ZENTRUM ANGEWANDTE CHEMIE
Institut für Technische Chemie
Institut für Technische Chemie, Callinstr. 3, 30167 Hannover

Correlation analysis of fermentation data

P.F.O. Lindner, University of Hannover/D; K. Hellmuth, Chr. Hansen GmbH, Nienburg/D;
B. Hitzmann, University of Hannover/D

Introduction

On-line and off-line data from a fermentation on industrial scale have been collected over a long period of time (66 processes). Among the data are measurements of exhaust gas concentration (O_2 and CO_2), substrate and product concentration, temperature, pressure, power input (stirrer), viscosity and others. The data was analysed with statistical methods. The main focus of this work was the evaluation by correlation analysis. The aim was to find connections between the process variables, and dependencies of the overall yield and the losses during down-stream on process parameters. With the aid of these results an optimization of process yield and stability was desired.

Methods

Before the collected data can be used in a correlation analysis a preprocessing step is necessary. The goal of this step was to reduce the amount of data but keep as much information from each measurement as possible. For each process variable a different preprocessing method had to be developed. One simple approach, that involves the calculation of mean and maximum values as well as linear regression is shown in Fig.1. The result of data preprocessing is the matrix M containing 63 characteristic values for each process.

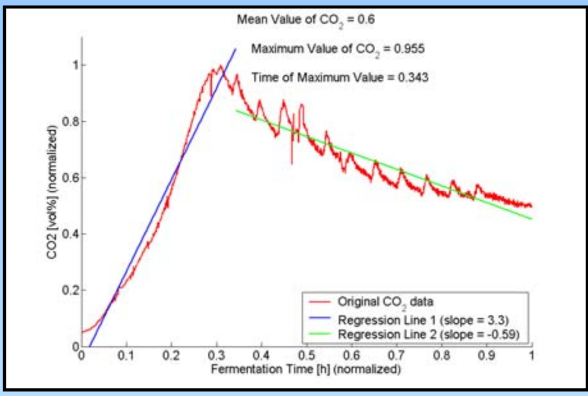

Figure1. Data preprocessing of a CO_2 measurement reduces over 1500 data points to 5 characteristic values: mean, max, time of max and two slopes.

$$r_{xy} = \frac{\mathrm{cov}(x,y)}{std(x)\,std(y)}$$ Correlation coefficient of x and y

$$\mathrm{cov}(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)}$$ Covariance of x and y

$$std(x) = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$$ Standard deviation of x

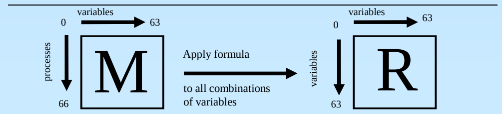

Figure2. The correlation matrix R is calculated from M by applying the formula to each combination of variables.

Using this matrix the correlation coefficients for each combination of two variables were calculated (see Fig.2). Lines and columns in M that contain missing values in one of the two variables were eliminated before each calculation.

The data preprocessing, correlation analysis and visualization was done with Matlab (Version 6.5).

Results

A small portion of the correlation matrix R is given in Table 1. It shows that the losses of product during down-stream do not correlate with any of the process parameters. The product yield correlates with the substrate feed, citric acid production and the broth viscosity. Weaker correlations can be found with pO2, exhaust gas concentration, stirrer speed and the substrate feeding profile. Many variables that correlate with the yield are related to cell growth and activity and are not eligible for a process optimization. In this context the most promising starting-point for an optimization is broth viscosity. Because the viscosity itself correlated to variables derived from the shape of the substrate feed (number, height and length, and time between the feed pulses) an optimization of the yield by controlling the broth viscosity throught a modified substrate feed seems to be feasible. Another option to get an increase in product yield is to increase the power input to the fermentation broth (stirrer speed).

	Product yield	Losses	mean concen-tration of CO_2 (exhaust)	Total substrate feed	Total ammonia feed	Slope of ammonia feed	Velocity of citric acid production	mean broth viscosity
Yield	1	-0,25	0,34	0,63	0,25	0,23	0,64	-0,68
Losses	-0,25	1	-0,3	-0,07	-0,11	-0,15	0,13	-0,2
CO_2 (mean)	0,34	-0,3	1	0,67	0,55	0,65	-0,03	-0,49
CO_2 (max)	0,21	-0,17	0,9	0,65	0,53	0,59	-0,07	-0,56
Total substrate feed	0,63	-0,07	0,67	1	0,63	0,6	0,39	-0,67
Slope of substrate feed	0,35	-0,14	0,83	0,68	0,65	0,8	0,3	-0,38
Total ammonia feed	0,25	-0,11	0,55	0,63	1	0,96	0,29	-0,33
Airflow (mean)	-0,09	0,23	-0,3	-0,18	-0,2	-0,16	-0,15	-0,25
Pressure (mean)	-0,07	0,31	-0,31	-0,26	-0,41	-0,35	-0,09	0,03
pH (mean)	0,03	0,17	0,17	0,34	0,25	0,25	0,32	0,15
pH (max)	-0,38	-0,02	-0,22	-0,34	-0,21	-0,19	-0,31	0,62
pO_2 (mean)	-0,47	0,23	-0,26	-0,62	-0,79	-0,61	-0,41	0,54
Temperature (mean)	0,02	-0,07	0,39	0,32	0,65	0,76	0,38	-0,05
Stirrer Speed (mean)	0,37	0,25	0,08	0,26	-0,21	-0,16	0,39	-0,51
Viscosity (mean)	-0,68	-0,2	-0,49	-0,67	-0,33	-0,27	-0,51	1
Viscosity (max)	-0,72	-0,41	-0,48	-0,66	-0,18	-0,15	-0,58	0,73
Citric Acid (max)	0,68	0,14	-0,06	0,39	0,18	0,18	0,94	-0,39
Velocity (Citric Acid)	0,64	0,13	-0,03	0,39	0,29	0,32	1	-0,51
Length of feed pulse (mean)	-0,44	0,12	-0,78	-0,75	-0,65	-0,77	-0,38	0,41
Height of feed pulse (mean)	0,14	0,07	0,22	0,08	0,19	0,21	0,45	-0,38
Number of feed pulses	0,47	-0,11	0,44	0,82	0,46	0,42	0,22	-0,51
Time btw. pulse 1 and 2	-0,31	-0,18	-0,04	-0,22	0,06	0,11	-0,46	0,52

Table1. Excerpt from the correlation matrix R.

While increasing the stirrer speed did not give an increase in product yield in the fermentation process, a modified substrate feed to control broth viscosity improved product yield significantly.