Optimization of the Müller-Rochow Synthesis of Silanes

Bangert, Patrick [1) 2) 3)]
[1)] algorithmica technologies GmbH, Bremen, Germany
[2)] Advanced International Research Institute on Industrial Optimization gGmbH, Bremen, Germany
[3)] Department of Mathematics, University College London, London, United Kingdom

Abstract

The chemical industry would like to increase the *profitability* of its plants via increasing the production yield. This may be achieved via engineering changes (cost intensive) or via operational changes. To determine the optimal operational changes, we can employ physical modeling or automated machine learning methods. The former method depends on theoretical knowledge and human intelligence, and has proven in practice to be time-consuming, inflexible and idealistic. For the first time, we demonstrate that it is possible to create an accurate, adaptive and complete mathematical model of a complex chemical reaction at the industrial scale (Müller-Rochow synthesis) quickly, using *automated machine learning methods* that give a precise description of the interaction of human operators and the plant via the plant's equipment. The basis for the model is the historical data from the data historian. The resulting differential equations represent the process well enough to be able to compute in real-time what set-points should be changed so as to obtain the maximum profitability of the reactor at any time given the current market prices of the various end products and considering the numerous interdependencies and boundary conditions that exist. *This method requires no engineering changes to be made to the plant and requires nominal human effort to implement. We conclude that a profitability increase of approximately 6% is possible using these methods via an overall yield increase (5.1%) and a selectivity increase (2.9%) for most profitable end product.*

Introduction

Silanes are chemical compounds that are based on silicon and hydrogen. Important for industrial use are the methyl chloride silicones. Industrially, these are principally produced using the Müller-Rochow Synthesis (MRS), which is the following reaction, $2\ CH_3Cl + Si \rightarrow (CH_3)_2SiCl_2$. There are various hypotheses regarding the catalytic mechanism of the reaction but there is no generally accepted theory. Regardless of this, the reaction is widely used in large-scale industrial facilities to produce Di methyl chloride silanes $(CH_3)SiCl_2$ and Tri methyl chloride silanes $(CH_3)SiCl_3$, which we will refer to below as Di and Tri.

Practically, the reaction is carried out using silicon that is available in powder form in particle sizes between 45 and 250 µm with a purity of higher than 97%. The most common catalyst is copper and the promoters are a combination of zinc, tin, phosphorus and various other elements. The reaction is carried out at about 300°C and between 0.5 and 2 bars overpressure. In a fluidized bed reactor the silicon powder encounters chloromethane gas from below. The product leaving the reactor contains the desired end product but also unused methyl chloride that has to be separated in a condenser. The mixture of a variety of silicones is now separated in a rectification where the desired Di and Tri are split off from the other methyl chloride silanes that are mostly waste. These desired end products can now be hydrolyzed into various silicones. The final products of this process can be practically used as lubricants in cars, creams for cosmetics, flexible rubber piping, paint for various applications, isolating paste for buildings and in a variety of other applications.

Unfortunately, the reaction produces several by-products that are unwanted. The selectivity of the process measures how much of the total end product is of the various types; for example a Di-selectivity of 80% indicates that 80% of the total end product is in the form of Di.

The market value of Di is highest among the different end products and so we would want to maximize the Di-selectivity. However, the selectivity is influenced in part by the addition of catalyst. The relationship between increasing catalyst and increasing selectivity is a matter of folklore in this area. As part of this research project, no relationship whatsoever could be discovered within the range, 1% – 3% catalyst addition, studied.

As such the most economical selectivity is not, in fact, the maximum that could be chemically reached. We desire an economic maximum here.

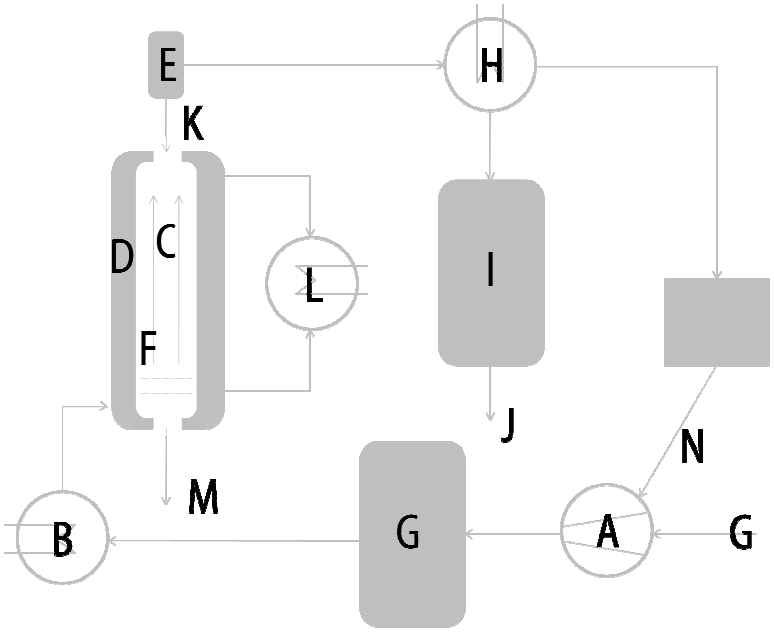

Figure 1.: *Sketch for the Müller-Rochow-Synthesis plant including: (A) Compressor, (B) Vaporizer, (C) Fluidized bed reactor, (D) Cooling jacket, (E) Cyclone, (F) Silicon / Copper (catalyst), (G) Methyl chloride, (H) Condenser, (I) Raw Silane, (J) To the distillation, (K) Silicon / Copper dust, (L) Heat exchanger, (M) Remainder, and (N) Backflow methyl chloride.*

Due to the fact that no generally accepted theory of catalytic mechanics exists, there is considerable debate and experimentation in an industrial setting on the correct use of the catalyst and promoters in order to get optimum performance. The question specifically is: In what circumstances should what amount of what element be added to the reaction? An important component of finding an answer to this question is what the desired outcome of adding these substances is. In an industrial setting, the commercial environment supplies us with some additional variables such as market prices and supply and demand variations. Finally, we establish that the desired outcome is a maximum of profitability. Whatever combination of catalysts, promoters and end products is required, they will be taken and it is the purpose of an optimization to compute this at any time.

Abstract Picture

The production plant is in some sense a black box. We can affect this box actively by putting raw material or catalyst in it or changing some temperatures. This box outputs some end product and also produces a number of measurements that we may record. There is no generally accepted theory of the reaction. Thus, we will treat the plant as a black box, i.e. all that we know about it are the things we can measure by using sensors. Then we will use the mathematical theory of machine learning to establish a mathematical model of the plant from the measured data alone without adding any human knowledge at all. The basic process may be seen in figure 2.

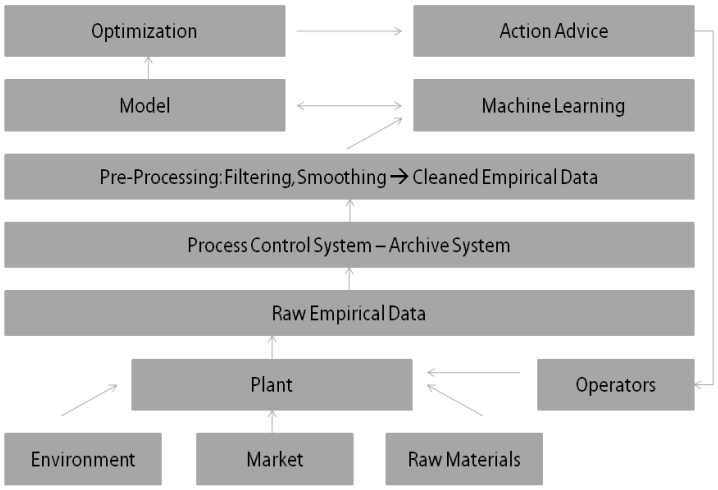

Figure 2.: *A flowchart of the processing steps. We begin with the influence of the environment (e.g. outside air temperature), the market (e.g. prices) and raw materials (e.g. quality) on the plant. There is an obvious interaction of the plant with the operators. The plant produces empirical data that is being stored in the control system and archive system. The raw data is cleaned in a pre-processing step before being passed to the machine learning algorithm that produces the mathematical model. The model is used by an optimization method that outputs action advice to the operators. The reaction of the operators, or the change of the external conditions, then induces a continuous circle of further computations and advice.*

The general black box that we will use to hold the Müller-Rochow synthesis has five principal features:

1. There are various slots into which you feed raw materials such as silicon, copper and so on.

2. There are some pipes where the various end products come out of the box.

3. The box has a few dials and buttons with which you can act upon the system. These will be called the *controllable* variables, $\vec{c}$.

4. The box has various gauges that display some information about the inside of the box such as various temperatures and pressures. These variables change in dependence upon the controllable ones but cannot directly be controlled and thus will be called *semi-controllable* variables, $\vec{s}$.

5. The box also has gauges that display some information about the external world such as market prices for end products or the outside air temperature. As these variables are determined by the

external world, we have no influence over them at all. These will therefore be called *uncontrollable* variables, $\vec{u}$.

Any industrial facility records the values measured by all the gauges and dials in an archive system that is capable of describing the state of the box over a long history. As the underlying chemistry has not changed over time, we therefore have a large collection of "input signals" (controllable) into the unknown process alongside their corresponding "output signals" (semi-controllable) in dependence upon the boundary conditions or constraints (uncontrollable), which, mathematically, are also a form of input signal. This experimental data should allow us to design a mathematical description of the process that would take the form of several coupled partial differential equations [1]. Formally speaking, these equations look like $\vec{s} = f(\vec{c}; \vec{u})$. Mathematically speaking, the uncontrollable variables assume the role of parameters (and hence follow the semi-colon in the notation) in this function.

Discovering this function is the principal purpose of this approach and it is a very complex task. One of the most intriguing features is that all three sets of variables are time-dependent and the process itself has a memory. Thus the output of the process now depends on the last few minutes of one variable and the last few hours of another. These memories of the process must be correctly modeled in order for this function to represent the process well enough to use it as a basis for decision making [2]. In order not to clutter the mathematical notation, we will be skipping the dependence upon time that is explicitly assumed as added to every variable in the paper.

In order to do optimization, we need to define a goal g to maximize, which is a function of the process variables, $g = g(\vec{c}, \vec{s}; \vec{u})$. Using the above modeling approach, the goal becomes $g(\vec{c}, f(\vec{c}; \vec{u}); \vec{u})$, i.e. the goal is now a function of *only* the controllable variables and the uncontrollable *parameters*. The entire exercise is done to produce the function $f(\cdots)$ and to make parameters out of the uncontrollable measurements. Now we have a function of true variables – entities that we can change at will.

Optimization theory can be applied to this in order to find the optimal point $\hat{c}$ at which the goal function assumes a maximum, $g_{max} = g(\hat{c}, f(\hat{c}; \vec{u}); \vec{u})$. As the location of the optimal point is computed in dependence upon the goal function as described above, it becomes clear that the optimal point is, in fact, a function of the uncontrollable measurements, $\hat{c} = \hat{c}(\vec{u})$. Now we have the optimal point at any moment in time. We simply determine the uncontrollable measurements by observation and compute the optimal point that depends only upon these measurements [3].

Thus we arrive at our final destination: The correct operational response $\vec{r}$ at any moment in time is thus the difference between the current operational point $\vec{c}$ and the optimal controllable point $\hat{c}(\vec{u})$, i.e. $\vec{r} = \hat{c}(\vec{u}) - \vec{c}$. This response $\vec{r}$ is what we report to the control room personnel and we request them to implement. Ideally, the plant is already at the optimal point in which case the response $\vec{r}$ is the null vector and nothing needs to be done. As a result of the plant personnel performing the response $\vec{r}$, the optimal point will be attained and an increase of the goal function value will be observed; this increase is $\Delta g = g_{max} - g(\vec{c}, f(\vec{c}; \vec{u}); \vec{u})$, which we can easily compute and report as well [4]. The relative (percentage) increase $\Delta g_{rel} = \Delta g / g(\vec{c}, f(\vec{c}; \vec{u}); \vec{u})$ has been found, in this example, to be approximately 6%, see results section for details.

Please note carefully that the response $\vec{r} = \hat{c}(\vec{u}) - \vec{c}$ is a time-dependent response even though we have skipped this dependency in the notation. Thus, we do not necessarily proceed from the current point $\vec{c}$ to the optimal point $\hat{c}(\vec{u})$ in one step. Most often it is important to carefully negotiate the plant from the current to the optimal point and this journey may take a macroscopic amount of time – sometimes several hours.

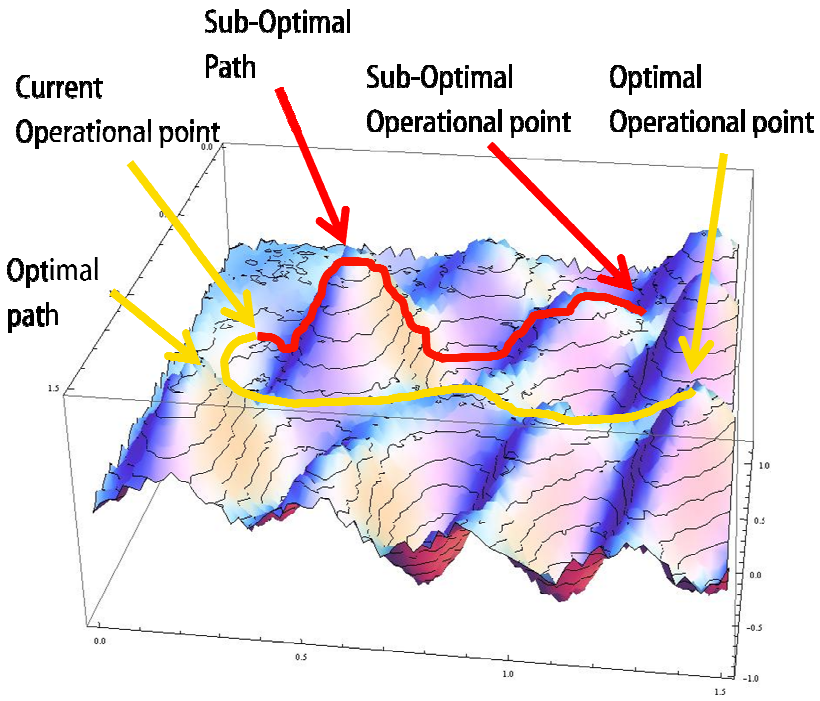

Figure 3.: *The dependency of the goal on two controllable variables. The upper path displays the reaction of a human operator and the lower path displays the reaction of the optimization system. The paths are different and arrive at a different destination. The optimized path is better than the human determined path by approximately 5% as measured by the goal function.*

Figure 3 displays this problem graphically using real data taken from the current process. The two axes on the horizontal plane indicate two controllable variables and the vertical axis displays the goal function. We can easily see that the change in a controllable variable can produce a dramatic change in the goal. The two paths displayed represent the reactions to the current situation by a human operator (the upper path) and the computer program (the lower path). The initially begin on the left at the current operational point. Because of their differing operational philosophies, the path deviates and eventually arrives at different final states. This is a practical example of the human operator making decisions that he believes are best but that are, in fact, not the best possible.

Computationally we plan the best path between the current point and the optimal point using the calculus of variations as developed in physics in the form of Feynman path integrals. This method is capable of determining a path through a multi-dimensional space that is optimal in the integrated goal function value over the whole course of the path. This theory goes beyond the scope of this paper.

Mathematical Process

In terms of mathematical operations, there are two critical methods involved there. First, we must create the function $\vec{s} = f(\vec{c}; \vec{u})$ from experimental data [5]. Second, we must find $\hat{c}$ from $g_{max} = g(\hat{c}, f(\hat{c}; \vec{u}); \vec{u})$.

The first step is made by methods from a field of mathematics called machine learning [1]. This branch has the goal of determining functions from data by using particularly flexible functional templates as a basis and computing the best possible values for the parameters of this template such that the function arrived at fits the data in the best possible way (in the least-squares sense). Particular emphasis must be

paid to the fact that the process has memory and thus depends upon its own history. A special class of methods, known as recurrent neural networks, is particularly good at capturing this time-dependence and so they are used here [2]. It would be beyond the scope of this paper to describe how these are arrived at.

The second step is a multivariate optimization computation. This does not involve time as we are attempting to find the optimal point at the present time. Therefore, this problem can be solved using any global optimization methodology [6]. We must be careful not to use a local method as this goal function will have a great many local minima in practice. From the remaining choices for methods, we apply simulated annealing. The precise manner in which to set up this seemingly simple algorithm is also beyond the scope of this paper [7].

Done properly, both steps can be automated such that the computer continuously keeps track of the optimum point and alerts the user to necessary actions in order to keep the physical process at the computed optimum. The methods of machine learning work in such as way as to update the model with every new measurement. Therefore, the model validates itself over time as it always checks its performance against experimental verification and alters itself if necessary. Even if changes are made to the process, the model will learn them autonomously after some time has passed.

Specific Design

For the specific current application, the molecules are produced in three separate reactors and then brought together for shipment. We are to optimize the global performance of the plant but are able to make changes for each reactor separately. We now outline the variables that were under our influence. These should be understood as being per reactor.

In this case, the controllable variables $\vec{c}$ were the following: Temperature of the reactor, amount of raw material to the jet mill, steam pressure to the jet mill, amount of Methylene Chloride (MeCl) to the reactor, pressure of the reactor and others relating to the processes before the synthesis itself.

The uncontrollable variables $\vec{u}$ were the following: X-ray fluorescence spectroscopy measurements on 17 different elements. The semi-controllable variables $\vec{s}$ are the other variables that are measured in the system. In total, there were almost 1000 variables measured at different cadences.

The goal function is the financial gain of the reaction. We compute the input raw materials and the output end products. Each amount is multiplied with the currently relevant financial cost or revenue. The final goal is thus the added value to the product provided by the synthesis. We desire this to assume a maximum. This function is dominated by two effects: Di is the most valuable end product and so we wish to maximize its selectivity and the overall yield represents the profit margin and so we wish to maximize it also. Possible conflicts between these criteria are resolved by their respective contributions to the overall financial goal. In the results, we will focus on these two factors.

Results

The results reported here were obtained in an experimental period lasting three months and encompassing three reactors. During one period, the optimization was used and catalyst added only when deemed useful by the human operator; we call this the *evaluation* period. During a second period, the optimization was fully utilized including for adding catalyst; we call this the *usage* period. During a third period, the optimization was not used at all; we call this the *reference* period.

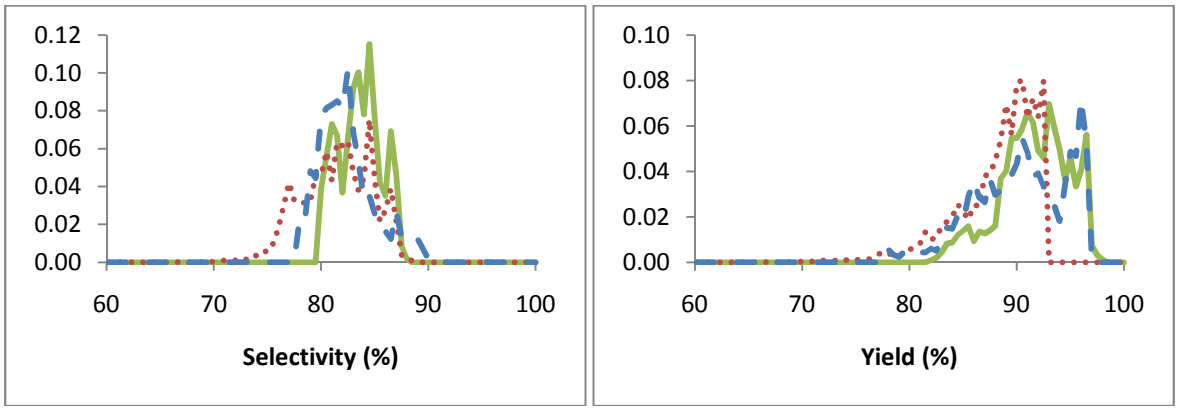

Figure 4.: *The probability distribution functions for selectivity and yield of Di for periods in which the optimization was not used (dotted), used for controllable variables without the catalyst (dashed) and used fully without restrictions (solid).*

We may observe the results in figure 4. In each graph, the dotted line is the reference period, the dashed line is the evaluation period and the continuous line is the usage period. What is being displayed is the probability distribution function of the observed values. This way of presenting the results allows immediate statistical assessment of the result instead of presenting a time-series.

It is apparent, from the images alone, that we increase the selectivity and the yield with more use of the optimization and that we decrease the variance in both selectivity and yield as well. Numerically, the results are displayed in table 1. Decreasing the variance is desirable because it yields a more stable reaction over the long term and thus produces its output more uniformly over time.

We may conclude that the selectivity can be increased by approximately 2.9% and the yield by approximately 5.1% absolute. Together these two factors yield an increase in profitability of approximately 6% in the plant. *We emphasize that this profitability increase of 6% has been made possible through a change of operator behavior only (as assisted by the computational optimization) and no capital expenditures were necessary.*

	Selectivity (%)	Yield (%)
Reference	79.8 ± 3.6	86.6 ± 4.2
Evaluation	79.9 ± 2.5	89.7 ± 4.3
Usage	82.7 ± 1.9	91.7 ± 3.2

Table 1: *The results numerically displayed. For both selectivity and yield, we compute the mean ± the standard deviation for all three periods.*

The practical setup of this optimization took approximately two days of time for the operating personnel. The computation time for the computer to construct the necessary functions is about one month. The computer interfaces for input and output of the data are standardized in the industry and can thus be applied readily without delay. Thus, within about one month, the model can be fully operational without occupying the operators for much time. The approach is thus practicable in a real industrial plant.

References

[1] Bishop, C.M.: Pattern Recognition and Machine Learning. Heidelberg: Springer 2006
[2] Mandic, D., Chambers, J.: Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability. Hoboken: Wiley 2001
[3] Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes. Cambridge: Cambridge University Press 2010
[4] Bangert, P.D., Czernitzky, J.-A. (2010): *Increase of Overall Combined-Heat-and-Power (CHP) Efficiency via Mathematical Modeling*. Proceedings of the VGB PowerTech e.V. Conference "Fachtagung Dampferzeuger, Industrie- und Heizkraftwerke 2010"
[5] Bangert, P.D. (2010): *Two-Day Advance Prediction of a Blade Tear on a Steam Turbine of a Coal Power Plant*. 175--182. in Link, M. (2010): Schwingungsanalyse & Identifikation -- Proceedings of the Conference of 23./24.03.2010 in Leonberg, Germany. (VDI-Berichte No. 2093) ISBN: 978-3180920931.
[6] Bangert, P.D. (2009): *Braids and Knots*. 1--73. in Berger, M.A. et. al. (2009): Lectures on Topological Fluid Mechanics. (Springer Verlag, Berlin) ISBN: 978-3642008368.
[7] Bangert, P.D. (2005): *Optimizing Simulated Annealing*. Vol. 3. 198-202. Proceedings of SCI 2005 -- The 9th World Multi-Conference on Systemics, Cybernetics and Informatics from 10.-13.07.2005 in Orlando, FL, USA.

Optimization of the Müller-Rochow Synthesis of Silanes

Dr. Patrick Bangert, CEO
algorithmica technologies GmbH
Außer der Schleifmühle 67, 28203 Bremen, Germany
Tel: +49 (0) 421 337-4646
email: p.bangert@algorithmica-technologies.com
www.algorithmica-technologies.com