Nonlinearity Analysis of Variables for Modelling and Control

Esko K. Juuso

Control Engineering, Environmental and Chemical Engineering, Faculty of Technology, P.O.Box 4300, FI-90014 University of Oulu, Finland, (e-mails: esko.juuso@oulu.fi)

Abstract: Nonlinearities become essential in various systems when the operating area widens. The linear models are special cases for narrow areas. The behaviour is often asymmetric and can become gradually steeper or flatter depending on the case. These nonlinear effects can be analysed from data distributions for chosen operating areas. Further extensions require recursive analysis. The widely used Gaussian distribution is seldom valid for a wide area. The variable specific scaling can be presented with two second order polynomial defined by five parameters interpreted as the operating point and four corner points of the feasible range. These parameters define the shape factors which may require adjusting to fill the only requirement that the functions need to be monotonously increasing. Alternative constraints provide good solutions for combining expert knowledge with the data-based analysis. If the nonlinear behaviour is analysed correctly, only linear interactions are needed in the models. As the analysis is based on the same methodology, different applications can be combined by using appropriate process data. The smooth operation and high quality of products is the main goal of all these applications, and this can be achieved by combining these indicators with process control in the same way as it has been one for smaller indicators used in condition monitoring and process control. Different parts of the methodology have been tested in versatile applications. The main benefit is that the same structures can be used in various applications since the scaling functions take care of linking to the nonlinear real world.

Keywords: intelligent models, nonlinear scaling, statistical analysis, nonlinear systems

1. INTRODUCTION

Pieces of the informative and reliable datasets are selected in such a way that the data may contain measurement sets from several experiment periods. The *multiple-experiment* sets and selected data periods must be handled appropriately, especially in dynamic modelling. Feedback effects, narrow operating areas and unknown disturbances cause problems in modelling. *Designed experiments* are needed if the data material is not sufficient for modelling (Hinkelmann and Kempthorne, 2008). In industrial applications, the primary goal is to extract the maximum amount of unbiased information from as few (costly) observations as possible.

Normalisation or *scaling* of the data is needed for measurements with considerably different magnitudes. Widely used *min-max normalisation* matches the values between the minimum and maximum to the range [0, 1]. The operating point c_j is fixed in *z-score*,

$$p_j = \frac{x_j - c_j}{\Delta c_j}. \qquad (1)$$

which is calculated about the arithmetic mean, $c_j = \overline{x_j}$, by using the standard deviation of the variable $\Delta c_j = \sigma_j$, transforms the values to a distribution with mean of 0 and standard deviation 1. The arithmetic means and standard deviations are optimal for normal distributions.

Data distributions should be taken into account in estimating the centre point c_j and developing the scaling functions. The geometric mean and harmonic mean are useful when the sample is distributed log-normal or heavily skewed. The median and trimmed mean are two measures that are resistant (robust) to outliers. The trimmed mean ignores a small percentage of the highest and lowest values of a sample when determining the centre of the sample. Scaling with the *median* and *median absolute deviation*, i.e. $c_j = median(x_j)$ and $\Delta c_j = median(|x_j - median(x_j)|)$, provides a solution, which is insensitive to outliers and the points in the extreme tails of the distribution. *Decimal scaling*, where the values are scaled by $10^{log_{10} \max(x_j)}$, suits for cases where the ranges of the variables vary by a logarithmic factor. Minimum and maximum values are very sensitive to outliers.

The *outliers*, which are unusually large disturbances caused for example by temporary sensor or transmitter failures, should be removed from the data. This can be done by examining more thoroughly the data corresponding to the unusually large residual values. An observation is often considered as an outlier if the absolute value $|p_j|$ obtained by (1) is greater than 3. The Joliffe method has been introduced to detect observations that do not confirm with the correlation structure of the data (Fortuna et al., 2007; Warne et al., 2004). A survey of outlier detection methods is reported in (Englund and Verikas, 2005). As statistical inspection of process data tend to remove peaks which can carry precious information about system dynamics, all available information, including expert knowledge and input-output relationships, should be used (Fortuna et al., 2007).

Nonlinear activation functions, log-sigmoid and hyperbolic tangent, are used to generate the neuron outputs from the sum of the weighted inputs and the bias. These functions have been modified to improve the normalisation of the matching scores in multimodal biometric systems (Jain et al., 2005; Snelick et al., 2005). The *log-sigmoid function*, $(1+exp(-2p_j))^{-1}$, can be used for nonlinear scaling from the z-score values p_j to the range $[0, 1]$. The *double sigmoid* function extends this with different linear characteristics in the intervals $[c_j - \Delta c_j^-,\ c_j]$ and $[c_j,\ c_j + \Delta c_j^+]$. The operating point c_j and the edges Δc_j^- and Δc_j^+ are tuned. The sigmoid function is related to the *hyperbolic tangent* $tanh(\frac{1}{2}p_j)$, which scales to the range $[-1, 1]$. The functions introduced by Snelick et al. (2005) are based on the scaling of the min-max normalised values with two functions, which are quadratic, logistic and combined linear and quadratic: the inflection point is in the range [0, 1].

The *clustering algorithms* can be used for compressing large datasets for modelling: the cluster centres will replace the corresponding datapoints. *Interpolation* is needed if measurements are not frequent enough or if the sampling period is not constant, e.g. various laboratory measurements are based on samples taken infrequently compared to the on-line measurements. In practice, some measurements are missing because of failures in sensors or in data acquisition. These values are either reported as missing or recognised as erroneous values. Missing data can be replaced by using imputation with constants, e.g. the feature or class mean (Enders, 2010). Outliers are handled in the same way but with extra care as their difference from the acceptable values can be fairly small. For large data sets, missing values are simply left out, since the imputation may bias the data. Multiple solutions based on clustering or model-based correction form a basis for iteration.

Quality control systems are developed

- to make quality control more effective and closer to real time,
- to identify calibration, measurement and communication errors as close to the observation source as possible,
- to focus on automatic quality control algorithms development,
- to develop a comprehensive flagging system to indicate data quality level,
- to make it easier for data users to identify suspicious and erroneous data, and to highlight corrected values.

Numerous methods are used real-time and non real-time for the spatial and temporal checks of meteorological data (Vejen et al., 2002).

Nonlinear effects can be presented with various functions but the large systems become highly complex combinations of special modules. In fuzzy set systems, the meanings of the variables are shown with a set of membership functions and the interactions between labels are handled with fuzzy rules.

This paper focus on nonlinearity analysis to find unified solutions for modelling and control applications (Section 2). Proposed parametric methodologies are compared with several statistical distributions (Section 3). The methodologies open new possibilities for different types of applications discussed in Section 5. Conclusions and future research are presented in Section 6.

2. NONLINEARITY ANALYSIS

The nonlinearity analysis is based on the data distributions in the operating area of the (sub)system. Data values are transformed to dimensionless scaled values, also called linguistic values, are set to be within a real-valued interval [-2,2]. The basic scaling approach presented in (Juuso, 2004) has been improved later: a new constraint handling was introduced in (Juuso, 2009), and a new skewness based methodology was presented for signal processing in (Juuso and Lahdelma, 2010).

The generalised data-driven analysis extends solutions with dimensionless features and indicators. The resulting nonlinear scaling functions are compact solutions for variable specific nonlinearity handling.

2.1 Fuzzy systems

The origin in fuzzy set systems is seen variable specific feasible ranges which are defined by membership functions. Membership functions for finer partitions can be generated with the scaling functions (Juuso et al., 1993). The support area is defined by the minimum and maximum values of the variable, i.e. the support area is $[\min(x_j), \max(x_j)]$ for each variable $j, j = 1, \ldots, m$. The central tendency value, c_j, divides the support area into two parts, and the core area is defined by the central tendency values of the lower and the upper part, $(c_l)_j$ and $(c_h)_j$, correspondingly. This means that the core area of the variable j defined by $[(c_l)_j, (c_h)_j]$ is within the support area.

In early applications, the corner points were extracted from existing rule-based fuzzy systems or defined manually. The fuzzy labels were understood as membership locations corresponding values of the membership definitions within the range [-2,2].

2.2 Data-driven analysis

All the parameters are defined together in the data-driven approach. The analysis of the corner points and the centre point were earlier based on the arithmetic means or medians of the corresponding data sets (Juuso, 2004).

The nonlinearity analysis has been later extended to generalised norms defined by

$$||^{\tau}M_j^p||_p = (^{\tau}M_j^p)^{1/p} = [\frac{1}{N}\sum_{i=1}^{N}(x_j)_i^p]^{1/p}, \qquad (2)$$

where $p \neq 0$, is calculated from N values of a sample, τ is the sample time. With a real-valued order $p \in \Re$ this norm can be used as a central tendency value if $||^{\tau}M_j^p||_p \in \Re$, i.e. $x_j > 0$ when $p < 0$, and $x_j \geq 0$ when $p > 0$. The norm (2) is calculated about the origin, and it combines two trends:

a strong increase caused by the power p and a decrease with the power $1/p$. All the norms have same dimensions as x_j. The norm (2) is a Hölder mean, also known as the power mean. The generalised norm for absolute values $|x_j|$ was introduced for signal analysis in (Lahdelma and Juuso, 2008a).

For variables with only negative values, the norm is the opposite of the norm obtained for the absolute values. If a variable has both positive and negative values, each norm is an average of two norms obtained where the data sets are made positive and negative by subtracting a value $x_L < min((x_j))$ and a value $x_H > max(x_j))$, respectively. (Juuso, 2011b)

The generalised norm values increase with increasing order, i.e.

$$({}^{\tau}M_j^p)^{1/p} \leq ({}^{\tau}M_j^q)^{1/q}, \tag{3}$$

if $p < q$. The increase is monotonous if all the signals are not equal. The arithmetic mean, the harmonic mean and the root-mean-square (rms) are special cases where the order p is 1, -1 and 2, respectively. Norms from the minimum to the maximum corresponding the orders $-\infty \leq p < \infty$ are presented by (2), i.e. the definition includes the l_p norms defined for $1 \leq p < \infty$. The geometric mean is obtain from (2) when the order $p \to 0$.

The computation of the norms can be divided into the computation of equal sized sub-blocks, i.e. the norm for several samples can be obtained as the norm of the norms of the individual samples:

$$||^{K_s\tau}M_j^p||_p = \{\frac{1}{K_s}\sum_{i=1}^{K_s}[({}^{\tau}M_j^p)_i^{1/p}]^p\}^{1/p} \tag{4}$$

where K_s is the number of samples $\{x_j\}_{i=1}^N$. In automation and data collection systems, the sub-blocks are normally used for arithmetic mean ($p = 1$).

2.3 Dimensionless features

Distributions of the data can be analysed with dimensionless features obtained by normalising the moments $\underline{M}_j^k$, for example by standard deviation σ_j:

$$\gamma_k = \frac{{}^{\tau}\underline{M}_j^k}{\sigma_j^k} = \frac{1}{N\sigma_j^k}\sum_{i=1}^{N}[(x_j)_i - c_j]^k, \tag{5}$$

where the moment $\underline{M}_j^k$ is obtained about some central value, usually arithmetic mean. Variance σ_j^2 is the second moment $\underline{M}_j^2$. The feature γ_3 is called the coefficient of skewness, or briefly skewness, and the feature γ_4 as the coefficient of kurtosis. The skewness is a measure of asymmetry: $\gamma_3 = 0$ for a symmetric distribution. If $\gamma_3 > 0$, the skewness is called positive skewness and the distribution has a long tail to the right, and vice versa if $\gamma_3 < 0$. The kurtosis is a measure of the concentration of the distribution near its mean. The generalised moment for absolute values $|x_j|$ was introduced for signal analysis in (Lahdelma and Juuso, 2008b).

The normalised moments (5) are generalised by using the generalised norm (2) as the central value. The standard deviation σ_j, which is calculated about the origin, is used to obtain a dimensionless feature. Juuso and Lahdelma (2010) introduced a new approach based on using the generalised skewness γ_3^p for defining the central tendency value and the core area. The central tendency value is chosen by the point where the skewness changes from positive to negative, i.e. $\gamma_3^p = 0$. Then the data set is divided into two parts: a lower part and an upper part.

The same analysis is done for these two data sets. The estimates of the corner points, $(c_l)_j$ and $(c_h)_j$, are the points where $\gamma_3^p = 0$ for the lower and upper data sets, respectively. Since the search of these points is performed by using the order of the moment, the resulting orders $(p_l)_j$, $(p_0)_j$ and $(p_h)_j$ are good estimates when additional data sets are used. The norm values can be recursively updated with (4), and a new search for the orders is done only if the values change considerably (Juuso, 2011b).

In practical applications, the data points do not always cover the whole area of operation, e.g. only the close neighbourhood of the normal operation point may be covered, or we would like to extend the model of upper part later to the lower part. Only one part may be in use in fault diagnosis. Expert knowledge is used in extending the feasible range or selecting the methodologies.

Process data often contains outliers, which must be removed before generating the feasible area, because the procedure described above is sensitive to them. This is the idea in medians and trimmed means, which are used for the data samples containing outliers. A good estimate for the support area can be obtained with the generalised norms (2) with large negative and large positive orders since these features are less sensitive to the outliers than the minimum and maximum values. Discarding values at the high and low end can be used together with the generalised norms if there are obvious outliers. Trimming does not need to be the same for the low and high values.

The operating area of each variable is defined by a feasible range represented with a trapezoidal membership function whose corner points are $\min(x_j)$, $(c_l)_j$, $(c_h)_j$ and $\max(x_j)$. Warnings and alarms can be generated directly from the degrees of membership of the complement.

2.4 Nonlinear scaling functions

A nonlinear scaling function is defined as a (nonlinear) mapping of variable values inside its range to a range $[-2, 2]$, denoted as *linguistic range*. It more or less describes the distribution of variable values over its range which includes the normal operation in the range $[-1, 1]$ and the areas with warnings and alarms. The values X_j are called *linguistic values* since the scaling idea originates from the fuzzy set systems: values -2, -1, 0, 1 and 2 can be associated to the linguistic labels, e.g.

$$\{very\ low,\ low,\ normal,\ high,\ very\ high\} \tag{6}$$

are defined with membership functions The number of membership functions is not limited to five: the values between these integers correspond to finer partitions of the fuzzy set system. The early applications of the linguistic equations used only integer values (Juuso, 1999).

In present systems, membership definitions are used in a continuous form consisting of two second order polynomials:

$$\begin{aligned} x_j &= f_j^-(X_j),\ X_j \in [-2,0), \\ x_j &= f_j^+(X_j),\ X_j \in [0,2]. \end{aligned} \tag{7}$$

The functions should be monotonous, increasing functions in order to result in realisable systems. The lower part function is defined by values corresponding linguistic levels -2, -1 and 0, and the upper part function by values corresponding linguistic levels 0, 1 and 2. The upper and lower parts should overlap at the linguistic value 0. (Juuso, 2004)

Five parameters define the coefficients of the second order polynomials,

$$\begin{aligned} f_j^-(X_j) &= a_j^- X_j^2 + b_j^- X_j + c_j,\ X_j \in [-2,0), \\ f_j^+(X_j) &= a_j^+ X_j^2 + b_j^+ X_j + c_j,\ X_j \in [0,2]. \end{aligned} \tag{8}$$

The scaling function is asymmetrical when the coefficients in the upper and lower part are different. The centre point, c_j, defines the operating point. Four linear equations are needed for solving the other coefficients:

$$\begin{aligned} 4a_j^- - 2b_j^- + c_j &= \min(x_j), \\ a_j^- - b_j^- + c_j &= (c_l)_j, \\ a_j^+ + b_j^+ + c_j &= (c_h)_j, \\ 4a_j^+ + 2b_j^+ + c_j &= \max(x_j). \end{aligned} \tag{9}$$

In order to keep the functions monotonous and increasing, the derivatives of functions f_j^- and f_j^+ should always be positive (Fig. 1). As a second order polynomial has either a minimum or a maximum point, this requirement is fulfilled only if these points are outside the ranges $(-2,0)$ and $(0,2)$ for functions f_j^- and f_j^+, respectively. The derivatives,

$$\begin{aligned} D_j^- &= 2\ a_j^- X_j + b_j^-,\ X_j \in [-2,0), \\ D_j^+ &= 2\ a_j^+ X_j + b_j^+,\ X_j \in [0,2], \end{aligned} \tag{10}$$

are corrected to positive in the areas $(-2,0)$ and $(0,2)$, respectively, by changing the coefficients of the polynomials (Juuso, 2004). The membership definitions are continuous functions but derivatives can have discontinuities in the centre point.

The functions are monotonous and increasing if the ratios,

$$\begin{aligned} \alpha_j^- &= \frac{(c_l)_j - \min(x_j)}{c_j - (c_l)_j}, \\ \alpha_j^+ &= \frac{\max(x_j) - (c_h)_j}{(c_h)_j - c_j}, \end{aligned} \tag{11}$$

are both in the range $[\frac{1}{3},\ 3]$, see (Juuso, 2009). If needed, the ratios are corrected by modifying the core $[(c_l)_j, (c_h)_j]$ and/or the support $[\min(x_j), \max(x_j)]$. Errors are checked independently for f_j^- and f_j^+: each error can always be corrected either by moving the corner of the core or the support In some cases, good results can also be obtained by moving c_j in the range defined by If these constraints allow a non-empty range, the maximum of the lower limits and the minimum of the upper limit are chosen to define the limits for continuous definitions (Fig. 2).

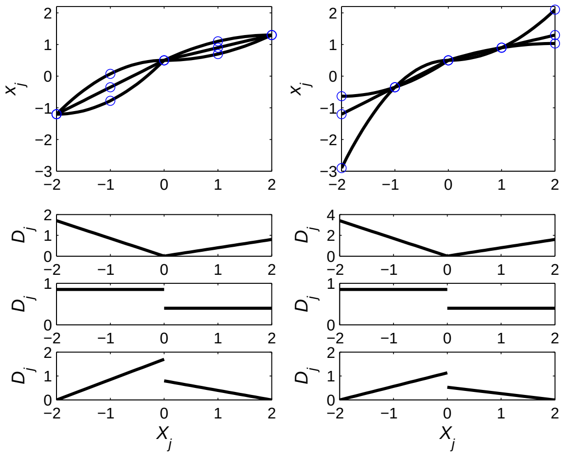

Fig. 1. Feasible shapes of the membership definitions f_j and corresponding derivatives D_j: coefficients adjusted with the core (left) and the support (right). Derivatives are presented in three groups: (1) decreasing and increasing, (2) linear and linear, and (3) increasing and decreasing. (Juuso, 2009).

The coefficients of the polynomials can be represented by

$$\begin{aligned} a_j^- &= \frac{1}{2}(1-\alpha_j^-)\ \Delta c_j^-, \\ b_j^- &= \frac{1}{2}(3-\alpha_j^-)\ \Delta c_j^-, \\ a_j^+ &= \frac{1}{2}(\alpha_j^+ - 1)\ \Delta c_j^+, \\ b_j^+ &= \frac{1}{2}(3-\alpha_j^+)\ \Delta c_j^+, \end{aligned} \tag{12}$$

where $\Delta c_j^- = c_j - (c_l)_j$ and $\Delta c_j^+ = (c_h)_j - c_j$. Membership definitions may contain linear parts if some coefficients α_j^- or α_j^+ equals to one (Fig. 1).

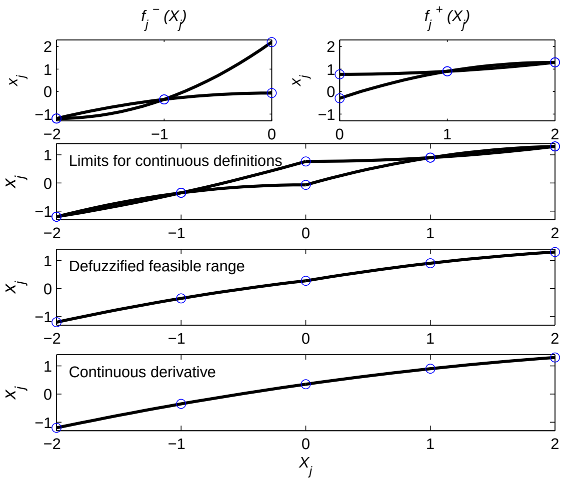

Fig. 2. Membership definitions in the core: coefficients adjusted with the centre point c_j.

The centre point is not known if the feasible range is defined manually. It can be calculated by defuzzifying the feasible range with the centre of gravity: For strongly asymmetrical feasible ranges, this value may be outside the core (Juuso, 2004). The requirement can be fulfilled by modifying the corner points. Additional constraints can be taken into account, e.g. a good solution can be to use a locally linear function in the neighbourhood of the centre point. Then a continuous derivative is chosen at the centre point. This can be achieved by modifying the centre point or the corner points of the feasible range. There can be several acceptable modifications, for which the ratios (11) remain in the range $[\frac{1}{3}, 3]$.

Monotonously increasing membership definitions can be constructed by adjusting the centre point c_j, the core $[(c_l)_j, (c_h)_j]$ and the support $[\min(x_j), \max(x_j)]$. An easier way for manual approach was introduced in (Juuso, 2009): first define the centre point c_j, then the core by choosing the ratios (11) from the range $[\frac{1}{3}, 3]$, and finally calculate the support $[\min(x_j), \max(x_j)]$. The norms are used together with the generalised skewness in the data-driven approach to define the centre and corner points. The ratios (11), which are checked in all data-driven cases, are also guiding the manual construction of the membership definitions.

For each variable, the membership definitions are configured with five parameters, which can be presented with three consistent sets. The working point (centre point) c_j belongs to all these sets, where the other parameters are:

- the corner points $\{\min(x_j), (c_l)_j, (c_h)_j, \max(x_j)\}$ are good for visualisation;
- the parameters $\{\alpha_j^-, \Delta c_j^-, \alpha_j^+, \Delta c_j^+\}$ suit for tuning;
- the coefficients $\{a_j^-, b_j^-, a_j^+, b_j^+\}$ are used in the calculations.

The upper and lower parts of the scaling functions can be convex or concave independently. Also, simplified functions can be used: a linear membership definition needs only two parameters: c_j and $b_j = b_j^+ = b_j^-$ or $\Delta c_j = \Delta c_j^+ = \Delta c_j^-$, since $\alpha_j^+ = \alpha_j^- = 1$ and $a_j^+ = a_j^- = 0$; an asymmetrical linear definition has $\Delta c_j^+ \neq \Delta c_j^-$ and $b_j^+ \neq b_j^-$. Local linear functions defined by are used if appropriate.

3. STATISTICAL DISTRIBUTIONS

In data-based analysis, the nonlinear scaling functions are based on data samples. The parameters obtained by statistical analysis depend strongly on the statistical distribution. The functions extend the normalisation and scaling solutions from the symmetric special case defined by the z-score (1), where $c_j = ||^\tau M_j^1||_1$ and $\Delta c_j = \sigma_j = ||^\tau M_j^2||_2$, i.e. generalised norms (2) with orders $p = 1$ and $p = 2$, respectively. Other special cases, geometric mean ($p = 0$) and harmonic mean ($p = -1$), are used in defining the centre of the sample for log-normal or heavily skewed data. Trimmed or truncated means, medians and median absolute deviations are generally recommended for the cases with outliers. The generalised norms can also be trimmed by discarding values at the high and low end. For heavily skewed data, the discarding limits are defined by the norms with high positive and negative orders, respectively.

In the skewness based approach presented above, all the parameters are analysed from the data. As expected, the c_j is close to the arithmetic mean ($p = 1$) when the sample is taken from a normal distribution. Normalisation with the z-score is the first phase since the core is symmetrical, i.e. $\Delta c_j^+ = \Delta c_j^- = \frac{1}{2}\, ||^\tau M_j^2||_2$. The resulting shape factors are equal, $\alpha_j^- = \alpha_j^+ = 3$, and the support is $[c_j - 2\sigma_j,\ c_j + 2\sigma_j]$. The size of the random sample effects on the analysis: the centre point is correctly obtained from a small sample ($N = 1000$), and also the core is fairly accurate. The limits of the support area and the shape factors require larger samples, e.g. 10000 points provides fairly good estimate, but 50000 points are required for highly accurate estimates. Only a slight adjustment of the core or preferably the support is needed for these samples.

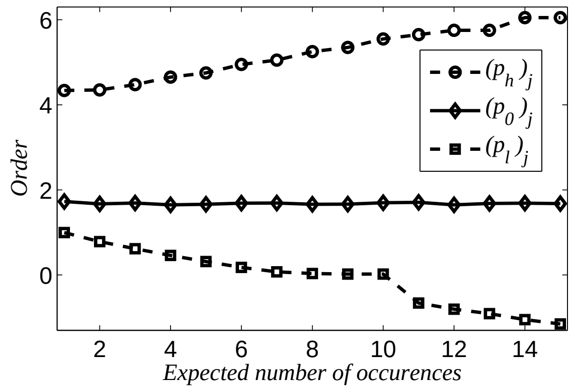

Fig. 3. Orders of the norms (Poisson)

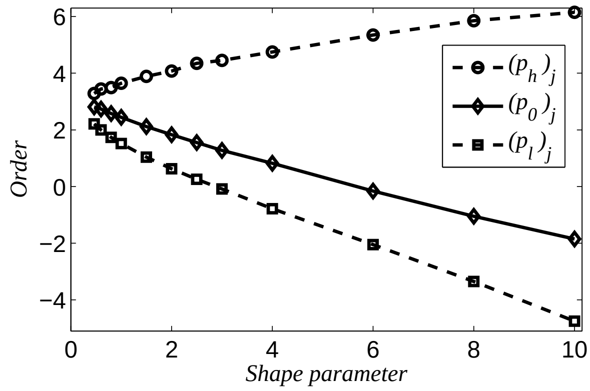

Fig. 4. Orders of the norms (Weibull, $\lambda_W = 2$).

The scaling functions become asymmetrical about the centre c_j in random samples of Poisson and Weibull distributions. Orders of the norms, $\{(p_l)_j, (p_0)_j, (p_h)_j\}$, and shape factors, $\{\alpha_j^-, \alpha_j^+\}$, show strong variations in these asymmetrical distributions (Figs. 3 - 6). For the Poisson distribution, the order $(p_0)_j$ is almost constant, 1.68 ± 0.03 when the expectation number $\lambda_P \geq 2$, and $(p_0)_j = 1.73$ when $\lambda_P = 1$ (Fig. 3). For the Weibull distribution, the order $(p_0)_j$ decreases smoothly from 2.8 to -1.85 when the shape parameter increases from one to ten (Fig. 4). The order range $[(c_l)_j, (c_h)_j]$ increases for both: from [1, 4.34] to $[-1.15, 6.05]$ for Poisson and from [2.2, 3.2] to $[-4.75, 6.15]$ for Weibull distributions whose scale parameter $\lambda_W = 3$.

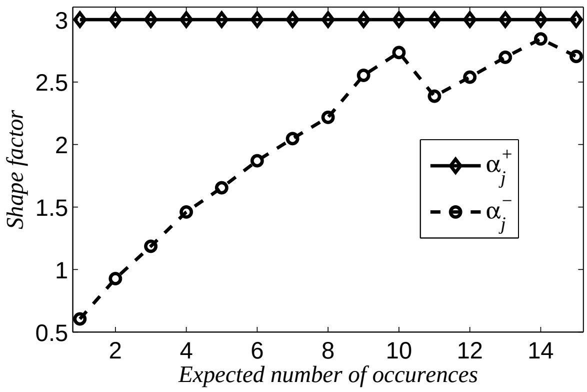

Fig. 5. Shape factors (Poisson).

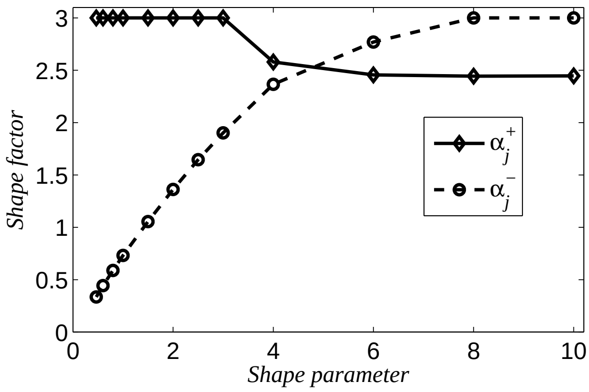

Fig. 6. Shape factors (Weibull, $\lambda_W = 2$)).

Poisson distributions have the same shape factor α_j^+ as the normal distributions, but the shape factor α_j^- increases from 0.6 to almost 3 when the expectation number λ_P increases from one to ten (Fig. 5). The core also is asymmetrical: $\Delta c_j^+ > \Delta c_j^-$. The difference is high, when λ_P is small, and becomes negligible, when $\lambda_P > 10$ (Fig. 5). Weibull distributions are very asymmetrical when the shape parameter κ is small: $\Delta c_j^+ \gg \Delta c_j^-$, $\alpha_j^- \approx \frac{1}{3}$ and $\alpha_j^+ = 3$, when $\kappa = 0.5$ (Fig. 6). This exponential distribution becomes more symmetrical when κ increases, but becomes again asymmetrical for higher κ values (Fig. 6). The Poisson distributions have only integer values, which causes irregular changes in orders $(c_l)_j$ and α_j^- obtained from random samples.

For all these distributions, the core area becomes wider than in the previous approaches where the mean or the median were used. Higher sensitivity around the centre point was already detected in (Juuso and Lahdelma, 2010). High positive and negative orders are used in selecting the limits for the core area if small deviations are not important. Asymmetrical scaling functions can be obtained by analysing the upper and the lower part separately.

The scaling functions consisting of two second order polynomials operate well for versatile distributions, and various sigmoid functions can be interpreted as special cases. The centre points, which define the operating point of the model, can be defined manually. For the error, the derivative of error, the sum of error, the original error and the change of control, the centre point is zero. Also, the core and support areas can be defined manually for any membership definition. Monotonous increase needs to be checked for the manually defined functions.

The shape factors define the type of the feasible range: narrow and wide cores correspond to high and low shape factors, respectively. Also, an asymmetric core, i.e. the core can be narrow on one side of the centre point c_j and wide on the other side, is allowed. The support can depend strongly on the number of points as seen in the comparisons of different statistical distributions. Expert knowledge and physical limitations can be used in selecting the shape factors α_j^- and α_j^+. The factors can be set to three if the data set is fairly limited and there is no specific additional knowledge. Linear scaling functions, i.e. $\alpha_j^- = \alpha_j^+ = 1$ are used if the material is very limited.

4. NATURAL LANGUAGE

The values within the range [-2, 2] obtained by the nonlinear scaling are also called as linguistic values since they can be interpreted with linguistic terms. The linguistic terms can be interpreted as fuzzy numbers: for example values -2, -1, 0, 1 and 2 can be associated to the linguistic labels (6) which can be made sharper or wider with powering modifiers 'extremely', 'very', 'more or less' and 'roughly', and then processed with the conjunction, disjunction and negation. Applications can have specific labels to make understanding easier, and the number of labels are not limited to these examples. The labels are only for information, the calculations are done with the numbers.

5. TYPES OF APPLICATIONS

The nonlinear scaling approach expands the operating areas in many applications. The following areas are examples where compact solutions have been developed. Severity criteria are checked with the scaled values are the same for all variables. Indicators, models and control can be combined in applications (Juuso, 2018).

Intelligent indicators are the first applications of the combinations of the generalised norms and nonlinear scaling. Even single norms or indicators can replace and outperform large rule-based systems. Several indicators can be combined as a weighted sum. The severity criteria are the same for these combined indicators as well. (Juuso and Lahdelma, 2010)

Statistical process control (SPC) is an important area in utilizing data. The generalised SPC introduced in (Juuso, 2015) expands the SPC from Gaussian to non-Gaussian data sets. The analysis methods are suitable for a large set of statistical distributions. Categorical information can be studied with the same approach by using manual definitions, which means that also mixed cases can be handled. The limits can be updated in short run SPC since they are defined by the nonlinear scaling approach. The limits can even change gradually. The GSPC does not need any interruptions and even recursive approaches are possible. In these systems, the control levels are defined uniformly for the scaled values.

DOI: 10.3384/ecp212.065

Modelling and simulation is extended to nonlinear systems by combining the nonlinear scaling and linear equations. The models can be adapted to different operating conditions by changing the parameters of the scaling functions (Juuso, 2020).

Intelligent LE controllers can use linear controller structures in nonlinear systems, The controllers can be adapted to different operating conditions by changing the parameters of the scaling functions. (Juuso, 2011b)

Temporal analysis provides indirect measurements and detection of trend episodes for high level control. For any variable, a *trend index* is calculated as a difference of the means of the scaled values obtained for a short and a long time period, respectively. The index value is in the linguistic range $[-2, 2]$ representing the strength of both decrease and increase of the variable x_j. The same analysis can be used for detecting temporal changes of any indicators (Juuso, 2011a).

6. CONCLUSIONS AND FUTURE RESEARCH

This paper summarizes the main parts of the nonlinear scaling approach. Highly nonlinear asymmetrical data can be utilized in appropriate way. There is no need to assume Gaussian data outside its operating area. Different parts of the methodology has been tested in versatile applications. The main benefit is the analysis of the nonlinear behaviour. Different applications can extend the use of linear structures by enhancing them with the nonlinear scaling to take care about linking to the nonlinear real world. Future research continues this with more detailed analysis of applicability.

REFERENCES

Enders, C.K. (2010). *Applied Missing Data Analysis.* Guilford Press, New York.

Englund, C. and Verikas, A. (2005). A hybrid approach to outlier detection in the offset lithographic printing process. *Engineering Applications of Artificial Intelligence*, 18(6), 759–768.

Fortuna, L., Graziani, S., Rizzo, A., and Xibilia, M.G. (2007). *Soft Sensors for Monitoring and Control of Industrial Processes.* Advances in Indusrial Control. Springer, New York. 270 pp.

Hinkelmann, K. and Kempthorne, O. (2008). *Design and Analysis of Experiments: Introduction to Experimental Design.* John Wiley & Sons, New York, 2nd edition.

Jain, A., Nandakumara, K., and Ross, A. (2005). Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12), 2270–2285.

Juuso, E. and Lahdelma, S. (2010). Intelligent scaling of features in fault diagnosis. In *7th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, CM 2010 - MFPT 2010, 22-24 June 2010, Stratford-upon-Avon, UK*, volume 2, 1358–1372. BINDT. ISBN=978-1-61839-013-4.

Juuso, E.K. (1999). Fuzzy control in process industry: The linguistic equation approach. In H.B. Verbruggen, H.J. Zimmermann, and R. Babuška (eds.), *Fuzzy Algorithms for Control, International Series in Intelligent Technologies*, volume 14 of *International Series in Intelligent Technologies*, 243–300. Kluwer, Boston. doi: 10.1007/978-94-011-4405-6_10.

Juuso, E.K. (2004). Integration of intelligent systems in development of smart adaptive systems. *International Journal of Approximate Reasoning*, 35(3), 307–337. doi: 10.1016/j.ijar.2003.08.008.

Juuso, E.K. (2009). Tuning of large-scale linguistic equation (LE) models with genetic algorithms. In *Adaptive and Natural Computing Algorithms, ICANNGA 2009, Lecture Notes in Computer Science*, volume 5495, 161–170. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-04921-7_17.

Juuso, E.K. (2011a). Intelligent trend indices in detecting changes of operating conditions. In *2011 UKSim 13th International Conference on Modelling and Simulation*, 162–167. IEEE Computer Society. doi: 10.1109/UKSIM.2011.39.

Juuso, E.K. (2011b). Recursive tuning of intelligent controllers of solar collector fields in changing operating conditions. *IFAC Proceedings Volumes*, 44(1), 12282–12288. doi:10.3182/20110828-6-IT-1002.03621.

Juuso, E.K. (2015). Generalised statistical process control GSPC in stress monitoring. *IFAC-PapersOnline*, 48(17), 207–212. doi:10.1016/j.ifacol.2015.10.104.

Juuso, E.K. (2018). Smart adaptive big data analysis with advanced deep learning. *Open Engineering*, 8(1), 403–416. doi:10.1515/eng-2018-0043.

Juuso, E.K. (2020). Expertise and uncertainty processing with nonlinear scaling and fuzzy systems for automation. *Open Engineering*, 10(1), 712–720. doi: 10.1515/eng-2020-0080.

Juuso, E.K., Bennavail, J., and Singh, M. (1993). Hybrid knowledge-based system for managerial decision making in uncertainty environment. In N.P. Carreté and M.G. Singh (eds.), *Qualitative Reasoning and Decision Technologies, Proceedings of the IMACS International Workshop on Qualitative Reasoning and Decision Technologies -QUARDET'93, Barcelona, June 16 - 18, 1993*, 234–243. CIMNE, Barcelona.

Lahdelma, S. and Juuso, E. (2008a). Signal processing in vibration analysis. In *5th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, CM 2008 - MFPT 2008, 15-18 July 2008, Edinburgh, UK*, 867–878. BINDT.

Lahdelma, S. and Juuso, E. (2008b). Vibration analysis of cavitation in Kaplan water turbines. *IFAC Proceedings Volumes*, 41(2), 13420–13425. doi:10.3182/20080706-5-KR-1001.02273.

Snelick, R., Uludag, U., Mink, A., and Jain, M.I.A. (2005). Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3), 450–455.

Vejen, F., Jacobsson, C., Fredriksson, U., Moe, M., Andresen, L., Hellsten, E., Rissanen, P., Pálsdóttir, and Arason (2002). Quality control of meteorological observations automatic methods used in the nordic countries. Oslo, Norway.

Warne, K., Prasad, G., Rezvani, S., and Maguire, L. (2004). Statistical and computational intelligence techniques for inferential model development: a comparative evaluation and a novel proposition for fusion. *Engineering Applications of Artificial Intelligence*, 17, 871–930.