DIGITAL PRODUCT DEVELOPMENT AT BASF

1 *Laboratory employee tests the wetting properties of a silicon wafer.*

One of the biggest challenges in the chemical industry for product developers is to spend the least possible effort to come up with a low-cost chemical compound that meets certain cost and quality properties.

Problem description, sample project, and software development

An example is the production of surface coatings. Various properties are desirable depending on the area of use of the coating. In addition to surface protection, for example, certain smoothness or optical properties may be required. Frequently, the objectives are competing, so that a suitable compromise must be found.

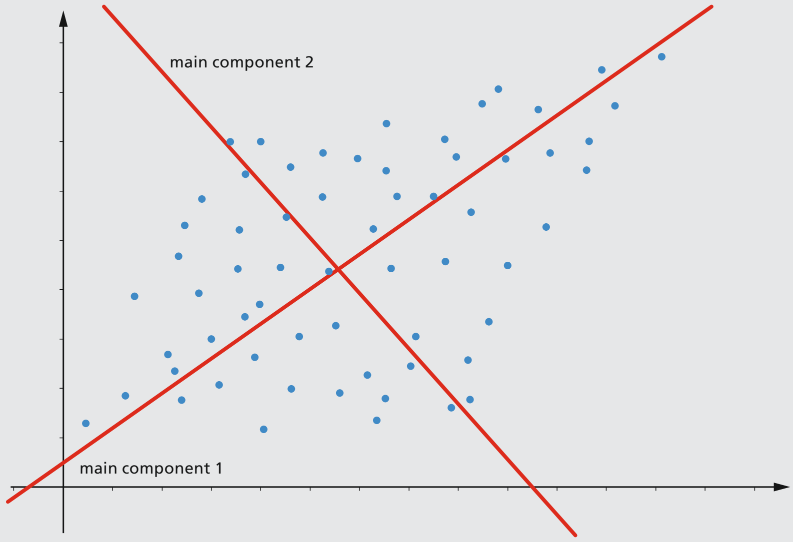
Below, we present a typical workflow for a new research project: First, one defines the objective functions and design variables before starting the first series of experiments. Based on the initial results, one then generates a mathematical model to predict the appropriate design specifications. These are used to start a new series of experiments and this process continues until a satisfactory compound is found.

Our department has developed a software tool to assist the chemists throughout this process. It starts with the analysis and visualization of the data, continues with the modeling of the individual target variables and supports in finding best compromises and in planning of a new experiment.

Machine learning and modeling

The mathematical model chosen to describe the desired functions depending on the design variables is of major significance. The processes in chemical manufacturing are often highly complicated and difficult to model. Together with BASF SE, we are developing a tool to address this issue using machine learning methods. A particularly complex challenge is selecting a suitable model, as this is essential for the quality of the optimized solution.

Various methods are used to measure the suitability of the model, such as cross-validation. This method uses part of the data for training and the remaining part for the validation of the model. For example, it is used in selecting the components for linear regression models in order to avoid over-adaptation of the data.



2

Specifically, it is about improving model quality by filtering out the low-impact components that are random in nature. However, in very few cases do purely data-driven models lead to the goal. That is why we include the user's expert knowledge in the modeling process.

Optimization and experimental design

As mentioned above, a key element is the planning of new experiments. This activity is often time consuming and costly, which makes efficient design of experiments even more important. Our tool helps to reduce the number of attempts required and saves valuable resources for the user.

The problem is solved as follows: The user navigates within a range towards the most attractive target functions. This requires a multi-criteria optimization to be performed in advance. The new experiments are planned within the boundaries of this range. The tool also supports forward planning based on the model, which means the user can directly test particularly promising individual recipes. The user also receives information about the uncertainty of the prediction in the form of confidence intervals.

Web architecture in a big data system

The software is implemented as a web solution with a modern database, which is able to manage a large amount of data. By doing so, a simple and computer-independent use of the tool on mobile devices such as tablets is ensured. At the same time, team productivity is increased because all data are stored centrally and all users always have access to the latest version.

2 *The figure shows a data set with two principal components in a two-dimensional space. Principal component analysis finds the directions which best explain the variance of the data.*