

Comprehensive Analysis Competence and Innovative Approaches for Sustainable Chemical Production

Joerg Appel^{*a}, Corrado Colombo^b, Urs Dätwyler^c, Yun Chen^d, and Nimet Kerimoglu^a

Abstract: Humanity currently sees itself facing enormous economic, ecological, and social challenges. Sustainable products and production in specialty chemistry are an important strategic element to address these megatrends. In addition to that, digitalization and global connectivity will create new opportunities for the industry. One aspect is examined in this paper, which shows the development of comprehensive analysis of production networks for a more sustainable production in which the need for innovative solutions arises. Examples from data analysis, advanced process control and automated performance monitoring are shown. These efforts have significant impact on improved yields, reduced energy and water consumption, and better product performance in the application of the products.

Keywords: Advanced control · Big data · Energy · Process monitoring · Water · Yield improvement

1. Introduction

Humankind currently sees itself facing enormous economic and social changes which are represented in global megatrends like urbanization, globalization, resources and energy consumption, *etc.* Therefore, tremendous challenges in protecting the ecological system are created. The specialty chemicals industry has to provide a decisive contribution to the successful management of these global challenges by an efficient use of resources and energy.^[1,2] Innovative solutions from this industrial sector are leading to a significant contribution to sustainability in their applications. New raw materials based on biological feedstocks and innovative process technologies are developed in order to increase resource efficiency and to reduce the carbon footprint of the next generation's products and processes.^[3]

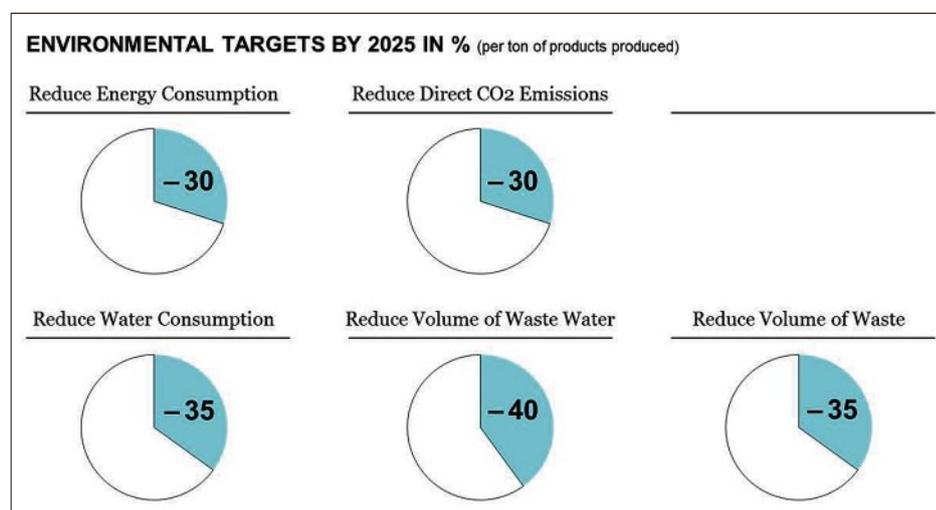
Chemical companies have published their sustainability policy and defined environmental targets for 2025, for example as set out by Clariant in Fig. 1.

Besides these step changing innovations, it is a strategic goal of chemical companies like Clariant^[4] to incrementally improve their existing production assets for higher chemical yields and lower energy and water consumption. The regional supply of specialty chemicals adapted to local customer needs is performed in a global production network. Optimizing these networks is a complex task and requires a well-organized approach in which best practices are shared and transferred throughout the network. This requires a thorough analysis and a clear definition of the best approaches based on key performance indicators, which represent a plant's operation in a balanced and robust way, avoiding an optimization to a local optimum.

Through these analyses and the representation of the plant's performance

in terms of resource efficiency, existing performance gaps are visible even for the current existing best practice. Innovative improvements in plant control and monitoring are promising opportunities to further improve the sustainability of existing production plants. Digitalization, which is developing at a rapid pace, and further developments in sensor technologies^[5] are creating new platforms for innovative solutions in performance monitoring and model-based plant control.

In the following chapters we describe the overarching approaches for process analysis *via* the use of big data analysis towards the implementation of advanced process control. With this comprehensive approach Clariant is developing its global production network into an eco-friendly and sustainable system.



Correspondence: Dr. J. Appel^{}

E-mail: joerg.appel@clariant.com

^aClariant Produkte (Deutschland) GmbH, 84508 Burgkirchen, Germany

^bClariant Prodotti Italia, 20037 Novara, Italy

^cClariant International AG, 4132 Muttenz, Switzerland

^dClariant Produkte (Deutschland) GmbH, 65929 Frankfurt, Germany

Fig. 1. Clariant's environmental targets 2025 are seen as a driver for sustainability and growth.^[4]

2. Process Analysis in Global Production Networks

2.1 YEE (Yield – Energy – Environment)

Analyzing a global network of different production sites requires dedication, expertise and the right methodology. Therefore, Clariant decided to build-up a dedicated team for the analysis of production plants with a focus on improving yield and energy consumption as well as reducing the impact on the environment. The analysis follows a two-step approach whereby the YEE specialists perform an intensive diagnostics of potential improvement areas together with the chemical engineering expert and the operations team. An important success factor of this work is the personal presence of the experts in the plant to get the information first hand. After the identification of improvement potentials, an implementation phase starts, in which the operations team realizes the optimization together with dedicated improvement project managers.

During the last four years, high benefit potential has been diagnosed in about 25 plants in 15 different production sites, which resulted in more than 600 improvement measures. A large extent of these measures has already been realized. Corresponding significant benefit has already been realized and tracked in a comprehensive database. However, apart from the financial impact, there is a positive impact on the environment as well. Using less raw material, consuming less energy and reducing emissions and waste are directly linked to Clariant's sustainability policy and essential to reach Clariant's environmental targets.

The initiative started with diagnostics of a large continuous plant and was then performed in multiproduct batch productions located similarly at many different sites globally. Savings result particularly from yield improvement and energy efficiency. In principle these potentials are defined as losses compared to a potential optimum. They are classified in operational and design losses. Scale-up effects can be seen as an additional contribution to yield losses. The principle form of a yield loss bridge is illustrated in Fig. 2.

Fig. 3 shows the result of one of those detailed analyses as a long bridge of operational losses. Some of these losses can be solved by adapting the design of the corresponding apparatus. The choice of the right apparatus and the proof of the new process have to be demonstrated on pilot scale or even in a production trial. This is realized by process engineering experts who are also involved in the implementation project teams.

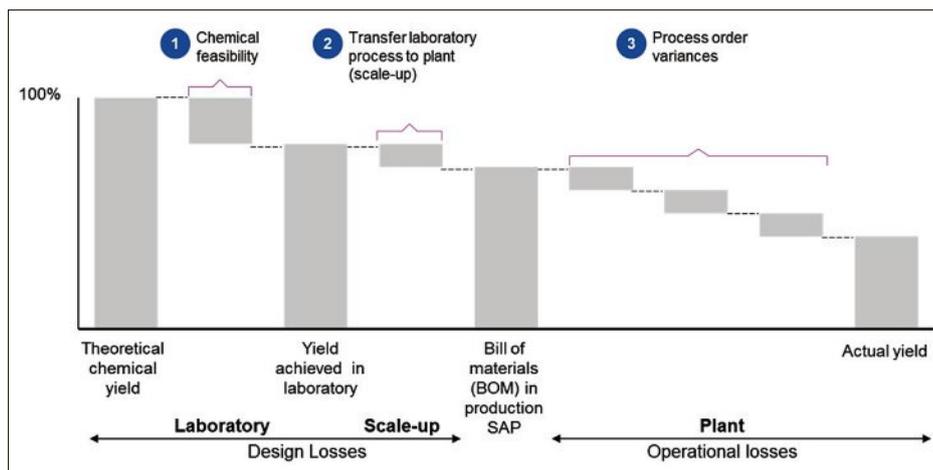


Fig. 2. Generic loss bridge from a YEE analysis.

Having this portfolio of ideas and potential improvements, the realization is carried out taking into account the effort for the individual improvement measures.

Yield improvements in sold out production lines have double impact on benefits. On top of material savings, additional production volumes enable growth and generate additional margin.

In the area of energy, there is a close collaboration with the internal Competence Center Energy & Utilities. The YEE team and the energy experts jointly conduct diagnostics. Areas of improvements are identified by a trend analysis of the specific energy consumption. Specific values are most often related to production volumes. Internal benchmarking by identifying the best demonstrated practice (BDP) is also used to access untapped potential.^[6] Minimum BDP target after a YEE diagnostics is usually the reliable achievement of the best 25% range (> third quartile) of yield and specific energy consumption *etc.* observed in the last few years. Afterwards, these figures have to be improved by con-

tinuous improvement projects. Special kinds of continuous improvements could be the installation of an advanced control and monitoring system. In sections 2 and 3, we will come back to the application and a further development of the best demonstrated practice curves. Fig. 4 shows one example of a best demonstrated practice curve. Here the steam consumption per production throughput is shown.

It is obvious that the specific energy consumption is a function of the load of the plant. The best demonstrated practice curve shows which energy consumption the plant had already realized in the past. It does not however give an explanation how these optimal operating points have been reached. The curve helps to estimate how big the benefit could be if the operations team would search for the reason which will result in generally lower energy consumption.

Throughout the diagnostics and the implementation of the optimizations best practices are collected and then summarized in so-called Blue Prints, which are

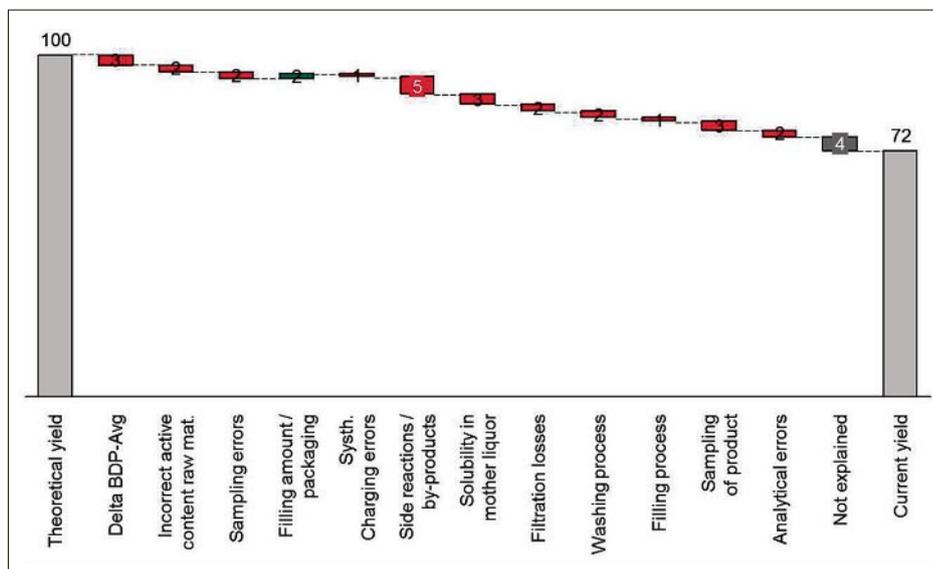


Fig. 3. Detailed loss bridge from a real site diagnosis.

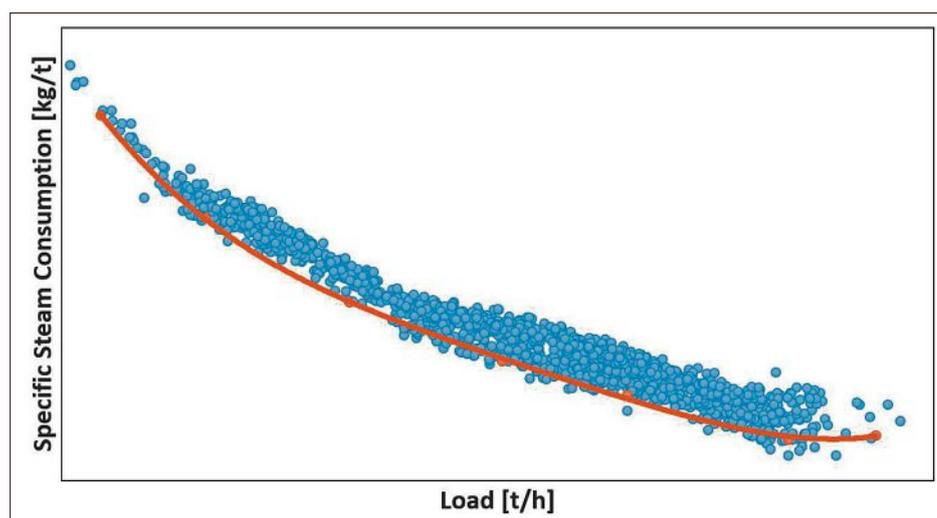


Fig. 4. Load specific energy consumption as function of steam consumption per ton of product produced. The orange curve indicates the BDP curve.

descriptions and instructions for optimum solutions. These procedures are then published to relevant sites all over the world in order to reach a scalability of the approach in the company.

Frequently, yield and energy are directly linked. Solvent losses can be reduced for instance by an appropriate cooling system. Yield can be improved through more intense cooling due to higher selectivity of the reaction. In these cases, the YEE team helps to quantify the impact and calculates the net benefit. Through this holistic approach, the overall economical optimum is identified. Multidimensional models derived from process data analysis serve as the mathematical representation for this optimization.

Yield improvements also immediately impact the environmental situation through reducing waste and emissions. The YEE team also looks at water consumption as water is becoming a more and more scarce natural resource. Water consumption is thus a key parameter in Clariant's environmental targets 2025. YEE will support the achievement of the ambitious sustainability targets by diagnostics of the current situation and ideations for improvements.

The number of YEE projects has significantly increased during the last three years and this allows increasing global benchmarking at comparable production technologies. There are many activities also beyond the YEE team to which operation teams across the globe and all business units and sites are contributing. There is definitely still a lot of potential which needs to be identified and tackled to reach the environmental targets. It is increasingly necessary to go deeper into technical detail and broader along the value chain during the analysis. The YEE approach is therefore included in an integrated end-to-end view of the value chain. Optimizing inter-

faces between different responsibility areas would probably provide the most gains and most of the future benefits.

Performing analysis over long value chains and considering the effect of interfaces will increase the amount of data tremendously. Therefore, it is crucial to have an efficient data handling system in place and to be able to extract significant information from the data. Data infrastructure, the capability to analyze and interpret the data and the acceptance of the methodologies are the three pillars for a successful application of data mining in the future process optimization activities.^[7]

2.2 Innovative Data Management and Big Data Analytics

The analysis of chemical processes in a global network of production units allows the generation of a chemical processes landscape, best practice approaches and also an overview of the data structure around the different plants. This gives an overview of the availability of process relevant data for improvement projects. It also shows where the infrastructure is not sufficiently developed for the application of novel technologies from the field of digitalization. In general, it has to be distinguished which kind of installation is the next best step towards a higher data integration in production. Parallel to the data availability the acceptance of the technologies in the work force is developing. The acceptance is a crucial success factor for the implementation of data-intensive technologies.

With increasing digitalization of processes, the volume and complexity of data in multiple data sources becomes a technical challenge, but also an opportunity to make a company more successful. The transformation of data to insights that drive decisions or to identify improvement potential or new business opportunities are

the main targets of 'Big Data', as it goes beyond standard reporting by flexible combination of data from multiple sources in meaningful ways at any time.

A chemical company owns a huge amount of data, stored mostly in different computer systems, which have been designed in order to accomplish a well-defined function. It is self-evident that we have to see all these as a source for 'Big-Data' analytics. It is big in the sense of frequency, *e.g.* when each minute thousands of sensor measurements are stored from a plant control system in a data historian warehouse. It is also big in the sense of volume, variety and diversity of the information type and data structure.

The questions which arise from this scenario are:

- Is the information used in a comprehensive way?
- What would be the opportunities if we could transform the information into knowledge?
- Are there new business opportunities by linking data sources and creating transparency?

The overarching goal is an accelerated and improved decision-making process.

Together with approaches described in the first sections of this article, a data treatment competence allowing a detailed analysis of production processes as well as an automated performance analysis was built up within Clariant. A major task of this work is to give the data set a structure, which makes it accessible for statistical, machine learning and data modelling techniques. In literature^[8,9] there are also activities which try to handle the unstructured nature of the data sets but this has not been followed yet in the present approach. The structure of this data cube approach is described in Fig. 5.

Data from different sources like a data historian system, enterprise resource planning systems ERP, process control systems, quality assurance and analytics are stored in a new database, where they can be treated independently from the daily business. *Via* certain key variables like the batch, or lot-number, the product name or the production unit each data set can be assigned to a certain product and so a certain production in a specific plant.

In Fig. 6 the workflow from data acquisition to result implementation is depicted.

Having accomplished this data acquisition and organization step a crucial part of the data treatment has to be performed. Normally the data sets are of different structures. Batch-related information is an individual set of numbers and facts for different batches. For example the production unit, the raw material amounts, the type of raw materials are discrete information.

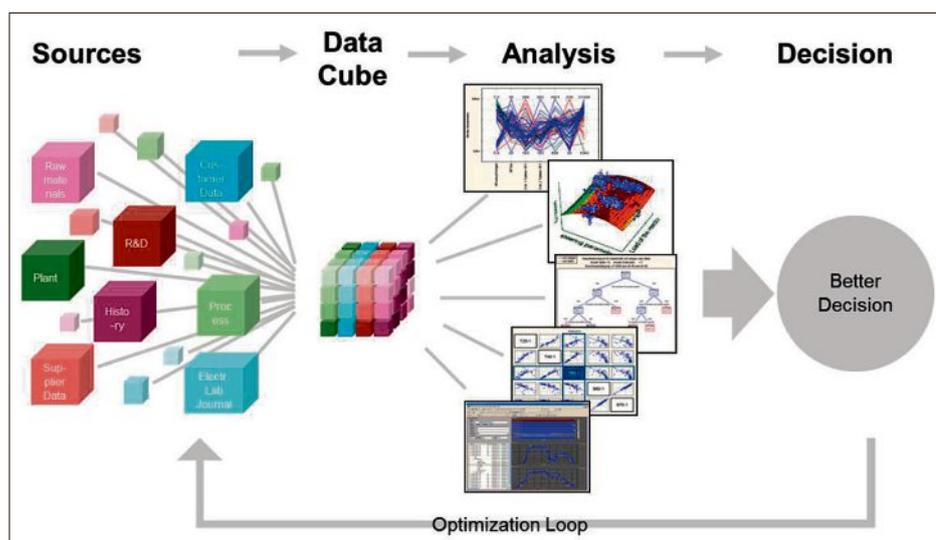


Fig. 5. Data cube approach for comprehensive data analysis and an improved decision making process.

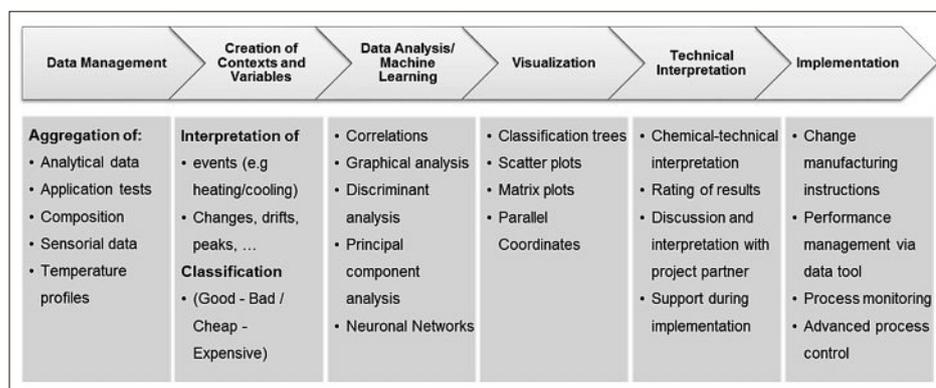


Fig. 6. Work-flow for data analysis projects.

There is the need to combine this information with the time-dependent profiles from sensors in the plant. A combination can be reached by assigning contexts to the relevant measured profiles. These contexts can be complete patterns, maximum or minimum values, mean values or slopes during different phases of a batch. With this approach it is possible to characterize each curve in a data historian system. Also this step needs to be automated and it has to be repeated for all the profiles that are detected in a batch system. It might appear attractive to select variable by intuition in order to reduce the effort for analysis and speed-up the process. Nevertheless, this includes the danger to follow a wrong and misleading interpretation, which might be based on misinterpretation of operational experience. Only if the algorithm is applied to all variables the observation of unexpected correlation of process parameters in the subsequent mathematical analysis is possible. Fig. 7 demonstrates the idea of profiling, context generation and saving the results in a database.

If the data profiles come from different sources such as different plants, the newly

created so-called data cube allows to slice out a single product, which has been produced in different plants. Having the data of both plants available allows the search for performance differences and for the

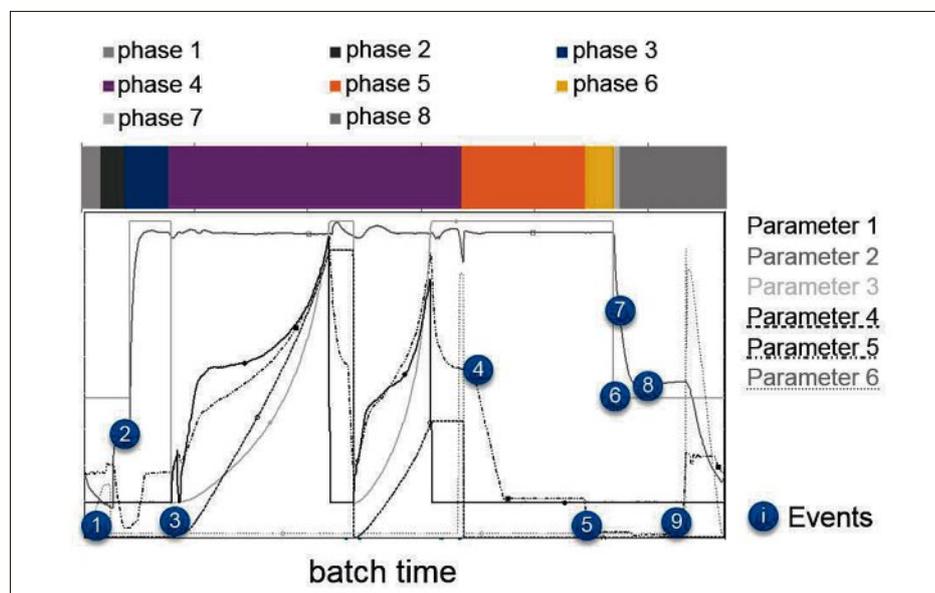


Fig. 7. Batch time analysis, events and phases are the fundament of the data table for analysis.

explanations of these differences. If plant design data has been incorporated to the analysis even the single design criterion can be identified, which leads to the performance difference of the two plants. In one example a straightforward optimization measure could be derived directly by comparing temperature profiles from two different plants in one plot.

All the efforts described so far increase the size and complexity of the data set. Introducing a classification for continuous variables describing the general location of the data entry compared to the full data set additionally adds complexity but will allow the application of the full toolset for analysis. Often the interpretation of the results is improved by the newly introduced classes because it is in the nature of human beings to interpret clusters with the help of an 'if .. then' statement rather than dealing with real numbers and their functional relations.^[10] The result of such an analysis can be displayed in a decision tree to generate understanding and give guidance, when a decision is needed. In Fig. 8 an example of such a decision tree is shown.

Data analysis starts with a smartly formulated question, which allows the data mining expert to apply the right methodology to the data set. It is very important to understand the nature of the problem. From this knowledge the right tools for analysis and optimization can be chosen.^[11,12]

Generally it makes sense to start each analysis with a visualization of the data. Very often differences can be quickly seen visually and a technical interpretation brings the expert from one graph to the next in order to support or to disprove a theory behind his/her observations. Generally the graphical examination is powerful and fast but the approach is limited to the number

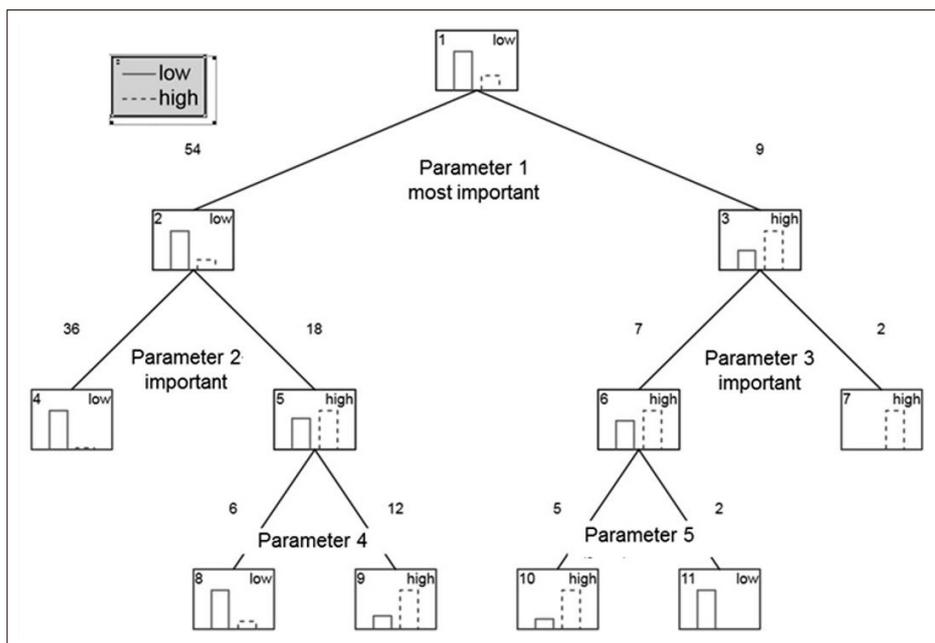


Fig. 8. Example for the result representation as a decision tree. The importance of a parameter is calculated from its strength to separate the two groups of interest from each other.

of dimensions, which can be depicted and which can be interpreted by the user.

Large multivariate analyses require the application of the corresponding tools like principle component analysis or discriminant analysis or many other similar techniques in order to identify correlations, clusters and functional dependencies. All methods have in common that they can be used to search for clusters in the variables in order to reduce the dimension of the problem and to derive the right variables for model building.

Building data-driven models should also follow the concept that the model approach which is chosen should fit to the data type, data quality and size as well as the accuracy of the expected response. The subsequent application of the model is also an important factor. For instance, an advanced process control methodology requires a much higher accuracy and reliability compared to a trend analysis of the expected future outcomes of experiments.

An important factor in industrial approaches with data driven models is the application of real data, which are not cleaned or produced by a dedicated design of experiment. While outliers can be detected and removed quite easily another important factor is the redundancy in the data. The complexity of modelling approach, which is equal to the number of degrees of freedom, has to correspond to the number of meaningful operating points. Meaningful in this context means that the operating points show significant variation and are not redundant. The higher the nonlinearity of the model, the higher the required number of independent points. The use of artificial neuronal networks^[13]

is very efficient for modelling, because they can be defined in different complexity without predefining the type of non-linearity.

Models which have shown comparable sensitivities at different complexity are a reliable basis for advanced process control.

3. Advanced Process Control with Data-driven Models

In principle advanced process control refers to various techniques used in addition to the standard control methods in industrial process control systems. In the context of the current paper multivariable model predictive control is in the focus

of the discussion. Model predictive control can be realized by using deterministic models^[14,15] where mostly the model parameters have to be adapted automatically in order to reach the required accuracy for process control.^[16] These models have the advantage that they have a strong capability of prediction even to regions where the plant has not yet been operated. The disadvantage is the relatively high development and maintenance effort for the models. Data-driven models can be built very fast and automated but have the disadvantage that they lose reliability when the parameter space reaches regions for which the models are not built.

A combination of both approaches, deterministic and data-driven, can be found in hybrid models. They combine both advantages, reduced complexity and prediction capability. The deterministic modelling part is applied where reliable models are already available.^[17]

As an example an industrial implementation of neuronal network-based advanced control system is shown in Fig. 9.

In this example the dosing of an additive to a reactor system is shown. The additive can be seen as a catalyst for the reaction. Since the reaction does not reach full conversion the raw materials are fed back to the reactor after the stream went through a work-up procedure. The product and the by-products are removed from the intermediate product stream. After an intensive data mining the significant parameters for model building have been identified. The model was trained and the capability to predict the sensitivities have been checked over long periods of time. This work has been done in collaboration with the Atlan-tec Systems GmbH in Willich.^[18] After the technical implementation of the model and the software environment from Atlan-tec

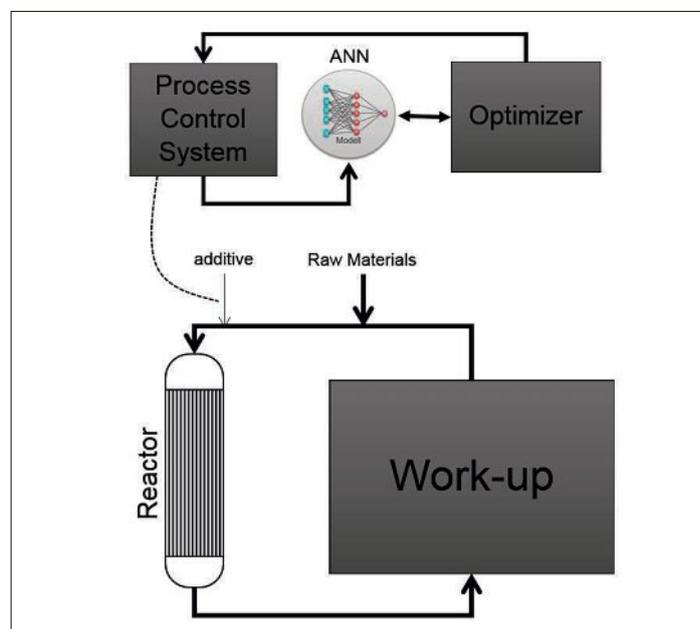


Fig. 9. Implementation of the artificial neuronal network in the environment of the process control system.

a test period was performed to adapt the dynamics of the model to the dynamics of the changes at the plant's boundary conditions. The advanced control system led the plant to a more stable operation with higher technical performance.

In most plants the optimum performance is not a fix value but a function of time. The changes in performance can be caused by fouling, catalyst deactivation or seasonal effects. Therefore it is necessary to evaluate if an autonomous control system really reaches the optimum performance or if adaptations are needed to improve the model itself.

4. Automated Performance Monitoring

Additional to the described aging and fouling effects with very long time scales, unpredictable outside disturbances on short timescales can occur as well. Therefore, a performance monitoring should be installed to validate if the system is on the best trajectory. Depending on the quantity and frequency of the outside disturbances, significant deviations from the optimal performance can arise when the data-driven model has not yet been trained with the effect of the occurring event or the right input parameter is missing in the model.

A systematic and automated performance monitoring algorithm is necessary to realize deviations early enough to keep the system at its best performance. Since advanced control systems try to improve the existing good plant performance, it is often hard to detect if a system is working at the optimal operating point. This means that additional to the long time scale effects a very accurate observation of the performance is necessary in order to distinguish aging from weak control of the process.

The mentioned challenge of an accurate process control to an optimum performance requires excellent sensor technologies, a smart control strategy and an evaluation of the benefit and the robustness of the system. In the project CONSENS (Integrated CONTROL and SENSing)^[5] funded by HORIZON2020, the European Union's Research and Innovation Program, these topics have been investigated. Goal of the project is to advance the continuous production of high-value products that meet high quality demands in flexible intensified continuous plants by introducing novel online sensing equipment and closed-loop control and performance monitoring of the key product parameters. The project is focusing on flexible continuous plants but the results in the areas of sensing, control, and performance monitoring will be transferable to other large-scale processes. In this context, Clariant is leading the work

package which deals with tools for design evaluation and monitoring. One part of this work package is the on-line monitoring of the performance of a plant. For demonstration purposes data from an existing large-scale production plant has been used. The implemented advanced control system directly uses measured data from on-line sensors of the product quality. From this, the performance of the plant can be derived directly. As described in section 2 the load of the plant is of course one of the most important performance factors. Fig. 10 depicts that the plant shows significant load changes and also stable load levels lasting several weeks.

The algorithm of the performance monitoring methodology includes the following steps:

1. Data pre-treatment
2. Clustering of load levels
3. Creating sub-clusters to reduce aging effects
4. Cross validation of the performance of the sub-clusters
5. Derivation of a curve for the expected best performance for each cluster
6. Evaluation of the distance between actual and expected performance

In the data pre-treatment step outliers are removed from the data set. Outliers are also the time spans of load change because they are not significant compared to the overall time. The first level of clustering are the load levels. Here the sensitivity of the cluster algorithm has to be adjusted to have enough different load levels to eliminate the load effect within the cluster. On the other hand, the individual cluster needs to be large enough to perform an expected best performance calculation as shown in Fig. 11. Criteria for this algorithm are the intra- and inter-cluster distances. The sub-clusters in Fig. 12 are created by choosing a small and constant time discretization.

The step size is a function of aging rate of the complete cluster and a large enough amount of data to create a statistically significant regression model for the cross validation. The cross validation itself is a prediction of the performance of one sub-cluster based on the other sub-clusters. The results of such a cross-validation is shown in Fig. 12. By this way the expected best performance curve is derived for each cluster and can be used to evaluate the overall performance of the plant shown in Fig. 13.

The evaluation of the actual performance compared to the expected performance is an indicator if the advanced control systems runs properly or if maintenance is necessary either to adapt model parameters (e.g. training periods or parameter weights) or if a significant change of the model development is necessary.

The described effort demonstrates how yield and energy consumption can be improved by applying a high degree of automation and hence the defined optimization potentials can be made sustainable.

5. Conclusion

The optimization of global production networks is a very important effort to achieve sustainable chemical production. In the first step it is important to identify the most interesting improvement potentials and try to harvest the 'low hanging fruits' as fast as possible. One important point here is best practice sharing, communication and a dedicated focused team work. It was observed that after a certain very successful time period, optimization potentials can only be found further by analyzing long value chains and the interfaces involved. Due to the rising complexity it is obvious that big data sets are the basis for these analyses and new tools and methodologies have to be applied to ensure

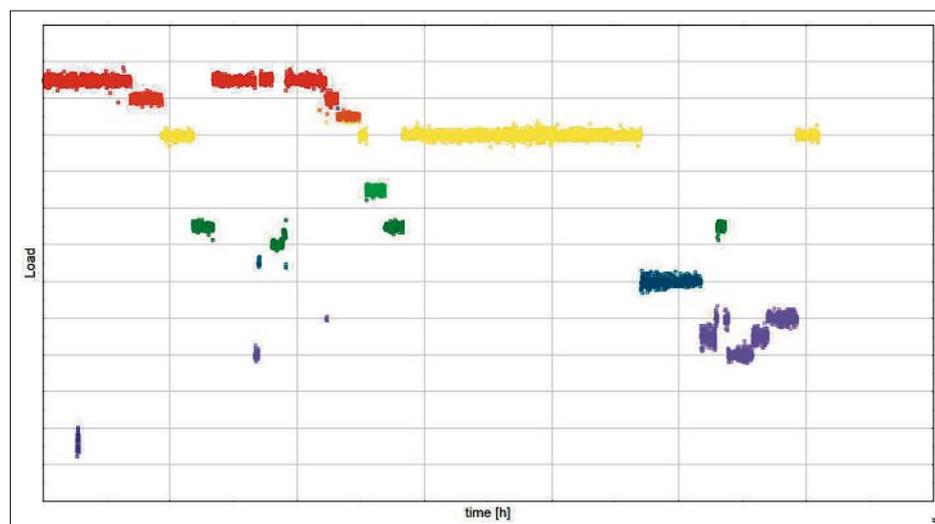


Fig. 10. Clustered load levels in the plant after data pre-treatment.

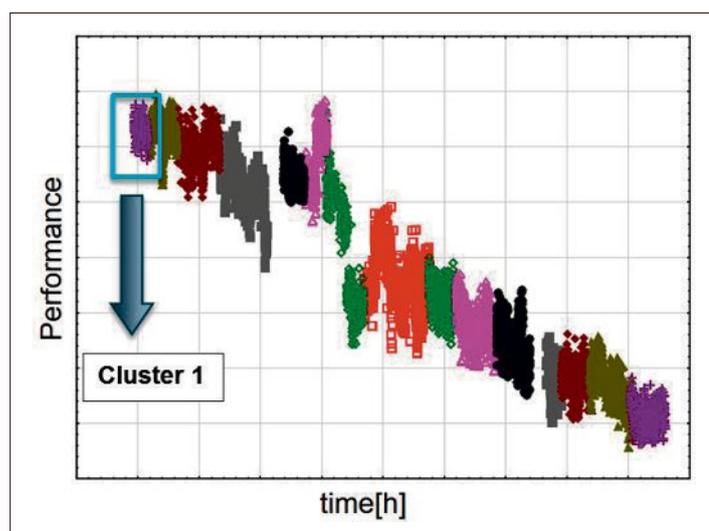


Fig. 11. Result of the cluster analysis for the load shown as a time dependent performance curve.

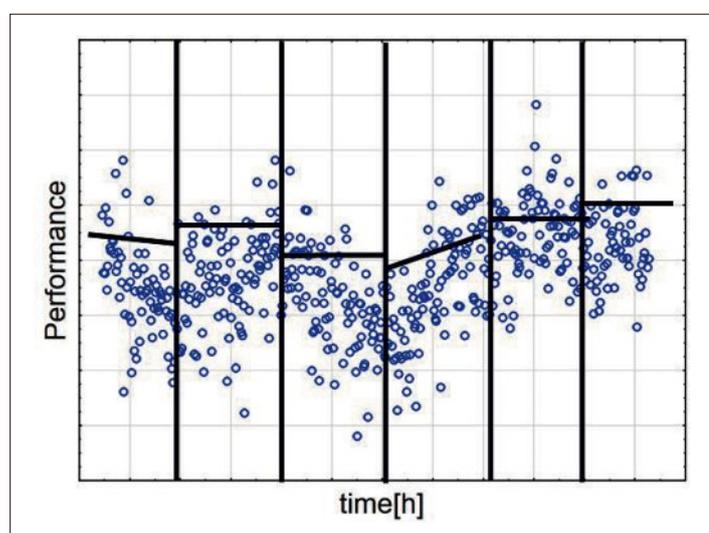


Fig. 12. Sub-clustering and cross validation to obtain expected best performance of Cluster 1 (cluster 1 is highlighted in Fig. 10).

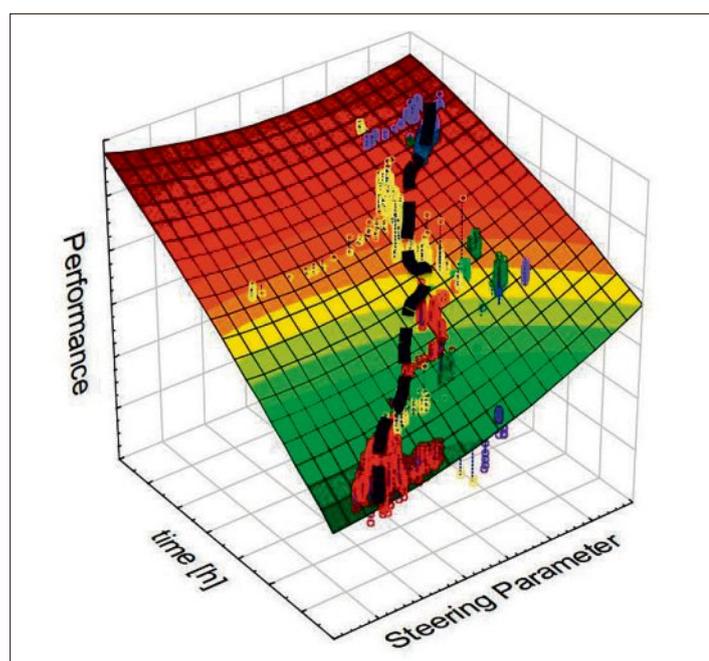


Fig. 13. Overall expected best performance (black dots) with respect to actual performance.

efficiency. Searching in unstructured data is a promising approach but requires development and training. In this article the focus lies on the work with structured data

explicitly prepared for data mining and modelling. Depending on the level of digitalization and network integration the data driven models can be used for advance pro-

cess control and automated performance monitoring.

The described pathway from onsite diagnostics, *via* big data analytics and advanced control towards automated performance monitoring is an innovative approach. It will gain more and more importance through the increasing level of digitalization in industry. Increasing and extensive horizontal and vertical integration of data sources create the platforms for the application of modern tools for data processing. An important success factor for these technologies is the competence in interpreting the modelling and/or data mining results. A second important success factor, the men machine interface, has to be kept in mind and actively addressed during development and implementation.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 636942.

Disclaimer

Any dissemination of results must indicate that it reflects only the author's view and that the *Commission* is not responsible for any use that may be made of the information it contains.

Received: June 23, 2016

- [1] ProcessNet-FutureWorkshop:Resultsataglance, <http://processnetschaffitzukunft.wordpress.com/ergebnisse-des-workshops-im-uberblick/>
- [2] C.-J. Klases, *Chemistry & More* **2014**, 4, 18.
- [3] O. Wachsen, V. Bacher, A. Geisbauer, J. Appel, V. Niebuhr, O. Lade, *Chem. Ing. Tech.* **2015**, 87, 683.
- [4] Clariant Sustainability Report **2015**.
- [5] CONSENS Project, www.consens-spire.eu, **2016**.
- [6] C. O'Dell, C. J. Grayson, *California Management Review* **1998**, 40, 154.
- [7] K. Bunse, M. Vodicka, P. Schönsleben, M. Brühlhart, F. O. Ernst, *J. Cleaner Prod.* **2011**, 19, 667.
- [8] I. H. Witten, E. Frank, M. A. Hall, 'Data Mining, Practical Machine Learning Tools and Techniques', 3rd ed., Morgan Kaufmann Publishers, **2011**.
- [9] C. M. Bishop, 'Pattern Recognition and Machine Learning', Springer Science + Business Media, **2006**.
- [10] R. Otte, V. Otte, V. Kaiser, 'Data Mining für die industrielle Praxis', Carl Hanser Verlag München, **2004**.
- [11] Y. C. Ho, D. L. Pepyne, *J. Optimization Theory and Applications* **2002**, 115, 549.
- [12] Y. C. Ho, Q.-C. Zhao, D. I. Pepyne, *IEEE Trans. Automatic Control* **2003**, 48, 783.
- [13] L. Fausett, 'Fundamentals of Neural Networks: Architectures, Algorithms and Applications', Prentice-Hall, **1994**.
- [14] S. Lucia, T. Finkler, S. Engell, *J. of Process Control* **2013**, 23, 1306.
- [15] W. Gao, S. Engell, *Comp. Chem. Engin.* **2005**, 29, 1401.
- [16] C. Y. Chen, B. Joseph, *Ind. Engin. Chem. Res.* **1987**, 26, 1924.
- [17] J. Appel, O. Wachsen, *Chem. Eng. Sci.* **2007**, 62, 4910.
- [18] Atlan-Tec GmbH, www.atlan-tec.de, **2016**.