Evaluation and development of strategies for pooling in preparative chromatography

By: Michael Persson

Superviser: Niklas Andersson

Examiner: Bernt Nilsson

Abstract

Computer simulation and modelling was used to simulate a real time environment in preparative chromatography to evaluate the performance of three different strategies for pooling control for different levels of robustness. The pooling problem was based around a separation case where three different insulin species were to be separated while disturbances to the modulators potassium chloride, ethanol and sample load could be imposed on the system. The simulation was created with the assumption that the only measurement available would be the UV absorbance at the process outlet. The first strategy implemented was a time based method where the cut points were determined by offline optimization and would then be static. The second strategy started with offline optimization which was used to determine the UV absorbance measurement at the optimal cut points, these UV absorbances were then used to determine cut placement in real time simulation. The last strategy was a predictive method which made estimations of the concentration profiles in the column based on the gathered measurements and subsequently used these estimations to continuously make new optimized pooling decisions in the real time simulation. The parameters investigated for the evaluation of the performance were the process yield, purity and number of batch failures due to unmet purity requirements. The time based strategy showed the best performance when only load disturbances were present and the prediction based strategy showed the best performance when only disturbances to the modulators were present. The UV based strategy had a large percentage of batch failures for all disturbance cases, the strategy only had moderate success at the highest levels of robustness used in this thesis. Results also indicate that the type of disturbance distribution used could play a part in which strategy shows the best performance. The predictive strategy fared better in cases where latin hypercube sampling was used for the disturbance distribution while the time based strategy showed better performance for a normal random disturbance distribution.

Populärvetenskaplig sammanfattning

I detta examensarbete undersöktes olika strategier för att kontrollera upptag av produkter från kromatografisk separation. Prestandan för dessa strategier jämfördes även med en nyutvecklad strategi vars kontrollbeslut baserades på förutsägelser.

Att utveckla nya läkemedel är en kostsam process samtidigt som bara 15-30% av de produkter som utvecklas blir godkända för lansering. Detta medför höga krav på att resten av produktionskedjan är kostnadseffektiv. På grund av höga krav på kvalitét och renhet så står rening för en stor del av kostnaderna i läkemedelsproduktion. En av de mest använda metoderna för rening av läkemedel är kromatografisk separation.

Kromotografi används i syfte att separera olika ämnen i en lösning från varandra och processen skulle kunna liknas vid att anordna ett lopp mellan löpare, cyklister och bilister på olika banor. Om banan är en motorväg så kan det väntas att alla bilister kommer i mål först följt av alla cyklister och sen alla löpare sist. Är det en väldigt kort bana så kommer inte bilisterna hinna få så stort försprång och vissa bilar kanske till och med kommer i mål efter några av löparna. Om banan däremot är lång kommer uppdelningen däremot bli väldigt tydlig. Det är denna typ av uppdelning man vill uppnå i en kromatografisk process. Om olika ämnen generellt kommer i mål vid olika tider så kan man hitta ett tidsintervall då endast en typ av ämne kommer ut och på så vis få en ren produkt.

Valet av det här tidsintervallet kallas poolning och det jag har undersökt i mitt examensarbete är hur det tidsintervallet kan styras för att maximera den mängd produkt man kan få ut utan att påverka renheten negativt.

Det är nämligen så att i ett kromatografiskt system så kan det bara observeras hur loppet mellan de olika ämnena går precis vid målgången och oftast kan man bara mäta hur mycket som går i mål men inte vilken typ de tillhör. Detta betyder att om oväntade saker händer på banan så att alla tävlande inte kommer ut vid de väntade tiderna, så kan det vara svårt att bestämma i vilket tidsintervall som produkt ska plockas ut. Det finns således ett behov av att ha ett bra kontrollsystem för att undvika beslut som leder till dålig produktkvalitet och slöseri på råvaror.

De olika strategierna provades för olika typer av oväntade händelser för att undersöka hur väl de kunde hanteras. Det visade sig att ingen av strategierna var enskilt bäst för alla fall. Olika strategier klarade av att hantera olika typer av oväntade händelser med varierande resultat.

Att veta hur olika strategier klarar av att bestämma tidsintervallet för att plocka ut produkt från en kromatografisk process är värdefullt för att kunna maximera lönsamheten i sin produktion. I mitt examensarbete undersökte jag tre olika strategier för att bestämma detta intervall. Den första strategin tittade på ett standardfall för målgång och bestämde tidsintervallet baserat på det. Den andra strategin tittade på standardfallet men istället för att bara ta tidsintervallet rakt av så kollade den på hur mycket av ämnena som uppmättes precis vid ändpunkterna i tidsintervallet och använda dessa för att bestämma nya tidsintervall för det riktiga fallet. Den sista strategin som provades använde mätningarna på mängden som gick i mål för att gissa hur loppet gick under tiden loppet var igång. Baserat på dessa gissningar försökte den aktivt hitta det bästa tidsintervallet under loppets gång.

Table of Contents

1. Introduction

The pharmaceutical industry today faces several challenges. Investment costs for the development of new pharmaceutical drugs are high while only a small percentage of all new drugs get approved (Hagel, Jagschies, & Sofer, 2008). The industry is very competitive and any increase in productivity and yield can be very important for the profitability of a process. Downstream processing is a crucial part of the production process due to the high demand on patient safety and product quality but this also means that a large part of the production cost stems from this part of the process. The optimization of the downstream processing equipment can therefore be of great interest (Westerberg, 2012).

One of most commonly used methods for purification in the pharmaceutical industry is chromatography. Chromatography has the advantage of being able to separate substances that are chemically very similar and otherwise couldn't be separated using other methods (Hagel et al., 2008). One of the bigger downsides of preparative chromatography is that it is a batch process which puts certain limitations on process design and productivity. Chromatographic processes are highly nonlinear and due to limitations in the ability to monitor these processes, optimal control can be difficult to achieve. This can lead to losses in the potential yield alternatively batch failure due to purity requirements not being met. Developing smart ways of controlling these kinds of processes is therefore of great interest since using such control could have serious potential for improving performance (Westerberg, 2012).

1.1. Aim
The aim of this master thesis paper was to evaluate different strategies used for pooling in preparative chromatography using MATLAB to simulate chromatography and pooling control in a real time environment. The focus was put on general performance during specific types of disturbances. The performance of a process is judged based on the resulting yield and the ability to stay within defined purity constraints. Furthermore the goal was to develop a proof of concept prototype for a predictive pooling strategy and compare its performance with the performance of existing technology.

2. Theory/Background

2.1. Chromatography
Chromatography is a broad concept and covers a lot of different types of applications. Chromatographic processes can be divided into two subcategories, analytical chromatography and preparative chromatography. The main difference between these two applications is the goal of the separation process. If the goal is to gain information about the sample then it is called analytical chromatography, if the goal is to extract material then it is called preparative chromatography. Furthermore chromatographic processes can differ both in scale and in the physical driving force behind the separation (Shirazi, 2006). A chromatogram is the visual representation of the concentration at the outlet of a chromatographic system as a function of time or eluent volume (these concepts can be used interchangeably, eluent volume is described below). An example of a chromatogram with three different substances can be seen in figure 1 (Ettre, 1993).
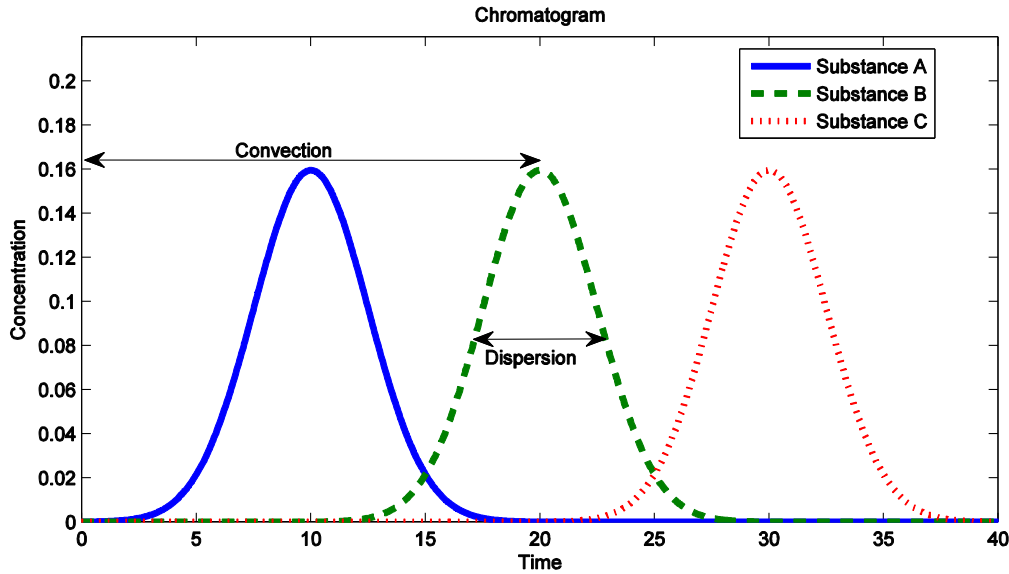
Figure 1 Representation of the concentration of three different substances at the outlet of a chromatographic system as a function of time, also called a chromatogram. The effects that convection and dispersion have on the system are represented with arrows.

A chromatographic system can be divided into three parts, the solute which is the mixture that the chromatographic system aims to separate, the eluent which is a mobile phase in which the solute can move and the stationary phase, usually the packing material in the column, which acts to slow down the individual components of the solute. Differences in the individual tendencies of the substances to slow down is what causes the actual separation in the process (Ettre, 1993).

2.1.1. Adsorption Chromatography
An adsorption process refers to the binding of molecules to a surface through physical or chemical interactions such as van der Waals forces or covalent binding. Adsorption Chromatography is a method by which substances in a homogeneous molecular mixture can be separated based on their individual affinity to the adsorbent of the chromatographic system. Substances with higher affinity will bind to the surface to a greater extent and therefore get delayed in the column while substances with less affinity will pass through more easily (Ettre, 1993).

2.2. Pooling/cut strategies (chromatographic control)
Pooling in preparative chromatography is the retrieval of substances from a chromatographic separation. The decision of where to put the fractionation cut points that decide which substances are retrieved is the pooling problem. The captured pool is the mixture of substances caught within the fractionation cut points, these concepts are illustrated in figure 2. The correct placement of cut points is vital for the economics of a chromatographic system since it has a great effect on the process parameters purity and yield. When making the pooling decisions the objective is most often to maximize the yield of the process defined in equation 1

$$Y = \frac{N_{i,pool}\ (mol)}{N_{i,max}\ (mol)} \tag{1}$$

where Y is yield, $N_{i,pool}$ is the amount of substance i in the pool and $N_{i,max}$ is the maximum amount of substance i that could theoretically be obtained. The maximization of the process yield will often come at the expense of purity (p) as defined in equation 2

6

$$p = \frac{N_{i,pool}\ (mol)}{\sum N_{j,pool}\ (mol)} \qquad\qquad (2)$$

since increasing yield will often mean moving the cut points to make the captured pool wider, including more of the impurities in the process, as can be seen in figure 2.
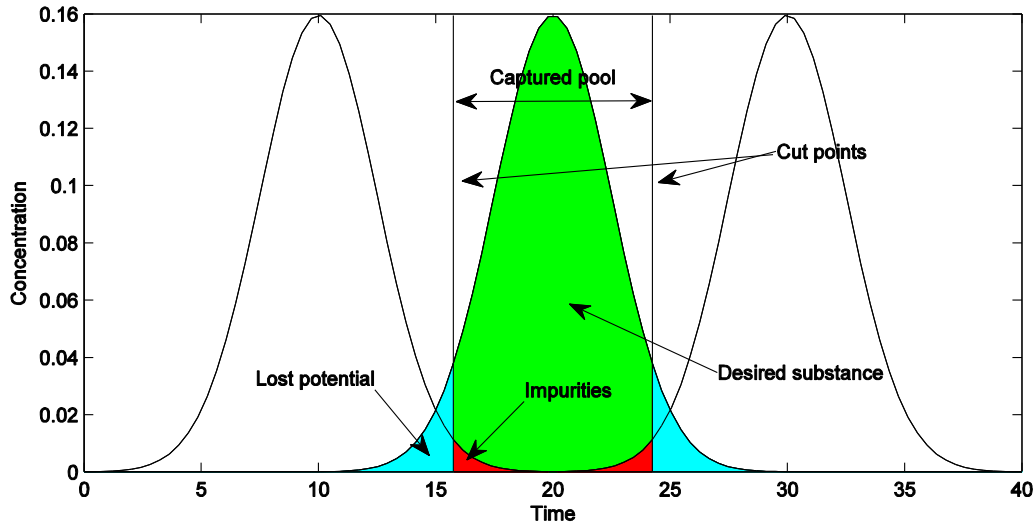

Figure 2 Illustration of a chromatogram with fractionation cut points. The green area represents the desired substance captured within the pool, the red area represents the impurities captured within the pool and the cyan area represents desired substance that did not get captured within the pool.

Traditionally the cut placement decisions are made offline, i.e. before the actual chromatographic separation is performed, based on past data and process understanding. An advantage of this kind of method is that it is non-invasive and does not require any complex control systems to operate. The use of online real time decision making, i.e. pooling decisions being made during the time of the actual separation, has been suggested by (Brestrich, Briskot, Osberghaus, & Hubbuch, 2014).This approach however requires the implementation of more sophisticated process control equipment and software, but in theory has the potential ability to make more flexible decisions when disturbances are present.

3. Methods

All the work presented in this thesis was performed using computers to simulate chromatographic columns and the control of them. The code created for the project can be divided into three different categories, modelling and simulation of chromatographic systems, strategies for chromatographic pooling and statistical analysis of performance during system disturbances.

3.1. Modelling/Cases

It was necessary to create models which could be used to produce chromatograms that displayed general behaviour due to process disturbances in a feasible manner. This kind of data is necessary to be able to properly test the performance of the pooling strategies evaluated in this project. Modelling was done with two different approaches, the first was using Gaussian curves as an abstract general case and the second was fitting a competitive Langmuir kinetic dispersive model to experimental data obtained from (Johansson et al., 2015).

7

3.1.1. Gaussian curves

Chromatographic systems produce chromatograms, due to the physical nature of chromatography these concentration profiles tend to display behaviour similar to Gaussian distributions. Gaussian curves can therefore be very useful for describing chromatograms (Jönsson, 1978), the equation for a Gaussian curve can be seen in equation 3

$$(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)} \tag{3}$$

where μ is the mean and represents the central position of the peak. It has the effect that convection has in a physical system as seen in figure 1. σ is the standard deviation and represents the broadening of the peak, it has the effect that dispersion has in a physical system as seen in figure 1. This means that Gaussian distributions can be used to create generalized simulations of chromatograms. These chromatograms might not be useful for determining how physical disturbances affect process outputs but can be useful as a tool for studying the effects of more abstracted disturbances such as changes in elution time, peak height or peak width. The main way that Gaussian chromatographic models were used in this thesis was as a fitting model for deconvolution which is explained in more detail in section 3.3.5..

3.1.2. Competitive Langmuir/kinetic dispersive model

The calibration of the model was based on data obtained from (Johansson et al., 2015). The study by (Johansson et al., 2015) was a separation of three different insulin species (insulin aspart, insulin desB30 and insulin ester) in a Tricorn chromatographic column, from GE Healthcare, packed with absorbent from FeF Chemicals A/S. The experiments were performed under isocratic conditions with varying concentrations of two different modulators (Potassium chloride (KCl) and Ethanol (EtOH)). After evaluation of the data from this study, three data sets were selected as the basis for a calibration used in this thesis.

There are different models that can be used for simulation of chromatographic systems. For this thesis a lumped rate model was chosen. The lumped rate model simplifies the column equation by not taking pore diffusion in the stationary phase into account and instead viewing all such behaviour as adsorption kinetics. The lumped rate model can be seen in equation 4 (Westerberg, 2012)

$$\frac{\partial c}{\partial t} = D_L \frac{\partial^2 c}{\partial z^2} - u_{lin} \frac{\partial c}{\partial z} - F \frac{\partial q}{\partial t} \tag{4}$$

where c is the concentration in the bulk and q is the concentration in the stationary phase. D_L is the axial dispersion coefficient, u_{lin} is the velocity of the interstitial liquid and F is the volume ratio between stationary and mobile phases. Three different parts contribute to equation, $D_L \cdot \partial^2 c/\partial z^2$ represents the diffusion in the system, $u_{lin} \cdot \partial c/\partial z$ represents convection in the system and $F \cdot \partial q/\partial t$ represents the kinetics of the system (Westerberg, 2012).

A competitive Langmuir model was used to describe the adsorption kinetics of the system. It is based on the Langmuir isotherm seen in equation 5

$$q = \frac{q_{max}K_{eq}c}{1+K_{eq}c} \tag{5}$$

This equation describes the equilibrium of adsorbed substance on the surface q with respect to the concentration of the substance in the bulk c. Where q_{max} is the maximum possible concentration in the stationary phase and $K_{eq} = A/q_{max}$ where A is the partitioning coefficient. The competitive aspect comes in from the fact that there is more than one substance adsorbing to the surface. This is taken into

account by using equation 6 to calculate q_{free} which is the total amount of free adsorption sites (Westerberg, 2012).

$$q_{free} = q_{max}\left(1 - \sum_{j=1}^{N}\left(\frac{q_j}{q_{max,j}}\right)\right) \tag{6}$$

To be able to simulate non-equilibrium conditions adsorption and desorption coefficients are introduced, k_{ads} and k_{des}. If $K_{eq} = k_{ads}/k_{des}$ then equation 7, seen below, can be used to describe the adsorption kinetics (Westerberg, 2012).

$$\frac{\partial q_i}{\partial t} = k_{des,i}\left(K_{eq}c_i q_{max,i}\left(1 - \sum_{j=1}^{N}\frac{q_j}{q_{max,j}}\right) - q_i\right) \tag{7}$$

In the case studied two types of modulators where used KCl and EtOH, to account for their effect on the process exponents were added as can be seen in equation 8 and 9 (Westerberg, 2012).

$$k_{ads} = k_{ads0}e^{\gamma c_{EtOH}} \tag{8}$$

$$k_{des} = k_{des0}(c_{KCl})^{\beta} \tag{9}$$

3.2. Statistical analysis

Statistical analysis of the different control strategies was performed to determine their performance for different kinds of disturbances. The actual analysis was structured in a way where the types of disturbances, their magnitude and the way they were generated could be decided externally. A chromatographic system was then simulated for each of the disturbance profiles generated and the individual chromatograms from these simulations were sent to be evaluated by the control strategies. Each control strategy was evaluated with different levels of robustness, as defined in section (x).

3.2.1. Disturbances

Disturbances are a reality of the physical world, variations of some sort are present in all manner of processes and chromatography is no exception. Process disturbances in chromatography can be divided into two categories, sporadic and autocorrelated. Sporadic disturbances are made up of factors that mostly depend on external sources such as variations in the composition of the solute or the eluent. Autocorrelated factors are mostly related to the properties of the column and are due to for example degradation of column efficiency (Nagrath, Bequette, & Cramer, 2003). The focus was put on sporadic disturbances in this thesis, although some of the concepts and results could be applicable to the handling of autocorrelated disturbances.

Three different parameters were altered when creating disturbances for the evaluation of strategies in this thesis. The parameters were the concentrations of the modulators EtOH and KCL respectively, and the total sample load introduced to the system. These disturbances were chosen since they affect the system in characteristic ways, the modulators affect the retention of the substances in the system i.e. time of elution and the load affects the height of the peaks. By having disturbances that representing these more general concepts it is possible to generalize the results from the strategy evaluation since the strategies do not discriminate between the origins of a disturbance.

Two different ways of creating distributed disturbance samples were used. The first was normal distribution or Gaussian distribution. The MATLAB function "normrnd" was used to generate this kind of sample distribution. The method was latin hypercube sample (LHS), explained in further detail below.

3.2.2. Latin hypercube sampling

LHS is a method by which sample points can be generated in a multidimensional space while avoiding overlapping sample points. The principle is that a given square grid with sample points is a latin square if each row and column in the grid only contain one sample point. This principle translates in the same way when additional dimensions are added. The benefit of this method is that a well distributed sample that covers a great area of possibilities is obtained while minimizing the amount of actual sample points (Olsson, Sandberg, & Dahlblom, 2003), an example of a latin square grid can be seen in figure 3. In this thesis the MATLAB function "lhsdesign" was used to create distributed samples for different combinations of system disturbances.
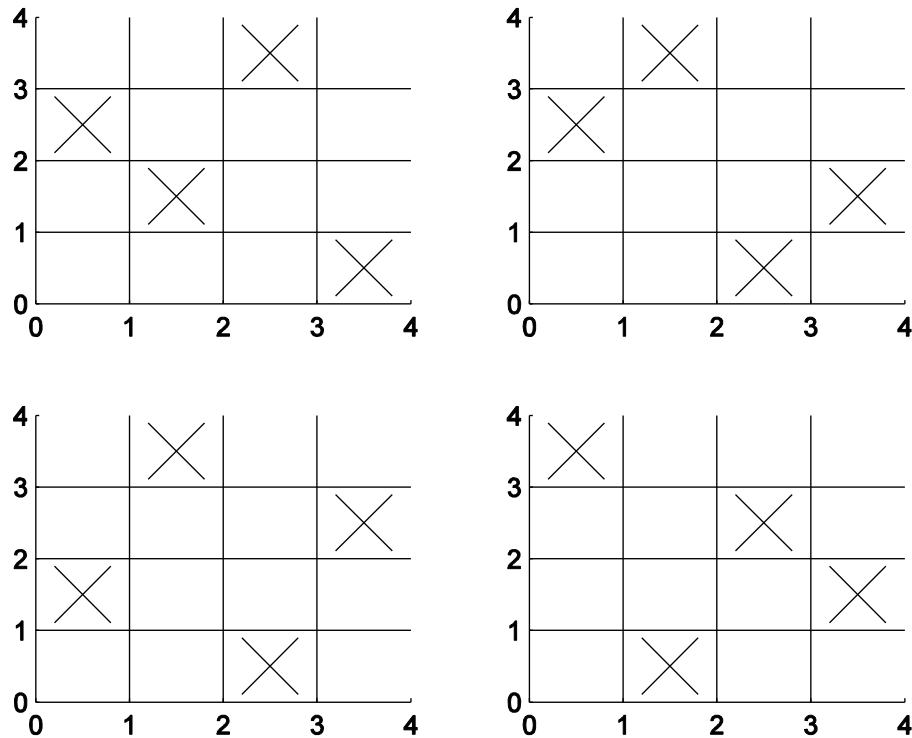

Figure 3 Four examples of latin squares, the underlying concept is to create point distributions within a grid with no overlap in either rows or columns. This concept can be extended to multiple dimensions.

The use of latin hypercube sampling can be advantageous for identifying particular weaknesses of the different strategies since it will cover the entire spectrum of possibilities equally.

3.3. Strategies

The general approach to pooling strategy design in this project was based on a couple of assumptions. The first assumption was that data for a nominal case would be available. It was also assumed that some kind of measurement that could be used for real time decisions would be available (e.g. UV-absorbance, conductivity etc.). Each strategy was implemented in a way where it would act as if it were functioning in real time. Three different strategies were implemented for testing of pooling control. The strategies were a time based approach where cut points were predetermined, an approach where specific UV absorbance measurements triggered pooling decisions and an approach where UV absorbance measurements were used to make predictions about the chromatogram and this in turn was used as a basis for dynamic optimization of the cut points. It is worth mentioning that no data was available for the how these methods would be implemented in real applications, this means that the implementations used in this thesis were based on personal interpretations.

3.3.1. Robustness

Process variations are a reality of systems where disturbances are present, a systems ability to cope with such variations without failing is called the robustness of the system. When optimizing chromatographic systems, yield and purity often become contradictory goals, one can only be increased at the expense of the others. Since chromatography is often used for the purification of high value compounds, such as pharmaceuticals, maximizing yield is a high priority for process economy. At the same time such specialty chemicals often have very high demands on purity due to their nature as specialty chemicals. This leads to operating points near the boundary value of the purity demand when optimizing for process economy. The problem with such an operating point is that it is not very robust, small disturbances can have a great impact on the process leading to purity demands not being met which in some cases can mean the failure of an entire batch. To ensure the robustness of the process some of the potential yield has to be sacrificed so that the purity demands are met even when there are variations in the process (Westerberg, 2012).

3.3.2. Pooling cut optimization

The main tool used to obtain optimal pooling cut points was the program *simplexpooling*, developed at LTH, it is specifically designed to make optimal pooling decisions on well-defined systems. When given the concentration profiles of all individual components at the outflow in a chromatographic system it can utilize different optimization algorithms to find cut points optimal with respect to process yield. Different constraints can also be imposed on the system such as minimum yield or purity. For this project only purity constraints were used when searching for optimal cut points.

3.3.3. Time based cuts

The principle utilized in the time based cut strategy is that a nominal case is identified and the cut points are obtained by offline optimization using *simplexpooling*. The cut times obtained from *simplexpooling* were then used without changes regardless of the disturbances imposed on the system. Robustness was taken into account by the initial optimization, higher robustness was obtained by increasing the purity constraints in *simplexpooling*.

3.3.4. UV based cuts

The UV based cut strategy works on the principle that when a certain UV absorbance measure is met pooling starts or stops. In this project this was implemented in a very similar manner to how time based cuts were implemented. First a nominal case was identified and an offline optimization was performed. But instead of just using the optimal cut times directly the UV absorbance at these times was identified. These UV absorbances were then used in the pooling decision routine. When an UV absorbance corresponding to the optimized UV absorbance was measured, pooling started or stopped, a schematic illustration describing the UV based strategy can be seen in figure 4. Since UV absorbance is a lumped measurement it does not discriminate between which substances caused the UV absorbance and depending to the nature of the separation problem the same UV absorbance might be measured more than one time in the same chromatogram. As can be seen in figure 4 the UV measured at the optimal time of the first cut occurs twice before the optimal time. To avoid activating pooling in the wrong area of the chromatogram the routine was set to only make measurements and decisions during limited time spans. The allowed time span for cut decisions was defined as a percentage of the time difference between the optimized cut times before and after the optimized cut times. Furthermore to avoid getting no pooling at all, in the case that the desired UV absorbance was never measured, a cut would automatically be placed in the end of the allowed decision time span.
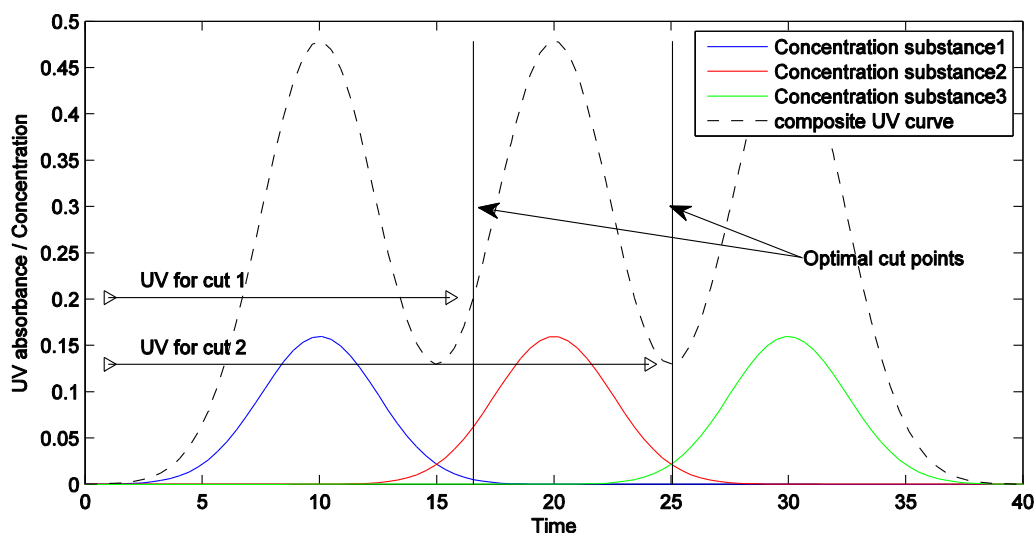
Figure 4 An illustration of how the UV based strategy works. Note that the cut points in this figure are placed arbitrarily just to show an example. The composite UV curve is a representation of the measured signal.

Robustness was handled much the same as in the time based cut strategy, higher robustness was obtained by an increase of the purity constraints in *simplexpooling* in the offline optimization.

3.3.5. Prediction based cuts

The main difference between the prediction based cut strategy and the previously described strategies was the implementation of simulated online real time cut decisions. To be able to make pooling optimization decisions in real time, information about the concentration profiles of the individual components is needed. The problem in a lot of different real applications of detectors in chromatographic systems is that only lumped measurements such as UV absorbance are available. This problem could be solved with novel solutions in chromatographic detectors such as Diode array detectors used with partial least squares regression to identify the concentration of individual substances (Brestrich et al., 2014), but in this thesis it was assumed that only lumped sum measurements would be available.

To get around the problem of lumped measurements in this project, the lumped measurements were deconvoluted using least squares fitting. Given the assumption that there was a nominal case, Gaussian curves were first fitted to the individual concentration peaks of each substance. Later when receiving data from another simulation the Gaussian curve parameters obtained from the nominal case were used as initial guesses for estimating Gaussian curve parameters to fit the data measured in in the simulation. By trying to fit a Gaussian curve to each peak in the lumped measurement an estimation of the individual concentration profiles was obtained.

This also served as a very basic prediction method that produced estimates of the chromatogram that were then used by *simplexpooling* to make the actual pooling decisions. The idea of the method was to simulate a real time environment by mimicking measurement sampling and then letting the strategy make a decisions in every theoretical sample point. The nominal case was used as a reference value and was overwritten with the simulated disturbed case for each new sample point, this simply meant that the chromatogram consisted of data from the disturbed case up to the point of the current sample time and after that consisted of data from the nominal reference point. An example of how this might look can be seen in figure 5.A. After every new sample the chromatogram was deconvoluted, the deconvolution served as the predictor since it was basing its estimation on a combination of nominal and disturbed data, an example of such deconvoluted chromatograms can be seen in figure 5.B.
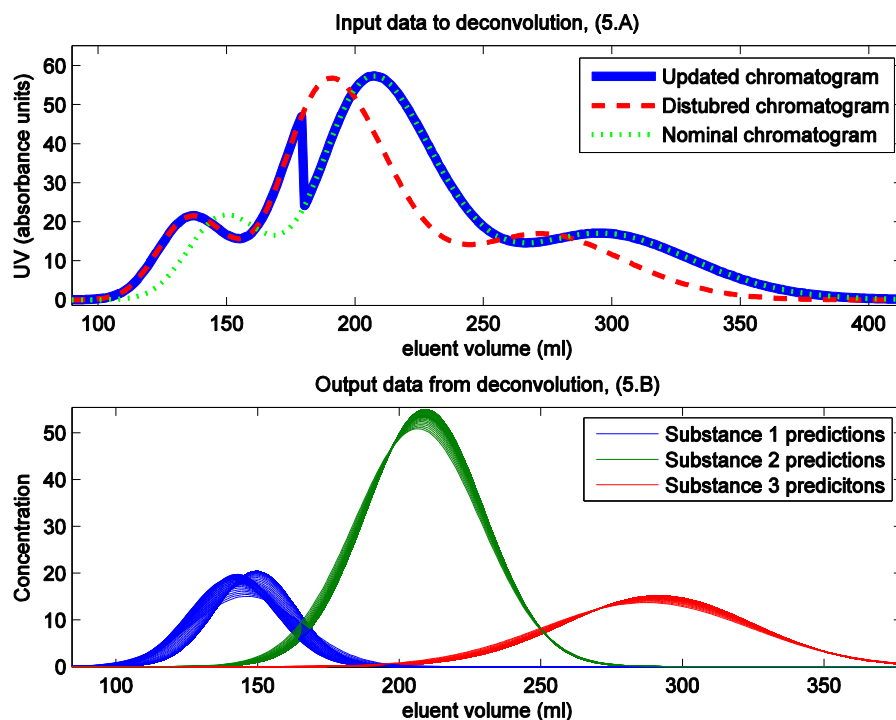
Figure 5.A shows how the data from the disturbed chromatogram and the nominal chromatogram are combined before being sent to deconvolution, note the boundary between disturbed and nominal data in the updated chromatogram at around 180 ml eluent volume. Figure 5.B shows all predictions made up to the point of the latest data update. Note that predictions for each individual substance are made simultaneously.

After deconvolution *simplexpooling* was then used to find optimal cut points for the estimated data. If the cut point obtained from optimization corresponded to the current sample time, pooling would start and likewise when the next cut point corresponded to the time, pooling would stop.

Robustness in this case was handled in a way very similar to the previous cases. Robustness was increased by increasing the purity requirements in *simplexpooling*.

4. Results

4.1. Calibration

After implementation of the kinetic dispersive model the model parameters were calibrated to fit with the experimental data obtained from the insulin separation case. Three sets of experimental data were used for the calibration and the fitting was performed for all these cases simultaneously, the fitting obtained can be seen in figure 6. The calibration was considered sufficiently accurate when the systems response to disturbances showed the same general tendencies as the experimental data i.e. the peaks moved in the same directions as could be observed from the experimental data when altering parameters.
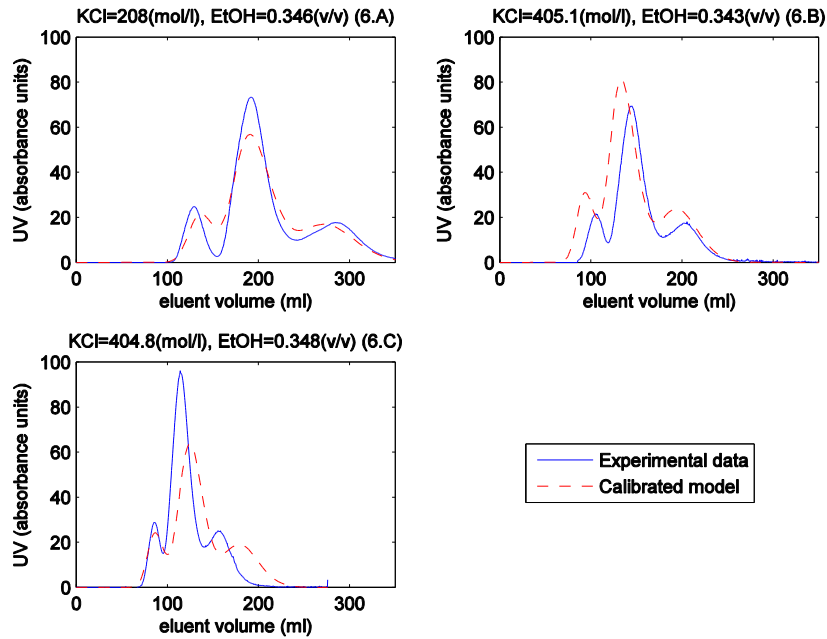
Figure 6(A,B,C) show the chromatograms used for the calibration of the parameters of the model used for simulations in this thesis, with the respective fitting obtained. The text above each figure describe the conditions for each data set used in the calibration.

After the calibration was complete simulations with each individual disturbance were performed. The actual effect that disturbances had on the simulated system can be seen in figure 7. As can be observed both of the modulators, KCl and EtOH, have very similar effects on the system. Increasing them decreases the retention of all substances in the column making elution time shorter and conversely decreasing then increases retention. Increases to the sample load result in heightening of the entire profile and decreases result in lowering of the profile.
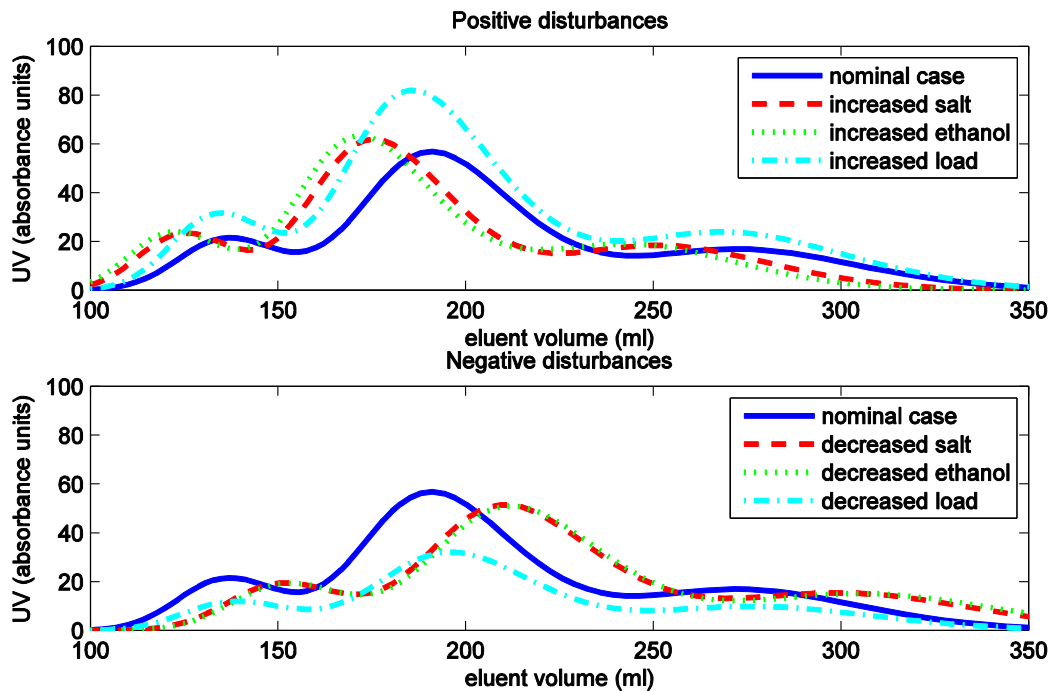

Figure 7 results from simulation, UV absorbance as a function of time for different disturbances.

The conditions chosen for the nominal case can be seen in table 1, this case is based on the same conditions as can be seen in figure 6.A.

Table 1 this table contains the operating conditions for the case chosen as the nominal operating point.

Salt concentration	208 mol/m³
Ethanol concentration	0.346 V/V
Load	0.1522 ml

4.2. Sample distributions

Two different methods were used to create distributed disturbance samples, normal random distribution and LHS. Two examples of what the distributions look like can be seen in figure 8. As can be seen in these figures LHS creates evenly distributed samples while the normal random distribution creates samples with an intensity maximum around the nominal point.
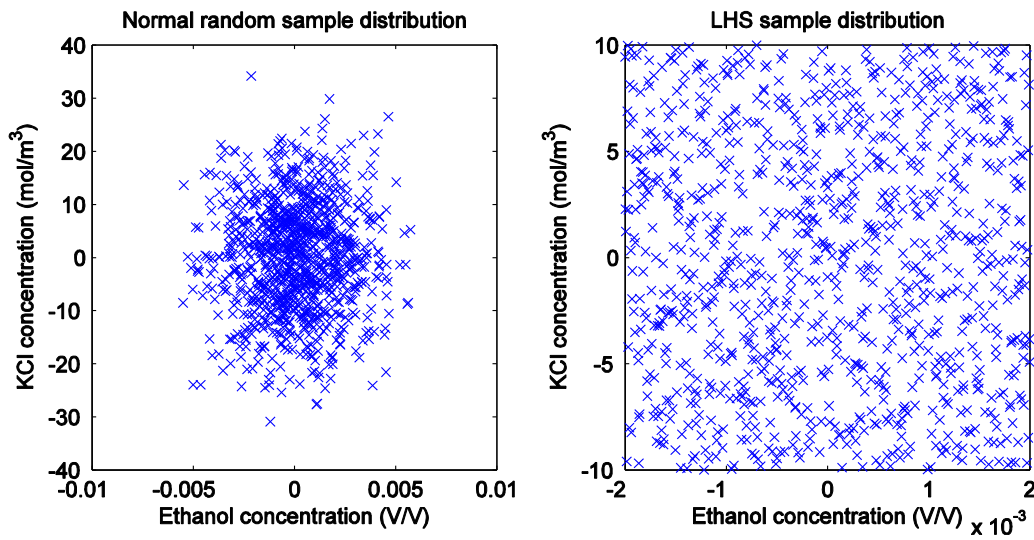

Figure 8 the figures represent examples of sample distribution for two different disturbances. Normal random distribution was used to produce the sample distribution seen in the leftmost figure while LHS was used to produce the distribution seen in the rightmost figure.

4.3. Strategy evaluation results

The pooling cut strategies were tested for many different conditions and combinations of disturbances. All simulation sets were performed with 1000 sample points and a purity requirement of $p = 0.95$.

The results presented below are from one of these simulation sets. The data set used had a normal random sample distribution for all three disturbances with the standard deviations $\sigma_{load} = 0.02$ for sample load, $\sigma_{KCl} = 10$ for the concentration of KCl and $\sigma_{EtOH} = 0.002$ for the concentration of EtOH. For every strategy three specific cases are shown to represent the cut strategies behaviour at the nominal case and two extremes. The extreme points represent all disturbances giving positive addition and negative addition respectively to the process conditions. The points were chosen from the sample set to meet these requirements.

4.3.1. Time based cut strategy

The nature of the time based cut strategy is to be static, the actual cut points never change from case to case. As can be expected this strategy works very well for the nominal case since the actual cut points were optimized for the nominal case. As seen in table 2 the purity for the nominal case never drops below the requirement. As expected the yield decreases as the robustness is increased but the yield is still very close to the theoretical maximum with a yield of 92.7% of the maximum possible yield even for 2% robustness. The fact that the purities are not exactly consistent with purity requirements used for

the optimization is due to the way that the theoretical sampling is performed in the code, only making it possible to perform pooling cuts at discrete points along the volume axis. The chromatograms and the actual pooling cut points made by the strategy can be seen alongside the optimal cut placements, made by *simplexpooling* given all information, in figure 9.

Because of the static nature of the strategy the performance noticeably decreases when it is introduced to systems with large disturbances, as illustrated by the positive and negative disturbance cases in table 2. Since no online decisions are being made the cuts do not adapt and performance losses follow. This is most clearly evidenced by the fact that strategies fall well below the required purity no matter the robustness giving an effective yield of 0 for the cases. This can also be observed in the positive and negative disturbance cases in figure 9 where the cut placements are far away from the theoretical optimum.

Table 2 results from nominal and extreme cases for the time based strategy, r represents robustness. Yield is defined as 0 if the purity requirement of $p = 0.95$ is not met, the value given in parenthesis is the yield captured in the pool regardless of batch failure.

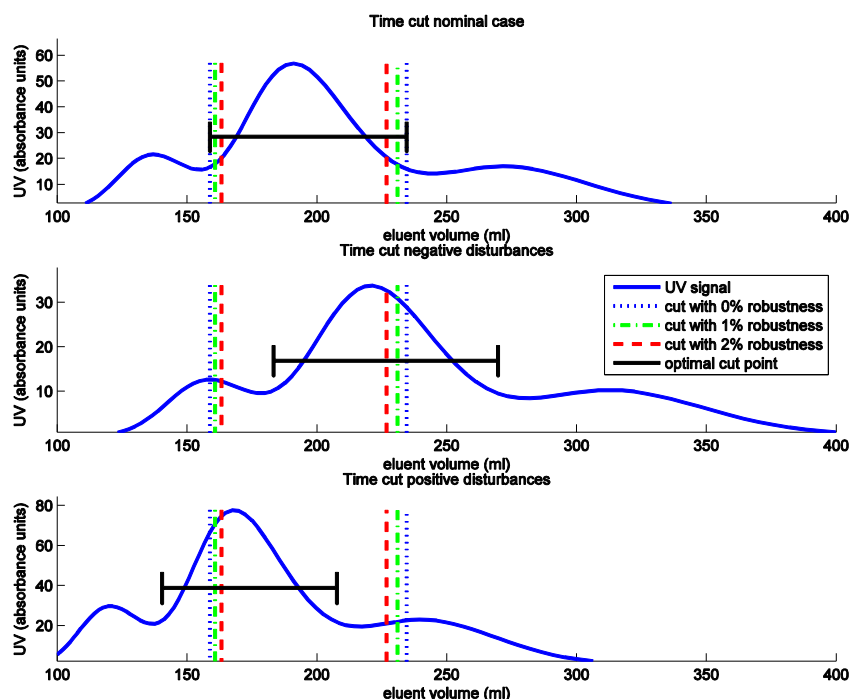| Strategy/Robustness	Purity	Yield	Yield/maxYield
Nominal case			
Timecut r=0%	0.956	0.925	0.987
Timecut r=1%	0.965	0.900	0.961
Timecut r=2%	0.973	0.868	0.927
Negative disturbances			
Timecut r=0%	0.839	0 (0.640)	0.684
Timecut r=1%	0.844	0 (0.586)	0.627
Timecut r=2%	0.848	0 (0.530)	0.566
Positive disturbances			
Timecut r=0%	0.837	0 (0.742)	0.792
Timecut r=1%	0.848	0 (0.695)	0.742
Timecut r=2%	0.860	0 (0.646)	0.689

Figure 9 representations of chromatograms for nominal and extreme cases with cut placements made by the time based strategy as well as the optimal pooling cut points.

16

The performance of the time based cut strategy in yield, purity and number of passed and failed batches can be seen in figure 10. It is compared to the theoretical maximum, which was determined using *simplexpooling*, for the same disturbances. The average performance for the entire set is presented in section 4.3.4. and can be seen in table 7.
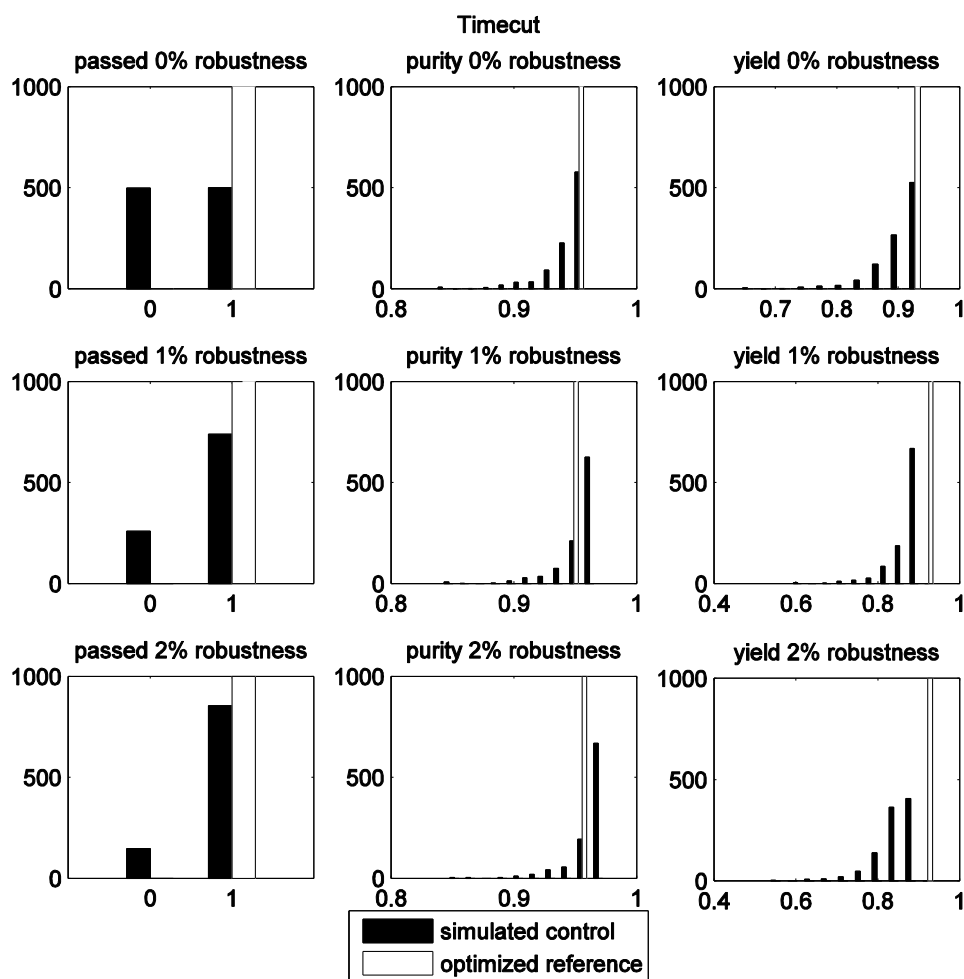

Figure 10 histograms showing the distribution of performance for the time based strategy for different levels of robustness alongside the theoretical maximum performance for the same simulation set. The parameters displayed are number of passed batches (where passed batches are represented by 1 and failed batches are represented by 0), batch purity and batch yield (without consideration for batch failures).

4.3.2. UV based cut strategy

The UV based cut strategy was designed so that cut points would be placed when a UV absorbance measurement corresponding to the UV absorbance at the optimal cut time was encountered. However UV absorbance is a lumped measurement and does not discriminate between which substance caused the UV absorbance. This means that the same UV absorbance signal can and will occur several times in the same chromatogram. In the implementation of the strategy this was solved by imposing a limited time span where the strategy was actually active. The results from three different cases from the same simulation run, as represented for the time based cut strategy, can be seen in figure 11. The problem with encountering the same UV to early can be seen in the nominal case where only the cut placement using the highest robustness actually managed to make a pooling decision within the boundaries of the purity requirements. This is also reflected in figure 12 when comparing the amount of batch failures between the different robustness levels, where lower percentage of the cases with low robustness passed compared to the cases with 2% robustness.

Strategy/Robustness	Purity	Yield	Yield/maxYield
Nominal case			
UVcut r=0%	0.908	0 (0.962)	1.032
UVcut r=1%	0.891	0 (0.953)	1.021
UVcut r=2%	0.968	0.887	0.951
Negative disturbances			
UVcut r=0%	0.967	0.855	0.914
UVcut r=1%	0.967	0.855	0.914
UVcut r=2%	0.969	0.792	0.846
Positive disturbances			
UVcut r=0%	0.464	0 (0.273)	0.291
UVcut r=1%	0.498	0 (0.273)	0.291
UVcut r=2%	0.818	0 (0.258)	0.275
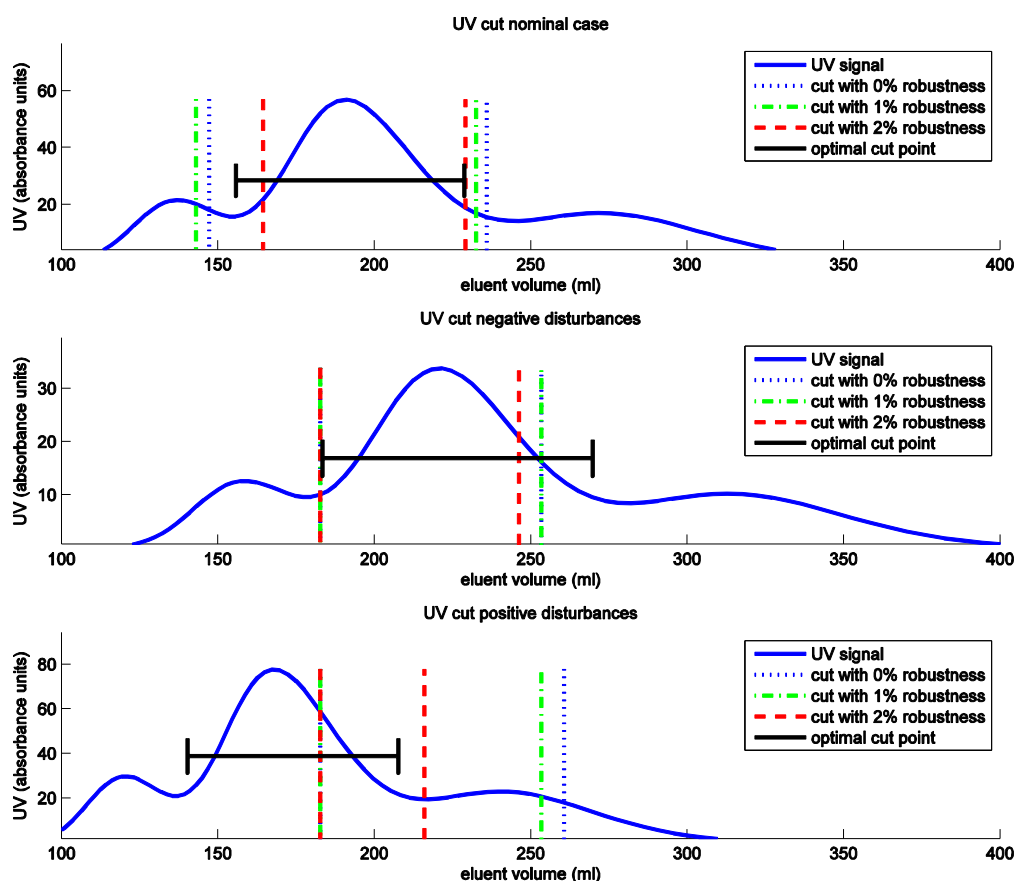
Figure 11 representations of chromatograms for nominal and extreme cases with cut placements made by the UV based strategy as well as the optimal pooling cut points.

The fact that the UV based strategy manages to make as good pooling decisions in the negative disturbance case as it does, as can be seen in table 3, is most likely an exception due to lucky circumstances when taking the data in figure 12 into consideration, where it can be seen that the method very consistently performs badly.

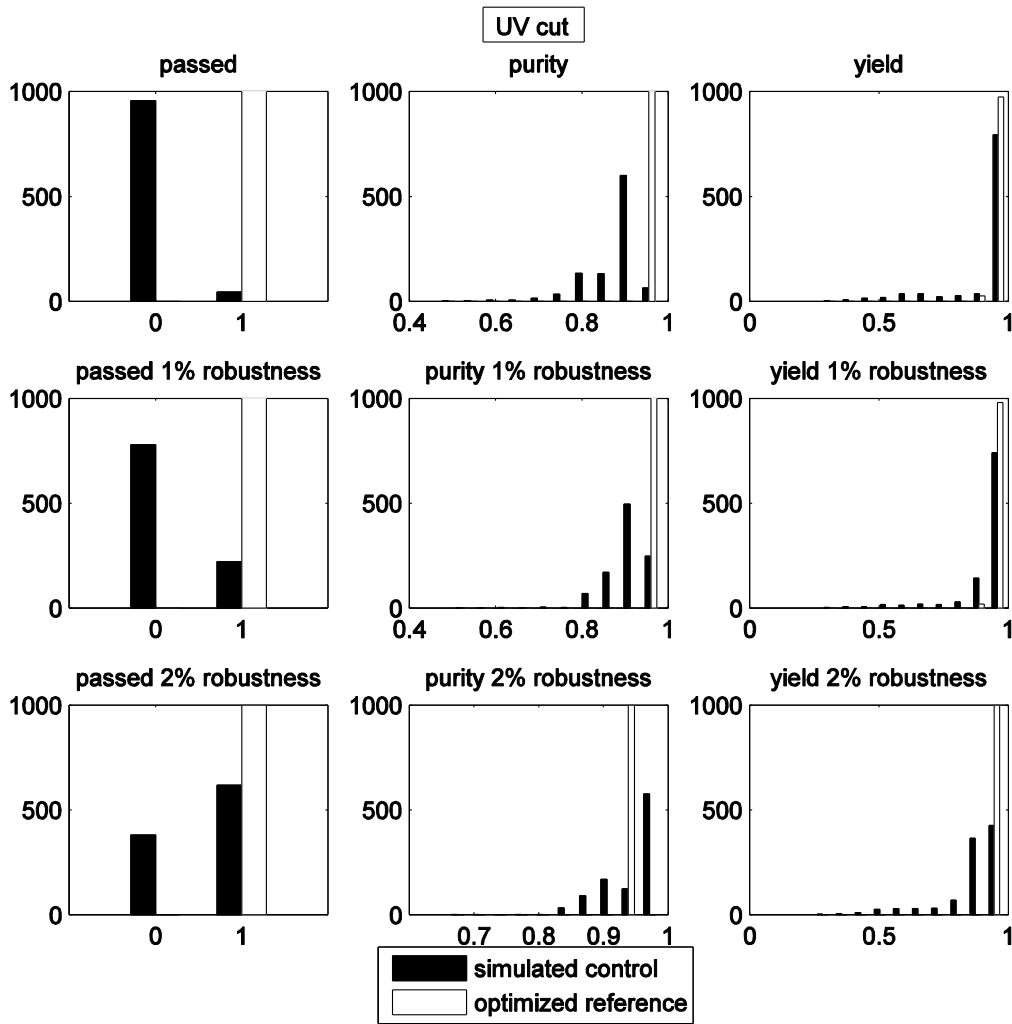*Figure 12 histograms showing the distribution of performance for the UV based strategy for different levels of robustness alongside the theoretical maximum performance for the same simulation set. The parameters displayed are number of passed batches (where passed batches are represented by 1 and failed batches are represented by 0), batch purity and batch yield (without consideration for batch failures).*

4.3.3. Prediction based cut strategy

The prediction based cut strategy used an updating estimation of the chromatogram to make predictions of where the optimal cut points would be. The original estimation is based on the nominal case and as could be expected the pooling strategy produces a yield very close to the theoretical maximum for the nominal case. For the cases with disturbances the results are mixed as illustrated by the negative and positive disturbance cases in table 4. For the case with negative disturbances the strategy fails to find cut points which satisfy the purity demand and for the positive disturbances the purity demand is satisfied for all levels of robustness but the yield is relatively low as can be seen in table 4 with a yield of 71.7% of the theoretical maximum yield for a robustness of 0%.

Table 4 results from nominal and extreme cases for the prediction based strategy, r represents robustness.

| Strategy/Robustness	Purity	Yield	Yield/maxYield
Nominal case			
Predictcut r=0%	0.957	0.918	0.984
Predictcut r=1%	0.970	0.881	0.944
Predictcut r=2%	0.976	0.832	0.892
Negative disturbances			
Predictcut r=0%	0.823	0 (0.703)	0.752
Predictcut r=1%	0.764	0 (0.600)	0.641
Predictcut r=2%	0.821	0 (0.600)	0.641
Positive disturbances			
Predictcut r=0%	0.960	0.672	0.717
Predictcut r=1%	0.967	0.623	0.665
Predictcut r=2%	0.973	0.569	0.607
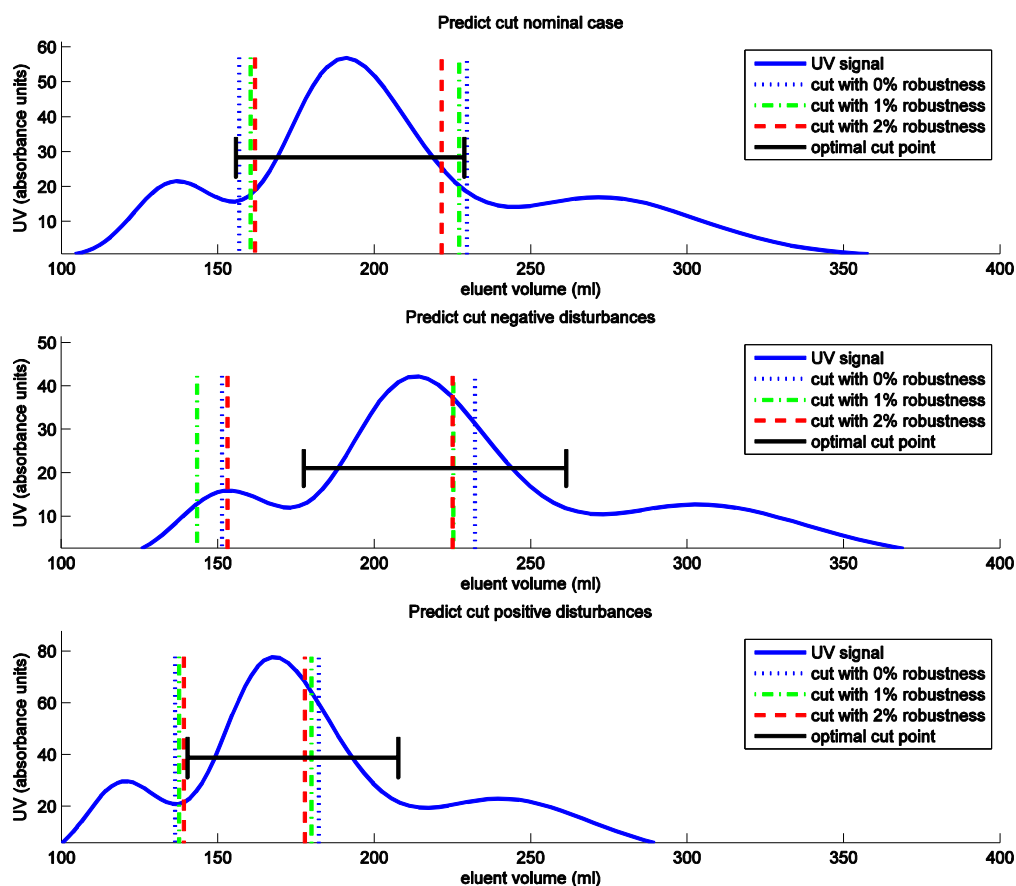
Figure 13 representations of chromatograms for nominal and extreme cases with cut placements made by the prediction based strategy as well as the optimal pooling cut points.

As can be seen in figure 14 the predictive pooling strategy shows great consistence in making pooling decisions that meet the required purity demands, surpassing the other methods on all levels of robustness for the amount of passed batches. On the other hand the strategy seems to tend to make more conservative decisions, exemplified by the positive disturbance case in figure 13. As can be seen in the yield histograms in figure 14 the average yield suffers from this. It should be noted that the reason for

some of the points showing a value of zero for purity is due to a safety measure put in to the code to prevent crashes during simulations. If the strategy failed during evaluation the purity and yield were automatically put to zero.
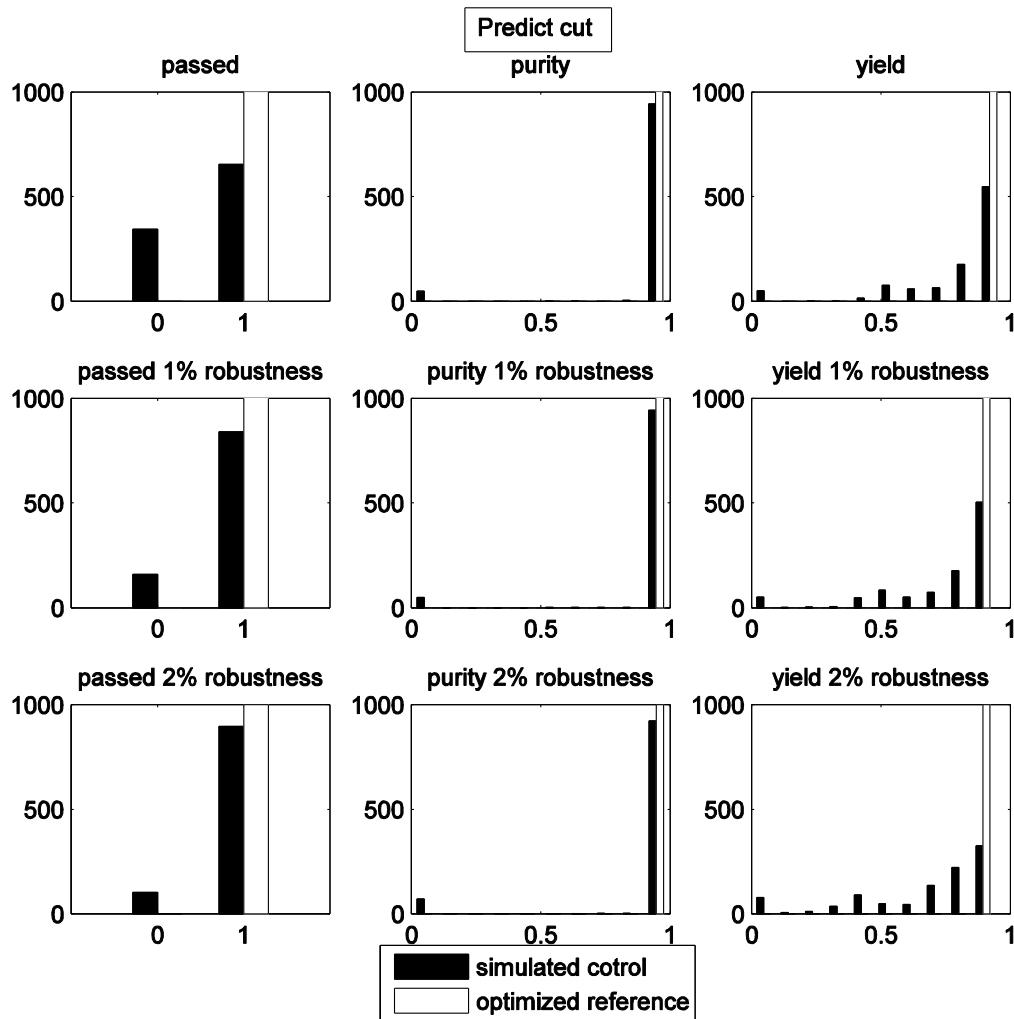

Figure 14 histograms showing the distribution of performance for the prediction based strategy for different levels of robustness alongside the theoretical maximum performance for the same simulation set. The parameters displayed are number of passed batches (where passed batches are represented by 1 and failed batches are represented by 0), batch purity and batch yield (without consideration for batch failures).

4.3.4. Comparison

When comparing the different strategies it is evident that they excel at different types of disturbances. For simulation sets where disturbances were only introduced to the modulators, with a standard deviation of $\sigma_{KCl} = 10$ for the concentration of KCl, $\sigma_{EtOH} = 0.002$ for the concentration of EtOH and a normal random sample distribution, the predictive cut strategy outperformed both of the other methods for all levels of robustness. This can be seen in table 5 which presents the average performance of the strategies for the data set with these conditions.

Set 13	Tc r=0%	Tc r=1%	Tc r=2%	Uc r=0%	Uc r=1%	Uc r=2%	Pc r=0%	Pc r=1%	Pc r=2%
Yield	0.433	0.622	0.696	0	0.090	0.725	0.855	0.807	0.726
Yield/Yield$_{max}$	0.462	0.664	0.744	0	0.096	0.775	0.914	0.863	0.776
Avgpurity	0.941	0.951	0.960	0.897	0.895	0.958	0.954	0.959	0.963

The yield presented in tables 5 to 8 is defined as the total average yield for all cases with the yield of failed batches set to zero.

In contrast to this the time based cut strategy showed the highest average yield when simulating a set with disturbances only to the load, with a standard deviation of $\sigma_{load} = 0.02$ for sample load and a normal random sample distribution, as can be seen in table 6.

Table 6 results from simulation set with disturbances to sample load with a normal random sample distribution, Tc stands for time based strategy, Uc stands for UV based strategy and Pc stands for prediction based strategy. Robustness is represented by r. The highlighted results are the ones that show the best performance.

Set 17	Tc r=0%	Tc r=1%	Tc r=2%	Uc r=0%	Uc r=1%	Uc r=2%	Pc r=0%	Pc r=1%	Pc r=2%
Yield	0.924	0.899	0.867	0.034	0.106	0.489	0.608	0.714	0.726
Yield/Yield$_{max}$	0.987	0.961	0.927	0.037	0.113	0.522	0.650	0.763	0.775
Avgpurity	0.956	0.965	0.973	0.894	0.906	0.947	0.913	0.917	0.897

Another thing that could be observed from the results was that the average performance was affected by the type of sample distribution used for the simulations. When comparing results obtained from a simulation set

Two data sets using different sample distributions were compared, the first using a normal random distribution with the standard deviations $\sigma_{load} = 0.02$ for sample load, $\sigma_{KCl} = 10$ for the concentration of KCl and $\sigma_{EtOH} = 0.002$ for the concentration of EtOH and the second using LHS with the standard deviations $\sigma_{load} = 0.012$, $\sigma_{KCl} = 12$ and $\sigma_{EtOH} = 0.002$. When comparing the two sets the strategy that showed the best performance differs between the different types of sample distribution even though the standard deviations are roughly the same. The time based strategy shows the best performance when the disturbances have normal random distribution, as can be seen in table 7, while the predictive approach shows better performance for disturbances with a LHS distribution as can be seen in table 8. This indicates that the expected disturbance distribution can be of importance when choosing strategies for optimal performance. It is however worth noting that the case with normal random distribution had a higher standard deviation for load disturbances which the time based strategy generally handles better as pointed out previously.

Table 7 results from simulation set with disturbances to all parameters with a normal random sample distribution, Tc stands for time based strategy, Uc stands for UV based strategy and Pc stands for prediction based strategy. Robustness is represented by r. The highlighted results are the show the best performance.

Set 14	Tc r=0%	Tc r=1%	Tc r=2%	Uc r=0%	Uc r=1%	Uc r=2%	Pc r=0%	Pc r=1%	Pc r=2%
Yield	0.460	0.656	0.725	0.037	0.187	0.514	0.521	0.652	0.647
Yield/Yield$_{max}$	0.492	0.702	0.775	0.040	0.200	0.550	0.557	0.697	0.692
Avgpurity	0.943	0.953	0.961	0.876	0.906	0.942	0.907	0.914	0.902

Table 8 results from simulation set with disturbances to all parameters with a LHS distribution, Tc stands for time based strategy, Uc stands for UV based strategy and Pc stands for prediction based strategy. Robustness is represented by r. The highlighted results are the show the best performance.

Set 9	Tc r=0%	Tc r=1%	Tc r=2%	Uc r=0%	Uc r=1%	Uc r=2%	Pc r=0%	Pc r=1%	Pc r=2%
Yield	0.387	0.564	0.643	0.009	0.152	0.568	0.617	0.713	0.676
Yield/Yield$_{max}$	0.414	0.603	0.688	0.001	0.162	0.607	0.660	0.761	0.722
Avgpurity	0.938	0.948	0.957	0.875	0.900	0.940	0.945	0.935	0.935

5. Discussion

The fact that the time based cut strategy outperforms the other pooling strategies for cases where only load disturbances are present can be attributed to the fact that a disturbance to the load will have minimal impact on the elution time of the individual peaks. This means that the optimal cut points will move to a very small extent. Even if a higher load might lead to higher concentrations of impurities in the pool, this is counteracted by the increase in the concentration of the desired substance leaving the purity of the pool at approximately the same level. This behaviour is of course a result of how the load disturbance was designed for the simulation, where a change to the load proportionally changes the concentration of each substance. Having disturbances with variability in the individual concentrations of the different substances might have produced different results.

It is worth mentioning that the deconvolution used for the predictive pooling strategy did not have the ability to alter the total concentration of substances in its predictions. Adding these parameters might have increased the performance of the predictive strategy for load related disturbances. The reason for not implementing them however was the risk of decreasing the accuracy of the predictions due to the increased freedom it would give in fitting.

The fact that the predictive cut strategy showed performance surpassing that of the time based strategy when introduced to disturbances to the modulators was also expected. Since changes to the modulators alter the elution times, the optimal cut points will shift to follow the position of the peaks. This is where one would expect an adaptive method as the predictive strategy to be better than a static method as the time based strategy. These results highlight the strengths of this type of approach to chromatographic control.

Another interesting aspect of the results is the consistently poor performance of the UV based method. This is most likely more due to factors in the implementation of the strategy rather than a reflection of viability of this approach in real life. One factor that might play a part in this is that the definition of robustness is somewhat unfair when comparing the methods. As can be seen in the results the performance of the UV based strategy sometimes increases several times over when comparing lower and higher levels of robustness. Since the method is based on searching for a specific measurement within a fixed timespan, altering the timespan can produce very different results. Changing the robustness changes the timespan and as mentioned the strategy becomes much more viable at higher robustness levels. This indicates that the decision timespans might have been poorly optimized and changing the timespan implementation might produce performance results very different to the results presented in this thesis. Another way of improving the performance of the UV method would be to completely change the approach used to create the decision space. For example, instead of using predetermined timespans, an implementation where the gradient of the UV absorbance is measured to determine the relative position of the chromatogram could be used as a criterion for the decision space. This kind of implementation would of course be very case specific but could nonetheless prove more efficient.

5.1. Validity and sources of error

5.1.1. Experimental data

This thesis has focused on determining the performance of pooling strategies for general types of disturbances. But even with this broad approach determining the validity of the results for general cases is difficult. The focus has basically been on one case study with data taken from a study with the primary focus of determining the effect of different modulators on elution time. The range of deviation for the disturbances in the simulations was chosen to try and give a realistic representation of how a chromatographic process might behave, but depending on the actual system, substances or separation problem the magnitude and type of disturbances might differ radically. This gives some validity to the general approach but applications of these methods to any other case will require further work.

5.1.2. Simulation uncertainty

With any work that is strictly based on computer simulations there is always a risk of missing aspects of a problem that might only become evident when performing actual experiments. These could be technical aspects of real time measurements such as interference from noise which has not been accounted for in the implementation of the strategies.

5.2. Future work

There are several areas that could be worked on to further improve understanding of this subject. Improvements to the implementation of the simulation could increase the feasibility of the results produced. Further development could also improve performance of the individual strategies and help highlight their individual strengths and weaknesses.

By expanding the simulation study with more cases the legitimacy of the results would increase and light could be shed on the generality of the usefulness of the different strategies. The case currently used in the simulation is a relatively simple separation case with clearly defined peaks of Gaussian nature. This does not necessarily represent the behaviour of cases with different kinetics or cases with different conditions imposed on them.

One of the areas that could be the most interesting to investigate would be alternative implementations of predictive cut strategies. The predictive method used in this thesis was a very basic implementation but still showed great promise for certain types of cases. The implementation of a feed forward system, where information gathered prior to the separation could be used to fine tune the prediction model before each batch run, could help improve capability of the predictive approach immensely. There are also other approaches to prediction that could be investigated such as machine learning.

Something that really should be tested further is the validity of these methods in practice with real experimental setups utilizing the suggested methods. This would help illuminate problems that might appear when working with physical systems that might not become apparent in simulations. The evidence presented from such a study would also be stronger for the purpose of determining the viability of the methods suggested.

6.   Conclusions

After evaluation of the results it can be concluded that performance for different strategies can differ radically for different sets of conditions. The type of disturbance as well as the type of sample distribution has an effect on which strategy shows the best performance. The time based strategy shows the best average performance for load disturbances and mixed disturbances when using normal random sample distribution. The predictive strategy shows the best performance for modulator disturbances and mixed disturbances when using a LHS distribution. The UV based strategy showed poor performance in almost all cases, only at higher robustness levels did it show performances in acceptable ranges. This might indicate inferiority of the UV based approach general but is more likely due to flaws in the implementation of the strategy.

7. References

Brestrich, N., Briskot, T., Osberghaus, A., & Hubbuch, J. (2014). A tool for selective inline quantification of co-eluting proteins in chromatography using spectral analysis and partial least squares regression. *Biotechnology and Bioengineering*, *111*(7), 1365–1373. http://doi.org/10.1002/bit.25194

Ettre, L. S. (1993). Nomenclature for chromatography (IUPAC Recommendations 1993). *Pure and Applied Chemistry*. http://doi.org/10.1351/pac199365040819

Hagel, L., Jagschies, G., & Sofer, G. (2008). *Handbook of Process Chromatography*. *Handbook of Process Chromatography*. Elsevier. http://doi.org/10.1016/B978-012374023-6.50003-8

Johansson, K., Frederiksen, S. S., Degerman, M., Breil, M. P., Mollerup, J. M., & Nilsson, B. (2015). Combined effects of potassium chloride and ethanol as mobile phase modulators on hydrophobic interaction and reversed-phase chromatography of three insulin variants. *Journal of Chromatography A*, *1381*, 64–73. http://doi.org/10.1016/j.chroma.2014.12.081

Jönsson, J. Å. (1978). Shape of the chromatographic elution curve derived from the normal distribution of sample in the column. *Journal of Chromatography A*, *150*(1), 11–16. http://doi.org/10.1016/S0021-9673(01)92091-6

Nagrath, D., Bequette, B. W., & Cramer, S. M. (2003). Evolutionary operation and control of chromatographic processes. *AIChE Journal*, *49*(1), 82–95. http://doi.org/10.1002/aic.690490109

Olsson, A., Sandberg, G., & Dahlblom, O. (2003). On Latin hypercube sampling for structural reliability analysis. *Structural Safety*, *25*(1), 47–68. http://doi.org/10.1016/S0167-4730(02)00039-5

Shirazi, G. & F. &. (2006). *Fundamentals of Preparative and Nonlinear Chromatography, 2nd Edition* (2nd ed.).

Westerberg, K. (2012). *Modeling for Quality and Safety in Biopharmaceutical Production Processes*. Lund University.