Model processes and cavitation indicators for a quantitative description of an ultrasonic cleaning vessel: Part II – Multivariate data analysis

Christian Koch *, Matthias Jüschke

Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig, Germany

A R T I C L E   I N F O

Article history:
Received 10 June 2011
Received in revised form 20 December 2011
Accepted 28 December 2011
Available online 5 January 2012

Keywords:
Cavitation
Ultrasound
Multivariate data analysis
Erosion
Sonochemiluminescence

A B S T R A C T

A multivariate data analysis of cavitation indicators and parameters was carried out to improve the quantitative characterization of cavitation processes used in manufacturing and medical applications. The indicators were obtained from four model measurement methods applied to a 45 kHz cleaning vessel. Together with experimental data such as temperature and electrical input power they form the data basis of a factor analysis. The loadings of three factors were calculated and the indicators, the parameters, and finally the data were depicted in factor space. The factors show relations between the variables and several overlapping indicators and parameters were identified. The coordinates of the data (data scores) indicate tendencies within the data and the assessment of the factors allows the finding of hidden relations. Using the factor analysis three representing indicators or parameters specific to the application case are identified which can be used for a complete description of the process. This characterization method can favourably be applied in quality management systems.

1. Introduction

Cavitation is applied in many technical [1–3] and medical [4–7] fields where fluids are used for processing, cleaning, or as an imaging medium. A quantitative description of the processes is often required for optimization purposes or, in particular, for quality management. If, for example, a cleaning process step is included in a line production, the quality management system requires a specific description of all important process details for a potential replacement of the device in the case of malfunctioning. Since many parameters influence cavitation the outcome of such a cavitation-based process varies strongly with the application conditions and often has a random behaviour. Thus, no widely accepted process description methods or even standards have been established hindering the comparability and the reliability of manufacturing.

A possible strategy to find a general methodology for the description of cavitation effects is the determination of experimental indicators or numbers from cavitation induced effects. A variety of methods has been presented, for example, the generation of holes and dents in an aluminium foil [8–11] or the erosion of passivating layers on metal electrodes [12] for example made of aluminium [13]. These techniques rely on the mechanical impact of cavitation but also the chemical effects have been detected by electro-chemical probes [14–17] or model reactions [18–20]. Another alternative is to measure the sound field and to calculate particular spectral numbers such as the subharmonic or the noise power [21–26]. Many of these indicators arise, however, from very specific determination methods, and the relation to practical cavitation effects and outcomes is often not yet clear or proven.

In the first part of this study [27] a combination of model processes was developed and tested that could describe cavitation processes. The indicators were chosen to be closely related to major cavitation effects often used in applications. Several significantly different indicators were selected to cover a wide range of cavitation effects. It was shown that many properties of the cleaning process could be described by using these indicators in parallel [27]. It became, however, unclear as to what extent the indicators cover different properties of cavitation or whether they overlap. In addition, the relations between the indicators and their dependence on experimental parameters could not be investigated which is important for the question of which indicator could be represented by another one.

Cavitation is a statistical process and the quantitative output indicators such as cleaning quality or sonochemical yield often show random behaviour. Thus, the determination of quantitative parameters lacks reliability and repeatability [11,20,22] and many measurements are necessary to obtain ensemble average values. In general, statistical means could help to give further insight into data structure and to obtain reliable results. Methods of the multivariate data analysis are able to reveal structure and hidden relations within the data set and to sort and form useful subsets, cluster or principal variables. In this paper a factor analysis [28,29] was applied to data obtained from measurements of the
and several parameters set during the experiments. The analysis calculates common factors of all variables which were used as new axes in a factor space where all data were assigned with new coordinates. This procedure allows the revealing of relations and causalities between the variables (indicators and parameters) within the data set. It finds indicators which overlap and which are redundant in a process description. The factors represent particular properties of the cavitation experiment and can be estimated by indicators or parameters which are closely related. Finally, this leads to a data reduction because all important measurands can be described in terms of the factors, i.e. representative indicators or parameters. Under the assumption of knowing these relations, only the representing indicators or parameters need to be measured for a process description, for example after a replacement of the cleaning bath, which significantly reduces the measurement effort.

2. Determination of spatially resolved cavitation indicators

2.1. Measurement set-up

All measurements were carried out in a commercially available ultrasonic cleaning vessel (TI-H-5, Elma GmbH, Germany) of 4 L volume. The working frequency was $f_{0w} = 45$ kHz and the transducers were driven by an external amplifier (the Tamp Proline 1800, Musikhaus Thomann e.K., Burgebrach, Germany) [27]. The excitation signal was a non-modulated sine-wave generated by a synthesizer (3326A, Hewlett-Packard Company, Palo Alto, USA) and fed into the amplifier. The voltage at the transducers was measured via a resistor network and the excitation current into the transducers was detected by a current probe (P6021, Tektronix Inc., USA). The temperature of the de-ionized water in the vessel was measured by a 4-wired PT100 element and could be controlled with an accuracy of 1 °C by means of a flow system [27]. The O$_2$ content (determined by a HQ 30d sensor, Hach-Lange GmbH, Düsseldorf, Germany) was set by starting the measurements with degassed water and waiting for a particular time till approaching the next O$_2$ content value by naturally returning of air into the water. Seven measurement points (see Fig. 1) were chosen in the vessel to obtain results at very active and more passive areas in the vessel.

For clarity of terms, a parameter defines a measurand which is set by the experimenter during the measurement, for example the temperature or the synthesizer voltage. An indicator means a value or number measured or determined during the experiment. Note, that both are used as variables in the factor analysis.

2.2. Measurement of sound field data

The sound field was measured using a hydrophone (TC 4013, Reson, Denmark) which was positioned by a three-axis positioning system (TL 78, Micro-controle, France) controlled by a computer [27]. From the time-dependent data, the spectrum was calculated and several indicators were derived: the amplitude of the fundamental at $f_{0w}$, the subharmonic at $f_{0w}/2$, the ultraharmonic at $3f_{0w}$/2, the second harmonic at $2f_{0w}$, and a noise value integrating the noise spectral density in a frequency range from 100 kHz to 200 kHz. In addition, the rms-value of the sound pressure was determined from the time-dependent data.

2.3. Determination of erosion indicator

For quantitative determination of erosion an aluminium foil technique was used [11]. A 15 µm thick foil (aluminium alloy 1200, Korrf AG, Switzerland) was mounted in a frame and placed in the running cleaner for 20 s. The loaded foil was imaged using a flatbed scanner and purpose made software detected the dents, the burrs, and the holes in the foil [11,30]. Counting all cavitation events in a detection area with the size of the hydrophone dimensions and weighting the dents, burrs and holes yields the erosion indicator [27]. The detection area was chosen to account for spatial averaging effects, which were most prominent during the sound field measurements.

2.4. Determination of sonochemical indicator

The chemical effect of cavitation was concerned with the simple model reaction of the oxidation of iodine ions [20,31]. The test solution of 0.5 M potassium iodide (Sigma–Aldrich, Germany) was poured into a thin-walled plastic test tube which was taken to the different measurement points in the bath [27]. The local concentration of the test fluid in the tube allowed a spatially resolved measurement. After ultrasound exposure the fluid was analysed using a ultra-violet transmission spectrometer (SpectroFlex 6600, WTW GmbH, Germany) and the amount of produced I$_2$ ions was detected. The change in extinction was used as an indicator. For comparison with excitation parameters, it was not normalized to the electrical power of the transducers.

2.5. Measurement of luminescence

Luminescence was detected in a luminol solution (10 mg luminol were solved in 500 ml NaOH base of about pH 10.5) brought into the vessel. The solution was also poured into a plastic thin-walled test tube for a spatially resolved measurement. The emitted light from the nearly transparent test tube was detected by an EMCCD camera (iXon 885, Andor Technology, Northern Ireland) and compared to the background noise from the surroundings. The light intensity was averaged over the volume of the test tube. The shutter time was chosen in a way assuring the dynamic range of the detector being used in the full range.

2.6. Course of measurement cycles

Three measurement cycles were carried out for obtaining the experimental data (Table 1). In the first cycle air-saturated de-ionized water was used and the temperature was kept constant at 20 ± 1 °C. In the second cycle measurements at three different temperatures (10 °C, 20 °C, and 35 °C) were made using a different number of driving voltage settings. In the third and last cycle all parameters (temperature $T$, driving voltage $U$, and O$_2$ content $c_{O_2}$) were varied within a certain range (10–35 °C, 1.75–2.75 V, and 2.5–7 mg/L). The measurements were made within subsequent sub-cycles by increasing the temperature step by step from the starting value of 10 °C at a fixed O$_2$-content or raising the O$_2$-content from 2.5 mg/L at a fixed temperature. For practical reasons the $T$ and $c_{O_2}$ setting courses were not mixed. It was also not possible to measure all indicators at the same time because exchanging the sensors needed some time in which the O$_2$ content
The data matrix and all parameters and indicators are included as variables. The analysis tries to find overlapping of variables and calculates common factors that represent joint properties of the variables.

The number of three factors. was at least 90%. In nearly all calculations this was achieved by a number of factors was chosen in such a way, that communality was at least 90%. The explained part is called communality. The number of factors. Summing up all specific variances yields a meaningful of modelling. There are no clear rules for setting the number of measurements desirable for minimizing the technical effort. The erosion and the sonochemistry indicator are close together but the specific variances are large showing strong effect of the common factors but electrical power, fundamental and rms-sound pressure seem to be slightly separated from sub- and ultraharmonic. Fundamental and rms-sound pressures were redundant and often only one of them was calculated later on. Extending the measurement to the whole bath increases the spreading but the variables are still quite close together (Fig. 3). As already mentioned in the accompanying paper [27] the bath showed a comparatively homogeneous behaviour in the sense that general measurement results at different measurement points coincide even when the cavitation activity is very different. Although some individual measurement results may show larger differences this is a general trend which justifies a reduction of the number of measurements desirable for minimizing the technical effort. The erosion and the sonochemistry indicator are close together but the specific variances are large showing strong spreading of the data.

The wider the parameters are varied the larger is the variability of the variables in factor space. In Fig. 4 the $O_2$ content was varied between 2.5 mg/L and 7 mg/L and the complete upper half-space of factor space is used. The $O_2$ content and the temperature are close to factor 2 but in opposite direction which means that they have the opposite influence on the data set. This conclusion is highly relevant because the specific variances of both variables are small. Luminescence is close to the chemical indicator as could be expected; also the subharmonic and the electrical power can be found in this range. The fundamental has a difference to the measured electrical power which was found in many data sets.

Not only the variables can be analysed in the factor space plot. All data points, i.e. measurement results can also be visualized in the factor space. They can be assigned to the factors and, for example, changes of parameters can be observed by moving the data points in the factor space. This technique can be used to find general tendencies.

### 3. Multivariate analysis of experimental data

#### 3.1. Method

Multivariate data analysis is a powerful tool to investigate large data sets. It includes structure finding methods for data where no relations can be assumed a priori and structure proving methods when a hypothesis is already known. Since the cavitation data obtained from the experiments do not imply a structure a priori, a factor analysis was applied for analysing and arranging the data. For the numerical implementation, the routine of the statistical toolbox of Matlab was used.

For a factor analysis the input data have to be arranged as variables and all parameters and indicators are included as variables. The data matrix $X = x_{ij}$ contains $V$ variables in $V$ columns ($1 \leq k \leq V$) of length $N$ which is the number of measurements ($1 \leq i \leq N$). In the case of cavitation data, for example, in the first column the driving voltages of every measurement were set, in the second the temperature values and so on. For all calculations the data were normalized within each column by

$$Z = z_{ik} = \frac{x_{ik} - \bar{X}}{\sigma(k)}$$

where $\sigma(k)$ denotes the standard deviation of the $k$th column. The analysis tries to find overlapping of variables and calculates common factors that represent joint properties of the variables. The number of factors $F$ has to be chosen in advance (see below). Often a common factor indicates a causal relation between variables giving a deeper insight into the basis of the process. The factors span a new coordinate system (factor space) and both the variables and the data can be depicted in this new system. The coordinates of the variables in the factor space are called loadings in a matrix $A = a_{jk}$, where $1 \leq j \leq F$ and the coordinates of the experimental (original) data are the scores $P = p_{ij}$. The scores were calculated from the fundamental lemma of factor analysis

$$P = AZ$$

where the dash means the transpose of the matrix.

For each variable $z_{ik}$ a specific variance is calculated which describes the component due to independent random variability. A small specific variance indicates that most of the behaviour of the variable is explained by the common factors. The variances depend on the number of factors which can be chosen at the beginning of modelling. There are no clear rules for setting the number of factors. Summing up all specific variances yields a measure to assess the percentage of the total variability not explained by all factors; the explained part is called communality [29]. The number of factors was chosen in such a way, that communality was at least 90%. In nearly all calculations this was achieved by a number of three factors.

The results of the calculation can beneficially be visualized in an $F$-dimensional plot of factor space for a better overview. The variables were depicted as lines and the data as points. Variables which are close together have a high similarity because they were described by the same factor combination and a causal relation can be assumed. The similarity or correlation can be quantitatively expressed by the angle between the variable vectors in the factor space:

$$s = \frac{(\bar{a}_j - \bar{a}_m) \cdot (\bar{a}_j - \bar{a}_m)}{\| \bar{a}_j - \bar{a}_m \|}$$

$$\bar{a}_j = (a_{ij}), \quad \bar{a}_m = (a_{mj}), \quad 1 \leq j \leq F$$

where $s$ is the cosine of the angle as a measure of similarity, the brackets in the numerator denote the scalar product and $k$ and $m$ are arbitrary but fixed numbers of two variables. Since each axis represents a factor, variables which are close to an axis have a high similarity with the factor and can be used as representative of the factor, i.e. a common property of the data set for a global interpretation.

All data points, i.e. measurement results can also be visualized in the factor space. They can be assigned to the factors and, for example, changes of parameters can be observed by moving the data points in the factor space. This technique can be used to find general tendencies.

#### 3.2. Application of factor analysis to cavitation data

A factor analysis was applied to experimental data from the three measurement cycles. Fig. 2 shows the factor space plot for a measurement of cycle 2 at the point M4 at three different temperatures (10 °C, 20 °C, and 35 °C) in gas-saturated water. What are depicted are the data as points and all variables as vectors whose lengths correspond to the specific variances given in Fig. 2 caption. All variables form a small cluster and it is difficult to separate the effect of the common factors but electrical power, fundamental and rms-sound pressure seem to be slightly separated from sub- and ultraharmonic. Fundamental and rms-sound pressures were redundant and often only one of them was calculated later on.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental conditions of the three measurement cycles.</td>
</tr>
<tr>
<td>Driving voltage (V)</td>
</tr>
<tr>
<td>Cycle one</td>
</tr>
<tr>
<td>Cycle two</td>
</tr>
<tr>
<td>Cycle three</td>
</tr>
</tbody>
</table>
lie in the lower half space, the medium power values in the centre and the high-power values in the upper half space. They “move” in the direction of factor three which is mainly represented by the fundamental. Within the cluster the arrow shows the direction of increasing temperature. It points mainly along the axis of factor two which is indeed represented by the temperature. This can lead to the conclusion that high-power and high-temperature systems are mainly determined by temperature and fundamental pressure.

3.3. Correlation between indicators and parameters

The angle between the variables can quantitatively describe the overlap of variables which is a robust indicator for similarity of the variables. If the angle is small, it can be assumed that the variables have similar causes and they are redundant during measurement.

This can reduce the measurement effort. In the following, several variables were analysed with respect to similarity. As an example Fig. 6 shows the results for the subharmonic and the ultraharmonic. Five different measurements at different measurement positions and under various conditions were analysed. The dark bar gives the similarity s and the grey one the sum of the specific variances. It is clearly seen that the similarity is always high with good reliability indicated by small variances.

Fig. 7 compares the sonochemical with the erosion indicator. Differences between the measurements are obvious. Although expected, a general close relation cannot be concluded and the analysis needs to be done for every application case.
3.4. Factor-based data reduction for description of an arbitrary application case

For a quantitative description, the cavitation outcome or its indicators are to be expressed by measurands or parameters obtained from an investigating analysis. The most obvious approach is to use a fitting procedure of experimental data but it is not known in advance which indicator or parameter is best suited for a description of another one. A factor analysis can help to identify useful describing measurands and parameters and to reduce the number of necessary measurements because only matching variables need to be determined.

An indicator can successfully be described by another one if they have similar causality or a close internal relation. Since a factor analysis searches for common structures, variables, i.e. indicators or parameters with high similarity (small angle between the vectors in the factor space) have a high potential for mutual expression. Fig. 8 shows an example of the measurement of the factor analysis given in Fig. 5. Two variables, the subharmonic and the erosion indicator were depicted in dependence on the noise power which is an easily achievable parameter. To assess the quality of the description, a polynomial fit of the order of 1, 2, or 3 was applied where the software took the best one by optimizing the $r^2$ parameter, which is calculated from the sum of the squares of the distances of the points from the best-fit curve. In the case of similar variables the subharmonic can be fitted with $r^2 = 0.94$ but the more distant erosion only very poorly with $r^2 = 0.09$ as immediately expected from the factor analysis.

In many application cases, however, it is not possible to find an easily accessible measurand with high similarity or the outcome of several effects or indicators needs to be described. In these cases a factor analysis can be used to find a limited number of appropriate indicators and parameters. These representing variables can be combined to a global parameter which can be used for finding quantitative relations for example by fitting. Since the factors represent common properties of the application under investigation it can be concluded that not more than $F$ variables are needed for a description. It is useful to choose variables which are close to the factor axes such as for example noise power, temperature and fundamental in Fig. 5. These are the representing variables $\tilde{a}_{k(n)} = (\tilde{a}_{d(j)})$, where $1 \leq j \leq F$, $1 \leq n \leq F$ and $k(n)$ means $F$ fixed numbers between 1 and $F$, and $\tilde{d}_k = (\tilde{d}_d)$ is the variable to be described. For these variables, weighting factors are calculated from similarities which take into account that not all representing variables are equally relevant for the variable to be described:

$$
\tilde{w} = \begin{pmatrix}
\tilde{w}_1 \\
\vdots \\
\tilde{w}_F
\end{pmatrix} = \begin{pmatrix}
\tilde{a}_{d(1)} \\
\vdots \\
\tilde{a}_{d(F)}
\end{pmatrix}
$$

Fig. 6. Similarity and combined variance of the two variables subharmonic and ultraharmonic in different measurement situations, M3, M4, etc. define the measurements points, where several points are given, measurements at all these points are included. Med power: measurement at medium power for all 7 measurement points.

Fig. 7. Similarity and combined variance of the two variables erosion and chemical indicator in different measurement situations, M3, M4, etc. define the measurements points, where several points are given, measurements at all these points are included. Medium power: measurement at medium power at all 7 measurement points, T35: measurement at 35 °C at all measurement points.

Fig. 8. Description of subharmonic and erosion by noise power for the measurement and factor analysis example of Fig. 5.
of 0.63 which is a good result for technical cavitation applications. The specific variances could amount to within a wide range of numbers, sometimes they exceeded 0.4. Although from a statistical point of view such high values are not acceptable, in the case of the cavitation data a useful conclusion could be obtained. The methods of Section 3.4 could often be successfully applied for variables with high specific variances and no general limit for an acceptance or rejection of a specific variance value was found plausible. To test the influence of the choice of variables, several factor analyses were made with a reduced but varying subset of variables for the same data. In no case did the specific variance of any variable differ more than 5% and the influence of variable choice seems to be of minor importance.

The distribution of variable vectors in the factor space depends on the application conditions of cavitation. Every experimental situation requires its own factor analysis, similarity assessment and a particular decision about describing variables. Within one description the reliability was satisfactory although a quantitative proof goes beyond the scope of this study. The question to be solved in practice is the definition of the limits of an experimental situation within which one factor analysis would be valid. The more measurement values are available, the clearer are the borderlines and the more “stable” is the factor analysis. This remains valid when the range of parameters is large. Thus, both strategies could be successful, if they had a wide range of parameters with a large number of measurements and a possibly higher number of factors or clear, defined measurement conditions with a smaller amount of variability.

The data can be depicted in the factor space by the scores which are calculated from Eq. (2). The spreading and the composition of the point cloud in factor space provide information about whether the factors homogeneously describe the data or whether subsets of data have different allocations to the factors. If data with a particularly fixed parameter form a clearly defined cluster in factor space it can easily be found which factors are concerned or dominant and they may be different from that of other separated clusters. Thus, the balance between the factors may change within one data set giving further insight into tendencies within the data.

Factor analysis allows the reduction of variables which results in a possible reduction of measurement effort. In most of the experimental situations during this work, the description of typical cavitation output indicators was possible with three representing variables. Knowing these relations in advance, only these three parameters or indicators need to be determined for a characterization of the process. In this study no experiments could, however, be carried out to validate this assumption on a long time scale. It seems to be likely that the choice of representing indicators or parameters is quite universal and hardly dependent on time.
Therefore, they can be used for the characterization of cavitation processes, for example, in quality management procedures, for a more reliable manufacturing of goods.

Acknowledgements

This work was supported by the DECHEMA as a member of the AiF in the Industrielle Gemeinschaftsforschung (IGF) program. The authors thank Christoph Kling and Klaus Jenderka for valuable discussions.

References