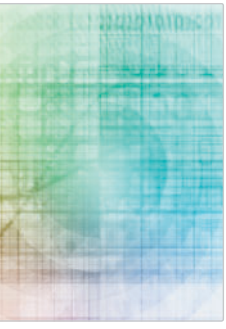


Methods for Identifying Out-of-Trend Data in Analysis of Stability Measurements—Part I: Regression Control Chart

Máté Mihalovits and Sándor Kemény

The regulation of pharmaceutical stability studies still lacks universally accepted techniques regarding the identification of out-of-trend data. Three methods have been suggested for identifying out-of-trend data in pharmaceutical stability studies: the regression control chart, the by-time-point method, and the slope control chart. In Part I of this article series, the regression control chart method is investigated, and an improved approach is suggested. The method is illustrated using realistic data. In Part II, the by-time-point method and the multivariate control chart method will be discussed.

In the course of stability studies performed in the pharmaceutical industry, stability profiles are obtained for batches by measuring the change of certain attributes of the pharmaceutical preparations over time. The samples from the batches are kept at certain conditions (e.g., temperature, humidity) for the time of the study and measured at certain time points. The purpose of stability studies is to establish shelf-life of products. At the registration of the medicine, the method of recording the stability profile is regulated by the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) (1). The same method may be used to monitor ongoing production processes.

Certain variation in data may well be explained by uncertainty. Larger deviations are considered out-of-trend (OOT) results. OOT is to be distinguished from out-of-specification (OOS); the latter means that the result is outside of the allowed range given by the specification. As noted by the Pharmaceutical Research and Manufacturers of America Chemistry, Manufacturing, and Control (PhRMA CMC) Statistics and Stability Expert Teams (2), "The procedures described for detecting OOT results can be viewed as an alarm or alert system, showing that some kind of action is needed. In other words, at each stability time point when a new result is collected, one should determine whether the result is in agreement with what is expected and if not, take the appropriate action."

A stability result at a single time-point may not follow the expected trend in two ways: either in comparison with earlier batches or with respect to previous results collected from the observed batch. The OOT phenomenon is to be detected as soon as possible, because these results indicate errors either in the process of production or in the analytical measurement. If an OOT point is detected, one should find the source of error and fix it. Further actions may be required if the problem is found to be with the production process.

Submitted: March 22, 2017
Accepted: June 27, 2017

The PhRMA CMC Statistics and Stability Expert Teams began a dialogue toward building up a system, supported by statistical calculations, to identify OOT results in stability measurements (2). The methods suggested in this 2003 seminal paper were supported by example calculations in an article by Torbovska and Trajkovic-Jolevska (3). Also, statistically less efficient, but more user-friendly methods were suggested by the PhRMA CMC Statistics and Stability Expert Teams in a later paper published in 2005 (4). In this article, the authors reflect on the regression control chart method, described in reference 2 and used in reference 3, and attempt to refine the method. The amount of active substance (assay) of the pharmaceutical product was used as an example, but the method presented may be used for any other attributes that follow a linear trend over time.

Detecting OOT results

Three methods are outlined and illustrated by the PhRMA CMC Statistics and Stability Expert Teams (2):

- Regression control chart method
- By-time-point method
- Slope control chart method.

In the regression control chart method, data are compared within a batch, while in the by-time-point method, data are compared to historical data of earlier batches at the same time point. In the slope control chart method, a feature (the slope) of the stability profile of the observed batch is compared to that of earlier batches.

The technique of decision is the comparison of a value to a statistical interval. Three kinds of statistical intervals are used: confidence, prediction, and tolerance. Users are sometimes confused about which interval should be used in certain situations (5).

The confidence interval is the range in which a parameter (e.g., the expected value) occurs with a certain probability. The prediction interval is the range in which a future observation is found with some specified probability. The tolerance range contains a specified part of future observations (content) at a certain probability (confidence). There is a connection between the latter two intervals: a prediction interval to contain a single future observation with 1-alpha probability is equivalent to a tolerance interval to contain, on average, 1-alpha proportion of the population (6). Calculating the confidence and prediction range is straightforward; it is found in standard textbooks and taught in basic statistics courses. The tolerance range is more difficult to calculate, and sometimes, only approximate methods are available.

The interval is characterized by its acceptance probability. For quality engineering, usually a 99.73% confidence level is used in control charts. This may not be the proper choice in OOT analysis of stability studies. The higher the confidence level, the lower the chance of identifying data as OOT while it is actually not the case (i.e., type I error, "false positive alarm"), while the chance of accepting an OOT data as non-OOT is increased (i.e., type II error, "false negative case"). This latter probability is not explicitly stated, however. In the pharmaceutical industry (and also from a regulatory viewpoint), type II error (not realizing that the data is actually OOT) presents a more serious threat, but such errors can be avoided if the confidence level is decreased. The 95% confidence level is used in this paper.

Identifying OOT data by regression control chart

The peculiarity of identifying OOT phenomenon with respect to stability studies is that there is a trend in data anyway, because the attributes change with time due to chemical and physical changes, and OOT is to be stated in relation to this obvious trend. In this method, a straight line model is used, however, one should justify the assumption of linearity of a stability profile.

Control charts. Control charts (7) are widely used in the industry to monitor the production process. The regular usage of the charts can be separated into two phases (8). In the first phase (Phase I), parameters of the distribution (typically mean and standard deviation) of the reference data are estimated. In the second phase (Phase II), certain statistics (e.g., sample mean, range, etc.) of the observed sample are plotted and examined. In the example used by the authors, the observed sample contains only a single datapoint, thus the statistic to be plotted is the individual value (individuals control chart). The distribution parameters obtained in Phase I are used to construct lower and upper control limits (i.e., lower and upper quantiles) for the statistic (i.e., the individual value) and added to the chart. The interval between the control limits of the individuals control chart contains the allowable value of the new data of the observed process with a certain probability, when the process is in control. Typically, this probability is 0.9973, which conforms to the $\pm 3\sigma$ range (3 is taken from z-table). This range is valid only if the reference sample size in Phase I is theoretically infinite, because the method (calculation of quantiles) assumes the perfect knowledge of the distribution. If the distribution is not perfectly known (e.g., the parameters are estimated only), the 0.99865 and 0.00135 quantiles are at best approximate only. This method is proposed by Shewhart for industrial processes and proved to be useful there for distinguishing assignable and common causes, thus they serve as statistical process monitoring (9). The typical application requires at least 25 samples in the reference set. For the authors' study, this method is valid only if the sample size in the reference set of Phase I is large enough, then the 0.9973 probability for the $\pm 3\sigma$ range is reached as a limit. This approach will be referred to as the Shewhart method in this paper. In the current situation, however, the Shewhart method is not considered reliable because the reference data set is small.

In the case where the sample size is small, one should consider the uncertainty of expected value and variance. The earlier is taken into account by including the term $\sigma/\sqrt{n}$

into the standard deviation, while the latter is taken into account by using *t*-distribution instead of *z*. Thus, the width of the valid control limits in the authors' typical case (small reference set) are $\pm ts\sqrt{(1+\frac{1}{n})}$, which is the prediction interval.

The prediction limits are used in Phase II of the individuals control chart to check whether the stability is further maintained. If a point is found to be OOT, one may choose from two routes, depending on whether data analysis is in real time or not. If the data analysis is performed in real time, the OOT suspicion, if arisen, is followed by the repetition of analytical measurement. If the non-OOT nature is justified (by re-measurement), the stability study is continued, neglecting the point. If the OOT nature is justified, a corrective action follows the detection, and thus, the original process is ended.

The other route is followed if the analysis of data is performed not in real time but upon collecting more data. Here, the original process is continued, and the point that is later found to be OOT cannot be re-measured and should be removed from the data set. If the OOT point is not removed from further analysis, the acceptance range will be widened, hindering detection of further OOT points. If a point is found to be non-OOT, the point is added to the historical data set, new control limits are calculated, and a new control chart is made for the next observed data.

Regression control chart. The regression control chart is used to monitor processes where there is a systematic change over time (10). The "parameters" of the chart are a regression line (i.e., an expected value changing with time) constructed by the previous data from Phase I and the variance of residuals, and both are considered as known in the Shewhart method. If the new point is found to be within the control limits, the point is accepted; if not, the point is rejected and declared as OOT.

Use of regression control chart method, original proposal

The PhRMA CMC Statistics and Stability Expert Teams (2) suggest fitting a regression line to historical data of the observed batch and extrapolating the line to the time point of the new data. Control limits are calculated for the expected value by *expected value*±*ks* at the new time point. The *k* is taken from a table of normal distribution at a desired significance level. The residual standard error (*s*) is calculated either by only the regression line of the observed batch, or by regression lines of historical batches. In the latter case, a common slope is assumed, and different intercepts are allowed for the historical batches. If the observed point is out of the interval, it is identified as OOT. It is also mentioned in the paper (2) that using prediction or tolerance interval is a better but more complex method. The limits calculated there (and in reference 2) using the *expected value*±*ks* formula are called the Shewhart limits (SL) here.

Suggested use of regression control chart method

The prediction interval with *t*-distribution is to be used if one asks about a single new observation, because both the expected value and the variance are unknown and estimated only from a small sample. The method requires points in the observed batch as reference data to construct the first regression line. One should decide how many points from the beginning of the observed batch should be considered as the reference data set. Moreover, one should assume that those reference data are non-OOT. Here, the first three points of the observed batch are considered to be the reference data, and the fourth point is the first to be analyzed by the method.

The prediction interval for the measured variable (y) at the new time point (x^*) is calculated in **Equation 1**:

$$\hat{Y} - t_{\alpha/2} s_{y^*-\hat{Y}} < y^* < \hat{Y} + t_{\alpha/2} s_{y^*-\hat{Y}} \qquad \text{[Eq. 1]}$$

where $\hat{Y}$ is the predicted value of y at x^*, $t_{\alpha/2}$ is the one-sided upper critical t value at $\alpha/2$ one-sided level (with $(n-2)$ degrees of freedom, where n is the number of points used to construct the regression line), y^* is the new measured value at x^* time point, and $s_{y^*-\hat{Y}}$ is the sample standard deviation calculated for the $y^*-\hat{Y}$ variable. The right side of the inequality is the upper prediction limit (UPL), while the left side is the lower prediction limit (LPL). If the inequality is satisfied, that is the y* is within the limits, the data is accepted, otherwise it is OOT.

$\hat{Y}$ is calculated through the extrapolation of the regression line to the x^* new time point by **Equation 2**:

$$\hat{Y} = b_0 + b_1 x^* \qquad \text{[Eq. 2]}$$

where b_0 and b_1 are the intercept and slope of the regression line (created with reference data), respectively.

$s_{y^*-\hat{Y}}$ is calculated by **Equation 3**:

$$s_{y^*-\hat{Y}} = s_r \sqrt{\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{j=1}^{k}(x_j - \bar{x})^2}\right)} \qquad \text{[Eq. 3]}$$

where s_r is the residual standard deviation, n is the number of points used to construct the regression line, and $\bar{x}$ is the mean of the x_j values (time points) of data (without x^*) used to estimate the regression line. The third term in the bracket is the reason of constriction of the band in the middle.

As the PhRMA CMC Statistics and Stability Expert Teams highlights (2), the residual standard error may be calculated using earlier batches as well, not only from the recent one in study. The use of earlier batches may follow several different routes. The authors of reference 2 used parallel lines (i.e., the common slope for all historical batches). The authors of this article are not convinced about the usefulness of this assumption, however. If common slope is forced for historical batches, the fit is deteriorated. This situation would lead to greater residual error than would be achieved using the

Table I: Stability data set.

Amount of active substance (%)									
Time (month)	Batch I	Batch II	Batch III	Batch IV	Batch V	Batch VI	Batch VII	Batch VIII	Batch IX
0	97.6	98.4	100.9	98.7	98.8	100.5	100.3	101.5	100.9
3	97.7	99.4	98.2	95.8	97.5	96.5	99.7	100.1	97.3
6	97.7	96.2	98.5	96.7	97.5	96	98.6	99.5	97.7
9	96.9	97.3	94.6	97.5	98.9	96.3	98.3	99.6	98.4
12	94	95.3	96.9	94.7	97.5	98.3	96.8	98.3	96.5
18	96.5	94.9	96.3	93.7	96.5	94.1	96.7	95.2	99.5
24	96	97.5	95.8	93.1	96	92.5	96.3	97.1	96
36	92.1	92.7	92.3	91.3	92	89.5	93.9	93.8	93.7

best linear fits (letting different slopes). On the other hand, if separate lines are fitted to historical batches, the residual standard errors obtained may be pooled. The pooled error has higher degrees of freedom than that obtained from a single batch, thus, it is a better estimate to the variance, leading to better estimated control limits. If every group has the same number of data, the pooled sample variance is the arithmetic average of the individual sample variances:

$$s^2_{p(res)} = \frac{\sum s_i^2}{p} \qquad \text{[Eq. 4]}$$

where $s^2_{p(res)}$ is the pooled sample variance, ${s_i}^2$ is the squared standard deviation of residuals for the *ith* historical batch, and p_i is the number of historical batches involved.

For this purpose, regression lines were fit to each historical batch and residual standard errors (s_i) were calculated. The pooling is justified only if the variation of data of different batches has the same variance, which can be tested by the Bartlett and Levene tests for example. For the authors' data, the hypothesis of homogeneity of variances is accepted (details of the tests are not shown here).

For calculation of the prediction interval with pooled residual standard error, in **Equation 3**, is substituted for s_r. The degrees of freedom of $s^2_{p(res)}$ is $p(n-2)$, thus a different $t_{\alpha/2}$ should be taken from t-table when calculating the control limits.

For illustration purposes along with prediction limits, the Shewhart limits (α=0.05), confidence limits (α=0.05), and tolerance limits (P=0.99 as content, γ=0.95 as confidence level) are calculated as well. None of these are appropriate to use in the current situation (where only a single new

MARTIN BARRAUD/OJO IMAGES/GETTY IMAGES

Table II: Prediction limits, Shewhart limits, confidence limit, tolerance limits, regression control chart method using pooled standard deviation $s_{p(res)}$. LPL is lower prediction limit. UPL is upper prediction limit. LSL is lower Shewhart limit. USL is upper Shewhart limit. LCL is lower confidence limit. UCL is upper confidence limit. LTL is lower tolerance limit. UTL is upper tolerance limit.

Time (months)	$S_{p(res)}$	Observed data	LPL	UPL	LSL	USL	LCL	UCL	LTL	UTL
0	-	100.9	-	-	-	-	-	-	-	-
3	-	97.3	-	-	-	-	-	-	-	-
6	-	97.7	-	-	-	-	-	-	-	-
9	1.2	98.4	91.0	99.8	93.1	97.8	91.8	99.1	89.9	101.0
12	1.2	96.5	93.0	100.6	91.5	96.2	93.8	99.8	91.7	101.9
18	1.2	99.5	91.0	99.1	88.3	93.0	91.8	98.3	89.8	100.4
24	1.2	96.0	88.3	98.8	85.1	89.8	88.8	98.2	92.6	102.7
36	1.2	93.7	89.3	97.9	78.7	83.4	90.1	97.2	90.3	100.7

data point [y^*] is observed), except for the prediction limits. These prediction limits are discussed in detail below.

Calculation of the Shewhart limits:

$$SL = \hat{y}^* \pm z_{\alpha/2}\,\sigma_r \qquad \text{[Eq. 5]}$$

where $z_{\alpha/2}$ is the critical value of standard normal distribution at two-sided $\alpha/2$ level, and σ_r is the square root of variance of residuals, equal to s_r. As the known variance is assumed, there is no room for pooling. One may use the square root of the pooled standard deviation, however, as a substitute of the σ_r. The latter is falsely assumed to be known because of the small sample size as explained earlier. When the square root of the pooled estimated variance is substituted, σ_r is taken to be equal to $S_{p(res)}$.

Confidence limits are calculated in **Equation 6**:

$$CL = \hat{y}^* \pm t_{\alpha/2}\, s_r/\sqrt{n} \qquad \text{[Eq. 6]}$$

For the calculations with pooled standard deviation, s_r is substituted with $S_{p(res)}$ in **Equation 6** and degrees of freedom of $s^2_{p(res)}$ is used to obtain the t-score.

Tolerance limits can be calculated by **Equation 7**:

$$TL = \hat{y}^* \pm k_1 s_r \qquad \text{[Eq. 7]}$$

Where k_1 is calculated by **Equation 8** (11):

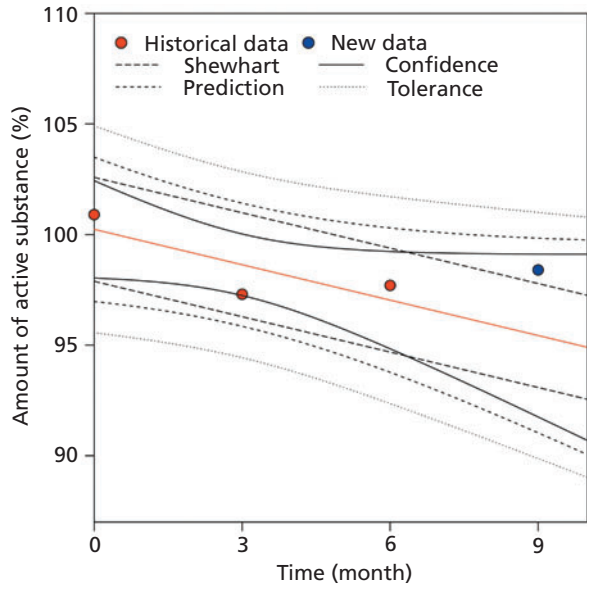
$$k_1 = \sqrt{\frac{\nu \chi^2_{P,1}(\frac{1}{n'})}{\chi^2_{1-\gamma,\nu}}} \qquad \text{[Eq. 8]}$$

where n' is the effective number of observations, ν is the degrees of freedom of s_r^2 (n-2 actually), $\chi^2_{1-\gamma,\nu}$ is the critical value of the chi-square distribution at a one-sided 1-γ level with ν degrees of freedom, $\chi^2_{P,1}(\delta)$ is the critical value of non-central-chi-square distribution at one-sided P level with ν degrees of freedom, and δ is the argument in the non-central-chi-square function. Actually, its value here is 1/n'.

n' can be calculated by **Equation 9** (12):

$$n' = \frac{n\sum(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2 + n(x^* - \bar{x})^2} \qquad \text{[Eq. 9]}$$

Figure 1: Statistical intervals of regression control chart.



This equation is simplified to **Equation 10**:

$$n' = \frac{s_r^2}{s_{\hat{Y}}^2} \qquad \text{[Eq. 10]}$$

Where

$$s_{\hat{Y}}^2 = s_r \sqrt{\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{j=1}^{k}(x_j - \bar{x})^2}\right)} \qquad \text{[Eq. 11]}$$

For the calculations using pooled standard deviation, s_r is substituted with $S_{p(res)}$ in **Equation 10** and **Equation 11** whenever a new effective number (n') of observations is to be obtained. The proper substitution is performed for degrees of freedom as well in **Equation 8**. Also, in **Equation 8**, a new $\chi^2_{P,1}(\frac{1}{n'})$ is calculated with the new n', which is calculated by **Equation 10**.

Example: Regression control chart

The data in **Table I** are from a realistic stability data set from a pharmaceutical manufacturing process (details are not given due to proprietary reason). The amount of active substance of the medicine was recorded over 36 months.

Batch IX is the object of these calculations; the rest of the columns refer to historical batches. The regression line was created from the first three points of the observed batch; it was then extrapolated to the new point (i.e., the ninth month), and the prediction interval (LPL, UPL) is calculated along with the other intervals that should not be used to identify OOT points (see **Table II**).

The limits calculated with pooled standard deviation are given in **Table II**. Using **Equation 4**, the calculated s_(p(res))^2 is 1.438. **Figure 1** shows the fitted line, the data points up to the nine-month time point, the 95% Shewhart interval (SI), the 95% confidence interval (CI), the 95% prediction interval (PI), and the 95% tolerance interval (TI) containing 99% of future observations.

More Peer-Reviewed Research

For more peer-reviewed articles, go to PharmTech.com to read the following:

How to Monitor HPMC Concentration Through Conductivity Measurement
The authors demonstrate that a proper measure of HPMC solution concentration is its electrical conductivity rather than its viscosity.
www.pharmtech.com/how-monitor-hpmc-concentration-through-conductivity-measurement

Detection of Lumps in Powder Blends by Inline NIR
This study shows that the presence of API lumps can be detected by inline NIR, and elaborates on why NIR sensor dimensions and actual measured sample volume by the NIR sensor are important variables for adequate interpretation of obtained results.
www.pharmtech.com/detection-lumps-powder-blends-inline-nir

Analysis of Total and Transferrin-Bound Iron from Serum Samples
Incorporating structural constraints into pharmacokinetic studies of iron sucrose formulations.
www.pharmtech.com/analysis-total-and-transferrin-bound-iron-serum-samples

A Forensic Microscopy Approach for Identifying Subvisible Particulates in a Sterile Ophthalmic Solution
In this study, the authors describe a forensic microscopy approach to characterize particles that were visually observed during stressed stability testing of an ophthalmic solution formulation.
www.pharmtech.com/forensic-microscopy-approach-identifying-subvisible-particulates-sterile-ophthalmic-solution

FIGURE 1 IS COURTESY OF THE AUTHORS.

The confidence interval is very narrow in the range of historical data (i.e., for the first three points); it allows an even tighter interval than the Shewhart limits. The uncertainty of the parameters of the regression line is the reason behind the bend in the limits. The farther the new point is from the historical data, the wider the interval is given. As the Shewhart method assumes known parameters of the regression line, the band of that is not bent; the limits are parallel to the regression line.

Having accepted the nine-month time point, the calculations are repeated, but now the line is fitted to both the initial three data points and the accepted data (nine-month time point) while the new data at the 12-month time point is observed. Having continued the calculation practice, the 18-month time point is found to be out of the prediction interval, as 99.5 > UPL; therefore, it is declared as OOT. The outlier point is then left out from the examination of later points if the study is to be continued, but intervention is clearly required. The grey coloring indicates that the observed data in the row of the grey pair of boxes is OOT considering the limits specified in the columns. The mistakenly used confidence interval would give the same qualitative result in the current situation, while with tolerance interval approach, all the points would be accepted, and with the Shewhart interval method, every point would be OOT.

Conclusion

Having realized that the regulation of pharmaceutical stability studies is still lacking universally accepted techniques for the identification of OOT data, three methods were suggested earlier by PhRMA CMC group. The regression control chart method discussed in this paper is a possible way to identify OOT data within a batch. To be statistically more rigorous, the earlier suggested approach is improved. As only a small sample is available, the use of prediction interval with t-distribution is justified. Incorrectly using the tolerance interval instead would unnecessarily widen the acceptance interval, while the unjustified use of the Shewhart procedure would result in too narrow an interval.

References

1. ICH, Q1A(R2) *Stability Testing of New Drug Substances and Products*, Step 4 version (2003).
2. PhRMA CMC Statistics and Stability Expert Teams, *Pharm. Tech.,* 27 (4) 38–52 (2003).
3. A. Torbovska and S. Trajkovic-Jolevska, *Pharm. Tech. Eur,* 37 (6) (2013).
4. PhRMA CMC Statistics and Stability Expert Teams, *Pharm. Tech.,* 29 (10) (2005).
5. S. De Gryze, I. Langhans, and M. Vandebroek, *Chemometrics and Intelligent Laboratory Systems,* 87 (2) 147–154 (2007).
6. F. Proschan, *Journal of American Statistical Association,* 48 (263) 550–564 (1953).
7. W.A. Jensen et al., *Journal of Quality Technology,* 38 (4) 349–364 (2006)
8. T.P. Ryan, *Statistical Methods For Quality Improvement* (John Wiley & Sons, NY, 3rd ed., 1989), pp. 89-95.
9. W.A. Shewhart, *Economic control of quality of manufactured product.* (ASQ Quality Press, Milwauke, Wisconsin, 1931)
10. B.J. Mandel, *Journal of Quality Technology,* 1 (1) 1–9 (1969)
11. K. Krishnamoorthy, T. Mathew, *Statistical Tolerance Regions: Theory, Applications, and Computation* (John Wiley & Sons, New Jersey, 2009), pp. 70.
12. W.A. Wallis, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, Berkeley, Calif., 1951), pp. 43–51. **PT**

Máté Mihalovits is a PhD student, mihalovits@mail.bme.hu; and **Sándor Kemény*** is an emeritus professor, Tel.: +36 309936307, kemeny@mail.bme.hu, both at Budapest University of Technology and Economics, Faculty of Chemical Technology and Biotechnology, Department of Chemical and Environmental Process Engineering Hungary, 1111. Budapest, Műegyetem rakpart 3.

* To whom all correspondence should be addressed.