European journal of operational research 266 (2018) 269–277 doi: 10.1016/j.ejor.2017.08.058

Rayleigh copula for describing impedance data – with application to condition monitoring of proton exchange membrane fuel cells

Pavle Boškoski[a,*], Andrej Debenjak[a], Biljana Mileva Boshkoska[b,a]

[a] *Jožef Stefan Institute, Jamova cesta 39, SI-1000 Ljubljana, Slovenia*
[b] *Faculty of Information Studies, Ljubljanska cesta 31a, SI-8000 Novo mesto, Slovenia*

Abstract

Proton exchange membrane (PEM) fuel cells are considered as one of the power sources of the future. However, there are still challenges facing hydrogen technologies, such as costs, efficiency and reliability. Accurate and timely quantification of the fuel cell's performance indices has two main benefits: (i) it enables conducting appropriate mitigation control actions, and (ii) time evolution of the indices can be exploited for prognostics. This in turn contributes to the overall safety, prolonged durability, and increased reliability of fuel cell systems.

Condition deterioration of PEM fuel cells affects their electrochemical impedance characteristic, thus a multitude of condition monitoring approaches are based on electrochemical impedance spectroscopy (EIS). By employing recent results of statistical characterisation of impedance components, it becomes possible to describe the complete impedance characteristic through a multivariate distribution that belongs to the exponential family of distributions. As a result, the complete process of condition monitoring of fuel cells can be performed by quantifying the changes of the multivariate distribution with respect to the nominal one. In the presented approach, the multivariate distribution is built through Rayleigh bivariate copula and D-Vine algorithm. It is shown that the selected copula is the optimal choice for modelling impedance characteristics of PEM fuel cells. The performance of the approach was evaluated and validated on an industrial grade 8.5 kW PEM fuel cell system subjected to the flooding and drying faults.

Keywords: OR in Energy, Rayleigh copula, PEM fuel cell, Condition monitoring, Water management faults

1. Introduction

A proton exchange membrane (PEM) fuel cell is an electrochemical device that converts chemical energy of hydrogen directly into electrical energy. An outstanding advantage of fuel cells is its high energy efficiency compared to conventional internal combustion machines. In the commercially available PEM fuel cell-based

*Corresponding author.
Email address: pavle.boskoski@ijs.si (Pavle Boškoski)

power systems, the net efficiency of the chemical-to-electrical energy conversion at nominal output power varies between 45 and 60%, which is substantially higher compared to the conventional energy converting devices [1, 2]. In addition to the high conversion efficiency, a PEM fuel cell-based power system generates no greenhouse gases and emits only low level noise. Low temperature of operation (up to max. ∼85°C), ability to use oxygen from the air, short start-up and shut-down times, and high power density are the features that promote PEM fuel cells as a technology of the future [3].

PEM fuel cells are considered to be the most suitable for setups of up to a few hundreds of kW of power, both for stationary and transportation applications. However, a major obstacle in reaching high commercialisation level of PEM fuel cell power systems remains its relatively low reliability of operation and insufficient lifetime. It is well recognised that when prognostics and health management (PHM) information of a technical system is applied wisely, it provides benefits in term of reliability and durability of the system [4–6]. Furthermore, an online information about the condition indices of an electrochemical system can be directly included into their exploitation scheme thus improving their the overall efficiency [7, 8].

The main source of information regarding current condition of PEM fuel cells is the electrochemical impedance spectroscopy (EIS) data [9]. The bulk of the research is, however, focused predominantly on the measurement and estimation aspects. Whereas proper statistical analysis of impedance data is usually neglected. This paper provides in-depth analysis of the statistical properties of impedance by employing the concepts of multivariate random variables. As a result, the probability of observing a particular impedance characteristic can be treated as an overall condition indicator for PEM fuel cells.

In the field of PEM fuel cell and wider, stochastic excitation signals have already been employed for performing impedance spectroscopy [10–14]. Employing stochastic excitation signals has two benefits compared to the traditional small-amplitude sinusoidal excitation [9, 15, 16] or multi-sine excitation [17–19]. The most apparent benefit is the shorter probing time which is due to the broadband excitation. The second one is the possibility to employ various powerful approaches for statistical signal processing and analysis.

Statistical analysis requires specifying probability distribution for each impedance component as well as for the impedance characteristic as a whole. Boškoski and Debenjak [12] provided a solution for the former requirement by analytically specifying the probability density functions (PDFs) of particular impedance components. For the latter problem, a multidimensional cumulative distribution function (CDF) and/or PDF have to be specified, which describe the complete impedance characteristic. This is far more difficult problem.

Copula functions offer a way of building multidimensional CDFs based on marginal the distributions of its components. Such approaches have been already proposed for the concepts of condition monitoring [20, 21]. Since the marginal distributions of the impedance amplitude (i.e., the distributions of the impedance amplitude at individual frequencies) are known to be a ratio of two Rayleigh random variables [12], it is justified to select a copula function that belongs to the exponential copula family. Exploiting the results

of Zeng et al. [22] for bivariate copulas, this paper proposes a solution for constructing the multivariate PDF that encompasses the complete impedance characteristic based on the Rayleigh copula and the D-Vine approach.

Specifying the multivariate PDF of the impedance characteristic has significant practical merits in the context of condition monitoring of PEM fuel cells. The initial fault-free condition of a PEM fuel cell under observation is fully captured by the parameters of the multivariate distribution of its impedance. As a result, the changes in the impedance, induced by faults, can be quantified by calculating the probability of observing a particular impedance characteristic. This offers a way of performing fault detection and fault evaluation without the need of any prior characterisation of the fuel cell under various fault modes. Values near the mode of the PDF represent fault-free state, whereas values in the tail regions indicate faulty states.

The paper is organised as follows. Section 2 presents the essentials of the EIS signal processing and PDF parameter estimation processes as developed in [10, 12]. Section 3 provides the necessary theoretical background of copula functions and D-Vine algorithm for multidimensional structures. All the necessary relations required for building the multivariate PDF using Rayleigh copula are derived in Section 4. It is shown that the selected multivariate Rayleigh-based PDF optimally describes the impedance characteristic as a whole. Finally in Section 5, the proposed multidimensional distribution is employed for condition monitoring of a PEM fuel cell system under various water management faults.

2. EIS with continuous wavelet transform (CWT) and discrete random binary sequence (DRBS)

EIS is an active characterisation tool that requires external excitation of the electrochemical device under test. A carefully selected stochastic excitation signal enables in-depth statistical analysis, whereas the selection of signal processing technique has a profound influence on the impedance results' accuracy. In order to employ statistical methodology, the EIS is enhanced by using DRBS as excitation signal and complex CWT with Morlet mother wavelet as signal processing tool [10]. This section contains the essentials for performing EIS with stochastic excitation. For more comprehensive explanation on the methodology refer to [10, 12, 13].

2.1. Discrete Random Binary Sequence

The value of DRBS can either be a or $-a$[1]. The changes in value occur only at discrete points in time $k\lambda$ ($k \in \mathbb{N}^0$), where λ is the length of the time interval. Time interval λ is the time between two consecutive points where the signal can change its value. The generating process for DRBS has the following form:

$$X(t) = \sum_n a_n p(t - n\lambda - \alpha),\ \alpha \sim \mathcal{U}[0, \lambda], \tag{1}$$

[1] In a given period of time, the probability of the number of changes of the signal's value is distributed according to Poisson distribution [23, pp. 161–162].

where $p(t)$ is the probability describing the binary change of amplitude a and α is random phase shift, making process $X(t)$ stationary. The power spectral density of the DRBS defined by (1) is:

$$\Phi_X^d(\omega) = a^2\lambda \left| \frac{\sin\left(\frac{\omega\lambda}{2}\right)}{\frac{\omega\lambda}{2}} \right|^2 \tag{2}$$

It has zeros exactly at integer multiples of the frequency $1/\lambda$. The useful frequency band f_B is approximately determined by the -3 dB frequency limit as [24]:

$$f_B \approx \frac{1}{3\lambda} \tag{3}$$

2.2. Continuous Wavelet Transform

Wavelet transform is a time-frequency analysis tool that provides variable time-frequency resolution through the concept of scaling. It is defined through an orthogonal set of functions with compact support called wavelets $\psi(t)$. The wavelet function can be scaled and translated by introducing two parameters s and u as:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right). \tag{4}$$

CWT of a square integrable function $f(t) \in \mathbf{L}^2(\mathbb{R})$ is defined as [25]:

$$Wf(s,u) = \int_{-\infty}^{\infty} f(t)\psi_{u,s}^*(t)\,dt, \tag{5}$$

where $\psi_{u,s}(t)$ is defined by (4). For the application at hand the complex Morlet wavelet was selected as the mother wavelet. The complex Morlet wavelet is defined as [26]:

$$\psi(t) = \pi^{-\frac{1}{4}}\left(e^{-j\omega_0 t} - e^{-\frac{\omega_0^2}{2}}\right)e^{-\frac{t^2}{2}}, \tag{6}$$

where ω_0 is referred to as wavelet's central frequency and is usually set to a value so that the ratio of the highest two peaks of the wavelet is approximately $\frac{1}{2}$:

$$\omega_0 = \pi\left(\frac{2}{\ln 2}\right)^{\frac{1}{2}} \approx 5.336 \tag{7}$$

Impedance through complex wavelet coefficients

The result of the CWT analysis of the voltage $u(t)$ and current $i(t)$, with the Morlet wavelet, is a set of complex wavelet coefficients:

$$\begin{aligned} Wi(t,f) &= \Re\{Wi(t,f)\} + j\Im\{Wi(t,f)\}, \\ Wu(t,f) &= \Re\{Wu(t,f)\} + j\Im\{Wu(t,f)\}. \end{aligned} \tag{8}$$

Finally, the impedance is calculated as the following ratio:

$$Z(t,f) = \frac{Wu(t,f)}{Wi(t,f)}. \tag{9}$$

The ratio (9) provides the value for both the phase and the impedance amplitude for every time-frequency pair.

2.3. Probability Distribution of the Impedance Amplitude

Having a stationary stochastic excitation with zero mean, such as DRBS, makes the spectral components of the measured signals a complex circular Gaussian random variables [27]. Consequently, their corresponding modules $|Wu(t, f_0)|$ and $|Wi(t, f_0)|$ of wavelet coefficients (8) are distributed with Rayleigh distribution as:

$$\begin{aligned} |Wu(t, f_0)| &\sim \text{Rayleigh}(\sigma_u) \\ |Wi(t, f_0)| &\sim \text{Rayleigh}(\sigma_i). \end{aligned} \tag{10}$$

The PDF of the module of the ratio (9) can be calculated as a ratio of two dependent complex circular Gaussian random variables as [28, 29]:

$$f_{|Z|}(|z|) = \frac{2\sigma_i^2\sigma_u^2(1-\rho^2)|z|(\sigma_i^2|z| + \sigma_u^2)}{[(\sigma_i^2|z|^2 + \sigma_u^2)^2 - 4\rho^2\sigma_u^2\sigma_i^2|z|^2]^{3/2}}, |z| \geq 0, \tag{11}$$

where $|\rho| \leq 1$ is a complex correlation coefficient. The PDF (11) is positively skewed and defined on the non-negative semi axis. The CDF that corresponds to (11) reads [28, 29]:

$$F_{|Z|}(z) = \frac{1}{2} - \frac{\sigma_u^2 - z^2\sigma_i^2}{2\sqrt{4z^2\,(1-\rho^2)\,\sigma_i^2\sigma_u^2 + (\sigma_u^2 - z^2\sigma_i^2)^{\,2}}}, \tag{12}$$

where $z \geq 0$. PDF (11) depend on three parameters σ_u, σ_i and the correlation coefficient ρ. These parameters can be easily estimated through the calculated wavelet coefficients as:

$$\begin{aligned} E\{Wu(t, f_0)Wu(t, f_0)^*\} &= \frac{\sigma_u^2}{2} \\ E\{Wi(t, f_0)Wi(t, f_0)^*\} &= \frac{\sigma_i^2}{2} \\ E\{Wi(t, f_0)Wu(t, f_0)^*\} &= \frac{\sigma_u\sigma_i}{2}\rho. \end{aligned} \tag{13}$$

3. Copula functions

The PDF (11) describes the impedance components for a single frequency. In order to encompass the whole frequency range, one has to define a multivariate PDF. Estimating parameters of a multidimensional PDF from data is a difficult task, predominantly due to the amount of required data. Using the traditional kernel based estimation methods, for a d dimensional PDF, one would need N^d data points, where N is the amount required for estimating the marginal distribution in one dimension [30]. Furthermore, the N^d data points should be distributed throughout the observed hyperspace thus capturing the joint PDF shape as accurately as possible. These requirements are rarely met in particular when using experimental data acquired from a system in a fixed operating point.

Copula functions resolve the problem of massive data sets by performing estimation of the joint PDF in two phases. First, the parameters of the d marginals are estimated for each dimension separately. Afterwards,

the copula function's parameters θ are estimated using the original data set. Typically, the copula functions are bivariate functions coupling two random variables. Coupling more than two random variables can be performed in two ways either by defining multivariate copula or by building tree-like hierarchical structures based on bivariate copulas. Due to the flexibility and simplified derivation, the latter strategy is the most commonly employed.

The way how the hierarchical structures are built depends on the copula family. The most popular family of copula functions are the so-called Archimedean copulas. However, the number of original PDFs that have an Archimedean copula in closed form is limited. This is the case with Rayleigh-like distribution, which is crucial in the analysis of impedance values. A solution to this problem are the so-called hierarchical Vine copulas structures [31], which are used in the subsequent analysis. In order to use these structures, several supporting functions have to be derived based on the selected copula function. In what follows, the basic of the Vine copula algorithm is presented accompanied with all the necessary derivations of the supporting functions based on the selected Rayleigh copula.

3.1. Vine copulas

Vine copula algorithm is based on copulas that represent PDF by employing the Bayes' rule. The joint PDF of n dependent random variables can be written as:

$$f(x_1, \ldots, x_n) = f(x_1|x_2, \ldots, x_n) \cdots f(x_{n-2}|x_{n-1}, x_n) f(x_{n-1}|x_n) f(x_n) \tag{14}$$

This decomposition is unique for the specified order of random variables (unless the joint PDF is permutation symmetric). Based on the Sklar's theorem [32], the joint CDF can be written as:

$$F(x_1, \ldots, x_n) = C(F(x_1), F(x_2), \ldots, F(x_n)|\boldsymbol{\theta}), \tag{15}$$

where $C(u_1, \ldots, u_n)$ is the copula CDF function, $u_i = F(x_i)$, $F(x_i)$ is the CDF of x_i and $\boldsymbol{\theta}$ is the set of parameters that have to be estimated. Here and in what follows, for simplified notation, the parameters $\boldsymbol{\theta}$ are omitted when specifying the copula CDFs and PDFs.

For continuous and strictly increasing CDFs, the joint PDF can be derived as the n^{th} partial derivative of the joint CDF resulting into:

$$f(x_1, \ldots, x_n) = c(F(x_1), F(x_2), \ldots, F(x_n)) \times f(x_1) \cdots f(x_n), \tag{16}$$

where $c(u_1, \ldots, u_n)$ is the copula PDF. For a bivariate case (16) becomes:

$$f(x_1, x_2) = c(F(x_1), F(x_2)) f(x_1) f(x_2). \tag{17}$$

Consequently, the bivariate copula PDF directly represents the conditional density as:

$$f(x_1|x_2) = c(F(x_1), F(x_2)) f(x_1). \tag{18}$$

Generalising the joint PDF (16) by using bivariate copula as replacement for the conditional distributions lead to the following relation:

$$f(x|\mathbf{v}) = c_{x|\mathbf{v}}(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j}))f(x|\mathbf{v}_{-j}), \tag{19}$$

where $\mathbf{v}$ is the $n-1$ dimensional vector, v_j represents one arbitrary dimension and $\mathbf{v_{-j}}$ represents $n-2$ dimensional vector with the j^{th} component excluded. For every j the conditional CDF can be calculated as [33]:

$$F(x|\mathbf{v}) = \frac{\partial C_{x,v_j|\mathbf{v}_j}\left(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j})\right)}{\partial F(v|\mathbf{v}_{-j})}. \tag{20}$$

Relation (20) is required in the process of maximum likelihood estimation and is usually referred to as function $h(x, v)$.

3.2. D-Vine copula structure

The decomposition (14) can be performed in many different ways, based on the ordering of the random variables. Bedford and Cooke [34] have introduced a graphical way of representing the possible decomposition of the multidimensional PDF based on bivariate copulas. These approach is called regular vine. Two variations of these approach offer a more practical way of making the multidimensional decomposition: the canonical vine and D-Vine [35]. The presented results are obtained by employing the D-Vine decomposition.

The D-Vine decomposition of a five-dimensional PDF is shown in Figure 1. The input random variables are represented with the circles as the leaves of the tree, whereas the bivariate copulas are represented with edges. Generally there are $n!$ different ways of combining the input variables. By having a symmetric copulas, these number reduces down to $\frac{n(n-1)}{2}$.

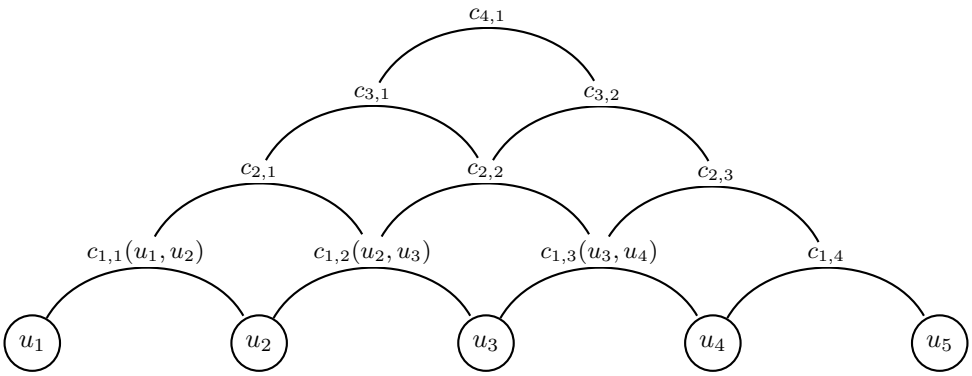

Figure 1: Structure of the selected D-vine copula hierarchy.

3.3. Maximum likelihood estimation algorithm for D-Vine copulas

The modelling process with D-Vine copulas, or for that matter, with any type of copula structures is a two step process. The first step is determining the families of bivariate copulas that will be employed

when building the appropriate tree-like structure. The second step is estimation of the set of parameters θ that are associated with each bivariate copula. It should be noted that when building a tree-like copula decomposition one can mix various copula families if necessary.

There are many ways how to determine the most suitable bivariate copula family for the problem in hand [31]. However, as shown in Section 2, the marginal PDFs are Rayleigh, thus the subsequent analysis is performed by using Rayleigh bivariate copula as the main building block.

The estimation of the corresponding set of parameters θ can be performed using a pre-determined maximum likelihood estimation algorithm that is independent of the choice of bivariate copula family. The algorithm requires recursive calculation of likelihood values and evaluation of the $h(x, v|\theta)$ function (20).

4. Rayleigh copula and the corresponding $h(x, v)$ functions

The Rayleigh copula CDF can be obtained through

$$C(u,v) = (1-\rho)\int_0^{a_1}\int_0^{a_2} e^{-z_1-z_2} I_0(2\sqrt{\rho z_1 z_2})dz_1 dz_2, \tag{21}$$

where $I_k(x)$ is the modified Bessel's function of first kind of order k. Using the transformation [36, Eq. (2.14)], the double integral can be transformed into:

$$C(u,v) = 1 + e^{\rho a_2 - a_2}\left(e^{-a_1}\int_0^{\rho a_2} e^{-s} I_0\left(2\sqrt{s a_1}\right) ds - 1\right) - e^{-a_1}\int_0^{a_2} e^{-s} I_0\left(2\sqrt{\rho a_1 s}\right) ds, \tag{22}$$

where

$$\begin{aligned} a_1 &= -\frac{\log(1-u)}{1-\rho} \\ a_2 &= -\frac{\log(1-v)}{1-\rho}. \end{aligned} \tag{23}$$

In order to employ the D-Vine copula maximum likelihood estimation algorithm, the corresponding $h(\cdot)$ functions read as:

$$h_x(y) = F_{Y|X}(y|x) = \frac{\partial C(u,v)}{\partial v}. \tag{24}$$

$$h_y(x) = F_{X|Y}(x|Y) = \frac{\partial C(u,v)}{\partial u}. \tag{25}$$

The notation of $h(\cdot)$ functions are adopted in order to be inline with the original copula vine algorithm as defined by Aas et al. [31].

The partial derivative of (24) reads as:

$$\begin{aligned} \frac{\partial C(u,v)}{\partial v} =& 1 - (1-u)^{\frac{1}{1-\rho}}(1-v)^{\frac{\rho}{1-\rho}} I_0\left(2\sqrt{\rho a_1 a_2}\right) \\ & - (1-u)^{\frac{1}{1-\rho}}\int_0^{\rho a_2} e^{-s} I_0\left(2\sqrt{s a_1}\right) ds \end{aligned} \tag{26}$$

The integral in (26) can be evaluated through the Marcum Q-function as:

$$\int_0^{qa_2} e^{-s} I_0\left(2\sqrt{sa_1}\right)\, ds = \left[1 - Q_1\left(\sqrt{2a_1}, \sqrt{2\rho a_2}\right)\right] e^{a_1}, \tag{27}$$

where

$$Q_M(a,b) = \int_b^{\infty} x \left(\frac{x}{a}\right)^{M-1} \exp\left(-\frac{x^2+a^2}{2}\right) I_{M-1}(ax)\, dx. \tag{28}$$

Similarly, the partial derivative of (25) reads as:

$$\begin{aligned}\frac{\partial C(u,v)}{\partial u} =& \frac{(1-u)^{\frac{\rho}{1-\rho}} \overbrace{\int_0^{a_2} e^{-s} I_0\left(2\sqrt{\rho s a_1}\right)\, ds}^{(30)}}{1-\rho} - \frac{\rho(1-u)^{\frac{1}{1-\rho}}}{(1-\rho)(1-u)} \overbrace{\int_0^{a_2} \frac{e^{-s} s I_1\left(2\sqrt{\rho s a_1}\right)}{\sqrt{\rho s a_1}}\, ds}^{(31)} \\ &+ \frac{(1-u)^{\frac{1}{1-\rho}}(1-v)}{(1-\rho)(1-u)} \left(\underbrace{\int_0^{\rho a_2} \frac{e^{-s} s I_1\left(2\sqrt{s a_1}\right)}{\sqrt{s a_1}}\, ds}_{(32)} - \underbrace{\int_0^{\rho a_2} e^{-s} I_0\left(2\sqrt{s a_1}\right)\, ds}_{(27)} \right).\end{aligned} \tag{29}$$

All of the integrals involving the modified Bessel's functions can be solved through the Marcum Q-function by using the following relations:

$$\int_0^{a_2} \frac{e^{-s} s I_1\left(2\sqrt{a_1 \rho s}\right)}{\sqrt{a_1 \rho s}}\, ds = \left[1 - Q_2\left(\sqrt{2\rho a_1}, \sqrt{2a_2}\right)\right] e^{\rho a_1} \tag{30}$$

$$\int_0^{a_2} e^{-s} I_0\left(2\sqrt{\rho a_1 s}\right) ds = \left[1 - Q_1\left(\sqrt{2\rho a_1}, \sqrt{2a_2}\right)\right] e^{\rho a_1} \tag{31}$$

$$\int_0^{\rho a_2} \frac{e^{-s} s I_1\left(2\sqrt{a_1 s}\right)}{\sqrt{a_1 s}}\, ds = \left[1 - Q_2\left(\sqrt{2a_1}, \sqrt{2\rho a_2}\right)\right] e^{a_1}. \tag{32}$$

The derivation of the substitution integrals is given in Appendix A. Introducing the Marcum Q-function significantly simplifies the numerical evaluations of the derived relations. As a result, the parameter estimation can be done in a computationally efficient manner.

4.1. Selection of the D-Vine structure

The final task when building the D-Vine copula structure is to determine which of the $n!$ possible trees is the most suitable for the problem at hand. The selection determines the order in which the input random variables are coupled during the process of building the multidimensional PDF. Selecting the most appropriate structure can be obtained either by using goodness-of-fit tests and testing each of the possible combinations or by incorporating an experts knowledge. The former approach is rather time consuming, taking into account that the number of possible tree combinations increases by $n!$. Therefore, for the case of PEM fuel cells condition monitoring, the latter approach is selected.

Having EIS data as the main source of information, the goal is to capture its structure in the best possible manner. For fuel cells, the impedance characteristics has two distinct lobes whose shape and location determine the condition of the fuel cell under test. As a result, the subsequent analysis was performed by selecting five frequencies that cover the measured frequency band from 0.8 up to 150 Hz. The frequency interval was divided in a dyadic manner. The selected frequencies are shown to be sufficient for the purpose of capturing the main shape properties of the impedance characteristics.

The usual practice is to build the tree like structure by starting with those components that exhibit the biggest dependence structure [35]. The dependence values were calculated for all possible pairs of bivariate Rayleigh copulas using a data set from fault free operation of the PEM fuel cell under observation. The results are given in Table 1. Due to symmetry only the upper triangular elements are shown.

Table 1: Dependence values among the selected five frequency components

	f_1	f_2	f_3	f_4	f_5
f_1		0.4285	0.1536	0.0569	0.0001
f_2			0.4232	0.1468	0.0126
f_3				0.3437	0.0695
f_4					0.4581

The values shown in Table 1 offer additional insight of the impedance characteristic of the PEM fuel cell under test. As expected, the frequencies that are closer to each other show higher degree of dependence. The biggest dependence is between the highest frequency components, thus $c_{4,5}$ is the first copula. The next component is the one that has the highest dependence with either the 5^{th} or the 4^{th} component, i.e., the component f_3. Consequently, the order in which frequency components enter in the process of building the D-Vine was started the most dependent ones. After applying the maximum likelihood algorithm for the complete structure the θ values for each of the copula nodes are shown in Figure 2.

These maximum likelihood estimates of the $\boldsymbol{\theta}$ parameters reflect the dependence structure among the frequency components. The parameter $\theta_{1,2}$, which couples the frequency components f_3 and f_5 has quite low values, indicating low dependence. The same effect is visible for node $\theta_{3,1}$. Copula at this node "connects" the high and low frequency arcs of the PEM fuel cell impedance characteristic. Since $\theta_{3,1} = 0$, the resulting copula is the independence one, i.e., the joint PDF is a product of the corresponding copulas.

4.2. Comparison with other copula functions

Comparing the performance among copula functions is performed by employing various goodness-of-fit measures. The most commonly used ones are Akaike information criterion (AIC) or Bayesian information

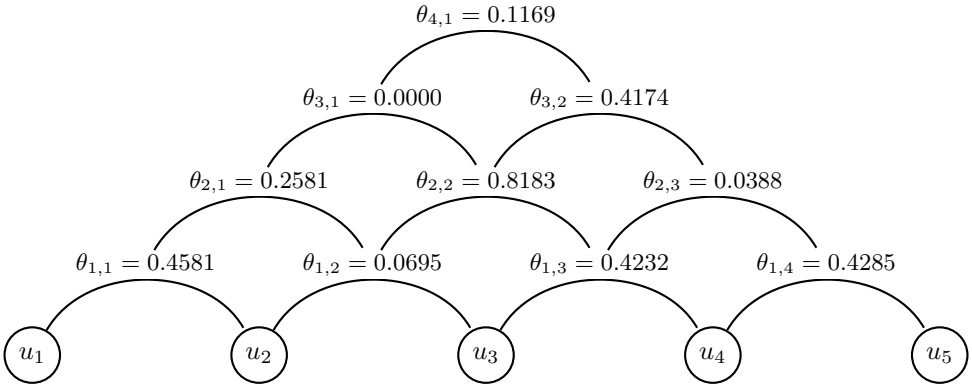

Figure 2: Estimated θ values for the D-Vine structure.

criterion (BIC):

$$\begin{aligned} AIC &= -2\log(\mathcal{L}(\mathbf{U}|\boldsymbol{\theta})) + 2k \\ BIC &= -2\log(\mathcal{L}(\mathbf{U}|\boldsymbol{\theta})) + k\log(n), \end{aligned} \tag{33}$$

where $\mathcal{L}(\mathbf{U}|\boldsymbol{\theta})$ is the maximum likelihood of the copula model based on the observation $\mathbf{U}$ and parameters $\boldsymbol{\theta}$. The parameter k represents the cardinality of $\boldsymbol{\theta}$ and n is the number of observations i.e., the cardinality of $\mathbf{U}$. The bivariate copula functions usually have the similar number of parameters, and the evaluation is performed using the same number of observations n. Consequently, the last terms in the AIC and BIC in (33) will be constant regardless of the selected copula function. As a result the model selection is performed solely on the maximum likelihood value $\mathcal{L}(\mathbf{U}|\boldsymbol{\theta})$.

A more robust approach is to select the copula family through the concepts of mutual information [37]. The mutual information describes the mutual dependence of two random variables as:

$$I(X,Y) = \iint_{x\in X, y\in Y} f_{X,Y}(x,y)\log\left(\frac{f_{X,Y}(x,y)}{f_X(x)f_Y(y)}\right)\,dxdy. \tag{34}$$

Having copula function the mutual information can be calculated as [38]:

$$I(X,Y) = \iint_{[0,1]^2} c(u,v)\log\left[c(u,v)\right]\,dudv, \tag{35}$$

where $u = F_X^{-1}(x)$ and $v = F_Y^{-1}(y)$. By using the data from the fault-free operation of the PEM fuel cell under examination the calculated mutual information values are shown in Table 2. It is clearly visible that the selection of Rayleigh copula is justified as it was initially assumed based on the marginal PDFs of the impedance values.

5. Results

The evaluation of the proposed methodology was performed on a commercial PEM fuel cell system Hydrogenics HyPM HD 8. The stack consists of 80 cells, each with surface area of 200 cm^2, providing

Table 2: Mutual information values for various copula families estimated on EIS data

	Copula family				
	Rayleigh	Gauss	Student-t	Clayton	Gumbel
Mutual Information	**0.1003**	0.0036	0.051	6.756×10^{-4}	1.16×10^{-11}

8.5 kW of electric power in total. The fuel cell system operates on pure hydrogen and ambient air. Besides the fuel cell system, the experimental setup also included an electronic load, function generator and data acquisition equipment (block diagram of the setup is shown Figure 3).

The DRBS excitation waveform was generated by the external function generator and superimposed to the DC current of the fuel cell stack by the electronic load. The current and voltage signals were measured using a custom-designed voltage monitor [11]. The signals were sampled with frequency of $f_s = 5$ kHz. Individual measurement sections (data acquisition segments) were repeated every 40 seconds. In this way, a set of 105 measurements were acquired.

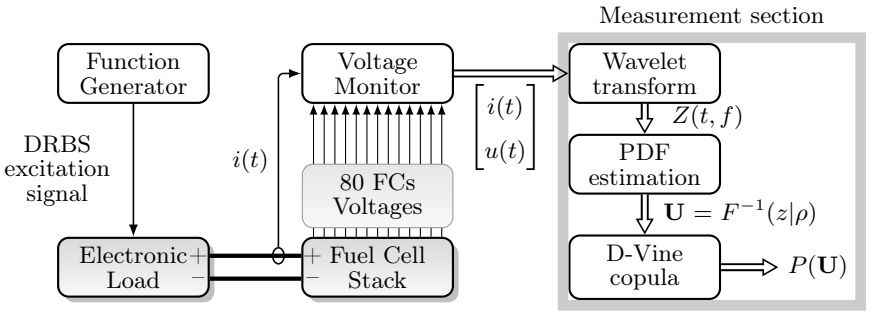

Figure 3: Block diagram of the experimental setup and signal processing steps.

5.1. Measurement section explained

The first step of the measurement section is transforming the acquired voltage $u(t)$ and electrical current $i(t)$ into wavelet coefficients (8). Afterwards, the parameters σ_u, σ_i and ρ are estimated through (13). The parameters are estimated for each of the selected frequencies f_i, where $i \in [1, \ldots, N_f]$. Therefore the result is $n \times N_f$ matrix, where n is the number of samples.

The impedance values are than transformed into a $n \times N_f$ matrix $\mathbf{U}$ of uniformly distributed random variables. The transformation is performed by using the inverse CDF, i.e., by applying inverse transform sampling for each sample separately. The last step is the evaluation of the D-Vine structure based on the initially estimated parameters $\boldsymbol{\theta}$. The result is a set of n probabilities of observing particular impedance values, one for each acquired sample. This process is shown in Figure 4.

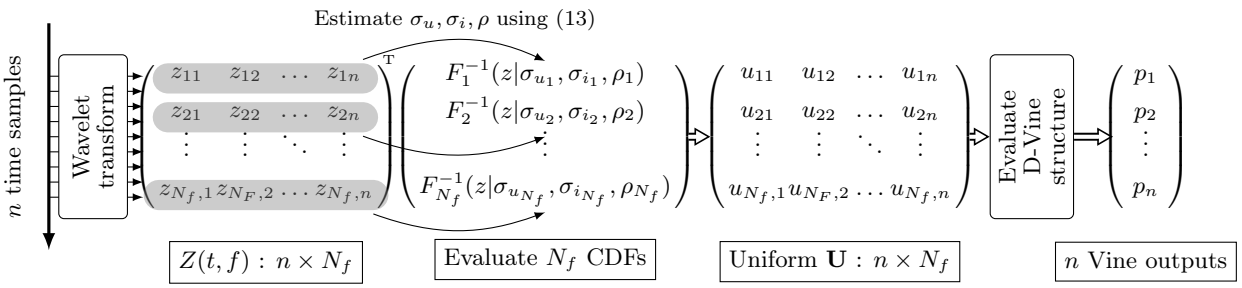

Figure 4: Calculation steps for each measurement section.

It should be noted that the D-Vine parameters $\boldsymbol{\theta}$ are estimated from data acquired from nominal operation, which in our case was the very first measurement. The $\boldsymbol{\theta}$ parameters are estimated using the D-Vine maximum likelihood estimation algorithm [35]. For the experiment at hand the estimated $\boldsymbol{\theta}$ values are presented in Section 4.1. Every subsequent evaluation of the D-Vine structure uses these initially estimated values of $\boldsymbol{\theta}$. The output $p(U_i)$ represents the copula output for each measured data point.

An alternative approach would include estimating new copula parameters for each measurement section and then quantifying the alteration of the resulting multidimensional PDFs. Such an approach has two main drawbacks. First, the estimation of the D-Vine structure parameters i.e., the $\boldsymbol{\theta}$ parameters, is time consuming thus in our approach this process is performed only ones. Second, using any form of comparison of two PDFs (such as Kullback-Leibler divergence or similar approaches) requires either close form solution of the PDFs or numerical evaluation of the divergence measures. In either way, this would be computationally demanding and highly impractical for online operation.

5.2. Experimental profile

Throughout the experiment, the fuel cell system was kept at constant operating and environmental conditions. Airflow temperature was kept constant at 50°C with stoichiometry 2.5. Grade 5 hydrogen at a constant temperature of 20°C was used as fuel. The fuel cell operating point was set to 70 A DC, resulting in the fuel cell stack output voltage of 55 V. The amplitude of the superposed DRBS waveform was set to 2 A. As such, the peak-to-peak amplitude was 4 A and therefore small enough not to cause difficulties due to non-linearity of the PEM fuel cells.

Two types of water management faults were included in the evaluation process, i.e., flooding of the cells and drying of the membranes. The rationale behind the selected fault modes is twofold. First, water management faults can be induced by manipulating the humidity of the inlet air, while preserving constant operating conditions. Second and more important fact is the reversibility of such faults. Therefore, the changes in the fuel cell condition can be performed in a timely manner while at the same time preventing any substantial damage of the system.

The experiment went through five stages:

1. nominal inlet air humidity for 11 minutes (9% which at 50°C corresponds to 7.5 g/m^3 - grams of water per cubic meter of air)
2. dried inlet air for 10 minutes (5% which at 50°C corresponds to 4.1 g/m^3)
3. nominal inlet air humidity for 10 minutes
4. 100% humid inlet air for 30 minutes (100% which at 50°C corresponds to 83 g/m^3) and
5. nominal inlet air for the last measurements.

The interim inlet air with nominal humidity was fed in order to cancel the influence of the previous fault and monitor the recovery of the fuel cell.

5.3. Time evolution of the copula output

Based on the procedure described in Section 5.1, the copula output $p(\mathbf{U}_i)$ was calculated for each measurement section. The results of the D-Vine copula PDF evaluation are shown in Figure 5. The box plot represents the values of the D-Vine structure for each sample of the corresponding measurement section. The boxes depict the interquartile interval and points are the median values.

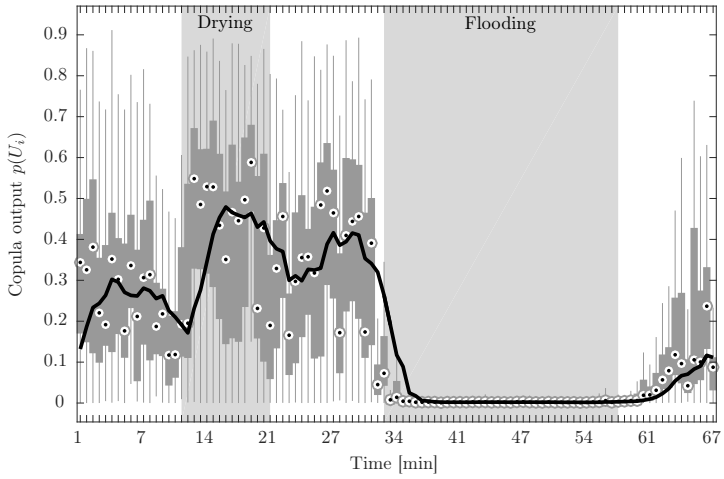

Figure 5: Evolution of the condition indicator (points are the median values and the line is its time average).

In the first 10 minutes the PEM fuel cell was operated under nominal conditions. The copula output during this period can be regarded as the reference one. During the period with dry inlet air, the copula output shifts towards the upper tail, with respect to the initial fault free region. Feeding saturated inlet air at 100% humidity, causes even more significant change in the D-Vine values, thus indicating a highly unlikely impedance. By restoring the water management balance, the D-Vine output returns back to the nominal values.

5.4. Discussion of the results

From the results shown in Figure 5, it is apparent that the influence of the dry inlet air is significantly smaller than the one imposed by the humid air. There are two main reasons for this discrepancy. The used fuel cell system was designed to operate without an external humidifier unit, therefore inducing drying fault is rather challenging. Furthermore, in order to prevent any local excessive drying and thus damaging the membranes, this phase was kept relatively short, i.e., 10 minutes. Due to these factors it was possible to achieve only mild drying of the fuel cell stack, hence the small deviation of the copula output.

Notwithstanding the mild drying, its effects are visible even after the humidity of the inlet air was returned to normal. As described by Ji and Wei [39], at low current densities, back diffusion will prevail on electroosmotic drag, while at high current densities, electro-osmotic drag will prevail over back diffusion and thus the anode will tend to dry out, even if the cathode is well hydrated.

This analysis shows that the copula output can be directly related to the fault's severity. Consequently, the proposed approach can be employed for the process of performing fault detection and fault evaluation. Furthermore, any type of fault affects the cell's system and consequently its impedance characteristic. This influence will influence the copula's output too. Consequently, the proposed solution can be applied for any approach that operates with impedance data regardless of the system's nature.

5.5. Comparison with previous research

Copula functions have already been applied for condition monitoring of PEM fuel cells in the study by Mileva Boshkoska et al. [21]. Unlike the present approach, the original study was performed using Archimedean copulas. Due to the straightforward parameter estimation method, based on the so-called generator functions, Archimedean copulas are readily implemented in various numerical packages. The weakest point of that approach is suboptimal selection of the copula family. Despite having suboptimal copula, the results confirmed the applicability of copula based approaches for condition monitoring of PEM fuel cells.

The analysis presented in this paper goes one step further. The governing factor in the selection of copula family is the nature of the random variables at hand. The PDF of the impedance components is a ratio of two Rayleigh distributed random variables (as shown in Section 3.3). For PDF that belong to the family of exponential distributions, which is the case for the Rayleigh distribution, it is not possible to construct an Archimedean copula (a generator function does not exist in the closed form). Addressing this issue, the present analysis is based on the Rayleigh copula. Being a non-Archimedean type imposes significant difficulties in the process of parameter estimation and numerical evaluation. The presented material contains all the necessary derivations required for the numerical implementation of the parameter estimation algorithm for the Rayleigh copula.

Both studies, the present one and the study of Mileva Boshkoska et al. [21], were evaluated with the same data set. As shown by the estimated mutual information values, presented in Table 2, the Rayleigh copula significantly outperforms the two Archimedean copulas that were employed in the first study.

6. Conclusion

With stochastic excitation signals, the impedance amplitude is a ratio of two Rayleigh random variables. As a result, the complete impedance characteristic itself can be analysed through the framework of multivariate random variables, where each dimension represents the impedance at single frequency. The multivariate analysis is performed through the concepts of copula functions.

Since the nature of the underlying marginal distributions stems from the Rician family, the analysis was performed using the Rayleigh copula. As there are no available implementations for such an approach, this paper contains all the necessary mathematical derivations required for numerical implementation of the proposed algorithm.

The output of the proposed copula structure is a probability of observing a particular impedance characteristic vis-á-vis the nominal one. Therefore, the output can be treated as a unit-free condition indicator. The performance of the condition indicator was tested on a 8.5 kW industrial PEM fuel cell for detection and evaluation of fooding and drying faults with various severity.

There are several practical benefits offered by the copula based statistical condition indicator. Online condition monitoring of PEM fuel cells can be performed without any prior characterisation of the device under various fault modes. The tuning is done using data from nominal operation. The value of the condition indicator can be directly related to the fault severity. Finally, since the statistical properties of the condition indicator are known, the calculation of the fault alarm thresholds can be based on a desired probability of false alarm.

Appendix A. Derivation of substitution integral (30)

By entering substitution

$$s = \frac{x^2}{2} \tag{A.1}$$

the integral (30) becomes

$$\begin{aligned}\int_0^{\sqrt{2a_2}} \frac{e^{-\frac{x^2}{2}} s I_1\left(2\sqrt{a_1\rho s}\right)}{\sqrt{a_1\rho s}}\, ds &= \int_0^{\sqrt{2a_2}} e^{-\frac{x^2}{2}} \frac{x^2}{\sqrt{2\rho a_1}} I_1\left(x\sqrt{2\rho a_1}\right)\, dx \times e^{-\frac{2\rho a_1}{2}} e^{\frac{2\rho a_1}{2}} \\ &= e^{\frac{2\rho a_1}{2}} \int_0^{\sqrt{2a_2}} e^{-\frac{x^2+2\rho a_1}{2}} \frac{x^2}{\sqrt{2\rho a_1}} I_1\left(x\sqrt{2\rho a_1}\right)\, dx\end{aligned} \tag{A.2}$$

The last part of (A.2) has the required form of the Marcum Q-function (28). The integral limits can be adjusted by taking into account that the Marcum Q-function is the survival function of the noncentral χ^2 distribution.

The survivor function $s(x)$ for an arbitrary distribution specified by its CDF $F_x(x)$ and PDF $f_X(x)$ is defined as

$$s(x) = \int_x^\infty f_X(t)\,dt = 1 - F_X(x). \tag{A.3}$$

Therefore, the relation $1 - Q_M(a, b)$ is a CDF of a non-central χ^2 distribution with $k = 2M$ moments with parameters a and b. Being a CDF, makes the last integral in (A.2) equal to $1 - Q_2(a, b)$ where $a = \sqrt{2\rho a_1}$ and $b = \sqrt{2a_2}$. Similar derivation can be performed for the remaining substitution integrals needed for calculation of functions $h_y(x)$ and $h_x(y)$.

Acknowledgments

The authors acknowledge the project L2-7668 was financially supported by the Slovenian Research Agency. The authors also acknowledge the financial support of the Slovenian Research Agency through Research Programme P2-0001. The authors would like to thank the anonymous referees for very detailed and helpful remarks.

[1] V. Jovan, M. Perne, J. Petrovčič, An assessment of the energetic flows in a commercial PEM fuel-cell system, Energy Convers. Manag. 51 (12) (2010) 2467–2472, doi:10.1016/j.enconman.2010.04.014.

[2] B. Pregelj, A. Debenjak, G. Dolanc, J. Petrovčič, A diesel-powered fuel cell APU – reliability issues and mitigation approaches, IEEE Transactions on Industrial Electronics PP (99) (2017) 1–11, doi:10.1109/TIE.2017.2674628.

[3] F. Barbir, PEM Fuel Cells: Theory and Practice, Elsevier, 2005.

[4] A. Heng, S. Zhang, A. Tan, J. Mathew, Rotating machinery prognostics: State of the art, challenges and opportunities, Mechanical Systems and Signal Processing 23 (3) (2009) 724–739, doi:10.1016/j.ymssp.2008.06.009.

[5] J. Zhang, J. Lee, A review on prognostics and health monitoring of Li-ion battery, Journal of Power Sources 196 (15) (2011) 6007–6014, doi:10.1016/j.jpowsour.2011.03.101.

[6] W. Wang, H. Wang, Preventive replacement for systems with condition monitoring and additional manual inspections, European Journal of Operational Research 247 (2) (2015) 459 – 471, ISSN 0377-2217.

[7] M. Hu, J. D. Weir, T. Wu, Decentralized operation strategies for an integrated building energy system using a memetic algorithm, European Journal of Operational Research 217 (1) (2012) 185 – 197, ISSN 0377-2217.

[8] P. Jochem, M. Schönfelder, W. Fichtner, An efficient two-stage algorithm for decentralized scheduling of micro-CHP units, European Journal of Operational Research 245 (3) (2015) 862 – 874, ISSN 0377-2217.

[9] X.-Z. Yuan, C. Sons, H. Wang, J. Zhang, Electrochemical Impedance Spectroscopy in PEM Fuel Cells, Fundamentals and Applications, Springer, London, 2010.

[10] A. Debenjak, P. Boškoski, B. Musizza, J. Petrovčič, Đ. Juričić, Fast measurement of proton exchange membrane fuel cell impedance based on pseudo-random binary sequence perturbation signals and continuous wavelet transform, Journal of Power Sources 254 (2014) 112–118, doi:http://dx.doi.org/10.1016/j.jpowsour.2013.12.094.

[11] A. Debenjak, J. Petrovčič, P. Boškoski, B. Musizza, Đani Juričić, Fuel cell condition monitoring system based on interconnected DC-DC converter and voltage monitor, IEEE Transactions on Industrial Electronics 62 (8) (2015) 5293–5305, doi:10.1109/TIE.2015.2434792.

[12] P. Boškoski, A. Debenjak, Optimal selection of Proton Exchange Membrane fuel cell condition monitoring thresholds, Journal of Power Sources 268 (2014) 692–699.

[13] J. Schoukens, M. Vaes, R. Pintelon, Linear System Identification in a Nonlinear Setting: Nonparametric Analysis of the Nonlinear Distortions and Their Impact on the Best Linear Approximation, IEEE Control System Magazine 36 (3) (2016) 38–69, doi:10.1109/MCS.2016.2535918.

[14] R. Pintelon, J. Schoukens, System identification: A Frequency Domain Approach, IEEE Press, 2001.

[15] A. Lasia, Electrochemical Impedance Spectroscopy and its Applications, Springer-Verlag, New York, doi:10.1007/978-1-4614-8933-7, 2014.

[16] S. Wasterlain, D. Candusso, F. Harel, D. Hissel, X. François, Development of new test instruments and protocols for the diagnostic of fuel cell stacks, Journal of Power Sources 196 (12) (2011) 5325–5333.

[17] C. Brunetto, A. Moschetto, G. Tina, PEM fuel cell testing by electrochemical impedance spectroscopy, Electronic Power Systems Research 79 (2008) 17–26.

[18] C. de Beer, P. S. Barendse, P. Pillay, Fuel Cell Condition Monitoring Using Optimized Broadband Impedance Spectroscopy, IEEE Transactions on Industrial Electronics 62 (8) (2015) 5306–5316.

[19] N. Katayama, S. Kogoshi, Real-Time Electrochemical Impedance Diagnosis for Fuel Cells Using a DC-DC Converter, IEEE Transactions on Energy Conversion 30 (2) (2015) 707–713.

[20] S. Wu, Construction of asymmetric copulas and its application in two-dimensional reliability modelling, European Journal of Operational Research 238 (2) (2014) 476 – 485, ISSN 0377-2217.

[21] B. Mileva Boshkoska, P. Boškoski, A. Debenjak, Đani Juričić, Dependence among complex random variables as a fuel cell condition indicator, Journal of Power Sources 284 (2015) 566–573, doi:10.1016/j.jpowsour.2015.03.044.

[22] X. Zeng, J. Ren, Z. Wang, S. Marshall, T. Durrani, Copulas for statistical signal processing (Part I): Extensions and generalization, Signal Processing 94 (2014) 691 – 702, ISSN 0165-1684.

[23] R. Isermann, M. Münchhof, Identification of Dynamic Systems: An Introduction with Applications, Advanced textbooks in control and signal processing, Springer-Verlag Berlin Heidelberg, ISBN 9783540788799, doi:http://dx.doi.org/10.1007/978-3-540-78879-9, 2011.

[24] W. Davies, System identification for self-adaptive control, John Wiley & Sons, New York, ISBN 9780471198857, 1970.

[25] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, Elsevier Academic Press, 3 edn., ISBN 9780080922027, 2008.

[26] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, ISBN 9780898712742, 1992.

[27] P. O. Amblard, M. Gaeta, J. L. Lacoume, Statistics for complex variables and signals – Part I: Variables, Signal Processing 53 (1996) 1–13.

[28] M. K. Simon, Probability Distributions Involving Gaussian Random Variables, Springer Science, 2006.

[29] P. Boškoski, A. Debenjak, B. Mileva Boshkoska, Fast Electrochemical Impedance Spectroscopy As a Statistical Condition Monitoring Tool, SpringerBriefs in Applied Sciences and Technology, Springer International Publishing, doi:10.1007/978-3-319-53390-2, 2017.

[30] B. W. Silverman, Densiy Estimation for Statistics and Data Analysis, Chapman and Hall, 1986.

[31] K. Aas, C. Czado, A. Frigessi, H. Bakken, Pair-copula constructions of multiple dependence, Technical report SAMBA/24/06, Norwegian Computing Center, 2006.

[32] A. Sklar, Distributions with Fixed Marginals and Related Topics - Random Variables, Distribution functions, and Copulas - A Personal Look Backward and Forward, vol. 28, Institute of Mathematical Statistics, Hayward, CA., 1996.

[33] H. Joe, Multivariate Models and Dependence Consepts, Chapman and Hall, 1997.

[34] T. Bedford, R. M. Cooke, Vines: A new graphical model for dependent random variables, The Annals of Statistics 30 (4)

(2002) 1031–1068.

[35] D. Schirmacher, E. Schirmacher, Multivariate dependence modeling using pair copulas, Tech. Rep., Society of Acturaries, 2008.

[36] N. M. Temme, A Double Integral Containing the Modified Bessel Function: Asymptotic and Computation, Mathematics of Computation 47 (176) (1986) 683–691.

[37] X. Zeng, J. Ren, M. Sun, S. Marshall, T. Durrani, Copulas for statistical signal processing (Part II): Simulation, optimal selection and practical applications, Signal Processing 94 (2014) 681 – 690, ISSN 0165-1684.

[38] X. Zeng, T. S. Durrani, Estimation of mutual information using copula density function, Electronics Letters 47 (8) (2011) 493–494, ISSN 0013-5194, doi:10.1049/el.2011.0778.

[39] M. Ji, Z. Wei, A Review of Water Management in Polymer Electrolyte Membrane Fuel Cells, Energies 2 (4) (2009) 1057–1106.