Process Optimization: With Data or Without?

The chemical and pharmaceutical industry must increase continuously the quality of its products and its productivity. This requires a regular process optimization. For this purpose, the methods of industrial statistics constitute proven tools because they allow gaining a thorough comprehension of the processes out of the corresponding data. However, which method can be recommended depending on the amount of data available?

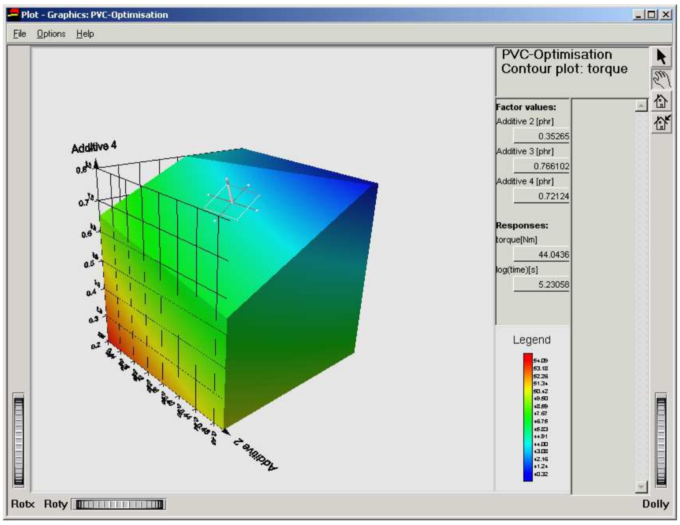
When the structure of a process is defined, its parameters “only need” to be ideally adjusted to get the process optimized. Common response variables are then the yield and some quality characteristics, for example the pureness of the final product. The parameters considered – the so-called factors – are mostly temperatures, pressure values, operation durations, used amounts… Occasionally, categorical factors appear as well, like the centrifuge used or the vendor of a raw material.

The search for the ideal combination of the parameters decomposes into two phases. Firstly, one identifies those factors which, alone or considering their interactions, have a significant influence on e.g. the yield. The actual optimization, which only starts after this screening phase, determines the best value for each of these important factors.

No data available

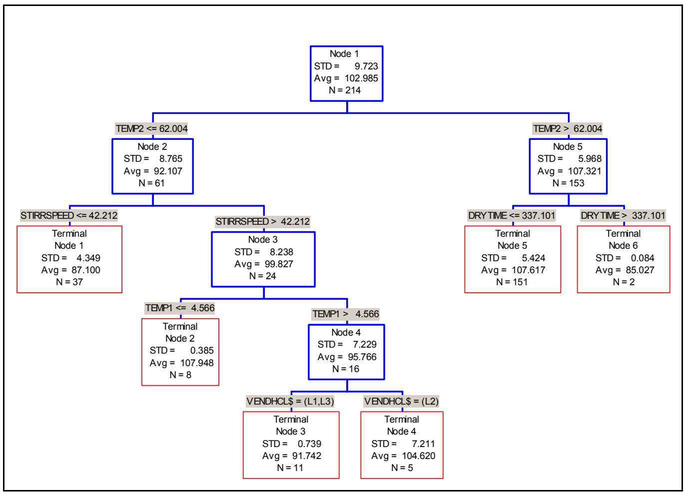
If no process data is available yet, for instance at the beginning of the development, Statistical Design of Experiments is the technique of choice because it ensures an effective data collection for process optimization. Precisely, it generates on the basis of the factors and their variation ranges a minimal list of optimally distributed experimental points, so that a reliable calculation of the ideal process settings out of the test data becomes possible.

In one of numerous applications, this method was used to determine the additive composition that resulted in optimal torque values when kneading a PVC mixture. If required, particularities of the problem setting can then be considered, amongst others by means of restrictions like “the total amount of two given additives does not exceed a preset limit”.



1: An axis is assigned to each additive. The dimension of the response variable is represented by the colour scale.

The practical application of Design of Experiments is significantly supported by powerful software solutions. For example, STAVEX suitably guides the user from the problem definition to the result analysis over the determination of the experimental design. Even if preliminary experiments have been performed without having recourse to an experimental design, it is possible to use them by extending the existing experiment list to a suitable design. Furthermore, numerous 2D to 4D graphical representations make sure that the results can be interpreted easily. For the example above, Figure 1 shows how the torque behaves depending on the amounts of the three additives.



2: The tree from the CART analysis gives insight into the process.

As well in case of an information flood

On the contrary to the situation just discussed, huge amounts of data, which are too seldom analyzed, are generated in production environments. Even if they are not perfect due to errors, such data graveyards hide core information for process optimization, for instance if the yield fluctuates too intensely from batch to batch. Modern data mining methods like classification and regression trees enable to easily extract from the data flood the essential dependencies between the response variables and the factors.

The functioning of this technique can be illustrated on the basis of a yield optimization example for a process which consists of a two-phase reaction, a centrifugation and a drying. Firstly, the factor whose variation best explains the fluctuation of the yield is determined: in Figure 2 it is the temperature of the second reaction phase. The threshold of this temperature – here 62°C – which provides the clearest separation between “high yield” and “low yield” batches is then identified. In this way, the data set gets divided into two parts, on which the same principle is applied. The connections in the resulting tree show, amongst others, that the best yield values are achieved when the temperature of the second reaction phase is below 62°C, the stirring speed above 42 min^{-1} and the temperature of the first reaction phase less than 4.5°C.

User-friendly software tools are indispensable for the implementation of this technique. Here the package CART combines computational efficiency, clear visualization (see Figure 2) and flexibility. Consequently, even huge amounts of process data can be analyzed without effort.

Conclusion

Independently of the amount of process data, statistical methods can greatly contribute to process optimization. Who nevertheless believes that statistics can only be applied with much effort or by experts, unintentionally reduces his own competitiveness.

Author: Dr. Philippe Solot, AICOS Technologies Ltd.

AICOS Technologies Ltd. Tel.: +41 61 686 98 77 Fax: +41 61 686 98 88 E-mail: info @ aicos.com Web: www.aicos.com