kurowicka³⁵ and Brechmann et al.³⁶ Truncated vine copulas are helpful as they can be constructed by using only pair-wise copulas and a lower number of conditional pair-wise copulas. This method is applied by considering a vine copula structure where all pair-wise copulas with conditioning set equal to or larger than K are replaced with independence copulas. Various approaches exist for obtaining the optimal K , as discussed by Kurowicka³⁵ and Brechmann et al.³⁶

To summarise in this article, we propose a modelling framework applicable to the 6D joint distribution. Our proposed model makes use of directional statistics to account for the cyclicity of the rotational variables and is constructed based on truncated vine copulas. The pair copula decomposition concept of vine copulas represents the dependence structure as a combination of L-L, C-C and C-L pairs modelled by their respective copulas. This allows us to assess the dependencies in the joint distribution. An advantage of using vine copulas is the flexibility to build multivariate distributions via bivariate copulas that model the dependence between pairs of random variables. The truncation of the vine copula assists with the computation effort required to estimate our model.

3 Data and motivation

Let X , Y and Z denote the linear (translational) variables defined on \mathbb{R} , where X is the displacement in the x -axis, Y is the displacement in the y -axis and Z is the displacement in the z -axis. Let Θ_X , Θ_Y and Θ_Z denote the circular (rotational) variables defined on the unit circle \mathbb{S}^1 , where Θ_X is the angular displacement about the x -axis, Θ_Y is the angular displacement

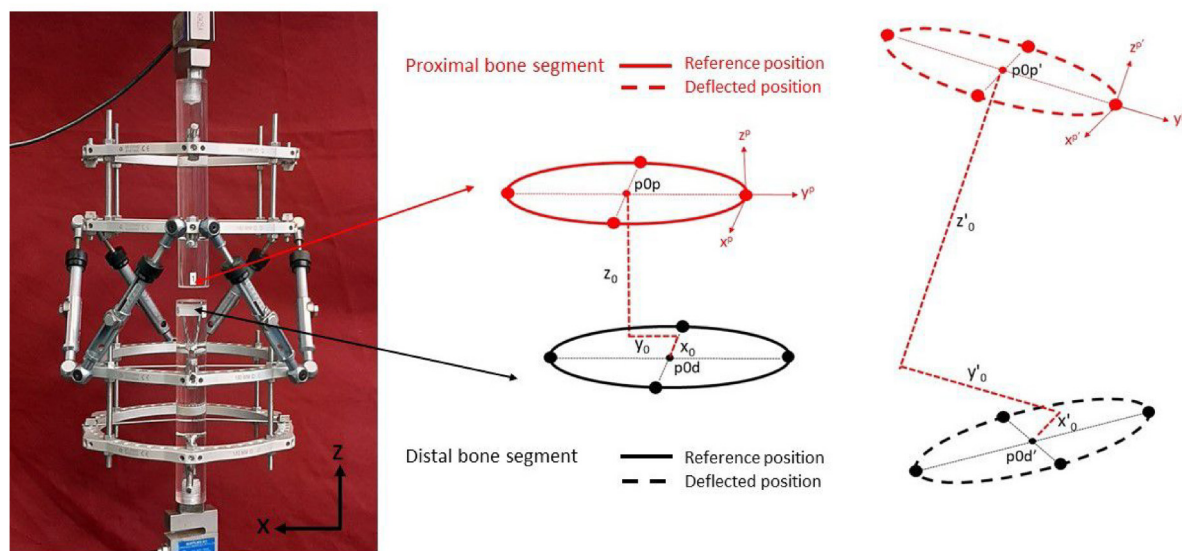


Figure 1. External fixator on left. Fracture ends at the reference position in middle. Fracture ends at the deflected position on right.

Table 1. Descriptive statistics of the translational variables for each configuration.

Configuration	Variable	Mean	sd	IQR	Skewness	Kurtosis
1	<i>X</i>	-0.2708	0.4475	0.5192	-0.3707	-0.0128
	<i>Y</i>	0.2948	0.4565	0.5646	-0.4791	0.2707
	<i>Z</i>	5.1039	3.6230	6.9932	0.0084	-1.4289
2	<i>X</i>	0.0455	0.4760	0.5915	-0.7638	2.573
	<i>Y</i>	1.5551	2.2361	3.1523	1.0108	-0.0973
	<i>Z</i>	6.7993	4.4999	8.9947	-0.0029	-1.5106

sd: standard deviation; IQR: interquartile range.

Table 2. Descriptive statistics of the rotational variables for each configuration.

Configuration	Variable	\bar{r}	$\bar{\theta}$	V_{Θ}	\hat{s}	\hat{k}
1	Θ_X	0.9999	0.0011	0.0143	0.3286	-1.1368
	Θ_Y	0.9999	-0.0056	0.0127	-0.1026	0.2556
	Θ_Z	0.9999	0.0008	0.0119	0.0048	-1.2235
2	Θ_X	0.9992	-0.0642	0.0397	2.7537	6.0765
	Θ_Y	0.9999	0.0166	0.0165	-0.0424	0.1532
	Θ_Z	0.9998	0.0194	0.0209	-1.7600	1.4948

about the *y*-axis and Θ_z is the angular displacement about the *z*-axis. In this study, the motion of the fracture is quantified by the displacement of the proximal and distal fracture ends. The translational and rotational motion is measured relative to an axis system fixed to, and rotating with, the proximal bone segment (i.e. $x^p - y^p - z^p$ in Figure 1). The translational motion is defined by considering the centroid of the two fracture surfaces (i.e. $p0p$ and $p0d$ in Figure 1). Each motion component is calculated by subtracting the reference position from the deflected position.

For this application, we consider the fracture displacement data comprising of the translational and rotational variables for two different constructs of an Ilizarov ring fixator which we refer to as configuration 1 and configuration 2. In Table 1, the descriptive statistics for the translational (linear) variables of configuration 1 and configuration 2 are given; specifically, the mean, standard deviation (sd), interquartile range (IQR), skewness and kurtosis.

In Table 2, the values of the main circular statistics¹⁵ are given for the rotational variables of configuration 1 and configuration 2. Specifically, the mean resultant length (\bar{r}), mean direction ($\bar{\theta}$), circular standard deviation (V_{Θ}), circular skewness (\hat{s}) and circular kurtosis (\hat{k}) are shown. The unit of measurement for the circular variables is considered as radians.

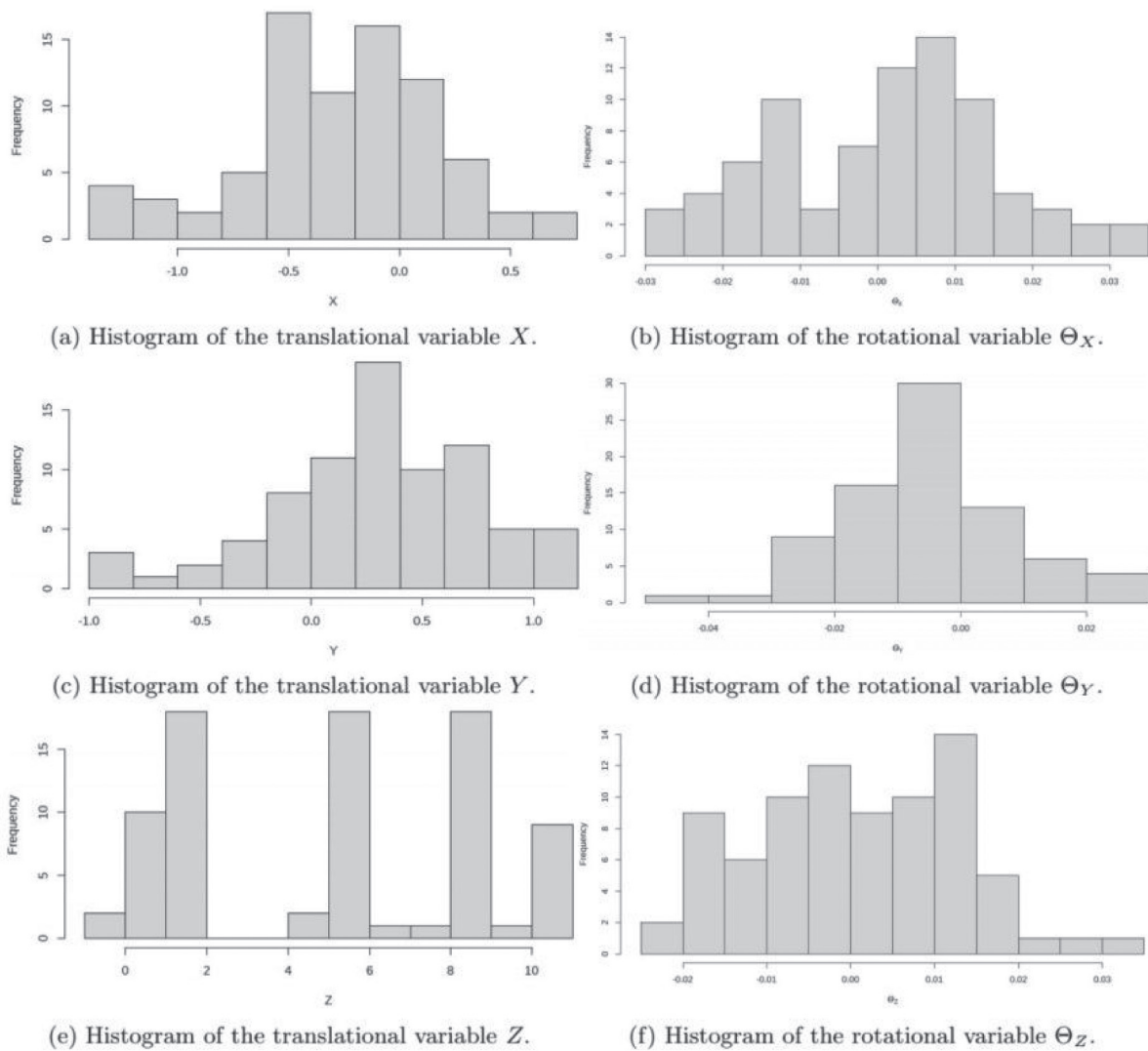


Figure 2. Histograms of the data of the translational (left) and rotational (right) variables for configuration 1.

In Figures 2 and 3, the data plots of the translation and rotational variables for configuration 1 and configuration 2 are provided, respectively. From Figures 2 and 3, bimodality and slight skewness are observed in some of the data plots indicating the need for a model that can account for multimodality, for example, the use of finite mixture models for specific variables. Based on these model requirements, a copula approach where the marginal distributions can be specified separately from the dependence structure is a useful technique for building a joint model. From the results in Table 2 and the histograms in Figures 2 and 3, we note that the rotational variables are highly concentrated which raises questions on the assumption of the periodicity of the variables. However, due to the nature of the variables we consider them to be circular in the modelling approach as the concentration might differ for various manufacturers and configurations but the nature and domain of the variable remains unchanged. As pointed out by Mardia,¹⁷ the underlying geometry is the main driver motivating the use of statistics for non-Euclidean variables. Hence, directional statistical techniques are considered for this study.

Pearson’s correlation coefficient was considered for the paired translational variables. The bivariate relationship between the translational and rotational variables is measured by the linear-circular correlation coefficient,¹⁵ $\rho_{x,\theta}$, as

$$\rho_{x,\theta} = \sqrt{\frac{\rho_{xc}^2 + \rho_{xs}^2 - 2\rho_{xc}\rho_{xs}\rho_{cs}}{1 - \rho_{cs}^2}} \tag{1}$$

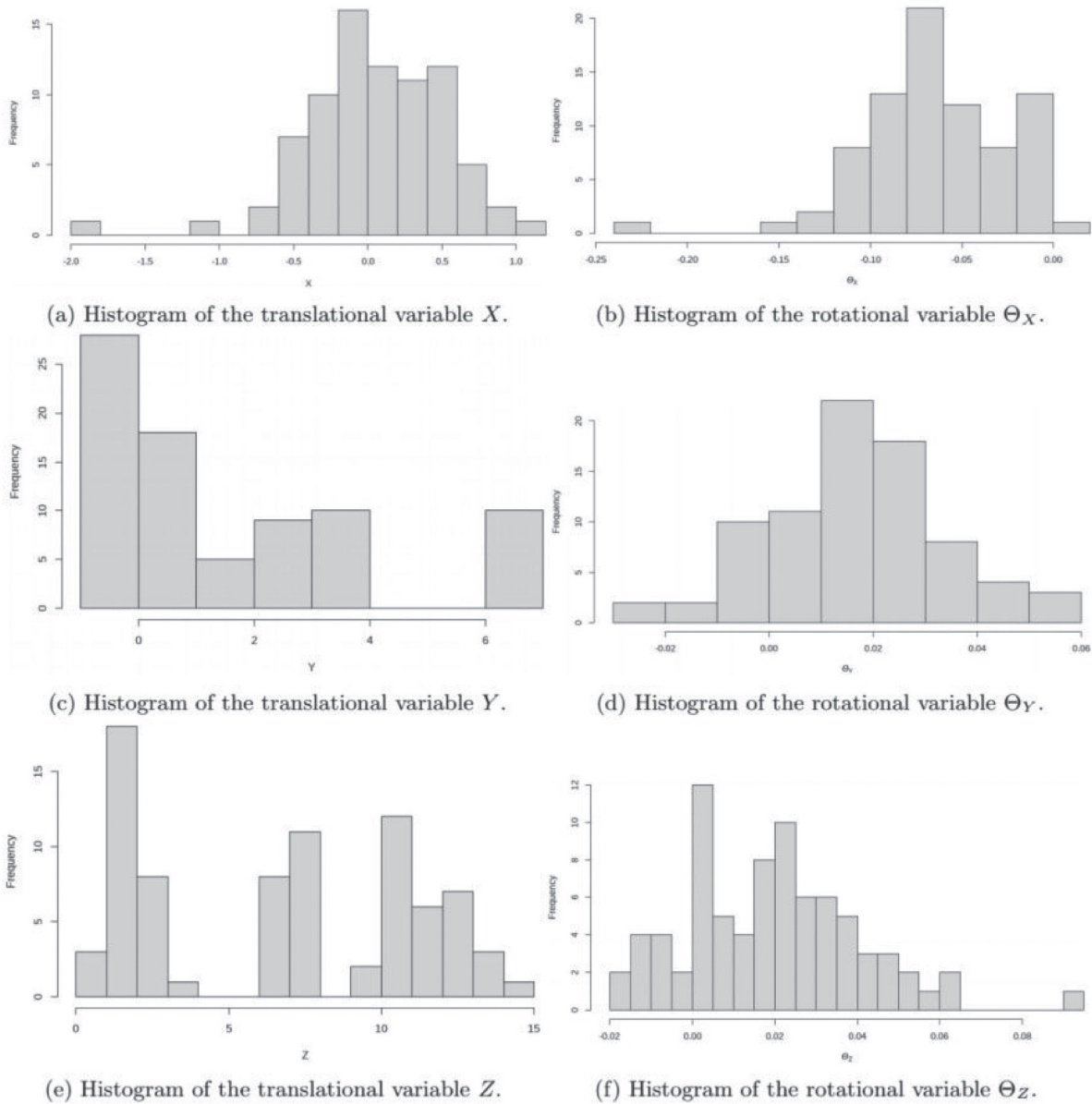


Figure 3. Histograms of the data of the translational (left) and rotational (right) variables for configuration 2.

where $\rho_{xc} = \text{cor}(x, \cos \theta)$, $\rho_{xs} = \text{cor}(x, \sin \theta)$ and $\rho_{cs} = \text{cor}(\cos \theta, \sin \theta)$ are the sample correlation coefficients. For the paired rotational variables, the correlation coefficient¹⁵ is defined as follows:

$$\rho_{\theta,\phi} = \sqrt{\frac{(\rho_{cc}^2 + \rho_{cs}^2 + \rho_{sc}^2 + \rho_{ss}^2) + 2(\rho_{cc}\rho_{ss} + \rho_{cs}\rho_{sc})\rho_1\rho_2 - 2(\rho_{cc}\rho_{cs} + \rho_{sc}\rho_{ss})\rho_2 - 2(\rho_{cc}\rho_{sc} + \rho_{cs}\rho_{ss})\rho_1}{(1 - \rho_1^2)(1 - \rho_2^2)}}, \quad (2)$$

where $\rho_{cc} = \text{cor}(\cos \theta, \cos \phi)$, $\rho_{ss} = \text{cor}(\sin \theta, \sin \phi)$, $\rho_{cs} = \text{cor}(\cos \theta, \sin \phi)$, $\rho_{sc} = \text{cor}(\sin \theta, \cos \phi)$, $\rho_1 = \text{cor}(\cos \theta, \sin \theta)$ and $\rho_2 = \text{cor}(\cos \phi, \sin \phi)$ are the sample correlation coefficients. The correlation plots for configuration 1 and configuration 2 are given in Figures 4 and 5, respectively. From Figures 4 and 5 and Tables 3 and 4, it is observed that the assumption of independence is violated by the data. Thus, a joint dependent model is required.

A significance test was performed to evaluate the necessity of accounting for a dependence structure between the variables. Tables 3 and 4 provide the significant (at a 5% significance level) correlation coefficients for configuration 1 and



Figure 4. Correlation plot of the paired translational variables, paired rotational variables and translation-rotational variables for configuration 1.

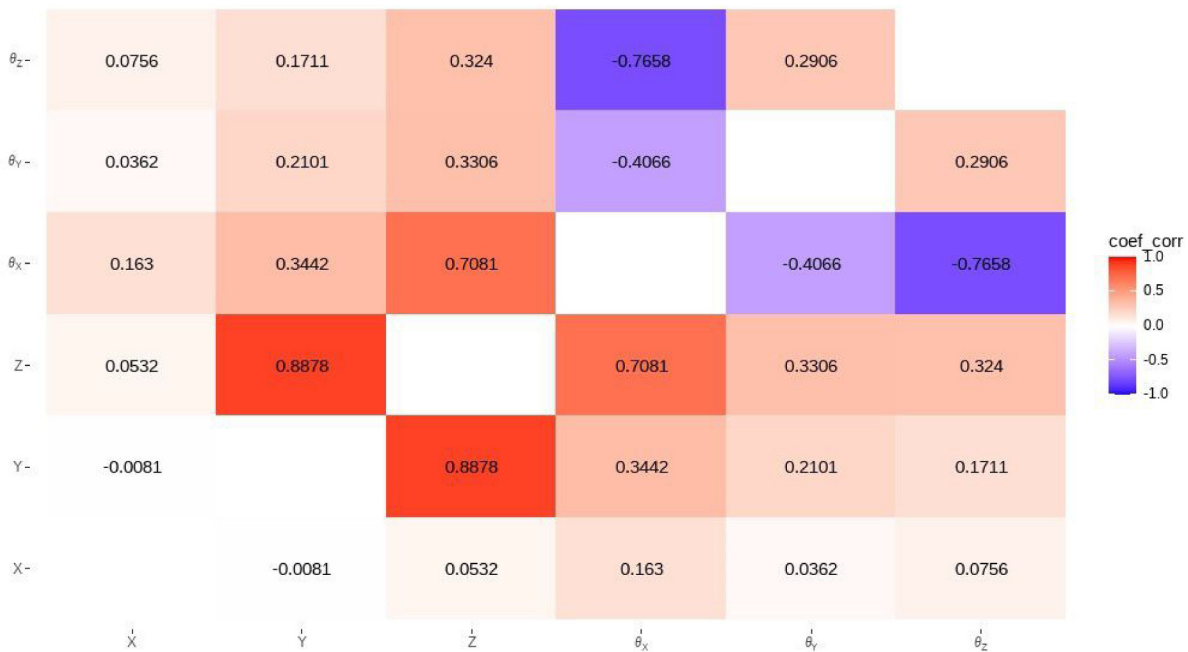


Figure 5. Correlation plot of the paired translational variables, paired rotational variables and translation-rotational variables for configuration 2.

configuration 2, respectively, where the null hypothesis assumes independency among the respective variables. The results from the correlation tests further emphasise the need for a joint model that accounts for dependencies.

4 Methodology

In this section, we define the proposed modelling framework applicable to the 6D joint distribution for modelling the displacement of a fracture site. The six variables, of which three are translational and three are rotational, are taken into

Table 3. Paired variables for configuration 1 with correlation coefficients significant at a 5% significance level.

Paired variables		ρ	p-value
X	Y	0.3662	0.0017
Y	Z	0.5758	0.000
Y	Θ_X	0.1196	0.0001
Y	Θ_Y	0.078	0.0024
X	Θ_Z	0.0961	0.0006
Z	Θ_Z	0.1001	0.0004

Table 4. Paired variables for configuration 2 with correlation coefficients significant at a 5% significance level.

Paired variables		ρ	p-value
Y	Z	0.8878	0.000
Θ_X	Θ_Y	-0.4066	0.0218
Θ_X	Θ_Z	-0.7658	0.0007
X	Θ_X	0.163	0.000
Y	Θ_X	0.3442	0.000
Z	Θ_X	0.7081	0.000
Y	Θ_Y	0.2101	0.000
Z	Θ_Y	0.3306	0.000
X	Θ_Z	0.0756	0.0029
Y	Θ_Z	0.1711	0.000
Z	Θ_Z	0.324	0.000

account in this framework. The linear and circular probability density functions (PDFs) considered are given as well as the copula functions.

4.1 Linear and circular distributions

If X, Y and Z denote the linear (translational) variables defined on \mathbb{R} , let the PDFs be denoted as $f_X(x), f_Y(y)$ and $f_Z(z)$, respectively. If Θ_X, Θ_Y and Θ_Z denote the circular (rotational) variables defined on the unit circle S^1 , let the PDFs be denoted as $f_{\Theta_X}(\theta_x), f_{\Theta_Y}(\theta_y)$ and $f_{\Theta_Z}(\theta_z)$, respectively. The cumulative distribution function (CDF) of the variables is represented as $F_X(x), F_Y(y), F_Z(z)$ for the linear variables and $F_{\Theta_X}(\theta_x), F_{\Theta_Y}(\theta_y), F_{\Theta_Z}(\theta_z)$ for the circular variables.

Based on the preliminary analysis of the data, and the histograms of the variables provided in Figures 2 and 3, for the translational variables, the normal (N) distribution and a two component mixture of the normal (MN) distribution were considered to be most appropriate. A two component mixture of the normal distribution was considered to accommodate for the multimodality observed in the data. The expression for a finite mixture model is given as

$$f_X(x) = \sum_{j=1}^m \omega_j f(x|\beta_j) \tag{3}$$

where $j = 1, \dots, m$, m is the number of mixture components, β_j is the parameter of the j th component of the finite mixture distribution, $0 < \omega_j \leq 1$ with $\sum_{j=1}^m \omega_j = 1$, where ω_j represents the mixing proportions.

For the rotational variables, we considered various circular distributions and found the wrapped Cauchy (wC) and a two component mixture of the wC (MwC) to be most appropriate. The peakedness of the wC makes it an appealing choice for this study. The PDF of the wC¹⁵ is given by

$$f_{\Theta}(\theta) = \frac{1}{2\pi} \frac{1 - \kappa^2}{1 + \kappa^2 - 2\kappa \cos(\theta - \mu)} \tag{4}$$

where $\theta \in (-\pi, \pi]$, the concentration parameter $\kappa \in (0, 1)$ and the mean direction parameter $\mu \in [-\pi, \pi)$.

To build our model, we consider the defined linear and circular distributions as the marginal distributions for our framework based on the analysis of the data (see Figures 2 and 3 and Tables 1 and 2). Various other distributions may also be considered depending on the complexity of a dataset.

4.2 Copulas

The copula approach allows us to consider the marginal distributions separately from the dependence structure between the variables. Consider $F_{X_1, X_2, \dots, X_d}(x_1, x_2, \dots, x_d)$ and $f_{X_1, X_2, \dots, X_d}(x_1, x_2, \dots, x_d)$ to be the joint CDF (JCDF) and joint PDF (JPDF) of the d -dimensional random vector (X_1, X_2, \dots, X_d) , respectively. Then we define the copula, C , as

$$F_{X_1, X_2, \dots, X_d}(x_1, x_2, \dots, x_d) = C(u_1, u_2, \dots, u_d) \tag{5}$$

and thus

$$f_{X_1, X_2, \dots, X_d}(x_1, x_2, \dots, x_d) = c(u_1, u_2, \dots, u_d) \prod_{j=1}^d f_{X_j}(x_j) \tag{6}$$

where $u_j = F_{X_j}(x_j), j = 1, 2, \dots, d$ with $x_j \in \mathbb{R}, u_j \in [0, 1]$ and $f_{X_j}(x_j)$ the marginal PDF of each variable.

In the linear domain, various multivariate copula functions have been defined.²⁵ Since in directional statistics copula functions are limited to the bivariate case, we consider the use of vine copulas.

For the L-L pair copula, we consider the conventional Farlie-Gumbel-Morgenstern (FGM) defined as follows:

$$c_{X_1, X_2}(u_1, u_2) = 1 + \alpha(1 - 2u_1)(1 - 2u_2) \tag{7}$$

where $\alpha \in [-1, 1]$.

For the C-L pair copula, we consider the most common function proposed by Johnson and Wehrly,²⁴ the JW copula, defined as

$$c_{\Theta, X}(u_1, u_2) = 2\pi g(\eta) \tag{8}$$

where

$$\eta = 2\pi(u_1 - qu_2) \tag{9}$$

and $\eta \in [0, 2\pi]$ is a circular random variable with $g(\cdot)$ defined as a circular PDF with $q \in \{-1, 1\}$.

Jones et al.²⁷ proposed a similar approach to Johnson and Wehrly²⁴ to obtain a bivariate circular copula. The resulting copula function simplifies to a circular PDF where the argument is defined as in (9).

4.3 Proposed model

Based on the concept of vine copulas, our proposed model is built using the L-L, C-L and C-C pair copulas and their respective marginal distributions.

In Figure 6, the illustrated canonical vine structure considered is given. The paired variables that form the basis of the vine copula structure are denoted as $ZX, ZY, Z\Theta_X, Z\Theta_Y$ and $Z\Theta_Z$ for Tree 1 in Figure 6. The same notation follows for

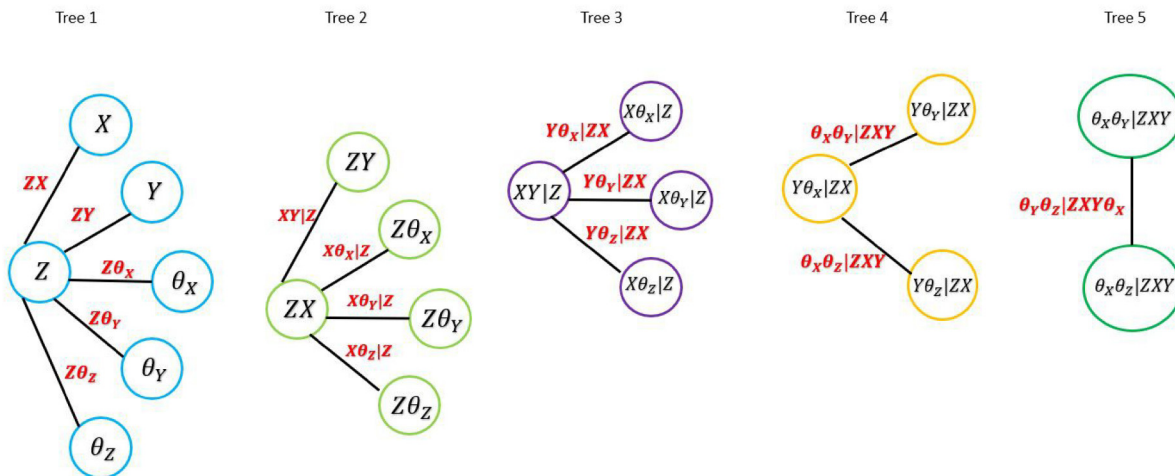


Figure 6. A schematic diagram of the canonical vine structure considered for the proposed model.

Trees 2 to 5 to indicate the paired variables. From the structure in Figure 6, the resulting JPDF can be extracted as follows:

$$\begin{aligned}
 f(x, y, z, \theta_x, \theta_y, \theta_z) = & f(x)f(y)f(z)f(\theta_x)f(\theta_y)f(\theta_z) \\
 & \times c_{xz}(F_X, F_Z)c_{yz}(F_Y, F_Z)c_{z\theta_x}(F_Z, F_{\theta_x}) \\
 & \times c_{z\theta_y}(F_Z, F_{\theta_y})c_{z\theta_z}(F_Z, F_{\theta_z}) \\
 & \times c_{xy|z}(F_{X|Z}, F_{Y|Z})c_{x\theta_x|z}(F_{X|Z}, F_{\theta_x|Z}) \\
 & \times c_{x\theta_y|z}(F_{X|Z}, F_{\theta_y|Z})c_{x\theta_z|z}(F_{X|Z}, F_{\theta_z|Z}) \\
 & \times c_{y\theta_x|z}(F_{Y|Z}, F_{\theta_x|Z})c_{y\theta_y|z}(F_{Y|Z}, F_{\theta_y|Z}) \\
 & \times c_{y\theta_z|z}(F_{Y|Z}, F_{\theta_z|Z})c_{\theta_x\theta_y|z}(F_{\theta_x|Z}, F_{\theta_y|Z}) \\
 & \times c_{\theta_x\theta_z|z}(F_{\theta_x|Z}, F_{\theta_z|Z})c_{\theta_y\theta_z|z}(F_{\theta_y|Z}, F_{\theta_z|Z})
 \end{aligned} \tag{10}$$

It is important to note that different vine structures may lead to different results. As mentioned by Aas et al.,³³ the decomposition should be selected by determining which pair relationships are most important. Based on expert opinion, in this application, we considered a structure where the variable, Z, is identified as most important and is linked to all the variables and is considered the conditional variable.

For the parameter estimation of (10) the method of maximum likelihood estimation (MLE) is considered. Closed-form expressions for the MLEs cannot be obtained due to the complex functional form. Advanced optimisation algorithms are required to numerically compute the maximum likelihood and thus the MLEs.

5 Application

In this section, we illustrate the validity of our proposed model by considering the fracture displacement data discussed in Section 3.

The flexibility offered by vine copulas comes at a computational cost for high dimensions. As a result, we consider a *simplified* form of the joint PDF given in (10) namely a truncated vine copula. For the purposes of this study, we considered $K = 2$ for the truncation of our vine copula. Our motivation for the choice of $K = 2$ stems from the dependence structure desired, the correlations observed from the data analysis as well as the practical implications of the relationships between the variables and their interactions. The authors considered a pragmatic approach instead of a heuristic approach, in which the objective is to select the most appropriate truncated vine copula within the constraints of time and computational effort, rather than to identify the optimal copula. The objective is to precisely describe the initial $K = 2$ trees in the truncated vine copula model in order to account for the primary dependencies. Subsequently, independence copulas for the higher order trees will be examined.

To compute our joint distribution model we first need to specify the marginal and copula functions for each variable and variable-pair, respectively. Based on preliminary analysis of the data (as illustrated in Section 3) and models defined in Section 4, for the linear variables we consider the N distribution as well as the MN distribution. For the circular variables we consider the wC distribution as well as the MwC distribution. A combination of different marginal functions is considered for each configuration termed *cases*. Table 5 provides a summary of the different cases evaluated for each configuration. We consider two cases, A and B, for each configuration (1 and 2). The cases are labelled with the case option and configuration, for example, the second case of configuration 1 is denoted as case B1. The cases were chosen in consultation with domain experts and the need for a parsimonious model.

Table 5. Combination of marginal and copula functions considered for each configuration.

Configuration	Case	Function										
		X	Y	Z	Θ_X	Θ_Y	Θ_Z	ZX	ZY	Z Θ_X	Z Θ_Y	Z Θ_Z
1	A1	N	N	N	wC	wC	wC	FGM	FGM	wC (JW)	wC (JW)	wC (JW)
	B1	N	N	N	MwC	wC	MwC	FGM	FGM	wC (JW)	wC (JW)	wC (JW)
2	A2	N	N	N	wC	wC	wC	FGM	FGM	wC (JW)	wC (JW)	wC (JW)
	B2	N	MN	N	wC	wC	MwC	FGM	FGM	wC (JW)	wC (JW)	wC (JW)

N: normal; MN: mixture of the normal; wC: wrapped Cauchy; MwC: mixture of the wrapped Cauchy; FGM: Farlie-Gumbel-Morgenstern; JW: Johnson and Wehrly.

Table 6. The number of parameters (p), maximised log-likelihood (MLL), Akaike information criterion (AIC) and Bayesian information criterion (BIC) values for the different combinations of marginal and copula functions.

Configuration	Case	p	MLL	AIC	BIC
1	A1	17	169.4548	-304.9096	-264.4152
	B1	23	170.6057	-295.2114	-240.4248
2	A2	17	37.7123	-41.4246	-0.9302
	B2	23	103.2792	-160.5584	-105.7718

For the finite mixture models, the expectation-maximisation (EM) algorithm may be utilised. Due to the high-dimensional space of the parameter set, advanced optimisation algorithms such as the particle swarm optimisation (PSO)³⁷ and genetic algorithm (GA)³⁸ were used to efficiently estimate the parameters of the joint model.

In Table 6, the performance measures, for the different cases specified in Table 5, are provided. For the performance evaluation, two goodness-of-fit metrics are applied to evaluate the models. The Akaike information criterion (AIC) estimates the relative amount of information lost by a given model. The Bayesian information criterion (BIC) is widely used for model selection and is similar to the AIC, however, the BIC is more strict in its penalisation of model complexity. These two metrics are defined as follows:

$$\text{AIC} = 2p - 2 \ln(\hat{L})$$

and

$$\text{BIC} = p \ln(n) - 2 \ln(\hat{L})$$

where p is the number of estimated parameters in the model, n is the total number of data points, and \hat{L} is the maximum value of the likelihood function for a specific model. Based on the performance measures in Table 6, we can conclude that case A1 and case B2 are the best models for configuration 1 and configuration 2, respectively.

Remark 1. *It is important to note that the results of the proposed model cannot be compared with existing methods (defined in the linear domain) as the models are defined on different manifolds. Due to the nature of the rotational variables, implementing an approach that incorporates directional statistics is crucial for accurate modelling.*

To illustrate the validity of the proposed model in comparison to the conventional use of an independent model, we consider a likelihood ratio test (LRT). We consider the model under the null hypothesis to be the independent model and case B1 and case B2 (more complex model) to be the model under the alternative hypothesis, respectively. For the independent model, we consider the same distributions as specified for case B1 and case B2 for the six variables ($X, Y, Z, \Theta_X, \Theta_Y$ and Θ_Z), respectively, and independence copulas for all the trees. Thus, the joint distribution under the null hypothesis reduces to the product of the specified marginal distributions only. For configuration 1, independent model versus case B1, we obtain an LRT value of 80.4512 and reject the null hypothesis at a 5% significance level (p -value = 6.66×10^{-16}). For configuration 2, independent model versus case B2, we obtain an LRT value of 22.1448 and reject the null hypothesis at a 5% significance level (p -value = 0.0005). Thus, we can conclude that the joint dependent model is a more suitable fit for the data.

6 Conclusion

In this article, we propose a modelling framework applicable to the 6D joint distribution. This model accounts for the cyclicity of the rotational variables by means of directional statistics as well as accounts for a dependence structure between the variables. The framework is constructed based on vine copulas. The pair copula decomposition concept of vine copulas represents the dependence structure as a combination of C-L, C-C and L-L pairs modelled by their respective copulas. This allows us to assess the dependencies in the joint distribution. An advantage of using vine copulas is the flexibility to build multivariate distributions via bivariate copulas that model the dependence between pairs of random variables. For efficient estimation, a truncation of the vine copula was considered. The analysis of this data motivates the need for a dependence structure to be accounted for when modelling this type of data. From the results of the real data application, the advantage of applying the joint dependence model is observed. Based on the LRT, we can conclude that the joint dependent model is a better choice for modelling the fracture displacements and will thus be more informative for evaluations of these devices

and the design thereof. The proposed modelling framework can be adjusted for other practical cases depending on the desired dependency structure required and the relationship between the variables. The primary goal of an external fixator is injury rehabilitation. Fracture healing is inevitably influenced by the complex interplay of biology and biomechanics – that is, inter-fragmentary motion and biomechanics. The construct of an external fixator is determined by the configuration of the hardware. The construct's configuration may lead to different inter-fragmentary motions and it is therefore important to accurately model and understand the motions in play to obtain the most appropriate construct for optimal healing. The modelling framework proposed in this article provides a more accurate view for fracture displacements thus leading to improved evaluations and design of these devices, thus, aligning with the United Nation's Sustainable Development Goal (SDG) 3 to promote good health and well-being.

Acknowledgements

The authors would like to thank the anonymous reviewers for their insightful comments which led to an improvement in this paper.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was based upon research supported in part by the National Graduate Academy for Mathematical and Statistical Sciences (NGA) of South Africa; National Research Foundation (NRF) of South Africa, Reference: SRUG2204203965, grant no. 120839 and Reference: RA171022270376, grant no. 119109; DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), South Africa and STATOMET at the Department of Statistics at the University of Pretoria. The research of M. Arashi is based upon research funded by Iran National Science Foundation, grant no. 4015320 as well as the National Research Foundation (NRF) of South Africa, Reference: RA211204653274, grant no. 151035.

ORCID iD

Priyanka Nagar  <https://orcid.org/0000-0002-1775-8905>

References

1. Glatt V, Evans CH and Tetsworth K. A concert between biology and biomechanics: the influence of the mechanical environment on bone healing. *Front Physiol* 2017; **7**: 678.
2. Lobst C, Ferreira N and Kold S. A review and comparison of hexapod external fixators: current concept review. *J Pediatr Orthop Soc North Am* 2023; **5**: 627.
3. Fenton C, Henderson D, Samchukov M, et al. Comparative stiffness characteristics of Ilizarov-and hexapod-type external frame constructs. *Strat Trauma Limb Reconstruction* 2021; **16**: 138.
4. Gessmann J, Citak M, Jettkant B, et al. The influence of a weight-bearing platform on the mechanical behavior of two Ilizarov ring fixators: tensioned wires vs. half-pins. *J Orthop Surg Res* 2011; **6**: 1–11.
5. Henderson DJ, Rushbrook JL, Stewart TD, et al. What are the biomechanical effects of half-pin and fine-wire configurations on fracture site movement in circular frames? *Clin Orthop Relat Res* 2016; **474**: 1041–1049.
6. Corona PS, Pujol O, Vicente M, et al. Outcomes of two circular external fixation systems in the definitive treatment of acute tibial fracture related infections. *Injury* 2022; **53**: 3438–3445.
7. Watts A, Sadekar V, Moulder E, et al. A comparative evaluation of the time to frame removal for tibia fractures treated with hexapod and Ilizarov circular frames. *Injury* 2023; **54**: 996–1003.
8. Rivest LP. A directional model for the statistical analysis of movement in three dimensions. *Biometrika* 2001; **88**: 779–791.
9. Rivest LP, Baillargeon S and Pierrynowski M. A directional model for the estimation of the rotation axes of the ankle joint. *J Am Stat Assoc* 2008; **103**: 1060–1069.
10. Pataky TC and Challis JH. Using directional statistics to test hypotheses regarding rigid body attitude: comparison to univariate and multivariate cardan angle tests. *J Biomech* 2020; **111**: 109976.
11. Telschow FJ, Pierrynowski MR and Huckemann SF. Functional inference on rotational curves under sample-specific group actions and identification of human gait. *Scand J Stat* 2021; **48**: 1256–1276.
12. Ley C and Verdebout T. *Applied directional statistics: modern methods and case studies*. Boca Raton: CRC Press, 2018.
13. SenGupta A and Arnold BC. *Directional statistics for innovative applications: a bicentennial tribute to Florence Nightingale*. Singapore: Springer, 2022.
14. Ley C and Verdebout T. *Modern directional statistics*. Boca Raton: Chapman and Hall/CRC Press, 2017.
15. Mardia KV and Jupp PE. *Directional statistics*. vol. 494. Chichester: John Wiley & Sons, 2009.
16. Pewsey A and García-Portugués E. Recent advances in directional statistics. *Test* 2021; **30**: 1–58.
17. Mardia KV. Fisher's legacy of directional statistics, and beyond to statistics on manifolds. *arXiv preprint arXiv:240517919*, 2024.

18. Wang ZW, Zhang WM, Zhang YF, et al. Circular-linear-linear probabilistic model based on vine copulas: an application to the joint distribution of wind direction, wind speed, and air temperature. *J Wind Eng Indus Aerod* 2021; **215**: 104704.
19. Zheng XW, Li HN and Li C. Damage probability analysis of a high-rise building against wind excitation with recorded field data and direction effect. *J Wind Eng Indus Aerod* 2019; **184**: 10–22. DOI: <https://doi.org/10.1016/j.jweia.2018.11.018>. <https://www.sciencedirect.com/science/article/pii/S0167610518306147>.
20. Solari S and Ángel Losada M. Simulation of non-stationary wind speed and direction time series. *J Wind Eng Indus Aerod* 2016; **149**: 48–58. DOI: <https://doi.org/10.1016/j.jweia.2015.11.011>. <https://www.sciencedirect.com/science/article/pii/S0167610515002822>.
21. Leguey I, Larrañaga P, Bielza C, et al. A circular-linear dependence measure under Johnson–Wehrly distributions and its application in Bayesian networks. *Inf Sci (Ny)* 2019; **486**: 240–253. DOI: <https://doi.org/10.1016/j.ins.2019.01.080>. <https://www.sciencedirect.com/science/article/pii/S0020025519300581>.
22. Mardia K and Sutton T. A model for cylindrical variables with applications. *J R Stat Soc: Ser B (Methodological)* 1978; **40**: 229–233.
23. Luengo-Sanchez S, Bielza C and Larrañaga P. Hybrid Gaussian and von Mises model-based clustering. In *Proceedings of the Twenty-Second European Conference on Artificial Intelligence*. ECAI'16, NLD: IOS Press. ISBN 9781614996712, pp.855–862. DOI:10.3233/978-1-61499-672-9-855. <https://doi.org/10.3233/978-1-61499-672-9-855>.
24. Johnson RA and Wehrly TE. Some angular-linear distributions and related regression models. *J Am Stat Assoc* 1978; **73**: 602–606.
25. Nelsen RB. *An introduction to copulas*. New York: Springer Science & Business Media, 2007.
26. Joe H. *Dependence modeling with copulas*. Boca Raton: CRC Press, 2014.
27. Jones M, Pewsey A and Kato S. On a class of circular copulas for circular distributions. *Ann Inst Stat Math* 2015; **67**: 843–862.
28. Kato S, Pewsey A and Jones M. Circulars from Fourier series. Technical report, Technical report 7, School of Mathematics and Statistics, Open University, 2018.
29. Kato S, Ley C and Loizidou S. The trivariate wrapped Cauchy copula—a multi-purpose model for angular data. *arXiv preprint arXiv:240110824*, 2024.
30. Joe H. $m(m-1)/2$ bivariate dependence parameters. *Distrib Fixed Marginals Related Topics* 1996; **28**: 120.
31. Bedford T and Cooke RM. Probability density decomposition for conditionally dependent random variables modeled by vines. *Ann Math Artif Intell* 2001; **32**: 245–268.
32. Bedford T and Cooke RM. Vines—a new graphical model for dependent random variables. *Ann Stat* 2002; **30**: 1031–1068.
33. Aas K, Czado C, Frigessi A, et al. Pair-copula constructions of multiple dependence. *Ins: Math Econo* 2009; **44**: 182–198.
34. Heredia-Zavoni E and Montes-Iturrizaga R. Modeling directional environmental contours using three dimensional vine copulas. *Ocean Eng* 2019; **187**: 106102.
35. Kurowicka D. Optimal truncation of vines. In *Dependence modeling: Vine copula handbook*. World Scientific, 2010.
36. Brechmann EC, Czado C and Aas K. Truncated regular vines in high dimensions with application to financial data. *Canad J Stat* 2012; **40**: 68–85.
37. Wang D, Tan D and Liu L. Particle swarm optimization algorithm: an overview. *Soft comput* 2018; **22**: 387–408.
38. Chatterjee S, Laudato M and Lynch LA. Genetic algorithms and their statistical applications: an introduction. *Comput Stat Data Anal* 1996; **22**: 633–651.