

With data-based models and design of experiments towards successful products - Concept of the product design workbench

Thomas Mrziglod*, Arne Ohrenberg
Bayer Technology Services GmbH
D-51368 Leverkusen, Germany

Abstract

To systematically support the industrial product development process, the so called product design workbench has been developed. The integrated software system supports the typical workflow consisting of experiment design, data analysis, development of data-based models and formulation optimization. Thereby, specifically adopted artificial neural networks or hybrid models can be used in order to precisely describe the complex property formulation relationship.

Keywords: product development, data-based models, artificial neural networks, hybrid models, design of experiments, problem-specific models, database inquiry

1. Introduction

The increasing complexity of production processes as well as innovation and cost pressures pose new challenges to modern product development. The integration of data acquisition, data analysis, modelling and optimization as well as tailored methods for the design of experiments is a central market requirement. However, all integration efforts are lost, if the individual methods cannot cope with the ever-increasing complexity of real-life product development processes.

Hidden information contained in available data treasures can be extracted systematically into data-based models with the help of artificial neural networks. Usually, models with different complexities are necessary to attain an accurate solution depending on the underlying product design problem. Especially for this class of problems it is important to use the appropriate subset of data. Here product development is understood in the sense of finding the right formulation, i.e. the optimal intermediates and their concentration ranges as well as suitable process parameters, to satisfy a given property profile. Possible objectives are to replace expensive components by cheaper ingredients without changing the main product properties, to meet modified customer demands, or to develop completely new products.

* Author to whom correspondence should be addressed: Thomas.Mrziglod@bayertechnology.com

To systematically satisfy the resulting difficulties of an industrial product development process, Bayer Technology Services has designed the so called product design workbench.

2. Concept

The Product Design Workbench is a flexible, integrated software system basically supporting the steps design of experiments, data analysis, model creation and formulation optimization. Hence overall the system configuration and the individual module functionalities can be easily tailored to the specific requirements of the application context.

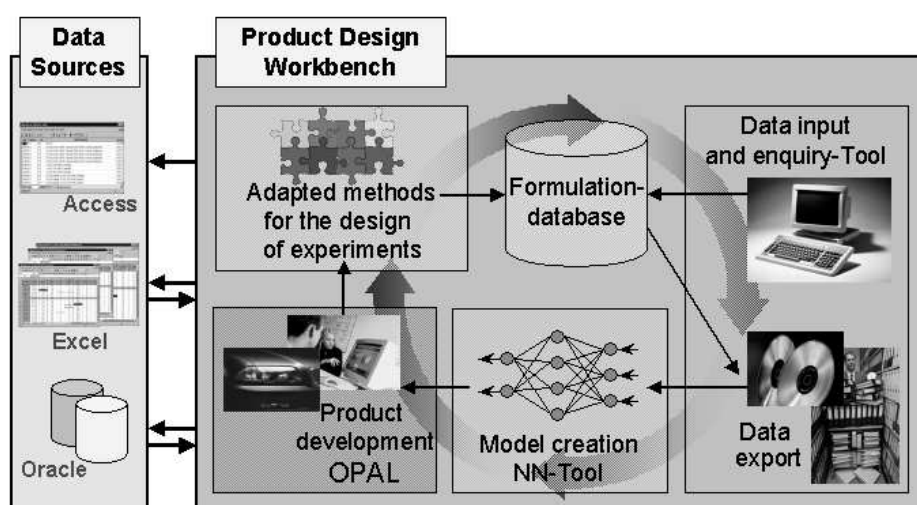


Figure 1. Typical product design workflow

A typical example for such a product design workflow may be described as follows: There are existing data sources containing the formulations already tested - process parameters as well as the measured product properties. With the help of an adapted interface, these data can be imported into the internal formulation database of the workbench. Using the integrated inquiry module, the product developer is able to search for the experiments which are suitable for a given product design problem. Usually, this will be done in an iterative way. The resulting data set is processed as a flat table which is then used to generate a data-based model. Thereby the product developer can find exactly the subset of data leading to an optimal model for a specific product design problem.

The model development process itself is based on artificial neural networks or hybrid models. Neural networks are some kind of complex non-linear regression systems. A hybrid model is a systematic combination of neural networks with physical or chemical knowledge. The user module *OPAL* allows to predict the property profile of virtual formulations and to calculate new formulations with a given property profile. In case the available data set does not lead to a sufficiently precise model, the data set can be

completed with further experiments, using the integrated design of experiments module which is adapted to the hybrid modelling methods employed.

3. Modules

The internal formulation database represents a central module of the product design workbench. The logical data model is kept rather general to cover formulation data from completely different application areas. Therefore, the database can be applied as a formulation data management system for a complete operational area. In principle the underlying database system can be chosen, e.g. Oracle or Access. A specific component supports the data import into and export from the database in a standardized way. In addition, an adopted interface is used to connect the internal database with available data sources.

3.1 Suitable formulations

The inquiry tool allows the product developer to find the adequate data in the database. Typical queries require or permit certain ingredients or require a lower bound for the number of experiments containing a component or parameter. In the same manner the suitable data can be restricted to formulations where the components or the properties meet some specifications. The complete query is specified with the help of a graphical user interface and the corresponding SQL statement is generated automatically. As a first result the user gets statistical information (including bar charts) about the resulting data subset, e.g. the number of formulations, their ingredients and concentration ranges as well as the measured properties. Using this information, the query is refined until the ratio of the number of influencing variables (as a measure of problem complexity) and number of suitable data points is as small as possible.

3.2 Model development process

The resulting data subset is passed on as a flat table to the model development module. Depending on the application area, artificial neural networks or hybrid models are used. In many situations neural networks are a sufficiently powerful tool to describe the formulation property relationship. For this case a commercial neural network component is integrated to support the product developer. For the non-expert user the algorithm is largely parameterized, ensuring to generate accurate models (depending on the data quality) in almost all situations.

A significant improvement is achieved by the additional use of a priori structural information about the underlying relationship. The central idea is that this information is used to reduce the solution manifold of the system. This approach shall herein be called *Structured Hybrid Modelling*. A structured hybrid model (SHM) therefore consists of three components:

- rigorous submodels describing the i/o (input/output)-relation of those subprocesses which are well understood
- black-box submodels for those subprocesses where no rigorous model is available
- a model flowsheet describing the i/o-structure of all the submodels, joining them together by mapping the i/o-structure of the real process.

In theoretical and practical investigations since 1995 (Schuppert, 2000) it has been shown that by the use of hybrid models the extrapolation behaviour can be improved

dramatically. During a project over several years at Bayer, hybrid modelling was advanced to a standard technique (Mogk et al., 2002). This modelling technique has been completely integrated into the workbench system.

3.3 What if scenarios

Using the neural network or hybrid model the properties of virtual formulations are predicted with the help of the user module OPAL (Optimization Prediction Analysis Lab). In addition OPAL provides visual insight into the high-dimensional relationship and graphical means for formulation optimization. Moreover, it is possible to calculate ideal formulations to meet a given property profile with some kind of Monte Carlo method in combination with a gradient-based optimization method. To this end, a multi-objective optimization problem with a number of inequality side conditions (to meet the given property profile) is transformed into a single-objective optimization problem in the sense of the goal attainment method. Then the Monte Carlo type method is used to incorporate additional discrete side conditions (e.g. to restrict the number of used ingredients or the number of ingredients structured in different groups, or nonnumeric parameters). A limited number of best results are in the second step used as initial estimates for a gradient based optimization, considering only the continuous side conditions. Particularly with regard to getting a uniform distribution of the candidate points in case of the Monte Carlo method, an enormous effort has been put into taking the various formulation side conditions into account.

3.4 Closing data gaps

The prediction capability of a data based model can be only as good as the underlying data. To close gaps in the data a specific component for the design of experiments in the context of neural networks and hybrid models has been developed and integrated into the workbench system: new experiments are planned such that they are as different as possible with respect to each other and with respect to the available data. Thereby regions in the formulation space are preferred where, based on a first model, good quality properties can be expected. To this end, in practical applications complex side conditions similar to the optimization case have to be taken into account.

3.5 Data mining

Data mining is a generic term for algorithms and techniques which can transfer data into information (Fayyad et al, 1997). That is, data mining comprises methods which can extract rules, patterns or other structural information from data. In addition, it is possible to identify new relationships which were undiscovered before in the data. Typical data mining techniques are, e.g., clustering algorithms, separation methods, decision trees, and association rules.

In order to assess the quality of data mining rules, Bayer Technology Services has developed proprietary algorithms, e.g. for rule analysis and rule statistics. To make data mining results reliable, it is essential to apply different approaches and check the resulting rules for mutual compatibility.

The generated rules, patterns, or structure relationships can be used to support the product developer during the data query phase, in development of enhanced neural networks or hybrid models, or in a better targeted experiment design.

3.6 Evolutionary Optimization

Evolutionary Optimization can be applied as an experiment design strategy to optimize or to find better formulations, e.g., for catalysts (Duff et al, 2004). As the name implies, evolutionary optimization is based on genetic or, in general, evolutionary algorithms. These algorithms are very robust and they easily handle large parameter spaces as well as discrete variables. The latter characteristics are typical for problems from the area of high throughput experimentation such as the search of novel catalysts.

4. Case studies

4.1 Catalyst discovery

A new catalyst was searched in order to modify a hydrogenation reaction. The catalysts to be comprised of up to two elements which could be selected from a group of six elements. The concentration values of the catalyst components can have four possible levels. Other parameters are the reaction temperature, the reaction time, the calcination temperature, the calcination atmosphere, the ratio of the two reactants and support. All in all the parameter space is described by roughly twelve variables which are continuous or discrete factors. The objectives are five parameters such as yield, selectivity, amount of special by-products as well as ratios of special products and by-products. For the catalyst discovery roughly 300 old experiments were available. From this data pool 14 data points were extracted and used as a priori information.

To close gaps in the data to derive an accurate neural network model the first three experiment iterations were designed by an evolutionary strategy (using the module ECCO, Duff et al, 2004). The first two iterations comprise of 50 experiments each. During the third iteration 20 further experiments were carried out. The evaluation of this experiments was supported by using data mining techniques. With the old data sets and the results of these 120 new experiments for each objective a neural network model was developed. Using this neural network a model-based experiment design was carried out to optimize promising catalysts. The experiment design strategy is described above. After additional 30 experiments several leads could be identified which satisfy the given target requirements.

4.2 Optimization of plastic properties

An established synthetic product should be optimized with respect to its properties, specifically, the transparency should be improved. The aim was to modify the present formulation via concentration variation or exchange of the ingredients. Other properties should not be modified and should be retained within the given specification range.

In this study the formulations were compositions out of ten components. Two of the ten components could be replaced with others. During the optimization process seven objectives such as mechanical properties and two side constraints, e.g. a summation condition for the concentrations of the components used, were considered. An optimal formulation was discovered after two workbench iterations and all in all 50 experiments were used.

5. Discussion

The main focus of the workbench concept is to support the product developer with respect to a systematic product design. Therefore, the software system is configured specifically to meet the requirements of the given problem class and the needs of the user. In this context, configuration means both the special selection of the implemented modules and the adaptation of the modules and software components. An example for the former aspect is the use of a rigorous model instead of artificial neural networks. The latter point means that the software components can be simplified for the use in routine problems or that more functionality of components or algorithms can be provided for the use as an expert system. Finally, the problem decides which methods or modules should be applied.

The idea of the described system is to support the product developer with the help of a model-based product development. Therefore, it is important to use problem-specific models which are based on selected data from the present data pool. These models can be understood as a kind of “road map” to find new or enhanced products. During the product development procedure, these road maps are improved or transferred into more detailed maps which only describe sections in the formulation-property space. For this reason the presented system is not only an atlas, it is a kind of navigation system for a systematic product development.

The product design workbench has been proved to be of great value in many different application projects. Moreover, the system could be applied to a broad range of problems which are not restricted to the polymers or plastics industry. Food, cosmetics, and detergent industry have a high potential to increase efficiency by using the product design workbench.

References

- Duff, G.D., Ohrenberg, A., Voelkening, S., Boll, M., 2004, A Screening Workflow for Synthesis and Testing of 10,000 Heterogeneous Catalysts per Day - Lessons Learned, *Macromol. Rapid Commun.* 25, 169-177.
- Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R., Eds., 1997, *Advances in Knowledge Discovery and Data Mining*, AAAI Press, Cambridge, Massachusetts.
- Mogk, G., Mrziglod, Th., Schuppert, A., 2002, Application of hybrid models in chemical industry, *European Symposium on Computer Aided Process Engineering - 12*, Elsevier Science B.V.
- Schuppert, A., 2000, Extrapolability of structured hybrid models: a key to optimization of complex processes, *Proceedings of EquaDiff '99*, B.Fiedler, K.Gröger, J.Sprekels Eds., 1135-1151

Acknowledgements

The financial support of the Bundesministerium für Bildung und Forschung is greatly appreciated.