Statistical Characterization of Some Electrical and Mechanical Phenomena by a Neural Probability Density Function Estimation Technique

Simone Fiori and Roberto Rossi

Affiliation:
Facoltà di Ingegneria, Università di Perugia
Polo Didattico e Scientifico del Ternano
Loc. Pentima bassa, 21, I-05100 Terni (Italy)

E-mail: fiori@unipg.it

Pages: 29, **Figures and Tables**: 14, **References**: 43

Accepted for publication in:
Neural Networks World

Submitted: July 30, 2003. Revised: January 14, 2004.
Accepted: April 7, 2004

Statistical Characterization of Some Electrical and Mechanical Phenomena by a Neural Probability Density Function Estimation Technique

Simone Fiori and Roberto Rossi

Abstract

The present paper concerns the estimation of probability density functions using the particular parameterized class of distribution functions implemented by a single non-linear neuron, introduced in the previous contribution [12]. The estimation procedure is applied to the statistical characterization of some electrical and mechanical phenomena.

1 Introduction

Probability density function (PDF) approximation by neural systems is an active and challenging research field, because it conjugates the intrinsic relevance of PDF estimation problem with the well-known functional-approximation ability exhibited by artificial neural systems.

Generally speaking, the problem of modeling a probability density function, given a finite number of data points, plays a central role in the description of a structured data-space, where the points are not uniformly distributed but their density changes from region to region. Density estimation is an important problem studied in statistical literature [39] as well as in pattern recognition [2]. As meaningful examples, it is worth citing the applications to linear system analysis [26], blind signal processing by neural networks and neuromorphic adaptive filters [9, 10, 11, 13, 16], physics and chemistry [4, 23, 24, 43], engineering data classification [2] and in economical sciences [3, 35, 42].

A typical problem is PDF approximation by sample analysis from incomplete data [5]: It regards the estimation of the PDF of data-sets when some particular features of the true PDF are obtained from the available samples. The term 'incomplete' in general denotes the practical situation where the amount of available data is insufficient to provide a complete description of the statistical properties of the phenomena that produced the data. It is interesting to just mention the dual problem to data scarcity, that arises in those situations where there is an overwhelming amount of logged data [31].

The most widespread approaches to PDF estimation are based on formalisms developed in statistics to describe nearly-Gaussian distributions [5]; two such

formulations are often found in the scientific literature: The cumulant or *Edgeworth expansion* and the quasi-moment or *Gram-Charlier expansion*. Other well-known approaches rely on simple histogram method [2], on the mixture-of-kernels technique (leading e.g. to the EM-algorithm) [2, 6] and on flexible neural structures [2, 12, 30, 37].

In [12], the first Author introduced and discussed a density estimation technique based on an unsupervised adaptive-activation-function (FAN) neuron model.

The first mathematical model of a neuron was proposed by McCulloch and Pitts in 1943. The underlying idea that this model tries to capture is that the response function of a neuron is a weighted sum of its inputs squashed by a nonlinear function. Much research work has been carried out in the field of neural networks since that time. In particular, the studies on non-linear models have been promoted after the observation that the response of biological neurons exhibits non-linear behavior and some evidences suggest that non-linear information processing takes place in the dendritic trees [7, 29]. Consequently, highly non-linear formal neurons are required to model these cells. A detailed survey of the advancements in this field was given in [8], while a more recent review was proposed in [40].

In [12], we demonstrated the usefulness of the proposed model and discussed its features, in comparison to Gram-Charlier, maximum-entropy and kernel-based PDF expansion/estimation approaches. The conclusions of the previous contributions were that the FAN models may be superior to both expansion-based and kernel-based estimation approaches. In particular, in [12], we showed with practical experiments as well as analytical considerations that the results achieved with expansion-based methods may be not very good when the data under analysis are described by probability density functions that are not very close to a Gaussian, thus the Gram-Charlier approximation may fail; also, the results obtained with EM-type algorithms may be not much satisfactory, especially considering the high computational cost exhibited by the EM algorithm.

The present paper has essentially a practice-and-experience nature: Our main aim is to discuss applications of the mentioned PDF-estimation technique to the statistical characterization of some electrical and mechanical phenomena. To this aim, it is necessary to experimentally find the best set of parameters that allow the neural algorithm to converge to the proper estimates. The obtained results allow us to acquire knowledge on practical PDF estimation on real-world incomplete data through the FAN system.

2 Study Motivation and FAN Model

Let us consider the random variable x, ranging in $[a, b]$, whose PDF is to be estimated.[1] In order to develop the following considerations, we make the technical hypothesis on the probability density function to be approximated that it is zero-mean, namely $E_x[x] = 0$. In the event that the random variable x does not meet this restriction, it is always possible to replace it with its normalized version. Also, with $f_x(x)$ it is denoted the probability density function of the random variable x.

In the following we discuss in general the problem of PDF estimation in order to motivate the neural-based estimation choice. We also briefly recall from [12] the neuron model under analysis and its principal features.

2.1 Study Motivation

The motivation of the present experimental study has its roots in the practical difficulty in the numerical estimation of statistical quantities, such as the empirical differential (Shannon) entropy, in an incomplete-data setting. The usefulness of reliable entropy-estimation techniques is known in different scientific fields, such as for instance in the psychophysics of taste [34], in cellular automata analysis and natural text compression [38], in supervised and unsupervised neural network learning theory [9] and in data-analysis by projection pursuit [18].

We chose to illustrate such difficulty in the present section with two groups of examples, for which we try to tackle in a purely numerical way the problem of empirical entropy estimation. In particular, we consider the simplest method for estimating a probability density distribution, based on samples histogram: It consists in partitioning the empirical data range into equally spaced bins and estimating the PDF at every bin-center value as the normalized number of samples that fall inside that bin.

In the first group, we consider two random variables whose probability density functions are known analytically and whose differential entropies can be computed in closed form, namely, a zero-mean Gaussian distribution $f_G(x)$ and a zero-mean Laplacean distribution $f_L(x)$. These write, respectively, as:

$$f_G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \; f_L(x) = \frac{\lambda}{2} e^{-\lambda|x|} \; , \tag{1}$$

[1] From a practical point of view, even PDFs endowed with an unlimited support are supposed to be estimated within a closed (although conveniently wide) interval.

where σ^2 denotes the variance of the Gaussian distribution and $\lambda > 0$ is the dispersion parameter of the Laplacean distribution.

The entropy of a distribution $f_x(x)$ is defined as:

$$H_x \stackrel{\text{def}}{=} -\int_a^b f_x(x) \log f_x(x) dx \ . \tag{2}$$

The exact differential entropies pertaining to the Gaussian and Laplacean distributions write, respectively, as:

$$H_G = \log \sqrt{2\pi \mathrm{e} \sigma^2}, \ H_L = 1 - \log \frac{\lambda}{2} \ . \tag{3}$$

As these toy random variables possess known and closed-form distributions, we can generate as many samples as we please and investigate the behavior of histogram-based empirical-entropy estimation when the number of samples is large ($20,000$ samples) and in the case of incomplete data (200 samples). This investigation is accomplished by the analysis of the behavior of the estimated entropy $H_x[N_{bin}]$ as a function of the number of considered bins N_{bin}.

As approximation of $f_x(x)$ we consider its histogram-based estimate $h_x(n)$, with $2 \leq n \leq N_{bin}$ and N_{bin} being the number of bins that the interval $[a, b]$ is partitioned into. Let us denote by Δ_x the width of the partition $(b-a)/N_{bin}$. With this approximation for the PDF, the entropy may also be approximated as follows:

$$H_x[N_{bin}] \cong -\sum_{n=1}^{N_{bin}} h_x(n) \log h_x(n) \Delta_x \ . \tag{4}$$

Special attention should be paid to the fact that, for some n, the histogram value $h_x(n)$ could be null; in this case we extend the limit $\lim_{\epsilon \to 0} \epsilon \log \epsilon = 0$ and consider $0 \log 0 = 0$ in the summation (4).

The obtained results are shown in the Figure 1. The reported entropy curves clearly show that, in the presence of few samples, when the number of bins is very small the entropy estimate is very poor, due to the resulting poorly-structured shape of the estimating histogram. The more the number of bins increases, the more the estimated entropy approaches the true value. Note that – in general – having N_{bin} tend to infinity may not guarantee convergence to the true value: The estimation bias may tend to zero but the variance can increase. This is the familiar bias-variance trade-off problem known from e.g. statistical density estimation theory [39]. As a matter of fact, N_{bin} should be a tuning parameter that can be optimized for better estimation results, whose adjusting process is, however, extremely difficult unless toy problems are dealt with. In the presence of many data-samples, the estimated entropy stabilizes around the true value,

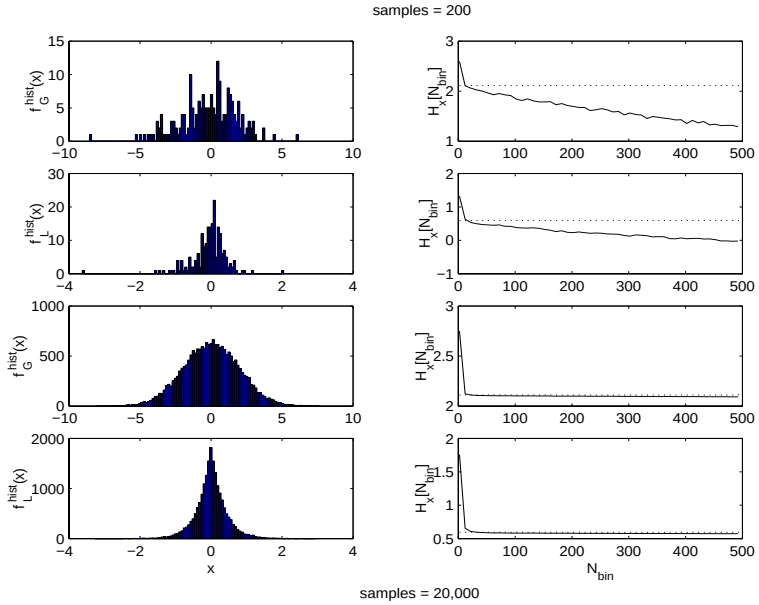

Figure 1: Numerical entropy estimation of Gaussian and Laplacean random variables as a function of the number of histogram-bins. Top four pictures: 200 available data-samples. Bottom four pictures: $20,000$ available data-samples. Left panels: Histograms of samples-values distribution. Right panels:

making the empirical estimation reliable enough, while in presence of limited amounts of data, the function $H_x[N_{bin}]$ changes for every value of the number of bins used for histogram estimation.

It is important to note that the limited-data-amount condition is highly probable in some practical applications, thus the poor reliability of the histogram-based entropy-estimate in presence of incomplete data may seriously flaw any statistical analysis of this kind.

In order to confirm the previous analysis using real-world data, as a second group of examples, we consider one data set drawn from each of the three data-classes utilized in the sequel of the paper, namely a electromagnetic-compatibility (EMC), a surface-roughness (SRM) and a composite-material analysis (PP) data-set (explained with details in section 3.1). The results of estimation are shown in the Figure 2.

These data-sets contain a limited amount of samples, and the numerical analysis confirms that, because of this intrinsic limitation, the empirical entropy estimate is completely unreliable, in that it results impossible to find out a 'stable' value.

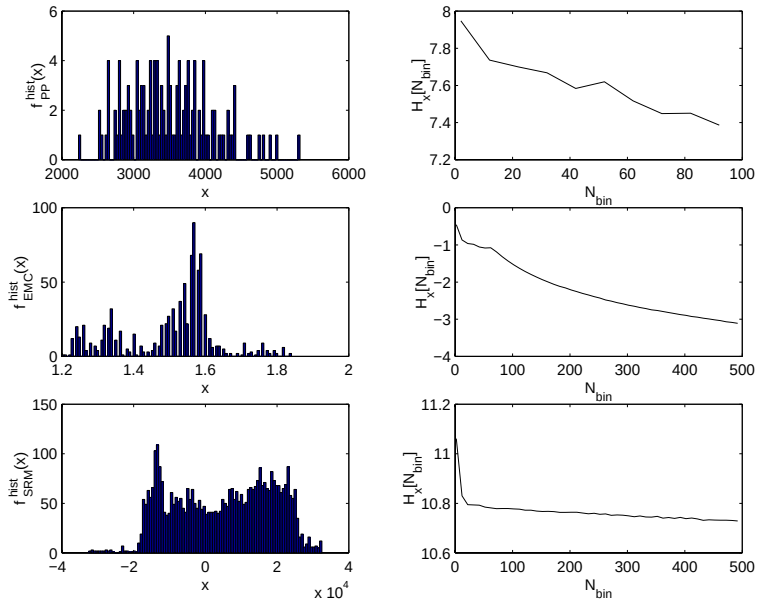

Figure 2: Numerical entropy estimation of real-world random variables as a function of the number of histogram-bins.

As an useful side-note to the last results, it might be important to observe that, even with a limited amount of data, for neural-system training purpose we might think to augment artificially the cardinality of the data by simply duplicating the available ones (this could be necessary if small learning stepsize values are required which imply many training iterations for the learning procedure to converge properly). It is worth noting that this operation does not change data's information content.

2.2 Neural PDF estimation by FAN model

The problem of estimating $f_x(x)$ is closely related to the estimation of the cumulative distribution function $F_x(x)$. In principle, estimating $f_x(x)$ or $F_x(x)$ or any other statistical function is completely equivalent, however, $F_x(x)$ is the only one among these which is a bounded saturating (sigmoidal) function, thus it appears to be naturally estimable by tunable neurons that classically provide saturating input-output transference[2].

A way of estimating $f_x(x)$ relies on the 'flattening' property of the function

[2]The equivalence holds if some minimal regularity conditions are met. Otherwise, estimating f_x and F_x is not completely equivalent, in fact, it is possible for F_x to exist and f_x not to exist.

$F_x(x)$: It is in fact well known [36] that warping the variable x by the function $F_x(x)$, that is computing $y = F_x(x) \in [0, 1]$, results in a uniformly distributed random variable or, equivalently, a variable with maximum entropy H_y. As a consequence, when $F_x(x)$ is approximated by a parametric function $y = \Psi(x; \mathbf{p})$ with parameters in $\mathbf{p}$, a maximum entropy estimator is obtained by maximizing H_y with respect to $\mathbf{p}$.

Let us denote by $f_y(y; \mathbf{p})$ the probability density function of the random variable y; it is important to note that it depends upon the value of the parameters of the parametric function. The differential entropy of the random variable y is defined by:

$$H_y(\mathbf{p}) \stackrel{\text{def}}{=} -\int_0^1 f_y(y; \mathbf{p}) \log f_y(y; \mathbf{p}) dy \ . \tag{5}$$

It relates to the differential entropy of the random variable x by means of the fundamental formula $f_y = f_x/|\psi|$, $\psi(x; \mathbf{p}) \stackrel{\text{def}}{=} \Psi'(x; \mathbf{p})$, provided that $\Psi(x; \cdot)$ is a monotonic function for every $x \in [a, b]$. Using that substitution in the entropy formula (5), yields:

$$H_y = -\int_a^b \frac{f_x}{|\psi|} \log \left(\frac{f_x}{|\psi|} \right) |\psi| dx = H_x + \int_a^b f_x \log |\psi| dx \ , \tag{6}$$

As mentioned, our aim is to maximize the entropy of the neuron's response, which is equivalent to maximizing the quantity $\Gamma_H \stackrel{\text{def}}{=} H_y - H_x$, referred to as *entropy gap* between the original and the warped random processes.

In the theoretical paper [12], the following input-output description for an adaptive-activation-function neuron (FAN) was assumed:

$$y = \Psi(x; c, \mathbf{p}) \stackrel{\text{def}}{=} \sigma[p(x; c, \mathbf{p})] \ , \ p(x; c, \mathbf{p}) \stackrel{\text{def}}{=} c + \int_0^x R(\sum_{n=0}^{N} p_n \xi^n) d\xi \ , \tag{7}$$

where $\sigma(\cdot)$ denotes a positive sigmoidal function ranging in $[0,\ 1]$ and $R(\cdot)$ is a positive integrable function. The learnable parameters are c and $\mathbf{p}$.

This representation has the following advantages:

- The function $\Psi(x; c, \mathbf{p})$ ranges in $[0, 1]$ and is always monotonically non-decreasing, thus it represents a valid cumulative function.
- The computation of the optimal value of $\mathbf{p}$ may be carried out as an unconstrained maximization problem, because the function $\Gamma_H(c, \mathbf{p})$ is always limited from above by $-H_x$.

After FAN neuron learning is completed, that is $c = c^*$ and $\mathbf{p} = \mathbf{p}^*$, the function ψ provides an approximation of f_x, namely $\hat{f}_x(x) = \psi(x, c^*, \mathbf{p}^*)$. The mentioned

structure provides valid PDF estimation, in fact it is readily verified that:

$$\psi(x;c,\mathbf{p}) = \sigma'[p(x;c,\mathbf{p})]R(\sum_{n=0}^{N} p_n x^n) \geq 0 \ . \tag{8}$$

In order to find a vector of parameters $\mathbf{p}$ that maximizes the entropy of the neuron response $y(t)$, a set of learning equations derived by the gradient steepest ascent method was considered, where it is understood that the parameters-space is endowed with the standard Euclidean metric. The optimal value of the response entropy, corresponding to a completely-flat distribution of y in $[0,1]$, is $H_y^* = 0$.

As the partial derivative of the entropy coincides with the partial derivative of the entropy-gap, that is easier to compute, the proposed neuron may learn through a maximum-entropy-gap principle. The entropy gap in this case is equal to:

$$\Gamma_H(c,\mathbf{p}) = \int_a^b f_x(x) \log|\psi(x;c,\mathbf{p})| dx \ , \tag{9}$$

therefore the computation of the partial derivatives of $\Gamma_H(c,\mathbf{p})$ with respect to c and $\mathbf{p}$ requires computing $\frac{1}{\psi}\frac{\partial\psi}{\partial c}$ and $\frac{1}{\psi}\frac{\partial\psi}{\partial \mathbf{p}}$.

We considered the setting $\sigma(u) = 0.5 + 0.5\mathrm{erf}(u)$ and $R(u) = \frac{1}{2}u^2$; in this case, the relevant quantities for neuron's learning write:

$$p(x;c,\mathbf{p}) = c + \frac{1}{2}\sum_{n=0}^{N}\sum_{m=0}^{N} \frac{p_n p_m}{n+m+1} x^{n+m+1} \ , \tag{10}$$

$$\frac{1}{\psi}\frac{\partial\psi}{\partial c} = -2p(x;c,\mathbf{p}) \ , \tag{11}$$

$$\frac{1}{\psi}\frac{\partial\psi}{\partial p_r} = -2p(x;c,\mathbf{p})\sum_{n=0}^{N}\frac{p_n x^{n+r+1}}{n+r+1} + \frac{2x^r}{\sum_{n=0}^{N} p_n x^n} \ , \tag{12}$$

while the approximated PDF has expression[3] (after learning $c = c^*$ and $\mathbf{p} = \mathbf{p}^*$):

$$\hat{f}_x^{FAN}(x) = \frac{1}{2\sqrt{\pi}} e^{-p^2(x;c^*,\mathbf{p}^*)} (\sum_{n=0}^{N} p_n^* x^n)^2 \ . \tag{13}$$

It is worth noting that the expression involved in equations (10)-(12) may be efficiently implemented in a high-level programming language such as MATLAB, as shown in the Appendix of paper [12], by the help of some useful numerical structure such as the Hilbert matrix and component-wise (Hadamard) operation mode.

It is also worth noting that a discrete-time *stochastic* parameter estimation technique is to be employed here in order to progressively adapt and refine the

[3]Note that the expression given in [12] lacks of a factor $\frac{1}{2}$.

values of neuron's polynomial part. In particular, the learning equations read:

$$\Delta c = -\eta \frac{1}{\psi} \frac{\partial \psi}{\partial c} \ , \ \Delta p_r = -\eta \frac{1}{\psi} \frac{\partial \psi}{\partial p_r} \ , \ r = 1, \ 2, \ ..., \ N \ , \tag{14}$$

where $\eta > 0$ is the learning step-size.

3 Experiments on real-world data

We consider data coming from composite materials analysis, electromagnetic compatibility measures and surface roughness measures and try to estimate their statistical distributions through the FAN neuron. In order to objectively assess the behavior of the selected models in density estimation, we defined and discuss two performance indices.

3.1 Data description

The specification of the data is the subject of the present subsection. Some comments on data format are also given.

3.1.1 Reinforced composite data

The first data-sets come from measures of the vegetal fibers' diameter and length values in reinforced polypropylene matrices.

Short conventional fibers (glass, nylon, etc.) have been extensively used over the last years as reinforcements of polymeric matrices. They are incorporated into plastics with the main aim of improving its features and of reducing the cost of the final products. The growing interest in the use of natural vegetable fibers as reinforcement of polymer-based composites is mainly due to their renewable origin, relative high specific strength, light weight and low price [1, 14].

In contrast with synthetic fibers, whose properties can be easily and univocally determined, natural fibers are characterized by a large dispersion of their characteristics: Such features make it advisable to employ a statistical approach to define their properties and those of the related composites.

The vegetal fibers are added to the polypropylene to reinforce its structure: The fibers have initially constant length and diameter, but after production they spread out and find within the composite with different length and diameter values [28]. The micro- and macro-mechanical features of the composite slab depend on the statistical distribution of these parameters [33] and may be evaluated e.g. through the Halpin-Tsai equations [28]. An interesting macro-mechanical parameter is, for instance, the elasticity of the slab (Young module

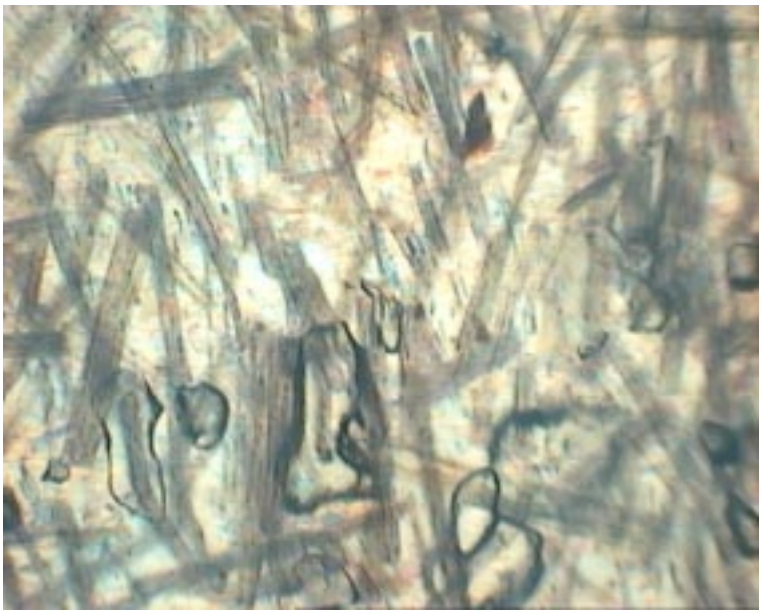

Figure 3: Picture of a polypropylene matrix with several jute fibers.

Y). The elasticity is a random parameter itself, whose distribution depends on the distribution of the geometric parameters of the fibers embedded in the composite, and the mechanical properties of the slab may be given in terms of the distribution of the random variable Y.

A typical matrix of polypropylene composite is depicted in the Figure 3, where several randomly oriented discontinuous jute fibers are visible. The bubbles are due to the process of high-temperature fusion of the polypropylene necessary to embed the fibers in the matrix and to the subsequent cooling.

The available data in this class consist of three data sets pertaining to Young module measures of three kind of polypropylene composite: PP4LF with 40% of volume of flax fibers (126 samples), PP4SF with 40% of volume of sisal fibers (129 samples) and PP4YF with 40% of volume of jute fibers (130 samples).

3.1.2 Electromagnetic compatibility data

The second data-sets come from measures of environmental electromagnetic compatibility. The available data express the power-level of electromagnetic field measured at a certain location at different times. The statistical characterization of these levels is useful in applications such as electromagnetic pollution monitoring [19, 20] and telecommunication systems design/optimization [27].

Electromagnetic radiation is nowadays recognized as a pollutant of the en-

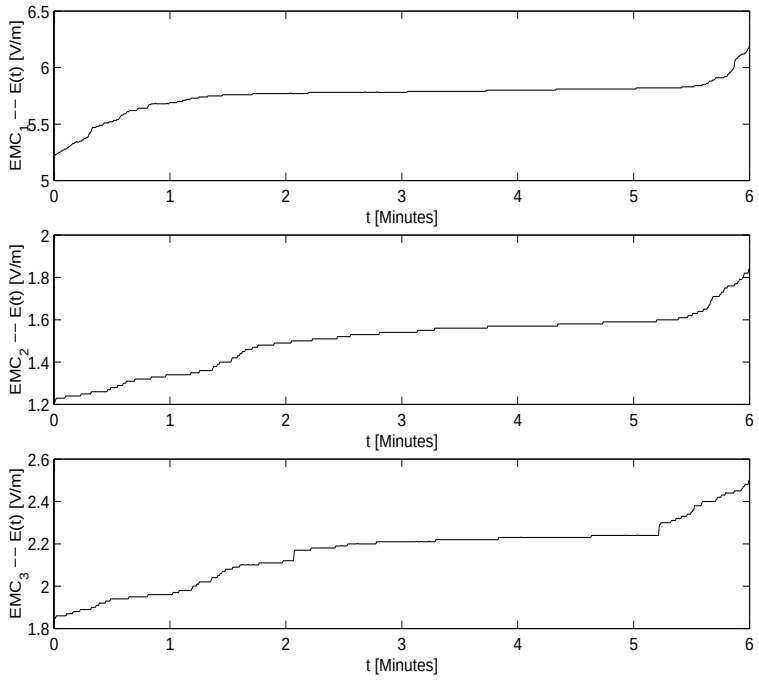

Figure 4: Three electrical field intensity measures versus time.

vironment. Electromagnetic pollution comes from e.g. power lines, electrical blankets, cellular phones, televisions, computers and microwave ovens. For instance, mobile-phones stations emit microwave frequencies, while computer monitors' cathode ray tubes emit waves ranging from extremely low frequencies to X-rays. Although electric fields can be shielded by conducting materials, magnetic fields can penetrate almost any material, including the human body. Reliable measurement systems and data-processing techniques are thus necessary for electromagnetic pollution monitoring [15, 19, 20, 25].

Usually the applications that make use of electromagnetic measures exploit empirical mean and variance parameters extracted from the available data [19, 20, 41], which are however completely meaningful only when the distribution of the measured values is Gaussian. In general, evidence shows that this condition does not hold in practice, thus PDF estimation may reveal important additional information.

The available data in this class pertain to measures taken in three different places in the city of L'Aquila (Italy) and consist of three data sets: EMC_1 (886 samples), EMC_2 (894 samples) and EMC_3 (872 samples). The available discrete-time measure-sequences are depicted versus time in the Figure 4. It is

worth noting that the EMC_1 data present the largest values of measured fields, because they have been taken very close to an emission station (in the proximity of an array of cellular-phone broadcasting antennas).

3.1.3 Surface roughness data

The third data-sets come from measures of surface roughness. The aim of rough surface characterization breaks into two sub-tasks: Surface description and surface modeling. A surface may be described with various levels of detail: A single characteristic may be sufficient for some applications even if for texture discrimination, for instance, a much more specific description is necessary, such as the statistical characterization. In fact, statistical techniques have proved to be helpful for establishing correlation between a surface's features and its intended functions [21].

An important statistical function in rough surface analysis is termed 'Amplitude Distribution Function' (ADF), that is the probability density function $f_z(z)$ of the quota z of the surface, with respect to a reference level, collected over a grid of measurement points on a specimen. The 'Bearing Ratio Curve' is also considered as an important descriptive function: It is the cumulative distribution function $F_z(z)$ associated to the density $f_z(z)$. The availability of a good estimate of the probability density function of the quota allows a reliable estimation of the skewness of the distribution. The skewness is a simple measure of the asymmetry of the ADF, or, equivalently, it measures the asymmetry of the variation of a profile about its mean line. The skewness parameter correlates with e.g. porosity, load carrying capability and other characteristics of surfaces produced by processes other than conventional machining. Surfaces with a positive skewness, such as turned surfaces, exhibit high spikes that protrude above a flatter average. Surfaces with negative skewness, such as porous surfaces, exhibit deep valleys in a smoother plateau. More random (e.g. ground) surfaces have a quasi-zero skewness. Also, the kurtosis of the distribution is another ADF-shape parameter that is worth considering. The kurtosis relates to the uniformity of the ADF or, equivalently, to the spikiness of the profile. Sparse distributions, exhibiting large negative kurtosis values, relate to a highly irregular surface, while peaked distributions around the reference value relate to a more regular surface.

There is a significant amount of information encapsulated in the shape of the bearing ratio curve for a surface: Efforts in surface finishing rely seriously on accurate determination of bearing ratio curve [22, 32].

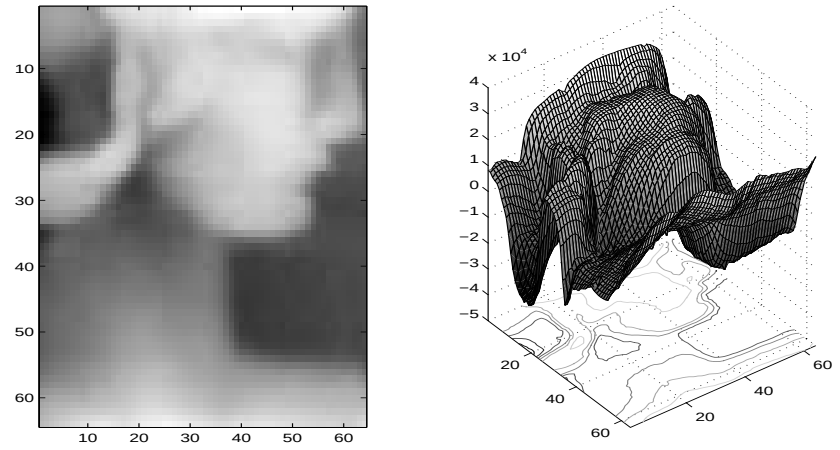

Figure 5: A coin examined with a surface-roughness probe. Left: Image of roughness. Right: Three-dimensional representation of the data along with level-contour.

A meaningful example of surface roughness measure, where the subject of the measures is clearly recognizable, is depicted in the Figure 5. The data pertain to a 64×64 grid.

The available data in this class consist of two data sets: SRM_1 ($2,500$ samples) and SRM_2 ($4,096$ samples).

3.1.4 On data format

As a general comment on the data, it is important to recall that, to the end of PDF estimation, the ordering of the samples in the data batches is completely meaningless. In fact, the PDF subsumes the distribution of the values of a data-set, not their ordering structure like the natural time-odering in the time domain appearing e.g. in the electromagnetic compatibility data-sets.

This appears as an important consideration when the data come after possibly unknown tampering that occurs during data storage. In the present work, this is the case for the reinforced composite data, that were ordered in ascending fashion by the operator who effected the measurements and the surface roughness data-set SRM_1, that initially appeared as a 50×50 array of numbers and that was stored into a 500×10 array (for some unknown reason).

A similar observation applies to the format of the data's support. The EMC and PP4(L-S-Y)F are one-dimensional data with one-dimensional support, while the SRM data are one-dimensional data with bi-dimensional support. To the aim of probability density function estimation, the dimension of the support is

irrelevant, provided that the dimension of the data (i.e. the number of random variables whose statistical description is sought for) is equal to one.

3.2 Estimation quality indices: Description and discussion

In the present section, two estimation quality indices are described and discussed, both analytically and numerically.

3.2.1 Indices description

In order to objectively measure the discrepancy between the true probability density function $f_x(x)$ and the estimate $\hat{f}_x(x)$ provided by the learnt neuron, we defined two indices.

The first index is a square probability distance (SPD), defined as:

$$SPD^* \stackrel{\text{def}}{=} \int_a^b [f_x(x) - \hat{f}_x(x)]^2 dx \ . \tag{15}$$

The second index is the pseudo-distance provided by the Kullback-Leibler divergence (KLD):

$$KLD^* \stackrel{\text{def}}{=} \int_a^b f_x(x) \log \frac{f_x(x)}{\hat{f}_x(x)} dx \ . \tag{16}$$

In the present experimental setting the true PDFs are not known in closed form, thus the above closed-form indices cannot be computed. We should thus resort to numerical approximation, by using again the settings introduced in section 2.1. Note that the partition of the interval $[a, b]$ into bins, as introduced in section 2.1, gives rise to discrete values of the variable x, indicated here with x_n, thus the estimated probability density function may be discretized accordingly by computing the values $\hat{f}_x(n) \stackrel{\text{def}}{=} \hat{f}_x(x_n)$. The quantities (15) and (16) may thus be approximated by:

$$SPD \stackrel{\text{def}}{=} \sum_{n=1}^{N_{bin}} [h_x(n) - \hat{f}_x(n)]^2 \Delta_x \ , \tag{17}$$

$$KLD \stackrel{\text{def}}{=} \sum_{n=1}^{N_{bin}} h_x(n) [\log h_x(n) - \log \hat{f}_x(n)] \Delta_x \ . \tag{18}$$

As already mentioned in section 2.1, for some n the value $h_x(n)$ could be null and it is therefore convenient to extend the limit $\lim_{\epsilon \to 0} \epsilon \log \epsilon = 0$ with $0 \log 0 = 0$ in the summation (18).

3.2.2 Discussion of SPD and KLD indices and entropic criterion

In the present context the SPD and KLD indices have been defined and exploited for density-function-estimation quality analysis. The two indices are sensitive to different features of the approximated PDF, but it is interesting to note that the SPD and KLD indices are somewhat related, as explained in [17]. Let us denote by $f_x(x)$ the true density function and by $\hat{f}_x(x;\mathbf{p})$ the estimated function of parameters in $\mathbf{p}$. Let us also define the functional compactness of a probability density function as its L_2 norm:

$$\|f\|_2^2 \stackrel{\text{def}}{=} \int_{\mathbb{R}} f^2(x)dx \ .$$

From definition (15) we thus have:

$$SPD^*(\mathbf{p}) = \|f\|_2^2 + \|\hat{f}_x(\mathbf{p})\|_2^2 - 2E_x[\hat{f}_x(x;\mathbf{p})] \ ,$$

therefore the optimal parameter-vector is sought for in order to find:

$$\text{argmin}_{\mathbf{p}} SPD^*(\mathbf{p}) = \text{argmin}_{\mathbf{p}}\{\|\hat{f}_x(\mathbf{p})\|_2^2 - 2E_x[\hat{f}_x(x;\mathbf{p})]\} \ , \tag{19}$$

On the other hand, from definition (16) we have:

$$KLD^*(\mathbf{p}) = -H_x - E_x[\log \hat{f}_x(x;\mathbf{p})] \ ,$$

where H_x denotes again the differential entropy of the random variable x. In this case the optimal parameter-vector value is sought for in order to find:

$$\text{argmin}_{\mathbf{p}} KLD^*(\mathbf{p}) = \text{argmin}_{\mathbf{p}}\{-E_x[\log \hat{f}_x(x;\mathbf{p})]\} \ . \tag{20}$$

Now we may invoke Jensen's inequality $E_x[\log \hat{f}_x(x;\mathbf{p})] \geq \log E_x[\hat{f}_x(x;\mathbf{p})]$ in (20) to show that:

$$\text{min}_{\mathbf{p}} KLD^*(\mathbf{p}) \leq \text{min}_{\mathbf{p}}\{-\log E_x[\hat{f}_x(x;\mathbf{p})]\} \ . \tag{21}$$

From (19) and (21) it readily follows that the minimization of SPD is equivalent to the minimization of an upper-bound of the KLD under maximal compactness of the estimated density.

Also, it is worth mentioning that another PDF-estimation performance index that might be considered is the neuron's response differential entropy H_y: From the basic theory recalled in section 2.2 it is in fact known that the optimal value of the neuron response entropy is 0, therefore the quantity $|H_y|$ itself would behave as a meaningful error function. However, the exact computation of H_y or even a reliable estimation of the neuron-response differential entropy

is inherently impossible: Equation (6) shows that the entropy under discussion computes as the sum of the stimulus entropy and the entropy gap; even if the latter can be estimated numerically, we already know the entropy of the input H_x cannot be reliably estimated. These notes explain why, ultimately, the learning cost function H_y cannot be assumed as a valid estimation performance index.

3.2.3 Numerical analysis of SPD and KLD indices suitability

The conclusions of the above discussion are that the KLD* and SPD* indices take into account different aspects of PDF estimation quality, and that the KLD and SPD are only numerical approximations to the exact indices.

In order to illustrate these two faces of the index suitability question, we present two numerical experiments performed on the same Gaussian data set formed by 5,000 samples: As the true PDF in this case is known, we may compute both the SPD/KLD and SPD*/KLD* pairs and compare their values.

The approximation results shown in the Figure 6 pertain to a quite badly-converged neuron (that was forced to learn very slowly by selecting $\eta = 0.00005$). In this experiment, the slow convergence is testified by the slow growing of the entropy-gap value.

We may observe from the behavior of the local SPD and KLD indices (that is, the un-integrated square difference of equation (15) and the weighted-log-ratio of equation (16), respectively) that they indeed emphasize different aspects of the dissimilarities among the true and estimated PDFs.

The values of the computed indices, in this experiments, are: SPD = 0.3274, SPD* = 0.3217, KLD = 2.5431, KLD* = 2.4663. In this case, the (large) values of the indices approximated through the histogram reference and the exact values of the indices are close to each other.

The approximation results shown in the Figure 7 pertain instead to a well-converged neuron (FAN parameters: $N = 1$ and $\eta = 0.0005$).

Again, we note from the behavior of the local SPD and KLD indices that they take into account different aspects of the dissimilarities among the true and estimated PDFs.

The values of the computed indices, in this experiments, are: SPD = 2.9426×10^{-3}, SPD* = 1.7877×10^{-3}, KLD = 1.4259×10^{-2}, KLD* = 0.8684×10^{-2}. It is interesting to note that, in this case, namely when the values of the indices are quite low, that testifies a good convergence of the neural PDF estimation algorithm, the values of the indices approximated through the histogram reference

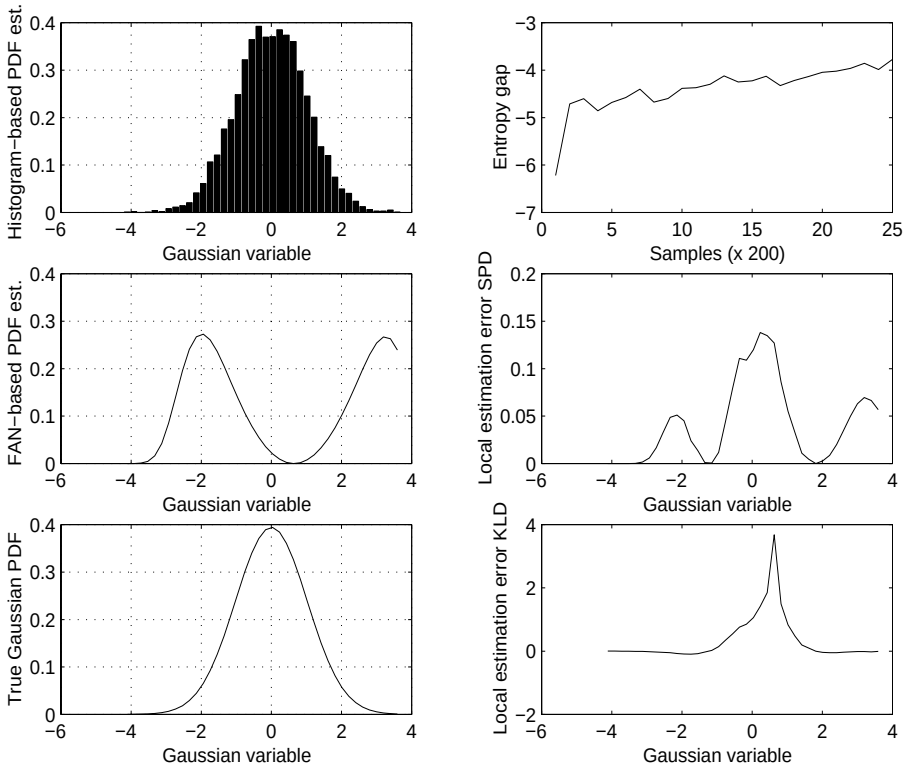

Figure 6: Estimation results on a Gaussian data set, obtained with $N = 1$ and $\eta = 0.00005$. Left-hand column: Histogram-based PDF estimate (40 bins), FAN-based estimate and true PDF. Right-hand column: Entropy gap during learning, local SPD and local KLD.

and the exact values of the indices differ of a factor of about 1/2. Nevertheless, by comparing the indices values observed in the first and the second experiment, it is readily found that lower values of the approximated estimation errors correspond to better estimation results.

As a further investigation on the same subject, we considered the average behavior of the four indices over 30 independent realizations of the Gaussian data-set but with a variable number of generated samples (here denote by N_s). The histograms are always computed over 40 bins. In this case, the quality of the PDF estimation through the histogram method changes (see e.g. again the Figure 1) and it is possible to evaluate the quality of the indices based on such estimation versus the true indices. The results are reported in the Table 1. These results confirm that – on average – the KLD and SPD indices 'follow' quite closely the corresponding exact indices, especially when the neuron is still

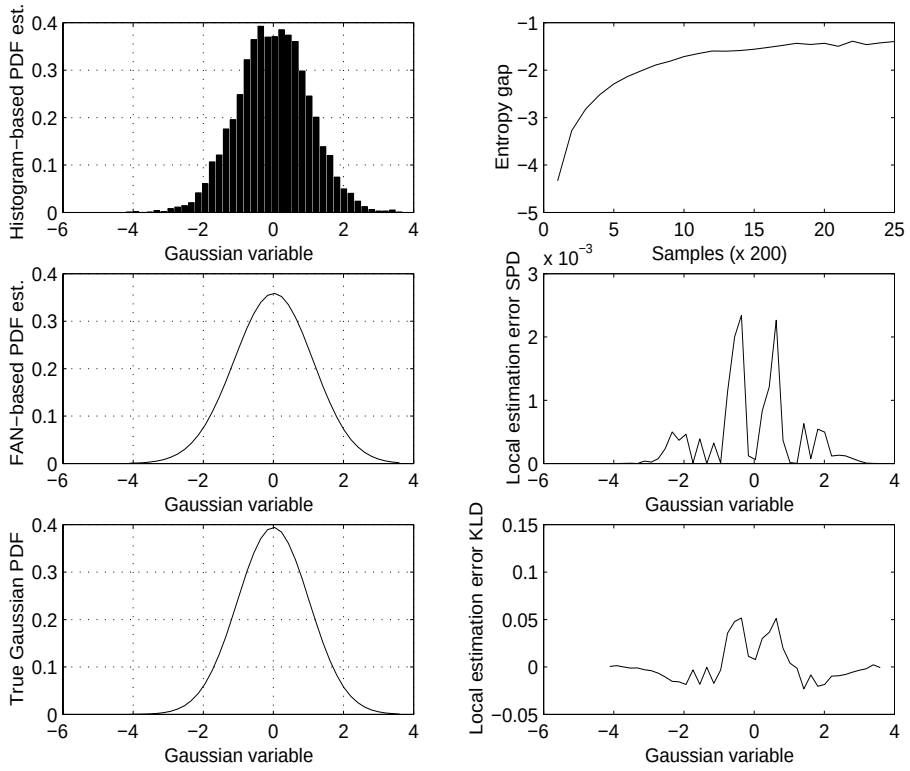

Figure 7: Estimation results on a Gaussian data set, obtained with $N = 1$ and $\eta = 0.0005$. Left-hand column: Histogram-based PDF estimate (40 bins), FAN-based estimate and true PDF. Right-hand column: Entropy gap during learning, local SPD and local KLD.

far from converging to the proper configuration. In any case, it seems that the approximated indices tend to over-estimate the real error.

From the above numerical analysis, we may conclude that the choice of the SPD and KLD indices is meaningful for the FAN-model comparison task.

3.3 Experimental results and discussion

The FAN neuron PDF estimation ability has been tested on the described data sets. It is important to notice that in order to better train the neuron we augmented the cardinality of the data by simply duplicating the available ones, thus the information content and the histograms are not modified, while the estimated probability density function provided by the trained neurons is better specified.

Index	$N_s = 200$	$N_s = 500$	$N_s = 1,000$	$N_s = 5,000$	$N_s = 10,000$
SPD_{ave}	0.2573	0.2060	0.1377	0.0505	9.90×10^{-3}
SPD^*_{ave}	0.2231	0.1930	0.1331	0.0496	9.50×10^{-3}
KLD_{ave}	2.8055	1.7575	0.8941	0.2736	0.0647
KLD^*_{ave}	2.7002	1.6213	0.8652	0.2630	0.0597

Table 1: Average behavior of the four indices over 30 independent realizations of the Gaussian data-set with a variable number of generated samples N_s. These results pertain to the structural parameters values $N = 1$ and $\eta = 0.0001$.

The results are reported as the values of the indices averaged over 20 independent trials – obtained by making the neuron starting from randomly-selected initial parameter values – along with the standard deviations of the achieved indices values; thus the values SPD_{ave} and SPD_{std} as well as KLD_{ave} and KLD_{std} are reported in tabular format.

Both the SPD and the KLD indices should be interpreted as estimation error measures and the more their mean value SPD_{ave} and KLD_{ave} in the experiments is small, the more the selected neuron-topology is suitable for the estimation task. The standard deviations SPD_{std} and KLD_{std} of these indices measure the stability of topology's internal state learnt: The more the standard deviations are small, the more the neuron final internal state is independent of the learning initial conditions. The 'NC' value reported in the following tables means that at least in one of the twenty trials the algorithm did not reach convergence.

In the experiments, the neuron has been tested with four different values of the learning step-size η and four different values of the polynomial-order-indicator N.

Table 2 refers to the results obtained with two of the three data-sets of the composite material category that gave rise to the best and the worst approximation results. First, it is important to note that the values of the two indices SPD and KLD increase in the same way as the learning step-size η decreases for every polynomial-order-indicator value N. Thus we can note that, for $N = 1, 2$, the best value of the learning step-size, according to the results reported on the Table 2, is $\eta = 0.001$. In the same way, we can also observe that increasing further the polynomial-order-indicator value increases the possibility that the algorithm does not reach convergence, as shown by the presence of the 'NC' value on the Table.

From the above results, we may employ the PDF estimation algorithm to

DATA	N	η	SPD_{ave}	SPD_{std}	KLD_{ave}	KLD_{std}
PP4LF	1	0.001	0.0209	4.78×10^{-12}	0.0888	2.84×10^{-10}
		0.0001	0.1118	0.1358	0.6062	0.7456
		0.00001	0.1920	0.1237	1.2618	0.8165
		0.000001	0.2851	0.0049	3.8557	0.7364
	2	0.001	0.0201	2.61×10^{-10}	0.0835	6.46×10^{-10}
		0.0001	0.0227	6.09×10^{-4}	0.0860	0.0018
		0.00001	0.2607	0.0362	1.0452	0.5375
		0.000001	0.3266	0.0167	2.7395	0.4732
	3	0.001	NC	NC	NC	NC
		0.0001	0.3780	0.0265	0.1856	0.1908
		0.00001	0.2684	0.0959	0.8906	0.4767
		0.000001	0.4326	0.0551	3.3661	0.4978
	4	0.001	NC	NC	NC	NC
		0.0001	NC	NC	NC	NC
		0.00001	0.3467	0.1234	1.0925	0.4610
		0.000001	0.4305	0.0412	2.5711	0.1717
PP4YF	1	0.001	0.0349	1.27×10^{-11}	0.1452	9.43×10^{-12}
		0.0001	0.0363	3.17×10^{-4}	0.1470	5.95×10^{-4}
		0.00001	1.1791	5.9×10^{-4}	1.0125	0.0072
		0.000001	0.3147	0.0187	3.4205	0.8221
	2	0.001	0.0379	1.68×10^{-9}	0.1309	1.19×10^{-10}
		0.0001	0.0415	1.62×10^{-4}	0.1373	2.18×10^{-4}
		0.00001	0.2521	0.0300	0.9933	0.2295
		0.000001	0.3417	0.0191	3.5899	0.9473
	3	0.001	NC	NC	NC	NC
		0.0001	0.0718	0.0522	0.2259	0.1824
		0.00001	0.2692	0.0899	1.2180	0.7406
		0.000001	0.4424	0.0364	3.6866	0.6256
	4	0.001	NC	NC	NC	NC
		0.0001	NC	NC	NC	NC
		0.00001	0.2382	0.0257	0.7571	0.1007
		0.000001	0.4319	0.0323	2.8827	0.4378

Table 2: Numerical results of statistical characterization of data pertaining to the composite material category PP4LF and PP4YF.

obtain an estimate of the distribution of the values of the polymer composite data. Figure 8 shows the results obtained with the best values found in the Table 2 for the PP4LF data, namely $\eta = 0.001$ and $N = 2$.

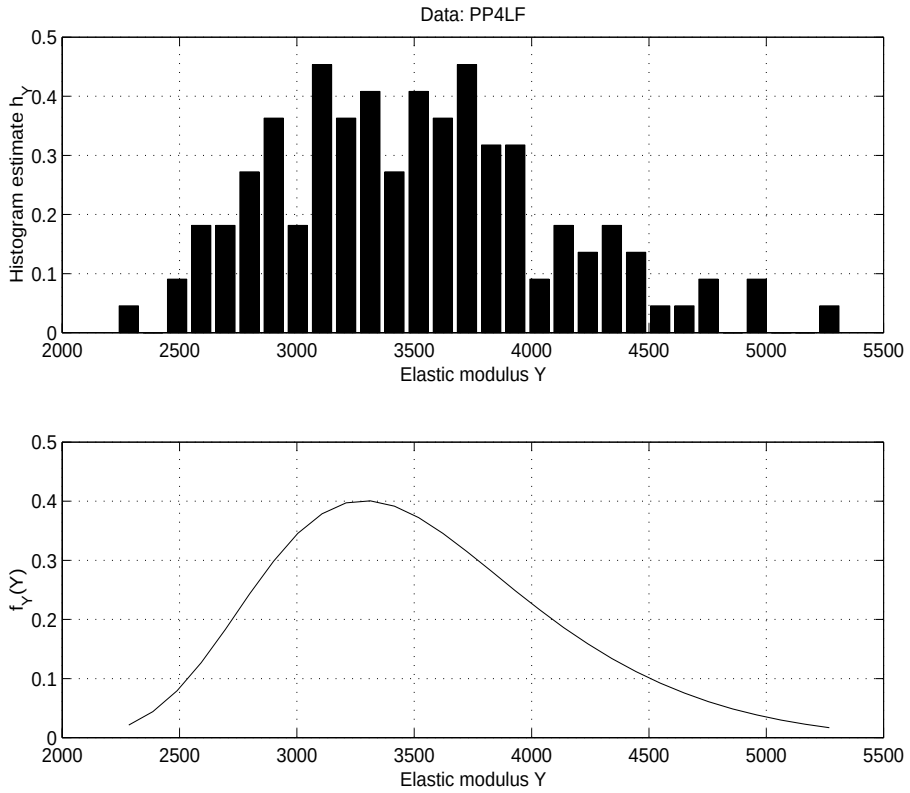

Figure 8: Estimation results on PP4LF data. Top: Histogram-based PDF estimate. Bottom: FAN-based PDF estimate.

Table 3 refers instead to the results obtained with two of the three data-sets of the electromagnetic compatibility category that gave rise to the best and the worst approximation results.

An estimate of the distribution of the values of the electromagnetic compatibility data is reported in the Figure 9 which shows the results obtained with the best values found in the Table 3 for the CEM_2 data, namely $\eta = 0.0001$ and $N = 1$.

Table 4 refers to the results obtained with the two data-sets of the surface-roughness-measure category.

As a last example of the obtained results, an estimate of the distribution of the values of the surface roughness data is reported in the Figure 10 which pertains to the best values found in the Table 3 for the SRM_1 data, namely

DATA	N	η	SPD_{ave}	SPD_{std}	KLD_{ave}	KLD_{std}
EMC_2	1	0.001	0.2155	3.51×10^{-9}	0.8810	1.05×10^{-8}
		0.0001	0.1876	1.24×10^{-4}	0.3636	2.26×10^{-5}
		0.00001	0.3239	0.0011	1.1993	0.0089
		0.000001	0.4623	0.0026	3.4951	0.2330
	2	0.001	0.3066	1.19×10^{-7}	1.3935	4.15×10^{-7}
		0.0001	0.2092	6.11×10^{-4}	0.3875	8.7×10^{-4}
		0.00001	0.4314	0.0047	0.9922	0.0120
		0.000001	0.4725	0.0139	2.8317	0.1117
	3	0.001	NC	NC	NC	NC
		0.0001	0.1758	0.0018	0.3449	0.0026
		0.00001	0.4505	0.1074	1.0073	0.2245
		0.000001	0.5569	0.0894	2.9705	0.1421
	4	0.001	NC	NC	NC	NC
		0.0001	0.1773	0.0029	0.5153	0.0056
		0.00001	0.4878	0.0045	1.0525	0.0106
		0.000001	0.8088	0.0010	2.6247	0.1033
EMC_3	1	0.001	0.5706	4.27×10^{-8}	1.1418	1.17×10^{-7}
		0.0001	0.5464	1.79×10^{-4}	0.6754	7.18×10^{-5}
		0.00001	0.7192	0.0656	1.7497	0.3455
		0.000001	0.8170	6.71×10^{-4}	4.0836	0.3356
	2	0.001	0.6605	4.78×10^{-7}	1.6718	5.06×10^{-7}
		0.0001	0.5769	8.28×10^{-4}	0.6933	1.10×10^{-3}
		0.00001	0.7707	0.0059	1.2531	0.0124
		0.000001	0.8049	0.0065	3.3872	0.4859
	3	0.001	NC	NC	NC	NC
		0.0001	0.5510	0.0019	0.6671	2.80×10^{-3}
		0.00001	0.8252	0.0927	1.4561	0.3361
		0.000001	0.9801	0.4250	3.4491	0.2405
	4	0.001	NC	NC	NC	NC
		0.0001	0.5465	0.0030	0.7665	0.0038
		0.00001	0.8477	0.0085	1.3395	0.014
		0.000001	1.1232	0.0826	2.9010	0.1699

Table 3: Numerical results of statistical characterization of data pertaining to the electromagnetic compatibility categories EMC_2 and EMC_3.

DATA	N	η	SPD_{ave}	SPD_{std}	KLD_{ave}	KLD_{std}
SRM_1	1	0.001	0.0170	7.97×10^{-4}	0.0775	0.0015
		0.0001	0.0179	2.79×10^{-4}	0.0765	2.77×10^{-4}
		0.00001	0.2060	0.1764	1.5891	1.6852
		0.000001	0.2650	0.0177	3.0857	0.6635
	2	0.001	0.0019	0.0014	0.0615	0.0022
		0.0001	0.0301	0.0025	0.0708	0.0028
		0.00001	0.1964	0.0538	0.5763	0.1457
		0.000001	0.2610	0.0053	2.6530	0.3514
	3	0.001	NC	NC	NC	NC
		0.0001	0.0301	0.0062	0.0744	0.0111
		0.00001	0.2632	0.1338	0.7794	0.4251
		0.000001	0.3138	0.0654	2.8267	0.5488
	4	0.001	NC	NC	NC	NC
		0.0001	0.0486	0.0529	0.1278	0.1669
		0.00001	0.3102	0.0066	0.7071	0.0098
		0.000001	0.5236	0.0877	2.3151	0.4304
SRM_2	1	0.001	0.0616	0.0150	0.1729	0.0463
		0.0001	0.0595	0.0131	0.2080	0.1520
		0.00001	0.1590	0.0493	1.0639	0.3029
		0.000001	0.2606	0.0611	3.1843	1.0542
	2	0.001	0.0153	0.0019	0.0690	0.0037
		0.0001	0.0133	0.0012	0.0763	0.0014
		0.00001	0.2248	0.0039	0.6877	0.0147
		0.000001	0.2820	0.0080	2.6688	0.0777
	3	0.001	NC	NC	NC	NC
		0.0001	0.0289	0.0293	0.1228	0.1603
		0.00001	0.2252	0.0197	0.8751	0.4571
		0.000001	0.3299	0.0978	2.6692	0.2603
	4	0.001	NC	NC	NC	NC
		0.0001	0.0166	0.0038	0.0556	0.0086
		0.00001	0.3286	0.0133	0.7826	0.032
		0.000001	0.6315	0.0812	2.4097	0.2872

Table 4: Numerical results of statistical characterization of data pertaining to the surface-roughness-measure categories SRM_1 and SRM_2.

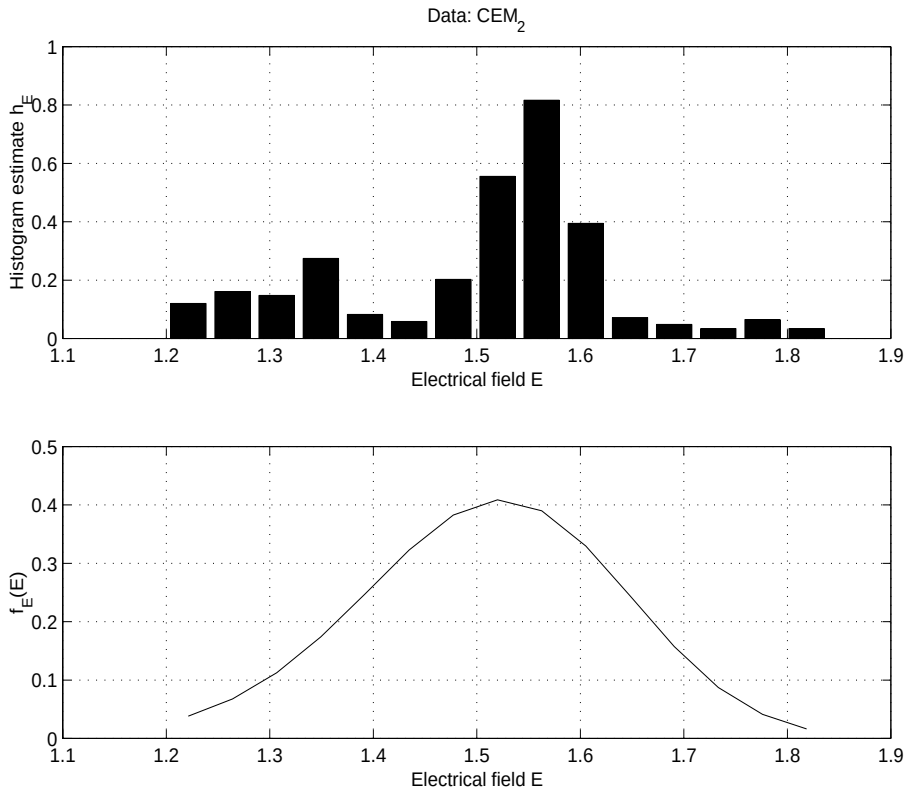

Figure 9: Estimation results on CEM_2 data. Top: Histogram-based PDF estimate. Bottom: FAN-based PDF estimate.

$\eta = 0.001$ and $N = 2$.

If we compare the results in the Tables 3 and 4, we can note that, according to the results in Table 2, the best values of the four indices SPD_{ave}, SPD_{std}, KLD_{ave} and KLD_{std} are those we obtained for $N = 1, 2$ and $\eta = 0.001$ and $\eta = 0.0001$. We conclude that for a limited amount of data those values are the optimal polynomial degrees and learning step-sizes. In fact, for greater polynomial degrees we have observed the tendency of the neuron not to attain convergence while for smaller learning step-sizes the values of the four indices tend to worsen.

4 Conclusion

The aim of the present contribution was to propose the estimation of the probability density function of electrical- and mechanical-type data. A FAN neuron model, recalled from a previous research paper, was considered and the choice

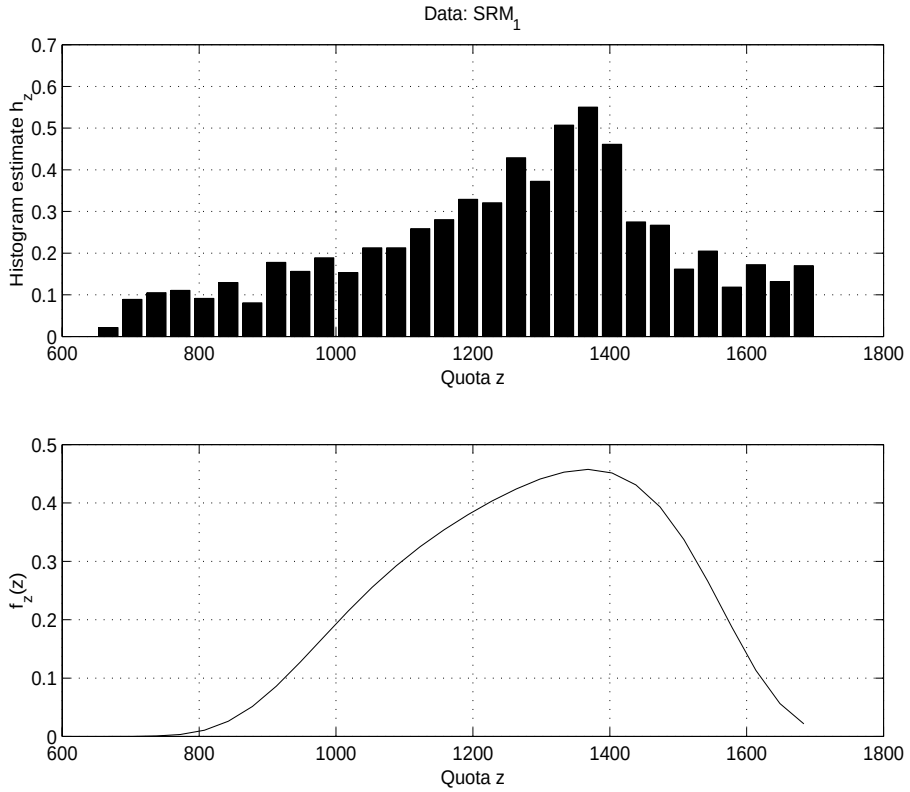

Figure 10: Estimation results on SRM_1 data. Top: Histogram-based PDF estimate. Bottom: FAN-based PDF estimate.

of its structural complexity and learning parameter in relation to its probability density function estimation ability, was addressed.

In particular, the model selection was performed by the comparison, through two objective measures, of the PDF estimation capability of the different models on a large series of real-world data about electromagnetic compatibility, surface roughness analysis and mechanical elasticity measures on polymer composites.

The presented investigation allowed us to acquire incremental knowledge on practical estimation of the probability density function of real-world incomplete data.

In particular, a close examination of the obtained numerical result reveals that the best parameters values for the considered FAN models are low polynomial order-indicator N and relatively high values of learning step-size; these parameters-values make the learning ability of the FAN model good and relatively independent of the initial state of learning.

Given the large number of considered data-sets and their high variability,

which concerns the distributions of the values and the number of samples among the different data-batches, is can be figured that the above conclusion are quite general and the same good parameter values would be good for every data-set.

5 Acknowledgment

We wish to gratefully thank Dr. Antonio Faba (University of Perugia, Italy) for the data on electromagnetic compatibility, Dr. Nicola Senin (University of Parma, Italy) for the data on surface roughness and the colleagues of the Polymer Technology Group at Perugia University for the data on reinforced composites. This research work has been carried out when the student RR followed the course of "Modellistica Elettrica dei Materiali" at the Faculty of Engineering of the Perugia University.

References

[1] J. Biagiotti, S. Fiori, L. Torre, M.A. López-Manchado and J.M. Kenny, *Mechanical Properties of Polypropylene Matrix Composites Reinforced with Natural Fibers: A Statistical Approach*, Polymer Composites. Accepted for publication

[2] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995

[3] P.W. Buchen and M. Kelly, *The maximum entropy distribution of an asset inferred from option prices*, Journal of Finance and Quantitative Analysis, Vol. 31, pp. 143 – 159, 1996

[4] P. Coppens, In *International Tables for Crystallography*, Vol. B: Reciprocal Space (U. Shmueli, ed.), Section 1.2.11. Dordrecht: Kluwer Academic Publishers, 1993

[5] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991

[6] A.P. Dempster, N.M. Laird and D.B. Rubin, *Maximum likelihood for incomplete data via the EM algorithm*, Journal of the Royal Statistical Mathematical Society, Vol. B39, No. 1, pp. 1 – 38, 1977

[7] R. Durbin and D.E. Rumelhart, *Product units: A computationally powerful and biologically plausible extension to backpropagation network*, Neural Computation, Vol. 1, pp. 133 – 142, 1990

[8] J.A. Feldman and D.H. Ballard, *Connectionist models and their perspectives*, Computer Science, Vol. 6, pp. 205 – 254, 1982

[9] S. Fiori, *Blind Signal Processing by the Adaptive Activation Function Neurons*, Neural Networks, Vol. 13, No. 6, pp. 597 – 611, Aug. 2000

[10] S. Fiori, *A Theory for Learning by Weight Flow on Stiefel-Grassman Manifold*, Neural Computation, Vol. 13, No. 7, pp. 1625 – 1647, July 2001

[11] S. Fiori, *A Contribution to (Neuromorphic) Blind Deconvolution by Flexible Approximated Bayesian Estimation*, Signal Processing, Vol. 81, No. 10, pp. 2131 – 2153, Sept. 2001

[12] S. Fiori, *Probability Density Function Learning by Unsupervised Neurons*, International Journal of Neural Systems, Vol. 11, No. 5, pp. 399 – 417, Oct. 2001

[13] S. Fiori, *A Theory for Learning Based on Rigid Bodies Dynamics*, IEEE Trans. on Neural Networks, Vol. 13, No. 3, pp. 521 – 531, May 2002

[14] S. Fiori, *Non-Symmetric PDF Estimation by Artificial Neurons: Application to Statistical Characterization of Reinforced Composites*, IEEE Trans. on Neural Networks, Vol. 14, No. 4, pp. 959 – 962, July 2003

[15] S. Fiori, L. Albini, A. Faba, E. Cardelli and P. Burrascano, *Numerical Modeling for the Localization and the Assessment of Electromagnetic Field Sources*, IEEE Trans. on Magnetics, Vol. 39, No. 3, pp. 1638 – 1641, May 2003

[16] S. Fiori, *Analysis of Modified 'Bussgang' Algorithms (MBA) for Channel Equalization*, IEEE Trans. on Circuits and Systems – Part I. Accepted for publication

[17] M. Girolami and C. He, *Probability Density Estimation from Optimally Condensed Data Samples*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 25, No. 10, pp. 1253 – 1264, Oct. 2003

[18] A. Hyvärinen, *New approximations of differential entropy for independent component analysis and projection pursuit*, Report A47, Dept. of Com-

puter Science, Helsinki University of Technology (Finland), Aug. 1997 (partially published in Advances in Neural Information Processing Systems 10, pp. 273 – 279, 1998)

[19] IEEE/ANSI C95.1-1991, *IEEE Standard for Safety Levels with Respect to Human Exposure to Ratio Frequency Electromagnetic Fields, 3 kHz to 300 GHz*, 1991

[20] ICNIRP, *Guidelines for limiting exposure to time-varying electric, magnetic and electromagnetic fields (up to 300 GHz)*, Health Physics, No. 4, Vol. 74, Apr. 1998

[21] ISO Standard 4287/1, *Surface Roughness - Terminology - Part 1: Surface and Its Parameters*, 1984

[22] ISO 4288, *Geometrical product specifications (GPS) - Surface texture: Profile method - Rules and procedures for the assessment of surface texture*, 1996

[23] C.K. Johnson and H.A. Levy, *International Tables for X-ray Crystallography*, Vol. IV, pp. 311 – 336. Birmingham: Kynoch Press, 1974

[24] W.F. Kuhs, *Acta Crystallographica*, Vol. A48, pp. 80 – 98, 1992

[25] P. Lacomme, J.C. Marchais, J.P. Hardange and E. Normant, *Air and space-borne radar systems*, IEE Radar, Sonar, Navigation and Avionics Series, Co-published with William Andrew Inc., USA, 2001

[26] K. Letsch and R. Matzner, *The Relation between Input and Output Probability Density for Discrete-Time Linear Systems - An Approximational Approach*, Proc. of the 31st annual Conference on Information Sciences and Systems, Vol. I, pp 460 – 461, (Baltimore - USA), March 1997

[27] H.-P. Lin, D.-B. Lin, R.-T. Juang, C.-Y. Ko and Y. Wang, *Background Noise Floor Measurements and Cells Planning for WCDMA System*, Proc. of the GlobeCom 2002 Conference (paper#811)

[28] P.M. Mallick, *Fiber-Reinforced Composites*, Marcel Dekker Inc., N.Y. 1993

[29] B.W. Mel, *Information processing in dendritic trees*, Neural Computation, Vol. 6, pp. 1031 – 1085, 1994

[30] G. Miller and D. Horn, *Probability Density Estimation Using Entropy Maximization*, Neural Computation, Vol. 10, pp. 1925 – 1938, 1998

[31] P. Mitra, C.A. Murthy and S.K. Pal, *Density-based Multiscale Data Condensation*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 6, pp. 734 – 747, June 2002

[32] L. Mummery, *Surface Texture Analysis - The Handbook*, Hommelwerke, 1990

[33] L. E. Nielsen, *Mechanical Properties of Polymers and Composite*, Vol. 2, Marcel Dekker (New York), 1974

[34] K.H. Norwich, *The psychophysics of taste from the entropy of the stimulus*, Perception & Psychophysics, Vol. 35, No. 3, pp. 269 – 278, 1984

[35] H. White, *An efficient algorithm to compute maximum entropy densities*, Econometric Reviews, Vol. 18, No. 2, pp. 127 – 140, 1999

[36] V.K. Rohatgi, *Statistical Inference*, J. Wiley & Sons, New York, 1984

[37] Z. Roth and Y. Baram, *Multidimensional Density Shaping by Sigmoids*, IEEE Trans. on Neural Networks, Vol. 7, No. 5, pp. 1291 – 1298, Sept. 1996

[38] T. Schürmann and P. Grassberg, *Entropy estimation of symbol sequences*, CHAOS, Vol. 6, No. 3, pp. 414 – 427, 1996

[39] D. Scott, *Multivariate Density Estimation*, J. Wiley & Sons, 1992

[40] M.W. Spratling and G.M. Hayes, *Learning Synaptic Clusters for Nonlinear Dendritic Processing*, Neural Processing Letters, Vol. 11, No. 1, pp. 17 – 27, Feb. 2000

[41] G. Swartwout, *Electromagnetic Pollution Solutions*, Aerai Publications, 1991

[42] A. Zellner and R.A. Highfield, *Calculation of maximum entropy distributions and approximation of marginal posterior distributions*, Journal of Econometrics, Vol. 37, pp. 195 – 209, 1988

[43] U.H. Zucker and H.H. Shulz, *Acta Crystallographica*, Vol. A38, pp. 568 – 576, 1982